Hindawi
Journal of Mathematics
Volume 2023, Article ID 8874880, 20 pages
https://doi.org/10.1155/2023/8874880

Research Article

Asymptotic Normality of Nonparametric Kernel Regression Estimation for Missing at Random Functional Spatial Data

Fatimah Alshahrani,[1] **Ibrahim M. Almanjahie**,[2] **Tawfik Benchikh**,[3] **Omar Fetitah**,[3] **and Mohammed Kadi Attouch**[3]

[1]*Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia*
[2]*Department of Mathematics, College of Science, King Khalid University, Abha 62223, Saudi Arabia*
[3]*Statistical and Stochastic Process Laboratory, Djillali Liabes University, Sidi Bel Abbes, Algeria*

Correspondence should be addressed to Mohammed Kadi Attouch; attou_kadi@yahoo.fr

Received 11 May 2023; Revised 25 August 2023; Accepted 28 August 2023; Published 13 September 2023

Academic Editor: Jun Fan

This study investigates the estimation of the regression function using the kernel method in the presence of missing at random responses, assuming spatial dependence, and complete observation of the functional regressor. We construct the asymptotic properties of the established estimator and derive the probability convergence (with rates) as well as the asymptotic normality of the estimator under certain weak conditions. Simulation studies are then presented to examine and show the performance of our proposed estimator. This is followed by examining a real data set to illustrate the suggested estimator's efficacy and demonstrate its superiority. The results show that the proposed estimator outperforms existing estimators as the number of missing at random data increases.

1. Introduction

In several domains of current research, including environmental sciences, geography, econometrics, microbiology, geophysics, climates, and other applied fields, the analysis of massive volumes of data with a spatial argument is frequently required (geographical location). To describe these processes, you need to find the relationship between random variables in one area, in terms of correlation, and those in nearby areas. This step is considered one of the most essential parts of analyzing spatial data. Recently, the new statistical branch, called functional data analysis (FDA), gave a new dynamism to theoretical and methodological improvements and the diversification of application domains. Such improvements have been possible as computer tools' storage capabilities have increased, allowing them to store and analyze large amounts of data. We mention the monographs [1] for the practical aspects, reference [2] for the theoretical elements, and reference [3] for a nonparametric study as reference works on the issue. For the most recent contributions in this area, readers can consult the book [4] as well as several bibliographic reviews in [5, 6]. In this context, functional regression is an essential component of the FDA because it links its regressor X to the scalar variable Y. The authors in [7] discussed and determined the initial findings for estimating the regression function (in semimetric space). It should be clear to the engaging readers that such topic theories and methods in this field of study are well-established; for examples, see monograph [3] and the references included, as well as [8, 9].

Incorporating spatial statistics with functional data analysis is made possible by functional data coupled with geographical dependencies and spatial functional statistics. This combination extends the FDA method to analyze a sample of functions obtained at different regional sites (functional data with spatial correlation). Both the theoretical and practical aspects of statistics stand to benefit from this combination (for some recent, advanced, and noteworthy citations on the topic, see [10]). The model of spatial functional regression is considered and explored by [11]. The authors constructed the rates of almost sure convergence using the nonparametric kernel method in functional regression. Then, the authors in [12] determined the asymptotic normality of robust regression simultaneously. We take note that the spatial functional regression is a particular case of two widely recognized spatial dependence models that have garnered significant interest in the analysis of lattice data, known as the spatial autoregressive (SAR)-dependent variable and the spatial autoregressive error (SAE), where the model error is SAR. These models extend the concept of regression from a time series framework to a spatial context (see [13] for a large discussion).

Complete data analysis is also the topic of all the work listed. Unfortunately, this topic has not received better attention in many applications, such as the analysis of survival statistics. Specifically, the problem lies in finding the best way to replace missing data and control the accuracy of such an imputation; this topic has received extensive studies and treatment in the multivariate case (see, for example, [14–16]). Imputation techniques for missing responses commonly used involve kernel regression imputation, linear regression imputation, and so on. There are many studies on regression functions with missing data and related statistical conclusions in the statistical literature when the predictor variables are finite-dimensional. For parametric regression, we quote [17, 18], and for nonparametric regression with a kernel, we cite [19, 20]. Also, references [21–23] examined the case in which some observations on the covariates are missing at random (MAR), whereas the observations on the scalar response are entirely observed. Thus, reference [24] examined the presence of missing data in the robust regression model while the author in [25] studied MAR regression using the response variable and predictors (covariates).

Very little research has been conducted on the properties of the functional nonparametric regression model for the missing data when the predictors are functional. It was first suggested by the authors in [8], who estimated the mean of a MAR scalar response using an i.i.d. functional sample and observed predictors. They extended the result in [19] and demonstrated the asymptotic characteristics of the regression operator estimate when the functional regressor is totally observed and some responses are randomly missed. Later, the authors in [26] established the asymptotic properties of the regression function, considering cases when the explanatory variables are functional, stationary, and ergodic with MAR response. The local linear estimation method and the k-nearest neighbor (k-NN) technique were used by the authors in [27] for estimating the regression function when the regressor and response variables are functional and scalar, respectively. Still, the latter observed fewer MARs, while the authors in [6] constructed the nonparametric quantile-regression estimate for the functional data with MAR response. The authors in [28] suggest and compare different methods for estimating spatial autoregressive panel models with randomly missing data in the dependent variable.

As far as we know, no previous studies have been conducted on nonparametric regression based on functional spatial data with MAR response. Hence, our goal is to investigate the kernel method to estimate the regression function based on spatially dependent data, and the response is MAR.

We structure this paper by introducing the considered spatial model in Section 2 (as in (1)) and explicitly generating the estimate of $m(.)$ utilizing MAR. We outline, in Section 3, the notations and several assumptions behind the considered model. Section 4 shows the main theoretical result of our study. The strong proof supporting our findings is presented in Section 5, where the latter involves the evaluation of our method using both simulation and real data application. It includes a comparison between the typical spatial nonparametric functional model and its incomplete counterpart, demonstrating the superiority of our method. Finally, our conclusion is stated at the end.

2. The Estimates and the Spatial Model

Denote by $\{Z_{\mathbf{i}} = (X_{\mathbf{i}}; Y_{\mathbf{i}}); \mathbf{i} \in \mathbb{Z}^N, N \geq 1\}$ a measurable strictly stationary spatial process defined over a probability space $(\Omega, \mathscr{A}, \mathbf{P})$ with identical distribution as $Z = (X, Y)$, where X is a functional random variable valued in a separable semimetric space $(\mathscr{E}, d(.,.))$ and Y is a real-valued and integrable variable. We suppose that the process can be observed in the rectangular region $\mathscr{I}_{\mathbf{n}} = \{\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{Z}^N, 1 \leq i_k \leq n_k, k = 1 \ldots, N\}$, with a sample size of $\widehat{\mathbf{n}} = n_1 \times \cdots \times n_N$ where $\mathbf{n} = (n_1, \ldots, n_N)$. Suppose moreover that, for $l = 1 \ldots, N$, n_l approaches infinity at the same rate: $C_1 < |n_j/n_k| < C_2$ for some $0 < C_1 < C_2 < \infty$. The term "site" will be used to refer to a point $\mathbf{i}$. If $\min_{k=1\ldots,N}(n_k) \longrightarrow \infty$, we shall write $\mathbf{n} \longrightarrow \infty$.

The nonparametric spatial regression model is as follows:

$$Y_{\mathbf{i}} = m(X_{\mathbf{i}}) + \varepsilon_{\mathbf{i}}, \quad \mathbf{i} \in \mathbb{Z}^N, \tag{1}$$

where the function $m(.)$ is an unknown and the random errors $\varepsilon_{\mathbf{i}}$ are centered independent and identically distributed with $\mathbb{E}(\varepsilon_{\mathbf{i}} \mid X_{\mathbf{i}}) = 0$ and unknown finite variance $\sigma^2 = \mathrm{var}(\varepsilon_{\mathbf{i}})$.

We know that (see [11]) the spatial kernel regression estimator of $m(.) = \mathbb{E}(Y_{\mathbf{i}} \mid X_{\mathbf{i}})$ is obtained as

$$\widehat{m}_{\mathbf{n}}(x) = \frac{\sum_{\mathbf{i} \in \mathscr{I}_{\mathbf{n}}} Y_{\mathbf{i}} K_{\mathbf{i}}}{\sum_{\mathbf{i} \in \mathscr{I}_{\mathbf{n}}} K_{\mathbf{i}}}, \quad x \in \mathscr{E}, \tag{2}$$

where $K_{\mathbf{i}} = K(a_{\mathbf{n}}^{-1} d(x, X_{\mathbf{i}}))$ with K as the kernel function and $a_{\mathbf{n}}$ as a sequence of decreasing bandwidths as $\mathbf{n}$ approaches infinity.

Our contribution is distinguished by the fact that we tackle the issue of incomplete data. In particular, we examine the situation when the response observations (Y values) are MAR, but the independent variable (X) values are all observed. If $Y_{\mathbf{i}}$ does not contain all of the required elements, then we say that something is missing. For simplicity, we refer to δ as a real random variable and take into account the sample $\delta_{\mathbf{i}}$ where $\delta_{\mathbf{i}} = 1$ if the value $Y_{\mathbf{i}}$ is known; otherwise, $\delta_{\mathbf{i}} = 0$. Thus, we receive the following missing information:

$$\{(X_{\mathbf{i}}, Y_{\mathbf{i}}, \delta_{\mathbf{i}}), \mathbf{i} \in \mathcal{I}_{\mathbf{n}}\}. \tag{3}$$

The conditional probability $\pi(x)$ of the observable response Y given the explanatory X is typically unknown; hence, it is assumed that the random variable $\delta_{\mathbf{i}}$ follows Bernoulli distribution where

$$\mathbb{P}\left(\delta_{\mathbf{i}} = 1 \,\middle|\, X_{\mathbf{i}} = x, Y_{\mathbf{i}} = y\right) = \mathbb{P}(\delta = 1 \mid X = x) = \pi(x). \tag{4}$$

Under this assumption, we provide the estimator of $m(.)$ using the sample $\{(X_{\mathbf{i}}, Y_{\mathbf{i}}, \delta_{\mathbf{i}}), \mathbf{i} \in \mathcal{I}_{\mathbf{n}}\}$, by

$$\widetilde{m}_{\mathbf{n}}(x) = \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \delta_{\mathbf{i}} Y_{\mathbf{i}} K_{\mathbf{i}}}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \delta_{\mathbf{i}} K_{\mathbf{i}}} = \frac{\widetilde{f}_{\mathbf{n}}(x)}{\widetilde{g}_{\mathbf{n}}(x)}, \tag{5}$$

with

$$\begin{aligned} \widetilde{f}_{\mathbf{n}}(x) &= \frac{1}{\widehat{\mathbf{n}}\mathbb{E}\left(K\left(h_{\mathbf{n}}^{-1} d(x, X)\right)\right)} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \delta_{\mathbf{i}} Y_{\mathbf{i}} K_{\mathbf{i}}, \quad x \in \mathcal{E}, \\ \widetilde{g}_{\mathbf{n}}(x) &= \frac{1}{\widehat{\mathbf{n}}\mathbb{E}\left(K\left(h_{\mathbf{n}}^{-1} d(x, X)\right)\right)} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \delta_{\mathbf{i}} K_{\mathbf{i}}, \quad x \in \mathcal{E}, \end{aligned} \tag{6}$$

where $K_{\mathbf{i}} = K(h_{\mathbf{n}}^{-1} d(x, X_{\mathbf{i}}))$ is the kernel function and the bandwidths $h_{\mathbf{n}}$ are a series that tends to zero as $\mathbf{n}$ approaches infinity.

Recall that our primary objective is to investigate the asymptotic normality of our estimator when the process $Z_{\mathbf{i}}$ is strictly stationary, which satisfies the following α-mixing condition.

There exists $\varphi(t)$ a real function that tends to 0 as t goes to ∞, such that for finite cardinals subsets $E, E' \subset \mathbb{Z}^N$:

$$\begin{aligned} \alpha\left(\mathfrak{B}(E), \mathfrak{B}\left(E'\right)\right) &= \sup_{\left\{(A,B)\in\mathfrak{B}(E)\times\mathfrak{B}\left(E'\right)\right\}} \{|P(A\cap B) - P(A)P(B)|\} \\ &\le \varphi\left(d'\left(E, E'\right)\right)\psi\left(\operatorname{Card}(E), \operatorname{Card}\left(E'\right)\right), \end{aligned} \tag{7}$$

where $d'(E, E')$ represents the distance in Euclidean terms between E and E', Card (E) (resp., Card (E')) is the cardinality of E (resp., E'), $\mathfrak{B}(S) = \mathfrak{B}(Z_{\mathbf{i}}, \mathbf{i} \in S \subset \mathbb{Z}^N)$, is the σ-fields that are generated by the random variables $Z_{\mathbf{i}}$, and $\psi: \mathbb{Z}^2 \longrightarrow \mathbb{R}^+$ is a symmetric nondecreasing function. We assume that the two functions ψ and ϕ satisfy the following conditions:

$$\forall (r, s) \in \mathbb{Z}^2, \psi(r, s) \le C \min(r, s), \quad \text{for some } C > 0, \tag{8}$$

and

$$\sum_{i=1}^{\infty} i^{\gamma} (\varphi(i)) < \infty, \quad \text{for some } \gamma > 0. \tag{9}$$

Note that condition (8) can be replaced by

$$\forall (r, s) \in \mathbb{Z}^2, \psi(r, s) \le C(r + s + 1)^{\lambda}, \quad \text{for some } \lambda \ge 1. \tag{10}$$

Note that many stochastic processes satisfy the mixing conditions (8) and (9) (see [29] for some examples).

3. Notations and Hypotheses

Foremost, for $x \in \mathcal{E}$, we denote $B(x, h) = \{x' \in \mathcal{E} / d(x, x') < h\}$ and $\phi_x(h) = \mathbb{P}(X \in B(x, h))$ called small ball probability. The proposed predictor's consistency outcomes are established under the following assumptions:

(H1): we suppose that $\forall \mathbf{i} \ne \mathbf{j} \in \mathbb{Z}^N$, and the probabilistic joint distribution $\nu_{\mathbf{ij}}$ of $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$ fulfills $\forall x \in \mathcal{E}$:

$$\begin{aligned} &\text{for some } 1 < a < \gamma N^{-1} \\ &\sup_{\mathbf{i}\ne\mathbf{j}} \nu_{\mathbf{ij}} = \sup_{\mathbf{i}\ne\mathbf{j}} \mathbb{P}\left[\left(X_{\mathbf{i}}, X_{\mathbf{j}}\right) \in (B(x, h) \times B(x, h))\right] \le C(\phi_x(h))^{a+1/a}. \end{aligned} \tag{11}$$

(H2): $K: \mathbb{R} \longrightarrow \mathbb{R}^+$ is assumed to be a differentiable function supported on the interval [0, 1]. Its K' derivative function exists, as well as there are two constants, C_3 and C_4, such that

$$-\infty < C_3 < K'(t) < C_4 < 0, \quad \text{for } t \in [0, 1]. \tag{12}$$

(H3): there exist constants $C > 0$, $\kappa > 0$, and $C > 0$, such that

$$|m(x_1) - m(x_2)| \le C\, d(u, v)^{\kappa}, \quad \text{for all } x_1, x_2 \in \mathcal{E}. \tag{13}$$

(H4): there exist differentiable nonnegative functions τ and f where

$$\phi_x(h) = f(h) \times \tau(x) + o(f(h)). \quad (14)$$

(H5): the bandwidth $h_{\mathbf{n}}$ is defined as follows: $h_{\mathbf{n}}$ tend to 0 as $\mathbf{n}$ tends to ∞, and for all t in the interval [0, 1], we have

$$\lim_{h_{\mathbf{n}} \longrightarrow 0} \frac{\phi_x(t h_{\mathbf{n}})}{\phi_x(h_{\mathbf{n}})} = \beta_x(t). \quad (15)$$

(H6): we assume that

(i) $\mathbb{E}[Y^l \mid X = x] < M_l(x) < \infty$, for some $l \geq 2$, with M_l denoting a continuous function
(ii) $\forall \mathbf{i} \neq \mathbf{j}$, $\mathbb{E}[Y_{\mathbf{i}} Y_{\mathbf{j}} \mid (X_{\mathbf{i}}, X_{\mathbf{j}})] < \infty$
(iii) $\exists \varepsilon > 0$, such that $\mathbb{E}|Y_{\mathbf{1}}|^{2+\varepsilon} < \infty$

(H7): let $V_2(x) = \text{var}(Y_{\mathbf{i}} \mid X_{\mathbf{i}})$ and $V_s(u) = \mathbb{E}[|Y_{\mathbf{i}} - m(x)|^s \mid X_{\mathbf{i}} = u]$, with $s > 2$. We suppose that the functions $V_2(.)$ and $V_s(.)$ are continuous functions near x, i.e.,

$$\sup_{\{u: d(x,u) \leq h\}} \left| V_k(u) - V_k(x) \right| = o(1), \quad k \geq 2 \text{ since } h \text{ tends toward } 0. \quad (16)$$

(H8): we also suppose that $\pi(.)$ is a continuous function near x, i.e.,

$$\sup_{\{u: d(x,u) \leq h\}} |\pi(u) - \pi(x)| = o(1), \quad \text{when } h \longrightarrow 0. \quad (17)$$

(H9): there exists $1 > \theta > N/\gamma$, in such a way that $\widehat{\mathbf{n}}^{(\theta-1)/(2N+1)} \leq \phi_x(h)$.

Comments on the assumptions:

(i) In this regard, our conditions are quite standard. The conditions (ϕ) are identical to those employed by the authors in [3]. The above assumptions are common analyses in nonparametric statistics for functional regression models. The assumption (H1) specifies the behavior of the joint distribution of the couple (i, j) with respect to its margin and permits us to present an explicitly asymptotic variance term (measure the local dependence of the observations). Local dependence condition (H6) (respectively, (H7)) is a classical condition in kernel estimation based on nonstrictly stationary-dependent data (see, for example, [30]). The assumption (H6) (respectively, (H7)) controls the local dependence (respectively, the local identical distribution), whereas the mixing condition regulates the dependence of distant sites.
(ii) It is important to note that the condition (H3) defines the nonparametric space of our model. Once more, it is possible to proceed without making the assumption of Hölder condition. Instead, we can make a regularity assumption that is less restrictive on the nonparametric model. Nevertheless, the convergence rate of the bias term is also impacted by any limitation imposed on this assumption. In this context, imposing a more stringent condition on the model leads to an enhanced rate of convergence, while conversely relaxing the condition results in a slower rate of convergence. Specifically, if we substitute the Hölder condition with a continuity assumption, the convergence rate becomes slower, accompanied by a bias term of order o(1). In summary, it can be stated that hypothesis **(H3)** is formulated in a broad manner, enabling the examination of the nonparametric aspect of the model's convergence rate through the bias term.
(iii) In infinite-dimensional spaces, the assumption definition of $\phi_x(z)$ and (H4) is known as the "concentration property." For numerous instances, the small ball probability $\phi_x(h)$ can be approximated, around zero, as the product of two independent functions $f(x)$ and $\tau(h)$ (see, for example, reference [31] for the diffusion process, reference [32] for a Gaussian measure, and reference [33] for a general Gaussian process). The most common result found in the research literature has the form $\phi_x(h) \sim f(x)\tau(h)$, where $\tau(h) = h^\gamma \exp(-C/h^\gamma)$ with $\gamma \geq 0$ and $p \geq 0$. It corresponds to the Ornstein–Uhlenbeck and general diffusion processes ($p = 2$ and $\gamma = 0$ for such processes) and the fractal processes ($\gamma > 0$ and $p = 0$ for such processes). This class of processes also meets the requirements of condition (H5). It should be noted that these concepts are closely related to the proximity measure d that is taken into consideration, and all the instances described previously involve d being standard norms (such as the Hölder norm or supremum norm, for example). Multiple continuous time processes (see, for example, reference [32] for a Gaussian process) are used to test the hypothesis (H5).
(iv) The hypotheses (H7)–(H9) display the local continuous conditions required to establish the main results and consolidate the results. In fact, conditional expectation requires that if V_s is continuous for some $s > 2$, then V_2 is also continuous. The assumption (H8) in FDA MAR models is typical (see, for example, [27]).

4. Theoretical Results

We can now present our main results. It is important to note that these results extend the case of full data obtained by [11]. The following result gives the probability convergence of the regression kernel estimator with MAR.

Theorem 1. *Under hypotheses (H1)–(H9), (7), (9), and* $\widehat{\mathbf{n}}(\phi_x(h_{\mathbf{n}})/\log(\widehat{\mathbf{n}}))$ *tend to* ∞ *when* $\mathbf{n}\longrightarrow\infty$*. If we quote*

$$B_{\mathbf{n}}(x) := -\frac{\left(m(x)\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x)) - \mathbb{E}\left(\widetilde{f}_{\mathbf{n}}(x)\right)\right)}{\left(\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x))\right)}, \tag{18}$$

then

$$\sqrt{\widehat{\mathbf{n}}\left(\frac{\phi_x(h_{\mathbf{n}})}{\log(\widehat{\mathbf{n}})}\right)}\left|\widetilde{m}_{\mathbf{n}}(x) - m(x) - B_{\mathbf{n}}(x)\right| \text{ converge in probability to 0.} \tag{19}$$

In addition, if $\widehat{\mathbf{n}}h_{\mathbf{n}}^{\kappa}(\phi_x(h_{\mathbf{n}})/\log(\widehat{\mathbf{n}}))$ tend to 0, as $\mathbf{n}\longrightarrow\infty$, then

$$\sqrt{\widehat{\mathbf{n}}\left(\frac{\phi_x(h_{\mathbf{n}})}{\log(\widehat{\mathbf{n}})}\right)}\left|\widetilde{m}_{\mathbf{n}}(x) - m(x)\right| \text{ converge in probability to 0.} \tag{20}$$

The following result illustrates the asymptotic normality convergence of the regression kernel estimator with MAR.

Theorem 2. *For the hypotheses (H1) through (H9), if additionally* $\widehat{\mathbf{n}}(\phi_x(h_{\mathbf{n}}))$ *tend to* ∞*, as* $\mathbf{n}$ *go to* $+\infty$*, consequently, we have*

$$\sqrt{\widehat{\mathbf{n}}\phi_x(h_{\mathbf{n}})}\left|\widetilde{m}_{\mathbf{n}}(x) - m(x) - B_{\mathbf{n}}(x)\right| \xrightarrow{D} \mathcal{N}\left(0,\sigma^2(x)\right), \tag{21}$$

with $\sigma^2(x) = \beta_2/\beta_1^2 V_2(x)/\pi(x)\tau(x)$ *and* $\beta_k = -\int_0^1 \beta_x(t)(K^k)'(t)dt$*, for* $k = 1, 2$.

In addition, if $\widehat{\mathbf{n}}h_{\mathbf{n}}^{2\kappa}(\phi_x(h_{\mathbf{n}}))$ tend to 0, as $\mathbf{n}$ go toward infinity, we have

$$\sqrt{\widehat{\mathbf{n}}\phi_x(h_{\mathbf{n}})}\left|\widetilde{m}_{\mathbf{n}}(x) - m(x)\right| \xrightarrow{D} \mathcal{N}\left(0,\sigma^2(x)\right). \tag{22}$$

The proof of Theorems 1 and 2 is established on the following decomposition:

$$\widetilde{m}_{\mathbf{n}}(x) - m(x) - B_{\mathbf{n}}(x) = \frac{B_{\mathbf{n}}(x)\left(\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x)) - \widetilde{g}_{\mathbf{n}}(x)\right) + Q_{\mathbf{n}}(x)}{\widetilde{g}_{\mathbf{n}}(x)}, \tag{23}$$

where $Q_{\mathbf{n}}(x) := m(x)(\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x)) - \widetilde{g}_{\mathbf{n}}(x)) - (\mathbb{E}(\widetilde{f}_{\mathbf{n}}(x)) - \widetilde{f}_{\mathbf{n}}(x))$.

Theorems 1 and 2, thus, are immediate consequence of the following Lemmas.

Lemma 3. *Based on the assumptions (H1), (H2), and (H8), we can obtain for any* $x \in \mathscr{E}$ *that*

$$\widetilde{g}_{\mathbf{n}}(x) \xrightarrow{P} \pi(x), \quad \text{as } \mathbf{n}\longrightarrow\infty. \tag{24}$$

Lemma 4. *Suppose conditions (H1)–(H5) hold, then*

$$B_{\mathbf{n}}(x) = O\left(h_{\mathbf{n}}^{\kappa}\right), \text{ in probability}, \tag{25}$$

and

$$\sqrt{\widehat{\mathbf{n}}\phi_x(h_{\mathbf{n}})}B_{\mathbf{n}}(x)\left(\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x)) - \widetilde{g}_{\mathbf{n}}(x)\right) \text{ converge in probability to 0 as } \mathbf{n} \text{ tends to } \infty. \tag{26}$$

Lemma 5. *Based on assumptions (H1)–(H9), denote* $V(x) = \beta_2/\beta_1^2\pi(x)V_2(x)/\tau(x)$*, then we have*

$$\left(\frac{\widehat{\mathbf{n}}\phi_x(h_{\mathbf{n}})}{V(x)}\right)^{1/2} Q_{\mathbf{n}}(x) \xrightarrow{D} \mathcal{N}(0,1), \quad \text{as } \mathbf{n} \text{ tends to } \infty. \tag{27}$$

5. Simulation and Application Results

This section's primary purpose is to evaluate the excellent behavior of our estimator for various missing rates and sample sizes and to demonstrate the efficacy of this approach in comparison to the conventional one.

5.1. Simulation Study. We establish here the significance of our proposed predictor by evaluating its performance in numerical experiments. The introduced predictor is compared to the conventional kernel technique, which ignores missing data. In order to determine the finite sample performance of the introduced estimator $\widehat{r}$, we conducted a simulation study derived from the observations $(X_{\mathbf{i}}, Y_{\mathbf{i}}, \delta_{\mathbf{i}}) \in (\mathscr{E}\times\mathbb{R}\times\{0,1\})$. Note that $\mathbf{i} = (i_1, i_2)$ is defined using $1 \le i_1 \le n_1$, $1 \le i_2 \le n_2$, and $\forall \mathbf{i} \in \mathbb{Z}^2$. The model was created in the following manner:

$$\forall t \in [0,1], X_{\mathbf{i}}(t) = \cos\left(2\pi A_{\mathbf{i}}t\right) + B_{\mathbf{i}}t, \tag{28}$$

and

$$Y_i = m(X_{\mathbf{i}}) + \varepsilon_{\mathbf{i}}, \tag{29}$$

with $m(Z) = 5/\int_0^1 |Z(t)|dt$.

Next, we refer to $\mathrm{GRF}(m, \sigma^2, s)$ as a stationary Gaussian random field with mean m and the functional covariance is given as

$$C(l) = \sigma^2 \exp\left(-\left(\frac{\|l\|}{s}\right)^2\right), \quad l \in \mathbb{R}^2. \tag{30}$$

Then, put $A = D * \sin(G/2 + .5), B = \mathrm{GRF}(2.5, 5, 3)$, $\varepsilon = \mathrm{GRF}(0, .1, 5)$, $G = \mathrm{GRF}(0, 5, 3)$, $D_{\mathbf{i}} = 1/n_1 \times n_2 \sum_{\mathbf{j}} \exp(-\|\mathbf{i} - \mathbf{j}\|/a)$, and

$$\left(D_{(\mathbf{i},\mathbf{j})} = \frac{1}{n_1 \times n_2} \sum_{1 \le j_1, j_2 \le 25} \exp\left(-\frac{\|(i_1, i_2) - (j_1, j_2)\|}{a}\right)\right), \tag{31}$$

where the latter function is designed to ensure and adjust spatial mixing conditions. We simulated model (29) and used the missing method, as described by [8], where

$$p(z) = \mathbb{P}\left(\delta = \frac{1}{Z} = z\right) = \mathrm{expit}\left(2\kappa \int_0^1 z^2(t)dt\right). \tag{32}$$

Note that $\mathrm{expit}(u) = e^u/(1 + e^u)$, $\forall u \in \mathbb{R}$. The above formula contains a parameter denoted by κ, which controls the level of dependence between the functional curve X and the variable δ. In order to maintain the value of $p(x)$, we calculate

$$\overline{\delta} = 1 - \frac{1}{n_1 \times n_2} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \delta_{(i_1, i_2)}. \tag{33}$$

Figure 1 depicts the simulated functional curves.

Regarding the parameters involved in the implementation of the estimator A, we would like to emphasize that a quadratic kernel has been taken into consideration, given by $K(t) = 3/2(1 - t^2)1_{[0,1]}(t)$. Similar to [3], the bandwidth parameter is defined by

$$d(X_{\mathbf{i}}, X_{\mathbf{j}}) = \sqrt{\int_0^1 \left(X'_{\mathbf{i}}(t) - X'_{\mathbf{j}}(t)\right)^2 dt}, \quad \forall X_{\mathbf{i}}, X_{\mathbf{j}} \in \mathscr{E}. \tag{34}$$

The location observations $\mathbf{i}$ and $\mathbf{j}$ with $\|\mathbf{i} - \mathbf{j}\| < 15$ are spatially dependent and almost independent when $\|\mathbf{i} - \mathbf{j}\| \ge 15$, since the model (in these conditions) is based on Gaussian random fields with covariance function C and scale $s = 5$. Our observations are, therefore, a combination of dependent and independent observations (see Figure 2). Therefore, decreasing the value of a is all that is required to abandon independence (our results are based on $a = 0.5$).

The primary purpose of this comparison is to examine our proposed estimator, MAR ($\widehat{m}_{\mathbf{n}}(x)$), with the naive estimator, denoted by MARV ($\widetilde{m}_{\mathbf{n}}(x)$), and the complete data estimator, denoted by ECD ($\widetilde{m}_{n,l}(x)$) and proposed by [11]. To assess the effectiveness of the proposed estimator, $(X_{\mathbf{i}}, Y_{\mathbf{i}}, \delta_{\mathbf{i}})_{\mathbf{i}}$ was divided into two subsets at random:

(i) $(X_{\mathbf{i}}, Y_{\mathbf{i}}, \delta_{\mathbf{i}})_{\mathbf{i} \in I}$, the learning sample

(ii) $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i} \in I'}$, the test sample

We use the training sample to determine the smoothing parameters $h_{k_{\mathrm{opt}}}$ for the $k-$ NN cross-validation operations. The bandwidth corresponding to the optimal number of neighbors generated by a cross-validation technique is denoted by $h_{k_{\mathrm{opt}}}$:

$$h_k = \min\left\{\frac{h \in \mathbb{R}^+}{\sum_{\mathbf{i} \in I} \mathbb{1}_{B(x,h)}(Z_{\mathbf{i}})} = k\right\}, \tag{35}$$

with

$$\begin{aligned} k_{opt} &= \underset{k}{\operatorname{argmin}}\, \mathrm{CV}(k), \\ \mathrm{CV}(k) &= \sum_{\mathbf{i} \in I} \left(Y_i - \widehat{m}_{\mathbf{n}}^{(-\mathbf{i})}(X_{\mathbf{i}})\right)^2, \end{aligned} \tag{36}$$

and also $\widehat{m}_{\mathbf{n}}^{(-\mathbf{i})}$ is the leave-one-out version of $\widehat{m}_{\mathbf{n}}$, evaluated by eliminating the i th datum from the initial sample (for additional information, see [3]).

In one sense, the accuracy of the estimate, $\widehat{m}_{\mathbf{n}}(\cdot)$ of $m(\cdot)$, was performed by using the mean square errors (MSEs):

$$\mathrm{MSE} = \frac{1}{\#\left(I'\right)} \sum_{\mathbf{i} \in I'} \left(\widehat{m}_{\mathbf{n}}(X_{\mathbf{i}}) - m(X_{\mathbf{i}})\right)^2, \tag{37}$$

where $\#(I')$ is the sample size used for testing. The results of the 3 various models are depicted in Figure 3, which compares the predicted values to the real values.

Consequently, Tables 1 and 2 exhibit the MSE and Bias for the MAR, MARV, and complete data models, respectively.

We evaluate the suggested estimator's performance in terms of bias as well. Using $M = 100$ replicates of the experiment, we can quantify the bias of the estimators of r by

$$\mathrm{Bias}(\widehat{m}_{\mathbf{n}}) = \frac{1}{M} \sum_{k=1}^{M} \widehat{m}_{\mathbf{n}}^{(k)}(x) - m(x), \tag{38}$$

where $\widehat{m}_{\mathbf{n}}^{(k)}(x)$ is the estimator of $m(x)$ for the replication k of the different proposed models. We summarize these results in Table 2.

When the missing data rate is minor, the naïve version provides a superior MSE, but as the rate rises, the MAR estimator provides a better estimate. This is shown in Tables 1 and 2. We also take note of the fact that when $\mathbf{n}$ increases, the MSE and bias reduce dramatically. The theoretical conclusions of Theorem 1 are consistent with such a numerical outcome. In addition, the bias is negligible in all cases and is always negative for all MAR settings.

5.2. Real Data Application. In this section, it can be stated that the stationarity hypothesis is fundamental to the nonparametric analysis of spatio-functional data, and that the proposed detrending method is an ideal method for ensuring this hypothesis.

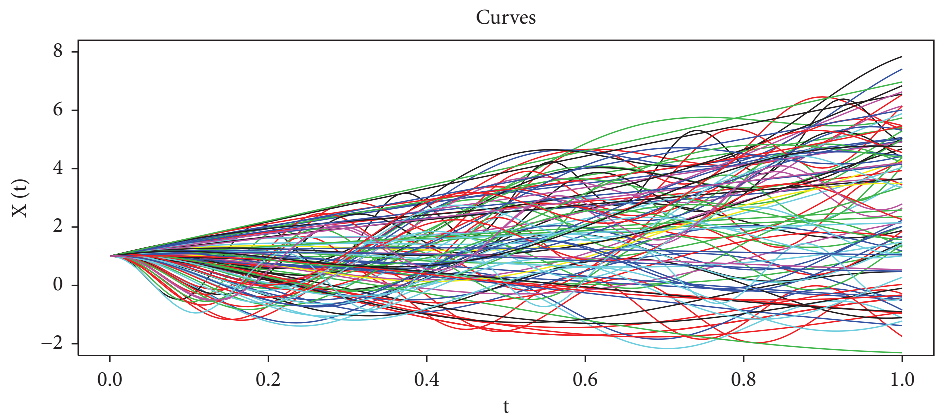

FIGURE 1: The functional curves $X_{\mathbf{i}}(t), t \in [0,1]$.

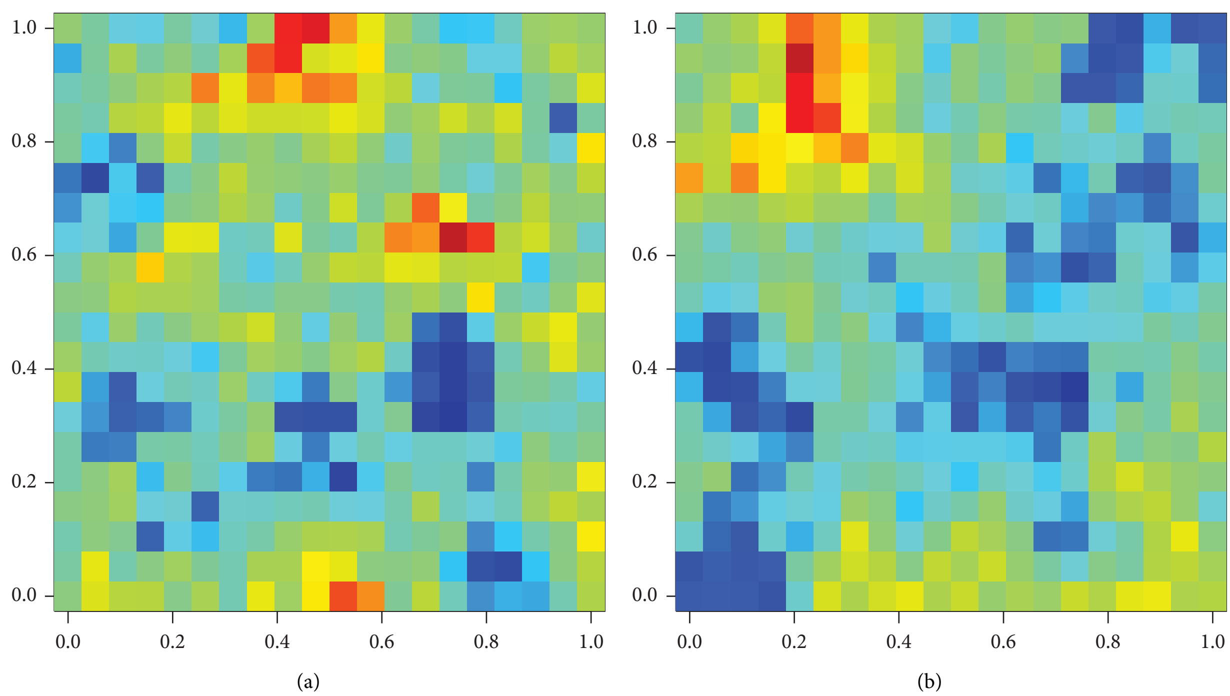

FIGURE 2: Random field simulation.

5.2.1. No Detrending Case. First, we have acquired a large dataset in 122 stations in the United States, containing the daily mean of ozone concentration for the day 11/08/2019 (measured in Parts per million). In addition to this, we have obtained a dataset that consists of the chronological hourly weather data pertaining to the temperature (measured in Fahrenheit Degrees) for the same day. These observations can be found at https://aqs.epa.gov/aqsweb/airdata/download_files.html#Raw. Figure 4 depicts the position of the 122 stations in the USA.

Given the daily temperature curve denoted by X, we are interested in the daily mean ozone concentration forecast Y (for the day of August 11, 2019). We suppose that the two variables are linked by

$$Y = m(X) + \epsilon. \tag{39}$$

Figure 5 provides 122 curves of hourly temperature measures of each station measured in Degrees Fahrenheit.

The functional explanatory $X_{\mathbf{i}}$ represents the daily temperature curve in the $\mathbf{i}$th station (specified geographically by the coordinates $\mathbf{i}$ = (Latitude; Longitude)), whereas $Y_{\mathbf{i}}$ is the ozone concentration in the same location. We implemented the theoretical findings from the previous section into actual data. Specifically, in the context of spatial functional prediction, we analyze the effectiveness of our constructed estimator with MAR data, which highlights the significance of taking spatial locations into consideration in this kind of data. Note that our data have some missing values (38 NaNs stations, about (31.15% missing data), since, in some stations, $Y_{\mathbf{i}}$ are not measured on some of the samples. Therefore, our sample is formed as

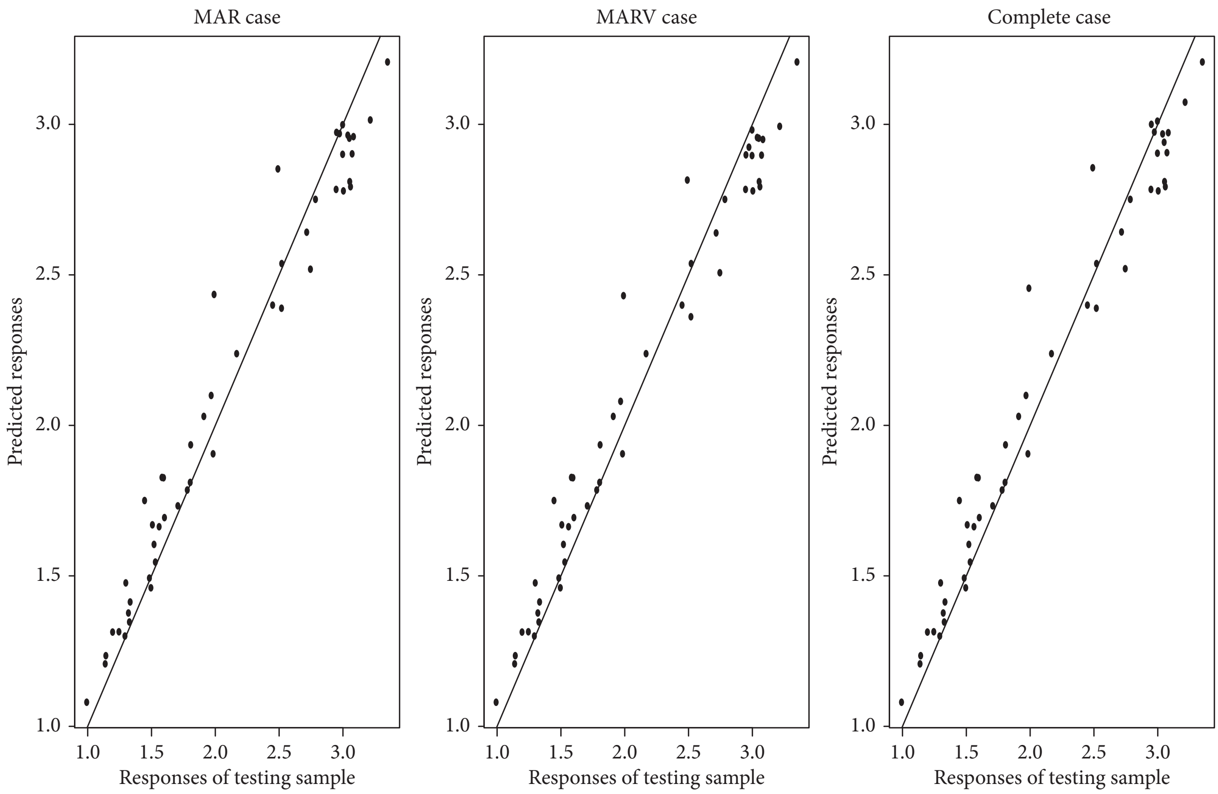

Figure 3: Plots of the predictions for the MAR, MARV, and complete data models.

Table 1: MSE (mean squared error) for complete data, MAR, and MARV models.

n_1	n_2	κ	$\overline{\delta}$	MAR (MSE)	MARV (MSE)	Complete (MSE)
10	10	0.5	0.1386	0.2234	0.2211	0.2097
		2.0	0.0393	0.2439	0.2433	0.2356
	20	0.5	0.1353	0.1463	0.1483	0.1285
		2.0	0.0280	0.1340	0.1361	0.1318
	30	0.5	0.1340	0.1267	0.1257	0.1058
		2.0	0.0217	0.1117	0.1128	0.1100
20	10	0.5	0.1431	0.1581	0.1552	0.1334
		2.0	0.0251	0.1656	0.1659	0.1608
	20	0.5	0.1344	0.1108	0.1106	0.0954
		2.0	0.0228	0.0956	0.0975	0.0937
	30	0.5	0.1302	0.0920	0.0894	0.0778
		2.0	0.0216	0.0794	0.0793	0.0781
30	10	0.5	0.1426	0.1346	0.1336	0.1233
		2.0	0.0262	0.1178	0.1187	0.1156
	20	0.5	0.1330	0.0824	0.0801	0.0720
		2.0	0.0253	0.0764	0.0767	0.0748
	30	0.5	0.1303	0.0552	0.0543	0.0467
		2.0	0.0215	0.0579	0.0584	0.0567

follows $(X_{\mathbf{i}}, Y_{\mathbf{i}}, \delta_{\mathbf{i}})$, where $\delta_{\mathbf{i}} = 1$ if $Y_{\mathbf{i}}$ is observed and 0 otherwise. Further, the quadratic function $K(u) = 1.5(1-u^2)1_{[0,1]}$ is used for defining the kernel of the model. In functional nonparametric regression, the choice of the pseudo-metric is a key decision since it significantly influences the type of model that is taken into consideration and the effectiveness of the estimation process adapted to this type of data. We use PCA-type semimetric, defined by

$$d_q^{\mathrm{PCA}}\left(X_{\mathbf{i}}, X_{\mathbf{j}}\right) = \sqrt{\sum_{k=1}^{q}\left(\int\left[X_{\mathbf{i}}(z) - X_{\mathbf{j}}(z)\right]\nu_k(z)\mathrm{d}z\right)^2}. \quad (40)$$

Here, we utilize $q = 4$ and choose the eigenfunction ν_k from the set of eigenfunctions of the empirical covariance operator:

Table 2: BIAS for a fixed x for complete data, MAR, and MARV models.

n_1	n_2	κ	$\overline{\delta}$	MAR (BIAS)	MARV (BIAS)	Complete (MSE)
10	10	0.5	0.1386	−0.0909	−0.0985	−0.0159
		2.0	0.0393	−0.0416	−0.0481	−0.0204
	20	0.5	0.1353	−0.0354	−0.0505	0.0063
		2.0	0.0280	−0.0048	−0.0075	0.0063
	30	0.5	0.1340	−0.0080	−0.0215	0.0334
		2.0	0.0217	−0.0040	−0.0086	0.0047
20	10	0.5	0.1431	−0.0420	−0.0543	−0.0042
		2.0	0.0251	−0.0029	−0.0083	0.0056
	20	0.5	0.1344	−0.0396	−0.0563	−0.0105
		2.0	0.0228	0.0051	0.0004	0.0121
	30	0.5	0.1302	−0.0155	−0.0333	0.0127
		2.0	0.0216	0.0036	−0.0016	0.0092
30	10	0.5	0.1426	−0.0599	−0.0737	−0.0203
		2.0	0.0262	−0.0057	−0.0101	0.0014
	20	0.5	0.1330	−0.0273	−0.0384	0.0000
		2.0	0.0253	0.0044	−0.0014	0.0093
	30	0.5	0.1303	−0.0265	−0.0420	−0.0027
		2.0	0.0215	0.0116	0.0068	0.0159

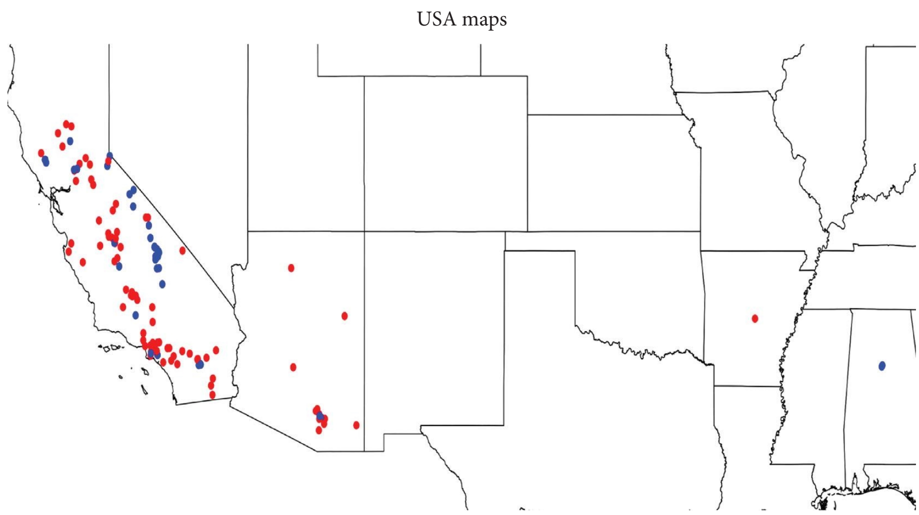

Figure 4: Observed stations' locations (the red points are the observed stations and the blue are the missing one).

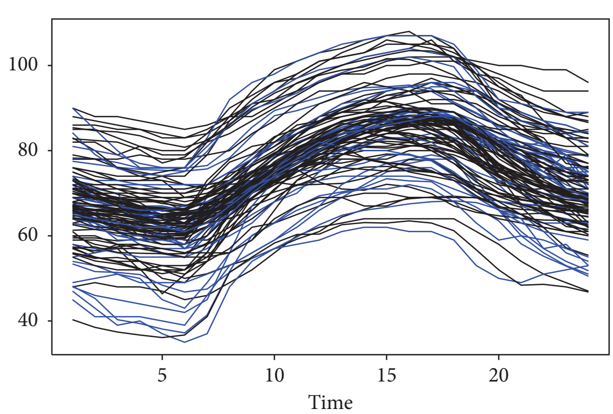

Figure 5: Daily temperature curves of 122 stations.

$$\Gamma_X^n(s,t) = \frac{1}{n}\sum_{i=1}^{n} X_{\mathbf{i}}(s) X_{\mathbf{j}}(t). \quad (41)$$

Finally, we examined at the cross-validation method's selection of the optimal bandwidth $h{:}\ = h_{n,K}$. Then, we divided our data $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}}$ into two subsets at random: test sample $(T_{\mathbf{i}}, X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in I'}$, (30 stations) and learning sample $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in I}$ (92 stations). The following definition outlines the mean square error (MSE), which we use as an accuracy indicator:

$$\mathrm{MSE} = \frac{1}{30}\sum_{\mathbf{i}\in I'} \left(Y_{\mathbf{i}} - \widetilde{Y}_{\mathbf{i}}\right)^2, \quad (42)$$

where $\widetilde{Y}_{\mathbf{i}}$ denotes the estimator's value.

To investigate the efficacy of our models further, we execute $M = 100$ independent repeats, which permit us to generate 100 values for MSE and depict their distribution using a boxplot. The boxplots of MSE of the prediction values are shown in Figure 6.

Now, in Figure 7, we show the 90% prediction ranges for the ozone concentrations of the 20 last data in the sample test. This result demonstrates that our asymptotic normality is effective.

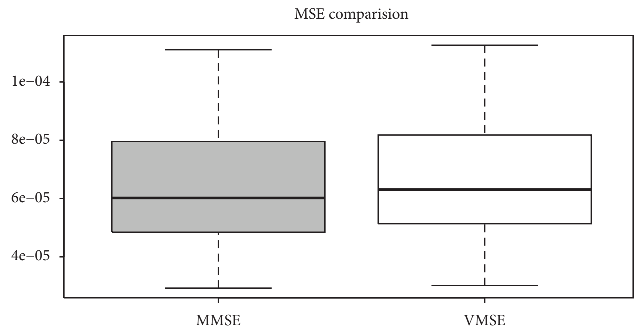

Figure 6: The boxplots of the predicted values' MSE.

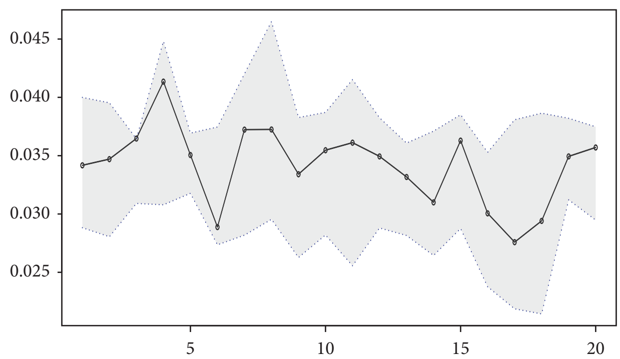

Figure 7: Extremes of the predicted values compared to the real values and confidence intervals. The true values are connected by the solid black line. The dashed blue curves connect the expected minimum and maximum values.

5.2.2. Detrending Case. As discussed in [34], the use of this type of spatial modeling requires prior preparation of the initial data in order to verify the stationarity hypothesis. The latter controls the spatial heterogeneity linked to a differentiation of the effects of space on the sampling units. To control this aspect, we adopt the algorithm proposed by [34] for the multivariate case in a finite-dimension for which the spatial heterogeneity of the two variables (explanatory and response) is modeled by the following regression:

$$\begin{aligned} \widetilde{X}_{\mathbf{i}} &= m_1(\mathbf{i}) + X_{\mathbf{i}}, \\ \widetilde{Y}_{\mathbf{i}} &= m_2(\mathbf{i}) + Y_{\mathbf{i}}. \end{aligned} \quad (43)$$

Thus, instead of the initial observations $(X_{\mathbf{i}}, Y_{\mathbf{i}}, \delta_{\mathbf{i}})_{\mathbf{i}}$, we compute the SPL and NP estimators from the statistics $(\widehat{X}_{\mathbf{i}}, \widehat{Y}_{\mathbf{i}}, \delta_{\mathbf{i}})_{\mathbf{i}}$. The latter are obtained by

$$\begin{aligned} \widehat{X}_{\mathbf{i}} &= \widetilde{X}_{\mathbf{i}} - \widehat{m}_1(\mathbf{i}), \\ \widehat{Y}_{\mathbf{i}} &= \widetilde{Y}_{\mathbf{i}} - \widehat{m}_2(\mathbf{i}), \end{aligned} \quad (44)$$

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

$$\begin{aligned} \widehat{m}_1(\mathbf{i}_0) &= \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \delta_{\mathbf{i}} X_{\mathbf{i}} H_1\left(\left\|\mathbf{i}_0 - \mathbf{i}\right\|/\lambda_{\mathbf{n}}\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \delta_{\mathbf{i}} H_1\left(\left\|\mathbf{i}_0 - \mathbf{i}\right\|/\lambda_{\mathbf{n}}\right)}, \\ \widehat{m}_2(\mathbf{i}_0) &= \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \delta_{\mathbf{i}} Y_{\mathbf{i}} H_2\left(\left\|\mathbf{i}_0 - \mathbf{i}\right\|/\gamma_{\mathbf{n}}\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \delta_{\mathbf{i}} H_2\left(\left\|\mathbf{i}_0 - \mathbf{i}\right\|/\gamma_{\mathbf{n}}\right)}, \end{aligned} \quad (45)$$

where H_1 and H_2 are kernel functions and $\lambda_{\mathbf{n}}$ and $\gamma_{\mathbf{n}}$ are the bandwidth parameters of the real regression. Such a step is called "detrending step" and is fundamental in the nonparametric analysis of spatial data. For our actual data set, we highlight the impact of this detrending step in practice. To do this, we compare the efficiency of our estimator in the two situations (with and without detrending). For this, we keep the same strategies as those used in the simulation example to select the parameters involved in the estimator. More precisely, we use the quadratic kernel on $(0, 1)$ and the PCA metric and the criterion CV to choose the smoothing parameter $h_{\mathbf{n}}$. Concerning the real regressions $m_1(.)$ and $m_2(.)$, we used the routine code npreg in the R-package np over $K = H_1 = H_2$. The feasibility of this is evaluated, by splitting randomly and several times (exactly 100 times) the data sample. Finally, we examine the importance of the proposed detrending procedure through the MSE in Figure 8 used in the simulation example.

6. Proofs of the Main Results

Throughout the rest of this paper, we define, respectively, $\lambda_{\mathbf{i}}(x)$ and the random variable $L_{\mathbf{i}}(x)$ by

$$\begin{aligned} \lambda_{\mathbf{i}}(x) &= \frac{\sqrt{\phi_x(h)}}{\mathbb{E}(K_1)}\left[\delta_{\mathbf{i}}(Y_{\mathbf{i}} - m(x))K_{\mathbf{i}}\right], \\ L_{\mathbf{i}}(x) &= \lambda_{\mathbf{i}}(x) - \mathbb{E}(\lambda_{\mathbf{i}}(x)). \end{aligned} \quad (46)$$

Lemma 6. *Based on the hypotheses of Theorem 2, we have, for all* $(\mathbf{i}, \mathbf{j})$*:*

(i) $\sum_{\mathbf{i}\neq\mathbf{j}} Cov(L_{\mathbf{i}}(x), L_{\mathbf{j}}(x)) = o(\widehat{\mathbf{n}})$

(ii) $1/\widehat{\mathbf{n}}\mathrm{Var}(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} L_{\mathbf{i}}(x)) \longrightarrow V(x) = \pi(x)V_2(x)\beta_2/\beta_1^2\tau(x)$ *as* $\mathbf{n} \longrightarrow \infty$

Proof. First, we have

$$\mathrm{Var}\left(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} L_{\mathbf{i}}(x)\right) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \mathrm{Var}(L_{\mathbf{i}}(x)) + \sum_{\mathbf{i}\neq\mathbf{j}} \mathrm{Cov}\left(L_{\mathbf{i}}(x), L_{\mathbf{j}}(x)\right). \quad (47)$$

Then, we quote $I_{\mathbf{n}}(x) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \mathrm{Var}(L_{\mathbf{i}}(x))$ and $R_{\mathbf{n}}(x) = \sum_{\mathbf{i}\neq\mathbf{j}} \mathrm{Cov}(L_{\mathbf{i}}(x), L_{\mathbf{j}}(x))$. For the variance term, we have $\mathrm{Var}(L_{\mathbf{i}}(x)) = \mathbb{E}(L_{\mathbf{i}}^2(x)) = \mathbb{E}(\lambda_{\mathbf{i}}^2(x)) - \mathbb{E}^2(\lambda_{\mathbf{i}}(x))$. Now, by conditioning on $X_{\mathbf{i}}$ and using (H3), (H8), and MAR assumption, we get

$$\mathbb{E}(\lambda_{\mathbf{i}}(x)) = \sqrt{\phi_x(h)}(m(x_{\mathbf{i}}) - m(x))\frac{\mathbb{E}(K_{\mathbf{i}})}{\mathbb{E}(K_1)}. \leq \sqrt{\phi_x(h)}h^{\kappa}(\pi(x) + o(1)). \quad (48)$$

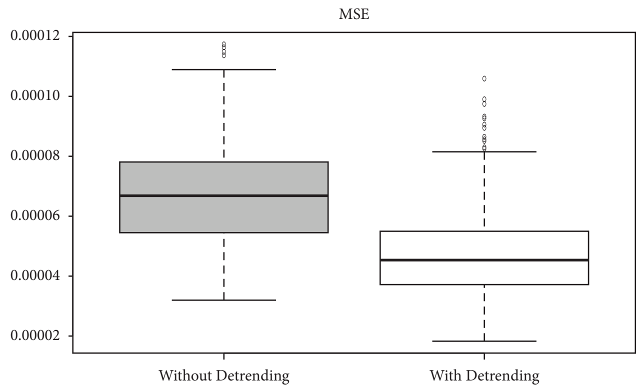

Figure 8: The boxplots of the predicted values' MSE without and with detrending.

Likewise, use MAR assumption and conditioning on $X_{\mathbf{i}}$ to get

$$\mathbb{E}\left(\lambda_{\mathbf{i}}^2(x)\right) = \phi_x(h)\left[\mathbb{E}\left(\pi(x_{\mathbf{i}})\left(r(x_{\mathbf{i}}) - r(x)\right)^2 \frac{K_{\mathbf{i}}^2}{\mathbb{E}^2(K_1)}\right)\right] + \phi_x(h)\left[\mathbb{E}\left(\pi(x_{\mathbf{i}})V_2(x_{\mathbf{i}})\frac{K_{\mathbf{i}}^2}{\mathbb{E}^2(K_1)}\right)\right]. \tag{49}$$

It follows by (H7) and (H8) that

$$\left[\mathbb{E}\left(\pi(x_{\mathbf{i}})V_2(x_{\mathbf{i}})\frac{K_{\mathbf{i}}^2}{\mathbb{E}^2(K_1)}\right)\right] = (\pi(x) + o(1))\left(V_2(x) + o(1)\right)\frac{\mathbb{E}\left(K_{\mathbf{i}}^2\right)}{\mathbb{E}^2(K_1)}, \tag{50}$$

and by (H3) and (H8) that

$$\mathbb{E}\left(\pi(x_{\mathbf{i}})\left(m(x_{\mathbf{i}}) - m(x)\right)^2 \frac{K_{\mathbf{i}}^2}{\mathbb{E}^2(K_1)}\right) \leq (\pi(x) + o(1))h^{2\kappa}\frac{\mathbb{E}\left(K_{\mathbf{i}}^2\right)}{\mathbb{E}^2(K_1)} = (\pi(x) + o(1))\, h^{2\kappa}\,\frac{\mathbb{E}\left(K_{\mathbf{i}}^2\right)}{\mathbb{E}^2(K_1)}. \tag{51}$$

Based on the following inequality:

$$\mathbb{E}\left(\lambda_{\mathbf{i}}^2(x)\right) \leq \phi_x(h)(\pi(x) + o(1))\left[\left(V_2(x) + o(1)\right) + h^{2\kappa}\right]\frac{\mathbb{E}\left(K_{\mathbf{i}}^2\right)}{\mathbb{E}^2(K_1)}, \tag{52}$$

we get

$$\begin{aligned}\mathbb{E}\left(L_{\mathbf{i}}(x)^2\right) \leq{}& \phi_x(h)(\pi(x) + o(1))\left(V_2(x) + o(1)\right)\frac{\mathbb{E}\left(K_{\mathbf{i}}^2\right)}{\mathbb{E}^2(K_1)} + h^{2\kappa}\phi_x(h)(\pi(x) + o(1))\frac{\mathbb{E}\left(K_{\mathbf{i}}^2\right)}{\mathbb{E}^2(K_1)}\\ &+ h^{2\kappa}\phi_x(h)(\pi(x) + o(1))\frac{\mathbb{E}^2(K_{\mathbf{i}})}{\mathbb{E}^2(K_1)},\end{aligned} \tag{53}$$

and

$$I_{\mathbf{n}}(x) \leq \widehat{\mathbf{n}}\phi_x(h)(\pi(x)+o(1))(V_2(x)+o(1))\frac{\mathbb{E}(K_{\mathbf{i}}^2)}{\mathbb{E}^2(K_{\mathbf{1}})} + +\widehat{\mathbf{n}}h^{2\kappa}(\pi(x)+o(1))\frac{\mathbb{E}(K_{\mathbf{i}}^2)}{\mathbb{E}^2(K_{\mathbf{1}})} + \widehat{\mathbf{n}}h^{2\kappa}\phi_x(h)(\pi(x)+o(1))\frac{\mathbb{E}^2(K_{\mathbf{i}})}{\mathbb{E}^2(K_{\mathbf{1}})}. \tag{54}$$

Furthermore, under (H1)-(H2) and (H4)-(H5), we know that

$$\frac{1}{\tau(x)\phi_x(h)}\mathbb{E}(K_{\mathbf{i}}^j) \longrightarrow \beta_j, \quad \text{for } j=1,2, \tag{55}$$

where β_j is given in Theorem 2. Thus, we get

$$\frac{1}{\widehat{\mathbf{n}}}I_{\mathbf{n}}(x) = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\mathbb{E}\left((L_{\mathbf{i}}(x)^2\right) \longrightarrow V(x) = \frac{\beta_2}{\beta_1^2}\frac{\pi(x)V_2(x)}{\tau(x)}, \quad \text{as } \mathbf{n}\longrightarrow\infty. \tag{56}$$

Let us now focus on the covariance term. As

$$\sum_{\mathbf{i}\neq\mathbf{j}}\operatorname{Cov}(L_{\mathbf{i}}(x), L_{\mathbf{j}}(x)) = \sum_{\mathbf{i}\neq\mathbf{j}}\mathbb{E}(L_{\mathbf{i}}(x)L_{\mathbf{j}}(x)), \tag{57}$$

and by some argument as above, we have

$$\sum_{\mathbf{i}\neq\mathbf{j}}\mathbb{E}(L_{\mathbf{i}}(x)L_{\mathbf{j}}(x)) \leq \frac{1}{\mathbb{E}^2(K_{\mathbf{1}})}h^{2\kappa}\phi_x(h)(\pi(x)+o(1))^2\sum_{\mathbf{i}\neq\mathbf{j}}\left(\mathbb{E}(K_{\mathbf{i}}K_{\mathbf{j}}) - \mathbb{E}(K_{\mathbf{i}})\mathbb{E}(K_{\mathbf{i}})\right). \tag{58}$$

This sum is divided into two distinct sums, one on the site E_1 and the other on the site E_2, with

$$E_1 = \{\mathbf{i}\neq\mathbf{j}\in\mathcal{I}_{\mathbf{n}}, \quad \text{such that } \|\mathbf{i}-\mathbf{j}\|\leq c_{\mathbf{n}}\},$$
$$E_2 = \{\mathbf{i}\neq\mathbf{j}\in\mathcal{I}_{\mathbf{n}}, \quad \text{such that } \|\mathbf{i}-\mathbf{j}\| > c_{\mathbf{n}}\}, \tag{59}$$

where $c_{\mathbf{n}}$ tend to $+\infty$ as $\mathbf{n}\longrightarrow\infty$ which will be given later.

Let

$$R_{\mathbf{n}}^1 = \sum_{E_1}\left|\mathbb{E}(K_{\mathbf{i}}K_{\mathbf{j}}) - \mathbb{E}(K_{\mathbf{i}})\mathbb{E}(K_{\mathbf{j}})\right|,$$
$$R_{\mathbf{n}}^2 = \sum_{E_2}\left|\mathbb{E}(K_{\mathbf{i}}K_{\mathbf{j}}) - \mathbb{E}(K_{\mathbf{i}})\mathbb{E}(K_{\mathbf{j}})\right|. \tag{60}$$

As $\mathbb{E}(K_{\mathbf{i}}K_{\mathbf{j}}) \leq C\mathbb{P}[(X_{\mathbf{i}}, X_{\mathbf{j}}) \in (B(x,h)\times B(x,h))]$ and $\mathbb{E}(K_{\mathbf{i}}) \leq C\mathbb{P}[X_{\mathbf{i}}\in B(x,h)]$, it follows, by assumption (H1), that

$$R_{\mathbf{n}}^1 \leq C\widehat{\mathbf{n}}c_{\mathbf{n}}^N\frac{1}{\mathbb{E}^2(K_{\mathbf{1}})}h^{2\kappa}\phi_x(h)(\pi(x)+o(1))\sum_{\mathbf{i}\neq\mathbf{j}}\phi_x(h)^{1+1/a}. \tag{61}$$

Now, using the boundedness of the random variable $K_{\mathbf{i}}$, we deduce from Lemma (3.3) in [30] that

$$\left|\mathbb{E}(K_{\mathbf{i}}K_{\mathbf{j}}) - \mathbb{E}(K_{\mathbf{i}})\mathbb{E}(K_{\mathbf{j}})\right| \leq C\varphi(\|\mathbf{i}-\mathbf{j}\|), \tag{62}$$

from where we have

$$\begin{aligned} R_{\mathbf{n}}^2 &\leq \frac{1}{\mathbb{E}^2(K_{\mathbf{1}})}Ch^{2\kappa}\phi_x(h)(\pi(x)+o(1))\sum_{\mathbf{i},\mathbf{j}\in E_2}\varphi(\|\mathbf{i}-\mathbf{j}\|) \\ &\leq \frac{1}{\mathbb{E}^2(K_{\mathbf{1}})}C\widehat{\mathbf{n}}h^{2\kappa}\phi_x(h)(\pi(x)+o(1))\sum_{\mathbf{i}:\|\mathbf{i}\|\geq c_{\mathbf{n}}}\varphi(\|\mathbf{i}\|) \\ &\leq \frac{1}{\mathbb{E}^2(K_{\mathbf{1}})}C\widehat{\mathbf{n}}h^{2\kappa}\phi_x(h)(\pi(x)+o(1))c_{\mathbf{n}}^{-\mathrm{Na}}\sum_{\mathbf{i}\neq\mathbf{j}}\sum_{\mathbf{i}:\|\mathbf{i}\|\geq c_{\mathbf{n}}}\|\mathbf{i}\|^{\mathrm{Na}}\varphi(\|\mathbf{i}\|). \end{aligned} \tag{63}$$

Then, by condition (9) and if $c_{\mathbf{n}} = (\phi_x(h))^{-1/\mathrm{Na}}$, we have

$$R_{\mathbf{n}} = R_{\mathbf{n}}^1 + R_{\mathbf{n}}^2 \leq C\widehat{\mathbf{n}}\frac{1}{\mathbb{E}^2(K_1)}h^{2\kappa}\phi_x(h)(\pi(x) + o(1)). \quad (64)$$

So, equations (59), (61), and (63) imply that $\sum_{\mathbf{i},\mathbf{j}}\mathrm{Cov}(L_{\mathbf{i}}(x), L_{\mathbf{j}}(x)) = o(\widehat{\mathbf{n}})$. □

Proof (Lemma 3). Clearly, $\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x)) \longrightarrow \pi(x)$ as $\mathbf{n} \longrightarrow \infty$. Indeed, by conditioning on $X_{\mathbf{i}}$ and when the MAR assumption holds, the assumptions (H2) and (H8) imply

$$\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x)) = (\pi(x) + o(1))\frac{\mathbb{E}(K_{\mathbf{i}})}{\mathbb{E}(K_1)} \longrightarrow \pi(x), \quad \text{as } \mathbf{n} \longrightarrow \infty. \quad (65)$$

Then, it suffices to show that $\mathrm{Var}(\widetilde{g}_{\mathbf{n}}(x))$ tends to 0 as $\mathbf{n} \longrightarrow \infty$. For this, let $\Lambda_{\mathbf{i}} = 1/\mathbb{E}(K_1)[\delta_{\mathbf{i}} K_{\mathbf{i}} - \mathbb{E}(\delta_{\mathbf{i}} K_{\mathbf{i}})]$ and $S_{\mathbf{n}} = \sum_{\mathbf{i}\in\mathcal{I}} \Lambda_{\mathbf{i}}$. Then, we have

$$\begin{aligned}\mathrm{Var}(\widetilde{g}_{\mathbf{n}}(x)) &= \mathrm{Var}\left(\frac{S_{\mathbf{n}}}{\widehat{\mathbf{n}}}\right) = \frac{1}{\widehat{\mathbf{n}}^2}\left[\sum_{\mathbf{i}\in\mathcal{I}}\mathrm{Var}(\Lambda_{\mathbf{i}}) + \sum_{\mathbf{i}\neq\mathbf{j}}\mathrm{Cov}\left(\Lambda_{\mathbf{i}}, \Lambda_{\mathbf{j}}\right)\right]\\ &= \frac{1}{\widehat{\mathbf{n}}^2}\left[\sum_{\mathbf{i}\in\mathcal{I}}\mathbb{E}\left(\Lambda_{\mathbf{i}}^2\right) + \sum_{\mathbf{i}\neq\mathbf{j}}\mathbb{E}\left(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}\right)\right].\end{aligned} \quad (66)$$

Using the same argument as in Lemma 6 with the same notations, we have on one side that

$$\begin{aligned}\mathbb{E}\left((\delta_{\mathbf{i}} K_{\mathbf{i}})^2\right) &= (\pi(x) + o(1))^2\mathbb{E}\left((K_{\mathbf{i}})^2\right) \leq C\,(\pi(x) + o(1))\,(\phi_x(h))^{a+1/a},\\ \mathbb{E}(\delta_{\mathbf{i}} K_{\mathbf{i}}) &= (\pi(x) + o(1))^2\mathbb{E}^2(K_{\mathbf{i}}) \leq C(\pi(x) + o(1))\phi_x(h).\end{aligned} \quad (67)$$

As $\mathbb{E}(\Lambda_{\mathbf{i}}) \leq 1/(\mathbb{E}(K_1))^2[\mathbb{E}((\delta_{\mathbf{i}} K_{\mathbf{i}})^2) + (\mathbb{E}(\delta_{\mathbf{i}} K_{\mathbf{i}}))^2]$, it follows that

$$\begin{aligned}\mathrm{Var}(\Lambda_{\mathbf{i}}) &\leq \frac{1}{(\mathbb{E}(K_1))^2}C(\pi(x) + o(1))\left((\phi_x(h))^2 + (\phi_x(h))^{a+1/a}\right)\\ &\leq \frac{1}{(\mathbb{E}(K_1))^2}C(\pi(x) + o(1))\left((\phi_x(h))^2 + \phi_x(h)\right)\\ &\leq \frac{C}{(\mathbb{E}(K_1))^2}(\pi(x) + o(1))\phi_x(h)),\end{aligned} \quad (68)$$

which implies that

$$I_{\mathbf{n}} = \sum_{\mathbf{i}\in\mathcal{I}}\mathbb{E}\left(\Lambda_{\mathbf{i}}^2\right) \leq \frac{\widehat{\mathbf{n}}}{(\mathbb{E}(K_1))^2}C(\pi(x) + o(1))(\phi_x(h)). \quad (69)$$

On the other hand, for $R_{\mathbf{n}} = \sum_{\mathbf{i}\neq\mathbf{j}}\mathbb{E}(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}})$ and using the same reasoning as in the previous lemma, we get

$$R_{\mathbf{n}} = \sum_{\mathbf{i}\neq\mathbf{j}}\left|\mathbb{E}\left(\left(\Lambda_{\mathbf{i}}, \Lambda_{\mathbf{j}}\right)\right| = \frac{1}{(\mathbb{E}(K_1))^2}(\pi(x) + o(1))\sum_{\mathbf{i}\neq\mathbf{j}}\left|\mathbb{E}\left(K_{\mathbf{i}}K_{\mathbf{j}}\right) - \mathbb{E}(K_{\mathbf{i}})\mathbb{E}\left(K_{\mathbf{j}}\right)\right|. \quad (70)$$

For the first sum on E_1, the definition of $c_{\mathbf{n}}$ and the assumptions (H1) and (H8) imply that

$$R_{\mathbf{n}}^1 = \sum_{E_1}\left|\mathbb{E}\left(\left(\Lambda_{\mathbf{i}}, \Lambda_{\mathbf{j}}\right)\right| \leq C\widehat{\mathbf{n}}(\pi(x) + o(1))(\phi_x(h))^{1+1/a}. \quad (71)$$

For the second sum on E_2 and by using the bounding of the random variables $K_{\mathbf{i}}$, we deduce from Lemma (3.3) in [30] that

$$R_{\mathbf{n}}^{2}=\sum_{E_2}\left|\left(\mathbb{E}\left(\Lambda_{\mathbf{i}},\Lambda_{\mathbf{j}}\right)\right)\right|\leq\frac{1}{\left(\mathbb{E}\left(K_{1}\right)\right)^{2}}C\left(\pi(x)+o(1)\right)^{2}\sum_{E_2}\varphi(\|\mathbf{i}-\mathbf{j}\|)$$
$$\leq\frac{1}{\left(\mathbb{E}\left(K_{1}\right)\right)^{2}}C\left(\pi(x)+o(1)\right)\widehat{\mathbf{n}}\sum_{\mathbf{i}:\ \|\mathbf{i}\|\geq c_{\mathbf{n}}}\varphi(\|\mathbf{i}\|)$$
$$\leq\frac{1}{\left(\mathbb{E}\left(K_{1}\right)\right)^{2}}C\widehat{\mathbf{n}}c_{\mathbf{n}}^{-\mathrm{Na}}\left(\pi(x)+o(1)\right)\sum_{\mathbf{i}:\ \|\mathbf{i}\|\geq c_{\mathbf{n}}}\|\mathbf{i}\|^{\mathrm{Na}}\varphi(\|\mathbf{i}\|). \tag{72}$$

Then, by condition (9) and the definition of $c_{\mathbf{n}}$ we have

$$R_{\mathbf{n}}=R_{\mathbf{n}}^{1}+R_{\mathbf{n}}^{2}\leq C\widehat{\mathbf{n}}\left(\pi(x)+o(1)\right)\phi_{x}(h). \tag{73}$$

Finally, from (66), (69), and (73), we deduce that

$$\operatorname{Var}\left(\widetilde{g}_{\mathbf{n}}(x)\right)=\operatorname{Var}\left(\frac{S_{\mathbf{n}}}{\widehat{\mathbf{n}}}\right)=O\left(\frac{\phi_{x}(h)}{\widehat{\mathbf{n}}}\right), \tag{74}$$

which implies that $\operatorname{Var}(\widetilde{g}_{\mathbf{n}}(x))\longrightarrow 0$ from where we get (24). □

Proof (Lemma 4). As $B_{\mathbf{n}}(x)=-m(x)\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x))-\mathbb{E}(\widetilde{f}_{\mathbf{n}}(x))/\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x))$ and according to Lemma 3, it suffices to demonstrate that

$$m(x)\mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right)-\mathbb{E}\left(\widetilde{f}_{\mathbf{n}}(x)\right)=o_{a.s}\left(h_{\mathbf{n}}^{\kappa}\right). \tag{75}$$

With the same steps of the proof of Lemma 4 and according to (H2) and (H8), it results that

$$m(x)\mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right)-\mathbb{E}\left(\widetilde{f}_{\mathbf{n}}(x)\right)=\frac{1}{\widehat{\mathbf{n}}\,\mathbb{E}\left(K_{1}\right)}\sum_{\mathbf{i}\in\mathcal{I}}\mathbb{E}\left(m(x)-Y_{\mathbf{i}}\right)\delta_{\mathbf{i}}K_{\mathbf{i}}$$
$$=\frac{1}{\widehat{\mathbf{n}}\,\mathbb{E}\left(K_{1}\right)}\sum_{\mathbf{i}\in\mathcal{I}}\mathbb{E}\left\{\mathbb{E}\left[\left(m(x)-Y_{\mathbf{i}}\right)\delta_{\mathbf{i}}K_{\mathbf{i}}\,\middle|\,X_{\mathbf{i}}\right]\right\}$$
$$=\frac{1}{\widehat{\mathbf{n}}\,\mathbb{E}\left(K_{1}\right)}\sum_{\mathbf{i}\in\mathcal{I}}\mathbb{E}\left[\left(m(x)-m\left(x_{\mathbf{i}}\right)\right)\pi\left(x_{\mathbf{i}}\right)K_{\mathbf{i}}\right]$$
$$\leq\sup_{u\in B(x,h)}|m(x)-m(u)|\pi\left(x_{\mathbf{i}}\right)$$
$$=o\left(h_{\mathbf{n}}^{\kappa}\right). \tag{76}$$

This completes the proof of (25). Now, for proofing (26) follows the same reasoning as in [26] to establish that

$$\sqrt{\widehat{\mathbf{n}}\phi_{x}\left(h_{\mathbf{n}}\right)}\left(\widetilde{g}_{\mathbf{n}}(x)-\mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right)\right)\xrightarrow{D}\mathcal{N}\left(0,\sigma_{0}^{2}\right),\text{ as }\mathbf{n}\longrightarrow\infty\text{ where }\sigma_{0}^{2}(x)=\frac{\beta_{2}}{\beta_{1}^{2}}\frac{\pi(x)}{\tau(x)}. \tag{77}$$

Then, (25) and the condition $\widehat{\mathbf{n}}\phi_{x}(h_{\mathbf{n}})\longrightarrow\infty$ as $\mathbf{n}\longrightarrow\infty$ complete the proof of (26) and therefore the proof of Lemma 4. The proof of (77) is identical to Lemma 5. □

Proof (Lemma 5). According to the fact that $Q_{\mathbf{n}}(x):=m(x)(\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x))-\widetilde{g}_{\mathbf{n}}(x))-(\mathbb{E}(\widetilde{f}_{\mathbf{n}}(x))-\widetilde{f}_{\mathbf{n}}(x))$, we have

$$\sqrt{\frac{\widehat{\mathbf{n}}\phi_x(h)}{\sigma^2(x)}}(Q_{\mathbf{n}}(x)) = \frac{1}{\sqrt{\widehat{\mathbf{n}}\sigma^2(x)}}S_{\mathbf{n}}, \quad \text{where } S_{\mathbf{n}} = \sum_{\mathbf{i}\in I_{\mathbf{n}}} L_{\mathbf{i}}. \tag{78}$$

Thus, the asymptotic normality of $1/\sqrt{\widehat{\mathbf{n}}\sigma^2(x)}S_{\mathbf{n}}$ implies the proof of Lemma 5.

As in [35], this normality is demonstrated by the blocking method. This method is defined by putting into large blocks and small blocks of the random variables $L_{\mathbf{j}}$ by

$$\begin{aligned}
U(\mathbf{n},\mathbf{j},1) &= \sum_{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+1 k=1,\ldots,N}^{j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}} L_{\mathbf{i}},\\
U(\mathbf{n},\mathbf{j},2) &= \sum_{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+1 k=1,\ldots,N-1}^{j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}} \sum_{i_N=j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1}^{(j_N+1)(p_{\mathbf{n}}+q_{\mathbf{n}})} L_{\mathbf{i}},\\
U(\mathbf{n},\mathbf{j},3) &= \sum_{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+1 k=1,\ldots,N-2}^{j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}} \sum_{i_{N-1}=j_{N-1}(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1}^{(j_{N-1}+1)(p_{\mathbf{n}}+q_{\mathbf{n}})} \sum_{i_N=j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+1}^{j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}} L_{\mathbf{i}},\\
U(\mathbf{n},\mathbf{j},4) &= \sum_{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+1 k=1,\ldots,N-2}^{j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}} \sum_{i_{N-1}=j_{N-1}(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1}^{(j_{N-1}+1)(p_{\mathbf{n}}+q_{\mathbf{n}})} \sum_{i_N=j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1}^{(j_N+1)(p_{\mathbf{n}}+q_{\mathbf{n}})} L_{\mathbf{i}},
\end{aligned} \tag{79}$$

and so on. Finally, the last two terms are

$$\begin{aligned}
U\left(\mathbf{n},\mathbf{j},2^{N-1}\right) &= \sum_{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1 k=1,\ldots,N-1}^{(j_k+1)(p_{\mathbf{n}}+q_{\mathbf{n}})} \sum_{i_N=j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+1}^{j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}} L_{\mathbf{i}},\\
U\left(\mathbf{n},\mathbf{j},2^{N}\right) &= \sum_{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1 k=1,\ldots,N}^{(j_k+1)(p_{\mathbf{n}}+q_{\mathbf{n}})} L_{\mathbf{i}},
\end{aligned} \tag{80}$$

where

$$\begin{aligned}
p_{\mathbf{n}} &= \left[\frac{(\widehat{\mathbf{n}}\phi_x(h))^{1/(2N)}}{s_{\mathbf{n}}}\right],\\
q_{\mathbf{n}} &= o\left(\left[\widehat{\mathbf{n}}(\phi_x(h))^{(1+2N)}\right]^{1/(2N)}\right),
\end{aligned} \tag{81}$$

with

$$s_{\mathbf{n}} = o\left(\left[\widehat{\mathbf{n}}(\phi_x(h))^{(1+2N)}\right]^{1/(2N)} q_{\mathbf{n}}^{-1}\right). \tag{82}$$

By (H9), it is simple to show that all sequences $q_{\mathbf{n}}$, $p_{\mathbf{n}}$, and $s_{\mathbf{n}}$ go to infinity.

In the following, we put $m_k = n_k/(p_{\mathbf{n}}+q_{\mathbf{n}})^N$ and, for each integer $i = 1,\ldots,2^N$, we define the random variable $W(i,\mathbf{n})$ by $W(i,\mathbf{n}) = \sum_{\mathbf{j}\in\mathcal{J}} U(\mathbf{n},\mathbf{j},i)$ with $\mathcal{J} = \{0,\ldots,m_1 - 1\}\times\cdots\times\{0,\ldots,m_N - 1\}$. Then, by verifying that

$$S_{\mathbf{n}} = \sum_{\mathbf{i}=1}^{2^N} W(i,\mathbf{n}), \tag{83}$$

the proof of Lemma 5 requires only

$$\frac{W(1,\mathbf{n})}{\sqrt{\widehat{\mathbf{n}}\sigma(x)^2}} \xrightarrow{D} \mathcal{N}(0,1), \tag{84}$$

and

$$\frac{1}{\sqrt{\widehat{\mathbf{n}}}}\left(\sum_{i=2}^{2N} W(i,\mathbf{n})\right) \xrightarrow{P} 0. \tag{85}$$

Clearly, to prove (85), we only need to show that

$$\frac{1}{\widehat{\mathbf{n}}}\mathbb{E}\left[\sum_{i=2}^{2^N} W(i,\mathbf{n})\right]^2 \longrightarrow 0. \tag{86}$$

For this, it suffices to notice that

$$\frac{1}{\widehat{\mathbf{n}}}\mathbb{E}\left[\sum_{i=2}^{2^N} W(i,\mathbf{n})\right]^2 = \frac{1}{\widehat{\mathbf{n}}}\left(\sum_{i=2}^{2^N}\mathbb{E}[W(i,\mathbf{n})]^2 + \sum_{i,j=2,\ldots,2^N, i\neq j}\mathbb{E}[W(i,\mathbf{n})W(\mathbf{n},j)]\right). \tag{87}$$

Then, for all $2 \le i, j \le 2^N$, by the Cauchy–Schwartz inequality, we get

$$\frac{1}{\hat{\mathbf{n}}}\mathbb{E}[W(i,\mathbf{n})W(j,\mathbf{n})] \le \left(\frac{1}{\hat{\mathbf{n}}}\mathbb{E}[W(i,\mathbf{n})]^2\right)^{1/2}\left(\frac{1}{\hat{\mathbf{n}}}\mathbb{E}[W(j,\mathbf{n})]^2\right)^{1/2}. \tag{88}$$

So, to obtain (86), it suffices to prove that

$$\frac{1}{\hat{\mathbf{n}}}\mathbb{E}[W(i,\mathbf{n})]^2 \longrightarrow 0, \quad \forall 2 \le i \le 2^N. \tag{89}$$

We will only demonstrate (89) when $i = 2$ since the other cases are similar. Start with enumerate $U(\mathbf{n},\mathbf{j},2)$ in the arbitrary way $\widehat{U}_1,\ldots,\widehat{U}_M$ and write

$$\begin{aligned}\mathbb{E}[W(2,\mathbf{n})]^2 &= \mathbb{E}\left[\sum_{i=1}^{M}\widehat{U}_i\right]^2 = \sum_{i=1}^{M}\operatorname{Var}\left[\widehat{U}_i\right] + \sum_{i=1}^{M}\sum_{j=1 i\neq j}^{M}\operatorname{Cov}\left(\widehat{U}_i,\widehat{U}_j\right)\\ &= A_1 + A_2.\end{aligned} \tag{90}$$

First, as $(X_{\mathbf{i}},Y_{\mathbf{i}})$ is stationary, then we have

$$\begin{aligned}\operatorname{Var}\left[\widehat{U}_i\right] &= \operatorname{Var}\left[\sum_{\substack{i_k=1\\ k=1,\ldots,N-1}}^{p_{\mathbf{n}}}\sum_{i_N=1}^{q_{\mathbf{n}}} L_{\mathbf{i}}\right]\\ &= p_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\operatorname{Var}\left[L_{\mathbf{i}}\right]\\ &\quad + \sum_{\substack{i_k=1,\\ k=1,\ldots,N-1}}^{p_{\mathbf{n}}}\sum_{i_N=1}^{q_{\mathbf{n}}}\sum_{\substack{j_k=1\\ k=1,\ldots,N-1\\ \mathbf{i}\neq\mathbf{j}}}^{p_{\mathbf{n}}}\sum_{j_N=1}^{q_{\mathbf{n}}}\mathbb{E}\left[L_{\mathbf{i}}L_{\mathbf{j}}\right].\end{aligned} \tag{91}$$

According to Lemma 6 and equation (53), we have $\operatorname{Var}[L_1] \longrightarrow V(x)$. Moreover, we employ Lemma (3.3) in [30] to get

$$\left|\mathbb{E}\left(L_{\mathbf{i}}L_{\mathbf{j}}\right)\right| \le C\phi_x(h)^{-1}\varphi(\|\mathbf{i}-\mathbf{j}\|). \tag{92}$$

So, we deduce that

$$\begin{aligned}\operatorname{Var}\left[\widehat{U}_i\right] &\le Cp_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\left(V(x) + \phi_x(h)^{-1}\sum_{i_k=1,k=1,\ldots,N-1}^{p_{\mathbf{n}}}\sum_{i_N=1}^{q_{\mathbf{n}}}(\varphi(\|\mathbf{i}\|))\right)\\ &\le Cp_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\phi_x(h)^{-1}\sum_{i_k=1,k=1,\ldots,N-1}^{p_{\mathbf{n}}}\sum_{i_N=1}^{q_{\mathbf{n}}}(\varphi(\|\mathbf{i}\|)).\end{aligned} \tag{93}$$

Consequently, we have

$$\frac{1}{\hat{\mathbf{n}}}A_1 \le \frac{1}{\hat{\mathbf{n}}}CMp_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\phi_x(h)^{-1}\sum_{i=q_{\mathbf{n}}}^{\infty} i^{N-1}\varphi(i). \tag{94}$$

From the definition of M and $p_{\mathbf{n}}$ and the fact that

$$\left(p_{\mathbf{n}}+q_{\mathbf{n}}\right)^N p_{\mathbf{n}}^{N-1}q_{\mathbf{n}} = \left(p_{\mathbf{n}}+q_{\mathbf{n}}\right)^{N-1}p_{\mathbf{n}}^{N}\left(\frac{q_{\mathbf{n}}}{p_{\mathbf{n}}}\right) \le \frac{q_{\mathbf{n}}}{p_{\mathbf{n}}}, \tag{95}$$

we have

$$\begin{aligned}\frac{1}{\hat{\mathbf{n}}}CMp_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\phi_x(h)^{-1} &= \frac{1}{\hat{\mathbf{n}}}\hat{\mathbf{n}}\left(p_{\mathbf{n}}+q_{\mathbf{n}}\right)^{-N}p_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\phi_x(h)\\ &\le \left(\frac{q_{\mathbf{n}}}{p_{\mathbf{n}}}\right)\phi_x(h)^{-1}\\ &= q_{\mathbf{n}}\left(\hat{\mathbf{n}}\phi_x(h)\right)^{-1/2N}\phi_x(h)^{-1}s_{\mathbf{n}}\\ &= q_{\mathbf{n}}\left(\hat{\mathbf{n}}\phi_x(h)^{(1+2N)}\right)^{-1/2N}s_{\mathbf{n}}.\end{aligned} \tag{96}$$

Then, by the fact that $s_{\mathbf{n}} = o([\widehat{\mathbf{n}}\phi_x(h)^{(1+2N)}]^{1/(2N)} q_{\mathbf{n}}^{-1})$, this last term converges to 0. Moreover, by (9) with $\gamma > N$, we have

$$\sum_{i=1}^{\infty} i^{N-1}\varphi(i) < \infty. \tag{97}$$

Then, we deduce that

$$\frac{1}{\widehat{\mathbf{n}}}A_1 \longrightarrow 0. \tag{98}$$

For the evaluation of A_2, it suffices to notice by a simple calculation that the sites of r.v.'s $L_{\mathbf{i}}$ which intervenes in the two variables $\widehat{U}_i$ and $\widehat{U}_j$ with $i \neq j$ are spaced by distance of $q_{\mathbf{n}}$ at least. So, (92) and the stationarity of the process imply that

$$\begin{aligned} A_2 &\leq \sum_{j_k=1 k=1,\ldots,N}^{n_k} \sum_{i_k=1 k=1,\ldots,N \|\mathbf{i}-\mathbf{j}\|>q_{\mathbf{n}}}^{n_k} \mathbb{E}\left(L_{\mathbf{i}}L_{\mathbf{j}}\right) \\ &\leq C\phi_x(h)^{-1}\widehat{\mathbf{n}} \sum_{i_k=1,k=1,\ldots,N,\|\mathbf{i}\|>q_{\mathbf{n}}}^{n_k} \varphi(\|\mathbf{i}\|), \end{aligned} \tag{99}$$

and then we can obtain

$$\frac{1}{\widehat{\mathbf{n}}}A_2 \leq C\phi_x(h)^{-1} \sum_{i=q_{\mathbf{n}}}^{\infty} i^{N-1}\varphi(i). \tag{100}$$

However, by assumption (H9) and the definition of $q_{\mathbf{n}}$, we observe that

$$\phi_x(h)^{-1}\sum_{i=q_{\mathbf{n}}}^{\infty} i^{N-1}\varphi(i) \leq \phi_x(h)^{-1}\sum_{i=q_{\mathbf{n}}}^{\infty} i^{N-1-\gamma} \leq \phi_x(h)^{-1}\int_{q_{\mathbf{n}}}^{\infty} t^{N-1-\gamma}\mathrm{d}t = C\phi_x(h)^{-1}q_{\mathbf{n}}^{N-\gamma} \longrightarrow 0, \tag{101}$$

which implies that

$$\frac{1}{\widehat{\mathbf{n}}}A_2 \longrightarrow 0. \tag{102}$$

The proof of (85) is, therefore, completed.

It is sufficient to demonstrate the three claims as follows in order to prove (84):

$$Q_1 \equiv \left|\mathbb{E}[\exp[iuW(1,\mathbf{n})]] - \prod_{j_k=0,k=1,\ldots,N}^{r_k-1} \mathbb{E}[\exp[iuU(\mathbf{n},\mathbf{j},1)]]\right| \text{tend to } 0, \tag{103}$$

$$Q_2 \equiv \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{j}\in\mathcal{J}} \mathbb{E}[U(\mathbf{n},\mathbf{j},1)]^2 \longrightarrow V(x), \tag{104}$$

and

$$Q_3 \equiv \widehat{\mathbf{n}}^{-1}\sum_{\mathbf{j}\in\mathcal{J}} \mathbb{E}\left[(U(\mathbf{n},\mathbf{j},1))^2 \mathbf{1}_{\left\{|U(\mathbf{n},\mathbf{j},1)|>\epsilon(V(x)\widehat{\mathbf{n}})^{1/2}\right\}}\right] \longrightarrow 0, \quad \text{for all } \epsilon > 0. \tag{105}$$

Proof of (103). Let us enumerate the r.v. $U(\mathbf{n},\mathbf{j},1)$, $\mathbf{j}\in\mathcal{J}$, in arbitrary manner $\widetilde{U}_1,\ldots,\widetilde{U}_T$ where $T=\prod_{k=1}^{N} r_k$. Then, to prove (103), we will use Lemma 3 in [36] applied to the variables $(\exp(iu\widetilde{U}_1),\ldots,\exp(iu\widetilde{U}_T))$. As $|\prod_{s=j+1}^{T}\exp[iu\widetilde{U}_s]|\le 1$, then

$$\begin{aligned}
Q_1 &= \left|\mathbb{E}[\exp[iuW(\mathbf{n},1)]] - \prod_{j_k=0,k=1,\ldots,N}^{r_k-1}\mathbb{E}[\exp[iuU(\mathbf{n},\mathbf{j},1)]]\right| \\
&= \left|\mathbb{E}\left[\prod_{j_k=0,k=1,\ldots,N}^{r_k-1}\exp[iuU(\mathbf{n},\mathbf{j},1)]\right] - \prod_{j_k=0,k=1,\ldots,N}^{r_k-1}\mathbb{E}[\exp[iuU(\mathbf{n},\mathbf{j},1)]]\right| \\
&\le \sum_{k=1}^{T-1}\sum_{j=k+1}^{M}\left|\mathbb{E}(\exp[iu\widetilde{U}_k]-1)\left(\exp\left[iu\widetilde{U}_j\right]-1\right)\prod_{s=j+1}^{T}\exp[iu\widetilde{U}_s] - \mathbb{E}(\exp[iu\widetilde{U}_k]-1)\mathbb{E}\left(\exp\left[iu\widetilde{U}_j\right]-1\right)\prod_{s=j+1}^{M}\exp[iu\widetilde{U}_s]\right| \\
&= \sum_{k=1}^{T-1}\sum_{j=k+1}^{M}\left|\mathbb{E}(\exp[iu\widetilde{U}_k]-1)\left(\exp\left[iu\widetilde{U}_j\right]-1\right) - \mathbb{E}(\exp[iu\widetilde{U}_k]-1)\mathbb{E}\left(\exp\left[iu\widetilde{U}_j\right]-1\right)\right| \times \left|\prod_{s=j+1}^{T}\exp[iu\widetilde{U}_s]\right| \\
&\le \sum_{k=1}^{T-1}\sum_{j=k+1}^{T}\left|\mathbb{E}(\exp[iu\widetilde{U}_k]-1)\left(\exp\left[iu\widetilde{U}_j\right]-1\right) - \mathbb{E}(\exp[iu\widetilde{U}_k]-1)\mathbb{E}\left(\exp\left[iu\widetilde{U}_j\right]-1\right)\right|.
\end{aligned} \tag{106}$$

For $\mathbf{j}\in\mathcal{J}$, we denote $\widetilde{I}_j$ the set involved with $\widetilde{U}_j=\sum_{\mathbf{i}\in\widetilde{I}(1,\mathbf{n},\mathbf{j})}L_{\mathbf{i}}$: $\widetilde{I}_j=\{\mathbf{i}\colon j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+1\le i_k \le j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}};k=1,\ldots,N\}$. Then, each these of sites $\widetilde{I}_{1\le j\le T}$ contains $p_{\mathbf{n}}^N$ sites and they are at least a distance apart $q_{\mathbf{n}}$. Applying Lemma (3.3) in [30], we get

$$\left|\mathbb{E}[\exp[iu\widetilde{U}_k]-1]\left(\exp\left[iu\widetilde{U}_j\right]-1\right) - \mathbb{E}[\exp[iu\widetilde{U}_k]-1]\mathbb{E}\left[\exp\left[iu\widetilde{U}_j\right]-1\right]\right| \le C\varphi\left(d\left(\widetilde{I}_j,\widetilde{I}_k\right)\right)p_{\mathbf{n}}^N. \tag{107}$$

Hence, it follows from (9), assumption (H9), and the definition of $q_{\mathbf{n}}$ that

$$\begin{aligned}
Q_1 &\le Cp_{\mathbf{n}}^N\sum_{k=1}^{T-1}\sum_{j=k+1}^{T}\varphi\left(d\left(\widetilde{I}_j,\widetilde{I}_k\right)\right) \\
&\le Cp_{\mathbf{n}}^N T\sum_{k=2}^{T}\varphi(d(\widetilde{I}_1,\widetilde{I}_k)) \\
&\le Cp_{\mathbf{n}}^N T\sum_{i=1}^{\infty}\sum_{k:iq_{\mathbf{n}}\le d\left(\widetilde{I}_1,\widetilde{I}_k\right)<(i+1)q_{\mathbf{n}}}\varphi(d(\widetilde{I}_1,\widetilde{I}_k)) \\
&\le Cp_{\mathbf{n}}^N T\sum_{i=1}^{\infty} i^{N-1}\varphi(iq_{\mathbf{n}}) \\
&\le C\widehat{\mathbf{n}}q_{\mathbf{n}}^{-\delta}\sum_{i=1}^{\infty} i^{N-1-\delta}\longrightarrow 0.
\end{aligned} \tag{108}$$

□

Proof of (104). Recall that

$$\begin{aligned}
\frac{1}{\widehat{\mathbf{n}}}\mathbb{E}[W(\mathbf{n},1)]^2 &= \frac{1}{\widehat{\mathbf{n}}}\sum_{j_k=0,k=1,\ldots,N}^{r_k-1}\mathbb{E}[U(\mathbf{n},\mathbf{j},1)]^2 \\
&+ \frac{1}{\widehat{\mathbf{n}}}\sum_{j_k=0,k=1,\ldots,N}^{r_k-1}\ \sum_{\substack{i_k=0,k=1,\ldots,N \\ i_k\ne j_k \text{ for some } k}}^{r_k-1} Cov[U(\mathbf{n},\mathbf{j},1),U(\mathbf{n},\mathbf{i},1)].
\end{aligned} \tag{109}$$

By using the same arguments previously for A_2, we show that the covariance tends to zero. Therefore, the limit in (104) is the same as the limit of $1/\widehat{\mathbf{n}}\mathbb{E}(W(\mathbf{n},1))^2$.

For this, let $S_{\mathbf{n}}'=W(\mathbf{n},1)$ and $S_{\mathbf{n}}''=\sum_{i=2}^{2^N}W(\mathbf{n},i)$, thus

$$\frac{1}{\widehat{\mathbf{n}}}\mathbb{E}\left[S_{\mathbf{n}}'^2\right]^2 = \frac{1}{\widehat{\mathbf{n}}}\mathbb{E}\left[S_{\mathbf{n}}^2\right] + \frac{1}{\widehat{\mathbf{n}}}\mathbb{E}[S_{\mathbf{n}}'']^2 - 2\frac{1}{\widehat{\mathbf{n}}}\mathbb{E}[S_{\mathbf{n}}S_{\mathbf{n}}''], \tag{110}$$

where

$$S_{\mathbf{n}} = S_{\mathbf{n}}' + S_{\mathbf{n}}''. \tag{111}$$

Equations (56) and (86) imply, respectively, that $\widehat{\mathbf{n}}^{-1}\mathbb{E}[S_{\mathbf{n}}^2]=\widehat{\mathbf{n}}^{-1}\mathrm{Var}(S_{\mathbf{n}})\longrightarrow V(x)$ and $\widehat{\mathbf{n}}^{-1}\mathbb{E}[S_{\mathbf{n}}''^2]\longrightarrow 0$. In addition, by Cauchy–Schwartz's inequality, we have

$$\left|\frac{1}{\widehat{\mathbf{n}}}\mathbb{E}[S_{\mathbf{n}}S_{\mathbf{n}}'']\right| \le \frac{1}{\widehat{\mathbf{n}}}\mathbb{E}\left|S_{\mathbf{n}}S_{\mathbf{n}}''\right| \le \left(\frac{1}{\widehat{\mathbf{n}}}\mathbb{E}[S_{\mathbf{n}}]^2\right)^{1/2}\left(\frac{1}{\widehat{\mathbf{n}}}\mathbb{E}[S_{\mathbf{n}}'']^2\right)^{1/2}. \tag{112}$$

It follows that $1/\widehat{\mathbf{n}}\mathbb{E}[S_{\mathbf{n}}'^2]^2\longrightarrow V(x)$ which concludes the proof of (104).

Hence, we get

$$\frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{j}\in\mathcal{J}}\mathbb{E}[U(\mathbf{n},\mathbf{j},1)]^2\longrightarrow V(x),\quad \text{as } \mathbf{n}\longrightarrow\infty. \tag{113}$$

□

Proof of (105). It is clear that $|U(\mathbf{n},\mathbf{j},1)|\le Cp_{\mathbf{n}}^N 1/\sqrt{\phi_x(h)}$ from the fact that we have $|L_{\mathbf{i}}|\le C1/\sqrt{\phi_x(h)}$. Thus, we deduce that

$$Q_3 \le C p_{\mathbf{n}}^{2N} \frac{1}{\widehat{\mathbf{n}}\sqrt{\phi_x(h)}} \sum_{j_k=0,k=1,\dots,N}^{r_k-1} \mathbb{P}\left[|U(\mathbf{n},\mathbf{j},1)| > \epsilon\left(\sqrt{\widehat{\mathbf{n}}V(x)}\right)\right]. \tag{114}$$

On the other hand, as $p_{\mathbf{n}} = \left[(\widehat{\mathbf{n}}\phi_x(h))^{1/(2N)}/s_{\mathbf{n}}\right]$ and $s_{\mathbf{n}} \longrightarrow \infty$, then

$$\begin{aligned}\frac{|U(\mathbf{n},\mathbf{j},1)|}{\left(\sqrt{\widehat{\mathbf{n}}V(x)}\right)} &\le C p_{\mathbf{n}}^{N}\left(\sqrt{\widehat{\mathbf{n}}\phi_x(h)}\right)\\ &= C\left(s_{\mathbf{n}}\right)^{-N} \longrightarrow 0.\end{aligned} \tag{115}$$

So, for $\mathbf{n}$ large enough, it comes that, for all $\mathbf{j} \in \mathcal{J}$, $\mathbb{P}\left[U(\mathbf{n},\mathbf{j},1) > \epsilon\left(\sqrt{\widehat{\mathbf{n}}V(x)}\right)\right] = 0$, where $\epsilon > 0$. Then, setting $Q_4 = 0$ gives the proof. □

Proof (Theorem 1). To prove (19), it suffices to use the fact that

$$\widetilde{m}_{\mathbf{n}}(x) - m(x) - B_{\mathbf{n}}(x) = \frac{B_{\mathbf{n}}(x)\left(\mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right) - \widetilde{g}_{\mathbf{n}}(x)\right) + Q_{\mathbf{n}}(x)}{\widetilde{g}_{\mathbf{n}}(x)}. \tag{116}$$

Then, according to Lemma 5 and the equations (24) and (26), we have

$$\sqrt{\widehat{\mathbf{n}}\phi_x(h_{\mathbf{n}})}\left[\frac{B_{\mathbf{n}}(x)\left(\mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right) - \widetilde{g}_{\mathbf{n}}(x)\right) + Q_{\mathbf{n}}(x)}{\widetilde{g}_{\mathbf{n}}(x)}\right] = O_p(1), \tag{117}$$

which implies that

$$\sqrt{\frac{\widehat{\mathbf{n}}\phi_x(h_{\mathbf{n}})}{\log(\widehat{\mathbf{n}})}}\left|\widetilde{m}_{\mathbf{n}}(x) - m(x) - B_{\mathbf{n}}(x)\right| \text{ converge in probability to } 0. \tag{118}$$

For the proof of (20), it suffices to use the fact that

$$\widetilde{m}_{\mathbf{n}}(x) - m(x) = \widetilde{m}_{\mathbf{n}}(x) - m(x) - B_{\mathbf{n}}(x) + B_{\mathbf{n}}(x), \tag{119}$$

and then, according to (19) and (25) and the condition $\widehat{\mathbf{n}}h_{\mathbf{n}}^{2\kappa}\left(\phi_x(h_{\mathbf{n}})/\log(\widehat{\mathbf{n}})\right) \longrightarrow 0$, (20) is obtained. □

Proof (Theorem 2). Based on the decomposition (116), by Lemmas 3–5 and Slutsky's theorem, we obtain (21). Using again Slutsky's theorem and decomposition (119) and Lemmas 3–5 gives (27) and thus Theorem 2 follows. □

7. Conclusion

This study examines a functional regression model when responses are missing at random and spatial dependence is present. We construct a Nadaraya–Watson kernel estimator for the nonparametric component based on insufficient data and derive the estimator's asymptotic properties, such as probability convergence (with rates) and asymptotic normality under certain weak conditions. Simulation analysis and real data application are performed to illustrate the finite sample behaviors of the suggested estimator. We also consider the missing mechanism to be missing at random based on our small investigation. This issue of nonignorable missing data, which has been extensively researched in traditional statistical analysis, has received little attention in functional data setup.

Data Availability

The hourly pollution data used to support the findings of this study have been deposited in the (hourly_44201_2019.zip) repository (https://aqs.epa.gov/aqsweb/airdata/download_files.html#Raw).

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors thank and extend their appreciation to the funders of this work. This work was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R358), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia; and the Deanship of Scientific Research at King Khalid University through the Research Groups Program (grant number R.G.P.1/177/44).

References

[1] J. Ramsay and B. Silverman, *Functional Data Analysis*, Spinger-Verlag, New York, NY, USA, 2nd edition, 2005.
[2] D. Bosq, "Linear processes in function spaces: theory and applications," *Lectures Notes in Statistics*, Vol. 149, Springer, Berlin, Germany, 2000.
[3] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis. Theory and Practice*, Springer Series in Statistics, New York, NY, USA, 2006.
[4] T. Hsing and R. L. Eubank, *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, John Wiley and Sons, Hoboken, NJ, USA, 2015.
[5] A. Goia and P. Vieu, "An introduction to recent advances in high/infinite dimensional statistics," *Journal of Multivariate Analysis*, vol. 146, pp. 1–6, 2016.
[6] N. Ling and P. Vieu, "Nonparametric modelling for functional data. selected survey and tracks for future," *Statistics*, vol. 52, no. 4, pp. 934–949, 2018.
[7] F. Ferraty and P. Vieu, "Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés," *Comptes Rendus de l'Academie des Sciences-Series I: Mathematics*, vol. 330, no. 2, pp. 139–142, 2000.
[8] F. Ferraty, M. Sued, and P. Vieu, "Mean estimation with data missing at random for functional covariables," *Statistics*, vol. 47, no. 4, pp. 688–706, 2013.
[9] S. Dabo-Niang and N. Rhomari, "Kernel regression estimation in a Banach space," *Journal of Statistical Planning and Inference*, vol. 139, no. 4, pp. 1421–1434, 2009.
[10] J. Mateu and E. Romano, "Advances in spatial functional statistics," *Stochastic Environmental Research and Risk Assessment*, vol. 31, pp. 1–6, 2017.
[11] S. Dabo-Niang, M. Rachdi, and A.-F. Yao, "Kernel regression estimation for spatial functional random variables," *Far East Journal of Theoretical Statistics*, vol. 37, no. 2, pp. 77–113, 2011.

[12] M. Attouch, B. Chouaf, and A. Laksaci, “Nonparametric M-estimation for functional spatial data,” *Communications for Statistical Applications and Methods*, vol. 19, no. 1, pp. 193–211, 2012.

[13] E. Malikov and Y. Sun, “Semiparametric estimation and testing of smooth coefficient spatial autoregressive models,” *Journal of Econometrics*, vol. 199, no. 1, pp. 12–34, 2017.

[14] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics, Hoboken, NJ, USA, 3rd edition, 2020.

[15] R. Trivellore, *Missing Data Analysis in Practice*, Chapman and Hall, Taylor and Francis Group, Milton Park, UK, 2015.

[16] J. W. Graham, *Missing Data Analysis and Design*, Springer, New York, NY, USA, 2020.

[17] M. A. Ali and M. S. Abu-Salih, “On estimation of missing observations in linear regression models,” *The Indian Journal of Statistics*, vol. 50, no. B, pp. 404–411, 1988.

[18] H. R. Siepman and S.-S. Yang, “Generalized least squares estimation of multivariate nonlinear models with missing data,” *Communications in Statistics- Theory and Methods*, vol. 23, no. 6, pp. 1565–1579, 1994.

[19] P. E. Cheng, “Nonparametric estimation of mean functionals with data missing at random,” *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 81–87, 1994.

[20] C. K. Chu and P. E. Cheng, “Nonparametric regression estimation with missing data,” *Journal of Statistical Planning and Inference*, vol. 48, no. 1, pp. 85–99, 1995.

[21] T. Nittner, “Missing at random (MAR) in nonparametric regression, a simulation experiment,” *Statistical Methods and Applications*, vol. 12, no. 2, pp. 195–210, 2003.

[22] S. Efromovich, “Nonparametric regression with responses missing at random,” *Journal of Statistical Planning and Inference*, vol. 141, no. 12, pp. 3744–3752, 2011.

[23] S. Efromovich, “Nonparametric regression with predictors missing at random,” *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 306–319, 2011.

[24] G. Boente, W. Gonzalez-Manteiga, and A. Perez-Gonzalez, “Robust nonparametric estimation with missing data,” *Journal of Statistical Planning and Inference*, vol. 139, no. 2, pp. 571–592, 2009.

[25] S. Efromovich, “Nonparametric regression with missing data,” *WIREs Computational Statistics*, vol. 6, no. 4, pp. 265–275, 2014.

[26] N. Ling, L. Liang, and P. Vieu, “Nonparametric regression estimation for functional stationary ergodic data with missing at random,” *Journal of Statistical Planning and Inference*, vol. 162, pp. 75–87, 2015.

[27] M. Rachdi, A. Laksaci, Z. Kaid, A. Benchiha, and F. A. Al Awadhi, “k-Nearest neighbors local linear regression for functional and missing data at random,” *Statistica Neerlandica*, vol. 75, no. 1, pp. 42–65, 2021.

[28] W. Wang and L. F. Lee, “Estimation of spatial panel data models with randomly missing data in the dependent variable,” *Regional Science and Urban Economics*, vol. 43, no. 3, pp. 521–538, 2013.

[29] X. Guyon, *Spatial Statistics and Modeling*, Springer Series in Statistics, Berlin, Germany, 2010.

[30] M. Carbon, L. T. Tran, and B. Wu, “Kernel density estimation for random fields (density estimation for random fields),” *Statistics and Probability Letters*, vol. 36, no. 2, pp. 115–125, 1997.

[31] E. Meyer-Wolf and O. Zeitouni, “The probability of small Gaussian ellips ODS,” *Annals of Probability*, vol. 21, no. 1, pp. 14–24, 1993.

[32] V. I. Bogachev, “Gaussian measures,” *Mathematical Surveys and Monographs*, p. 433, American Mathematical Society, Providence, RI, USA, 2009.

[33] W. V. Li and Q.-M. Shao, “Gaussian processes: inequalities, small ball probabilities and applications,” *Handbook of Statistics*, vol. 19, pp. 533–597, 2001.

[34] M. Hallin, Z. Lu, and K. Yu, “Local linear spatial quantile regression,” *Bernoulli*, vol. 15, no. 3, pp. 659–686, 2009.

[35] L. T. Tran, “Kernel density estimation on random fields,” *Journal of Multivariate Analysis*, vol. 34, no. 1, pp. 37–53, 1990.

[36] B. S. Nakhapetyan, “An approach to proving limit theorems for dependent random variables,” *Theory of Probability and Its Applications*, vol. 32, no. 3, pp. 535–539, 1988.