

## Statistica Sinica Preprint No: SS-2021-0415

<b>Title</b>	Statistical Inference for Mean Function of Longitudinal Imaging Data over Complicated Domains
<b>Manuscript ID</b>	SS-2021-0415
<b>URL</b>	<a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a>
<b>DOI</b>	10.5705/ss.202021.0415
<b>Complete List of Authors</b>	Qirui Hu and Jie Li
<b>Corresponding Authors</b>	Jie Li
<b>E-mails</b>	li-j17@tsinghua.org.cn

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

Qirui Hu<sup>1</sup>, Jie Li<sup>2</sup>

<sup>1</sup>*Tsinghua University, Beijing 100084, China.*

<sup>2</sup>*Renmin University of China, Beijing 100872, China.*

*Abstract:* We propose a novel procedure for estimating the mean function of longitudinal imaging data with inherent spatial and temporal correlation. We depict the dependence between temporally ordered images using a functional moving average, and use flexible bivariate splines over triangulations to handle the irregular domain of images which is common in imaging studies. We establish both the global and the local asymptotic properties of the bivariate spline estimator for the mean function, 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 though all images were entirely observed without errors. We use Monte Carlo simulation experiments to demonstrate the finite-sample performance of the proposed method, the results of which strongly corroborate the asymptotic theory. The proposed method is further illustrated by analyzing two seawater potential temperature data sets.

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

## 1. Introduction

Imaging data are generated by decomposing an image into many small areas, called pixels, with a value used to express its gray scale. Longitudinal imaging data, collected from a series of repeated observations of the same subject over some extended time frame, frequently appear in the fields of medicine, meteorology, geography, and environmental science, for example continuous observations of tomography imaging or remote sensing images. Analyses of longitudinal imaging data provide new opportunities for detecting the dynamic changes in one subject over time, but are complicated by the spatial correlation between pixels within a single image and temporal correlation between sequentially ordered images.

The method used most often to analyze 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. However, although their information criteria are highly accurate in terms of choosing spatial and temporal parametric correlation functions, the risk of model misspecification and poor performance in terms of inference remain inevitable. George et al. (2016) described how to use the aforementioned model in practice, and apply it to data from a longitudinal cardiac imaging study. In the study, a handful of successive im-

---

ages with a small number of spatial locations were collected daily, monthly, or even yearly, at limited times. Today, longitudinal imaging data usually comprise orders of magnitude 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 seawater potential temperature are recorded on dense regular grids (see Figure 1(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 four randomly selected images shown in Figure 2. The ultrahigh dimension of the data is problematic for unstructured correlation matrices, leading to the traditional model losing its effect. Therefore, we require a practical, computationally efficient, and theoretically reliable method for analyzing large-scale longitudinal imaging data.

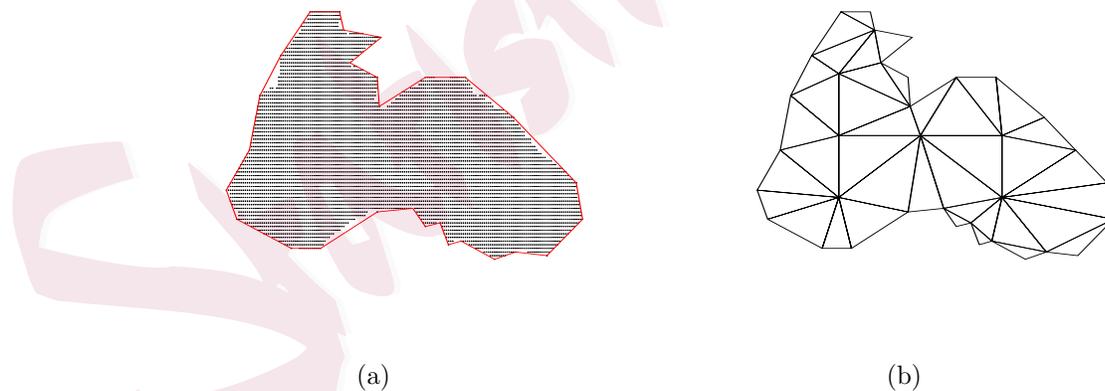
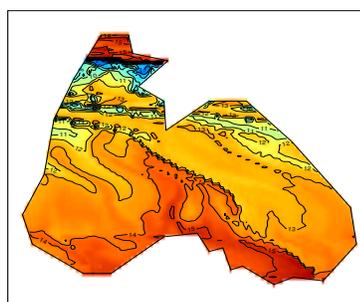
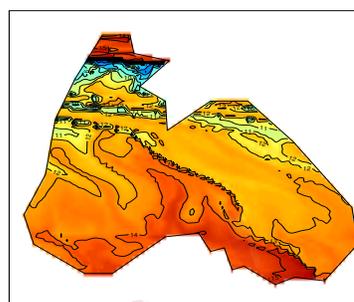


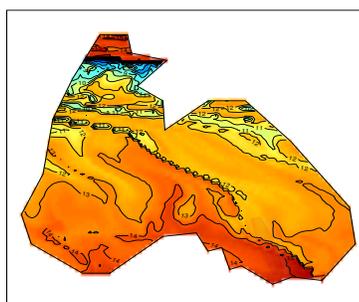
Figure 1: (a) Plots of observation locations (black dots) in the Black Sea; (b) Triangulation on the Black Sea domain.



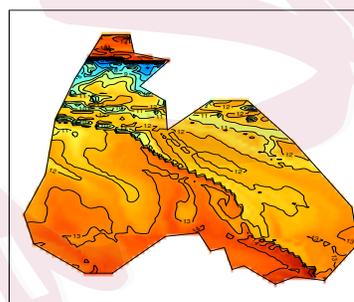
(a) 00:30 on December 9, 2020



(b) 17:30 on December 12, 2020



(c) 12:30 on December 17, 2020



(d) 09:30 on December 20, 2020

Figure 2: Four randomly sampled images for the hourly surface temperature of the Black Sea.

Functional data analysis provides a novel and powerful approach to dealing with imaging data. Instead of imposing a spatial structure directly, it views imaging data as realizations of random fields, which naturally captures the spatial correlations between 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 include a time dimension in the smoother, which causes additional computational complexity and a failure to derive the global mean surface. Furthermore, statistical inferences cannot be conducted, owing to a lack of theory. Kokoszka and Reimherr (2019) reviewed recent developments related to inferences for functions defined at spatial locations. They considered time series of functions defined at irregularly distributed spatial points or on a grid, namely, spatially indexed functional time series. In contrast to their research, we focus on 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 consist of a collection of  $n$  temporally ordered images  $\{\eta_t(\cdot)\}_{t=1}^n$  on a two-dimensional (2D) bounded domain  $\Omega$ , where  $\Omega$  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 are discrete values of a regular grid of pixels from the fields  $\{\eta_t(\cdot)\}_{t=1}^n$ , plus random errors. Because most imaging data are recorded by automated instruments, we assume the pixel locations are dense regular grids  $\mathbf{x}_{ij} \in \Omega$ , for  $i = 1, \dots, M$ ,  $j = 1, \dots, N_i$ , which forms an  $M$ -row array with  $N_i$  points in the  $i$ th row, see Figure 1(a) and Figure 3(a). A similar data setting is considered by Yu et al. (2021). Let  $Y_{t,ij} = Y_t(\mathbf{x}_{ij})$  be the observation of the  $t$ th image at location  $\mathbf{x}_{ij}$ . Then the data set  $\{(Y_{t,ij}, \mathbf{x}_{ij})\}$ , for  $t = 1, \dots, n$ ,  $i = 1, \dots, M$ ,

---

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

$$Y_{t,ij} = \eta_t(\mathbf{x}_{ij}) + \sigma(\mathbf{x}_{ij}) \varepsilon_{t,ij}, \quad (1.1)$$

where  $\varepsilon_{t,ij}$  are independent and identically distributed (i.i.d.) random errors with mean zero and variance one, and  $\sigma^2(\cdot)$  is the variance function of the measurement errors.

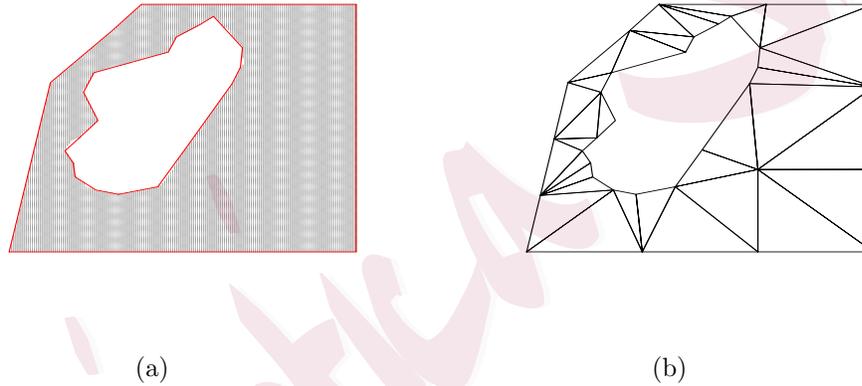


Figure 3: (a) Plots of observation locations (black dots) in the Madagascar surrounding sea; (b) Triangulation on the Madagascar surrounding sea domain.

In longitudinal imaging data analysis, a fundamental issue lies in estimating the of mean function  $m(\cdot)$ , defined as  $m(\cdot) = \mathbb{E}\{\eta(\cdot)\}$ . One challenge is that lots of imaging data are collected over complicated domains, even with gaps and holes (see Figure 4), leading to the problem of “leakage” across complex boundaries for some traditional smoothing methods, such as tensor product smoothing, kernel smoothing,

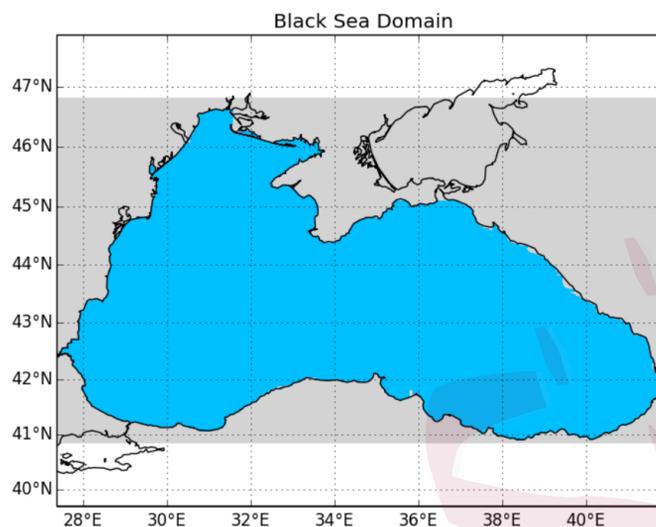


Figure 4: The Black Sea spatial domain (blue surface) for validation metrics.

and wavelet smoothing. Bivariate splines on triangulations, introduced in Lai and Schumaker (2007), are effective in overcoming the poor boundary estimation, and preserve important features (shape and smoothness) of the imaging data. Any 2D geometric domain can be represented as a polygon that is decomposed into triangles using triangulation. Bivariate splines are widely used, owing to their computational ease and convenient representation, with flexible degrees and various smoothness; see Lai and Wang (2013), Zhou and Pan (2014), and Ferraccioli et al. (2021) for applications in various statistical areas. Wang et al. (2020) proposed a consistent mean function estimator for imaging data, based on bivariate splines over triangulations. One serious limitation is that they restrict images  $\{\eta_t(\cdot)\}_{t=1}^n$  to be i.i.d. copies, which

---

is not the case in longitudinal imaging data. To model the time-ordered 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( $\infty$ ) series  $\{\xi_t(\cdot)\}_{t=1}^{\infty}$ , as in Li and Yang (2022+). Specifically,  $\xi_t(\cdot)$  satisfies the following equation:

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

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

Under the above dependence structure, we propose a bivariate spline estimator for the mean function  $m(\cdot)$ . Theorem 2 states 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 us to construct an asymptotically correct simultaneous confidence corridor (SCC) of the mean function  $m(\cdot)$ , under some mild conditions. The SCC is vital for evaluating the variability and testing the 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. The results of our 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 our main theoretical results for the SCC constructed from a bivariate spline estimator. In Section 4, we discuss procedures for implementing the proposed SCC, and in Section 5, we present the findings of our extensive simulation studies. In Section 6, we apply the proposed method to two hourly seawater potential temperature data sets. All technical proofs and some additional simulation results 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  $G(\mathbf{x}, \mathbf{x}') = \text{Cov}\{\eta(\mathbf{x}), \eta(\mathbf{x}')\}$ , for  $\mathbf{x}, \mathbf{x}' \in \Omega$ . The identically distributed random fields  $\{\eta_t(\cdot)\}_{t=1}^n$  are decomposed as  $\eta_t(\cdot) = m(\cdot) + \xi_t(\cdot)$ , where each  $\xi_t(\mathbf{x})$  can be viewed as a small-scale variation of  $\mathbf{x}$  on the  $t$ th image, and is assumed to be a strictly stationary  $L^2$  process, with  $\mathbb{E}\xi_t(\mathbf{x}) =$

## 2.1 Functional moving average model

0 and covariance  $G(\mathbf{x}, \mathbf{x}') = \text{Cov} \{ \xi_t(\mathbf{x}), \xi_t(\mathbf{x}') \}$ , for  $\mathbf{x}, \mathbf{x}' \in \Omega$ . Mercer's lemma entails the decomposition of its covariance function  $G(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^{\infty} \lambda_k \psi_k(\mathbf{x}) \psi_k(\mathbf{x}')$ , where  $\{ \lambda_k \}_{k=1}^{\infty}$  are a series of decreasing positive eigenvalues, and  $\{ \psi_k(\cdot) \}_{k=1}^{\infty}$  are the corresponding eigenfunctions, forming an orthogonal basis of  $L^2(\Omega)$ , such that  $\sum_{k=1}^{\infty} \lambda_k < \infty$  and  $\int G(\mathbf{x}, \mathbf{x}') \psi_k(\mathbf{x}') d\mathbf{x}' = \lambda_k \psi_k(\mathbf{x})$ .

Then, for any  $t \in \mathbb{Z}$ , the zero-mean field  $\xi_t(\mathbf{x})$ , for  $\mathbf{x} \in \Omega$ , allows the general Karhunen–Loève representation  $\xi_t(\mathbf{x}) = \sum_{k=1}^{\infty} \xi_{tk} \phi_k(\mathbf{x})$ , in which the rescaled eigenfunctions  $\{ \phi_k(\cdot) \}_{k=1}^{\infty}$ , called functional principle components (FPC), satisfy  $\phi_k = \sqrt{\lambda_k} \psi_k$  and  $\int \{ \eta(\mathbf{x}) - m(\mathbf{x}) \} \phi_k(\mathbf{x}) d\mathbf{x} = \lambda \xi_k$ , for  $k \geq 1$ . The random coefficients  $\xi_{tk}$  are uncorrelated over  $k$ , with mean zero and variance one, referred to as FPC scores. Note that although 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; a detailed estimation procedure is given in Section 4.

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

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

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

## 2.1 Functional moving average model

white noise  $\{\zeta_t(\cdot)\}_{t=-\infty}^{\infty}$  allows for its own Karhunen–Loève representation  $\zeta_t(\cdot) = \sum_{k=1}^{\infty} \zeta_{t,k} \phi_k(\cdot)$ , where  $\{\zeta_{t,k}\}_{t=-\infty, k=1}^{\infty, \infty}$  are uncorrelated random variables with mean zero and variance one. Together with (1.2), we have

$$\begin{aligned} \xi_t(\cdot) &= \sum_{t'=0}^{\infty} A_{t'} \left\{ \sum_{k=1}^{\infty} \zeta_{t-t',k} \phi_k(\cdot) \right\} = \sum_{t'=0}^{\infty} \sum_{k=1}^{\infty} a_{t',k} \zeta_{t-t',k} \phi_k(\cdot) \\ &= \sum_{k=1}^{\infty} \left( \sum_{t'=0}^{\infty} a_{t',k} \zeta_{t-t',k} \right) \phi_k(\cdot). \end{aligned} \quad (2.1)$$

Note that because  $\xi_t(\cdot) = \sum_{k=1}^{\infty} \xi_{tk} \phi_k(\cdot)$  absolutely almost surely, by the 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  $\xi_{tk}$  has variance one, we assume  $\sum_{t=0}^{\infty} a_{t,k}^2 \equiv 1$ , for  $k = 1, 2, \dots$ , which is reasonably analogous to what is assumed in numerical MA( $\infty$ ).

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

$$\begin{aligned} Y_{t,ij} &= m(\mathbf{x}_{ij}) + \xi_t(\mathbf{x}_{ij}) + \sigma(\mathbf{x}_{ij}) \varepsilon_{t,ij} \\ &= m(\mathbf{x}_{ij}) + \sum_{k=1}^{\infty} \xi_{tk} \phi_k(\mathbf{x}_{ij}) + \sigma(\mathbf{x}_{ij}) \varepsilon_{t,ij}, \end{aligned} \quad (2.2)$$

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

$$\xi_t(\cdot) = \sum_{k=1}^{\infty} \xi_{tk} \phi_k(\cdot), \quad \xi_{tk} = \sum_{t'=0}^{\infty} a_{t',k} \zeta_{t-t',k} \quad a.s. \quad (2.3)$$

## 2.2 Bivariate spline estimator

Had the  $n$  images  $\{\eta_t(\cdot)\}_{t=1}^n$  been entirely observed over  $\Omega$ , an intuitive estimator for the mean function  $m(\cdot)$  in (2.2) would be the sample mean

$$\bar{m}(\mathbf{x}) = n^{-1} \sum_{t=1}^n \eta_t(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad (2.4)$$

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

Triangulation is used to process data distributed over difficult domains with complex boundaries and/or interior holes. Denote by  $T$  a triangle that is a convex hull of three points, not located in one line. A triangulation of  $\Omega$  is a collection of  $H$  triangles  $\Delta = \{T_1, \dots, T_H\}$ , with  $\Omega = \bigcup_{h=1}^H T_h$ , provided that any nonempty intersection between a pair of triangles in  $\Delta$  is either a shared vertex or a shared edge. Given a triangle  $T \in \Delta$ , let  $|T|$  be its longest edge length, and  $\rho_T$  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  $\pi_T$  is small, the triangles are relatively uniform, in the sense

## 2.2 Bivariate spline estimator

that all angles of the triangles in  $\Delta$  are relatively the same. Denote the size of  $\Delta$  by  $|\Delta| = \max\{|T|, T \in \Delta\}$ , that is, the length of the longest edge of all triangles in  $\Delta$ .

For any triangle  $T \in \Delta$  and any fixed point  $\mathbf{x} \in \Omega$ , let  $b_1$ ,  $b_2$ , and  $b_3$ , be the barycentric coordinates of  $\mathbf{x}$  relative to  $T$ . The Bernstein basis polynomials of degree  $d$  relative to triangle  $T$  are defined as  $B_{ijk}^{T,d}(\mathbf{x}) = (i!j!k!)^{-1}d!b_1^i b_2^j b_3^k$ ,  $i + j + k = d$ , and are 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  $\Omega$ . Given  $\Delta$ , let  $S_d^r(\Delta) = \{s \in C^r(\Omega), s|_T \in \mathbb{P}_d(T), T \in \Delta\}$  be a spline space of degree  $d$  and smoothness  $r$  over  $\Delta$ , 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 a degree are continuous.

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

$$\mathbf{X} = \{\mathbf{B}(\mathbf{x}_{11}), \dots, \mathbf{B}(\mathbf{x}_{1N_1}), \dots, \mathbf{B}(\mathbf{x}_{MN_M})\}^\top = \left[ \{\mathbf{B}(\mathbf{x}_{ij})\}_{i=1, j=1}^{M, N_i} \right]^\top. \quad (2.5)$$

The  $t$ th unknown random field  $\eta_t(\mathbf{x})$  can be estimated using bivariate splines

## 2.2 Bivariate spline estimator

by  $\eta_t(\mathbf{x}) = \mathbf{B}^\top(\mathbf{x})\boldsymbol{\gamma}_t$ , where  $\boldsymbol{\gamma}_t^\top = (\gamma_{t1}, \dots, \gamma_{tp})$  is the spline coefficient vector. We show that the smoothness constraint in the derivative can be expressed as 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  $S_d^r(\Delta)$ . For a more detailed description of  $\mathbf{H}$ , refer to Section B.2 of the Supplementary Material of Yu et al. (2020). Thus,  $\eta_t(\mathbf{x})$  is obtained by the solving the following least squares problem:

$$\hat{\eta}_t(\mathbf{x}) = \arg \min_{g(\cdot) \in S_d^r(\Delta)} \sum_{i=1}^M \sum_{j=1}^{N_i} \{Y_{t,ij} - g(\mathbf{x}_{ij})\}^2, \quad (2.6)$$

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

To remove the constraint, we consider the QR decomposition of  $\mathbf{H}^\top$ :  $\mathbf{H}^\top = \mathbf{Q}\mathbf{R} = (\mathbf{Q}_1 \mathbf{Q}_2) \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{pmatrix}$ , where  $\mathbf{Q}$  is orthogonal and  $\mathbf{R}$  is upper triangular, and 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

$$\sum_{i=1}^M \sum_{j=1}^{N_i} \{Y_{t,ij} - \mathbf{B}(\mathbf{x}_{ij})\mathbf{Q}_2\boldsymbol{\beta}_t\}^2. \quad (2.7)$$

Denote  $\tilde{\mathbf{B}}(\mathbf{x}) = \mathbf{Q}_2^\top \mathbf{B}(\mathbf{x})$ ,  $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{Q}_2$ , 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  $\hat{\beta}_t = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \mathbf{Y}_t$ ,  $\hat{\gamma}_t = \mathbf{Q}_2 \hat{\beta}_t$ . Thus, the estimator of  $\eta_t(\cdot)$  is  $\hat{\eta}_t(\mathbf{x}) = \tilde{\mathbf{B}}(\mathbf{x})^\top (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \mathbf{Y}_t$ , and one can estimate the unknown mean function  $m(\cdot)$  as

$$\hat{m}(\cdot) = n^{-1} \sum_{t=1}^n \hat{\eta}_t(\cdot). \quad (2.8)$$

### 3. Main results

#### 3.1 Technical assumptions

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

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

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  $\nu$ th partial derivative of  $g$  with respect to the variable  $x$ . In addition, denote a class of Lipschitz continuous functions by  $\text{Lip}(\Omega, L) = \{g(\mathbf{x}) : |g(\mathbf{x}) - g(\mathbf{x}')| \leq L |\mathbf{x} - \mathbf{x}'|, \forall \mathbf{x}, \mathbf{x}' \in \Omega, L > 0\}$ .

To study the asymptotic properties of the bivariate spline estimator  $\hat{m}(\cdot)$ , we need the following 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  $\sigma(\cdot) \in \text{Lip}(\Omega, L)$ ,

### 3.1 Technical assumptions

for some  $L > 0$ , and there exist some positive constants  $M_\sigma, c_G$ , and  $C_G$ , such that  $\sup_{\mathbf{x} \in \Omega} |\sigma(\mathbf{x})| \leq M_\sigma, c_G \leq G(\mathbf{x}, \mathbf{x}) \leq C_G, \mathbf{x} \in \Omega$ .

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

(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  $n \rightarrow \infty, \sum_{k=1}^{\infty} \|\phi_k\|_{\infty, \Omega} = \mathcal{O}(n^{-1/2}), \sum_{k=1}^{k_n} |\phi_k|_{d+1, \Omega, \infty} |\Delta|^{d+1} = \mathcal{O}(1)$ , and  $k_n = \mathcal{O}(n^\alpha)$ , for some  $\alpha > 0$ .

(A5) There exist 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 i.i.d.  $N(0, 1)$  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  $\mathbb{Z}^+ \rightarrow \mathbb{Z}^+$  with the following property:

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

### 3.1 Technical assumptions

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

(A5') The i.i.d. 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 the FPC score white noise  $\{\zeta_{tk}\}_{t=1,k=1}^{n,\infty}$  is finite. There exist constants  $r_1 > 4 + 2\alpha$ ,  $r_2 > (2 + \theta)/\omega$ , such that for  $1 \leq t \leq n$ ,  $1 \leq i \leq M$ ,  $1 \leq j \leq N_i$ , and  $1 \leq k \leq \infty$ ,  $\mathbb{E}\zeta_{tk}^{r_1} + \mathbb{E}\varepsilon_{t,ij}^{r_2} < \infty$ .

(A6) The triangulations are  $\pi$ -quasi-uniform: that is, there exists a positive constant  $\pi$  such that  $(\min_{T \in \Delta} \rho_T)^{-1} |\Delta| \leq \pi$ . The smoothness parameter  $r$  satisfies  $d \geq 3r + 2$  for  $d$  in Assumption (A1). The size of the triangulations  $|\Delta|$  satisfies  $|\Delta|^{-1} = N^\gamma d_N$ , for some  $\gamma > 0$ , with  $d_N + d_N^{-1} = \mathcal{O}(\log^\vartheta N)$ , for some  $\vartheta > 0$  as  $N \rightarrow \infty$ , and for  $d$  in Assumption (A1),  $\theta$  in Assumption (A3),  $\beta_2$  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 bivariate spline smoothers in the nonparametric estimation literature, and controls the size of the bias of the estimator for  $m(\cdot)$ , and can be relaxed by requiring only  $m(\cdot) \in C^0(\Omega)$  if the imaging data have sharp edges; see Wang et al. (2020). Assumption (A2) ensures that variance function should be uniformly bounded. As-

### 3.1 Technical assumptions

sumption (A3) requires that sample size  $n$  grows not faster than the power  $\theta$  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 for the 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 the triangulations.

**Remark 1.** The assumptions above are quite mild and satisfied easily in many practical situations. One simple and reasonable setup for the parameters  $d$ ,  $\theta$ ,  $\omega$ ,  $\gamma$ , 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 our implementation, in Section 4.

**Remark 2.** The pixel locations  $\{\mathbf{x}_{ij}\}_{i=1,j=1}^{M,N_i}$  can be relaxed to vary over subjects (namely, time) as  $\{\mathbf{x}_{t,ij}\}_{t=1,i=1,j=1}^{n,M_t,N_{ti}}$ , as long as the dense condition (\*) in Assumption (A4) is replaced with  $\min_{1 \leq t \leq n} \|\mathbf{x}_{t,f_{t,1}(x)f_{t,2}(x)} - \mathbf{x}_{t,f_{t,1}(y)f_{t,2}(y)}\|_2 = \mathcal{O}(N^{-1/2})$ , with corresponding functions  $f_{t,1}, f_{t,2}$ , for  $t = 1, \dots, n$ . In this scenario, the main theoretical results, including Theorem 1 and Theorem 2, still hold because the order of the smoothing bias does not change. However, in implementation, the evaluation matrix  $\mathbf{X}$  varies over subjects, making it difficult to compute the spectral decomposition of  $G_\varphi(\mathbf{x}, \mathbf{x}')$  defined in (3.1) because  $G_\varphi(\mathbf{x}, \mathbf{x}')$  cannot be simplified as (4.2). Moreover, the triangulation selection should be conducted over each image under the setting of

---

### 3.2 Asymptotic properties of $\tilde{m}(x)$ and $\hat{m}(x)$

varying pixel locations, causing an additional heavy computational burden. Therefore, we assume the longitudinal imaging data are collected at the same locations over time, without loss of generality.

**Remark 3.** From Assumptions (A3) and (A6), the upper bound of  $\theta$  is  $\theta < (d + 2)/(3 + d) < 1$ , which implies that 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), where the convergence rate is  $(nh)^{-1/2}$  with the bandwidth  $h \rightarrow 0$ , which is 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 to maintain the convergence rate  $n^{-1/2}$ , we impose additional requirements on other parameters, which guarantees that smoothing over each image has a negligible effect. However, under the sparse functional data setting, where the number of observations in each trajectory has a finite expectation, we need to pool all observations together to estimate its mean function, leading to different asymptotic results.

### 3.2 Asymptotic properties of $\tilde{m}(x)$ and $\hat{m}(x)$

Denote

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

### 3.2 Asymptotic properties of $\tilde{m}(x)$ and $\hat{m}(x)$

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

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

where

$$G_\varphi(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^\infty \phi_k(\mathbf{x})\phi_k(\mathbf{x}') \left\{ 1 + 2 \sum_{t=0}^\infty \sum_{t'=t+1}^\infty a_{tk}a_{t'k} \right\}, \quad \mathbf{x}, \mathbf{x}' \in \Omega. \quad (3.1)$$

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  $\varphi(\mathbf{x})$  over  $\Omega$ , that is,

$$P \left[ \sup_{\mathbf{x} \in \Omega} |\varphi(\mathbf{x})| \leq Q_{1-\alpha} \right] = 1 - \alpha. \quad (3.2)$$

The following theorem presents the local and global asymptotic properties of the infeasible estimator  $\bar{m}(\cdot)$  in (2.4).

**Theorem 1.** *Under Assumptions (A1) and (A3)–(A5), for  $\alpha \in (0, 1)$ , as  $n \rightarrow \infty$ , the infeasible estimator  $\bar{m}(\cdot)$  converges at the  $\sqrt{n}$  rate to  $m(\cdot)$  with asymptotic covariance*

### 3.2 Asymptotic properties of $\tilde{m}(x)$ and $\hat{m}(x)$

function  $G_\varphi(\mathbf{x}, \mathbf{x}')$ , and thus

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

**Remark 4.** The convergence rate  $n^{-1/2}$  in Theorem 1 is optimal. Cai and Yuan (2011) considered a smoothing spline estimator of the  $p$ -times differentiable mean function when  $\eta_t(x)$  is a univariate process, and showed that its minmax optimal rate is of order  $N^{-p} + n^{-1/2}$  in the  $L_2$ -norm. Bosq (2000) found that the convergence rate of the central limit theorem in functional time series is  $n^{-1/2}$ . Under our setting of a high sampling frequency, the sample size  $n$  is controlled by the number of pixels  $N$ , from Assumptions (A3) and (A6), that is,  $n \ll N$ , and thus the optimal rate remains  $n^{-1/2}$ , and does not depend on  $N$ . That is also why the uniform convergence rate is the same as the pointwise convergence rate.

The following theorem shows that the difference between the bivariate-spline estimator  $\hat{m}(\cdot)$  in (2.8) and the infeasible estimator  $\bar{m}(\cdot)$  is uniformly bounded at the  $\mathcal{O}_p(n^{-1/2})$  rate, which enables us to construct an SCC based on  $\hat{m}(\cdot)$ .

**Theorem 2.** Under Assumptions (A1)–(A6), the bivariate spline estimator  $\hat{m}(\cdot)$  is oracally efficient, that is, it is asymptotically equivalent to  $\bar{m}(\cdot)$  up to order

$\mathcal{O}_p(n^{-1/2})$ ,

$$\sup_{\mathbf{x} \in \Omega} n^{1/2} |\bar{m}(\mathbf{x}) - \hat{m}(\mathbf{x})| = \mathcal{O}_p(1).$$

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

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

$$\hat{m}(\mathbf{x}) \pm G_\varphi(\mathbf{x}, \mathbf{x})^{1/2} Q_{1-\alpha} n^{-1/2}, \quad \mathbf{x} \in \Omega, \quad (3.3)$$

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

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

### 3.3 Extension to nonlinear processes

Noting that the classic MA( $\infty$ ) is a rather broad category, the FMA( $\infty$ ) in (2.3) can approximate a large class of stationary processes, but is restricted to linear processes. Thus, it would be worth extending FMA( $\infty$ ) to nonlinear functional processes. In what follows, we derive the theoretical extension. However, in the remainder of the paper, we focus on FMA( $\infty$ ) for its simple representation and straightforward theoretical properties.

---

Rewrite (2.3) as

$$\xi_t(\cdot) = \sum_{k=1}^{\infty} \xi_{tk} \phi_k(\cdot), \quad \xi_{tk} = F_k(\zeta_{t,k}, \zeta_{t-1,k}, \dots) \quad a.s., \quad (3.4)$$

where  $F_k$ ,  $k \in \mathbb{N}$  are measurable functions from  $\mathbb{R}^{\mathbb{Z}}$  to  $\mathbb{R}$ . It is easy to see that  $\xi_t(\cdot)$  in (3.4) is a nonlinear process with flexible structures. Following Wu (2005), the physical dependence measure is defined as

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

where  $\xi_{tk, \{0\}}$  is identical to  $\xi_{tk}$ , except we replace  $\zeta_{0,k}$  with its i.i.d. copy in (3.4). The next theorem states the asymptotic properties under a nonlinear process setting.

**Theorem 3.** *Under Assumptions (A1)–(A6) and (A5'), if  $\alpha$  in Assumption (A4) satisfies  $\alpha < 1/4$  and  $\sup_{k \in \mathbb{N}} \Delta_{t,k,r_1} = \rho_a^t$ , Theorems 1 and 2 still hold under the nonlinear functional process setting as (3.4), with the corresponding limiting covariance function*

$$G_{\varphi}^*(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^{\infty} \lambda_k^* \phi_k(\mathbf{x}) \phi_k(\mathbf{x}'),$$

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

#### 4. Implementation

## 4.1 Triangulation selection

Triangulation is crucial because bivariate spline fitting can be sensitive to the triangulation selection. Lai and Schumaker (2007) recommend several approaches, such as maxmin-angle triangulations or Delaunay triangulations, but there is no optimal triangulation method. As Yu et al. (2020) note, enough triangles are required to present the domain features, but after reaching the required minimum number of triangles, further increasing the number of triangles usually makes little difference to the fitting process, even leading to empty triangles that do not contain any pixels. 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 the triangulations  $|\Delta|$  needs to satisfy  $|\Delta|^{-1} = N^\gamma d_N$ , for some  $\gamma > 0$ , with  $d_N + d_N^{-1} = \mathcal{O}(\log^\vartheta N)$ , for some  $\vartheta > 0$ . Most widely used triangulation methods can guarantee this condition. We recommend  $|\Delta|^{-1} = cN^{3/16} \log \log N$ , 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  $K$  also measures the size of the triangulations, because there exist that  $K \asymp \lceil |\Delta|^{-1} \rceil$ , where  $\lceil a \rceil$  denotes the integer part of  $a$ , with  $|\Delta| = \sqrt{2}/K$  under the unit square domain as a special case. As  $K$  increases, the fineness of the triangulation increases. We suggest selecting  $K$  from the integers in  $[0.1N^{3/16} \log \log N, N^{3/16} \log \log N]$ . Of the triangulations indexed by

$K$ , we choose the one with the minimal MISE of the estimator  $\widehat{m}(\cdot)$  in (2.8), which is defined as

$$\text{MISE}(K) = \int_{\Omega} \mathbb{E} \{m(\mathbf{x}) - \widehat{m}(\mathbf{x})\}^2 d\mathbf{x}.$$

Because the explicit form of  $\text{MISE}(K)$  is tedious (see Ma (2014)), we propose computing it using discretization and summation, that is,

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

where  $L$  is the number of pre-simulations, with default value 20. Figures 5 to 7 show triangulations on three different domains (square, regular 12 polygon, and regular 12 polygon with a square hole), with  $K = 3, 4, 5$ .

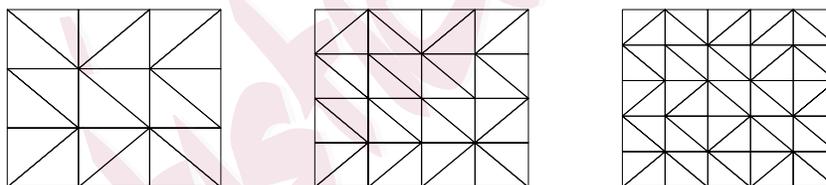


Figure 5: Triangulations on the square domain with  $K = 3$ (left),  $K = 4$ (median) and  $K = 5$ (right).

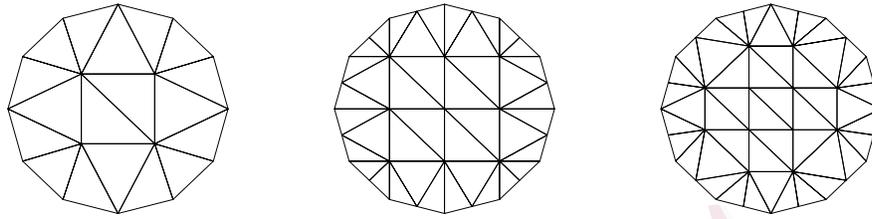


Figure 6: Triangulations on the regular 12 polygon domain with  $K = 3$ (left),  $K = 4$ (median) and  $K = 5$ (right).

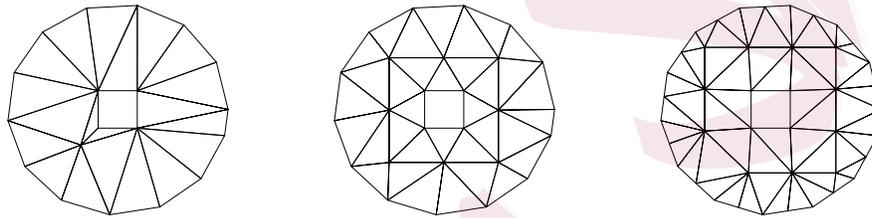


Figure 7: Triangulations on the regular 12 polygon with a square hole domain with  $K = 3$ (left),  $K = 4$ (median) and  $K = 5$ (right).

## 4.2 Covariance estimation

Denote  $\hat{\xi}_t(\mathbf{x}) = \hat{\eta}_t(\mathbf{x}) - \hat{m}(\mathbf{x})$ , for  $t = 1, \dots, n$ ,  $\mathbf{x} \in \Omega$ . To estimate the covariance function  $G_\varphi(\mathbf{x}, \mathbf{x}')$ , we divide  $\{\hat{\xi}_t(\cdot)\}_{t=1}^n$  into  $l$  groups in order, where each group has  $B = \lfloor n^m \rfloor$  samples, for some constants  $m > 0$  with  $l = \lfloor n/B \rfloor$ . Noting that  $\hat{G}_\varphi(\cdot, \cdot)$  is the limit of the covariance function of the process  $\sqrt{n} \{\bar{m}(\cdot) - \hat{m}(\cdot)\}$ , we use  $\hat{m}(\mathbf{x})$  to mimic  $m(\mathbf{x})$  and  $\sqrt{B} \hat{\delta}_j(\mathbf{x})$  to mimic the points from the process  $\sqrt{n} \{\bar{m}(\cdot) - \hat{m}(\cdot)\}$ ,

where

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

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

$$\widehat{G}_\varphi(\mathbf{x}, \mathbf{x}') = \frac{B}{l} \sum_{j=1}^l \left\{ \widehat{\delta}_j(\mathbf{x}) \widehat{\delta}_j(\mathbf{x}') - \bar{\delta}(\mathbf{x}) \bar{\delta}(\mathbf{x}') \right\}, \quad \mathbf{x}, \mathbf{x}' \in \Omega, \quad (4.1)$$

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

**Theorem 4.** *Under Assumptions (A1)–(A6), for a 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  $\widehat{G}_\varphi(\mathbf{x}, \mathbf{x}')$  of  $G_\varphi(\mathbf{x}, \mathbf{x}')$  is uniformly consistent in probability, that is,*

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

Throughout this section, we choose  $B = \lceil n^{1/5} \log \log n \rceil$ .

### 4.3 Estimating the percentile

Recalling that the solution of (2.7) is  $\widehat{\eta}_t(\mathbf{x}) = \widetilde{\mathbf{B}}(\mathbf{x})^\top \widehat{\boldsymbol{\beta}}_t$ , let  $\widehat{\boldsymbol{\beta}} = n^{-1} \sum_{t=1}^n \widehat{\boldsymbol{\beta}}_t$ . Then, the bivariate spline estimator  $\widehat{m}(\mathbf{x}) = \widetilde{\mathbf{B}}(\mathbf{x})^\top \widehat{\boldsymbol{\beta}}$ . Denote  $\widehat{\boldsymbol{\beta}}_{\delta_j} = B^{-1} \sum_{k=B(j-1)+1}^{Bj} (\widehat{\boldsymbol{\beta}}_k - \widehat{\boldsymbol{\beta}})$  and the matrix  $\widehat{g}_\varphi = l^{-1} \sum_{j=1}^l (\widehat{\boldsymbol{\beta}}_{\delta_j} \widehat{\boldsymbol{\beta}}_{\delta_j}^\top) - \left( l^{-1} \sum_{j=1}^l \widehat{\boldsymbol{\beta}}_{\delta_j} \right) \left( l^{-1} \sum_{j=1}^l \widehat{\boldsymbol{\beta}}_{\delta_j} \right)^\top$ . The co-

variance function estimator  $\widehat{G}_\varphi(\mathbf{x}, \mathbf{x}')$  allows the bivariate spline expansion as

$$\widehat{G}_\varphi(\mathbf{x}, \mathbf{x}') = \widetilde{\mathbf{B}}(\mathbf{x})^\top \widehat{g}_\varphi \widetilde{\mathbf{B}}(\mathbf{x}'). \quad (4.2)$$

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

$$\int_{\Omega} \widehat{G}_\varphi(\mathbf{x}, \mathbf{x}') \widehat{\psi}_{k,\varphi}(\mathbf{x}') d\mathbf{x}' = \widehat{\lambda}_{k,\varphi} \widehat{\psi}_{k,\varphi}(\mathbf{x}). \quad (4.3)$$

The next corollary is derived directly from Theorem 4.

**Corollary 2.** *Under the conditions in Theorem 4, the corresponding eigen-pairs  $\{\widehat{\lambda}_{k,\varphi}, \widehat{\psi}_{k,\varphi}(\mathbf{x})\}$ ,  $k \in \mathbb{N}$ , in (4.3) have uniform consistency in probability, that is, for  $k \in \mathbb{N}$ ,*

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

Note that the integration in the eigenequation (4.3) can be approximated using

### 4.3 Estimating the percentile

a discrete summation. Plugging in the covariance function estimator (4.2) leads to

$$N^{-1}\widehat{g}_\varphi\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}\widehat{\gamma}_k = \widehat{\lambda}_{k,\varphi}\widehat{\gamma}_k. \quad (4.4)$$

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 use the Cholesky decomposition,  $N^{-1}\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}} = L_{\widetilde{\mathbf{X}}}L_{\widetilde{\mathbf{X}}}^\top$ . Therefore, solving (4.4) is equivalent to solving  $L_{\widetilde{\mathbf{X}}}^\top\widehat{g}_\varphi L_{\widetilde{\mathbf{X}}}L_{\widetilde{\mathbf{X}}}^\top\widehat{\gamma}_k = \widehat{\lambda}_{k,\varphi}L_{\widetilde{\mathbf{X}}}^\top\widehat{\gamma}_k$ ; that is  $\widehat{\lambda}_{k,\varphi}$  and  $L_{\widetilde{\mathbf{X}}}^\top\widehat{\gamma}_k$  are the eigenvalues and unit eigenvectors, respectively, of  $L_{\widetilde{\mathbf{X}}}^\top\widehat{g}_\varphi L_{\widetilde{\mathbf{X}}}$ . Thus,  $\widehat{\gamma}_k$  can be obtained by multiplying  $(L_{\widetilde{\mathbf{X}}}^\top)^{-1}$  immediately after the unit eigenvectors of  $L_{\widetilde{\mathbf{X}}}^\top\widehat{g}_\varphi L_{\widetilde{\mathbf{X}}}$ . After that,  $\widehat{\psi}_{k,\varphi}(\mathbf{x})$  are obtained and  $\widehat{\phi}_{k,\varphi}(\mathbf{x}) = \widehat{\lambda}_{k,\varphi}^{1/2}\widehat{\psi}_{k,\varphi}(\mathbf{x})$ . Next, the truncated number  $\kappa$  of eigenfunctions is chosen using the following efficient criteria, that is,  $\kappa = \arg \min_{1 \leq v \leq T} \left\{ \sum_{k=1}^v \widehat{\lambda}_{k,\varphi} / \sum_{k=1}^T \widehat{\lambda}_{k,\varphi} > 0.95 \right\}$ , where  $\left\{ \widehat{\lambda}_{k,\varphi} \right\}_{k=1}^T$  are the first  $T$  estimated positive eigenvalues.

We then simulate  $\widehat{\zeta}_b(\mathbf{x}) = \widehat{G}_\varphi(\mathbf{x}, \mathbf{x}')^{-1/2} \sum_{k=1}^{\kappa} Z_{k,b} \widehat{\phi}_k(\mathbf{x})$ , where  $\{Z_{k,b}\}_{k=1,b=1}^{\kappa,b_M}$  are i.i.d. standard normal variables, and  $b_M$  is a preset large integer, with default value 1000. We take the maximal absolute value for each copy of  $\widehat{\zeta}_b(\mathbf{x})$ , and use 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

$$\widehat{m}(\mathbf{x}) \pm n^{-1/2} \widehat{G}_\varphi(\mathbf{x}, \mathbf{x})^{1/2} \widehat{Q}_{1-\alpha}, \quad \mathbf{x} \in \Omega. \quad (4.5)$$

---

## 5. Simulation Studies

In this section, we perform simulations to illustrate the finite-sample performance of the proposed method. The data are generated from the following model:

$$Y_{t,ij} = m(\mathbf{x}_{ij}) + \sum_{k=1}^7 \xi_{tk} \phi_k(\mathbf{x}_{ij}) + \sigma(\mathbf{x}_{ij}) \varepsilon_{t,ij}, \quad t = 1, \dots, n, \quad (5.1)$$

where  $\mathbf{x}_{ij} = (s_{ij}, t_{ij}) \in \Omega \subset [0, 1]^2$ , for  $i = 1, \dots, M, j = 1, \dots, N_i$ , and  $\sum_{i=1}^M N_i = N$ .

We consider three shapes of the domain  $\Omega$ : a square, a regular 12 polygon, which can be viewed as an approximation of a circle, and a regular 12 polygon with a square hole. The mean function  $m(\cdot)$  and eigenfunctions  $\phi_k(\cdot)$  are set as follows:

$$m(s, t) = 6 \sin(s + t) e^{-2(s+t)} + 3s \sin t,$$

$$\varphi_1(s, t) = \sin(\pi t/2) \sin(3\pi s/2),$$

$$\varphi_2(s, t) = \sin(3\pi t/2) \sin(\pi s/2),$$

$$\varphi_3(s, t) = \sin(3\pi t/2) \sin(3\pi s/2),$$

$$\varphi_4(s, t) = \sin(5\pi t/2) \sin(3\pi s/2),$$

$$\varphi_5(s, t) = \sin(3\pi t/2) \sin(5\pi s/2),$$

$$\varphi_6(s, t) = \sin(5\pi t/2) \sin(5\pi s/2),$$

$$\varphi_7(s, t) = \sin(5\pi t/2) \sin(7\pi s/2),$$

---

$$\varphi_k(s, t) = 0, \quad k \geq 8.$$

To guarantee the orthogonality of eigenfunctions, we use Schmidt orthogonalization and obtain  $\varphi_k^*(s, t)$ , for  $k = 1, 2, \dots$ . It is obvious that  $\varphi_k^*(s, t) = \varphi_k(s, t)$  in the square domain case, whereas  $\varphi_k^*(s, t)$  is a linear combination of  $\{\varphi_k(s, t)\}_{k=1}^\infty$  in the other situations. Then, let  $\phi_k(s, t) = \sqrt{\lambda_k} \varphi_k^*(s, t)$ , with  $\lambda_k = 2^{-(k-1)/2}$ , and the FPC scores  $\{\xi_{tk}\}_{t=1, k=1}^{n, 7}$  are generated from (2.3), where  $\{\zeta_{tk}\}_{t=-5, k=1}^{n, 7}$  are i.i.d.  $N(0, 1)$  variables and  $a_{0k} = 0.8$ ,  $a_{1k} = a_{2k} = 0.4$ ,  $a_{3k} = a_{4k} = a_{5k} = a_{6k} = -0.1$ , and  $a_{tk} = 0$ , for  $t \geq 2$ ,  $k = 1, \dots, 7$ .

We generate homoscedastic measurement errors  $\sigma(\mathbf{x}) = 0.1$  and heteroscedastic measurement errors  $\sigma(\mathbf{x}) = 0.1(5 - \exp(-(s+t)))/(5 + \exp(-(s+t)))$ . The errors  $\{\varepsilon_{t,ij}\}_{t=1, i=1, j=1}^{n, M, N_i}$  are i.i.d. with normal, uniform and Laplace distributions. The number of pixels  $N$  is 10000 and 20000, respectively, and the number of images  $n$  is taken to be  $\lceil N^{1/4} \log N (\log \log N)^2 \rceil$ .

Throughout this section, the mean function is estimated using bivariate splines in the space  $S_d^r(\Delta)$ , with  $d = 5$  and  $r = 1$ , which approaches the full approximation power asymptotically; see Lai and Schumaker (2007). Tables 1 and 2 display the empirical coverage rate, namely, the percentage of the 500 replications of the true mean function  $m(\cdot)$  covered by the bivariate spline SCCs (4.5) at the  $N$  points  $\{\mathbf{x}_{ij}\}_{i=1, j=1}^{M, N_i}$ . In both scenarios, the coverage rate of the SCC becomes closer to the

nominal confidence level as the sample size increases, which is a positive confirmation of the asymptotic theory.

To visualize the SCCs for the mean function, Figures 8 to 16 show the estimated mean functions and their 95% SCCs for the true mean function  $m(\cdot)$ , with  $\sigma(\mathbf{x}) = 0.1$ ,  $\varepsilon_{t,ij} \sim N(0, 1)$ , and  $N = 10000, 20000, 40000$ , respectively, on the 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 covered entirely by the upper and lower corridors.

Table 1: Coverage frequencies from 500 replications based on homoscedastic errors  $\sigma(s, t) \equiv 0.1$ .

Domain $\Omega$	$N$	10000			20000		
	Distribution of $\varepsilon$	normal	uniform	Laplace	normal	uniform	Laplace
Square	$\alpha = 0.10$	0.858	0.856	0.874	0.894	0.884	0.904
	$\alpha = 0.05$	0.916	0.916	0.932	0.942	0.952	0.946
	$\alpha = 0.025$	0.954	0.960	0.958	0.974	0.974	0.974
	$\alpha = 0.01$	0.984	0.984	0.982	0.994	0.984	0.994
Regular 12 polygon	$\alpha = 0.10$	0.738	0.724	0.720	0.894	0.898	0.884
	$\alpha = 0.05$	0.858	0.846	0.848	0.952	0.956	0.948
	$\alpha = 0.025$	0.926	0.926	0.928	0.976	0.976	0.980
	$\alpha = 0.01$	0.980	0.968	0.968	0.994	0.988	0.994
Regular 12 polygon with a square hole	$\alpha = 0.10$	0.838	0.842	0.850	0.904	0.900	0.902
	$\alpha = 0.05$	0.928	0.928	0.944	0.95	0.948	0.956
	$\alpha = 0.025$	0.972	0.962	0.974	0.976	0.974	0.978
	$\alpha = 0.01$	0.984	0.980	0.984	0.984	0.990	0.992

## 6. Real-Data Analysis

In this section, we apply the proposed SCCs to two seawater potential temperature data sets observed on a typically complicated domain. Seawater potential temper-

Table 2: Coverage frequencies from 500 replications based on heteroscedastic errors  $\sigma(s, t) = 0.1(5 - \exp(-(s + t))) / (5 + \exp(-(s + t)))$ .

Domain $\Omega$	$N$	10000			20000		
	Distribution of $\varepsilon$	normal	uniform	Laplace	normal	uniform	Laplace
Square	$\alpha = 0.10$	0.858	0.866	0.878	0.898	0.902	0.904
	$\alpha = 0.05$	0.936	0.932	0.932	0.948	0.952	0.946
	$\alpha = 0.025$	0.970	0.964	0.966	0.976	0.974	0.976
	$\alpha = 0.01$	0.984	0.976	0.988	0.992	0.990	0.992
Regular 12 polygon	$\alpha = 0.10$	0.736	0.740	0.746	0.880	0.898	0.892
	$\alpha = 0.05$	0.850	0.862	0.870	0.940	0.940	0.944
	$\alpha = 0.025$	0.926	0.944	0.934	0.970	0.970	0.968
	$\alpha = 0.01$	0.962	0.970	0.970	0.992	0.992	0.990
Regular 12 polygon with a square hole	$\alpha = 0.10$	0.866	0.862	0.870	0.898	0.904	0.896
	$\alpha = 0.05$	0.926	0.934	0.924	0.954	0.954	0.952
	$\alpha = 0.025$	0.968	0.976	0.958	0.976	0.976	0.976
	$\alpha = 0.01$	0.988	0.988	0.986	0.992	0.988	0.990

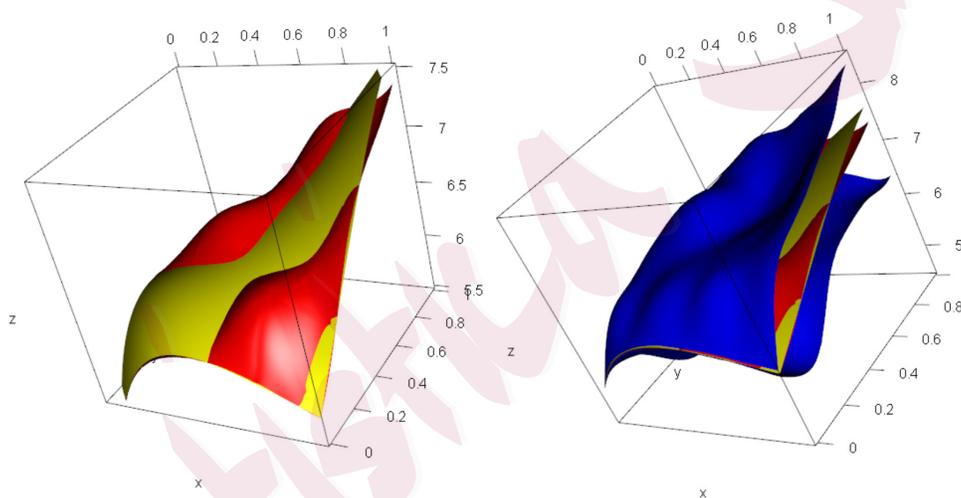


Figure 8: Plot of true mean function (yellow surface), bivariate spline estimator (red surface) and its 95% simultaneous confidence corridors (blue surfaces) on the square domain with the number of pixels  $N = 10000$ .

ature is an important factor in marine hydrological conditions, and is often used as a principal indicator when studying the properties and movement of water masses.

Investigating the temporal and spatial distribution and changing laws related to sea

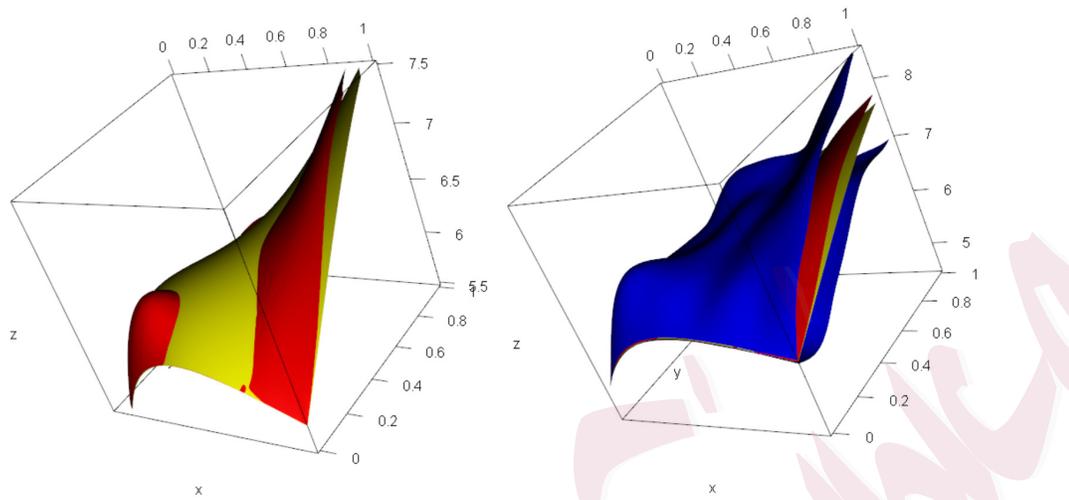


Figure 9: Plot of true mean function (yellow surface), bivariate spline estimator (red surface) and its 95% simultaneous confidence corridors (blue surfaces) on the square domain with the number of pixels  $N = 20000$ .

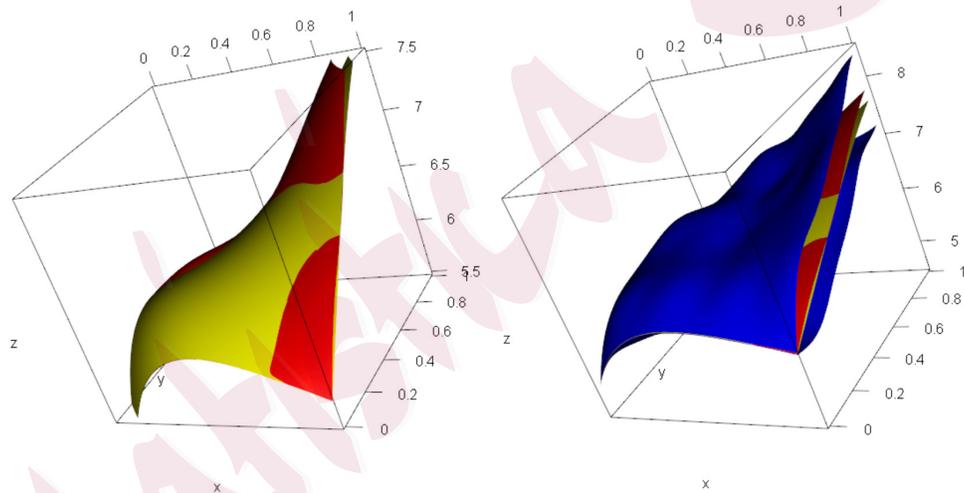


Figure 10: Plot of true mean function (yellow surface), bivariate spline estimator (red surface) and its 95% simultaneous confidence corridors (blue surfaces) on the square domain with the number of pixels  $N = 40000$ .

temperature are significant for marine fishing, aquaculture, and marine operations.

The data sets used in our analysis are from the CMEMS global analysis and fore-

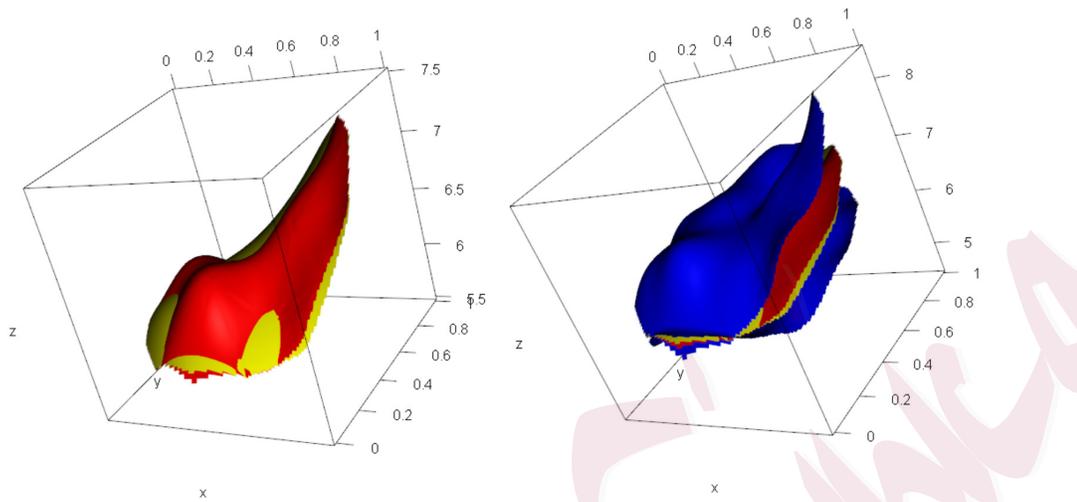


Figure 11: Plot of true mean function (yellow surface), bivariate spline estimator (red surface) and its 95% simultaneous confidence corridors (blue surfaces) on the regular 12 polygon domain with the number of pixels  $N = 10000$ .

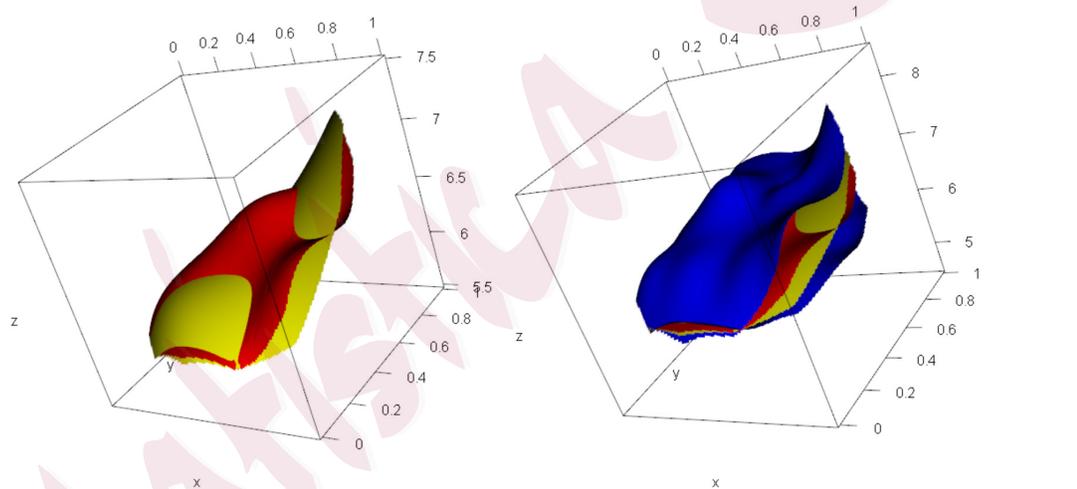


Figure 12: Plot of true mean function (yellow surface), bivariate spline estimator (red surface) and its 95% simultaneous confidence corridors (blue surfaces) on the regular 12 polygon domain with the number of pixels  $N = 20000$ .

cast product, available at <https://resources.marine.copernicus.eu>. CMEMS collects rough data, such as 3D potential temperature, salinity, and currents, bottom poten-

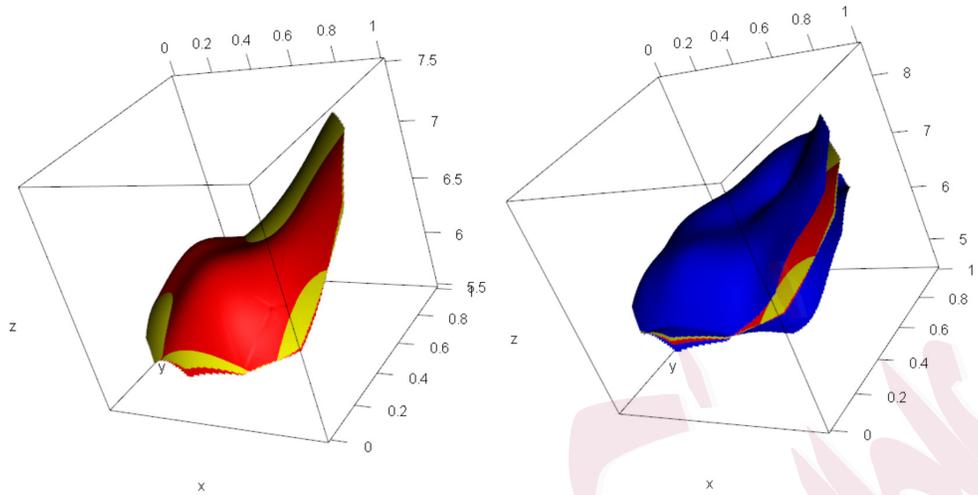


Figure 13: Plot of true mean function (yellow surface), bivariate spline estimator (red surface) and its 95% simultaneous confidence corridors (blue surfaces) on the regular 12 polygon domain with the number of pixels  $N = 40000$ .

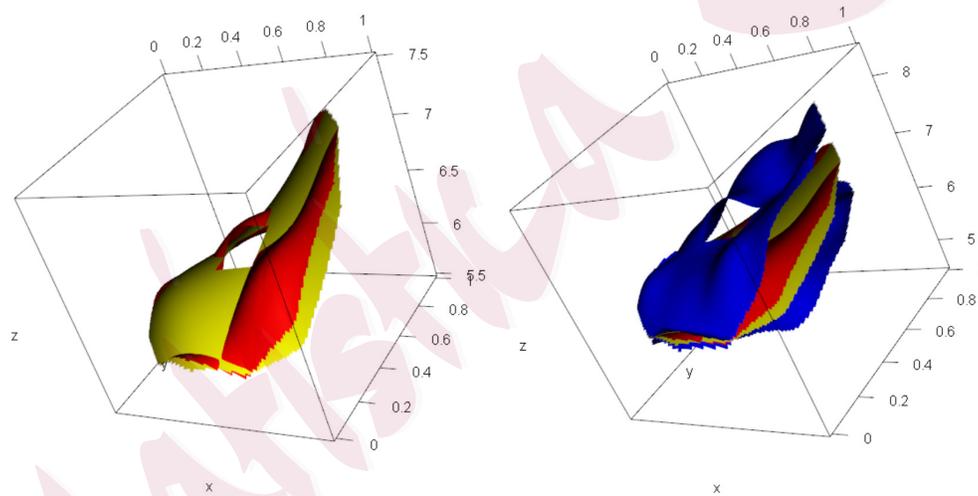


Figure 14: Plot of true mean function (yellow surface), bivariate spline estimator (red surface) and its 95% simultaneous confidence corridors (blue surfaces) on the regular 12 polygon with a square hole domain with the number of pixels  $N = 10000$ .

tial temperature, or mixed layer thickness, and then transform it using an algorithm.

All data are recorded globally on a standard grid at  $1/12$  degree (approximately 8

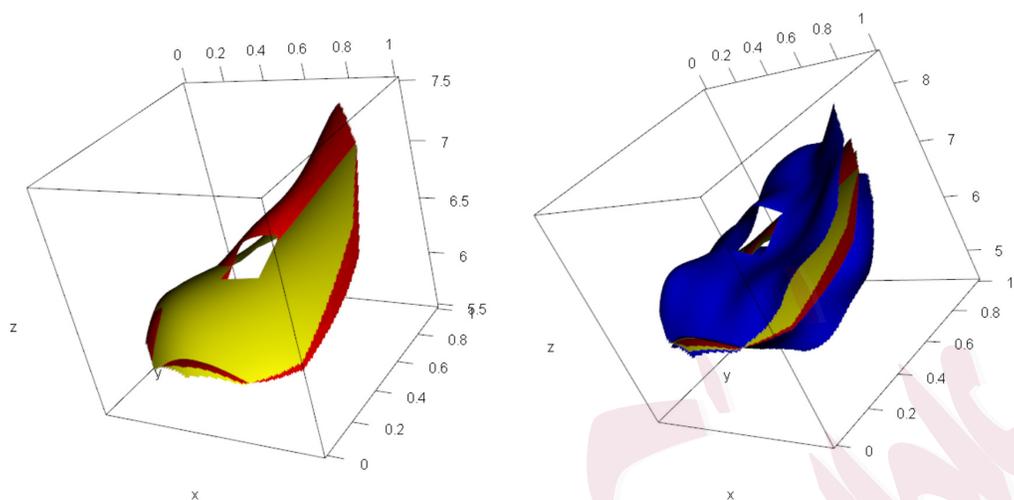


Figure 15: Plot of true mean function (yellow surface), bivariate spline estimator (red surface) and its 95% simultaneous confidence corridors (blue surfaces) on the regular 12 polygon with a square hole domain with the number of pixels  $N = 20000$ .

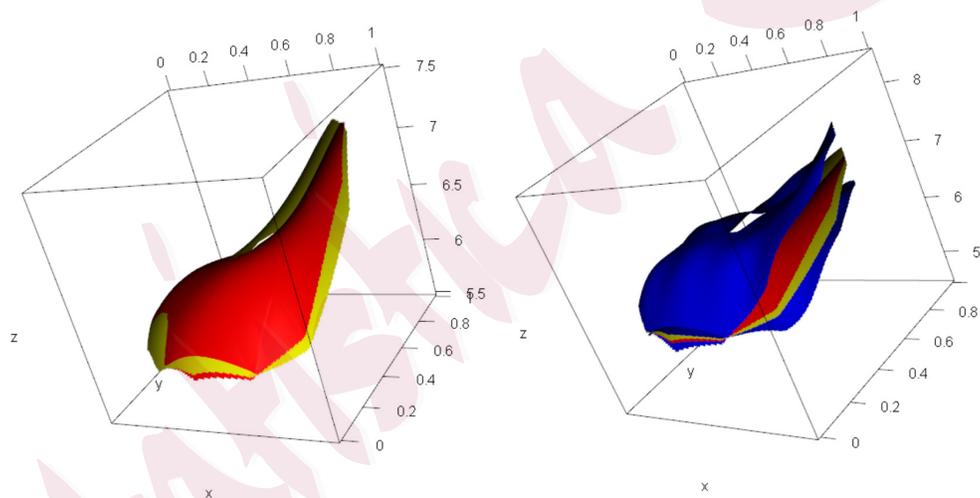


Figure 16: Plot of true mean function (yellow surface), bivariate spline estimator (red surface) and its 95% simultaneous confidence corridors (blue surfaces) on the regular 12 polygon with a square hole domain with the number of pixels  $N = 40000$ .

km) 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  $26.8^{\circ}\text{E}$  to  $42.2^{\circ}\text{E}$  and  $40.5^{\circ}\text{N}$  to  $47.6^{\circ}\text{N}$ ; see the equirectangular projection map in Figure 4. 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 1(a) show the observed data locations. Each hourly observed temperature data of the Black Sea can naturally be regraded as an 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 the seawater potential temperature data, and serves as a preliminary step for further data analysis. We use bivariate splines with the smoothing parameter  $r = 1$  and  $d = 5$  for the estimation of mean function. Figure 1(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 from (4.5) are displayed in Figures 17 and 18, respectively. Figure 17 shows that the average sea surface water temperature decreases from a low latitude to a high latitude, corroborating classic oceanographic theory.

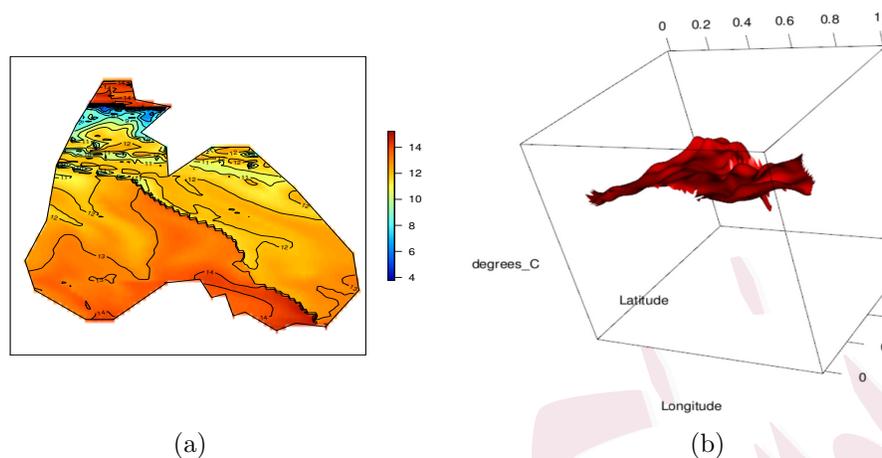


Figure 17: (a) Contour map for the bivariate spline estimator for mean function; (b) 3D plot of the bivariate spline estimator for mean function.

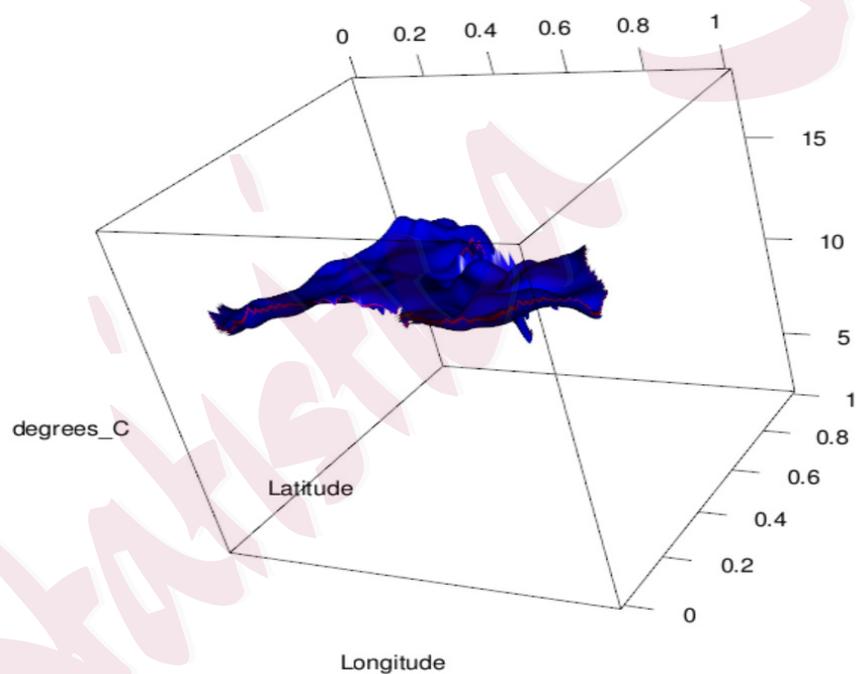


Figure 18: Plots of the bivariate spline estimator for mean function (middle red surface) and 95% SCC (upper and lower blue surfaces).

## 6.2 Madagascar

Madagascar is an island country in the Indian Ocean off the coast of East Africa. We investigate the potential temperature of the sea around Madagascar, ranging from  $41.0^\circ\text{E}$  to  $55.0^\circ\text{E}$  and  $11.0^\circ\text{S}$  to  $30.0^\circ\text{S}$ . Similarly to the previous case, hourly potential temperature is measured on standard grids every  $1/12$  degree; see Figure 3(a) for the pixel locations. The domain of the sea surrounding Madagascar is more complicated, owing 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  $n = 840$  time-ordered images, with  $N = 26151$  pixels per image.

We also use bivariate splines with the smoothing parameter  $r = 1$  and  $d = 5$  to approximate its mean function. The triangulation on the domain of the sea surrounding Madagascar is shown in Figure 3(b), with 33 triangles and 32 vertices. Figures 19 and 20 present the estimated mean function and its corresponding 95%, respectively, SCC computed from (4.5). Figure 19 shows that there is always a higher sea surface temperature near land. This strongly demonstrates that our method is widely applicable and capable of handling a complex image domain, even with a hole.

## 7. Conclusion

We have examined longitudinal imaging data over complicated domains, referring to ordered images with numerous pixels collected at a high frequency over time

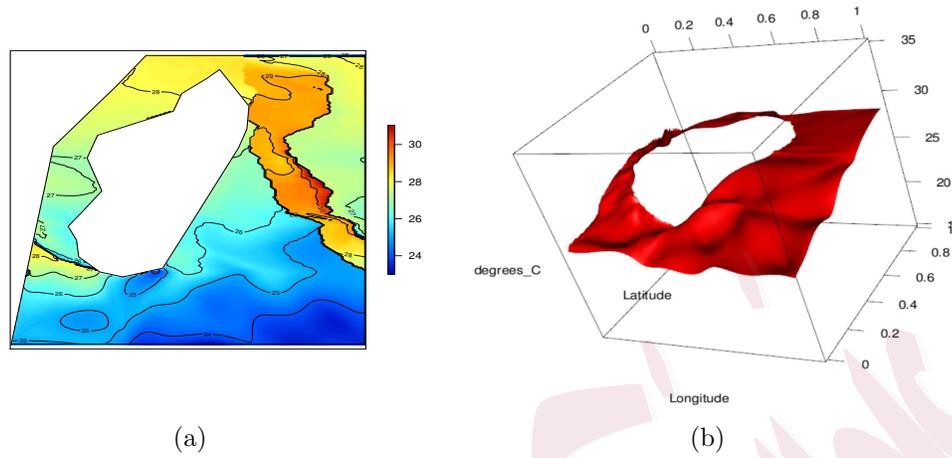


Figure 19: (a) Contour map for the bivariate spline estimator for mean function; (b) 3D plot of the bivariate spline estimator for mean function.

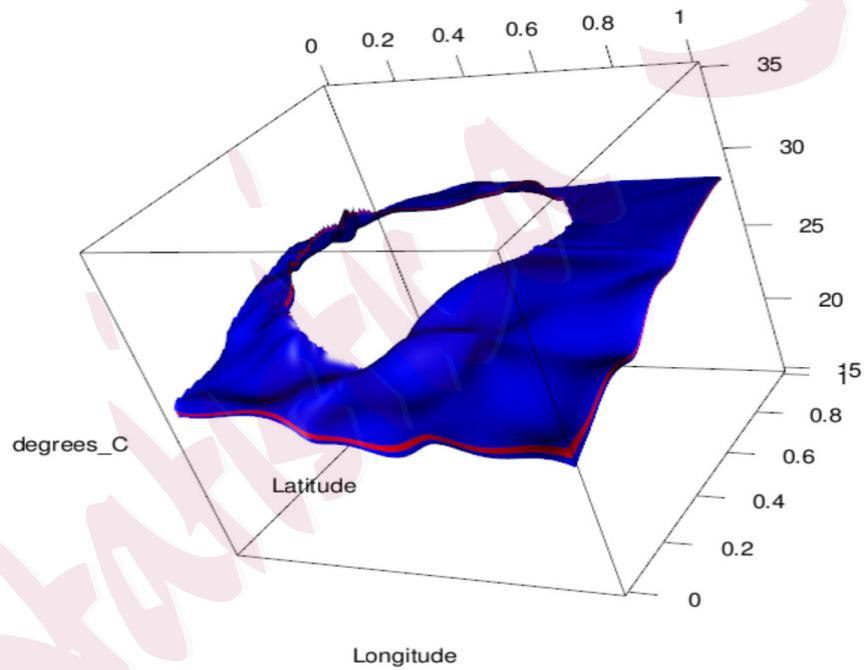


Figure 20: Plots of the bivariate spline estimator for mean function (middle red surface) and 95% SCC (upper and lower blue surfaces).

span. We propose an asymptotically correct and computationally efficient bivariate spline estimator for its mean function. We investigate both the global and the local asymptotic properties of the bivariate estimator, using SCCs to make inference on the true mean function. To the best of our knowledge, this is the first work to focus on large-scale longitudinal imaging data. The proposed method yields attractive inference results, and is free from an ultrahigh dimension and model misspecification. Our method can be applied widely to imaging data in geography, oceanography, and biomedical studies.

Several issues warrant further investigation. The data in Section 6 are considered to come from a horizontal plane, which ignores the curvature of the earth's surface. It may be more accurate to assume that the data are collected on a sphere. Spherical splines, introduced in Lai and Schumaker (2007), may be better able to approximate the aforementioned 3D imaging data. However, this is not an easy task, because of the elusive theory and heavy computational burden of high-resolution 3D images compared with 2D ones. In addition, we wish to extend the proposed methodology to functional regression models. How to construct SCCs of the functional coefficients in such models is also challenging, owing to the deeper asymptotic theory.

## Supplementary Material

The online Supplementary Material contains detailed proofs for the main results and some additional simulation results.

## Acknowledgments

This research was supported by the National Natural Science Foundation of China awarded 11771240 and 12026242. The authors are truly grateful to the editor, associate editor, two reviewers for their constructive comments and suggestions that led to significant improvements in the paper.

## 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.

---

REFERENCES

- 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.
- Csörgő, 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., DellItalia, 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.

---

REFERENCES

- 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 geoaddivitive 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: lijie\_stat@ruc.edu.cn