Statistica Sinica **29** (2019), 1233-1252
doi:https://doi.org/10.5705/ss.202016.0297

SEMIPARAMETRIC MODELING WITH NONSEPARABLE AND NONSTATIONARY SPATIO-TEMPORAL COVARIANCE FUNCTIONS AND ITS INFERENCE

Tingjin Chu, Jun Zhu and Haonan Wang

University of Melbourne, University of Wisconsin-Madison and Colorado State University

Abstract: In this study, we develop a new semiparametric approach to model geostatistical data measured repeatedly over time. In addition, we draw inferences 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 the likelihood of detrended data, and the computational algorithm is efficient owing to the dimension reduction. Extensions to spatio-temporal processes with general mean trends are also considered. Furthermore, the asymptotic properties of our proposed method are established, including consistency and asymptotic normality. A simulation study shows the sound finite-sample properties of the proposed method, and a real-data example is used to compare our method with alternative approaches.

Key words and phrases: Geostatistics, semiparametric methods, spatio-temporal processes.

1. Introduction

As 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 analyze the various types of spatio-temporal data, several statistical models and methods have been developed, including varying coefficients models (Lu et al. (2009)), hierarchical dynamic spatial models (Zhang, Yao and Tong (2003); Johannesson, Cressie and Huang (2007); Ghosh et al. (2010)), and filtering and dimensional reduction (Huang and Cressie (1996); Cressie, Shi and Kang (2010); Brynjarsdóttir and Berliner (2014)). For a further discussion on spatio-temporal statistics, see Cressie and Wikle (2011). In this study, we focus on spatio-temporal data in which the individual units are spatial sampling

locations, and at a given sampling location, measures are repeated over time. We propose a class of semiparametric models with nonseparable and nonstationary spatio-temporal covariance functions. We develop a new methodology for inferences that balance model flexibility and computational feasibility. We also establish the corresponding asymptotic properties.

When modeling spatio-temporal data, it is generally important to incorporate the spatio-temporal covariance. Separablility, such that the spatio-temporal covariance function is assumed to be a product of the spatial covariance and the 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 using a Fourier inversion and completely monotone functions, respectively. Although the above approaches allow for spatio-temporal nonseparability, the spatio-temporal processes are assumed to be stationary. To relax the stationarity assumption, Stein (2005) considered asymmetric models, that is, where the covariance function is spatially isotropic, but not symmetric spatio-temporally. Fuentes, Chen and Davis (2008) and Rodrigues and Diggle (2010) developed nonstationary and nonseparable models using a spectral representation and convolution, respectively. Although the above methods focus on building covariance models for spatio-temporal data, it is not always clear how model estimations and statistical inferences are to be carried out. There is a clear need for an additional statistical methodology for the analyzing of spatio-temporal data taken at regular or irregular sampling locations.

Nonparametric approaches are increasingly used for spatio-temporal modeling, and tend to be robust against covariance function misspecification. For example, extending the work of Gneiting (2002), Choi, Li and Wang (2013) proposed a nonparametric approximation of completely monotone functions in the construction of spatial and spatio-temporal covariance structures. Nonparametric methods may also alleviate the computational burden of estimating spatio-temporal covariance functions. Based on an originally parametric covariance function, Zhang, Sang and Huang (2015) proposed a nonparametric full-scale approximation, that 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 exploring 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 rare in semiparametric spatio-temporal statistics.

In particular, we model dependence over space and time using a Karhunen–Loève-type expansion that results in nonseparable and nonstationary spatio-temporal covariance functions. The model parameters are estimated using a two-step procedure based on likelihood, and the computational feasibility is enhanced further using dimension reduction. Extensions to spatio-temporal models with a general mean function (or trend) are also considered. Furthermore, 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, the first of their kind for semiparametric methods for inferences about spatio-temporal models with nonseparable and nonstationary covariance functions. A simulation study shows the sound finite-sample properties of the estimates, and a real-data example compares our method with several the existing approaches. Our model may also be applied to 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 their attention to separable models and the consistency of the sample means and empirical covariance operators, respectively (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 in the Supplementary Material.

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 errors, we model the spatio-

temporal response variable $y(s,t)$ by

$$y(s,t) = \mu(s,t) + \varepsilon(s,t) + \upsilon(s,t), \tag{2.1}$$

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 nonconvexity, but the domain needs to be continuous. In this and the following sections, we assume a zero-mean function $\mu(s,t) = 0$. We consider more general mean functions in Section 4. We further assume that the measurement error $\upsilon(s,t)$ follows an independent and identically distributed Gaussian distribution with mean zero and variance σ^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 $\{\xi_j(s) : s \in \mathcal{R}\}_{j=1}^{\infty}$ is assumed to be a sequence of independent Gaussian processes, and $\{\varphi_j(t)\}_{j=1}^{\infty}$ is a sequence of eigenfunctions. Thus, (2.1) can be rewritten as

$$y(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \xi_j(s)\varphi_j(t) + \upsilon(s,t). \tag{2.2}$$

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

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

We further assume that, for two locations $s \neq s'$, the 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); Hörmann and Kokoszka (2013)). Sufficient conditions to establish (2.2) are given in Appendix A of the Supplementary Material. Let $\lambda_j = \operatorname{var}\{\xi_j(s)\}$ denote the jth eigenvalue of the covariance function $\gamma_0(t,t')$. We further assume $\operatorname{cov}\{\xi_j(s), \xi_{j'}(s')\} = 0$, for $j \neq j'$, which ensures the 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

$$\gamma(t,s;t',s') = \sum_{j=1}^{\infty} \operatorname{cov}\{\xi_j(s), \xi_j(s')\}\varphi_j(t)\varphi_j(t'). \tag{2.3}$$

In spatio-temporal statistics, it is common practice to assume that the spatio-temporal process $\varepsilon(s,t)$ in model (2.1) is stationary over space and time. In contrast, the $\varepsilon(s,t)$ formulated here using the KL expansion encompasses spatio-temporal covariance functions that are nonstationary. Specifically, from (2.3), 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)). In addition, the spatio-temporal covariance function in (2.3) does not need to be separable in space and time (Cressie and Huang (1999); Fuentes, Chen and Davis (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 (2.2) by the first J components; that is, we assume

$$y(s,t) = \mu(s,t) + \xi(s)^T\varphi(t) + \upsilon(s,t), \tag{2.4}$$

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 θ_j (Cressie (1993)). Although 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 (2.4) 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 (2.4), assuming that $\mu(s,t)$ is known. We 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, $\breve{y}_{s_i} = (y(s_i,t_{i1}) - \mu(s_i,t_{i1}), \ldots, y(s_i,t_{im_i}) - \mu(s_i,t_{im_i}))^T$ denote the detrended data at sampling location s_i, $\breve{y} = (\breve{y}_{s_1}^T, \ldots, \breve{y}_{s_n}^T)^T$ denote the detrended data at all the sampling locations, and $N = \sum_{i=1}^n m_i$ denote the total number of observations. Furthermore, 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, $\Phi = \mathrm{diag}\{\Phi_1, \ldots, \Phi_n\}$ denote a block diagonal matrix of eigenfunctions at all the sampling locations, $\Lambda_{i,i'} =$

$(\mathrm{cov}\{\xi_j(s_i), \xi_{j'}(s_{i'})\})_{j,j'=1}^{J} = \mathrm{diag}\{\lambda_1\rho_1(\|s_i - s_{i'}\|;\theta_1), \ldots, \lambda_J\rho_J(\|s_i - s_{i'}\|;\theta_J)\}$ 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 the covariances between all pairs of sampling locations. It follows that the variance-covariance matrix of y is

$$\mathrm{cov}(y) = \Sigma = \Phi^T\Lambda\Phi + \sigma^2 I_N,$$

where, from (2.1), σ^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 (2.4) is

$$\ell(\lambda, \varphi(t), \theta_j, \sigma^2) = \frac{1}{2}\breve{y}^T\Sigma^{-1}\breve{y} + \frac{1}{2}\log\{\det(\Sigma)\} + \frac{N}{2}\log(2\pi), \tag{3.1}$$

where $\lambda = (\lambda_1, \ldots, \lambda_J)^T$ and $\varphi(t) = (\varphi_1(t), \ldots, \varphi_J(t))^T$. Maximizing (3.1) yields the maximum likelihood estimates of the eigenvalues λ, eigenfunctions $\varphi(t)$, q_i-dimensional correlation parameters θ_j, and measurement error variance σ^2. However, such a computation is intensive, if not infeasible, because 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 are established in Section 3.2.

In Step I of the estimation procedure, with

$$\widehat{\gamma}_0(t, t') = n^{-1}\sum_{i=1}^{n}\breve{y}(s_i, t)\breve{y}(s_i, t'),$$

where $\breve{y}(s_i, t) = y(s_i, t) - \mu(s_i, t)$, we estimate λ and $\varphi(t)$ as follows. First, let $\widetilde{\varphi}_1(t)$ be the maximizer of

$$\max_{\|f(t)\|_\alpha = 1}\int_0^T\int_0^T f(t)\widehat{\gamma}_0(t, t')f(t')dtdt', \tag{3.2}$$

where $\|f\|_\alpha = (f, f)_\alpha^{1/2}$, $(f, g)_\alpha = \int_0^T f(t)g(t)dt + \alpha\int_0^T f''(t)g''(t)dt$ 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 $\widetilde{\varphi}_1(t)$, defined as $\widehat{\varphi}_1(t) = \widetilde{\varphi}_1(t)/||\widetilde{\varphi}_1(t)||$, provides an estimate of $\varphi_1(t)$. For $j > 2$, let $\widetilde{\varphi}_j(t)$ be the maximizer of (3.2), subject to the constraints $(f, \widetilde{\varphi}_k)_\alpha = 0$, for $k < j$. Similarly, define $\widehat{\varphi}_j(t) = \widetilde{\varphi}_j(t)/\|\widetilde{\varphi}_j(t)\|$ as the estimate of $\varphi_j(t)$. Next, given $\widetilde{\varphi}_j(t)$, the estimate of λ_j is

$$\widehat{\lambda}_j = \int_0^T\int_0^T \widetilde{\varphi}_j(t)\widehat{\gamma}_0(t, t')\widetilde{\varphi}_j(t')dtdt', \quad j = 1, \ldots, J.$$

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

In Step II of the estimation procedure, we estimate θ_j and σ^2 by minimizing the negative log-likelihood function (3.1), given $\varphi(t) = \widehat{\varphi}(t)$ and $\lambda = \widehat{\lambda}$, as follows:

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

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

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

$$\det(\Sigma) = \det(\Phi^T \Lambda \Phi + \sigma^2 I_N) = \sigma^{2N} \det(\Lambda) \det\left(\frac{\Phi \Phi^T}{\sigma^2} + \Lambda^{-1} \right),$$

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 is adopted 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 $\xrightarrow{P}$ denote convergence in probability and $\xrightarrow{D}$ 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 $\mathcal{R}_{n^*}$ denote the spatial domain $\mathcal{R}$ 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 $\alpha \to 0$ as $n \to \infty$.

(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 $n \to \infty$.

(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 $n \to \infty$.

Conditions (A.1) and (A.2) establish the consistency of $\hat{\lambda}_j$ and $\hat{\varphi}_j(t)$ for the 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) refer to 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 Material.

In Theorem 1, we establish the consistency of the estimated eigenvalues $\widehat{\lambda}_j$ and the estimated eigenfunctions $\widehat{\varphi}_j$ obtained from Step I of the two-step procedure given in Section 3.1.

Theorem 1. *Under* (2.2) *and* (A.1)–(A.4)*, for each* j*, we have*

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

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

Theorem 1 assumes that the mean function is known, which we relax in Theorem 3.

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

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

whose (i, i')th component is the $((i-1)m+j, (i'-1)m+j)$th component of Λ. 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 θ_{j0} denotes the true value of the correlation parameter in $\rho_j(\cdot; \theta_j)$. Moreover, define $D_k\Lambda_j = \partial\Lambda_j/\partial\theta_{j,k}$, $D_{kk'}\Lambda_j = \partial^2\Lambda_j/\partial\theta_{j,k}\partial\theta_{j,k'}$, and $w_{j,kk'} = \mathrm{tr}\{(\Lambda_j + \sigma^2 I_n)^{-1}(D_k\Lambda_j)(\Lambda_j + \sigma^2 I_n)^{-1}(D_{k'}\Lambda_j)\}$, for $k, k' = 1, \ldots, q_j$ and $j = 1, \ldots, J$. Finally, with ϑ_l denoting the lth component of ϑ, 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 $t^*_{ll'} = \mathrm{tr}\{\Sigma^{-1}(D_l\Sigma)\Sigma^{-1}(D_{l'}\Sigma)\}$, for $l, l' = 1, \ldots, q$.

The regularity conditions for Theorem 2 are as follows.

(A.5) The correlation function $\rho_j(\cdot, \cdot; \theta_j)$ is twice differentiable with respect to

θ_j, with continuous second-order derivatives for $\theta_j \in \Omega_j$, where Ω_j is an open set.

(A.6) As $n \to \infty$, $\|\Lambda_j\|_2 = \mathcal{O}(1)$, $\|D_k\Lambda_j\|_2 = \mathcal{O}(1)$, and $\|D_{kk'}\Lambda_j\|_2 = \mathcal{O}(1)$, for $k, k' = 1, \ldots, q_j$ and $j = 1, \ldots, J$.

(A.7) For some $\delta > 0$, there exist positive constants C_k, such that $\|D_k\Lambda_j\|_F^{-2} \leq C_k n^{-1/2-\delta}$, for $k = 1, \ldots, q_j$ and $j = 1, \ldots, J$.

(A.8) For any $k, k' = 1, \ldots, q_j$, $a_{j,kk'} = \lim_{n\to\infty}\{w_{j,kk'}(w_{j,kk}w_{j,k'k'})^{-1/2}\}$ exists and $A_j = [a_{j,kk'}]_{k,k'=1}^{q}$ is nonsingular, for $j = 1, \ldots, J$.

(A.9) There exists a positive constant c_0, such that $\|\Lambda_j^{-1}\|_2 < c_0 < \infty$, for $j = 1, \ldots, J$.

(A.10) 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 the 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 the 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 the smoothness, growth, and convergence of the information matrix for spatio-temporal data (Sweeting (1980)).

Let $\ell''(\vartheta, \vartheta) = (\partial^2\ell(\lambda, \varphi(t), \theta_j, \sigma^2))/(\partial\vartheta\ \partial\vartheta^T)$ be the second-order derivatives of $\ell(\lambda, \varphi(t), \theta_j, \sigma^2)$ with respect to ϑ. Under (A.5)–(A.10), the asymptotic normality of $\widehat{\theta}_j$ is established in the following theorem.

Theorem 2. *Under* (2.4) *and* (A.5)–(A.10)*, we have*

$$H(\vartheta_0)^{1/2}(\widehat{\vartheta} - \vartheta_0) \xrightarrow{D} N(0, I_q),$$

as $n \to \infty$*, where* $q = \sum_{j=1}^{J} q_j$*,* $\widehat{\vartheta} = (\widehat{\theta}_1^T, \ldots, \widehat{\theta}_J^T)^T$*, and* $H(\vartheta_0) = E\{-\ell''(\vartheta_0, \vartheta_0)\}$ *is the information matrix for* ϑ*.*

Although Theorems 1 and 2 give the consistency and asymptotic normality of the parameter estimates, they are established for detrended spatio-temporal data. We relax this assumption 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 usually 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 $\bar{y}_s(t) = y(s,t) - \bar{\mu}(s,t)$ denote the detrended process. Furthermore, let $\bar{y}_{s_i} = (y(s_i,t_{i1}) - \bar{\mu}(s_i,t_{i1}),\ldots,y(s_i,t_{im_i}) - \bar{\mu}(s_i,t_{im_i}))^T$ 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 $\breve{y}(s,t)$, $\breve{y}_{s_i}$, and $\breve{y}$ with $\bar{y}(s,t)$, $\bar{y}_{s_i}$, and $\bar{y}$, respectively.

Specifically, in Step I of the modified two-step procedure, let $\bar{\gamma}_0(t,t') = n^{-1}\sum_{i=1}^n \bar{y}(s_i,t)\bar{y}(s_i,t')$. For $j=1$, let $\widetilde{\varphi}_1(t)$ be the maximizer of

$$\max_{\|f(t)\|_\alpha=1} \int_0^T \int_0^T f(t)\bar{\gamma}_0(t,t')f(t')dtdt'. \tag{4.1}$$

For $j>2$, let $\widetilde{\varphi}_j(t)$ be the maximizer of (4.1), subject to the constraints $(f,\widetilde{\varphi}_k)_\alpha = 0$ for $k<j$. After standardization, $\widehat{\varphi}_j(t) = \widetilde{\varphi}_j(t)/\|\widetilde{\varphi}_j(t)\|$ becomes the estimate of $\varphi_j(t)$. The estimate of λ_j is

$$\widehat{\lambda}_j = \int_0^T \int_0^T \widetilde{\varphi}_j(t)\bar{\gamma}_0(t,t')\widetilde{\varphi}_j(t')dtdt', \quad j=1,\ldots,J.$$

In Step II of the modified two-step procedure, the negative log-likelihood function (3.1), after substituting in the estimated mean function, takes the form

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

and is minimized with $\varphi(t) = \widehat{\varphi}(t)$ and $\lambda = \widehat{\lambda}$ held fixed. The resulting estimates are denoted as $\widehat{\theta}_j$ and $\widehat{\sigma}^2$.

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, and $\beta = (\beta_1,\ldots,\beta_p)^T$ denotes a p-dimensional vector of regression coefficients. The model (2.4) becomes

$$y(s,t) = x(s,t)^T\beta + \sum_{j=1}^J \xi_j(s)\varphi_j(t) + \upsilon(s,t). \tag{4.2}$$

Let $X(s_i) = (x(s_i,t_{i1}),\ldots,x(s_i,t_{im_i}))^T$ 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 (2.4) is

$$\ell(\lambda, \varphi(t), \beta, \theta_j, \sigma^2) = (y - X\beta)^T \Sigma^{-1} \frac{y - X\beta}{2} + \left(\frac{1}{2}\right) \log\{\det(\Sigma)\} + \left(\frac{N}{2}\right) \log(2\pi). \tag{4.3}$$

A practical choice of $\bar{\mu}(s,t)$ is $\bar{\mu}(s,t) = x(s,t)^T \bar{\beta}_{\text{ols}}$, where $\bar{\beta}_{\text{ols}} = (X^T X)^{-1} X^T y$ is the least squares estimate of β. 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 (4.3) with respect to β, θ_j, and σ^2, with $\varphi(t) = \widehat{\varphi}(t)$ and $\lambda = \widehat{\lambda}$ held fixed. The resulting estimates are denoted as $\widehat{\beta}$, $\widehat{\theta}_j$, and $\widehat{\sigma}^2$.

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 $E\{\bar{\mu}(s,t) - \mu(s,t)\}^4 \le c_n$, for $t \in [0,T]$, where c_n does not depend on s.

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

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

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

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

as $n \to \infty$.

For model (4.2), we establish the asymptotic properties of $\widehat{\beta}$ and $\widehat{\theta}_j$ from Step II, given λ, $\varphi_j(t)$, and σ^2. An additional regularity condition is assumed about the design matrix, which is standard for spatial linear models (Mardia and Marshall (1984)).

(A.12) The design matrix X has full rank p and is uniformly bounded in the max norm, with $\lim_{n\to\infty}(X^T X)^{-1} = 0$.

Let $\ell''(\beta,\beta) = (\partial^2 \ell(\lambda, \varphi(t), \beta, \theta_j, \sigma^2))/(\partial\beta\, \partial\beta^T)$ be the second-order derivative of $\ell(\lambda, \varphi(t), \beta, \theta_j, \sigma^2)$ with respect to β. The asymptotic normality of $\widehat{\beta}$ and

$\widehat{\theta}_j$ are established in Theorem 4.

Theorem 4. *Under the assumptions of* Theorem 2 *and* (A.12)*, we have*

$$\begin{pmatrix} H(\beta_0)^{1/2} & O_{p\times q} \\ O_{q\times p} & H(\vartheta_0)^{1/2} \end{pmatrix} \left\{ \begin{pmatrix} \widehat{\beta} \\ \widehat{\vartheta} \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \vartheta_0 \end{pmatrix} \right\} \xrightarrow{D} N(0, I_{p+q}),$$

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

The proof of the proposition and theorems above are provided in the Supplementary Material. The Gaussian assumption can be readily relaxed in Theorems 1 and 3, but not 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 standard normal distributions with a cross-covariate correlation of 0.5, and the regression coefficients are set to $\beta = (4, 3, 2, 1, 0, 0, 0)^T$. Each covariate is standardized to have sample mean zero and sample variance one and the response has a sample mean zero. Thus, there is no intercept in this model. The spatio-temporal 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 independent zero-mean stationary and isotropic Gaussian processes, with exponential covariance functions $\lambda_1 \exp(-d/r_1)$ and $\lambda_2 \exp(-d/r_2)$, respectively, for the spatial distance d, 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 orthonormal functions on $[0, 1]$, with normalization constants c_1 and c_2, respectively. The number of sampling locations is set to $n = 50$, 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, 20 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, we apply our method to estimate the regression coefficients β, spatial parameters (r_1, r_2), eigenvalues (λ_1, λ_2), eigenfunctions $(\varphi_1(t), \varphi_2(t))$, and measurement error variance σ^2. We compare our method 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, which accounts for temporal, but not spatial dependence (denoted as

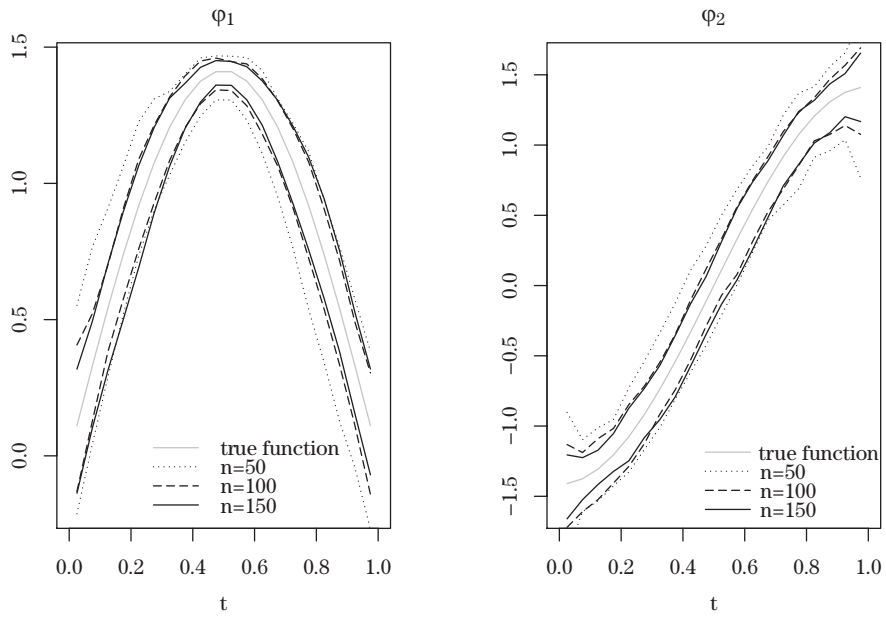

Figure 1. The 95% pointwise simulation intervals for $\varphi_1(t)$ (left) and $\varphi_2(t)$ (right) using our method. The true $\varphi_1(t)$ and $\varphi_2(t)$ are indicated as a gray solid line. The pointwise simulation intervals for $n = 50$, 100, and 150 are indicated as black dotted, dashed, and solid lines, respectively.

ALT_2). Moreover, the prediction in space and time is performed using all three approaches.

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

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

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

Table 1 reports the mean and the standard deviation of the regression coefficient estimates from 100 simulated data sets using the three approaches. The regression coefficient estimates have a 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 the ordinary

Table 1. The mean, and standard deviation (SD) of the regression coefficient estimates, and the 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 sizes $n = 50$, 100, and 150.

n	Method	β_1	β_2	β_3	β_4	β_5	β_6	β_7	MSPE_1	MSPE_2
	True Values	4.000	3.000	2.000	1.000	0.000	0.000	0.000	–	–
	SemiCov	4.001	3.000	1.997	0.996	0.006	−0.006	0.009	2.667	1.112
	SD	0.043	0.044	0.045	0.048	0.042	0.043	0.041	1.006	0.177
	SDm	0.045	0.045	0.045	0.045	0.045	0.045	0.045	–	–
50	ALT_1	4.005	3.005	1.996	0.992	0.006	−0.005	0.009	3.927	3.768
	SD	0.086	0.077	0.087	0.076	0.081	0.082	0.076	1.703	0.744
	ALT_2	4.001	3.000	1.997	0.996	0.006	−0.007	0.009	3.917	1.114
	SD	0.043	0.044	0.045	0.049	0.043	0.043	0.041	1.704	0.177
	SemiCov	4.001	2.992	1.998	1.007	0.002	0.000	−0.001	2.610	1.106
	SD	0.031	0.030	0.029	0.031	0.033	0.027	0.030	0.715	0.108
	SDm	0.031	0.031	0.031	0.031	0.031	0.031	0.031	–	–
100	ALT_1	3.996	2.991	2.001	1.011	0.005	−0.005	0.004	3.720	3.775
	SD	0.053	0.055	0.054	0.058	0.056	0.053	0.057	1.130	0.453
	ALT_2	4.001	2.992	1.998	1.007	0.002	0.000	−0.001	3.715	1.109
	SD	0.031	0.030	0.029	0.032	0.033	0.027	0.030	1.133	0.108
	SemiCov	4.001	3.004	1.997	0.997	−0.001	0.000	0.002	2.616	1.093
	SD	0.027	0.026	0.025	0.027	0.032	0.024	0.027	0.600	0.105
	SDm	0.025	0.025	0.025	0.025	0.025	0.025	0.026	–	–
150	ALT_1	4.005	3.007	1.991	1.001	−0.001	−0.007	0.005	3.836	3.847
	SD	0.043	0.054	0.051	0.048	0.049	0.045	0.045	0.914	0.414
	ALT_2	4.001	3.004	1.997	0.997	0.000	0.000	0.002	3.830	1.095
	SD	0.027	0.026	0.025	0.027	0.032	0.023	0.027	0.912	0.105

least squares in terms of yielding 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 $\text{sd}\{\widehat{\beta}\} = \text{diag}\{H(\widehat{\beta})^{-1}\}^{1/2}$, $\text{sd}\{\widehat{\lambda}\} = \text{diag}\{H(\widehat{\lambda})^{-1}\}^{1/2}$, and $\text{sd}\{\widehat{\vartheta}\} = \text{diag}\{H(\widehat{\vartheta})^{-1}\}^{1/2}$, where $\widehat{\beta}$, $\widehat{\lambda}$, and $\widehat{\vartheta}$ are the estimates from the simulated data set, and $H(\beta)$, $H(\lambda) = E\{-\ell''(\lambda, \lambda)\}$, and $H(\vartheta)$ are the information matrices for β, λ, and ϑ, 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 the nominal true standard deviations of the regression coefficient estimates.

Table 2. The mean and standard deviation (SD) of spatial-temporal coefficient estimates under our method (SemiCov) for sample size $n = 50$, 100, and 150.

n	Method	λ_1	r_1	λ_2	r_2	σ^2
	True Values	2.50	0.50	0.50	0.30	1.00
50	SemiCov	2.32	0.46	0.48	0.29	1.03
	SD	0.63	0.20	0.10	0.11	0.08
	SDm	0.65	0.18	0.12	0.12	0.05
100	SemiCov	2.31	0.47	0.52	0.32	1.02
	SD	0.45	0.14	0.09	0.11	0.05
	SDm	0.46	0.13	0.10	0.09	0.03
150	SemiCov	2.43	0.48	0.53	0.31	1.01
	SD	0.36	0.10	0.08	0.07	0.03
	SDm	0.41	0.11	0.08	0.07	0.03

For the prediction, we consider two scenarios. In Scenario 1, the 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

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

where c_0 is an N-dimensional vector with an ith element of $\operatorname{cov}\{y(s_0, t_0), y_i\}$, y_i is the ith element of y, $\Sigma = \operatorname{cov}(y)$, and $\widetilde{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y$ (Cressie (1993)). Because Σ and c_0 are usually unknown, the estimates of Σ and c_0 are substituted in, and an empirical best linear unbiased prediction is obtained.

To quantify the prediction error for the curve at location s_0, we use the mean integral squared error,

$$\mathrm{MISE}(s_0) = \int_0^T \{\widetilde{y}(s_0, t) - y(s_0, t)\}^2 dt.$$

In the simulation study, $T = 1$ and t_i are evenly distributed and, thus, $\mathrm{MISE}(s_0)$ is estimated by $m^{-1}\sum_{i=1}^m \{\widetilde{y}(s_0, t_{0i}) - y(s_0, t_{0i})\}^2$. Moreover, we generate 5, 10, and 15 additional curves for sample sizes $n = 50$, 100, and 150, respectively, and set them aside for prediction. Lastly, for the M unsampled curves at $s_{01}, \ldots, s_{0M}$, we define the first mean squared prediction error (MSPE_1) as $M^{-1}\sum_{i=1}^M \mathrm{MISE}(s_{0i})$.

In Scenario 2, we predict missing time points at a sampling location s_i, for $i = 1, \ldots, n$. For m unsampled time points $t_{0i} = (t_{0i,1}, \ldots, t_{0i,m})^T$ at each location s_i, we define a second mean squared prediction error (MSPE_2) as $(mn)^{-1}\sum_{i=1}^n \sum_{k=1}^m \{\widetilde{y}(s_i, t_{0i,k}) - y(s_i, t_{0i,k})\}^2$. For the simulation, predictions are made at $\{(s_i, t_{0i}) : t_{0i} = (0.25, 0.75)^T, i = 1, \ldots, n\}$.

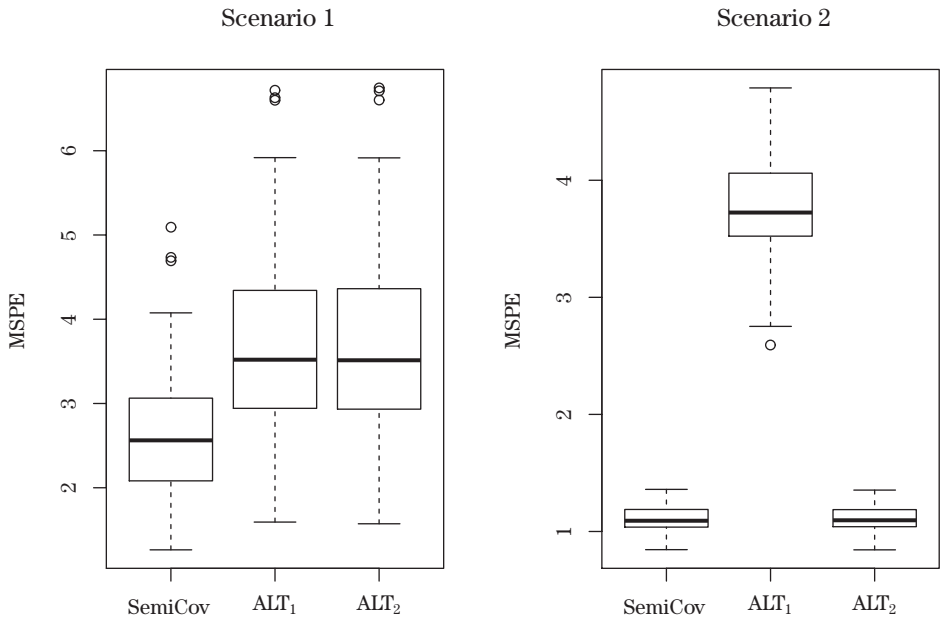

Figure 2. The mean squared prediction error (MSPE) for Scenario 1 (left) and Scenario 2 (right) under our method (SemiCov) and two alternative approaches (ALT_1 and ALT_2).

Table 1 shows the prediction performance for different sample sizes; the associated box-plots for $n = 100$ are given in Figure 2. The results in Scenario 1 show that ALT_2 outperforms ALT_1, while our SemiCov method outperforms both ALT_1 and ALT_2. This provides empirical evidence that incorporating the spatial correlation between locations can substantially improve the predictions 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 predicting at sampled locations with missing time points.

Finally, Table 2 reports the mean, standard deviation (SD), and mean standard error (SDm) of the estimates of the spatio-temporal parameters. The means of the 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

This example examines precipitation data (in inches per 24-hour period) for

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.

Method	Elevation	Slope	Aspect	B1M	B2M	B3M	B4M	B5M
SemiCov	0.190	0.003	0.006	0.091	0.006	0.032	−0.208	0.019
SE	0.019	0.010	0.009	0.062	0.026	0.043	0.072	0.037
ALT_1	0.046	0.047	0.010	−0.142	−0.079	0.052	−0.083	0.110
ALT_2	0.101	0.026	0.018	−0.145	−0.031	0.091	−0.077	0.004
	B6M	B7M	$MSPE_1$	$MSPE_2$				
SemiCov	0.016	−0.002	0.143	0.085				
SE	0.046	0.040	–	–				
ALT_1	−0.011	−0.009	0.474	0.137				
ALT_2	0.115	−0.047	0.473	0.077				

the period January to December, on a log-scale from 259 weather stations in Colorado (Reich and Davis (2008); Chu, Zhu and Wang (2011)), as shown in the left-hand panel of Figure 3. There are 10 covariates of interest, including elevation, slope, aspect, and seven spectral bands from satellite imagery (B1M through B7M). For model fitting, we use precipitation data for 10 months (excluding March and October) from 240 weather stations. Two types of prediction are considered. First, predictions are made for the remaining 19 weather stations, and the prediction results are summarized by $MSPE_1$. Second, the March and October precipitation data 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 suggests there is a spatio-temporal dependence for Colorado precipitation data. A data analysis is performed using our SemiCov method and two alternative approaches. For our SemiCov method, we choose the number of components $J = 2$, such that $\sum_{j=1}^{J} \widehat{\lambda}_j / \sum_{j=1}^{n} \widehat{\lambda}_j \geq 80\%$, as suggested by Zhu, Fan and Kong (2014). Furthermore, because there is multicollinearity among the covariates, we apply a forward selection using AIC. The resulting model contains two covariates, elevation and B4M. The results without the forward selection are reported in Table 3, and those with the forward selection are reported in Table 4. For the SemiCov method, there is strong evidence of an elevation and B4M effect, while there is moderate evidence of an effect of B1M. When predicting all time points at unsampled locations, our SemiCov method outperforms the two alternative methods. On the other hand, for the prediction at the two time points set aside at the sampling locations, both our method and the functional data

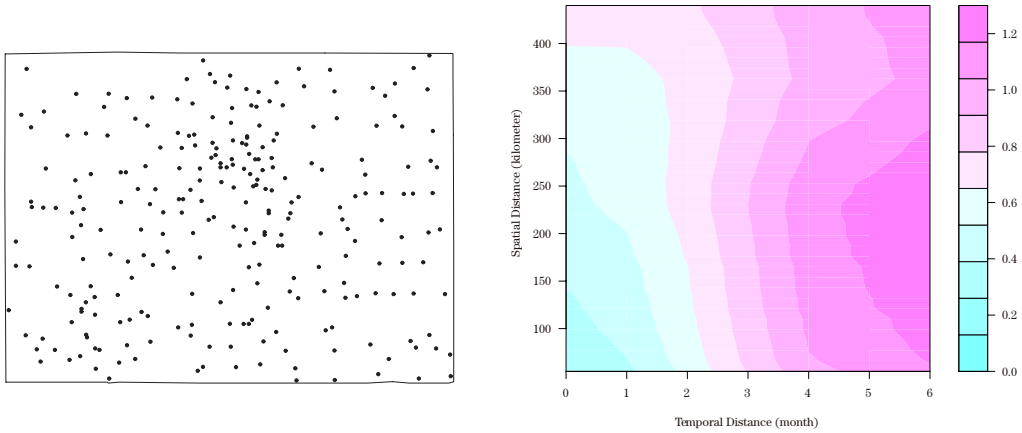

Figure 3. Map of locations of 259 weather stations in the Colorado precipitation data (left), and the empirical variogram over spatial and temporal lags (right).

Table 4. Precipitation data example with 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.

Method	Elevation	B4M	$MSPE_1$	$MSPE_2$
SemiCov	0.203	−0.059	0.140	0.085
SE	0.018	0.013	–	–
ALT_1	0.054	−0.173	0.469	0.140
ALT_2	0.101	−0.102	0.464	0.077

analysis outperform the ordinary least squares method, although the functional data analysis is slightly better than our method in this case.

Supplementary Material

The Supplementary Material contains the proofs of Proposition 1 and Theorems 1–4, as well as additional numerical studies.

Acknowledgements

We would like to thank a co-editor, an associate editor, and two referees for their helpful comments and suggestions. The research was partially supported by the National Natural Science Foundation of China (grant 11301536) for Tingjin Chu, the USGS CESU Award G16AC00344 for Jun Zhu, and the NSF grants DMS-1521746 and DMS-1737795 for Haonan Wang.

References

Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistican Association* **85**, 749–759.

Brynjarsdóttir, J. and Berliner, L. M. (2014). Dimension-reduced modeling of spatio-temporal processes. *Journal of the American Statistical Association* **109**, 1647–1659.

Choi, I., Li, B. and Wang, X. (2013). Nonparametric estimation of spatial and space-time covariance function. *Journal of Agricultural, Biological, and Environmental Statistics* **18**, 611–630.

Chu, T., Zhu, J. and Wang, H. (2011). Penalized maximum likelihood estiamtion and variable selection in geostatistics. *The Annals of Statistics* **39**, 2607–2625.

Cressie, N. (1993). *Statistics for Spatial Data*. revised edition, Wiley, New York.

Cressie, N. and Huang, H. C. (1999). Classes of nonseparable spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* **94**, 1330–1340.

Cressie, N., Shi, T. and Kang, E. L. (2010). Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics* **19**, 724–745.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-temporal Data*. Wiley, New York.

El Machkouri, M. (2007). Nonparametric regression estimation for random fields in a fixed-design. *Statistical Inference for Stochastic Processes* **10**, 29–47.

Fuentes, M., Chen, L. and Davis, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics* **19**, 487–507.

Ghanem, R. G. and Spanos, P. D. (1991). *Stochastic Finite Elements: A Special Approach*. Springer.

Ghosh, S. K., Bhave, P. V., Davis, J. M. and Lee, H. (2010). Spatio-temporal analysis of total nitrate concentrations using dynamic statistical models. *Journal of the American Statistical Association* **105**, 538–551.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for spacectime data. *Journal of the American Statistical Association* **97**, 590–600.

Gromenko, O. and Kokoszka, P. (2013). Nonparametric inference in small data sets of spatially indexed curves with application to ionospheric trend determination. *Computational Statistics and Data Analysis* **59**, 82–94.

Gromenko, O., Kokoszka, P., Zhu, L. and Sojka, J. (2012). Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *The Annals of Applied Statistics* **6**, 669–696.

Harville, D. (2008). *Matrix Algebra From A Statistician's Perspective*. Springer.

Hörmann, S. and Kokoszka, P. (2013). Consistency of the mean and the principal components of spatially distributed functional data. *Bernoulli* **19**, 1535–1558.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer.

Huang, H. C. and Cressie, N. (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis* **22**, 159–175.

Johannesson, G., Cressie, N. and Huang, H. C. (2007). Dynamic multi-resolution spatial models. *Environmental and Ecological Statistics* **14**, 5–25.

Lu, Z., Steinskog, D. J., Tjøstheim, D. and Yao, Q. (2009). Adaptively varying coefficient spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 859–880.

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

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F. and Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics* **24**, 579–599.

Paul, D. and Peng, J. (2011). Principal components analysis for sparsely observed correlated functional data using a kernel smoothing approach. *Electronic Journal of Statistics* **5**, 1960–2003.

Reich, R. M. and Davis, R. (2008). *Lecture Notes of Quantitative Spatial Analysis*. Colorado State University, Fort Collins, Colorado.

Rodrigues, A. and Diggle, P. (2010). A class of convolution-based models for spatio-temporal processes with non-separable covariance structure. *Scandinavian Journal of Statistics* **37**, 553–567.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics* **24**, 1–24.

Stein, M. L. (2005). Spacectime covariance functions. *Journal of the American Statistical Association* **100**, 310–321.

Sweeting, T. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *The Annals of Statistics* **8**, 1375–1381.

Zhang, B., Sang, H. and Huang, J. Z. (2015). Full-scale approximations of spatio-temporal covariance models for large datasets. *Statistica Sinica* **25**, 99–114.

Zhang, W., Yao, Q. and Tong, H. (2003). Smoothing for spatiotemporal models and its application to modeling muskrat-mink interaction. *Biometrics* **59**, 813–821.

Zhu, H., Fan, J. and Kong, L. (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association* **109**, 1084–1098.

School of Mathematics and Statistics, University of Melbourne, Parkville, VIC 3010, Australia.

E-mail: tingjin.chu@unimelb.edu.au

Department of Statistics, University of Wisconsin Madison, Madison WI 53706, USA.

E-mail: jzhu@stat.wisc.edu

Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA.

E-mail: wanghn@stat.colostate.edu

(Received June 2016; accepted September 2018)