Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 54 (3) (2025), 1164 – 1186
DOI : 10.15672/hujms.1640600

Research Article

Spatial kNN-Local linear estimation for semi-functional partial linear regression

Mohamed El Ouard Baouche[1], Tawfik Benchikh*[1,2], Omar Fetitah[1,3], Toufik Guendouzi[4]

[1] *Statistical and Stochastic Process Laboratory, Djillali Liabes University, Sidi Bel Abbes, Algeria*
[2] *Fauclty of Medecine, Djillali Liabes University, Sidi Bel Abbes, Algeria*
[3] *School of Computer Science of Sidi Bel Abbes (ESI-SBA), Sidi Bel Abbes, Algeria*
[4] *Laboratory of Stochastic Models, Statistics and Applications, Moulay Tahar university, Saida, Algeria*

Abstract

The objective of this paper is to investigate a semi-functional partial linear regression model for spatial data. The estimators are constructed using a k-nearest neighbors local linear method.Then, under suitable regularity conditions, we establish the asymptotic distribution of the parametric component and derive the uniform almost sure convergence rate for the nonparametric component. To assess the performance of the proposed estimators, we performed both simulation studies and real-data analyses. The results are compared with existing methods for semi-functional partial linear regression models using cross-validation. Specifically, we evaluate the predictive performance in terms of mean squared error and compare it against several benchmark estimators, including the kernel estimator, the local linear estimator and the kNN estimator. This practical study clearly demonstrates the feasibility and superiority of the local linear method estimator k-nearest neighbors over competing methods. This is evidenced by the lower mean squared error achieved by this estimator in both the simulation study and the real data application. These results indicate that this hybrid approach effectively addresses the common issue of bandwidth selection and yields estimators with reduced bias.

Mathematics Subject Classification (2020). 60G25, 62G05, 62G08,62G20,62F12

Keywords. Asymptotic normality, functional data analysis, kNN estimation, local linear estimation, partial linear regression

1. Introduction

In recent years, spatial statistics for functional data has become a major research topic in mathematics. It focuses on the modeling and analysis of data collected in a spatial order and valued in functional spaces. Such data arise in various application fields, including neuroimaging, epidemiology, chemistry, econometrics, oceanography, soil science, and environmental sciences. For an introduction to the subject, we refer to the monographs of

*Corresponding Author.
Email addresses: baouche.medouard@gmail.com (M. E. O. Baouche), benchikh.tawfik@gmail.com (T. Benchikh), o.fetitah@esi-sba.dz (O. Fetitah), tf.guendouzi@gmail.com (T. Guendouzi)
Received: 16.02.2025; Accepted: 15.05.2025

[53, 54], while for theoretical and practical aspects of functional data analysis, we refer to [37, 39, 42, 61]. In addition, statistical methods for analyzing functional data are currently undergoing intensive development. The reader may consult the survey articles by [9, 10] or the work of [58] for more recent advances and further references.

The purpose of analyzing spatially correlated functional data is to identify inherent spatial patterns that provide insight into the underlying spatial structure and dynamics of the studied phenomena. The aim is also to develop models that can be used for prediction or inference. In this context, nonparametric statistics provide a general framework for this study, particularly in nonparametric involving functional predictors, which are used to model the effect of a functional variable on a scalar response variable. For spatially correlated functional data, this type of model was first considered and explored by [30] using the non-parametric kernel method. Chouaf and Laksaci [29] proposed a functional estimator of the regression function based on the local linear method (LLM), while Saadaoui et al. [62] established the uniform almost complete convergence (with rate) of this estimator. A method of estimating spatial k-nearest neighbors for multivariate data was recently studied by [2]. An alternative approach based on robust kernel estimation was explored by [14, 15]. In the context of spatial functional data with responses missing at random, Alshahrani et al. [7] studied the kernel estimation of the regression function.

For a comprehensive discussion including state-of-the-art methodologies and applications in both parametric and non-parametric modeling of spatially correlated functional data, we refer readers to the survey articles by [56, 60] and the references therein. For reviews focused on nonparametric functional regression, see [37, 51]. However, in many practical applications, the response variable may depend simultaneously on a vector of scalar covariates and on one or more functional variables, commonly called hybrid data. In such cases, semiparametric models offer a compelling and flexible framework. These models benefit, on the one hand, from the flexibility of parametric regression models and, on the other hand, from the ability of nonparametric approaches to handle high-dimensional and infinite-dimensional data. For advanced reviews on semiparametric modeling with functional data, we refer the reader to [52, 63].

The semi-functional partial linear regression (SFPLR) is a semiparametric regression model that has received considerable attention in recent years. It is particularly suited for applications where understanding dynamic relationships among variables in temporal or spatial settings is essential. The model was introduced by [12] and is expressed as follows:

$$Y = \mathbf{X}^T \boldsymbol{\beta} + m(Z) + \varepsilon\,, \tag{1.1}$$

where Y is the scalar response variable, $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ is a p-vector of explanatory variables, β is an unknown$p$-dimensional parameter vector, Z is a functional explanatory variable, $m(.)$ is an unknown smooth functional operator, and ε are identically distributed random errors satisfying $\mathbb{E}(\epsilon) = 0$ and unknown variance $\sigma^2(\epsilon) < \infty$.

This model combines the strengths of a functional nonparametric component with the interpretability of a linear component for scalar covariates. Using parametric and nonparametric modeling techniques, the SFPLR model provides a robust framework for solving complex data analysis problems. Consequently, it has inspired a growing body of research focused on its estimation, theoretical properties, and practical applications.

Aneiros-Pérez and Vieu [12] proposed estimating the model parameters using the kernel method under the i.i.d. hypothesis, while Aneiros-Pérez and Vieu [11] extended the model to α-mixed data. Feng and Xue [35] introduced an estimation procedure based on the local linear method, and Ling et al. [48] employed the k-nearest neighbors (kNN) procedure to improve the efficiency of the estimators proposed by [12]. Boente and Vahnovan [23] investigated the properties of estimators derived from robust procedures, whereas Ling et al. [49] studied SFPLR models with randomly missing responses. More recently, Kedir et al. [45] explored a hybrid approach that combines the k-nearest neighbors method

with local linear estimation (kNN-LLE), still within the i.i.d. assumption. However, few research studies have focused on estimation in the semi-functional partial linear regression model for spatially dependent data.

An extension of the SFPLR model to spatial data, referred to as the spatial semifunctional partial linear regression (SFPLSR) model, was proposed by [20]. Using the kernel method, they established the asymptotic normality of the parametric component and derived the convergence rate in probability for the nonparametric component, considering a spatial index $i = (i_1, i_2) \in \mathbb{Z}^2$. Their work extends to the spatial functional setting the results previously obtained in [38]. [21] further explored the SFPLSR model under the setting of responses missing at random. It should be noted that the SFPLSR model can be viewed as a special case of the spatial semifunctional partial linear autoregressive model (SFPLAR) introduced in [64]. In this broader framework, quasi-maximum likelihood estimation was used for the parametric component, while the local linear estimation (LLE) method was applied to the nonparametric component, yielding convergence rates for its estimator. On the other hand, [65] studied the partial functional linear spatial autoregressive model, where estimators are constructed using functional principal component analysis (FPCA). In both of the aforementioned works, spatial dependence is modeled via a spatial weight matrix, which reflects the geographical configuration or contiguity of the observations. This matrix can be specified on the basis of decreasing geographical distance, economic proximity, or the structure of a social network. In particular, no stationarity assumption is imposed in either framework.

In this paper, we propose the use of the k-nearest neighbors combined with local linear estimation (LLE-kNN) for the semi-functional partial linear regression (SFPLR) model with spatially dependent data. This approach builds on the ideas introduced by [45] and developed in further detail in [5]. One of the main advantages of this method lies in its ability to reduce the bias term obtained with the classical kernel estimator. By combining the strengths of kNN and local linear estimation, the resulting estimator inherits favorable statistical properties from both techniques. Specifically, it produces a robust estimator with faster convergence and lower bias, while remaining straightforward to implement in practice. Moreover, the kNN-based smoothing approach naturally addresses the bandwidth selection problem, which is often a critical issue in nonparametric estimation. It should be noted that this kNN-LLE methodology was also employed by [32] in the context of semifunctional partial linear quantile regression. In fact, in practice, kNN methods offer several advantages over the traditional Nadaraya-Watson estimator. In particular, in the kNN framework, the bandwidth parameter is a random variable that depends on the distance between functional covariates. This introduces a local adaptivity feature that allows the estimator to adjust to heterogeneous data structures. Moreover, selecting the smoothing parameter k (i.e., the number of neighbors) involves considerably lower computational cost compared to bandwidth selection, as k typically takes values in a finite and manageable set. However, the theoretical analysis of kNN estimators remains more challenging due to the randomness and data dependence of the bandwidth. It is worth noting that the kNN method has recently been extended to the nonparametric functional context (see, for instance, [13, 22, 24, 43, 44, 47, 57] for recent advances and [3, 50] for comprehensive overviews). On the other hand, it is known that the local linear approach helps improve the bias term of the classical kernel method [34]. Due to this attractive mathematical efficiency, local functional linear modeling has become very popular in the analysis of nonparametric functional data in recent years. Baíllo and Grané [17] were the first to introduce this approach to estimate the functional regression operator using the Hilbert structure, and studied its asymptotic behavior under i.i.d. conditions. Later, Barrientos-Marin et al. [18] constructed a faster version using covariates with values in the Banach space for independent observations. The spatial version of the regression operator was given by [29]. Estimation of conditional models using the local linear approach was

demonstrated by [46] when the data are spatially dependent and functional in nature, while Demongeot et al. [31] studied the nonparametric local linear regression model when all variables are curves. They proved the quasi-complete (pointwise and uniform) consistency of the local linear estimator. More recent advances and references in LLE estimation in functional nonparametric regression can be found in [1, 19, 27].

Motivated by the advantages of the LLE-kNN approach, studies have been conducted in the field of Functional Data Analysis (FDA). We cite [16] investigated the estimation of the regression operator; Almanjahie et al. [5] focused on the estimation of the conditional cumulative distribution function for dependent data; and Almanjahie et al. [6] extended the methodology to the estimation of conditional expectations, distribution functions, and probability densities. The strong consistency of the kNN-LL estimator for the functional conditional density and mode was established by [28]. In addition, Rachdi [59] addressed the estimation of the regression operator when the response variable is missing at random, while Almanjahie et al. [4] considered the estimation of the conditional distribution function under the same missing data mechanism. Despite these developments, the literature on LLE-kNN estimation for functional spatial data remains scarce. To our knowledge, the only study to apply this methodology in a functional spatial setting is that of [8], who investigated its use to estimate conditional density and mode.

The objective is therefore to combine the strengths of both strategies in a spatially dependent functional data framework and to subsequently establish the almost sure convergence of the proposed estimator with respect to the number of neighbors. As mentioned previously, the main innovation lies in the integration of the local linear estimation (LLE) technique with the k-nearest neighbor (kNN) smoothing method to develop new estimators for the spatial SFPLR model. This hybrid approach solves the common problem of bandwidth selection and produces estimators with reduced bias. Nevertheless, this combination also presents new theoretical challenges. Specifically, in the kNN approach, the bandwidth parameter is inherently random, which complicates the analysis of the estimator's asymptotic properties.

The paper is organized as follows. We present our model in Section 2 whereas Section 3 is devoted to some notation and hypotheses necessary to obtain our results, which are given at the end of the section. Some simulation results and an application to real data are discussed in Section 4. Finally, the proofs of the different results are relegated to the last section.

2. The model and the local linear-kNN estimators

For $\mathbf{i} \in \mathbb{Z}^N$, $N \geq 1$ let $(Y_\mathbf{i}, \mathbf{X_i}, Z_\mathbf{i})$ be a $\mathbb{R} \times \mathbb{R}^p \times \mathcal{F}$ measurable strictly stationary spatial process defined over a probability space $(\Omega, \mathcal{A}, \mathbf{P})$, where $\mathcal{F}$ represents a functional semi-metric space equipped with a semi-metric d. We suppose that the $(Y_\mathbf{i}, \mathbf{X_i}, Z)$'s are (i.d.), which means identically distributed to (Y,$\mathbf{X}$,Z) and that the process is observed in the rectangular region expressed by $\mathcal{I}_\mathbf{n} = \{\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{Z}^N, 1 \leq i_l \leq n_l, l = 1 \ldots, N\}$ with a sample size of $\widehat{\mathbf{n}}$=$n_1 \times \ldots \times n_N$ where $\mathbf{n} = (n_1, \ldots, n_N)$. This assumption is obvious because the observation of data on regular lattice in $\mathbb{Z}^N$ corresponds to the same principle as in time series (where the observations are at times equally spaced in time), and this is also the case in practice because the processes are studied in a discrete manner. Remember that the term site is used to designate a vector $\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{Z}^N$.

The asymptotic behavior of this work is to consider the observation area asymptotically increasing while keeping the distance between observation positions to a minimum. For this, we assume that, for $l = 1 \ldots, N$, n_l tends towards infinity at the same rate: $C_1 < |n_j/n_k| < C_2$ for some $0 < C_1 < C_2 < \infty$ and we write that $\mathbf{n} \to \infty$ if $\min_{k=1\ldots,N}(n_k) \to \infty$. Furthermore, this asymptotic behavior is studied under the condition that the process $(Y_\mathbf{i}, \mathbf{X_i}, Z_\mathbf{i})$ is strictly stationary which satisfies the following α-mixing condition [26]: for

two σ-fields $\mathfrak{B}(E)$ and $\mathfrak{B}(E')$ generated by a functional random variable Z indexing by $\mathbf{i} \in E \subset \mathbb{Z}^N$ and by $\mathbf{i} \in E' \subset \mathbb{Z}^N$ respectively, where E, E' are two subsets with finite cardinals ($Card(.) < \infty$), there exists $\varphi(t) \downarrow 0$ as $t \to \infty$, such that:

$$\begin{aligned} \alpha(\mathfrak{B}(E), \mathfrak{B}(E')) &= \sup_{(A,B)\in\mathfrak{B}(E)\times\mathfrak{B}(E')} |P(A \cap B) - P(A)P(B)| \\ &\leq \Phi\left(d'(E, E')\right) \Psi\left(\mathrm{Card}(E), \mathrm{Card}(E')\right) . \end{aligned} \tag{2.1}$$

with $d'(E, E')$ means the Euclidean, and $\Psi : \mathbb{Z}^2 \longrightarrow \mathbb{R}^+$, is a symmetric positive function nondecreasing in each variable. We will be assumed that ψ satisfies

$$\forall (m, n) \in \mathbb{Z}^2, \ \Psi(m, n) \leq C \min(m, n) \, , \text{for some} \, C > 0, \tag{2.2}$$

Furthermore, as is often the case in spatial regression, we assume also assume that the process satisfies the following mixing condition:

$$\sum_{i=1}^{\infty} i^{\gamma}(\Phi(i)) < \infty \, , \text{for some} \, \gamma > 0. \tag{2.3}$$

Note that these conditions are verified by many spatial processes (for example, the spatial linear process), and the special case $N = 1$ corresponds to a strong mixture [33,40]. In what follows, we denote by Z a fixed curve in $\mathcal{F}$ and we indicate to the neighborhood of Z by $\mathcal{N}_Z$. Furthermore, we denote by $B(T, h) = \{Z^{'} \in \mathcal{F} \text{ such that } d(Z, Z^{'}) \leq h\}$ is the topological closed ball. As with all asymptotic results, in nonparametric functional statistics, it is necessary to control the local concentration of the marginal distributions of the functional observations. For this, it is assumed that the marginal distribution must satisfy the following condition: For any $h > 0$, the small ball probabilities $\varphi_Z(h) := \mathbb{P}(B(Z, h)) > 0$ is continuous and strictly increasing around 0 with $\varphi_Z(0) = 0$.

As mentioned in the introduction, our main objective in this work is to study the spatial co-variation between the response variable Y and the two explanatory variables X and Z according to a linear partial model. Recall that these links are generally modeled via the regression function $\mathbb{E}(Y_\mathbf{i}|X_\mathbf{i}, Z_\mathbf{i})$ expressed by a semi-parametric function , called semi-functional partial linear regression (SFPLR), in the following form

$$Y_\mathbf{i} = \sum_{s=1}^{p} X_\mathbf{i}^s \beta_s + g(Z_\mathbf{i}) + \epsilon_\mathbf{i} = \mathbf{X}_\mathbf{i}^T \boldsymbol{\beta} + g(Z_\mathbf{i}) + \epsilon_\mathbf{i} \quad \mathbf{i} \in \mathbb{Z}^N, \tag{2.4}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is unknown p-dimensional parameter vector, $m(.)$ is an unknown smooth functional operator and ε is the centered random error with finite unknown variance such that

$$\mathbb{E}(\epsilon_\mathbf{i}|X_\mathbf{i}^{(1)}, X_\mathbf{i}^{(2)}, \ldots, X_\mathbf{i}^{(p)}, Z_\mathbf{i}) = 0 \text{ and } \mathbb{E}(\epsilon_\mathbf{i}^2|X_\mathbf{i}^{(1)}, X_\mathbf{i}^{(2)}, \ldots, X_\mathbf{i}^{(p)}, Z_\mathbf{i}) < \infty.$$

Under the condition that $\mathbb{E}(Y_\mathbf{i}|Z_\mathbf{i} = Z)$ and $\mathbb{E}(\mathbf{X}_\mathbf{i}|Z_\mathbf{i} = Z)$ are known, by the expectation conditional, the least squares estimator (LSE) of $\boldsymbol{\beta}$ is given by

$$\overline{\boldsymbol{\beta}}_\mathbf{n} = \arg\min_{\boldsymbol{\beta}} \sum_{\mathbf{i}\in\mathcal{J}_\mathbf{n}} (\overline{Y}_\mathbf{i} - \overline{\mathbf{X}}_\mathbf{i}^T \boldsymbol{\beta})^2,$$

which is obtained by the following formula

$$\hat{\boldsymbol{\beta}}_\mathbf{n} = \left(\sum_{\mathbf{i}\in\mathcal{J}_\mathbf{n}} \overline{\mathbf{X}}_\mathbf{i} \overline{\mathbf{X}}_\mathbf{i}^T \right)^{-1} \sum_{\mathbf{i}\in\mathcal{J}_\mathbf{n}} \overline{\mathbf{X}}_\mathbf{i} \overline{Y}_\mathbf{i}. \tag{2.5}$$

where $\overline{Y}_\mathbf{i} = Y_\mathbf{i} - \mathbb{E}(Y_\mathbf{i}|Z_\mathbf{i} = Z)$ and $\overline{\mathbf{X}}_\mathbf{i} = \mathbf{X}_\mathbf{i} - \mathbb{E}(\mathbf{X}_\mathbf{i}|Z_\mathbf{i} = Z)$. However, as in the general case, $\mathbb{E}(Y_\mathbf{i}|Z_\mathbf{i} = Z)$ and $\mathbb{E}(\mathbf{X}_\mathbf{i}|Z_\mathbf{i} = Z)$ are unknown and must be estimated to apply the equation (2.5). In the following, we denote $g_\mathbf{X}(Z) = \mathbb{E}(\mathbf{X_1}|Z_\mathbf{1} = Z)$, $g_Y(Z) = \mathbb{E}(Y_\mathbf{1}|Z_\mathbf{1} = Z)$, $g_\mathbf{n}(Z) = g_Y(Z) - g_\mathbf{X}^T(Z)\boldsymbol{\beta}_n$ and assume that $g_\mathbf{X}$ and g_Y are smooth functions of Z.

Then, these functions can be estimated using nonparametric estimators. Under this, the spatial-kernel estimator [30] of $\widehat{g}_{\mathbf{X}}(Z)$ and $\widehat{g}_{Y}(Z)$ are defined by

$$\widehat{g}_{\mathbf{X}}(Z)=\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}} w_{\mathbf{n}}(Z,Z_{\mathbf{j}})\mathbf{X}_{\mathbf{j}}\,,\quad \text{and}\quad \widehat{g}_{Y}(Z)=\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}} w_{\mathbf{n}}(Z,Z_{\mathbf{j}})Y_{\mathbf{j}}, \tag{2.6}$$

where

$$w_{\mathbf{n}}(Z,Z_{\mathbf{j}})=\frac{K\left(d(Z,Z_{\mathbf{j}})h_{\mathbf{n}}^{-1}\right)}{\sum\limits_{\mathbf{i}\in\mathcal{J}_{\mathbf{n}}} K\left(d(Z,Z_{\mathbf{j}})h_{\mathbf{n}}^{-1}\right)},$$

with K denotes a real-valued kernel function and $h_{\mathbf{n}}$ a decreasing sequence of bandwidths that tends to zero as $\mathbf{n}$ tends to infinity.

From where, the kernel estimators of $\boldsymbol{\beta}_{\mathbf{n}}$ and $g_{\mathbf{n}}$ are defined by [20] as

$$\widehat{\boldsymbol{\beta}}_{\mathbf{n}}=\left(\sum_{\mathbf{i}\in\mathcal{J}_{\mathbf{n}}}\breve{X}_{\mathbf{i}}(\breve{X}_{\mathbf{i}})^{T}\right)^{-1}\left(\sum_{\mathbf{i}\in\mathcal{J}_{\mathbf{n}}}\breve{X}_{\mathbf{i}}\breve{Y}_{\mathbf{i}}\right) \tag{2.7}$$

and

$$\widehat{g}_{\mathbf{n}}(Z)=\widehat{g}_{Y}(Z)-\left(\widehat{g}_{\mathbf{X}}(Z)\right)^{T}\widehat{\boldsymbol{\beta}}_{\mathbf{n}}, \tag{2.8}$$

where

$$\breve{Y}_{\mathbf{i}}=Y_{\mathbf{i}}-\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}} w_{\mathbf{n}}(Z,Z_{\mathbf{j}})Y_{\mathbf{j}}\ ,\ \breve{X}_{\mathbf{i}}=X_{\mathbf{i}}-\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}} w_{\mathbf{n}}(Z,Z_{\mathbf{j}})X_{\mathbf{j}}.$$

The main purpose of this article is to using local linear approach weighted by the k-nearest neighbors smoothing procedure (kNN-LLE) to estimate $g_{\mathbf{X}}$ and g_{Y}. This approach was used for SFPLR models in the i.i.d. case by [45]. This study extends the existing approach to spatial type-dependent data. For this, we assume that, for all $Z_{\mathbf{l}}\in\mathcal{N}_{Z}$, the function $g_{\cdot}(Z_{\mathbf{l}})$ is locally approximated by

$$g_{\cdot}(Z_{\mathbf{l}})=a+b\varrho(Z_{\mathbf{l}},Z)+o(d(Z_{\mathbf{l}},Z)),$$

where $\varrho(\cdot,\cdot)$ is a known function from $\mathcal{F}^{2}$ into $\mathbb{R}$ such that $\varrho(Z,Z)=0$, and we denote by $h_{k}=h_{k,\mathbf{n}}$ the random sequence of positive real numbers such that

$$h_{k}=\min\{h\in\mathbb{R}^{+}\text{ such that }\textstyle\sum_{\mathbf{l}\in\mathcal{J}_{\mathbf{n}}}1_{B(Z,h)}(Z_{\mathbf{l}})=k\}.$$

Under this, the LLE-kNN estimator of $g_{\mathbf{X}}(Z)$ and $g_{Y}(Z)$ is defined in [29] and expressed by

$$\widetilde{g}_{\mathbf{X}}(Z)=\frac{\sum_{\mathbf{i},\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}W_{\mathbf{ij}}(Z,h_{k})\mathbf{X}_{\mathbf{j}}}{\sum_{\mathbf{i},\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}W_{\mathbf{ij}}(Z,h_{k})}=\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)\mathbf{X}_{\mathbf{j}}\,, \tag{2.9}$$

and

$$\widetilde{g}_{Y}(Z)=\frac{\sum_{\mathbf{i},\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}W_{\mathbf{ij}}(Z,h_{k})Y_{\mathbf{j}}}{\sum_{\mathbf{i},\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}W_{\mathbf{ij}}(Z,h_{k})}=\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)Y_{\mathbf{j}}, \tag{2.10}$$

where $W_{\mathbf{ij}}(Z,h_{k})=\varrho_{\mathbf{i}}(\varrho_{\mathbf{i}}-\varrho_{\mathbf{j}})K_{\mathbf{i}}K_{\mathbf{j}}$, with $\varrho_{\mathbf{i}}=\varrho(Z,Z_{\mathbf{i}})$ and $K_{\mathbf{i}}=K(h_{k}^{-1}d(Z,Z_{\mathbf{i}}))$. Note that K stands for the kernel function. Hence, an estimator of $\boldsymbol{\beta}$ after estimating $g_{\mathbf{X}}$ and g_{Y} is given by

$$\widetilde{\boldsymbol{\beta}}_{\mathbf{n}}=\left(\sum_{\mathbf{i}\in\mathcal{J}_{\mathbf{n}}}\widetilde{\mathbf{X}}_{\mathbf{i}}^{T}\widetilde{\mathbf{X}}_{\mathbf{i}}\right)^{-1}\sum_{\mathbf{i}\in\mathcal{J}_{\mathbf{n}}}\widetilde{\mathbf{X}}_{\mathbf{i}}^{T}\widetilde{Y}_{\mathbf{i}}, \tag{2.11}$$

where $\widetilde{\mathbf{X}}_{\mathbf{i}}=\mathbf{X}_{\mathbf{i}}-\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)\mathbf{X}_{\mathbf{j}}$ and $\widetilde{Y}_{\mathbf{i}}=Y_{\mathbf{i}}-\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)Y_{\mathbf{j}}$. Finally, a nonparametric estimator for $\widetilde{g}_{\mathbf{n}}$ in SFPLR can be obtained by (2.8) and (2.11)

$$\widetilde{g}_{\mathbf{n}}(Z)=\widetilde{g}_{Y}(Z)-\widetilde{g}_{\mathbf{X}}(Z)^{T}\widetilde{\boldsymbol{\beta}}_{\mathbf{n}}=\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)Y_{\mathbf{j}}-\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)\mathbf{X}_{\mathbf{j}}^{T}\widetilde{\boldsymbol{\beta}}_{n}. \tag{2.12}$$

It should be noted that the parameter k is unknown, so it is necessary to estimate it and for this we will use the cross-validation method. Precisely, we choose k according to the following cross-validation rule

$$k_{opt} = \arg \min_{k \in]k_{1,n}, k_{2,n}[} CV(k) = \arg \min_{k \in]k_{1,n}, k_{2,n}[} \sum_{i=1}^{n} \left(Y_i - \widetilde{Y}_{(-i)}^{kNN} \left((X_i, Z_{\mathbf{i}}) \right) \right)^2 \tag{2.13}$$

where $k_{1,n}$ and $k_{2,n}$ are two sequences of strictly positive integers and $\widetilde{Y}_{kNN-LL}^{(-\mathbf{i})}(X_{\mathbf{l}}, Z_{\mathbf{l}})$ is the values of the leave one-out kNN-LL estimator without observation $(X_{\mathbf{i}}, Z_{\mathbf{i}})$ after estimating $\beta_{\mathbf{n}}$. The existence of two suites $k_{1,n}$ and $k_{2,n}$ is guaranteed by the results of [55] (See [6]).

3. Hypotheses and the theoretical results

We introduce the following notations and pose the required hypotheses that are necessary to obtain our main results. In particular, to derive the asymptotic normality of the estimator $\widetilde{\boldsymbol{\beta}}_{\mathbf{n}}$, the uniform almost complete convergence of $\widetilde{g}_{\mathbf{n}}(Z)$ in some subset $\mathcal{S}$ of $\mathcal{F}$. We denote

$$g_{\mathbf{X_i}}^{(s)}(Z) = \mathbb{E}(\mathbf{X}_{\mathbf{i}}^{(s)} | Z_{\mathbf{i}} = Z), \quad (s = 1, \ldots, p),$$

$$\theta_{\mathbf{i}}^{(s)} = X_{\mathbf{i}}^{(s)} - g_{\mathbf{X_i}}^{(s)}(Z), \quad \boldsymbol{\theta}_{\mathbf{i}} = (\theta_{\mathbf{i}}^{(1)}, \ldots, \theta_{\mathbf{i}}^{(p)})^T .$$

The expressions of our estimators in (2.11) and (2.12) contain estimators of $\theta_{\mathbf{i}}$. Let $\mathcal{S}$ be a subset of $\mathcal{F}$, such that $\mathcal{S} \subset \bigcup_{l=1}^{d_{\mathbf{n}}} B(Z_l; r_{\mathbf{n}})$, where $d_{\mathbf{n}} > 1$ is some integer, $r_{\mathbf{n}}$ is a sequence of positive real numbers, and $Z_l \in \mathcal{F}$, $l = 1, \ldots, d_{\mathbf{n}}$ (this set can always be built). Subsequently, we assume that Z is valued in $\mathcal{S}$.

Technical assumptions

(A1): The spatial process $(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}}, Z_{\mathbf{i}})_{\mathbf{i}}$ satisfies the followings:

i) $\forall z \in \mathcal{S}$, there exist $C > 0$ such that, for some $1 < a < \gamma N^{-1}$,

$$\sup_{\mathbf{i} \neq \mathbf{j}} \mathbb{P}\left[(Z_{\mathbf{i}}, Z_{\mathbf{j}}) \in (B(z,h) \times B(z,h)) \right] \leq C(\varphi_z(h))^{\frac{a+1}{a}} .$$

ii For some $m \geq 4$, there exists $C' > 0$, $C'' > 0$, such that

- $\mathbb{E}(|Y_{\mathbf{i}}|^m | Z_{\mathbf{i}} = Z) < \varkappa_m^1(Z) < C' < \infty$ and $\sup_{\mathbf{i} \neq \mathbf{j}} \mathbb{E}\left[|Y_{\mathbf{i}} Y_{\mathbf{j}}| \left(Z_{\mathbf{i}}, Z_{\mathbf{j}} \right) \right] < \infty$,
- for $s = 1, \ldots, p$, $\mathbb{E}(|X_{\mathbf{i}}^{(s)}|^m | Z_{\mathbf{i}} = Z) \leq \varkappa_m^2(Z) < C'' \infty$ and $\sup_{\mathbf{i} \neq \mathbf{j}} \sup_{1 \leq s \leq p} \mathbb{E}\left[X_{\mathbf{i}}^{(s)} X_{\mathbf{j}}^{(s)} | \left(Z_{\mathbf{i}}, Z_{\mathbf{j}} \right) \right] < \infty$,

where $\varkappa_m^l(Z)$, $l = 1, 2$, are continuous functions on $\mathcal{S}$.

(A2): There exists a differentiable invertible function $\varphi(.)$, such that

i) for all $Z \in \mathcal{S}$: $0 < C_1 \varphi(h) \leq \varphi_Z(h) < C_2 \varphi(h)$, for $C_1 > 0$ and $C_2 > 0$.

ii) $\exists h_0 > 0$, such that for all $h_0 > h$, $\varphi'(h) < C_1$, where φ' denote the first derivative of φ.

iii) for all $t \in]0, 1[$:

$$\lim_{h_{\mathbf{n}} \to 0} \frac{\varphi(t\, h_{\mathbf{n}})}{\varphi(h_{\mathbf{n}})} = \iota(t).$$

iv) There exists $\eta > 0$, $\tau > 0$, $C_3, C_4 > 0$ such that: $C_3 \widehat{\mathbf{n}}^{(3-a/a+1)+\eta} \leq \varphi(h) \leq C_4 \widehat{\mathbf{n}}^{\tau}$, with $\eta < (a-3)/(a+1)$ and $\tau > 1$.

(A3): For some $C_5 < \infty$, there exists a positive number α, such that, for all u, $v \in \mathcal{S}$, the functions g and $g_{\mathbf{X_i}}^{(s)}$ are both in the set

$$\{ f : \mathcal{F} \to \mathbb{R}, |f(u) - f(v)| \leq C_5 d^{\alpha}(u, v) \}.$$

(A4): The kernel K has support $[0,1]$, which is Lipschitz function on $[0,1[$. Moreover, there exist two constants C_6 and C_7 such that $-\infty < C_6 < K'(.) < C_7 < 0$ for $K(1) > 0$.

(A5): The function ϱ satisfies the following three conditions:

- for $\forall z, z' \in \mathcal{F}$, there exists $0 < c_1 < c_2$, such that $c_1\, d(z,z') \leq \mid \varrho(z,z') \mid \leq c_2 d(z,z')$,
- $\forall z_1, z_2 \in \mathcal{S}$, there exists c_3, such that $|\varrho(z_1,z) - \varrho(z_2,z)| \leq C_3 d(z_1,z_2)$.
- For all sequence $h_n \to 0$, we have

$$h_{\mathbf{n}} \left(\int_{B(z,h_{\mathbf{n}})} \varrho(u,z) dP^z(u) \right) / \left(\int_{B(z,h_{\mathbf{n}})} \varrho^2(u,z) dP^z(u) \right) \to 0,$$

where dP^z denotes the probability distribution of the regressor Z.

(A6): The number of neighbors k is such that

$$k/\widehat{\mathbf{n}} \to 0 \text{ and } \frac{\log \widehat{\mathbf{n}}}{k} \to 0 \text{ as } \mathbf{n} \to \infty.$$

(A7): The subset $\mathcal{S}$ is such that, for $r_{\mathbf{n}} = O(\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}})$, the sequence $d_{\mathbf{n}}$ satisfies:

$$\frac{(\log \widehat{\mathbf{n}})^2}{k} < \log d_{\mathbf{n}} < \frac{k}{\log \widehat{\mathbf{n}}}\,,$$

and there exits $\varsigma > 1$ such that $\sum_{\mathbf{n}} d_{\mathbf{n}}^{1-\varsigma} < \infty$.

(A8): There exists a sequence of positive real numbers $\vartheta_{\mathbf{n}}$ such that

- $(\widehat{\mathbf{n}})^{1/2} \frac{\varphi(h)^{(2-\beta)/2\beta}}{\vartheta_{\mathbf{n}}} \to 0$
- $\sum_{\mathbf{n}} \widehat{\mathbf{n}} \vartheta_{\mathbf{n}}^{-m} < +\infty$,
- $\sum_{\mathbf{n}} \widehat{\mathbf{n}}^{1-\gamma/2N} \vartheta_{\mathbf{n}}^{\gamma/N} \left(\frac{\log \widehat{\mathbf{n}}}{\varphi(h)}\right)^{\gamma/2N} < \infty$.

(A9): Let $\mathbf{\Sigma} = \mathbb{E}\left[\left(\boldsymbol{\theta_1}(\boldsymbol{\theta_1})^T\right)\right]$ and let $R_{\mathbf{i}} = \boldsymbol{\theta}_{\mathbf{i}}\varepsilon_{\mathbf{i}}$ where $\mathbf{1}$ is the site spatial $(1,\ldots,1)$.

i) We assume that $\mathbf{\Sigma} = \mathbb{E}(\boldsymbol{\theta_1}\boldsymbol{\theta_1}^T)$ is an invertible matrix.

ii) we assume that the matrix $\mathbf{C} = \sum_{\mathbf{i}\in\mathbb{J}_{\mathbf{n}}} \mathbb{E}\left[\left(R_{\mathbf{0}} R_{\mathbf{i}}^T\right)\right]$ is positive definite, where $\mathbf{0}$ denote the spatial site $(0,\ldots,0)$.

We first recall that our hypotheses cover the three fundamental aspects of the asymptotic study in non-parametric functional data analysis: the type of functional data, the functional space and the parameters involved in local functional linear analysis in the context of uniform consistency including the kernel function, the smoothing parameter and the bifunctional operator ρ). Note that these assumptions are not the weakest possible conditions, but they are imposed to facilitate the proof of our results. In fact, hypothesis (A1) measures the local dependence between observations. Such a condition is necessary to achieve the same rate of convergence as in the i.i.d. case. Likewise, similar to the discussions in [44], assumption (A2) imposes the usual conditions on the probability of a small ball. Assumption (A2)(iii) concerns the standard concentration property of the functional variable, which is well documented as a key aspect in exploring the functional nature of the data. The variability of the small-ball probability is addressed by Assumption (A2)(iii), which is used to control the bias of non-parametric estimators. This assumption holds for several continuous-time processes (e.g., Gaussian processes, diffusion processes, and general Gaussian processes) and has been extensively discussed in the literature on nonparametric functional statistics [37]. These hypotheses could be weakened; however, the convergence rate would be affected by the presence of covariance terms. Assumptions (A6) and (A10) are the same as those used in [50] to obtain the uniform consistency rate of any estimator based on the kNN approach for the dependent data. The conditions in the set S are provided by assumption (A10). Naturally, these conditions also impose restrictions on the small-ball probability function ϕ, as expressed by assumption (A2). We can see [36,37] for examples of subsets S and functional spaces $\mathcal{F}$ where assumption (A10)

is satisfied. Assumption (A3) characterizes not only the functional space of our model but also allows us to evaluate the bias term in our asymptotic results. This is a technical assumption that enables the application of Bernstein's inequality to obtain almost complete convergence. Assumption (A4) concerns the kernel $K(.)$. It includes two types of kernels that are traditionally used in practice: box and continuous kernels. This assumption is satisfied by several kernels, such as the Epanechnikov kernel, Parzen kernels, triangular kernel, and others. Assumption (A5) has been introduced and discussed in [18] in the context of functional local linear regression. Readers can find in that reference several examples of bi-functional operators ρ that satisfy this condition. The moment integrability condition in assumption (A1)(ii) and the additional assumptions (A3) and (A11) are standard in the context of SFPLR. Finally, assumptions (A2)(iv) and (A8) are technical conditions introduced to simplify the proofs [62].

We are now in a position to give our asymptotic results. The first gives the asymptotic distribution of the estimator for the parametric component of the model ($\widehat{\boldsymbol{\beta}}_{\mathbf{n}}$).

Theorem 3.1. *When the assumptions* ***(A1)-(A9)****, 2.1, 2.2, 2.3 hold, if additionally the bandwidth parameter* $h_{\mathbf{n}}$ *and the function* $\varphi_z(h_{\mathbf{n}})$ *satisfies* $\widehat{\mathbf{n}}\frac{\varphi_x(h_{\mathbf{n}})}{\log(\widehat{\mathbf{n}})} \longrightarrow \infty$ *and* $\widehat{\mathbf{n}}h_{\mathbf{n}}^{\kappa}\frac{\varphi_x(h_{\mathbf{n}})}{\log(\widehat{\mathbf{n}})} \longrightarrow 0$*, when* $\mathbf{n} \longrightarrow \infty$*, we have:*

$$(\widehat{\mathbf{n}})^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\mathbf{n}} - \boldsymbol{\beta}\right) \xrightarrow{D} \mathcal{N}\left(0, \boldsymbol{\Sigma}^{-1}\mathbf{C}\left(\boldsymbol{\Sigma}^{-1}\right)^{T}\right). \tag{3.1}$$

The following results give the rate of uniform almost complete convergence for the nonparametric component.

Theorem 3.2. *Based on hypotheses of Theorem 3.1, we have*

$$\sup_{Z\in\mathcal{S}}|\widehat{g}_{\mathbf{n}}(z) - g(z)| = O\left(\varphi^{-1}\left(\frac{k_{\mathbf{n}}}{\widehat{\mathbf{n}}}\right)\right) + O_{a.co}\left(\sqrt{\frac{\log d_{\mathbf{n}}}{k_{\mathbf{n}}}}\right). \tag{3.2}$$

4. Computational study

The main objective of this section is to examine the behavior and practical implementation of the functional linear local kNN approach on finite samples generated by the SFPLR model (2.4), with particular attention to the influence of spatial correlation on the efficiency of estimators. Specifically, we compare the behavior of the mean squared error of prediction of the semifunctional partial linear regression model for the following estimators to highlight the superiority of this approach over other estimation methods:

- Spatial Semi-functional partial linear kernel regression (SFPL Kernel CV) introduced by [20],
- Spatial Semi-functional partial linear kernel kNN regression (SFPL Kernel KNN) proposed by [48],
- Spatial Semi-semi-functional partial linear Local-Linear kernel regression (SFPL Local.L CV), our estimator given by the equation (2.11) and (2.12) by replacing h_k with h_n (through the utilization of a cross-validation procedure),
- Spatial Semi-functional partial linear Local-Linear kNN regression (SFPL Local.L KNN), our estimator given by the equations (2.11) and (2.12).

Recall that the kernel CV estimator of the SFPL parameters is given by Equations (2.6), (2.7) and (2.8). with

$$w_{\mathbf{n}}(z, Z_{\mathbf{i}}) = \frac{K\left(d(z, Z_{\mathbf{i}})h_{\mathbf{n}}^{-1}\right)}{\sum_{\mathbf{i}\in\mathbb{J}_{\mathbf{n}}} K\left(d(z, Z_{\mathbf{i}})h_{\mathbf{n}}^{-1}\right)}, \tag{4.1}$$

and $h_{\mathbf{n}}$ being a sequence of bandwidths tending to zero as $\mathbf{n}$ tends to infinity, and the kernel K is a function from $\mathbb{R}^+$ to $\mathbb{R}^+$. The h_{opt} is the data-driven bandwidth obtained by a cross-validation procedure

$$h_{opt} = \arg\min_{h} CV(h)$$

where

$$\begin{aligned} CV(h) &= \sum_{\mathbf{i}\in\mathcal{I}} \left[Y_{\mathbf{i}} - \left(\widehat{g}_{\mathbf{n}}^{(-\mathbf{i})}(Z_{\mathbf{i}}) + \mathbf{X}_{\mathbf{i}}^T \widehat{\beta}_{\mathbf{n}} \right) \right]^2 \\ &= \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \left(Y_{\mathbf{i}} - \widehat{r}_{(-\mathbf{i})}^{CV}(X_{\mathbf{i}}, Z_{\mathbf{i}}) \right)^2, \end{aligned}$$

with $\widehat{g}_{\mathbf{n}}^{(-\mathbf{i})}(.)$ is the estimator of $g(.)$ based on the leave-one-out method calculated without observation $(X_{\mathbf{i}}, Z_{\mathbf{i}})$ after estimating $\beta_{\mathbf{n}}$.

The Kernel kNN estimator of the SFPL parameters is given by Equations (2.7) and (2.8) with

$$w_{\mathbf{n}}(z, Z_{\mathbf{i}}) = \frac{K\left(d(z, Z_{\mathbf{i}}) h_{k,\mathbf{n}}^{-1} \right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} K\left(d(z, Z_{\mathbf{i}}) h_{k,\mathbf{n}}^{-1} \right)}, \tag{4.2}$$

where $h_{k,\mathbf{n}}$ is a random bandwidth parameter defined as

$$h_{k,\mathbf{n}} = \min \left\{ h \in \mathbb{R}^+ \text{ such that } \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \mathbb{I}_{B(z,h)}(z_{\mathbf{i}}) = k \right\}.$$

The k-Nearest Neighbors technique is utilized to derive $h_{k_{opt},\mathbf{n}}$, which represents the bandwidth associated with the optimal number of neighbors, as determined by following cross-validation:

$$k_{opt} = \arg\min_{h} CV(k) \text{ with } CV(k) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \left(Y_{\mathbf{i}} - \widehat{Y}_{(-\mathbf{i})}^{kNN} \right)^2. \tag{4.3}$$

The $\widehat{Y}_{(-\mathbf{i})}^{kNN}$ is the leave-one-out values of the SFPL regression estimators calculated without observation $(X_{\mathbf{i}}, Z_{\mathbf{i}})$ after estimating $\beta_{\mathbf{n}}$.

4.1. Simulation study

We perform a simulation based on observations $(X_{\mathbf{i}}, Z_{\mathbf{i}}, Y_{\mathbf{i}}) \in (\mathbb{R}^2 \times \mathcal{E} \times \mathbb{R})$. In this case, we take $p = 2$, $\mathbf{i} = (i_1, i_2) \in \mathbb{Z}^2$ with $1 \le i_1 \le n_1$, $1 \le i_2 \le n_2$. The model was generated as follows:

$$Z_{\mathbf{i}}(t) = A_{\mathbf{i}} \cos(2\pi t), t \in [0,1], \quad X_{\mathbf{i}} = (X_{\mathbf{i}}^1, X_{\mathbf{i}}^2),$$

$$Y_i = X_{\mathbf{i}}^T \ \beta + g(Z_{\mathbf{i}}) + \varepsilon_{\mathbf{i}}. \tag{4.4}$$

Then, we simulated model with following assumptions:

- $\beta = (-1,\ 3)^T$ (^{T} mean is the transpose symbol),
- $X_{\mathbf{i}}^k \sim U(-1, 2)$, $k = 1, 2$,
- $A_{\mathbf{i}} = D_{\mathbf{i}} \times \left(2\cos(2G) + \exp\left(-4G^2\right)\right)$, with $G = GRF(0, 5, 3)$,
- $g(Z_{\mathbf{i}}) = \frac{A}{\pi^2} Z_{\mathbf{i}}^{(2)}$ ($Z^{(2)}$ denotes the second derivatives of a function Z),
- $\varepsilon = GRF(0, .1, 5)$,

where

- $GRF\left(m, \sigma^2, s\right)$ is a stationary Gaussian random field with mean m and covariance function defined by

$$C(l) = \sigma^2 \exp\left(-\left(\frac{\|l\|}{s}\right)^2\right), l \in \mathbb{R}^2 \text{ and } s > 0.$$

- $D_{\mathbf{i}} = \frac{1}{n_1 \times n_2} \sum_{\mathbf{j}} \exp\left(-\frac{\|\mathbf{i}-\mathbf{j}\|}{a}\right)$ $\left(D_{(\mathbf{i},\mathbf{j})} = \frac{1}{n_1 \times n_2} \sum_{1 \leq j_1, j_2 \leq 25} \exp\left(-\frac{\|(i_1,i_2)-(j_1,j_2)\|}{a}\right)\right)$.

The function D is here to ensure and control the spatial mixing condition even if using Gaussian random fields also brings some spatial dependency. The curve of $Z(t)$, following the values of a, is shown in Figure 1.

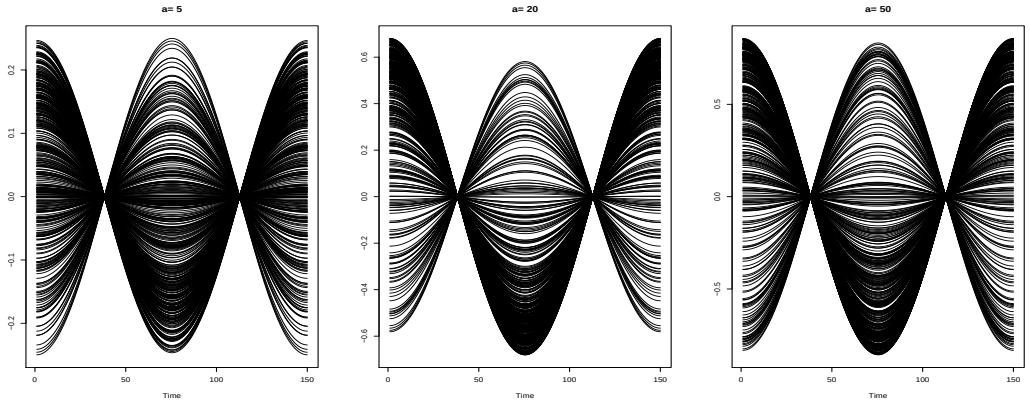


Figure 1. The curves $Z_{\mathbf{i}}$, for a=5; 20; 50.

The covariance function is presented in Figure 2.

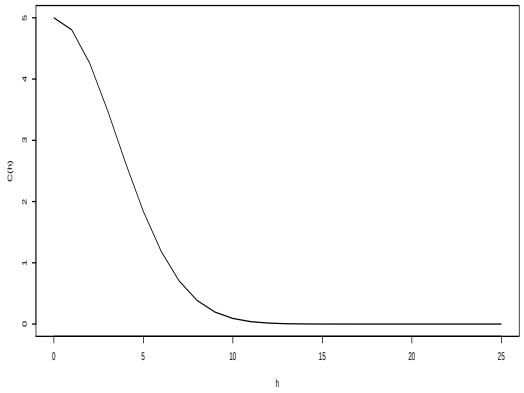


Figure 2. Covariance function with $\sigma^2 = 5$ and $s = 5$.

As seen in Figure 2, since the model is based on Gaussian random fields with covariance function C and scale $s = 5$, observations of sites $\mathbf{i}$ and $\mathbf{j}$ with $\|\mathbf{i}-\mathbf{j}\| < 15$ spatial dependence are almost independent of $\|\mathbf{i} - \mathbf{j}\| \geq 15$. So, our observations are a mixture of i.i.d. and dependent observations). Thus, to move away from independence, it suffices to lower the value of a. Random field simulation is presented in Figure 3

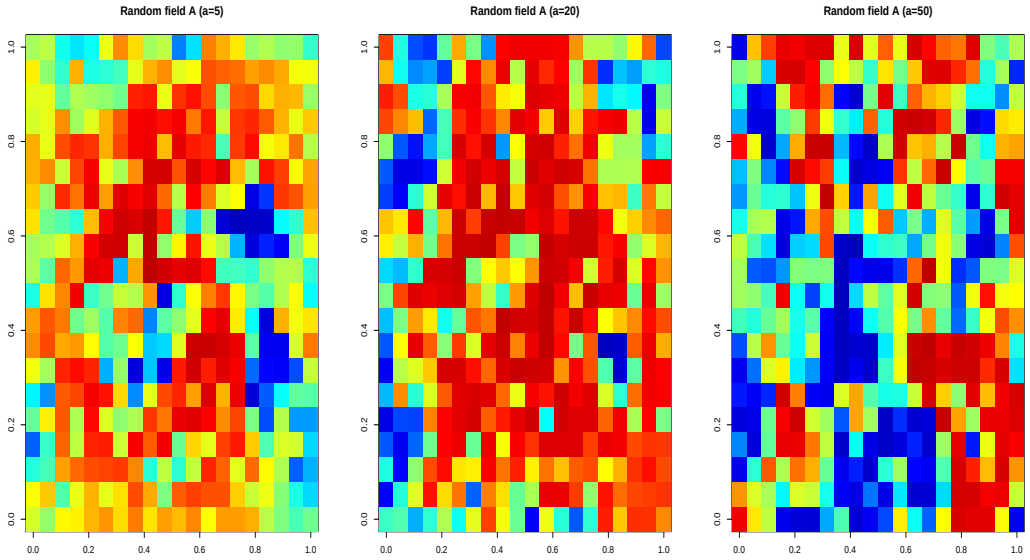


Figure 3. Random field simulation.

To compute our estimators, we use the class of semi-metrics d based on derivatives of sample curves, given by

$$d\left(Z_{\mathbf{i}}, Z_{\mathbf{j}}\right) = \sqrt{\int_0^1 \left(Z_{\mathbf{i}}^{(\kappa)}(t) - Z_{\mathbf{j}}^{(\kappa)}(t)\right)^2 dt}, \quad \text{for } \forall Z_{\mathbf{i}}, Z_{\mathbf{j}} \in \mathcal{E}.$$

We select the usual kernel function as $K(u) = \frac{3}{2}\left(1-u^2\right) 1_{[0,1]}(u)$. For the local-linear estimator, it is constructed by the same procedure proposed by [18] for which the locating function ϱ is defined by

$$\varrho(Z_{\mathbf{i}}, Z_{\mathbf{j}}) = \int_0^1 \theta(t)(Z_{\mathbf{i}}^{(\kappa)}(t) - Z_{\mathbf{j}}^{(\kappa)}(t))dt \tag{4.5}$$

where $Z^{(\kappa)}(t)$ denoting the κth derivative of the curve $Z_{\mathbf{i}}(t)$ and θ is the eigenfunction of the empirical covariance operator given by

$$\frac{1}{|\mathcal{I}_{\mathbf{n}}|}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}(Z_{\mathbf{i}}^{(\kappa)}(t) - \overline{Z^{(\kappa)}(t)})^t(Z_{\mathbf{i}}^{(\kappa)}(t) - \overline{Z^{(\kappa)}(t)}),$$

where $\overline{Z_{\mathbf{i}}^{(\kappa)}(t)} = \dfrac{1}{|\mathcal{I}_{\mathbf{n}}|}\displaystyle\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Z_{\mathbf{i}}^{(\kappa)}(t)$, associated with the q-greatest eigenvalue.

In this simulation study, we take the following parameters $\kappa = 2$, $q = 5$, and $|\mathcal{I}_{\mathbf{n}}| = n_1 \times n_2$.

In order to check the performance of the proposed estimator, denoted by $\widehat{r}(x,z) = x^T\widehat{\beta} + \widehat{g}_{m,n}(z)$ with $(z,x) \in (\mathcal{E}\times\mathbb{R}^2)$, we randomly split our data $(X_{\mathbf{i}}, Z_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}}$ into two subsets such as the learning sample $(X_{\mathbf{i}}, Z_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}$ and the test sample $(X_{\mathbf{i}}, Z_{\mathbf{i}}, Y_{\mathbf{i}}) \in \mathcal{I}_{\mathbf{n}'}$. The training sample was used to choose the smoothing parameters $h_{k_{opt}}$ and h_{opt} for k-Nearest Neighbors and cross-validation CV procedures.

The performance of the models depends on the parameters used in the estimation process. In fact, bandwidth parameters play a critical role in nonparametric estimation, affecting all asymptotic properties, and in particular the rate of convergence. As noted in Section 2, in our study, the k-Nearest Neighbors technique is utilized to derive $h_{k_{opt}}$, which represents the bandwidth associated with the optimal number of neighbors, as determined

by cross-validation in equation (2.13).

$$k_{opt} = \arg\min_{k} CV(k) = \arg\min_{k} \sum_{i=1}^{n} \left(Y_i - \widetilde{Y}_{(-i)}^{kNN-LL}((X_i, Z_{\mathbf{i}})) \right)^2,$$

where $\widetilde{Y}_{(-i)}^{kNN-LL}((X_i, Z_{\mathbf{i}}))$ is the values of the leave-one-out kNN-LL estimator without observation $(X_{\mathbf{i}}, Z_{\mathbf{i}})$ after estimating $\beta_{\mathbf{n}}$.

We use the mean square error (MSE) for the SFPLR model studied by [20]), the kNN estimator studied by [48] and the LLE-kNN estimator as an accuracy measure and is defined as follows:

$$MSE_{kernel-CV,kernel-kNN} = \frac{1}{\#(\mathcal{J}'_{\mathbf{n}})} \sum_{\mathrm{i}\in\mathcal{J}'_{\mathbf{n}}} \left[Y_{\mathbf{i}} - \left(X_{\mathbf{i}}^T \widehat{\beta} + \widehat{g}_n(Z_{\mathbf{i}}) \right) \right]^2,$$

$$MSE_{LL-CV,LL-kNN} = \frac{1}{\#(\mathcal{J}'_{\mathbf{n}})} \sum_{\mathrm{i}\in\mathcal{J}'_{\mathbf{n}}} \left[Y_{\mathbf{i}} - \left(X_{\mathbf{i}}^T \widetilde{\beta} + \widetilde{g}_n(Z_{\mathbf{i}}) \right) \right]^2,$$

where $\#(\mathcal{J}'_{\mathbf{n}})$ is the size of testing sample $\mathcal{J}'_{\mathbf{n}}$.

The prediction results are presented in Figure 4, where the predicted values are plotted against the true values for the SFPLR models using four different estimation methods: The SFPL Kernel CV regression, SFPL Kernel kNN regression, SFPL Local.L CV regression and SFPL Local-Linear kNN regression. The results are shown for various values of parameter a. The predictors appear to yield satisfactory results, with a slight advantage observed for the SFPL Local Linear kNN regression. This observation is supported by MSE computations. The MSE prediction (MSEP) errors are summarized in Table 1 and illustrated in Figure 5, which displays boxplots of the MSEP values for the testing sample in the four proposed regression estimation methods. Both Table 1 and Figure 5 demonstrate that the local linear kNN estimator consistently achieves the most accurate predictions. It outperforms the alternative methods by producing significantly lower MSE values, underscoring the robustness and reliability of the proposed approach.

Table 1. The MSE values of the models

n_1 ↓	Model → n_2 ↓	SFPL Kernel CV	SFPL Local.L CV	SFPL Kernel KNN	SFPL Local.L KNN
10	10	0.1757	0.17115	0.10695	0.10405
	20	0.1556	0.10745	0.0878	0.0839
	50	0.1356	0.0873	0.0826	0.07185
20	10	0.13665	0.0967	0.08465	0.0808
	20	0.12585	0.0785	0.0756	0.06905
	50	0.1191	0.0775	0.07395	0.06545
50	10	0.12915	0.08715	0.078	0.071
	20	0.11605	0.07745	0.0691	0.06355
	50	0.11	0.0741	0.06465	0.0633

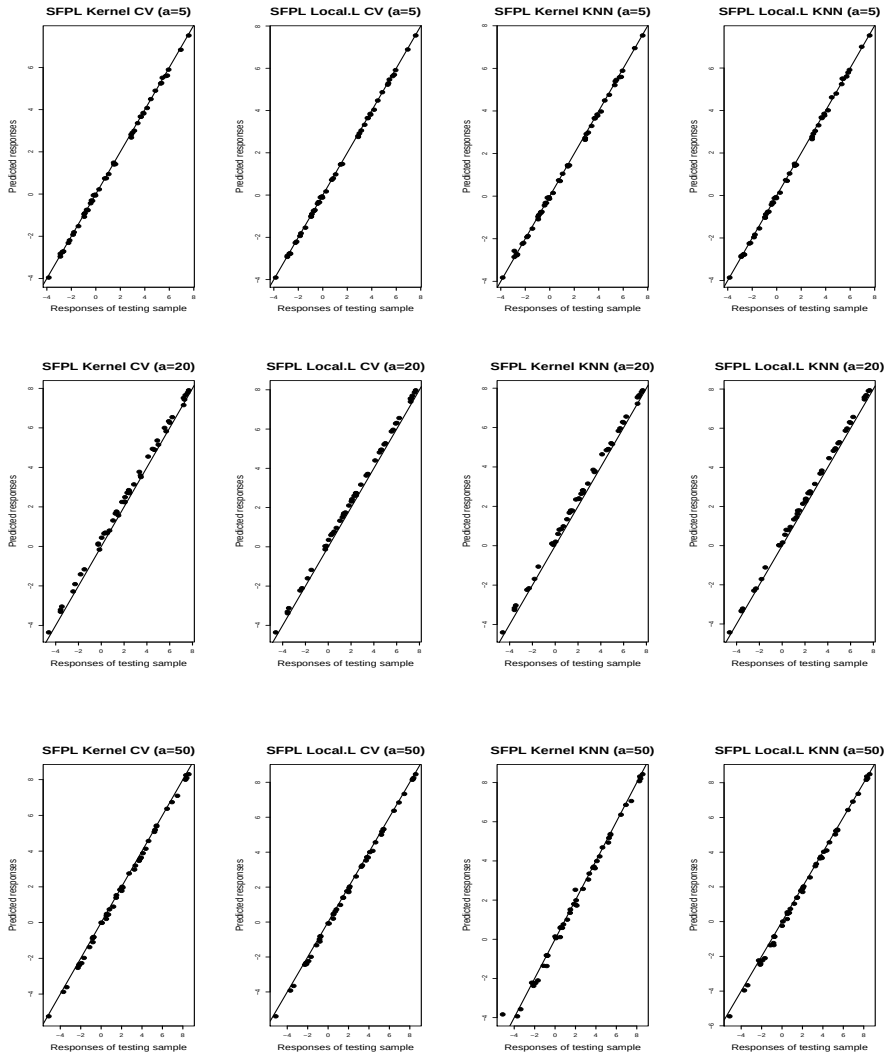
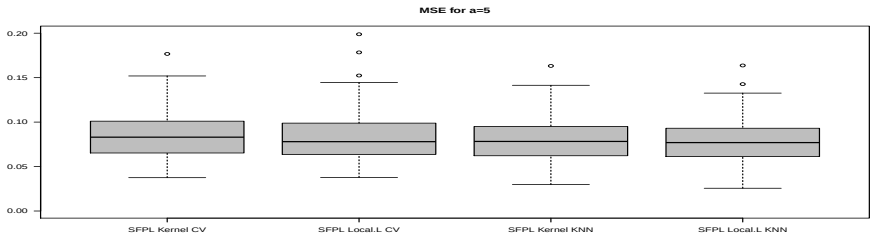


Figure 4. Predictions of the model (4.1) with different values of a.

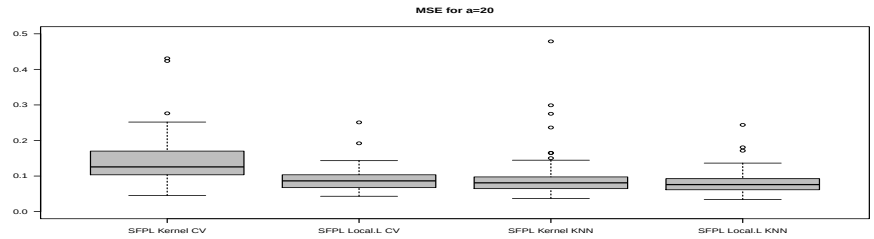
Figure 5 displays boxplots constructed from the MSE obtained for the models and the different values of a.

Table 2. The values of the estimator $\widetilde{\beta}$ for the SFPLR-LL kNN model according to the sizes (n_1, n_2).

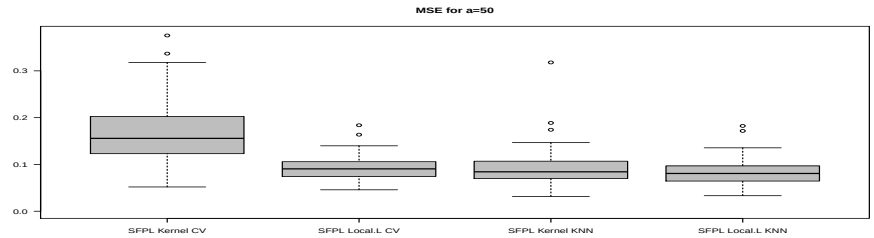
$n_2 \rightarrow$ n_1 $\downarrow$	10	25	50
10	(-0.9165934 , 3.047369)	(-1.0371385 , 2.965457)	(-0.9865957 , 3.014751)
20	(-1.0128823 , 3.014244)	(-1.0085904 , 3.011950)	(-1.0057499 , 3.008370)
50	(-0.9960193 , 3.007497)	(-0.9961910 , 2.996413)	(-1.0033046 , 2.998738)



(a) $a = 5$



(b) $a = 20$



(c) $a = 50$

Figure 5. The box plots of the MSE for $a = 5$, $a = 20$ and $a = 50$.

According to Table 2, we can observe the convergence of the estimators $\widehat{\beta}$ to the true values β as n_1 and n_2 increases.

4.2. Application

The objective of this section is to assess the effectiveness of the SFPLR model using our proposed estimators on a real data set consisting of particle pollution indices. The source of these data is the AriaWeb information system, managed by CSI Piemonte and Regione Piemonte, and our analysis is obtained from 34 monitoring sites using gravimetric instruments recorded during the winter season from October 2005 to March 2006 (daily measurements including $T = 182$ days). This analysis focuses on pollution levels, revealing higher concentrations in lowland areas near urban centers and lower concentrations near the Alps. For more detailed information about the dataset, we refer the reader to [25]. To identify relevant covariates, a preliminary regression analysis was performed, leading to the selection of the following explanatory variables:

- $X^1 = HMIX(s)$: maximum daily mixing height (in meters),
- $X^2 = EMI(s)$: daily primary aerosol emission rates (in g/s),
- $X^3 = PREC(s)$: total daily precipitation (in millimeters),
- $Z = TEMP$: the average daily temperature (in Kelvin $°K$).

Specifically, we assume that the observations follow the SFPLR model (2.4), where the response variable is $Y = PM_{10}(s)$ (in $\mu g/m^3$) (for each $s = 1, ..., 182$) represents pollution levels, the functional predictor $Z_{\mathbf{i}}(t)$ represents the daily mean temperature curve recorded at the **i**th station, with its precise location determined by the coordinates $\mathbf{i} = (UTMX; UTMY)$, $Z = TEMP(t); t = 1, ..., 182$, and the parametric part is: $X = (X^1, X^2, X^3)$, (for 182 days). Figure 6 provides the curves of the functional variable $Z_{\mathbf{i}}$.

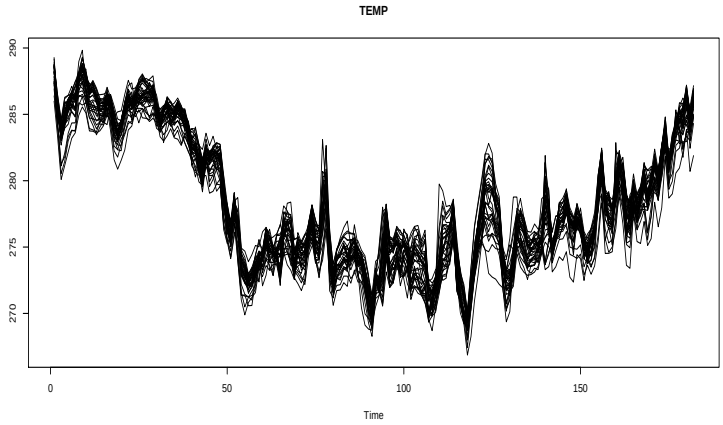


Figure 6. Temperature curves Z.

It is clear that the data exhibit a trend and are therefore non-stationary. Therefore, implementing this spatial modeling approach requires preliminary data pre-processing to ensure the stationarity assumption is met. This step is essential to address the spatial heterogeneity arising from variations in spatial effects between sampled units. To answer this, we will use an "**detrending** step" introduced by [41] which is designed for the multivariate case of the three variables (reponse, functional and vectorial explanatory). This algorithm is defined by the following regression:

$$\widetilde{X}_{\mathbf{i}} = m_1(\mathbf{i}) + X_{\mathbf{i}}, \quad \widetilde{Z}_{\mathbf{i}} = m_2(\mathbf{i}) + Z_{\mathbf{i}} \quad \text{and} \quad \widetilde{Y}_{\mathbf{i}} = m_3(\mathbf{i}) + Y_{\mathbf{i}}.$$

Thus, instead of the initial observations $(X_{\mathbf{i}}, Z_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}}$, we compute the SFPLR estimator from the statistics $(\widehat{X}_{\mathbf{i}}, \widehat{Z}_{\mathbf{i}}, \widehat{Y}_{\mathbf{i}})_{\mathbf{i}}$. The latter are obtained by

$$\widehat{X}_{\mathbf{i}} = \widetilde{X}_{\mathbf{i}} - \widehat{m}_1(\mathbf{i}), \quad \widehat{Z}_{\mathbf{i}} = \widetilde{Z}_{\mathbf{i}} - \widehat{m}_2(\mathbf{i}) \quad \text{and} \quad \widehat{Y}_{\mathbf{i}} = \widetilde{Y}_{\mathbf{i}} - \widehat{m}_3(\mathbf{i}),$$

and $\widehat{m}_1(.), \widehat{m}_2$ and $\widehat{m}_3$ are the kernel estimators of the regression functions $m_1(.), m_2(.)$ and $m_3(.)$ which are expressed by

$$\widehat{m}_1(\mathbf{i}_0) = \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} X_{\mathbf{i}} H_1\left(\|\mathbf{i}_0 - \mathbf{i}\|/h_{\mathbf{n}}^1\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} H_1\left(\|\mathbf{i}_0 - \mathbf{i}\|/h_{\mathbf{n}}^1\right)}, \widehat{m}_2(\mathbf{i}_0) = \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Z_{\mathbf{i}} H_2\left(\|\mathbf{i}_0 - \mathbf{i}\|/h_{\mathbf{n}}^2\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} H_2\left(\|\mathbf{i}_0 - \mathbf{i}\|/h_{\mathbf{n}}^2\right)}$$

$$\text{and } \widehat{m}_3(\mathbf{i}_0) = \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} H_3\left(\|\mathbf{i}_0 - \mathbf{i}\|/h_{\mathbf{n}}^3\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} H_3\left(\|\mathbf{i}_0 - \mathbf{i}\|/h_{\mathbf{n}}^3\right)},$$

where the functions H_j, $j = 1, 2, 3$ represent kernel functions, while $h_{\mathbf{n}}^j$, $j = 1, 2, 3$ are the bandwidth parameters associated with the actual regression.

We employ the same methodology as employed in the simulation example for the selection of the estimator's parameters to conduct this analysis. Specifically, we use the quadratic kernel on the interval $(0, 1)$ in combination with the PCA metric and the k-NN criterion to determine the smoothing parameter $h_{\mathbf{n}}$. For real regressions $m_1(.)$, $m_2(.)$, and $m_3(.)$, we use the `npreg` routine in the `R`-package `np`, with $K = H_1 = H_2 = H_3$. To assess the feasibility of this approach, we randomly split the data sample multiple times

(exactly 100 times). The data is divided into two subsets: a learning sample consisting of 24 observations and a test sample containing 10 observations.

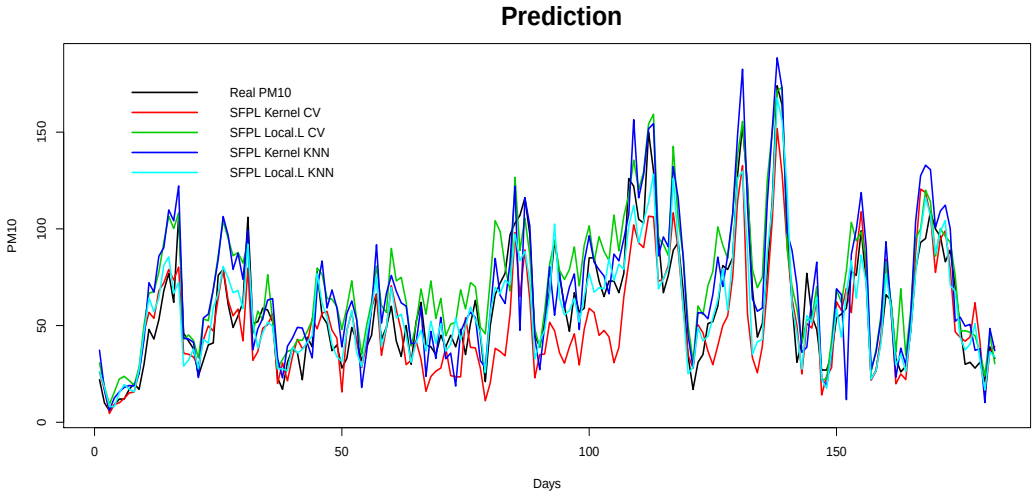


Figure 7. Prediction of the testing sample of the PM_{10} for $s = 1, ..., 182$ using the four models.

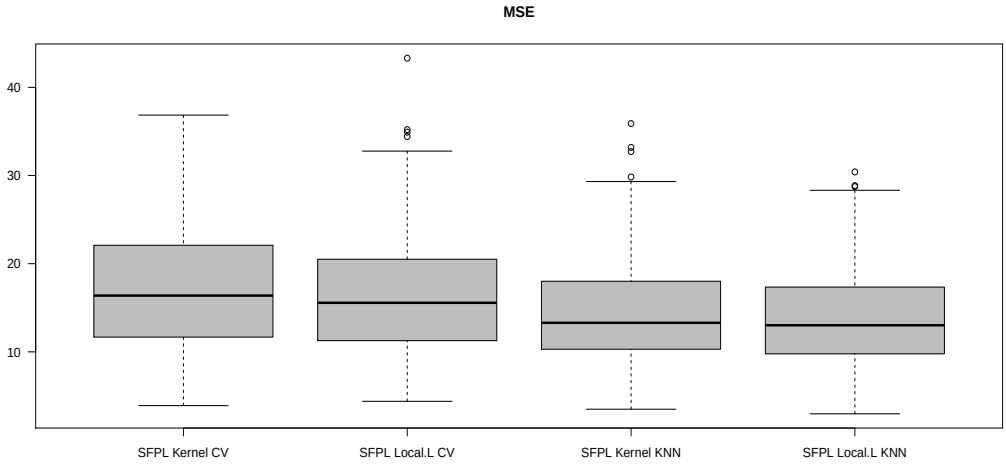


Figure 8. Boxplot of MSE values of the models.

In summary, this study applies the SFPLR model to analyze the particulate pollution indices collected from 34 monitoring stations. As shown in Figure 7, the model successfully captures both spatial heterogeneity and functional relationships between pollution levels and explanatory variables, demonstrating its practical relevance. Furthermore, the use of the local linear kNN method leads to a substantial improvement in prediction accuracy. This finding is supported by the reduction in mean square error (MSE), which is consistent with the results obtained in the simulation study. To evaluate the effectiveness of the proposed detrending procedure, we examine its impact on MSE values. The analysis (Figure 8) underscores the role of detrending in enhancing the practical performance of the estimators. Notably, the MSE is significantly lower when the local linear kNN approach is combined with detrending, highlighting the benefit of incorporating this preprocessing step.

5. Conclusion

This contribution addresses the spatial semi-functional partial linear regression (SFPLR) model, which integrates the strengths of partial linear regression and the flexibility of functional data analysis, while explicitly accounting for spatial dependence. Such models are particularly useful in applications where understanding complex relationships between variables across space or time is essential. The main contribution of this work is the integration of local linear estimation (LLE) with the k-nearest neighbor (kNN) smoothing method to develop novel estimators for the spatial SFPLR model. This hybrid approach effectively tackles the common issue of bandwidth selection in nonparametric estimation and results in estimators with reduced bias. We then establish the asymptotic distribution for the parametric component, as well as the uniform almost complete convergence rate for the nonparametric component. Finally, we assess the finite-sample performance of the proposed estimators through simulations and real data analysis, comparing them with existing methods for SFPLR models, including the kernel estimator without kNN smoothing, kernel estimator with kNN smoothing and the local linear estimator (LLE) without kNN smoothing. The results clearly demonstrate that the LLE-kNN estimator outperforms its competitors. This superiority is evidenced by the lowest mean squared error (MSE) obtained across both simulated and real datasets.

As mentioned in the Introduction, the key feature of the proposed estimator lies in its construction, which combines two major non-parametric approaches: the local linear method and kNN smoothing. In addition to its strong theoretical properties, the estimator proves to be highly practical, fast, robust, and more accurate than competing alternatives. The advantages of the LLE-kNN approach are two-fold. First, although the convergence rate of the proposed estimator is aligned with that of existing methods in the SFPLR framework, it significantly improves the bias component. The use of the local linear method not only reduces computational cost, but also enhances implementation efficiency, leading to substantial gains in predictive performance. Second, the incorporation of kNN smoothing offers an elegant and effective solution to the complex problem of bandwidth selection, a long-standing challenge in nonparametric statistics. Naturally, a remaining difficulty lies in determining an appropriate rule for selecting the optimal smoothing parameter and identifying the relevant subset for optimization. However, by reformulating the problem in terms of selecting an integer $k \in \{1, \ldots, n\}$, the kNN approach simplifies this task while maintaining high performance. To the best of our knowledge, this is the first study to propose a location-adaptive semiparametric framework tailored specifically to functional data that exhibit spatial dependence.

As a direction for future research, our results could be compared with those obtained using the semi-functional partial linear spatial autoregressive (SFPSAR) model, as proposed in [64]. In their approach, the quasi-maximum likelihood estimation is employed for estimating the parametric component, while the local linear estimation method is used for the nonparametric component. Such a comparison would provide further insight into the strengths and limitations of each modeling strategy.

Acknowledgements

The authors are indebted to the Editor-in-Chief and the referees for their very valuable comments and suggestions which led to a considerable improvement of the manuscript.

Author contributions. The authors contributed approximately equally to this work. All authors have read and agreed to the final version of the manuscript. Formal analysis, M.O. Baouche; Validation, O. Fetitah; Writing review & editing, T. Benchikh and T. Guendouzi.

Conflict of interest statement. The authors declare no conflict of interest.

Funding. This research was funded by Thematic Research Agency in Science and Technology (ATRST) for funding this work through research groups program under the project number PRFU, C00L03UN220120220002.

Data availability. The data used in this study are available through the link https://www.csipiemonte.it/en

References

[1] M. Abeidallah, B. Mechab and T. Merouan, *Local linear estimate of the point at high risk: spatial functional data case*, Commun. Stat. Theory Methods **49**, 25612584, 2020.

[2] M.S. Ahmed, M. N'diaye, M. Kadi Attouch and S. Dabo-Niange, *k-nearest neighbors prediction and classification for spatial data*, J. Spat. Econ. **4** (12), 2023.

[3] I. M. Almanjahie, K. A. Assiri, A. Laksaci and Z. Chikr Elmezouar, *The k nearest neighbors smoothing of the relative-error regression with functional regressor*, Commun. Stat. Theory Methods **51**, 41964209, 2022.

[4] I.M Almanjahie, W.M. Alahmari and A. Laksaci, *The k nearest neighbors local linear estimator of functional conditional density when there are missing data*. Hacet. J. Math. Stat. **51** (3), 914-931, 2022.

[5] I.M. Almanjahie, Z. Chikr-Elmezouar, A. Laksaci and M. Rachdi, *kNN local linear estimation of the conditional cumulative distribution function: Dependent functional data case*. C. R. Acad. Sci. Paris, Ser. I **356**, 10361039, 2018.

[6] I. M. Almanjahie, W. Mesfer, A. Laksaci and M. Rachdi *Computational aspects of the k NN local linear smoothing for some conditional models in high dimensional statistics*, Commun. Stat. Simul. Comput. **52** (7), 2985-3005, 2023.

[7] F. Alshahrani, I.M. Almanjahi, T. Benchikh, O. Fetitah and M.K. Attouch, *Asymptotic normality of nonparametric kernel regression estimation for missing at random functional spatial data*, Journal of Mathematics, 2023, https://doi.org/10.1155/2023/8874880.

[8] F. Alshahrani, W. Bouabsa, I.M. Almanjahie and M.K. Attouch *kNN local linear estimation of the conditional density and mode for functional spatial high dimensional data*. AIMS Math. **8** (7), 1584415875, 2023.

[9] G. Aneiros-Pérez, I. Horová, M. Hušková and P. Vieu, *Editorial for the Special Issue on Functional Data Analysis and Related Fields*, J. Multivariate Anal. **189**, 2022. M

[10] G. Aneiros-Pérez, R. Cao and P. Vieu, *Editorial on the special issue on functional data analysis and related topics. Computational Statistics*, **34**, 447-450, 2019.

[11] G. Aneiros-Pérez G. and P. Vieu, *Nonparametric time series prediction. A semi-functional partial linear modeling*. J. Multivariate Anal. **99**, 834-857, 2008.

[12] G. Aneiros-Pérez G. and P. Vieu, *Semi-functional partial linear regression*, Stat. Probab. Lett. **76** (11), 1102-1110, 2006.

[13] M.K Attouch M and T. Benchikh, *Asymptotic distribution of robust k-nearest neighbour estimator for functional nonparametric models*, Mat. Vesnik **644**, 275-285, 2012.

[14] M.K. Attouch, B. Chouaf and A. Laksaci, *Nonparametric M-estimation for functional spatial data*, Commun. Stat. Appl. Methods **19**, 193-211, 2012.

[15] M.K. Attouch, A. Gheriballah and A. Laksaci, *Robust nonparametric estimation for functional spatial regression*. In F. Ferraty editor, Recent Advances in Functional Data Analysis and Related Topics, Contrib. Stat. 27-31. Physica-Verlag HD, 2011.

[16] M.K. Attouch, A. Laksaci and F. Rafaa, *Estimation locale linéaire de la régression non paramétrique fonctionnelle par la méthode des k plus proches voisins*. Comptes Rendus. Mathématique **355** (7), 824-829, 2017.

[17] A. Baíllo and A. Grané, *Local linear regression for functional predictor and scalar response*, J. Multivariate Anal. **100** (1), 102-111, 2009.
[18] J. Barrientos-Marin J, F. Ferraty and P. Vieu, *Locally modelled regression and functional data*. J Nonparametr Stat. **22** (5-6), 617-632, 2009.
[19] F. Belarbi, S. Chemikh and A. Laksaci, *Local linear estimate of the nonparametric robust regression in functional data*, Stat. Probabil. Lett. **134**, 128133, 2018.
[20] M. Benallou, M.K. Attouch, T. Benchikh and O. Fetitah, *Asymptotic results of semi-functional partial linear regression estimate under functional spatial dependency*. Commun. Stat. - Theory Methods **51** (20), 7172-7192, 2021.
[21] T. Benchikh, I.M. Almanjahie, O. Fetitah and M.K. Attouch, *Estimation for spatial semi-functional partial linear regression model with missing response at random*, Demonstr. Math. **58**, 2025, doi:10.1515/dema-2025-0108.
[22] G. Biau, F. Cérou, and A. Guyader. *Rates of convergence of the functional k-Nearest Neighbor estimate*. IEEE Trans. Inf. Theory **56** (4), 2034-2040, 2010.
[23] G. Boente and A. Vahnovan, *Robust estimators in semi-functional partial linear regression models*. J. Multivariate Anal. **154** (C), 59-84, 2017.
[24] F. Burba, F. Ferraty and P. Vieu, *k-Nearest Neighbour method in functional nonparametric regression*. J. Nonparametr. Stat. **21** (4), 453-469, 2009.
[25] M. Cameletti, R. Ignaccolo and S. Bande, *Comparing spatio-temporal models for particulate matter in Piemonte*. Environmetrise **22** (8), 985-996, 2011.
[26] M. Carbon, L.T. Tran and B. Wu, *Kernel density estimation for random fields (density estimation for random fields)*. Stat Probab Lett. **36** (2), 115-125, 1997.
[27] A. Chahad, L. Ait-Hennani, A. Laksaci, *Functional local linear estimate for functional relative-error regression*, J. Stat. Theory Pract. **11**, 771789, 2017.
[28] Z. Chikr-Elmezouar, I.M. Almanjahie, A. Laksaci and M. Rachdi M. *FDA: strong consistency of the k nn local linear estimation of the functional conditional density and mode*. J. Nonparametric Stat. **31**, 175195, 2019.
[29] A. Chouaf and A. Laksaci, *On the functional local linear estimate for spatial regression*, Stat. Risk Model **29**, 189-214, 2013.
[30] S. Dabo-Niang, M. Rachdi M. and A.F. Yao, *Kernel regression estimation for spatial functional random variables*, Far East J. Theor. Stat. **37** (2), 77-113, 2011.
[31] J. Demongeot, A. Naceri, A. Laksaci and M. Rachdi, *Local linear regression modelization when all variables are curves*. Statist. Probab. Lett. **121**, 37-44, 2017.
[32] H. Ding, Z. Lu, J. Zhang and R.n Zhang, *Semi-functional partial linear quantile regression*, Stat Probab Lett. **142**, 92-101, 2018.
[33] P. Doukhan, *Mixing Properties and Examples. In: Lecture Notes in Statistics*, **85**, Springer-Verlag, New York, 1994.
[34] J. Fan and I. Gijbels. *Local polynomial modelling and its applications*, London: Chapman and Hall, 1996.
[35] S. Feng and L. Xue, *Partially functional linear varying coefficient model*. Statistics **50** (4), 717-732, 2016.
[36] Ferraty F, Laksaci A, Tadj A, Vieu P. *Rates of uniform consistency for nonparametric estimates with functional variables*. J Stat Plan Infer **140** (2), 335352, 2010.
[37] F. Ferraty and P. Vieu, *Nonparametric functional data analysis. Theory and Practice*. Springer Series in Statistics. New York, 2006.
[38] J. T. Gao and, Z. Lu and D. Tjφstheim D., *Estimation in semiparametric spatial regression*. Ann. Stat. **34** (3), 1395-1435, 2066.
[39] S. Greven and F. Scheipl, *A general framework for functional regression modelling*. Stat. Model. **17** (1-2), 1-35, 2017.
[40] X. Guyon, *Random Fields on a Network-Modeling, Statistics and Applications*, Springer, New-York, 1995.

[41] M. Hallin, Z. Lu and K. Yu, *Local Linear Spatial Quantile Regression*, Bernoulli **15** (3), 659-686, 2009.
[42] T. Hsing and R.L. Eubank, *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, John Wiley and Sons, 2015.
[43] L. Kara-Zaitri, A. Laksaci, M. Rachdi and P. Vieu, *Uniform in bandwidth consistency for various kernel estimators involving functional data*, J. Nonparametr. Stat. **29** (1), 85-107, 2017.
[44] N. Kudraszow and P. Vieu P, *Uniform consistency of kNN regressors for functional variables*, Statist. Probab. Lett. **83** (8), 1863-1870, 2013.
[45] N.H. Kedir, T. Benchikh, A. Naceri and O. Fetitah, *Local linear-kNN smoothing for semi-functional partial linear regression.* Hacet. J. Math. Stat. **53** (2):537-555, 2024.
[46] A. Laksaci, M. Rachdi, S. Rahmani, *Spatial modelization: local linear estimation of the conditional distribution for functional data*, Spatial Stat. **6**, 123, 2013.
[47] T. Laloë, *A k-nearest neighbor approach for functional regression.* Stat. Probab. Lett. **78**, 11891193, 2008.
[48] N. Ling N, G. Aneiros-Pérez and P. Vieu, *knn estimation in functional partial linear modeling.* Statist. Papers **61** (1), 423-444, 2020.
[49] N. Ling, R. Kan, P. Vieu and S. Meng, *Semi-functional partially linear regression model with responses missing at random.* Metrika **82** (1), 39-70, 2019.
[50] N. Ling, S. Meng and P. Vieu, *Uniform consistency rate of kNN regression estimation for functional time series data*, J. Nonparametr. Stat. **31** (2),451-468, 2019.
[51] N. Ling and P. Vieu, *Nonparametric modelling for functional data: selected survey and tracks for future.* Statistics **52** (4), 934-949, 2018.
[52] N. Ling and P. Vieu, *On semiparametric regression in functional data analysis.* WIRES Comput. Stat. **12** (6), 20-30, 2020.
[53] J. Mateu and E. Romano, *Advances in spatial functional statistics.* Stoch. Environ. Res. Risk. Assess. *31*, 1-6, 2017.
[54] J. Mateu, R. Giraldo, *Geostatistical Functional Data Analysis*, Wiley Series in Probability and Statistics, 1st Edition, 2021.
[55] A. Naceri, A. Laksaci and M. Rachdi, *Exact quadratic error of the local linear regression operator estimator for functional covariates.* In Functional statistics and applications, 79-90, Springer Cham Heidelberg, New York, 2019.
[56] M. Ndiaye, S. Dabo-Niang, P. Ngom, *Nonparametric Prediction for Spatial Dependent Functional Data Under Fixed Sampling Design*, Rev. Colomb. Estad. **45** (2), 391-428, 2022.
[57] S. Novo, G. Aneiros, and P. Vieu. *A kNN procedure in semiparametric functional data analysis*, Statist. Probab. Lett. **171**, 2021.
[58] M. Rachdi M. *Functional Data Analysis: Theory and Applications to Different Scenarios.* Mathematics, an Open Access Journal by MDPI 2023. https://www.mdpi.com/journal/mathematics/special issues/45POZ9BG9S.
[59] M. Rachdi, A. Laksaci, K. Kaid, A. Benchiha A and F. Al-Awadh, *k-Nearest neighbors local linear regression for functional and missing data at random.* Stat. Neerl. **75** (1), 42-65, 2021.
[60] M. Rachdi, A. Laksaci and N. M. Al-Kandari, *Expectile regression for spatial functional data analysis (sFDA)*, Metrika **85**, 627655, 2022.
[61] J. Ramsay and B. Silverman, *Functional Data Analysis*, Second Edition, Spinger-Verlag, New York, 2005.
[62] A. Saadaoui, F Benaissa and A. Chouaf, *On the local linear estimation of a generalized regression function with spatial functional data*, Commun. Stat. - Theory Methods **52** (21), 2023.
[63] J. L. Wang, J. M. Chiou, and H. G. Müller, *Functional data analysis.* Annu. Rev. Stat. Appl. **3** (1), 257-295, 2016.

[64] Y. Li and C. Ying, *Semi-functional partial linear spatial autoregressive model*, Commun. Stat. - Theory Methods **50** (24), 2021.
[65] Y. Hu, S. Wu and S. Feng, *Estimation in functional partially linear spatial autoregressive model*, Hacet. J. Math. Stat. **53** (4), 1196 1217, 2024.

APPENDIX

Proofs section

The proof of asymptotic results are given briefly because they follow from the same ideas as in [11, 20].

Proof of Theorem 3.1 The proof of Theorem 1 is established on the following decomposition

$$\begin{aligned}(\widehat{\mathbf{n}})^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\mathbf{n}}-\boldsymbol{\beta}\right) &= \left(\frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\widetilde{X}_{\mathbf{i}}(\widetilde{X}_{\mathbf{i}})^{T}\right)^{-1}\frac{1}{\sqrt{\widehat{\mathbf{n}}}}\left(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}R_{\mathbf{i}}+\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\theta_{\mathbf{i}}\left(\Delta_{\mathbf{i}}^{(0)}-\Delta_{\mathbf{i}}^{T}\boldsymbol{\beta}\right)\right.\\ &\left.+\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\Delta_{\mathbf{i}}\varepsilon_{\mathbf{i}}+\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\Delta_{\mathbf{i}}\left(\Delta_{\mathbf{i}}^{(0)}-\Delta_{\mathbf{i}}^{T}\boldsymbol{\beta}\right)\right)\end{aligned} \tag{5.1}$$

where

$$\Delta^{(s)}(z)=\mathbb{E}[X_{\mathbf{i}}^{(s)}|Z_{\mathbf{i}}=z]-\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\mathbf{W}_{j}(Z_{\mathbf{l}})X_{\mathbf{j}}^{(s)}\,,s=1,\cdots,p\,.$$

$$\Delta_{\mathbf{i}}^{(0)}=\mathbb{E}[Y_{\mathbf{i}}|Z_{\mathbf{i}}]-\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)Y_{\mathbf{j}}\,,$$

with

$$\Delta_{\mathbf{i}}=(\Delta_{\mathbf{i}}^{(1)},\cdots,\Delta_{\mathbf{i}}^{(p)})^{T}\,.$$

Thus, is an immediate consequence of the Cauchy-Schwartz inequality, central limit theorem and the following Lemmas

Lemma 5.1. *Under assumptions (A1)-(A8), we have*

$$\sup_{Z\in\mathcal{S}}\left|\Delta^{(s)}(z)\right|=O\left(\varphi^{-1}\left(\frac{k_{\mathbf{n}}}{\widehat{\mathbf{n}}}\right)\right)+O_{a.co}\left(\sqrt{\frac{\log d_{\mathbf{n}}}{k_{\mathbf{n}}}}\right). \tag{5.2}$$

$$\sup_{Z\in\mathcal{S}}\left|\Delta_{\mathbf{i}}^{(0)}\right|=O\left(\varphi^{-1}\left(\frac{k_{\mathbf{n}}}{\widehat{\mathbf{n}}}\right)\right)+O_{a.co}\left(\sqrt{\frac{\log d_{\mathbf{n}}}{k_{\mathbf{n}}}}\right). \tag{5.3}$$

Lemma 5.2. *Under the conditions (A1)-(A10), we have*

$$\frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\widetilde{X}_{\mathbf{i}}(\widetilde{X}_{\mathbf{i}})^{T}\longrightarrow\boldsymbol{\Sigma}\ a.s. \tag{5.4}$$

Remark 5.3. The lemma 5.1 extends Theorem 4.1 of [62] using the kNN-LL approach. The proof of this theorem is the same as that used in this reference. It is not given here.

Proof of Theorem 3.2
From the fact that

$$\begin{aligned}\widehat{g}(Z) &= \textstyle\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)Y_{\mathbf{j}}-\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)\mathbf{X}_{\mathbf{j}}^{T}\widetilde{\boldsymbol{\beta}}_{n}\\ &= \textstyle\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)\left(g(Z_{\mathbf{j}})+\varepsilon_{\mathbf{j}}\right)-\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}}\mathbf{W}_{\mathbf{j}}(Z)\mathbf{X}_{\mathbf{j}}^{T}\left(\widetilde{\boldsymbol{\beta}}_{n}-\boldsymbol{\beta}_{n}\right),\end{aligned} \tag{5.5}$$

we have

$$\begin{aligned}\sup_{Z\in\mathcal{S}}|\widehat{g}(Z)-g(Z)| &\leq \sup_{Z\in\mathcal{S}}|\textstyle\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}\mathbf{W_j}(Z)\left(g(Z_{\mathbf{j}})+\varepsilon_{\mathbf{j}}\right)-g(Z)|+\\ &\quad \sup_{Z\in\mathcal{S}}|\textstyle\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{n}}}\mathbf{W_j}(Z)\mathbf{X_j}^T|\|\widehat{\beta}_n-\beta)\| \\ &\leq \sup_{Z\in\mathcal{S}}|S_1|+\sup_{Z\in\mathcal{S}}|S_2|\,.\end{aligned} \tag{5.6}$$

From Lemma 5.1, we have

$$\sup_{Z\in\mathcal{S}}|S_1| = O\left(\varphi^{-1}\left(\frac{k_{\mathbf{n}}}{\widehat{\mathbf{n}}}\right)\right)+O_{a.co}\left(\sqrt{\frac{\log d_{\mathbf{n}}}{k_{\mathbf{n}}}}\right). \tag{5.7}$$

On the other side, we have

$$\begin{aligned}\sup_{Z\in\mathcal{S}}|S_2| &\leq \sup_{Z\in\mathcal{S}}|\textstyle\sum_{j=1}^{n}\mathbf{W_j}(\xi_l)\mathbf{X_j}|\|\widehat{\beta}_n-\beta\|\\ &\leq \sup_{Z\in\mathcal{S}}|\textstyle\sum_{j=1}^{n}\mathbf{W_j}(Z_{\mathbf{l}})\left(\mathbf{X_j}\right)-\mathbb{E}\left(\mathbf{X}_l/Z_{\mathbf{l}}\right)|\,|\widehat{\beta}_{\mathbf{n}}-\beta|+\\ &\quad \sup_{Z\in\mathcal{S}}|\mathbb{E}(\mathbf{X}_l/Z_{\mathbf{l}})|\,||\widehat{\beta}_{\mathbf{n}}-\beta||\,.\end{aligned}$$

Then, from Theorem 3.1, we have $||\widehat{\beta}_n-\beta||\to 0$ and according to the fact that $\sup_{Z\in\mathcal{S}}|\mathbb{E}(\mathbf{X}_l/\xi_l)|<\infty$, 5.1 implies that

$$\sup_{Z\in\mathcal{S}}|S_2| = O\left(\varphi^{-1}\left(\frac{k_{\mathbf{n}}}{\widehat{\mathbf{n}}}\right)\right)+O_{a.co}\left(\sqrt{\frac{\log d_{\mathbf{n}}}{k_{\mathbf{n}}}}\right). \tag{5.8}$$

So by using Equations 5.6, 5.7 and 5.8, we have

$$\sup_{Z\in\mathcal{S}}|\widehat{m}(\xi)-m(\xi)| = O\left(\varphi^{-1}\left(\frac{k_{\mathbf{n}}}{\widehat{\mathbf{n}}}\right)\right)+O_{a.co}\left(\sqrt{\frac{\log d_{\mathbf{n}}}{k_{\mathbf{n}}}}\right).$$