Statistica Sinica Preprint No: SS-2021-0415	
Title	Statistical Inference for Mean Function of Longitudinal Imaging Data over Complicated Domains
Manuscript ID	SS-2021-0415
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0415
Complete List of Authors	Qirui Hu and Jie Li
Corresponding Authors	Jie Li
E-mails	li-j17@tsinghua.org.cn
Notice: Accepted version subject to English editing.	

Statistica Sinica

STATISTICAL INFERENCE FOR MEAN FUNCTION OF LONGITUDINAL IMAGING DATA OVER COMPLICATED DOMAINS

Qirui Hu[1], Jie Li[2]

[1] *Tsinghua University, Beijing 100084, China.*

[2] *Renmin University of China, Beijing 100872, China.*

Abstract: Motivated by longitudinal imaging data possessing inherent spatial and temporal correlation, we propose a novel procedure to estimate its mean function. Functional moving average is applied to depict the dependence among temporally ordered images and flexible bivariate splines over triangulations are utilized to handle the irregular domain of images which is common in imaging studies. Both global and local asymptotic properties of the bivariate spline estimator for mean function are established with simultaneous confidence corridors (SCCs) as a theoretical byproduct. Under some mild conditions, the proposed estimator and its accompanying SCCs are shown to be consistent and oracle efficient as if all images were entirely observed without errors. The finite sample performance of the proposed method through Monte Carlo simulation experiments strongly corroborates the asymptotic theory. The proposed method is further illustrated by analyzing two sea water potential temperature data sets.

Key words and phrases: Bivariate splines, Spatiotemporal, Imaging data, Oracle efficiency,

1. Introduction

Recent years have witnessed a surge in imaging data as digital technology advanced considerably. Imaging data is generated by decomposing the image into many small areas, called pixels, with a value to express its gray scale. Longitudinal imaging data, which is collected through a series of repeated observations of the same subject over some extended time frame, frequently appears in the fields of medicine, meteorology, geography and environmental science, such as continuous observations of tomography imaging or remote sensing images. The analysis of longitudinal imaging data provides new opportunities to detect the dynamic change of one subject over time, but it is always intricate due to the spatial correlation among pixels within a single image and temporal correlation among sequentially ordered images.

Most commonly used method of analyzing longitudinal imaging data concentrates on linear regression models with correlated errors. George and Aban (2015) proposed a linear model with a separable parametric spatiotemporal error structure. Although they found information criteria was highly accurate at choosing spatial and temporal parametric correlation functions, the risk of model misspecification and poor performance in inference remained inevitable. George et al. (2016) described how to use the above model in practice and applied it for longitudinal cardiac imaging study. They restricted that a handful successive images with a small number of spatial locations were collected daily, monthly or even yearly in limited times, but now longitudinal imaging data usually comes in the form of magnitude order greater numbers of spatial (thousands of pixels) and temporal (multiple measures per day or hour) observations. One interesting example is the continued recording of the surface temperature of the Black sea. Hourly sea water potential temperature is recorded on dense regular grids (see Figure 14(a)) every 1/12 degree both longitude and latitude over 360 consecutive hours. This produces 360 sequentially ordered images, each consisting of 6583 pixels, with 4 randomly selected images shown in Figure 15. The ultrahigh dimension of the data poses great threat to unstructured correlation matrices, making the traditional model lose its effect. Therefore, a practical, computationally efficient and theoretically reliable method is urgently called for to analyze such large-scale longitudinal imaging data.

Functional data analysis provides a novel and powerful approach to dealing with imaging data. Instead of imposing spatial structure directly, it views imaging data as realizations of random fields, which naturally captures the spatial correlation among pixels. French and Kokoszka (2020) developed a spatiotemporal sandwich smoother based on radial basis functions and B-splines to fit large spatiotemporal data sets. They involved time dimension in the smoother, which caused additional computational complexity and failure in derivation of the global mean surface, and statistical inference could also not be conducted due to the lack of theory. Kokoszka and Reimherr (2019) reviewed recent developments related to inference for functions defined at spatial locations and considered time series of functions defined at irregularly distributed spatial points or on a grid, namely spatially indexed functional time series. Different from their research object, what we focus on are temporally indexed images, that is longitudinal imaging data with higher dimensions and more complex structures.

From the perspective of functional data analysis, longitudinal imaging data consists of a collection of n temporally ordered images $\{\eta_t(\cdot)\}_{t=1}^n$ on a two-dimensional bounded domain Ω, where Ω can be divided into several disjoint convex sets and the t-th image $\eta_t(\cdot)$ is a continuous stochastic field equal in distribution to a standard field $\eta(\cdot)$. However, the actual observed data is discrete values of a regular grid of pixels from fields $\{\eta_t(\cdot)\}_{t=1}^n$ plus random errors. Since most imaging data is recorded by some automated instruments, we assume the pixel locations are dense regular grids $\boldsymbol{x}_{ij} \in \Omega$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, which forms an M-row array with N_i points in the i-th row, see Figure 14(a) and Figure 18(a). The similar data setting was also considered in Yu et al. (2021). Let $Y_{t,ij} = Y_t(\boldsymbol{x}_{ij})$ be the observation of the t-th image at location $\boldsymbol{x}_{ij}$, then the data set $\{(Y_{t,ij}, \boldsymbol{x}_{ij})\}$, $t = 1, \ldots, n$, $i = 1, \ldots, M$,

$j = 1, \ldots, N_i$, can be modeled as

$$Y_{t,ij} = \eta_t\left(\boldsymbol{x}_{ij}\right) + \sigma\left(\boldsymbol{x}_{ij}\right)\varepsilon_{t,ij}, \tag{1.1}$$

in which $\varepsilon_{t,ij}$ are iid random errors with mean 0 and variance 1 and $\sigma^2(\cdot)$ is the variance function of measurement errors.

In longitudinal imaging data analysis, a fundamental issue lies in the estimation of mean function $m(\cdot)$, defined as $m(\cdot) = \mathbb{E}\{\eta(\cdot)\}$. One challenge is that lots of imaging data is collected over complicated domains even with gaps and holes (see Figure 13), leading to the problem of "leakage"across complex boundary for some traditional smoothing methods, such as tensor product smoothing, kernel smoothing or wavelet smoothing. Bivariate splines on triangulations introduced in Lai and Schumaker (2007) are effective tools to overcome the poor boundary estimation and preserve important features (shape and smoothness) of imaging data. Any two-dimensional geometric domain can be represented as a polygon which is decomposed into triangles through triangulation. Bivariate splines are widely used due to computational ease as well as convenient representation with flexible degrees and various smoothness, see Lai and Wang (2013), Zhou and Pan (2014) and Ferraccioli et al. (2021) for their applications in various statistical areas. Wang et al. (2020) proposed a consistent mean function estimator of imaging data based on bivariate splines over triangulations. One serious limitation is that they restricted images $\{\eta_t(\cdot)\}_{t=1}^{n}$ to be iid copy, apparently not the case of longitudinal imaging data. To model the timely ordered and dependent images, we embed the de-meaned stochastic fields $\xi_t(\cdot)$, defined as $\xi_t(\cdot) = \eta_t(\cdot) - m(\cdot)$, into the functional moving average infinity or FMA(∞) series $\{\xi_t(\cdot)\}_{t=1}^{\infty}$ as in Li and Yang (2022+). Precisely, $\xi_t(\cdot)$ satisfies the following equation:

$$\xi_t(\cdot) = \sum_{t'=0}^{\infty} A_{t'}\zeta_{t-t'}(\cdot), \ t = 0, \pm 1, \pm 2, \ldots \tag{1.2}$$

in which the $A_{t'}$'s are bounded linear operators $\mathcal{L}^2(\Omega) \rightarrow \mathcal{L}^2(\Omega)$, playing the role of scalar coefficients in classic MA(∞) and $\{\zeta_t(\cdot)\}_{t=-\infty}^{\infty}$ are orthonormal zero mean stochastic fields, called strong functional white noises in Bosq (2000). It should be noted that the classic MA(∞) is a broad category, which includes the widely used causal ARMA(p, q), thus including AR(p) and MA(q) as a special case. Actually, lots of stationary functional time series can be approximated by m-dependent series in L^2 sense.

Under the above dependence structure, we propose a bivariate spline estimator for the mean function $m(\cdot)$. It is established in Theorem 2 that the bivariate spline estimator is asymptotically equivalent to the infeasible "oracle" estimator obtained as if all images were totally observed without measurement errors. This oracle efficiency allows for the construction for asymptotically correct SCC of the mean function $m(\cdot)$

under some mild conditions. SCC serves as a vital tool for evaluating the variability and testing global behavior of functions, see Cao et al. (2016), Cao et al. (2012), Choi and Reimherr (2017), Gu et al. (2014), Gu and Yang (2015), Ma et al. (2012), Wang et al. (2020) and Yu et al. (2021) for related theory and applications. Simulation studies suggest that the proposed SCC is computationally efficient with the correct coverage frequency for finite samples.

The rest of the paper is organized as follows. Section 2 describes the functional moving average model and bivariate spline estimator for the mean function. Section 3 states main theoretical results on SCC constructed from bivariate spline estimator. Procedures to implement the proposed SCC are given in Section 4 with details. Section 5 presents the findings of extensive simulation studies. In Section 6, we apply the proposed method to two hourly sea water potential temperature data sets. All figures and tables in simulation and real data application, as well as technical proofs are included in the Supplementary Material.

2. Model and Estimation Method

2.1 Functional moving average model

Denote the covariance function of $\eta(\cdot)$ as $G(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}\left\{\eta(\boldsymbol{x}), \eta\left(\boldsymbol{x}'\right)\right\}$, $\boldsymbol{x}, \boldsymbol{x}' \in \Omega$. The identically distributed random fields $\{\eta_t(\cdot)\}_{t=1}^{n}$ are decomposed as $\eta_t(\cdot) = m(\cdot) + \xi_t(\cdot)$, where each $\xi_t(\boldsymbol{x})$ can be viewed as a small-scale variation of $\boldsymbol{x}$ on the

t-th image, and is assumed to be a strictly stationary L^2 process with $\mathbb{E}\xi_t(\boldsymbol{x}) = 0$ and covariance $G(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}\{\xi_t(\boldsymbol{x}), \xi_t(\boldsymbol{x}')\}$, $\boldsymbol{x}, \boldsymbol{x}' \in \Omega$. Mercer's lemma entails the decomposition of its covariance function $G(\boldsymbol{x}, \boldsymbol{x'}) = \sum_{k=1}^{\infty} \lambda_k \psi_k(\boldsymbol{x})\psi_k(\boldsymbol{x'})$, where $\{\lambda_k\}_{k=1}^{\infty}$ are a series of decreasing positive eigenvalues and $\{\psi_k(\cdot)\}_{k=1}^{\infty}$ are corresponding eigenfunctions, forming an orthogonal basis of $L^2(\Omega)$, such that $\sum_{k=1}^{\infty} \lambda_k < \infty$ and $\int G(\boldsymbol{x}, \boldsymbol{x'})\psi_k(\boldsymbol{x'})d\boldsymbol{x'} = \lambda_k \psi_k(\boldsymbol{x})$.

Then for any $t \in \mathbb{Z}$, the zero-mean field $\xi_t(\boldsymbol{x}), \boldsymbol{x} \in \Omega$, allows general Karhunen-Loève representation $\xi_t(\boldsymbol{x}) = \sum_{k=1}^{\infty} \xi_{tk}\phi_k(\boldsymbol{x})$, in which the rescaled eigenfunctions $\{\phi_k(\cdot)\}_{k=1}^{\infty}$, called functional principle components (FPC), satisfy that $\phi_k = \sqrt{\lambda_k}\psi_k$ and $\int\{\eta(\boldsymbol{x}) - m(\boldsymbol{x})\}\phi_k(\boldsymbol{x})d\boldsymbol{x} = \lambda\xi_k$ for $k \geq 1$. The random coefficients ξ_{tk} are uncorrelated over k, with mean 0 and variance 1, referred to as FPC scores. It is worthy to note that though the sequences $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\phi_k(\cdot)\}_{k=1}^{\infty}$ do exist in mathematics, they are unknown and unobservable in practice, with the detailed estimating procedure given in Section 4.

To make the FMA(∞) model better fit the data structure, operators $A_{t'}$ are assumed to be of diagonal form

$$A_{t'}\left\{\sum_{k=1}^{\infty} c_k\phi_k(\cdot)\right\} = \sum_{k=1}^{\infty} a_{t'k}c_k\phi_k(\cdot),\ a_{t'k} \in \mathbb{R},\ k = 1, 2, \ldots,\ t' = 0, 1, \ldots$$

with arithmetically decaying MA coefficients $|a_{t'k}| < C_a t'^{\rho_a}$ for constants $C_a > 0$

and $\rho_a \in (-\infty, -1)$, which is a rather loose requirement. The strong functional white noises $\{\zeta_t(\cdot)\}_{t=-\infty}^{\infty}$ allows for its own Karhunen-Loève representation $\zeta_t(\cdot) = \sum_{k=1}^{\infty} \zeta_{t,k}\phi_k(\cdot)$, in which $\{\zeta_{t,k}\}_{t=-\infty,k=1}^{\infty,\infty}$ are uncorrelated random variables with mean 0 and variance 1. Together with (1.2), one has

$$\xi_t(\cdot) = \sum_{t'=0}^{\infty} A_{t'} \left\{ \sum_{k=1}^{\infty} \zeta_{t-t',k}\phi_k(\cdot) \right\} = \sum_{t'=0}^{\infty}\sum_{k=1}^{\infty} a_{t',k}\zeta_{t-t',k}\phi_k(\cdot)$$
$$= \sum_{k=1}^{\infty} \left(\sum_{t'=0}^{\infty} a_{t',k}\zeta_{t-t',k} \right) \phi_k(\cdot). \quad (2.3)$$

Note that $\xi_t(\cdot) = \sum_{k=1}^{\infty} \xi_{tk}\phi_k(\cdot)$ absolutely almost surely by Karhunen-Loève expansion, it follows that the FPC score $\xi_{tk} = \sum_{t'=0}^{\infty} a_{t'k}\zeta_{t-t',k}$ almost surely. To ensure that ξ_{tk} has variance 1, we assume $\sum_{t=0}^{\infty} a_{tk}^2 \equiv 1, k = 1, 2, \ldots$, reasonably analogous to what is assumed in numerical MA(∞).

In summary, for $1 \le t \le n, 1 \le i \le M, 1 \le j \le N_i$, raw data $\{(Y_{t,ij}, \boldsymbol{x}_{ij})\}$of FMA($\infty$) can be written as

$$Y_{t,ij} = m(\boldsymbol{x}_{ij}) + \xi_t(\boldsymbol{x}_{ij}) + \sigma(\boldsymbol{x}_{ij})\varepsilon_{t,ij}$$
$$= m(\boldsymbol{x}_{ij}) + \sum_{k=1}^{\infty} \xi_{tk}\phi_k(\boldsymbol{x}_{ij}) + \sigma(\boldsymbol{x}_{ij})\varepsilon_{t,ij}, \quad (2.4)$$

where for $1 \leq t \leq n, k = 1, 2, \ldots$

$$\xi_t(\cdot) = \sum_{k=1}^{\infty} \xi_{tk}\phi_k(\cdot), \quad \xi_{tk} = \sum_{t'=0}^{\infty} a_{t'k}\zeta_{t-t',k} \quad a.s. \tag{2.5}$$

2.2 Bivariate spline estimator

Had the n images $\{\eta_t(\cdot)\}_{t=1}^n$ been entirely observed over Ω, an intuitive estimator for the mean function $m(\cdot)$ in (2.4) is the sample mean

$$\overline{m}(\boldsymbol{x}) = n^{-1}\sum_{t=1}^{n} \eta_t(\boldsymbol{x}), \ \boldsymbol{x} \in \Omega, \tag{2.6}$$

which is infeasible due to the finite pixel grids and measurement errors. However, it does suggest us to replace the unobservable $\eta_t(\cdot)$ with some suitable estimator $\widehat{\eta}_t(\cdot)$ and get the plug-in estimator $\widehat{m}(\cdot) = n^{-1}\sum_{t=1}^{n} \widehat{\eta}_t(\cdot)$. Bivariate splines that are piecewise polynomial functions over a 2D triangulated domain are employed to approximate each image $\eta_t(\cdot)$. In the following, we briefly introduce some elementary knowledge about triangulation techniques and bivariate splines.

Triangulation is a powerful weapon for processing data distributed over difficult domains with complex boundaries and/or interior holes. Denote by T a triangle which is a convex hull of three points not located in one line. A triangulation of Ω is a collection of H triangles $\triangle = \{T_1, \ldots, T_H\}$ with $\Omega = \bigcup_{h=1}^{H} T_h$ provided that any nonempty intersection between a pair of triangles in $\triangle$ is either a shared vertex or a shared edge. Given a triangle $T \in \triangle$, let $|T|$ be its longest edge length, and ρ_T be the radius of the largest disk inscribed in T. The shape parameter of T is defined as the ratio $\pi_T = |T|/\rho_T$. When π_T is small, the triangles are relatively uniform in the sense that all angles of triangles in $\triangle$ are relatively the same. Denote the size of $\triangle$ by $|\triangle| = \max\{|T|, T \in \triangle\}$, namely the length of the longest edge of all triangles in $\triangle$.

For any triangle $T \in \triangle$ and any fixed point $\boldsymbol{x} \in \Omega$, let b_1, b_2 and b_3 be the barycentric coordinates of $\boldsymbol{x}$ relative to T. The Bernstein basis polynomials of degree d relative to triangle T are defined as $B_{ijk}^{T,d}(\boldsymbol{x}) = (i!j!k!)^{-1}d!b_1^i b_2^j b_3^k$, $i + j + k = d$ and used to represent the bivariate splines. For an integer $r \geq 0$, let $C^r(\Omega)$ be the collection of all r-th continuously differentiable functions over Ω. Given $\triangle$, let $S_d^r(\triangle) = \{s \in C^r(\Omega), s|_T \in \mathbb{P}_d(T), T \in \triangle\}$ be a spline space of degree d and smoothness r over $\triangle$, where $s|_T$ is the polynomial piece of spline s restricted on triangle T, and $\mathbb{P}_d$ is the space of all polynomials of degree less than or equal to d. Bivariate splines on the triangulation T are piecewise polynomials defined on T satisfying additional smoothness conditions that the derivatives up to certain degree are continuous.

Let $\{B_\ell\}_{\ell=1}^p$ be the set of degree-d bivariate Bernstein basis polynomials for $S_d^r(\triangle)$ and the vector $\mathbf{B}(\boldsymbol{x}_{ij}) = \{B_1(\boldsymbol{x}_{ij}), \ldots, B_p(\boldsymbol{x}_{ij})\}^\top$. Denote by $\mathbf{X}$ the evaluation matrix of the Bernstein polynomial basis, then $\mathbf{X}$ can be written as

$$\mathbf{X} = \{\mathbf{B}(\boldsymbol{x}_{11}), \ldots, \mathbf{B}(\boldsymbol{x}_{1N_1}), \ldots, \mathbf{B}(\boldsymbol{x}_{MN_M})\}^\top = \left[\{\mathbf{B}(\boldsymbol{x}_{ij})\}_{i=1,j=1}^{M,N_i}\right]^\top \tag{2.7}$$

The t-th unknown random field $\eta_t(\boldsymbol{x})$ can be estimated via bivariate splines by $\eta_t(\boldsymbol{x}) = \mathbf{B}^\top(\boldsymbol{x})\boldsymbol{\gamma}_t$, where $\boldsymbol{\gamma}_t^\top = (\gamma_{t1}, \ldots, \gamma_{tp})$ is the spline coefficient vector. It is shown that the smoothness constraint in the derivative can be expressed by a linear equation system on the coefficient vector $\boldsymbol{\gamma}_t$: $\mathbf{H}\boldsymbol{\gamma}_t = 0$, where $\mathbf{H}$ is a $(p - p_0) \times p$ matrix determined by the smoothness constraints, p_0 is the dimension of $\mathbb{P}_d(T)$ and p is the dimension of $S_d^r(\triangle)$. For a more detailed description of $\mathbf{H}$, please refer to Section B.2 of the Supplementary Material of Yu et al. (2020). Thus $\eta_t(\boldsymbol{x})$ is obtained by the solving the following least square problem

$$\widehat{\eta}_t(\boldsymbol{x}) = \underset{g(\cdot)\in S_d^r(\triangle)}{\arg\min} \sum_{i=1}^{M}\sum_{j=1}^{N_i} \{Y_{t,ij} - g(\boldsymbol{x}_{ij})\}^2, \tag{2.8}$$

subject to $\mathbf{H}\boldsymbol{\gamma}_t = 0$.

To remove the constraint, we consider the QR decomposition of $\mathbf{H}^\top$: $\mathbf{H}^\top = \mathbf{Q}\mathbf{R} = (\mathbf{Q_1}\mathbf{Q_2})\binom{\mathbf{R_1}}{\mathbf{R_2}}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is upper triangular, the submatrix $\mathbf{Q}_1$ is the first m columns of $\mathbf{Q}$, where m is the rank of $\mathbf{H}$, and $\mathbf{R}_2$ is a matrix of zeros. The constraint $\mathbf{H}\boldsymbol{\gamma}_t = 0$ can be ensured by reparametrizing $\boldsymbol{\gamma}_t = \mathbf{Q}_2\boldsymbol{\beta}_t$ for some $\boldsymbol{\beta}_t$, then the minimization problem is converted to a conventional unrestricted problem:

$$\sum_{i=1}^{M}\sum_{j=1}^{N_i}\left\{Y_{t,ij}-\mathbf{B}\left(\boldsymbol{x}_{ij}\right)\mathbf{Q}_2\boldsymbol{\beta}_t\right\}^2. \tag{2.9}$$

Denote $\widetilde{\mathbf{B}}(\boldsymbol{x}) = \mathbf{Q}_2^\top\mathbf{B}(\boldsymbol{x})$, $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{Q}_2$ and $\mathbf{Y}_t = \left(\{Y_{t,ij}\}_{i=1,j=1}^{M,N_i}\right)^\top$. Applying elementary algebra, the solution is given by $\widehat{\boldsymbol{\beta}}_t = \left(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^\top\mathbf{Y}_t$, $\widehat{\boldsymbol{\gamma}}_t = \mathbf{Q}_2\widehat{\boldsymbol{\beta}}_t$. Thus, the estimator of $\eta_t(\cdot)$ is $\widehat{\eta}_t(\boldsymbol{x}) = \widetilde{\mathbf{B}}(\boldsymbol{x})^\top\left(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^\top\mathbf{Y}_t$ and one can estimate the unknown mean function $m(\cdot)$ as

$$\widehat{m}(\cdot) = n^{-1}\sum_{t=1}^{n}\widehat{\eta}_t(\cdot). \tag{2.10}$$

3. Main results

3.1 Technical assumptions

Suppose that Ω is a bounded domain in $\mathbb{R}^2$, for any function g over Ω, denote by $\|g\|_{\infty,\Omega} = \sup_{\boldsymbol{x}\in\Omega}|g(\boldsymbol{x})|$. For $d \geq 0$, the associated Sobolev space is defined by functions with

$$W^{d,\infty}(\Omega) = \left\{g : |g|_{k,\infty,\Omega} < \infty, 0 \leq k \leq d\right\},$$

where $|g|_{k,\infty,\Omega} = \max_{\nu+\mu=k}\|D_x^\nu D_y^\mu g\|_{\infty,\Omega}$ and $D_x^\nu g$ represents the ν-th partial derivative of g with respect to variable x. Also denote a class of Lipschitz continuous functions by $\mathrm{Lip}(\Omega, L) = \{g(\boldsymbol{x}) : |g(\boldsymbol{x}) - g(\boldsymbol{x}')| \leq L\,|\boldsymbol{x} - \boldsymbol{x}'|, \forall \boldsymbol{x}, \boldsymbol{x}' \in \Omega, L > 0\}$.

To study the asymptotic properties of the bivariate spline estimator $\widehat{m}(\cdot)$, we need some technical assumptions.

(A1) The mean function $m(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega)$ for some integer $d \geq 1$.

(A2) The standard deviation function of the measurement errors $\sigma(\cdot) \in \mathrm{Lip}(\Omega, L)$ for some $L > 0$, and there exist some positive constants M_σ, c_G, C_G, such that $\sup_{\boldsymbol{x}\in\Omega} |\sigma(\boldsymbol{x})| \leq M_\sigma, c_G \leq G(\boldsymbol{x}, \boldsymbol{x}) \leq C_G, \boldsymbol{x} \in \Omega$.

(A3) There exists a constant $\theta > 0$, such that as $N \to \infty$, $n = n(N) \to \infty$, $n = \mathcal{O}\left(N^\theta\right)$.

(A4) For $k \geq 1$, $\phi_k(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega)$ with $\sum_{k=1}^{\infty} |\phi_k|_{d+1,\Omega,\infty} < +\infty$ and for some integer $d \geq 1$; for increasing positive integers $\{k_n\}_{n=1}^{\infty}$, as $n \to \infty$, $\sum_{k_n+1}^{\infty} \|\phi_k\|_{\infty,\Omega} = \mathcal{o}(n^{-1/2})$, $\sum_{k=1}^{k_n} |\phi_k|_{d+1,\Omega,\infty} |\triangle|^{d+1} = \mathcal{o}(1)$ and $k_n = \mathcal{O}(n^\alpha)$ for some $\alpha > 0$.

(A5) There are constants $C_0, C_1, C_2 \in (0, +\infty), \gamma_1, \gamma_2 \in (1, +\infty), \beta_1 \in (0, 1/2), \beta_2 \in (0, \omega)$, $I_n \asymp n^\iota$ with $\max\{(-\alpha - 3)/(2\rho_a + 1), -(2r_1 - 2\beta_1 r_1 + 4 + \alpha)/r_1\rho_a\} < \iota < 1$, where $a_n \asymp b_n$ means a_n and b_n are asymptotically equivalent, and iid $N(0,1)$ variables $\{Z_{tk,\zeta}\}_{t=-I_n+1,k=1}^{n,k_n}$, $\{Z_{t,ij,\varepsilon}\}_{t=1,i=1,j=1}^{n,M,N_i}$ such that

$$\mathbb{P}\left\{\max_{1\leq k\leq k_n} \max_{-I_n+1\leq\tau\leq n} \left|\sum_{t=-I_n+1}^{\tau} \zeta_{tk} - \sum_{t=-I_n+1}^{\tau} Z_{tk,\zeta}\right| > C_0 n^{\beta_1}\right\} < C_1 n^{-\gamma_1},$$
$$\mathbb{P}\left\{\max_{1\leq t\leq n} \max_{1\leq\tau\leq N} \left|\sum_{k=1}^{\tau} \varepsilon_{t,f_1(k)f_2(k)} - \sum_{k=1}^{\tau} Z_{t,f_1(k)f_2(k),\varepsilon}\right| > N^{\beta_2}\right\} < C_2 N^{-\gamma_2},$$

where f_1, f_2 are functions $\mathbb{Z}^+ \to \mathbb{Z}^+$ with the following property

$$\left\| \boldsymbol{x}_{f_1(x) f_2(x)} - \boldsymbol{x}_{f_1(y) f_2(y)} \right\|_2 = \mathcal{O}(N^{-1/2}) \tag{$*$}$$

where $x, y \in \{1, \ldots, N\}$, $|x - y| \leq 1$, $f_1(1) = 1, f_1(N) = M, f_2(1) = 1$ or N_1 and $f_2(N) = 1$ or N_M.

(A5') The iid variables $\{\varepsilon_{t,ij}\}_{t=1,i=1,j=1}^{n,M,N_i}$ are independent of $\{\zeta_{tk}\}_{t=1,k=1}^{n,\infty}$. The number of distinct distributions for FPC score white noises $\{\zeta_{tk}\}_{t=1,k=1}^{n,\infty}$ is finite. There exist constants $r_1 > 4 + 2\alpha, r_2 > (2 + \theta)/\omega$ such that for $1 \leq t \leq n, 1 \leq i \leq M, 1 \leq j \leq N_i, 1 \leq k \leq \infty$, $\mathbb{E}\xi_{tk}^{r_1} + \mathbb{E}\varepsilon_{t,ij}^{r_2} < \infty$.

(A6) The triangulations are π-quasi-uniform, that is, there exists a positive constant π such that $(\min_{T\in\triangle} \rho_T)^{-1} |\triangle| \leq \pi$. The smoothness parameter r satisfies $d \geq 3r + 2$ for d in Assumption (A1). The size of triangulations $|\triangle|$ satisfies $|\triangle|^{-1} = N^{\gamma} d_N$ for some $\gamma > 0$, with $d_N + d_N^{-1} = \mathcal{O}\left(\log^{\vartheta} N\right)$ for some $\vartheta > 0$ as $N \to \infty$, and for d in Assumption (A1), θ in Assumption (A3), β_2 in Assumption (A5) and r_1 in Assumption (A5')

$$\frac{\theta}{d+1}\left(\frac{2}{r_1} + \frac{1}{2}\right) < \gamma < 1/2 - \theta/2 - \beta_2.$$

A few comments on the regularity conditions are in order. Assumption (A1) is typical for the bivariate spline smoother in nonparametric estimation literature, which controls the size of the bias of the estimator for $m(\cdot)$ and can be relaxed by only requiring $m(\cdot) \in C^0(\Omega)$ if the imaging data has sharp edges, see Wang et al. (2020). Assumption (A2) ensures the variance function should be uniformly bounded. Assumption (A3) requires that sample size n grows not faster than power θ of the number N of pixels per image. The collective bounded smoothness of the principal components is provided in Assumption (A4). Assumption (A5) presents a strong approximation of estimation errors and the strong white noise $\{\zeta_t(\cdot)\}_{t=-\infty}^{\infty}$, which can be guaranteed by a more elementary Assumption (A5'). Assumption (A6) suggests using more uniform triangulations with smaller shape parameters and specifies the size of triangulations.

Remark 1. The assumptions above are quite mild and can be easily satisfied in many practical situations. One simple and reasonable setup for above parameters d, θ, ω, γ and d_N is as follows: $d = 5$, $\theta = 1/4$, $\omega = 1/6$, $\gamma = 3/16$ and $d_N = \log\log N$. These constants are used as defaults in implementation, see Section 4.

Remark 2. It is worth noticing that the pixel locations $\{\boldsymbol{x}_{ij}\}_{i=1,j=1}^{M,N_i}$ can be relaxed to vary over subjects (namely time) as $\{\boldsymbol{x}_{t,ij}\}_{t=1,i=1,j=1}^{n,M_t,N_{ti}}$, as long as the dense condition $(*)$ in Assumption (A4) is replaced by $\min_{1\le t\le n} \left\|\boldsymbol{x}_{t,f_{t,1}(x)f_{t,2}(x)} - \boldsymbol{x}_{t,f_{t,1}(y)f_{t,2}(y)}\right\|_2 = \mathcal{O}(N^{-1/2})$, with corresponding functions $f_{t,1}, f_{t,2}$, $t = 1,\ldots,n$. In this scenario, the main theoretical results, including Theorem 1 and Theorem 2, still hold since the order of smoothing bias does not change. But in implementation, the evaluation matrix $\mathbf{X}$ would vary over subjects, making it hard to compute the spectral decomposition of $G_\varphi(\boldsymbol{x}, \boldsymbol{x}')$ defined in 3.11 since $G_\varphi(\boldsymbol{x}, \boldsymbol{x}')$ can not be simplified as (4.16). Moreover, the triangulation selection should be conducted over each image under the setting of varying pixel locations, causing additional heavy computation burden. Therefore, we assume the longitudinal imaging data to be collected at the same locations over time without loss of generality.

Remark 3. From Assumptions (A3) and (A6), it is straightforward that the upper bound of θ is $\theta < (d+2)/(3+d) < 1$, which implies the number of pixels N in each image should not be much smaller than the sample size n. This is quite different from the sparse setting considered in Zheng et al. (2014), whose convergence rate is $(nh)^{-1/2}$ with the bandwidth $h \to 0$, slower than $n^{-1/2}$. Under our dense setting where N tends to infinity, we first smooth over each image and then take the average to estimate the mean function. To control the error brought by smoothing and maintain the convergence rate $n^{-1/2}$, we impose additional requirements on other parameters, which guarantees that smoothing over each image has a negligible impact. However, under the sparse functional data setting where the number of observations in each trajectory has finite expectation, one needs to pool all observations together to estimate its mean function, leading to totally different asymptotic results.

3.2 Asymptotic properties of $\widetilde{m}(x)$ and $\widehat{m}(x)$

Denote

$$\varphi(\boldsymbol{x}) = \frac{\sum_{k=1}^{\infty}\sum_{t=1}^{\infty} a_{tk}\phi_k(\boldsymbol{x})\, U_k}{\mathrm{Var}^{1/2}\left\{\sum_{k=1}^{\infty}\sum_{t=1}^{\infty} a_{tk}\phi_k(\boldsymbol{x})\, U_k\right\}}, \quad \boldsymbol{x} \in \Omega,$$

where $\{U_k\}_{k=1}^{\infty}$ are iid $\mathrm{N}(0,1)$ random variables. Then $\varphi(\boldsymbol{x})$ is a Gaussian process with $\mathbb{E}\varphi(\boldsymbol{x}) \equiv 0$, $\mathbb{E}\varphi^2(\boldsymbol{x}) \equiv 1$, $\boldsymbol{x} \in \Omega$ and covariance function

$$\mathbb{E}\varphi(\boldsymbol{x})\varphi(\boldsymbol{x}') = G_{\varphi}(\boldsymbol{x},\boldsymbol{x}')\left\{G_{\varphi}(\boldsymbol{x},\boldsymbol{x})\, G_{\varphi}(\boldsymbol{x}',\boldsymbol{x}')\right\}^{-1/2}, \quad \boldsymbol{x},\boldsymbol{x}' \in \Omega,$$

where

$$G_{\varphi}(\boldsymbol{x},\boldsymbol{x}') = \sum_{k=1}^{\infty}\phi_k(\boldsymbol{x})\phi_k(\boldsymbol{x}')\left\{1 + 2\sum_{t=0}^{\infty}\sum_{t'=t+1}^{\infty} a_{tk}a_{t'k}\right\}, \quad \boldsymbol{x},\boldsymbol{x}' \in \Omega. \tag{3.11}$$

For any $\alpha \in (0,1)$, define $z_{1-\alpha/2}$ as the $100(1-\alpha/2)$-th percentile of the standard normal distribution. Denote by $Q_{1-\alpha}$ the $100(1-\alpha)$-th percentile of the absolute maxima distribution of $\varphi(\boldsymbol{x})$ over Ω, i.e.,

$$P\left[\sup_{\boldsymbol{x}\in\Omega}|\varphi(\boldsymbol{x})| \le Q_{1-\alpha}\right] = 1-\alpha. \tag{3.12}$$

The following theorem presents the local and global asymptotic properties of the infeasible estimator $\overline{m}(\cdot)$ in (2.6).

Theorem 1. *Under Assumptions (A1), (A3)-(A5), for* $\alpha \in (0,1)$, *as* $n \to \infty$, *the infeasible estimator* $\overline{m}(\cdot)$ *converges at the* $\sqrt{n}$ *rate to* $m(\cdot)$ *with asymptotic covariance function* $G_{\varphi}(\boldsymbol{x}, \boldsymbol{x}')$, *and thus*

$$\mathbb{P}\left\{\sup_{\boldsymbol{x}\in\Omega} n^{1/2} \left|\overline{m}(\boldsymbol{x}) - m(\boldsymbol{x})\right| G_{\varphi}(\boldsymbol{x}, \boldsymbol{x})^{-1/2} \leq Q_{1-\alpha}\right\} \to 1-\alpha,$$
$$\mathbb{P}\left\{n^{1/2} \left|\overline{m}(\boldsymbol{x}) - m(\boldsymbol{x})\right| G_{\varphi}(\boldsymbol{x}, \boldsymbol{x})^{-1/2} \leq z_{1-\alpha/2}\right\} \to 1-\alpha, \quad \boldsymbol{x} \in \Omega.$$

Remark 4. It is worth noting that the convergence rate $n^{-1/2}$ in Theorem 1 is optimal. Cai and Yuan (2011) considered the smoothing spline estimator of the p-times differentiable mean function when $\eta_t(x)$ is an univariate process, and showed that its minmax optimal rate is of the order $N^{-p} + n^{-1/2}$ in L_2-norm. Bosq (2000) derived that the convergence rate of Central Limit Theorem in functional time series is $n^{-1/2}$. Under our setting of high sampling frequency, the sample size n is controlled by the number of pixels N due to Assumptions (A3) and (A6), namely $n \ll N$, thus the optimal rate remains $n^{-1/2}$ and does not depend on N. That is also the reason why the uniform converge rate is the same as point-wise convergence rate.

The following theorem shows that the difference between the bivariate-spline estimator $\widehat{m}(\cdot)$ in (2.10) and the infeasible estimator $\overline{m}(\cdot)$ is uniformly bounded at the $o_p\left(n^{-1/2}\right)$ rate, which enables one to construct SCC based on $\widehat{m}(\cdot)$.

Theorem 2. *Under Assumptions (A1)–(A6), the bivariate spline estimator* $\widehat{m}(\cdot)$ *is* *oracally efficient, i.e., it is asymptotically equivalent to* $\overline{m}(\cdot)$ *up to order* $\mathcal{O}_p\left(n^{-1/2}\right)$

$$\sup_{\boldsymbol{x}\in\Omega} n^{1/2}\left|\overline{m}(\boldsymbol{x}) - \widehat{m}(\boldsymbol{x})\right| = \mathcal{O}_p\left(1\right).$$

Applying the above two theorems, we obtain both pointwise confidence interval and simultaneous confidence corridor for $m(\cdot)$.

Corollary 1. *Under Assumptions (A1)-(A6), for any* $\alpha \in (0,1)$*, as* $n \to \infty$*, an asymptotic* $100\,(1-\alpha)\,\%$ *correct confidence corridor for* $m(\cdot)$ *is*

$$\widehat{m}(\boldsymbol{x}) \pm G_{\varphi}\left(\boldsymbol{x},\boldsymbol{x}\right)^{1/2} Q_{1-\alpha} n^{-1/2}, \quad \boldsymbol{x}\in\Omega, \tag{3.13}$$

and an asymptotic $100\,(1-\alpha)\,\%$ *pointwise confidence interval for* $m(\boldsymbol{x})$ *is*

$$\widehat{m}(\boldsymbol{x}) \pm G_{\varphi}\left(\boldsymbol{x},\boldsymbol{x}\right)^{1/2} z_{1-\alpha/2} n^{-1/2}, \quad \boldsymbol{x}\in\Omega.$$

3.3 Extension to nonlinear process

Noting that the classic MA(∞) is a rather broad category, the FMA(∞) in (2.5) can approximate a large class of stationary processes, but restricted to linear process. As one referee pointed out, it is worth extending FMA(∞) to nonlinear functional processes. In what follows we derive the theoretical extension, but in the rest of our paper, we still focus on FMA(∞) for its simple representation and straightforward theoretical properties.

Rewrite (2.5) as

$$\xi_t(\cdot) = \sum_{k=1}^{\infty} \xi_{tk}\phi_k(\cdot), \quad \xi_{tk} = F_k(\zeta_{t,k}, \zeta_{t-1,k}, \ldots) \quad a.s. \tag{3.14}$$

where F_k, $k \in \mathbb{N}$ are measurable function from $\mathbb{R}^{\mathbb{Z}}$ to $\mathbb{R}$. It is easy to see that $\xi_t(\cdot)$ in (3.14) is a nonlinear process with flexible structures. Following Wu (2005), the physical dependence measure is defined as

$$\Delta_{t,k,r} = \left\|\xi_{tk} - \xi_{tk,\{0\}}\right\|_r,$$

where $\xi_{tk,\{0\}}$ is identical to ξ_{tk} except replacing $\zeta_{0,k}$ by its i.i.d. copy in (3.14). The next theorem states the asymptotic properties under nonlinear process setting.

Theorem 3. *Under Assumptions (A1)-(A6) and (A5'), if α in Assumption (A4) satisfies $\alpha < 1/4$ and $\sup_{k\in\mathbb{N}} \Delta_{t,k,r_1} = \rho_a^t$, then the statements of Theorem 1 and 2 still hold under the nonlinear functional process setting as (3.14), with corresponding limiting covariance function*

$$G^*_{\varphi}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{k=1}^{\infty} \lambda^*_k \phi_k(\boldsymbol{x})\phi_k(\boldsymbol{x}'),$$

where $\lambda_k^* = \lim_{n\to\infty} \mathrm{var}\left(\sum_{t=1}^{n} \xi_{tk}\right)/n,\ k \in \mathbb{N}$ *is the long run variance of* ξ_{tk}.

4. Implementation

4.1 Triangulation selection

Triangulation is crucial since bivariate spline fitting can be sensitive to the triangulation selection. Several approaches, such as maxmin-angle triangulations or Delaunay triangulations, are recommend for selecting the triangulation in Lai and Schumaker (2007), but there is no optimal method of triangulation in the literature. As Yu et al. (2020) pointed out, enough triangles are necessary to present domain features, but after reaching the required minimum number of triangles, further increase of the number of triangles usually makes little difference on the fitting process, even leading to the existence of empty triangles which do not contain any pixel. Thus we tend to choose a moderate number and use the R package Triangulation mentioned in Wang et al. (2020) to build the triangulated meshes.

Assumption (A6) in Section 3 states that the size of triangulations $|\triangle|$ needs to satisfy that $|\triangle|^{-1} = N^{\gamma} d_N$ for some $\gamma > 0$, with $d_N + d_N^{-1} = \mathcal{O}\left(\log^{\vartheta} N\right)$ for some $\vartheta > 0$. Most widely used triangulation methods can guarantee this condition. We recommend that $|\triangle|^{-1} = cN^{3/16} \log\log N$, where c is a tuning constant. The integer parameter K in the R package Triangulation controls the fineness of the triangulation and subsequent triangulation. The parameter K can also be used to measure the size of triangulations since there exists that $K \asymp \left[|\triangle|^{-1}\right]$, where $[a]$ denotes the integer part of a, with $|\triangle| = \sqrt{2}/K$ under the unit square domain as a special case. As K increases, the triangulation fineness increases. We suggest selecting K from the integers in $\left[0.1N^{3/16}\log\log N, N^{3/16}\log\log N\right]$. Among the triangulations indexed by K, we choose the one with the minimal MISE of the estimator $\widehat{m}(\cdot)$ in (2.10), which is defined as

$$\mathrm{MISE}(K) = \int_{\Omega} \mathbb{E}\left\{m(\boldsymbol{x}) - \widehat{m}(\boldsymbol{x})\right\}^2 d\boldsymbol{x}$$

Since the explicit form of MISE(K) is tedious (see Ma (2014)), we propose to compute it conveniently through discretization and summation, that is

$$\mathrm{MISE}(K) = \frac{1}{NL}\sum_{i=1}^{M}\sum_{j=1}^{N_i}\sum_{l=1}^{L}\left\{m(\boldsymbol{x}_{ij}) - \widehat{m}_l(\boldsymbol{x}_{ij})\right\}^2,$$

where L is the number of pre-simulations with default value 20. Figure 1 to 3 show triangulations on three different domains (square, regular 12 polygon and regular 12 polygon with a square hole) with $K = 3, 4, 5$.

4.2 Covariance estimation

Denote $\widehat{\xi}_t(\boldsymbol{x}) = \widehat{\eta}_t(\boldsymbol{x}) - \widehat{m}(\boldsymbol{x})$, $t = 1, \ldots, n$, $\boldsymbol{x} \in \Omega$. To estimate the covariance function $G_{\varphi}(\boldsymbol{x}, \boldsymbol{x}')$, one divides $\left\{\widehat{\xi}_t(\cdot)\right\}_{t=1}^{n}$ into l groups in order and each group has

$B = [n^m]$ samples for some constants $m > 0$ with $l = [n/B]$. Noting that $\widehat{G}_\varphi(\cdot,\cdot)$ is the limit of the covariance function of the process $\sqrt{n}\{\overline{m}(\cdot) - \widehat{m}(\cdot)\}$, we use $\widehat{m}(\boldsymbol{x})$ to mimic $m(\boldsymbol{x})$ and $\sqrt{B}\,\widehat{\delta}_j(\boldsymbol{x})$ to mimic the points from the process $\sqrt{n}\{\overline{m}(\cdot) - \widehat{m}(\cdot)\}$, where

$$\widehat{\delta}_j(\boldsymbol{x}) = B^{-1} \sum_{k=B(j-1)+1}^{Bj} \widehat{\xi}_k(\boldsymbol{x}), \quad j = 1, \ldots, l, \quad \boldsymbol{x} \in \Omega.$$

The estimator $\widehat{G}_\varphi(\boldsymbol{x}, \boldsymbol{x}')$ of $G_\varphi(\boldsymbol{x}, \boldsymbol{x}')$ is defined as

$$\widehat{G}_\varphi(\boldsymbol{x}, \boldsymbol{x}') = \frac{B}{l} \sum_{j=1}^{l} \left\{ \widehat{\delta}_j(\boldsymbol{x}) \widehat{\delta}_j(\boldsymbol{x}') - \overline{\widehat{\delta}}(\boldsymbol{x}) \overline{\widehat{\delta}}(\boldsymbol{x}') \right\}, \quad \boldsymbol{x}, \boldsymbol{x}' \in \Omega, \tag{4.15}$$

where $\overline{\widehat{\delta}}(\boldsymbol{x}) = l^{-1} \sum_{j=1}^{l} \widehat{\delta}_j(\boldsymbol{x}), \boldsymbol{x} \in \Omega$. The next theorem characterizes the uniform weak convergence of $\widehat{G}_\varphi(\boldsymbol{x}, \boldsymbol{x}')$.

Theorem 4. *Under Assumptions (A1)- (A6), for constant m that satisfies $-(1 + 2/r_1)/(\rho_a + 1/2) < m < \min\{(d+1)r_1/\theta - 4/r_1, (1/2 - \beta_2 - \gamma)/\theta - 2/r_2\}$, the estimator $\widehat{G}_\varphi(\boldsymbol{x}, \boldsymbol{x}')$ of $G_\varphi(\boldsymbol{x}, \boldsymbol{x}')$ is uniformly consistent in probability, i.e.*

$$\sup_{\boldsymbol{x}, \boldsymbol{x}' \in \Omega} \left| \widehat{G}_\varphi(\boldsymbol{x}, \boldsymbol{x}') - G_\varphi(\boldsymbol{x}, \boldsymbol{x}') \right| = o_p(1).$$

Throughout this section, we choose $B = \left[n^{1/5} \log\log n\right]$.

4.3 Estimating the percentile

Recalling that the solution of (2.9) is $\widehat{\eta}_t(\boldsymbol{x}) = \widetilde{\mathbf{B}}(\boldsymbol{x})^\top \widehat{\boldsymbol{\beta}}_t$, let $\widehat{\boldsymbol{\beta}} = n^{-1}\sum_{t=1}^{n}\widehat{\boldsymbol{\beta}}_t$, then the bivariate spline estimator $\widehat{m}(\boldsymbol{x}) = \widetilde{\mathbf{B}}(\boldsymbol{x})^\top \widehat{\boldsymbol{\beta}}$. Denote $\widehat{\boldsymbol{\beta}}_{\delta_j} = B^{-1}\sum_{k=B(j-1)+1}^{B_j}\left(\widehat{\boldsymbol{\beta}}_k - \widehat{\boldsymbol{\beta}}\right)$ and the matrix $\widehat{g}_\varphi = l^{-1}\sum_{j=1}^{l}\left(\widehat{\boldsymbol{\beta}}_{\delta_j}\widehat{\boldsymbol{\beta}}_{\delta_j}^T\right) - \left(l^{-1}\sum_{j=1}^{l}\widehat{\boldsymbol{\beta}}_{\delta_j}\right)\left(l^{-1}\sum_{j=1}^{l}\widehat{\boldsymbol{\beta}}_{\delta_j}\right)^T$. The covariance function estimator $\widehat{G}_\varphi(\boldsymbol{x},\boldsymbol{x}')$ allows the bivariate spline expansion as

$$\widehat{G}_\varphi(\boldsymbol{x},\boldsymbol{x}') = \widetilde{\mathbf{B}}(\boldsymbol{x})^\top \widehat{g}_\varphi \widetilde{\mathbf{B}}(\boldsymbol{x}') \tag{4.16}$$

For $k \geq 1$, we consider the following bivariate spline approximation for the eigenfunction $\widehat{\psi}_{k,\varphi}(\boldsymbol{x})$ of $\widehat{G}_\varphi(\boldsymbol{x},\boldsymbol{x}')$: $\widehat{\psi}_{k,\varphi}(\boldsymbol{x}) = \widetilde{\mathbf{B}}(\boldsymbol{x})^\top\widehat{\gamma}_k$, where $\widehat{\gamma}_k$ are coefficients satisfying that $N^{-1}\widehat{\gamma}_k^\top\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}\widehat{\gamma}_k = 1$. The estimates of eigenvalues $\lambda_{k,\varphi}$ and corresponding eigenfunctions $\psi_{k,\varphi}$ can be obtained by solving the following eigenequation,

$$\int_\Omega \widehat{G}_\varphi(\boldsymbol{x},\boldsymbol{x}')\widehat{\psi}_{k,\varphi}(\boldsymbol{x}')d\boldsymbol{x}' = \widehat{\lambda}_{k,\varphi}\widehat{\psi}_{k,\varphi}(\boldsymbol{x}). \tag{4.17}$$

The next corollary can be derived from Theorem 4 directly.

Corollary 2. *Under the conditions in Theorem 4, the corresponding eigen-pairs* $\left\{\widehat{\lambda}_{k,\varphi}, \widehat{\psi}_{k,\varphi}(\boldsymbol{x})\right\}$, $k \in \mathbb{N}$, *in (4.17) have uniform consistency in probability, i.e. for*

$k \in \mathbb{N}$,

$$\left|\widehat{\lambda}_{k,\varphi} - \lambda_{k,\varphi}\right| + \sup_{\boldsymbol{x} \in \Omega} \left|\widehat{\psi}_{k,\varphi}(\boldsymbol{x}) - \psi_{k,\varphi}(\boldsymbol{x})\right| = o_p(1).$$

Note that the integration in eigenequation (4.17) can be approximated by discrete summation, and plugging the covariance function estimator (4.16) leads to

$$N^{-1}\widehat{g}_{\varphi}\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}}\widehat{\gamma}_k = \widehat{\lambda}_{k,\varphi}\widehat{\gamma}_k. \tag{4.18}$$

To solve the above equation subject to $N^{-1}\widehat{\gamma}_k^{\top}\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}}\widehat{\gamma}_k = 1$, we utilize the Cholesky decomposition: $N^{-1}\widetilde{\mathbf{X}}^{\top}\widetilde{\mathbf{X}} = L_{\widetilde{\mathbf{X}}}L_{\widetilde{\mathbf{X}}}^{\top}$. Therefore solving (4.18) is equivalent to solve $L_{\widetilde{\mathbf{X}}}^{\top}\widehat{g}_{\varphi}L_{\widetilde{\mathbf{X}}}L_{\widetilde{\mathbf{X}}}^{\top}\widehat{\gamma}_k = \widehat{\lambda}_{k,\varphi}L_{\widetilde{\mathbf{X}}}^{\top}\widehat{\gamma}_k$, that is $\widehat{\lambda}_{k,\varphi}$ and $L_{\widetilde{\mathbf{X}}}^{\top}\widehat{\gamma}_k$ are the eigenvalues and unit eigenvectors of $L_{\widetilde{\mathbf{X}}}^{\top}\widehat{g}_{\varphi}L_{\widetilde{\mathbf{X}}}$. Thus $\widehat{\gamma}_k$ can be obtained by multiplying $\left(L_{\widetilde{\mathbf{X}}}^{\top}\right)^{-1}$ immediately after the unit eigenvectors of $L_{\widetilde{\mathbf{X}}}^{\top}\widehat{g}_{\varphi}L_{\widetilde{\mathbf{X}}}$. After that, $\widehat{\psi}_{k,\varphi}(\boldsymbol{x})$ are obtained and $\widehat{\phi}_{k,\varphi}(\boldsymbol{x}) = \widehat{\lambda}_{k,\varphi}^{1/2}\widehat{\psi}_{k,\varphi}(\boldsymbol{x})$. Next the truncated number κ of eigenfunctions is chosen by the following efficient criteria, i.e. $\kappa = \arg\min_{1 \le v \le T}\left\{\sum_{k=1}^{v}\widehat{\lambda}_{k,\varphi} / \sum_{k=1}^{T}\widehat{\lambda}_{k,\varphi} > 0.95\right\}$, where $\left\{\widehat{\lambda}_{k,\varphi}\right\}_{k=1}^{T}$ are the first T estimated positive eigenvalues.

One then simulates $\widehat{\zeta}_b(\boldsymbol{x}) = \widehat{G}_{\varphi}(\boldsymbol{x}, \boldsymbol{x}')^{-\frac{1}{2}}\sum_{k=1}^{\kappa} Z_{k,b}\widehat{\phi}_k(\boldsymbol{x})$, where $\{Z_{k,b}\}_{k=1,b=1}^{\kappa,b_M}$ are iid standard normal variables, and b_M is a preset large integer with default 1000. One takes the maximal absolute value for each copy of $\widehat{\zeta}_b(\boldsymbol{x})$ and uses the empirical quantile $\widehat{Q}_{1-\alpha}$ of these maximum values as an estimate of $Q_{1-\alpha}$.

Finally, the SCC for the mean function is computed as

$$\widehat{m}(\boldsymbol{x}) \pm n^{-1/2}\widehat{G}_{\varphi}\left(\boldsymbol{x},\boldsymbol{x}\right)^{1/2}\widehat{Q}_{1-\alpha}, \quad \boldsymbol{x} \in \Omega. \tag{4.19}$$

5. Simulation Studies

In this section, we carry out various simulations to illustrate the finite sample performance of the proposed method. The data is generated from the following model:

$$Y_{t,ij} = m\left(\boldsymbol{x}_{ij}\right) + \sum_{k=1}^{7} \xi_{tk}\phi_k\left(\boldsymbol{x}_{ij}\right) + \sigma\left(\boldsymbol{x}_{ij}\right)\varepsilon_{t,ij}, \quad t = 1,\ldots,n, \tag{5.20}$$

where $\boldsymbol{x}_{ij} = (s_{ij}, t_{ij}) \in \Omega \subset [0,1]^2$, $i = 1,\ldots,M$, $j = 1,\ldots,N_i$ and $\sum_{i=1}^{M} N_i = N$. We consider three different shapes of the domain Ω: square, regular 12 polygon, which can be viewed as an approximation of a circle, and regular 12 polygon with a square hole. The mean function $m(\cdot)$ and eigenfunctions $\phi_k(\cdot)$ are set as follows:

$$\begin{aligned}
m(s,t) &= 6\sin\left(s+t\right)e^{-2(s+t)} + 3s\sin t,\\
\varphi_1\left(s,t\right) &= \sin\left(\pi t/2\right)\sin\left(3\pi s/2\right),\\
\varphi_2\left(s,t\right) &= \sin\left(3\pi t/2\right)\sin\left(\pi s/2\right),\\
\varphi_3\left(s,t\right) &= \sin\left(3\pi t/2\right)\sin\left(3\pi s/2\right),
\end{aligned}$$

$$\varphi_4(s,t) = \sin(5\pi t/2)\sin(3\pi s/2),$$
$$\varphi_5(s,t) = \sin(3\pi t/2)\sin(5\pi s/2),$$
$$\varphi_6(s,t) = \sin(5\pi t/2)\sin(5\pi s/2),$$
$$\varphi_7(s,t) = \sin(5\pi t/2)\sin(7\pi s/2),$$
$$\varphi_k(s,t) = 0, \quad k \geq 8.$$

To guarantee the orthogonality of eigenfunctions, one can use Schmidt orthogonalization and get $\varphi_k^*(s,t)$, $k = 1, 2, \ldots$ It is obvious that $\varphi_k^*(s,t) = \varphi_k(s,t)$ in the square domain case, while $\varphi_k^*(s,t)$ is a linear combination of $\{\varphi_k(s,t)\}_{k=1}^{\infty}$ in other situations. Then let $\phi_k(s,t) = \sqrt{\lambda_k}\varphi_k^*(s,t)$ with $\lambda_k = 2^{-(k-1)/2}$ and FPC scores $\{\xi_{tk}\}_{t=1,k=1}^{n,7}$ are generated from (2.5), where $\{\zeta_{tk}\}_{t=-5,k=1}^{n,7}$ are iid $\mathrm{N}(0,1)$ variables and $a_{0k} = 0.8$, $a_{1k} = a_{2k} = 0.4$, $a_{3k} = a_{4k} = a_{5k} = a_{6k} = -0.1$, $a_{tk} = 0$ for $t \geq 2$, $k = 1, \ldots, 7$.

We generate the homoscedastic measurement errors $\sigma(\boldsymbol{x}) = 0.1$ and heteroscedastic measurement errors $\sigma(\boldsymbol{x}) = 0.1\,(5 - \exp(-(s+t)))\,/\,(5 + \exp(-(s+t)))$. The errors $\{\varepsilon_{t,ij}\}_{t=1,i=1,j=1}^{n,M,N_i}$ are iid with three different distributions: normal, uniform and Laplace distribution. The number of pixels N is 10000 and 20000 respectively, while the number of images n is taken to be $\left[N^{1/4}\log N(\log\log N)^2\right]$.

Throughout this section, the mean function is estimated by bivariate splines in space $S_d^r(\triangle)$ with $d = 5$ and $r = 1$, which can approach the full approximation power asymptotically, see Lai and Schumaker (2007). Tables 1 and 2 display the empirical coverage rate, namely the percentage out of the 500 replications of the true mean function $m(\cdot)$ being covered by the bivariate spline SCCs (4.19) at the N points $\{\boldsymbol{x}_{ij}\}_{i=1,j=1}^{M,N_i}$. It is shown that in both scenarios, the coverage rate of the SCC becomes closer to the nominal confidence level as the sample size increases, which reveals a positive confirmation of the asymptotic theory.

To visualize the SCCs for the mean function, Figure 4 to 12 show the estimated mean functions and their 95% SCCs for the true mean function $m(\cdot)$ with $\sigma(\boldsymbol{x}) = 0.1$, $\varepsilon_{t,ij} \sim \mathrm{N}(0,1)$ and $N = 10000, 20000, 40000$ respectively on three different domains. As expected, when N increases, the SCC becomes narrower and the bivariate spline estimators are closer to the true mean function. In all panels, the true mean function is entirely covered by upper and lower corridors.

6. Real Data Analysis

In this section, we apply the proposed SCCs to two sea water potential temperature data sets oberseved on typically complicate domain. Sea water potential temperature serves as one of the most important factors in marine hydrological conditions, which is often used as a principal indicator for studying the properties and the movement of water masses. Investigating the temporal and spatial distribution and changing laws of sea temperature is of great significance for marine fishing, aquaculture, and marine operations.

The data sets used in our analysis are from the CMEMS global analysis and forecast product, which is available at https://resources.marine.copernicus.eu. CMEMS collects the rough data, such as 3D potential temperature, salinity and currents, bottom potential temperature, or mixed layer thickness, and then transform it by some professional algorithm. All data is recorded globally on a standard grid at 1/12 degree (approximately 8km) and 50 standard levels.

6.1 Black Sea

The Black Sea is a marginal sea of the Atlantic Ocean lying between Europe and Asia, covering an area from 26.8°E to 42.2°E and 40.5°N to 47.6°N, see its equirectangular projection map in Figure 13. Hourly sea surface (at depth 0.494m) water potential temperature is recorded on standard grids every 1/12 degree both longitude and latitude from 00:30 on December 9, 2020, to 00:30 on December 24, 2020. The black dots in Figure 14(a) show the observed data locations. Each hourly observed temperature data of the Black Sea can be naturally regraded as a image. This results in longitudinal imaging data with $n = 360$ temporally ordered images and $N = 6583$ pixels in each image.

The mean function reflects the overall trend of sea water potential temperature data, and also serves as a preliminary step for further data analysis. We use bivariate splines with smooth parameter $r = 1$ and $d = 5$ for the estimation of mean function. Figure 14(b) presents the triangulation of the Black Sea domain, which contains 39 triangles with 35 vertices. The estimated mean function and its corresponding 95% SCC computed by (4.19) are displayed in Figure 16 and 17 respectively. It is shown in Figure 16 that the average sea surface water temperature decreases from low latitude to high latitude, which corroborates the classic oceanographic theory.

6.2 Madagascar surrounding sea

Madagascar is an island country in the Indian Ocean and off the coast of East Africa. We investigate the potential temperature of the sea around Madagascar, ranging from 41.0°E to 55.0°E and 11.0°S to 30.0°S. Similar to the previous case, hourly potential temperature is measured on the standard grids every 1/12 degree, see Figure 18(a) for pixel locations. It is clear the domain of Madagascar surrounding sea is more complicated due to the existence of a hole (Madagascar island). We focus on the data from 00:30 on December 9, 2020, to 00:30 on January 24, 2021. Hence, there are $n = 840$ timely ordered images with $N = 26151$ pixels per image.

We also utilize bivariate splines with smooth parameter $r = 1$ and $d = 5$ to approximate its mean function. Triangulation on the Madagascar surrounding sea domain is shown in Figure 18(b), with 33 triangles and 32 vertices. Figure 19 and 20 present the estimated mean function and its corresponding 95% SCC computed by

(4.19). We can see from Figure 19 that there is always higher sea surface temperature near land. This strongly demonstrates that our method is widely applicable and capable of handling the complex image domain even with a hole.

7. Concluding Remarks

Our paper investigates longitudinal imaging data over complicated domains, referring to ordered images with numerous pixels collected at a high frequency over a time span. We propose an asymptotically correct and computationally efficient bivariate spline estimator for its mean function. Both global and local asymptotic properties of the bivariate estimator are investigated, with SCCs to make inference on the true mean function. To our best knowledge, this is the first piece of work in large-scale longitudinal imaging data, which yields attractive inference results, and at meanwhile is free from ultra high dimension and model misspecification. There is no doubt that our method enjoys a wide range of application to imaging data in geography, oceanography and biomedical studies.

Some issues still warrant further investigation. The data in Section 6 is considered to come from a horizontal plane, which ignores the curvature of the earth's surface. It may be more accurate to assume the data collected on a sphere. Spherical splines introduced in Lai and Schumaker (2007) could be suitable to better approximate the aforementioned 3D imaging data. However, it is not an easy task because of the elusive theory and heavy computing burden of high-resolution 3D images compared to the 2D ones. In addition, it makes sense to extend the proposed methodology to functional regression models, which has been a cutting-edge research area in recent years. How to construct SCCs of the functional coefficients in such models is also challenging due to the deeper asymptotic theory.

References

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications.* Springer-Verlag, New York.

Cai, T. T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist* **39**, 2330–2355.

Cao, G., Wang, L., Li, Y. and Yang, L. (2016). Oracle-efficient confidence envelopes for covariance functions in dense functional data. *Stat Sin* **26**, 359-383

Cao, G., Yang, L. and Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *J Nonparametr Stat* **24**, 359-377.

Choi, H., and Reimherr, M. (2018). A geometric approach to confidence regions and bands for functional parameters. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **80**, 239-260.

Csőrgő, M., and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York-London.

French, J. P., and Kokoszka, P. S. (2020). A sandwich smoother for spatio-temporal functional data. *Spat. Stat.* **42**, 100413.

Ferraccioli, F., Arnone, E., Finos, L., Ramsay, J. O. and Sangalli, L. M. (2021). Nonparametric density estimation over complicated domains. *J R Stat Soc Series B*, **83**, 346-368.

George, B., Denney, Jr., Thomas, G.H., DelItalia, L. and Aban, I. (2016) Applying a spatiotemporal model for longitudinal cardiac imaging data. *Ann. Appl. Stat.* **1**, 527-548.

George, B., and Aban, I. (2015). Selecting a separable parametric spatiotemporal covariance structure for longitudinal imaging data. *Stat. Med.* **34**(1), 145-161.

Gu, L., Wang, L., Härdle, W. and Yang, L. (2014). A simultaneous confidence corridor for varying coefficient regression with sparse functional data. *TEST* **23**, 806-843.

Gu, L., and Yang, L. (2015). Oracally efficient estimation for single-index link function with simultaneous confidence band. *Electron. J. Stat.* **9**, 1540-1561.

Kokoszka, P., and Reimherr, M. (2019). Some recent developments in inference for geostatistical functional data. *Rev. Colombiana Mat.* **42**(1), 101-122.

Lai, M. J., and Schumaker, L. (2007). Cambridge University Press, Cambridge.

Lai, M. J., and Wang, L. (2013). Bivariate penalized splines for regression. *Stat Sin* **23**, 1399-1417.

Li, J. and Yang, L. (2022). Statistical inference for the mean function of stationary functional time series. *Stat Sin*. DOI: 10.5705/ss.202021.0107

Ma, S. (2014). A plug-in the number of knots selector for polynomial spline regression. *J Nonparametr Stat* **26** (3), 489-507.

Ma, S., Yang, L. and Carroll, R. (2012). A simultaneous confidence band for sparse longitudinal regression. *Stat Sin* **22**, 95-122.

Ramsay, J.,and Silverman, B. (2005). *Functional Data Analysis*, Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Wang, Y., Wang, G., Wang, L. and Ogden, R. T. (2020). Simultaneous confidence corridors for mean functions in functional data analysis of imaging data. *Biometrics* **76**(2), 427-437.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA*, **102**(40), 14150-14154.

Yu, S., Wang, G., Wang, L., Liu, C., and Yang, L. (2020). Estimation and inference for generalized geoadditive models. *J. Amer. Statist. Assoc.* **115**, 761-774.

Yu, S., Wang, G., Wang, L., and Yang, L. (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Stat Sin* **31**(3), 1463-1487.

Zheng, S., Yang, L. and Härdle, W. (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *J Am Stat Assoc* **109** 661–673.

Zhou, L. and Pan, H. (2014). Principal Component Analysis of Two-Dimensional Functional Data, *J. Comput. Graph. Statist.* **23**(3), 779-801.

Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China.

E-mail: hqr20@mails.tsinghua.edu.cn

School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: li-j17@tsinghua.org.cn