

Spatial regression estimation for functional data with spatial dependency

Titre: Estimation de la régression spatiale pour données fonctionnelles avec dépendance spatiale

Camille Ternynck¹

Abstract: We propose a nonparametric estimator of the regression function of a scalar spatial variable Y_i given a functional variable X_i . The specificity of the proposed estimator is to depend on two kernels in order to control both the distance between observations and spatial locations. Mean square consistency of this estimator is obtained when the sample considered is an α -mixing sequence. Lastly, numerical results are provided to illustrate the behavior of our estimator.

Résumé : Nous proposons un estimateur non paramétrique de la fonction de régression d'une variable spatiale, Y_i , scalaire conditionnellement à une variable, X_i , fonctionnelle. La spécificité de l'estimateur proposé est de dépendre de deux noyaux permettant de contrôler à la fois la distance entre les observations et les sites. La convergence en moyenne quadratique de cet estimateur est obtenue quand l'échantillon considéré est une séquence α -mélangeante. Pour terminer, des résultats numériques illustrent le comportement de notre estimateur.

Keywords: kernel regression estimation, spatial process, functional data

Mots-clés : estimation à noyau de la régression, processus spatial, données fonctionnelles

AMS 2000 subject classifications: 62G08, 62G20, 62H11, 62M30

1. Introduction

The spatial indexing, which provides geographical reference of data, is encountered in many subject areas such as oceanography, epidemiology, forestry survey and economy. As a consequence, the scientific research community is increasingly interested in analyzing spatial data and then in developing more and more efficient spatial statistical tools. Early spatial models appeared at the beginning of the 19th century and are mainly related to parametric spatial statistics modeling (see Ripley (1981); Cressie (1993); Guyon (1995); Anselin and Florax (1995); Chilès and Delfiner (1999) for more details on statistics for spatial data). The nonparametric methods are able to reveal structure in data that might be missed by classical parametric ones. Nowadays, a dynamic concerns the deployment of nonparametric methods to spatial statistics such as density estimation, regression, prediction . . . (e.g. Journel (1983); Tran (1990); Carbon et al. (1997); Biau and Cadre (2004); Hallin et al. (2009); Menezes et al. (2010)). However, most of nonparametric spatial contributions deal with univariate or multivariate data whereas recent advances of real-time measurement instruments and data storage resources led to the emergence of functional data. The studied objects can then be curves, not numbers or vectors. This kind of data is more and more frequently involved in statistical problems since the 1990's. For an introduction to this field, the

¹ University of Lille.

E-mail: camille.ternynck@univ-lille3.fr

reader is directed to the books of [Bosq \(2000\)](#); [Ramsay and Silverman \(2005\)](#); [Ferraty and Vieu \(2006\)](#).

Currently, the literature on spatial statistics for functional data is not extensive (see [Laksaci and Maref \(2009\)](#); [Nerini et al. \(2010\)](#); [Delicado et al. \(2010\)](#); [Dabo-Niang et al. \(2010\)](#); [Laksaci and Mechab \(2010\)](#); [Dabo-Niang et al. \(2011\)](#); [Attouch et al. \(2011\)](#); [Dabo-Niang et al. \(2012\)](#); [Dabo-Niang and Yao \(2013\)](#)) and is the baseline of this current work. Indeed, we are interested in estimating the nonparametric regression for functional data presenting spatial dependence. More particularly, this regression estimator aims at taking into account the spatial dependency directly in its construction. To the best of our knowledge, very little research deals with this issue. Among the nonparametric methods, the usual kernel density estimator (see [Rosenblatt \(1956\)](#)) is often used in order to estimate the regression operator. In [Menezes et al. \(2010\)](#), a nonparametric kernel prediction is considered for spatial stochastic processes when a stochastic sampling design is assumed for selection of random locations. The particularity of this predictor is to be constructed with a kernel function on the locations. In the kernel-type estimator suggested in [García-Soidán and Menezes \(2012\)](#), the dependence structure is reduced to the estimation of one indicator variogram, as a nonparametric alternative to Matheron's indicator variogram. [Wang et al. \(2012\)](#) proposed a local linear spatio-temporal prediction model, using a kernel weight function taking into account the distance between sites. The works of [Dabo-Niang et al. \(2013\)](#) and [Dabo-Niang et al. \(2014\)](#) proposed, respectively, a spatial density and regression estimators, for multivariate data, depending on two kernels, one of which controls the distance between observations and the other controls the spatial dependence structure. All these previous works concern real valued data. The spatial kernel density estimator proposed in [Dabo-Niang et al. \(2011\)](#) for functional data does not directly take into account the spatial dependency in the form of the estimator but the authors explained how this can be done by introducing a second kernel, based on distances between sites. Here, we combine these three last works since the regression operator is constructed from the kernel density estimator introduced in [Dabo-Niang et al. \(2013\)](#) and [Dabo-Niang et al. \(2014\)](#), when the explanatory variables are defined on \mathbb{R}^d , but here adapted to the functional data framework.

Denote the integer lattice points in the N -dimensional Euclidean space by \mathbb{Z}^N , $N \geq 1$. Consider a strictly stationary random field $\{X_{\mathbf{i}}, Y_{\mathbf{i}}\}$ indexed by \mathbf{i} in \mathbb{Z}^N whose elements have the same distribution as a variable (X, Y) and 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 takes values in a separable semi-metric space $(\mathcal{E}, d(\cdot, \cdot))$ (of eventually infinite dimension) (i.e. X is a functional random variable and d a semi-metric) and Y takes values in \mathbb{R} . We are interested in the regression model defined by $Y_{\mathbf{i}} = r(X_{\mathbf{i}}) + \varepsilon_{\mathbf{i}}$ where the noise $\varepsilon_{\mathbf{i}}$ is centered, α -mixing and independent of $X_{\mathbf{i}}$. Then, the main goal of this paper is to estimate the regression function $r(\cdot)$.

In the following, we will assume, without loss of generality, that the data are observed over a rectangular region, defined by $\mathcal{S}_{\mathbf{n}} := \{\mathbf{i} : \mathbf{i} \in (\mathbb{N}^*)^N, 1 \leq i_k \leq n_k, k = 1, \dots, N\}$. Such regions are used in the literature to estimate nonparametrically the spatial density ([Tran \(1990\)](#); [Biau and Cadre \(2004\)](#); [Wang and Wang \(2009\)](#)). Let us recall that, as in any nonparametric spatial density model (see, e.g., [El Machkouri \(2011\)](#)), the method proposed here remains valid when the observed region has a more general form (e.g. subset of a large family of lattices of \mathbb{R}^N or $\mathcal{S}_{\mathbf{n}} \subset \mathbb{R}^2$ is a closed convex domain with non-empty interior). Let $\hat{\mathbf{n}} := n_1 \times \dots \times n_N$ be the sample size. The letter C will be used to denote constants whose values are unimportant, $\|\cdot\|$ will denote

any norm over \mathbb{Z}^N and $B(x, \rho)$ the opened ball of center x and radius ρ . We will write $\mathbf{n} \rightarrow \infty$ if $\min_{k=1, \dots, n} n_k \rightarrow \infty$ and for all $1 \leq j, k \leq N$, for some constant $0 < C < \infty$, we assume $|n_j/n_k| < C$. This means that the number of observations on the rectangular region expands to infinity at the same rate along all directions. Such an expansion is called isotropic divergence. An other case could be considered, it is the less restrictive non-isotropic case where $\mathbf{n} \rightarrow \infty$ if $\min_{k=1, \dots, n} n_k \rightarrow \infty$. Note that the proof of the result obtained here is similar in the non-isotropic case.

Thereafter, we assume, without loss of generality, that $n_1 = n_2 = \dots = n_N = n$. For each site \mathbf{j} , let $k_{\mathbf{n}} = k_{\mathbf{n}, \mathbf{j}} = \sum_{\mathbf{i} \in \mathcal{S}_{\mathbf{n}}} \mathbf{1}_{[\|\mathbf{i}-\mathbf{j}\| \leq d_{\mathbf{n}}]}$ where $d_{\mathbf{n}} > 0$ is such that $d_{\mathbf{n}} \rightarrow \infty$ as $\mathbf{n} \rightarrow \infty$. Note that $k_{\mathbf{n}, \mathbf{j}}$ is the number of neighbors \mathbf{i} for which the distance between \mathbf{i} and \mathbf{j} is less or equal to distance $d_{\mathbf{n}}$. Taking the Euclidean distance and if $N = 2$, we have $k_{\mathbf{n}} \leq 4d_{\mathbf{n}}^2 - 4d_{\mathbf{n}} + 4$ which leads to $k_{\mathbf{n}} = O(d_{\mathbf{n}}^2)$ and $k_{\mathbf{n}} = o(d_{\mathbf{n}}^\eta)$, $\eta > 2$. Moreover, if $d_{\mathbf{n}} = o(\hat{\mathbf{n}}^\varepsilon)$, $0 < \varepsilon < 1$ then we have $k_{\mathbf{n}} = o(\hat{\mathbf{n}}^{2\varepsilon})$, see e.g. [Kelejian and Prucha \(2007\)](#).

Considering normalized sites, the proposed kernel regression estimator of r , for a fixed $x_{\mathbf{i}_0} \in (\mathcal{E}, d(\cdot, \cdot))$ located at a site \mathbf{i}_0 , is defined as

$$r_{\mathbf{n}}(x_{\mathbf{i}_0}) = \begin{cases} \frac{g_{\mathbf{n}}(x_{\mathbf{i}_0})}{f_{\mathbf{n}}(x_{\mathbf{i}_0})} & \text{if } f_{\mathbf{n}}(x_{\mathbf{i}_0}) \neq 0; \\ \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{S}_{\mathbf{n}}} Y_{\mathbf{i}} & \text{otherwise,} \end{cases}$$

where the functions $g_{\mathbf{n}}(x_{\mathbf{i}_0})$ and $f_{\mathbf{n}}(x_{\mathbf{i}_0})$ are defined by

$$g_{\mathbf{n}}(x_{\mathbf{i}_0}) = \sum_{\mathbf{i} \in \mathcal{S}_{\mathbf{n}}} \frac{1}{a_{\mathbf{n}, \mathbf{i}_0} \mathbb{E} \left[K_1 \left(\frac{d(x_{\mathbf{i}_0}, X_{\mathbf{i}})}{b_{\mathbf{n}}} \right) \right]} Y_{\mathbf{i}} K_1 \left(\frac{d(x_{\mathbf{i}_0}, X_{\mathbf{i}})}{b_{\mathbf{n}}} \right) K_{2, \rho_{\mathbf{n}}}(\|\mathbf{i}_0 - \mathbf{i}\|)$$

$$f_{\mathbf{n}}(x_{\mathbf{i}_0}) = \sum_{\mathbf{i} \in \mathcal{S}_{\mathbf{n}}} \frac{1}{a_{\mathbf{n}, \mathbf{i}_0} \mathbb{E} \left[K_1 \left(\frac{d(x_{\mathbf{i}_0}, X_{\mathbf{i}})}{b_{\mathbf{n}}} \right) \right]} K_1 \left(\frac{d(x_{\mathbf{i}_0}, X_{\mathbf{i}})}{b_{\mathbf{n}}} \right) K_{2, \rho_{\mathbf{n}}}(\|\mathbf{i}_0 - \mathbf{i}\|)$$

with $a_{\mathbf{n}, \mathbf{i}_0} = \sum_{\mathbf{j} \in \mathcal{S}_{\mathbf{n}}} K_{2, \rho_{\mathbf{n}}}(\|\mathbf{i}_0 - \mathbf{j}\|) = \sum_{\mathbf{j} \in \mathcal{S}_{\mathbf{n}}} K_2 \left(\rho_{\mathbf{n}}^{-1} \left\| \frac{\mathbf{i}_0 - \mathbf{j}}{\mathbf{n}} \right\| \right)$, (where $\frac{\mathbf{i}}{\mathbf{n}} = (\frac{i_1}{n}, \frac{i_2}{n}, \dots, \frac{i_N}{n})$). It can also be written that $K_{2, \rho_{\mathbf{n}}}(\|\mathbf{i}_0 - \mathbf{i}\|) = K_2 \left(\frac{\|\mathbf{i}_0 - \mathbf{i}\|}{n \rho_{\mathbf{n}}} \right)$. Moreover, K_1 and K_2 are kernels defined on \mathbb{R} , $b_{\mathbf{n}}$ and $\rho_{\mathbf{n}}$ are the bandwidths tending to zero. The estimator $f_{\mathbf{n}}(x_{\mathbf{i}_0})$ is a function of the number $k_{\mathbf{n}}$ for which distance $d_{\mathbf{n}}$ is chosen to be $d_{\mathbf{n}} = n \rho_{\mathbf{n}}$ with $k_{\mathbf{n}} \rightarrow \infty$ as $\mathbf{n} \rightarrow \infty$, $k_{\mathbf{n}} = O(d_{\mathbf{n}}^N) = O(\hat{\mathbf{n}} \rho_{\mathbf{n}}^N)$. Hereinafter, we assume that $k_{\mathbf{n}} = C_N d_{\mathbf{n}}^N + O(d_{\mathbf{n}}^\beta)$ as $d_{\mathbf{n}} \rightarrow \infty$, $0 < \beta < N$ and C_N is a constant that depends on N . This is based on the well-known problem of counting points with lattice coordinates in the N -dimensional ball (see the first point of Remark 1 for further explanations). Similar conditions on the number of observations \mathbf{i} in $\mathcal{S}_{\mathbf{n}}$ with $\left\| \frac{\mathbf{i}_0 - \mathbf{i}}{\mathbf{n}} \right\| \leq \rho_{\mathbf{n}}$ are used in [Wang and Wang \(2009\)](#) who studied a local linear fitting method for real spatio-temporal data using some weights. Then, in this latter article, additional conditions concern time characteristics.

Remark 1.

- To give some examples where the assumption on $k_{\mathbf{n}}$ is reasonable, consider q_N the number of standard lattice (in \mathbb{Z}^N) points contained in a closed ball $B(\mathbf{j}, d_{\mathbf{n}})$ that is $q_N = \text{Card}\{\mathbf{i} \in$

$\mathbb{Z}^N, \|\mathbf{i} - \mathbf{j}\| \leq d_{\mathbf{n}}$ where \mathbf{j} is any vector of \mathbb{R}^N . It is well known that

$$q_{\mathbf{n}} = \frac{\pi^{N/2}}{\Gamma(N/2 + 1)} d_{\mathbf{n}}^N + O(d_{\mathbf{n}}^{N-1}),$$

where $\Gamma(\cdot)$ is the Gamma function, see for instance [Mitchell \(1966\)](#); [Chamizo and Iwaniec \(1995\)](#); [Tsang \(2000\)](#); [Meyer \(2011\)](#). And notice that $k_{\mathbf{n}} = C_N q_{\mathbf{n}}$. In particular, if $N = 2$, $q_{\mathbf{n}} = \frac{\pi}{\Gamma(2)} d_{\mathbf{n}}^2 + O(d_{\mathbf{n}})$ or $q_{\mathbf{n}} = \frac{\pi}{\Gamma(2)} d_{\mathbf{n}}^2 + o(d_{\mathbf{n}}^{2/3})$.

- Instead of the previous functions $g_{\mathbf{n}}$ and $f_{\mathbf{n}}$, one can consider the simpler following versions

$$g_{\mathbf{n}}(x_{\mathbf{i}_0}) = \frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \mathbb{E} \left[K_1 \left(\frac{d(x_{\mathbf{i}_0}, X_1)}{b_{\mathbf{n}}} \right) \right]} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} K_1 \left(\frac{d(x_{\mathbf{i}_0}, X_{\mathbf{i}})}{b_{\mathbf{n}}} \right) K_{2, \rho_{\mathbf{n}}}(\|\mathbf{i}_0 - \mathbf{i}\|),$$

$$f_{\mathbf{n}}(x_{\mathbf{i}_0}) = \frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \mathbb{E} \left[K_1 \left(\frac{d(x_{\mathbf{i}_0}, X_1)}{b_{\mathbf{n}}} \right) \right]} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K_1 \left(\frac{d(x_{\mathbf{i}_0}, X_{\mathbf{i}})}{b_{\mathbf{n}}} \right) K_{2, \rho_{\mathbf{n}}}(\|\mathbf{i}_0 - \mathbf{i}\|).$$

Such functions allow the following result to remain valid with some minor changes in conditions on $k_{\mathbf{n}}$.

The rest of the paper is organized as follows. In Section 2, we provide the assumptions, state our main result and present an example of application of $r_{\mathbf{n}}(x_{\mathbf{i}_0})$ to prevision. To check the performance of our estimator, numerical results are reported in Section 3. Conclusion is given in Section 4 while proofs and technical lemmas are postponed in the Appendix section.

2. Assumptions and main result

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

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

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

$$\gamma(i) \leq Ci^{-\theta}, \text{ for some } \theta > 0,$$

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

$$\gamma(i) \leq C \exp(-si), \text{ for some } s > 0,$$

i.e. that $\gamma(i)$ tends to zero at an exponential rate. Concerning the function $\gamma(\cdot)$, for the sake of simplicity, we will only study the case where $\gamma(\cdot)$ tends to zero at a polynomial rate. However,

similar result to that of Theorem 1 could be obtained with $\gamma(\cdot)$ tending to zero at an exponential rate (which implies the polynomial case). Throughout the paper, it will be assumed that ψ satisfies either

$$\forall n, m \in \mathbb{N}, \quad \psi(n, m) \leq C \min(n, m) \quad \text{or} \quad \psi(n, m) \leq C(n + m + 1)^{\tilde{\beta}}$$

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

The consistency result of $r_{\mathbf{n}}$ is obtained under the following assumptions (A1-A6) on r , the kernels, the bandwidths and local dependence condition. We will denote by p the probability distribution of the (X_i) 's and by $p_{\mathbf{i}, \mathbf{j}}$ the joint probability distribution of (X_i, X_j) , for all \mathbf{i} and \mathbf{j} .

- **A1:** The kernels $K_i : \mathbb{R} \rightarrow \mathbb{R}^+$, $i = 1, 2$, are of integral 1 and are such that there exist two constants C_1 and C_2 with $0 < C_1 < C_2 < \infty$, such that

$$C_1 \mathbf{1}_{[0,1]}(t) \leq K_i(t) \leq C_2 \mathbf{1}_{[0,1]}(t).$$

- **A2:** $r(\cdot)$ is a Lipschitz function, that is $r \in Lip_{\mathcal{E}}$ where

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

- **A3: Local dependence condition** For all $\mathbf{i} \neq \mathbf{j} \in \mathbb{N}^N$, the joint probability distribution $p_{\mathbf{i}, \mathbf{j}}$ of X_i and X_j satisfies

$$\exists \varepsilon_1 \in (0, 1], \quad p_{\mathbf{i}, \mathbf{j}}(B(x_{\mathbf{i}_0}, b_{\mathbf{n}}) \times B(x_{\mathbf{j}_0}, b_{\mathbf{n}})) \leq C_4 (\varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}}))^{1+\varepsilon_1},$$

where $\varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}}) = \mathbb{P}[X \in B(x_{\mathbf{i}_0}, b_{\mathbf{n}})]$, called small ball probability in the literature (e.g. Ferraty and Vieu (2006)).

- **A4:** $\forall n, m \in \mathbb{N}$, $\psi(n, m) \leq C \min(n, m)$ and $\hat{\mathbf{n}} \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}})^{\theta_1} \rho_{\mathbf{n}}^{N\theta_1} \log \hat{\mathbf{n}}^{-\theta_1} \rightarrow \infty$ with the mixing coefficient $\theta > 4N$ and with $\theta_1 = \frac{2N - \theta}{4N - \theta}$.
- **A5:** $\forall n, m \in \mathbb{N}$, for some $\tilde{\beta} \geq 1$, $\psi(n, m) \leq C(n + m + 1)^{\tilde{\beta}}$ and $\hat{\mathbf{n}} \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}})^{\theta_1^*} \rho_{\mathbf{n}}^{N\theta_1^*} \log \hat{\mathbf{n}}^{-\theta_1^*} \rightarrow \infty$ with the mixing coefficient $\theta > N(3 + 2\tilde{\beta})$ and with $\theta_1^* = \frac{N - \theta}{N(3 + 2\tilde{\beta}) - \theta}$.
- **A6:** The variable Y is bounded almost surely and $|Y| < M$.

Remark 2.

- **Assumptions A1 and A2 allow to control the bias of the estimator.**
 - *Assumption A1 concerns the kernels K_i , $i = 1, 2$. More general kernels such as Gaussian or Silverman can also be used but for simplicity of calculations, we consider such kernels usually considered in nonparametric regression. For example, this condition is verified by e.g. triangular (Bartlett), biweight, circular (cosine), Epanechnikov, Parzen, Tukey-Hanning kernels.*

- A nonparametric assumption on the regression function is considered through hypothesis **A2**. In fact, this Lipschitz condition allows the precise rate of convergence to be found whereas a continuity-type model would give only convergence results. Assuming the continuity condition, one can obtain that

$$r_{\mathbf{n}}(x_{i_0}) - r(x_{i_0}) \xrightarrow{m.s.} 0 \quad \text{with } \widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_{x_{i_0}}(b_{\mathbf{n}}) \rightarrow \infty, b_{\mathbf{n}} \rightarrow 0 \text{ and } \rho_{\mathbf{n}} \rightarrow 0.$$

- Assumption **A3** concerns the local dependency and a consequence is

$$\begin{aligned} |p_{i,j}(B(x_{i_0}, b_{\mathbf{n}}) \times B(x_{i_0}, b_{\mathbf{n}})) - (\varphi_{x_{i_0}}(b_{\mathbf{n}}))^2| &\leq |C_4(\varphi_{x_{i_0}}(b_{\mathbf{n}}))^{1+\varepsilon_1} - (\varphi_{x_{i_0}}(b_{\mathbf{n}}))^2| \\ &\leq C_4(\varphi_{x_{i_0}}(b_{\mathbf{n}}))^{1+\varepsilon_1} \leq 1. \end{aligned}$$

As it is noticed in [Dabo-Niang et al. \(2011\)](#), this result can be linked with the classical local dependence condition met in the literature of real valued data when X and (X_i, X_j) admit, respectively, the densities f and $f_{i,j}$. Such assumption can be also found in [Ferraty and Vieu \(2006\)](#) (Chapter 11, page 163) and in [Dabo-Niang and Yao \(2013\)](#).

- Assumptions **A4** and **A5** concern the mixing dependency and are similar to those of [Carbon et al. \(1997\)](#).

The following theorem states the pointwise mean square convergence of the proposed regression function estimator, whose proof is given in Appendix. We will denote $\|r_{\mathbf{n}}(x_{i_0}) - r(x_{i_0})\|_2 = (\mathbb{E}[(r_{\mathbf{n}}(x_{i_0}) - r(x_{i_0}))^2])^{1/2}$.

Theorem 1. Under assumptions **A1-A3**, **A4** or **A5** and **A6**, we have

$$\|r_{\mathbf{n}}(x_{i_0}) - r(x_{i_0})\|_2 = O\left(b_{\mathbf{n}} + \sqrt{\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_{x_{i_0}}(b_{\mathbf{n}})}}\right).$$

Precisely, we have

$$\|r_{\mathbf{n}}(x_{i_0}) - r(x_{i_0})\|_2 = C_3 \times b_{\mathbf{n}} + (2C(2MC_2 + 2M\sqrt{C_4} + C_0) + 4M) \times \sqrt{\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_{x_{i_0}}(b_{\mathbf{n}})}},$$

where C depends on N (see e.g. [Chamizo and Iwaniec \(1995\)](#)) and is such that $k_{\mathbf{n}} \leq C\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N$ whereas C_0 is a constant depending on the constant appearing in Lemma 2.

This result permits to have a bound of the mean squared error of $r_{\mathbf{n}}(\cdot)$ that depends on $\rho_{\mathbf{n}}$. This is linked with the fact that $r_{\mathbf{n}}(\cdot)$ incorporates the dependence between sites compared to the result of [Dabo-Niang et al. \(2011\)](#). In the foregoing paper, the authors used the following functional regression estimator

$$r_{\mathbf{n}}^*(x) = \frac{\sum_{i \in \mathcal{J}_{\mathbf{n}}} Y_i K_1\left(\frac{d(X_i, x)}{b_{\mathbf{n}}^*}\right)}{\sum_{i \in \mathcal{J}_{\mathbf{n}}} K_1\left(\frac{d(X_i, x)}{b_{\mathbf{n}}^*}\right)}$$

where K_1 is a kernel and $b_{\mathbf{n}}^*$ is the corresponding bandwidth. They gave an uniform almost sure bound of $|r_{\mathbf{n}}^*(x) - r(x)|$ on a specific set \mathcal{C} , that is $O\left(b_{\mathbf{n}}^* + \sqrt{\frac{\log \widehat{\mathbf{n}}}{\Gamma(b_{\mathbf{n}}^*)\widehat{\mathbf{n}}}}\right)$ with $\Gamma(b_{\mathbf{n}}^*) = \sup_{x \in \mathcal{C}} \varphi_x(b_{\mathbf{n}})^*$.

For multivariate data, [Dabo-Niang et al. \(2013\)](#) focus on a rate of almost complete convergence of the density estimator $f_{\mathbf{n}}(v_{\mathbf{i}})$ of $f(v_{\mathbf{i}})$, for \mathbb{R} -valued spatial data $v_{\mathbf{i}}$, depending on two kernels. They obtained that

$$|f_{\mathbf{n}}(v_{\mathbf{i}}) - f(v_{\mathbf{i}})| = O\left(a_{\mathbf{n}} + \sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} \tau_{\mathbf{n}}^N a_{\mathbf{n}}^d}}\right), \quad \text{a. c.}$$

where $a_{\mathbf{n}}$ and $\tau_{\mathbf{n}}$ are the bandwidths corresponding to the kernels on the observations and on the sites respectively. This work is extended to regression estimation for multivariate data by [Dabo-Niang et al. \(2014\)](#).

Remark 3.

- *This current work is supported by a particular sampling scheme, which only includes deterministic designs for the spatial locations. For this reason, the bound of Theorem 1 shows a dissymmetric contribution of $b_{\mathbf{n}}$ and $\rho_{\mathbf{n}}$ on the risk even though both kernels K_1 and K_2 play symmetric roles. One can generalize this work to random spatial sample such as in [Menezes et al. \(2010\)](#) (for real-valued regression) and in [Kelejian and Prucha \(2007\)](#) (for spatial HAC estimation) and have a bound including $\rho_{\mathbf{n}}^\alpha$.*
- *Theorem 1 deals with local convergence (for a fixed $x_{\mathbf{i}_0}$) of the regression estimate but one can extend the obtained result to uniform one, on a set where corresponding sites are in a set S (that can be a subset of $\mathcal{I}_{\mathbf{n}}$ or a set larger than $\mathcal{I}_{\mathbf{n}}$) by considering $l_{\mathbf{n}} = \sup_{\mathbf{j} \in S} k_{\mathbf{n},\mathbf{j}}$.*

The remainder of this section focuses on the application of the proposed regression function through an example, namely the spatial prediction.

Application to spatial prediction

In spatial statistic, an important topic, encountered in the literature, concerns the spatial prediction. One of the most popular method is kriging, which was developed at the beginning of the 1950's and studied in the scope of geostatistics. More recently, some works proposed nonparametric predictors for spatial fields indexed by lattices. The first results in this direction are those of [Biau and Cadre \(2004\)](#) and concerned the kernel prediction of a strictly stationary random field indexed in $(\mathbb{N}^*)^N$. Later, [Dabo-Niang and Yao \(2007\)](#) contribute to [Biau and Cadre \(2004\)](#)'s investigations since they are interested in the kernel regression estimation and prediction of continuously indexed random fields. In [Menezes et al. \(2010\)](#), nonparametric kernel prediction is considered for spatial stochastic processes when a stochastic sampling design is assumed for selection of random locations. These contributions, but also [Dabo-Niang et al. \(2014\)](#), dealt with multivariate data. In [Dabo-Niang et al. \(2011\)](#), the authors stated that their work, dealing with the spatial regression estimator for functional data, offers some interesting perspectives of investigation, namely in spatial forecasting and real data problem. In continuation of these works, we propose a spatial prediction methodology dealing with functional data, taking explicitly into account the spatial locations.

In this application, we consider a \mathbb{R} -valued strictly stationary random spatial process $(\eta_{\mathbf{i}}, \mathbf{i} \in (\mathbb{R}^*)^N)$. This process is assumed to be bounded and observed over a subset $\mathcal{O}_{\mathbf{n}} \subset \mathcal{I}_{\mathbf{n}}$. We are interested in predicting $\eta_{\mathbf{i}_0}$ at an unobserved given location $\mathbf{i}_0 \in \mathcal{I}_{\mathbf{n}} \setminus \mathcal{O}_{\mathbf{n}}$. In practice, we expect

that $\eta_{\mathbf{i}_0}$ depends only on the values of the process in a bounded neighborhood $V_{\mathbf{i}_0} = \mathbf{i}_0 + V \subset \mathcal{O}_n$, where $\mathbf{0} = (0, 0, \dots, 0) \notin V$. Consequently, we can construct a function $\tilde{\eta}_{\mathbf{i}_0}$ from the observations in a continuous vicinity $V_{\mathbf{i}_0}$ of \mathbf{i}_0 and define $\tilde{\eta}_{\mathbf{i}} = \{\eta_{\mathbf{j}}, \mathbf{j} \in V_{\mathbf{i}} = \mathbf{i} + V \subset \mathbb{R}^N\}$ which belongs to the space of continuous and bounded functions. For more details on the choice of V , see [Dabo-Niang and Yao \(2007\)](#).

To achieve the forecasting at the site \mathbf{i}_0 , we propose to use the regression function estimator r_n suggested previously. Then, the value to be predicted of the field $(\eta_{\mathbf{i}})_{\mathbf{i} \in (\mathbb{R}^*)^N}$ at a site \mathbf{i}_0 becomes

$$\hat{\eta}_{\mathbf{i}_0} = r_n(\tilde{\eta}_{\mathbf{i}_0}) = \frac{\sum_{\mathbf{i} \in \mathcal{O}_n} \eta_{\mathbf{i}} K_1\left(\frac{d(\tilde{\eta}_{\mathbf{i}_0}, \tilde{\eta}_{\mathbf{i}})}{b_n}\right) K_{2, \rho_n}(\|\mathbf{i}_0 - \mathbf{i}\|)}{\sum_{\mathbf{i} \in \mathcal{O}_n} K_1\left(\frac{d(\tilde{\eta}_{\mathbf{i}_0}, \tilde{\eta}_{\mathbf{i}})}{b_n}\right) K_{2, \rho_n}(\|\mathbf{i}_0 - \mathbf{i}\|)}.$$

One can derive an asymptotic result such as mean square convergence of $\hat{\eta}_{\mathbf{i}_0}$ by considering a kernel regression estimate of functional spatial random variables continuously indexed. Having checked the theoretical behavior of our estimator and presented a potential application, we are going to study its practical behavior through some numerical results.

3. Numerical results

In this section, we study the performance of the proposed regression estimator through some simulations which point out the importance of taking into account the spatial locations of the data. We remind that our theoretical result is obtained under a mixing condition which can be considered by the kernel function on the locations. We compare our estimator with the one that ignores any spatial dependence in the structure of the regression estimate (see [Dabo-Niang et al. \(2011\)](#)). We consider a sample of dependent realizations of some spatial functional variables $X_{\mathbf{i}}$ with the same distribution as a random field X valued in some infinite dimensional semi-metric space $(\mathcal{E}, d(\cdot, \cdot))$. That is, on each site \mathbf{i} , we have a curve $X_{\mathbf{i}}$ such that $X_{\mathbf{i}} = \{X_{\mathbf{i}}(t), t \in [0, T]\}$. Before studying the numerical results, we propose a useful procedure for estimating the spatial regression function.

3.1. Procedure of estimation of $r(X_{\mathbf{j}})$, $\mathbf{j} \in \mathcal{J}_n$

1. Specify sets of bandwidths $S(b)$ and $S(\rho)$ of respectively K_1 and K_2 .
2. For each $b_n \in S(b)$ and $\rho_n \in S(\rho)$ and each $\mathbf{j} \in \mathcal{J}_n$, compute

$$r_n(X_{\mathbf{j}}) = \frac{\sum_{\substack{\mathbf{i} \in \mathcal{J}_n \\ \mathbf{i} \neq \mathbf{j}}} Y_{\mathbf{i}} K_1\left(\frac{d(X_{\mathbf{i}}, X_{\mathbf{j}})}{b_n}\right) K_2\left(\rho_n^{-1} \left\| \frac{\mathbf{i} - \mathbf{j}}{\mathbf{n}} \right\| \right)}{\sum_{\substack{\mathbf{i} \in \mathcal{J}_n \\ \mathbf{i} \neq \mathbf{j}}} K_1\left(\frac{d(X_{\mathbf{i}}, X_{\mathbf{j}})}{b_n}\right) K_2\left(\rho_n^{-1} \left\| \frac{\mathbf{i} - \mathbf{j}}{\mathbf{n}} \right\| \right)}$$

3. Compute $b_{n,opt}$ and $\rho_{n,opt}$ by applying a cross-validation procedure over $S(b)$ and $S(\rho)$. More precisely, consider the following minimization problem, i.e. determine $b_{n,opt}$ and $\rho_{n,opt}$ which minimize the mean squared error over the $\hat{\mathbf{n}}$ sites

$$\min_{b_n, \rho_n} \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{j} \in \mathcal{J}_n} (r_n(X_{\mathbf{j}}) - r(X_{\mathbf{j}}))^2$$

4. For each \mathbf{j} , compute $r_{\mathbf{n},opt}(X_{\mathbf{j}})$ corresponding to $b_{\mathbf{n},opt}$ and $\rho_{\mathbf{n},opt}$.

3.2. Simulation

This last procedure is used in the following simulation study dealing with $N = 2$. We consider observations $(X_{(i,j)}, Y_{(i,j)})$, $1 \leq i, j \leq 25$, such that

$$\begin{aligned} Y_{(i,j)} &= r(X_{(i,j)}) + \varepsilon_{(i,j)} \\ &= 4 \times A_{(i,j)}^2 + \varepsilon_{(i,j)} \end{aligned}$$

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

Case 1: $X_{(i,j)}(t) = A_{(i,j)}^2 \times (t - 0.5)^2 + A_{(i,j)} \times B_{(i,j)}$;

Case 2: $X_{(i,j)}(t) = A_{(i,j)} \times \cos(2\pi t)$,

where $A = (A_{(i,j)})$, $B = (B_{(i,j)})$ and $\varepsilon = (\varepsilon_{(i,j)})$ are random variables which will be specified according to the following considered model on $A = (A_{(i,j)})$. Several curve examples of $X_{(i,j)}(t)$, for each case, are drawn on Figure 1. More precisely, the figure on the left displays some curves simulated from Case 1 and that on the right concerns Case 2. In Case 1, an example of the function $r(\cdot)$ could be $r(X) = 2X''$ (where X'' denotes the second derivative of X with respect to t) whereas in Case 2, it could be $r(X) = \frac{A}{\pi^2} X''$ with $t = \frac{1}{2}$. We will denote by $GRF(m, \sigma^2, s)$ any stationary Gaussian Random Field with mean m and spatial covariance function defined by

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

Then, we define the two considered models on $A = (A_{(i,j)})$ by

Model A: $A_{i,j} = D_{i,j} \times (\sin(2G_{i,j}) + 2 \exp(-16G_{i,j}^2))$;

Model B: $A_{i,j} = D_{i,j} \times (2 \cos(2G_{i,j}) + \exp(-4G_{i,j}^2))$.

Here, the number of observations $\hat{\mathbf{n}}$ is equal to 25×25 , i.e. 625. The several fields are defined by $D_{i,j} = \frac{1}{625} \sum_{1 \leq m, t \leq 25} \exp\left(-\frac{\|(i,j)-(m,t)\|}{a}\right)$, $G_{i,j} = GRF(0, 5, 3)$, $B_{i,j} = GRF(2.5, 5, 3)$ and $\varepsilon_{i,j} = GRF(0, 0.1, 5)$. We note that the spatial dependence is controlled by the value of a . In fact, the greater a is, the weaker the spatial dependency is. According to this fact, we provide simulation results obtained with different values of a which are $a = 5, 20$ and 50 .

Along this part, the spatial regression is computed based on the kernels K_1 as the Epanechnikov kernel and K_2 as the Parzen kernel. The choice of the semi-metric $d(\cdot, \cdot)$ is important and depends on the information one gets on the data. Ferraty and Vieu (2006) present three families of semi-metrics. The first is built from functional principal component analysis (FPCA) and is adapted to rough curves. The second is built from the partial least square (PLS) approach and is relevant when one consider multivariate response. The last, based on derivatives, is well adapted in the presence of smooth curves. Specifically, it approximates L_2 metric between derivatives of the curves based on their B -spline representation. Note that other semi-metrics are encountered in the literature. However, according to Delsol (2008), the theoretical justification of the usefulness of a particular semi-metric is still an open problem. In this work, we consider a semi-metric between curves based on their first $q = 2$ derivatives because of the smoothness of the curves.

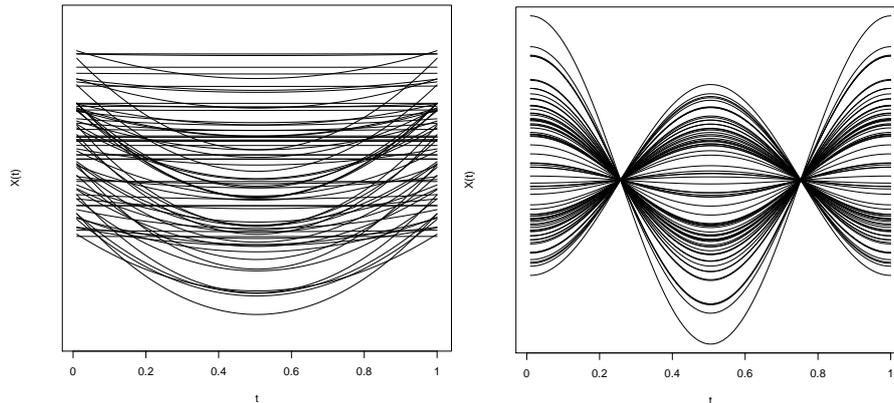


Figure 1: Some simulated curves of Case 1 (left) and Case 2 (right)

This semi-metric (between X_i and X_j) is defined by

$$\sqrt{\int \left(X_i^{(q)}(t) - X_j^{(q)}(t) \right)^2 dt}, \quad q = 0, 1, 2, \dots$$

where, for any q -times differentiable real function X , $X^{(q)}$ denotes the q th derivative of X (we refer, for example, to [Ferraty and Vieu \(2006\)](#) for the theoretical setting about semi-metrics used for functional nonparametric investigations). To confirm our semi-metric choice, we tested, in addition to the semi-metrics based on their first derivatives, two other semi-metrics (based on PCA and on Fourier’s decomposition) and different parameters such as the number of derivatives, principal components, basis, etc. It turns out that the results are similar or worse than those obtained considering a semi-metric between curves based on their first $q = 2$ derivatives.

Recall that, in the work of [Dabo-Niang et al. \(2011\)](#), a theoretical estimator of the spatial regression function for functional data is proposed. This estimator does not directly take into account the spatial locations. However, in the application section, the authors explained how this can be done using the k -nearest neighbors method. Then, in the simulation study, they proposed a procedure of estimation based on nearest neighbors. This combination looks like to the estimator $r_n(x_{i_0})$ introduced in this paper. The difference is that in [Dabo-Niang et al. \(2011\)](#) the k -nearest neighbors of a point \mathbf{i}_0 are considered in the pointwise regression estimation whereas, with our methodology, all the points in the ball of radius $\rho_{n,opt}$ and center \mathbf{i}_0 are considered.

To evaluate the performance of the proposed regression estimator, now denoted by $r_n^\sharp(\cdot)$ and to compare it with the one that does not directly take into account the distance between locations and denoted $r_n^*(\cdot)$ (the theoretical estimator introduced in [Dabo-Niang et al. \(2011\)](#)), each studied model is replicated 30 times. Recall that $r_n^\sharp(\cdot)$ and $r_n^*(\cdot)$ are defined by

$$r_n^\sharp(X_j) = \frac{\sum_{\substack{\mathbf{i} \in \mathcal{I}_n \\ \mathbf{i} \neq \mathbf{j}}} Y_i K_1 \left(\frac{d(X_i, X_j)}{b_n^\sharp} \right) K_2 \left(\rho_n^{-1} \left\| \frac{\mathbf{i} - \mathbf{j}}{\mathbf{n}} \right\| \right)}{\sum_{\substack{\mathbf{i} \in \mathcal{I}_n \\ \mathbf{i} \neq \mathbf{j}}} K_1 \left(\frac{d(X_i, X_j)}{b_n^\sharp} \right) K_2 \left(\rho_n^{-1} \left\| \frac{\mathbf{i} - \mathbf{j}}{\mathbf{n}} \right\| \right)} \quad \text{and} \quad r_n^*(X_j) = \frac{\sum_{\substack{\mathbf{i} \in \mathcal{I}_n \\ \mathbf{i} \neq \mathbf{j}}} Y_i K_1 \left(\frac{d(X_i, X_j)}{b_n^*} \right)}{\sum_{\substack{\mathbf{i} \in \mathcal{I}_n \\ \mathbf{i} \neq \mathbf{j}}} K_1 \left(\frac{d(X_i, X_j)}{b_n^*} \right)}.$$

At each replication k , we compute the mean squared error over the $\hat{\mathbf{n}}$ sites. The bandwidths used, different at each replication, are those obtained using the previous procedure 3.1. For the k^{th}

TABLE 1. Simulation results according to the models A and B, the cases 1 and 2 and the value of $a = 5, 20$ and 50 : the table gives the average mean squared errors (AMSE) for each situation and in brackets the corresponding standard deviation. The column entitled “ p -value” gives the p -value of a paired t -test performing in order to determine whether $AMSE^\sharp$ is significantly less than $AMSE^*$. The two last columns display the average coefficients of determination (AR^2).

Model	Case	a	AMSE [‡]	AMSE [*]	p -value	AR ^{2‡}	AR ^{2*}
A	1	5	0.034 (0.014)	0.095 (0.030)	3.92×10^{-14}	0.652	0.057
		20	0.041 (0.013)	0.097 (0.024)	3.30×10^{-17}	0.956	0.896
		50	0.060 (0.014)	0.100 (0.022)	3.66×10^{-13}	0.981	0.969
	2	5	0.007 (0.003)	0.093 (0.030)	3.94×10^{-16}	0.925	0.054
		20	0.036 (0.006)	0.097 (0.031)	6.84×10^{-13}	0.960	0.895
		50	0.058 (0.011)	0.100 (0.031)	1.12×10^{-09}	0.982	0.970
B	1	5	0.012 (0.004)	0.092 (0.029)	6.86×10^{-16}	0.914	0.361
		20	0.049 (0.008)	0.100 (0.029)	3.52×10^{-12}	0.994	0.988
		50	0.071 (0.014)	0.100 (0.025)	1.56×10^{-10}	0.998	0.997
	2	5	0.010 (0.001)	0.093 (0.030)	7.65×10^{-16}	0.926	0.356
		20	0.060 (0.010)	0.100 (0.031)	4.58×10^{-10}	0.993	0.988
		50	0.086 (0.017)	0.108 (0.031)	7.23×10^{-07}	0.997	0.996

replication, we define the mean squared error ($MSE^{(k)}$) by

$$MSE^{(k)} = \frac{1}{\hat{n}} \sum_{j \in \mathcal{J}_n} (r_{n,opt}^\dagger(X_j) - Y_j)^2, \quad \text{with } r_n^\dagger = r_n^\sharp \text{ or } r_n^*. \quad (1)$$

The obtained results are summarized in Table 1. For each situation (Model, Case and value of a), this table provides the average MSE over the 30 replications of Equation (1) and the corresponding standard deviation. The $AMSE^\sharp$'s (average mean squared error) column makes reference to the proposed estimator r_n^\sharp whereas the $AMSE^*$'s column corresponds to the estimator r_n^* which takes no account of location. Besides, we use a statistical hypothesis test rather than directly compare the average MSE accuracy. The column entitled “ p -value” gives, for each considered situation, the p -value of a paired t -test performing in order to determine if MSE^\sharp is significantly less than MSE^* (the alternative hypothesis is then $H_1: MSE^\sharp < MSE^*$). The two last columns give the average of the coefficients of determination over the 30 replications. Recall that a value of R^2 close to 1 means that the quality of estimation is reliable.

The first general point to make about this study is that, when $a = 5$, regardless the considered kind of model or case, the estimator r_n^\sharp leads to better results since the mean squared errors are significantly lower than with r_n^* . For instance, for Model A and Case 2, the average of the mean squared errors is 0.007 using the estimator r_n^\sharp and 0.093 with r_n^* . Moreover, it can be seen that the standard deviations are greater with r_n^* than with r_n^\sharp . Secondly, we note that when the value of a increases, $AMSE^\sharp$ is still higher than $AMSE^*$ but the difference becomes narrower. Consequently, the higher the value of a (less spatial dependency), the lower the difference between the results of the two estimators is. In other words, our estimator r_n^\sharp outperforms r_n^* when the spatial dependence is important. However, the two estimators tend to give similar performance in case of spatially independent fields. The low p -values (less than 7.23×10^{-07}) confirm that r_n^\sharp produces less errors than r_n^* . Nevertheless, the probability of erroneously rejecting the null hypothesis is highest when the value of a is equal to 20 or 50 rather than 5 (without one exception) since the p -value increases with the value of a . Finally, we may note that the R^2 criterion strengthens the previous comments.

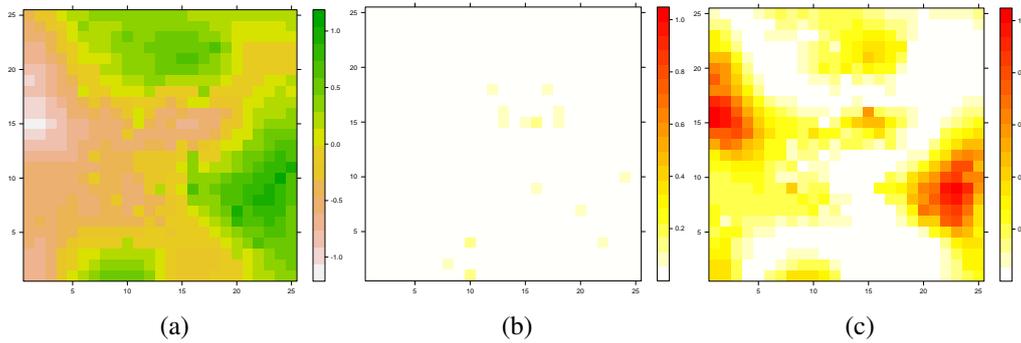


Figure 2: A simulated field considering Model A, Case 2 and $a = 5$ with (a) an image of the field Y ; (b) the squared errors using $r_n^\#$; (c) the squared errors using r_n^*

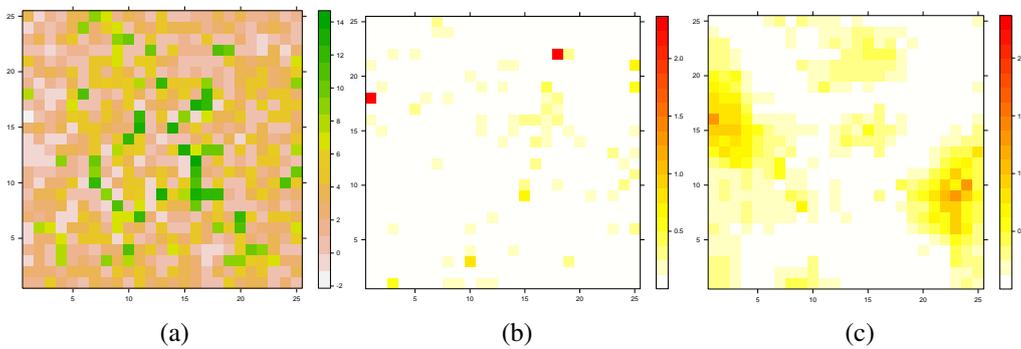


Figure 3: A simulated field considering Model B, Case 1 and $a = 20$ with (a) an image of the field Y ; (b) the squared errors using $r_n^\#$; (c) the squared errors using r_n^*

In fact, the values $AR^{2\#}$ are higher than AR^{2*} and the difference between them decreases as the value of a increases.

Insight into the performance of the two regression estimators can also be viewed from graphical outputs. In fact, Figures 2, 3 and 4 illustrate different situations. The first deals with spatially dependent data ($a = 5$) simulated from Model A and Case 2 of which a representation of Y is depicted in Figure 2a. Figures 2b and 2c show squared errors (more precisely, at each site \mathbf{j} , $[r_n(X_j) - Y_j]^2$ is represented) obtained using functions $r_n^\#$ and r_n^* , respectively. These two figures confirm that our methodology generates less errors than using the regression function r_n^* since the more colorful the representation is, the greater the error is. Figure 3 considers lower spatial dependence ($a = 20$) simulated from Model B and Case 1 for which the field Y is represented in Figure 3a. Figure 3b displays slightly less errors than in Figure 3c. Finally, Figure 4 gives summarized results of Model B and Case 2, with almost independent spatial data ($a = 50$). The two estimators seem to provide similar errors according to Figures 4b and 4c. It is not surprising to note that when a is high the two estimators produce similar results. In fact, in this situation, the bandwidths ρ_n are large and could take the maximal distance between observations. In short, the two estimators work in an almost identical manner in lack of spatial dependence.

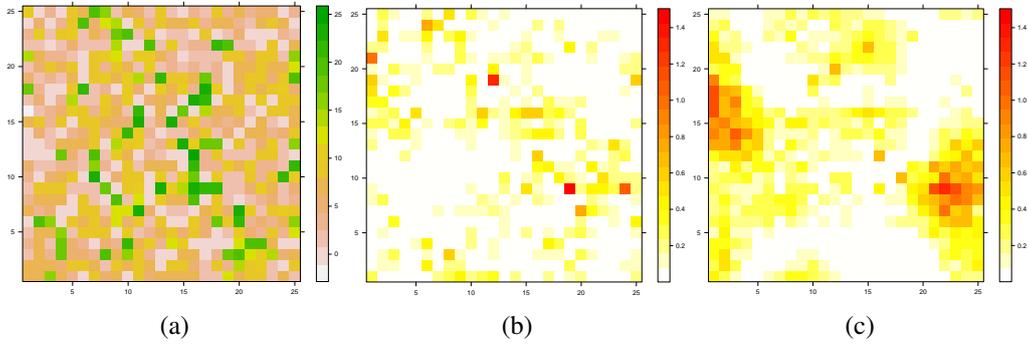


Figure 4: A simulated field considering Model B, Case 2 and $a = 50$ with (a) an image of the field Y ; (b) the squared errors using $r_n^\#$; (c) the squared errors using r_n^*

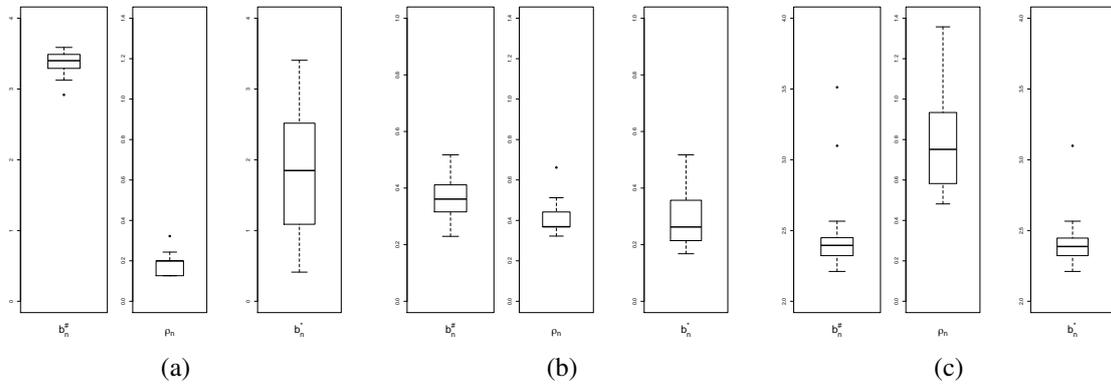


Figure 5: Boxplots of $b_{n,opt}^\#, \rho_{n,opt}$ and $b_{n,opt}^*$ respectively, over the 30 replications of the three following situations: (a) Model A, Case 2 and $a = 5$; (b) Model B, Case 1 and $a = 20$; (c) Model B, Case 2 and $a = 50$

Regarding the bandwidths selection, we carried out a cross-validation procedure. This selection is made differently, according to the situation, $r_n^\#$ and r_n^* . Firstly, with spatially dependent data ($a = 5$) the selected bandwidths ρ_n have the smallest values. This result was expected because when the spatial dependence is high, sites that are close together tend to be more related than sites that are far apart. From Model A and Case 2, the bandwidths $\rho_{n,opt}$, dealing with the kernel on the locations, are between 0.126 and 0.322. For the bandwidth linked to the distance between the observations (according to K_1), the selection differs according to the considered estimator. In fact, the values of $b_{n,opt}$ are widely lower considering r_n^* rather than $r_n^\#$. For more details on the values of the optimal bandwidths, through the 30 replications, Figure 5a displays the corresponding boxplots. Secondly, when $a = 20$, considering Model B and Case 1, the values of $\rho_{n,opt}$ are slightly higher than when $a = 5$ with values comprised between 0.322 and 0.662 (see Figure 5b). Finally, considering $a = 50$ with Model B and Case 2, the values of $\rho_{n,opt}$ are more scattered and higher than with $a = 5$ or 20 since it varies between 0.482 and 1.358 at each run (see Figure 5c). Moreover, for $a = 20$ and $a = 50$ the bandwidth selection of $b_{n,opt}$ is equivalent using $r_n^\#$ or r_n^* (see Figures 5b and 5c). In these situations, the value of $\rho_{n,opt}$ varies at each run while the locations do not change. In fact, contrary to the condition $a = 5$, the values of $X_{i,j}(t)$ are more scattered and then imply a change in the value of $\rho_{n,opt}$.

The previous study highlights the reliable performance of our estimator, particularly in presence of spatial dependence. But a disadvantage may be that the cross-validation procedure on the two parameters b_n and ρ_n is very time-consuming. To this end, we tried to deal with simulations considering a fixed bandwidth ρ_n as in Kelejian and Prucha (2007) where it is advised to take $d_n = n\rho_n = \lfloor \hat{n}^{1/4} \rfloor$ with $\lfloor \cdot \rfloor$ denotes the integer part. In our case, with $\hat{n} = 625$ sites, the corresponding bandwidths would be $\rho_n \approx 0.20$. It allows to save time and obtain results that are quite satisfactory when the spatial dependence is high. More precisely, when $a = 5$ the results are similar or slightly worse than those obtained by the cross-validation procedure on the two parameters: it is explained by the fact that the cross-validation procedure chooses a value of ρ_n close to 0.20 (different at each replication). Nevertheless, the fixed bandwidth $\rho_n = 0.20$ produces better results than using the estimator r_n^* . Note that the results depend largely on the spatial dependence structure considered. However, the results are worse with weaker spatial dependence ($a = 20$ or 50). In fact, in some cases (depending on the spatial dependency) the performance obtained by fixing ρ_n (according to the sample size \hat{n} as above) is poorer than those obtained using the estimator r_n^* . In this case, the cross-validation procedure on the two parameters remains necessary.

4. Conclusion

This work proposes a new method to model spatial regression function for functional random fields. Our main theoretical contribution was to derive the convergence in mean square. One can see the proposed methodology as a good alternative to the classical kernel approach for functional spatial data. More precisely, it is apparent that the proposed approach is particularly well adapted to the spatial regression estimation, for functional data, in presence of spatial dependence. This good behavior is observed both from an asymptotic point of view and from a simulation study. However, in case of low spatial dependence, the two estimators, herein called $r_n^\#$ and r_n^* , produce similar results.

In addition, this work offers very interesting perspectives of investigation. First of all, a future

work will be tied up to the uniform convergence of our estimator. Then, we could improve the choice of b_n and ρ_n which is outside the scope of this paper. For further study, we could investigate this new approach using local linear spatial regression (see, for example, Hallin et al. (2004)). Also, an adaptation of this method to issues such as the spatial conditional mode or quantile regression estimation could be developed. Application of the proposed regression estimator to real data, and more particularly to data collected by the French Research Institute for Exploitation of the Sea (Ifremer) during the campaign IBTS (International Bottom Trawl Survey), will be investigated. Moreover, an other perspective is the study of regression estimation for continuous indexed spatial functional fields $\{Z_i, \mathbf{i} \in \mathbb{R}^N\}$ that can be applied to spatial prediction.

Acknowledgement

The author is grateful to two anonymous referees and the Associate Editor for their careful reading of the first version of this paper, and for their insightful and constructive comments.

5. Appendix

5.1. Some preliminary results for the proofs

Lemma 1. (Carbon et al. (2007)) Let the sets S_1, S_2, \dots, S_k containing each m sites and such that, for all $i \neq j$, and for $1 \leq i, j \leq k$, $\text{dist}(S_i, S_j) \geq \delta_0$. Let W_1, W_2, \dots, W_k a sequence of random variables with real values and measurable respectively with respect to $\mathcal{B}(S_1), \dots, \mathcal{B}(S_k)$. Let be W_i with values in $[a, b]$. There exists a sequence of independent random variables $W_1^*, W_2^*, \dots, W_k^*$ such that W_i^* has the same distribution as W_i and satisfies:

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

Lemma 2. (Carbon et al. (1997)) 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,

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

For bounded random variables with probability 1, we have:

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

In the following, we will often use the notation $K_{1i} = K_1 \left(\frac{d(x_{i_0}, X_i)}{b_n} \right)$ and $K_{2i} = K_{2, \rho_n}(\|\mathbf{i}_0 - \mathbf{i}\|)$. Let $W_{ni} = \frac{K_{1i}K_{2i}}{\sum_{k \in \mathcal{J}_n} K_{1k}K_{2k}}$ with the convention $0/0 = 0$, then $\sum_{i \in \mathcal{J}_n} W_{ni} = 0$ or 1 . Thus, we have

$$r_n(x_{i_0}) = \begin{cases} \sum_{i \in \mathcal{J}_n} W_{ni} Y_i & \text{if } \sum_{i \in \mathcal{J}_n} W_{ni} = 1; \\ \frac{1}{n} \sum_{i \in \mathcal{J}_n} Y_i & \text{otherwise.} \end{cases}$$

Lemma 3. Under hypothesis **A2**, we have

$$\mathbb{E}^{1/2} \left[\sum_{\mathbf{i} \in \mathcal{I}_n} W_{\mathbf{ni}} \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}}) - r(x_{\mathbf{i}_0}) \right]^2 = O(b_n).$$

Lemma 4. Under hypotheses **A1**, **A3**, **A4** or **A5** and **A6**, we have

$$\mathbb{E}^{1/2} \left[\sum_{\mathbf{i} \in \mathcal{I}_n} W_{\mathbf{ni}} (Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})) \right]^2 = O \left(\frac{1}{\widehat{\mathbf{n}} \rho_n^N \varphi_{x_{\mathbf{i}_0}}(b_n)} \right)^{1/2}.$$

Sketch of the proof for Lemma 4: The expression $W_{\mathbf{ni}}(Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}}))$ is decomposed in the sum of two terms, for which it is sufficient to show that:

1. $\|e_n(x_{\mathbf{i}_0})\|_2 = \left\| \frac{1}{a_{n,\mathbf{i}_0} \mathbb{E}[K_{1\mathbf{i}}]} \sum_{\mathbf{i} \in \mathcal{I}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} [Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})] \right\|_2 = O(\widehat{\mathbf{n}} \rho_n^N \varphi_{x_{\mathbf{i}_0}}(b_n))^{-1/2}$. To obtain this result, we let $\xi_{\mathbf{i}} = K_{1\mathbf{i}} K_{2\mathbf{i}} [Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})]$ and study $\mathbb{E}(\sum_{\mathbf{i} \in \mathcal{I}_n} \xi_{\mathbf{i}}^2) = \sum_{\mathbf{i} \in \mathcal{I}_n} \mathbb{E}[\xi_{\mathbf{i}}^2] + \sum_{\mathbf{i}, \mathbf{k} \in S} \mathbb{E}[\xi_{\mathbf{i}} \xi_{\mathbf{k}}] + \sum_{\mathbf{i}, \mathbf{k} \in S^c} \mathbb{E}[\xi_{\mathbf{i}} \xi_{\mathbf{k}}]$ with $S = \{\mathbf{i}, \mathbf{k} \in \mathcal{I}_n, \|\mathbf{i} - \mathbf{k}\| \leq D_n\}$ and denote by S^c the complementary of S . Moreover D_n is a sequence of real numbers tending to ∞ as $\widehat{\mathbf{n}} \rightarrow \infty$.
2. $\mathbb{P} \left[\sum_{\mathbf{i} \in \mathcal{I}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} \leq \frac{a_{n,\mathbf{i}_0} u}{2} \right] = O(\widehat{\mathbf{n}} \rho_n^N \varphi_{x_{\mathbf{i}_0}}(b_n))^{-1/2}$ using the well-known spatial block decomposition (Tran (1990)), Markov and Bernstein inequalities and Lemmas 1 and 6, with $u = \mathbb{E}[K_{1\mathbf{i}}]$.

Lemma 5. Under the hypotheses of Lemma 4, we have

$$\mathbb{E}^{1/2} \left[\frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_n} Y_{\mathbf{i}} - r(x_{\mathbf{i}_0}) \right]^2 = O \left(\frac{1}{\widehat{\mathbf{n}} \rho_n^N \varphi_{x_{\mathbf{i}_0}}(b_n)} \right)^{1/2}.$$

Lemma 6. Under the hypotheses **A1** and **A3**, we have

$$I_n(x_{\mathbf{i}_0}) + R_n(x_{\mathbf{i}_0}) = O \left(\frac{1}{\widehat{\mathbf{n}} \rho_n^N \varphi_{x_{\mathbf{i}_0}}(b_n)} \right).$$

where $I_n(x_{\mathbf{i}_0}) = \sum_{\mathbf{i} \in \mathcal{I}_n} \mathbb{E}[(\Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}))^2]$ and $R_n(x_{\mathbf{i}_0}) = \sum_{\mathbf{i}, \mathbf{k} \in \mathcal{I}_n} \sum_{\mathbf{i} \neq \mathbf{k}} |\mathbb{E}[\Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}) \Lambda_{\mathbf{k}}(x_{\mathbf{i}_0})]|$ with $\Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}) = \frac{1}{a_{n,\mathbf{i}_0} \mathbb{E}[K_{1\mathbf{i}}]} [K_{1\mathbf{i}} K_{2\mathbf{i}} - \mathbb{E}(K_{1\mathbf{i}} K_{2\mathbf{i}})]$.

5.2. Proofs

5.2.1. Proof of Theorem 1

We study the expression $\|r_n(x_{\mathbf{i}_0}) - r(x_{\mathbf{i}_0})\|_2 = (\mathbb{E}[|r_n(x_{\mathbf{i}_0}) - r(x_{\mathbf{i}_0})|^2])^{1/2}$.

$$\begin{aligned} r_n(x_{\mathbf{i}_0}) - r(x_{\mathbf{i}_0}) &= \left(\sum_{\mathbf{i} \in \mathcal{I}_n} W_{\mathbf{ni}} \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}}) - r(x_{\mathbf{i}_0}) \right) \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{I}_n} W_{\mathbf{ni}}=1]} + \left(\sum_{\mathbf{i} \in \mathcal{I}_n} W_{\mathbf{ni}} (Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})) \right) \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{I}_n} W_{\mathbf{ni}}=1]} \\ &\quad + \left(\frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_n} Y_{\mathbf{i}} - r(x_{\mathbf{i}_0}) \right) \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{I}_n} W_{\mathbf{ni}}=0]} \end{aligned}$$

$$\|r_n(x_{\mathbf{i}_0}) - r(x_{\mathbf{i}_0})\|_2 \leq \mathbb{E}^{1/2}[\mathbf{A}]^2 + \mathbb{E}^{1/2}[\mathbf{B}]^2 + \mathbb{E}^{1/2}[\mathbf{C}]^2$$

applying Minkowski's inequality. The terms on the right-hand-side of the previous equation are dealt in the Lemmas 3, 4 and 5 respectively. \square

5.2.2. Proof of Lemma 3

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

by assumptions **A1** and **A2** (Lipschitz condition). \square

5.2.3. Proof of Lemma 4

Let $G = (\sum_{\mathbf{i} \in \mathcal{J}_n} W_{\mathbf{n}\mathbf{i}} [Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})]) \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{J}_n} W_{\mathbf{n}\mathbf{i}}=1]} = \left(\frac{e_n(x_{\mathbf{i}_0})}{f_n(x_{\mathbf{i}_0})} \right) \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{J}_n} W_{\mathbf{n}\mathbf{i}}=1]}$ with

$$e_n(x_{\mathbf{i}_0}) = \frac{1}{a_{\mathbf{n},\mathbf{i}_0} \mathbb{E}[K_{1\mathbf{i}}]} \sum_{\mathbf{i} \in \mathcal{J}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} [Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})] \quad \text{and} \quad f_n(x_{\mathbf{i}_0}) = \frac{1}{a_{\mathbf{n},\mathbf{i}_0} \mathbb{E}[K_{1\mathbf{i}}]} \sum_{\mathbf{i} \in \mathcal{J}_n} K_{1\mathbf{i}} K_{2\mathbf{i}}.$$

Note that $\forall \mathbf{i}: 0 \leq |Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})| \leq 2M$, then, $|G| \leq \sum_{\mathbf{i} \in \mathcal{J}_n} W_{\mathbf{n}\mathbf{i}} 2M \leq 2M$.

$$|G| = |G| \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{J}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} > c]} + |G| \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{J}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} \leq c]} \leq \frac{|e_n(x_{\mathbf{i}_0})|}{f_n(x_{\mathbf{i}_0})} \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{J}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} > c]} + 2M \times \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{J}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} \leq c]}$$

where c is a given constant. Let us take $c = \frac{a_{\mathbf{n},\mathbf{i}_0} u}{2}$ with $u = \mathbb{E}[K_{1\mathbf{i}}] \leq C \times \varphi_{x_{\mathbf{i}_0}}(b_n)$ since by assumption **A1**, we have $C_1 \varphi_{x_{\mathbf{i}_0}}(b_n) \leq \mathbb{E}[K_{1\mathbf{i}}] \leq C_2 \varphi_{x_{\mathbf{i}_0}}(b_n)$. If $\sum_{\mathbf{i} \in \mathcal{J}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} > c = \frac{a_{\mathbf{n},\mathbf{i}_0} u}{2}$ then $f_n(x_{\mathbf{i}_0}) > \frac{a_{\mathbf{n},\mathbf{i}_0} u}{2 a_{\mathbf{n},\mathbf{i}_0} \mathbb{E}[K_{1\mathbf{i}}]} > \frac{1}{2}$. Consequently, $\|G\|_2 \leq 2 \|e_n(x_{\mathbf{i}_0})\|_2 + 2M \left(\mathbb{P} \left[\sum_{\mathbf{i} \in \mathcal{J}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} \leq \frac{a_{\mathbf{n},\mathbf{i}_0} u}{2} \right] \right)^{1/2}$.

$$\diamond \quad \|e_n(x_{\mathbf{i}_0})\|_2 = \frac{1}{a_{\mathbf{n},\mathbf{i}_0} \mathbb{E}[K_{1\mathbf{i}}]} \left[\mathbb{E} \left(\sum_{\mathbf{i} \in \mathcal{J}_n} \xi_{\mathbf{i}} \right)^2 \right]^{1/2} \quad \text{where } \xi_{\mathbf{i}} = K_{1\mathbf{i}} K_{2\mathbf{i}} \theta_{\mathbf{i}} \text{ with } \theta_{\mathbf{i}} = Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})$$

Let D_n be a sequence of real numbers tending to ∞ as $\hat{\mathbf{n}} \rightarrow \infty$. Set $S = \{\mathbf{i}, \mathbf{k} \in \mathcal{J}_n, \|\mathbf{i} - \mathbf{k}\| \leq D_n\}$ and denote by S^c the complementary of S . Let $V_j = \left\{ \mathbf{i}, \left\| \frac{\mathbf{i}-\mathbf{j}}{\hat{\mathbf{n}}} \right\| \leq \rho_n \right\}$ with $\text{Card}(V_j) = k_n = C_N \hat{\mathbf{n}} \rho_n^N + O((\hat{\mathbf{n}} \rho_n^N)^\beta)$, see the definition of $r_n(x_{\mathbf{i}_0})$. First, we are interested in

$$\mathbb{E} \left(\sum_{\mathbf{i} \in \mathcal{J}_n} \xi_{\mathbf{i}} \right)^2 = \sum_{\mathbf{i} \in \mathcal{J}_n} \mathbb{E} [\xi_{\mathbf{i}}^2] + \sum_{\mathbf{i}, \mathbf{k} \in S} \mathbb{E} [\xi_{\mathbf{i}} \xi_{\mathbf{k}}] + \sum_{\mathbf{i}, \mathbf{k} \in S^c} \mathbb{E} [\xi_{\mathbf{i}} \xi_{\mathbf{k}}]$$

$$\begin{aligned} \bullet \sum_{i \in \mathcal{I}_n} \mathbb{E} [\xi_i^2] &\leq \sum_{i \in \mathcal{I}_n} \mathbb{E} [(K_{1i}K_{2i}|\theta_i|)^2] \leq 4M^2 \sum_{i \in \mathcal{I}_n} K_{2i}^2 \mathbb{E} [(K_{1i})^2] \leq 4M^2 \times k_n \mathbb{E} [(K_{1i})^2] \\ &\leq 4M^2 \times C_2^2 \times k_n \varphi_{x_{i_0}}(b_n) = O(\widehat{n} \rho_n^N \varphi_{x_{i_0}}(b_n)), \end{aligned}$$

since $C_1^2 \varphi_{x_{i_0}}(b_n) \leq \mathbb{E} [K_{1i}^2] \leq C_2^2 \varphi_{x_{i_0}}(b_n)$.

$$\bullet \sum_{i, k \in S} \mathbb{E} [\xi_i \xi_k] \leq 4M^2 \sum_{i, k \in S} K_{2i} K_{2k} \mathbb{E} [K_{1i} K_{1k}] \leq 4M^2 \sum_{i, k \in S} K_{2i} K_{2k} \mathbb{P} [(X_i, X_k) \in B(x_{i_0}, b_n) \times B(x_{i_0}, b_n)]$$

By assumption **A3**, we have

$$\begin{aligned} \sum_{i, k \in S} \mathbb{E} [\xi_i \xi_k] &\leq 4M^2 C_4 \sum_{i, k \in S} \mathbf{1}_{[0,1]} \left(\rho_n^{-1} \left\| \frac{\mathbf{i}_0 - \mathbf{i}}{\mathbf{n}} \right\| \right) \mathbf{1}_{[0,1]} \left(\rho_n^{-1} \left\| \frac{\mathbf{i}_0 - \mathbf{k}}{\mathbf{n}} \right\| \right) (\varphi_{x_{i_0}}(b_n))^{1+\varepsilon_1} \\ &\leq 4M^2 C_4 \sum_{i, k \in V_{i_0}} \mathbf{1}_{[0,1]} \left(\frac{\|\mathbf{i} - \mathbf{k}\|}{D_n} \right) (\varphi_{x_{i_0}}(b_n))^{1+\varepsilon_1} \\ &\leq 4M^2 C_4 \sum_{i \in V_{i_0}} \sum_{\mathbf{u} \in V_{i_0}} \mathbf{1}_{\{\mathbf{u}; \|\mathbf{u}\| \leq D_n\}} (\varphi_{x_{i_0}}(b_n))^{1+\varepsilon_1} \leq 4M^2 C_4 k_n D_n^N (\varphi_{x_{i_0}}(b_n))^{1+\varepsilon_1} \end{aligned}$$

$$\bullet \sum_{i, k \in S^c} \mathbb{E} [\xi_i \xi_k] \leq \sum_{i, k \in S^c} |\mathbb{E} (K_{1i} K_{2i} K_{1k} K_{2k} Y_i Y_k - K_{1i} K_{2i} \mathbb{E} [Y_i | X_i] K_{1k} K_{2k} \mathbb{E} [Y_k | X_k])|$$

and, since the function K_1 and K_2 are bounded, we get by applying Lemma 2

$$\begin{aligned} \sum_{i, k \in S^c} \mathbb{E} [\xi_i \xi_k] &\leq C \sum_{i, k \in S^c} \{\psi(1, 1) \gamma(\|\mathbf{i} - \mathbf{k}\|)\} \leq C \sum_{i, k \in S^c \cap V_{i_0}} \gamma(\|\mathbf{i} - \mathbf{k}\|) \leq C 2^N \sum_{\mathbf{k} \in V_{i_0}} \sum_{\substack{\mathbf{k} - \mathbf{u} \in V_{i_0}, \\ \|\mathbf{u}\| > D_n}} \gamma(\|\mathbf{u}\|) \\ &\leq C k_n \sum_{\|\mathbf{i}\| > D_n} \gamma(\|\mathbf{i}\|). \end{aligned}$$

Since $\sum_{\|\mathbf{i}\| > D_n} \gamma(\|\mathbf{i}\|) \leq C \sum_{\|\mathbf{i}\| > D_n} \|\mathbf{i}\|^{-\theta} \leq C \sum_{\|\mathbf{i}\| > D_n} \|\mathbf{i}\|^{-\theta} \|\mathbf{i}\|^{-N} \|\mathbf{i}\|^N$ and $\|\mathbf{i}\| > D_n$, $\|\mathbf{i}\|^{-N} \leq (D_n)^{-N}$, we have

$$C \sum_{\|\mathbf{i}\| > D_n} \|\mathbf{i}\|^{-\theta} \|\mathbf{i}\|^{-N-\varepsilon_1} \|\mathbf{i}\|^{N+\varepsilon_1} \leq C (D_n)^{-N-\varepsilon_1} \sum_{\|\mathbf{i}\| > D_n} \|\mathbf{i}\|^{-\theta} \|\mathbf{i}\|^{N+\varepsilon_1} \leq C D_n^{-N-\varepsilon_1} \sum_{\|\mathbf{i}\| > D_n} \|\mathbf{i}\|^{N+\varepsilon_1-\theta}$$

and then $\sum_{i, k \in S^c} \mathbb{E} [\xi_i \xi_k] \leq C k_n D_n^{-N-\varepsilon_1} \sum_{\|\mathbf{i}\| > D_n} \|\mathbf{i}\|^{N+\varepsilon_1-\theta}$. The fact that $\theta > 4N > N + 1$ leads to choose $D_n = (\varphi_{x_{i_0}}(b_n))^{\frac{-\varepsilon_1}{N} + a}$ with $a > 0$ and such that $Na \leq \varepsilon_1 - \frac{N}{N+\varepsilon_1}$. In fact, these conditions lead to

$$D_n^{-(N+\varepsilon_1)} = \varphi_{x_{i_0}}(b_n) (\varphi_{x_{i_0}}(b_n))^{\frac{-(N+\varepsilon_1)(Na-\varepsilon_1)-N}{N}} = O\left(\varphi_{x_{i_0}}(b_n)\right)$$

since $\frac{-(N+\varepsilon_1)(Na-\varepsilon_1)-N}{N} > 0$. Moreover, this choice of D_n implies that

$$\sum_{i, k \in S} \mathbb{E} [\xi_i \xi_k] \leq 4M^2 C_4 k_n D_n^N (\varphi_{x_{i_0}}(b_n))^{1+\varepsilon_1} \leq 4M^2 C_4 k_n (\varphi_{x_{i_0}}(b_n))^{1+Na} = O(\widehat{n} \rho_n^N \varphi_{x_{i_0}}(b_n))$$

Consequently, $\left[\mathbb{E} \left(\sum_{\mathbf{i} \in \mathcal{S}_n} \xi_{\mathbf{i}} \right)^2 \right]^{1/2} = O(\widehat{\mathbf{n}} \rho_{\mathbf{n}}^N \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}}))^{1/2}$ and $\|e_{\mathbf{n}}(x_{\mathbf{i}_0})\|_2 = O\left(\widehat{\mathbf{n}} \rho_{\mathbf{n}}^N \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}})\right)^{-1/2}$.
 Second, we deal with

$$\begin{aligned} \diamond \quad P &= \mathbb{P} \left[\sum_{\mathbf{i} \in \mathcal{S}_n} K_{1\mathbf{i}} K_{2\mathbf{i}} \leq \frac{a_{\mathbf{n}, \mathbf{i}_0} u}{2} \right] = \mathbb{P} \left[\sum_{\mathbf{i} \in \mathcal{S}_n} (K_{1\mathbf{i}} K_{2\mathbf{i}} - \mathbb{E}(K_{1\mathbf{i}} K_{2\mathbf{i}})) \leq \frac{-a_{\mathbf{n}, \mathbf{i}_0} u}{2} \right] \\ &\leq \mathbb{P} \left[\frac{1}{a_{\mathbf{n}, \mathbf{i}_0} u} \left| \sum_{\mathbf{i} \in \mathcal{S}_n} (K_{1\mathbf{i}} K_{2\mathbf{i}} - \mathbb{E}(K_{1\mathbf{i}} K_{2\mathbf{i}})) \right| \geq \frac{1}{2} \right] \leq \mathbb{P} [|S_{\mathbf{n}}(x_{\mathbf{i}_0})| \geq \varepsilon] \end{aligned}$$

with $S_{\mathbf{n}}(x_{\mathbf{i}_0}) = \sum_{\mathbf{i} \in \mathcal{S}_n} \Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}) = \sum_{\mathbf{i} \in \mathcal{S}_n} \frac{1}{a_{\mathbf{n}, \mathbf{i}_0} u} (K_{1\mathbf{i}} K_{2\mathbf{i}} - \mathbb{E}(K_{1\mathbf{i}} K_{2\mathbf{i}}))$. 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 $\Lambda_{\mathbf{i}}(x_{\mathbf{i}_0})$ can be grouped into $2^N q_1 \dots q_N$ cubic blocks of side b . Let, and so on. Noticing that

$$\begin{aligned} U(1, \mathbf{n}, x_{\mathbf{i}_0}, \mathbf{j}) &= \sum_{\substack{i_k=2j_k b+1, \\ k=1, \dots, N.}}^{(2j_k+1)b} \Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}), \\ U(2, \mathbf{n}, x_{\mathbf{i}_0}, \mathbf{j}) &= \sum_{\substack{i_k=2j_k b+1, \\ k=1, \dots, N-1.}}^{(2j_k+1)b} \sum_{i_N=(2j_N+1)b+1}^{2(j_N+1)b} \Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}), \\ U(3, \mathbf{n}, x_{\mathbf{i}_0}, \mathbf{j}) &= \sum_{\substack{i_k=2j_k b+1, \\ k=1, \dots, N-2.}}^{(2j_k+1)b} \sum_{i_{N-1}=(2j_{N-1}+1)b+1}^{2(j_{N-1}+1)b} \sum_{i_N=2j_N b+1}^{(2j_N+1)b} \Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}), \\ U(4, \mathbf{n}, x_{\mathbf{i}_0}, \mathbf{j}) &= \sum_{\substack{i_k=2j_k b+1, \\ k=1, \dots, N-2.}}^{(2j_k+1)b} \sum_{i_{N-1}=(2j_{N-1}+1)b+1}^{2(j_{N-1}+1)b} \sum_{i_N=(2j_N+1)b+1}^{(2j_N+1)b} \Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}) \end{aligned}$$

and so on. Noticing that

$$\begin{aligned} U(2^{N-1}, \mathbf{n}, x_{\mathbf{i}_0}, \mathbf{j}) &= \sum_{\substack{i_k=(2j_k+1)b+1, \\ k=1, \dots, N-1.}}^{2(j_k+1)b} \sum_{i_N=2j_N b+1}^{(2j_N+1)b} \Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}) \\ U(2^N, \mathbf{n}, x_{\mathbf{i}_0}, \mathbf{j}) &= \sum_{\substack{i_k=(2j_k+1)b+1, \\ k=1, \dots, N.}}^{2(j_k+1)b} \Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}) \end{aligned}$$

for each integer $1 \leq l \leq 2^N$, we define $T(\mathbf{n}, x_{\mathbf{i}_0}, l) = \sum_{\substack{j_k=0 \\ k=1, \dots, N.}}^{q_k-1} U(l, \mathbf{n}, x_{\mathbf{i}_0}, \mathbf{j})$. We obtain $S_{\mathbf{n}}(x_{\mathbf{i}_0}) = \sum_{l=1}^{2^N} T(\mathbf{n}, x_{\mathbf{i}_0}, l)$. For $\varepsilon > 0$, $P \leq \mathbb{P} \left(\left| \sum_{l=1}^{2^N} T(\mathbf{n}, x_{\mathbf{i}_0}, l) \right| > \varepsilon \right) \leq 2^N \mathbb{P} \left(|T(\mathbf{n}, x_{\mathbf{i}_0}, 1)| > \frac{\varepsilon}{2^N} \right)$. We enumerate in arbitrary manner the $\widehat{q} = q_1 \times \dots \times q_N$ terms $U(1, \mathbf{n}, x_{\mathbf{i}_0}, \mathbf{j})$ of the sum $T(\mathbf{n}, x_{\mathbf{i}_0}, 1)$, and refer to them as $W_1, \dots, W_{\widehat{q}}$. Note that $U(1, \mathbf{n}, x_{\mathbf{i}_0}, \mathbf{j})$ is a measurable σ -algebra generated by $X_{\mathbf{i}}$, with \mathbf{i} such that $2j_k b + 1 \leq i_k \leq (2j_k + 1)b$, $k = 1, \dots, N$. 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_2 and K_1 are bounded, we can write $|W_l| \leq C \frac{b^N}{a_{\mathbf{n}, \mathbf{i}_0} u}$ with $C = \|K_1\|_\infty \|K_2\|_\infty$ (where $\|\cdot\|_\infty$ is the sup norm). Lemma 1 insures the existence of some random variables $W_1^*, W_2^*, \dots, W_{\hat{q}}^*$ such that

$$\sum_{l=1}^{\hat{q}} \mathbb{E}|W_l - W_l^*| \leq 2\hat{q}C \frac{b^N}{a_{\mathbf{n}, \mathbf{i}_0} u} \psi((\hat{q}-1)b^N, b^N) \gamma(b) \leq 2C \frac{\hat{\mathbf{n}}}{2^N b^N} \frac{b^N}{a_{\mathbf{n}, \mathbf{i}_0} u} \psi(\hat{\mathbf{n}}, b^N) \gamma(b).$$

Markov inequality allows us to write

$$\mathbb{P}\left(\sum_{l=1}^{\hat{q}} |W_l - W_l^*| > \frac{\varepsilon}{2^{N+1}}\right) \leq 2C \frac{\hat{\mathbf{n}}}{2^N b^N} \frac{b^N}{a_{\mathbf{n}, \mathbf{i}_0} u} \psi(\hat{\mathbf{n}}, b^N) \gamma(b) \frac{2^{N+1}}{\varepsilon},$$

and by Bernstein inequality, we have

$$\mathbb{P}\left(\sum_{l=1}^{\hat{q}} |W_l^*| > \frac{\varepsilon}{2^{N+1}}\right) \leq 2 \exp\left\{\frac{-\varepsilon^2 / (2^{N+1})^2}{4 \sum_{l=1}^{\hat{q}} \mathbb{E}(W_l^{*2}) + \frac{2\varepsilon}{2^{N+1}} \frac{b^N}{a_{\mathbf{n}, \mathbf{i}_0} u} C}\right\}$$

which leads to

$$P \leq 2^{N+1} \exp\left\{\frac{-\varepsilon^2 / (2^{N+1})^2}{4 \sum_{l=1}^{\hat{q}} \mathbb{E}(W_l^{*2}) + 2^{-N} C \varepsilon \frac{b^N}{a_{\mathbf{n}, \mathbf{i}_0} u}}\right\} + 2^{N+1} C \frac{\hat{\mathbf{n}}}{2^N b^N} \frac{b^N}{a_{\mathbf{n}, \mathbf{i}_0} u} \psi(\hat{\mathbf{n}}, b^N) \gamma(b) \frac{2^{N+1}}{\varepsilon}$$

Let $\delta > 0$, $\varepsilon = \varepsilon_{\mathbf{n}} = \delta \left(\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}}) \rho_{\mathbf{n}}^N}\right)^{1/2}$ and $b = \left(\frac{\hat{\mathbf{n}} \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}}) \rho_{\mathbf{n}}^N}{\log \hat{\mathbf{n}}}\right)^{\frac{1}{2N}}$. Since the variables W_l and W_l^* have the same distributions, we have $\sum_{l=1}^{\hat{q}} \mathbb{E}W_l^{*2} = \sum_{l=1}^{\hat{q}} \text{Var}(W_l^*) = \sum_{l=1}^{\hat{q}} \text{Var}(W_l) \leq I_{\mathbf{n}}(x_{\mathbf{i}_0}) + R_{\mathbf{n}}(x_{\mathbf{i}_0})$, and according to Lemma 6, we have $\sum_{l=1}^{\hat{q}} \mathbb{E}W_l^{*2} \leq O\left([\hat{\mathbf{n}} \rho_{\mathbf{n}}^N \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}})]^{-1}\right)$. Then,

$$P \leq 2^{N+1} \exp\left\{\frac{-\varepsilon^2}{2^{2N+2} \left(4 \frac{C}{\hat{\mathbf{n}} \rho_{\mathbf{n}}^N \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}})} + C 2^{-N} \varepsilon \frac{b^N}{a_{\mathbf{n}, \mathbf{i}_0} u}\right)}\right\} + 2^{N+2} C \frac{\hat{\mathbf{n}}}{a_{\mathbf{n}, \mathbf{i}_0} u} \psi(\hat{\mathbf{n}}, b^N) b^{-\theta} \varepsilon^{-1}$$

Since $C_1 k_{\mathbf{n}} \leq a_{\mathbf{n}, \mathbf{i}_0} \leq C_2 k_{\mathbf{n}}$ and $k_{\mathbf{n}} = C_N d_{\mathbf{n}}^N + O(d_{\mathbf{n}}^\beta)$, $\beta < N$, we have

$$\begin{aligned} P &\leq 2^{N+1} \exp\left\{\frac{-\delta^2 \frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}}) \rho_{\mathbf{n}}^N}}{\frac{2^{2N+4} C}{\hat{\mathbf{n}} \rho_{\mathbf{n}}^N \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}})} + \frac{C 2^{N+2} \delta}{\hat{\mathbf{n}} \rho_{\mathbf{n}}^N \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}})}}\right\} + 2^{N+2} C \frac{\hat{\mathbf{n}}}{a_{\mathbf{n}, \mathbf{i}_0} u} \psi(\hat{\mathbf{n}}, b^N) b^{-\theta} \delta^{-1} \left(\frac{\hat{\mathbf{n}} \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}}) \rho_{\mathbf{n}}^N}{\log \hat{\mathbf{n}}}\right)^{1/2} \\ &\leq 2^{N+1} \exp\{\log \hat{\mathbf{n}}^{-a}\} + 2^{N+2} C \delta^{-1} \frac{\hat{\mathbf{n}}}{a_{\mathbf{n}, \mathbf{i}_0} u} \psi(\hat{\mathbf{n}}, b^N) \left(\frac{\hat{\mathbf{n}} \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}}) \rho_{\mathbf{n}}^N}{\log \hat{\mathbf{n}}}\right)^{\frac{N-\theta}{2N}} \\ &\leq C \hat{\mathbf{n}}^{-a} + 2^{N+2} C \delta^{-1} \frac{\hat{\mathbf{n}}}{a_{\mathbf{n}, \mathbf{i}_0} \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}})} \psi(\hat{\mathbf{n}}, b^N) \left(\frac{\hat{\mathbf{n}} \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}}) \rho_{\mathbf{n}}^N}{\log \hat{\mathbf{n}}}\right)^{\frac{N-\theta}{2N}} \end{aligned}$$

with $a = \frac{\delta^2}{2^{2N+4}C + C_N 2^{N+2}\delta} > 0$. Note that $\widehat{\mathbf{n}}^{1-a} \varphi_{x_{i_0}}(b_{\mathbf{n}}) \rho_{\mathbf{n}}^N$ tends to 0 for $a > 1$ and then $C\widehat{\mathbf{n}}^{-a} = o\left([\widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})\rho_{\mathbf{n}}^N]^{-1}\right)$. 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. From assumptions on $\psi(n, m)$, two cases arise.

First case: $\psi(n, m) \leq C \min(n, m), \forall n, m \in \mathbb{N}$

$$\begin{aligned} \widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})\rho_{\mathbf{n}}^N C 2^{N+2}\delta^{-1} \frac{\widehat{\mathbf{n}}}{a_{\mathbf{n}, i_0} \varphi_{x_{i_0}}(b_{\mathbf{n}})} b^N \left(\frac{\widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})\rho_{\mathbf{n}}^N}{\log \widehat{\mathbf{n}}} \right)^{\frac{N-\theta}{2N}} &\leq \widehat{\mathbf{n}}\rho_{\mathbf{n}}^N C 2^{N+2}\delta^{-1} \frac{\widehat{\mathbf{n}}}{a_{\mathbf{n}, i_0}} \left(\frac{\widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})\rho_{\mathbf{n}}^N}{\log \widehat{\mathbf{n}}} \right)^{\frac{2N-\theta}{2N}} \\ &\leq \widehat{\mathbf{n}}\rho_{\mathbf{n}}^N C_N 2^{N+2}\delta^{-1} \frac{1}{\rho_{\mathbf{n}}^N} \left(\frac{\widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})\rho_{\mathbf{n}}^N}{\log \widehat{\mathbf{n}}} \right)^{\frac{2N-\theta}{2N}} \\ &\leq C_N \left[\widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})^{\frac{2N-\theta}{4N-\theta}} \rho_{\mathbf{n}}^{\frac{N(2N-\theta)}{4N-\theta}} \log \widehat{\mathbf{n}}^{\frac{\theta-2N}{4N-\theta}} \right]^{\frac{4N-\theta}{2N}} \end{aligned}$$

which tends to 0 according to hypothesis **A4**.

Second case: $\psi(n, m) \leq C(n+m+1)^{\tilde{\beta}}, \forall n, m \in \mathbb{N}$. Note that $\psi(\widehat{\mathbf{n}}, b^N) \leq C(\widehat{\mathbf{n}} + b^N + 1)^{\tilde{\beta}} \leq C\widehat{\mathbf{n}}^{\tilde{\beta}}$.

$$\begin{aligned} \widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})\rho_{\mathbf{n}}^N C 2^{N+2}\delta^{-1} \frac{\widehat{\mathbf{n}}}{a_{\mathbf{n}, i_0} \varphi_{x_{i_0}}(b_{\mathbf{n}})} \widehat{\mathbf{n}}^{\tilde{\beta}} \left(\frac{\widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})\rho_{\mathbf{n}}^N}{\log \widehat{\mathbf{n}}} \right)^{\frac{N-\theta}{2N}} &\leq \widehat{\mathbf{n}}\rho_{\mathbf{n}}^N C_N 2^{N+2}\delta^{-1} \frac{1}{\rho_{\mathbf{n}}^N} \widehat{\mathbf{n}}^{\tilde{\beta}} \left(\frac{\widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})\rho_{\mathbf{n}}^N}{\log \widehat{\mathbf{n}}} \right)^{\frac{N-\theta}{2N}} \\ &\leq C_N \left[\widehat{\mathbf{n}}\varphi_{x_{i_0}}(b_{\mathbf{n}})^{\frac{N-\theta}{N(3+2\tilde{\beta})-\theta}} \rho_{\mathbf{n}}^{\frac{N(N-\theta)}{N(3+2\tilde{\beta})-\theta}} \log \widehat{\mathbf{n}}^{\frac{\theta-N}{N(3+2\tilde{\beta})-\theta}} \right]^{\frac{N(3+2\tilde{\beta})-\theta}{2N}} \end{aligned}$$

which tends to 0 according to hypothesis **A5**. Therefore, in the two cases, we have

$$\mathbb{P} \left[\sum_{\mathbf{i} \in \mathcal{J}_{\mathbf{n}}} K_{1i} K_{2i} \leq \frac{a_{\mathbf{n}, i_0} u}{2} \right] = o \left(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_{x_{i_0}}(b_{\mathbf{n}}) \right)^{-1}$$

Consequently, $\|G\|_2 = o \left(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_{x_{i_0}}(b_{\mathbf{n}}) \right)^{-1/2}$ □

5.2.4. Proof of Lemma 5

Since Y_i and $r(\cdot)$ are bounded, we have

$$\begin{aligned} \mathbb{E}^{1/2}[\mathbf{C}] &\leq \mathbb{E}^{1/2} \left[\frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{J}_{\mathbf{n}}} Y_i - r(x_{i_0}) \mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{J}_{\mathbf{n}}} w_{ni} = 0]} \right] \leq 2M \mathbb{E}^{1/2} \left[\mathbf{1}_{[\sum_{\mathbf{i} \in \mathcal{J}_{\mathbf{n}}} w_{ni} = 0]} \right] \\ &\leq 2M \left(\mathbb{P} \left[\sum_{\mathbf{i} \in \mathcal{J}_{\mathbf{n}}} K_{1i} K_{2i} = 0 \right] \right)^{1/2} \leq 2M \left(\mathbb{P} \left[\sum_{\mathbf{i} \in \mathcal{J}_{\mathbf{n}}} K_{1i} K_{2i} \leq \frac{a_{\mathbf{n}, i_0} u}{2} \right] \right)^{1/2} = o \left(\frac{1}{\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N \varphi_{x_{i_0}}(b_{\mathbf{n}})} \right)^{1/2}, \end{aligned}$$

using Lemma 4. □

5.2.5. Proof of Lemma 6

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

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

Then, we have $I_{\mathbf{n}}(x_{\mathbf{i}_0}) = O \left([\widehat{\mathbf{n}} \rho_{\mathbf{n}}^N \varphi_{x_{\mathbf{i}_0}}(b_{\mathbf{n}})]^{-1} \right)$. We now treat the term $R_{\mathbf{n}}(x_{\mathbf{i}_0})$. Since the functions $K_1(\cdot)$ and $K_2(\cdot)$ are bounded, we get by applying Lemma 2

$$|\mathbb{E} [\Lambda_{\mathbf{i}}(x_{\mathbf{i}_0}) \Lambda_{\mathbf{k}}(x_{\mathbf{i}_0})]| \leq C \frac{K_{2\mathbf{i}} K_{2\mathbf{k}}}{a_{\mathbf{n}, \mathbf{i}_0}^2 u^2} \psi(1, 1) \gamma(\|\mathbf{i} - \mathbf{k}\|).$$

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

References

- Anselin, L. and Florax, R. J. G. M. (1995). *New Directions in Spatial Econometrics*. Advances in Spatial Science. Springer.
- Attouch, M. K., Gheriballah, A., and Laksaci, A. (2011). Robust nonparametric estimation for functional spatial regression. In Ferraty, F., editor, *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, pages 27–31. Physica-Verlag HD.
- Biau, G. and Cadre, B. (2004). Nonparametric spatial prediction. *Statistical Inference for Stochastic Processes*, 7(3):327–349.
- Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag New York Inc.
- Carbon, M., Francq, C., and Tran, L. T. (2007). Kernel regression estimation for random fields. *Journal of Statistical Planning and Inference*, 137(3):778–798.
- Carbon, M., Tran, L. T., and Wu, B. (1997). Kernel density estimation for random fields (density estimation for random fields). *Statistics & Probability Letters*, 36(2):115–125.
- Chamizo, L. F. and Iwaniec, H. (1995). On the sphere problem. *Revista matemática iberoamericana*, 11(2):417–430.
- Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Applied Probability and Statistics. John Wiley & Sons, Inc.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*, volume 110 of *Wiley Series in Probability and Statistics*. Wiley-Interscience, revised edition.
- Dabo-Niang, S., Hamdad, L., Ternynck, C., and Yao, A.-F. (2013). A kernel spatial density estimation allowing for the analysis of spatial clustering: application to Monsoon Asia Drought Atlas data. *In revision*.
- Dabo-Niang, S., Kaid, Z., and Laksaci, A. (2012). On spatial conditional mode estimation for a functional regressor. *Statistics & Probability Letters*, 82(7):1413–1421.
- Dabo-Niang, S., Rachdi, M., and 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., and Yao, A.-F. (2014). Spatial regression for multivariate data taking spatial characteristics into consideration and applications. *Preprint*.
- Dabo-Niang, S. and Yao, A.-F. (2007). Kernel regression estimation for continuous spatial processes. *Mathematical Methods of Statistics*, 16(4):298–317.
- Dabo-Niang, S. and 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., and 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.
- Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics*, 21(3-4):224–239.
- Delsol, L. (2008). *Régression sur variable fonctionnelle: Estimation, tests de structure et Applications*. PhD thesis, Université Paul Sabatier-Toulouse III.
- 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.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. Springer.
- García-Soidán, P. and Menezes, R. (2012). Estimation of the spatial distribution through the kernel indicator variogram. *Environmetrics*, 23(6):535–548.
- Guyon, X. (1995). *Random Fields on a Network: Modeling, Statistics, and Applications*. Probability and its Applications. Springer-Verlag.
- Hallin, M., Lu, Z., and Tran, L. T. (2004). Local linear spatial regression. *The Annals of Statistics*, 32(6):2469–2500.
- Hallin, M., Lu, Z., and Yu, K. (2009). Local linear spatial quantile regression. *Bernoulli*, 15(3):659–686.
- Journel, A. G. (1983). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, 15(3):445–468.
- Kelejian, H. H. and Prucha, I. R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140(1):131–154.
- Laksaci, A. and Maref, F. (2009). Estimation non paramétrique de quantiles conditionnels pour des variables fonctionnelles spatialement dépendantes. *Comptes Rendus Mathématique*, 347(17-18):1075–1080.
- Laksaci, A. and Mechab, B. (2010). Estimation non paramétrique de la fonction de hasard avec variable explicative fonctionnelle: cas des données spatiales. *Revue Roumaine de Mathématiques Pures et Appliquées*, 55(1):35–51.
- Menezes, R., García-Soidán, P., and Ferreira, C. (2010). Nonparametric spatial prediction under stochastic sampling design. *Journal of Nonparametric Statistics*, 22(3):363–377.
- Meyer, A. (2011). On the number of lattice points in a small sphere. In *WCC 2011 - Workshop on coding and cryptography*, pages 463–472.
- Mitchell, W. (1966). The number of lattice points in a k -dimensional hypersphere. *Mathematics of Computation*, 20(94):300–310.
- Nerini, D., Monestiez, P., and Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis*, 101(2):409–418.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, second edition.
- Ripley, B. D. (1981). *Spatial Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837.
- Tran, L. T. (1990). Kernel density estimation on random fields. *Journal of Multivariate Analysis*, 34(1):37–53.
- Tsang, K.-M. (2000). Counting lattice points in the sphere. *Bulletin of the London Mathematical Society*, 32(6):679–688.
- Wang, H. and Wang, J. (2009). Estimation of the trend function for spatio-temporal models. *Journal of Nonparametric Statistics*, 21(5):567–588.
- Wang, H., Wang, J., and Huang, B. (2012). Prediction for spatio-temporal models with autoregression in errors. *Journal of Nonparametric Statistics*, 24(1):217–244.