DOI 10.1007/s10986-023-09586-z
*Lithuanian Mathematical Journal, Vol. 63, No. 1, January, 2023, pp. 13–30*

# On estimation and prediction in spatial functional linear regression model

**Stéphane Bouka** [a], **Sophie Dabo-Niang** [b,c], **and Guy Martial Nkiet** [a]

[a] Université des Sciences et Techniques de Masuku, BP 913 Franceville, Gabon
[b] Laboratoire Paul Painlevé, CNRS UMR 8524, Université de Lille, France
[c] INRIA-MODAL, F-59000 Lille, France
(e-mail: stephane.bouka@univ-masuku.org; sophie.dabo@univ-lille.fr; guymartial.nkiet@univ-masuku.org)

Received January 18, 2022; revised November 8, 2022

**Abstract.** We consider a spatial functional linear regression, where a scalar response is related to a square-integrable spatial functional process. We use a smoothing spline estimator for the functional slope parameter and establish a finite sample bound for variance of this estimator. Then we give the optimal bound of the prediction error under mixing spatial dependence. Finally, we illustrate our results by simulations and by an application to ozone pollution forecasting at nonvisited sites.

*MSC:* 62J05, 62M30

*Keywords:* functional linear regression, spatial functional process, mixing spatial dependence

## 1 Introduction

Functional data analysis (FDA) is a field that brings together statistical methods allowing us to process data that are digitized points of curves representing, for example, the evolution of random phenomena over time. It has had, over the last two decades, an extensive development which allowed the introduction of methods well suited for the analysis of large and complex data with a space and/or time-dynamic component that abound in a number of disciplines, such as environmental sciences, neuroimaging, and genomic, epidemiology, hydrology. In these domains, we are often interested in studying relationships between a real-valued response variable and an explanatory variable of functional nature. Since the seminal paper of [17] on a functional linear model with a scalar response, several types of functional linear models have been developed for different purposes. Estimation of the slope function is a crucial issue, which has been addressed in different ways in the literature: [12] described and compared estimation procedures based on partial least squares, ridge regression, and principal component regression, [5] introduced an estimator based on principal components analysis, [6] proposed an estimator based on B-spline expansion of the functional coefficient, whereas [9] prolonged this work by using a smoothing splines approach, [19] used a Fourier basis expansion under the assumption that the slope function and the explanatory variable are periodic, and [8] proposed a thresholded projection estimator with tuning parameter selected by minimizing a stochastic penalized contrast function. Another tackled problem is that of prediction of the response given an unsampled value of the explanatory variable, but it is

0363-1672/23/6301-0013 

related to the previous estimation problem since prediction is usually achieved by replacing the slope function by its estimator in the model. The optimal mean-square convergence rates of predictor were determined in [4] and [9] for estimators of the slope function based, respectively, on principal component analysis and on a smoothing spline approach. All these works consider the case where the sample consists of independent observations of the involved variables. However, in a number of disciplines, such as environmental sciences, agronomy, or mining, the data have an inherent spatial component in addition to their functional nature. An example in meteorological sciences is provided by long time series of meteorological variables recorded at each point of a monitoring network [14]. In this case the observations are no more independent but are rather spatially dependent. Spatial statistics is a field that has emerged over the last ten years to analyze this kind of data by assuming specific dependence structures in the data, and as it was pointed out in [14], it includes methods that use FDA methods to model spatial big data, including functional data. Nevertheless, less attention has been paid to functional linear regression with spatially dependent data. There are some researches on functional spatial linear prediction using kriging methods (see, e.g., [1, 2, 13, 14, 15, 16], and [21]). Spatial autoregressive functional models were considered in [22, 23]. Prediction based on spatial linear regression model with derivatives is tackled in [3], and the used methodology for determining this prediction is based on the moment method combined with the one of regularization by two sequences decreasing to zero. However, the theoretical rate of convergence of its prediction error is not optimal. It then highlights the interest of considering another method that would allow us to obtain the optimal prediction rate. In this paper, we consider the following spatial functional linear regression model:

$$Y_{\mathbf{i}} = \beta_0 + \int_0^1 \beta(t) X_{\mathbf{i}}(t)\,\mathrm{d}t + \epsilon_{\mathbf{i}}, \quad \mathbf{i} \in \mathcal{I}_{\mathbf{n}} = \{1, \ldots, n\}^d,\ d \geqslant 1, \tag{1.1}$$

where $Y_{\mathbf{i}}$ is an $\mathbb{R}$-valued random variable, $\beta_0$ is an unknown constant, $X_{\mathbf{i}}$ is a random function belonging to the space $\mathcal{F} = L^2([0,1])$ of square-integrable functions endowed with seminorm, $\beta$ is an unknown function representing the slope function, and $\epsilon_{\mathbf{i}}$ is a centered random spatial noise, independent of $X_{\mathbf{i}}$ and with known variance $\sigma_\epsilon^2$. This model is just an extension of the usual functional linear regression model to the case of spatially dependent data, observed on a grid $\mathcal{I}_{\mathbf{n}}$ of points in $\mathbb{Z}^d$ and with a specified dependence structure. We are first interested in estimation of $\beta_0$ and $\beta$ by using a smoothing spline approach as in [9], when $\mathrm{Cov}(\epsilon_{\mathbf{i}}, \epsilon_{\mathbf{j}}) = \sigma_\epsilon^2 \exp(-a\|\mathbf{i}-\mathbf{j}\|_2)$, where $a$ is some known positive constant, and $\|\cdot\|_2$ stands for the Euclidean norm on $\mathbb{Z}^d$ (see [11]). We then obtain the convergence rate for the resulting estimator $\widehat{\beta}$ of $\beta$ in a specified $L^2$ sense. Next, we are interested in prediction at a non-visited site, computed from the aforementioned estimators and we obtain an optimal convergence rate for the resulting predictor. The rest of the paper is organized as follows. Section 2 presents the spline estimator that will be used. Assumptions and main results, namely the convergence rates for the proposed estimator and predictor, are stated in Section 3. A simulation study is given in Section 4, whereas an application to ozone pollution forecasting at a nonvisited site is given in Section 5. The proofs of the main results are postponed to Section 6.

## 2 Smoothing spline estimation of a slope function

In this section, we give an estimator of $\beta$ in (1.1) by using an approach similar to that of [9]. Since this procedure of estimation does not take into account the nature of the dependence of the data, we obtain an estimator that has the same form than that of [9]. We assume that the random functions $X_{\mathbf{i}}$ are observed at $p$ equidistant points $t_1, \ldots, t_p \in I := [0,1]$ with $t_j = j/p$ for all $j = 1, \ldots, p$. By using the lexicographic order we rewrite the sample as $\{(X_{\mathbf{i}_i}, Y_{\mathbf{i}_i})\}_{1 \leqslant i \leqslant n^d}$. Put $\mathbf{Y} = (Y_{\mathbf{i}_1} - \overline{Y}, \ldots, Y_{\mathbf{i}_{n^d}} - \overline{Y})^{\mathrm{T}}$, where $u^{\mathrm{T}}$ denotes the transposed of $u$, and $\overline{Y} = (1/n^d)\sum_{\ell=1}^{n^d} Y_{\mathbf{i}_\ell}$, and consider the $n^d \times p$ matrix $\mathbf{X}$ with general term $X_{\mathbf{i}_i}(t_j) - \overline{X}(t_j)$ for $i = 1, \ldots, n^d$, $j = 1, \ldots, p$. Let $\mathbf{P}_m$ be a $p \times p$ projection matrix projecting into the linear space $\mathbf{F}_m$ defined by

$$\mathbf{F}_m := \left\{ \mathbf{z} = (z_1, \ldots, z_p)^{\mathrm{T}} \in \mathbb{R}^p \setminus z_j = \sum_{\ell=1}^{m} \theta_{\beta,\ell} t_j^{\ell-1},\ j = 1, \ldots, p,\ \theta_{\beta,\ell} \in \mathbb{R} \right\},$$

where the $\theta_{\beta,\ell}$ satisfy

$$\sum_{j=1}^{p} \left[ \beta(t_j) - \sum_{\ell=1}^{m} \theta_{\beta,\ell} t_j^{\ell-1} \right]^2 = \min_{\theta_1,\ldots,\theta_m} \sum_{j=1}^{p} \left[ \beta(t_j) - \sum_{\ell=1}^{m} \theta_{\ell} t_j^{\ell-1} \right]^2.$$

Let $\mathbf{D}(t) = (D_1(t), \ldots, D_p(t))^{\mathrm{T}}$ be a functional basis of the $p$-dimensional linear space $NS^m(t_1, \ldots, t_p)$ of natural splines of order $2m$ with knots at $t_1, \ldots, t_p$, and let $\widetilde{\mathbf{D}}$ be the $p \times p$ matrix with general term $D_i(t_j)$ for $i, j = 1, \ldots, p$. We consider two $p \times p$ matrices $\mathbf{B}_m$ and $\mathbf{A}_m$ defined by

$$\mathbf{B}_m = \widetilde{\mathbf{D}} \big( \widetilde{\mathbf{D}}^{\mathrm{T}} \widetilde{\mathbf{D}} \big)^{-1} \left[ \int_0^1 \mathbf{D}^{(m)}(t) \mathbf{D}^{(m)}(t)^{\mathrm{T}}\, \mathrm{d}t \right] \big( \widetilde{\mathbf{D}}^{\mathrm{T}} \widetilde{\mathbf{D}} \big)^{-1} \widetilde{\mathbf{D}}^{\mathrm{T}},$$

where $\mathbf{D}^{(m)}(t) = (D_1^{(m)}(t), \ldots, D_p^{(m)}(t))^{\mathrm{T}}$ with $D_j^{(m)}(t)$ standing for the $m$th derivative of the spline function $D_j(t)$ for $j = 1, \ldots, p$, and

$$\mathbf{A}_m = \mathbf{P}_m + p\mathbf{B}_m.$$

Then the estimator $\widehat{\beta}$ of $\beta$ is obtained by computing the least-squares linear estimator of the coefficients of $\widehat{\boldsymbol{\beta}}$ introduced in (2.2) with respect to the spline basis and is given by

$$\widehat{\beta}(t) = \mathbf{D}(t)^{\mathrm{T}} \big( \widetilde{\mathbf{D}}^{\mathrm{T}} \widetilde{\mathbf{D}} \big)^{-1} \widetilde{\mathbf{D}}^{\mathrm{T}} \widehat{\boldsymbol{\beta}} \tag{2.1}$$

with

$$\widehat{\boldsymbol{\beta}} = \frac{1}{n^d} \left( \frac{1}{n^d p} \mathbf{X}^{\mathrm{T}} \mathbf{X} + \rho \mathbf{A}_m \right)^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{Y}, \tag{2.2}$$

where $\rho > 0$ is a smoothing parameter. For estimating the intercept $\beta_0$, we take $\widehat{\beta}_0 = \overline{Y} - \langle \widehat{\beta}, \overline{X} \rangle$, where $\overline{X} = \{ \overline{X}(t),\, t \in [0,1] \}$ with $\overline{X}(t) = (1/n^d) \sum_{\ell=1}^{n^d} X_{\mathbf{i}_\ell}(t)$, $\langle \cdot, \cdot \rangle$ denotes the usual inner product of $L^2([0,1])$, and $\|\cdot\|$ is its associated norm.

## 3 Assumptions and main results

In this section, we first introduce the assumptions needed to obtain the main results of the paper and then theorems that give the rate of convergence of the estimator $\widehat{\beta}$ and also that of the prediction at a nonvisited site.

### 3.1 Assumptions

ASSUMPTION 1. $\beta$ is $m$ times differentiable, and $\beta^{(m)}$ belongs to $L^2([0,1])$.

ASSUMPTION 2. There exists $\kappa \in ]0,1[$ such that for every $\delta_1 > 0$, there exists a constant $C_1 > 0$ such that for all $(t,s) \in [0,1]^2$,

$$\mathrm{P}\big( |X(t) - X(s)| \leqslant C_1 |t-s|^{\kappa} \big) \geqslant 1 - \delta_1.$$

ASSUMPTION 3. For $C_2 \in \mathbb{R}_+^*$ and all $r \in \mathbb{N}^*$, there exist an $r$-dimensional linear subspace $\mathcal{L}_r$ of $L^2([0,1])$ and $q \in ]0,1[$ such that

$$\mathrm{E}\Big(\inf_{f\in\mathcal{L}_r}\sup_t \big|X(t)-f(t)\big|^2\Big) \leqslant C_2 r^{-2q}.$$

ASSUMPTION 4. For all $\ell = 1,2,\dots$ and all $\mathbf{i}_i, \mathbf{i}_j \in \mathcal{I}_\mathbf{n}$,

$$\mathrm{E}\big(\langle X_{\mathbf{i}_i}-\mathrm{E}(X),\, \zeta_\ell\rangle\langle X_{\mathbf{i}_j}-\mathrm{E}(X),\, \zeta_\ell\rangle\big) = \lambda_\ell \Psi_\ell\big(\|\mathbf{i}_i-\mathbf{i}_j\|_2\big) \quad \text{and} \quad \Psi_\ell(0)=1,$$

where $\sum_{\ell=1}^{+\infty}\lambda_\ell^{1/4}<\infty$, $\sum_{\ell\geqslant 1}\lambda_\ell\Psi_\ell(t)=g(t)$, and $g$ and $\Psi_\ell$ are known $\mathbb{R}_+$-valued decreasing functions such that $\sum_{t=1}^{\infty} t^{d-1}g(t)<\infty$. In addition, for any $(j,\ell)\in\mathbb{N}^*$such that $j\neq\ell$, we have

$$\mathrm{Var}\left(\frac{1}{n^d}\sum_{i=1}^{n^d}\langle X_{\mathbf{i}_i}-\mathrm{E}(X),\, \zeta_j\rangle\langle X_{\mathbf{i}_i}-\mathrm{E}(X),\, \zeta_\ell\rangle\right) \leqslant \frac{C_3}{n^d}\mathrm{E}\big(\langle X-\mathrm{E}(X),\, \zeta_j\rangle^2\big)\mathrm{E}\big(\langle X-\mathrm{E}(X),\, \zeta_\ell\rangle^2\big),$$

where $0<C_3<\infty$, and $\{\zeta_j\}_{j\in\mathbb{N}^*}$ is a complete orthonormal system of eigenfunctions of the operator $\Gamma$ from $L^2([0,1])$ to itself defined by

$$\Gamma u := \mathrm{E}\big(\langle u,\, X-\mathrm{E}(X)\rangle\big(X-\mathrm{E}(X)\big)\big) = \mathrm{E}\left(\int_0^1 \big(X(t)-\mathrm{E}(X)(t)\big)u(t)\,\mathrm{d}t\big(X-\mathrm{E}(X)\big)\right),$$

each $\zeta_j$ being associated with the $j$th largest eigenvalue $\lambda_j$, and $\mathrm{E}(X)=\{\mathrm{E}(X)(t),\ t\in[0,1]\}$.

Assumptions 1–4 are technical conditions similar to those considered in [9]. If $X$ is an almost surely $R_1$-times continuously differentiable random function that for all $R_3>0$, satisfies

$$\mathrm{E}\Big(\sup_{|t-s|\leqslant R_3}\big|X^{(R_1)}(t)-X^{(R_1)}(s)\big|^2\Big) \leqslant C' R_3^{2q} \tag{3.1}$$

for some $R_1=0,1,2,\dots$, $0<q<1$, $0<C'<\infty$. Then by the Jackson inequality we have the inequality

$$\inf_{f\in\mathcal{L}_r}\sum_{j=1}^{p}\big(X(t_j)-f(t_j)\big)^2 \leqslant C'' r^{-2R_1}\sup_{|t-s|\leqslant 1/r}\big|X^{(R_1)}(t)-X^{(R_1)}(s)\big|^2$$

with probability 1 for some $0<C''<+\infty$. Here $\mathcal{L}_r$ is the space of all polynomials of order $r$ on $[0,1]$. So if Assumption 2 is replaced by relation (3.1) with $R_1=0$, then Assumption 3 holds. Besides, even if $X$ is not smooth, Assumption 3 may yet be satisfied for a large value of $q$, for instance, by Brownian motions. Assumption 4 is satisfied when $\Lambda_{ij}=\langle X_{\mathbf{i}_i}-\mathrm{E}(X),\zeta_j\rangle$ and $\Lambda_{i\ell}=\langle X_{\mathbf{i}_i}-\mathrm{E}(X),\zeta_\ell\rangle$ are independent for $j\neq\ell$, and $\{\Lambda_{ij}, \mathbf{i}_i\in\mathcal{I}_\mathbf{n}\}$, $j\geqslant 1$ are stationary Gaussian random fields such that $\mathrm{E}(\Lambda_{ij}\Lambda_{kj})=\lambda_j\Psi_j(\|\mathbf{i}_i-\mathbf{i}_k\|_2)$ (see the assumptions of Proposition 8 in [18]). Two examples of correlation functions that satisfy the first condition of Assumption 4 are:

- The powered exponential model

$$\Psi_j(t) = \exp\left[-\left(\frac{t}{b_j}\right)^{b_0}\right] \quad \text{with } \sup_{j\geqslant 1} b_j<\infty,$$

where $0<b_0\leqslant 2$, and $b_j$, $j\geqslant 1$ are some positive constants;

- The Matern class

$$\Psi_j(t) = t^\nu K_\nu\left(\frac{t}{b_j}\right),$$

where $\nu$ is a positive constant, and the modified Bessel function $K_\nu$ decays monotonically and approximatively exponentially fast.

So, if $j > \ell$ (i.e., $\lambda_j^2 < \lambda_j\lambda_\ell$) and $i \neq k$, then from Example 8 of [18] we have

$$\begin{aligned}
\operatorname{Cov}(\Lambda_{ij}\Lambda_{i\ell}, \Lambda_{kj}\Lambda_{k\ell}) &= \lambda_j^2\Psi_j + \lambda_\ell^2\Psi_\ell + \lambda_j\lambda_\ell\frac{\Psi_j+\Psi_\ell}{2} - \left(\lambda_j^{3/2}\Psi_j + \lambda_\ell^{3/2}\Psi_\ell\right)\sqrt{\lambda_j+\lambda_\ell} \\
&< \lambda_\ell^2\left[\Psi_\ell - \left(\Psi_\ell + \left(\frac{\lambda_j}{\lambda_\ell}\right)^{3/2}\Psi_j\right)\left(1+\frac{\lambda_j}{\lambda_\ell}\right)^{1/2}\right] + \lambda_j\lambda_\ell\frac{3\Psi_j+\Psi_\ell}{2} \\
&< \frac{3\Psi_j+\Psi_\ell}{2}\mathrm{E}\left[\left\langle X-\mathrm{E}(X),\, \zeta_j\right\rangle^2\right]\mathrm{E}\left[\left\langle X-\mathrm{E}(X),\, \zeta_\ell\right\rangle^2\right],
\end{aligned}$$

where $\Psi_j := \Psi_j(\|\mathbf{i}_i - \mathbf{i}_k\|_2)$.

ASSUMPTION 5. $\|X_\mathbf{i}\| < M_2$ almost surely, where $M_2$ is some strictly positive constant. Moreover, the process $\{Z_\mathbf{i} = (X_\mathbf{i}, Y_\mathbf{i}),\ \mathbf{i} \in \mathbb{Z}^d\}$ is strongly polynomially mixing, that is, $\alpha_{1,\infty}(n) = O(n^{-\theta}) \to 0$ as $n \to +\infty$ for some $\theta > 0$, where

$$\alpha_{1,\infty}(n) = \sup\left\{\alpha\left(\sigma(Z_\mathbf{i}), F_H\right),\ \mathbf{i} \in \mathbb{Z}^d,\ H \subset \mathbb{Z}^d,\ \delta\left(H, \{\mathbf{i}\}\right) \geqslant n\right\}, \tag{3.2}$$

$\alpha$ being the $\alpha$-mixing coefficient given, for two sub-$\sigma$-algebras $\mathcal{U}$ and $\mathcal{V}$ of $\mathcal{A}$, by $\alpha(\mathcal{U},\mathcal{V}) = \sup\{|\mathrm{P}(A\cap B) - \mathrm{P}(A)\mathrm{P}(B)|,\ A \in \mathcal{U},\ B \in \mathcal{V}\}$, $F_H = \sigma(Z_\mathbf{i},\ \mathbf{i} \in H)$, and the distance $\delta$ is defined for any subsets $H_1$ and $H_2$ of $\mathbb{Z}^d$ by $\delta(H_1, H_2) = \min\{\|\mathbf{i}-\mathbf{j}\|_2,\ \mathbf{i} \in H_1,\ \mathbf{j} \in H_2\}$ with $\|\mathbf{i}-\mathbf{j}\|_2 = [\sum_{k=1}^d (i_k - j_k)^2]^{1/2}$ for $\mathbf{i}$, $\mathbf{j}$ in $\mathbb{Z}^d$.

The $\alpha$-mixing condition in Assumption 5 is a classical assumption. As pointed out in [18, p. 1540], the $\alpha$-mixing condition is suitable if we need more delicate results. Here it is needed to establish the optimal rate of the prediction error at a nonvisited site: it is the main difference between this work and [9], where for $d = 1$, it is assumed that $X_{n+1}$ is independent of $X_1, \dots, X_n$. The boundedness in Assumption 5 has already been made in some works (see, e.g., [20]). Here we need it for obtaining the following relation:

$$\mathrm{E}\left[\left\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\right\rangle^4\right] \leqslant 4M_2^2\mathrm{E}\left[\left\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\right\rangle^2\right] = 4M_2^2\lambda_\ell.$$

Details on the importance of this boundedness are given in the proof of Theorem 2. However, this boundedness can be relaxed by a moment assumption, that is,

$$\mathrm{E}\left[\|X\|^6\right] < M_2,$$

so by the Cauchy–Schwarz inequality we have

$$\begin{aligned}
\left\|\left\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\right\rangle^2\right\|_2^2 &= \mathrm{E}\left[\left\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\right\rangle^4\right] \\
&\leqslant \left\{\mathrm{E}\left[\left\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\right\rangle^2\right]\right\}^{1/2}\left\{\mathrm{E}\left[\left\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\right\rangle^6\right]\right\}^{1/2} \\
&\leqslant \lambda_\ell^{1/2}\left\{\mathrm{E}\left[\left\|X_{\mathbf{i}_0} - \mathrm{E}(X)\right\|^6\right]\right\}^{1/2} = O\left(\lambda_\ell^{1/2}\right).
\end{aligned}$$

We will then obtain the same result by assuming that $\sum_{\ell=1}^{+\infty} \lambda_\ell^{1/8} < +\infty$.

### 3.2 Main results

We consider the seminorm $\|\cdot\|_\Gamma$ defined by

$$\|u\|_\Gamma^2 := \langle \Gamma u, u\rangle, \quad u \in L^2\big([0,1]\big),$$

and the discretized empirical seminorm defined for any $\mathbf{u} \in \mathbb{R}^p$ as

$$\|\mathbf{u}\|_{\Gamma_{n,p}}^2 := \frac{1}{p}\mathbf{u}^{\mathrm{T}}\left(\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\mathbf{u}.$$

The following theorem gives a bound of the variance of the estimator. In this theorem, $\mathrm{E}_\epsilon$ refers to the conditional expectation given $X_{\mathbf{i}_1}, \dots, X_{\mathbf{i}_{n^d}}$.

**Theorem 1.** *Under Assumptions* 1 *and* 5*, for all* $\rho > n^{-2md}$*, if the eigenvalues* $\lambda_{x,1} \geqslant \lambda_{x,2} \geqslant \cdots \geqslant \lambda_{x,p} \geqslant 0$ *of* $1/(n^d p)\mathbf{X}^{\mathrm{T}}\mathbf{X}$ *satisfy* $\sum_{j=r+1}^{p} \lambda_{x,j} \leqslant C \cdot r^{-2q}$ *with* $C > 0$*,* $q > 0$*, and* $r := \lfloor \rho^{-1/(2m+2q+1)} \rfloor$*, then*

$$\mathrm{E}_\epsilon\big(\|\widehat{\boldsymbol{\beta}} - \mathrm{E}_\epsilon(\widehat{\boldsymbol{\beta}})\|_{\Gamma_{n,p}}^2\big) \leqslant \frac{c}{n^d}\big(m + \lfloor \rho^{-1/(2m+2q+1)} \rfloor (2 + C \cdot C_0)\big),$$

*where* $C_0 > 0$*,* $c > 0$*, and* $\lfloor x \rfloor$ *is the integer part of* $x$*.*

*Remark 1.* Let $\mathcal{P}_r$ be a $p \times p$ projection matrix projecting onto the $r$-dimensional subspace $\mathcal{L}_{r,p}$ defined as $\mathcal{L}_{r,p} = \{z \in \mathbb{R}^p \mid z = (f(t_1), \dots, f(t_p))^{\mathrm{T}},\ f \in \mathcal{L}_r\}$, and let $\mathbf{I}_p$ denote the $p \times p$ identity matrix. Since

$$\sum_{j=r+1}^{p} \lambda_{x,j} \leqslant \inf_{\mathcal{P}_r} \mathrm{Tr}\left((\mathbf{I}_p - \mathcal{P}_r)\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X}\right) = \frac{1}{n^d p}\sum_{\ell=1}^{n^d} \inf_{f \in \mathcal{L}_r} \sum_{j=1}^{p} \big(X_{\mathbf{i}_\ell}(t_j) - \overline{X}(t_j) - f(t_j)\big)^2,$$

by Assumption 3 for any $\delta_1 > 0$, there exists $C_{\delta_1} > 0$ such that

$$\mathrm{P}\left(\sum_{j=r+1}^{p} \lambda_{x,j} \leqslant C_{\delta_1} r^{-2q}\right) \geqslant 1 - \delta_1.$$

Besides, the difficulty encountered in the proof of this theorem is to bound $\sum_{j=1,\, j\neq i}^{n^d} \mathrm{Cov}(\epsilon_{\mathbf{i}_i}, \epsilon_{\mathbf{i}_j})$, which is not the case in the independent data case considered in [9]. For solving that, we consider an exponential covariance model. However, others spatial covariance models can be used, for instance, Matérn's spatial covariance model or Gaussian covariance model.

Using Theorem 1 and arguing as in [9], we obtain the following:

**Corollary 1.** *Let the assumptions of Theorem* 1 *together with Assumptions* 2–4 *be satisfied. Moreover, suppose that* $n^d p^{-2\kappa} = O(1)$*,* $\rho \to 0$*, and* $1/(n^d \rho) \to 0$ *as* $n, p \to \infty$*. Then we have*

$$\|\widehat{\beta} - \beta\|_\Gamma^2 = O_p\big(\rho + \big(n^d \rho^{1/(2m+2q+1)}\big)^{-1} + n^{-d(2q+1)/2}\big).$$

Next, we give a bound for prediction error at a nonvisited site $\mathbf{i}_0$ such that $\delta(\{\mathbf{i}_0\}, \mathcal{I}_{\mathbf{n}}) \geqslant \lfloor n^{4d/\theta} \rfloor$. It is sufficient to choose $\theta$ large for doing the prediction at any nonvisited site. We consider the prediction $\widehat{Y}_{\mathbf{i}_0}$ and the "theoretical" prediction $Y_{\mathbf{i}_0}^*$ at a nonvisited site $\mathbf{i}_0 \in \mathbb{Z}^d$ such that $(X_{\mathbf{i}_0}, Y_{\mathbf{i}_0})$ has the same distribution as $(X, Y)$. In fact,

$$\widehat{Y}_{\mathbf{i}_0} = \widehat{\beta}_0 + \langle \widehat{\beta}, X_{\mathbf{i}_0}\rangle \quad \text{and} \quad Y_{\mathbf{i}_0}^* = \beta_0 + \langle \beta, X_{\mathbf{i}_0}\rangle.$$

We give a bound of the prediction error between $\widehat{Y}_{\mathbf{i}_0}$ and $Y_{\mathbf{i}_0}^*$ in the following theorem.

**Theorem 2.** *Suppose that assumptions of Corollary* 1 *hold with* $\alpha_{1,\infty}(u) = O(u^{-\theta})$, $\theta > d$. *If* $2q \geqslant 1$, $\rho \sim n^{-d(2m+2q+1)/(2m+2q+2)}$, *and* $p$ *is chosen sufficiently large compared to* $n^d$, *then*

$$\mathrm{E}\big[(\widehat{Y}_{\mathbf{i}_0} - Y^*_{\mathbf{i}_0})^2\big] = O\big(n^{-d(2m+2q+1)/(2m+2q+2)}\big).$$

*Remark 2.*

(i) If $d = 1$, we obtain the same rate as that obtained in [9]. In this sense, this work is an extension from i.i.d. data to the dependent functional times series. The $O\big(n^{-d(2m+2q+1)/(2m+2q+2)}\big)$ is the optimal rate of convergence of the prediction error when the predictor is a spatially dependent random function and the response is a spatial scalar variable, whereas [3] obtained a rate that may be quite close to $O(\sqrt{\log n/n^d})$. This work can be easily extended to the spatial grid

$$\big\{\mathbf{i} = (i_1, \dots, i_d) \in \mathbb{Z}^d \;\big|\; 1 \leqslant i_k \leqslant n_k,\ k = 1, \dots, d\big\}.$$

(ii) If $Y^*_{\mathbf{i}_0}$ is replaced by $Y_{\mathbf{i}_0}$, then we obtain

$$\mathrm{E}\big[(\widehat{Y}_{\mathbf{i}_0} - Y_{\mathbf{i}_0})^2\big] = \mathrm{E}\big[(\widehat{Y}_{\mathbf{i}_0} - Y^*_{\mathbf{i}_0})^2\big] + \sigma_\epsilon^2 - 2\mathrm{E}\big[\epsilon_{\mathbf{i}_0}(\widehat{Y}_{\mathbf{i}_0} - Y^*_{\mathbf{i}_0})\big] \leqslant 2\big\{\mathrm{E}\big[(\widehat{Y}_{\mathbf{i}_0} - Y^*_{\mathbf{i}_0})^2\big] + \sigma_\epsilon^2\big\}.$$

## 4 A simulation study

In this section, we present the results of simulations made to evaluate the performances of the proposed method for prediction in model (1.1). We compute the prediction errors from simulated spatial data in $\mathbb{Z}^2$. Using the lexicographic order, we generate a sample $\{(X_{\mathbf{i}_\ell}, Y_{\mathbf{i}_\ell})\}_{1\leqslant\ell\leqslant(n+6)^2}$ as follows: we consider the 19th first elements $B_1, \dots, B_{19}$ of the Fourier basis. For $k = 1, \dots, 19$, we generate a vector $(\xi_{\mathbf{i}_1,k}, \dots, \xi_{\mathbf{i}_{(n+6)^2},k})^{\mathrm{T}}$ using the R function $\mathrm{rtmvnorm}$, which uses the singular value decomposition to simulate truncated Gaussian vectors with mean 0 and the $(n+6)^2 \times (n+6)^2$ covariance matrix $\Sigma^1$ with general term $\Sigma^1_{ij} = \mathrm{Cov}(\xi_{\mathbf{i}_i,k}, \xi_{\mathbf{i}_j,k}) = \lambda_k \exp(-a\|\mathbf{i}_i - \mathbf{i}_j\|_2)$, where $a = 1, 200$, $\lambda_k = 1$ for $k = 1, \dots, 19$, and $\lambda_k = 0$ for $k \geqslant 20$. The truncation limit is taken as the square $[0,1]^{(n+6)^2}$. Then the components of this vector are $\alpha$-mixing dependent (see [7]). Notice that when $a = 200$, there is approximately no spatial autocorrelation between the components. The process $\{\xi_{\mathbf{i}_i,k}, \mathbf{i}_i \in \mathbb{Z}^2\}$ is said to be strongly correlated when $a = 1$. For $\ell = 1, \dots, (n+6)^2$, we take $X_{\mathbf{i}_\ell}(t) = \sum_{k=1}^{19} \xi_{\mathbf{i}_\ell,k} B_k(t)$, where $\{B_k\}$ is defined by $B_k(t) = \sqrt{2}\sin((t - 1/2)k\pi/2)$ when $k$ is even and $B_k(t) = \sqrt{2}\cos((t-1/2)(k-1)\pi/2)$ when $k$ is odd. Considering 366 equispaced points into $[0,1]$, we compute each $Y_{\mathbf{i}_\ell}$ by approximating the integral in the spatial functional linear regression model (SFLR) defined in (1.1) by the rectangular method. That gives $Y_{\mathbf{i}_\ell} = (1/365)\sum_{j=1}^{365} \beta(t_j) X_{\mathbf{i}_\ell}(t_j) + \epsilon_{\mathbf{i}_\ell}$, where $t_j = j/365$, $j = 1, \dots, p = 365$, the vector $(\epsilon_{\mathbf{i}_1}, \dots, \epsilon_{\mathbf{i}_{n^2}})^{\mathrm{T}}$ is generated by using the R function $\mathrm{mvrnorm}$, which simulates a vector from a normal distribution $\mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \Sigma^1)$, and $\beta(t) = [\sin(2\pi t^3)]^3$ is the true function. Our sample is split into two following subsamples:

- a sample $(X_{\mathbf{i}_\ell}, Y_{\mathbf{i}_\ell})_{\mathbf{i}_\ell}$, $\mathbf{i}_\ell \in \{1, \dots, n\}^2$, $\ell = 1, \dots, n^2$, is used to compute the estimator $\widehat{\beta}$ of $\beta$;
- a sample $(X_{(i,j)}, Y_{(i,j)})_{(i,j)}$, $(i,j) \in \{n+1, \dots, n+6\} \times \{1, \dots, n\}$, is used to compute the prediction error between $\widehat{Y}_{(n+k,j)}$ and $Y_{(n+k,j)}$ at the nonvisited sites $(n+k, j)$, $k = 1, \dots, 6$ and $j = 1, \dots, n$.

The estimator $\widehat{\beta}$ of $\beta$ in model (1.1) is computed from formulas (2.1) and (2.2) with $m = 2$ (cubic smoothing splines), and $\rho$ is obtained from the generalized cross validation given by

$$\mathrm{GCV}(\rho) = \frac{\sum_{\ell=1}^{n^2} (\widehat{Y}_{\mathbf{i}_\ell} - Y_{\mathbf{i}_\ell})^2}{n^2(1 - n^{-2}Tr(H_\rho))},$$

where

$$H_\rho = \frac{1}{n^2}\mathbf{X}\bigg(\frac{1}{n^2 p}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \rho\mathbf{A}_m\bigg)\mathbf{X}^{\mathrm{T}}.$$

**Table 1.** The mean and standard deviation of MSEP based on 100 replications.

| | | | $a = 1$ | | $a = 200$ | |
|---|---|---|---|---|---|---|
| $\sigma_\epsilon$ | $n^2$ | Model | $m$ | $sd$ | $m$ | $sd$ |
| 0.1 | 25 | SFLR | 0.0349 | 0.0035 | 0.0267 | 0.0027 |
| | | SFLRD | 0.0303 | 0.0030 | 0.0216 | 0.0022 |
| | 100 | SFLR | 0.0281 | 0.0028 | 0.0280 | 0.0028 |
| | | SFLRD | 0.0122 | 0.0012 | 0.0110 | 0.0011 |
| | 225 | SFLR | 0.0240 | 0.0024 | 0.0280 | 0.0028 |
| | | SFLRD | 0.0113 | 0.0011 | 0.0109 | 0.0011 |

We assess performance of our method through calculation of the mean squared error of predictions (MSEP):

$$\text{MSEP} = \frac{1}{6n} \sum_{j=1}^{n} \sum_{k=1}^{6} (\widehat{Y}_{(n+k,j)} - Y_{(n+k,j)})^2,$$

where $\widehat{Y}_{(n+k,j)}$ is the prediction of $Y_{(n+k,j)}$ at the nonvisited sites $(n+k, j)$, $k = 1, \dots, 6$ and $j = 1, \dots, n$. We take $n = 5, 10, 15$, $\sigma_\epsilon = 0.1$, and $a = 1, 200$ over 100 replications, and we compare the prediction errors obtained by the proposed approach in this paper with that obtained from the spatial functional linear regression model with derivatives (SFLRD) studied in [3]. Especially, we write the SFLRD as

$$Y_{\mathbf{i}_\ell} = \int_0^1 \gamma_1(t) X_{\mathbf{i}_\ell}(t)\, \mathrm{d}t + \int_0^1 \gamma_2(t) X'_{\mathbf{i}_\ell}(t)\, \mathrm{d}t + \epsilon_{\mathbf{i}_\ell},$$

where $\gamma_1$ and $\gamma_2$ are two functions to estimate, $X'_{\mathbf{i}_\ell}$ stands for the first derivative of $X_{\mathbf{i}_\ell}$ and is computed from the function “fdata.deriv” of the $R$ fda package. The estimate ($\widehat{\gamma}_1$, $\widehat{\gamma}_2$) of the pair ($\gamma_1$,$\gamma_2$) is obtained from the moment method combined with the regularization sequences approach allowing us to inverse empirical covariance operators (see [3]). These regularization sequences $w$ and $\phi$ are obtained from the cross-validation based on the evaluation of the mean standard error of prediction (CVMSEP):

$$\text{CVMSEP}(w, \phi) = \frac{1}{n^2} \sum_{\ell=1}^{n^2} \left( Y_{\mathbf{i}_\ell} - \widetilde{Y}_{\mathbf{i}_\ell}^{(-\ell)}(w, \phi) \right)^2,$$

where $\widetilde{Y}_{\mathbf{i}_\ell}^{(-\ell)}(w, \phi)$ denotes the prediction of $Y_{\mathbf{i}_\ell}$ for a given $(w, \phi)$, calculated without the $\ell$th observation $(X_{\mathbf{i}_\ell}, Y_{\mathbf{i}_\ell})$. For each of both these methods, the values of the regularization sequences or penalty parameter are given for any fixed $\sigma_\epsilon$, $n^2$, and $a$. The mean $m$ and the standard deviation $sd$ of MSEP are calculated over 100 replications with different values of $\sigma_\epsilon$, $n^2$, and $a$. The obtained results are presented in Table 1.

Table 1 reveals that as the sample size increases, the prediction errors from both models decrease. This means that our prediction method can well fit to the spatial functional linear regression model. Moreover, in both these models with fixed $n$ (i.e., $n^2 = 100, 225$) and $\sigma_\epsilon = 0.1$, the prediction errors of strongly autocorrelated processes (i.e., $a = 1$) are similar to those approximately nonautocorrelated (i.e., $a = 200$). Finally, for all sample sizes ($n^2 = 25, 100, 225$), the prediction errors are similar for both models.

## 5 Application to ozone pollution forecasting at the nonvisited sites

In this section, we apply our methodology to predict the level of ozone pollution at the nonvisited sites of California state. For that, we consider the data available on internet site `https://www.epa.gov/outdoor-air-quality-data`.

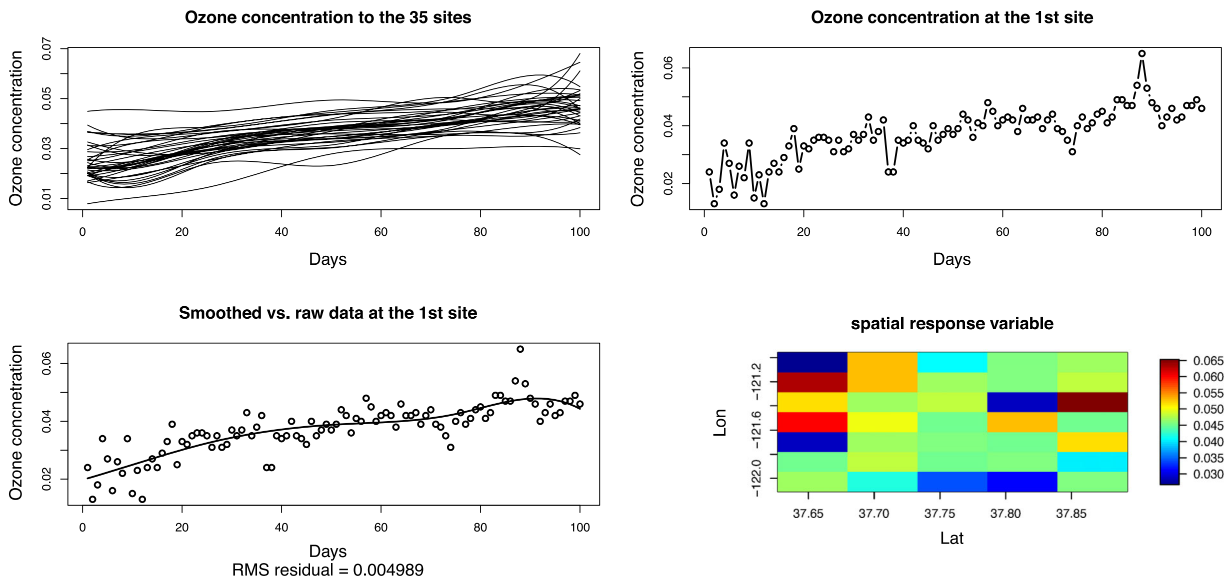


**Figure 1.** Ozone pollution data: daily ozone concentration at the 35 stations (left and top), daily ozone concentration at the first station (right and top), smoothed and raw of daily ozone concentration data at the first station (left and down), ozone pollution level (right and down).

The explanatory functional variables

$$\{X_{s_i}(t),\ t = 1, \dots, 100,\ s_i = (Latitude, Longitude)_i,\ i = 1, \dots, 51\}$$

correspond to the measurements of ozone concentration obtained for the $p = 100$ first days, from 1 January 2021 to 12 April 2021 on each of $n^2 = 51$ sites. The response variables

$$\{Y_{s_i},\ s_i = (Latitude, Longitude)_i,\ i = 1, \dots, 35\}$$

correspond to the measurements of ozone concentration obtained on 13 April 2021 on each of 35 first stations. For evaluating the performances of our method, we yet compare it to that of SFRD defined in the previous section. So from both methodologies we predict

$$\{Y_{s_i},\ s_i = (Latitude, Longitude)_i,\ i = 36, \dots, 51\},$$

which correspond to the measurements of ozone concentration obtained on 13 April 2021 on 16 other sites assumed nonvisited at this date. In what follows, we illustrate our data by graphics (see Fig. 1).

In these graphics, we see that these data correspond to our study. In what follows, we present the results of our study.

Both graphics in Fig. 2 present a small difference confirmed by computation (see Table 2) of the prediction error (PE) given by

$$\mathrm{PE} = \sqrt{\sum_{i=36}^{51} (Y_{s_i} - \widehat{Y}_{s_i})^2}.$$

We see a very minor advantage for prediction obtained from the SFLR model studied in this paper.

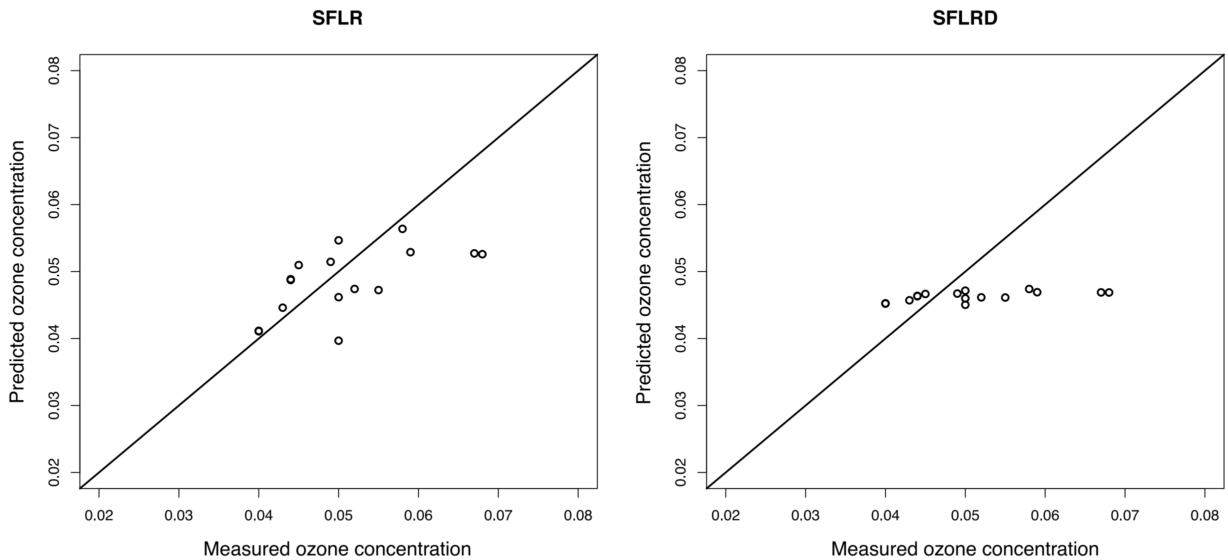


**Figure 2.** Predicted values of ozone concentration, from SLFR (left) and SLFRD (right), versus the measured values.

**Table 2.** Prediction error computed from both methods.

| | SFLR ($nbasis = 7$, $\rho = 0.09$) | SFLRD ($w = 0.01$, $\phi = 0.11$) |
|---|---|---|
| Prediction error (PE) | 0.0282 | 0.038 |

## 6 Conclusions

In this paper, we proposed to study asymptotic properties of a prediction at nonvisited site computed from a smoothing spline estimator of the slope function in a spatial functional linear regression model, where a scalar response is related to a square-integrable spatial functional process. The originality of the proposed method is that we consider spatially dependent data. We established the optimal convergence rate of the estimation and prediction errors when the considered processes are stationary isotropic. The simulation study and application to ozone pollution revealed that the proposed prediction fits well with the spatial functional linear regression model. Besides, the SFLR method produces equivalent predictions with the SFLRD method for the large sample sizes. However, the presented methodology in this paper has a minor advantage over the SFLRD method, since its theoretical rate of convergence is better than that given in [3]. We can see the proposed methodology as a good alternative to [9] when available data are spatially dependent and support model (1.1).

## 7 Proofs

### 7.1 Preliminary lemmas

Letting

$$\mathcal{M} = \left(\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \rho\mathbf{A}_m\right)^{-1}\left(\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X}\right), \tag{7.1}$$

we have the following:

**Lemma 1.** *Let Assumption* 1 *be satisfied. If* $\rho > n^{-2md}$ *and the eigenvalues* $\lambda_{x,1} \geqslant \lambda_{x,2} \geqslant \cdots \geqslant \lambda_{x,p} \geqslant 0$ *of* $1/(n^d p)\mathbf{X}^{\mathrm{T}}\mathbf{X}$ *satisfy* $\sum_{j=r+1}^{p} \lambda_{x,j} \leqslant Cr^{-2q}$ *with* $C > 0$, $q > 0$, *and* $r := \lfloor \rho^{-1/(2m+2q+1)} \rfloor$, *then*

$$\mathrm{Tr}(\mathcal{M}^2) \leqslant \mathrm{Tr}(\mathcal{M}) \leqslant \bigl(m + \lfloor \rho^{-1/(2m+2q+1)} \rfloor (2 + C \cdot C_0)\bigr),$$

*where* $\mathrm{Tr}(\cdot)$ *stands for the matrix trace.*

*Proof.* Since $\mathbf{A}_m$ is a symmetric nonnegative matrix, its has a square root, denoted by $\mathbf{A}_m^{1/2}$, which is also a symmetric nonnegative matrix. Denoting by $\mathbf{A}_m^{-1/2}$ the inverse of $\mathbf{A}_m^{1/2}$ and by $\mathbf{I}_p$ the $p \times p$ identity matrix, we have

$$\mathcal{M} = \mathbf{A}_m^{-1/2} \left( \frac{1}{n^d p} \mathbf{A}_m^{-1/2} \mathbf{X}^{\mathrm{T}} \mathbf{X} \mathbf{A}_m^{-1/2} + \rho \mathbf{I}_p \right)^{-1} \left( \frac{1}{n^d p} \mathbf{A}_m^{-1/2} \mathbf{X}^{\mathrm{T}} \mathbf{X} \right).$$

Then from the spectral decomposition

$$\frac{1}{n^d p} \mathbf{A}_m^{-1/2} \mathbf{X}^{\mathrm{T}} \mathbf{X} \mathbf{A}_m^{-1/2} = \sum_{\ell=1}^{p} \mu_\ell \, u_\ell u_\ell^{\mathrm{T}},$$

where $\mu_\ell$ are the nonnegative eigenvalues, and $\{u_\ell\}_{1 \leqslant \ell \leqslant p}$ is an orthonormal basis of $\mathbb{R}^p$ consisting of eigenvectors, it follows that

$$\mathcal{M} = \sum_{\ell=1}^{p} \sum_{k=1}^{p} \frac{\mu_k}{\mu_\ell + \rho} \, \mathbf{A}_m^{-1/2} \, u_\ell u_\ell^T u_k u_k^{\mathrm{T}} \mathbf{A}_m^{1/2} = \sum_{\ell=1}^{p} \frac{\mu_\ell}{\mu_\ell + \rho} \, \mathbf{A}_m^{-1/2} \, u_\ell u_\ell^{\mathrm{T}} \mathbf{A}_m^{1/2}.$$

Therefore, since $\mathrm{Tr}(\mathbf{A}_m^{-1/2} u_\ell u_\ell^{\mathrm{T}} \mathbf{A}_m^{1/2}) = \mathrm{Tr}(u_\ell^{\mathrm{T}} \mathbf{A}_m^{1/2} \mathbf{A}_m^{-1/2} u_\ell) = \mathrm{Tr}(u_\ell^{\mathrm{T}} \, u_\ell) = 1$, we deduce that $\mathrm{Tr}(\mathcal{M}) = \sum_{\ell=1}^{p} \mu_\ell/(\mu_\ell + \rho)$. Finally,

$$\begin{aligned}\mathrm{Tr}(\mathcal{M}^2) &= \mathrm{Tr} \left( \sum_{\ell=1}^{p} \sum_{k=1}^{p} \left( \frac{\mu_\ell}{\mu_\ell + \rho} \right) \left( \frac{\mu_k}{\mu_k + \rho} \right) \mathbf{A}_m^{-1/2} \, u_\ell u_\ell^T u_k u_k^{\mathrm{T}} \mathbf{A}_m^{1/2} \right) \\ &= \sum_{\ell=1}^{p} \left( \frac{\mu_\ell}{\mu_\ell + \rho} \right)^2 \leqslant \sum_{\ell=1}^{p} \frac{\mu_\ell}{\mu_\ell + \rho} = \mathrm{Tr}(\mathcal{M}).\end{aligned}$$

On the other hand, from a reasoning similar to the proof of Theorem 1 in [9] (see from (6.1) to the end of the proof) we get $\mathrm{Tr}(\mathcal{M}) \leqslant (m + \lfloor \rho^{-1/(2m+2q+1)} \rfloor (2 + C \cdot C_0))$, ending the proof. □

The following lemma will be useful for proving Theorem 2. Its proof is is similar to that of Lemma 1 in [10, p. 871].

**Lemma 2.** *Assume that* (3.2) *holds. Let* $\mathcal{L}_r(\mathcal{A})$ *be the class of* $\mathcal{A}$*-measurable random functions* $X$ *satisfying* $\|X\|_r = (\mathrm{E}(|X|^r))^{1/r} < +\infty$. *Let* $r,\ s,\ h$ *be positive constants such that* $r^{-1} + s^{-1} + h^{-1} = 1$. *Then for any* $X \in \mathcal{L}_r(\mathcal{B}(S))$ *and* $Y \in \mathcal{L}_s(\mathcal{B}(S'))$, *we have*

$$\bigl|\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\bigr| \leqslant 10 \|X\|_r \|Y\|_s \bigl\{\alpha_{1,\infty}\bigl(\delta(S, S')\bigr)\bigr\}^{1/h}.$$

The following three lemmas will also be useful in the proofs of Theorems 1 and 2.

**Lemma 3.** *Let the assumptions of Theorem* 1 *together with Assumptions* 2 *and* 3 *be satisfied. If* $n^d p^{-2\kappa} = O(1)$*,* $\rho \to 0$*, and* $1/(n^d\rho) \to 0$*, as* $n, p \to +\infty$*, then we have*

$$\|\widehat{\beta} - \beta\|^2 = O_p(1) \quad \textit{and} \quad \|\widehat{\beta}\|^2 = O_p(1).$$

*Proof.* The proof of this lemma is done in two steps:

(i) $(1/p)\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\beta}} = O_p(1)$;

(ii) $|\|\widehat{\beta}\|^2 - (1/p)\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}| = O_p(1/p)$.

The proof of (ii) is very similar to that of Theorem 2 in [9, p. 59], and we have

$$\left|\|\widehat{\beta}\|^2 - \frac{1}{p}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right| = O_p\left(\frac{1}{p}\left(1 + \frac{p^{-2\kappa}}{\rho} + \frac{1}{n^d \rho^{(2m+2q+2)/(2m+2q+1)}}\right)\right).$$

It remains to prove (i). We have

$$\frac{1}{p}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\beta}} \leqslant A + B + C,$$

where

$$A = \frac{3}{p}\boldsymbol{\beta}^{\mathrm{T}}\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X}\left(\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \rho\mathbf{A}_m\right)^{-2}\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta},$$

$$B = \frac{3}{n^d p}\mathbf{d}^{\mathrm{T}}\mathbf{X}\left(\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \rho\mathbf{A}_m\right)^{-2}\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{d},$$

$$C = \frac{3}{n^d p}\boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{X}\left(\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \rho\mathbf{A}_m\right)^{-2}\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\boldsymbol{\epsilon},$$

where

$$\mathbf{d} = (d_1 - \overline{d}, \ldots, d_{n^2} - \overline{d})^{\mathrm{T}}$$

with

$$d_\ell = \int_0^1 \beta(t)X_{\mathbf{i}_\ell}(t)\,\mathrm{d}t - \frac{1}{p}\sum_{j=1}^{p}\beta(t_j)X_{\mathbf{i}_\ell}(t_j), \qquad \overline{d} = \int_0^1 \beta(t)\overline{X}(t)\,\mathrm{d}t - \frac{1}{p}\sum_{j=1}^{p}\beta(t_j)\overline{X}(t_j),$$

and

$$\boldsymbol{\epsilon} = (\epsilon_{\mathbf{i}_1} - \overline{\epsilon},\ \ldots,\ \epsilon_{\mathbf{i}_{n^d}} - \overline{\epsilon})^{\mathrm{T}} \quad \text{with } \overline{\epsilon} = \frac{1}{n^d}\sum_{\ell=1}^{n^d}\epsilon_{\mathbf{i}_\ell}.$$

By a reasoning similar to that in Theorem 2 in [9, p. 24] we get

$$A \leqslant \frac{3}{p}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta} = O(1) \quad \text{and} \quad B = O_p\left(\frac{p^{-2\kappa}}{\rho}\right).$$

Finally, putting

$$\varXi = \frac{1}{n^d p}\mathbf{X}\left(\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \rho\mathbf{A}_m\right)^{-2}\mathbf{X}^{\mathrm{T}},$$

we have

$$\mathrm{E}_\epsilon(C) = \frac{3}{n^d}\mathrm{E}_\epsilon\big(\boldsymbol{\epsilon}^{\mathrm{T}} \Xi \boldsymbol{\epsilon}\big) = \frac{3}{n^d}\sum_{i=1} \Xi_{ii}\mathrm{E}\big(\tau_i^2\big) + \frac{3}{n^d}\sum_{i=1}^{n^d}\sum_{\substack{j=1\\ j\neq i}}^{n^d} \Xi_{ij}\mathrm{E}(\tau_i\tau_j),$$

where $\tau_i = \epsilon_{\mathbf{i}_i} - \overline{\epsilon}$. From (7.2) we have

$$\frac{3}{n^d}\sum_{i=1} \Xi_{ii}\mathrm{E}\big(\tau_i^2\big) \leqslant \frac{6K}{n^d}\operatorname{Tr}(\Xi) \leqslant \frac{6K}{n^d}\operatorname{Tr}\big[(\rho\mathbf{A}_m)^{-1}\big] = O\bigg(\frac{1}{n^d\rho}\bigg).$$

Note that $\Xi = \Sigma\Sigma^{\mathrm{T}}$, where

$$\Sigma = \frac{1}{\sqrt{n^d p}}\mathbf{X}\bigg(\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \rho\mathbf{A}_m\bigg)^{-1}.$$

Then

$$|\Xi_{ij}| = \Bigg|\sum_{k=1}^{n^d} \Sigma_{ik}\Sigma_{jk}\Bigg| \leqslant \frac{1}{2}\sum_{k=1}^{n^d}\big(\Sigma_{ik}^2 + \Sigma_{jk}^2\big) = \frac{1}{2}(\Xi_{ii} + \Xi_{jj}),$$

and using (7.4) we obtain

$$\begin{aligned}\Bigg|\frac{3}{n^d}\sum_{i=1}^{n^d}\sum_{\substack{j=1\\ j\neq i}}^{n^d} \Xi_{ij}\mathrm{E}(\tau_i\tau_j)\Bigg| &\leqslant \frac{3}{n^d}\sum_{i=1}^{n^d}\sum_{\substack{j=1\\ j\neq i}}^{n^d} |\Xi_{ij}|\big|\mathrm{E}(\tau_i\tau_j)\big| \leqslant \frac{3}{n^d}\sum_{i=1}^{n^d} \Xi_{ii}\sum_{\substack{j=1\\ j\neq i}}^{n^d}\big|\mathrm{E}(\tau_i\tau_j)\big| \\ &\leqslant \frac{3K_2}{n^d}\operatorname{Tr}(\Xi) \leqslant \frac{3K_2}{n^d}\operatorname{Tr}\big[(\rho\mathbf{A}_m)^{-1}\big] = O\bigg(\frac{1}{n^d\rho}\bigg).\end{aligned}$$

Therefore

$$\frac{1}{p}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\beta}} = O_p\bigg(1 + \frac{p^{-2\kappa}}{\rho} + \frac{1}{n^d\rho}\bigg).$$

So from (i) and (ii) we deduce that $\|\widehat{\beta}\|^2 = O_p(1)$ and $\|\widehat{\beta} - \beta\|^2 = O_p(1)$. □

**Lemma 4.** *We have*

$$\mathrm{E}\Bigg[\Bigg(\frac{1}{n^d}\sum_{\ell=1}^{n^d}\epsilon_{\mathbf{i}_\ell}\Bigg)^2\Bigg] = O\bigg(\frac{1}{n^d}\bigg).$$

*Proof.* We have

$$\begin{aligned}\mathrm{E}\Bigg[\Bigg(\frac{1}{n^d}\sum_{\ell=1}^{n^d}\epsilon_{\mathbf{i}_\ell}\Bigg)^2\Bigg] &= \frac{1}{n^{2d}}\sum_{\ell=1}^{n^d}\mathrm{E}\big(\epsilon_{\mathbf{i}_\ell}^2\big) + \frac{1}{n^{2d}}\sum_{i=1}^{n^d}\sum_{\substack{j=1\\ j\neq i}}^{n^d}\operatorname{Cov}(\epsilon_{\mathbf{i}_i}, \epsilon_{\mathbf{i}_j}) \\ &= \frac{\sigma_\epsilon^2}{n^d} + \frac{\sigma_\epsilon^2}{n^{2d}}\sum_{i=1}^{n^d}\sum_{\substack{j=1\\ j\neq i}}^{n^d}\exp\big(-a\|\mathbf{i}_i - \mathbf{i}_j\|_2\big) = O\bigg(\frac{1}{n^d}\bigg).\end{aligned}$$ □

**Lemma 5.** *Under Assumptions* 4 *and* 5*, we have*

$$\mathrm{E}\big[\|\overline{X}-\mathrm{E}(\overline{X})\|^2\big]=O\bigg(\frac{1}{n^d}\bigg).$$

*Proof.* From Assumptions 4 and 5 we have

$$\begin{aligned}\mathrm{E}\big[\|\overline{X}-\mathrm{E}(\overline{X})\|^2\big]&=\mathrm{E}\Bigg[\frac{1}{n^{2d}}\sum_{i=1}^{n^d}\|X_{\mathbf{i}_i}-\mathrm{E}(X)\|^2+\frac{1}{n^{2d}}\sum_{\substack{i,j=1\\ i\neq j}}^{n^d}\big\langle X_{\mathbf{i}_i}-\mathrm{E}(X),\,X_{\mathbf{i}_j}-\mathrm{E}(X)\big\rangle\Bigg]\\&=O\bigg(\frac{1}{n^d}\bigg)+\frac{1}{n^{2d}}\sum_{\substack{i,j=1\\ i\neq j}}^{n^d}\sum_{k\geqslant 1}\mathrm{E}\big[\big\langle X_{\mathbf{i}_i}-\mathrm{E}(X),\,\zeta_k\big\rangle\big\langle X_{\mathbf{i}_j}-\mathrm{E}(X),\,\zeta_k\big\rangle\big]\\&=O\bigg(\frac{1}{n^d}\bigg)+\frac{1}{n^{2d}}\sum_{\substack{i,j=1\\ i\neq j}}^{n^d}\sum_{k\geqslant 1}\lambda_k\Psi_k\big(\|\mathbf{i}_i-\mathbf{i}_j\|\big),\end{aligned}$$

and thus

$$\mathrm{E}\big[\|\overline{X}-\mathrm{E}(\overline{X})\|^2\big]\leqslant\frac{K_1}{n^d}+\frac{K_2}{n^d}\sum_{t=1}^{+\infty}t^{d-1}\sum_{k\geqslant 1}\lambda_k\Psi_k(t)=O\bigg(\frac{1}{n^d}\bigg),$$

where $K_1$ and $K_2$ are some positive constants. □

### 7.2 Proof of Theorem 1

Putting

$$\Theta=\mathbf{X}\bigg(\frac{1}{n^dp}\mathbf{X}^{\mathrm{T}}\mathbf{X}+\rho\mathbf{A}_m\bigg)^{-1}\bigg(\frac{1}{n^dp}\mathbf{X}^{\mathrm{T}}\mathbf{X}\bigg)\bigg(\frac{1}{n^dp}\mathbf{X}^{\mathrm{T}}\mathbf{X}+\rho\mathbf{A}_m\bigg)^{-1}\mathbf{X}^{\mathrm{T}},$$

we have

$$\begin{aligned}\mathrm{E}_\epsilon\big(\|\widehat{\boldsymbol\beta}-\mathrm{E}_\epsilon(\widehat{\boldsymbol\beta})\|^2_{\Gamma_{n,p}}\big)&=\frac{1}{p}\mathrm{E}_\epsilon\Bigg(\frac{1}{n^{2d}}\boldsymbol\epsilon^{\mathrm{T}}\mathbf{X}\bigg(\frac{1}{n^dp}\mathbf{X}^{\mathrm{T}}\mathbf{X}+\rho\mathbf{A}_m\bigg)^{-1}\\&\quad\times\bigg(\frac{1}{n^dp}\mathbf{X}^{\mathrm{T}}\mathbf{X}\bigg)\bigg(\frac{1}{n^dp}\mathbf{X}^{\mathrm{T}}\mathbf{X}+\rho\mathbf{A}_m\bigg)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol\epsilon\Bigg)\\&=\frac{1}{n^{2d}p}\Bigg(\sum_{i=1}^{n^d}\Theta_{ii}\mathrm{E}\big(\tau_i^2\big)+\sum_{i=1}^{n^d}\sum_{\substack{j=1\\ j\neq i}}^{n^d}\Theta_{ij}\mathrm{E}(\tau_i\tau_j)\Bigg),\end{aligned}$$

where $\tau_i=\epsilon_{\mathbf{i}_i}-\overline{\epsilon}$ with $\overline{\epsilon}=n^{-d}\sum_{j=1}^{n^d}\epsilon_{\mathbf{i}_j}$. We deduce from the stationarity and Lemma 4 that

$$\mathrm{E}\big(\tau_i^2\big)\leqslant 2\big[\sigma_\epsilon^2+\mathrm{E}\big(\overline{\epsilon}^2\big)\big]\leqslant K\bigg(1+\frac{1}{n^d}\bigg),\tag{7.2}$$

where $K$ is a positive constant. Clearly, $\sum_{i=1}^{n^d} \Theta_{ii} = \mathrm{Tr}(\Theta) = n^d p \,\mathrm{Tr}(\mathcal{M}^2)$, where $\mathcal{M}$ is defined in (7.1). Then from Lemma 1 it follows that

$$\sum_{i=1}^{n^d} \Theta_{ii} \leqslant n^d p\bigl(m + \rho^{-1/(2m+2q+1)}(2 + C \cdot C_0)\bigr). \tag{7.3}$$

Then from (7.2) and (7.3) we deduce that

$$\frac{1}{n^{2d}p} \sum_{i=1}^{n^d} \Theta_{ii} \mathrm{E}\bigl(\tau_i^2\bigr) \leqslant \frac{K}{n^d}\left(1 + \frac{1}{n^d}\right)\bigl(m + \rho^{-1/(2m+2q+1)}(2 + C \cdot C_0)\bigr),$$

On the other hand,

$$\mathrm{E}(\tau_i\tau_j) = \mathrm{E}(\epsilon_{\mathbf{i}_i}\epsilon_{\mathbf{i}_j}) - \frac{2}{n^d}\sigma_\epsilon^2 - \frac{1}{n^d}\sum_{\substack{k=1\\k\neq i}}^{n^d} \mathrm{E}(\epsilon_{\mathbf{i}_i}\epsilon_{\mathbf{i}_k}) - \frac{1}{n^d}\sum_{\substack{k=1\\k\neq j}}^{n^d} \mathrm{E}(\epsilon_{\mathbf{i}_j}\epsilon_{\mathbf{i}_k}) + \mathrm{E}\bigl(\bar{\epsilon}^2\bigr).$$

Then from Lemma 4 we obtain $|\mathrm{E}(\tau_i\tau_j)| \leqslant |\mathrm{E}(\epsilon_{\mathbf{i}_i}\epsilon_{\mathbf{i}_j})| + K_1/n^d$ and

$$\sum_{\substack{j=1\\j\neq i}}^{n^d} \bigl|\mathrm{E}(\tau_i\tau_j)\bigr| \leqslant \sum_{\substack{j=1\\j\neq i}}^{n^d} \bigl|\mathrm{E}(\epsilon_{\mathbf{i}_i}\epsilon_{\mathbf{i}_j})\bigr| + K_1 \leqslant K_2, \tag{7.4}$$

where $K_1$ and $K_2$ are positive constants. Note that $\Theta = \mathbf{F}^2$, where

$$\mathbf{F} = \bigl(n^d p\bigr)^{-1/2}\mathbf{X}\left(\frac{1}{n^d p}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \rho\mathbf{A}_m\right)^{-1}\mathbf{X}^{\mathrm{T}};$$

then

$$|\Theta_{ij}| = \left|\sum_{k=1}^{n^d} \mathbf{F}_{ik}\mathbf{F}_{kj}\right| \leqslant \frac{1}{2}\sum_{k=1}^{n^d}\bigl(\mathbf{F}_{ik}^2 + \mathbf{F}_{kj}^2\bigr) = \frac{1}{2}(\Theta_{ii} + \Theta_{jj}),$$

and putting

$$S = \frac{1}{n^{2d}p}\sum_{i=1}^{n^d}\sum_{\substack{j=1\\j\neq i}}^{n^d} \Theta_{ij}\mathrm{E}(\tau_i\tau_j),$$

we deduce from this inequality and from (7.3) and (7.4) that

$$\begin{aligned}|S| &\leqslant \frac{1}{n^{2d}p}\sum_{i=1}^{n^d}\sum_{\substack{j=1\\j\neq i}}^{n^d} |\Theta_{ij}|\bigl|\mathrm{E}(\tau_i\tau_j)\bigr| \leqslant \frac{1}{n^{2d}p}\sum_{i=1}^{n^d}\Theta_{ii}\sum_{\substack{j=1\\j\neq i}}^{n^d}\bigl|\mathrm{E}(\tau_i\tau_j)\bigr| \\ &\leqslant \frac{K_2}{n^{2d}p}\sum_{i=1}^{n^d}\Theta_{ii} \leqslant \frac{K_2}{n^d}\bigl(m + \rho^{-1/(2m+2q+1)}(2 + C \cdot C_0)\bigr).\end{aligned}$$

### 7.3 Proof of Theorem 2

Since $\beta_0 = \mathrm{E}(Y) - \langle \beta, \mathrm{E}(X)\rangle$ and $\widehat{\beta}_0 = \overline{Y} - \langle \widehat{\beta}, \overline{X}\rangle$, we have

$$\widehat{Y}_{\mathbf{i}_0} - Y^*_{\mathbf{i}_0} = \overline{Y} - \mathrm{E}(Y) - \big\langle \widehat{\beta},\, \overline{X} - \mathrm{E}(X)\big\rangle + \big\langle \widehat{\beta} - \beta,\, X_{\mathbf{i}_0} - \mathrm{E}(X)\big\rangle,$$

and thus

$$\begin{aligned}\mathrm{E}\big[(\widehat{Y}_{\mathbf{i}_0} - Y^*_{\mathbf{i}_0})^2\big] \leqslant{}& 4\mathrm{E}\big[\big(\overline{Y} - \mathrm{E}(Y)\big)^2\big] + 4\mathrm{E}\big[\big\langle \widehat{\beta},\, \overline{X} - \mathrm{E}(X)\big\rangle^2\big] \\ &+ 2\mathrm{E}\big[\big\langle \widehat{\beta} - \beta,\, X_{\mathbf{i}_0} - \mathrm{E}(X)\big\rangle^2\big].\end{aligned}$$

However, from Lemmas 3 and 5 we have

$$\mathrm{E}\big[\big\langle \widehat{\beta},\, \overline{X} - \mathrm{E}(X)\big\rangle^2\big] = O\big(n^{-d}\big),$$

and from Lemmas 4 and 5 we have

$$\mathrm{E}\big[\big(\overline{Y} - \mathrm{E}(Y)\big)^2\big] \leqslant 2\mathrm{E}\big(\overline{\epsilon}^2\big) + 2\mathrm{E}\big[\big\langle \beta,\, \overline{X} - \mathrm{E}(X)\big\rangle^2\big] = O\big(n^{-d}\big).$$

Besides, we have

$$\begin{aligned}B &:= \mathrm{E}\big[\big\langle \widehat{\beta} - \beta,\, X_{\mathbf{i}_0} - \mathrm{E}(X)\big\rangle^2\big] \\ &= \sum_{j\geqslant 1} \mathrm{E}\big[\langle \widehat{\beta} - \beta,\, \zeta_j\rangle^2 \big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle^2\big] \\ &\quad + \sum_{j\neq \ell} \mathrm{E}\big[\langle \widehat{\beta} - \beta,\, \zeta_j\rangle\langle \widehat{\beta} - \beta,\, \zeta_\ell\rangle\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\big\rangle\big] \\ &= B_1 + B_2.\end{aligned}$$

On the one hand,

$$\begin{aligned}B_1 &= \sum_{j\geqslant 1} \mathrm{Cov}\big(\langle \widehat{\beta} - \beta,\, \zeta_j\rangle^2, \big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle^2\big) \\ &\quad + \sum_{j\geqslant 1} \mathrm{E}\big(\langle \widehat{\beta} - \beta,\, \zeta_j\rangle^2\big)\mathrm{E}\big(\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle^2\big) \\ &= B_3 + B_4.\end{aligned}$$

Since $X_{\mathbf{i}_0}$ has the same distribution as that of $X$, from Assumption 4 we have

$$\begin{aligned}B_4 &= \sum_{j\geqslant 1} \mathrm{E}\big(\langle \widehat{\beta} - \beta,\, \zeta_j\rangle^2\big)\mathrm{E}\big(\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle^2\big) \\ &= \mathrm{E}\left[\left\langle \widehat{\beta} - \beta,\, \sum_{j\geqslant 1} \lambda_j \zeta_j \otimes \zeta_j(\widehat{\beta} - \beta)\right\rangle\right] = \mathrm{E}\big[\big\langle \widehat{\beta} - \beta,\, \mathit{\Gamma}(\widehat{\beta} - \beta)\big\rangle\big] \\ &= \mathrm{E}\big[\|\widehat{\beta} - \beta\|^2_{\mathit{\Gamma}}\big],\end{aligned}$$

where $\otimes$ stands for the tensor product defined by $(f \otimes g)(h) = \langle f, h\rangle g$. From Lemmas 2 and 3 with $\delta(\{\mathbf{i}_0\}, \mathcal{I}_\mathbf{n}) \geqslant \lfloor n^{4d/\theta} \rfloor$ we have

$$\begin{aligned}
B_3 &= \sum_{j\geqslant 1} \mathrm{Cov}\big(\langle \widehat{\beta} - \beta,\, \zeta_j\rangle^2, \big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle^2\big) \\
&\leqslant K \sum_{j\geqslant 1} \big\|\langle \widehat{\beta} - \beta,\, \zeta_j\rangle^2\big\|_4 \big\|\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle^2\big\|_2 \big\{\alpha_{1,\infty}\big(\delta\big(\{\mathbf{i}_0\}, \mathcal{I}_\mathbf{n}\big)\big)\big\}^{1/4} \\
&\leqslant \frac{K_1}{n^d} \sum_{j\geqslant 1} \lambda_j^{1/2} = O\left(\frac{1}{n^d}\right),
\end{aligned}$$

where $K$ and $K_1$ are positive constants. On the other hand, from Lemmas 2 and 3 with $\delta(\{\mathbf{i}_0\}, \mathcal{I}_\mathbf{n}) \geqslant \lfloor n^{4d/\theta} \rfloor$ and Assumption 4 we have

$$\begin{aligned}
B_2 &= \sum_{j\neq \ell} \mathrm{Cov}\big(\langle \widehat{\beta} - \beta,\, \zeta_j\rangle\langle \widehat{\beta} - \beta,\, \zeta_\ell\rangle, \big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\big\rangle\big) \\
&\leqslant K_2 \sum_{j\neq \ell} \big\|\langle \widehat{\beta} - \beta,\, \zeta_j\rangle\langle \widehat{\beta} - \beta,\, \zeta_\ell\rangle\big\|_4 \big\|\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\big\rangle\big\|_2 \\
&\quad \times \big\{\alpha_{1,\infty}\big(\delta\big(\{\mathbf{i}_0\}, \mathcal{I}_\mathbf{n}\big)\big)\big\}^{1/4} \\
&\leqslant \frac{K_3}{n^d} \sum_{j\neq \ell} \big\{\mathrm{E}\big[\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_j\big\rangle^2\big]\big\}^{1/4} \big\{\mathrm{E}\big[\big\langle X_{\mathbf{i}_0} - \mathrm{E}(X),\, \zeta_\ell\big\rangle^2\big]\big\}^{1/4} = O\left(\frac{1}{n^d}\right),
\end{aligned}$$

where $K_2$ and $K_3$ are positive constants. It follows that

$$\mathrm{E}\big[(\widehat{Y}_{\mathbf{i}_0} - Y^*_{\mathbf{i}_0})^2\big] \leqslant 2\mathrm{E}\big[\|\widehat{\beta} - \beta\|_\Gamma^2\big] + O\left(\frac{1}{n^d}\right).$$

Applying Corollary 1 with $2q \geqslant 1$, $\rho \sim n^{-d(2m+2q+1)/(2m+2q+2)}$, we obtain the result of Theorem 2.

## References


1. M. Bohorquez, R. Giraldo, and J. Mateu, Optimal sampling for spatial prediction of functional data, *Stat. Methods Appl.*, **25**:39–54, 2016.
2. M. Bohorquez, R. Giraldo, and J. Mateu, Multivariate functional random fields: Prediction and optimal sampling, *Stoch. Environ. Risk Assess.*, **31**:53–70, 2017.
3. S. Bouka, S. Dabo-Niang, and G.M. Nkiet, On estimation in spatial functional regression with derivatives, *C. R. Acad. Sci. Paris, Sér. I*, **356**:558–562, 2018.
4. T.T. Cai and P. Hall, Prediction in functional linear regression, *Ann. Stat.*, **34**:2159–2179, 2006.
5. H. Cardot, F. Ferraty, and P. Sarda, Functional linear model, *Stat. Probab. Lett.*, **45**:11–22, 1999.
6. H. Cardot, F. Ferraty, and P. Sarda, Spline estimators for the functional linear model, *Stat. Sin.*, **13**:571–591, 2003.
7. R. Chen, On the rate of strong mixing in stationary Gaussian random fields, *Stud. Math.*, **101**:183–191, 1992.
8. F. Comte and J. Johannes, Adaptive functional linear regression, *Ann. Stat.*, **40**:2765–2797, 2012.
9. C. Crambes, A. Kneip, and P. Sarda, Smoothing splines estimators for functional linear regression, *Ann. Stat.*, **37**: 35–72, 2009.

10. C. M. Deo, A note on empirical processes of strong mixing sequences, *Ann. Probab.*, **1**:870–875, 1973.

11. M. Francisco-Fernandez and J.D. Opsomer, Smoothing parameter selection methods for nonparametric regression with spatially correlated errors, *Can. J. Stat.*, **33**:279–295, 2005.

12. I.E. Frank and J.H. Friedman, A statistical view of some chemometrics regression tools, *Technometrics*, **35**:109–135, 1993.

13. R. Giraldo, Cokriging based on curves, prediction and estimation of the prediction variance, *InterStat*, **2**:1–30, 2014.

14. R. Giraldo, S. Dabo-Niang, and S. Martínez, Statistical modeling of spatial big data: An approach from a functional data analysis perspective, *Stat. Probab. Lett.*, **136**:126–129, 2018.

15. R. Giraldo, P. Delicado, and J. Mateu, Ordinary kriging for function-valued spatial data, *Environ. Ecol. Stat.*, **18**: 411–426, 2011.

16. R. Giraldo, J. Mateu, and P. Delicado, geofd: An R package for function-valued geostatistical prediction, *Rev. Colomb. Estad.*, **35**:383–405, 2012.

17. T. Hastie and C. Mallows, A statistical view of some chemometrics regression tools: Discussion, *Technometrics*, **35**: 140–143, 1993.

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

19. Y. Li and T. Hsing, On rates of convergence in functional linear regression, *J. Multivariate Anal.*, **98**:1782–1804, 2007.

20. A. Mas and B. Pumo, Functional linear regression with derivatives, *J. Nonparametric Stat.*, **21**:19–40, 2009.

21. D. Nerini, P. Monestiez, and C. Manté, Cokriging for spatial functional data, *J. Multivariate Anal.*, **101**:409–418, 2010.

22. M.D. Ruiz-Medina, Spatial autoregressive and moving average Hilbertian processes, *J. Multivariate Anal.*, **102**:292–305, 2011.

23. M.D. Ruiz-Medina, Spatial functional prediction from spatial autoregressive Hilbertian processes, *Environmetrics*, **23**:119–128, 2012.