Title: Semiparametric Modeling with Nonseparable and Nonstationary Spatio-Temporal Covariance Functions and Its Inference

Manuscript ID: SS-2016-0297

URL: http://www.stat.sinica.edu.tw/statistica/

DOI: 10.5705/ss.202016.0297

Complete List of Authors: Tingjin Chu, Jun Zhu and Haonan Wang

Corresponding Author: Tingjin Chu

E-mail: tingjin_chu@outlook.com

Notice: Accepted version subject to English editing.
Semiparametric Modeling with Nonseparable and Nonstationary Spatio-Temporal Covariance Functions and Its Inference

Tingjin Chu¹, Jun Zhu² and Haonan Wang³

¹School of Mathematics and Statistics, University of Melbourne
²Department of Statistics and Department of Entomology, University of Wisconsin-Madison
³Department of Statistics, Colorado State University

Abstract

In this paper, we develop a new semiparametric approach to modeling geostatistical data repeatedly taken over time and drawing inference about the parameters and components of the underlying spatio-temporal process. Dependence in time and across space is modeled semiparametrically, giving rise to a class of nonseparable and nonstationary spatio-temporal covariance functions. A two-step procedure is devised to estimate the model parameters based on likelihood for detrended data, and the computational algorithm is efficient due to dimension reduction. Extensions to spatio-temporal processes with general mean trends are also considered. Further, the asymptotic properties of our proposed method are established including consistency.
and asymptotic normality. A simulation study shows sound finite-sample properties of the proposed method and a real data example illustrates our method in comparison with alternative approaches.

**Keywords**: Geostatistics, semiparametric methods, spatio-temporal processes

## 1 Introduction

As the real-time monitoring technologies continue to advance, data over space and time are becoming more abundant. Collectively known as spatio-temporal data, these data arise in many scientific fields with different data formats and goals to address. To analyze various types of spatio-temporal data, different statistical models and methods have been developed, such as varying coefficients models (Lu et al., 2009), hierarchial dynamic spatial models (Zhang et al., 2003; Johannesson et al., 2007; Ghosh et al., 2010), filtering and dimensional reduction (Huang and Cressie, 1996; Cressie et al., 2010; Brynjarsdóttir and Berliner, 2014). For more details about spatio-temporal statistics, see Cressie and Wikle (2011).

In this paper, we focus on spatio-temporal data such that the individual units are spatial sampling locations and at a given sampling location, repeated measures are taken over time. We propose a class of semiparametric models with nonseparable and nonstationary spatio-temporal covariance functions. We develop new methodology for inferences that balance model flexibility and computational feasibility, as well as establish the corresponding asymptotic properties.

For modeling spatio-temporal data, it is generally important to incorporate spatio-temporal covariance. Separability, such that the spatio-temporal covariance function is
assumed to be a product of spatial covariance and temporal covariance, is a convenient assumption, but can be overly restrictive in many applications. Thus, various nonseparable spatio-temporal covariance functions have been proposed. For example, Cressie and Huang (1999) and Gneiting (2002) constructed nonseparable spatio-temporal covariance functions by Fourier inversion and completely monotone functions, respectively. While the above approaches allow spatio-temporal nonseparability, the spatio-temporal processes are assumed to be stationary. To relax the stationarity assumption, Stein (2005) considered asymmetric models, that is, the covariance function is spatially isotropic but not symmetric spatio-temporally. Fuentes et al. (2008) and Rodrigues and Diggle (2010) developed nonstationary and nonseparable models via spectral representation and convolution, respectively. While the above methods concern covariance model building for spatio-temporal data, it is not always clear how model estimation and statistical inference are to be carried out. There is a clear need for further development of statistical methodology for the analysis of spatio-temporal data taken at regular or irregular sampling locations.

Nonparametric approaches are increasingly used for spatio-temporal modeling, which tend to be robust against covariance function misspecification. For example, extending the work of Gneiting (2002), Choi et al. (2013) proposed a nonparametric approximation of completely monotone functions in the construction of spatial and spatio-temporal covariance structures. Nonparametric methods may also alleviate computational burden in the estimation of spatio-temporal covariance functions. With an originally parametric covariance function, Zhang et al. (2015) proposed a nonparametric full scale approximation, which applies reduced rank techniques and sparse matrix algorithms to enhance computational
efficiency, although theoretical backing is not given. We believe that there is considerable value to further develop nonparametric or semiparametric methods and explore their theoretical properties in spatio-temporal statistics. Here, we adopt a semiparametric approach to modeling and drawing inferences about the spatio-temporal mean function and covariance function. Moreover, the theoretical properties of our new methods are established, which seem to be still rare in semiparametric spatio-temporal statistics.

In particular, we model dependence over space and time by a Karhunen-Loève type expansion which results in nonseparable and nonstationary spatio-temporal covariance functions. The model parameters are estimated by a two-step procedure based on likelihood and the computational feasibility is further enhanced by dimension reduction. Extensions to spatio-temporal models with a general mean function (or, trend) are also considered. Further, the asymptotic properties of our proposed method, such as consistency and asymptotic normality, are investigated and established. These theoretical results are, to the best of our knowledge, first of its kind for semiparametric methods to infer about spatio-temporal models with nonseparable and nonstationary covariance functions. A simulation study shows sound finite-sample properties of the estimates and a real data example illustrates our method in comparison with some of the existing approaches. Our model may also be regarded as spatially correlated functional data, although without replicates at each sampling location (Paul and Peng, 2011; Gromenko et al., 2012; Hörmann and Kokoszka, 2013). However, Gromenko et al. (2012) did not establish the theoretical properties of their method, whereas Paul and Peng (2011) and Hörmann and Kokoszka (2013) restricted attention to, respectively, separable models and the consistency of sample means and
empirical covariance operators only (see also Horváth and Kokoszka (2012)).

The remainder of the paper is organized as follows. We propose a nonseparable and nonstationary spatio-temporal covariance model in Section 2. We develop an estimation procedure for the detrended spatio-temporal data in Section 3.1, and in Section 4.1, we extend the results to spatio-temporal data with a general mean trend. The theoretical properties of our methodology are established as theorems in Sections 3.2 and 4.2. Numerical examples using simulated and real data are given in Section 5. The technical details including theorem proofs are provided as supplementary materials.

2 Semiparametric Spatio-Temporal Model Formulation

Let $\mathcal{R}$ denote a spatial domain of interest in $\mathbb{R}^d$ with $d \geq 1$, and let $[0, T]$ denote the time interval of interest with $0 < T < \infty$. Taking into account possible spatio-temporal correlation and measurement error, we model the spatio-temporal response variable $y(s, t)$ by

$$y(s, t) = \mu(s, t) + \varepsilon(s, t) + \upsilon(s, t), \tag{1}$$

where $\mu(s, t) = E\{y(s, t)\}$ is a fixed spatio-temporal mean function, $\varepsilon(s, t)$ is a zero-mean spatio-temporal random process, and $\upsilon(s, t)$ is a zero-mean measurement error, where $s \in \mathcal{R}$ and $t \in [0, T]$. For the spatial domain of interest $\mathcal{R}$, our method allows some irregularity such as non-convexity, but the domain needs to be continuous. In this and the following sections, we will assume a zero-mean function $\mu(s, t) = 0$ and consider more general mean
functions in Section 4. We also assume that the measurement error \( \nu(s,t) \) follows iid Gaussian distribution with mean 0 and variance \( \sigma^2 \), independent of \( \varepsilon(s,t) \).

To model the spatio-temporal random process \( \varepsilon(s,t) \), we assume that it is a zero-mean Gaussian process with a Karhunen-Loève (KL) type expansion (Ghanem and Spanos, 1991). That is, \( \varepsilon(s,t) = \sum_{j=1}^{\infty} \xi_j(s) \varphi_j(t) \), where \( \{\xi_j(s) : s \in \mathcal{R}\} \) is assumed to be a sequence of independent Gaussian processes, and \( \{\varphi_j(t)\} \) is a sequence of eigenfunctions. Thus, (1) can be rewritten as

\[
y(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \xi_j(s) \varphi_j(t) + \nu(s,t).
\]

In general, we may assume that, for a given \( s \in \mathcal{R} \), \( \varepsilon(s,t) \in L^2[0,T] \) is a square integrable random function and is modeled by a stochastic (not necessarily Gaussian) process with mean zero and a spatio-temporal covariance function denoted as

\[
\gamma(t,s; t', s') = \text{cov}\{\varepsilon(s,t), \varepsilon(s', t')\}, \quad s, s' \in \mathcal{R}, \quad t, t' \in [0,T].
\]

We also assume that, for two locations \( s \neq s' \), the two curves \( \varepsilon(s,t) \) and \( \varepsilon(s',t') \) have the same (possibly nonstationary) temporal covariance function \( \gamma_0(t,t') \) (see, e.g., Gromenko et al., 2012; H"ormann and Kokoszka, 2013). Sufficient conditions to establish (2) are given in Appendix A of the supplementary materials. Let \( \lambda_j = \text{var}\{\xi_j(s)\} \) denote the \( j \)th eigenvalue of the covariance function \( \gamma_0(t,t') \). We further assume \( \text{cov}\{\xi_j(s), \xi_{j'}(s')\} = 0 \) for \( j \neq j' \), which ensures positive definiteness of the covariance function \( \gamma(t,s; t', s') \) and enhances the computational feasibility (Gromenko and Kokoszka, 2013). From the KL expansion, we can write the spatio-temporal covariance function as

\[
\gamma(t,s; t', s') = \sum_{j=1}^{\infty} \text{cov}\{\xi_j(s), \xi_j(s')\} \varphi_j(t) \varphi_j(t').
\]
In spatio-temporal statistics, it is a common practice to assume that the spatio-temporal process \( \varepsilon(s, t) \) in model (1) is stationary over space and time, whereas \( \varepsilon(s, t) \) formulated here via the KL expansion encompasses spatio-temporal covariances functions that are nonstationary. Specifically, from (3), it is clear that \( \varepsilon(s, t) \) does not need to be stationary in space or time, but can be stationary in space if \( \xi_j(s) \) is a stationary spatial process for all \( j \) (Cressie, 1993). The spatio-temporal covariance function in (3) also does not need to be separable in space and time (Cressie and Huang, 1999; Fuentes et al., 2008), but is separable if \( \lambda_j = 0 \) for \( j \geq 2 \). Moreover, if any \( \xi_j(s) \) is a nonstationary spatial process, \( \varepsilon(s, t) \) is nonstationary in both space and time.

Next, we approximate (2) by the first \( J \) components; that is, we assume

\[
y(s, t) = \mu(s, t) + \xi(s)^T \varphi(t) + v(s, t),
\]

where \( \xi(s) = (\xi_1(s), \ldots, \xi_J(s))^T \) and \( \varphi(t) = (\varphi_1(t), \ldots, \varphi_J(t))^T \) for \( s \in \mathcal{R} \) and \( t \in [0, T] \).

We further assume that the spatial covariance function of \( \xi_j(s) \) is \( \lambda_j \rho_j(\|s - s'\|; \theta_j) \), where \( \rho_j(\cdot; \theta_j) \) is a spatial correlation function in the Matérn family with a \( q_i \)-dimensional vector of correlation parameters \( \theta_j \) (Cressie, 1993). While there are multiple ways to model a spatial process, we have chosen the Matérn family because it is theoretically sound and is a popular choice in practice. That is, the spatio-temporal covariance structure specified in (4) is semiparametric.
3 Covariance Estimation and Theoretical Properties

3.1 A Two-step Estimation Procedure

We now turn to the estimation of the covariance function in (4) assuming that \( \mu(s, t) \) is known. We will relax this assumption and consider a general mean trend in Section 4.

Suppose data are observed at \( n \) sampling locations \( s_1, \ldots, s_n \), and at sampling location \( s_i \), \( y(s_i, t) \) is observed at \( m_i \) time points \( t_{i1}, \ldots, t_{im_i} \). Let \( y_{s_i} = (y(s_i, t_{i1}), \ldots, y(s_i, t_{im_i}))^T \) denote the observed data at sampling location \( s_i \), \( y = (y_{s_1}^T, \ldots, y_{s_n}^T)^T \) denote the observed data at all the sampling locations, \( \tilde{y}_{s_i} = (y(s_i, t_{i1}) - \mu(s_i, t_{i1}), \ldots, y(s_i, t_{im_i}) - \mu(s_i, t_{im_i}))^T \) denote the detrended data at sampling location \( s_i \), \( \tilde{y} = (\tilde{y}_{s_1}^T, \ldots, \tilde{y}_{s_n}^T)^T \) denote the detrended data at all the sampling locations, and \( N = \sum_{i=1}^n m_i \) denote the total number of observations. Further, let \( \Phi_i = (\varphi(t_{i1}), \ldots, \varphi(t_{im_i})) \) denote a \( J \times m_i \) matrix of eigenfunctions at sampling location \( s_i \), \( \Phi = \text{diag}\{\Phi_1, \ldots, \Phi_n\} \) denote a block diagonal matrix of eigenfunctions at all the sampling locations, \( \Lambda_{i,i'} = (\text{cov}\{\xi_j(s_i), \xi_{j'}(s_{i'})\})_{j,j'=1}^J = \text{diag}\{\lambda_1 \rho_1(\|s_i - s_{i'}\|; \theta_1), \ldots, \lambda_J \rho_J(\|s_i - s_{i'}\|; \theta_J)\} \) denote a \( J \times J \) diagonal matrix for the covariance between sampling locations \( s_i \) and \( s_{i'} \), and \( \Lambda = [\Lambda_{i,i'}]_{i,i'=1}^n \) denote an \( n \times n \) block matrix for covariances between all pairs of sampling locations. It follows that the variance-covariance matrix of \( y \) is

\[
\text{cov}(y) = \Sigma = \Phi^T \Lambda \Phi + \sigma^2 I_N,
\]

where recall from (1) that \( \sigma^2 \) is the measurement error variance, and \( I_N \) is the identity matrix with rank \( N \).
Thus, the negative log-likelihood function of the parameters in (4) is

\[ \ell(\lambda, \varphi(t), \theta_j, \sigma^2) = \frac{1}{2} \hat{y}^T \Sigma^{-1} \hat{y} + \frac{1}{2} \log \{ \det(\Sigma) \} + \frac{N}{2} \log(2\pi), \]

(5)

where \( \lambda = (\lambda_1, \ldots, \lambda_J)^T \) and recall from (4) that \( \varphi(t) = (\varphi_1(t), \ldots, \varphi_J(t))^T \). Maximizing (5) would give the maximum likelihood estimates of the eigenvalues \( \lambda \), eigenfunctions \( \varphi(t) \), \( q_i \)-dimensional correlation parameters \( \theta_j \), and measurement error variance \( \sigma^2 \). However, such computation is intensive if not infeasible, since there are \( J + \sum_{j=1}^J q_j + 1 \) unknown parameters and \( J \) unknown functions involved. To overcome this challenge, we develop a two-step procedure that is likelihood-based but computationally more feasible than the maximum likelihood estimation. The theoretical properties of the resulting estimates will be established in Section 3.2.

In Step I of the estimation procedure, with

\[ \hat{\gamma}_0(t, t') = n^{-1} \sum_{i=1}^n \hat{y}(s_i, t) \hat{y}(s_i, t') \]

where \( \hat{y}(s, t) = y(s, t) - \mu(s, t) \), we estimate \( \lambda \) and \( \varphi(t) \) as follows. First, let \( \tilde{\varphi}_1(t) \) be the maximizer of

\[ \max_{\|f(t)\|_\alpha = 1} \int_0^T \int_0^T f(t) \hat{\gamma}_0(t, t') f(t') dt dt' \]

(6)

where \( \|f\|_\alpha = (f, f)_\alpha^{1/2} \) and \( (f, g)_\alpha = \int_0^T f(t)g(t)dt + \alpha \int_0^T f''(t)g''(t)dt \) is an inner product, and \( \alpha > 0 \) controls the smoothness of the resulting maximizer. When \( \alpha = 0 \), we denote \( \|f\| = (f, f)^{1/2} \) and \( (f, g) = \int_0^T f(t)g(t)dt \). Consequently, the standardized \( \tilde{\varphi}_1(t) \), defined as \( \tilde{\varphi}_1(t) = \tilde{\varphi}_1(t)/\|\tilde{\varphi}_1(t)\| \), provides an estimate of \( \varphi_1(t) \). For \( j > 2 \), let \( \tilde{\varphi}_j(t) \) be the maximizer of (6) subject to the constraints \( (f, \tilde{\varphi}_k) = 0 \) for \( k < j \). Similarly, define
\( \tilde{\varphi}_j(t) = \varphi_j(t) / \| \varphi_j(t) \| \) to be the estimate of \( \varphi_j(t) \). Next, given \( \tilde{\varphi}_j(t) \), the estimate of \( \lambda_j \) is

\[
\hat{\lambda}_j = \int_0^T \int_0^T \tilde{\varphi}_j(t) \tilde{\gamma}_0(t, t') \tilde{\varphi}_j(t') dt dt', \quad j = 1, \ldots, J.
\]

The computation in Step I above can be carried out by an algorithm for smoothed functional principal component analysis via a basis expansion (Silverman, 1996). Although this algorithm is fast to compute and the consistency of the resulting estimates is well-established when data are drawn independently from a stochastic process, its applicability and the estimation properties for spatio-temporal data have not been adequately studied before. We will pursue this in Section 3.2.

In Step II of the estimation procedure, we estimate \( \theta_j \) and \( \sigma^2 \) by minimizing the negative log-likelihood function (5) given \( \varphi(t) = \hat{\varphi}(t) \) and \( \lambda = \hat{\lambda} \) as follows.

\[
\ell(\theta_j, \sigma^2 | \hat{\lambda}, \hat{\varphi}(t)) = \frac{1}{2} \hat{y}^T \Sigma^{-1} \hat{y} + \frac{1}{2} \log \{ \det(\Sigma) \} + \frac{N}{2} \log(2\pi).
\]

The resulting estimates are denoted as \( \hat{\theta}_j \) and \( \hat{\sigma}^2 \). The computational complexity is of order \( \mathcal{O}(N^3) \), where \( N \) is the total number of observations, due to the inversion and determinant calculation of an \( N \times N \) matrix \( \Sigma \) which is still intensive for large \( N \). Thus, we further improve the computational efficiency by Sherman-Morrison-Woodbury formula and Sylvester’s determinant theorem (Harville, 2008)

\[
\Sigma^{-1} = (\Phi^T \Lambda \Phi + \sigma^2 I_N)^{-1} = \sigma^{-2} I_N - \sigma^{-2} \Phi^T \{ \Phi \Phi^T + \sigma^2 \Lambda^{-1} \}^{-1} \Phi,
\]

\[
\det(\Sigma) = \det(\Phi^T \Lambda \Phi + \sigma^2 I_N) = \sigma^{2N} \det(\Lambda) \det(\Phi \Phi^T / \sigma^2 + \Lambda^{-1}),
\]

which reduce the computational complexity to a smaller order of \( \mathcal{O}(J^3 n^3) \), where \( n \) is the number of sampling locations. A similar approach has been taken by Nychka et al.
(2015), although our method is semiparametric for spatio-temporal processes, and theirs is nonparametric for spatial-only processes.

3.2 Theoretical Properties

We now investigate the asymptotic properties of the estimates obtained from the proposed two-step procedure in Section 3.1. Let $P \xrightarrow{}$ denote convergence in probability and $D \xrightarrow{}$ denote convergence in distribution. We consider the increasing domain asymptotics such that the distance between any two sampling sites is greater than a constant. We assume the following regularity conditions for Theorem 1. Let $R_{n^*}$ denote the spatial domain $R$ at the $n^*$-th stage of the asymptotics.

(A.1) The eigenvalues of $\gamma_0(t, t')$ satisfy $\lambda_1 > \lambda_2 > \cdots > 0$.

(A.2) The smoothness parameter satisfies $\alpha \rightarrow 0$ as $n \rightarrow \infty$.

(A.3) $\sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{j=1}^{\infty} \sum_{j'=1}^{\infty} E\{\xi_j(s_i)\xi_j(s_{i'})\} E\{\xi_{j'}(s_i)\xi_{j'}(s_{i'})\} = o(n^2)$ as $n \rightarrow \infty$.

(A.4) $\sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{j=1}^{\infty} [E\{\xi_j(s_i)\xi_j(s_{i'})\}]^2 = o(n^2)$ as $n \rightarrow \infty$.

Conditions (A.1) and (A.2) have been assumed to establish consistency of $\hat{\lambda}_j$ and $\hat{\varphi}_j(t)$ for functional data analysis (Silverman, 1996) and are assumed here for spatio-temporal data. For a spatio-temporal separable covariance function with $J = 1$, Condition (A.1) can be relaxed to $\lambda_1 > \lambda_2 > \cdots \lambda_M > 0$, $\lambda_{M+1} = \cdots = 0$, for some $M > 0$. Conditions (A.3) and (A.4) are about the covariance function of the spatial random process $\xi_j(s)$ and hold for commonly used spatial covariance functions, some of which are shown in the supplementary materials.
In the following Theorem 1, we establish the consistency for the estimated eigenvalues \(\hat{\lambda}_j\) and the estimated eigenfunctions \(\hat{\varphi}_j\), obtained from Step I of the two-step procedure given in Section 3.1.

**Theorem 1.** Under (2) and (A.1)–(A.4), for each \(j\), we have

\[
\hat{\lambda}_j \xrightarrow{P} \lambda_j, \quad (\hat{\varphi}_j, \varphi_j)^2 \xrightarrow{P} 1,
\]

as \(n \to \infty\), where \((\hat{\varphi}_j, \varphi_j) = \int_0^T \hat{\varphi}_j(t) \varphi_j(t) dt\).

Theorem 1 is under the assumption that the mean function is known, which will be relaxed in Theorem 3.

Next, we establish the asymptotic properties for the estimates of the spatial parameters \(\theta_j\) from Step II, given \(\lambda, \varphi_j(t)\), and \(\sigma^2\). For an \(n \times n\) matrix \(A\), let \(\mu_i(A)\) denote the \(i\)th largest eigenvalue of \(A\) and \(\|A\|_2 = \max_{i=1,\ldots,n}\{\mu_i(A^T A)\}^{1/2}\) denote the spectral norm of \(A\). Define

\[
\Lambda_j = \left[ \text{cov}\{\xi_j(s_i), \xi_j(s_{i'})\} \right]_{i,i'=1}^n = \left[ \lambda_j \rho_j(\|s_i - s_{i'}\|; \theta_j) \right]_{i,i'=1}^n,
\]

whose \((i, i')\)th component is the \(((i - 1)m + j, (i' - 1)m + j)\)th component of \(\Lambda\). With \(q = \sum_{j=1}^J q_j\), let \(\vartheta = (\theta_1^T, \ldots, \theta_J^T)^T\) denote a \(q\)-dimensional vector and define \(\vartheta_0 = (\theta_{10}^T, \ldots, \theta_{J0}^T)^T\), where \(\theta_{j0}\) denotes the true value of the correlation parameter in \(\rho_j(\cdot; \theta_j)\). Moreover, define \(D_k \Lambda_j = \partial \Lambda_j / \partial \theta_{j,k}, D_{k'k} \Lambda_j = \partial^2 \Lambda_j / \partial \theta_{j,k} \partial \theta_{j,k'}\), \(w_{j,k,k'} = \text{tr}\{(\Lambda_j + \sigma^2 I_n)^{-1}(D_k \Lambda_j)(\Lambda_j + \sigma^2 I_n)^{-1}(D_{k'} \Lambda_j)\}\), for \(k, k' = 1, \ldots, q_j\) and \(j = 1, \ldots, J\). Finally, with \(\vartheta_l\) denoting the \(l\)th component of \(\vartheta\), define \(D_l \Sigma = \partial \Sigma / \partial \vartheta_l, D_{ll'} \Sigma = \partial^2 \Sigma / \partial \vartheta_l \partial \vartheta_{l'}, D_l \Lambda = \partial \Lambda / \partial \vartheta_l, D_{ll'} \Lambda = \partial^2 \Lambda / \partial \vartheta_l \partial \vartheta_{l'}, \) and \(t_{ll'}^* = \text{tr}\{\Sigma^{-1}(D_l \Sigma)\Sigma^{-1}(D_{l'} \Sigma)\}\), for \(l, l' = 1, \ldots, q\).

The regularity conditions for Theorem 2 are as follows.
The correlation function \( \rho_j(\cdot, \cdot; \theta_j) \) is twice differentiable, with respect to \( \theta_j \), with continuous second-order derivatives for \( \theta_j \in \Omega_j \), where \( \Omega_j \) is an open set.

As \( n \to \infty \), \( \| A_j \|_2 = \mathcal{O}(1) \), \( \| D_k A_j \|_2 = \mathcal{O}(1) \), \( \| D_{kk'} A_j \|_2 = \mathcal{O}(1) \) for \( k, k' = 1, \ldots, q_j \) and \( j = 1, \ldots, J \).

For some \( \delta > 0 \), there exist positive constants \( C_k \) such that \( \| D_k A_j \|_F^2 \leq C_k n^{-1/2-\delta} \) for \( k = 1, \ldots, q_j \) and \( j = 1, \ldots, J \).

For any \( k, k' = 1, \ldots, q_j \), \( a_{j, kk'} = \lim_{n \to \infty} \{ w_{j, kk'} w_{j, kk'}^{-1/2} \} \) exists and \( A_j = [a_{j, kk'}]_{k, k' = 1}^q \) is nonsingular, for \( j = 1, \ldots, J \).

There exists a positive constant \( c_0 \), such that \( \| A_j^{-1} \|_2 < c_0 < \infty \), for \( j = 1, \ldots, J \).

As \( n \to \infty \), \( m_i = \mathcal{O}(1) \) and \( \Phi_i^T \Phi_i = m_i I_{m_i} \).

Conditions (A.5)–(A.9) are standard assumptions made about Gaussian random fields in spatial linear models to ensure smoothness, growth and convergence of the information matrix (Mardia and Marshall, 1984). For spatio-temporal data, we assume (A.5)–(A.9) for \( \xi_j(s) \), a spatial Gaussian process, in the Karhunen-Loève type expansion. Condition (A.10) is based on the orthonormality of eigenfunctions \( \varphi_j(t) \), that is, \( \int_0^T \varphi_j(t) \varphi_j'(t) \, dt = 1 \) if \( j = j' \) and 0 otherwise. Together with (A.5)–(A.9), (A.10) ensures smoothness, growth and convergence of the information matrix for spatio-temporal data (Sweeting, 1980).

Let \( \ell''(\vartheta, \vartheta) = \frac{\partial^2 \ell(\lambda, \varphi(t), \theta_j, \sigma^2)}{\partial \vartheta \partial \vartheta^T} \) be the second-order derivatives of \( \ell(\lambda, \varphi(t), \theta_j, \sigma^2) \) with respect to \( \vartheta \). Under (A.5)–(A.10), the asymptotic normality of \( \hat{\theta}_j \) is established in the following theorem.
Theorem 2. Under (4) and (A.5)–(A.10), we have
\[ H(\vartheta_0)^{1/2}(\hat{\vartheta} - \vartheta_0) \xrightarrow{D} N(0, I_q), \]
as \( n \to \infty \), where \( q = \sum_{j=1}^{J} q_j \), \( \hat{\vartheta} = (\hat{\theta}_1^T, \ldots, \hat{\theta}_J^T)^T \), and \( H(\vartheta_0) = E\{-\ell''(\vartheta_0, \vartheta_0)\} \) is the information matrix for \( \vartheta \).

Although Theorems 1 and 2 give consistency and asymptotic normality of the parameter estimates, they are established for detrended spatio-temporal data — an assumption we will relax in Section 4.

4 Extensions to Spatio-Temporal Data with Trend

4.1 A Modified Two-step Estimation Procedure

In geostatistics, the mean function tends to vary over space and thus \( \mu(s, t) = E\{y(s, t)\} \) depends on location \( s \). In addition, \( \mu(s, t) \) is generally unknown and needs to be estimated.

There are various methods to estimate \( \mu(s, t) \) such as kernel smoothing which often yields consistent estimates (Altman, 1990). Here, we let \( \bar{\mu}(s, t) \) denote an estimated mean trend and \( \bar{\mu}(s, t) = y(s, t) - \bar{\mu}(s, t) \) denote the detrended process. Further, let \( \bar{y}_{si} = (y(s_i, t_{i1}) - \bar{\mu}(s_i, t_{i1}), \ldots, y(s_i, t_{i_m}) - \bar{\mu}(s_i, t_{i_m}))^T \) denote the detrended data, and \( \bar{y} = (\bar{y}_{s_1}^T, \ldots, \bar{y}_{s_n}^T)^T \) denote the detrended data at all the sampling locations. The two-step estimation procedure developed in Section 3.1 for detrended data can be applied here by replacing \( \bar{y}(s, t), \bar{y}_{si} \) and \( \bar{y} \) with \( y(s, t), \bar{y}_{si} \) and \( \bar{y} \), respectively.

Specifically, in Step I of the modified two-step procedure, let \( \gamma_0(t, t') = n^{-1} \sum_{i=1}^{n} \bar{y}(s_i, t)\bar{y}(s_i, t') \).
For \( j = 1 \), let \( \tilde{\varphi}_1(t) \) be the maximizer of
\[
\max_{\|f(t)\|_a = 1} \int_0^T \int_0^T f(t)\gamma_0(t, t')f(t') dt dt'
\] (7)

For \( j > 2 \), let \( \tilde{\varphi}_j(t) \) be the maximizer of (7) subject to the constraints \((f, \tilde{\varphi}_k) = 0\) for \( k < j \). After standardization, \( \hat{\varphi}_j(t) = \tilde{\varphi}_j(t)/\|\tilde{\varphi}_j(t)\| \) becomes the estimate of \( \varphi_j(t) \). The estimate of \( \lambda_j \) is
\[
\hat{\lambda}_j = \int_0^T \int_0^T \tilde{\varphi}_j(t)\gamma_0(t, t')\tilde{\varphi}_j(t') dt dt', \quad j = 1, \ldots, J.
\]

In Step II of the modified two-step procedure, the negative log-likelihood function (5) with a plug-in estimated mean function takes on the form
\[
\ell(\lambda, \varphi(t), \theta_j, \sigma^2) = \frac{1}{2} y^T \Sigma^{-1} y + \frac{1}{2} \log \{ \det(\Sigma) \} + \frac{N}{2} \log(2\pi),
\]
and is minimized with \( \varphi(t) = \hat{\varphi}(t) \) and \( \lambda = \hat{\lambda} \) held fixed. The resulting estimates are denoted as \( \hat{\theta}_j \) and \( \hat{\sigma}^2 \).

Next, we consider a linear regression for the mean function, \( \mu(s, t) = x(s, t)^T \beta \), where
\[
x(s, t) = (x_1(s, t), \ldots, x_p(s, t))^T
\]
denotes \( p \) covariate functions at location \( s \) and time \( t \), \( \beta = (\beta_1, \ldots, \beta_p)^T \) denotes a \( p \)-dimensional vector of regression coefficients. The model (4) becomes
\[
y(s, t) = x(s, t)^T \beta + \sum_{j=1}^J \xi_j(s)\varphi_j(t) + \nu(s, t).
\] (8)

Let \( X(s_i) = (x(s_i, t_{i1}), \ldots, x(s_i, t_{im}))^T \) denote an \( m_i \times p \) design matrix at sampling location \( s_i \) and \( X = (X(s_1)^T, \ldots, X(s_n)^T)^T \) denote an \( N \times p \) design matrix. Thus, the negative log-likelihood function of the parameters in model (4) is
\[
\ell(\lambda, \varphi(t), \beta, \theta_j, \sigma^2) = (y - X\beta)^T \Sigma^{-1} (y - X\beta) / 2 + (1/2) \log \{ \det(\Sigma) \} + (N/2) \log(2\pi).
\] (9)
A practical choice of $\bar{\mu}(s, t)$ is $x(s, t)^T \bar{\beta}_{\text{ols}}$, where $\bar{\beta}_{\text{ols}} = (X^T X)^{-1} X^T y$ is the least squares estimate of $\beta$. Then, Step I of the modified two-step estimation procedure can be carried out as before. In Step II, however, we minimize the negative log-likelihood function (9) with respect to $\beta$, $\theta_j$, and $\sigma^2$, with $\varphi(t) = \hat{\varphi}(t)$ and $\lambda = \hat{\lambda}$ held fixed. The resulting estimates are denoted as $\hat{\beta}$, $\hat{\theta}_j$, and $\hat{\sigma}^2$.

4.2 Theoretical Properties

Now, we consider the asymptotic properties of the parameter estimates obtained from the modified two-step procedure above under additional regularity conditions. First, we assume the following about the fourth moment of the estimated mean function.

(A.11) As $n \to \infty$, there exists a sequence $c_n \to 0$, such that $E\{\bar{\mu}(s, t) - \mu(s, t)\}^4 \leq c_n$ for $t \in [0, T]$, where $c_n$ does not depend on $s$.

For kernel smoothing estimate $\hat{\mu}(s, t)$, (A.11) can be verified to hold under certain conditions (El Machkouri, 2007).

The following Theorem 3 establishes the consistency for the parameter estimates in Step I of the modified estimation procedure, in the case of an unknown mean function $\mu(s, t)$.

**Theorem 3.** Under the assumptions of Theorem 1 and (A.11), we have

$$\hat{\lambda}_j \xrightarrow{P} \lambda_j, \quad (\hat{\varphi}_j, \varphi_j)^2 \xrightarrow{P} 1,$$

as $n \to \infty$.

For model (8), we establish the asymptotic properties of $\hat{\beta}$ and $\hat{\theta}_j$ from Step II, given $\lambda$, $\varphi_j(t)$, and $\sigma^2$. An additional regularity condition is assumed about the design matrix, which is standard for spatial linear models (Mardia and Marshall, 1984).
The design matrix $X$ has full rank $p$ and is uniformly bounded in max norm with
$$\lim_{n\to\infty}(X^TX)^{-1} = 0.$$ 

Let $\ell''(\beta, \beta) = \frac{\partial^2\ell(\lambda, \varphi(t), \beta, \theta, \sigma^2)}{\partial\beta\partial\beta^T}$ be the second-order derivatives of $\ell(\lambda, \varphi(t), \beta, \theta, \sigma^2)$ with respect to $\beta$. The asymptotic normality of $\hat{\beta}$ and $\hat{\theta}_j$ are established in Theorem 4.

**Theorem 4.** Under the assumptions of Theorem 2 and (A.12), we have
$$\begin{pmatrix} H(\beta_0)^{1/2} & O_{p\times q} \\ O_{q\times p} & H(\vartheta_0)^{1/2} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\vartheta} \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \vartheta_0 \end{pmatrix} \xrightarrow{D} N(0, I_{p+q}),$$

as $n \to \infty$, where $H(\beta_0) = E\{-\ell''(\beta_0, \beta_0)\}$ is the information matrix for $\beta$, and $O_{p\times q}$ denotes a $p \times q$ zero matrix.

The proof of the proposition and theorems above are given as supplementary materials.

The Gaussian assumption can be readily relaxed in Theorems 1 and 3, but it does not appear to be the case for Theorems 2 and 4.

## 5 Numerical Examples

### 5.1 Simulation Study

A simulation study is conducted to investigate the finite-sample properties of our spatio-temporal semiparametric covariance (SemiCov) method developed in Sections 2–4. First, the covariates $x(s, t)$ are generated from the standard normal distributions with a cross-covariate correlation of 0.5, and the regression coefficients are set to be $\beta = (4, 3, 2, 1, 0, 0, 0)^T$. Each covariate is standardized to have sample mean 0 and sample variance 1 and the response has a sample mean 0. Thus, there is no intercept in this model. The spatio-temporal
Figure 1: The 95% pointwise simulation intervals for $\varphi_1(t)$ (left) and $\varphi_2(t)$ (right) using our method. The true $\varphi_1(t)$ and $\varphi_2(t)$ are drawn in gray solid line. The pointwise simulation intervals for $n=50$, $100$ and $150$ are drawn by black dotted, dashed and solid lines, respectively.

The process is defined as $\varepsilon_1(s, t) = \xi_1(s)\varphi_1(t) + \xi_2(s)\varphi_2(t)$, where $\xi_1(s)$ and $\xi_2(s)$ are two independent zero-mean stationary and isotropic Gaussian processes with an exponential covariance function $\lambda_1 \exp(-d/r_1)$ and $\lambda_2 \exp(-d/r_2)$ for spatial distance $d$, respectively, with $\lambda_1 = 2.5$, $r_1 = 0.5$, $\lambda_2 = 0.5$ and $r_2 = 0.3$. Moreover, $\varphi_1(t) = c_1 \cos(\pi t)$ and $\varphi_2(t) = c_2 \sin(\pi t)$ are two orthonormal functions on $[0, 1]$ with normalization constants $c_1$ and $c_2$. The number of sampling locations is set to be $n=50$, $100$ and $150$ and the locations are randomly distributed within the spatial domain $\mathcal{R} = [0, l] \times [0, l]$, where $l = 2^{-1}n^{1/2}$. At each sampling location, twenty time points are set at $t_i = (2i-1)/(2m)$ for $i = 1, \ldots, m$ and $m = 20$. For each sample size $n$, $100$ data sets are simulated.

For each simulated data set, our method is applied to estimate the regression coefficients $\beta$, the spatial parameters $(r_1, r_2)$, the eigenvalues $(\lambda_1, \lambda_2)$, the eigenfunctions $(\varphi_1(t), \varphi_2(t))$, and the measurement error variance $\sigma^2$. Our method is compared with two alternative methods, namely, the ordinary least squares which ignores both spatial and temporal dependence (denoted as ALT$_1$) and the functional data analysis that accounts for temporal
but not spatial dependence (denoted as ALT$_2$). Moreover, the prediction in space and time is performed using all these three approaches.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Method</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>MSPE$_1$</th>
<th>MSPE$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>True Values</td>
<td>4.000</td>
<td>3.000</td>
<td>2.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>SemiCov</td>
<td>4.001</td>
<td>3.000</td>
<td>1.997</td>
<td>0.996</td>
<td>0.006</td>
<td>-0.006</td>
<td>0.009</td>
<td>2.667</td>
<td>1.112</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.043</td>
<td>0.044</td>
<td>0.045</td>
<td>0.048</td>
<td>0.042</td>
<td>0.043</td>
<td>0.041</td>
<td>1.006</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>SDm</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>ALT$_1$</td>
<td>4.005</td>
<td>3.005</td>
<td>1.996</td>
<td>0.992</td>
<td>0.006</td>
<td>-0.005</td>
<td>0.009</td>
<td>3.927</td>
<td>3.768</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.031</td>
<td>0.030</td>
<td>0.029</td>
<td>0.031</td>
<td>0.033</td>
<td>0.027</td>
<td>0.030</td>
<td>0.715</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>SDm</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>100</td>
<td>SemiCov</td>
<td>4.001</td>
<td>2.992</td>
<td>1.998</td>
<td>1.007</td>
<td>0.002</td>
<td>0.000</td>
<td>-0.001</td>
<td>2.610</td>
<td>1.106</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.031</td>
<td>0.030</td>
<td>0.029</td>
<td>0.031</td>
<td>0.033</td>
<td>0.027</td>
<td>0.030</td>
<td>0.715</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>SDm</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>0.031</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>ALT$_1$</td>
<td>3.996</td>
<td>2.991</td>
<td>2.001</td>
<td>1.011</td>
<td>0.005</td>
<td>-0.005</td>
<td>0.004</td>
<td>3.720</td>
<td>3.775</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.053</td>
<td>0.055</td>
<td>0.054</td>
<td>0.058</td>
<td>0.056</td>
<td>0.053</td>
<td>0.057</td>
<td>1.130</td>
<td>0.453</td>
</tr>
<tr>
<td></td>
<td>SDm</td>
<td>0.031</td>
<td>0.030</td>
<td>0.029</td>
<td>0.032</td>
<td>0.033</td>
<td>0.027</td>
<td>0.030</td>
<td>1.133</td>
<td>0.108</td>
</tr>
<tr>
<td>150</td>
<td>SemiCov</td>
<td>4.001</td>
<td>3.004</td>
<td>1.997</td>
<td>0.997</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.002</td>
<td>2.616</td>
<td>1.093</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.027</td>
<td>0.026</td>
<td>0.025</td>
<td>0.027</td>
<td>0.032</td>
<td>0.024</td>
<td>0.027</td>
<td>0.600</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>SDm</td>
<td>0.025</td>
<td>0.025</td>
<td>0.025</td>
<td>0.025</td>
<td>0.025</td>
<td>0.025</td>
<td>0.026</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>ALT$_1$</td>
<td>4.005</td>
<td>3.007</td>
<td>1.991</td>
<td>1.001</td>
<td>-0.001</td>
<td>-0.007</td>
<td>0.005</td>
<td>3.836</td>
<td>3.847</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.043</td>
<td>0.054</td>
<td>0.051</td>
<td>0.048</td>
<td>0.049</td>
<td>0.045</td>
<td>0.045</td>
<td>0.914</td>
<td>0.414</td>
</tr>
<tr>
<td></td>
<td>SDm</td>
<td>0.027</td>
<td>0.026</td>
<td>0.025</td>
<td>0.027</td>
<td>0.032</td>
<td>0.023</td>
<td>0.027</td>
<td>0.912</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>ALT$_2$</td>
<td>4.001</td>
<td>3.004</td>
<td>1.997</td>
<td>0.997</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>3.830</td>
<td>1.095</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.027</td>
<td>0.026</td>
<td>0.025</td>
<td>0.027</td>
<td>0.032</td>
<td>0.023</td>
<td>0.027</td>
<td>0.912</td>
<td>0.105</td>
</tr>
</tbody>
</table>

Table 1: The mean, standard deviation (SD) of regression coefficient estimates and mean square prediction errors under the proposed SemiCov method, ALT$_1$ and ALT$_2$, as well as the mean estimated standard deviation (SDm) under the SemiCov method for sample size $n = 50$, 100 and 150.

Figure 1 gives the 95% pointwise simulation intervals for the eigenfunctions $\varphi_j(t)$, defined as

$$\left[\frac{1}{2}\left\{\hat{\varphi}_j^{(97)}(t) + \hat{\varphi}_j^{(98)}(t)\right\}, \frac{1}{2}\left\{\hat{\varphi}_j^{(2)}(t) + \hat{\varphi}_j^{(3)}(t)\right\}\right],$$

where $\hat{\varphi}_j^{(i)}(t)$ is the $i$th largest value of $\{\hat{\varphi}_j(t) : i = 1, \ldots, 100\}$, and $\hat{\varphi}_j(t)$ is the estimate.
for \( \varphi_j(t) \) from the \( j \)th simulated data set. The results show that the true eigenfunctions are captured by the 95% pointwise simulation intervals. Moreover, the pointwise simulation intervals become narrower as the number of sampling locations \( n \) increases, which supports the theory that the estimates of \( \varphi_1(t) \) and \( \varphi_2(t) \) are consistent.

Table 1 reports the mean and standard deviation of regression coefficient estimates from 100 simulated data sets using the three approaches. The regression coefficient estimates have lower bias and the standard deviations become smaller as the number of sampling locations \( n \) increases. Moreover, both our SemiCov method and the functional data analysis outperform ordinary least squares in terms of smaller standard deviations. This suggests that incorporating spatio-temporal structures can greatly improve the estimation of regression coefficients.

The standard errors of the regression coefficients, eigenvalues, and spatial parameter estimates can be obtained using the information matrix in Theorem 4. That is, for each simulated data set, define \( \text{sd}\{\hat{\beta}\} = \text{diag}\{H(\hat{\beta})^{-1}\}^{1/2} \), \( \text{sd}\{\hat{\lambda}\} = \text{diag}\{H(\hat{\lambda})^{-1}\}^{1/2} \), and \( \text{sd}\{\hat{\vartheta}\} = \text{diag}\{H(\hat{\vartheta})^{-1}\}^{1/2} \), where \( \hat{\beta} \), \( \hat{\lambda} \) and \( \hat{\vartheta} \) are the estimates from the simulated data set, \( H(\beta) \), \( H(\lambda) = E\{-\ell''(\lambda, \lambda)\} \) and \( H(\vartheta) \) are the information matrix for \( \beta \), \( \lambda \) and \( \vartheta \), respectively. From the 100 simulated data sets, the mean of the standard errors (SDm) is computed for our SemiCov method and presented in Table 1. The results show that these means are close to nominal true standard deviations of regression coefficient estimates.

For prediction, we consider two scenarios. In Scenario 1, prediction is carried out for multiple time points \( t_{01}, \ldots, t_{0m} \) at an unsampled location \( s_0 \). It is straightforward to show
that the best linear unbiased prediction (BLUP) of $y(s_0, t_0)$ is

$$\tilde{y}(s_0, t_0) = x(s_0, t_0)^T \tilde{\beta} + c_0^T \Sigma^{-1}(y - X \tilde{\beta}),$$

where $c_0$ is an $N$-dimensional vector whose $i$th element is $\text{cov}\{y(s_0, t_0), y_i\}$, $y_i$ is the $i$th element of $y$, $\Sigma = \text{cov}(y)$, and $\tilde{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y$ (Cressie, 1993). Since $\Sigma$ and $c_0$ are usually unknown, the estimates of $\Sigma$ and $c_0$ are plugged in and an empirical best linear unbiased prediction is obtained.

To quantify the prediction error for the curve at locations $s_0$, a mean integral squared error,

$$\text{MISE}(s_0) = \int_0^T \{\tilde{y}(s_0, t) - y(s_0, t)\}^2 dt,$$

is used. In the simulation study, $T = 1$ and $t_i$ are evenly distributed and thus, $\text{MISE}(s_0)$ is estimated by $m^{-1} \sum_{i=1}^m \{\tilde{y}(s_0, t_{0i}) - y(s_0, t_{0i})\}^2$. Moreover, we generate 5, 10, and 15 additional curves for sample size $n = 50, 100, \text{and} 150$, respectively, and set them aside for prediction. In the end, for $M$ unsampled curves at $s_{01}, \ldots, s_{0M}$, we define the first mean squared prediction error ($\text{MSPE}_1$) as $M^{-1} \sum_{i=1}^M \text{MISE}(s_{0i})$.

![Scenario 1 and Scenario 2 graphs]

Figure 2: The mean squared prediction error (MSPE) for Scenario 1 (left) and Scenario 2 (right) under our method (SemiCov) and two alternative approaches (ALT$_1$ and ALT$_2$).

In Scenario 2, prediction is made about time points that are missing at a sampling loca-
Table 2: The mean, standard deviation (SD) of spatial-temporal coefficients estimates under our method (SemiCov) for sample size \( n = 50, 100 \) and 150.

<table>
<thead>
<tr>
<th>( n )</th>
<th>Method</th>
<th>( \lambda_1 )</th>
<th>( r_1 )</th>
<th>( \lambda_2 )</th>
<th>( r_2 )</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>True Values</td>
<td>2.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.30</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>SemiCov</td>
<td>2.32</td>
<td>0.46</td>
<td>0.48</td>
<td>0.29</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.63</td>
<td>0.20</td>
<td>0.10</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>SDm</td>
<td>0.65</td>
<td>0.18</td>
<td>0.12</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>100</td>
<td>SemiCov</td>
<td>2.31</td>
<td>0.47</td>
<td>0.52</td>
<td>0.32</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.45</td>
<td>0.14</td>
<td>0.09</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>SDm</td>
<td>0.46</td>
<td>0.13</td>
<td>0.10</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>150</td>
<td>SemiCov</td>
<td>2.43</td>
<td>0.48</td>
<td>0.53</td>
<td>0.31</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.36</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>SDm</td>
<td>0.41</td>
<td>0.11</td>
<td>0.08</td>
<td>0.07</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The prediction performance for different sample sizes are given in Table 1 and the box plots of them for \( n = 100 \) are given in Figure 2. The results in Scenario 1 show that ALT\(_2\) performs better than ALT\(_1\), while our SemiCov method outperforms both ALT\(_1\) and ALT\(_2\), which provides empirical evidence that incorporating spatial correlation between different locations can substantially improve prediction at unsampled locations. In Scenario 2, our SemiCov method and ALT\(_2\) both outperform ALT\(_1\), and our SemiCov method is slightly better than ALT\(_2\) for prediction at sampled locations with missing time points.

Finally, Table 2 reports the mean, the standard deviation (SD), the mean standard error (SDm) of the estimates for spatio-temporal parameters. The means of the spatio-temporal parameter estimates approach the true values and the standard deviations become smaller.
as the sample size increases. Moreover, the mean standard error of the spatio-temporal parameter estimates is fairly close to the nominal true standard deviation. For our SemiCov method, the simulated data sets are fitted with the number of eigenfunctions $J = 2$.

5.2 Data Example

Figure 3: Map of locations of 259 weather stations in the Colorado precipitation data (left), and the empirical variogram over spatial and temporal lags (right).

The data consists of precipitation (in inches per 24-hour period) from January to December on the log-scale from 259 weather stations in Colorado (Reich and Davis, 2008; Chu et al., 2011), as shown in the left-hand panel of Figure 3. There are ten covariates of interest, including elevation, slope, aspect, and seven spectral bands from a satellite imagery (B1M through B7M). For model fitting, precipitation of 10 months (excluding precipitation
Table 3: Precipitation data example without forward selection: Regression coefficient estimates and mean squared prediction errors under our method (SemiCov) and two alternative approaches (ALT\(_1\) and ALT\(_2\)), along with the standard errors (SE) for our SemiCov method.

of March and October) at 240 weather stations are used. Two types of prediction are considered. First, prediction is made at the remainder 19 weather stations and the prediction results are summarized by MSPE\(_1\). Second, the March and October precipitation for the 240 weather stations are predicted and the results are summarized by MSPE\(_2\).

In the right-hand panel of Figure 3, the empirical variogram over spatial and temporal lags is presented, and suggests that there is a spatio-temporal dependence for Colorado precipitation data. Data analysis is performed by our SemiCov method and two alternative approaches. For our SemiCov method, we choose the number of components \( J = 2 \) such that \( \sum_{j=1}^J \hat{\lambda}_j / \sum_{j=1}^n \hat{\lambda}_j \geq 80\% \), as suggested by Zhu et al. (2014). Furthermore, since there is multicollinearity among the covariates, a forward selection through AIC is applied. The resulting model contains two covariates elevation and B4M. The results without forward selection are reported in Table 3, while the results with forward selection are reported in Table 4. With SemiCov method, there is strong evidence of an elevation and B4M effect, while there is moderate evidence of an effect of B1M. For prediction of all time points at
<table>
<thead>
<tr>
<th>Method</th>
<th>Elevation</th>
<th>B4M</th>
<th>MSPE₁</th>
<th>MSPE₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>SemiCov</td>
<td>0.203</td>
<td>-0.059</td>
<td>0.140</td>
<td>0.085</td>
</tr>
<tr>
<td>SE</td>
<td>0.018</td>
<td>0.013</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>ALT₁</td>
<td>0.054</td>
<td>-0.173</td>
<td>0.469</td>
<td>0.140</td>
</tr>
<tr>
<td>ALT₂</td>
<td>0.101</td>
<td>-0.102</td>
<td>0.464</td>
<td>0.077</td>
</tr>
</tbody>
</table>

Table 4: Precipitation data example with forward selection: Regression coefficient estimates and mean squared prediction errors under our method (SemiCov) and two alternative approaches (ALT₁ and ALT₂), along with the standard errors (SE) for our SemiCov method.

unsampled locations, our SemiCov method outperforms the two alternative methods. On the other hand, for prediction at the two time points set aside at sampling locations, both our method and the functional data analysis outperform the ordinary least squares method, while the functional data analysis is slightly better than our method in this case.

References


ing data with jump discontinuities. Journal of the American Statistical Association,
109:1084–1098.