Revista Colombiana de Estadística - Theorical Statistics
July 2022, volume 45, issue 2, pp. 391–428
http://doi.org/10.15446/rce.v45n2.98957

Nonparametric Prediction for Spatial Dependent Functional Data Under Fixed Sampling Design

Predicción no paramétrica para datos funcionales dependientes del espacio bajo un diseño de muestreo fijo

Mamadou Ndiaye[1,a], Sophie Dabo-Niang[2,b], Papa Ngom[3,c]

[1]Higher Polytechnic School (ESP), Cheikh Anta Diop University (UCAD), Dakar, Senegal

[2]Department of Mathematics, University of Lille, Villeneuve d'ascq, France

[3]Department of Mathematics and Computer Science, Faculty of Sciences and Technologies, Cheikh Anta Diop University, Dakar, Senegal

Abstract

In this work, we consider a nonparametric prediction of a spatio-functional process observed under a non-random sampling design. The proposed predictor is based on functional regression and depends on two kernels, one of which controls the spatial structure and the other measures the proximity between the functional observations. It can be considered, in particular, as a supervised classification method when the variable of interest belongs to a predefined discrete finite set. The mean square error and almost complete (or sure) convergence are obtained when the sample considered is a locally stationary α-mixture sequence. Numerical studies were performed to illustrate the behavior of the proposed predictor. The finite sample properties based on simulated data show that the proposed prediction method outperforms the classical predictor which not taking into account the spatial structure.

***Key words*:** Functional dependent data; Fixed design; Non-parametric prediction; Supervised classification.

Resumen

En este trabajo consideramos una predicción no paramétrica de un proceso espacial y funcional observado bajo un diseño de muestreo no aleatorio. El predictor propuesto se basa en la regresión funcional y depende de dos núcleos, uno de los cuales controla la estructura espacial y el otro mide

[a]Ph.D. E-mail: mamadoundiaye397@gmail.com
[b]Pr. E-mail: sophie.dabo@univ-lille.fr
[c]Pr. E-mail: papa.ngom@ucad.edu.sn

la proximidad entre las observaciones funcionales. Esta metodología puede considerarse, en particular, como una nueva herramienta de clasificación supervisada cuando la variable de interés pertenece a un conjunto finito discreto predefinido. El error cuadrático medio y la convergencia casi completa (o certera) se obtienen cuando la muestra considerada es una secuencia α-mixta localmente estacionaria. Además, en este estudio se han realizado estudios numéricos para ilustrar el comportamiento de nuestro predictor. Esta aplicación mediante simulación de un modelo numérico muestra que el método de predicción propuesto supera al predictor clásico que no tiene en cuenta la estructura espacial.

***Palabras clave*:** Clasificación supervisada; Datos funcionales dependientes; Diseño fijo; Predicción no paramétrica.

1. Introduction

Functional data analysis (FDA) deals with the analysis and theory of data that are in the form of functions, curves, images and shapes, or more general complex objects (Wang et al., 2016).

In the last decade, FDA has undergone a large development in a wide variety of fields such as ecology, Embling et al. (2012),Yen et al. (2014),Yan et al. (2015),Yates et al. (2021); medicine, Sørensen et al. (2013), Oshinubi et al. (2022), Wu & Li (2022) or environmental sciences, Giraldo et al. (2011), Torres et al. (2011), Dabo-Niang et al. (2010); monitoring networks of the weather and pollutants see Escabias et al. (2005), Ignaccolo et al. (2013), among others. It has been applied, for modeling purposes, in many areas that require spatial statistics; a branch of mathematics that studies spatially dependent processes.

Functional data can be recored both in temporal and in spatial setting. The analysis of processing of spatially distributed information that is measured, in continuously way, in time or in space/space-time uses functional modeling. It have contributed to the development of new mathematical theories in functional statistics; with the emergence of continuous and spatially dependent data. It gave birth to a branch of mathematics which is a combination of spatial statistics and functional statistics. Indeed, Ruiz-Medina (2011) recognizes the necessity of new developments in spatial correlated functional data and extends the real-valued spatial autoregressive model and the spatial moving average model to stochastic processes taking values in Hilbert spaces. The dimension reduction methodology based on eigenfunctions basis of the auto-covariance operator has been used in Ruiz-Medina M (2012) and Ruiz-Medina et al. (2015). FDA applies tools on multivariate spatial statistics to address questions about prediction at un-sampled locations, classifying spatial curves with unsupervised classification methods or/and discrimination rules, estimating relationships between a primary variable and independent variables.

They are many situations in which one may wish to study the link between two variables, with the main goal is to able to predict new values of one of them, or classify it in k-given homogenous groups inside of each one. For example, in marine biology, it is interesting to see the effect of the environment or other ecology parameter to the variability of biomass and spatial distribution of one specie or a group of species in marine wildlife fauna. In many ecological studies, counts or biomass of interacting species are collected from several sites. Such data are often very sparse, high-dimensional and include highly correlated responses, and the main aim of the statistical analysis is to understand relationships among such multiple, correlated responses Niku et al. (2019), Niku et al. (2021). In Environmental & socio-economic studies, some aquatic species are useful as indicators to study ecosystem health and habitat quality. Amphibian species have been considered as useful ecological indicators. Soltysiak et al. (2016) use a Machine-learning methods in the classification of water bodies. The later use Amphibian species as indicators of environmental contamination. In ecosystemic approach, better understanding interaction between species and environment helps to monitoring fisheries in the context of climate change.

In oil research, we can be interested in the prediction of the physical parameters of oil layers by taking into account other parameters available in oil fields Baouche (2015). The explanatory relationship between variables (response variable and co-variables) is widely studied, in prediction and classification problems, using classical FDA methods. Functional analysis in spatial statistics thus makes it possible to build, among other things, regression, prediction and classification methods, see Ferraty & Vieu (2006), Mateu & Romano (2017), Cuesta-Albertos et al. (2017), Li et al. (2018), Ballari et al. (2018), Chen et al. (2019), among others. However, this study is relatively limited from a theoretical and practical point of view. In addition, it relates, in large part, to parametric/and semi-parametric models Menafoglio et al. (2013), Zhang (2019), Menafoglio (2021), Menafoglio et al. (2022). The high intrinsic dimensionality of these data poses challenges both for theory and computation, that vary with how the functional data were sampled. This means that parametric/and semi-parametric models are not suitable in this new spatio-functional context. Therefore, the modeling of these complex problems uses non-parametric regression models as alternative methods especially when the relationship between the two variables of interest is not linear.

Nonparametric prediction problem, in the spatial setting, has been widely studied in the literature when variables are of finite dimension Dabo-Niang et al. (2016).

During the first half of the 20th century, spatial prediction was widely studied in the scope of geostatistics, commonly known as Kriging. The latter is a spatial linear interpolation method Cressie (1993). Krigging has been applied to a various number of areas including marine biology to evaluate fish abundances Rivoirard et al. (2000). Other methods have also been applied such as K-function Ripley (1987), Heppell et al. (1999), Gardner et al. (2008), Lefort et al. (2011), SDMs model using conventional statistical methods as Generalized linear models and Generalized additive models Young & Carr (2015), Luan et al. (2018), Pollock et al. (2014).

Nowadays there is a dynamic on the development of non-parametric methods for spatial prediction and classification. The first results in this direction are those of Biau and Cadre Biau & Cadre (2004), on kernel prediction of a strictly stationary random field indexed in $(\mathbb{N}^*)^N$. Later, Dabo-Niang & Yao (2007) have contributed to Biau and Cadre's Biau & Cadre (2004) investigation since they were interested in kernel regression estimation and prediction for continuously indexed strictly stationary random fields. In Dabo-Niang et al. (2016), non-parametric prediction for spatial multivariate setting is considered. The latter deals with fixed design and controls the geographic proximity of the spatial sites. In Menezes et al. (2010), non-parametric kernel prediction is considered for spatial stochastic processes when a stochastic sampling design is assumed for the sample locations. In Rachdi et al. (2021) two main aspects of the statistical analysis, namely the parametric and nonparametric approaches are considered to construct and compare four estimators of the regression quantile. Precisely, using a parametric approach, they construct two estimators that are based respectively on the B-spline smoothing and the PCA regression. The other two estimators are constructed using a non-parametric approach, namely the local constants method and the local linear method. Then, they establish the asymptotic properties of the four constructed estimators. Ternynck (2014) considered a functional spatial regression model estimated for strict stationary processes using a kernel method accounting the spatial proximity of locations.

Discrimination kernel rule has been investigated extensively in the literature particularly for independent or time-series data Paredes & Vidal (2006), Devroye et al. (1994), Devroye & Wagner (1982), Hastie & Tibshirani (1996) among others, see the monograph of Biau & Devroye (2015) for more details. Recently, Younso (2017) has addressed a discrimination kernel rule for multivariate strictly stationary spatial processes $(X_{\mathbf{i}} \in \mathbb{R}^d)_{\mathbf{i}\in\mathbb{N}^N}$ and binary spatial classes $(Y_{\mathbf{i}} \in \{0,1\})_{\mathbf{i}\in\mathbb{N}^N}$. Ahmed et al. (2019) proposed a spatial k-nearest neighbors classification rule for multivariate data.

In the functional context, Ferraty & Vieu (2006) investigated a non-parametric prediction and a discrimination kernel rule for independent data and dependent non-spatial data. In Cuevas et al. (2007) and Cuesta-Albertos et al. (2017) a DD^G-classification and a Depth classification rules are proposed. In Xiaoying et al. (2021), a classification of functional data is discussed. Firstly, it preprocesses the abnormal curve based on the centrality and externality of the functional data depth; then combines the functional data non-parametric classification method to calculate the posterior probability value of the given curve belonging to each category, and classify the unknown curve according to the principle of maximum posterior probability, see Xiaoying et al. (2021).

In this work[1], we extend the regression estimate of Ternynck (2014) to a non-strictly stationary sequences and establish the uniform convergence, in addition we investigate a kernel classification rule. Namely, our aim is to propose both kernel method spatial prediction and, in particular, a supervised classification approach in a spatio-functional setting. The propoed model is based on the fact that the response variable of interest is real-valued while the covariate is functional nature. The sample used comes from to a spatial sequence locally and

[1] this come from chapter 4 of this thesis Ndiaye et al. (2020)

identically distributed, contrary to that used in the work of Ternynck (2014). The originality of the suggested method takes advantages of that previously considered by Ternynck (2014), Dabo-Niang et al. (2014), in a more general context. In fact, the present contribution goes along the direction of spatial proximity structure into a kernel predictor and takes advantages of these previous works Francisco-Fernandez & Opsomer (2005), Francisco-Fernández et al. (2012), Dabo-Niang et al. (2014), Ternynck (2014).

2. Regression Model and Predictor

Denote the integer lattice points in the N-dimensional Euclidean space by $\mathbb{Z}^N$, $N \geq 1$. We consider a spatial process $\{Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}}), \mathbf{i} \in \mathbb{Z}^N\}$ defined over some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

A point in bold $\mathbf{i} = (i_1, \dots, i_N) \in \mathbb{Z}^N$ will be referred as a site. Suppose $X_{\mathbf{i}}$ takes values in a separable semi-metric space $(\mathcal{E}, d(\cdot, \cdot))$ (of eventually infinite dimension) (i.e. $X_{\mathbf{i}}$ is a functional random variable and $d(.,.)$ a semi-metric) and $Y_{\mathbf{i}}$ takes values in $\mathbb{R}$. In the following, $\|\cdot\|$ will denote any norm in $\mathbb{R}^d$ or $\mathbb{R}^N$ (there will be no ambiguity since the vectors of $\mathbb{R}^N$ are in bold), C and C' will indicate some arbitrary positive constants that may vary from line to line, for each real u, $\lfloor u \rfloor$ will indicate the integer part of u. Moreover, we write $u_{\mathbf{n}} = O(v_{\mathbf{n}})$ means that $\exists\, C$ such that $|u_{\mathbf{n}}/v_{\mathbf{n}}| \leq C$ as $v_{\mathbf{n}} \to \infty$ and $u_{\mathbf{n}} = o(v_{\mathbf{n}})$ means that $|u_{\mathbf{n}}/v_{\mathbf{n}}| \to 0$ as $v_{\mathbf{n}} \to \infty$, where $\mathbf{n} \in \mathbb{R}^N$.

As it is classically assumed in the literature, the process under study $\{Z_{\mathbf{i}}\}$ is observable over the rectangular domain $\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} = (i_1, \dots, i_N), 1 \leq i_k \leq n_k,\ k = 1, \dots, N\}$, where a point $\mathbf{i} \in \mathbb{Z}^N$ refers to a site. Let $\mathbf{n} = (n_1, \dots, n_N)$ and pose $\widehat{\mathbf{n}} = n_1 \times \cdots \times n_N$, the sample size. From now on, we assume, for seek of simplicity, that $n_1 = n_2 = \ldots = n_N = n$ El Machkouri (2007), El Machkouri & Stoica (2010), El Machkouri (2011), but the following results can be extended to a more general framework. We write $\mathbf{n} \to \infty$ if $n \to \infty$. For each site $\mathbf{i}_0$, let $k_{\mathbf{n}} = k_{\mathbf{n},\mathbf{i}_0} = \sum 1_{[\|\mathbf{i}-\mathbf{i}_0 \leq d_{\mathbf{n}}\|]}$ denote the number of neighbors $\mathbf{i}$ for which the distance between $\mathbf{i}_0$ is less than or equal to distance $d_{\mathbf{n}} > 0$ such that $d_{\mathbf{n}} \to \infty$ as $\mathbf{n} \to \infty$. This last assumes the proximity between locations (eventually) increases as the sample size increases.

We do not suppose strict stationarity. We will assume that the variables $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}}$ are locally identically distributed (see for instance Klemelä (2008) who considered density estimation for locally identically time-series data): a sufficient number of $(X_{\mathbf{i}}, Y_{\mathbf{i}})$ has a distribution close to that of a couple (X, Y).

The main goal is to predict the spatial process $\{Y_{\mathbf{i}}, \mathbf{i} \in \mathbb{Z}^N\}$ in some unobserved locations, particularly at an unobserved site $\mathbf{i_0} \in \mathcal{I}_{\mathbf{n}}$ under the information that can be drawn on $X_{\mathbf{i_0}}$ and observations $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i} \in \mathcal{O}_{\mathbf{n}}}$, where $\mathcal{O}_{\mathbf{n}}$ is the observed spatial set of finite cardinality tending to ∞ as $\mathbf{n} \to +\infty$ and contained in $\mathcal{I}_{\mathbf{n}}$, with $\mathbf{i_0} \notin \mathcal{O}_{\mathbf{n}}$. Let $(X_{\mathbf{i_0}}, Y_{\mathbf{i_0}})$ be of same distribution as (X, Y)).

One may imagine that when $\mathbf{i}$ is close to some $\mathbf{i_0}$, and if there is enough sites $\mathbf{i}$ closed to $\mathbf{i_0} \notin \mathcal{I}_{\mathbf{n}}$, then sequence $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}}$ may be used to predict $Y_{\mathbf{i_0}}$, under the condition $X_{\mathbf{i}_0} = x_{\mathbf{i_0}}$, denoted x in the following with abuse of notation.

We suppose that the spatial process satisfies the following non-parametric regression model:

$$Y_{\mathbf{i}} := r(X_{\mathbf{i}}) + \varepsilon_{\mathbf{i}} \tag{1}$$

where $r(.) = \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}} = .)$, is assumed to be independent of $\mathbf{i}$, the noise $\varepsilon_{\mathbf{i}}$ is centered, α-mixing and independent of $X_{\mathbf{i}}$. Let $\mathbb{E}|Y_{\mathbf{i}}| < \infty$.
Let the following regression estimator where we assume that the observed region $\mathcal{O}_{\mathbf{n}}$ is the rectangular domain $\mathcal{I}_{\mathbf{n}}$:

$$r_{\mathbf{n}}(x) = \begin{cases} \dfrac{g_{\mathbf{n}}(x)}{f_{\mathbf{n}}(x)}, \text{if } f_{\mathbf{n}}(x) \neq 0, \\ \\ \dfrac{1}{\widehat{\mathbf{n}}} \displaystyle\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}} \quad \text{otherwise,} \end{cases} \tag{2}$$

where the functions $f_{\mathbf{n}}$ and $g_{\mathbf{n}}$ are defined, respectively, by

$$f_{\mathbf{n}}(x) = \frac{1}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_1\left(\frac{d(x, X_{\mathbf{i}})}{b_{\mathbf{n}}}\right) K_{2,\rho_{\mathbf{n}}}\left(\|\mathbf{i}_0 - \mathbf{i}\|\right),$$

and

$$g_{\mathbf{n}}(x) = \frac{1}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}} K_1\left(\frac{d(x, X_{\mathbf{i}})}{b_{\mathbf{n}}}\right) K_{2,\rho_{\mathbf{n}}}\left(\|\mathbf{i_0} - \mathbf{i}\|\right),$$

with $a_{\mathbf{n}} = \displaystyle\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_{2,\rho_{\mathbf{n}}}\left(\|\mathbf{i_0} - \mathbf{i}\|\right) \mathbb{E}\left[K_1\left(\frac{d(x, X_{\mathbf{i}})}{b_{\mathbf{n}}}\right)\right]$,
where $K_{2,\rho_{\mathbf{n}}}\left(\|\mathbf{i_0} - \mathbf{i}\|\right) = K_2\left(\frac{\|(\mathbf{i_0}-\mathbf{i})/\mathbf{n}\|}{\rho_{\mathbf{n}}}\right) = K_2\left(\frac{\|\mathbf{i_0}-\mathbf{i}\|}{n\rho_{\mathbf{n}}}\right)$ $\left(\frac{\mathbf{i}}{\mathbf{n}} = (\frac{i_1}{n}, \frac{i_2}{n}, \dots \frac{i_N}{n})\right)$, $b_{\mathbf{n}}$ and $\rho_{\mathbf{n}}$ are bandwidths tending to zero; K_1 and K_2 are kernels, defined in hypothesis **H1**.

Hereinafter, we assume that $k_{\mathbf{n}} = C_N d_{\mathbf{n}}^N + O(d_{\mathbf{n}}^\beta)$ as $d_{\mathbf{n}} \to +\infty$, $0 < \beta < N$ and C_N is a constant that depends on N.

Taking the Euclidean distance and if $N = 2$ (square grid), we have $k_{\mathbf{n}} \leq 4d_{\mathbf{n}}^2 - 4d_{\mathbf{n}} + 4$ which leads to $k_{\mathbf{n}} = O(d_{\mathbf{n}}^2) = O(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^2)$.

Recall that the main application of the above regression estimate is the prediction of the unobserved value $Y_{\mathbf{i}_0}$ at a location $\mathbf{i}_0$ using a sufficient number of observations $(X_{\mathbf{i}}, Y_{\mathbf{i}})$ available at neighbor locations. For that, let the sample set of sites be $\mathcal{O}_{\mathbf{n}} = \mathcal{I}_{\mathbf{n}} \setminus \{\mathbf{i}_0\}$, with optimal bandwidths $b_{\mathbf{n}}^\sharp$ and $\rho_{\mathbf{n}}^\sharp$ (detailed in the prediction procedure of Section 5). The predictor is defined as:

$$\widehat{Y}_{\mathbf{i}_0}^\sharp = \frac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}} K_1\left(\frac{d(x,X_{\mathbf{i}})}{b_{\mathbf{n}}^\sharp}\right) K_{2,\rho_{\mathbf{n}}^\sharp}\left(\|\mathbf{i}_0 - \mathbf{i}\|\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_1\left(\frac{d(x,X_{\mathbf{i}})}{b_{\mathbf{n}}^\sharp}\right) K_{2,\rho_{\mathbf{n}}^\sharp}\left(\|\mathbf{i}_0 - \mathbf{i}\|\right)}, \tag{3}$$

if the denominator is not null otherwise the predictor is equal to the empirical mean. The accuracy of (3) will be compared with the following one that does not take into account the spatial structure:

$$\widehat{Y}^{\star}_{\mathbf{i}_0} = \frac{\sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} Y_\mathbf{i} K_1\left(\frac{d(x,X_\mathbf{i})}{b^{\star}_\mathbf{n}}\right)}{\sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} K_1\left(\frac{d(x,X_\mathbf{i})}{b^{\star}_\mathbf{n}}\right)}, \tag{4}$$

with $b^{\star}_\mathbf{n}$ an optimal bandwidth detailed in Section 5.1.
Remark that equation (4) is based on the classical non-parametric regression estimator Dabo-Niang et al. (2011) without the second kernel on the locations.

$$r^{cl}_\mathbf{n}(x) = \frac{\sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} Y_\mathbf{i} K_1\left(\frac{d(x,X_\mathbf{i})}{b_\mathbf{n}}\right)}{\sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} K_1\left(\frac{d(x,X_\mathbf{i})}{b_\mathbf{n}}\right)}. \tag{5}$$

3. Large Sample Properties and Assumptions

We first introduce some mixing assumptions. In fact, to take into account the spatial dependency, we assume that the process $\{Z_\mathbf{i} = (X_\mathbf{i}, Y_\mathbf{i}),\ \mathbf{i} \in \mathbb{Z}^N\}$ satisfies a mixing condition defined in Carbon et al. (1997) as follows: there exists a function $\chi(t) \searrow 0$ as $t \to \infty$, such that

$$\begin{aligned} \alpha(\sigma(S),\sigma(S')) &= \sup\{|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A\in\sigma(S), B\in\sigma(S')\},\\ &\leq \psi(\mathrm{Card}(S),\mathrm{Card}(S'))\chi(\mathrm{dist}(S,S')), \end{aligned}$$

where $\mathrm{dist}(S,S')$ is the Euclidean distance between the two finite sets of sites S and S', $\mathrm{Card}(S)$ denotes the cardinality of the set S, $\sigma(S)$ (resp. $\sigma(S')$) denotes the σ-fields generated by $\{Z_\mathbf{i}, \mathbf{i} \in S\}$ (resp. $\{Z_\mathbf{i}, \mathbf{i} \in S'\}$) and $\psi(\cdot)$ is a positive symmetric function nondecreasing in each variable. We recall that the process is said to be strongly mixing if $\psi \equiv 1$. As usual, we will assume that one of both following conditions on $\chi(i)$ is verified. These conditions are defined by

$$\chi(i) \quad \leq \quad Ci^{-\theta}, \text{ for some } \theta > 0, \tag{6}$$

i.e. that $\chi(i)$ tends to zero at a polynomial rate, or

$$\chi(i) \quad \leq \quad C\exp(-si), \text{ for some } s > 0, \tag{7}$$

i.e. that $\chi(i)$ tends to zero at an exponential rate. Concerning the function $\chi(\cdot)$, for the sake of simplicity, we will only study the case where $\chi(\cdot)$ tends to zero at a polynomial rate. However, similar asymptotic results may be obtained with $\chi(\cdot)$ tending to zero at an exponential rate (which implies the polynomial case). Throughout the paper, it will be assumed that ψ satisfies either

$$\forall n, m \in \mathbb{N}, \quad \psi(n,m) \leq C\min(n,m), \tag{8}$$

or

$$\psi(n,m) \leq C(n+m+1)^{\kappa} \tag{9}$$

for some $C > 0$, and some $\kappa \geq 1$. Such functions $\psi(n, m)$ can be found, for instance, in Tran (1990), Carbon et al. (1997), Hallin et al. (2004), Biau & Cadre (2004), Dabo-Niang & Yao (2013).

Let $u_{\mathbf{n}} = \prod_{i=1}^{N} (\log n_i)(\log \log n_i)^{1+\varepsilon}$ for $\varepsilon > 0$, then $\sum_{\mathbf{n} \in \mathbb{N}} 1/\hat{\mathbf{n}} u_{\mathbf{n}} < \infty$.

We will denote by $p_{\mathbf{i}}$ the probability distribution of $X_{\mathbf{i}}$ and by $p_{\mathbf{i},\mathbf{j}}$ the joint probability distribution of $(X_{\mathbf{i}}, X_{\mathbf{j}})$, for all $\mathbf{i}$ and $\mathbf{j}$. The small ball probabilities are denoted by $\varphi_{\mathbf{i},x}(b_{\mathbf{n}}) = \mathbb{P}[X_{\mathbf{i}} \in B(x, b_{\mathbf{n}})]$, with $\varphi_{\mathbf{i},x}(b_{\mathbf{n}})$ tending to zero when $b_{\mathbf{n}}$ goes to zero (see e.g. Ferraty & Vieu (2006) for more details).

For any real-valued random variable Z and integer $p \in \mathbb{N}^*$, let $\|Z\|_p = (\mathbb{E}[|Z|^p])^{1/p}$.

The mean square consistency result of $r_{\mathbf{n}}$ is obtained under the following assumptions on r, the kernel, the bandwidth and local dependence condition.

H1:
- K_1 is defined from $\mathbb{R}^+$ to $\mathbb{R}^+$, and we assume that there exist two constants C_{11} and C_{12} with $0 < C_{11} < C_{12} < \infty$, such that

$$C_{11} \mathbb{I}_{[0,1]}(t) \leq K_1(t) \leq C_{12} \mathbb{I}_{[0,1]}(t).$$

- K_2 is a bounded nonnegative function defined from $\mathbb{R}^+$ to $\mathbb{R}^+$, and we assume that there exist constants C_{21}, C_{22} and ρ such that

$$C_{21} \mathbb{I}_{\{\|s\| \leq \rho\}} \leq K_2(\| s \|) \leq C_{22} \mathbb{I}_{\{\|s\| \leq \rho\}}, \ \forall\, s \in \mathbb{R}^N, \quad 0 < C_{21} \leq C_{22} < \infty, \rho > 0. \tag{10}$$

H2: r is a Lipschitz function, that is $r \in Lip_{\mathcal{E}}$ where

$$Lip_{\mathcal{E}} = \{f : \mathcal{E} \to \mathbb{R}, \exists\, C_3 \in \mathbb{R}_*^+, \forall\, (x, x') \in \mathcal{E}^2, |f(x) - f(x')| < C_3 d(x, x')\}.$$

H3: (i) **Local dependence condition:** For all $\mathbf{i} \neq \mathbf{j} \in \mathbb{N}^N$, $\mathbf{i}, \mathbf{j} \in \mathcal{V}_{\mathbf{i}_0}$ the joint probability distribution $p_{\mathbf{i},\mathbf{j}}$ of $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$ satisfies

$$\exists\, \varepsilon \in (0, 1], p_{\mathbf{i},\mathbf{j}}(B(x, b_{\mathbf{n}}) \times B(x, b_{\mathbf{n}})) \leq C_4 (\varphi_{\mathbf{i},x}(b_{\mathbf{n}}) \varphi_{\mathbf{j},x}(b_{\mathbf{n}}))^{\frac{1+\varepsilon}{2}},$$

for some constant $C_4 > 0$, where $\mathcal{V}_{\mathbf{i}_0} = \{\mathbf{i} \in \mathcal{O}_{\mathbf{n}},\ \|\frac{\mathbf{i} - \mathbf{i}_0}{\mathbf{n}}\| < \rho_{\mathbf{n}}\}$.

(ii) **Small ball probabilities:** For all $\mathbf{i}$ and x, there exist a function $\varphi_x(h) > 0$ tending to zero as h goes to zero such that

$$\sup_{\mathbf{i} \in \mathcal{V}_{\mathbf{i}_0}} |\varphi_{\mathbf{i},x}(h) - \varphi_x(h)| = o(1).$$

Remark 1. These assumptions are standard in the context of spatial non-parametric modeling. Indeed, Assumptions **H1** and **H2** allow to control the bias of the estimator. Assumption **H1** is satisfied, for instance, by several kernels with compact support such as triangular (Bartlett), biweight, triweight, Epanechnikov, Parzen kernels. The Lipschitz condition **H2** allows to obtain rate of convergence whereas a continuity-type model would give only convergence results. Local dependence condition **H3**(i) is a classical condition in kernel estimation of dependent sequences non-necessarily strictly stationary (Bosq, 1998; Carbon et al., 1997; Masry, 2005).

In order to control the constraints on the bandwidth sequence due to the mixing coefficients with polynomial decreasing rate (6), we define

$$\gamma_1 = \frac{2N-\theta}{4N-\theta} \text{ and } \gamma_1^* = \frac{N-\theta}{N(3+2\kappa)-\theta}.$$

The following result gives a bound of the mean squared error of $r_{\mathbf{n}}$

Theorem 1. *Assume that assumptions **H1-H3** hold with* $|Y_{\mathbf{i}}| \leq M$.

1. *If* (8) *is satisfied and*

$$\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})^{\gamma_1} \rho_{\mathbf{n}}^{N\gamma_1} (\log \widehat{\mathbf{n}})^{-\gamma_1} \to \infty \text{ with } \theta > 4N,$$

   or

2. *if* (9) *is satisfied and*

$$\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})^{\gamma_1^*} \rho_{\mathbf{n}}^{N\gamma_1^*} (\log \widehat{\mathbf{n}})^{-\gamma_1^*} \to \infty \text{ with } \theta > (3+2\kappa)N,$$

   then

$$\|r_{\mathbf{n}}(x) - r(x)\|_2 \quad = \quad O\left(b_{\mathbf{n}} + \sqrt{\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})}}\right).$$

Precisely, we have

$$\|r_{\mathbf{n}}(x) - r(x)\|_2 = C_3 \times b_{\mathbf{n}} + \left(2C(2MC_{22} + 2M\sqrt{C_4} + C_0) + 4M\right) \times \sqrt{\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})}},$$

where C *depends on* N *whereas* C_0 *is a constant depending on the constant appearing in Lemma 1.*

Remark 2. The conditions on the bandwidth in Theorem 1 are technical assumptions, which appear (in the proofs when studying the asymptotic behavior of the estimator) in the particular case where the mixing coefficient is such that χ tends to zero at a polynomial rate, for some examples, see Neaderhouser (1980), Rosenblatt (1985). Each of these conditions is related to a specific case of mixing in the spatial context and are used respectively in Neaderhouser (1980) and Takahata (1983).

3.1. Uniform Almost Complete Convergence

We consider a set $\mathcal{D}$ such that $\mathcal{D} \subset \bigcup_{k=1}^{v_{\mathbf{n}}} B_k$ where $B_k = B(x_k, \ell_{\mathbf{n}})$ (note that such set can always be built), $v_{\mathbf{n}} > 0$ is some integer, $x_k \in \mathcal{E}$, $k = 1, \ldots, v_{\mathbf{n}}$, and $\ell_{\mathbf{n}} > 0$. We assume that:

H4 there exist $\Gamma_{\mathbf{i}}(b_{\mathbf{n}}) = \sup_{x\in\mathcal{D}} \varphi_{\mathbf{i},x}(b_{\mathbf{n}})$, $\Gamma(b_{\mathbf{n}}) = \sup_{x\in\mathcal{D}} \varphi_x(b_{\mathbf{n}})$ non increasing positive functions such that:

(i) $\lim_{\mathbf{n}\to\infty}\Gamma_{\mathbf{i}}(b_{\mathbf{n}}) = \lim_{\mathbf{n}\to\infty}\Gamma(b_{\mathbf{n}}) = 0$, and

$$\sup_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} |\Gamma_{\mathbf{i}}(b_{\mathbf{n}}) - \Gamma(b_{\mathbf{n}})| = o(1),$$

(ii) $\lim_{\mathbf{n}\to\infty} \frac{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \Gamma(b_{\mathbf{n}})}{\log \widehat{\mathbf{n}}} \to \infty$,

(iii) $v_{\mathbf{n}} = \widehat{\mathbf{n}}^{\beta}$ for some $\beta > 0$.

H5 Local dependence condition: For and $\mathbf{i} \neq \mathbf{j} \in \mathbb{N}^N$, $\mathbf{i},\mathbf{j} \in \mathcal{V}_{\mathbf{i}_0}$, the joint probability distribution $p_{\mathbf{i},\mathbf{j}}$ of $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$ satisfies

$$\exists\, \varepsilon \in (0,1], p_{\mathbf{i},\mathbf{j}}(B(x,b_{\mathbf{n}}) \times B(x,b_{\mathbf{n}})) \leq C_3''(\Gamma(b_{\mathbf{n}}))^{1+\varepsilon}, \text{ for all } x \in \mathcal{D}.$$

H6 There exists $s > 2$ and $C > 0$ such that for $\mathbf{i},\mathbf{j} \in \mathcal{V}_{\mathbf{i}_0}$,

(i) $\sup_{\mathbf{i}} \mathbb{E}\left(|Y_{\mathbf{i}}|^s \mid X_{\mathbf{i}}\right) < C$,

(ii) $\sup_{\mathbf{i},\mathbf{j}} \mathbb{E}\left(|Y_{\mathbf{i}}Y_{\mathbf{j}}|\,|X_{\mathbf{i}}, X_{\mathbf{j}}\right) < C$ for some constant $C > 0$.

Let us introduce the following functions of the mixing coefficient which is related to the conditions on the bandwidth and the moment of the functional covariate:

$$\begin{aligned}
\theta_1 &= \frac{2s(N-\theta)}{2Ns(\beta+2)+\theta(2-s)}, \quad \theta_2 = \frac{(\theta-2N)s}{2Ns(\beta+2)+\theta(2-s)},\\
\theta_3 &= \frac{2(Ns+\theta)}{2Ns(\beta+2)+\theta(2-s)} \quad \theta_1^* = \frac{s(-N-\theta)}{N(2s\beta+2s\kappa+s+2)+\theta(2-s)},\\
\theta_2^* &= \frac{s(\theta-N)}{N(2s\beta+2s\kappa+s+2)+\theta(2-s)} \quad \theta_3^* = \frac{2(N+\theta)}{N(2s\beta+2s\kappa+s+2)+\theta(2-s)}.
\end{aligned}$$

The following theorem gives an uniform almost sure convergence of the regression estimate.

Theorem 2. *Assume that assumptions* ***H1–H6*** *hold.*

(i) If (8) is satisfied and

$$\widehat{\mathbf{n}}\Gamma(b_{\mathbf{n}})^{\theta_1}\rho_{\mathbf{n}}^{N\theta_1}\left(\log\widehat{\mathbf{n}}\right)^{\theta_2} u_{\mathbf{n}}^{\theta_3} \to \infty \text{ with } \theta > 2Ns(\beta+2)/\left(s-2\right), \tag{11}$$

(ii) or if (9) is satisfied and

$$\widehat{\mathbf{n}}\Gamma(b_{\mathbf{n}})^{\theta_1^*}\rho_{\mathbf{n}}^{N\theta_1^*}\left(\log\widehat{\mathbf{n}}\right)^{\theta_2^*} u_{\mathbf{n}}^{\theta_3^*} \to \infty \text{ with } \theta > N(2s\beta+2s\kappa+s+2)/\left(s-2\right), \tag{12}$$

then

$$\sup_{x\in\mathcal{D}} |r_{\mathbf{n}}(x) - r(x)| = O\left(b_{\mathbf{n}} + \sqrt{\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\,\rho_{\mathbf{n}}^N\Gamma(b_{\mathbf{n}})}}\right) \ a.s.$$

Recall that Dabo-Niang et al. (2011) gave an uniform almost sure bound of their regression estimate on a specific set $\mathcal{C}$ that is $O\left(b_{\mathbf{n}}^{\star}+\sqrt{\frac{\log\widehat{\mathbf{n}}}{\Gamma(b_{\mathbf{n}}^{\star})\widehat{\mathbf{n}}}}\right)$ with $\Gamma(b_{\mathbf{n}}^{\star})=\sup_{x\in\mathcal{C}}\varphi_x(b_{\mathbf{n}}^{\star})$ when the considered process is strictly stationary.

Corollary 1. *Under the conditions of Theorem 2, one can derive an almost sure consistency of the predictor,*

$$\left|\widehat{Y}_{\mathbf{i}_0}-Y_{\mathbf{i}_0}\right| \underset{\mathbf{n}\to\infty}{\longrightarrow} 0 \quad \textit{almost surely}, \tag{13}$$

where

$$\widehat{Y}_{\mathbf{i}_0}=\frac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}} K_1\left(\frac{d(x,X_{\mathbf{i}})}{b_{\mathbf{n}}}\right) K_{2,\rho_{\mathbf{n}}}\left(\|\mathbf{i}_0-\mathbf{i}\|\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_1\left(\frac{d(x,X_{\mathbf{i}})}{b_{\mathbf{n}}}\right) K_{2,\rho_{\mathbf{n}}}\left(\|\mathbf{i}_0-\mathbf{i}\|\right)}, \tag{14}$$

is the predictor of $Y_{\mathbf{i}_0}$ at a location $\mathbf{i}_0$.

4. Application to Supervised Classification Issue

The goal of supervised classification or discrimination is to predict a feature Y lying in a discrete finite set $\{1,\ldots,M\}$, with the help of a variable of interest X. When $M=2$, the problem becomes a binary classification. That may occur when one want to model absence or presence of some phenomena, abundance or not, specie overfishing or not. When $M>2$, we have categorical classification problem. In a context of supervised classification, we aim to predict at a given location, an unknown discrete variable Y given a functional observed variable X. The unknown nature Y is called a *class*, the functional variable X belongs to $(\mathcal{E}, d(\cdot,\cdot))$.
Let $\mathbf{i}_0\in\mathcal{I}_{\mathbf{n}}$ be the location where we want to predict the class using a sample of spatial dependent observations $(X_{\mathbf{i}},Y_{\mathbf{i}})_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}}$, $\mathcal{O}_{\mathbf{n}}\subset\mathcal{I}_{\mathbf{n}}\backslash\mathbf{i}_0$. In the following, we describe a non-parametric spatial functional classification rule. This is done through a kernel estimator derived from the regression estimate (2).

General classification rule (Bayes rule): Given a function x at some station $\mathbf{i}_0\in\mathcal{I}_{\mathbf{n}}$, namely $x=X_{\mathbf{i}_0}$ the purpose is to estimate the $|M|$ posterior probabilities,

$$p_j(x)=P(Y=j/X=x),\ j=1,\ldots,M.$$

Once that the $|M|$ probabilities are estimated $(\widehat{p}_1(x),\ldots,\widehat{p}_M(x))$, the classification rule consists of assigning an incoming functional observation x to the class with highest estimated posterior probability:

$$\widehat{y}(x)=\arg\max_{j\in\{1,\ldots,M\}}\widehat{p}_j(x). \tag{15}$$

Remark that

$$p_j(x)=\mathbb{E}(1_{[Y=j]}|X=x), \tag{16}$$

with $1_{[Y=j]}$ equals to 1, if $Y = j$ and 0 elsewhere. Then, estimations of the posterior probabilities can be expressed as:

$$\widehat{p}_{\mathbf{n},j}(x) = \widehat{p}_{b_{\mathbf{n}},\rho_{\mathbf{n}},j}(x) = \sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} W^{\sharp}_{\mathbf{n},\mathbf{i}_0}(x) 1_{[Y_{\mathbf{i}}=j]}, \tag{17}$$

where

$$W^{\sharp}_{\mathbf{n},\mathbf{i}_0}(x) = \frac{K_1\left(\frac{d(x,X_{\mathbf{i}})}{b_{\mathbf{n}}}\right) K_{2,\rho_{\mathbf{n}}}\left(\|\mathbf{i}_0 - \mathbf{i}\|\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_1\left(\frac{d(x,X_{\mathbf{i}})}{b_{\mathbf{n}}}\right) K_{2,\rho_{\mathbf{n}}}\left(\|\mathbf{i}_0 - \mathbf{i}\|\right)}. \tag{18}$$

As explained in Ferraty & Vieu (2006), the discrimination problem can be viewed as a prediction one since it is related to estimation of conditional expectation of indicator variable (class). So, the asymptotic results stated in the prediction setting remain valid in the discrimination context. Then we state the following theorems; the first gives the point-wise almost complete convergence of the estimator of posterior probabilities whereas the second one gives the uniform almost complete convergence.

Theorem 3. *Under conditions of Theorem 1 and assumption on the continuity on the model (i.e $p_j \in Lip_{\mathcal{E}}$: see Ferraty & Vieu (2006) for this assumption), we have, for $j = 1, \dots, M$,*

$$\|\widehat{p}_{\mathbf{n},j}(x) - p_j(x)\|_2 = O\left(b_{\mathbf{n}} + \sqrt{\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})}}\right). \tag{19}$$

Theorem 4. *Under assumptions of Theorem 2 and $p_j \in Lip_{\mathcal{E}}$, see Ferraty & Vieu (2006) for this assumption, we have, for $j = 1, \dots, M$,*

$$\sup_{x\in\mathcal{D}} |\widehat{p}_{\mathbf{n},j}(x) - p_j(x)| = O\left(b_{\mathbf{n}} + \sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\, \rho_{\mathbf{n}}^N \Gamma(b_{\mathbf{n}})}}\right), \text{ almost surely.}$$

5. Finite Sample Properties With Simulated Data

In this section, we study the performance of the proposed predictor towards some numerical experiment which highlight its importance. The proposed predictor is compared with the classical kernel method which does not take into account the spatial dependency (proximity between locations) proposed in the strict stationary case Biau & Cadre (2004), Dabo-Niang & Yao (2007). Let us first describe the prediction procedure. It allows to compute optimal bandwidths using *leave one cross-validation* approach based on the regression model (1).

5.1. Procedure of Prediction

The choice of bandwidth (even in finite or infinite dimensional setting) is a crucial question in non-parametric estimation. We propose to choose the optimal bandwidths using *leave one cross-validation* approach.

Step 1
Specify sets bandwidths S_1 and S_2 for respectively K_1 and K_2.

Step 2
For each $b_{\mathbf{n}} \in S_1$, $\rho_{\mathbf{n}} \in S_2$ and $\mathbf{i}_0 \in \mathcal{I}_{\mathbf{n}}$, compute equation (2).

Step 3
Compute optimal bandwidths $b_{\mathbf{n}}^{\sharp}$ and $\rho_{\mathbf{n}}^{\sharp}$ by applying a cross-validation procedure over S_1 and S_2. More precisely, consider the following minimization problem i.e determine $b_{\mathbf{n}}$ and $\rho_{\mathbf{n}}$ that minimizing the mean squared error over the $\hat{\mathbf{n}}$ sites

$$\min_{b_{\mathbf{n}} \in S_1, \rho_{\mathbf{n}} \in S_2} \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}_0 \in \mathcal{I}_{\mathbf{n}}} (\widehat{Y}_{\mathbf{i}_0} - Y_{\mathbf{i}_0})^2 \tag{20}$$

and denote them $b_{\mathbf{n}}^{\sharp}$ and $\rho_{\mathbf{n}}^{\sharp}$.

The same procedure is applied to equation (5) for computing $b_{\mathbf{n}}^{\star}$ by replacing $r_{\mathbf{n}}(.)$ by $r_{\mathbf{n}}^{cl}(.)$ in equation (20) minimizing with respect to S_1.

Step 4

For each site $\mathbf{i}_0$, predict $Y_{\mathbf{i}_0}$ by:

- computing the proposed predictor $\widehat{Y}_{\mathbf{i}_0}^{\sharp}$ using $b_{\mathbf{n}}^{\sharp}$ and $\rho_{\mathbf{n}}^{\sharp}$, see equation (3)
- computing $\widehat{Y}_{\mathbf{i}_0}^{\star}$, the one that does not takes into account the spatial proximity, using $b_{\mathbf{n}}^{\star}$, see equation (4)
- comparing $\widehat{Y}_{\mathbf{i}_0}^{\sharp}$ and $\widehat{Y}_{\mathbf{i}_0}^{\star}$, through their prediction errors respectively, using equation (25).

5.2. Simulations Studies

In the following, we let $N = 2$, to illustrate our results, we have done some of simulations based on observations $(X_{(i,j)}, Y_{(i,j)}), 0 \leq i, j \leq 25$ such that $\forall\, i, j$,

$$Y_{(i,j)} = r(X_{(i,j)}) + \varepsilon_{(i,j)} \tag{21}$$

$$= 4A_{(i,j)}^2 + \varepsilon_{(i,j)}, \tag{22}$$

and for $t \in [0, 1]$, $X_{(i,j)}(t)$ is defined according to the following model:

$$X_{i,j}(t) = A_{i,j}^2 * (t - 0.5)^2. \tag{23}$$

Where $A = (A_{i,j})$ and $\varepsilon = (\varepsilon_{i,j})$ are random variables which will be specified later on. Several curves examples of $X_{(i,j)}(t)$, are drawn on Figure **1**(left down panel). An example of the function $r(.)$ could be $r(X) = 2X''$ (where f'' denotes the second derivatives of a function f).

The model (21) is simulated with spatial dependency structure. Thereafter, we denote by $GRF(m, \sigma^2, s)$ a stationary Gaussian random field with mean m and covariance function defined by $C(h) = \sigma^2 \exp(-(\frac{\|h\|}{s}))$, $h \in \mathbb{R}^2$ and $s > 0$.

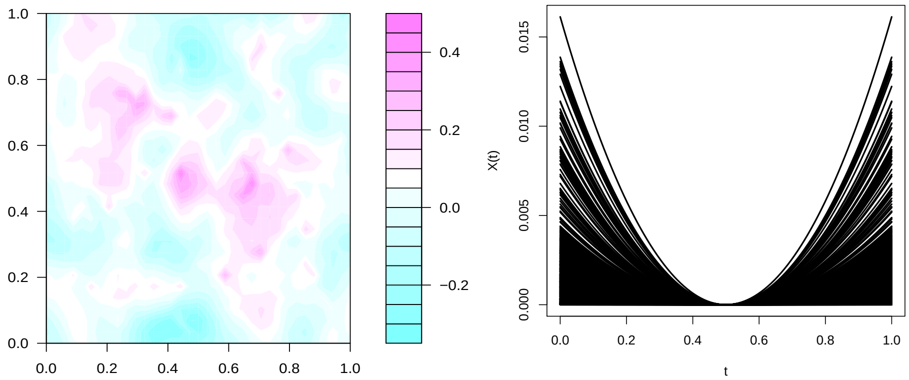

FIGURE 1: Some of simulation when $a = 5$; right panel: field Y; left panel: simulated curves for model 23.

We simulate the model 21 based on $A = (A_{i,j})$. Then, we define the following version for A. $A_{i,j} = D_{i,j}(\sin(2G_{i,j})+2\exp(-16G_{i,j}^2))$; $\varepsilon_{i,j} = GRF(0, .1, 5), G_{i,j} = GRF(0, 5, 3)$ and $D_{\mathbf{i}} = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{j}} \exp\left(-\frac{\|\mathbf{i}-\mathbf{j}\|}{a}\right)$.

$(D_{(i,j)} = \frac{1}{25\times 25}\sum_{1\leq m,t\leq 25} \exp\left(-\frac{\|(i,j)-(m,t)\|}{a}\right))$. The function D is here to ensure and control the spatial mixing condition (even if using the Gaussian Random Fields also brings some spatial dependency). Indeed, our model can be seen verifying a mixing condition with $\alpha(h) \to 0$ at exponential rate. Then, the greater is a, the weaker is the spatial dependency. Furthermore, if $a \to \infty$, $D_{\mathbf{i}} \to 1$. Simulations have done with different values of a which are $a = 5, 10, 20$ and grid size($\widehat{\mathbf{n}} = 35 \times 30 = 1050$).

Along this part, the spatial prediction is computed based several kernels K_1(for observations) and K_2(for sites) respectively. The choice of the semi-metric $d(.,.)$ is important and depends on the information one gets on the data. We consider a semi-metric between curves(observations) based on their first $q = 2$ derivatives. This latter is presented in Ferraty & Vieu (2006).

The construction of the proposed predictor $\widehat{Y}_{\mathbf{i}_0}$ is based on the regression estimator $r_{\mathbf{n}}(.)$. We study its performance and compare it with the one that does not directly take into account the distance between locations noted by $r^{cl}_{\mathbf{n}}(.)$; each studied model is replicated 50 times. Remind that $Y^{\sharp}(.)$ and $Y^{\star}(.)$ are defined by:

$$Y^{\sharp}_{\mathbf{j}}(X_{\mathbf{j}}) = \frac{\sum_{\substack{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}\\ \mathbf{i}\neq\mathbf{j}}} Y_{\mathbf{i}} K_1\left(\frac{d(X_{\mathbf{j}},X_{\mathbf{i}})}{b_{\mathbf{n}^\sharp}}\right) K_{2,\rho^{\sharp}_{\mathbf{n}}(\|\mathbf{j}-\mathbf{i}\|)}}{\sum_{\substack{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}\\ \mathbf{i}\neq\mathbf{j}}} K_1\left(\frac{d(X_{\mathbf{j}},X_{\mathbf{i}})}{b^{\sharp}_{\mathbf{n}}}\right) K_{2,\rho^{\sharp}_{\mathbf{n}}(\|\mathbf{j}-\mathbf{i}\|)}} \text{ and } Y^{\star}_{\mathbf{n}}(X_{\mathbf{j}}) = \frac{\sum_{\substack{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}\\ \mathbf{i}\neq\mathbf{j}}} Y_{\mathbf{i}} K_1\left(\frac{d(X_{\mathbf{j}},X_{\mathbf{i}})}{b^{\star}_{\mathbf{n}}}\right)}{\sum_{\substack{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}\\ \mathbf{i}\neq\mathbf{j}}} K_1\left(\frac{d(X_{\mathbf{j}},X_{\mathbf{i}})}{b^{\star}_{\mathbf{n}}}\right)}. \quad (24)$$

At each replication k, we compute the mean squared error over the $\widehat{\mathbf{n}}$ sites. The bandwidths used at each replication are those obtained using the previous procedure 5.1. For the k^{th} replication, we define the mean squared error of predictions ($MSE_{(k)}^{(+)}$) by:

$$MSE_{(k)}^{(+)} = \frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{n}}} (Y_{\mathbf{n},opt}^{+}(X_{\mathbf{j}}) - Y_{\mathbf{j}})^2 \text{ with } Y_{\mathbf{n},opt}^{+} = Y_{\mathbf{n}}^{\sharp} \text{ or } Y_{\mathbf{n}}^{\star}. \quad (25)$$

The obtained results are summarized in following Tables 1, 2, 3 which give the average mean squared errors (AMSE), standard deviation, the average coefficients of determination ($AR^{\star}$). The last column gives the p-value of a paired t-test performing in order to determine if $MSE^{\sharp}$ is significantly less than $MSE^{\star}$ (the alternative hypothesis is then $H1$: $MSE^{\sharp} < MSE^{\star}$). The quality of estimation is measured by coefficient of determination. We recall that a value of R^2 close to 1 means that the quality of estimation is reliable.

In regardless of any considered cases of the spatial dependency, measured by parameter a, for all kernels, the estimator $r_{\mathbf{n}}^{\sharp}(.)$ leads to better results since the $AMSE^{\sharp}$ is significantly lower than $AMSE^{\star}$. In the other hand we note that the standard deviation of $AMSE^{\sharp}$ is smaller than $AMSE^{\star}$'s one, for all considered cases. In addition, even if the spatial dependency becomes low, $AMSE^{\star}$ stills higher and relatively constant, while $AMSE^{\sharp}$ varies constantly. That highlight, that $r_{\mathbf{n}}^{\sharp}(.)$ method is more adapted to a local data structure and local stationarity(local dependency). Finally we note $AR^{2\sharp}$, is higher than $AR^{2\star}$ for all considered cases, but the difference between them decreases as the value of a increases(less spatial dependency).

6. Conclusion

In this work, we propose a new non-parametric spatial predictor for a local strictly stationary spatial process in a functional setting. The proposed predictor becomes a new method of supervised classification when response variable Y belong to a discret set. The originality of the proposed method is to take into account both the distance between sites and that between functional observations. In the setting of prediction, we give an extension of the recent work of Dabo-Niang et al. (2016) on spatial kernel predictor of a local stationary multivariate process. In the context of Supervised Classification, we contribute first to the kernel discrimination rule of Younso (2017) for multivariate strictly stationary spatial processes and on the other hand the kernel discrimination rule of Ferraty & Vieu (2006) for functional observations. We provide asymptotic results on the predictor. The numerical results show that proposed predictor method outperforms the classical kernel predictor.

Table 1: Simulation results according several kernels, with $a = 5$, $nxy = 1050$

Kernel1	Kernel2	$AMSE^{\sharp}$	Var($AMSE^{\sharp}$)	$AMSE^{\star}$	Var($AMSE^{\star}$)	$AR^{2\sharp}$	$AR^{2\star}$	P-value
Trianglaire	**Triangular**	0,00018	$1,1\,10^{-09}$	0,0080	$2,2\,10^{-06}$	$0,9998$	$0,9905$	****
	Biweight	0,00017	$1,3\,10^{-09}$	0,0081	$2,9\,10^{-06}$	$0,9998$	$0,9902$	****
	Triweight	0,00011	$3,7\,10^{-10}$	0,0079	$2,5\,10^{-06}$	$0,9999$	$0,9903$	****
	Parzen	0,00006	$1,0\,10^{-10}$	0,0081	$3,2\,10^{-06}$	$0,9999$	$0,9906$	****
	Epanechnikov	0,00032	$2,9\,10^{-09}$	0,0077	$1,8\,10^{-06}$	$0,9996$	$0,9907$	****
	Gauss	0,00195	$9,8\,10^{-09}$	0,0079	$2,2\,10^{-06}$	$0,9976$	$0,9904$	****
Biweight	**Triangular**	0,00037	$4\,10^{-09}$	0,0080	$2,4\,10^{-06}$	$0,9996$	$0,9907$	****
	Biweight	0,00020	$1,2\,10^{-09}$	0,0083	$2,3\,10^{-06}$	$0,9998$	$0,9902$	****
	Triweight	0,00019	$9,1\,10^{-10}$	0,0078	$1,8\,10^{-06}$	$0,9998$	$0,9910$	****
	Parzen	0,00012	$4,8\,10^{-10}$	0,0081	$2,4\,10^{-06}$	$0,9999$	$0,9903$	****
	Epanechnikov	0,00218	$1,6\,10^{-07}$	0,0083	$2,6\,10^{-06}$	$0,9975$	$0,9904$	****
	Gauss	0,00007	$2,0\,10^{-10}$	0,0084	$3,5\,10^{-06}$	$0,9999$	0,9898	****
Triweight	**Triangular**	0,00032	$3,6\,10^{-09}$	0,0081	$2,4\,10^{-06}$	$0,9996$	$0,9903$	****
	Biweight	0,00018	$9,8\,10^{-10}$	0,0077	$2\,10^{-06}$	$0,9998$	$0,9911$	****
	Triweight	0,00017	$9,7\,10^{-10}$	0,0080	$2,9\,10^{-06}$	$0,9998$	$0,9908$	****
	Parzen	0,00010	$4\,10^{-10}$	0,0082	$2,2\,10^{-06}$	$0,9999$	$0,9903$	****
	Epanechnikov	0,00183	$9,6\,10^{-09}$	0,0079	$2,1\,10^{-06}$	$0,9979$	$0,9910$	****
	Gauss	0,00006	$2,1\,10^{-10}$	0,0081	$2\,10^{-06}$	$0,9999$	$0,9902$	****
Parzen	**Triangular**	0,00027	$3,3\,10^{-09}$	0,0077	$2\,10^{-06}$	$0,9997$	$0,9908$	****
	Biweight	0,00015	$1,3\,10^{-09}$	0,0080	$5\,10^{-06}$	$0,9998$	$0,9908$	****
	Triweight	0,00014	$5,5\,10^{-10}$	0,0078	$3\,10^{-06}$	$0,9998$	$0,9909$	****
	Parzen	0,00009	$3\,10^{-10}$	0,0083	$2\,10^{-06}$	$0,9999$	$0,9904$	****
	Epanechnikov	0,00162	$1,1\,10^{-07}$	0,0078	$2\,10^{-06}$	$0,9980$	$0,9903$	****
	Gauss	0,00005	$1\,10^{-10}$	0,0081	$2,9\,10^{-06}$	$0,9999$	$0,9905$	****
Epanechnikov	**Triangular**	0,00044	$6,3\,10^{-09}$	0,0083	$3,2\,10^{-06}$	$0,9995$	0,9895	****
	Biweight	0,00025	$2,4\,10^{-09}$	0,0086	$3,5\,10^{-06}$	$0,9997$	$0,9900$	****
	Triweight	0,00024	$1,8\,10^{-09}$	0,0084	$2,7\,10^{-06}$	$0,9997$	$0,9901$	****
	Parzen	0,00015	$6\,10^{-10}$	0,0083	$2\,10^{-06}$	$0,9998$	$0,9903$	****
	Epanechnikov	0,00247	$1\,10^{-07}$	0,0083	$1\,10^{-06}$	$0,9971$	$0,9903$	****
	Gauss	0,00008	$2,8\,10^{-10}$	0,0083	$2\,10^{-06}$	$0,999$	$0,9905$	****
Gauss	**Triangular**	0,00067	$6,6\,10^{-09}$	0,0091	$2\,10^{-06}$	$0,9992$	0,9887	****
	Biweight	0,00041	$3,8\,10^{-09}$	0,0086	$3\,10^{-06}$	$0,9995$	0,9896	****
	Triweight	0,00040	$3,5\,10^{-09}$	0,0090	$3\,10^{-06}$	$0,9995$	0,9892	****
	Parzen	0,00026	$1,1\,10^{-09}$	0,0089	$3\,10^{-06}$	$0,9997$	0,9896	****
	Epanechnikov	0,00319	$4\,10^{-07}$	0,0085	$3\,10^{-06}$	$0,996$	$0,990$	****
	Gauss	0,00015	$5,9\,10^{-10}$	0,0088	$2,8\,10^{-06}$	$0,9998$	$0,9895$	****

TABLE 2: Simulation results according several kernels, with $a = 10$, $nxy = 1050$

Kernel1	Kernel2	$AMSE^\sharp$	Var($AMSE^\sharp$)	$AMSE^\star$	Var($AMSE^\star$)	$AR^{2\sharp}$	$AR^{2\star}$	P-value
Triangular	**Triangular**	0,0005	$6,4\,10^{-09}$	0,0085	$2,2\,10^{-06}$	0,990	0,840	****
	Biweight	0,0005	$3,8\,10^{-09}$	0,0087	$1,5\,10^{-06}$	0,991	0,837	****
	Triweight	0,0003	$2,0\,10^{-09}$	0,0085	$1,3\,10^{-06}$	0,994	0,845	****
	Epanechnikov	0,0009	$1,5\,10^{-08}$	0,0085	$1,5\,10^{-06}$	0,983	0,839	****
	Gaussian	0,0035	$2,6\,10^{-07}$	0,0086	$2,2\,10^{-06}$	0,936	0,842	****
	Parzen	0,0002	$7,0\,10^{-10}$	0,0085	$1,5\,10^{-06}$	0,997	0,839	****
Biweight	**Triangular**	0,0006	$9,0\,10^{-09}$	0,0086	$2,9\,10^{-06}$	0,989	0,841	****
	Parzen	0,0002	$9,9\,10^{-10}$	0,0088	$1,9\,10^{-06}$	0,996	0,841	****
	Biweight	0,0006	$6,5\,10^{-09}$	0,0089	$1,7\,10^{-06}$	0,989	0,829	****
	Triweight	0,0004	$2,1\,10^{-09}$	0,0088	$1,7\,10^{-06}$	0,993	0,836	****
	Epanechnikov	0,0010	$1,6\,10^{-08}$	0,0086	$1,5\,10^{-06}$	0,981	0,838	****
	Gaussian	0,0037	$2,3\,10^{-07}$	0,0088	$1,8\,10^{-06}$	0,932	0,838	****
Triweight	**Triangular**	0,0005	$7,1\,10^{-09}$	0,0085	$2,8\,10^{-06}$	0,990	0,842	****
	Parzen	0,0002	$5,0\,10^{-10}$	0,0085	$1,5\,10^{-06}$	0,997	0,839	****
	Biweight	0,0005	$5,4\,10^{-09}$	0,0088	$1,6\,10^{-06}$	0,991	0,831	****
	Parzen	0,0002	$5,0\,10^{-10}$	0,0085	$1,5\,10^{-06}$	0,997	0,839	****
	Gaussian	0,0034	$4,4\,10^{-07}$	0,0085	$2,8\,10^{-06}$	0,937	0,842	****
	Triweight	0,0003	$1,6\,10^{-09}$	0,0085	$1,5\,10^{-06}$	0,994	0,839	****
Epanechnikov	**Triangular**	0,0007	$1,2\,10^{-08}$	0,0087	$2,9\,10^{-06}$	0,987	0,838	****
	Biweight	0,0007	$8,5\,10^{-09}$	0,0091	$1,7\,10^{-06}$	0,987	0,827	****
	Triweight	0,0004	$3,5\,10^{-09}$	0,0089	$1,7\,10^{-06}$	0,992	0,833	****
	Parzen	0,0002	$1,4\,10^{-09}$	0,0089	$1,9\,10^{-06}$	0,996	0,838	****
	Epanechnikov	0,0012	$2,2\,10^{-08}$	0,0087	$1,6\,10^{-06}$	0,978	0,835	****
	Gaussian	0,0040	$2,9\,10^{-07}$	0,0089	$1,8\,10^{-06}$	0,925	0,835	****
Parzen	**Triangular**	0,0004	$5,1\,10^{-09}$	0,0083	$2,7\,10^{-06}$	0,992	0,845	****
	Biweight	0,0004	$4,1\,10^{-09}$	0,0087	$1,6\,10^{-06}$	0,992	0,834	****
	Triweight	0,0003	$1,1\,10^{-09}$	0,0085	$1,6\,10^{-06}$	0,995	0,841	****
	Parzen	0,0001	$5,7\,10^{-10}$	0,0085	$1,8\,10^{-06}$	0,997	0,845	****
	Epanechnikov	0,0008	$1,0\,10^{-08}$	0,0083	$1,4\,10^{-06}$	0,986	0,842	****
	Gaussian	0,0030	$1,6\,10^{-07}$	0,0085	$1,7\,10^{-06}$	0,944	0,842	****
Gaussian	**Triangular**	0,0010	$2,4\,10^{-08}$	0,0090	$3,1\,10^{-06}$	0,981	0,833	****
	Biweight	0,0009	$1,4\,10^{-08}$	0,0093	$1,8\,10^{-06}$	0,982	0,822	****
	Triweight	0,0006	$5,7\,10^{-09}$	0,0092	$1,8\,10^{-06}$	0,988	0,829	****
	Parzen	0,0004	$2,6\,10^{-09}$	0,0091	$2,0\,10^{-06}$	0,993	0,834	****
	Epanechnikov	0,0016	$3,9\,10^{-08}$	0,0090	$1,7\,10^{-06}$	0,970	0,831	****
	Gaussian	0,0048	$3,8\,10^{-07}$	0,0091	$1,9\,10^{-06}$	0,910	0,831	****

Table 3: Simulation results according several kernels, with $a = 20$, $nxy = 1050$

Kernel1	Kernel2	$AMSE^{\sharp}$	Var($AMSE^{2\sharp}$)	$AMSE^{\star}$	Var($AMSE^{\star}$)	$AR^{2\sharp}$	$AR^{2\star}$	P-value
Epanechnikov	**Triangular**	0,0007	$1,39\,10^{-08}$	0,0087	$3,00\,10^{-06}$	0,998	0,981	****
	Biweight	0,0007	$8,58\,10^{-09}$	0,0091	$1,69\,10^{-06}$	0,999	0,980	****
	Triweight	0,0004	$2,71\,10^{-09}$	0,0089	$1,80\,10^{-06}$	0,999	0,980	****
	Parzen	0,0002	$1,36\,10^{-09}$	0,0089	$2,01\,10^{-06}$	0,997	0,981	****
	Epanechnikov	0,0012	$2,05\,10^{-08}$	0,0087	$1,64\,10^{-06}$	0,991	0,981	****
	Gauss	0,0041	$2,98\,10^{-07}$	0,0089	$1,94\,10^{-06}$	0,997	0,981	****
Triangular	**Triangular**	0,0005	$7,90\,10^{-09}$	0,0085	$2,86\,10^{-06}$	0,991	0,981	****
	Biweight	0,0005	$5,31\,10^{-09}$	0,0089	$1,62\,10^{-06}$	0,998	0,980	****
	Triweight	0,0003	$1,54\,10^{-09}$	0,0087	$1,72\,10^{-06}$	0,998	0,981	****
	Parzen	0,0002	$7,90\,10^{-10}$	0,0085	$2,86\,10^{-06}$	0,999	0,981	****
	Gauss	0,0036	$2,26\,10^{-07}$	0,0085	$1,55\,10^{-06}$	0,992	0,981	****
	Epanechnikov	0,0009	$1,27\,10^{-08}$	0,0085	$1,55\,10^{-06}$	0,998	0,981	****
Biweight	**Triangular**	0,0006	$9,71\,10^{-09}$	0,0086	$2,90\,10^{-06}$	0,999	0,981	****
	Biweight	0,0006	$6,45\,10^{-09}$	0,0089	$1,64\,10^{-06}$	0,999	0,981	****
	Triweight	0,0004	$1,90\,10^{-09}$	0,0088	$1,74\,10^{-06}$	0,997	0,980	****
	Parzen	0,0002	$9,46\,10^{-10}$	0,0088	$1,94\,10^{-06}$	0,992	0,981	****
	Epanechnikov	0,0010	$1,51\,10^{-08}$	0,0086	$1,57\,10^{-06}$	0,998	0,981	****
	Gauss	0,0037	$2,48\,10^{-07}$	0,0088	$1,89\,10^{-06}$	0,997	0,981	****
Triweight	**Triangular**	0,0005	$7,94\,10^{-09}$	0,0085	$2,24\,10^{-06}$	0,992	0,981	****
	Biweight	0,0005	$3,62\,10^{-09}$	0,0088	$1,52\,10^{-06}$	0,998	0,981	****
	Triweight	0,0003	$2,02\,10^{-09}$	0,0086	$1,39\,10^{-06}$	0,999	0,981	****
	Parzen	0,0002	$6,66\,10^{-10}$	0,0082	$1,62\,10^{-06}$	0,999	0,982	****
	Epanechnikov	0,0009	$1,55\,10^{-08}$	0,0085	$1,56\,10^{-06}$	0,998	0,982	****
	Gauss	0,0034	$2,54\,10^{-07}$	0,0086	$2,29\,10^{-06}$	0,992	0,981	****
Parzen	**Triangular**	0,0004	$4,24\,10^{-09}$	0,0085	$1,81\,10^{-06}$	0,998	0,981	****
	Biweight	0,0004	$4,00\,10^{-09}$	0,0084	$1,50\,10^{-06}$	0,999	0,982	****
	Triweight	0,0003	$1,95\,10^{-09}$	0,0084	$2,17\,10^{-06}$	0,999	0,982	****
	Parzen	0,0001	$4,89\,10^{-10}$	0,0084	$1,35\,10^{-06}$	0,999	0,982	****
	Epanechnikov	0,0008	$9,41\,10^{-09}$	0,0084	$1,46\,10^{-06}$	0,998	0,981	****
	Gauss	0,0032	$2,21\,10^{-07}$	0,0086	$1,48\,10^{-06}$	0,993	0,981	****
Gauss	**Triangular**	0,0011	$1,50\,10^{-08}$	0,0090	$1,69\,10^{-06}$	0,997	0,980	****
	Biweight	0,0010	$1,96\,10^{-08}$	0,0090	$3,14\,10^{-06}$	0,997	0,980	****
	Triweight	0,0007	$6,94\,10^{-09}$	0,0093	$1,79\,10^{-06}$	0,998	0,989	****
	Parzen	0,0004	$1,34\,10^{-09}$	0,0092	$2,05\,10^{-06}$	0,999	0,980	****
	Epanechnikov	0,0016	$3,57\,10^{-08}$	0,0090	$1,69\,10^{-06}$	0,996	0,980	****
	Gauss	0,0050	$4,67\,10^{-07}$	0,0092	$1,86\,10^{-06}$	0,989	0,980	****

Acknowledgements

This paper originates from chapter 4 of the Ph.D. thesis entitled "Contribution to spatial and functional statistics: Modeling spatio-temporal of the fishery resources of Senegal" deal with spatial and functional statistics Ndiaye et al. (2020).

[Received: November 2021 — Accepted: May 2022]

References

Ahmed, M. S., Ndiaye, M., Attouch, M. & Dabo-Niang, S. (2019), 'k-nearest neighbors prediction and classification for spatial data', *preprinted* .

Ballari, D., Giraldo, R., Campozano, L. & Samaniego, E. (2018), 'Spatial functional data analysis for regionalizing precipitation seasonality and intensity in a sparsely monitored region: Unveiling the spatio-temporal dependencies of precipitation in ecuador', *International Journal of Climatology* **38**(8), 3337–3354.

Baouche, R. (2015), Prédiction des Paramètres Physiques des Couches Pétrolifères par Analyse des Réseaux de Neurones et Analyse Faciologique., PhD thesis, université M'hamed Bougara. Boumerdès.

Biau, G. & Cadre, B. (2004), 'Nonparametric spatial prediction', *Statistical Inference for Stochastic Processes* **7**(3), 327–349.

Biau, G. & Devroye, L. (2015), *Lectures on the nearest neighbor method*, Springer.

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes: Estimation and prediction*, Vol. 110 of *Lecture Notes in Statist.*, 2nd edn, Springer-Verlag, New York.

Carbon, M., Tran, L. T. & Wu, B. (1997), 'Kernel density estimation for random fields', *Statistics & Probability Letters* **36**(2), 115–125.

Chen, W., Pourghasemi, H. R., Zhang, S. & Wang, J. (2019), A comparative study of functional data analysis and generalized linear model data-mining techniques for landslide spatial modeling, *in* 'Spatial Modeling in GIS and R for Earth and Environmental Sciences', Elsevier, pp. 467–484.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, Vol. 110 of *Wiley Series in Probability and Statistics*, revised edn, Wiley-Interscience.

Cuesta-Albertos, J. A., Febrero-Bande, M. & de la Fuente, M. O. (2017), 'The dd^G-classifier in the functional setting', *Test* **26**(1), 119–142.

Cuevas, A., Febrero, M. & Fraiman, R. (2007), 'Robust estimation and classification for functional data via projection-based depth notions', *Computational Statistics* **22**(3), 481–496.

Dabo-Niang, S., Hamdad, L., Ternynck, C. & Yao, A.-F. c. c. (2014), 'A kernel spatial density estimation allowing for the analysis of spatial clustering: application to Monsoon Asia Drought Atlas data', *Stoch. Environ. Res. Risk Assess* **28**(8), 2075–2099.

Dabo-Niang, S., Rachdi, M. & Yao, A.-F. (2011), 'Kernel regression estimation for spatial functional random variables', *Far East Journal of Theoretical Statistics* **37**(2), 77–113.

Dabo-Niang, S., Ternynck, C. & Yao, A.-F. (2016), 'Nonparametric prediction of spatial multivariate data', *Journal of Nonparametric Statistics* **28**(2), 428–458.

Dabo-Niang, S. & Yao, A.-F. (2007), 'Kernel regression estimation for continuous spatial processes', *Mathematical Methods of Statistics* **16**(4), 298–317.

Dabo-Niang, S. & Yao, A.-F. (2013), 'Kernel spatial density estimation in infinite dimension space', *Metrika* **76**(1), 19–52.

Dabo-Niang, S., Yao, A.-F., Pischedda, L., Cuny, P. & Gilbert, F. (2010), 'Spatial mode estimation for functional random fields with application to bioturbation problem', *Stochastic Environmental Research and Risk Assessment* **24**(4), 487–497.

Devroye, L., Gyorfi, L., Krzyzak, A. & Lugosi, G. (1994), 'On the strong universal consistency of nearest neighbor regression function estimates', *The Annals of Statistics* .

Devroye, L. & Wagner, T. J. (1982), '8 nearest neighbor methods in discrimination', *Handbook of Statistics* .

El Machkouri, M. (2007), 'Nonparametric regression estimation for random fields in a fixed-design', *Stat. Inference Stoch. Process.* **10**(1), 29–47.

El Machkouri, M. (2011), 'Asymptotic normality of the Parzen–Rosenblatt density estimator for strongly mixing random fields', *Statistical Inference for Stochastic Processes* **14**(1), 73–84.

El Machkouri, M. & Stoica, R. (2010), 'Asymptotic normality of kernel estimates in a regression model for random fields', *J. Nonparametr. Stat.* **22**(8), 955–971.

Embling, C. B., Illian, J., Armstrong, E., van der Kooij, J., Sharples, J., Camphuysen, K. C. J. & Scott, B. E. (2012), 'Investigating fine-scale spatio-temporal predator-prey patterns in dynamic marine ecosystems: a functional data analysis approach', *Journal of Applied Ecology* **49**(2), 481–492.

Escabias, M., Aguilera, A. & Valderrama, M. (2005), 'Modeling environmental data by functional principal component logistic regression', *Environmetrics: The official journal of the International Environmetrics Society* **16**(1), 95–107.

Ferraty, F. & Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer.

Francisco-Fernandez, M. & Opsomer, J. D. (2005), 'Smoothing parameter selection methods for nonparametric regression with spatially correlated errors', *Canadian Journal of Statistics* **33**(2), 279–295.

Francisco-Fernández, M., Quintela-del Río, A. & Fernández-Casal, R. (2012), 'Nonparametric methods for spatial regression. an application to seismic events', *Environmetrics* **23**(1), 85–93.

Gardner, B., Sullivan, P. J., Morreale, S. J. & Epperly, S. P. (2008), 'Spatial and temporal statistical analysis of bycatch data: patterns of sea turtle bycatch in the North Atlantic', *Canadian Journal of Fisheries and Aquatic Sciences* **65**(11), 2461–2470.

Giraldo, R., Delicado, P. & Mateu, J. (2011), 'Ordinary kriging for function-valued spatial data', *Environmental and Ecological Statistics* **18**(3), 411–426.

Hallin, M., Lu, Z. & Tran, L. T. (2004), 'Local linear spatial regression', *The Annals of Statistics* **32**(6), 2469–2500.

Hastie, T. & Tibshirani, R. (1996), 'Discriminant adaptive nearest neighbor classification and regression', *Advances in Neural Information Processing Systems* .

Heppell, S. S., Crowder, L. B. & Menzel, T. R. (1999), Life table analysis of long-lived marine species with implications for conservation and management, *in* 'American Fisheries Society Symposium', Vol. 23, pp. 137–148.

Ignaccolo, R., Ghigo, S. & Bande, S. (2013), 'Functional zoning for air quality', *Environmental and ecological statistics* **20**(1), 109–127.

Klemelä, J. (2008), 'Density estimation with locally identically distributed data and with locally stationary data', *J. Time Ser. Anal.* **29**(1), 125–141. http://dx.doi.org/10.1111/j.1467-9892.2007.00547.x

Lefort, R., Fablet, R., Berger, L. & Boucher, J.-M. (2011), 'Spatial statistics of objects in 3-d sonar images: application to fisheries acoustics', *IEEE Geoscience and Remote Sensing Letters* **9**(1), 56–59.

Li, X., Ghosal, S. et al. (2018), 'Bayesian classification of multiclass functional data', *Electronic Journal of Statistics* **12**(2), 4669–4696.

Luan, J., Zhang, C., Xu, B., Xue, Y. & Ren, Y. (2018), 'Modelling the spatial distribution of three portunidae crabs in haizhou bay, China', *PloS one* **13**(11), e0207457.

Masry, E. (2005), 'Nonparametric regression estimation for dependent functional data: asymptotic normality', *Stochastic Process. Appl.* **115**(1), 155–177.

Mateu, J. & Romano, E. (2017), 'Advances in spatial functional statistics'.

Menafoglio, A. (2021), Spatial statistics for distributional data in bayes spaces: From object-oriented kriging to the analysis of warping functions, *in* 'Advances in Compositional Data Analysis', Springer International Publishing, pp. 207–224.

Menafoglio, A., Pigoli, D. & Secchi, P. (2022), 'Mathematical foundations of functional kriging in Hilbert spaces and riemannian manifolds', *Geostatistical Functional Data Analysis* pp. 27–54.

Menafoglio, A., Secchi, P. & Rosa, M. D. (2013), 'A universal kriging predictor for spatially dependent functional data of a Hilbert space', *Electronic Journal of Statistics* **7**(none).

Menezes, R., García-Soidán, P. & Ferreira, C. (2010), 'Nonparametric spatial prediction under stochastic sampling design', *Journal of Nonparametric Statistics* **22**(3), 363–377.

Ndiaye, M., Dabo-Niang, S., Ngom, P., Ciré Elimane, S. & Fall, M. (2020), Contribution to spatial and functional statistics : Modelingspatio-temporal of the fishery resources of Senegal, PhD thesis, Ecole Doctorale Mathématiques et Informatique de l'Université Cheikh Anta Diop de Dakar.

Neaderhouser, C. C. (1980), 'Convergence of block spins defined by a random field', *J. Statist. Phys.* **22**(6), 673–684.

Niku, J., Hui, F. K., Taskinen, S. & Warton, D. I. (2019), 'gllvm: Fast analysis of multivariate abundance data with generalized linear latent variable models in r', *Methods in Ecology and Evolution* **10**(12), 2173–2182.

Niku, J., Hui, F. K., Taskinen, S. & Warton, D. I. (2021), 'Analyzing environmental-trait interactions in ecological communities with fourth-corner latent variable models', *Environmetrics* **32**(6), e2683.

Oshinubi, K., , Ibrahim, F., Rachdi, M. & Demongeot, J. (2022), 'Functional data analysis: Application to daily observation of COVID-19 prevalence in france', *AIMS Mathematics* **7**(4), 5347–5385.

Paredes, R. & Vidal, E. (2006), 'Learning weighted metrics to minimize nearest-neighbor classification error', *IEEE Transactions on Pattern Analysis and Machine Intelligence* .

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A. & McCarthy, M. A. (2014), 'Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (jsdm)', *Methods in Ecology and Evolution* **5**(5), 397–406.

Rachdi, M., Laksaci, A. & Al-Awadhi, F. A. (2021), 'Parametric and nonparametric conditional quantile regression modeling for dependent spatial functional data', *Spatial Statistics* **43**, 100498.

Ripley, B. (1987), Spatial point pattern analysis in ecology, *in* 'Develoments in Numerical Ecology', Springer, pp. 407–429.

Rivoirard, J., Simmonds, J., Foote, K., Fernandes, P. & Bez, N. (2000), *Geostatistics for estimating fish abundance*, Wiley Online Library.

Rosenblatt, M. (1985), *Stationary sequences and random fields*, Birkhauser, Boston.

Ruiz-Medina, M. (2011), 'Spatial autoregressive and moving average Hilbertian processes', *Journal of Multivariate Analysis* **102**(2), 292–305.

Ruiz-Medina, M. D., Anh, V. V., Espejo, R. M., Angulo, J. M. & Frias, M. P. (2015), 'Least-squares estimation of multifractional random fields in a Hilbert-valued context', *Journal of Optimization Theory and Applications* **167**(3), 888–911.

Ruiz-Medina M, E. R. (2012), 'Spatial autoregressive functional plug-in prediction of ocean surface temperature', *Stoch Environ Res Risk Assess 26(3):335–344* .

Soltysiak, M., Blachnik, M. & Dabrowska, D. (2016), 'Machine-learning methods in the classification of water bodies', *Environmental & Socio-economic Studies* **4**(2), 34–42.

Sørensen, H., Goldsmith, J. & Sangalli, L. M. (2013), 'An introduction with medical applications to functional data analysis', *Statistics in medicine* **32**(30), 5222–5240.

Takahata, H. (1983), 'On the rates in the central limit theorem for weakly dependent random fields', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64**(4), 445–456.

Ternynck, C. (2014), 'Spatial regression estimation for functional data with spatial dependency', *SFDS,155, 2* .

Torres, J. M., Nieto, P. G., Alejano, L. & Reyes, A. (2011), 'Detection of outliers in gas emissions from urban areas using functional data analysis', *Journal of hazardous materials* **186**(1), 144–149.

Tran, L. T. (1990), 'Kernel density estimation on random fields', *Journal of Multivariate Analysis* **34**(1), 37–53.

Wang, J.-L., Chiou, J.-M. & Müller, H.-G. (2016), 'Functional data analysis', *Annual Review of Statistics and Its Application* **3**, 257–295.

Wu, H. & Li, Y.-F. (2022), 'Clustering spatially correlated functional data with multiple scalar covariates', *IEEE Transactions on Neural Networks and Learning Systems* pp. 1–15.

Xiaoying, W., Qian, S. & Jialiang, G. (2021), Research on nonparametric classification method of functional data, *in* '2021 2nd International Conference on Education, Knowledge and Information Management (ICEKIM)', IEEE.

Yan, F., Liu, L., Li, Y., Zhang, Y., Chen, M. & Xing, X. (2015), 'A dynamic water quality index model based on functional data analysis', *Ecological Indicators* **57**, 249–258.

Yates, M. C., Derry, A. M. & Cristescu, M. E. (2021), 'Environmental RNA: A revolution in ecological resolution?', *Trends in Ecology & Evolution* **36**(7), 601–609.

Yen, J. D. L., Thomson, J. R., Paganin, D. M., Keith, J. M. & Nally, R. M. (2014), 'Function regression in ecology and evolution: FREE', *Methods in Ecology and Evolution* **6**(1), 17–26.

Young, M. & Carr, M. H. (2015), 'Application of species distribution models to explain and predict the distribution, abundance and assemblage structure of nearshore temperate reef fishes', *Diversity and Distributions* **21**(12), 1428–1440.

Younso, A. (2017), 'On the consistency of a new kernel rule for spatially dependent data', *Statistics & Probability Letters* .

Zhang, H. (2019), Topics in functional data analysis and machine learning predictive inference, PhD thesis, Iowa State University.

Appendix A. Proofs of Theorems 1, 2, 3 and 4

Appendix A.1. Some Preliminary Results for the Proofs

Lemma 1. *Carbon et al. (1997) Let the sets $S_1, S_2, \ldots, S_k$ containing each m sites and such that, for all $i \neq j$, and for $1 \leq i, j \leq k$, $dist(S_i, S_j) \geq \delta_0$. Let $W_1, W_2, \ldots, W_k$ a sequence of random variables with real values and measurable respectively with respect to $\mathcal{B}(S_1), \ldots, \mathcal{B}(S_k)$. Let be W_l with values in $[a, b]$. There exists a sequence of independent random variables $W_1^*, W_2^*, \ldots, W_k^*$ such that W_l^* has the same distribution as W_l and satisfies:*

$$\sum_{l=1}^{k} \mathbb{E}|W_l - W_l^*| \leq 2k(b-a)\psi((k-1)m, m)\chi(\delta_0).$$

Lemma 2. *Tran (1990) Denote by $\mathcal{L}_r(\mathcal{F})$ the class of $\mathcal{F}$-measurable random variables X which satisfy: $\|X\|_r = (\mathbb{E}|X|^r)^{1/r} < \infty$. Suppose that $X \in \mathcal{L}_r(\mathcal{B}(E))$, $Y \in \mathcal{L}_r(\mathcal{B}(E'))$, $1 \leq r, s, t < \infty$ and $\frac{1}{r} + \frac{1}{s} + \frac{1}{t} = 1$. Then,*

$$|\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y| \leq C\|X\|_r\|Y\|_s\{\psi(Card(E), Card(E'))\chi(dist(E, E'))\}^{1/t}.$$

For bounded random variables with probability 1, we have:

$$|\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y| \leq C\{\psi(Card(E), Card(E'))\chi(dist(E, E'))\}.$$

In the following, we will often use the notation $K_{\mathbf{i}}(x) = K_{1\mathbf{i}} K_{2\mathbf{i}}$ and $W_{\mathbf{ni}}(x) = \dfrac{K_{\mathbf{i}}(x)}{\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} K_{\mathbf{j}}(x)}$ with $K_{1\mathbf{i}} = K_1\left(\frac{d(x,X_{\mathbf{i}})}{b_{\mathbf{n}}}\right)$ and $K_{2\mathbf{i}} = K_{2,\rho_{\mathbf{n}}}(\|\mathbf{i}_0 - \mathbf{i}\|)$. By convention, we set $0/0 = 0$, then $\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} W_{\mathbf{ni}}(x) = 0$ or 1. Thus, we have

$$r_{\mathbf{n}}(x) \quad = \quad \begin{cases} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} W_{\mathbf{ni}}(x) Y_{\mathbf{i}} & \text{if} \quad \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} W_{\mathbf{ni}}(x) = 1; \\ \frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} & \text{otherwise.} \end{cases}$$

Let us use the following decomposition:

$$r_{\mathbf{n}}(x) - r(x) = \frac{1}{f_{\mathbf{n}}(x)}\left[(g_{\mathbf{n}}(x) - \mathbb{E}(g_{\mathbf{n}}(x))) - (r(x) - \mathbb{E}(g_{\mathbf{n}}(x)))\right] \tag{26}$$

$$-\frac{r(x)}{f_{\mathbf{n}}(x)}\left[f_{\mathbf{n}}(x) - 1\right]$$

Lemma 3. *Under hypotheses **H1-H3**, we have*

$$\mathbb{E}^{1/2}\left[\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)\mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}}) - r(x)\right]^2 \quad = \quad O(b_{\mathbf{n}}).$$

Lemma 4. *Under the conditions of Theorem 1, we have*

$$\mathbb{E}^{1/2}\left[\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)(Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}}))\right]^2 \quad = \quad O\left(\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})}\right)^{1/2}.$$

Lemma 5. *Under the conditions of Theorem 1, we have*

$$\mathbb{E}^{1/2}\left[\frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} Y_{\mathbf{i}} - r(x)\right]^2 \quad = \quad O\left(\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})}\right)^{1/2}.$$

Define

$$\Lambda_{\mathbf{i}}(x) = \frac{1}{a_{\mathbf{n}}}\left[K_{\mathbf{i}}(x) - \mathbb{E}(K_{\mathbf{i}}(x))\right],$$

$$I_{\mathbf{n}}(x) = \sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} \mathbb{E}\left[(\Lambda_{\mathbf{i}}(x))^2\right] \text{ and } R_{\mathbf{n}}(x) = \sum_{\mathbf{i},\mathbf{k}\in\mathcal{O}_{\mathbf{n}}}\sum_{\mathbf{i}\neq\mathbf{k}} \left|\mathbb{E}\left[\Lambda_{\mathbf{i}}(x)\Lambda_{\mathbf{k}}(x)\right]\right|.$$

Lemma 6. *Under the conditions of Theorem 1, we have*

$$I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x) \quad = \quad O\left(\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})}\right).$$

Appendix A.2. Proofs

Appendix A.2.1. Proof of Theorem 1

Because of the local stationarity defined in assumption (**H3**), the decomposition of $r_{\mathbf{n}}(x) - r(x)$ makes sense over $\mathcal{V}_{\mathbf{i}_0}$. Thus we have

$$\begin{aligned} r_{\mathbf{n}}(x) - r(x) &= \left(\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)\mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}}) - r(x)\right)\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)=1\right\}} \\ &+ \left(\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)(Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}}))\right)\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)=1\right\}} \\ &+ \left(\frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} Y_{\mathbf{i}} - r(x)\right)\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)=0\right\}} := \mathbf{A}+\mathbf{B}+\mathbf{C}. \end{aligned}$$

Applying Minkowski's inequality, we get

$$\|r_{\mathbf{n}}(x) - r(x)\|_2 \leq \mathbb{E}^{1/2}[\mathbf{A}]^2 + \mathbb{E}^{1/2}[\mathbf{B}]^2 + \mathbb{E}^{1/2}[\mathbf{C}]^2. \tag{27}$$

Therefore, Theorem 1 follows from (27) and Lemmas 3, 4 and 5. $\square$

Appendix A.2.2. Proof of Lemma 3

By the Lipschitz condition on Assumption **H2**, there exists a constant $C_3 > 0$ such that

$$\begin{aligned} \mathbb{E}^{1/2}[\mathbf{A}]^2 &\leq \mathbb{E}^{1/2}\left[\left(\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)|r(X_{\mathbf{i}}) - r(x)|\right)\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)=1\right\}}\right]^2 \\ &\leq \mathbb{E}^{1/2}\left[\left(\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)(C_3\times d(X_{\mathbf{i}},x))\right)\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)=1\right\}}\right]^2 \\ &\leq C_3\mathbb{E}^{1/2}\left[\times\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)b_{\mathbf{n}}\right]^2. \end{aligned}$$

Thus, the local stationarity assumption **H3** implies

$$C_3\mathbb{E}^{1/2}\left[\times\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)b_{\mathbf{n}}\right]^2 = O(b_{\mathbf{n}}). \ \square$$

Appendix A.2.3. Proof of Lemma 4

Define

$$\begin{aligned} G(x) &= \left(\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)[Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})]\right)\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)=1\right\}} \\ &:= \frac{e_{\mathbf{n}}(x)}{f_{\mathbf{n}}(x)}\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)=1\right\}}, \end{aligned}$$

where

$$e_{\mathbf{n}}(x) = \frac{1}{a_{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x)[Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})] \quad \text{and} \qquad f_{\mathbf{n}}(x) = \frac{1}{a_{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x).$$

Note that, since $Y_{\mathbf{i}}$ is bounded, we have $\forall\mathbf{i}$, $0 \leq |Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})| \leq 2M$. It follows that $|G(x)| \leq 2M$ and

$$\begin{aligned} |G(x)| &= |G(x)|\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x)>c\right\}} + |G(x)|\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x)\leq c\right\}} \\ &\leq \frac{|e_{\mathbf{n}}(x)|}{f_{\mathbf{n}}(x)}\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x)>c\right\}} + 2M \times \mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x)\leq c\right\}}, \end{aligned}$$

where c is a given constant. Let us take $c = \frac{a_{\mathbf{n}}}{2}$, if $\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x) > c = \frac{a_{\mathbf{n}}}{2}$ then $f_{\mathbf{n}}(x) > \dfrac{a_{\mathbf{n}}}{2a_{\mathbf{n}}} > \dfrac{1}{2}$. It follows that

$$\|G(x)\|_2 \leq 2\|e_{\mathbf{n}}(x)\|_2 + 2M\left(\mathbb{P}\left[\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x) \leq \frac{a_{\mathbf{n}}}{2}\right]\right)^{1/2},$$

and

$$\|e_{\mathbf{n}}(x)\|_2 = \frac{1}{a_{\mathbf{n}}}\left[\mathbb{E}\left(\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \xi_{\mathbf{i}}\right)^2\right]^{1/2},$$

where

$$\xi_{\mathbf{i}} = K_{\mathbf{i}}(x)\left[Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})\right].$$

To prove Lemma 4, we have to show that

$$\|e_{\mathbf{n}}(x)\|_2 = O(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}}))^{-1/2}, \tag{28}$$

and

$$\mathbb{P}\left[\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x) \leq \frac{a_{\mathbf{n}}}{2}\right] \leq O\left(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})\right)^{-1/2}. \tag{29}$$

Observe that, by Assumptions **H1** and **H3**, we have

$$\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\xi_{\mathbf{i}}^2\right] \leq \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[K_{\mathbf{i}}^2(x)\left[Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})\right]^2\right] = 4M^2 \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{2\mathbf{i}}^2 \mathbb{E}[K_{1\mathbf{i}}]^2 \leq 4M^2 C_{22}^2 k_{\mathbf{n}} \varphi_x(b_{\mathbf{n}})$$

$$= O(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})).$$

Now, let $d_{\mathbf{n}}$ be a sequence of real numbers tending to ∞ as $\mathbf{n}\to\infty$ and set

$$S = \{(\mathbf{i},\mathbf{k}) \in \mathcal{V}_{\mathbf{i}_0}^2, \quad \|\mathbf{i}-\mathbf{k}\| \leq d_{\mathbf{n}}\} \text{ and } \; S^c = \{(\mathbf{i},\mathbf{k}) \in \mathcal{V}_{\mathbf{i}_0}^2, \quad \|\mathbf{i}-\mathbf{k}\| > d_{\mathbf{n}}\}.$$

First, see that $\mathbb{E}\left(\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \xi_{\mathbf{i}}\right)^2 = \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}[\xi_{\mathbf{i}}^2] + \sum_{\mathbf{i},\mathbf{k}\in S} \mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{k}}\right] + \sum_{\mathbf{i},\mathbf{k}\in S^c} \mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{k}}\right]$

Using Assumption **H3**, we have

$$\begin{aligned}
\sum_{\mathbf{i},\mathbf{k}\in S} \mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{k}}\right] &\leq 4M^2 \sum_{\mathbf{i},\mathbf{k}\in S} \mathbb{E}\left[K_{\mathbf{i}}(x)K_{\mathbf{k}}(x)\right] \\
&\leq 4M^2 \sum_{\mathbf{i},\mathbf{k}\in S} K_{2\mathbf{i}}K_{2\mathbf{k}}\mathbb{P}\left[(X_{\mathbf{i}},X_{\mathbf{k}}) \in B(x,b_{\mathbf{n}}) \times B(x,b_{\mathbf{n}})\right] \\
&\leq 4M^2 C_4 \sum_{\mathbf{i},\mathbf{k}\in S} \mathbf{1}_{[0,1]}\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{i}_0-\mathbf{i}}{\mathbf{n}}\right\|\right)\mathbf{1}_{[0,1]}\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{i}_0-\mathbf{k}}{\mathbf{n}}\right\|\right)\varphi_x(b_{\mathbf{n}})^{1+\varepsilon} \\
&\leq 4M^2 C_4 \sum_{\mathbf{i},\mathbf{k}\in V_{\mathbf{i}_0}} \mathbf{1}_{[0,1]}\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{i}_0-\mathbf{i}}{d_{\mathbf{n}}}\right\|\right)\varphi_x(b_{\mathbf{n}})^{1+\varepsilon} \\
&\leq 4M^2 C_4 \sum_{\mathbf{i}\in V_{\mathbf{i}_0}} \sum_{\mathbf{i}-\mathbf{u}\in V_{\mathbf{i}_0}} \mathbf{1}_{\{\mathbf{u};\|\mathbf{u}\|\leq d_{\mathbf{n}}\}}\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{i}_0-\mathbf{i}}{d_{\mathbf{n}}}\right\|\right)\varphi_x(b_{\mathbf{n}})^{1+\varepsilon} \\
&\leq 4M^2 C_4 k_{\mathbf{n}} d_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}}))^{1+\varepsilon}.
\end{aligned}$$

Since K_1 and K_2 are bounded, applying Lemma 2, we get

$$\sum_{\mathbf{i},\mathbf{k}\in S^c} \mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{k}}\right] \leq C \sum_{\mathbf{i},\mathbf{k}\in S^c} \{\psi(1,1)\chi(\|\mathbf{i}-\mathbf{k}\|)\} \leq C \sum_{\mathbf{i},\mathbf{k}\in S^c\cap V_{\mathbf{i}_0}} \chi(\|\mathbf{i}-\mathbf{k}\|) \leq C2^N \sum_{\mathbf{k}\in V_{\mathbf{i}_0}} \sum_{\substack{\mathbf{k}-\mathbf{u}\in V_{\mathbf{i}_0}\\ \|\mathbf{u}\|>d_{\mathbf{n}}}} \chi(\|\mathbf{i}\|)$$

$$\leq Ck_{\mathbf{n}} \sum_{\|\mathbf{i}\|>d_{\mathbf{n}}} \chi(\|\mathbf{i}\|).$$

Since

$$\sum_{\|\mathbf{i}\|>d_{\mathbf{n}}} \chi(\|\mathbf{i}\|) \leq C \sum_{\|\mathbf{i}\|>d_{\mathbf{n}}} \|\mathbf{i}\|^{-\theta} \leq C \sum_{\|\mathbf{i}\|>d_{\mathbf{n}}} \|\mathbf{i}\|^{-\theta}\|\mathbf{i}\|^{-N}\|\mathbf{i}\|^N,$$

and $\|\mathbf{i}\| > d_{\mathbf{n}}, \|\mathbf{i}\|^{-N} \leq (d_{\mathbf{n}})^{-N}$, we have

$$C \sum_{\|\mathbf{i}\|>d_{\mathbf{n}}} \|\mathbf{i}\|^{-\theta}\|\mathbf{i}\|^{-N-\varepsilon}\|\mathbf{i}\|^{N+\varepsilon} \quad \leq \quad Cd_{\mathbf{n}}^{-N-\varepsilon} \sum_{\|\mathbf{i}\|>d_{\mathbf{n}}} \|\mathbf{i}\|^{N+\varepsilon-\theta}.$$

Then,

$$\sum_{\mathbf{i},\mathbf{k}\in S^c} \mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{k}}\right] \leq Ck_{\mathbf{n}} d_{\mathbf{n}}^{-N-\varepsilon} \sum_{\|\mathbf{i}\|>d_{\mathbf{n}}} \|\mathbf{i}\|^{N+\varepsilon-\theta}.$$

Choosing $d_{\mathbf{n}} = (\varphi_x(b_{\mathbf{n}}))^{\frac{-\varepsilon}{N}+a}$ with $a > 0$ such that $Na \leq \varepsilon - \frac{N}{N+\varepsilon}$ lead to

$$d_{\mathbf{n}}^{-(N+\varepsilon)} \quad = \quad \varphi_x(b_{\mathbf{n}})(\varphi_x(b_{\mathbf{n}}))^{\frac{-(N+\varepsilon)(Na-\varepsilon)-N}{N}} = O\left(\varphi_x(b_{\mathbf{n}})\right),$$

Since $\frac{-(N+\varepsilon)(Na-\varepsilon)-N}{N} > 0$, Moreover, this choice of $d_{\mathbf{n}}$ implies that

$$\begin{aligned}\sum_{\mathbf{i},\mathbf{k}\in S} \mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{k}}\right] \quad &\leq \quad 4M^2 C_4 k_{\mathbf{n}} d_{\mathbf{n}}^N (\varphi_x(b_{\mathbf{n}}))^{1+\varepsilon} \\ &\leq 4M^2 C_4 k_{\mathbf{n}} (\varphi_x(b_{\mathbf{n}}))^{1+Na} = O(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})).\end{aligned}$$

Then, we deduce that

$$\mathbb{E}\left(\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \xi_{\mathbf{i}}\right)^2 \quad = \quad \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\xi_{\mathbf{i}}^2\right] + \sum_{\mathbf{i},\mathbf{k}\in S} \mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{k}}\right] + \sum_{\mathbf{i},\mathbf{k}\in S^c} \mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{k}}\right] = O\left(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})\right).$$

Consequently,

$$\left[\mathbb{E}\left(\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \xi_{\mathbf{i}}\right)^2\right]^{1/2} \quad = O(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}}))^{1/2}$$

and $\|e_{\mathbf{n}}(x)\|_2 = O\left(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_x(b_{\mathbf{n}})\right)^{-1/2}$ since by Assumptions **H1** and **H3**, $a_{\mathbf{n}} \geq C_{11}^2 k_{\mathbf{n}} \varphi_{\mathbf{i},x}(b_{\mathbf{n}})$.
Next, for (29), define

$$S_{\mathbf{n}}(x) = \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \Lambda_{\mathbf{i}}(x) = \frac{1}{a_{\mathbf{n}}}\left[f_{\mathbf{n}}(x) - \mathbb{E}(f_{\mathbf{n}}(x))\right].$$

Then, we have

$$\begin{aligned}\mathbb{P}\left[\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x) \leq \frac{a_{\mathbf{n}}}{2}\right] \quad &= \quad \mathbb{P}\left[\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} (K_{\mathbf{i}}(x) - \mathbb{E}(K_{\mathbf{i}}(x))) \leq \frac{-a_{\mathbf{n}}}{2}\right] \\ &\leq \quad \mathbb{P}\left[\frac{1}{a_{\mathbf{n}}}\left|\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} (K_{\mathbf{i}}(x) - \mathbb{E}(K_{\mathbf{i}}(x)))\right| \geq \frac{1}{2}\right] \\ &\leq \quad \mathbb{P}\left[|S_{\mathbf{n}}(x)| \geq \varepsilon\right], \text{for } \mathbf{n} \text{ large enough.}\end{aligned}$$

We will now introduce the spatial blocks decomposition introduced by Tran (1990) which will be useful afterwards. Without loss of generality, we suppose that $n_k = 2bq_k$, for $1 \leq k \leq N$. The random variables $\Lambda_{\mathbf{i}}(x)$ can be grouped into

$2^N q_1 \dots q_N$ cubic blocks of side b. Let,

$$
\begin{aligned}
U(1,\mathbf{n},x,\mathbf{j}) &= \sum_{\substack{i_k=2j_kb+1,\\ k=1,\dots,N.}}^{(2j_k+1)b} \Lambda_{\mathbf{i}}(x),\\
U(2,\mathbf{n},x,\mathbf{j}) &= \sum_{\substack{i_k=2j_kb+1,\\ k=1,\dots,N-1.}}^{(2j_k+1)b} \sum_{i_N=(2j_N+1)b+1}^{2(j_N+1)b} \Lambda_{\mathbf{i}}(x),\\
U(3,\mathbf{n},x,\mathbf{j}) &= \sum_{\substack{i_k=2j_kb+1,\\ k=1,\dots,N-2.}}^{(2j_k+1)b} \sum_{i_{N-1}=(2j_{N-1}+1)b+1}^{2(j_{N-1}+1)b} \sum_{i_N=2j_Nb+1}^{(2j_N+1)b} \Lambda_{\mathbf{i}}(x),\\
U(4,\mathbf{n},x,\mathbf{j}) &= \sum_{\substack{i_k=2j_kb+1,\\ k=1,\dots,N-2.}}^{(2j_k+1)b} \sum_{i_{N-1}=(2j_{N-1}+1)b+1}^{2(j_{N-1}+1)b} \sum_{i_N=(2j_N+1)b+1}^{(2j_N+1)b} \Lambda_{\mathbf{i}}(x)
\end{aligned}
$$

and so on. Noting that

$$
\begin{aligned}
U(2^{N-1},\mathbf{n},x,\mathbf{j}) &= \sum_{\substack{i_k=(2j_k+1)b+1,\\ k=1,\dots,N-1.}}^{2(j_k+1)b} \sum_{i_N=2j_Nb+1}^{(2j_N+1)b} \Lambda_{\mathbf{i}}(x)\\
U(2^N,\mathbf{n},x,\mathbf{j}) &= \sum_{\substack{i_k=(2j_k+1)b+1,\\ k=1,\dots,N.}}^{2(j_k+1)b} \Lambda_{\mathbf{i}}(x)
\end{aligned}
$$

for each integer $1 \le l \le 2^N$, we define $T(\mathbf{n},x,l) = \sum_{\substack{j_k=0\\ k=1,\dots,N.}}^{q_k-1} U(l,\mathbf{n},x,\mathbf{j})$. We obtain $S_{\mathbf{n}}(x) = \sum_{l=1}^{2^N} T(\mathbf{n},x,l)$. For $\varepsilon > 0$, $P \le \mathbb{P}\left(\left|\sum_{l=1}^{2^N} T(\mathbf{n},x,l)\right| > \varepsilon\right) \le 2^N \mathbb{P}\left(|T(\mathbf{n},x,1)| > \frac{\varepsilon}{2^N}\right)$. We enumerate in arbitrary manner the $\widehat{q} = q_1 \times \cdots \times q_N$ terms $U(1,\mathbf{n},x,\mathbf{j})$ of the sum $T(\mathbf{n},x,1)$, and refer to them as $W_1,\dots,W_{\widehat{q}}$. Note that $U(1,\mathbf{n},x,\mathbf{j})$ is a measurable σ-algebra generated by $X_{\mathbf{i}}$, with $\mathbf{i}$ such that $2j_kb+1 \le i_k \le (2j_k+1)b$, $k = 1,\dots,N$. For all $l = 1,\dots,\widehat{q}$, the sets of the sites in W_l are separated by a distance of at least equal to b. In addition, since K_1 and K_2 write $|W_l| \le C\frac{b^N}{a_{\mathbf{n}}}$ with $C = \|K_1\|_\infty \|K_2\|_\infty$ (where $\|\cdot\|_\infty$ is the sup norm). Lemma 1 insures the existence of some random variables $W_1^*, W_2^*, \dots, W_{\widehat{q}}^*$ such that

$$
\begin{aligned}
\sum_{l=1}^{\widehat{q}} \mathbb{E}|W_l - W_l^*| &\le 2\widehat{q}C\frac{b^N}{a_{\mathbf{n}}}\psi((\widehat{q}-1)b^N, b^N)\chi(b)\\
&\le 2C\frac{\widehat{\mathbf{n}}}{2^N b^N}\frac{b^N}{a_{\mathbf{n}}}\psi(\widehat{\mathbf{n}}, b^N)\chi(b).
\end{aligned}
$$

Markov inequality allows us to write

$$
\mathbb{P}\left(\sum_{l=1}^{\widehat{q}} |W_l - W_l^*| > \frac{\varepsilon}{2^{N+1}}\right) \le 2C\frac{\widehat{\mathbf{n}}}{2^N b^N}\frac{b^N}{a_{\mathbf{n}}}\psi(\widehat{\mathbf{n}}, b^N)\chi(b)2^{N+1}\varepsilon^{-1},
$$

and by Bernstein inequality, we have

$$\mathbb{P}\left(\sum_{l=1}^{\widehat{q}}|W_l^*| > \frac{\varepsilon}{2^{N+1}}\right) \quad \leq \quad 2\exp\left\{\frac{-\varepsilon^2/(2^{N+1})^2}{4\sum_{l=1}^{\widehat{q}}\mathbb{E}(W_l^{*2}) + \frac{2\varepsilon}{2^{N+1}}\frac{b^N}{a_{\mathbf{n}}}C}\right\}$$

which leads to

$$\begin{aligned}\mathbb{P}\left[|S_{\mathbf{n}}(x)| \geq \varepsilon\right] \quad &\leq \quad 2^{N+1}\exp\left\{\frac{-\varepsilon^2/(2^{N+1})^2}{4\sum_{l=1}^{\widehat{q}}\mathbb{E}(W_l^{*2}) + 2^{-N}C\varepsilon\frac{b^N}{a_{\mathbf{n}}}}\right\} \\ &\quad +2^{N+1}C\frac{\widehat{\mathbf{n}}}{2^N b^N}\frac{b^N}{a_{\mathbf{n}}}\psi(\widehat{\mathbf{n}}, b^N)\chi(b)2^{N+1}\varepsilon^{-1}.\end{aligned}$$

Let $\delta > 0$, $\varepsilon = \varepsilon_{\mathbf{n}} = \delta\left(\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}\right)^{1/2}$ and $b = \left(\frac{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\frac{1}{2N}}$.

Since the variables W_l and W_l^* have the same distributions, we have $\sum_{l=1}^{\widehat{q}}\mathbb{E}W_l^{*2} = \sum_{l=1}^{\widehat{q}}\operatorname{var}(W_l^*) = \sum_{l=1}^{\widehat{q}}\operatorname{var}(W_l) \leq I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x)$, and according to Lemma 6, we have $\sum_{l=1}^{\widehat{q}}\mathbb{E}W_l^{*2} \leq O\left([\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})]^{-1}\right)$. Then,

$$\begin{aligned}\mathbb{P}\left[|S_{\mathbf{n}}(x)| \geq \varepsilon\right] \quad &\leq \quad 2^{N+1}\exp\left\{\frac{-\varepsilon^2}{2^{2N+2}\left(4\frac{C}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})} + C2^{-N}\varepsilon\frac{b^N}{a_{\mathbf{n}}}\right)}\right\} \\ &\quad +2^{N+2}C\frac{\widehat{\mathbf{n}}}{a_{\mathbf{n}}}\psi(\widehat{\mathbf{n}}, b^N)b^{-\theta}\varepsilon^{-1}.\end{aligned}$$

Since $C_1''k_{\mathbf{n}}\varphi_x(b_{\mathbf{n}}) \leq a_{\mathbf{n}} \leq C_2''k_{\mathbf{n}}\varphi_x(b_{\mathbf{n}})$, where C_1''andC_2'' are positive constant and $k_{\mathbf{n}} = O(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N)$, we have

$$\begin{aligned}\mathbb{P}\left[|S_{\mathbf{n}}(x)| \geq \varepsilon_{\mathbf{n}}\right] \quad &\leq \quad 2^{N+1}\exp\left\{\frac{-\delta^2\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}}{\frac{2^{2N+4}C}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})} + \frac{C2^{N+2}\delta}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}}\right\} \\ &\quad +2^{N+2}C\frac{\widehat{\mathbf{n}}}{a_{\mathbf{n}}}\psi(\widehat{\mathbf{n}}, b^N)b^{-\theta}\delta^{-1}\left(\frac{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{1/2} \\ &\leq \quad C2^{N+1}\exp\left\{\log\widehat{\mathbf{n}}^{-a}\right\} \\ &\quad +2^{N+2}C\delta^{-1}\frac{\widehat{\mathbf{n}}}{a_{\mathbf{n}}}\psi(\widehat{\mathbf{n}}, b^N)\left(\frac{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\frac{N-\theta}{2N}} \\ &\leq \quad C\widehat{\mathbf{n}}^{-a} + 2^{N+2}C\delta^{-1}\frac{\widehat{\mathbf{n}}}{a_{\mathbf{n}}}\psi(\widehat{\mathbf{n}}, b^N)\left(\frac{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\frac{N-\theta}{2N}} \\ &:= \quad C\widehat{\mathbf{n}}^{-a} + C2^{N+2}\delta^{-1}D_{\mathbf{n}},\end{aligned}$$

with $a = \dfrac{\delta^2}{2^{2N+4}C + C2^{N+2}\delta} > 0$. Note that $\widehat{\mathbf{n}}^{1-a}\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})$ tends to 0 for $a > 1$ and then $C\widehat{\mathbf{n}}^{-a} = o\left([\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})]^{-1}\right)$. Moreover $a > 1$ if and only if $\delta > 2^{N+1}C(1 + \sqrt{4C}) > 2^{N+1}C$ (with $\delta > 0$). Now, we treat the second term.

When (8) is satisfied, *i.e.* $\psi(n,m) \leq C\min(n,m)$, $\forall n,m \in \mathbb{N}$, we have

$$\begin{aligned}
\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})2^{N+2}C\delta^{-1}D_{\mathbf{n}} &\leq \widehat{\mathbf{n}}\rho_{\mathbf{n}}^N 2^{N+2}C\delta^{-1}\frac{\widehat{\mathbf{n}}}{a_{\mathbf{n}}}\left(\frac{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\frac{2N-\theta}{2N}} \\
&\leq \widehat{\mathbf{n}}\rho_{\mathbf{n}}^N 2^{N+2}C\delta^{-1}\frac{1}{\rho_{\mathbf{n}}^N}\left(\frac{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\frac{2N-\theta}{2N}} \\
&\leq C\left[\widehat{\mathbf{n}}\left(\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})\right)^{\frac{2N-\theta}{4N-\theta}}(\log\widehat{\mathbf{n}})^{\frac{\theta-2N}{4N-\theta}}\right]^{\frac{4N-\theta}{2N}}
\end{aligned}$$

which tends to 0 as $\mathbf{n} \to 0$ since $\theta > 4N$.

When (9) is satisfied, *i.e.* $\psi(n,m) \leq C(n+m+1)^{\kappa}$, $\forall n,m \in \mathbb{N}$, and note that $\psi(\widehat{\mathbf{n}}, b^N) \leq C(\widehat{\mathbf{n}} + b^N + 1)^{\kappa} \leq C\widehat{\mathbf{n}}^{\kappa}$, we have

$$\begin{aligned}
\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})C2^{N+2}\delta^{-1}\frac{\widehat{\mathbf{n}}}{a_{\mathbf{n}}}\widehat{\mathbf{n}}^{\kappa}\left(\frac{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\frac{2N-\theta}{2N}} &\leq \widehat{\mathbf{n}}\rho_{\mathbf{n}}^N 2^{N+2}C\delta^{-1}\frac{1}{\rho_{\mathbf{n}}^N}\widehat{\mathbf{n}}^{\kappa}\left(\frac{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\frac{2N-\theta}{2N}} \\
&\leq C\left[\widehat{\mathbf{n}}\left(\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})\right)^{\frac{N-\theta}{N(3+2\kappa)-\theta}}(\log\widehat{\mathbf{n}})^{\frac{\theta-N}{N(3+2\kappa)-\theta}}\right]^{\frac{N(3+2\kappa)-\theta}{2N}}
\end{aligned}$$

which tends to 0 as $\mathbf{n} \to$ since $\theta > N(3+2\kappa)$. Therefore, (29) follows, which conclude the proof of Lemma 4. □

Appendix A.2.4. Proof of Lemma 5

Since $Y_{\mathbf{i}}$ and r are bounded, we have

$$\begin{aligned}
\mathbb{E}^{1/2}[\mathbf{C}] &\leq \mathbb{E}^{1/2}\left[\left|\frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} Y_{\mathbf{i}} - r(x)\right|\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)=0\right\}}\right] \\
&\leq 2M\mathbb{E}^{1/2}\left[\mathbf{1}_{\left\{\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} W_{\mathbf{ni}}(x)=0\right\}}\right] = 2M\left(\mathbb{P}\left[\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x) = 0\right]\right)^{1/2} \\
&\leq 2M\left(\mathbb{P}\left[\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{i}}(x) \leq \frac{a_{\mathbf{n}}}{2}\right]\right)^{1/2} = O\left(\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})}\right)^{1/2},
\end{aligned}$$

by Lemma 4. □

Appendix A.2.5. Proof of Lemma 6

Firstly, we deal with $I_{\mathbf{n}}(x) = \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\left(\frac{1}{a_{\mathbf{n}}}K_{\mathbf{i}}(x)\right)^2\right] - \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}}\left(\frac{1}{a_{\mathbf{n}}}\mathbb{E}(K_{\mathbf{i}}(x))\right)^2$.

$$\begin{aligned}\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\left(\frac{1}{a_{\mathbf{n}}}K_{\mathbf{i}}(x)\right)^2\right] &\leq C\frac{1}{a_{\mathbf{n}}^2}\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} K_{\mathbf{2i}}^2\mathbb{E}\left[K_{\mathbf{1i}}^2(x)\right]\\ &\leq C\frac{1}{a_{\mathbf{n}}^2}\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} k_{\mathbf{n}}\varphi_x(b_{\mathbf{n}})\\ &\leq \frac{C}{k_{\mathbf{n}}\varphi_x(b_{\mathbf{n}})} = O\left([\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})]^{-1}\right),\end{aligned}$$

for $\mathbf{n}$ sufficiently large.
Then, we have $I_{\mathbf{n}}(x) = O\left([\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})]^{-1}\right)$. We now treat the term $R_{\mathbf{n}}(x)$. Since the functions $K_1(.)$ and $K_2(.)$ are bounded, applying Lemma 1, we get

$$|\mathbb{E}\left[\Lambda_{\mathbf{i}}(x)\Lambda_{\mathbf{k}}(x)\right]| \leq C\frac{K_{2\mathbf{i}}K_{2\mathbf{k}}}{a_{\mathbf{n}}^2}\psi(1,1)\gamma(\|\mathbf{i}-\mathbf{k}\|).$$

Let $E_{\mathbf{n}}$ be a sequence of real numbers tending to ∞ as $\widehat{\mathbf{n}} \to \infty$. Set $T = \{\mathbf{i},\mathbf{k} \in \mathcal{V}_{\mathbf{i}_0}, \|\mathbf{i}-\mathbf{k}\| \leq E_{\mathbf{n}}\}$ and denote by T^c the complementary of T. Let $R_{\mathbf{n}}^{(1)} = \sum_{\mathbf{i},\mathbf{k}\in T}|\mathbb{E}\left[\Lambda_{\mathbf{i}}(x)\Lambda_{\mathbf{k}}(x)\right]|$ and $R_{\mathbf{n}}^{(2)} = \sum_{\mathbf{i},\mathbf{k}\in T^c}|\mathbb{E}\left[\Lambda_{\mathbf{i}}(x)\Lambda_{\mathbf{k}}(x)\right]|$. Hence, $R_{\mathbf{n}}(x) \leq R_{\mathbf{n}}^{(1)} + R_{\mathbf{n}}^{(2)}$. Moreover, using the same arguments as in the proof of Lemma 4, we have $I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x) = O\left([\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\varphi_x(b_{\mathbf{n}})]^{-1}\right)$. □

Appendix A.2.6. Proof of Theorem 2

Recall that $K_{\mathbf{i}}(x) = K_{1\mathbf{i}}K_{2\mathbf{i}}$. Set $T_{\mathbf{n}} = (\hat{\mathbf{n}}u_{\mathbf{n}})^{1/s}$ where

$$u_{\mathbf{n}} = \prod_{i=1}^{N}(\log n_i)(\log\log n_i)^{1+\varepsilon},$$

and define

$$\begin{aligned} g_{\mathbf{n}}(x) &= \frac{1}{a_{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}}K_{\mathbf{i}}(x), \qquad f_{\mathbf{n}}(x) = \frac{1}{a_{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_{\mathbf{i}}(x),\\ \widetilde{g}_{\mathbf{n}}(x) &= \frac{1}{a_{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} Y_{\mathbf{i}}\mathbb{1}_{\{Y_{\mathbf{i}}\leq T_{\mathbf{n}}\}}K_{\mathbf{i}}(x).\end{aligned}$$

Then, we can write

$$r_{\mathbf{n}}(x) - r(x) = -\frac{r(x)}{f_{\mathbf{n}}(x)}A_1(x) + \frac{1}{f_{\mathbf{n}}(x)}\left[A_2(x) + A_3(x) + A_4(x)\right], \qquad (30)$$

where

$$\begin{aligned} A_1(x) &= f_{\mathbf{n}}(x) - 1, \\ A_2(x) &= \mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right) - r(x), \\ A_3(x) &= \widetilde{g}_{\mathbf{n}}(x) - \mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right), \\ A_4(x) &= g_{\mathbf{n}}(x) - \widetilde{g}_{\mathbf{n}}(x). \end{aligned}$$

Therefore Theorem 2 follows from (30) and Lemmas 7, 8, 9, 12. □

Lemma 7. *Under assumptions **H1-H4** and* **H6***,*

$$\sup_{x\in\mathcal{D}} \left|\mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right) - r(x)\right| = O\left(b_{\mathbf{n}} + \sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \Gamma(b_{\mathbf{n}})}}\right).$$

Proof of Lemma 7

Since

$$\begin{aligned} &\mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right) - r(x) \\ &= \frac{1}{a_{\mathbf{n}}\varphi_x(b_{\mathbf{n}})} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\left(Y_{\mathbf{i}} - Y_{\mathbf{i}}\mathbb{1}_{\{|Y_{\mathbf{i}}|>T_{\mathbf{n}}\}}\right) K_{\mathbf{i}}(x)\right] - r(x) \\ &= \frac{1}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\mathbb{E}\left(Y_{\mathbf{i}}|X_{\mathbf{i}}\right) K_{\mathbf{i}}(x)\right] - \frac{1}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[Y_{\mathbf{i}}\mathbb{1}_{\{|Y_{\mathbf{i}}|>T_{\mathbf{n}}\}} K_{\mathbf{i}}(x)\right] - r(x) \\ &= \frac{1}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\left(r(X_{\mathbf{i}}) - r(x)\right) K_{\mathbf{i}}(x)\right] - \frac{1}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[Y_{\mathbf{i}}\mathbb{1}_{\{|Y_{\mathbf{i}}|>T_{\mathbf{n}}\}} K_{\mathbf{i}}(x)\right], \end{aligned}$$

we have

$$\begin{aligned} \left|\mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right) - r(x)\right| &\leq \frac{1}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\left|r(X_{\mathbf{i}}) - r(x)\right| K_{\mathbf{i}}(x)\right] \\ &+ \frac{1}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\left|Y_{\mathbf{i}}\right| \mathbb{1}_{\{|Y_{\mathbf{i}}|>T_{\mathbf{n}}\}} K_{\mathbf{i}}(x)\right] := I + II. \end{aligned}$$

Using assumptions **H1** and **H2**, we have

$$|r(X_{\mathbf{i}}) - r(x)| \leq \sup_{u\in B(x,b_{\mathbf{n}})} |r(x) - r(u)| = O(b_{\mathbf{n}}), \text{ so that } I = O\left(b_{\mathbf{n}}\right).$$

For II, since $s > 2$, using Assumption **H4** and **H6**, we can write

$$\begin{aligned} II &\leq \frac{T_{\mathbf{n}}^{1-s}}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\left|Y_{\mathbf{i}}\right|^s K_{\mathbf{i}}(x)\right] \leq \frac{T_{\mathbf{n}}^{1-s}}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left[\mathbb{E}\left(\left|Y_{\mathbf{i}}\right|^s |X_{\mathbf{i}}\right) K_{\mathbf{i}}(x)\right] \\ &\leq CT_{\mathbf{n}}^{1-s} = o\left(\left(\widehat{\mathbf{n}}u_{\mathbf{n}}\right)^{-1/2}\right) = o\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \Gamma(b_{\mathbf{n}})}}\right), \end{aligned}$$

which conclude the proof of Lemma 7. □

Lemma 8. *If Assumption* ***(H6)*** *(i) holds, then*

$$\sup_{x\in\mathcal{D}} |g_{\mathbf{n}}(x) - \widetilde{g}_{\mathbf{n}}(x)| = 0$$

for sufficiently large $\mathbf{n}$.

Proof of Lemma 8

Recall that $T_{\mathbf{n}} = (\hat{\mathbf{n}} u_{\mathbf{n}})^{1/s}$ and note that

$$g_{\mathbf{n}}(x) - \widetilde{g}_{\mathbf{n}}(x) = \frac{1}{a_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} Y_{\mathbf{i}} \mathbb{1}_{\{|Y_{\mathbf{i}}|>T_{\mathbf{n}}\}} K_{\mathbf{i}}(x).$$

By the Markov inequality, $\mathbb{P}\left(|Y_{\mathbf{i}}| > T_{\mathbf{n}}\right) \leq T_{\mathbf{n}}^{-s}\mathbb{E}|Y_{\mathbf{i}}|^s$ for any $\mathbf{i} \in \mathbb{Z}^N$. Therefore

$$\sum_{\mathbf{n}\in\mathbb{Z}^N} \mathbb{P}\left(|Y_{\mathbf{n}}| > T_{\mathbf{n}}\right) \leq C \sum_{\mathbf{n}\in\mathbb{Z}^N} \frac{1}{\hat{\mathbf{n}} u_{\mathbf{n}}} < \infty.$$

The Borel-Cantelli lemma ensures that almost surely $|Y_{\mathbf{i}}| \leq T_{\mathbf{n}}$ for sufficiently large $\mathbf{n}$. Since $T_{\mathbf{n}} \to \infty$ as $\mathbf{n} \to \infty$, we have almost surely $|Y_{\mathbf{i}}| < T_{\mathbf{n}}$ for all $\mathbf{i} \in \mathcal{V}_{\mathbf{i}_0}$ and for $\mathbf{n}$ sufficiently large enough, and thus the conclusion follows. □

Lemma 9. *Under the assumptions of Theorem 2,*

$$\sup_{x\in\mathcal{D}} |\widetilde{g}_{\mathbf{n}}(x) - \mathbb{E}\left(\widetilde{g}_{\mathbf{n}}(x)\right)| = O\left(\left(\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\Gamma(b_{\mathbf{n}})}\right)^{1/2}\right) \; a.s$$

Define

$$\widetilde{\Lambda}_{\mathbf{i}}(x) \quad = \quad Y_{\mathbf{i}} \mathbb{1}_{\{|Y_{\mathbf{i}}|\leq T_{\mathbf{n}}\}} K_{\mathbf{i}}(x) - \mathbb{E}\left(Y_{\mathbf{i}} \mathbb{1}_{\{|Y_{\mathbf{i}}|\leq T_{\mathbf{n}}\}} K_{\mathbf{i}}(x)\right),$$

$$\widetilde{I}_{\mathbf{n}}(x) = \frac{1}{a_{\mathbf{n}}^2} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \mathbb{E}\left(\widetilde{\Lambda}_{\mathbf{i}}(x)^2\right) \text{ and } \widetilde{R}_{\mathbf{n}}(x) = \frac{1}{a_{\mathbf{n}}^2} \sum_{\mathbf{i}\neq\mathbf{j}} \left|\mathbb{E}\left[\widetilde{\Lambda}_{\mathbf{i}}(x)\widetilde{\Lambda}_{\mathbf{j}}(x)\right]\right|. \tag{31}$$

Then, arguing as in the proof of Lemma 6 with $\varphi_{\mathbf{i},x}(b_{\mathbf{n}})$ replacing by $\Gamma(b_{\mathbf{n}})$, one can prove under assumptions **H1-H2**, **H4–H6** that,

$$\widetilde{I}_{\mathbf{n}}(x) + \widetilde{R}_{\mathbf{n}}(x) = O\left(\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\Gamma(b_{\mathbf{n}})}\right) \quad \text{for any } x \in \mathcal{D}. \tag{32}$$

Let us define

$$\Omega_{\mathbf{n}} = \sqrt{\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\Gamma(b_{\mathbf{n}})}} \text{ and choose } \ell_{\mathbf{n}} \leq C\Omega_{\mathbf{n}}\varphi_x(b_{\mathbf{n}})\rho_{\mathbf{n}}^N\Gamma(b_{\mathbf{n}})T_{\mathbf{n}}^{-1} \text{ for some constant } C > 0.$$

We suppose that the compact set $\mathcal{D}$ is covered with $v_{\mathbf{n}}$ cubes B_k having sides of length $\ell_{\mathbf{n}}$ and centered at x_k. We have

$$\sup_{x\in\mathcal{D}} |\widetilde{g}_{\mathbf{n}}(x) - \mathbb{E}(\widetilde{g}_{\mathbf{n}}(x))| \leq Q_{1\mathbf{n}} + Q_{2\mathbf{n}} + Q_{3\mathbf{n}}, \tag{33}$$

where

$$\begin{aligned} Q_{1\mathbf{n}} &= \max_{1\leq k\leq v_{\mathbf{n}}} \sup_{x\in B_k} |\widetilde{g}_{\mathbf{n}}(x) - \widetilde{g}_{\mathbf{n}}(x_k)|, \\ Q_{2\mathbf{n}} &= \max_{1\leq k\leq v_{\mathbf{n}}} \sup_{x\in B_k} |\mathbb{E}(\widetilde{g}_{\mathbf{n}}(x_k)) - \mathbb{E}(\widetilde{g}_{\mathbf{n}}(x))|, \\ Q_{3\mathbf{n}} &= \max_{1\leq k\leq v_{\mathbf{n}}} \sup_{x\in B_k} |\widetilde{g}_{\mathbf{n}}(x_k) - \mathbb{E}(\widetilde{g}_{\mathbf{n}}(x_k))|. \end{aligned}$$

Lemma 10. *Under Assumptions* **H1**, **H2** *and* **H4**, $Q_{1\mathbf{n}} = O(\Omega_{\mathbf{n}})$ *and* $Q_{2\mathbf{n}} = O(\Omega_{\mathbf{n}})$ *a.s.*

Proof of Lemma 10

By Assumptions **H1**, **H2** and **H4**, for all $x \in B_k$,

$$\begin{aligned} |\widetilde{g}_{\mathbf{n}}(x) - \widetilde{g}_{\mathbf{n}}(x_k)| &\leq a_{\mathbf{n}}^{-1}\varphi_x(b_{\mathbf{n}})^{-1}\rho_{\mathbf{n}}^{-N}\Gamma(b_{\mathbf{n}})^{-1}T_{\mathbf{n}}\|x - x_k\| \\ &\leq C\varphi_{\mathbf{i},x}(b_{\mathbf{n}})^{-1}\rho_{\mathbf{n}}^{-N}\Gamma(b_{\mathbf{n}})^{-1}T_{\mathbf{n}}\ell_{\mathbf{n}} = O(\Omega_{\mathbf{n}})\, a.s. \end{aligned}$$

and Lemma 10 follows. □

Next, we have to show that

$$Q_{3\mathbf{n}} = O(\Omega_{\mathbf{n}}) \text{ a.s.} \tag{34}$$

Define

$$\widetilde{S}_{\mathbf{n}}(x) = a_{\mathbf{n}}^{-2}\varphi_x(b_{\mathbf{n}})^{-2} \sum_{\mathbf{i}\in\mathcal{V}_{\mathbf{i}_0}} \widetilde{\Lambda}_{\mathbf{i}}(x) = \widetilde{g}_{\mathbf{n}}(x) - \mathbb{E}(\widetilde{g}_{\mathbf{n}}(x)).$$

Define also $\widetilde{U}(i,\mathbf{n},x,\mathbf{j})$ and $\widetilde{T}(\mathbf{n},x,i)$ to be the same as $U(i,\mathbf{n},\mathbf{j},x)$ and $T(\mathbf{n},i,x)$ in the proof of Lemma 4 except with $\Lambda_{\mathbf{j}}$ replacing by $\widetilde{\Lambda}_{\mathbf{j}}$. Arguing that $\widetilde{S}_{\mathbf{n}}$ is a finite sum of the $\widetilde{T}(\mathbf{n},x,i)$, then showing (34) is equivalent to show that

$$\max_{1\leq k\leq v_{\mathbf{n}}} \left|\widetilde{T}(\mathbf{n},x_k,1)\right| = O(\Omega_{\mathbf{n}}) \text{ a.s.} \tag{35}$$

By same arguments as in Lemma 4, $\widetilde{T}(\mathbf{n},1,x)$ is the sum of $\widehat{q} = q_1 \times \cdots \times q_N$ of the $\widetilde{U}(i,\mathbf{n},\mathbf{j},x)$'s which are measurable with σ-field generated by $X_{\mathbf{i}}$, where $\mathbf{i}$ belong to the set of sites which are separated by a distance at least p. Enumerate these random variables as $Z_1,\dots,Z_{\widehat{q}}$ and approximate them by the independent random variables $Z_1^*,\dots,Z_{\widehat{q}}^*$ as was done in Lemma 1. Define

$$p \sim \Omega_{\mathbf{n}}^{-1/N} T_{\mathbf{n}}^{-1/N},$$

and

$$\widetilde{\beta}_{\mathbf{n}} = T_{\mathbf{n}}\rho_{\mathbf{n}}^{-N}\Gamma(b_{\mathbf{n}})^{-1}\psi(\widehat{\mathbf{n}},p^N)p^{-\theta}\Omega_{\mathbf{n}}^{-1}.$$

Lemma 11. *Under assumptions of Theorem 2, there exist two positive constants A and C such that, for any $\lambda > 0$,*

$$\mathbb{P}\left(\max_{1\le k\le v_{\mathbf{n}}}\left|\widetilde{T}(\mathbf{n},x_k,i)\right| > \lambda\Omega_{\mathbf{n}}\right) \le C\widehat{\mathbf{n}}^{\beta}\left[\widehat{\mathbf{n}}^{-A}+\widetilde{\beta}_{\mathbf{n}}\right].$$

Proof of Lemma 11

Since $\widetilde{T}(\mathbf{n},x,i) = \sum_{i=1}^{\widehat{q}} Z_i$, we have, for any $\lambda > 0$,

$$\mathbb{P}\left(\left|\widetilde{T}(\mathbf{n},x,i)\right| > \lambda\Omega_{\mathbf{n}}\right) \le \mathbb{P}\left(\sum_{i=1}^{\widehat{q}}|Z_i - Z_i^*| > \lambda\Omega_{\mathbf{n}}/2\right) + \mathbb{P}\left(\left|\sum_{i=1}^{\widehat{q}} Z_i^*\right| > \lambda\Omega_{\mathbf{n}}/2\right).$$

By the boundedness of the functions K_1 and K_2 respectively, we have

$$|Z_{\mathbf{i}}| \le Cp^N T_{\mathbf{n}} a_{\mathbf{n}}^{-1}\varphi_x(b_{\mathbf{n}})^{-1} \le CT_{\mathbf{n}}p^N\left(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\Gamma(b_{\mathbf{n}})\right)^{-1}.$$

Note that $\widehat{\mathbf{n}} = 2^N p^N \widehat{q}$. Therefore Markov inequality gives: for any $\lambda > 0$,

$$\mathbb{P}\left(\sum_{i=1}^{\widehat{q}}|Z_i - Z_i^*| > \lambda\Omega_{\mathbf{n}}\right) \le 2\widehat{q}p^N T_{\mathbf{n}}\left(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N\Gamma(b_{\mathbf{n}})\right)^{-1}\psi(\widehat{\mathbf{n}},p^N)\chi(p)\lambda^{-1}\Omega_{\mathbf{n}}^{-1} \le C\widetilde{\beta}_{\mathbf{n}}.$$

By Lemma 32, we get, for any $\lambda > 0$, there exists a constant $C > 0$ such that

$$\mathbb{P}\left(\left|\sum_{i=1}^{\widehat{q}} Z_i^*\right| > \lambda\Omega_{\mathbf{n}}\right) \le C\widehat{\mathbf{n}}^{-A},$$

and the conclusion follows. □

Proof of Lemma 9 Note that by the Fubini's theorem, it can be seen that $\sum_{\mathbf{n}\in\mathbb{Z}^N} 1/(\widehat{\mathbf{n}}u_{\mathbf{n}}) < \infty$. By (33), Lemma 10 , and Lemma 11, proving Lemma 9 is equivalent to show that

$$\widehat{\mathbf{n}}u_{\mathbf{n}}\widehat{\mathbf{n}}^{\beta-A} \to 0 \text{ and } \widehat{\mathbf{n}}u_{\mathbf{n}}\widehat{\mathbf{n}}^{\beta}\widetilde{\beta}_{\mathbf{n}} \to 0 \text{ as } \mathbf{n}\to\infty. \tag{36}$$

Note that, the first part of (36) holds by choosing A such that $A > \beta+2$. For its second part, when (8) is satisfied, $\psi(\widehat{\mathbf{n}},p^N) = p^N$ for $\mathbf{n}$ large enough. Then

$$\begin{aligned}\widehat{\mathbf{n}}^{\beta+1}u_{\mathbf{n}}\widetilde{\beta}_{\mathbf{n}} &\le C\widehat{\mathbf{n}}^{\beta}\left(\widehat{\mathbf{n}}u_{\mathbf{n}}\right)^{1/s+1}\rho_{\mathbf{n}}^{-N}\Gamma(b_{\mathbf{n}})^{-1}\Omega_{\mathbf{n}}^{(\theta-2N)/N}\left(\widehat{\mathbf{n}}u_{\mathbf{n}}\right)^{(\theta-N)/sN}\\ &= C\widehat{\mathbf{n}}^{\beta+1/s+1+(\theta-N)/(sN)+(2N-\theta)/(2N)}\rho_{\mathbf{n}}^{-\frac{\theta}{2}}\Gamma(b_{\mathbf{n}})^{\frac{-\theta}{2N}}\left(\log\widehat{\mathbf{n}}\right)^{\frac{\theta-2N}{2N}}u_{\mathbf{n}}^{\frac{sN+\theta}{sN}}\\ &= C\left[\widehat{\mathbf{n}}\rho_{\mathbf{n}}^{N\theta_1}\Gamma(b_{\mathbf{n}})^{\theta_1}\left(\log\widehat{\mathbf{n}}\right)^{\theta_2}u_{\mathbf{n}}^{\theta_3}\right]^{\frac{2sN(\beta+2)+\theta(2-s)}{2sN}},\end{aligned}$$

which goes to zero when $\theta > \left(2Ns(\beta+2)\right)/(s-2)$.
Similarly, when (9) is satisfied, we have $\psi(\widehat{\mathbf{n}},p^N) \le C\widehat{\mathbf{n}}^{\kappa}$ for $\mathbf{n}$ large enough. Then,

$$\begin{aligned}\widehat{\mathbf{n}}^{\beta+1}u_{\mathbf{n}}\widetilde{\beta}_{\mathbf{n}} &\le C\widehat{\mathbf{n}}^{\beta+\kappa}\rho_{\mathbf{n}}^{-N}\Gamma(b_{\mathbf{n}})^{-1}T_{\mathbf{n}}^{1+\theta/N}\Omega_{\mathbf{n}}^{\frac{\theta-N}{N}}\\ &= C\widehat{\mathbf{n}}^{\beta+\kappa+(N+\theta)/(sN)+(N-\theta)/(2N)}\left(\rho_{\mathbf{n}}^N\Gamma(b_{\mathbf{n}})\right)^{\frac{-N-\theta}{2N}}\left(\log\widehat{\mathbf{n}}\right)^{\frac{\theta-N}{2N}}u_{\mathbf{n}}^{\frac{N+\theta}{sN}}\\ &= C\left[\widehat{\mathbf{n}}\left(\rho_{\mathbf{n}}^N\Gamma(b_{\mathbf{n}})\right)^{\theta_1^*}\left(\log\widehat{\mathbf{n}}\right)^{\theta_2^*}u_{\mathbf{n}}^{\theta_3^*}\right]^{\frac{N(2s\beta+2s\kappa+s+2)+\theta(2-s)}{2sN}},\end{aligned}$$

which goes to zero when $\theta > (N(2s\beta + 2s\kappa + s + 2))/(s-2)$ and Lemma 9 follows. □

Lemma 12. *Under Assumptions **H1**, **H2**, **H4** and **H5**,*

1. *if* (8) *is satisfied and*

$$\widehat{\mathbf{n}}\left(\rho_{\mathbf{n}}^{N}\Gamma(b_{\mathbf{n}})\right)^{\theta_4}(\log\widehat{\mathbf{n}})^{\theta_5}u_{\mathbf{n}}^{\theta_6}\to\infty \text{ with } \theta > 2N(\beta+2),$$

2. *or if* (9) *is satisfied and*

$$\widehat{\mathbf{n}}\left(\rho_{\mathbf{n}}^{N}\Gamma(b_{\mathbf{n}})\right)^{\theta_4^*}(\log\widehat{\mathbf{n}})^{\theta_5^*}u_{\mathbf{n}}^{\theta_6^*}\to\infty \text{ with } \theta > N(2\beta+2\kappa+3),$$

then,

$$\sup_{x\in\mathcal{D}}|f_{\mathbf{n}}(x)-1| = O\left(\left(\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^{N}\Gamma(b_{\mathbf{n}})}\right)^{1/2}\right) \ a.s,$$

where

$$\theta_4 = \frac{\theta}{\theta - 2N(\beta+2)} \quad \theta_5 = \frac{\theta - 2N}{2N(\beta+2)-\theta} \quad \theta_6 = \frac{2N}{2N(\beta+2)-\theta},$$
$$\theta_4^* = \frac{-N-\theta}{N(2\beta+2\kappa+3)-\theta} \quad \theta_5^* = \frac{\theta-N}{N(2\beta+2\kappa+3)-\theta} \quad \theta_6^* = \frac{2N}{N(2\beta+2\kappa+3)-\theta}.$$

Proof of Lemma 12

To prove Lemma 12, just adapt the arguments considered in the proof of Lemma 9 to the case where $Y_{\mathbf{i}} \equiv 1$ and $T_{\mathbf{n}} = 1$.

Appendix A.2.7. Proof of Theorem 3

This result is derived directly from the proof of Theorem 1 where the regression r is replaced with p_j and for the particular response variable $1_{[Y=j]}$, we remark that $p_j(x) = \mathbb{E}(1_{[Y=j]}|X=x) = P(Y=j|X=x)$.

Appendix A.2.8. Proof of Theorem 4

This result is derived directly from the proof of Theorem 2 when the regression r is replaced with the posterior probability p_j and for the particular response variable $1_{[Y=j]}$ allows us to get the result.