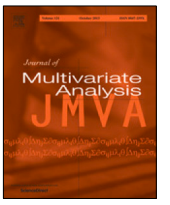
Journal of Multivariate Analysis 180 (2020) 104664

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva



Surface functional models

Ziqi Chen [a,*], Jianhua Hu [b], Hongtu Zhu [c]

[a] School of Statistics, Key Laboratory of Advanced Theory and Application in Statistics and Data Science - MOE, East China Normal University, Shanghai 200062, PR China
[b] Department of Medicine, Columbia University, NY 10032, USA
[c] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

ARTICLE INFO

Article history:
Received 8 January 2020
Received in revised form 19 July 2020
Accepted 19 July 2020
Available online 30 July 2020

AMS 2010 subject classifications:
primary 62H12
secondary 62G05
62G20
62H35
62M10
62M30

Keywords:
Covariance structure
Efficiency
Surface functional response
Time-spatial process
Uniform convergence
Weighting schemes

ABSTRACT

The aim of this paper is to develop a new framework of surface functional models (SFM) for surface functional data which contains repeated observations in two domains (typically, time-location). The primary problem of interest is to investigate the relationship between a response and the two domains, where the numbers of observations in both domains within a subject may be diverging. The SFMs are far beyond the multivariate functional models with two-dimensional predictor variables. Unprecedented complexity presented in the surface functional models, such as possibly distinctive sampling designs and the dependence between the two domains, makes our models more complex than the existing ones. We provide a comprehensive investigation of the asymptotic properties of the local linear estimator of the mean function based on a general weighting scheme, including equal weight (EW), direction-to-denseness weight (DDW) and subject-to-denseness weight (SDW), as special cases. Moreover, we can mathematically categorize the surface data into nine cases according to the sampling designs (sparse, dense, and ultra-dense) of both the domains, essentially based on the relative order of the number of observations in each domain to the sample size. We derive the specific asymptotic theories and optimal bandwidth orders in each of the nine sampling design cases under all the three weighting schemes. The three weighting schemes are compared theoretically and numerically. We also examine the finite-sample performance of the estimators through simulation studies and an autism study involving white-matter fiber skeletons.

1. Introduction

There has been a great interest in the analysis of massive functional data. A particular problem that has received much attention is mean function estimation, for example, Yao et al. [29], Zhang and Wang [31]. Most of the existing methods have focused on the case of independently sampled curve functions, typically functional data in time-domain only, that is, observations are only made over discrete time points within each subject. Such data have been termed curve data [10,23]. Along this line, it is common to pre-smooth the observations of each subject for bias removal and reconstruction of its random curve prior to mean function estimation, when the number of observations N_i for the ith subject is larger than some power of the number of subjects n which corresponds to the case of "dense" data [8,12,22,30]. On the other hand, [21,24,29] pooled all subjects to borrow information across them for mean function estimation and inference, when the number of observations within a subject is bounded by a finite positive number or follows a fixed

* Corresponding author.
E-mail address: zqchen@fem.ecnu.edu.cn (Z. Chen).

https://doi.org/10.1016/j.jmva.2020.104664
0047-259X/

distribution, which is called the case of "sparse" data. Therefore, the estimation and inference procedures are distinct between the "dense" and "sparse" cases. In practice, both cases can be encountered and, even worse, it can be difficult to distinguish between the two to decide which methodology to use [15]. Realizing this practical challenge, [31] intended to address it by classifying the sampled curves into three types of sparse, dense, and ultra-dense data according to the relative order of N_i to n, and by providing a unified approach to handle all the types with the corresponding asymptotic properties.

Along with the advance of technologies in many scientific areas, such as environmental and biomedical research, however, massive data have been generated from multi-domain processes. As an example of biomedical imaging, including functional magnetic resonance imaging (fMRI), electroencephalography (EEG), diffusion tensor imaging (DTI), and positron emission tomography (PET), imaging features related to the pathophysiology and pathogenesis of a disease are collected over a sequence of repeated scan times simultaneously at multiple regions or sites that are often spatially dependent [7,9,14,19,32,33]. We refer to such data as surface functional/longitudinal data. In this problem, spatial–temporal processes are involved. An important attribute of such multi-domain processes is that the sampling designs of different characteristics are not necessarily identical. For example, a standard scenario is called longitudinal functional data in the mentioned biomedical imaging applications, when it could be sparse along the temporal axis but dense along the spatial axis. It is of great interest to investigate the relationship between the response and the two domains (time-spatial) for surface functional models. For example, we may use 64 electrodes to collect the EEG signals on scalp over 90 s for n patients. The mean function describes how the average value of EEG signals changes with time and the location of scalp. The covariance function depicts the covariance between two EEG signals varying with time and location where the two EEG signals are recorded. In the real data analysis, we show the changing pattern of average fractional anisotropy (FA) value with age and location along the corpus callosum for DTI data.

There exists some literature on the estimation of mean regression functions for longitudinal functional models, which fall into a general functional mixed effects modeling framework [4,11,16,18,32]. Guo [11] pioneered in introducing functional mixed effects models for correlated functional data, while [18] and subsequent work by this group proposed general functional mixed effects models with multiple levels of random effect functions as well as curve-to-curve deviations. Chen and Müller [4] described a double functional principal component analysis method that relies on mild assumptions and provided the asymptotic results, however, only in two cases where the recordings of the curves are sparsely or ultra-densely designed while the sampling for each curve along the time coordinate is ultra-dense. Zhu et al. [32] conducted a systematic and theoretical analysis of local linear estimates of mean regression functions for a class of functional mixed effects models developed for longitudinal functional responses, see also references therein. Liebl [16] considered inference for the mean regression functions of covariate adjusted functional data based on local linear estimators.

The aim of this paper is to investigate the estimating procedure and the corresponding asymptotic properties of the mean function estimator for surface functional data. The major contributions of this work are as follows:

(a) We develop a general framework to systematically analyze two-domain surface functional data. The challenges stem from:

(i) the numbers of observations collected over two domains are likely increasing with the sample size n;

(ii) a complex correlation structure within and across the two domains;

(iii) possibly great influence of within-subject dependence on the variance of the mean estimator;

(iv) likely completely different characteristics of the two domains, including sampling designs.

(b) We comprehensively consider all the nine scenarios resulted from the combinations of three sampling designs [31] (sparse, dense, and ultra-dense) of the two domains. The partition of surface functional data according to the two-domain sampling design is based on the relative orders of both the numbers of observations collected over two domains, simultaneously, to the sample size n, which is much more complex than that of the curve functional data [2,15,31].

(c) We provide a unified nonparametric mean function estimation based on local linear smoothers and the unifying theoretical platform that can handle all the nine sampling design scenarios (i.e., allow the magnitude of the number of observations in each domain relative to n to vary freely). We establish the specific asymptotic results for all the nine sampling design scenarios under the three proposed weighting schemes, respectively, including equal weight (EW), direction-to-denseness weight (DDW) and subject-to-denseness weight (SDW). Moreover, we derive the optimal bandwidth orders for estimating mean function for each two-domain sampling design case under all the three weighting schemes.

(d) We compare the EW, DDW and SDW schemes both theoretically and numerically.

As a result, the proposed surface functional model framework is far beyond all existing results for multivariate nonparametric models, functional/longitudinal models and longitudinal functional models.

The rest of the article is organized as follows. In Section 2, we introduce the framework of surface functional models (SFM) and the corresponding nonparametric mean function estimator. It is followed by the formal theoretical results of the estimator presented in Section 3. We investigate the finite-sample performance of the proposed estimator via simulation studies and conduct the analysis of a diffusion tensor imaging data in an autism study in Section 4. At the end, we provide the concluding remarks in Section 5. Some Lemmas are deferred to Appendix A.

2. Methodology

We start with some notation and definition. Without loss of generality, let $Z(t,s)$ be a surface stochastic process defined on $[a,b]\times[c,d]$, where t and s, respectively, denote the time and spatial axes. Suppose that we have n random time-spatial surfaces $Z_1(t,s),\ldots,Z_n(t,s)$ corresponding to n subjects, each of which is an independent realization of $Z(t,s)$. The $Z_i(t,s)$ for the ith subject contains the observations at N_i time points $\times$ M_i positions. That is, we have the total of N_iM_i observations for subject i. In many situations, one can only observe the process $Z(t,s)$ intermittently and with measurement errors. Let Y_{ijk} be the observation of the random surface Z_i of subject i made at (T_{ij},S_{ik}) for $i\in\{1\ldots,n\}$, $j\in\{1,\ldots,N_i\}$, and $k\in\{1,\ldots,M_i\}$. It is assumed that we have

$$Y_{ijk}=Z_i(T_{ij},S_{ik})+\epsilon_{ijk}, \tag{1}$$

where measurement errors ϵ_{ijk}'s are assumed to be independently and identically distributed (i.i.d.) copies of ϵ with mean zero and variance $\sigma^2>0$. Let $\mu(t,s):=\mathrm{E}\{Z(t,s)\}$, and we can express $Z(t,s)$ as

$$Z(t,s)=\mu(t,s)+U(t,s),$$

where $U(t,s)$ is the stochastic part of $Z(t,s)$ and used to characterize the within-surface dependence with $\mathrm{E}\{U(t,s)\}=0$ for any $t\in[a,b]$ and $s\in[c,d]$. Consequently, model (1) can be rewritten as

$$Y_{ijk}=\mu(T_{ij},S_{ik})+U_i(T_{ij},S_{ik})+\epsilon_{ijk},$$

which is the proposed surface functional model. Without loss of generality, we assume T_{ij}'s and S_{ik}'s are, respectively, i.i.d. copies of independent random variables T and S. We define the covariance function of $Z(t,s)$ (or $U(t,s)$) as

$$R(t,s;t',s'):=\mathrm{Cov}\{Z(t,s),Z(t',s')\}=\mathrm{Cov}\{U(t,s),U(t',s')\}.$$

Our question of interest is to estimate the mean function of the surface process $\mu(t,s)$ based on the observations $\{Y_{ijk},\ T_{ij},\ S_{ik}\}$ for $i\in\{1\ldots,n\}$, $j\in\{1,\ldots,N_i\}$, and $k\in\{1,\ldots,M_i\}$. For this purpose, we adopt the local linear smoothing technique [6,26] because of several attractive features, including conceptual simplicity, small bias, good asymptotic efficiency, and well-known ability for automatic boundary correction in the independent case. Extensions of using this technique to the correlated case include [29,31], which demonstrated its excellent theoretical and numerical properties.

We denote the mean estimator as $\hat{\mu}(t,s)=\hat{\beta}_0$, where

$$(\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2)=\operatorname{argmin}\sum_{i=1}^{n}\xi_i\sum_{j=1}^{N_i}\sum_{k=1}^{M_i}\Big\{Y_{ijk}-\beta_0-\beta_1(T_{ij}-t)-\beta_2(S_{ik}-s)\Big\}^2K_{h_1}(T_{ij}-t)K_{h_2}(S_{ik}-s),$$

where $K_{h_i}(\cdot)=K(\cdot/h_i)/h_i$ $(i\in\{1,2\})$ for a kernel function $K(\cdot)$ with h_1 and h_2 being the bandwidth parameters. h_1 and h_2 may be estimated by the leave-one-out cross validation method [6]. The weights ξ_i's satisfy $\sum_{i=1}^{n}\xi_iN_iM_i=1$.

We consider three ways of constructing the weights ξ_i. A simple way is to assign the equal weight to all the subjects (abbreviated as EW), that is, $\xi_i=1/(\sum_{i=1}^{n}N_iM_i)$ [28,31]. EW tends to overly impose the influence of subject i on the optimizer when the time points and/or positions of this subject are dense [15]. Therefore, we also consider other two popular strategies, so called direction-to-denseness weight (DDW) and subject-to-denseness weight (SDW), to assign a smaller weight to a subject who has denser time points and/or positions. Specifically, the DDW scheme $\xi_i=1/\{(\sum_{i=1}^{n}N_i)M_i\}$ is preferred when the location axis has dense data and $\xi_i=1/\{N_i(\sum_{i=1}^{n}M_i)\}$ will be chosen when the time axis has dense data, whereas the SDW scheme $\xi_i=1/(nN_iM_i)$ could be used when both time and location axes have dense data. The EW and SDW schemes used for our surface functional model estimation are respectively consistent with the so-called equal weight per observation scheme used in [29] and the equal weight per subject scheme in [15] in the case of curve functional data; also see [2,31]. Without loss of generality, for DDW, we only consider the case with $\xi_i=1/\{(\sum_{i=1}^{n}N_i)M_i\}$. Such DDW scheme is a mixture of the equal weight per observation scheme and the equal weight per subject scheme, and is expected to be preferable when the time domain has sparse data and the location axis has dense data. In practice, the weighting scheme may be selected by using the leave-one-out cross validation method. See the Supplementary Materials for details.

According to the relative orders of the number of time points and the number of positions to the number of subjects n, both the time and location axes can be partitioned into sparse, dense, and ultra-dense cases. When $\overline{N}/n^{1/4}\to 0$ with $\overline{N}=\sum_{i=1}^{n}N_i/n$, $\overline{N}/n^{1/4}\to C$ with $0<C<\infty$ being a constant, and $\overline{N}/n^{1/4}\to\infty$, the time axis is defined as "Sparse", "Dense", and "Ultra-Dense", respectively [2,31]. This definition still applies to the location axis. Therefore, we may partition the surface functional data into nine cases as in Table 1. We will systematically investigate the applicability of our proposed approach for all these 9 types of surface functional data. Our estimation method for surface functional models does not need to distinguish which types of sampling designs to encounter in practice. In contrast, most existing methods can only handle individual sampling design scenarios for curve functional data [12,29,30].

We will provide a comprehensive presentation of the asymptotic properties of the estimator for all the sampling design cases on a unified platform. The novelties of our theoretical and numerical results include:

(i) We derive the conditional bias, the conditional variance, the asymptotic normality, and the uniform convergence properties of the estimator $\hat{\mu}(t,s)$ for the proposed surface functional models. These results are on a unified platform for all

Table 1
Partition the surface functional data into nine cases.

Sparse+Sparse	Sparse+Dense	Sparse+Ultra-Dense
Dense+Sparse	Dense+Dense	Dense+Ultra-Dense
Ultra-Dense+Sparse	Ultra-Dense+Dense	Ultra-Dense+Ultra-Dense

Table 2
Asymptotic results of $\hat{\mu}_{EW}(t,s)$ for the different sampling cases. S:Sparse; D:Dense; UD:Ultra-Dense; Bandwidth Order: the optimal bandwidth order selected using the AMSE (asymptotic mean-squared error) criterion; Bias: the dominating conditional bias of $\hat{\mu}_{EW}(t,s)$; Variance: the dominating conditional variance of $\hat{\mu}_{EW}(t,s)$; Rate: Convergence rate; $\overline{N} = O(n^{\alpha})$ and $\overline{M} = O(n^{\beta})$.

(α,β)	Case	Bandwidth order	Bias	Variance	Rate
$5\alpha-\beta-1<0$ $5\beta-\alpha-1<0$	S+S	$h_1 \asymp (n\overline{N}\ \overline{M})^{-1/6}$ $h_2 \asymp (n\overline{N}\ \overline{M})^{-1/6}$	$B_{1n}+B_{2n}$	A_{4n}	$n^{-(1+\alpha+\beta)/3}$
$\alpha<1/4$ $5\beta-\alpha-1=0$	S+D	$h_1 \asymp (n\overline{N})^{-1/5}$ $h_2 \asymp (n\overline{N})^{-1/5}$	$B_{1n}+B_{2n}$	$A_{3n}+A_{4n}$	$n^{-2(1+\alpha)/5}$
$\alpha<1/4$ $5\beta-\alpha-1>0$	S+UD	$h_1 \asymp (n\overline{N})^{-1/5}$ $h_2 \asymp (n\overline{N}\ \overline{M}^{5/4})^{-4/25}$	B_{1n}	A_{3n}	$n^{-2(1+\alpha)/5}$
$\alpha=1/4$ $\beta=1/4$	D+D	$h_1 \asymp n^{-1/4}$ $h_2 \asymp n^{-1/4}$	$B_{1n}+B_{2n}$	$\sum_{l=1}^{4} A_{ln}$	$n^{-1/2}$
$\alpha=1/4$ $\beta>1/4$	D+UD	$h_1 \asymp n^{-1/4}$ $h_2 \asymp (n\overline{M})^{-1/5}$	B_{1n}	$A_{1n}+A_{3n}$	$n^{-1/2}$
$\alpha>1/4$ $\beta>1/4$	UD+UD	$h_1 \asymp (n\overline{N})^{-1/5}$ $h_2 \asymp (n\overline{M})^{-1/5}$	$B_{1n}+B_{2n}$	A_{1n}	$n^{-1/2}$

the nine sampling designs under a general weighting scheme (See Theorems 1, 3, and 5). Moreover, these characteristics are different among the nine sampling designs.
(ii) We explicitly specify the partition of the 9 types of sampling design according to the relative orders of both $\overline{N}$ and $\overline{M}$ to the number of subjects n, where $\overline{N} := \sum_{i=1}^{n} N_i/n$ and $\overline{M} := \sum_{i=1}^{n} M_i/n$, for EW scheme, see Fig. 1 and Table 2. Similar partitions are derived when the DDW and SDW schemes are adopted.
(iii) We obtain the optimal bandwidth orders for each sampling design of surface functional data, which are more complicated than those in the existing literature for the nonparametric regression and curve functional regression, see Tables 2–4.
(iv) We make detailed comparisons among the EW, DDW and SDW schemes both theoretically and numerically, and provide the guideline on which one is preferred over the other two for the different types of sampling design, see Sections 3.2 and 4. Specifically, EW, DDW and SW are the most preferable schemes for "Sparse+Sparse" Case, "Sparse+Ultra-Dense" Case and "Ultra-Dense+Ultra-Dense" Case in terms of estimating efficiency, respectively.

Remark 1. Our proposed two-domain surface functional models differ significantly from one-domain multivariate functional/longitudinal models with 2−dimensional predictor variables, given by

$$Y_{ij} = Z_i(T_{ij}, S_{ij}) + \epsilon_{ij},$$

where Y_{ij} is the scalar observation of the random surface Z_i of subject i observed at (T_{ij}, S_{ij}) for $i \in \{1\ldots,n\}$ and $j \in \{1,\ldots,N_i\}$. The $Z_i(\cdot)$ are independent realizations of the underlying process $\{Z(\cdot)\}$ and ϵ_{ij}'s are i.i.d. copies of ϵ with mean zero and variance $\sigma^2 > 0$.

3. Theoretical results

In this section, we systematically investigate the asymptotic properties of $\hat{\mu}(t,s)$ proposed in Section 2. Let $\sigma_K^2 := \int u^2K(u)du$, $\|K\| := \sqrt{\int K(u)^2du}$, and $f_T(t)$ and $f_S(s)$, respectively, denote the densities of T and S. We denote $a(n) \prec b(n)$ to mean $\limsup_{n\to\infty} a(n)/b(n) = 0$, $a(n) \preceq b(n)$ to mean $\limsup_{n\to\infty} a(n)/b(n) < \infty$, and $a(n) \asymp b(n)$ to mean $a(n) \preceq b(n)$ and $b(n) \preceq a(n)$. Denote $\overline{NM} := n^{-1}\sum_{i=1}^{n} N_iM_i$, $\overline{(NM)}_H := (n^{-1}\sum_{i=1}^{n}(N_iM_i)^{-1})^{-1}$, $N_{S_k} := \sum_{i=1}^{n} N_i^k$, $M_{S_k} := \sum_{i=1}^{n} M_i^k$, $\overline{N}_{S_k} := N_{S_k}/n$, $\overline{M}_{S_k} := M_{S_k}/n$ for any $k \geq 2$, and $U_{ijk} := U_i(T_{ij}, S_{ik})$.

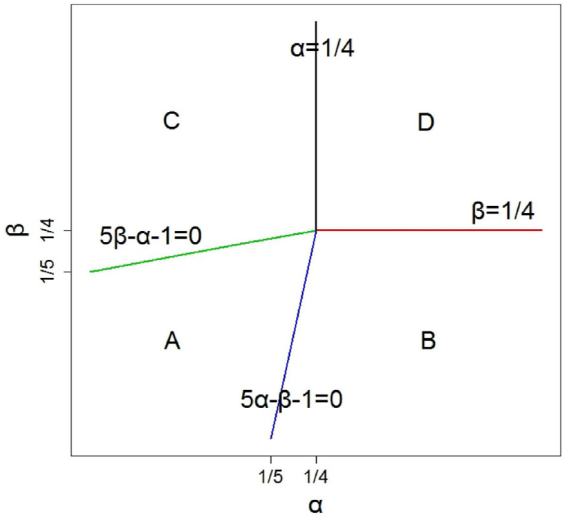


Fig. 1. Partition of surface functional data into nine categories based on the relative orders of both $\overline{N} = O(n^{\alpha})$ and $\overline{M} = O(n^{\beta})$ to n for $\hat{\mu}_{EW}(t, s)$ including A: S+S; B: UD+S; C: S+UD; D: UD+UD; Green Line: S+D; Red Line: UD+D; Blue Line: D+S; Black Line: D+UD; and Point (1/4, 1/4): D+D, where "S" denotes Sparse; "D" denotes Dense and "UD" denotes Ultra-Dense. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Asymptotic normality of $\hat{\mu}_{EW}(t, s)$ for the six sampling designs. S:Sparse; D:Dense; UD:Ultra-Dense; Bandwidth Order: the optimal bandwidth order selected using the AMSE (asymptotic mean-squared error) criterion; Bias and Variance: the bias and variance in the asymptotic normal distribution of $\hat{\mu}_{EW}(t, s)$; $\overline{N} = O(n^{\alpha})$ and $\overline{M} = O(n^{\beta})$.

(α, β)	Case	Bandwidth order	Bias	Variance
$5\alpha - \beta - 1 < 0$ $5\beta - \alpha - 1 < 0$	S+S	$h_1 \asymp (n\overline{N}\ \overline{M})^{-1/6}$ $h_2 \asymp (n\overline{N}\ \overline{M})^{-1/6}$	$B_{1n} + B_{2n}$	A^*_{4n}
$\alpha < 1/4$ $5\beta - \alpha - 1 = 0$	S+D	$h_1 \asymp (n\overline{N})^{-1/5}$ $h_2 \asymp (n\overline{N})^{-1/5}$	$B_{1n} + B_{2n}$	$A_{3n} + A^*_{4n}$
$\alpha < 1/4$ $5\beta - \alpha - 1 > 0$	S+UD	$h_1 \asymp (n\overline{N})^{-1/5}$ $h_2 \asymp (n\overline{N}\ \overline{M}^{5/4})^{-4/25}$	B_{1n}	A^*_{3n}
$\alpha = 1/4$ $\beta = 1/4$	D+D	$h_1 \asymp n^{-1/4}$ $h_2 \asymp n^{-1/4}$	$B_{1n} + B_{2n}$	$\sum_{l=1}^{4} A^*_{ln}$
$\alpha = 1/4$ $\beta > 1/4$	D+UD	$h_1 \asymp n^{-1/4}$ $h_2 \asymp (n\overline{M})^{-1/5}$	B_{1n}	$A^*_{1n} + A^*_{3n}$
$\alpha > 1/4$ $\beta > 1/4$	UD+UD	$h_1 \asymp (n\overline{N})^{-1/5}$ $h_2 \asymp (n\overline{M})^{-1/5}$	0	A^*_{1n}

3.1. Conditional bias and conditional variance

Assumption 1.
(A) Kernel function:
(A1) $K(\cdot)$ is a symmetric probability density function on $[-1, 1]$ and

$$\int u^2 K(u)du < \infty, \quad \int K(u)^2 du < \infty.$$

(A2) $K(\cdot)$ is Lipschitz continuous: that is, there exists a $0 < L < \infty$ such that $|K(u) - K(v)| \leq L|u - v|$ for any $u, v \in [-1, 1]$. This implies $\sup_{u\in[-1,1]} |K(u)| \leq M_K$ for a constant M_K.
(B) Time points, location points and true model:
(B1) $\{T_{ij}, i \in \{1, \ldots, n\}; j \in \{1, \ldots, N_i\}\}$ and $\{S_{ik}, i \in \{1, \ldots, n\}; k \in \{1, \ldots, M_i\}\}$ are, respectively, i.i.d. copies of independent random variables T and S. T and S are both defined on $[0, 1]$ with densities $f_T(t)$ and $f_S(s)$, respectively.
(B2) $f_T(t)$ and $f_S(s)$ are both bounded from below and above:

$$0 < m_T \leq \inf_{t\in[0,1]} f_T(t) \leq \sup_{t\in[0,1]} f_T(t) \leq M_T < \infty,$$

Table 4
Uniform convergence results of $\hat{\mu}_{EW}(t,s)$ for the different sampling designs. S:Sparse; D:Dense; UD:Ultra-Dense; Bandwidth: the optimal bandwidth order selected by minimizing the uniform convergence rate; Rate: Uniform Convergence Rate; $0 < C < \infty$.

	Case	Bandwidth	Rate
$\ln(n)\overline{N}^5/(n\overline{M}) \to 0$ $\ln(n)\overline{M}^5/(n\overline{N}) \to 0$	S+S	$h_1 \asymp \{n\overline{N}\ \overline{M}/\ln(n)\}^{-1/6}$ $h_2 \asymp \{n\overline{N}\ \overline{M}/\ln(n)\}^{-1/6}$	$\{\ln(n)/(n\overline{N}\ \overline{M})\}^{1/3}$
$\ln(n)\overline{N}^4/n \to 0$ $\ln(n)\overline{M}^5/(n\overline{N}) \to C$	S+D	$h_1 \asymp \{n\overline{N}/\ln(n)\}^{-1/5}$ $h_2 \asymp \{n\overline{N}/\ln(n)\}^{-1/5}$	$\{\ln(n)/(n\overline{N})\}^{2/5}$
$\ln(n)\overline{N}^4/n \to 0$ $\ln(n)\overline{M}^5/(n\overline{N}) \to \infty$	S+UD	$h_1 \asymp \{n\overline{N}/\ln(n)\}^{-1/5}$ $h_2 \asymp \{n\overline{N}\ \overline{M}^{5/4}/\ln(n)\}^{-4/25}$	$\{\ln(n)/(n\overline{N})\}^{2/5}$
$\ln(n)\overline{N}^4/n \to C$ $\ln(n)\overline{M}^4/n \to C$	D+D	$h_1 \asymp \{n/\ln(n)\}^{-1/4}$ $h_2 \asymp \{n/\ln(n)\}^{-1/4}$	$\{\ln(n)/n\}^{1/2}$
$\ln(n)\overline{N}^4/n \to C$ $\ln(n)\overline{M}^4/n \to \infty$	D+UD	$h_1 \asymp \{n/\ln(n)\}^{-1/4}$ $h_2 \asymp \{n\overline{M}/\ln(n)\}^{-1/5}$	$\{\ln(n)/n\}^{1/2}$
$\ln(n)\overline{N}^4/n \to \infty$ $\ln(n)\overline{M}^4/n \to \infty$	UD+UD	$h_1 \asymp \{n\overline{N}/\ln(n)\}^{-1/5}$ $h_2 \asymp \{n\overline{M}/\ln(n)\}^{-1/5}$	$\{\ln(n)/n\}^{1/2}$

$$0 < m_S \le \inf_{s\in[0,1]} f_S(s) \le \sup_{s\in[0,1]} f_S(s) \le M_S < \infty.$$

Both $f_T(t)$ and $f_S(s)$ are twice continuously differentiable.
(B3) U is independent of (T,S) and ϵ is independent of (T,S) and U.
(B4) $\partial^2\mu(t,s)/\partial t^2$, $\partial^2\mu(t,s)/\partial t\partial s$, and $\partial^2\mu(t,s)/\partial s^2$ are continuous on $[0,1]^2$.
(B5) All the second-order partial derivatives of $R(t,s;t',s')$ exist and are bounded on $[0,1]^4$.
(C) N_i's, M_i's, bandwidths and weights:
(C1) $h_1 \to 0$ and $h_2 \to 0$.
(C2) $\sum_{i=1}^n \xi_i^2 N_i(N_i-1)M_i(M_i-1) \to 0$, $\sum_{i=1}^n \xi_i^2 N_i(N_i-1)M_i/h_2 \to 0$, $\sum_{i=1}^n \xi_i^2 M_i(M_i-1)N_i/h_1 \to 0$, and $\sum_{i=1}^n \xi_i^2 N_i M_i/(h_1h_2) \to 0$.
(C3)

$$\liminf_n \min\left\{\overline{NM}/(\overline{N}\ \overline{M}),\quad n\overline{N}_{S_2}\overline{M}/(\sum_{i=1}^n N_i^2M_i)\right\} > 0,$$

$$\liminf_n \min\left\{n\overline{M}_{S_2}\overline{N}/(\sum_{i=1}^n N_iM_i^2),\quad n\overline{N}_{S_2}\overline{M}_{S_2}/(\sum_{i=1}^n N_i^2M_i^2)\right\} > 0,$$

$$\limsup_n \max\left\{\overline{N}_{S_2}/(\overline{N})^2, \overline{M}_{S_2}/(\overline{M})^2, n\overline{N}_{S_2}\overline{M}/(\sum_{i=1}^n N_i^2M_i)\right\} < \infty,$$

$$\limsup_n \max\left\{n\overline{M}_{S_2}\overline{N}/(\sum_{i=1}^n N_iM_i^2), n\overline{N}_{S_2}\overline{M}_{S_2}/(\sum_{i=1}^n N_i^2M_i^2)\right\} < \infty.$$

(C4)

$$\liminf_n \min\left\{n\overline{N}\{\overline{M}_H(\sum_{i=1}^n N_i/M_i)\}^{-1}, n\overline{N}_{S_2}\{\overline{M}_H(\sum_{i=1}^n N_i^2/M_i)\}^{-1}\right\} > 0,$$

$$\limsup_n \max\left\{\overline{N}_{S_2}/(\overline{N})^2, n\overline{N}/\{\overline{M}_H(\sum_{i=1}^n N_i/M_i)\}, n\overline{N}_{S_2}/\{\overline{M}_H(\sum_{i=1}^n N_i^2/M_i)\}\right\} < \infty.$$

(C5) $\liminf_n \overline{(NM)}_H/(\overline{N}_H\overline{M}_H) > 0$ and $\limsup_n \overline{(NM)}_H/(\overline{N}_H\overline{M}_H) < \infty$.

Remark 2. (i) Assumption (A) can be found in [31].
(ii) The assumptions on $\{T_{ij}\}$'s and $f_T(t)$ and those on $\{S_{ij}\}$'s and $f_S(t)$ coincide with those made on the time points [20,31] and on the spatial points [5]. Assumptions (B3), (B4) and (B5) are consistent with Assumptions (B2), (B3) and (B4) in [31], respectively. See also [15,20,26].
(iii) Assumptions (C1) and (C2) are routine and consistent with Assumption (C1a) in [31].

(iv) If $\{N_i\}_{i=1}^n$ and $\{M_i\}_{i=1}^n$ are independently and identically distributed (i.i.d) copies of positive integer-valued random variables N and M, respectively, with N being independent of M, then we have

$$\limsup_n n\overline{N}_{S_2}\overline{M}/(\sum_{i=1}^n N_i^2M_i) = \liminf_n n\overline{N}_{S_2}\overline{M}/(\sum_{i=1}^n N_i^2M_i) = 1.$$

Moreover, $\limsup_n \overline{N}_{S_2}/(\overline{N})^2 < \infty$ can be found in [31]. Thus, it may indicate that Assumptions (C3), (C4) and (C5), designed for EW, DDW and SDW schemes, respectively, are reasonable.

We first state the asymptotic conditional bias and conditional variance of the mean estimator based on a general weighting scheme. Without loss of generality, the random variables T and S are assumed to be defined on $[0, 1]$.

Theorem 1 (*Conditional Bias and Conditional Variance*). *Let (t, s) be a fixed point in the interior of $[0, 1]^2$. Suppose that Assumptions (A), (B), (C1), and (C2) hold. Let $X = \{T_{ij}, S_{ik}, i \in \{1, \ldots, n\}; j \in \{1, \ldots, N_i\}; k \in \{1, \ldots, M_i\}\}$, we have*

$$\mathrm{E}\{\hat{\mu}(t, s) - \mu(t, s)|X\} = \frac{1}{2}\sigma_K^2 h_1^2 \frac{\partial^2\mu(t, s)}{\partial t^2} + \frac{1}{2}\sigma_K^2 h_2^2 \frac{\partial^2\mu(t, s)}{\partial s^2} + o_p(h_1^2 + h_2^2) := B_{1n} + B_{2n} + o_p(h_1^2 + h_2^2)$$

and $\mathrm{Var}\{\hat{\mu}(t, s)|X\} = \Gamma_n\{1 + o_p(1)\}$, *where*

$$\begin{aligned}\Gamma_n &= \sum_{i=1}^n \xi_i^2 N_i(N_i - 1)M_i(M_i - 1)R(t, s; t, s) + \sum_{i=1}^n \xi_i^2 N_i(N_i - 1)M_i\|K\|^2 R(t, s; t, s)/\{h_2 f_S(s)\} \\ &\quad + \sum_{i=1}^n \xi_i^2 M_i(M_i - 1)N_i\|K\|^2 R(t, s; t, s)/\{h_1 f_T(t)\} + \sum_{i=1}^n \xi_i^2 N_i M_i\|K\|^4\{R(t, s; t, s) + \sigma^2\}/\{h_1 h_2 f_T(t) f_S(s)\} \\ &:= A_{1n} + A_{2n} + A_{3n} + A_{4n} = \sum_{l=1}^4 A_{ln}.\end{aligned}$$

Proof. We have

$$\hat{\mu}(t, s) = (1, 0, 0)\left(\sum_{i=1}^n X_i^\top W_i X_i\right)^{-1} \sum_{i=1}^n X_i^\top W_i Y_i, \tag{2}$$

where X_i, W_i, and Y_i are, respectively, $M_iN_i \times 3$, $(N_iM_i) \times (N_iM_i)$, and $(N_iM_i) \times 1$ matrices. Specifically,

$$\begin{aligned}Y_i &= (Y_{i11}, \ldots, Y_{i1M_i}, Y_{i21}, \ldots, Y_{i2M_i}, \ldots, Y_{iN_i1}, \ldots, Y_{iN_iM_i})^\top, \\ X_i &= \begin{pmatrix} 1 & \cdots & 1 & \ldots & 1 & \ldots & 1 \\ T_{i1} - t & \ldots & T_{i1} - t & \ldots & T_{iN_i} - t & \ldots & T_{iN_i} - t \\ S_{i1} - s & \ldots & S_{iM_i} - s & \ldots & S_{i1} - s & \ldots & S_{iM_i} - s \end{pmatrix}^\top, \\ W_i &= \xi_i \times \mathrm{diag}\Big(K_{h_1}(T_{i1} - t)K_{h_2}(S_{i1} - s), \ldots, K_{h_1}(T_{i1} - t)K_{h_2}(S_{iM_i} - s), \\ &\qquad \ldots, K_{h_1}(T_{iN_i} - t)K_{h_2}(S_{i1} - s), \ldots, K_{h_1}(T_{iN_i} - t)K_{h_2}(S_{iM_i} - s)\Big).\end{aligned}$$

First, we consider the term $\sum_{i=1}^n X_i^\top W_i X_i$ in (2) as follows. Let $\dot{f}_T(t) = df_T(t)/dt$ and $\dot{f}_S(t) = df_S(t)/dt$. By Assumptions (A), (B1), (B2), (C1) and (C2), it follows from the law of large number and properties of kernel smoothing technique that we have

$$\sum_{i=1}^n X_i^\top W_i X_i = \begin{pmatrix} a_{11n} & a_{12n} & a_{13n} \\ a_{12n} & a_{22n} & a_{23n} \\ a_{13n} & a_{23n} & a_{33n} \end{pmatrix} := A_n, \tag{3}$$

where $a_{11n} = f_T(t)f_S(s) + o_p(1)$, $a_{12n} = h_1^2\sigma_K^2\dot{f}_T(t)f_S(s) + o_p(h_1^2)$, $a_{13n} = h_2^2\sigma_K^2 f_T(t)\dot{f}_S(s) + o_p(h_2^2)$, $a_{22n} = h_1^2\sigma_K^2 f_T(t)f_S(s) + o_p(h_1^2)$, $a_{23n} = O_p(h_1^2h_2^2)$, and $a_{33n} = h_2^2\sigma_K^2 f_T(t)f_S(s) + o_p(h_2^2)$. Because $\det(A_n) = h_1^2h_2^2\sigma_K^4 f_T^3(t)f_S^3(s)\{1 + o_p(1)\}$, A_n is nonsingular based on Assumptions (A) and (B2). Let A_n^* be the adjoint of matrix A_n. By Assumption (B2), We have

$$\begin{aligned}(1, 0, 0)\left(\sum_{i=1}^n X_i^\top W_i X_i\right)^{-1} &= \big(\{\det(A_n)\}^{-1}, 0, 0\big)A_n^* \\ &= \left(\frac{1}{f_T(t)f_S(s)} + o_p(1), -\frac{\dot{f}_T(t)}{f_T^2(t)f_S(s)} + o_p(1), -\frac{\dot{f}_S(s)}{f_T(t)f_S^2(s)} + o_p(1)\right).\end{aligned} \tag{4}$$

Second, we consider the term $\sum_{i=1}^{n} X_i^\top W_i Y_i$ and rewrite it as

$$\sum_{i=1}^{n}\xi_i\sum_{j=1}^{N_i}\sum_{k=1}^{M_i}K_{h_1}(T_{ij}-t)K_{h_2}(S_{ik}-s)Y_{ijk}\left(1,T_{ij}-t,S_{ik}-s\right)^\top. \tag{5}$$

Define $\mu_i=(\mu_{i11},\ldots,\mu_{i1M_i},\ldots,\mu_{iN_i1},\ldots,\mu_{iN_iM_i})^\top$, where $\mu_{ijk}=\mu(T_{ij},S_{ik})$. We have

$$\begin{aligned}
&\mathrm{E}\Big(\sum_{i=1}^{n}X_i^\top W_iY_i\Big|X\Big)=\sum_{i=1}^{n}X_i^\top W_i\mu_i \qquad (6)\\
&=\sum_{i=1}^{n}X_i^\top W_iX_i\left(\mu(t,s),\frac{\partial\mu(t,s)}{\partial t},\frac{\partial\mu(t,s)}{\partial s}\right)^\top+\sum_{i=1}^{n}\xi_i\sum_{j=1}^{N_i}\sum_{k=1}^{M_i}K_{h_1}(T_{ij}-t)K_{h_2}(S_{ik}-s)\left(1,T_{ij}-t,S_{ik}-s\right)^\top\\
&\quad\times\left\{\frac{1}{2}(T_{ij}-t)^2\frac{\partial^2\mu(t,s)}{\partial t^2}+\frac{1}{2}(S_{ik}-s)^2\frac{\partial^2\mu(t,s)}{\partial s^2}+(T_{ij}-t)(S_{ik}-s)\frac{\partial^2\mu(t,s)}{\partial t\partial s}\right\}\left\{1+o_p(1)\right\}:=I_{1n}+I_{2n}.
\end{aligned}$$

By Assumptions (A), (B1), (B2), (B3), (B4), (C1) and (C2), the first entry of I_{2n}, which is a 3−dimensional vector, equals $2^{-1}\sigma_K^2 f_T(t)f_S(s)\{h_1^2\partial^2\mu(t,s)/\partial t^2+h_2^2\partial^2\mu(t,s)/\partial s^2\}+o_p(h_1^2+h_2^2)$. The second and third entries of I_{2n} are both $o_p(h_1^2+h_2^2)$. Combining (4) and (6) yields

$$\mathrm{E}\{\hat{\mu}(t,s)-\mu(t,s)|X\}=\frac{1}{2}\sigma_K^2\left\{h_1^2\frac{\partial^2\mu(t,s)}{\partial t^2}+h_2^2\frac{\partial^2\mu(t,s)}{\partial s^2}\right\}+o_p(h_1^2+h_2^2).$$

Third, with some calculations, we get

$$\mathrm{Var}(\hat{\mu}(t,s)|X)=(1,0,0)\left(\sum_{i=1}^{n}X_i^\top W_iX_i\right)^{-1}\left(\sum_{i=1}^{n}X_i^\top W_i\mathrm{cov}(Y_i|X)W_iX_i\right)\left(\sum_{i=1}^{n}X_i^\top W_iX_i\right)^{-1}(1,0,0)^\top.$$

The (1, 1)th entry of $\sum_{i=1}^{n}X_i^\top W_i\mathrm{cov}(Y_i|X)W_iX_i$ equals

$$\begin{aligned}
&\sum_{i=1}^{n}\xi_i^2\sum_{j=1}^{N_i}\sum_{k=1}^{M_i}\sum_{l=1}^{N_i}\sum_{v=1}^{M_i}g_{ijklv}:=J_{1n}+J_{2n}+J_{3n}+J_{4n} \qquad (7)\\
&=\sum_{i=1}^{n}\xi_i^2\sum_{j\neq l}\sum_{k\neq v}g_{ijklv}+\sum_{i=1}^{n}\xi_i^2\sum_{j\neq l}\sum_{k=v}g_{ijklv}+\sum_{i=1}^{n}\xi_i^2\sum_{j=l}\sum_{k\neq v}g_{ijklv}+\sum_{i=1}^{n}\xi_i^2\sum_{j=l}\sum_{k=v}g_{ijklv},
\end{aligned}$$

where $g_{ijklv}=K_{h_1}(T_{ij}-t)K_{h_2}(S_{ik}-s)K_{h_1}(T_{il}-t)K_{h_2}(S_{iv}-s)\mathrm{cov}(Y_{ijk},Y_{ilv}|X_i^*)$ with $X_i^*=\{(T_{ij},S_{ik}):j\in\{1,\ldots,N_i\};k\in\{1,\ldots,M_i\}\}$. It follows from Assumptions (A), (B1), (B2), (B3), (B5), (C1) and (C2) that we have, for $j\neq l$ and $k\neq v$,

$$\begin{aligned}
J_{1n}&=f_T(t)^2f_S(s)^2\sum_{i=1}^{n}\xi_i^2N_i(N_i-1)M_i(M_i-1)\mathrm{cov}(Y_{ijk},Y_{ilv}|T_{ij}=t,T_{il}=t,S_{ik}=s,S_{iv}=s)\{1+o_p(1)\}\\
&=f_T(t)^2f_S(s)^2R(t,s;t,s)\sum_{i=1}^{n}\xi_i^2N_i(N_i-1)M_i(M_i-1)\{1+o_p(1)\},\\
J_{2n}&=\frac{\|K\|^2f_T(t)^2f_S(s)}{h_2}\sum_{i=1}^{n}\xi_i^2N_i(N_i-1)M_i\mathrm{cov}(Y_{ijk},Y_{ilk}|T_{ij}=t,T_{il}=t,S_{ik}=s)\{1+o_p(1)\}\\
&=\frac{\|K\|^2}{h_2}f_T(t)^2f_S(s)R(t,s;t,s)\sum_{i=1}^{n}\xi_i^2N_i(N_i-1)M_i\{1+o_p(1)\},\\
J_{3n}&=\frac{\|K\|^2f_T(t)f_S(s)^2}{h_1}\sum_{i=1}^{n}\xi_i^2N_iM_i(M_i-1)\mathrm{cov}(Y_{ijk},Y_{ijv}|T_{ij}=t,S_{ik}=s,S_{iv}=s)\{1+o_p(1)\}\\
&=\frac{\|K\|^2}{h_1}f_T(t)f_S(s)^2R(t,s;t,s)\sum_{i=1}^{n}\xi_i^2N_iM_i(M_i-1)\{1+o_p(1)\},
\end{aligned}$$

and

$$J_{4n} = \frac{\|K\|^4 f_T(t) f_S(s)}{h_1 h_2} \sum_{i=1}^{n} \xi_i^2 N_i M_i \mathrm{cov}(Y_{ijk}, Y_{ijk} | T_{ij} = t, S_{ik} = s)\{1 + o_p(1)\}$$
$$= \frac{\|K\|^4}{h_1 h_2} f_T(t) f_S(s)\{R(t, s; t, s) + \sigma^2\} \sum_{i=1}^{n} \xi_i^2 N_i M_i \{1 + o_p(1)\}.$$

Denote the (j, k)th entry of $\sum_{i=1}^{n} X_i^\top W_i \mathrm{cov}(Y_i|X) W_i X_i$ as

$$\left(\sum_{i=1}^{n} X_i^\top W_i \mathrm{cov}(Y_i|X) W_i X_i \right)[j, k].$$

We can show that

$$\left(\sum_{i=1}^{n} X_i^\top W_i \mathrm{cov}(Y_i|X) W_i X_i \right)[j, k] \prec \left(\sum_{i=1}^{n} X_i^\top W_i \mathrm{cov}(Y_i|X) W_i X_i \right)[1, 1]$$

with probability goes to one as n goes to infinity, for either $j \neq 1$ or $k \neq 1$. For instance,

$$\left(\sum_{i=1}^{n} X_i^\top W_i \mathrm{cov}(Y_i|X) W_i X_i \right)[1, 2] = O_p \left\{ h_1^2 \times \left(\sum_{i=1}^{n} X_i^\top W_i \mathrm{cov}(Y_i|X) W_i X_i \right)[1, 1] \right\}.$$

Finally, we have $\mathrm{Var}(\hat{\mu}(t, s)|X) = \Gamma_n\{1 + o_p(1)\}$, which completes the proof of Theorem 1. □

We use the well-known local linear smoothing technique [6,26] to prove Theorem 1. The leading term of the conditional bias of $\hat{\mu}(t, s)$ is $B_{1n}+B_{2n}$, which is consistent with that of the traditional two-dimensional local linear estimator [6,26]. The variance term is more complex and represents our main new contribution. The first term of the condition variance, A_{1n}, comes from the covariances between those pairs of observations observed at different time points and different positions (i.e., $\mathrm{cov}(Y_{ijk}, Y_{ilv})$ with $j \neq l$ and $k \neq v$). The A_{2n} corresponds to the covariances between those pairs at different time points and the same position (i.e., $\mathrm{cov}(Y_{ijk}, Y_{ilv})$ with $j \neq l$ and $k = v$), A_{3n} accounts for the covariances between the pairs of observations at the same time point and different positions (i.e., $\mathrm{cov}(Y_{ijk}, Y_{ilv})$ with $j = l$ and $k \neq v$), and the last term corresponds to the marginal variances of all the observations. In the derivation, we need to deal with the within-surface dependence and the two-domain (time and location) correlation structure of the surface functional data that makes the variance of $\hat{\mu}(t, s)$ more complex than that of the curve functional data (Theorem 3.1 of [31] and Theorem 2 of [2]).

Remark 3. When the number of locations for subject i (i.e., M_i) is 1 for $i \in \{1, \ldots, n\}$, the surface functional data can be viewed as the curve functional data observed at different locations across all subjects. In this special case, $\Gamma_n = A_{2n} + A_{4n}$, and the variance term is of a similar form as that of the curve functional model. Distinctive from the curve functional models, we intend to estimate the two-dimensional function $\mu(t, s)$ and borrow information across subjects by assigning to them the corresponding weight $K_{h_2}(S_{i1} - s)$. Our result indicates a larger variance of the mean function estimator compared to that of the curve functional model [31] as $n \to \infty$. Specifically, the variance Γ_n is at the order of the variance in [31] divided by h_2, see also, [16].

The optimal bandwidths are, respectively, proportional to $n^{-1/5}$ and $n^{-1/6}$ for the traditional one-dimensional and two-dimensional nonparametric kernel regression [6,26]. We investigate the optimal bandwidth orders for $\hat{\mu}(t, s)$ in our surface functional model framework. The optimal bandwidth orders are calculated via minimizing the asymptotic mean-squared errors. The sophisticated variance structure presented in our framework leads to distinctive and more intricate results compared to the existing literature. We consider the optimal bandwidth orders for the EW scheme. The orders of the optimal bandwidths vary with the "sparsity" levels of both the time and location domains. See the Appendix A for detailed derivations. Let $\overline{N} := \sum_{i=1}^{n} N_i/n = O(n^\alpha)$ and $\overline{M} := \sum_{i=1}^{n} M_i/n = O(n^\beta)$ with $\alpha \geq 0$ and $\beta \geq 0$. We summarize the results in Table 2. For example, when $5\alpha - \beta - 1 < 0$ and $5\beta - \alpha - 1 < 0$ (e.g., all N_i and M_i are bounded), the optimal bandwidths h_1 and h_2 satisfy $h_1 \asymp n^{-(1+\alpha+\beta)/6}$ and $h_2 \asymp n^{-(1+\alpha+\beta)/6}$. That is, the two optimal bandwidths have the identical order of the total number of observations to the power of $-1/6$. This result is consistent with that of the conventional two-dimensional nonparametric kernel regression. On the other hand, when $\min(\alpha, \beta) > 1/4$ (i.e., the data in both time and location directions are sufficiently dense), the optimal bandwidths h_1 and h_2 are of the orders $n^{-(1+\alpha)/5}$ and $n^{-(1+\beta)/5}$, respectively, which follow the marginal optimality rule. Consistent results can be obtained for DDW and SDW schemes (See Corollary 2).

We now utilize Theorem 1 to derive the result specific to each of the aforementioned nine sampling design scenarios, categorized according to sparse, dense, and ultra-dense sampling of both the time and location domains in Section 2. We only consider the six cases in Table 2, because the properties of the Dense+Sparse, Ultra-Dense+Sparse and Ultra-Dense+Dense cases are comparable to those of the Sparse+Dense, Sparse+Ultra-Dense and Dense+Ultra-Dense cases, respectively. The full results for all the nine cases are put in Table 1 in the Supplementary Materials. Let $\alpha \geq 0$ and $\beta \geq 0$.

For EW, we set $\overline{N} = O(n^{\alpha})$ and $\overline{M} = O(n^{\beta})$; for DDW, we set $\overline{N} = O(n^{\alpha})$ and $\overline{M}_H := (n^{-1}\sum_{i=1}^{n} M_i^{-1})^{-1} = O(n^{\beta})$; for SDW, we set $\overline{N}_H := (n^{-1}\sum_{i=1}^{n} N_i^{-1})^{-1} = O(n^{\alpha})$ and $\overline{M}_H = O(n^{\beta})$. We use $\hat{\mu}_{EW}(t,s)$, $\hat{\mu}_{DDW}(t,s)$, and $\hat{\mu}_{SDW}(t,s)$ to denote $\hat{\mu}(t,s)$ derived using the EW, DDW, and SDW schemes, respectively. We summarize the asymptotic results of the six sampling categories in the following Corollary. We note that the partition of the time-location domains is performed according to the relative orders of both $\overline{N}$ and $\overline{M}$ to n for $\hat{\mu}_{EW}(t,s)$ (See, Fig. 1), and the dominating terms in the conditional variance across the six sampling designs are different.

Corollary 2. *Let (t,s) be a fixed element in the interior of $[0,1]^2$. Suppose all the assumptions of Theorem 1 hold.*
(i) For EW, it is also assumed that Assumption (C3) also holds. The conditional bias and conditional variance corresponding to the six sampling schemes in Table 2 hold. For instance, since the "Sparse+Sparse" case corresponds to $5\alpha - \beta - 1 < 0$ and $5\beta - \alpha - 1 < 0$, we have $h_1 \asymp (n\overline{N}\ \overline{M})^{-1/6}$, $h_2 \asymp (n\overline{N}\ \overline{M})^{-1/6}$, and

$$\mathrm{E}\{\hat{\mu}_{EW}(t,s) - \mu(t,s)|X\} = B_{1n} + B_{2n} + o_p(h_1^2 + h_2^2), \quad \mathrm{Var}\{\hat{\mu}_{EW}(t,s)|X\} = A_{4n}\{1 + o_p(1)\}.$$

(ii) It is assumed that Assumption (C4) also holds. For DDW, the optimal bandwidth orders and partition of the surface functional data, and conditional bias and conditional variance of $\hat{\mu}_{DDW}(t,s)$ for each of the six sampling scenarios in Table 2 hold by replacing $\xi_i = 1/(\sum_{i=1}^{n} N_i M_i)$ and $\overline{M}$ in (a) by $\xi_i = 1/\{(\sum_{i=1}^{n} N_i)M_i\}$ and $\overline{M}_H$, respectively.
(iii) It is assumed that Assumption (C5) also holds. For SDW, the optimal bandwidth orders and partition of the surface functional data, and conditional bias and conditional variance of $\hat{\mu}_{SDW}(t,s)$ for each of the six sampling scenarios in Table 2 hold by replacing $\xi_i = 1/(\sum_{i=1}^{n} N_i M_i)$, $\overline{N}$, and $\overline{M}$ in (a) by $\xi_i = 1/(nN_iM_i)$, $\overline{N}_H$, and $\overline{M}_H$, respectively.

Remark 4. Corollary 2 provides an intuitive criterion on how to define the nine sampling designs of surface functional data for the EW scheme. See Table 2 and Fig. 1. For example, the "Sparse+Sparse" category is defined by $5\alpha - \beta - 1 < 0$ and $5\beta - \alpha - 1 < 0$. The same partition holds for $\hat{\mu}_{DDW}(t,s)$ with $\overline{N} = O(n^{\alpha})$ and $\overline{M}_H = O(n^{\beta})$ and $\hat{\mu}_{SDW}(t,s)$ with $\overline{N}_H = O(n^{\alpha})$ and $\overline{M}_H = O(n^{\beta})$.

Remark 5. In the "Ultra-Dense+Ultra-Dense" Case, the dominating conditional bias is B_{1n} if $\alpha < \beta$, B_{2n} if $\alpha > \beta$, and $B_{1n} + B_{2n}$ if $\alpha = \beta$. Further, the order of the asymptotic conditional bias is smaller than that of the conditional variance because $h_1^2 + h_2^2 \prec n^{-1/2}$ holds for this case. For all the other scenarios, the asymptotic conditional bias and conditional variance are of the identical order.

The convergence rate is defined to be the order of $|\mathrm{Bias}| + \sqrt{\mathrm{Variance}}$, where "Bias" and "Variance" represent the asymptotic conditional bias and conditional variance, respectively. The results for $\hat{\mu}_{EW}(t,s)$ are given in Table 2. We present the following conclusions: (i) In the "Sparse+Sparse" case, the result is comparable to that of the traditional two-dimensional local linear smoother for independent data, specifically, the convergence rate is of the order between $n^{-1/3}$ and $n^{-1/2}$, however, can never reach $n^{-1/2}$; (ii) In the "Sparse+Dense" and "Sparse+Ultra-Dense" cases, the convergence rates are of the order between $n^{-2/5}$ and $n^{-1/2}$ and can never attain $n^{-1/2}$; (iii) In the "Dense+Dense", "Dense+Ultra-Dense" and "Ultra-Dense+Ultra-Dense" cases, the rate of $n^{-1/2}$ can be attained. These suggest that both the time and the location domains for surface functional models need to have sufficiently dense data to attain the fast rate of $n^{-1/2}$. Consistent results can be obtained for $\hat{\mu}_{DDW}(t,s)$ and $\hat{\mu}_{SDW}(t,s)$.

3.2. Asymptotic normality

Define

$$Q_n := \sum_{i=1}^{n} \xi_i^2 N_i(N_i-1)M_i(M_i-1) + \sum_{i=1}^{n} \xi_i^2 N_i(N_i-1)M_i/h_2 + \sum_{i=1}^{n} \xi_i^2 M_i(M_i-1)N_i/h_1 + \sum_{i=1}^{n} \xi_i^2 N_i M_i/(h_1h_2),$$

and

$$\begin{aligned} G_n := \max\Bigg\{ & \sum_{i=1}^{n} \xi_i^3 N_i M_i/(h_1^2h_2^2), \sum_{i=1}^{n} \xi_i^3 N_i M_i(M_i-1)/(h_1^2h_2), \sum_{i=1}^{n} \xi_i^3 N_i(N_i-1)M_i/(h_1h_2^2), \\ & \sum_{i=1}^{n} \xi_i^3 N_i M_i(M_i-1)(M_i-2)/h_1^2, \sum_{i=1}^{n} \xi_i^3 N_i(N_i-1)(N_i-2)M_i/h_2^2, \sum_{i=1}^{n} \xi_i^3 N_i(N_i-1)M_i(M_i-1)/(h_1h_2), \\ & \sum_{i=1}^{n} \xi_i^3 N_i(N_i-1)M_i(M_i-1)(M_i-2)/h_1, \sum_{i=1}^{n} \xi_i^3 N_i(N_i-1)(N_i-2)M_i(M_i-1)/h_2, \\ & \sum_{i=1}^{n} \xi_i^3 N_i(N_i-1)(N_i-2)M_i(M_i-1)(M_i-2) \Bigg\}. \end{aligned}$$

Assumption 2.
(C) N_i's, M_i's, bandwidths and weights:
(C6) $G_n Q_n^{-3/2} \to 0$.
(C7) $\min\{1/\{\sum_{i=1}^n \xi_i^2 N_i(N_i-1)M_i(M_i-1)\}, h_2/\{\sum_{i=1}^n \xi_i^2 N_i(N_i-1)M_i\}, h_1/\{\sum_{i=1}^n \xi_i^2 M_i(M_i-1)N_i\}, h_1h_2/\{\sum_{i=1}^n \xi_i^2 N_iM_i\}\}$ $\max\{h_1^6, h_2^6\} \to 0$.
(D) The stochastic part and ϵ:
(D1) $\sup_{t,s\in[0,1]} \mathrm{E}|U(t,s)|^3 < \infty$ and $\mathrm{E}|\epsilon|^3 < \infty$.

Remark 6. Assumptions (C6) and (D1) are needed for ensuring the Lyapunov condition for asymptotic normality and are consistent with Assumptions (C2a) and (C3a) in [31], respectively. Assumption (C7) is consistent with the assumption in Theorem 3.1 of [31].

We state the unified asymptotic normality result of $\hat{\mu}(t,s)$, which is applicable to all the nine sampling designs and the three weighting schemes.

Theorem 3 (*Asymptotic Normality*)**.** *Let (t,s) be a fixed element in the interior of $[0,1]^2$. Suppose that Assumptions (A), (B), (C1), (C2), (C6), (C7) and (D1) hold, we have*

$$\Gamma_n^{-1/2}\left\{\hat{\mu}(t,s) - \mu(t,s) - B_{1n} - B_{2n} + o_p(h_1^2+h_2^2)\right\} \to_L \mathcal{N}(0,1).$$

Proof. Define

$$\mu_i = (\mu_{i11}, \ldots, \mu_{i1M_i}, \ldots, \mu_{iN_i1}, \ldots, \mu_{iN_iM_i})^\top,$$

where $\mu_{ijk} = \mu(T_{ij}, S_{ik})$, and

$$\delta_i = \left(U_i(T_{i1},S_{i1}) + \epsilon_{i11}, \ldots, U_i(T_{i1},S_{1M_i}) + \epsilon_{i1M_i}, \ldots, U_i(T_{iN_i},S_{i1}) + \epsilon_{iN_i1}, \ldots, U_i(T_{iN_i},S_{iM_i}) + \epsilon_{iN_iM_i}\right)^\top.$$

By (4), we have

$$\begin{aligned}\hat{\mu}(t,s) &= (1,0,0)\left(\sum_{i=1}^n X_i^\top W_i X_i\right)^{-1}\sum_{i=1}^n X_i^\top W_i\mu_i + \left(\frac{1}{f_T(t)f_S(s)}, -\frac{\dot{f}_T(t)}{f_T^2(t)f_S(s)}, -\frac{\dot{f}_S(s)}{f_T(t)f_S^2(s)}\right)\sum_{i=1}^n X_i^\top W_i\delta_i\left\{1+o_p(1)\right\}\\ &:= R_{1n} + R_{2n}\end{aligned}$$

First, it follows from (4) and (6) that we have

$$R_{1n} = \mu(t,s) + \frac{1}{2}\sigma_K^2\left\{h_1^2\frac{\partial^2\mu(t,s)}{\partial t^2} + h_2^2\frac{\partial^2\mu(t,s)}{\partial s^2}\right\} + o_p(h_1^2+h_2^2).$$

Second, as n goes to ∞, we can show $\Gamma_n^{-1/2}R_{2n} \to_L \mathcal{N}(0,1)$. $\sum_{i=1}^n X_i^\top W_i\delta_i$ is a 3−dimensional vector and the first entry of $\sum_{i=1}^n X_i^\top W_i\delta_i$ is equal to

$$r_{1n} := \sum_{i=1}^n \xi_i \sum_{j=1}^{N_i}\sum_{k=1}^{M_i} K_{h_1}(T_{ij}-t)K_{h_2}(S_{ik}-s)\left(U_{ijk}+\epsilon_{ijk}\right).$$

Then, it follows from Lyapunov central limit theorem that $\{\mathrm{Var}(r_{1n})\}^{-1/2}r_{1n}$ converges in distribution to $\mathcal{N}(0,1)$ as $n\to\infty$. Specifically, the Lyapunov condition is satisfied due to Assumptions (A), (B), (C1), (C2), (C6), and (D1). Note that $\int K(u)^3du < \infty$ holds due to Assumption (A). By the law of total variance, we have $\mathrm{Var}(r_{1n}) = \mathrm{E}\{\mathrm{Var}(r_{1n}|X)\} + \mathrm{Var}\{\mathrm{E}(r_{1n}|X)\}$. Because $\mathrm{E}(r_{1n}|X) = 0$ holds due to Assumption (B3), we have $\mathrm{Var}(r_{1n}) = \mathrm{E}\{\mathrm{Var}(r_{1n}|X)\}$. By the similar arguments in (7), we have that the variance of r_{1n} is $f_T^2(t)f_S^2(s)\Gamma_n\{1+o(1)\}$. Note that $f_T(t)f_S(s)$ is bounded from zero. By Slutsky's theorem, we have

$$\frac{\Gamma_n^{-1/2}r_{1n}}{f_T(t)f_S(s)} \to_L \mathcal{N}(0,1).$$

Denote the second and third entries of $\sum_{i=1}^n X_i^\top W_i\delta_i$ as r_{2n} and r_{3n}, respectively. Because we have $\mathrm{Var}(r_{2n}) = o(\Gamma_n)$ and $\mathrm{Var}(r_{3n}) = o(\Gamma_n)$, both $\Gamma_n^{-1/2}r_{2n}$ and $\Gamma_n^{-1/2}r_{3n}$ converge in probability towards zero by Chebyshev's inequality. Since $\dot{f}_T(t)/\{f_T^2(t)f_S(s)\}$ and $\dot{f}_S(s)/\{f_T(t)f_S^2(s)\}$ are bounded due to Assumption (B2), we have

$$\left|\frac{\dot{f}_T(t)}{f_T^2(t)f_S(s)}\Gamma_n^{-1/2}r_{2n}\right| + \left|\frac{\dot{f}_S(s)}{f_T(t)f_S^2(s)}\Gamma_n^{-1/2}r_{3n}\right| = o_p(1).$$

Thus, it follows from Slutsky's theorem that $\Gamma_n^{-1/2}R_{2n}$ converges in distribution to $N(0,1)$ as n goes to infinity. This completes the proof of Theorem 3. □

We use the Lyapunov central limit theorem to prove Theorem 3. In the proof, the covariance structure within and across the two-domain (time and location) of the surface functional data increases complexity compared with the asymptotic normality results in the curve functional data (Theorem 3.1 of [31] and Theorem 2 of [2]). This Theorem provides a unified framework of the asymptotic normality of $\hat{\mu}(t,s)$ across all nine sampling designs. We see that the unified result is valid for the three weighting schemes.

Define

$$A^*_{1n} := \sum_{i=1}^{n} \xi_i^2 N_i^2 M_i^2 R(t,s;t,s), A^*_{2n} := \sum_{i=1}^{n} \xi_i^2 N_i^2 M_i \|K\|^2 R(t,s;t,s)/\{h_2 f_S(s)\},$$

$$A^*_{3n} := \sum_{i=1}^{n} \xi_i^2 M_i^2 N_i \|K\|^2 R(t,s;t,s)/\{h_1 f_T(t)\}, \quad \text{and} \quad A^*_{4n} := A_{4n}.$$

Corollary 4. *Let (t,s) be a fixed element in the interior of $[0,1]^2$. Suppose that all the Assumptions of Theorem 3 hold.*
(i) *EW: It is assumed that Assumption (C3) also holds. The asymptotic normality of the six sampling designs holds with their corresponding biases and variances summarized in Table 3. For instance, for the "Sparse+Sparse" case corresponding to $5\alpha - \beta - 1 < 0$ and $5\beta - \alpha - 1 < 0$, $h_1 \asymp (n\overline{N}\ \overline{M})^{-1/6}$ and $h_2 \asymp (n\overline{N}\ \overline{M})^{-1/6}$, we have*

$$A^{*-1/2}_{4n}\{\hat{\mu}_{EW}(t,s) - \mu(t,s) - B_{1n} - B_{2n}\} \to_L \mathcal{N}(0,1).$$

(ii) *DDW: It is assumed that Assumption (C4) holds. The asymptotic normal distribution of $\hat{\mu}_{DDW}(t,s)$ is also valid for each of the six sampling designs by replacing $\xi_i = 1/(\sum_{i=1}^n N_i M_i)$ and $\overline{M}$ in Table 3 by $\xi_i = 1/\{(\sum_{i=1}^n N_i)M_i\}$ and $\overline{M}_H$, respectively.*
(iii) *SDW: It is assumed that Assumption (C5) holds. The asymptotic normal distribution of $\hat{\mu}_{SDW}(t,s)$ for each of the six sampling designs by replacing $\xi_i = 1/(\sum_{i=1}^n N_i M_i)$, $\overline{N}$, and $\overline{M}$ in Table 3 by $\xi_i = 1/(nN_iM_i)$, $\overline{N}_H$, and $\overline{M}_H$, respectively.*

The dominating asymptotic biases (variances) in the limiting normal distribution for the six cases are consistent with those of the conditional biases (variance) in Section 3.1, respectively. We put the full results of all the nine sampling designs in Table 2 in the Supplementary Materials. The biases of $\hat{\mu}_{EW}(t,s)$, $\hat{\mu}_{DDW}(t,s)$ and $\hat{\mu}_{SDW}(t,s)$ in their limiting normal distributions are all $B_{1n} + B_{2n}$ for the "Sparse+Sparse", "Sparse+Dense", "Dense+Sparse", and "Dense+Dense" cases. In the "Sparse+Ultra-Dense" and "Dense+Ultra-Dense" cases, $\hat{\mu}_{EW}(t,s)$, $\hat{\mu}_{DDW}(t,s)$ and $\hat{\mu}_{SDW}(t,s)$ all have the bias of B_{1n}. The "Ultra-Dense+Sparse" and "Ultra-Dense+Dense" cases have the bias of B_{2n}. The "Ultra-Dense+Ultra-Dense" case has zero bias. These results indicate that the "Ultra-Dense" feature of the time domain (or the location domain) leads to disappearing of B_{1n} (or B_{2n}) in the bias term, and thus a large number of observations in both the time and location domains in the "Ultra-Dense+Ultra-Dense" case make both B_{1n} and B_{2n} disappear.

Remark 7. Case "Ultra-Dense+Ultra-Dense" falls in the parametric paradigm where the limiting normal distribution has zero bias and convergence rate $1/\sqrt{n}$.

We compare the estimation efficiency of the three weighting schemes for the mean function estimation in terms of asymptotic variance in the limiting normal distribution. We give the following conclusions:
(i) For Case "Sparse+Sparse", EW produces generally more efficient estimator than DDW and SDW. The asymptotic variance is A^*_{4n} and it follows from Cauchy–Schwarz inequality that $\left\{\sum_{i=1}^n N_i M_i\right\}\left\{\sum_{i=1}^n (N_i M_i)^{-1}\right\} \geq n^2$ and $\left(\sum_{i=1}^n N_i\right)^2 \leq \left(\sum_{i=1}^n N_i/M_i\right)\left(\sum_{i=1}^n N_i M_i\right)$ for any series $\{N_i\}_{i=1}^n$ and $\{M_i\}_{i=1}^n$, so the asymptotic variance of EW is smaller (not larger) than those of DDW and SDW;
(ii) For Case "Sparse+Ultra-Dense", DDW $\xi_i = 1/\{(\sum_{i=1}^n N_i)M_i\}$ outperforms both EW and SDW, because the asymptotic variance is A^*_{3n}, $\overline{N} \geq \overline{N}_H$ and $(\sum_{i=1}^n N_i M_i)^2 \leq (\sum_{i=1}^n N_i)(\sum_{i=1}^n N_i M_i^2)$. Similarly, for Case "Ultra-Dense+Sparse", DDW $\xi_i = 1/\{N_i(\sum_{i=1}^n M_i)\}$ performs the best;
(iii) For Case "Ultra-Dense+Ultra-Dense", SDW is the best weighting scheme, because the asymptotic variance is A^*_{1n}, $(\sum_{i=1}^n N_i M_i)^2 \leq n\sum_{i=1}^n N_i^2 M_i^2$ and $(\sum_{i=1}^n N_i)^2 \leq n\sum_{i=1}^n N_i^2$.
These results here are consistent with those in [2,31]. Specifically, the "Ultra-Dense" feature of time (or spatial) direction encourages us to assign the weight N_i^{-1} (or M_i^{-1}) to subject i in order to achieve an efficient mean estimator (i.e., an estimator with relative small asymptotic variance). On the other hand, the equal weight is preferable for the "Sparse" feature for such purpose.

3.3. Uniform convergence

This section concerns the uniform convergence property of $\hat{\mu}(t,s)$.

Assumption 3.
(C) N_i's, M_i's, bandwidths and weights:
(C8) $\ln(n)\max\{\sum_{i=1}^n \xi_i^2 M_i(M_i-1)N_i/h_1, \sum_{i=1}^n \xi_i^2 N_i M_i/(h_1h_2)\} \to 0$ and $\ln(n)\max\{\sum_{i=1}^n \xi_i^2 N_i(N_i-1)M_i(M_i-1), \sum_{i=1}^n \xi_i^2 N_i (N_i-1)M_i/h_2\} \to 0$.

(C9) $\sup_n(n\max_i N_iM_i\xi_i) \le B < \infty$.
(C10)

$$\ln(n)(h_1Q_n)^{-1}\max_i\{\xi_i^2M_i(M_i-1)N_i\} \to 0,\quad \ln(n)(h_2Q_n)^{-1}\max_i\{\xi_i^2N_i(N_i-1)M_i\} \to 0,$$
$$\ln(n)(h_1h_2Q_n)^{-1}\max_i\{\xi_i^2N_iM_i\} \to 0,\quad \ln(n)Q_n^{-1}\max_i\{\xi_i^2N_i(N_i-1)M_i(M_i-1)\} \to 0.$$

(D) The stochastic part and ϵ:
(D2) (a) There exists $\delta > 2$ such that

$$n\left\{\sum_{i=1}^n \xi_i^2N_i(N_i-1)M_i(M_i-1)h_1^2h_2^2 + \sum_{i=1}^n \xi_i^2N_i(N_i-1)M_ih_1^2h_2 + \sum_{i=1}^n \xi_i^2M_i(M_i-1)N_ih_1h_2^2 + \sum_{i=1}^n \xi_i^2N_iM_ih_1h_2\right\}\left\{\frac{\ln(n)}{n}\right\}^{2/\delta-1} \to \infty, \tag{8}$$

$$\mathrm{E}(U^{*\delta}) < \infty, \quad \text{and} \quad \mathrm{E}|\epsilon|^\delta < \infty, \tag{9}$$

where $U^* = \sup_{t,s\in[0,1]}|U(t,s)|$;
(b) Otherwise, it is assumed that
(g-subgaussian) there exists a constant $g \in (0,+\infty)$ such that, $\mathrm{E}e^{zU^*} \le e^{g^2z^2/2}$ for all $z \in R$;
(b-subgaussian) there exists a constant $b \in (0,+\infty)$ such that, $\mathrm{E}e^{z\epsilon} \le e^{b^2z^2/2}$ for all $z \in R$.

Remark 8. (i) Assumptions (C8) and (C9) are needed for ensuring uniform convergence and are consistent with Assumptions (C1c) and (C3c) in [31], respectively.
(ii) Assumption (C10) excludes the case where a few subjects dominate the variance term. It is a standard assumption. For instance, let us consider the SDW scheme as an example. If $\{N_i\}_{i=1}^n$ and $\{M_i\}_{i=1}^n$ are i.i.d. copies of positive integer-valued random variables N and M, respectively, then it is valid by using some algebraic calculations and Chebyshev's inequality.

Remark 9. Assumption (D2) is needed for proving the uniform convergence of our proposed estimators. The g-subgaussian assumption of U^* and b-subgaussian assumption of ϵ are stronger than (9). When (8) holds, the subgaussian assumptions can be weakened to the moment conditions (9). Note that Assumption (D2)(a) is consistent with Assumption (C2c) in [31].

Theorem 5 (*Uniform Convergence*). *Under Assumptions (A), (B), (C1), (C8), (C9), (C10) and (D2), we have*

$$\sup_{(t,s)\in[0,1]^2}|\hat\mu(t,s)-\mu(t,s)| = O\left(h_1^2+h_2^2+\left[\ln(n)\left\{\sum_{i=1}^n \xi_i^2N_i(N_i-1)M_i(M_i-1) + \sum_{i=1}^n \xi_i^2N_i(N_i-1)M_i/h_2 + \sum_{i=1}^n \xi_i^2M_i(M_i-1)N_i/h_1 + \sum_{i=1}^n \xi_i^2N_iM_i/(h_1h_2)\right\}\right]^{1/2}\right) \quad a.s. \tag{10}$$

The result can be proved by using Bernstein's inequality, Borel–Cantelli's Lemma, (2), (3) and (5) and Lemmas 2 and 3. In the proof, the derivation of the bias is essentially identical to that in the classical two-dimensional nonparametric regression, but there is an additional challenge of dealing with the within-surface dependence compared with the classical nonparametric regression. On the right-hand side of (10), the upper bound for the bias $\sup_{t,s\in[0,1]}|\mathrm{E}\hat\mu(t,s)-\mu(t,s)|$ is $O(h_1^2+h_2^2)$, whereas the sum of all the other terms is the upper bound of $\sup_{t,s\in[0,1]}|\hat\mu(t,s)-\mathrm{E}\hat\mu(t,s)|$. Specifically, $\ln(n)\sum_{i=1}^n\{\xi_i^2N_i(N_i-1)M_i(M_i-1)\}$ stems from the covariance of the paired observations across different time points and different positions, $\ln(n)\sum_{i=1}^n\{\xi_i^2N_i(N_i-1)M_i\}/h_2$ arises from that at different time points and the same positions, $\ln(n)\sum_{i=1}^n\{\xi_i^2M_i(M_i-1)N_i\}/h_1$ is from that at the same time points and different positions, and $\ln(n)\sum_{i=1}^n\{\xi_i^2N_iM_i\}/(h_1h_2)$ is obtained from the marginal variances of all the observations. Compared to the uniform convergence rate [15,31] in curve functional models, the deviation of the second bound for our surface functional models is more complex.

Corollary 6. *Under the Assumptions of Theorem 5,*
(i) *EW: Assumption (C3) is also valid, the uniform convergence rates in the six sampling design cases are summarized in Table 4. For instance, for the "Sparse+Sparse" case: when* $\ln(n)\overline{N}^5/(n\overline{M}) \to 0$, $\ln(n)\overline{M}^5/(n\overline{N}) \to 0$, $h_1 \asymp \{n\overline{N}\ \overline{M}/\ln(n)\}^{-1/6}$ *and* $h_2 \asymp \{n\overline{N}\ \overline{M}/\ln(n)\}^{-1/6}$ *hold, we have*

$$\sup_{(t,s)\in[0,1]^2}|\hat\mu(t,s)-\mu(t,s)| = O\left(\{\ln(n)/(n\overline{N}\ \overline{M})\}^{1/3}\right) \quad a.s.$$

(ii) *DDW: Assumption (C4) also holds, the uniform convergence rates can be obtained for* $\hat\mu_{DDW}(t,s)$ *for the six sampling designs by replacing* $\xi_i = 1/(\sum_{i=1}^n N_iM_i)$ *and* $\overline{M}$ *in Table 4 by* $\xi_i = 1/\{(\sum_{i=1}^n N_i)M_i\}$ *and* $\overline{M}_H$, *respectively.*
(iii) *SDW: Assumption (C5) also holds, the uniform convergence rates can be obtained for* $\hat\mu_{SDW}(t,s)$ *for the six sampling designs by replacing* $\xi_i = 1/(\sum_{i=1}^n N_iM_i)$, $\overline{N}$, *and* $\overline{M}$ *in Table 4 by* $\xi_i = 1/(nN_iM_i)$, $\overline{N}_H$, *and* $\overline{M}_H$, *respectively.*

For the three weighting schemes, the uniform convergence rates in the six sampling design cases are summarized in Corollary 6. The full results for all the nine cases are shown in Table 3 in the Supplementary Materials. We have the following conclusions: The uniform convergence rate in the "Sparse+Sparse" case, which can never reach $\{\ln(n)/n\}^{1/2}$, is comparable to that of the traditional two-dimensional local linear smoother for independent data; In the "Sparse+Dense" and "Sparse+Ultra-Dense" cases, the convergence rates are consistent with that in the "Sparse" case for curve functional data (Case 1 of Corollary 5.2, [31]), which is always slower than $\{\ln(n)/n\}^{1/2}$; The "Dense+Dense", "Dense+Ultra-Dense" and "Ultra-Dense+Ultra-Dense" cases show the common uniform convergence rate of $\{\ln(n)/n\}^{1/2}$, which is also the rate for the "Dense" and "Ultra-Dense" cases for curve functional data (Cases 2 and 3 of Corollary 5.2, [31]).

4. Numerical studies

4.1. Study 1

To examine the finite sample performance of the proposed estimators, we performed the following simulation studies. We generated data from the following Time-Line surface functional model:

$$Y_{ijk} = X_i(T_{ij}, S_{ik}) + \epsilon_{ijk} = \mu(T_{ij}, S_{ik}) + U_i(T_{ij}, S_{ik}) + \epsilon_{ijk}, \text{ for } j \in \{1, \ldots, N_i\}, k \in \{1, \ldots, M_i\}, i \in \{1, \ldots, n\},$$

where $\mu(t, s) = 0.5\exp(t/4 - s)$, $U_i(t, s) = 2\cos(\pi(t + s))B_{i1} + 2\sin(\pi(t + s))B_{i2}$, $T_{ij} \sim^{iid} \mathcal{N}(0, 1)$, $S_{ik} \sim^{iid} \mathcal{N}(0, 1)$, $B_{i1} \sim^{iid} \mathcal{N}(0, 1/4)$, $B_{i2} \sim^{iid} \mathcal{N}(0, 1/9)$, and $\epsilon_{ijk} \sim^{iid} \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.1$ or $\sigma = 1$. We are interested in estimating $\mu(t, s)$. Let $\lfloor x \rfloor$ be the largest integer not greater than x. We considered the following three cases of (N_i, M_i):
Case 1. N_i and M_i are both i.i.d. from the discrete uniform distribution on the set $\{1, 2, 3\}$;
Case 2. N_i and M_i are i.i.d. from the discrete uniform distribution on the set $\{1, 2, 3\}$ and on the interval $[\lfloor n^{1/2} \rfloor, \lfloor 2n^{1/2} \rfloor]$, respectively;
Case 3. N_i and M_i are both i.i.d. from the discrete uniform distribution on the interval $[\lfloor n^{1/2} \rfloor, \lfloor 2n^{1/2} \rfloor]$;

Case 1, 2 and 3 can be regarded as Sparse+Sparse Case, Sparse+Ultra-Dense Case and Ultra-Dense+Ultra-Dense Case, respectively. For each case, we considered sample sizes $n = 100$ and $n = 300$ and applied our estimation method to $Q = 500$ simulated datasets. We used the Gaussian kernel function $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$. We can use the leave-one-out cross validation method to select the bandwidths [6]. However, the procedure is computationally intensive. We followed the well-known Silverman's rule of thumb [13] to choose the bandwidth parameters in the simulation studies for simplicity.

We set $\xi_i = 1/(\sum_{i=1}^n N_i M_i)$ for EW, $\xi_i = 1/\{(\sum_{i=1}^n N_i)M_i\}$ for DDW, and $\xi_i = 1/(nN_iM_i)$ for SDW. Let both $\{T_1, \ldots, T_E\}$ and $\{S_1, \ldots, S_F\}$ be the equidistant partition on $[-1, 1]$ with the grid length of 0.2, resulting in $E = F = 11$. Then, we calculated the standard deviations (SD) of Q estimators of $\mu(T_l, S_v)$ based on the Q datasets, for $l \in \{1, \ldots, E\}$ and $v \in \{1, \ldots, F\}$. We defined MSD as the mean of the $E \times F$ standard deviations. We defined the empirical mean integrated squared error (EMISE) and empirical mean supremum absolute error (EMSAE): $\text{EMISE}(\hat{\mu}) = Q^{-1}\sum_{q=1}^Q (EF)^{-1}\sum_{l=1}^E\sum_{v=1}^F\{\hat{\mu}^{(q)}(T_l, S_v) - \mu(T_l, S_v)\}^2$, $\text{EMSAE}(\hat{\mu}) = Q^{-1} \times \sum_{q=1}^Q \max_{l,v} |\hat{\mu}^{(q)}(T_l, S_v) - \mu(T_l, S_v)|$, where $\hat{\mu}^{(q)}(\cdot, \cdot)$ is the estimator of $\mu(\cdot, \cdot)$ based on the qth dataset for $q \in \{1, \ldots, Q\}$.

We reported the results in Table 5. We observe that EW has smaller MSD, EMISE and EMSAE compared with DDW and SDW in Case 1 ("Sparse+Sparse" Case). DDW and SDW outperform in Case 2 ("Sparse+Ultra-Dense" Case) and Case 3 ("Ultra-Dense+Ultra-Dense" Case), respectively. The results are consistent with our theoretical findings. Moreover, SDW seems to perform the worst when the time/location domain has sparse data as in Cases 1 and 2 and EW has the largest MSD, EMISE and EMSAE among the three weighting schemes when both the time and location directions have dense enough data as in Case 3.

4.2. Study 2

We see the numerical performance of the local linear estimators when the spatial domain is multivariate (e.g., 2-dimensional) in this subsection. We generated data from the following Time-Plane surface functional model:

$$Y_{ijk} = X_i(T_{ij}, \mathbf{S}_{ik}) + \epsilon_{ijk} = \mu(T_{ij}, \mathbf{S}_{ik}) + U_i(T_{ij}, \mathbf{S}_{ik}) + \epsilon_{ijk}, \text{ for } j \in \{1, \ldots, N_i\}, \ k \in \{1, \ldots, M_i\}, \ i \in \{1 \ldots, n\},$$

where $\mu(t, \mathbf{s}) = 0.5\exp(t/4 - s_1 - s_2)$ with $\mathbf{s} = (s_1, s_2)^T$, $U_i(t, \mathbf{s}) = 2\cos(\pi(t + s_1 + s_2))B_{i1} + 2\sin(\pi(t + s_1 + s_2))B_{i2}$, $T_{ij} \sim^{iid} \mathcal{N}(0, 1)$, $\mathbf{S}_{ik} = (S_{ik1}, S_{ik2})^T$ with $S_{ik1} \sim^{iid} \mathcal{N}(0, 1)$ and $S_{ik2} \sim^{iid} \mathcal{N}(0, 1)$, $B_{i1} \sim^{iid} \mathcal{N}(0, 1/4)$, $B_{i2} \sim^{iid} \mathcal{N}(0, 1/9)$, and $\epsilon_{ijk} \sim^{iid} \mathcal{N}(0, 0.01)$. We considered the following four cases of (N_i, M_i):
Case 1*. N_i and M_i are both i.i.d. from the discrete uniform distribution on the set $\{1, 2, 3\}$;
Case 2*. N_i and M_i are i.i.d. from the discrete uniform distribution on the set $\{1, 2, 3\}$ and on the interval $[\lfloor n^{3/4} \rfloor, \lfloor 2n^{3/4} \rfloor]$, respectively;
Case 3*. N_i and M_i are i.i.d. from the discrete uniform distribution on the set $[\lfloor n^{1/2} \rfloor, \lfloor 2n^{1/2} \rfloor]$ and on the interval $\{1, 2, 3\}$, respectively;
Case 4*. N_i and M_i are i.i.d. from the discrete uniform distribution on the interval $[\lfloor n^{1/2} \rfloor, \lfloor 2n^{1/2} \rfloor]$ and $[\lfloor n^{3/4} \rfloor, \lfloor 2n^{3/4} \rfloor]$, respectively.

Table 5
The MSD, EMISE and EMSAE of $\hat{\mu}$ for EW, DDW and SDW schemes for the time-line surface functional models. The schemes EW $\xi_i = 1/(\sum_{i=1}^{n} N_i M_i)$, DDW $\xi_i = 1/\{(\sum_{i=1}^{n} N_i)M_i\}$, and SDW $\xi_i = 1/(nN_iM_i)$.

	Scheme	$\sigma = 0.1$			$\sigma = 1$		
		Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
Sample size = 100							
MSD	EW	0.1290	0.0940	0.0685	0.1865	0.1090	0.0745
	DDW	0.1327	0.0926	0.0676	0.1946	0.1084	0.0742
	SDW	0.1382	0.0949	0.0667	0.2057	0.1137	0.0734
EMISE	EW	0.0195	0.0096	0.0052	0.0379	0.0127	0.0061
	DDW	0.0206	0.0093	0.0051	0.0412	0.0126	0.0060
	SDW	0.0223	0.0099	0.0049	0.0459	0.0139	0.0059
EMSAE	EW	0.3454	0.2426	0.1711	0.4892	0.2843	0.1890
	DDW	0.3568	0.2393	0.1700	0.5109	0.2826	0.1887
	SDW	0.3673	0.2444	0.1680	0.5349	0.2972	0.1875
Sample size = 300							
MSD	EW	0.0888	0.0628	0.0458	0.1254	0.0717	0.0472
	DDW	0.0911	0.0617	0.0447	0.1311	0.0707	0.0465
	SDW	0.0929	0.0636	0.0438	0.1390	0.0730	0.0459
EMISE	EW	0.0092	0.0042	0.0023	0.0175	0.0054	0.0024
	DDW	0.0097	0.0041	0.0022	0.0193	0.0053	0.0023
	SDW	0.0103	0.0044	0.0021	0.0215	0.0057	0.0022
EMSAE	EW	0.2460	0.1616	0.1084	0.3359	0.1872	0.1139
	DDW	0.2503	0.1596	0.1072	0.3489	0.1853	0.1126
	SDW	0.2554	0.1649	0.1058	0.3656	0.1896	0.1122

Table 6
The MSD, EMISE and EMSAE of $\hat{\mu}$ for EW, DDW1, DDW2 and SDW schemes for the time-plane surface functional models. The schemes EW $\xi_i = 1/(\sum_{i=1}^{n} N_i M_i)$, DDW1 $\xi_i = 1/\{(\sum_{i=1}^{n} N_i)M_i\}$, DDW2 $\xi_i = 1/\{(\sum_{i=1}^{n} M_i)N_i\}$, and SDW $\xi_i = 1/(nN_iM_i)$.

	Scheme	Case 1*	Case 2*	Case 3*	Case 4*
MSD	EW	0.1187	0.0573	0.0869	0.0406
	DDW1	0.1287	0.0568	0.0946	0.0403
	DDW2	0.1255	0.0608	0.0854	0.0402
	SDW	0.1358	0.0603	0.0931	0.0399
EMISE	EW	0.0317	0.0054	0.0217	0.0031
	DDW1	0.0342	0.0053	0.0231	0.0031
	DDW2	0.0334	0.0058	0.0214	0.0031
	SDW	0.0361	0.0058	0.0229	0.0030
EMSAE	EW	0.6524	0.2585	0.5536	0.2000
	DDW1	0.6746	0.2581	0.5672	0.1996
	DDW2	0.6687	0.2666	0.5495	0.1994
	SDW	0.6948	0.2663	0.5645	0.1990

In each case, we generated $Q = 500$ datasets, each with a sample size of $n = 300$. For the time-plane surface functional models, Cases 1*, 2*, 3* and 4* can be regarded as Sparse+Sparse Case, Sparse+Ultra-Dense Case, Ultra-Dense+Sparse, and Ultra-Dense+Ultra-Dense Case, respectively. We considered schemes EW $\xi_i = 1/(\sum_{i=1}^{n} N_i M_i)$, DDW1 $\xi_i = 1/\{(\sum_{i=1}^{n} N_i)M_i\}$, DDW2 $\xi_i = 1/\{(\sum_{i=1}^{n} M_i)N_i\}$, and SDW $\xi_i = 1/(nN_iM_i)$. Let $\{T_1, \ldots, T_E\}$, $\{S_{11}, \ldots, S_{1F}\}$ and $\{S_{21}, \ldots, S_{2F}\}$ all be the equidistant partition on $[-1, 1]$ with the grid length of 0.2. We calculated the standard deviations (SD) of Q estimators of $\mu(T_l, S_{1v_1}, S_{2v_2})$ based on the Q datasets for $l = 1, \ldots, E$, $v_1 = 1, \ldots, F$ and $v_2 = 1, \ldots, F$. We defined MSD as the mean of the $E \times F^2$ standard deviations. We defined the EMISE and the EMSAE similarly as in Section 4.1. We summarized the results in Table 6. We observed that EW, DDW1, DDW2 and SW are the most favorable schemes for Case 1*, Case 2*, Case 3*, and Case 4*, respectively.

We will state the conditional bias and conditional variance, asymptotic normality and uniform convergence properties for surface functional models when the spatial domain is multivariate in future works. We note that, similar to Section 3.2, for multivariate spatial domain, the EW, DDW1, DDW2 and SW schemes are the most favorable schemes for "Sparse+Sparse", "Sparse+Ultra-Dense", "Ultra-Dense+Sparse" and "Ultra-Dense+Ultra-Dense" cases in terms of estimating efficiency, respectively.

4.3. Data analysis

We illustrate our proposed method via the application to a data set extracted from the national database for autism research (NDAR) (https://ndar.nih.gov/), which is an NIH-funded research data repository that aims at accelerating progress in autism spectrum disorders (ASD) research through data sharing, data harmonization, and research result reporting. A total of 416 MRI scans were selected from 253 normal children (126 males and 127 females). Among them, 137 children have 1 scan, 78 children have 2 scans, 34 subjects have 3 scans, 3 children have 5 scans, and 1 child has 6 scans.

First, we performed two important steps to process the diffusion tensor imaging (DTI) data: one is to implement a weighted least squares estimation method [1] to construct the diffusion tensors, and the other is to use an FMRIB Software Library (FSL) Tract-Based Spatial Statistics (TBSS) pipeline [27] to register DTIs from multiple subjects to create a mean image and a mean skeleton. Specifically, by using FMRIB Software Library, we obtained the maps of fractional anisotropy (FA) for all the subjects from the DTI posterior to Eddy current correction and automatic brain extraction. FA maps were then fed into the TBSS tool, which is also a part of the FMRIB Software Library. In the TBSS analysis, the FA data of all the subjects were aligned into a common space by non-linear registration, and the mean FA image was created and thinned to obtain a mean FA skeleton, which represents the centers of all the white matter (WM) tracts shared by the group. Subsequently, each subject's aligned FA data sets were projected onto this skeleton. The FA values were extracted at each grid point across multiple time points measured by ln(ages) (1 to 6 time points) along the selected 45 fiber tracts (arclength) for all 253 infants. Although several DTI fiber tracts were tracked, we focus on the corpus callosum to illustrate the applicability of our method.

The aim of this data analysis is to investigate the nonparametric relationship between FA value and ln(age) as well as location along the corpus callosum. We applied our surface functional model (1) to characterize the data with Y, T and S denoting FA value, ln(ages), and arclength, respectively. As mentioned in Section 2, we use the local linear regression to obtain $\hat{\mu}(t, s)$. We present the plots of FA value against ln(age) and arclength, respectively, in Fig. 2 for the schemes EW $\xi_i = 1/(\sum_{i=1}^{n} N_i M_i)$, DDW $\xi_i = 1/\{(\sum_{i=1}^{n} N_i)M_i\}$ and SDW $\xi_i = 1/(nN_iM_i)$. The three weighting schemes appear to yield similar results. First, FA seems to attain the smallest value when the arclength is around 15.5 and the ln(age) is around 4.7. Second, FA appears to increase with ln(age) for a given arclength. Third, FA achieves the largest when the arclengths are around 0 and around 48. Moreover, FA tends to first decrease and then increase with the turning point being around 35 as the arclength increases, for a given ln(age).

5. Conclusion

In this paper, we propose the surface functional models (SFM) framework and estimate the mean functions based on local linear smoothers. The SFM is commonly encountered in brain imaging studies, where imaging features, such as fractional anisotropy, are observed over different brain locations at multiple time points. The SFM are significantly different from the multivariate functional models with two-dimensional predictor variables. The covariance structure within and across the two domains (time-spatial) of surface functional data is much more complex than that of the single-domain curve functional data [15,31]. In our work, we provide the conditional bias, the conditional variance, the asymptotic normality, and the uniform convergence properties of the estimator $\hat{\mu}(t, s)$ for the surface functional models on a unified platform. The other major contribution of our work is that, based on the EW, DDW and SDW schemes, respectively, we provide the explicit definition of the nine sampling design scenarios, and derive the optimal bandwidth orders for each scenario. We would like to emphasize that the asymptotic results for the mean function estimator are specific to the sampling designs of the surface functional data. Most importantly, we perform an extensive comparisons among the EW, DDW and SDW schemes both theoretically and empirically, and provide the suggestions on which one is preferred over the other two for the different sampling designs, see Section 3.2. Specifically, we suggest to use EW, DDW and SW schemes for "Sparse+Sparse" Case, "Sparse+Ultra-Dense" Case and "Ultra-Dense+Ultra-Dense" Case, respectively. These results will be extended to the SFM with multivariate spatial domain formally in the future. Our work present in this paper focuses on mean function estimation in the new framework of surface functional models. We defer the covariance function estimation method in the Supplementary Materials. As another future research direction, we will systematically study the theoretical properties and the finite sample performance of the covariance function estimator as well as functional principal component analysis (FPCA) using the estimated covariance function.

CRediT authorship contribution statement

Ziqi Chen: Formal analysis, Software, Investigation, Writing - original draft. **Jianhua Hu:** Supervision, Visualization, Writing - review & editing. **Hongtu Zhu:** Supervision, Conceptualization, Methodology, Validation, Data curation, Writing - review & editing.

Acknowledgments

The authors thank the Editor, Associate Editor and the reviewers for helpful comments and suggestions, which have led to improvements of this article. The research of Dr. Chen was supported by NSFC, China grant 11871477. The research of Dr. Zhu was supported by NIH, USA grants MH116527 and MH086633. The research of Dr. Hu was supported by NIH/NCI, USA grants 5 P30 CA013696-43, R01AI143886 and R01CA219896.

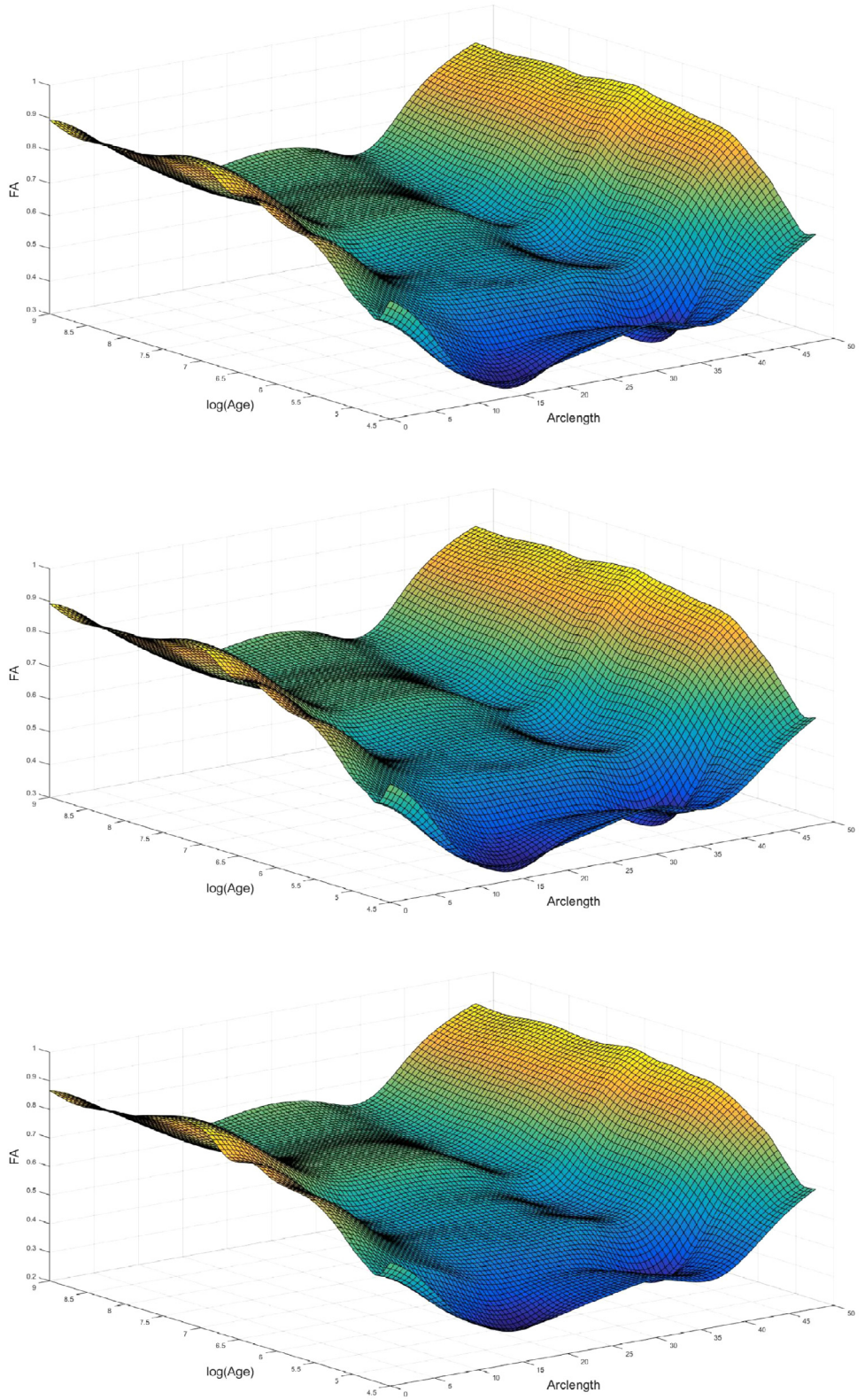


Fig. 2. Data analysis. The FA value varies as a function of ln(age) and arclength. The first, the second and the third ones are $\hat{\mu}_{EW}(t,s)$, $\hat{\mu}_{DDW}(t,s)$ and $\hat{\mu}_{SDW}(t,s)$, respectively.

Appendix A

Optimal bandwidth orders and partition of the surface functional data

We first consider the optimal bandwidth orders and partition of the functional data for the EW weighting scheme under Assumptions (A), (B), (C1), (C2) and (C3). Let $\overline{N} = O(n^{\alpha})$ and $\overline{M} = O(n^{\beta})$ with $\alpha \geq 0$ and $\beta \geq 0$. For any (t, s) in the interior of $[0, 1]^2$, the asymptotic mean-squared error is of order

$$n^{-1} + n^{-(1+\beta)}h_2^{-1} + n^{-(1+\alpha)}h_1^{-1} + n^{-(1+\alpha+\beta)}h_1^{-1}h_2^{-1} + h_1^4 + h_2^4.$$

Thus, the optimal bandwidths satisfy $h_1^5 \asymp n^{-(1+\alpha)} + n^{-(1+\alpha+\beta)}h_2^{-1}$ and $h_2^5 \asymp n^{-(1+\beta)} + n^{-(1+\alpha+\beta)}h_1^{-1}$. If we set

$$n^{-(1+\alpha)} \preceq n^{-(1+\alpha+\beta)}h_2^{-1} \quad \text{and} \quad n^{-(1+\beta)} \preceq n^{-(1+\alpha+\beta)}h_1^{-1}, \tag{11}$$

then we have $h_1^5 \asymp n^{-(1+\alpha+\beta)}h_2^{-1}$ and $h_2^5 \asymp n^{-(1+\alpha+\beta)}h_1^{-1}$. Thus, the optimal bandwidths satisfy

$$h_1 \asymp n^{-(1+\alpha+\beta)/6} \quad \text{and} \quad h_2 \asymp n^{-(1+\alpha+\beta)/6}. \tag{12}$$

Eq. (11) is equivalent to that $n^{\beta}h_2 \to 0$ and $n^{\alpha}h_1 \to 0$ hold. It follows from (12) that we have

$$n^{\beta}h_2 = n^{(5\beta-\alpha-1)/6} \to 0 \quad \text{and} \quad n^{\alpha}h_1 = n^{(5\alpha-\beta-1)/6} \to 0.$$

This means that $5\beta - \alpha - 1 < 0$ and $5\alpha - \beta - 1 < 0$ hold. That is, both α and β are small. This is the so-called "Sparse+Sparse" case. The optimal bandwidth orders and partition of the functional data of the other eight cases can be similarly discussed.

Second, the optimal bandwidth orders and partition of the functional data for the DDW scheme under Assumptions (A), (B), (C1), (C2) and (C4) can be derived similarly. In this case, we set $\overline{N} = O(n^{\alpha})$ and $\overline{M}_H = O(n^{\beta})$ with $\alpha \geq 0$ and $\beta \geq 0$.

Third, the optimal bandwidth orders and partition of the functional data for the SDW scheme under Assumptions (A), (B), (C1), (C2) and (C5) can be derived similarly. In this case, we set $\overline{N}_H = O(n^{\alpha})$ and $\overline{M}_H = O(n^{\beta})$ with $\alpha \geq 0$ and $\beta \geq 0$.

Lemmas

Lemma 1 (*Rivasplata [25]*). *If X is g-subgaussian, then for any $\alpha \in R$, αX is $|\alpha g|$-subgaussian. If X_1 and X_2 are random variables such that X_i is g_i-subguassian ($i = 1,2$), then $X_1 + X_2$ is $g_1 + g_2$-subguassian.*

Remark 10. Lemma 1 holds even when X_1 and X_2 are dependent (or correlated) by its proof in [25].

The following lemma shows the rate of uniform convergence for

$$L_n(t, s) = h_1h_2 \sum_{i=1}^{n} \xi_i \sum_{j=1}^{N_i} \sum_{k=1}^{M_i} K_{h_1}\left(T_{ij} - t\right) K_{h_2}\left(S_{ik} - s\right) U_{ijk}.$$

Lemma 2. *Under Assumptions (A), (B), (C1), (C8), (C9), (C10) and (D2), we have*

$$\sup_{t,s\in[0,1]} |L_n(t, s) - \mathrm{E}L_n(t, s)| = O\left(\left[\ln(n) \left\{ \sum_{i=1}^{n} \xi_i^2 N_i M_i h_1 h_2 + \sum_{i=1}^{n} \xi_i^2 N_i(N_i - 1) M_i h_1^2 h_2 \right.\right.\right.$$
$$\left.\left.\left. + \sum_{i=1}^{n} \xi_i^2 M_i(M_i - 1) N_i h_1 h_2^2 + \sum_{i=1}^{n} \xi_i^2 N_i(N_i - 1) M_i(M_i - 1) h_1^2 h_2^2 \right\} \right]^{1/2} \right) \quad a.s.$$

Proof. We define a_n as

$$\left[\ln(n) \left\{ \sum_{i=1}^{n} \xi_i^2 N_i(N_i - 1) M_i(M_i - 1) h_1^2 h_2^2 + \sum_{i=1}^{n} \xi_i^2 N_i(N_i - 1) M_i h_1^2 h_2 + \sum_{i=1}^{n} \xi_i^2 M_i(M_i - 1) N_i h_1 h_2^2 + \sum_{i=1}^{n} \xi_i^2 N_i M_i h_1 h_2 \right\} \right]^{1/2}.$$

It follows from Assumption (C8) that there are two constants $\gamma_1 > 0$ and $\gamma_2 > 0$ such that $n^{\gamma_1}h_1a_n \to \infty$ and $n^{\gamma_2}h_2a_n \to \infty$ hold. Let $\chi(\gamma_1)$ and $\varsigma(\gamma_2)$ be equidistance partitions on [0,1] with grid length $n^{-\gamma_1}$ and $n^{-\gamma_2}$, respectively. Therefore, $\sup_{t,s\in[0,1]} |L_n(t, s) - \mathrm{E}L_n(t, s)|$ is upper bounded by

$$\sup_{t\in\chi(\gamma_1), s\in\varsigma(\gamma_2)} |L_n(t, s) - \mathrm{E}L_n(t, s)| + D_{1n} + D_{2n},$$

where

$$D_{1n} = \sup_{|t-t'|<n^{-\gamma_1},|s-s'|<n^{-\gamma_2}} |L_n(t,s) - L_n(t',s')|,\ D_{2n} = \sup_{|t-t'|<n^{-\gamma_1},|s-s'|<n^{-\gamma_2}} |\mathrm{E}L_n(t,s) - \mathrm{E}L_n(t',s')|.$$

Step 1: To prove $\max\{D_{1n}, D_{2n}\} = o(a_n)$, *a.s.* Specifically, we have

$$\begin{aligned} D_{1n} &\le \sup_{|t-t'|<n^{-\gamma_1},|s-s'|<n^{-\gamma_2}} h_1h_2 \sum_{i=1}^{n} \xi_i \sum_{j=1}^{N_i}\sum_{k=1}^{M_i} \left| K_{h_1}\left(T_{ij}-t\right) K_{h_2}\left(S_{ik}-s\right) - K_{h_1}\left(T_{ij}-t'\right) K_{h_2}\left(S_{ik}-s'\right) \right| U_{ijk}, \\ &\le \sum_{i=1}^{n} \xi_i \sum_{j=1}^{N_i}\sum_{k=1}^{M_i} U_{ijk} M_K L\left(\frac{n^{-\gamma_1}}{h_1} + \frac{n^{-\gamma_2}}{h_2}\right) \le \sum_{i=1}^{n} \xi_i N_i M_i \sup_{t,s} |U_i(t,s)| M_K L\left(\frac{n^{-\gamma_1}}{h_1} + \frac{n^{-\gamma_2}}{h_2}\right) \\ &\le n \max_i\{\xi_i N_i M_i\} \left(\frac{1}{n}\sum_{i=1}^{n} \sup_{t,s} |U_i(t,s)|\right) M_K L\left(\frac{n^{-\gamma_1}}{h_1} + \frac{n^{-\gamma_2}}{h_2}\right). \end{aligned}$$

By Assumptions (C9) and (D2), we have $\sup_n n \max_i(\xi_i N_i M_i) \le B < \infty$ and $\mathrm{E}(\sup_{t,s} |U_i(t,s)|) < \infty$. We also have $n^{\gamma_1}h_1 a_n \to \infty$ and $n^{\gamma_2}h_2 a_n \to \infty$. Thus, $\max\{D_{1n}, D_{2n}\} = o(a_n)$, a.s.

Step 2: To prove $\sup_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)} |L_n(t,s) - \mathrm{E}L_n(t,s)| = O(a_n)$ *a.s.* Firstly, we prove the result under Assumption (D2)(a). Let $A_n = a_n\{n/\ln(n)\}$ and the truncated $L_n(t,s)$ be

$$L_n^*(t,s) = h_1h_2 \sum_{i=1}^{n} \xi_i \sum_{j=1}^{N_i}\sum_{k=1}^{M_i} K_{h_1}\left(T_{ij}-t\right) K_{h_2}\left(S_{ik}-s\right) U_{ijk} I\left(\left|U_{ijk}\right| \le A_n\right),$$

where $I(\cdot)$ is the indicator function. Then, $\sup_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)} |L_n(t,s) - \mathrm{E}L_n(t,s)|$ is upper bounded by

$$\sup_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)} |L_n^*(t,s) - \mathrm{E}L_n^*(t,s)| + E_{1n} + E_{2n},$$

where

$$E_{1n} = \sup_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)} h_1h_2 \sum_{i=1}^{n} \xi_i \sum_{j=1}^{N_i}\sum_{k=1}^{M_i} K_{h_1}\left(T_{ij}-t\right) K_{h_2}\left(S_{ik}-s\right) U_{ijk} I\left(\left|U_{ijk}\right| > A_n\right),$$

$$E_{2n} = \sup_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)} h_1h_2 \sum_{i=1}^{n} \xi_i \sum_{j=1}^{N_i}\sum_{k=1}^{M_i} \mathrm{E}\left\{K_{h_1}\left(T_{ij}-t\right) K_{h_2}\left(S_{ik}-s\right) U_{ijk} I\left(\left|U_{ijk}\right| > A_n\right)\right\}.$$

Moreover,

$$h_1h_2 \sum_{i=1}^{n} \xi_i \sum_{j=1}^{N_i}\sum_{k=1}^{M_i} K_{h_1}\left(T_{ij}-t\right) K_{h_2}\left(S_{ik}-s\right) U_{ijk} I\left(\left|U_{ijk}\right| > A_n\right) \le BM_K^2 A_n^{1-\delta}\left(n^{-1}\sum_{i=1}^{n} \sup_{t,s} |U_i(t,s)|^{\delta}\right),$$

which yields $\max\{E_{1n},\ E_{2n}\} = o(a_n)$, a.s. by using Assumptions (A), (C9) and (D2)(a).

We can rewrite $L_n^*(t,s) - \mathrm{E}L_n^*(t,s) = \sum_{i=1}^{n}\{W_{in}(t,s) - \mathrm{E}W_{in}(t,s)\}$, where

$$W_{in}(t,s) = h_1h_2\xi_i \sum_{j=1}^{N_i}\sum_{k=1}^{M_i} K_{h_1}(T_{ij}-t)K_{h_2}(S_{ik}-s)U_{ijk} I(\left|U_{ijk}\right| \le A_n).$$

We have $|W_{in}(t,s) - \mathrm{E}W_{in}(t,s)| \le 2BM_K^2 A_n/n$. Moreover, there exists an $M_V > 0$ such that $\mathrm{E}\{W_{in}(t,s) - \mathrm{E}W_{in}(t,s)\}^2 \le M_V c_{in}$, where c_{in} is given by $\xi_i^2\{N_i(N_i-1)M_i(M_i-1)h_1^2h_2^2 + N_i(N_i-1)M_ih_1^2h_2 + M_i(M_i-1)N_ih_1h_2^2 + N_iM_ih_1h_2\}$. For $M^* > 0$ large enough, using Bernstein's inequality, it follows that

$$\begin{aligned} \Pr\left(\sup_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)} |L_n^*(t,s) - \mathrm{E}L_n^*(t,s)| > M^*a_n\right) &\le n^{\gamma_1+\gamma_2} \Pr\left(|L_n^*(t,s) - \mathrm{E}L_n^*(t,s)| > M^*a_n\right) \\ &\le 2n^{\gamma_1+\gamma_2} \exp\left(-\frac{M^{*2}a_n^2/2}{M_V a_n^2/\ln(n) + 2BM_K^2 M^* A_n a_n/(3n)}\right) \\ &\le 1/n^2. \end{aligned}$$

By Borel–Cantelli's lemma, we have $\sup_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)} |L_n^*(t,s) - \mathrm{E}L_n^*(t,s)| = O(a_n)$ *a.s.* and the proof is complete.

Secondly, we prove the result under Assumption (D2)(b). Let $b_n = \sum_{i=1}^n c_{in}$. U^* is g-subguassian, thus U_{ijk}'s are all g-subguassian by the equivalence of exponential estimates in [17] (page 67). By Lemma 1, we have

$$v_{in}(t,s) := h_1h_2\xi_i \sum_{j=1}^{N_i}\sum_{k=1}^{M_i} K_{h_1}(T_{ij}-t)K_{h_2}(S_{ik}-s)U_{ijk}$$

is $\xi_i N_i M_i M_K^2 g$-subguassian, is thus $BM_K^2 g/n$-subguassian. By the equivalence of exponential estimates in [17] (pp. 67, 68), for $z>0$ and $z\sqrt{\mathrm{Var}(v_{in}(t,s))}\to 0$, we can show that by Chen and Leng [3]

$$\ln\{\mathrm{E}\exp(zv_{in}(t,s))\} \le z^2\mathrm{Var}(v_{in}(t,s)).$$

There exists an $M_V^*>0$ such that $\mathrm{Var}(v_{in}(t,s)) < M_V^* c_{in}$ holds for all $(t,s)\in[0,1]^2$. Note that $\sqrt{M_V^*}M^*a_n\sqrt{c_{in}}/(2M_V^*\sum_{i=1}^n c_{in})\to 0$ holds for all i by Assumption (C10). Thus, for large enough $M^*>0$, we have

$$\begin{aligned}\Pr\left[\{L_n(t,s)-\mathrm{E}L_n(t,s)\} > M^*a_n\right] &\le \exp(-zM^*a_n)\prod_{i=1}^n \mathrm{E}\exp\{zv_{in}(t,s)\}\\ &\le \exp(-zM^*a_n)\exp\left\{M_V^*\left(\sum_{i=1}^n c_{in}\right)z^2\right\} \end{aligned} \tag{13}$$

for all $(t,s)\in[0,1]^2$. Thus, (13) is maximized as $z = M^*a_n/(2M_V^*\sum_{i=1}^n c_{in})$ and the maximizer equals

$$\exp\left(-\frac{M^{*2}a_n^2}{4M_V^*\sum_{i=1}^n c_{in}}\right).$$

Similarly, for all $(t,s)\in[0,1]^2$,

$$\Pr\left[\{L_n(t,s)-\mathrm{E}L_n(t,s)\} < -M^*a_n\right] \le \exp\left(-\frac{M^{*2}a_n^2}{4M_V^*\sum_{i=1}^n c_{in}}\right).$$

Thus, for all $(t,s)\in[0,1]^2$, we have

$$\Pr\left\{|L_n(t,s)-\mathrm{E}L_n(t,s)| > M^*a_n\right\} \le 2\exp\left\{-M^{*2}a_n^2/\left(4M_V^*b_n\right)\right\}.$$

Therefore for large enough $M^*>0$, we get

$$\begin{aligned}\Pr\left(\sup_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)}|L_n(t,s)-\mathrm{E}L_n(t,s)| > M^*a_n\right) &\le n^{\gamma_1+\gamma_2}\max_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)}\Pr\left(|L_n(t,s)-EL_n(t,s)| > M^*a_n\right)\\ &\le 2n^{\gamma_1+\gamma_2}\exp\{-M^{*2}a_n^2/\left(4M_V^*b_n\right)\} \le \frac{1}{n^2}.\end{aligned}$$

By Borel–Cantelli's lemma, it follows that

$$\sup_{t\in\chi(\gamma_1),s\in\varsigma(\gamma_2)}|L_n(t,s)-\mathrm{E}L_n(t,s)| = O(a_n) \qquad a.s.,$$

which completes the proof. □

Define

$$L_{1n}(t,s) = h_1h_2\sum_{i=1}^n \xi_i\sum_{j=1}^{N_i}\sum_{k=1}^{M_i} K_{h_1}(T_{ij}-t)K_{h_2}(S_{ik}-s)(T_{ij}-t)U_{ijk},$$

$$L_{2n}(t,s) = h_1h_2\sum_{i=1}^n \xi_i\sum_{j=1}^{N_i}\sum_{k=1}^{M_i} K_{h_1}(T_{ij}-t)K_{h_2}(S_{ik}-s)(S_{ik}-s)U_{ijk}.$$

Using the similar arguments as in the proof of Lemma 2 yields Lemma 3.

Lemma 3. *Under Assumptions (A), (B), (C1), (C8), (C9), (C10) and (D2), we have*

$$\begin{aligned}\sup_{t,s\in[0,1]}|L_{1n}(t,s)-\mathrm{E}L_{1n}(t,s)| = o\Bigg(\Bigg[\ln(n)\Bigg\{&\sum_{i=1}^n \xi_i^2N_iM_ih_1h_2 + \sum_{i=1}^n \xi_i^2N_i(N_i-1)M_ih_1^2h_2\\ &+\sum_{i=1}^n \xi_i^2M_i(M_i-1)N_ih_1h_2^2 + \sum_{i=1}^n \xi_i^2N_i(N_i-1)M_i(M_i-1)h_1^2h_2^2\Bigg\}\Bigg]^{1/2}\Bigg) \qquad a.s.\end{aligned}$$

and

$$\sup_{t,s\in[0,1]} |L_{2n}(t,s) - \mathrm{E}L_{2n}(t,s)| = o\left(\left[\ln(n) \left\{ \sum_{i=1}^{n} \xi_i^2 N_i M_i h_1 h_2 + \sum_{i=1}^{n} \xi_i^2 N_i (N_i - 1) M_i h_1^2 h_2 + \sum_{i=1}^{n} \xi_i^2 M_i (M_i - 1) N_i h_1 h_2^2 + \sum_{i=1}^{n} \xi_i^2 N_i (N_i - 1) M_i (M_i - 1) h_1^2 h_2^2 \right\} \right]^{1/2} \right) \quad a.s.$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2020.104664.

References

[1] P.J. Basser, J. Mattlello, D. Lebihan, Estimation of the effective self-diffusion tensor from the nmr spin echo, J. Magn. Reson. Ser. B 103 (1994) 247–254.
[2] Z. Chen, Q. Gao, B. Fu, H. Zhu, Monotone nonparametric regression for functional/longitudinal data, Statist. Sinica 29 (2019) 2229–2249.
[3] Z. Chen, C. Leng, Dynamic covariance models, J. Amer. Statist. Assoc. 111 (2016) 1196–1207.
[4] K. Chen, H.G. Müller, Modeling repeated functional observations, J. Amer. Statist. Assoc. 107 (2012) 1599–1609.
[5] N. Cressie, Statistics for Spatial Data, Wiley, 1993.
[6] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Chapman & Hall/CRC, New York, 1996.
[7] L. Fass, Imaging and cancer: A review, Mol. Oncol. 2 (2008) 115–152.
[8] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer Verlag, New York, 2006.
[9] K.J. Friston, Modalities, modes, and models in functional neuroimaging, Science 326 (2009) 399–403.
[10] T. Gasser, A. Kneip, Searching for structure in curve samples, J. Amer. Statist. Assoc. 90 (1995) 1179–1188.
[11] W. Guo, Functional mixed effects models, Biometrics 58 (2002) 121–128.
[12] P. Hall, H.G. Müller, J.L. Wang, Properties of principal component methods for functional and longitudinal data analysis, Ann. Statist. 34 (2006) 1493–1517.
[13] W. Härdle, M. Müller, S. Sperlich, A. Werwatz, Nonparametric and Semiparametric Models, in: Springer Series in Statistics, Springer, New York, 2004.
[14] I. Heywood, S. Cornelius, S. Carver, An Introduction to Geographical Information Systems, third ed., Prentice Hall, New York, 2006.
[15] Y. Li, T. Hsing, Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data, Ann. Statist. 38 (2010) 3321–3351.
[16] D. Liebl, Inference for sparse and dense functional data with covariate adjustments, J. Multivariate Anal. 170 (2019) 315–335.
[17] Z. Lin, Z. Bai, Probability Inequalities, Springer, New York, 2010.
[18] J.S. Morris, R.J. Carroll, Wavelet-based functional mixed models, J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (2006) 179–199.
[19] E. Niedermeyer, F.L. Da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Williams & Wilkins, Baltimore, 2004.
[20] A. Pagan, A. Ullah, Nonparametric Econometrics, Cambridge University Press, Cambridge, 1999.
[21] D. Paul, J. Peng, Consistency of restricted maximum likelihood estimators of principal components., Ann. Statist. 37 (2009) 1229–1271.
[22] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.
[23] J. Rice, B. Silverman, Estimating the mean and covariance structure nonparametrically when the data are curves, J. R. Stat. Soc. Ser. B Stat. Methodol. 53 (1991) 233–243.
[24] J.A. Rice, C.O. Wu, Nonparametric mixed effects models for unequally sampled noisy curves, Biometrics 57 (2001) 253–259.
[25] O. Rivasplata, Subgaussian random variables: An expository note, 2012, Technical report.
[26] D. Ruppert, M.P. Wand, Multivariate locally weighted least squares regression, Ann. Statist. 22 (1994) 1346–1370.
[27] S.M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T.E. Nichols, C. Mackay, K.E. Watkins, O. Ciccarelli, M.Z. Cader, P.M. Matthews, T.E. Behrens, Tractbased spatial statistics: voxelwise analysis of multi-subject diffusion data, NeuroImage 31 (2006) 1487–1505.
[28] X. Yan, X. Pu, Y. Zhou, X. Xun, Convergence rate of principal component analysis with local linear smoother for functional data under a unified weighing scheme, Statist. Theory Relat. Fields 4 (2020) 55–65.
[29] F. Yao, H.G. Müller, J.L. Wang, Functional data analysis for sparse longitudinal data, J. Amer. Statist. Assoc. 100 (2005) 577–590.
[30] J.T. Zhang, J. Chen, Statistical inferences for functional data, Ann. Statist. 35 (2007) 1052–1079.
[31] X. Zhang, J.L. Wang, From sparse to dense functional data and beyond, Ann. Statist. 44 (2016) 2281–2321.
[32] H. Zhu, K. Chen, X. Luo, Y. Yuan, J.L. Wang, Fmem: functional mixed effects models for longitudinal functional responses, Statist. Sinica 29 (2019) 2007–2033.
[33] H. Zhu, R. Li, L. Kong, Multivariate varying coefficient model for functional responses, Ann. Statist. 40 (2012) 2634–2666.