Sankhyā A : The Indian Journal of Statistics
https://doi.org/10.1007/s13171-020-00213-4
© 2020, Indian Statistical Institute

Spatio-Temporal Expanding Distance Asymptotic Framework for Locally Stationary Processes

Tingjin Chu
University of Melbourne, Parkville, Australia
Jialuo Liu and Haonan Wang
Colorado State University, Fort Collins, USA
Jun Zhu
University of Wisconsin-Madison, Madison, USA

Abstract

Spatio-temporal data indexed by sampling locations and sampling time points are encountered in many scientific disciplines such as climatology, environmental sciences, and public health. Here, we propose a novel spatio-temporal expanding distance (STED) asymptotic framework for studying the properties of statistical inference for nonstationary spatio-temporal models. In particular, to model spatio-temporal dependence, we develop a new class of locally stationary spatio-temporal covariance functions. The STED asymptotic framework has a fixed spatio-temporal domain for spatio-temporal processes that are globally nonstationary in a rescaled fixed domain and locally stationary in a distance expanding domain. The utility of STED is illustrated by establishing the asymptotic properties of the maximum likelihood estimation for a general class of spatio-temporal covariance functions. A simulation study suggests sound finite-sample properties and the method is applied to a sea-surface temperature dataset.

AMS (2000) *subject classification*. Primary 62F12; Secondary 62M30.
Keywords and phrases. Covariance functions, Nonstationary processes, Random fields, Spatial statistics, Spatio-temporal statistics

1 Introduction

Spatio-temporal data are widely encountered and analyzed in many scientific disciplines, such as climatology (see, e.g., Cressie 2018; Kuusela and Stein 2018), environmental sciences (see, e.g., Liang et al. 2015; Porcu et al. 2018), and public health (see, e.g., Ludwig et al. 2017). While there are a myriad of statistical modeling and methods for analyzing spatio-temporal data (see, e.g., Sherman 2011; Cressie and Wikle 2011), there appear to be limited tools for studying the theoretical properties of these statistical techniques. The purpose of this paper is to fill some of this void in spatio-temporal statistics by proposing a novel asymptotic framework for data sampled in space and time.

For spatio-temporal data, spatio-temporal covariance functions have been proposed and employed to model spatio-temporal dependence. For example, Cressie and Huang (1999) and Gneiting (2002) constructed fully parametric nonseparable spatio-temporal covariance functions using spectral density and completely monotone functions. Stein (2005) developed spatially isotropic but asymmetric spatio-temporal models by taking the derivatives of spatially isotropic fully symmetric models. These spatio-temporal covariance models assume stationarity in both space and time, which could be restrictive in practice. For time series data, various nonstationary models have been developed including locally stationary processes (see, e.g., Dahlhaus 1997; Zhou and Wu 2009; Vogt 2012; Dahlhaus 2012) and mixing conditions (see, e.g., Fan and Yao 2003; Chang et al. 2015). For spatial data, nonstationary covariance functions have also been developed, such as kernel-based spatial convolution and spectral-based local stationarity (see, e.g., Higdon 1998; Fuentes 2002; Paciorek and Schervish 2006; Gelfand et al. 2010). More recently, Hsing et al. (2016) suggested a class of locally intrinsic stationary (LIS) covariance functions, which includes a variety of nonstationary models and has sound theoretical properties. However, there are very limited results on local stationarity for spatio-temporal processes. Nonstationary spatio-temporal processes have been considered, such as nonstationary models via spectral representation (Fuentes et al., 2008; Guinness and Fuentes, 2015) and a moving-window approach to estimating a locally stationary spatio-temporal Gaussian process (Kuusela and Stein, 2018), although the asymptotic properties of model estimation and inference are unexplored. We believe that advances are in need for studying the theoretical properties of locally stationary processes.

Asymptotic frameworks have played an important role in establishing the asymptotic properties of parameter estimates and their inference in spatial statistics. In an increasing domain asymptotic framework, the spatial domain expands while the smallest distance among the spatial sampling locations remains constant (see, e.g., Mardia and Marshall 1984; Cressie and Lahiri 1993; Yao and Brockwell 2006; Chu et al. 2011). In an infill asymptotic framework, the spatial sampling locations become denser in a fixed spatial domain (see, e.g., Ying 1993; Stein 1999; Zhang 2004; Loh

2005). A mixed asymptotic framework has both an expanding spatial domain and denser sampling locations (see, e.g., Hall and Patil 1994; Lahiri 2003; Lu and Tjøstheim 2014; Bandyopadhyay and Rao 2017). However, asymptotic frameworks are underdeveloped for spatio-temporal statistics, leaving the asymptotic properties of many methods for nonstationary spatio-temporal data unclear. Yet, it is non-trivial to extend the existing asymptotic frameworks for spatial processes to spatio-temporal processes due to the uni-directionality of time and a lack of stationarity. For instance, in an increasing domain asymptotic framework, the "local" behavior of a covariance function is a challenge to study, whereas in an infill asymptotic framework, some of the parameters in the covariance function are not consistently estimable. Recently, Bandyopadhyay et al. (2017) considered a spatio-temporal domain asymptotics with the increasing temporal domain and the mixed spatial domain for Fourier analysis in spectral domain.

Here, to incorporate the local stationarity, we propose a novel spatio-temporal expanding distance (STED) asymptotic framework in a fixed spatio-temporal domain. Let $\mathcal{R}$ denote a spatial domain of interest in $\mathbb{R}^d$ and $\mathcal{T}$ denote a temporal domain of interest in $\mathbb{R}$. In the one-dimensional space ($d = 1$), the sampling locations are for example $i = 1, \ldots, n$ at the nth stage in an increasing domain asymptotic framework, but are $1/n, \ldots, (n-1)/n, 1$ in an infill asymptotic framework. In nonlinear time series, a popular approach is to rescale the actual time to i/n (Fan and Yao, 2003), which we extend to a rescaled spatio-temporal domain such that both the spatial domain and the temporal domain are bounded.

Besides asymptotic framework, the asymptotic properties of parameter estimates also depends on the nature of spatio-temporal dependence. Here, we take the LIS framework for spatial processes as impetus (Hsing et al., 2016), and develop a class of locally stationary spatio-temporal covariance functions that vary across space and over time within the rescaled fixed spatio-temporal domain. The resulting spatio-temporal covariance functions are locally stationary in a distance expanding spatio-temporal domain; that is, they can be approximated locally by stationary covariance functions of actual distances in the spatio-temporal domain. Such a class of spatio-temporal covariance functions is quite general and flexible as we will demonstrate. Furthermore, our proposed STED asymptotic framework is not a generalization of the mixed asymptotic framework in spatial statistics, but rather a potentially useful tool for studying the properties of statistical inference for spatio-temporal processes that are globally nonstationary in a rescaled fixed domain and locally stationary in a distance expanding domain.

The remainder of the paper is organized as follows. We develop a class of locally stationary spatio-temporal processes in Section 2 and propose the STED asymptotic framework for fixed spatio-temporal domain in Section 3. For illustration, we consider spatio-temporal models and derive the theoretical properties of the corresponding maximum likelihood estimation in Section 4. Simulation studies are conducted in Section 5.1, and our method is applied to a sea-surface temperature data for illustration in Section 5.2. The technical details including theorem proofs and remarks are given in Appendices A.1, A.2, A.3 and A.4.

2 Local Stationarity in Space and Time

For the spatial domain of interest $\mathcal{R} \subset \mathbb{R}^d$ and the temporal domain of interest $\mathcal{T} \subset \mathbb{R}$, we consider a zero-mean spatio-temporal random process $\{Y(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathcal{R}, t \in \mathcal{T}\}$, and at stage n of a spatio-temporal asymptotic framework, the covariance function is defined as $\gamma_n((\boldsymbol{s}, t), (\boldsymbol{s}', t')) = \mathrm{Cov}(Y(\boldsymbol{s}, t), Y(\boldsymbol{s}', t'))$, where $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{R}$ and $t, t' \in \mathcal{T}$. To draw inference for the covariance functions, stationarity is generally assumed, which can be restrictive and may not hold in practice. Here we consider a new class of spatio-temporal covariance functions, allowing more flexibility than stationary processes to the extent of local stationarity.

We let the stage of the spatio-temporal asymptotics, n, appear as either a left superscript or a right subscript of a quantity that depends on n. We also let $\{A_n\}$ and $\{B_n\}$ denote two sequences of positive numbers and let $\|\cdot\|$ denote the Euclidean norm in $\mathbb{R}^d$. The following are conditions for defining locally stationary spatio-temporal covariance functions.

(LS.1). There exists a sequence of functions $g_n(\cdot, \cdot, \boldsymbol{s}, t)$ such that

$$\left|\gamma_n\left((\boldsymbol{s}, t), (\boldsymbol{s}', t')\right) - g_n\left(\boldsymbol{s}' - \boldsymbol{s}, t' - t, \boldsymbol{s}, t\right)\right| = \mathcal{O}\left(\|\boldsymbol{s}' - \boldsymbol{s}\| + |t' - t| + \rho_n\right)$$

uniformly for all $(\boldsymbol{s}, t), (\boldsymbol{s}', t') \in \mathcal{R} \times \mathcal{T}$, where $\{\rho_n\}$ is a sequence of positive numbers such that $\rho_n \to 0$ as $n \to \infty$. In addition, there exists a function g such that

$$\lim_{n\to\infty} \left|g_n\left(\boldsymbol{s}' - \boldsymbol{s}, t' - t, \boldsymbol{s}, t\right) - g(\boldsymbol{u}_1, u_2, \boldsymbol{s}, t)\right| \to 0, \text{ as } n \to \infty$$

uniformly for all $(\boldsymbol{s}, t), (\boldsymbol{s}', t') \in \mathcal{R} \times \mathcal{T}$, where $\boldsymbol{u}_1 = A_n(\boldsymbol{s}' - \boldsymbol{s})$ and $u_2 = B_n(t' - t)$.

(LS.2). Define $g(\boldsymbol{s}, t) = g(\boldsymbol{0}, 0, \boldsymbol{s}, t)$ with $g(\boldsymbol{u}_1, u_2, \boldsymbol{s}, t)$ given in (LS.1), and $g(\boldsymbol{s}, t)$ satisfies $|g(\boldsymbol{s}, t) - g(\boldsymbol{s}', t')| \leq C_1\|\boldsymbol{s} - \boldsymbol{s}'\| + C_2|t - t'|$ for all $(\boldsymbol{s}, t), (\boldsymbol{s}', t') \in \mathcal{R} \times \mathcal{T}$, where $C_1, C_2 > 0$ are constants.

(LS.3). There exist two positive nonincreasing functions γ_0 and γ_1 satisfying $\int_0^\infty u^{d-1}\gamma_0(u)du < \infty$ and $\int_0^\infty \gamma_1(u)du < \infty$ such that $|\gamma_n((\boldsymbol{s},t),(\boldsymbol{s}+\boldsymbol{u}_1/A_n, t+u_2/B_n))| \leq \gamma_0(\|\boldsymbol{u}_1\|)\gamma_1(|u_2|)$ for all n and $\|\boldsymbol{u}_1\|, |u_2| \in [0,\infty)$ such that $(\boldsymbol{s},t),(\boldsymbol{s}+\boldsymbol{u}_1/A_n, t+u_2/B_n) \in \mathcal{R}\times\mathcal{T}$.

Here, (LS.1)–(LS.2) can be viewed as a generalization of (W3)–(W4) for spatial processes in Hsing et al. (2016) to our spatio-temporal processes. In particular, (LS.1) describes *local stationarity* in the sense that the covariance function γ_n can be approximated by a function g_n, which is allowed to vary with location $\boldsymbol{s}$ and time t, rather than merely determined by spatial and temporal lags. Such approximation is adequate in a neighborhood of $(\boldsymbol{s},t)$, provided that $(\boldsymbol{s},t)$ and $(\boldsymbol{s}',t')$ are sufficiently close. Furthermore, the function g characterizes the limiting behavior of g_n with respect to the scaled spatial and temporal lags at the rates of A_n and B_n, respectively. (LS.2) imposes some mild restrictions on the covariance structure at the zero lag in space and time, whereas (LS.3) is a constraint on the decay rate of the covariance function in space and time.

The above definition of locally stationary covariance functions is satisfied by a variety of covariance functions. For illustration, we introduce a class of parametric covariance functions denoted as $\gamma_n((\boldsymbol{s},t),(\boldsymbol{s}',t');\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a $q\times 1$ vector of parameters, which includes the following generalized spatio-temporal Matérn covariance function

$$\gamma_n((\boldsymbol{s},t),(\boldsymbol{s}',t');\boldsymbol{\theta}) = \begin{cases} \frac{D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2\theta_3^{d/2}2^{1-\nu}}{(\theta_1^2u_2^2+1)^\nu(\theta_1^2u_2^2+\theta_3)^{d/2}\Gamma(\nu)} m(\boldsymbol{u}_1,u_2;\boldsymbol{\theta})^\nu K_\nu\left\{m(\boldsymbol{u}_1,u_2;\boldsymbol{\theta})\right\}, & \text{if } \|\boldsymbol{u}_1\|>0,\\ \frac{D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2\theta_3^{d/2}}{(\theta_1^2u_2^2+1)^\nu(\theta_1^2u_2^2+\theta_3)^{d/2}}, & \text{if } \|\boldsymbol{u}_1\|=0, |u_2|>0,\\ D(\boldsymbol{s},t)^2\sigma^2+\tau^2, & \text{if } \|\boldsymbol{u}_1\|=0, |u_2|=0, \end{cases} \tag{2.1}$$

where $m(\boldsymbol{u}_1,u_2;\boldsymbol{\theta}) = \theta_2\left(\frac{\theta_1^2u_2^2+1}{\theta_1^2u_2^2+\theta_3}\right)^{1/2}\|\boldsymbol{u}_1\|$, $\boldsymbol{\theta}=(\theta_1,\theta_2,\theta_3,\sigma^2,\tau^2)^\top$ is a vector of spatio-temporal parameters with a scaling parameter in time $\theta_1>0$, a scaling parameter in space $\theta_2>0$, a separability parameter $\theta_3\geq 1$, and are two variance components σ^2 and τ^2. In addition, $\boldsymbol{u}_1=\varrho_{1,n}(\boldsymbol{s}'-\boldsymbol{s})$ is the spatial lag scaled to the spatially expanding domain, and $u_2=\varrho_{2,n}(t'-t)$ is the temporal lag scaled to the temporally expanding domain, where $\varrho_{1,n}$ and $\varrho_{2,n}$ are two sequences of positive real numbers. Further, $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν, $\nu>0$ is a smoothness parameter assumed to be known, and $D(\boldsymbol{s},t)$ is a positive spatio-temporal function such that $D(\boldsymbol{0},0)=1$ and $D(\boldsymbol{s},t)^2\sigma^2+\tau^2$ is the variance of $Y(\boldsymbol{s},t)$. By Cressie and Huang (1999) and Gneiting (2002), it can be shown that (2.1) is a positive definite function and therefore, a valid covariance function. The class of spatio-temporal covariance functions (2.1) is generally nonseparable and nonstationary. In the special case of $D(\boldsymbol{s}, t) \equiv 1$ for all $\boldsymbol{s} \in \mathcal{R}$ and $t \in \mathcal{T}$, (2.1) reduces to a class of stationary, but still nonseparable, spatio-temporal covariance functions, which was introduced by Cressie and Huang (1999). Furthermore, (2.1) is separable only when $\theta_3 = 1$ and $D(\boldsymbol{s}, t)$ is separable for all $\boldsymbol{s} \in \mathcal{R}$ and $t \in \mathcal{T}$.

Let $\gamma_{n,k}(\cdot, \cdot; \boldsymbol{\theta}) = \partial \gamma_n(\cdot, \cdot; \boldsymbol{\theta})/\partial \theta_k$ and $\gamma_{n,kk'}(\cdot, \cdot; \boldsymbol{\theta}) = \partial^2 \gamma_n(\cdot, \cdot; \boldsymbol{\theta})/\partial \theta_k \partial \theta_{k'}$ denote the first- and second-order partial derivatives of $\gamma_n(\cdot, \cdot; \boldsymbol{\theta})$, respectively, with respect to θ_k and $\theta_{k'}$ for $1 \le k, k' \le q$. We consider the following additional conditions for developing the locally stationary parametric spatio-temporal covariance functions.

(LS.4). The covariance function $\gamma_n(\cdot, \cdot; \boldsymbol{\theta})$ is bounded and is twice continuously differentiable with respect to $\boldsymbol{\theta}$ in an open set.

(LS.5). There exist two positive nonincreasing functions γ_2 and γ_3 with $\int_0^\infty u^{d-1} \gamma_2(u) du < \infty$ and $\int_0^\infty \gamma_3(u) du < \infty$ such that $\max\{|\gamma_{n,k}((\boldsymbol{s}, t), (\boldsymbol{s} + \boldsymbol{u}_1/A_n, t + u_2/B_n))|, |\gamma_{n,kk'}((\boldsymbol{s}, t), (\boldsymbol{s} + \boldsymbol{u}_1/A_n, t + u_2/B_n))|\} \le \gamma_2(\|\boldsymbol{u}_1\|)\gamma_3(|u_2|)$ for all n and $\|\boldsymbol{u}_1\|, |u_2| \in [0, \infty)$ with $(\boldsymbol{s}, t), (\boldsymbol{s} + \boldsymbol{u}_1/A_n, t + u_2/B_n) \in \mathcal{R} \times \mathcal{T}$ and $1 \le k, k' \le q$.

In the above, (LS.4) is a standard assumption to ensure the smoothness of the covariance function, whereas (LS.5) restricts the decay rates of the first- and second-order partial derivatives of the covariance function with respect to covariance parameters by spatial lag and temporal lag.

With $A_n = \varrho_{1,n}$ and $B_n = \varrho_{2,n}$, we establish that the generalized spatio-temporal Matérn covariance function (2.1) satisfies (LS.1)–(LS.5).

Proposition 1. *Let $D(\boldsymbol{s}, t)$ be some positive known function with $D(\mathbf{0}, 0) = 1$ and $|D(\boldsymbol{s}, t) - D(\boldsymbol{s}', t')| \le \widetilde{C}_1 \|\boldsymbol{s} - \boldsymbol{s}'\| + \widetilde{C}_2 |t - t'|$ for all $(\boldsymbol{s}, t), (\boldsymbol{s}', t') \in \mathcal{R} \times \mathcal{T}$, where $\widetilde{C}_1, \widetilde{C}_2 > 0$ are constants. Then the generalized spatio-temporal Matérn covariance function* (2.1) *satisfies conditions (LS.1)–(LS.5).*

The proof of Proposition 1 is given in Appendix A.1. In general, if $D(\boldsymbol{s}, t)$ is parametric and twice continuously differentiable with respect to the parameters, we can incorporate those parameters into $\boldsymbol{\theta}$ and Proposition 1 will still hold. In addition, it can be shown that a class of generalized exponential covariance functions satisfies (LS.1)–(LS.5) (see details in Appendix A.2).

3 Spatio-Temporal Expanding Distance Asymptotic Framework

We now develop a novel spatio-temporal asymptotic framework under which the asymptotic properties of statistical inference can be investigated.

We consider a fixed spatial domain with continuous spatial indexes and a fixed temporal domain with continuous temporal indexes. Without loss of generality, we assume that the spatial domain is $\mathcal{R} = [0,1]^d$ and the temporal domain is $\mathcal{T} = [0,1]$ at all stages. We further assume that at stage n, N_n spatio-temporal sampling points are observed at $({}^n\boldsymbol{s}_1, {}^nt_1), \ldots, ({}^n\boldsymbol{s}_{N_n}, {}^nt_{N_n})$, where N_n tends to infinity as $n \to \infty$. For ease of notation, henceforth we suppress n in the left superscript of $({}^n\boldsymbol{s}_i, {}^nt_i)$.

We denote the smallest distance between the jth sampling points and the other sampling points in space and time as $\delta_{j,n} = \min\{\|\boldsymbol{s}_i - \boldsymbol{s}_j\| : 1 \leq i \leq N_n, \boldsymbol{s}_i \neq \boldsymbol{s}_j\}$ and $\zeta_{j,n} = \min\{|t_i - t_j| : 1 \leq i \leq N_n, t_i \neq t_j\}$, respectively. Let $\delta_n = \max_{1\leq j\leq N_n} \delta_{j,n}$ and $\zeta_n = \max_{1\leq j\leq N_n} \zeta_{j,n}$ denote the maximum smallest distance in space and in time, respectively.

We assume that, for all n,

(A.1). $\delta_n / \min_{1\leq j\leq N_n} \delta_{j,n} \leq c_1$,

(A.2). $\zeta_n / \min_{1\leq j\leq N_n} \zeta_{j,n} \leq c_2$,

(A.3). $\delta_n^d A_n^d \zeta_n B_n \geq c_3$,

where c_1, c_2 and c_3 are some positive constants independent of n. Here, we refer to (A.1)–(A.3) as an (A_n, B_n)-rate spatio-temporal expanding distance (STED) asymptotic framework in a fixed spatio-temporal domain.

For the STED asymptotic framework, (A.1)–(A.2) ensure bounded mesh ratios in both the spatial and temporal domain, whereas (A.3) provides a lower bound of δ_n and ζ_n. Together with (A.1) and (A.2), the cardinality of a neighborhood of any sampling point is decided by the minimal distances δ_n and ζ_n as well as A_n and B_n. Consequently, the sampling design is such that δ_n and ζ_n cannot decrease too fast or too slowly. That is, the number of observations in the neighborhood cannot be too few with insufficient information for parameter estimation or too many with too much redundant information. Thus, the local stationarity in space and time given in Section 2 is closely connected to the spatio-temporal sampling points in the STED asymptotic framework developed here. An alternative way to ensure a reasonable sampling design is to require the density function of spatial and time to be bounded (Lu and Tjøstheim, 2014).

Let $\widetilde{\mathcal{R}}_n = A_n\mathcal{R}$ and $\widetilde{\mathcal{T}}_n = B_n\mathcal{T}$ denote the spatial and temporal domains of interest at stage n, and the STED asymptotic framework can also be conceptualized through $\widetilde{\mathcal{R}}_n$ and $\widetilde{\mathcal{T}}_n$. That is, the (A_n, B_n)-rate spatio-temporal expanding distance (STED) asymptotic framework includes the following three sampling patterns. First, if $\delta_n^d A_n^d = \mathcal{O}(1)$ and $\zeta_n B_n = \mathcal{O}(1)$, the proposed asymptotic framework is equivalent to an increasing domain asymptotics in both space and time. Second, if $\delta_n^d A_n^d \to 0$ and $\zeta_n B_n \to \infty$, it is equivalent to an increasing domain asymptotics in time and a mixed asymptotics in space. Third, if $\delta_n^d A_n^d \to \infty$ and $\zeta_n B_n \to 0$, it is equivalent to an increasing domain asymptotics in space and a mixed asymptotics in time.

4 Illustration of Theoretical Development

Consider the following spatio-temporal model,

$$y(\boldsymbol{s}, t) = \varepsilon_1(\boldsymbol{s}, t) + \varepsilon_2(\boldsymbol{s}, t), \quad \boldsymbol{s} \in \mathcal{R},\ t \in \mathcal{T}, \tag{4.1}$$

where $\varepsilon_1(\boldsymbol{s}, t)$ is a Gaussian spatio-temporal error process and $\varepsilon_2(\boldsymbol{s}, t)$'s are independently and identically distributed Gaussian errors with mean 0 and variance τ^2, independent of $\varepsilon_1(\boldsymbol{s}, t)$.

The spatio-temporal covariance function of $y(\boldsymbol{s}, t)$ is denoted as $\gamma_n((\boldsymbol{s}, t), (\boldsymbol{s}', t'); \boldsymbol{\theta})$ for $(\boldsymbol{s}, t), (\boldsymbol{s}', t') \in \mathcal{R} \times \mathcal{T}$, where $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters. Recall that the data are observed at N_n points $(\boldsymbol{s}_1, t_1), \ldots, (\boldsymbol{s}_{N_n}, t_{N_n})$ sampled under the STED asymptotic framework (A.1)–(A.3) in Section 3. Let $\boldsymbol{y} = (y(\boldsymbol{s}_1, t_1), \ldots, y(\boldsymbol{s}_{N_n}, t_{N_n}))^\top$ denote an $N_n \times 1$ vector of the response variables and let ${}^n\boldsymbol{\Gamma}(\boldsymbol{\theta}) = [\gamma_n((\boldsymbol{s}_i, t_i), (\boldsymbol{s}_j, t_j); \boldsymbol{\theta})]_{i,j=1}^{N_n}$ denote an $N_n \times N_n$ covariance matrix of $\boldsymbol{y}$. For ease of notation, we omit the stage n in the left superscript of ${}^n\boldsymbol{\Gamma}$. Therefore, the log-likelihood function of $\boldsymbol{\theta}$ under (4.1) is

$$\ell(\boldsymbol{\theta}) = -(N_n/2)\log(2\pi) - (1/2)\log\{\det \boldsymbol{\Gamma}(\boldsymbol{\theta})\} - (1/2)\boldsymbol{y}^\top \boldsymbol{\Gamma}(\boldsymbol{\theta})^{-1}\boldsymbol{y}. \tag{4.2}$$

Denote the maximizer of Eq. 4.2 as $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$. Next, we will establish the asymptotic properties of $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ as an illustration of the STED asymptotic framework (A.1)–(A.3) in Section 3 under local spatio-temporal stationarity (LS.1)–(LS.5) defined in Section 2.

The following regularity conditions are assumed.

(C.1). Let $\boldsymbol{\Gamma}_k = \partial \boldsymbol{\Gamma}/\partial \theta_k$, for some $\iota > 0$, there exist positive constants D_k such that $\|\boldsymbol{\Gamma}_k\|_F^{-2} \le D_k N_n^{-1/2-\iota}$ for $k = 1, \ldots, q$.

(C.2). There exists a constant C^*, such that $\|\boldsymbol{\Gamma}^{-1}\|_2 < C^* < \infty$.

(C.3). Let $t_{kk'} = \mathrm{tr}(\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_k\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{k'})$, for $k, k' = 1, \ldots, q$. For sufficiently large n, $\boldsymbol{A}_n = (a_{kk'})_{k,k'=1}^q$ is nonsingular, where $a_{kk'} = \{t_{kk'}(t_{kk}t_{k'k'})^{-1/2}\}$.

(C.4). There exists a non-singular matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ which satisfies $N_n^{-1}\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) \longrightarrow \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$, as $n \to \infty$, where $\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) = E\left\{-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right\}$.

Here, (C.1) imposes a lower bound on the first-order partial derivatives of the covariance matrix and (C.2) is a constraint on the smallest eigenvalue of the covariance matrix. Both (C.1) and (C.2) are requirements on the covariance function. (C.3) ensures nonsingularity in the limit and the elements of $\widehat{\boldsymbol{\theta}}$ are not asymptotically linearly dependent. (C.4) is a standard condition for information matrix, and will be used together with (A.1)–(A.3) and (C.1)–(C.3) to establish a central limit theorem of $\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. Additionally in Appendix A.3, we provide sufficient conditions for (C.1) and show that the generalized spatio-temporal Matérn and exponential covariance functions satisfy (C.1) under a proper sampling design.

Let $\xrightarrow{p}$ and $\xrightarrow{D}$ denote convergence in probability and in distribution, respectively, as $n \to \infty$. We first establish a result about the spatio-temporal sampling design, which is fundamental for establishing the asymptotic properties of locally stationary processes. The proof of Theorem 1 is given in Appendix A.4.

Theorem 1. *Under (A.1)–(A.3), (LS.1)–(LS.5), and (C.1)–(C.3), we have,*

$$\left.\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{D} N(0, \boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}_0)) \quad \textit{and} \quad \left.\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{p} \boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}_0).$$

Theorem 1 establishes a central limit theorem of $\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, and the convergence of $\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$ at the true value $\boldsymbol{\theta}_0$ for local spatio-temporal stationary covariance functions. By Theorem 1 above, Theorem 1 of Sweeting (1980), and Theorem 2 of Mardia and Marshall (1984), we have the following asymptotic results.

Theorem 2. *Under (A.1)–(A.3), (LS.1)–(LS.5), and (C.1)–(C.4), there exists, with probability tending to one, a local maximizer ${}^n\widehat{\boldsymbol{\theta}}$ of $\ell(\boldsymbol{\theta})$ such that $\|{}^n\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = \mathcal{O}_p(N_n^{-1/2})$. Moreover, the local maximizer ${}^n\widehat{\boldsymbol{\theta}}$ is asymptotic normal; as $n \to \infty$, $N_n^{1/2}({}^n\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)^{-1})$.*

Theorem 2 established consistency and asymptotic normality under the proposed spatio-temporal framework, and can be used as a guideline of parameter estimation for spatio-temporal datasets. Although spatio-temporal data can be viewed as extending spatial data by adding a temporal domain, the extension is usually not straightforward and a careful examination is often needed. For asymptotic framework, it is well known that there are three asymptotic frameworks in spatial statistics: increasing domain asymptotics, infill domain asymptotics, and mixed asymptotics. However, this division does not apply directly to spatio-temporal data, as mentioned in Section 3.

In fact, if the proposed spatio-temporal framework is projected to the space domain, it can be any of the above three frameworks. To illustrate the above point, the following example is provided.

Example. *For two integers a, b, and let $\lfloor a/b \rfloor$ and $\langle a/b \rangle$ denote the quotient and reminder of a divided by b. For the nth stage, let $\bar{\boldsymbol{s}}(i_1, i_2) = \left(\frac{i_1}{p_n+1}, \frac{j_2}{p_n+1}\right)$ and $\bar{t}(i_3) = \frac{i_3}{q_n+1}$, for $i_1, i_2 = 1, \ldots, p_n$, and $i_3 = 1, \ldots, q_n$. The ith spatial and temporal observation is $(\boldsymbol{s}_i, t_i) = (\bar{\boldsymbol{s}}(i_1, i_2), \bar{t}(i_3))$, where $i_1 = \langle \langle i/p_n^2 \rangle / p_n \rangle$, $i_2 = \lfloor \langle i/p_n^2 \rangle \rfloor + 1$, and $i_3 = \lfloor i/p_n^2 \rfloor + 1$, for $i = 1, \ldots, p_n^2 q_n$.*

Next, we show the above spatio-temporal framework can be any of the three spatial asymptotics framework, when it is projected to the spatial domain. First, it can be calculated that the density of the projected spatial locations is p_n^2/A_n^2. In this example, we consider the case that q_n is bounded and p_n increases at the rate of $N_n^{1/2}$. If A_n is bounded and B_n increases at the rate of N_n, the resulting density is N_n and the framework is the infill domain framework. If A_n increases at the rate of $N_n^{1/2}$ and B_n is bounded, the resulting framework is the increasing domain. If A_n increases at the rate of N_n^{α} with $\alpha \in (0, 1/2)$, the resulting framework is the mixed asymptotics.

It is known that some estimates of the covariance function are not consistent under the infill asymptotics. Theorem 2 suggests that even for these datasets, when the temporal dimension is added following the proposed spatio-temporal framework, the consistency and asymptotic normality of parameter estimates can be achieved. On the other hand, if the temporal correlation is ignored, and the spatio-temporal data are treated as if they were spatial data, the resulting estimates of parameter can be inconsistent. Besides the spatio-temporal framework, it is also important to specify the proper spatio-temporal covariance functions. Theorem 2 ensures that for spatio-temporal covariance functions satisfying (LS.1)–(LS.5), consistency and asymptotic normality of the parameter estimates are ensured, which include locally stationary covariance functions, as well as stationary covariance functions (Cressie and Huang, 1999; Gneiting, 2002). In particular, by Proposition 1, Theorem 2 holds for the generalized spatio-temporal Matérn covariance function. In the following section, we provide a simulation study that suggests sound finite-sample properties.

5 Numerical Examples

5.1. Simulation Study We conduct a simulation study to investigate the finite sample performance of $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ in Section 4. First, N_s sampling locations, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_s}$, are generated within the spatial domain $[0, 1]^2$. At

Table 1: Sample mean, sample standard deviation (SD), average information-based standard deviation (SDm) of covariance parameters with $N_n = 806, 1644, 2449$

		$N_n = 806$			$N_n = 1644$			$N_n = 2449$		
	Truth	Mean	SD	SDm	Mean	SD	SDm	Mean	SD	SDm
COV-1										
σ^2	9.0	8.973	0.525	0.529	9.012	0.373	0.385	8.974	0.317	0.313
c	0.2	0.205	0.078	0.077	0.197	0.047	0.049	0.193	0.044	0.041
c_s	1.0	1.023	0.197	0.200	1.008	0.123	0.122	0.994	0.097	0.095
c_t	1.0	1.028	0.220	0.204	1.008	0.133	0.135	0.986	0.108	0.109
COV-2										
σ^2	9.0	8.941	1.528	1.555	9.093	1.100	1.103	9.085	0.865	0.881
c	0.2	0.233	0.120	0.105	0.202	0.060	0.062	0.196	0.052	0.050
a	1.0	0.996	0.117	0.105	1.002	0.071	0.071	1.010	0.058	0.059
b	1.0	1.001	0.139	0.135	1.005	0.089	0.088	1.008	0.069	0.070
d	1.0	1.018	0.240	0.238	0.999	0.165	0.162	0.992	0.127	0.129
COV-3										
σ^2	9.0	9.066	2.358	2.357	9.272	1.715	1.704	9.264	1.392	1.419
c	0.2	0.248	0.156	0.142	0.205	0.077	0.077	0.195	0.062	0.062
a	1.0	0.997	0.115	0.102	1.000	0.068	0.068	1.009	0.055	0.057
b	1.0	1.001	0.134	0.131	1.004	0.087	0.084	1.008	0.067	0.068
d	0.5	0.511	0.220	0.224	0.500	0.157	0.151	0.491	0.116	0.120
e	0.5	0.514	0.227	0.221	0.499	0.142	0.146	0.486	0.120	0.118
f	0.5	0.526	0.196	0.201	0.491	0.153	0.152	0.495	0.116	0.115

each sampling location, we consider time points $t_1, \dots, t_{N_t}$, where $t_i = (i - 1/2)/1000$ for $i = 1, \dots, 1000$, and each time point has a 0.04 probability of being sampled. Here, we set $\varrho_{1,n} = \sqrt{N_s}/2$ and $\varrho_{2,n} = N_t$. The spatio-temporal sampling points are generated once and remain fixed throughout the simulation study. We consider $N_s = 20, 40, 60$ sampling locations and the corresponding sample sizes are $N_n = 806, 1644, 2449$, respectively.

The spatio-temporal process $\varepsilon(\boldsymbol{s}, t)$ is generated from a zero-mean Gaussian process with one of three types of covariance functions. The first type is an exponential spatio-temporal covariance function

$$\gamma_n\{\varepsilon(\boldsymbol{s}_i, t_i), \varepsilon(\boldsymbol{s}_j, t_j)\} = \begin{cases} \sigma^2(1-c)\exp\{-\varrho_{1,n}\|\boldsymbol{s}_i - \boldsymbol{s}_j\|/c_s - \varrho_{2,n}|t_i - t_j|/c_t\}, & \text{if } i \neq j; \\ \sigma^2, & \text{if } i = j, \end{cases}$$

where, σ^2 is the variance of the error process, $c \in [0,1]$ is a nugget proportion such that $c\sigma^2$ is the nugget effect, and c_s and c_t are the positive range parameters in space and time, respectively. When there is only one spatial sampling location,, the covariance function is the same as an AR(1) model in time series. We set $\sigma^2 = 9.0,\ c = 0.2,\ c_s = 1$ and $c_t = 1$. The resulting spatio-temporal covariance function is stationary and separable in space and time, and is referred as `COV-1`.

The second type is a generalized spatio-temporal Matérn covariance function given in (2.1). We let the smoothness parameter be $\nu = 1/2$ and the separability parameter be $\theta_3 = 1$. Then, Eq. 2.1 is simplified to

$$\gamma_n\{\varepsilon(\boldsymbol{s}_i,t_i),\varepsilon(\boldsymbol{s}_j,t_j)\} = \begin{cases} D(\boldsymbol{s}_i,t_i)D(\boldsymbol{s}_j,t_j)\frac{\sigma^2}{(a^2|\varrho_{2,n}(t_i-t_j)|^2+1)^{3/2}}\exp\{-b\varrho_{1,n}\|\boldsymbol{s}_i-\boldsymbol{s}_j\|\}, & \text{if } i \neq j; \\ D(\boldsymbol{s}_i,t_i)D(\boldsymbol{s}_j,t_j)\sigma^2 + c\sigma^2, & \text{if } i = j. \end{cases} \tag{5.1}$$

Here, σ^2 is the variance of the error process, $c \in [0,1]$ is a nugget proportion such that $c\sigma^2$ is the nugget effect, and the range parameters in space and time are a and b, respectively. The nonstationarity of the covariance function is induced by $D(\boldsymbol{s}_i,t_i) = dt_i + 1$ with a change of variance over time. We set $\sigma^2 = 9, c = 0.2, a = 1, b = 1$, and $d = 1$. The resulting spatio-temporal covariance function is nonstationary but separable in space and time, and is referred as `COV-2`. For the third type of covariance functions, we also consider (5.1), but let $D(\boldsymbol{s}_i,t_i) = dt_i + es_{1i} + fs_{2i} + 1$. We set $\sigma^2 = 9.0, c = 0.2, a = 1, b = 1, d = 0.5, e = 0.5$ and $f = 0.5$. The resulting spatio-temporal covariance function is nonstationary and nonseparable in space and time, and is referred as `COV-3`.

For each combination of the sample size N_n and the covariance function, a total of 400 simulated data sets are generated. The sample mean, sample standard deviation (SD), and averaged information matrix based standard deviation (SDm) of covariance parameters are reported in Table 1. For all three types of covariance functions, as the sample size increases, the sample standard deviations of parameter estimates become smaller, supporting the consistency of the parameter estimates in Theorem 2. Moreover, the sample standard deviations of parameter estimates are close to the average information-based standard deviation, as indicated by the asymptotic normality in Theorem 2. For the nonstationary covariance functions (`COV-2` and `COV-3`), the sample mean has a significant bias for the nugget proportion parameter when the number of sampling locations is $N_s = 20$, likely because the covariance functions `COV-2` and `COV-3` are more complex than `COV-1`. As the sample size increases, the bias of the nugget proportion parameter

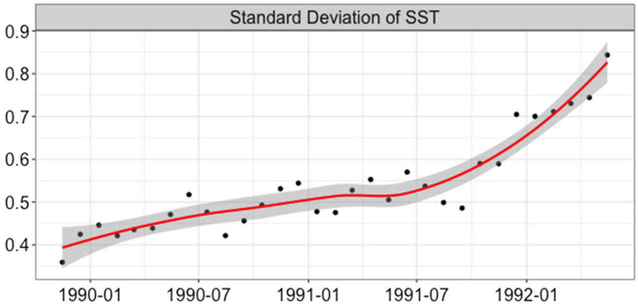

Figure 1: Standard deviations of sea-surface temperature anomalies over all locations from Nov 1989 to May 1992

becomes smaller for both `COV-2` and `COV-3`, which suggests that in practice, a larger sample size would be needed for more complex covariance functions.

5.2. Data Example We apply our method to a sea-surface temperature (SST) data from the National Oceanic and Atmospheric Administration (NOAA). In particular, we use temperature anomalies (i.e., the departure from the average) from November 1989 to May 1992, during which period the standard deviation over all locations is increasing over time, as shown in Fig. 1. This increasing standard deviation over time suggests that a non-stationary covariance function might be needed. In the analysis, we use a random sample of 1000 observations and hold out 10% for modeling evaluation.

We first remove the nonzero mean trend of SST over time using a smoothing spline. Three covariance functions are considered, including the exponential covariance function, referred as `COV-1`, and generalized spatio-temporal Matérn covariance function with two choices of $D(\boldsymbol{s}, t)$, $D_1(\boldsymbol{s}, t) \equiv 1$ and $D_2(\boldsymbol{s}, t) = dt + 1$, referred as `COV-2` and `COV-3`, respectively. The predictive performance is evaluated by the mean squared prediction error (MSPE)

$$\text{MSPE} = N_{\text{test}}^{-1} \sum_{i=1}^{N_{\text{test}}} (y_{i,\text{test}} - \widetilde{y}_{i,\text{test}})^2,$$

where $y_{i,\text{test}}$ is the ith observation in the test set, $\widetilde{y}_{i,\text{test}}$ is the predicted value at the ith observation of the test set, and N_{test} is the total number of observations in the test set.

We rescale the time to the unit interval and use an exploratory semi-variogram analysis to determine the order of range parameter in space and time. The preliminary analyses suggest that SST anomalies have correlation

Table 2: Parameter estimation and their standard deviations (SD) for models `COV-1`, `COV-2`, and `COV-3`, with mean squared prediction error (MSPE)

	`COV-1`		`COV-2`		`COV-3`	
	`estimate`	`SD`	`estimate`	`SD`	`estimate`	`SD`
$\varrho_{1,n} = 0.1$, $\varrho_{2,n} = 100$						
σ^2	0.294	0.022	0.305	0.023	0.158	0.023
r_s	1.266	0.156	0.755	0.09	0.761	0.089
r_t	11.028	1.547	0.116	0.01	0.118	0.01
d	—	—	—	—	0.697	0.157
MSPE	0.1113		0.1076		0.1074	
σ^2	0.294	0.022	0.305	0.023	0.158	0.023
r_s	0.633	0.078	1.51	0.179	1.521	0.179
r_t	5.514	0.774	0.233	0.021	0.237	0.021
d	—	—	—	—	0.697	0.157
MSPE	0.1113		0.1076		0.1074	

ranges on the order 20° in space and 0.01 in time. In practice, we choose $\varrho_{1,n}$ and $\varrho_{2,n}$ such that the ranges of the rescaled spatio-temporal space are in a normal range. In Table 2, we let $\varrho_{1,n} = 0.1$ and $\varrho_{2,n} = 100$. The parameter estimates of the covariance parameters and MSPE are summarized in Table 2, showing that both `COV-2` and `COV-3` have better performance compared to `COV-1` in terms of prediction accuracy. Although the range parameters in `COV-2` and `COV-3` are similar, the nonzero coefficient d in $D_2(\boldsymbol{s}, t)$ suggests that a nonstationary covariance function might be more appropriate. The second part of Table 2 gives results for a different rescaling choice of $\varrho_{1,n} = 0.05$ and $\varrho_{2,n} = 50$. It can be seen that $\varrho_{1,n}\widehat{r}_s$ and $\varrho_{2,n}\widehat{r}_t$ stay the same, which are actually the estimates of range parameter in the original domain. Moreover, MSPE stays the same, and therefore, in practice, the result is not sensitive to the choices of $\varrho_{1,n}$ and $\varrho_{2,n}$.

6 Discussion

In this paper, we introduced the (A_n, B_n)-rate spatio-temporal expanding distance (STED) asymptotic framework in a fixed spatio-temporal domain, which enables us to study asymptotic property of spatio-temporal covariance functions. The STED asymptotic framework is quite flexible and includes three different sampling patterns as mentioned in Section 2. If the framework is formulated in the spatio-temporal domain $\widetilde{\mathcal{R}}_n = A_n\mathcal{R}$ and $\widetilde{\mathcal{T}}_n = B_n\mathcal{T}$, the corresponding covariance functions are independent of the stage n. That is, for ${}^n\widetilde{\boldsymbol{s}}_i, {}^n\widetilde{\boldsymbol{s}}_j \in \widetilde{\mathcal{R}}_n$ and ${}^n\widetilde{t}_i, {}^n\widetilde{t}_j \in \widetilde{\mathcal{T}}_n$, the spatio-temporal covariance function $\gamma(({}^n\widetilde{\boldsymbol{s}}_i, {}^n\widetilde{t}_i), ({}^n\widetilde{\boldsymbol{s}}_j, {}^n\widetilde{t}_j)) = \mathrm{Cov}(Y({}^n\widetilde{\boldsymbol{s}}_i, {}^n\widetilde{t}_i), Y({}^n\widetilde{\boldsymbol{s}}_j, {}^n\widetilde{t}_j))$ does not depend on the stage n. Under the STED framework, the domains of interest $\mathcal{R}$ and $\mathcal{T}$ are fixed, while the covariance function changes across stages. Under an alternative framework, the domains of interest $\widetilde{\mathcal{R}}_n$ and $\widetilde{\mathcal{T}}_n$ change across stages, while the covariance functions are fixed. These two frameworks are essentially the same, and similar theoretical results can be obtained under the alternative asymptotic framework.

For the spatio-temporal covariance functions, a class of locally stationary spatio-temporal covariance functions is introduced, encompassing a wide range of spatio-temporal dependence. The generalized spatio-temporal Matérn covariance function serves as an example for locally stationary spatio-temporal covariance functions, and its property is investigated here. In addition, the proposed method can be applied to both separable and nonseparable covariance functions. Tests for separability assumption are possible for spatio-temporal data with repeated measures (Mitchell et al., 2005; 2006; Fuentes, 2006), but it remains an open question for spatio-temporal data without replicates. Our proposed covariance functions can also feature asymmetry if $D(\boldsymbol{s}, t)$ depends on both time and space. For spatio-temporal datasets, tests for stationarity assumption are available (see, e.g., Jun and Genton 2012; Bandyopadhyay et al. 2017), while testing for a particular parametric form of the covariance function, to the best of our knowledge, is not available. We leave these for future research.

Acknowledgements. The authors thank the Editor, the Associate Editor and the referees for their helpful comments. The research of Haonan Wang was partially supported by NSF grants DMS-1737795, DMS-1923142 and CNS-1932413. This work is in part supported by the U.S. Geological Survey under Grant/Cooperative Agreement No. G16AC00344. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Geological Survey.

Appendix A.1: Proof of Proposition 1

PROOF. First, it can be seen that the generalized spatio-temporal Matérn covariance function in Eq. 2.1 is bounded and twice continuously differentiable with respect to $\boldsymbol{\theta}$; thus, (LS.4) is satisfied. Next, we will show that the generalized spatio-temporal Matérn covariance function satisfies (LS.1) and (LS.2).

For any $\boldsymbol{u}_1$ and u_2, define

$$h(\boldsymbol{u}_1, u_2) = \begin{cases} \frac{\theta_3^{d/2} 2^{1-\nu}}{(\theta_1^2 u_2^2+1)^\nu (\theta_1^2 u_2^2+\theta_3)^{d/2}\Gamma(\nu)} m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta})^\nu K_\nu \left\{ m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta}) \right\}, & \text{if } \|\boldsymbol{u}_1\| > 0, \\ \frac{\theta_3^{d/2}}{(\theta_1^2 u_2^2+1)^\nu (\theta_1^2 u_2^2+\theta_3)^{d/2}}, & \text{if } \|\boldsymbol{u}_1\| = 0. \end{cases}$$

For any $\boldsymbol{s}$ and t, let

$$\begin{aligned} & g_n(\boldsymbol{s}' - \boldsymbol{s}, t' - t, \boldsymbol{s}, t) = g(\boldsymbol{u}_1, u_2, \boldsymbol{s}, t) \\ & = \begin{cases} D(\boldsymbol{s}, t)^2 \sigma^2 h(\boldsymbol{u}_1, u_2), & \text{if } \|\boldsymbol{u}_1\| > 0 \text{ or } |u_2| > 0, \\ D(\boldsymbol{s}, t)^2 \sigma^2 + \tau^2, & \text{otherwise.} \end{cases} \end{aligned}$$

Then, $\gamma_n((\boldsymbol{s}, t), (\boldsymbol{s}, t)) = g(\boldsymbol{0}, 0, \boldsymbol{s}, t)$. For all $(\boldsymbol{s}, t), (\boldsymbol{s} + \boldsymbol{u}_1/\varrho_{1,n}, t + u_2/\varrho_{2,n}) \in \mathcal{R} \times \mathcal{T}$ with $\|\boldsymbol{u}_1\| > 0$ or $|u_2| > 0$, we have

$$\begin{aligned} & |\gamma_n((\boldsymbol{s}, t), (\boldsymbol{s}', t')) - g(\boldsymbol{u}_1, u_2, \boldsymbol{s}, t)| = D(\boldsymbol{s}, t) h(\boldsymbol{u}_1, u_2) \sigma^2 |D(\boldsymbol{s}', t') - D(\boldsymbol{s}, t)| \\ & \leq D(\boldsymbol{s}, t) h(\boldsymbol{u}_1, u_2) \sigma^2 (\widetilde{C}_1 \|\boldsymbol{s} - \boldsymbol{s}'\| + \widetilde{C}_2 |t - t'|) = \mathcal{O}(\|\boldsymbol{s} - \boldsymbol{s}'\| + |t - t'|) \end{aligned}$$

uniformly since $D(\boldsymbol{s}, t)$ is bounded on $\mathcal{R} \times \mathcal{T}$ and $|h(\boldsymbol{u}_1, u_2)| \leq 1$. Thus, (LS.1) is satisfied.

For $g(\boldsymbol{s}, t)$ defined in (LS.2), we have $g(\boldsymbol{s}, t) = g(\boldsymbol{0}, 0, \boldsymbol{s}, t) = D(\boldsymbol{s}, t)^2 \sigma^2 + \tau^2$. Note that $\boldsymbol{s}' = \boldsymbol{s} + \boldsymbol{u}_1/\varrho_{1,n}$ and $t' = t + u_2/\varrho_{2,n}$, and we have $|g(\boldsymbol{s}, t) - g(\boldsymbol{s}', t')| = |D(\boldsymbol{s}, t)^2 - D(\boldsymbol{s}', t')^2| \sigma^2 = |D(\boldsymbol{s}, t) + D(\boldsymbol{s}', t')||D(\boldsymbol{s}, t) - D(\boldsymbol{s}', t')| \sigma^2 \leq |D(\boldsymbol{s}, t) + D(\boldsymbol{s}', t')|(\widetilde{C}_1 \|\boldsymbol{s} - \boldsymbol{s}'\| + \widetilde{C}_2 |t - t'|) \sigma^2$. Thus, (LS.2) holds by adjusting the constants.

Further, we will show that the generalized spatio-temporal Matérn covariance function (2.1) satisfies (LS.3) and (LS.5). For all $(\boldsymbol{s}, t), (\boldsymbol{s}', t') \in \mathcal{R} \times \mathcal{T}$, we have

$$\gamma_n((\boldsymbol{s}, t), (\boldsymbol{s}', t')) \leq (\max D(\boldsymbol{s}, t))^2 (\sigma^2 + \tau^2) h(\boldsymbol{u}_1, u_2).$$

Thus, to show (LS.3), it suffices to find $\gamma_0(\|\boldsymbol{u}_1\|)$ and $\gamma_1(|u_2|)$ to bound $h(\boldsymbol{u}_1, u_2)$. Moreover, straightforward calculation yields that all first- and second-order partial derivatives of γ_n, denoted by $\gamma_{n,k}$ and $\gamma_{n,kk'}$, can be bounded by the partial derivatives of $h(\boldsymbol{u}_1, u_2)$ up to some constant scales, which enables us to obtain $\gamma_2(\|\boldsymbol{u}_1\|)$ and $\gamma_3(|u_2|)$ in (LS.5).

In addition, we have

$$\begin{aligned} \frac{\partial h(\boldsymbol{u}_1, u_2)}{\partial \theta_1} &= \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\partial h(0, u_2)}{\partial \theta_1} m^\nu K_\nu(m) - h(0, u_2) m^\nu K_{\nu-1}(m) \frac{\partial m}{\partial \theta_1} \right), \\ \frac{\partial h(\boldsymbol{u}_1, u_2)}{\partial \theta_2} &= -\frac{2^{1-\nu}}{\Gamma(\nu)} h(0, u_2) m^\nu K_{\nu-1}(m) \frac{\partial m}{\partial \theta_2}, \end{aligned}$$

$$\frac{\partial h(\boldsymbol{u}_1, u_2)}{\partial \theta_3} = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\partial h(0, u_2)}{\partial \theta_3} m^\nu K_\nu(m) - h(0, u_2) m^\nu K_{\nu-1}(m) \frac{\partial m}{\partial \theta_3} \right),$$

where $m = m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta})$ and

$$\frac{\partial m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta})}{\partial \theta_1} = \frac{m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta}) \theta_1 u_2^2 (\theta_3 - 1)}{(\theta_1^2 u_2^2 + \theta_3)(\theta_1^2 u_2^2 + 1)},$$

$$\frac{\partial m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta})}{\partial \theta_2} = \frac{m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta})}{\theta_2}, \frac{\partial m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta})}{\partial \theta_3} = -\frac{m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta})}{2(\theta_1^2 u_2^2 + \theta_3)},$$

$$\frac{\partial h(0, u_2)}{\partial \theta_1} = -\frac{\theta_1 u_2^2 (2\nu(\theta_1^2 u_2^2 + \theta_3) + d(\theta_1^2 u_2^2 + 1))}{(\theta_1^2 u_2^2 + 1)(\theta_1^2 u_2^2 + \theta_3)} h(0, u_2),$$

$$\frac{\partial h(0, u_2)}{\partial \theta_3} = \frac{d \theta_3^{d/2-1} \theta_1^2 u_2^2}{2(\theta_1^2 u_2^2 + 1)^\nu (\theta_1^2 u_2^2 + \theta_3)^{d/2+1}} = \frac{d \theta_1^2 u_2^2}{2\theta_3(\theta_1^2 u_2^2 + \theta_3)} h(0, u_2).$$

For (LS.3), it can be seen that $m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta}) \leq \max \left\{ \theta_2 \theta_3^{-1/2}, \theta_2 \right\} \|\boldsymbol{u}_1\|$. Thus, we have

$$h(\boldsymbol{u}_1, u_2) \leq \frac{\theta_3^{d/2} 2^{1-\nu} \widetilde{m}(\boldsymbol{u}_1)^\nu K_\nu \left\{ \widetilde{m}(\boldsymbol{u}_1) \right\}}{(\theta_1^2 u_2^2 + 1)^\nu (\theta_1^2 u_2^2 + \theta_3)^{d/2} \Gamma(\nu)} \leq 1,$$

where $\widetilde{m}(\boldsymbol{u}_1) = \max \left\{ \theta_2 \theta_3^{-1/2}, \theta_2 \right\} \|\boldsymbol{u}_1\|$. We can see that, up to some constant scales,

$$h(\boldsymbol{u}_1, u_2) \leq (\widetilde{m}(\boldsymbol{u}_1)^\nu K_\nu \left\{ \widetilde{m}(\boldsymbol{u}_1) \right\}) \left(|u_2|^{-2\nu-d} \right) \equiv \gamma_0(\|\boldsymbol{u}_1\|) \gamma_1(|u_2|)$$

Here, $\gamma_0(\|\boldsymbol{u}_1\|)$ is a linear combination of a polynomial of $\|\boldsymbol{u}_1\|$ with degree ν and a modified Bessel function of the second kind and $\gamma_1(|u_2|)$ is a polynomial of $|u_2|$ with degree $-2\nu - d$.

For (LS.5), we focus on the first-order partial derivatives in detail and omit details for the second-order partial derivatives, as similar arguments can be applied. Straightforward calculation shows that the (absolute value of) partial derivatives of $h(\boldsymbol{u}_1, u_2)$ can be bounded by products of two positive functions, $\widetilde{\gamma}_2(\|\boldsymbol{u}_1\|)$ and $\widetilde{\gamma}_3(|u_2|)$. Moreover, $\widetilde{\gamma}_2(\|\boldsymbol{u}_1\|)$ is a linear combination of a polynomial of $\|\boldsymbol{u}_1\|$ with degree *at least* ν and a modified Bessel function of the second kind, and $\widetilde{\gamma}_3(|u_2|)$ is a polynomial of $|u_2|$ with degree at most $-2\nu - d$.

Since the partial derivatives of $h(\boldsymbol{u}_1, u_2)$ with respect to $\boldsymbol{\theta}$ is continuous in $\|\boldsymbol{u}_1\|$ and $|u_2|$, it is bounded if $\|\boldsymbol{u}_1\|$ and $|u_2|$ are bounded. To show (LS.3) and (LS.5), it suffices to show that, for $k, l > 0$, there exists $M > 0$ such that

(i) $\int_M^\infty u^k K_l(u)du < \infty$,

(ii) $u^k K_l(u)$ is bounded by a nonincreasing function on (M, ∞).

Since $d \geq 1$ and $k > 0$, u^{-2k-d} is bounded on (M, ∞) and $\int_M^\infty u^{-2k-d}du = M^{-2k-d+1}/(2k+d-1) < \infty$. The last two conditions hold. By the property of Bessel function, $K_l(u) \propto e^{-u}u^{-1/2}\{1+\mathcal{O}(1/u)\}$, as $|u| \to \infty$. We can find M_1, M_2 such that $K_l(u) \leq M_1 e^{-u}u^{-1/2}(1 + M_2/u)$, when $u \geq M_2$. So (i) holds since

$$\begin{aligned}\int_{M_2}^\infty u^k K_l(u)du \leq & \int_{M_2}^\infty M_1 u^{k-1/2}e^{-u}(1 + M_2/u)du \\ \leq & \; 2M_1 \int_{M_2}^\infty u^{k-1/2}e^{-u}du < 2M_1\Gamma(k + 1/2) < \infty.\end{aligned}$$

For (ii), we have $u^k K_l(u) \leq M_1 e^{-u}u^{k-1/2}(1+M_2/u) \leq 2M_1 e^{-u}u^{k-1/2}$, when $u \geq M_2$. Since $e^{-u}u^{k-1/2}$ is decreasing on $(k - 1/2, \infty)$, (ii) is satisfied.

Appendix A.2: Generalized Exponential Spatio-temporal Covariance Function

In this section, we show that the following exponential spatio-temporal covariance function used in a simulation study satisfies conditions (LS.1)–(LS.5). The covariance function can be written as

$$\begin{aligned}&\gamma_n((\boldsymbol{s}, t), (\boldsymbol{s}', t'); \boldsymbol{\theta}) \\ = &\begin{cases} D(\boldsymbol{s}, t)D(\boldsymbol{s}', t')\sigma^2 \exp\{-\|\boldsymbol{u}_1\|/c_s - |u_2|/c_t\}, & \text{if } \|\boldsymbol{u}_1\| > 0 \text{ or } |u_2| > 0; \\ D(\boldsymbol{s}, t)D(\boldsymbol{s}', t')\sigma^2 + \tau^2, & \text{otherwise,} \end{cases}\end{aligned}$$

where $\boldsymbol{\theta} = (c_s, c_t, \sigma^2, \tau^2)^\top$ is the vector of spatio-temporal parameters with the scaling parameter in space, $c_s \geq 0$, and the scaling parameter in time, $c_t \geq 0$. In addition, $\boldsymbol{u}_1 = \varrho_{1,n}(\boldsymbol{s} - \boldsymbol{s}')$ is the spatial lag scaled to the spatially expanding domain, and $u_2 = \varrho_{2,n}(t - t')$ is the temporal lag scaled to the temporally expanding domain, where $\varrho_{1,n}$ and $\varrho_{2,n}$ are positive constants. Further, $D(\boldsymbol{s}, t)$ is some fixed positive spatio-temporal function with $D(\boldsymbol{0}, 0) = 1$. Note that $D(\boldsymbol{s}, t)^2\sigma^2 + \tau^2$ is the variance of $Y(\boldsymbol{s}, t)$.

By arguments similar to Section Appendix, we show (LS.1), (LS.2) and (LS.4). For (LS.3), we can see that, for all $(\boldsymbol{s}, t), (\boldsymbol{s}', t') \in \mathcal{R} \times \mathcal{T}$,

$$\begin{aligned}\gamma_n((\boldsymbol{s}, t), (\boldsymbol{s}', t')) \;&\leq\; \{\max D(\boldsymbol{s}, t)\}^2(\sigma^2 + \tau^2) \exp\{-\|\boldsymbol{u}_1\|/c_s\} \exp\{-|u_2|/c_t\} \\ &=\; \gamma_0(\|\boldsymbol{u}_1\|)\gamma_1(|u_2|).\end{aligned}$$

Here, both $\gamma_0(\|\boldsymbol{u}_1\|)$ and $\gamma_1(|u_2|)$ are nonincreasing positive functions. Moreover, we have $\int_0^\infty e^{-u/c_s}du = 1/c_s < \infty$ and $\int_0^\infty e^{-u/c_t}du = 1/c_t < \infty$. Thus, (LS.3) is satisfied.

Further, we have

$$
\begin{aligned}
&\partial\gamma_n/\partial\tau^2 = 1_{\{\|\boldsymbol{u}_1\|=0,|u_2|=0\}},\\
&\partial\gamma_n/\partial\sigma^2 = D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\exp\{-\|\boldsymbol{u}_1\|/c_s - |u_2|/c_t\},\\
&\partial\gamma_n/\partial c_s = D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2\|\boldsymbol{u}_1\|\exp\{-\|\boldsymbol{u}_1\|/c_s - |u_2|/c_t\}/c_s^2,\\
&\partial\gamma_n/\partial c_t = D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2|u_2|\exp\{-\|\boldsymbol{u}_1\|/c_s - |u_2|/c_t\}/c_t^2,\\
&\partial^2\gamma_n/\partial\sigma^2\partial c_s = D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\|\boldsymbol{u}_1\|\exp\{-\|\boldsymbol{u}_1\|/c_s - |u_2|/c_t\}/c_s^2,\\
&\partial^2\gamma_n/\partial\sigma^2\partial c_t = D(\boldsymbol{s},t)D(\boldsymbol{s}',t')|u_2|\exp\{-\|\boldsymbol{u}_1\|/c_s - |u_2|/c_t\}/c_t^2,\\
&\partial^2\gamma_n/\partial c_s\partial c_t = D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2\|\boldsymbol{u}_1\||u_2|\exp\{-\|\boldsymbol{u}_1\|/c_s - |u_2|/c_t\}/(c_s^2c_t^2),\\
&\partial^2\gamma_n/\partial c_s^2 = D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2\|\boldsymbol{u}_1\|\exp\{-\|\boldsymbol{u}_1\|/c_s - |u_2|/c_t\}(\|\boldsymbol{u}_1\|/c_s^4 - 2/c_s^3),\\
&\partial^2\gamma_n/\partial c_t^2 = D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2|u_2|\exp\{-\|\boldsymbol{u}_1\|/c_s - |u_2|/c_t\}(|u_2|/c_t^4 - 2/c_t^3).
\end{aligned}
$$

Here, all the first- and second-order partial derivatives are continuous in $\|\boldsymbol{u}_1\|$ and $|u_2|$ and hence bounded when $\|\boldsymbol{u}_1\|$ and $|u_2|$ are bounded. In addition, they are bounded by a product of two functions $\widetilde{\gamma}_2(\|\boldsymbol{u}_1\|)$ and $\widetilde{\gamma}_3(|u_2|)$, where $\widetilde{\gamma}_2(u)$ and $\widetilde{\gamma}_3(u)$ are products of a polynomial of u and an exponential function of u. (LS.5) is satisfied, since for $k > 0$, u^ke^{-u} is nonincreasing on $[k,\infty)$, and $\int_k^\infty u^ke^{-u}du < \infty$.

Appendix A.3: A Remark on Assumption (C.1)

In this section, we provide two sufficient conditions for (C.1). Further, we will demonstrate that, if $\theta_3 > 1$, the generalized spatio-temporal Matérn covariance function (2.1) satisfies Assumption (C.1). The two sufficient conditions are stated as follows:

(E.1) For $1 \le k \le q$, $|\gamma_{n,k}((\boldsymbol{s},t),(\boldsymbol{s}',t'))|$ satisfies one of the following two conditions: (i) $|\gamma_{n,k}((\boldsymbol{s},t),(\boldsymbol{s},t))| > 0$; (ii) For $\|\boldsymbol{u}_1\|, |u_2| \in [M,\infty)$ for some constant $M > 0$ such that $(\boldsymbol{s},t),(\boldsymbol{s}+\boldsymbol{u}_1/A_n, t+u_2/B_n) \in \mathcal{R}\times\mathcal{T}$, we have $|\gamma_{n,k}((\boldsymbol{s},t),(\boldsymbol{s}+\boldsymbol{u}_1/A_n, t+u_2/B_n))| \ge C_3\exp(-C_4\|\boldsymbol{u}_1\| - C_5|u_2|)$ for all n, where $C_3, C_4, C_5 > 0$ are constants.

(E.2) (i) For any two given positive constants M_1, M_2, there exist M_1' and M_2' with $M_1 < M_1' < \infty$ and $M_2 < M_2' < \infty$ such that $\sum_i\sum_j 1(\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \in [M_1\delta_n, M_1'\delta_n])1(|t_i - t_j| \in [M_2\zeta_n, M_2'\zeta_n]) \ge C_6N_n^{1/2+\iota_1}$ for some $C_6 > 0$ and $\iota_1 > 0$. (ii) $A_n\delta_n = \mathcal{O}(b_n)$ and $B_n\zeta_n = \mathcal{O}(b_n)$, where $b_n = \log\log N_n$.

To see the sufficiency of (E.1) and (E.2), we first note that if $A_n\delta_n = \mathcal{O}(1)$ and $B_n\zeta_n = \mathcal{O}(1)$, then $\|{}^n\boldsymbol{\Gamma}_k\|_F^2 \ge CN_n^{1/2+\iota_1}$ for some $C > 0$. Thus, (C.1) is satisfied with $\iota = \iota_1$. If $A_n\delta_n \to \infty$ or $B_n\zeta_n \to \infty$, by (E.1)–(E.2), we have

$\|^n\mathbf{\Gamma}_k\|_F^2 \geq CN_n^{1/2+\iota_1'}$ for some $C > 0$ and any ι_1' such that $0 < \iota_1' < \iota_1$, so (C.1) is satisfied with $\iota = \iota_1'$.

Next, we will show that the generalized spatio-temporal Matérn covariance function (2.1) satisfies (E.1), when $\theta_3 > 1$. Since $\left|\frac{\partial\gamma_n((\boldsymbol{s},t),(\boldsymbol{s},t))}{\partial\sigma^2}\right| = D(\boldsymbol{s},t)^2 > 0$ and $\left|\frac{\partial\gamma_n((\boldsymbol{s},t),(\boldsymbol{s},t))}{\partial\tau^2}\right| = 1$, $\partial\gamma_n/\partial\sigma^2$ and $\partial\gamma_n/\partial\tau^2$ satisfy (E.1)(i).

Further, we show that $\partial\gamma_n/\partial\theta_i$ satisfies (E.1)(ii) for $i = 1, 2, 3$. Recall that for all $(\boldsymbol{s},t), (\boldsymbol{s}+\boldsymbol{u}_1/\varrho_{1,n}, t+u_2/\varrho_{2,n}) \in \mathcal{R}\times\mathcal{T}$ with $\|\boldsymbol{u}_1\| > 0$ or $|u_2| > 0$, we have

$$
\begin{aligned}
\left|\frac{\partial\gamma_n}{\partial\theta_1}\right| &= \frac{D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2 2^{1-\nu}\theta_1\theta_3^{d/2}u_2^2}{\Gamma(\nu)(\theta_1^2u_2^2+1)^{\nu+1}(\theta_1^2u_2^2+\theta_3)^{d/2+1}}\left\{(\theta_3-1)m^{\nu+1}K_{\nu-1}(m)\right. \\
&\quad \left. + \left(2\nu(\theta_1^2u_2^2+\theta_3)+d(\theta_1^2u_2^2+1)\right)m^\nu K_\nu(m)\right\}, \\
\left|\frac{\partial\gamma_n}{\partial\theta_2}\right| &= \frac{D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2 2^{1-\nu}\theta_3^{d/2}\{m^{\nu+1}K_{\nu-1}(m)\}}{\Gamma(\nu)\theta_2(\theta_1^2u_2^2+\theta_3)^{d/2}(\theta_1^2u_2^2+1)^\nu}, \\
\left|\frac{\partial\gamma_n}{\partial\theta_3}\right| &= \frac{D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2 2^{-\nu}\theta_3^{d/2-1}\{d\theta_1u_2^2m^\nu K_\nu(m)+\theta_3m^{\nu+1}K_{\nu-1}(m)\}}{\Gamma(\nu)(\theta_1^2u_2^2+\theta_3)^{d/2+1}(\theta_1^2u_2^2+1)^\nu}.
\end{aligned}
$$

Up to some constant scale, we have

$$
\begin{aligned}
\left|\frac{\partial\gamma_n}{\partial\theta_1}\right| &\geq |u_2|^{-2\nu-d-2}m^{\nu+1}K_{\nu-1}(m) + |u_2|^{-2\nu-d}m^\nu K_\nu(m) \\
&\geq |u_2|^{-2\nu-d-2}\left(\theta_2\theta_3^{-1/2}\|\boldsymbol{u}_1\|\right)^{\nu+1}K_{\nu-1}(\theta_2\|\boldsymbol{u}_1\|) \\
&\quad +|u_2|^{-2\nu-d}\left(\theta_2\theta_3^{-1/2}\|\boldsymbol{u}_1\|\right)^{\nu}K_{\nu}(\theta_2\|\boldsymbol{u}_1\|), \\
\left|\frac{\partial\gamma_n}{\partial\theta_2}\right| &\geq |u_2|^{-2\nu-d}m^{\nu+1}K_{\nu-1}(m) \\
&\geq |u_2|^{-2\nu-d}\left(\theta_2\theta_3^{-1/2}\|\boldsymbol{u}_1\|\right)^{\nu+1}K_{\nu-1}(\theta_2\|\boldsymbol{u}_1\|), \\
\left|\frac{\partial\gamma_n}{\partial\theta_3}\right| &\geq |u_2|^{-2\nu-d-2}m^{\nu+1}K_{\nu-1}(m) + |u_2|^{-2\nu-d}m^\nu K_\nu(m) \\
&\geq |u_2|^{-2\nu-d-2}\left(\theta_2\theta_3^{-1/2}\|\boldsymbol{u}_1\|\right)^{\nu+1}K_{\nu-1}(\theta_2\|\boldsymbol{u}_1\|) \\
&\quad +|u_2|^{-2\nu-d}\left(\theta_2\theta_3^{-1/2}\|\boldsymbol{u}_1\|\right)^{\nu}K_{\nu}(\theta_2\|\boldsymbol{u}_1\|),
\end{aligned}
$$

since $\theta_2\theta_3^{-1/2}\|\boldsymbol{u}_1\| \leq m(\boldsymbol{u}_1, u_2; \boldsymbol{\theta}) \leq \theta_2\|\boldsymbol{u}_1\|$. In addition, by property of Bessel function, $K_l(u) \propto e^{-u}u^{-1/2}\{1 + \mathcal{O}(1/u)\}$, as $|u| \to \infty$. We can find M_1, M_2 such that $K_l(u) \geq M_1 e^{-u} u^{-1/2}$, when $u \geq M_2$; thus, (E.1)(ii) follows.

Remark 1. It can be seen that the generalized exponential covariance function also satisfies (E.1). Note that $\left|\frac{\partial \gamma_n((\boldsymbol{s},t),(\boldsymbol{s},t))}{\partial \sigma^2}\right| = D(\boldsymbol{s},t)^2 > 0$ and $\left|\frac{\partial \gamma_n((\boldsymbol{s},t),(\boldsymbol{s},t))}{\partial \tau^2}\right| = 1$, and (E.1)(i) holds. Next, $\partial\gamma_n/\partial\sigma^2$, $\partial\gamma_n/\partial c_s$ and $\partial\gamma_n/\partial c_t$ are positive when $\|\boldsymbol{u}_1\| > 2c_s$ and $|u_2| > 2c_t$ and can be written as linear combinations of products of $|u_2|^j \exp(-a_1|u_2|)$ and $\|\boldsymbol{u}_1\|^k \exp(-a_2\|\boldsymbol{u}_1\|)$ for $j, k \geq 0$ and some constants $a_1, a_2 > 0$. Hence, (E.1)(ii) follows.

Appendix A.4: Proof of Theorem 1

Proof. To prove Theorem 1, it suffices to show that $\|\boldsymbol{\Gamma}\|_2 = \mathcal{O}(1)$, $\|\boldsymbol{\Gamma}_k\|_2 = \mathcal{O}(1)$ and $\|\boldsymbol{\Gamma}_{kk'}\|_2 = \mathcal{O}(1)$, for all $k, k' = 1, \ldots, q$ (Mardia and Marshall, 1984). Note that $\|\boldsymbol{A}\|_2 \leq \|\boldsymbol{A}\|_\infty$ for any positive definite matrix $\boldsymbol{A}$. We only need to show that $\|\boldsymbol{\Gamma}\|_\infty = \mathcal{O}(1)$, $\|\boldsymbol{\Gamma}_k\|_\infty = \mathcal{O}(1)$ and $\|\boldsymbol{\Gamma}_{kk'}\|_\infty = \mathcal{O}(1)$, for all $k, k' = 1, \ldots, q$.

For each i, let $\mathcal{A}_{1,i} = \{j : \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq C_{s,n}\}$ and $\mathcal{A}_{2,i} = \{j : |t_i - t_j| \leq C_{t,n}\}$. Let $a_1 = C_{s,n}/\delta_n$, $a_2 = \delta_n A_n$, $b_1 = C_{t,n}/\zeta_n$ and $b_2 = \zeta_n B_n$. Then,

$$\begin{aligned}
\|\boldsymbol{\Gamma}\|_\infty &= \max_{1\leq i\leq N_n} \sum_{j\in\mathcal{A}_{1,i}\cap\mathcal{A}_{2,i}} \Gamma_{ij} + \max_{1\leq i\leq N_n} \sum_{j\in\mathcal{A}_{1,i}^c\cap\mathcal{A}_{2,i}} \Gamma_{ij} \\
&\quad + \max_{1\leq i\leq N_n} \sum_{j\in\mathcal{A}_{1,i}\cap\mathcal{A}_{2,i}^c} \Gamma_{ij} + \max_{1\leq i\leq N_n} \sum_{j\in\mathcal{A}_{1,i}^c\cap\mathcal{A}_{2,i}^c} \Gamma_{ij} \\
&= (I_1) + (I_2) + (I_3) + (I_4),
\end{aligned}$$

where Γ_{ij} is the (i,j)th entry of $\boldsymbol{\Gamma}$.

Denote $\text{Card}(\mathcal{A})$ as the cardinality of the set $\mathcal{A}$, then

$$\begin{aligned}
(I_1) &\leq \|\boldsymbol{\Gamma}\|_{\max} \cdot \text{Card}(\mathcal{A}_{1,i} \cap \mathcal{A}_{2,i}) \leq \mathcal{O}\left(\frac{C_{s,n}^d C_{t,n}}{\delta_n^d \zeta_n}\right) = \mathcal{O}(a_1^d b_1), \\
(I_2) &\leq \text{Card}(\mathcal{A}_{2,i}) \sum_{m=\lfloor\left(\frac{C_{s,n}A_n}{b}\right)\rfloor} \mathcal{O}\left(\frac{m^{d-1}b^d}{\delta_n^d A_n^d}\right) \max_{mb\leq\|\boldsymbol{u}_1\|\leq(m+1)b} \gamma_0(\|\boldsymbol{u}_1\|) \\
&\leq \mathcal{O}\left(\frac{C_{t,n}}{\zeta_n \delta_n^d A_n^d}\right) \int_{C_{s,n}A_n}^\infty u^{d-1}\gamma_0(u)du = \mathcal{O}(b_1/a_2^d) \int_{a_1a_2}^\infty u^{d-1}\gamma_0(u)du, \\
(I_3) &\leq \text{Card}(\mathcal{A}_{1,i}) \sum_{m=\lfloor\left(\frac{C_{t,n}B_n}{b}\right)\rfloor} \mathcal{O}\left(\frac{b}{\zeta_n B_n}\right) \max_{mb\leq|u_2|\leq(m+1)b} \gamma_1(|u_2|)
\end{aligned}$$

$$\leq \quad \mathcal{O}\left(\frac{C_{s,n}^d}{\delta_n^d \zeta_n B_n}\right) \int_{C_{t,n}B_n}^{\infty} \gamma_1(u)du = \mathcal{O}(a_1^d/b_2) \int_{b_1 b_2}^{\infty} \gamma_1(u)du,$$

$$\begin{aligned}
(I_4) \quad \leq & \sum_{m=\lfloor\left(\frac{C_{s,n}A_n}{b}\right)\rfloor} \mathcal{O}\left(\frac{m^{d-1}b^d}{\delta_n^d A_n^d}\right) \max_{mb\leq\|\boldsymbol{u}_1\|\leq(m+1)b} \gamma_0(\|\boldsymbol{u}_1\|) \times \\
& \sum_{m'=\lfloor\left(\frac{C_{t,n}B_n}{b}\right)\rfloor} \mathcal{O}\left(\frac{b}{\zeta_n B_n}\right) \max_{m'b\leq|u_2|\leq(m'+1)b} \gamma_1(|u_2|) \\
\leq & \ \mathcal{O}\left(\frac{1}{\zeta_n B_n \delta_n^d A_n^d}\right) \int_{C_{s,n}A_n}^{\infty} u\gamma_0(u)du \int_{C_{t,n}B_n}^{\infty} \gamma_1(u)du \\
= & \ \mathcal{O}(1/a_2^d b_2) \int_{a_1 a_2}^{\infty} u^{d-1}\gamma_0(u)du \int_{b_1 b_2}^{\infty} \gamma_1(u)du.
\end{aligned}$$

To show $\|\boldsymbol{\Gamma}\|_\infty = \mathcal{O}(1)$, it suffices to show that

(i) $a_1^d b_1 \in [C_1, C_2]$ for some constants $C_1, C_2 > 0$,

(ii) $\mathcal{O}(1/a_1^d a_2^d) \int_{a_1 a_2}^{\infty} u^{d-1}\gamma_0(u)du = \mathcal{O}(1)$,

(iii) $\mathcal{O}(1/b_1 b_2) \int_{b_1 b_2}^{\infty} \gamma_1(u)du = \mathcal{O}(1)$.

Let $C_{s,n} = 1/A_n$ and $C_{t,n} = 1/B_n$. By (A.3), $a_1^d b_1 = (\delta_n^d A_n^d \zeta_n B_n)^{-1} \leq c_3^{-1} = \mathcal{O}(1)$, the above requirements are fulfilled. By (LS.5),

$$\max\{|\gamma_{n,k}((\boldsymbol{s},t),(\boldsymbol{s}',t');\boldsymbol{\theta})|, |\gamma_{n,kk'}((\boldsymbol{s},t),(\boldsymbol{s}',t');\boldsymbol{\theta})|\} \leq \gamma_2(0)\gamma_3(0),$$

uniformly for all n and $1 < k, k' < q$. Thus, we have $\|\boldsymbol{\Gamma}\|_2 = \mathcal{O}(1)$, and similar arguments can be applied to show that $\|\boldsymbol{\Gamma}_k\|_2 = \mathcal{O}(1)$ and $\|\boldsymbol{\Gamma}_{kk'}\|_2 = \mathcal{O}(1)$.

Together with (C.1)–(C.3) and by Theorem 1 of Sweeting (1980), we have the result of Theorem 1.

References

BANDYOPADHYAY, S. and RAO, S.S. (2017). A test for stationarity for irregularly spaced spatial data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**, 95–123.

BANDYOPADHYAY, S., JENTSCH, C. and SUBBA RAO, S. (2017). A spectral domain test for stationarity of spatio-temporal data. *J. Time Ser. Anal.* **38**, 326–351.

CHANG, J., GUO, B. and YAO, Q. (2015). High dimensional stochastic regression with latent factors, endogeneity and nonlinearity. *J. Econ.* **189**, 297–312.

CHU, T., ZHU, J. and WANG, H. (2011). Penalized maximum likelihood estiamtion and variable selection in geostatistics. *Ann. Stat.* **39**, 2607–2625.

CRESSIE, N. (2018). Mission co$_2$ntrol: a statistical scientist's role in remote sensing of atmospheric carbon dioxide (with discussion). *J. Am. Stat. Assoc.* **113**, 152–168.

CRESSIE, N. and HUANG, H.C. (1999). Classes of nonseparable spatio-temporal stationary covariance functions. *J. Am. Stat. Assoc.* **94**, 1330–1340.

CRESSIE, N. and LAHIRI, S. (1993). The asymptotic distribution of REML estimators. *J. Multivar. Anal.* **45**, 217–233.

CRESSIE, N. and WIKLE, C.K. (2011). *Statistics for Spatio-Temporal Data.* Wiley, New York.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Stat.* **25**, 1–37.

DAHLHAUS, R. (2012). *Locally stationary processes, 30.* Elsevier, Amsterdam, p. 351-413.

FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.

FUENTES, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* **89**, 197–210.

FUENTES, M. (2006). Testing for separability of spatial–temporal covariance functions. *Journal of Statistical Planning and Inference* **136**, 447–466.

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

GELFAND, A.E., DIGGLE, P.J., FUENTES, M. and GUTTORP, P. (2010). *Handbook of Spatial Statistics.* Boca RatonL, Chapman & Hall/CRC.

GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Am. Stat. Assoc.* **97**, 590–600.

GUINNESS, J. and FUENTES, M. (2015). Likelihood approximations for big nonstationary spatial temporal lattice data. *Stat. Sin.* **25**, 329–349.

HALL, P. and PATIL, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Theory Relat. Fields* **99**, 399–424.

HIGDON, D. (1998). A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environ. Ecol. Stat.* **5**, 173–190.

HSING, T., BROWN, T. and THELEN, B. (2016). Local intrinsic stationarity and its inference. *Ann. Stat.* **44**, 2058–2088.

JUN, M. and GENTON, M.G. (2012). A test for stationarity of spatio-temporal random fields on planar and spherical domains. *Stat. Sin.* **22**, 1737–1764.

KUUSELA, M. and STEIN, M.L. (2018). Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proceedings of the Royal Society A* **474**, 20180400.

LAHIRI, S. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhya, Series A* **65**, 356–388.

LIANG, X., ZOU, T., GUO, B., LI, S., ZHANG, H., ZHANG, S., HUANG, H. and CHEN, S.X. (2015). Assessing Beijing's PM2.5 pollution: severity, weather impact, apec and winter heating. *Proceedings of the Royal Society A: Mathematical, Phys. Eng. Sci.* **471**, 20150257.

LOH, W.L. (2005). Fixed-domain asymptotics for a subclass of Matérn-type gaussian random fields. *Ann. Stat.* **33**, 2344–2394.

LU, Z. and TJØSTHEIM, D. (2014). Nonparametric estimation of probability density functions for irregularly observed spatial data. *Journal of the American Statistical Association* **109**, 1546–1564.

LUDWIG, G., CHU, T., ZHU, J., WANG, H. and KOEHLER, K. (2017). Static and roving sensor data fusion for spatio-temporal hazard mapping with application to occupational exposure assessment. *The Annals of Applied Statistics* **11**, 139–160.

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

MITCHELL, M.W., GENTON, M.G. and GUMPERTZ, M.L. (2005). Testing for separability of space–time covariances. *Environmetrics: The official journal of the International Environmetrics Society* **16**, 819–831.

MITCHELL, M.W., GENTON, M.G. and GUMPERTZ, M.L. (2006). A likelihood ratio test for separability of covariances. *J. Multivar. Anal.* **97**, 1025–1043.

PACIOREK, C.J. and SCHERVISH, M.J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17**, 483–506.

PORCU, E., ALEGRIA, A. and FURRER, R. (2018). Modeling temporally evolving and spatially globally dependent data. *Int. Stat. Rev.* **86**, 344–377.

SHERMAN, M. (2011). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties.* Wiley, West Sussex.

STEIN, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York.

STEIN, M.L. (2005). Space-time covariance functions. *J. Am. Stat. Assoc.* **100**, 310–321.

SWEETING, T. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Stat.*, 1375–1381.

VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Stat.* **40**, 2601–2633.

YAO, Q. and BROCKWELL, P.J. (2006). Gaussian maximum likelihood estimation for ARMA models II: spatial processes. *Bernoulli* **12**, 403–429.

YING, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. Stat.* **21**, 1567–1590.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *J. Am. Stat. Assoc.* **99**, 250–261.

ZHOU, Z. and WU, W.B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Stat.* **37**, 2696–2729.

Publisher's Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Tingjin Chu
School of Mathematics and Statistics,
University of Melbourne, Parkville,
Victoria 3010, Australia
E-mail: tingjin.chu@unimelb.edu.au

Jialuo Liu
Haonan Wang
Department of Statistics, Colorado
State University, Fort Collins, CO
80523, USA

Jun Zhu
Department of Statistics and
Department of Entomology,
University of Wisconsin-Madison,
Madison, WI 53706, USA

Paper received: 9 June 2019.