A Linear Mixed Model Formulation for Spatio-Temporal Random Processes with Computational Advances for the Product, Sum, and Product-Sum Covariance Functions

Michael Dumelle[a,*], Jay M. Ver Hoef[b], Claudio Fuentes[a], Alix Gitelman[a]

[a]Department of Statistics, Oregon State University, Corvallis, OR 97331
[b]Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA Fisheries, Seattle, WA 98115

Abstract

To properly characterize a spatio-temporal random process, it is necessary to understand the process' dependence structure. It is common to describe this dependence using a single random error having a complicated covariance. Instead of using the single random error approach, we describe spatio-temporal random processes using linear mixed models having several random errors; each random error describes a specific quality of the covariance. This linear mixed model formulation is general, intuitive, and contains many commonly used covariance functions as special cases. We focus on using the linear mixed model formulation to express three covariance functions: product (separable), sum (linear), and product-sum. We discuss benefits and drawbacks of each covariance function and propose novel algorithms using Stegle eigendecompositions, a recursive application of the Sherman-Morrison-Woodbury formula, and Helmert-Wolf blocking to efficiently invert their covariance matrices, even when every spatial location is not observed at every time point. Via a simulation study and an analysis of temperature data in Oregon, USA, we assess computational and model performance of these covariance functions when estimated using restricted maximum likelihood (likelihood-based) and Cressie's weighted least squares (semivariogram-based). We end by offering guidelines for choosing among combinations of the covariance functions and estimation methods based on properties of observed data and the desired balance between computational efficiency and model performance.

Keywords: Correlation Function, Descriptive Model, Geostatistics, Restricted Maximum Likelihood, Semivariogram, Sherman-Morrison-Woodbury

1. Introduction

Spatio-temporal models are widely used to study random processes in several scientific fields, including climatology, ecology, environmental science, geography, geology, and others (see Cressie and Wikle (2011), Wikle et al. (2019), and references therein). Cressie and Wikle (2011) categorize spatio-temporal models into two broad classes: dynamic and descriptive. Dynamic models are built from conditional probability distributions; they capture the evolution of a spatio-temporal process using a Markovian framework. Although dynamic models offer a certain degree of flexibility, when the primary concern is describing the mean and dependence structures of a spatio-temporal process, it is common to use descriptive models. More formally, descriptive models are built by

---

*Corresponding author
Email address: dumellem@oregonstate.edu (Michael Dumelle)

specifying the first few moments of a probability distribution. In this paper, we build descriptive spatio-temporal models using a linear mixed modeling approach and show how several commonly used covariance functions are special cases of this general formulation.

Consider the spatio-temporal model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{y} \equiv \{y(\mathbf{s}_i, t_j)\}$ is a spatio-temporal process, $\mathbf{X} \equiv \{\mathbf{x}(\mathbf{s}_i, t_j)\}$ is a design matrix of covariates controlling the impact of $\boldsymbol{\beta}$ on $\mathbf{y}$, $\boldsymbol{\beta}$ is a vector of fixed effects specifying the mean (coarse-scale) of $\mathbf{y}$, and $\boldsymbol{\epsilon} \equiv \{\epsilon(\mathbf{s}_i, t_j)\}$ is the random error (fine-scale) of $\mathbf{y}$. The set $\{(\mathbf{s}_i, t_j)\}$ contains spatio-temporal locations in $\mathbb{S} \times \mathbb{T}$, where $\mathbb{S} \equiv \{\mathbf{s}_i : i = 1, \ldots, S\}$ is a set of spatial locations in $\mathbb{R}^2$, and $\mathbb{T} \equiv \{t_j : j = 1, \ldots, T\}$ is a set of time points in $\mathbb{R}^1$. If $\mathbf{y}$ is observed at every combination of the S spatial locations and T time points, then $\mathbf{y}$ has ST elements and $\{(\mathbf{s}_i, t_j)\} = \mathbb{S} \times \mathbb{T}$. If $\mathbf{y}$ is not observed at every combination of the S spatial locations and T time points, then $\mathbf{y}$ has fewer than ST elements and $\{(\mathbf{s}_i, t_j)\} \subset \mathbb{S} \times \mathbb{T}$.

The dependence structure of $\mathbf{y}$ in equation (1) is determined by $\boldsymbol{\epsilon}$. When $\boldsymbol{\epsilon}$ is second-order stationary in space and in time, the covariance between any two elements of $\mathbf{y}$ does not depend on their specific spatio-temporal locations; the covariance between these elements only depends on their spatial separation, $\mathbf{h}_s$, and their temporal separation, h_t. Even when $\boldsymbol{\epsilon}$ is second-order stationary in space and in time, it is challenging to generate classes of spatio-temporal covariance functions that are strictly positive definite. A covariance function is strictly positive definite if its associated covariance matrix is positive definite, while a covariance function is positive definite if its associated covariance matrix is only positive semi-definite. Strict positive definiteness is required for the covariance matrix to be invertible, and this inverse is often necessary to estimate $\boldsymbol{\beta}$ in equation (1), the parameters composing $\boldsymbol{\epsilon}$ in equation (1), or to make unique predictions at unobserved locations (Kriging). For a thorough review of strict positive definiteness and its implications on spatio-temporal covariance functions, see De Iaco and Posa (2018) and De Iaco et al. (2019). In this paper, we focus on three covariance functions that can be strictly positive definite: product (separable) (Posa, 1993; Haas, 1995; De Cesare et al., 1997), sum (linear) (Rouhani and Hall, 1989), and product-sum (De Cesare et al., 2001; De Iaco et al., 2001).

The product (separable) covariance function is

$$\mathcal{C}(\mathbf{h}_s, h_t) = \mathcal{C}_s(\mathbf{h}_s)\mathcal{C}(h_t), \tag{2}$$

where $\mathcal{C}_s(\mathbf{h}_s)$ is a spatial covariance function, and $\mathcal{C}_t(h_t)$ is a temporal covariance function. Product covariance functions are strictly positive definite when both $\mathcal{C}_s(\mathbf{h}_s)$ and $\mathcal{C}_t(h_t)$ are strictly positive definite (De Iaco et al., 2011). The product structure of this covariance function is restrictive and often unrealistic because of a proportionality implication. Viewed as a function of space (or similarly, time), the product covariance function at separate h_t (or similarly, $\mathbf{h}_s$) values are proportional. This further implies that for the product covariance function, no amount of spatial covariance is separate from temporal covariance: For example, as $\mathcal{C}_t(h_t)$ approaches zero, so does $\mathcal{C}(\mathbf{h}_s, h_t)$, irrespective of $\mathcal{C}_s(\mathbf{h}_s)$. Despite this drawback, product covariance functions are often used in practical applications even when their use is not physically justifiable (Gneiting et al., 2006). This is primarily because when $\{(\mathbf{s}_i, t_j)\} = \mathbb{S} \times \mathbb{T}$, obtaining the inverse of the product covariance matrix is computationally efficient. Unfortunately this computational efficiency is lost when $\{(\mathbf{s}_i, t_j)\} \subset \mathbb{S} \times \mathbb{T}$.

The sum (linear) covariance function is

$$\mathcal{C}(\mathbf{h}_s, \mathbf{h}_t) = \mathcal{C}_s(\mathbf{h}_s) + \mathcal{C}_t(\mathbf{h}_t). \tag{3}$$

Though free from the proportionality implication restricting the product covariance function, the sum covariance function has a different restriction: When both $\mathcal{C}_s(\mathbf{h}_s)$ and $\mathcal{C}_t(\mathbf{h}_t)$ are strictly positive definite, the sum covariance function is not guaranteed to be strictly positive definite, it is only guaranteed to be positive definite (Myers and Journel, 1990). Though likely a logical choice to model the covariance of many spatio-temporal processes, the practicality of the sum covariance function is largely diminished by the lack of strict positive definiteness. In contrast to the product covariance matrix, the sum covariance matrix does not have a computationally efficient inverse (if it exists) when $\{(\mathbf{s}_i, \mathbf{t}_j)\} = \mathbb{S} \times \mathbb{T}$.

A straightforward extension of the product and sum covariance functions yields the product-sum covariance function. The product-sum covariance function is

$$\mathcal{C}(\mathbf{h}_s, \mathbf{h}_t) = k_1 \mathcal{C}_s(\mathbf{h}_s)\mathcal{C}_t(\mathbf{h}_t) + k_2 \mathcal{C}_s(\mathbf{h}_s) + k_3 \mathcal{C}_t(\mathbf{h}_t), \tag{4}$$

where k_1, k_2, and k_3 are nonnegative weightings among the three components. Product-sum covariance functions are strictly positive definite when k_1 is positive and both $\mathcal{C}_s(\mathbf{h}_s)$ and $\mathcal{C}_t(\mathbf{h}_t)$ are strictly positive definite. Their flexible, intuitive form is free from the proportionality restriction of the product covariance function and the positive definite restriction of the sum covariance function. Because of this, product-sum covariance functions have been used to model many spatio-temporal processes in a variety of disciplines (De Iaco et al., 2015). Like the sum covariance matrix, the product-sum covariance matrix does not have a computationally efficient inverse when $\{(\mathbf{s}_i, \mathbf{t}_j)\} = \mathbb{S} \times \mathbb{T}$.

As illustrated by equations (2), (3), and (4), spatio-temporal covariances can involve complicated functions of several parameters. Rather than modeling these parameters using a single random error, we can isolate specific qualities of the covariance structure by incorporating several random errors that connect variance components to covariance functions. These random errors can be regarded as random effects in the linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1 \mathbf{u}_1 + \ldots + \mathbf{Z}_q \mathbf{u}_q + \boldsymbol{\nu}, \tag{5}$$

where $\mathbf{X}\boldsymbol{\beta}$ is the mean structure from equation (1), $\mathbf{u}_i$ is the ith random effect, $\mathbf{Z}_i$ is the design matrix corresponding to $\mathbf{u}_i$, and $\boldsymbol{\nu}$ is random error that is independent for each observation (i.e. completely independent random error). The $\mathbf{u}$ and $\mathbf{Z}$ terms from equation (5) are related to spatial and temporal locations in Section 2.

The rest of this paper is organized as follows. In Section 2, we describe spatio-temporal processes using a linear mixed model formulation. We link this linear mixed model formulation to the product (2), sum (3), and product-sum (4) covariance functions and introduce an adjustment to the sum covariance function, which we call the sum-with-error covariance function. The sum-with-error covariance function is more flexible than the sum covariance function, is strictly positive definite, and has a computationally efficient matrix inverse. In Section 3, we discuss covariance parameter estimation using likelihood-based and semivariogram-based methods and give an overview of spatio-temporal prediction (Kriging). In Section 4, we develop a novel algorithm used to efficiently invert product, sum-with-error, and product-sum covariance matrices, even when $\{(\mathbf{s}_i, \mathbf{t}_j)\} \subset \mathbb{S} \times \mathbb{T}$. Via a simulation study (Section 5) and an analysis of temperature data (Section 6), we compare the computational and model performance among the product, sum-

with-error, and product-sum covariance functions estimated using restricted maximum likelihood (likelihood-based) and Cressie's weighted least squares (semivariogram-based). Finally, in Section 7, we conclude with a general discussion and provide directions for future research.

2. A Linear Mixed Model Formulation for Spatio-Temporal Random Processes

The linear mixed model formulation in equation (5) is a general approach that can be used to model many spatio-temporal random processes. Building from Cressie and Wikle (2011, p. 304), consider the second-order stationary (in space and in time) linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta} + \boldsymbol{\omega} + \boldsymbol{\nu}, \tag{6}$$

where $\mathbf{y}$ is an $n \times 1$ response vector, $\mathbf{X}$ is an $n \times p$ design matrix of covariates, $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector of fixed effects, $\mathbf{Z}_s$ is an $n \times S$ design matrix whose rows reference unique spatio-temporal locations and columns reference S unique spatial locations, and $\mathbf{Z}_t$ is an $n \times T$ design matrix whose rows reference unique spatio-temporal locations and columns reference T unique time points. For a general spatio-temporal location $(\mathbf{s}_i, \mathbf{t}_j)$, the corresponding row in $\mathbf{Z}_s$ equals one in the ith column and zero elsewhere, and the corresponding row in $\mathbf{Z}_t$ equals one in the jth column and zero elsewhere. For a simple example of $\mathbf{Z}_s$ and $\mathbf{Z}_t$, consider $\mathbf{y} \equiv \{y(\mathbf{s}_1, \mathbf{t}_1), y(\mathbf{s}_2, \mathbf{t}_1), y(\mathbf{s}_3, \mathbf{t}_1), y(\mathbf{s}_1, \mathbf{t}_2), y(\mathbf{s}_2, \mathbf{t}_2)\}$. Then

$$\mathbf{Z}_s = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{Z}_t = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

The random effects in equation (6), $\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\omega}$, and $\boldsymbol{\nu}$, are zero-mean vectors. The vectors $\boldsymbol{\delta}, \boldsymbol{\tau}$, and $\boldsymbol{\omega}$ are the spatial, temporal, and spatio-temporal dependent random errors, respectively. The vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are the spatial and temporal independent random errors, respectively. The vector $\boldsymbol{\nu}$ is completely independent random error at each spatio-temporal location. The completely independent random error is common to standard linear regression models, and its inclusion in spatio-temporal models adds an extra layer of flexibility. We call equation (6) the spatio-temporal linear mixed model (spatio-temporal LMM).

Each random effect in the spatio-temporal LMM has a unique covariance: $\text{Cov}(\boldsymbol{\delta}) = \sigma_\delta^2 \mathbf{R}_s$, $\text{Cov}(\boldsymbol{\gamma}) = \sigma_\gamma^2 \mathbf{I}_s$, $\text{Cov}(\boldsymbol{\tau}) = \sigma_\tau^2 \mathbf{R}_t$, $\text{Cov}(\boldsymbol{\eta}) = \sigma_\eta^2 \mathbf{I}_t$, $\text{Cov}(\boldsymbol{\omega}) = \sigma_\omega^2 \mathbf{R}_{st}$, and $\text{Cov}(\boldsymbol{\nu}) = \sigma_\nu^2 \mathbf{I}_{st}$. The matrix subscripts, s, t, and st, indicate spatial-only, temporal-only, and spatio-temporal interaction components of the covariance, respectively. These matrix dimensions, as well as the dimensions of $\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\omega}$, and $\boldsymbol{\nu}$, follow directly from equation (6). Each component of these covariance matrices involves the product of a variance parameter and an $\mathbf{R}$ matrix or an $\mathbf{I}$ matrix. The $\mathbf{R}$ matrices model random errors having correlation and depend on range parameters controlling the correlation's behavior as a function of distance. Some examples of correlation functions commonly used to model the $\mathbf{R}$ matrices include the exponential, spherical, Gaussian, Matérn (Cressie, 1993, pp. 85-86, p. 94), or auto-regressive-integrated-moving-average (ARIMA) (Shumway and Stoffer, 2017, p. 134) functions. The variance parameters multiplied by the $\mathbf{R}$ matrices are commonly referred to as dependent random error variances or partial sills. The $\mathbf{I}$ matrices are identity matrices modeling independent random errors. The variance parameters multiplied by the $\mathbf{I}$ matrices are commonly referred to as independent random error variances or nuggets. It is worth noting that
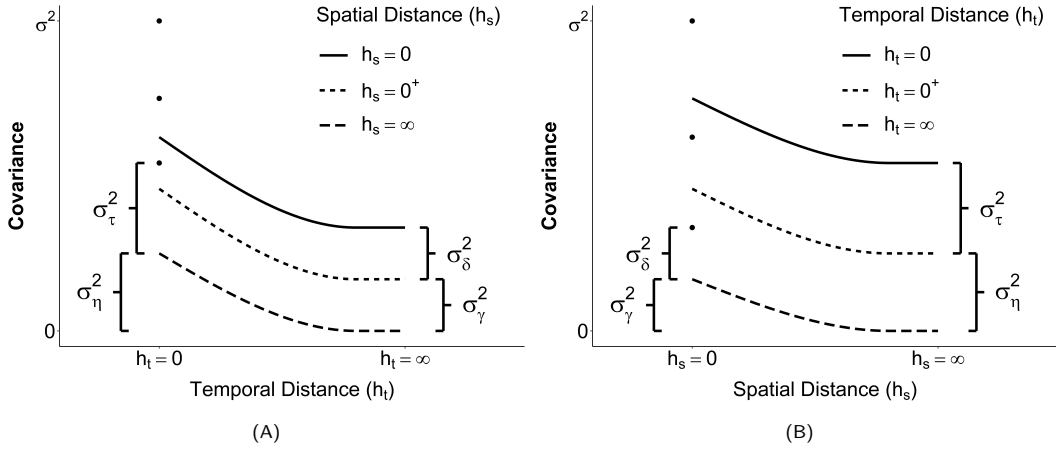
Figure 1: Covariance function behavior in the spatio-temporal LMM. In (A), the covariance function is viewed with temporal distance on the x-axis and spatial distance using line types. In (B), the covariance function is viewed with spatial distance on the x-axis and temporal distance using line types. Distances of 0^+ indicates a right limit approaching zero. The variance parameters σ_δ^2 (spatial dependent variance), σ_γ^2 (spatial independent variance), σ_τ^2 (temporal dependent variance), and σ_η^2 (temporal independent variance) are identified using brackets. The σ^2 parameter denotes the sum of all variance components (the overall variance).

equation (6) only requires specification of $\mathbf{y}$'s first two moments and does not inherently rely on any distributional assumptions.

Assuming mutual independence among the random effects in equation (6), the covariance matrix for $\mathbf{y}$, denoted by $\boldsymbol{\Sigma}$, is

$$\boldsymbol{\Sigma} = \sigma_\delta^2 \mathbf{Z}_s \mathbf{R}_s \mathbf{Z}_s' + \sigma_\gamma^2 \mathbf{Z}_s \mathbf{Z}_s' + \sigma_\tau^2 \mathbf{Z}_t \mathbf{R}_t \mathbf{Z}_t' + \sigma_\eta^2 \mathbf{Z}_t \mathbf{Z}_t' + \sigma_\omega^2 \mathbf{R}_{st} + \sigma_\nu^2 \mathbf{I}_{st}. \tag{7}$$

Several families of commonly used spatio-temporal covariance functions can be expressed as a special case of equation (7); Montero et al. (2015) and Porcu et al. (2019) provide thorough reviews of many of these families. For example, the Gneiting covariance function (Gneiting, 2002) can be obtained through specification of $\sigma_\omega^2 \mathbf{R}_{st}$ while setting all other variance parameters equal to zero. In further subsections, we show how the product, sum-with-error, and product-sum covariance functions are special cases of equation (7).

An understanding of equation (7) can be aided by through visualizations. The variance parameters can be uniquely identified through linear combinations of limiting cases of the $\mathbf{R}$ matrices, and Figure 1 shows how to clearly represent σ_δ^2, σ_γ^2, σ_τ^2, and σ_η^2. Further details and equivalent representations using semivariograms (semivariograms are discussed in Section 3.2) are provided in the supplementary material.

2.1. The Product Linear Mixed Model

Suppose the spatial dependence in $\mathbf{y}$ has covariance matrix $\mathbf{C}_s \equiv \sigma_s^2[(1-\pi_s)\mathbf{R}_s + \pi_s \mathbf{I}_s]$, where σ_s^2 is the overall spatial variance (sill), and π_s is the proportion of σ_s^2 attributable to independent random error (proportion of nugget variance). Analogously suppose the temporal dependence in $\mathbf{y}$ has covariance matrix $\mathbf{C}_t \equiv \sigma_t^2[(1-\pi_t)\mathbf{R}_t + \pi_t \mathbf{I}_t]$. The product covariance matrix of $\mathbf{y}$ is

$$\boldsymbol{\Sigma} = \mathbf{Z}_s \mathbf{C}_s \mathbf{Z}_s' \odot \mathbf{Z}_t \mathbf{C}_t \mathbf{Z}_t = \sigma_s^2 \sigma_t^2 (\mathbf{Z}_s \mathbf{R}_s^* \mathbf{Z}_s' \odot \mathbf{Z}_t \mathbf{R}_t^* \mathbf{Z}_t), \tag{8}$$

5

where $\odot$ denotes the Hadmard (element-wise) product, $\mathbf{R}_s^* = \mathbf{C}_s/\sigma_s^2$, and $\mathbf{R}_t^* = \mathbf{C}_t/\sigma_t^2$. The variance parameters in equation (8) are not identifiable individually, but their product is identifiable. Reparameterizing $\sigma_\omega^2 \equiv \sigma_s^2 \sigma_t^2$ to ensure identifiability of the variance parameter yields the following product covariance matrix:

$$\mathbf{\Sigma} = \sigma_\omega^2(\mathbf{Z}_s\mathbf{R}_s^*\mathbf{Z}_s' \odot \mathbf{Z}_t\mathbf{R}_t^*\mathbf{Z}_t) = \sigma_\omega^2\mathbf{R}_{st}, \tag{9}$$

where $\mathbf{R}_{st} = (\mathbf{Z}_s\mathbf{R}_s^*\mathbf{Z}_s' \odot \mathbf{Z}_t\mathbf{R}_t^*\mathbf{Z}_t)$. Written this way, it becomes clear that the covariance matrix in equation (9) is the covariance matrix of a special case of the spatio-temporal LMM. More specifically, this special case is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\omega}, \tag{10}$$

where $\mathrm{Cov}(\boldsymbol{\omega}) = \sigma_\omega^2\mathbf{R}_{st}$ and $\mathbf{R}_{st} = (\mathbf{Z}_s\mathbf{R}_s^*\mathbf{Z}_s' \odot \mathbf{Z}_t\mathbf{R}_t^*\mathbf{Z}_t)$. We define the model in equation (10) with the covariance matrix in equation (9) as the product linear mixed model (product LMM). Expanding equation (9) further details its product structure:

$$\begin{aligned}
\mathbf{\Sigma} = {} & \sigma_\omega^2(1 - \pi_s)(1 - \pi_t)(\mathbf{Z}_s\mathbf{R}_s\mathbf{Z}_s' \odot \mathbf{Z}_t\mathbf{R}_t\mathbf{Z}_t') \\
& + \sigma_\omega^2(1 - \pi_s)(\pi_t)(\mathbf{Z}_s\mathbf{R}_s\mathbf{Z}_s' \odot \mathbf{Z}_t\mathbf{I}_t\mathbf{Z}_t') \\
& + \sigma_\omega^2(\pi_s)(1 - \pi_t)(\mathbf{Z}_s\mathbf{I}_s\mathbf{Z}_s' \odot \mathbf{Z}_t\mathbf{R}_t\mathbf{Z}_t') \\
& + \sigma_\omega^2(\pi_s)(\pi_t)(\mathbf{Z}_s\mathbf{I}_s\mathbf{Z}_s' \odot \mathbf{Z}_t\mathbf{I}_t\mathbf{Z}_t'),
\end{aligned} \tag{11}$$

Equation (11) highlights the explicit dependence among the variance parameters, making clear the restrictions of this product structure. For example, when π_s tends towards zero, then $(1 - \pi_t)\pi_s\mathbf{R}_t \otimes \mathbf{I}_s$, a function of the temporal correlation, also tends towards zero.

2.2. The Sum-With-Error Linear Mixed Model

Another special case of the spatio-temporal LMM is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta}. \tag{12}$$

Equation (12) has covariance matrix

$$\mathbf{\Sigma} = \sigma_\delta^2\mathbf{Z}_s\mathbf{R}_s\mathbf{Z}_s' + \sigma_\gamma^2\mathbf{Z}_s\mathbf{Z}_s' + \sigma_\tau^2\mathbf{Z}_t\mathbf{R}_t\mathbf{Z}_t' + \sigma_\eta^2\mathbf{Z}_t\mathbf{Z}_t',$$

which is the matrix representation of the sum covariance function. As mentioned in Section 1, a significant drawback of the sum covariance function is that it is not guaranteed to be strictly positive definite. We can instead consider a model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta} + \boldsymbol{\nu}, \tag{13}$$

which has covariance matrix

$$\mathbf{\Sigma} = \sigma_\delta^2\mathbf{Z}_s\mathbf{R}_s\mathbf{Z}_s' + \sigma_\gamma^2\mathbf{Z}_s\mathbf{Z}_s' + \sigma_\tau^2\mathbf{Z}_t\mathbf{R}_t\mathbf{Z}_t' + \sigma_\eta^2\mathbf{Z}_t\mathbf{Z}_t' + \sigma_\nu^2\mathbf{I}_{st}. \tag{14}$$

We call the model in equation (13) with the covariance matrix in equation (14) the sum-with-error linear mixed model (sum-with-error LMM). We call the covariance function whose matrix representation is equation (14) the sum-with-error covariance function. The sum-with-error covariance function is certainly more flexible than the sum covariance function, as the sum co-

variance function is a special case of the sum-with-error covariance function. Furthermore, the sum-with-error covariance function is strictly positive definite. A proof of this statement is included in the supplementary material and relies on the positive semi-definiteness of the matrix $\sigma_\delta^2 \mathbf{Z}_s \mathbf{R}_s \mathbf{Z}_s' + \sigma_\gamma^2 \mathbf{Z}_s \mathbf{Z}_s' + \sigma_\tau^2 \mathbf{Z}_t \mathbf{R}_t \mathbf{Z}_t' + \sigma_\eta^2 \mathbf{Z}_t \mathbf{Z}_t'$ and the positive definiteness of the matrix $\sigma_\nu^2 \mathbf{I}_{st}$.

2.3. The Product-Sum Linear Mixed Model

Consider the full spatio-temporal LMM (6) with covariance matrix from equation (7). If $\mathbf{R}_{st}$ in equation (7) equals $(\mathbf{Z}_s \mathbf{R}_s \mathbf{Z}_s') \odot (\mathbf{Z}_t \mathbf{R}_t \mathbf{Z}_t')$, the covariance matrix is

$$\mathbf{\Sigma} = \sigma_\delta^2 \mathbf{Z}_s \mathbf{R}_s \mathbf{Z}_s' + \sigma_\gamma^2 \mathbf{Z}_s \mathbf{Z}_s' + \sigma_\tau^2 \mathbf{Z}_t \mathbf{R}_t \mathbf{Z}_t' + \sigma_\eta^2 \mathbf{Z}_t \mathbf{Z}_t + \sigma_\omega^2 (\mathbf{Z}_s \mathbf{R}_s \mathbf{Z}_s') \odot (\mathbf{Z}_t \mathbf{R}_t \mathbf{Z}_t') + \sigma_\nu^2 \mathbf{I}_{st}. \tag{15}$$

The model in equation (6) with the covariance matrix in equation (15) is the product-sum linear mixed model (product-sum LMM). The product-sum LMM combines components of both the product LMM and the sum-with-error LMM. It is not bound by the proportionality implication of the product LMM, and unlike the sum-with-error LMM, the product-sum LMM contains a product term. The product-sum LMM's covariance matrix in equation (15) does not always equal the matrix representation of the original product-sum covariance function in equation (4). In equation (15), the product term does not contain independent error components. But unlike the original product-sum covariance function, equation (15) enables complete variance component separation. This complete variance component separation makes it straightforward to isolate variance parameters and partition sources of error.

The product-sum LMM's covariance matrix could even be made more flexible by letting $\mathbf{R}_{st}$ equal $(\mathbf{Z}_s \tilde{\mathbf{R}}_s \mathbf{Z}_s') \odot (\mathbf{Z}_t \tilde{\mathbf{R}}_t \mathbf{Z}_t')$, where $\tilde{\mathbf{R}}_s$ and $\tilde{\mathbf{R}}_s$ are different from $\mathbf{R}_s$ and $\mathbf{R}_t$, respectively. Though this yields a more general family of covariance functions, we do not focus on this case in this paper for a few reasons. First, the original product-sum covariance function specifies the same spatial and temporal covariance functions for the sum and product terms. Second, interpreting and visualizing the behavior of relevant spatial and temporal correlation functions, as in Figure 1, is less intuitive when accommodating $\tilde{\mathbf{R}}_s$ and $\tilde{\mathbf{R}}_t$. Third, the additional parameters in $\tilde{\mathbf{R}}_s$ and $\tilde{\mathbf{R}}_t$ make estimation more challenging by increasing both the dimensionality of the optimization and the difficulty associated with identifying unique covariance function components. We discuss estimation in more detail next.

3. Parameter Estimation and Prediction

The covariance parameters and fixed effects of the models in Section 2 generally require estimation. In this section, we review parameter estimation using restricted maximum likelihood (REML) (Patterson and Thompson, 1971; Harville, 1977) and Cressie's weighted least squares (CWLS) (Cressie, 1985), two commonly used parameter estimation methods. We then review best linear unbiased prediction (Kriging). We end the section by discussing alternative estimation and prediction approaches.

3.1. Restricted Maximum Likelihood Estimation

Minus twice a profiled Gaussian log-likelihood is

$$-2l(\boldsymbol{\theta}|\mathbf{y}) = \ln |\mathbf{\Sigma}| + (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})\mathbf{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})' + c_1, \tag{16}$$

where $\mathbf{y}$ is a vector of response variables, $\boldsymbol{\theta}$ is a vector of covariance parameters, $\mathbf{\Sigma}$ is a covariance matrix, $|.|$ denotes the determinant operator, $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y}$, and c_1 is an additive

constant. Minimizing equation (16) yields $\hat{\boldsymbol{\theta}}_{ml}$, the maximum likelihood (ML) estimator of $\boldsymbol{\theta}$. In general, no closed-form solution exists for $\hat{\boldsymbol{\theta}}_{ml}$, and it must be solved for numerically. After obtaining $\hat{\boldsymbol{\theta}}_{ml}$, a closed form solution for the ML estimator of $\boldsymbol{\beta}$ exists: $\hat{\boldsymbol{\beta}}_{ml} \equiv (\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{ml}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{ml}^{-1}\mathbf{y}$, where $\hat{\boldsymbol{\Sigma}}_{ml}$ is $\boldsymbol{\Sigma}$ evaluated at $\hat{\boldsymbol{\theta}}_{ml}$ instead of $\boldsymbol{\theta}$. Unfortunately, $\hat{\boldsymbol{\theta}}_{ml}$ can be badly biased for $\boldsymbol{\theta}$ because equation (16) does not account for the simultaneous estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. To address this bias problem, Patterson and Thompson (1971) propose transforming equation (16) using random error contrasts, which results in a new likelihood: The restricted (or residual) Gaussian likelihood. Wolfinger et al. (1994) shows that minus twice the profiled restricted Gaussian log-likelihood is

$$-2l_R(\boldsymbol{\theta}|\mathbf{y}) = -2l(\boldsymbol{\theta}|\mathbf{y}) + \ln|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}| + c_2, \tag{17}$$

where c_2 is an additive constant. Minimizing equation (17) yields $\hat{\boldsymbol{\theta}}_{reml}$, the restricted maximum likelihood estimator (REML) of $\boldsymbol{\theta}$. Similar to ML estimation, $\hat{\boldsymbol{\theta}}_{reml}$ must generally be solved for numerically, but then a closed-form solution for the REML estimator of $\boldsymbol{\beta}$ exists: $\hat{\boldsymbol{\beta}}_{reml} \equiv (\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{reml}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{reml}^{-1}\mathbf{y}$. Equations (16) and (17) can also be further profiled by their overall variance, reducing the number of parameters in $\boldsymbol{\theta}$ requiring estimation by one (Wolfinger et al., 1994). Because of the bias problem in ML estimation, we focus only on REML estimation henceforth.

When $\mathbf{y}$ is Gaussian, $\hat{\boldsymbol{\theta}}_{reml}$ has several attractive statistical properties: It is computed from unbiased estimation equations (Heyde, 1994; Cressie and Lahiri, 1993, 1996); under appropriate regularity conditions, it is consistent, asymptotically efficient, and asymptotically Gaussian (Sweeting, 1980; Mardia and Marshall, 1984; Cressie and Lahiri, 1993); and its standard errors can be estimated using the expected or observed Hessian (Cressie and Lahiri, 1993). For REML estimation, model selection can be performed using likelihood-based statistics such as AIC (Akaike, 1974) or a likelihood-ratio test for nested models.

When $\mathbf{y}$ is not Gaussian, $\hat{\boldsymbol{\theta}}_{reml}$ is still computed from unbiased estimating equations (Heyde, 1994; Cressie and Lahiri, 1993, 1996). This is a crucial result because it implies $\mathbf{y}$ does not have to be Gaussian for $\hat{\boldsymbol{\theta}}_{reml}$ to be unbiased. To illustrate a familiar example, suppose $\mathbf{y}$ has mean $\mathbf{X}\boldsymbol{\beta}$ and common independent variance $\sigma^2\mathbf{I}$. The standard ordinary least squares estimate of σ^2 is $s^2 \equiv [(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})]/(n-p)$, where n is the sample size, p is the dimension of $\boldsymbol{\beta}$, and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. It is well known s^2 is unbiased for σ^2 even when the distribution of $\mathbf{y}$ is unknown. Furthermore, it is straightforward to show that the REML estimate of σ^2 is s^2. More generally, irrespective of $\mathbf{y}$'s distribution and under appropriate regularity conditions, $\hat{\boldsymbol{\beta}}_{reml}$ is unbiased, consistent, asymptotically Gaussian, and asymptotically efficient (Theil, 1971; Fuller and Battese, 1973; Schmidt, 1976; Judge et al., 1985; Schabenberger and Gotway, 2017). Together, REML's unbiased estimating equations for $\boldsymbol{\theta}$ and attractive asymptotic results for $\boldsymbol{\beta}$ highlight its usefulness when $\mathbf{y}$ is not Gaussian.

Though these aforementioned properties of $\hat{\boldsymbol{\theta}}_{reml}$ are certainly attractive, REML suffers from computational limitations as the sample size grows. This is because numerical minimization of equation (17) requires repeated inversion of $\boldsymbol{\Sigma}$, and the computational cost of inversion is cubic in the sample size.

3.2. Semivariogram-Based Estimation Using Cressie's Weighted Least Squares

Instead of covariance functions, spatio-temporal dependence can be described using semivariograms. The spatio-temporal semivariogram quantifies the variability in the differences among elements in $\mathbf{y}$ as a function of the spatial and temporal distances between these elements. Cressie and Wikle (2011) provide a thorough description and review of spatio-temporal semivariograms

8

and discuss the one-to-one correspondence between spatio-temporal covariance functions and semi-variograms when the random processes is second-order stationary in space and in time (we also provide this one-to-one correspondence in the supplementary material).

Starting with the spatio-temporal LMM in equation (6), define $\boldsymbol{\epsilon} \equiv \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. The spatio-temporal semivariogram for $\boldsymbol{\epsilon}$, denoted by $\gamma_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$, depends on the same parameter vector $\boldsymbol{\theta}$ that the covariance function does. Before estimating $\boldsymbol{\theta}$, however, we must first estimate $\gamma_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$. This is often accomplished by using an estimator that moment-matches $\gamma_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$ at a set of fixed spatial and temporal distance classes (Cressie and Wikle, 2011). This estimate of $\gamma_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$, denoted by $\hat{\gamma}_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$, is commonly referred to as the empirical semivariogram for $\boldsymbol{\epsilon}$. After calculating the empirical semivariogram, $\boldsymbol{\theta}$ can be estimated using a least squares approach. Least squares approaches estimate $\boldsymbol{\theta}$ by minimizing a sum of squares involving $\hat{\gamma}_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$ and $\gamma_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$. The least squares approach we focus on is Cressie's weighted least squares (CWLS), where numerical minimization of

$$\sum_i w_i [\hat{\gamma}_\epsilon(\mathbf{h}_s, \mathrm{h}_t)_i - \gamma_\epsilon(\mathbf{h}_s, \mathrm{h}_t)_i]^2 \tag{18}$$

yields $\hat{\boldsymbol{\theta}}_{cwls}$. In equation (18), i indexes the spatio-temporal distance classes used to compute $\hat{\gamma}_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$, $|N(\mathbf{h}_s, \mathrm{h}_t)|$ denotes the number of observations in the distance class, and $w_i = |N(\mathbf{h}_s, \mathrm{h}_t)_i| \gamma_\epsilon(\mathbf{h}_s, \mathrm{h}_t)_i^{-2}$. We focus on CWLS because it commonly used and computationally efficient. Reviews of other semivariogram-based estimation approaches are outlined by Cressie (1993), Lahiri et al. (2002), and Schabenberger and Gotway (2017).

In practice, we do not observe a realization of $\boldsymbol{\epsilon}$, and it must be estimated. One estimate, denoted by $\hat{\boldsymbol{\epsilon}}$, is the vector of ordinary least squares residuals. Unfortunately $\hat{\gamma}_{\hat{\epsilon}}(\mathbf{h}_s, \mathrm{h}_t)$ is an estimate of $\gamma_{\hat{\epsilon}}(\mathbf{h}_s, \mathrm{h}_t)$, not an estimate of $\gamma_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$. An implication is that $\hat{\gamma}_{\hat{\epsilon}}(\mathbf{h}_s, \mathrm{h}_t)$ is biased for $\gamma_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$, though this bias decreases as the sample size increases (Cressie, 1993, pp. 166-168).

After computing $\hat{\gamma}_{\hat{\epsilon}}(\mathbf{h}_s, \mathrm{h}_t)$ and using CWLS (18) to estimate $\hat{\boldsymbol{\theta}}_{cwls}$, empirical (feasible) generalized least squares (EGLS) can be used with $\hat{\boldsymbol{\theta}}_{cwls}$ to estimate $\boldsymbol{\beta}$: $\hat{\boldsymbol{\beta}}_{cwls} \equiv (\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{cwls}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{cwls}^{-1}\mathbf{y}$. This estimate has the same form as $\hat{\boldsymbol{\beta}}_{reml}$ – they only differ in the $\hat{\boldsymbol{\theta}}$ used to compute $\hat{\boldsymbol{\Sigma}}^{-1}$. After computing $\hat{\boldsymbol{\beta}}_{cwls}$, we can compute EGLS residuals and use them to recompute $\hat{\gamma}_{\hat{\epsilon}}(\mathbf{h}_s, \mathrm{h}_t)$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\beta}}_{cwls}$. This iterative re-estimation process can continue until some convergence criterion is satisfied, though Kitanidis (1993) and Ver Hoef and Cressie (2001) observed additional iterations generally had little impact on their model performance.

Under appropriate regularity conditions, $\hat{\boldsymbol{\theta}}_{cwls}$ is consistent and asymptotically Gaussian (Lahiri et al., 2002). Like $\hat{\boldsymbol{\beta}}_{reml}$, $\hat{\boldsymbol{\beta}}_{cwls}$ is unbiased and under appropriate regularity conditions is consistent, asymptotically Gaussian, and asymptotically efficient. The main computational burden of CWLS is calculating the empirical semivariogram, not minimizing equation (18) – equation (18) tends to be very computationally efficient. Once the empirical semivariogram has been calculated, comparing the fit of different dependence structures using CWLS estimation only requires separate minimizations of (18). To compare two dependence structures using REML, however, the REML likelihood must be maximized twice, which can be time consuming for large sample sizes due to the matrix inversions required.

Unfortunately there are some drawbacks to CWLS estimation. Most notably, CWLS estimation requires the specification of arbitrary spatial and temporal distance classes used to compute the empirical semivariogram, and different choices of distances classes impact parameter estimates and model performance. If calculating $\hat{\gamma}_\epsilon(\mathbf{h}_s, \mathrm{h}_t)$ requires averaging within distance classes, CWLS estimation also results in some loss of detail regarding the underlying process. Additionally, $\hat{\boldsymbol{\theta}}_{cwls}$

is only asymptotically efficient in special cases, and obtaining its standard errors is less straight-forward than for REML estimation (Lahiri et al., 2002).

3.3. Prediction

Spatio-temporal prediction is often the primary goal of a data analysis. It is usually desired that these predictions are in some sense optimal. In this context, we consider a predictor optimal if it is unbiased, linear, and has minimum variance within the class of all unbiased, linear predictors; this predictor is often referred to as a best linear unbiased predictor (BLUP) or a Kriging predictor. Consider the structure of $\mathbf{y}$ in the spatio-temporal LMM (6), and let $\dot{\mathbf{y}}$ be a vector of unobserved variables coming from the same distribution. When $\boldsymbol{\beta}$ is unknown and $\boldsymbol{\theta}$ is known, the (BLUP) of $\dot{\mathbf{y}}$ is given by

$$\dot{\mathbf{y}}_{\text{BLUP}} = \dot{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \dot{\mathbf{c}}\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}), \tag{19}$$

where $\dot{\mathbf{X}}$ is the design matrix of covariates at the prediction locations, $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$, and $\dot{\mathbf{c}} = \text{Cov}(\dot{\mathbf{y}}, \mathbf{y})$ (Cressie, 1993, p. 173). The covariance of $\dot{\mathbf{y}}_{\text{BLUP}}$ is given by

$$\text{Cov}(\dot{\mathbf{y}}_{\text{BLUP}}) = \dot{\boldsymbol{\Sigma}} - \dot{\mathbf{c}}\boldsymbol{\Sigma}^{-1}\dot{\mathbf{c}}' + \mathbf{H}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{H}', \tag{20}$$

where $\dot{\boldsymbol{\Sigma}}$ is the covariance matrix of $\dot{\mathbf{y}}$, and $\mathbf{H} = (\dot{\mathbf{X}} - \dot{\mathbf{c}}\boldsymbol{\Sigma}^{-1}\mathbf{X})$ (Cressie, 1993, p. 173). If $\boldsymbol{\theta}$ is unknown, the quantities in equations (19) and (20) involving $\boldsymbol{\theta}$ are instead evaluated at $\hat{\boldsymbol{\theta}}$. In this context, the predictor is instead referred to as an empirical best linear unbiased predictor (EBLUP) and is denoted by $\dot{\mathbf{y}}_{\text{EBLUP}}$.

3.4. Alternative Approaches to Estimation and Prediction

Though we focus on REML estimation, CWLS estimation, and best linear unbiased prediction, it is important to acknowledge that these estimation and prediction approaches require the inverse of an $n \times n$ matrix, where n is the sample size: REML requires several $n \times n$ inverses to iteratively estimate $\boldsymbol{\theta}$ and one $n \times n$ inverse to estimate $\boldsymbol{\beta}$, CWLS requires one $n \times n$ inverse to estimate $\boldsymbol{\beta}$, and best linear unbiased prediction requires one $n \times n$ inverse. Directly inverting an $n \times n$ matrix has computational cost $\mathcal{O}(n^3)$, which makes covariance matrix inversion more challenging as the sample size increases. This challenge has generated a wide array of research focused on alternative estimation and prediction approaches that avoid $n \times n$ matrix inversions, either by approximating the inverse or by using an approach that does not depend on the inverse. Heaton et al. (2019) provides a thorough overview of several of these alternatives and compares them on real and simulated data. Our novel, computationally efficient approach is described in Section 4 and leverages the structure of the product, sum-with-error, and product-sum covariance functions to compute exact covariance matrix inverses without having to invert any $n \times n$ matrices.

4. Efficient Covariance Matrix Inversion

As mentioned in Section 3.4, we propose a novel computational approach that leverages the structure of the product, sum-with-error, and product-sum covariance matrices to compute exact inverses without having to invert any $n \times n$ matrices. Our approach is purely algorithmic, which has two attractive implications. First, our approach solves the exact inverse of these co-variance matrices without requiring approximations. Second, our approach can even be combined with alternative approaches to further reduce computational costs associated with estimation and

prediction. For example, our efficient inversion algorithms could be combined with the low-rank representations used by Fixed-Rank Kriging (Cressie and Johannesson, 2008; Zammit-Mangion and Cressie, 2017). Next, we describe these algorithms and discuss how they change whether $\{(\mathbf{s}_i, \mathbf{t}_j)\} = \mathbb{S} \times \mathbb{T}$ or $\{(\mathbf{s}_i, \mathbf{t}_j)\} \subset \mathbb{S} \times \mathbb{T}$. We then provide some computational benchmarks.

4.1. Inverse Computations When $\{(\mathbf{s}_i, \mathbf{t}_j)\} = \mathbb{S} \times \mathbb{T}$

We previously mentioned that when $\{(\mathbf{s}_i, \mathbf{t}_j)\} = \mathbb{S} \times \mathbb{T}$, product covariance matrices have a Kronecker structure facilitating efficient inverse computation. Suppose the data are ordered by space within time: $\mathbf{y} \equiv \{y(\mathbf{s}_1, \mathbf{t}_1), y(\mathbf{s}_2, \mathbf{t}_1), \ldots, y(\mathbf{s}_1, \mathbf{t}_2), y(\mathbf{s}_2, \mathbf{t}_2), \ldots, y(\mathbf{s}_S, \mathbf{t}_T)\}$. In this context, the inverse of the product LMM's covariance matrix (9) can be expressed as

$$\mathbf{\Sigma}^{-1} = [(\mathbf{R}_s^*)^{-1} \otimes (\mathbf{R}_t^*)^{-1}]/\sigma_\omega^2. \tag{21}$$

Equation (21) is computationally efficient because it only requires one set of $\mathcal{O}(S^3)$ floating point operations (flops) and one set of $\mathcal{O}(T^3)$ flops for inverses; this is substantially less demanding than the $\mathcal{O}(S^3 T^3)$ flops required to compute the inverse using a algorithm that does not take advantage of the Kronecker structure, such as the Cholesky decomposition.

The sum-with-error LMM's covariance matrix (14) has an efficient form that can be exploited by an iterative application of the Sherman-Morrison-Woodbury formula (Sherman, 1949; Sherman and Morrison, 1950; Woodbury, 1950). It is first helpful to rewrite equation (14) using condensed notation:

$$\mathbf{\Sigma} = \mathbf{Z}_s \mathbf{\Sigma}_s \mathbf{Z}_s' + \mathbf{Z}_t \mathbf{\Sigma}_t \mathbf{Z}_t' + \sigma_\omega^2 \mathbf{I}_{st},$$

where $\mathbf{\Sigma}_s = \sigma_\delta^2 \mathbf{R}_s + \sigma_\gamma^2 \mathbf{I}_s$ and $\mathbf{\Sigma}_t = \sigma_\tau^2 \mathbf{R}_t + \sigma_\eta^2 \mathbf{I}_t$. The first step in computing $\mathbf{\Sigma}^{-1}$ requires computing $(\mathbf{Z}_t \mathbf{\Sigma}_t \mathbf{Z}_t' + \sigma_\nu^2 \mathbf{I}_{st})^{-1}$ using the Sherman-Morrison-Woodbury formula. This inverse, denoted by $\mathrm{SMW}(\sigma_\nu^2 \mathbf{I}_{st}^{-1}, \mathbf{\Sigma}_t, \mathbf{Z}_t)$, can be expressed as

$$\mathrm{SMW}(\sigma_\nu^2 \mathbf{I}_{st}^{-1}, \mathbf{\Sigma}_t, \mathbf{Z}_t) = (\sigma_\nu^2 \mathbf{I}_{st})^{-1} - (\sigma_\nu^2 \mathbf{I}_{st})^{-1} \mathbf{Z}_t (\mathbf{\Sigma}_t^{-1} + \mathbf{Z}_t'(\sigma_\nu^2 \mathbf{I}_{st})^{-1} \mathbf{Z}_t)^{-1} \mathbf{Z}_t'(\sigma_\nu^2 \mathbf{I}_{st})^{-1}.$$

Another application of the Sherman-Morrison-Woodbury formula can be used to compute $(\mathbf{Z}_s \mathbf{\Sigma}_s \mathbf{Z}_s' + \mathbf{Z}_t \mathbf{\Sigma}_t \mathbf{Z}_t' + \sigma_\omega^2 \mathbf{I}_{st})^{-1}$, which equals $\mathbf{\Sigma}^{-1}$. The entire algorithm is expressed compactly as

$$\mathbf{\Sigma}^{-1} = \mathrm{SMW}(\mathrm{SMW}(\sigma_\nu^2 \mathbf{I}_{st}^{-1}, \mathbf{\Sigma}_t, \mathbf{Z}_t), \mathbf{\Sigma}_s, \mathbf{Z}_s). \tag{22}$$

Equation (22) is computationally efficient because the inner $\mathrm{SMW}(\cdot)$ requires two sets of $\mathcal{O}(T^3)$ flops for inverses, and the outer $\mathrm{SMW}(\cdot)$ requires two sets of $\mathcal{O}(S^3)$ flops for inverses. Furthermore, the sparsity of $\mathbf{Z}_s$ and $\mathbf{Z}_t$ can be used to avoid direct multiplications in products involving these matrices. Because the sum covariance (3) matrix lacks $\sigma_\nu^2 \mathbf{I}_{st}$, its inversion (if it exists) requires $\mathcal{O}(S^3 T^3)$ flops because the Sherman-Morrison-Woodbury formula cannot be recursively applied.

The covariance of the product-sum LMM (15) has an efficient form that can be exploited by Stegle eigendecompositions (Stegle et al., 2011) and an iterative application of the Sherman-Morrison-Woodbury formula. Equation (15) can also be rewritten using condensed notation:

$$\mathbf{\Sigma} = \mathbf{Z}_s \mathbf{\Sigma}_s \mathbf{Z}_s' + \mathbf{Z}_t \mathbf{\Sigma}_t \mathbf{Z}_t' + \mathbf{\Sigma}_{st},$$

where $\mathbf{\Sigma}_s = \sigma_\delta^2 \mathbf{R}_s + \sigma_\gamma^2 \mathbf{I}_s$, $\mathbf{\Sigma}_t = \sigma_\tau^2 \mathbf{R}_t + \sigma_\eta^2 \mathbf{I}_t$, and $\mathbf{\Sigma}_{st} = \sigma_\omega^2 \mathbf{R}_t \otimes \mathbf{R}_s + \sigma_\nu^2 \mathbf{I}_{st}$. The first step in computing $\mathbf{\Sigma}^{-1}$ is using the Stegle eigendecomposition of $\mathbf{\Sigma}_{st}$. Let $\mathbf{U}_s \mathbf{P}_s \mathbf{U}_s'$ be the eigendecomposition of $\mathbf{R}_s$

11

and $\mathbf{U}_t\mathbf{P}_t\mathbf{U}_t'$ be the eigendecomposition of $\mathbf{R}_t$. Then $\boldsymbol{\Sigma}_{st}^{-1}$, denoted by $\mathrm{STE}(\boldsymbol{\Sigma}_{st})$, can be expressed as

$$\mathrm{STE}(\boldsymbol{\Sigma}_{st}) = (\mathbf{W}\mathbf{V}^{-1/2})(\mathbf{W}\mathbf{V}^{-1/2})', \tag{23}$$

where $\mathbf{W} = \mathbf{U}_t \otimes \mathbf{U}_s$ and $\mathbf{V} = \sigma_\omega^2 \mathbf{P}_t \otimes \mathbf{P}_s + \sigma_\nu^2 \mathbf{I}_t \otimes \mathbf{I}_s$. Because $\mathbf{P}_s$ and $\mathbf{P}_t$ are diagonal matrices of eigenvalues, $\mathbf{V}$ is diagonal, and computing $\mathbf{V}^{-1/2}$ is trivial. In the second part of the algorithm, the Sherman-Morrison-Woodbury formula is used to compute $(\mathbf{Z}_t\boldsymbol{\Sigma}_t\mathbf{Z}_t' + \boldsymbol{\Sigma}_{st})^{-1}$:

$$\mathrm{SMW}(\boldsymbol{\Sigma}_{st}^{-1}, \boldsymbol{\Sigma}_t, \mathbf{Z}_t) = \boldsymbol{\Sigma}_{st}^{-1} - \boldsymbol{\Sigma}_{st}^{-1}\mathbf{Z}_t(\boldsymbol{\Sigma}_t^{-1} + \mathbf{Z}_t'\boldsymbol{\Sigma}_{st}^{-1}\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\boldsymbol{\Sigma}_{st}^{-1}.$$

The third part of the algorithm uses another application of Sherman-Morrison-Woodbury formula to compute $(\mathbf{Z}_s\boldsymbol{\Sigma}_s\mathbf{Z}_s' + \mathbf{Z}_t\boldsymbol{\Sigma}_t\mathbf{Z}_t' + \boldsymbol{\Sigma}_{st})^{-1}$, which equals $\boldsymbol{\Sigma}^{-1}$. The entire inversion algorithm is compactly expressed as

$$\boldsymbol{\Sigma}^{-1} = \mathrm{SMW}(\mathrm{SMW}(\mathrm{STE}(\boldsymbol{\Sigma}_{st}), \boldsymbol{\Sigma}_t, \mathbf{Z}_t), \boldsymbol{\Sigma}_s, \mathbf{Z}_s). \tag{24}$$

Equation (24) is computationally efficient because $\mathrm{STE}(\cdot)$ requires one set of $\mathcal{O}(S^3)$ flops and one set of $\mathcal{O}(T^3)$ flops for eigendecompositions, the inner $\mathrm{SMW}(\cdot)$ requires two sets of $\mathcal{O}(T^3)$ flops for inverses, and the outer $\mathrm{SMW}(\cdot)$ requires two sets of $\mathcal{O}(S^3)$ flops for inverses, far fewer than $\mathcal{O}(S^3T^3)$ flops. The computational efficiency of equation (24) can be further improved by incorporating two additional tools. First, as with the sum-with-error LMM, the sparsity of $\mathbf{Z}_s$ and $\mathbf{Z}_t$ can be used to avoid direct multiplications in products involving these matrices. Second, $\boldsymbol{\Sigma}^{-1}$ can be multiplied on the right by $\mathbf{X}$, $\mathbf{y}$, and $\dot{\mathbf{c}} \equiv \mathrm{Cov}(\dot{\mathbf{y}}, \mathbf{y})$, where $\dot{\mathbf{y}}$ is a vector of new $\mathbf{y}$'s to predict. This operation makes it possible to avoid direct multiplication of the two $ST \times ST$ matrices in equation (23) while obtaining the products $\boldsymbol{\Sigma}^{-1}\mathbf{X}$, $\boldsymbol{\Sigma}^{-1}\mathbf{y}$, and $\boldsymbol{\Sigma}^{-1}\dot{\mathbf{c}}$. These products can be used for estimation and prediction without actually requiring $\boldsymbol{\Sigma}^{-1}$. We provide analogous forms of equations (21), (22), and (24) for log determinants in the supplementary material.

4.2. Inverse Computations When $\{(\mathbf{s}_i, \mathbf{t}_j)\} \subset \mathbb{S} \times \mathbb{T}$

It is common in practice to be missing at least one element of $\{(\mathbf{s}_i, \mathbf{t}_j)\}$ from $\mathbb{S} \times \mathbb{T}$. When $\{(\mathbf{s}_i, \mathbf{t}_j)\} \subset \mathbb{S} \times \mathbb{T}$, equation (22) is unaffected because it does not rely on Kronecker products. Equations (21) and (24), however, cannot be used directly because the efficient form for the inverse (or eigendecomposition) of a Kronecker product is lost. Next we show how to incorporate Helmert-Wolf blocking (Wolf, 1978) when $\{(\mathbf{s}_i, \mathbf{t}_j)\} \subset \mathbb{S} \times \mathbb{T}$ to retain most of the computational efficiency gained by equations (21) and (24).

Before illustrating the usefulness of Helmert-Wolf blocking, suppose $\mathbf{y}$ is now more general and contains elements we observe and elements we do not observe. In this context, $\mathbf{y}$ can always be partitioned by two components. The first component is $\mathbf{y}_o$, an $n_o \times 1$ vector of observable elements whose spatio-temporal locations form a subset of $\mathbb{S} \times \mathbb{T}$. The second component is $\mathbf{y}_u$, an $n_u \times 1$ vector of unobservable elements whose spatio-temporal locations are those in $\mathbb{S} \times \mathbb{T}$ missing from the spatio-temporal locations in $\mathbf{y}_o$. Because $\mathbf{y}_o$ and $\mathbf{y}_u$ are distinct, and the union of their spatio-temporal locations equals $\mathbb{S} \times \mathbb{T}$, it follows that $n_o + n_u = ST$. The covariance matrix of $\mathbf{y}$ can be written in block form:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{ou} \\ \boldsymbol{\Sigma}_{uo} & \boldsymbol{\Sigma}_{uu} \end{bmatrix}, \tag{25}$$

where $\boldsymbol{\Sigma}_{oo} = \mathrm{Cov}(\mathbf{y}_o, \mathbf{y}_u)$, $\boldsymbol{\Sigma}_{ou} = \mathrm{Cov}(\mathbf{y}_u, \mathbf{y}_o)$, $\boldsymbol{\Sigma}_{uo} = \mathrm{Cov}(\mathbf{y}_u, \mathbf{y}_o)$, and $\boldsymbol{\Sigma}_{uu} = \mathrm{Cov}(\mathbf{y}_u, \mathbf{y}_u)$. Generally, $\boldsymbol{\Sigma}_{ou}$, $\boldsymbol{\Sigma}_{uo}$, and $\boldsymbol{\Sigma}_{uu}$ are not useful because estimation and prediction only requires $\boldsymbol{\Sigma}_{oo}^{-1}$, the inverse of the covariance matrix of $\mathbf{y}_o$. But in this context, the block representation in equation (25) is relevant because it corresponds to a spatio-temporal process where $\{(\mathbf{s}_i, \mathrm{t}_j)\} = \mathbb{S} \times \mathbb{T}$. Though we have not observed $\mathbf{y}_u$, we do know its spatio-temporal locations are the elements of $\mathbb{S} \times \mathbb{T}$ missing from the spatio-temporal locations of $\mathbf{y}_o$. With a second-order stationarity assumption in space and in time, we can use the spatio-temporal locations of $\mathbf{y}_o$ and $\mathbf{y}_u$ to construct $\boldsymbol{\Sigma}_{ou}$, $\boldsymbol{\Sigma}_{uo}$, and $\boldsymbol{\Sigma}_{uu}$ and by consequence, $\boldsymbol{\Sigma}$. After ordering $\boldsymbol{\Sigma}$ by space within time, equations (21) or (24) can be applied. Then the space within time ordering can be undone, finally yielding $\boldsymbol{\Sigma}^{-1}$. Like $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^{-1}$ can be expressed in block form:

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_{oo} & \tilde{\boldsymbol{\Sigma}}_{ou} \\ \tilde{\boldsymbol{\Sigma}}_{uo} & \tilde{\boldsymbol{\Sigma}}_{uu} \end{bmatrix},$$

where the dimensions of the blocks in $\boldsymbol{\Sigma}^{-1}$ match the dimensions of the blocks in $\boldsymbol{\Sigma}$. Helmert-Wolf blocking enables recovery of $\boldsymbol{\Sigma}_{oo}^{-1}$ through the following matrix operation:

$$\boldsymbol{\Sigma}_{oo}^{-1} = \tilde{\boldsymbol{\Sigma}}_{oo} - \tilde{\boldsymbol{\Sigma}}_{ou} \tilde{\boldsymbol{\Sigma}}_{uu}^{-1} \tilde{\boldsymbol{\Sigma}}_{uo}. \tag{26}$$

The main computational burden in equation (26) is inversion of $\tilde{\boldsymbol{\Sigma}}_{uu}$, which must be computed directly. If n_u is small, this additional computation cost is minimal, and using equation (21) or (24) in combination with equation (26) is nearly as fast as directly applying equation (21) or (24). For the product-sum LMM, the multiplication of two $ST \times ST$ matrices can be avoided by computing $\boldsymbol{\Sigma}_{oo}^{-1} \mathbf{X}_o$, $\boldsymbol{\Sigma}_{oo}^{-1} \mathbf{y}_o$, and $\boldsymbol{\Sigma}_{oo}^{-1} \dot{\mathbf{c}}$ instead of $\boldsymbol{\Sigma}_{oo}^{-1}$. Finally, an analogous result for equation (26) exists for log determinants and is provided in the supplementary material.

4.3. Computational Benchmarks

We compared the algorithms from sections 4.1 and 4.2 to a Cholesky decomposition approach for various spatio-temporal sample sizes ranging from 1,000 to 15,000. Inversion times were recorded as the product of $\boldsymbol{\Sigma}^{-1}$ with a fixed effect design matrix, $\mathbf{X}$, and the product of $\boldsymbol{\Sigma}^{-1}$ with a response vector, $\mathbf{y}$ (recall $\boldsymbol{\Sigma}^{-1}\mathbf{X}$ and $\boldsymbol{\Sigma}^{-1}\mathbf{y}$ are required for estimation of $\boldsymbol{\beta}$). In our application, $\mathbf{X}$ has four columns and $\mathbf{y}$ has one column. These dimensions were chosen to match the dimensions of $\mathbf{X}$ and $\mathbf{y}$ in Sections 5 and 6. Other decompositions in combination with linear forward solves can be used for estimation of $\boldsymbol{\theta}$, estimation of $\boldsymbol{\beta}$, and prediction, but we focus on comparing to the Cholesky decomposition approach because it yields convenient forms for log determinants (required during REML estimation) and directly solves $\boldsymbol{\Sigma}^{-1}\mathbf{X}$ and $\boldsymbol{\Sigma}^{-1}\mathbf{y}$ in a straightforward manner.

The average inversion times for the Cholesky decomposition inversion approach were much slower than for our algorithms, especially at the larger sample sizes (figures 2A and 2B). Among our algorithms, the product was the fastest, followed by the sum-with-error and then the product-sum. The ratio of times for the Cholesky decomposition inversion approach to times using our algorithms increased with the sample size (Figure 2C). Together, figures 2A, 2B, and 2C show these product, sum-with-error, and product-sum algorithms can vastly increase the speed associated with REML estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, CWLS estimation of $\boldsymbol{\beta}$, and best linear unbiased prediction by more quickly acquiring the products involved with $\boldsymbol{\Sigma}^{-1}$. But these computational gains are especially important for REML estimation of $\boldsymbol{\theta}$ because of the repeated inversions of the covariance matrix required. For each of the three algorithms, suppose REML estimation requires 60 iterations to converge to a solution. Figure 2B implies that at a sample size near 15,000, estimating covariance
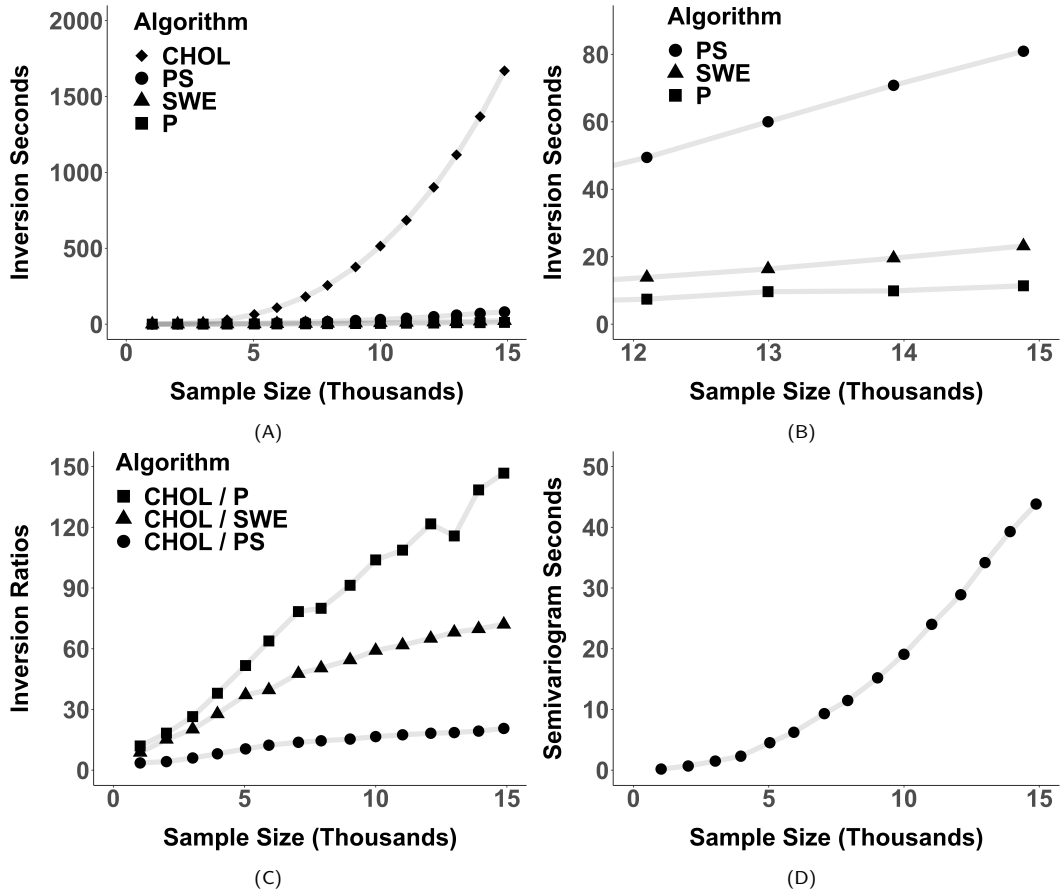
13

Figure 2: Computational benchmarks for various sample sizes. Total sample sizes were determined by taking all combinations of equal numbers of unique spatial locations and unique temporal locations and then removing a single spatio-temporal location so $\{(\mathbf{s}_i, t_j)\} \subset \mathbb{S} \times \mathbb{T}$. All plot symbols indicate the average of 100 repetitions at each sample size. (A) shows the average time (in seconds) for the product (P), sum-with-error (SWE), product-sum (PS), and Cholesky (CHOL) inversion algorithm approaches. In (A), discerning between the P, SWE, and PS algorithms is difficult, so (B) zooms in on both the x and y axes. (C) shows the ratios of average inversion approach times for the Cholesky decomposition relative to our algorithms. (D) shows the average computational time (in seconds) for the empirical semivariogram.

Table 1: Models, estimation methods, and abbreviations.

Model	Estimation Method	Abbreviation
Product LMM	Restricted Maximum Likelihood	P_{REML}
Product LMM	Cressie's Weighted Least Squares	P_{CWLS}
Sum-With-Error LMM	Restricted Maximum Likelihood	SWE_{REML}
Sum-With-Error LMM	Cressie's Weighted Least Squares	SWE_{CWLS}
Product-Sum LMM	Restricted Maximum Likelihood	PS_{REML}
Product-Sum LMM	Cressie's Weighted Least Squares	PS_{CWLS}
Independent Random Error	Ordinary Least Squares	IRE_{OLS}

parameters of the product, sum-with-error, and product-sum LMM's could take approximately 10 minutes, 20 minutes, and 80 minutes, respectively.

CWLS estimation of $\boldsymbol{\theta}$ does not require inverting a covariance matrix, but it does require calculating the empirical semivariogram. Computational times associated with the empirical semivariogram are provided in Figure 2D for the same sample sizes used to benchmark inversions. Unsurprisingly, the computational time grows with the sample size, increasing to just over a minute at a sample size of 15,000. At a sample size of 15,000, estimating covariance parameters of the product, sum-with-error, and product-sum LMM's could take roughly one minute.

It is clear that CWLS estimation is far more computationally efficient than REML estimation when considering REML often needs several inversion iterations (figures 2A and 2B) to converge to a solution. But REML often outperforms CWLS in the context of estimation and prediction (Zimmerman and Zimmerman, 1991; Lark, 2000; Ver Hoef and Cressie, 2001; Minasny and McBratney, 2005; Bevilacqua et al., 2012). In Sections 5 and 6, we explore the balance between performance and computational efficiency using simulated and real data, respectively.

5. Simulation Study

We used a simulation study to compare among the product, sum-with-error, and product-sum LMMs (dependent random error models), to compare restricted maximum likelihood (likelihood-based) and Cressie's weighted least squares (semivariogram-based) estimation, and to compare the dependent random error models to an independent random error model (estimated using ordinary least squares). The model and estimation method combinations of interest are summarized and given relevant abbreviations in Table 1.

In total, we considered four separate simulation scenarios. Before discussing the differences among the scenarios, we present their similarities. Each simulation scenario consisted of 2000 independent trials. In each trial, we selected 35 random spatial locations in $[0, 1] \times [0, 1]$ and 35 equally spaced time points in $[0, 1]$. The response variables, denoted by $\mathbf{y}$, were simulated at all combinations of these spatio-temporal locations and had a consistent mean structure. This mean structure was $\mathbf{X}\boldsymbol{\beta} \equiv \beta_0 + \beta_s \mathbf{x}_s + \beta_t \mathbf{x}_t + \beta_{st} \mathbf{x}_{st}$, where $\mathbf{x}_s, \mathbf{x}_t$, and $\mathbf{x}_{st}$ are covariates simulated independently from zero mean Gaussian distributions with unit variance: $\mathbf{x}_s$ varies through space but not time, $\mathbf{x}_t$ varies through time but not space, and $\mathbf{x}_{st}$ varies through space and time. For example, if $\mathbf{y}$ is daily maximum temperature, $\mathbf{x}_s$ may represent elevation, $\mathbf{x}_t$ may represent day-of-the-month, and $\mathbf{x}_{st}$ may represent precipitation. In all trials, each parameter in $\boldsymbol{\beta}$ was fixed at zero, implying the true variability in $\mathbf{y}$ was driven by only the random errors. Though we discuss the random errors in more detail in the next paragraph, their covariances always depended on the exponential spatial correlation, $\mathbf{R}_s \equiv \exp(-3\|\mathbf{h}_s\|/\phi)$, and the tent (linear-with-sill) temporal

15

Table 2: Random error structure in simulation scenarios 1 (S1) and 2 (S2).

	$\text{Cov}(\boldsymbol{\delta})$	$\text{Cov}(\boldsymbol{\gamma})$	$\text{Cov}(\boldsymbol{\tau})$	$\text{Cov}(\boldsymbol{\eta})$	$\text{Cov}(\boldsymbol{\omega})$	$\text{Cov}(\boldsymbol{\nu})$	Total Variance
Scenario 1	$4 \times \mathbf{R}_s$	$4 \times \mathbf{I}_s$	$4 \times \mathbf{R}_t$	$4 \times \mathbf{I}_s$	$10 \times \mathbf{R}_t \otimes \mathbf{R}_s$	$4 \times \mathbf{I}_{st}$	30
Scenario 2	$18 \times \mathbf{R}_s$	$0 \times \mathbf{I}_s$	$10 \times \mathbf{R}_t$	$0 \times \mathbf{I}_s$	$0 \times \mathbf{R}_t \otimes \mathbf{R}_s$	$2 \times \mathbf{I}_{st}$	30

correlation, $\mathbf{R}_t \equiv (1 - |\mathrm{h}_t|/\kappa) \times \mathrm{I}\{|\mathrm{h}_t| \leqslant \kappa\}$. In $\mathbf{R}_s$, $||\mathbf{h}_s||$ is the Euclidean spatial distance, and ϕ is the spatial range. In $\mathbf{R}_t$, $|\mathrm{h}_t|$ is the absolute temporal distance, κ is the temporal range, and $\mathrm{I}\{|\mathrm{h}_t| \leqslant \kappa\}$ is an indicator variable equaling one if $|\mathrm{h}_t| \leqslant \kappa$ and zero otherwise. These spatial and temporal range parameters were given values $\phi = 1/\sqrt{2}$ and $\kappa = 1/2$, implying observations with spatial distances larger than ϕ and temporal distances larger than κ were approximately uncorrelated.

Random errors in Scenario 1 and Scenario 2 were simulated from Gaussian distributions with covariance structures in Table 2. In Scenario 1, one-third of the total variance is from the spatio-temporal dependent random error, and the leftover variability is spread out evenly among the remaining random errors. We expected the sum-with-error LMM may struggle in this scenario relative to the product and product-sum LMM's because the sum-with-error LMM lacks a component modeling $\mathbf{R}_t \otimes \mathbf{R}_s$. In Scenario 2, the total variance is concentrated in the spatial dependent random error, temporal dependent random error, and completely independent random error components. We expected the product LMM may struggle in Scenario 2 relative to the sum-with-error and product-sum LMM's because the parameter confounding in the product-LMM (11) forces a completely independent error of zero if the proportion of spatial independent random error and temporal independent random error both equal zero. In Scenario 2, the proportion of spatial independent random error and temporal independent random error both equaled zero, but the completely independent random error equaled two.

Scenarios 3 and 4 had similar structures to Scenarios 1 and 2, respectively, but their errors were not Gaussian. For each trial in Scenario 3 (or similarly, Scenario 4), the random errors $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$, $\boldsymbol{\eta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\nu}$ were simulated using a Gaussian distribution and the covariance configuration from Scenario 1 (or similarly, Scenario 2). Each random error was then squared and rescaled so the sample variance of these squared errors matched the sample variance of the Gaussian errors. This squaring and rescaling gives the errors larger tails than would be expected under a Gaussian distribution.

Of the 1225 elements in $\mathbf{y}$, 1224 (n_o) elements were randomly treated as training data and used to estimate the covariance parameters and fixed effects. The remaining one (n_u) observation was treated as test data and used to compare against predictions made at its location.

We quantified the effectiveness of each model and estimation combination method by assessing fixed effect and prediction performance. Studying fixed effect performance was important because fixed effects are often used to understand scientific associations between response variables and covariates. Fixed effect performance was evaluated using type I error rates, mean bias, and root-mean-squared error. Studying prediction performance was important because it is often of interest to make predictions at unobserved locations. Prediction performance was evaluated using prediction interval coverage rates, mean prediction bias, and root-mean-squared-prediction error. We also quantified the computational cost associated with covariance parameter estimation, thereby providing an initial understanding of the balance between performance and computational feasibility for each model and estimation method combination.

Table 3: Type I error rates of $\hat{\beta}_s, \hat{\beta}_t$, and $\hat{\beta}_{st}$ for all model and estimation method combinations (Model$_{\text{Method}}$) in all four simulation scenarios (S1, S2, S3, S4).

Model$_{\text{Method}}$	S1			S2			S3			S4		
	$\hat{\beta}_s$	$\hat{\beta}_t$	$\hat{\beta}_{st}$	$\hat{\beta}_s$	$\hat{\beta}_t$	$\hat{\beta}_{st}$	$\hat{\beta}_s$	$\hat{\beta}_t$	$\hat{\beta}_{st}$	$\hat{\beta}_s$	$\hat{\beta}_t$	$\hat{\beta}_{st}$
P$_{\text{REML}}$	.139	.050	.110	.091	.400	.084	.093	.145	.052	.257	.277	.105
P$_{\text{CWLS}}$	.110	.167	.089	.129	.128	.393	.119	.198	.099	.093	.323	.184
SWE$_{\text{REML}}$	.058	.065	.046	.061	.058	.051	.061	.063	.041	.059	.093	.053
SWE$_{\text{CWLS}}$	.064	.091	.074	.056	.085	.067	.060	.117	.089	.063	.155	.064
PS$_{\text{REML}}$	.058	.058	.049	.063	.062	.048	.038	.070	.047	.061	.096	.054
PS$_{\text{CWLS}}$	.089	.089	.049	.057	.072	.077	.089	.153	.054	.065	.160	.131
IRE$_{\text{OLS}}$	.595	.523	.054	.695	.492	.046	.599	.461	.047	.702	.348	.061

Table 4: Root-Mean-Squared-Errors of $\hat{\beta}_s, \hat{\beta}_t$, and $\hat{\beta}_{st}$ for all model and estimation method combinations (Model$_{\text{Method}}$) in all four simulation scenarios (S1, S2, S3, S4).

Model$_{\text{Method}}$	S1			S2			S3			S4		
	$\hat{\beta}_s$	$\hat{\beta}_t$	$\hat{\beta}_{st}$	$\hat{\beta}_s$	$\hat{\beta}_t$	$\hat{\beta}_{st}$	$\hat{\beta}_s$	$\hat{\beta}_t$	$\hat{\beta}_{st}$	$\hat{\beta}_s$	$\hat{\beta}_t$	$\hat{\beta}_{st}$
P$_{\text{REML}}$	.514	.396	.068	.551	.199	.053	.506	.342	.062	.590	.181	.023
P$_{\text{CWLS}}$	.513	.400	.069	.488	.185	.059	.506	.342	.067	.573	.159	.026
SWE$_{\text{REML}}$	.501	.397	.086	.440	.145	.041	.506	.343	.097	.538	.151	.019
SWE$_{\text{CWLS}}$	.513	.400	.086	.468	.161	.041	.508	.345	.097	.564	.160	.019
PS$_{\text{REML}}$	.493	.391	.064	.442	.145	.041	.487	.337	.061	.537	.154	.017
PS$_{\text{CWLS}}$	.515	.399	.066	.470	.159	.042	.501	.344	.065	.565	.157	.019
IRE$_{\text{OLS}}$	.572	.483	.147	.687	.436	.135	.580	.431	.145	.740	.378	.135

5.1. Fixed Effect Performance

For type I error rates, the test statistic associated with each β parameter equals $|\hat{\beta}|/\hat{\text{SE}}(\hat{\beta})$, where $\hat{\text{SE}}(\cdot)$ denotes the estimated standard error. Though the null distributions of these test statistics are generally unknown, each was approximated by a zero mean Gaussian distribution with unit variance. As a result, Type I error rates were estimated at a significance level of 0.05 by calculating the rate at which the the test statistic exceeded 1.96, a standard cutoff for a two-sided Gaussian hypothesis test. An estimated type I error rate is valid if it is within $[0.04, 0.06]$, where the half-width of this interval approximately equals the margin of error for a 95% binomial confidence interval with probability (of rejection) equaling 0.05 and sample size of 2000. Mean bias was estimated as each $\hat{\beta}$'s average deviation from β across the 2000 trials. RMSE was estimated as the square root of each $\hat{\beta}$'s average squared deviation from β across the 2000 trials.

Type I error rates are summarized in Table 3 for β_s, β_t, and β_{st}. P$_{\text{REML}}$ and P$_{\text{CWLS}}$ type I error rates were generally too large, sometimes as high as 40%. SWE$_{\text{REML}}$, SWE$_{\text{CWLS}}$, PS$_{\text{REML}}$, and PS$_{\text{CWLS}}$ type I error rates were often valid in S1 and S2 (Gaussian errors), but they tended to be slightly larger and outside of the valid range in S3 and S4 (non-Gaussian errors). Furthermore, the SWE$_{\text{CWLS}}$ and PS$_{\text{CWLS}}$ type I error rates were usually larger than their REML counterparts. IRE$_{\text{OLS}}$ type I error rates were far too large for $\hat{\beta}_s$ and $\hat{\beta}_t$ but valid or nearly valid for $\hat{\beta}_{st}$.

Mean bias was approximately zero for all models and estimation methods in all four scenarios. As a result, we leave the table summarizing mean bias for the supplementary material.

Root-Mean-Squared-Errors (RMSE) are summarized in Table 4 for β_s, β_t, and β_{st}. In S1, S3, and S4, PS$_{\text{REML}}$ had the lowest (best) RMSE, followed closely by PS$_{\text{CWLS}}$ and then by the other

dependent error models. In S2, SWE_{REML}, SWE_{CWLS}, PS_{REML} has the same RMSE after rounding, followed closely by and PS_{CWLS} and then by the product LMM's. In S1 and S3, the deficiencies of the sum-with-error LMM are most apparent in $\hat{\beta}_3$'s RMSE, which was roughly 25%-50% higher than $\hat{\beta}_3$ RMSE for the other dependent error models. In S2 and S4, the deficiencies of the product LMM are most apparent in $\hat{\beta}_3$'s RMSE, which was roughly 30%-50% higher than $\hat{\beta}_3$ RMSE for the other dependent error models. Finally, IRE_{OLS} had the highest RMSE across all scenarios, and performance relative to the dependent error models was worst for $\hat{\beta}_{st}$. These results showed the relative discrepancies in RMSE among all models were most drastic for the $\hat{\beta}_{st}$ estimate associated with $\mathbf{x}_{st}$, the covariate varying across space and time.

Type I error rate and RMSE performance for $\hat{\beta}_3$ was much better than $\hat{\beta}_1$ or $\hat{\beta}_2$ performance for all models and estimation methods across all simulation scenarios. This likely occurred because the $\mathbf{x}_s$ and $\mathbf{x}_t$ covariates associated with $\hat{\beta}_s$ and $\hat{\beta}_t$ are patterned in space and in time, causing noticeable variance inflation in the dependent error models (Reich et al., 2006).

5.2. Prediction Performance

We quantified prediction performance by evaluating prediction interval coverage rates, mean prediction bias, and root-mean-squared-prediction error (RMSPE). Recall that in each simulation trial, we simulated $\mathbf{y}$ at all 1225 combinations of 35 spatial locations and 35 time points. Then a random location was chosen and its associated element in $\mathbf{y}$, denoted by y_u, was held out and used to evaluate prediction performance. We did explore holding out up to 100 $\mathbf{y}$ values in each trial, but there was little change in the prediction performance metrics compared to holding out a single value. This is likely because of the large number of simulation trials (2000).

We predicted y_u and its associated variance using equations (19) and (20) evaluated at $\hat{\boldsymbol{\theta}}$. Prediction interval coverage rates were estimated by calculating the rate at which each element of y_u is contained in its 95% Gaussian prediction interval. The estimated prediction interval coverage rate is valid if it is within $[0.94, 0.96]$, where the half-width of this interval approximately equals the margin of error for a 95% binomial confidence interval with probability (of coverage) equaling 0.95 and sample size of 2000. Mean prediction bias was estimated as each $\hat{y}_u$'s average deviation from y_u across the 2000 trials. RMSPE was estimated as the square root of the average of each $\hat{y}_u$'s squared deviation from y_u across the 2000 trials.

Prediction interval coverage rates are summarized in Table 5. In S1, coverage rates were a few percentage points too low for P_{CWLS} and SWE_{CWLS} but valid for the other model and estimation method combinations. In S2, coverage rates dropped for all model and estimation method combinations, and only SWE_{REML} and IRE_{OLS} had valid coverage rates. The coverage rate was especially low for P_{CWLS}, which tended to consistently overestimate the spatial and temporal range parameters and associated correlation. In S3 and S4, no models or estimation methods had valid coverage rates, and coverage rates generally decreased relative to their respective S1 and S2 counterparts. This drop in coverage rates was likely due to the non-Gaussian errors used to simulate the data while a Gaussian assumption was implied while constructing the prediction intervals. Coverage for P_{CWLS} increased in S4 relative to S2, likely because there was less overestimation of the strength of the spatial and temporal range parameters and associated correlation.

Mean prediction bias was approximately zero for all models and estimation methods in all four scenarios. As a result, we leave a table summarizing mean prediction bias for the supplementary material.

Root-Mean-Squared-Prediction-Errors (RMSPE) are also summarized in Table 5. In S1 and S3, PS_{REML} has the lowest (best) RMSPE, followed closely by P_{REML}, PS_{CWLS}, and P_{CWLS}. In these scenarios, RMSPE for SWE_{REML}, and SWE_{CWLS} was approximately 26% to 71% higher

18

Table 5: Coverage rate (Coverage) and root-mean-squared-prediction error (RMSPE) for all model and estimation method combinations (Model$_{\text{Method}}$) in all four simulation scenarios (S1, S2, S3, S4).

Model$_{\text{Method}}$	Coverage				RMSPE			
	S1	S2	S3	S4	S1	S2	S3	S4
P$_{\text{REML}}$	.950	.925	.928	.927	2.379	1.764	2.155	0.733
P$_{\text{CWLS}}$	.910	.661	.916	.873	2.474	1.848	2.247	0.856
SWE$_{\text{REML}}$	.949	.940	.931	.935	3.115	1.500	3.641	0.629
SWE$_{\text{CWLS}}$	.917	.912	.902	.931	3.114	1.506	3.645	0.654
PS$_{\text{REML}}$	.952	.938	.936	.935	2.342	1.504	2.130	0.630
PS$_{\text{CWLS}}$	.956	.901	.936	.878	2.388	1.516	2.295	0.705
IRE$_{\text{OLS}}$	.946	.955	.928	.931	5.215	4.707	5.419	4.885

Table 6: The average empirical semivariogram calculation seconds (SV Sec.), average covariance parameter estimation seconds (Est. Sec.), average total empirical semivariogram calculation and covariance parameter estimation seconds (Tot. Sec; the sum of SV Sec. and Est. Sec), and average REML iterations (REML Iter.) for all models and estimation methods (Model$_{\text{Method}}$) in S1.

Model$_{\text{Method}}$	SV Sec.	Est. Sec.	Tot. Sec.	REML Iter.
P$_{\text{REML}}$	NA	5.51	5.51	96.82
P$_{\text{CWLS}}$	0.31	0.08	0.39	NA
SWE$_{\text{REML}}$	NA	3.81	3.81	46.94
SWE$_{\text{CWLS}}$	0.31	0.18	0.49	NA
PS$_{\text{REML}}$	NA	11.00	11.00	58.77
PS$_{\text{CWLS}}$	0.31	0.26	0.57	NA
IRE$_{\text{OLS}}$	NA	0.01	0.01	NA

than RMSPE for the other dependent error models. Alternatively in S2 and S4, SWE$_{\text{REML}}$ has the lowest (best) RMSPE, followed closely by PS$_{\text{REML}}$, SWE$_{\text{CWLS}}$, and PS$_{\text{CWLS}}$. In these scenarios, RMSPE for P$_{\text{REML}}$, and P$_{\text{CWLS}}$ was approximately 5% to 36% higher than RMSPE for the other dependent error models. Finally, IRE$_{\text{OLS}}$ had the highest RMSPE across all scenarios.

5.3. Computational Performance

We quantified computational performance by measuring and contextualizing the average time required to estimate the covariance parameters for all models and estimation methods. These times are summarized in Table 6 for S1. In S1, covariance parameter estimation for P$_{\text{REML}}$ took an average of 5.51 seconds and 96.82 iterations, for SWE$_{\text{REML}}$ took an average of 3.81 seconds and 46.94 iterations, and for PS$_{\text{REML}}$ took an average of 11.00 seconds and 58.77 iterations. Though inverting the product covariance matrix is faster than inverting the sum-with-error covariance matrix, the extra inversions required resulted in the increased computational time for P$_{\text{REML}}$. Covariance parameter estimation for CWLS generally took around half a second; the majority of this computational burden was from calculating the empirical semivariogram (0.31 seconds). Irrespective of the LMM in question, CWLS estimation was much faster than REML estimation. All dependent error models took far longer to estimate than IRE$_{\text{OLS}}$, however.

Though computational performance tables for S2, S3, and S4 are saved for the supplementary material, the general takeaways are the same: The speed of REML estimation can vary greatly depending on how many iterations are required to converge to a solution, REML estimation is very slow relative to CWLS estimation, and the most computationally costly part of CWLS estimation
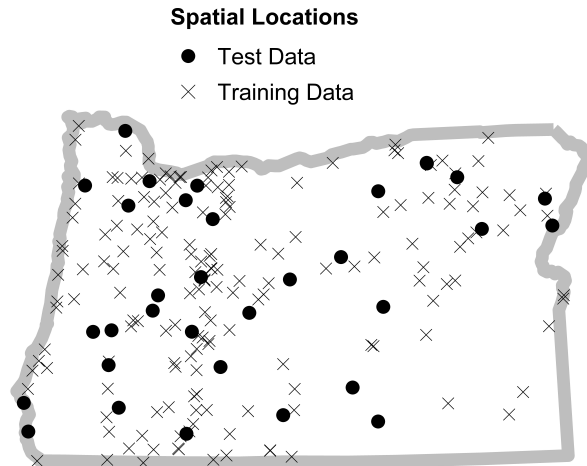
Figure 3: Oregon station locations in the training and test data observed for at least one day in July.

is calculation of the empirical semivariogram. All computations were performed on an Intel(R) Xeon(R) CPU E5-2690 v3 @ 2.60 GHz processor using a single core and 227 GB of available RAM. Further computational details are provided in the supplementary material.

6. Application: Oregon Daily Maximum Temperature

It is often of interest to study the effect of environmental variables on daily temperature patterns. Oregon is a wet, mountainous state in the United States of America that has varying climate regions and moderately warm summers. We used the same model and estimation method combinations from Section 5 (Table 1) to explain variation in daily maximum temperature (Farenheit) in Oregon during each day in July, 2019. Data were obtained through the National Oceanic and Atmospheric Administration's Global Historical Climate Network. To compute distances in terms of kilometers, we used a Transverse Mercator projection (Lambert, 1972).

Subsets of the full data were used as training data and test data. The training data contained observations from 33 randomly selected weather stations at all time points available. Some weather stations in the training data were not observed at every time point; the training data contained 972 of the 1023 possible observations. The training data were used to estimate the covariance parameters and some fixed effects. The test data contained 2000 observations that were randomly selected from the full data after removing observations from the training data. The test data were used to evaluate prediction performance. Unique spatial locations in the training and test data are shown in Figure 3.

We modeled daily maximum temperature having mean structure $\mathbf{X}\boldsymbol{\beta} \equiv \beta_0 + \beta_{elev}\mathbf{x}_{elev} + \beta_{day}\mathbf{x}_{day} + \beta_{prcp}\mathbf{x}_{prcp}$, where $\mathbf{x}_{elev}$ is weather station elevation (in meters above mean sea level), $\mathbf{x}_{day}$ is day-of-the-month, and $\mathbf{x}_{prcp}$ is daily precipitation (in millimeters). This mean structure matches the mean structure from the simulation study in Section 5: One covariate varied through space but not time (elevation), one covariate varied through time but not space (day-of-the-month), and one covariate varied through space and time (precipitation). The empirical spatio-temporal semivariogram suggested the exponential correlation was a reasonable choice to model the spatial and temporal correlations. The model and estimation methods from

Table 7: Coverage rates (Coverage), root-mean-squared-prediction-error (RMSPE), semivariogram calculation seconds (SV Sec.), covariance parameter estimation seconds (Est. Sec.), total semivariogram calculation and covariance parameter estimation seconds (Tot. Sec; the sum of SV Sec. and Est. Sec.), and REML iterations (REML Iter.) for all models and estimation methods ($Model_{Method}$) used to analyze the daily maximum temperature data.

$Model_{Method}$	Coverage	RMSPE	SV Sec.	Est. Sec.	Tot. Sec.	REML Iter.
P_{REML}	.895	6.097	NA	12.63	12.63	179
P_{CWLS}	.913	7.828	0.17	0.07	0.24	NA
SWE_{REML}	.965	4.826	NA	8.62	8.62	95
SWE_{CWLS}	.907	7.584	0.17	0.16	0.33	NA
PS_{REML}	.959	4.644	NA	12.79	12.79	80
PS_{CWLS}	.955	6.415	0.17	0.36	0.53	NA
IRE_{OLS}	.911	8.136	NA	0.01	0.01	NA

Section 5 (Table 1) were used to study estimation, prediction, and computational performance of these data.

All model and estimation method combinations found a strong, positive association between day-of-the-month and daily maximum temperature (all Gaussian-based p-values from $< .001$ to .038) and a strong, negative association between elevation and daily maximum temperature (all Gaussian-based p-values from $< .001$ to .043). IRE_{OLS} found a strong, negative association between precipitation and daily maximum temperature (Gaussian-based p-value $< .001$), but the dependent random error models found less evidence of this association (Gaussian-based p-values from .096 to .224). This discrepancy is likely related to the variance inflation in the dependent random error models discussed in Section 5.1, though in the simulation scenarios the inflation was most apparent for the parameters associated with the spatially patterned and temporally patterned covariates.

Test data prediction performance for each model and estimation method combination was summarized using familiar metrics: 95% Gaussian prediction interval coverage, mean prediction bias, and root-mean-squared-prediction-error (RMSPE). Table 7 shows PS_{REML} and PS_{CWLS} had coverage rates closest to 0.95, PS_{SWE} had coverage rates that were slightly too high, and P_{REML}, P_{CWLS}, SWE_{CWLS}, and IRE_{OLS} had coverage rates that were slightly too low. Among the dependent error models, PS_{REML} has the lowest RMSPE, followed in order by SWE_{REML}, P_{REML}, PS_{CWLS}, SWE_{CWLS}, and finally P_{CWLS}. The dependent random error models all had lower RMSPE than IRE_{OLS}.

Mean prediction bias was approximately zero for all models and estimation methods. As a result, we leave summarizing mean prediction bias for the supplementary material.

A fitted semivariogram using the model and estimation combination method yielding the lowest RMSPE, PS_{REML}, is provided in Figure 4. Based on Figure 4, this estimated process has spatial covariance decaying to zero at nearly 1,500 kilometers, has temporal covariance decaying to zero at nearly 4 days, and is primarily influenced by the spatial dependent random error.

Computational performance was also summarized in Table 7 for all model and estimation method combinations. Similar to Section 5, REML estimation (8.62 - 12.79 seconds) took far longer than CWLS estimation (0.24 - 0.53 seconds). All computations were performed on an Intel(R) Xeon(R) CPU E5-2690 v3 @ 2.60 GHz processor using a single core and 227 GB of available RAM. Further computational details are provided in the supplementary material.
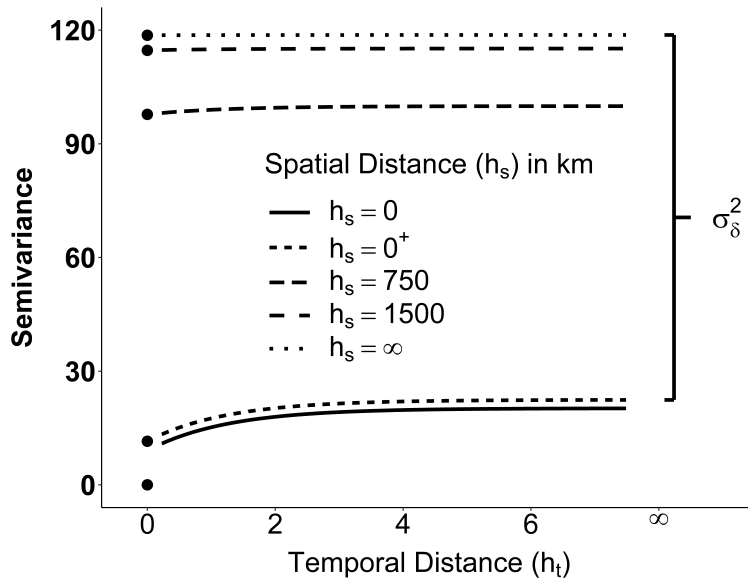
21

Figure 4: Fitted PS$_{\text{REML}}$ spatio-temporal semivariogram. The fitted semivariogram is viewed with temporal distance on the x-axis and spatial distance using line types. The spatial distance of 0^+ indicates a right limit approaching zero. The largest variance estimate, σ^2_δ (spatial dependent variance or spatial partial sill), is identified using a bracket.

7. Discussion

In this paper, we described spatio-temporal random processes using linear mixed models (LMMs). This approach builds upon the single random error formulation and partitions sources of spatial and temporal variability into distinct random effects. The spatio-temporal LMM's general, flexible framework accommodates many commonly used covariance functions as special cases. It can also be used to model spatio-temporal processes whose covariance functions are not second-order stationary or not isotropic (isotropy implies the covariance function behaves similarly in all directions), though we leave those types of covariance functions for future research. Building from the sum covariance function, we proposed the sum-with-error covariance function, which is strictly positive definite. Though there are many elegant spatio-temporal covariance functions and estimation methods in the literature, we focused on studying three familiar covariance functions (product, sum-with-error, and product-sum) and two familiar estimation methods (REML and CWLS).

Our novel algorithms manipulate the structure of the product, sum-with-error, and product-sum covariance matrices to efficiently invert them, even when every spatial location is not observed at every time point. Though these algorithms are most useful for extending the range of sample sizes for which REML estimation can applied, they are also useful for best linear unbiased estimation of $\boldsymbol{\beta}$ and best linear unbiased prediction (Kriging). Our algorithms can also be combined with other computationally efficient approaches that require inversion of a covariance matrix, such as Fixed Rank Kriging.

We found that in the processes we studied, REML estimation tended to outperform CWLS estimation, even when the processes' errors were not Gaussian. But the allure of this increased performance is diminished when considering the significantly increased computational burden associated with REML estimation. At the lower sample sizes, this computational performance gap may

22

not make much practical difference. For example, in S1, the longest average estimation time was 11 seconds. But as the sample size increases, REML's computational limitations become a more important consideration. Moreover, model comparison is easier using CWLS estimation because its main computational burden is calculating the empirical semivariogram, which can be reused to estimate covariance parameters of separate models. Estimating covariance parameters of separate models using REML estimation, however, requires separate optimization routines, each requiring repeated covariance matrix inversions. One advantage of model selection via REML, however, is that likelihood-based statistics (e.g. AIC) can be leveraged. An alternative approach for REML estimation is to inform model selection through inspection of the empirical semivariogram (or covariance function) based on the representations in Figure 1 and the supplementary material. These visualizations can aid in understanding how well each model may perform before estimation. For example, if the empirical covariance or semivariogram suggests σ_ω^2 (the spatio-temporal dependent error) is much smaller than the other variance parameters, the sum-with-error LMM will likely fit the data well and is more computationally efficient than the product-sum LMM.

We identified scenarios in which the sum-with-error LMM is inferior to the product and product-sum LMMs (S1, S3) and scenarios in which the product LMM is inferior to the sum-with-error and product-sum LMMs (S2, S4, temperature data). And while the product-sum LMM generally performed best, estimating its covariance parameters (using either estimation method) always had a higher computational cost than estimating the covariance parameters of the other LMMs.

It is important to emphasize that all spatio-temporal random processes have unique characteristics governing their covariance structure, and even among the model and estimation methods combinations discussed in this paper, there is certainly no combination that is uniformly best for all applications. The balance between model performance and computational efficiency must be at the core of a practitioner's decision making process when determining which model and estimation methods to implement.

8. Acknowledgments

10. Data and Code Availability

All data and code used for this paper are available on GitHub at https://github.com/michaeldumelle/DumelleEtAl2021STLMM. The model and estimation methods studied in this paper will be available as part of a forthcoming R package named stmodel.

11. Declarations of Interest

none.

12. Funding

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Bevilacqua, M., Gaetan, C., Mateu, J., Porcu, E., 2012. Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. Journal of the American Statistical Association 107, 268–280.

Cressie, N., 1985. Fitting variogram models by weighted least squares. Journal of the International Association for Mathematical Geology 17, 563–586.

Cressie, N., 1993. Statistics for Spatial Data, Revised Edition. Wiley.

Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70, 209–226.

Cressie, N., Lahiri, S.N., 1993. The asymptotic distribution of reml estimators. Journal of multivariate analysis 45, 217–233.

Cressie, N., Lahiri, S.N., 1996. Asymptotics for reml estimation of spatial covariance parameters. Journal of Statistical Planning and Inference 50, 327–341.

Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. Wiley.

De Cesare, L., Myers, D., Posa, D., 1997. Spatial-temporal modelling of so 2 in milan district. e. baafi (ed.), geostatistics wollongong '96 vol. 2.

De Cesare, L., Myers, D., Posa, D., 2001. Estimating and modeling space–time correlation structures. Statistics & Probability Letters 51, 9–14.

De Iaco, S., Myers, D., Posa, D., 2011. On strict positive definiteness of product and product–sum covariance models. Journal of Statistical Planning and Inference 141, 1132–1140.

De Iaco, S., Myers, D.E., Posa, D., 2001. Space–time analysis using a general product–sum model. Statistics & Probability Letters 52, 21–28.

De Iaco, S., Palma, M., Posa, D., 2015. Spatio-temporal geostatistical modeling for french fertility predictions. Spatial Statistics 14, 546–562.

De Iaco, S., Posa, D., 2018. Strict positive definiteness in geostatistics. Stochastic environmental research and risk assessment 32, 577–590.

De Iaco, S., Posa, D., Cappello, C., Maggio, S., 2019. Isotropy, symmetry, separability and strict positive definiteness for covariance functions: A critical review. Spatial statistics 29, 89–108.

Fuller, W.A., Battese, G.E., 1973. Transformations for estimation of linear models with nested-error structure. Journal of the American Statistical association 68, 626–632.

Gneiting, T., 2002. Nonseparable, stationary covariance functions for space–time data. Journal of the American Statistical Association 97, 590–600.

Gneiting, T., Genton, M.G., Guttorp, P., 2006. Geostatistical space-time models, stationarity, separability, and full symmetry.

Haas, T.C., 1995. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. Journal of the American Statistical Association 90, 1189–1199.

Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association 72, 320–338.

Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M., et al., 2019. A case study competition among methods for analyzing large spatial data. Journal of Agricultural, Biological and Environmental Statistics 24, 398–425.

Heyde, C., 1994. A quasi-likelihood approach to the reml estimating equations. Statistics & Probability Letters 21, 381–384.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., Lee, T.C., 1985. The Theory and Practice of Econometrics. John Wiley & Sons, New York.

Kitanidis, P.K., 1993. Generalized covariance functions in estimation. Mathematical Geology 25, 525–540.

Lahiri, S.N., Lee, Y., Cressie, N., 2002. On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. Journal of Statistical Planning and Inference 103, 65–85.

Lambert, J., 1972. Notes and comments on the composition of terrestrial and celestial maps, michigan geographical publication no. 8, translated by waldo tobler, department of geography, university of michigan. Ann Arbor .

Lark, R., 2000. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. European Journal of Soil Science 51, 717–728.

Mardia, K.V., Marshall, R.J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 71, 135–146.

Minasny, B., McBratney, A.B., 2005. The matérn function as a general model for soil variograms. Geoderma 128, 192–207.

Montero, J.M., Fernández-Avilés, G., Mateu, J., 2015. Spatial and Spatio-Temporal Geostatistical Modeling and Kriging. Wiley.

Myers, D.E., Journel, A., 1990. Variograms with zonal anisotropies and noninvertible kriging systems. Mathematical Geology 22, 779–785.

Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554.

Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. RFID Journal 8, 204–218.

Pebesma, E.J., 2004. Multivariable geostatistics in s: the gstat package. Computers & Geosciences 30, 683–691.

Porcu, E., Furrer, R., Nychka, D., 2019. 30 years of space–time covariance functions. Wiley Interdisciplinary Reviews: Computational Statistics , e1512.

Posa, D., 1993. A simple description of spatial-temporal processes. Computational Statistics & Data Analysis 15, 425–437.

Reich, B.J., Hodges, J.S., Zadnik, V., 2006. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics 62, 1197–1206.

Rouhani, S., Hall, T.J., 1989. Space-time kriging of groundwater data, in: Geostatistics. Springer, pp. 639–650.

Schabenberger, O., Gotway, C.A., 2017. Statistical methods for spatial data analysis. CRC press.

Schmidt, P., 1976. Econometrics. Marcel Dekker, New York.

Sherman, J., 1949. Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. The Annals of Mathematical Statistics 20, 621.

Sherman, J., Morrison, W.J., 1950. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. The Annals of Mathematical Statistics 21, 124–127.

Shumway, R.H., Stoffer, D.S., 2017. Time series analysis and its applications: with R examples, fourth edition. Springer Texts in Statistics.

Stegle, O., Lippert, C., Mooij, J.M., Lawrence, N.D., Borgwardt, K., 2011. Efficient inference in matrix-variate gaussian models with iid observation noise, in: Advances in neural information processing systems, pp. 630–638.

Sweeting, T.J., 1980. Uniform asymptotic normality of the maximum likelihood estimator. The Annals of Statistics , 1375–1381.

Theil, H., 1971. Principles of econometrics. John Wiley & Sons, New York.

Ver Hoef, J., Cressie, N., 2001. Spatial statistics: Analysis of field experiments, in: Schenier, S., Gurevitch, J. (Eds.), Design and Analysis of Ecological Experiments: Second Edition. Oxford University Press, pp. 289–307.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer.

Wikle, C.K., Zammit-Mangion, A., Cressie, N., 2019. Spatio-Temporal Statistics in R. volume 1. Chapman & Hall.

Wolf, H., 1978. The helmert block method, its origin and development, in: Proceedings of the Second International Symposium on Problems Related to the Redefinition of North American Geodetic Networks, pp. 319–326.

Wolfinger, R., Tobias, R., Sall, J., 1994. Computing gaussian likelihoods and their derivatives for general linear mixed models. SIAM Journal on Scientific Computing 15, 1294–1310.

Woodbury, M.A., 1950. Inverting modified matrices. Statistical Research Group, Memorandum report No. 42 106.

Zammit-Mangion, A., Cressie, N., 2017. Frk: An r package for spatial and spatio-temporal prediction with large datasets. arXiv preprint arXiv:1705.08105 .

Zimmerman, D.L., Zimmerman, M.B., 1991. A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. Technometrics 33, 77–91.

Highlights for "A Linear Mixed Model Formulation for Spatio-Temporal Random Processes with Computational Advances for the Product, Sum, and Product-Sum Covariance Functions"

Michael Dumelle[a,1], Jay Ver Hoef[b], Claudio Fuentes[a], Alix Gitelman[a]

[a]*Department of Statistics, Oregon State University, Corvallis, OR 97331*
[b]*Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA Fisheries, Seattle, WA 98115*

1. Highlights

- Described spatio-temporal random processes using linear mixed models and visualized the dependence

- Proposed the strictly positive definite sum-with-error covariance function

- Novel algorithm efficiently inverts product, sum-with-error, and product-sum covariance matrices

- Compared product, sum-with-error, and product-sum covariance function performance

- Compared restricted maximum likelihood and Cressie's weighted least squares estimation performance

---
1 Corresponding author
Email address: dumellem@oregonstate.edu (Michael Dumelle)

Author Contributions for "A Linear Mixed Model Formulation for Spatio-Temporal Random Processes with Computational Advances for the Product, Sum, and Product-Sum Covariance Functions"

Michael Dumelle[a,*], Jay M. Ver Hoef[b], Claudio Fuentes[a], Alix Gitelman[a]

[a]*Department of Statistics, Oregon State University, Corvallis, OR 97331*
[b]*Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA Fisheries, Seattle, WA 98115*

1. CRediT Statement

Michael Dumelle: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing (Original Draft), Writing (Review and Editing)

Jay M. Ver Hoef: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing (Original Draft), Writing (Review and Editing)

Claudio Fuentes: Conceptualization, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing (Original Draft), Writing (Review and Editing)

Alix Gitelman: Conceptualization, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing (Original Draft), Writing (Review and Editing)

---

*Corresponding author
Email address: `dumellem@oregonstate.edu` (Michael Dumelle)

Declarations of Interest for "A Linear Mixed Model Formulation for Spatio-Temporal Random Processes with Computational Advances for the Product, Sum, and Product-Sum Covariance Functions."

Michael Dumelle[a,*], Jay Ver Hoef[b], Claudio Fuentes[a], Alix Gitelman[a]

[a]*Department of Statistics, Oregon State University, Corvallis, OR 97331*
[b]*Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA Fisheries, Seattle, WA 98115*

Declarations of Interest: none.

---

[*]Corresponding author
Email address: dumellem@oregonstate.edu (Michael Dumelle)