KERNEL INVERSE REGRESSION FOR SPATIAL RANDOM FIELDS.

JEAN-MICHEL LOUBES[♭] AND ANNE-FRANÇOISE YAO [§]

ABSTRACT. In this paper, we propose a dimension reduction model for spatially dependent variables. Namely, we investigate an extension of the *inverse regression* method under strong mixing condition. This method is based on estimation of the matrix of covariance of the expectation of the explanatory given the dependent variable, called the *inverse regression*. Then, we study, under strong mixing condition, the weak and strong consistency of this estimate, using a kernel estimate of the *inverse regression*. We provide the asymptotic behaviour of this estimate. A spatial predictor based on this dimension reduction approach is also proposed. This latter appears as an alternative to the spatial non-parametric predictor.

KEYWORDS: Kernel estimator; Spatial regression; Random fields; Strong mixing coefficient; Dimension reduction; Inverse Regression.

1. INTRODUCTION

Spatial statistics includes any techniques which study phenomenons observed on spatial subset S of $\mathbb{R}^N$, $N \geq 2$ (generally, $N = 2$ or $N = 3$). The set S can be discret, continuous or the set of realization of a point process. Such techniques have various applications in several domains such as soil science, geology, oceanography, econometrics, epidemiology, forestry and many others (see for example [27], [11] or [18] for exposition, methods and applications).

Most often, spatial data are dependents and any spatial model must be able to handle this aspect. The novelty of this dependency unlike the time-dependency, is the lack of order relation. In fact, notions of past, present and futur does not exist in space and this property gives great flexibility in spatial modelling.

In the case of *spatial regression* that interests us, there is an abundant literature on *parametric models*. We refer for example to the spatial regression models with correlated errors often used in economics (see e.g. Anselin and Florax [2], Anselin and Bera [1], Song and Lee [29]) or to the spatial Generalized Linear Model (GLM) study in Diggle et al. [14] and Zhang [36]. Recall also the spatial Poisson regression methods which have been proposed for epidemiological data (see for example Diggle [13] or Diggle et al [14]).

§ Corresponding author : University Aix-Marseille 2, Campus de Luminy, case 901, 13288 Marseille cedex 09. anne-francoise.yao@univmed.fr.
♭ University Toulouse 3, Institut de Mathématiques de Toulouse.

Unlike the parametric case, the spatial regression on nonparametric setting have been studied by a few paper: quote for example Biau and Cadre [5], Lu and Chen [25], Hallin et al. [19], Carbon et al. [9], Tran and Yakowitz [32] and Dabo-Niang and Yao [12]. Their results show that, as in the *i.i.d.* case, the spatial nonparametric estimator of the regression function is penalized by the dimension of the regressor. This is the spatial counterpart of the well-known problem called "*the curse of dimensionality*". Recall that dimension reduction methods are classically used to overcome this issue. Observing an *i.i.d.* sample $Z_i = (X_i, Y_i)$ the aim is to estimate the regression function $m(x) = \mathbf{E}(Y|X = x)$. In the dimension reduction framework, one assumes that there exist Φ an orthonormal matrix $d \times D$, with D as small as possible, and $g : \mathbb{R}^D \to \mathbb{R}$, an unknown function such that the function $m(.)$ can be written as

$$m(x) = g(\Phi\, .X). \tag{1.1}$$

Model (1.1) conveys the idea that "less information on X" , $\Phi\, .X$; provides as much information on $m(.)$ as X. The function g is the regression function of Y given the D dimensional vector $\Phi.X$. Estimating the matrix Φ and then the function g (by nonparametric methods) provides an estimator which converges faster than the initial nonparametric estimator. The operator Φ is unique under orthogonal transformation. An estimation of this latter is done through an estimation of his range $\mathrm{Im}(\Phi^T)$ (where Φ^T is the transpose of Φ) called *Effective Dimensional Reduction space* (EDR).

Various methods for dimension reduction exist in the literature for *i.i.d* observations. For example we refer to the multiple linear regression, the generalized linear model (GLM) in [8], the additive models (see e.g. Hastie and Tibshirani [21]) deal with methods based on estimation of the gradient of the regression function $m(.)$ developped in for example in [22] or [35].

In this paper, we focus on the *inverse regression* method, proposed by Li [24]: if X *is such that for all vector b in $\mathbb{R}^d$, there exists a vector B of $\mathbb{R}^D$ such that* $\mathbf{E}(b^T X|\Phi.X) = B^T(\Phi.X)$ (this latter condition is satisfied as soon as X is elliptically distributed), then, if Σ denotes the variance of X, the space $\mathrm{Im}(\Sigma^{-1}\mathbf{var}(\mathbf{E}(X|Y))$ is included into the *EDR space*. Moreover, the two spaces coincide if the matrix $\Sigma^{-1}\mathbf{var}(\mathbf{E}(X|Y))$ is of full rank. Hence, the estimation of the *EDR space* is essentially based on the estimation of the covariance matrix of the *inverse regression* $\mathbf{E}(X|Y)$ and Σ which is estimated by using a classical empirical estimator. In his initial version, Li suggested an estimator based on the regressogram estimate of $\mathbf{E}(X|Y)$ but drawbacks of the regressogram lead other authors to suggest alternatives based on the nonparametric estimation of $\mathbf{E}X|Y$, see for instance [23] or [37] which enable to recover the optimal rate of convergence in $\sqrt{n}$.

This work is motivated by the fact that to our knowledge, there is no *inverse regression* method estimation for spatially dependent data under strong mixing condition. Note however that a dimension reduction method for supervised motion segmentation based on spatial-frequential analysis called *Dynamic Sliced Inverse Regression* (DSIR) has been proposed by Wu and Lu [34]. We propose here a spatial counterpart of the estimating method of [37] which uses kernel estimation of $\mathbf{E}X|Y$. Other methods based on other spatial estimators of $\mathbf{E}X|Y$ will be the subject of futher investigation.

As any spatial model, a spatial dimension reduction model must take into account spatial dependency. In this work, we focus on an estimation on model (1.1) for spatial dependent data under strong mixing conditions. The spatial kernel regression estimation of $\mathbf{E}X|Y$ being studied in [5, 10, 9].

An important problem in spatial modelling is that of spatial prediction. The aim being reconstruction of a random field over some domain from a set of observed values. It is such a problem that interest us in the last part of this paper. More precisely, we will use the properties of the *inverse regression* method to build a *dimension reduction predictor* which corresponds to the *nonparametric predictor* of [5]. It is an interesting alternative to parametric predictor methods such as the *krigging* methods (see e.g. [33], [11]) or spatial autoregressive model (see for example [11]) since it does not requires any underlying model. It only requires the knowledge of the number of the neighbors. We will see that the property of the *inverse regression* method provides a way of estimating this number.

This paper falls into the following parts. Section 2 provides some notations and assumptions on the spatial process, as well as some preliminar results on U-statistics. The estimation method and the consistency results are presented in Section 3. Section 4 uses this estimate to forecast a spatial process. Section 5 is devoted to Conclusion. Proofs and the technical lemmas are gathered in Section 6.

2. General setting and preliminary Results

2.1. **Notations and assumptions.** Throughout all the paper, we will use the following notations.
For all $b \in \mathbb{R}^d$, $b^{(j)}$ will denote the j^{th} component of the vector b;
a point in bold $\mathbf{i} = (i_1, ..., i_N) \in \mathbf{n} \in (\mathbb{N}^*)^N$ will be referred to as a site, we will set $\mathbf{1}_N = (\underbrace{1, ..., 1}_{N\,\text{times}})$; if $\mathbf{n} = (n_1, ..., n_N)$, we will set $\widehat{\mathbf{n}} = n_1 \times ... \times n_N$ and write $\mathbf{n} \to +\infty$ if $\min_{i=1,...,N} n_i \to +\infty$ and $\frac{n_i}{n_k} < C$ for some constant $C > 0$.
The symbol $\|.\|$ will denote any norm over $\mathbb{R}^d$, $\|u\|_\infty = \sup_x |u(x)|$ for some function u

and C an arbitrary positive constant. If A is a set, let $1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$.
The notation $W_{\mathbf{n}} = \mathcal{O}_p(V_{\mathbf{n}})$ (respectively $W_{\mathbf{n}} = \mathcal{O}_{a.s}(V_{\mathbf{n}})$) means that $W_{\mathbf{n}} = V_{\mathbf{n}} S_{\mathbf{n}}$ for a sequence $S_{\mathbf{n}}$, which is bounded in *probability* (respectively *almost surely*).

We are interested in some $\mathbb{R}^d \times \mathbb{R}$-valued stationary and measurable random field $Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}})$, $\mathbf{i} \in (\mathbb{N}^*)^N$, $(N,\, d \geq 1)$ defined on a probability space $(\Omega,\, \mathcal{A}, \mathbf{P})$. Without loss of generality, we consider estimations based on observations of the process $(Z_{\mathbf{i}},\, \mathbf{i} \in \mathbb{Z}^N)$ on some rectangular set $\mathcal{I}_{\mathbf{n}} = \left\{\mathbf{i} = (i_1, ..., i_N) \in \mathbb{Z}^N,\, 1 \leq i_k \leq n_k,\, k = 1, ..., N\right\}$ for all $\mathbf{n} \in (\mathbb{N}^*)^N$.
Assume that the $Z_{\mathbf{i}}$'s have the same distribution as (X, Y) which is such that:

- the variable Y has a density f.
- $\forall j = 1, ..., d$ each component $X^{(j)}$ of X, is such that the pair $(X^{(j)}, Y)$ admits an unknown density $f_{X^{(j)},Y}$ with respect to Lebesgue measure λ over $\mathbb{R}^2$ and each $X^{(j)}$ is integrable.

2.2. **Spatial dependency.**

As mentionned above, our model as any spatial model must take into account spatial dependence between values at differents locations. Of course, we could consider that there is a global linear relationships between locations as it is generally done in spatial linear modeling, we prefer to use a nonlinear spatial dependency measure. Actually, in many circumstances the spatial dependency is not necessary linear (see [3]). It is, for example, the classical case where one deals with the spatial pattern of extreme events such as in the economic analysis of poverty, in the environmental science,... Then, it is more appropriate to use a nonlinear spatial dependency measure such as positive dependency (see [3]) or strong mixing coefficients concept (see Tran [31]). In our case, we will measure the spatial dependency of the concerned process by means of $\alpha-$*mixing* and *local dependency measure*.

2.2.1. *Mixing condition :*

The field $(Z_{\mathbf{i}})$ is said to satisfy a *mixing condition* if:

- there exists a function $\mathcal{X} : \mathbb{R}^+ \to \mathbb{R}^+$ with $\mathcal{X}(t) \downarrow 0$ as $t \to \infty$, such that whenever $S,\, S' \subset (\mathbb{N}^*)^N$,

$$\begin{aligned} \alpha(\mathcal{B}(S), \mathcal{B}(S')) &= \sup_{A \in \mathcal{B}(S),\, B \in \mathcal{B}(S')} |P(B \cap C) - P(B)P(C)| \\ &\leq \psi(\text{Card}S,\, \text{Card}S')\, \mathcal{X}(\text{dist}(S, S')) \end{aligned} \tag{2.1}$$

where $\mathcal{B}(S)$(*resp.* $\mathcal{B}(S')$) denotes the Borel $\sigma-$fields generated by $(Z_{\mathbf{i}}, \mathbf{i} \in S)$ (*resp.* $(Z_{\mathbf{i}}, \mathbf{i} \in S')$), Card S (*resp.* CardS') the cardinality of S(*resp.* S'), dist(S, S') the Euclidean distance between S and S', and $\psi : \mathbb{N}^2 \to \mathbb{R}^+$ is a symmetric positive function nondecreasing in each variable. If $\psi \equiv 1$, then $Z_{\mathbf{i}}$ is called strong mixing. It is this latter case which will be tackled in this paper and for all $v \geq 0$, we have

$$\alpha(v) = \sup_{\mathbf{i},\mathbf{j} \in \mathbb{R}^N, \|\mathbf{i}-\mathbf{j}\|=v} \alpha\left(\sigma\left(Z_{\mathbf{i}}\right), \sigma\left(Z_{\mathbf{j}}\right)\right) \leq \mathcal{X}(v).$$

- The process is said to be *Geometrically Strong Mixing* (GSM) if there exists a non-negative constant $\rho \in [0, 1[$ such that for all $u > 0$, $\alpha(u) \leq C\rho^u$.

Remark. A lot of published results have shown that the mixing condition (2.1) is satisfied by many time series and spatial random processes (see e.g. Tran [31], Guyon [18], Rosenblatt [28], Doukhan [15]). Moreover, the results presented in this paper could be extended under additional technical assumptions to the case, often considered in the literature, where ψ satisfies:

$$\psi(\mathbf{i}, \mathbf{j}) \leq c \min(\mathbf{i}, \mathbf{j}), \quad \forall\, \mathbf{i}, \mathbf{j} \in \mathbb{N},$$

for some constant $c > 0$.

In the following, we will consider the case where $\alpha(u) \leq Cu^{-\theta}$, for some $\theta > 0$. But, the results can be easly extend to the *GSM* case.

2.2.2. *Local dependency measure.*

In order to obtain the same rate of convergence as in the *i.i.d* case, one requires an other dependency measure, called a *local dependency measure.* Assume that

- For $\ell = 1, ..., d$, there exits a constant $\Delta > 0$ such that the pairs $(X_{\mathbf{i}}^{(\ell)}, X_{\mathbf{j}})$ and $((X_{\mathbf{i}}^{(\ell)}, Y_{\mathbf{i}}), (X_{\mathbf{j}}^{(\ell)}, Y_{\mathbf{j}}))$ admit densities $f_{\mathbf{i},\mathbf{j}}$ and $g_{\mathbf{i},\mathbf{j}}$, as soon as dist$(\mathbf{i}, \mathbf{j}) > \Delta$, such that
$$|f_{\mathbf{i},\mathbf{j}}(x, y) - f(x) f(y)| \leq C, \ \ \forall x, y \in \mathbb{R}$$
$$|g_{\mathbf{i},\mathbf{j}}(u, v) - g(u) g(v)| \leq C, \ \ \forall u, v \in \mathbb{R}^2$$
for some constant $C \geq 0$.

Remark. The link between the two dependency measures can be found in Bosq [7].

Note that if the second measure (as is name point out) is used to control the local dependence, the first one is a kind of "asymptotic dependency" control.

2.3. **Results on U-statistics.**

Let $(X_n,\, n\geq 1)$ be a sequence of real-valued random variables with the same distribution as F. Let the functional:

$$\Theta(F)=\int_{\mathbb{R}^m} h(x_1,x_2,...,x_m)dF(x_1)...dF(x_m),$$

where $m\in\mathbb{N}$, $h(.)$ is some measurable function, called the kernel and F is a distribution function from some given set of distribution function. Without loss of generality, we can assume that $h(.)$ is invariable by permutation. Otherwise, the transformation $\frac{1}{m!}\sum_{1\leq i_1\neq i_2\neq\ldots\neq i_m\leq n} h(x_{i_1},...,x_{i_m})$ will provide a symmetric kernel.

A $U-$statistic with kernel $h(.)$ of degree m based on the sample $(X_i,\, 1\leq i\leq n)$ is a statistic defined by:

$$U_{\mathbf{n}} \;=\; \frac{(n-m)!}{n!}\sum_{1\leq i_1\neq i_2\neq\ldots\neq i_m\leq n} h(X_{i_1},...,X_{i_m})$$

It is said to be an $m-$order $U-$statistic. Let $h_1(x_1)=\int_{\mathbb{R}^{m-1}} h(x_1,x_2,...,x_m)\prod_{j=2}^m dF(x_j)$.

The next Lemma is a consequence of Lemma 2.6 of Sun & Chian [30].

Lemma 2.1. *Let $(X_n,\, n\geq 1)$ be a stationary sequence of strongly mixing random variables. If there exists a positive number δ and δ' $(0<\delta'<\delta)$ verifying $\gamma=\frac{6(\delta-\delta')}{(4+\delta)(2+\delta')}>1$ such that*

$$||h(X_1,...,X_m)||_{4+\delta}<\infty, \tag{2.2}$$

$$\int_{\mathbb{R}^m}|h(x_1,...,x_m)|^{4+\delta}\prod_{j=1}^m dF(x_j)<\infty, \tag{2.3}$$

and $\alpha(n)=\mathcal{O}(n^{-3(4+\delta')/(2+\delta')})$. Then,

$$U_n=\Theta(F)+\frac{2}{n}\sum_{i=1}^n(h_1(X_i)-\Theta(F))+\mathcal{O}_p(\frac{1}{n}).$$

To give strong consistency results, we need the following law of the iterated logarithm of U-statistics:

Lemma 2.2. *(Sun & Chian,* [30]*) Under the same conditions of the previous lemma, we have*

$$U_n-\Theta(F)=\frac{2}{n}\sum_{i=1}^n(h_1(X_i)-\Theta(F))+\mathcal{O}_{a.s}\left(\sqrt{\frac{\log\log n}{n}}\right).$$

Remark 2.3.

- In the following, we are dealing with a kernel $h(.) = K(\frac{.}{h_{\mathbf{n}}})$ which depends on $\mathbf{n}$. Actually, it is a classical approach to use $U-$statistics result to get some assymptotic results of kernel estimators, in the $i.i.d$ case, we refer for example Härdle and Stoker [20]. In fact, the dependence of $h_{\mathbf{n}}$ on $\mathbf{n}$ does not influence the asymptotical results presented here.

3. Estimation of the covariance of Inverse Regression Estimator

We suppose that one deals with a random field $(Z_{\mathbf{i}}, \, \mathbf{i} \in \mathbb{Z}^N)$ which, corresponds, in the spatial regression case, to observations of the form $Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}})$, $\mathbf{i} \in \mathbb{Z}^N$, $(N \geq 1)$ at different locations of a subset of $\mathbb{R}^N$, $N \geq 1$ with some dependency structure. Here, we are particularly interested with the case where the locations take place in lattices of $\mathbb{R}^N$. The general continuous case will be the subject of a forthcoming work.

We deal with the estimation of the matrix $\Sigma_e = \mathbf{varE}(X|Y)$ based on the observations of the process: $(Z_{\mathbf{i}}, \, \mathbf{i} \in \mathcal{I}_{\mathbf{n}})$; $\mathbf{n} \in (\mathbb{N}^*)^N$. In order to ensure the existence of the matrix $\Sigma = \mathbf{var}\, X$ and $\Sigma_e = \mathbf{var}\, \mathbf{E}(X|Y)$, we assume that $\mathbf{E}||X||^4 < \infty$. For sake of simplicity we will consider centered process so $\mathbf{E}X = 0$.

To estimate model (1.1), as previously mentioned, one needs to estimate the matrix $\Sigma^{-1}\Sigma_e$. On the one hand, we can estimate the variance matrix Σ by the empirical spatial estimator, whose consistency will be easily obtained. On the other hand, the estimation of the matrix Σ_e is delicate since it requires the study of the consistency of a suitable estimator of the (inverse) regression function of X given Y:

$$r(y) = \begin{cases} \frac{\varphi(y)}{f(y)} & \text{if } f(y) \neq 0; \\ \mathrm{E}Y & \text{if } f(y) = 0 \end{cases} \quad \text{where } \varphi(y) = \left(\int_{\mathbb{R}} x^{(i)} f_{X^{(i)},Y}(x^{(i)}, y) dx, \, 1 \leq i \leq d \right), y \in \mathbb{R}.$$

An estimator of the *inverse regression* function $r(.)$, based on $(Z_{\mathbf{i}}, \, \mathbf{i} \in \mathcal{I}_{\mathbf{n}})$ is given by

$$r_{\mathbf{n}}(y) = \begin{cases} \frac{\varphi_{\mathbf{n}}(y)}{f_{\mathbf{n}}(y)} & \text{if } f_{\mathbf{n}}(y) \neq 0, \\ \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} & \text{if } f_{\mathbf{n}}(y) = 0, \end{cases}$$

with for all $y \in \mathbb{R}$,

$$f_{\mathbf{n}}(y) = \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K\left(\frac{y - Y_{\mathbf{i}}}{h_{\mathbf{n}}} \right)$$

$$\varphi_{\mathbf{n}}(y) = \frac{1}{\widehat{\mathbf{n}} h_{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} X_{\mathbf{i}} K\left(\frac{y - Y_{\mathbf{i}}}{h_{\mathbf{n}}} \right),$$

where $f_{\mathbf{n}}$ is a kernel estimator of the density, $K : \mathbb{R}^d \to \mathbb{R}$ is a bounded integrable kernel such that $\int K(x)\, dx = 1$ and the bandwidth $h_{\mathbf{n}} \geq 0$ is such that $\lim_{n \to +\infty} h_{\mathbf{n}} = 0$.

The consistency of the estimators $f_{\mathbf{n}}$ and $r_{\mathbf{n}}$ has been studied by Carbon et al [10]. To prevent small-valued density observations y, we consider the following density estimator:

$$f_{e,\mathbf{n}}(y) = \max(e_{\mathbf{n}}, f_{\mathbf{n}}(y))$$

where $(e_{\mathbf{n}})$ is a real-valued sequence such that $\lim_{\mathbf{n}\to\infty} e_{\mathbf{n}} = 0$. Then, we consider the corresponding estimator of r

$$r_{e,\mathbf{n}}(y) = \frac{\varphi_{\mathbf{n}}(y)}{f_{e,\mathbf{n}}(y)}.$$

Finally, for $\overline{X} = \frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} X_{\mathbf{i}}$ we consider the estimator of Σ_e:

$$\Sigma_{e,\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}}\sum r_{e,\mathbf{n}}(Y_{\mathbf{i}})\, r_{e,\mathbf{n}}(Y_{\mathbf{i}})^T - \overline{X}\,\overline{X}^T.$$

We aim at proving the consistency of the empirical variance associated to this estimator.

Remark. Here, we consider as estimator of the density f, $f_{e,\mathbf{n}} = \max(e_{\mathbf{n}}, f_{\mathbf{n}})$, to avoid small values. There are other alternatives such as $f_{e,\mathbf{n}} = f_{\mathbf{n}} + e_{\mathbf{n}}$ or $f_{e,\mathbf{n}} = \max\{(f_{\mathbf{n}} - e_{\mathbf{n}}), 0\}$.

3.1. **Weak consistency.** In the following, for a fixed $\eta > 0$ and a random variable Z in $\mathbb{R}^d$, we will use the notation $\|Z\|_\eta = \mathbf{E}(||Z||^\eta)^{1/\eta}$.

In this section, we will make the following technical assumptions

$$\left\|\frac{r(Y)}{f(Y)}\right\|_{4+\delta_1} < \infty, \quad \text{for some } \delta_1 > 0 \tag{3.1}$$

and

$$\left\|\frac{r(Y)}{f(Y)}\mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}\right\|_2 = \mathcal{O}\left(\frac{1}{\hat{\mathbf{n}}^{\frac{1+\delta}{2}}}\right). \quad \text{for some } 1 > \delta > 0. \tag{3.2}$$

These assumptions are the spatial counterparts of respectively $\|r(Y)\|_{4+\delta} < \infty$ and $\left\|r(Y)\,\mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}\right\|_2 = \mathcal{O}\left(\frac{1}{\hat{\mathbf{n}}^{\frac{1}{4}+\delta}}\right)$ needed in the *i.i.d* case.

We also assume some regularity conditions on the functions: $K(.)$, $f(.)$ and $r(.)$:

- The kernel function $K(.) : \mathbb{R} \to \mathbb{R}^+$ is a $k-$order kernel with compact support and satisfying a Lipschitz condition $|K(x) - K(y)| \le C|x-y|$
- $f(.)$ and $r(.)$ are functions of $C^k(\mathbb{R})$ $(k \ge 2)$ such that $\sup_y |f^{(k)}(y)| < C_1$ and $\sup_y ||\varphi^{(k)}(y)|| < C_2$ for some constants C_1 and C_2,

Set $\Psi_{\mathbf{n}} = h_{\mathbf{n}}^k + \frac{\sqrt{\log\hat{\mathbf{n}}}}{\sqrt{\hat{\mathbf{n}}h_{\mathbf{n}}}}$.

Theorem 3.1. *Assume that* $\alpha(t) \le Ct^{-\theta}$, $t > 0$, $\theta > 2N$ *and* $C > 0$. *If* $E(||X||) < \infty$ *and* $\psi(.) = \mathbf{E}(||X||^2|Y = .)$ *is continuous. Then for a choice of* $h_{\mathbf{n}}$ *such that* $\widehat{\mathbf{n}}h_{\mathbf{n}}^3(\log\widehat{\mathbf{n}})^{-1} \to 0$

and $\widehat{\mathbf{n}}h_{\mathbf{n}}^{\theta_1}(\log \widehat{\mathbf{n}})^{-1} \to \infty$ *with* $\theta_1 = \frac{4N+\theta}{\theta-2N}$, *then, we get*

$$\Sigma_{e,\mathbf{n}} - \Sigma_e = \mathcal{O}_p\left(h_{\mathbf{n}}^k + \frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right)$$

Corollary 3.2. *Under Assumptions of Theorem 3.1 with* $h \simeq n^{-c_1}$, $e_n \simeq n^{-c_2}$ *for some positive constants* c_1 *and* c_2 *such that* $\frac{c_2}{k} + \frac{1}{4k} < c_1 < \frac{1}{2} - 2c_2$, *we have*

$$\Sigma_{e,\mathbf{n}} - \Sigma_e = o_p\left(\frac{1}{\sqrt{\hat{\mathbf{n}}}}\right).$$

Corollary 3.3. *(Central limit theorem) Under previous assumptions, we have*

$$\sqrt{\hat{\mathbf{n}}}\,(\Sigma_{e,\mathbf{n}} - \Sigma_e) \xrightarrow{\mathcal{L}} \Lambda$$

where Λ *is a zero-mean gaussian on the space of d-order matrix with covariance* $\mathbf{var}\left(r(Y)r(Y)^T\right)$.

3.2. **Strong consistency.**

Here we study the case where the response, Y takes values in some compact set. We replace the assumption $\left\|\frac{r(Y)}{f(Y)}\mathbf{1}_{\{f(Y)\leq e_{\mathbf{n}}\}}\right\|_2 = \mathcal{O}\left(\frac{1}{\hat{\mathbf{n}}^{\frac{1}{2}+\delta}}\right)$ by $\mathbf{E}\left(\exp\left(\|r(Y)\|\ \mathbf{1}_{\{f(Y)\leq e_{\mathbf{n}}\}}\right)\right) = \mathcal{O}\left(\hat{\mathbf{n}}^{-\xi}\right)$ for some $\xi > 0$. : $\mathbf{E}\exp\gamma||X|| < \infty$ for some constant $\gamma > 0$.

Theorem 3.4. *If* $(Z_{\mathbf{u}})$ *is GSM, for a choice of* $h_{\mathbf{n}}$ *such that* $\widehat{\mathbf{n}}h_{\mathbf{n}}^3(\log\widehat{\mathbf{n}})^{-1} \to 0$ *and* $\hat{\mathbf{n}}\,h_{\mathbf{n}}(\log\hat{\mathbf{n}})^{-2N-1} \to \infty$. *Assume also that* $\inf_S f(y) > 0$ *for some compact set* S, *then under the Assumptions of Lemma 2.1, we have:*

$$\Sigma_{e,\mathbf{n}} - \Sigma_e = \mathcal{O}_{a.s}\left(h_{\mathbf{n}}^k + \frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right).$$

Corollary 3.5. *Under previous Assumptions, with* $h_{\mathbf{n}} \simeq (\hat{\mathbf{n}})^{-c_1}$, $e_{\mathbf{n}} \simeq \hat{\mathbf{n}}^{-c_2}$ *for some positive constants* c_1 *and* c_2 *such that* $\frac{c_2}{k} + \frac{1}{4k} \leq c_1 < \frac{1}{2} - 2c_2$, *we get*

$$\Sigma_{e,\mathbf{n}} - \Sigma_e = o_{a.s}\left(\sqrt{\frac{\log\log\hat{\mathbf{n}}}{\hat{\mathbf{n}}}}\right).$$

As mentionned previously, the eigenvectors associated with the positive eigenvalues of $\Sigma_{\mathbf{n}}^{-1}\Sigma_{e,\mathbf{n}}$ provide an estimation of the EDR space. Classically, weak and strong consistency results concerning the estimation of the EDR space are obtained by using the previous consistency respectively of the Σ and Σ_e and the theory of perturbation as for example in [37].

4. Spatial inverse methode for spatial prediction

4.1. **Prediction of a spatial process.**

Let $(\xi_{\mathbf{n}}, \mathbf{n} \in (\mathbb{N}^*)^N)$ be a $\mathbb{R}-$valued strictly stationary random spatial process, assumed to be observed over a subset $\mathcal{O}_{\mathbf{n}} \subset \mathcal{I}_{\mathbf{n}}$ ($\mathcal{I}_{\mathbf{n}}$ is a rectangular region as previously defined for some $\mathbf{n} \in (\mathbb{N}^*)^N$). Our aim is to predict the square integrable value, $\xi_{\mathbf{i}_0}$, at a given site $\mathbf{i}_0 \in I_{\mathbf{n}} - \mathcal{O}_{\mathbf{n}}$. In practice, one expects that $\xi_{\mathbf{i}_0}$ only depends on the values of the process on a bounded vicinity set (as small as possible) $\mathcal{V}_{\mathbf{i}_0} \subset \mathcal{O}_{\mathbf{n}}$; i.e that the process $(\xi_{\mathbf{i}})$ is (at least locally) a Markov Random Field (MRF) according to some system of vicinity. Here, we will assume (without loss of generality) that the set of vicinity $(\mathcal{V}_{\mathbf{j}}, \mathbf{j} \in (\mathbb{N}^*)^N)$ is defined by $\mathcal{V}_{\mathbf{j}}$ of the form $\mathbf{j} + \mathcal{V}$ (call vicinity prediction in Biau and Cadre [5]). Then it is well known that the minimum *mean-square error of prediction* of $\xi_{\mathbf{i}_0}$ given the data in $\mathcal{V}_{\mathbf{i}_0}$ is

$$E(\xi_{\mathbf{i}_0} | \xi_{\mathbf{i}}, \mathbf{i} \in \mathcal{V}_{\mathbf{i}_0})$$

and we can consider as predictor any $d-$dimensional vector (where d is the cardinal of $\mathcal{V}$) of elements of $\mathcal{V}_{\mathbf{i}_0}$ concatenated and ordered according to some order. Here, we choose the vector of values of $(\xi_{\mathbf{n}})$ which correspond to the $d-$nearest neighbors: for each $\mathbf{i} \in \mathbb{Z}^N$, we consider that the predictor is the vector $\xi^d_{\mathbf{i}} = (\xi_{\mathbf{i}(k)}; 1 \leq k \leq d)$ where $\mathbf{i}(k)$ is the $k-$th nearest neighbor of $\mathbf{i}$. Then, our problem of prediction amounts to estimate :

$$m(x) = E(\xi_{\mathbf{i}_0} | \xi^d_{\mathbf{i}_0} = x).$$

For this purpose we construct the *associated process*:

$$Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}}) = (\xi^d_{\mathbf{i}}, \xi_{\mathbf{i}}), \ \mathbf{i} \in \mathbb{Z}^N$$

and we consider the estimation of $m(.)$ based on the data $(Z_{\mathbf{i}}, \in \mathcal{O}_{\mathbf{n}})$ and the model (1.1). Note that the linear approximation of $m(.)$ leads to linear predictors. The available literature on such spatial linear models (we invite the reader think of the *kriging method* or spatial auto-regressive method*)* is relatively abundant, see for example, Guyon [18], Anselin and Florax [2], Cressie [11], Wackernagel [33]. In fact, the linear predictor is the optimal predictor (in mimimun mean square error meaning) when the random field under study is *Gaussian*. Then, linear techniques for spatial predicition, give unsatisfactory results when the the process is not *Gaussian*. In this latter case, other approaches such as *log-normal kriging* or the *trans-Gaussian kriging* have been introduced. These methods consist in transforming the original data into a Gaussian distributed data. But, such methods lead to outliers which appear as an effect of the heavy-tailed densities of the data and cannot be delete. Therefore, a specific consideration is needed. This can be done by using, for example, a nonparametric model. That is what is proposed by Biau and Cadre

[5] where a predictor based on *kernel methods* is developped. But, This latter (the kernel nonparametric predictor) as all kernel estimator is submitted to the so-called *dimension curse* and then is penalized by d $(= \text{card}(\mathcal{V}))$, as highlighted in Section 1. Classically, as in Section 1, one uses dimension reduction such as the *inverse regression* method, to overcome this problem. We propose here an adaptation of the *inverse regression* method to get a *dimension reduction predictor* based on model (1.1):

$$\xi_{\mathbf{i}} = g(\Phi.\xi_{\mathbf{i}}^d). \tag{4.1}$$

Remark 4.1.

(1) To estimate this model, we need to check the SIR condition in the context of prediction i.e: *X is such that for all vector b in $\mathbb{R}^d$, there exists a vector B of $\mathbb{R}^D$ such that $\mathbf{E}(b^T X|\Phi.X) = B^T(\Phi.X)$,* is verify if the process $(\xi_{\mathbf{i}})$ is a spatial elliptically distributed process such as Gaussian random field.
(2) In the time series forecasting problem, "inverse regression" property can be an "handicap", since then, one needs to estimate the expectation of the *"future"* given the *"past"*. So, the process under study must be reversible. The flexibility that provide spatial modelling overcome this default since as mentioned in the introduction, the notion of past, present and future does not exist.

At this stage, one can use the method of estimation of the model (1.1) given in Section 1 to get a predictor. Unfortunately (as usually in prediction problem) d is unknown in practice. So, we propose to estimate d by using the fact that we are dealing both with a Markov property and *inverse regression* as follows.

4.2. **Estimation of the number of neighbors necessary for prediction.**

Note that we suppose that the underline process is a stationary Markov process with respect to the $d-$neighbors system of neighborhood, so the variables $\xi_{\mathbf{i}(k)}$ and $\xi_{\mathbf{i}}$ are independent as soon as $k > d$ and

$$\mathbf{E}(\xi_{\mathbf{i}(k)}|\xi_{\mathbf{i}} = y) = 0$$

(since $(\xi_{\mathbf{i}})$ is a stationary zero mean process).

Futhermore since our estimator (of model (1.1)) is based on estimation of $\mathbf{E}(X|Y = y) = \mathbf{E}(\xi_{\mathbf{i}}^d|\xi_{\mathbf{i}} = y) = (\mathbf{E}(\xi_{\mathbf{i}(k)}|\xi_{\mathbf{i}} = y);\, 1 \leq k \leq d)$, that allows us to keep only the neighbors $\xi_{\mathbf{i}(k)}$ for which $\mathbf{E}(\xi_{\mathbf{i}(k)}|\xi_{\mathbf{i}} = y) \neq 0$. Then, an estimation of d is obtained by estimation of $\text{argmin}_k \mathbf{E}(\xi_{\mathbf{i}(k)}|\xi_{\mathbf{i}} = y) = 0$. We propose the following algorithm to get this estimator.

Algorithm for estimation of d, the number of neighbors.

(1) Initialization: specify a parameter $\delta > 0$ (small) and fix a site $\mathbf{j}_0$; set $k = 1$.

(2) compute $r_{\mathbf{n}}^{(k)}(y) = \dfrac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}},\mathcal{V}_{\mathbf{j}_0}\subset\mathcal{O}_{\mathbf{n}}} \xi_{\mathbf{i}(k)}\, K_{h_{\mathbf{n}}}(y-\xi_{\mathbf{i}})}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}},\mathcal{V}_{\mathbf{j}_0}\subset\mathcal{O}_{\mathbf{n}}} K_{h_{\mathbf{n}}}(y-\xi_{\mathbf{i}})}$, the kernel estimate of $r^{(k)}(y) = \mathbf{E}(X^{(k)}|Y = y)$

(3) if $|(r_{\mathbf{n}}^{(k)}(y)| > \delta$, then $k = k+1$ and continue with Step 2; otherwise terminate and $d = k$.

Then, we can compute a predictor based on $d = k$:

4.3. **The dimension reduction predictor.**

To get the predictor, we suggest the following algorithm:

(1) compute

$$r_{\mathbf{n}}^{*}(y) = \frac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}},\mathcal{V}_{\mathbf{i}_0}\subset\mathcal{O}_{\mathbf{n}}} \xi_{\mathbf{i}}^{d}\, K_{h_{\mathbf{n}}}\left(y-\xi_{\mathbf{i}}\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}},\mathcal{V}_{\mathbf{i}_0}\subset\mathcal{O}_{\mathbf{n}}} K_{h_{\mathbf{n}}}\left(y-\xi_{\mathbf{i}}\right)}$$

(2) compute

$$\Sigma_{e,\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}},\mathcal{V}_{\mathbf{i}_0}\subset\mathcal{O}_{\mathbf{n}}} r_{e,\mathbf{n}}^{*}(Y_{\mathbf{i}})\, r_{e,\mathbf{n}}^{*}(Y_{\mathbf{i}})^{T} - \overline{X}\,\overline{X}^{T}.$$

(3) Do the *principal component analisys* of $\Sigma_{\mathbf{n}}^{-1}\Sigma_{e,\mathbf{n}}$ both to get a basis of $\mathrm{Im}(\Sigma_{\mathbf{n}}^{-1}\Sigma_{e,\mathbf{n}})$ and estimation of the D, the dimension of $\mathrm{Im}(\Phi)$ as suggested in the next remark

(4) compute the predictor:

$$\hat{\xi}_{\mathbf{i}_0} = g_{\mathbf{n}}^{*}(\Phi_{\mathbf{n}}^{*}.X_{\mathbf{i}_0}).$$

based on data $(Z_{\mathbf{i}}, \mathbf{i} \in \mathcal{O}_{\mathbf{n}})$; where $g_{\mathbf{n}}^{*}$ is the kernel estimate:

$$g_{\mathbf{n}}^{*}(x) = \frac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}},\mathcal{V}_{\mathbf{i}_0}\subset\mathcal{O}_{\mathbf{n}}} \xi_{\mathbf{i}}\, K_{h_{\mathbf{n}}}\left(\Phi_{\mathbf{n}}^{*}(x-\xi_{\mathbf{i}}^{d})\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}},\mathcal{V}_{\mathbf{i}_0}\subset\mathcal{O}_{\mathbf{n}}} K_{h_{\mathbf{n}}}\left(\Phi_{\mathbf{n}}^{*}(x-\xi_{\mathbf{i}}^{d})\right)} \forall x \in \mathbb{R}^{d}.$$

Remark 4.2.

(1) The problem of estimation of D in step (4) is a classical problem in dimension reduction problems. Several ways exist in the literature. One can for example use the eigenvalues representation of the matrix $\Sigma_{\mathbf{n}}^{-1}\Sigma_{e,\mathbf{n}}$, the measure of distance between spaces as in Li [24] or the selection rule of Ferré [16].

(2) Consitency on the convergence of $\hat{\xi}_{\mathbf{i}_0}$ to $\xi_{\mathbf{i}_0}$ can be obtained by sketching both result of Section 3 and results Biau and Cadre [5].

5. Conclusion

In this work, we have proposed two dimension reduction methods for spatial modeling. The first one is a dimension reduction for spatial regression. It is a natural extension of the idea of Li [24] (called Inverse Regression method) for spatially dependent variables under strong mixing condition. Then, on one hand, we can say that is a good alternative to spatial linear regression model since the link between the variables X and Y is not necessarly linear. Futhermore, as raises Li [24], any linear model can be seen as a particular case of model (1.1) with g being the identity function and $D = 1$. On the other hand, as in the *i.i.d.* case, it requieres less data for calculus than spatial non-parametric regression methods.

The second method that we have studied here deals with spatial prediction modelling. Indeed, it is more general than *kriging method* were the gaussian assumption on the X is needed. Here, we requier that X belongs to a larger class of random variables (that obey to Li [24]'s condition recalled in the introduction). Futhermore, our spatial prediction method has the ease of implementation property of the *inverse regression* methods. Then, for example, it allows to estimate the number of neighbors need to predict. That cannot do the non-parametric prediction method of Biau and Cadre [5].

We have presented here the theoretical framework of our techniques. The next step is to apply them on real data. It is the subjet of works under development.

6. Proofs and Technical Results

6.1. **Deviation Bounds .** To show the strong consistency results, we will use the following Bernstein type deviation inequality:

Lemma 6.1. *Let* $(\zeta_{\mathbf{v}},\, \mathbf{v} \in \mathbb{N}^N)$ *be a zero-mean real-valued random spatial process such that each* $\mathbf{v} \in (\mathbb{N}^*)^N$ *there exists* $c > 0$ *verifying*

$$\mathbf{E}\,|\zeta_{\mathbf{v}}|^k \le k!\, c^{k-2}\, \mathbf{E}\,|\zeta_{\mathbf{v}}|^2\,,\ \forall\, k \ge 2 \tag{6.1}$$

for some constant $c > 0$. *Let* $S_{\mathbf{n}} = \sum_{\mathbf{v}\in I_{\mathbf{n}}} \zeta_{\mathbf{v}}$. *Then for each* $r \in [1, +\infty]$ *and each* $\mathbf{n} \in (\mathbb{N}^*)^N$ *and* $\mathbf{q} \in (\mathbb{N}^*)^N$ *such that* $1 \le q_i \le \frac{n_i}{2}$ *and each* $\varepsilon > 0$,

(6.2)

$$P(|S_{\mathbf{n}}| > \hat{\mathbf{n}}\varepsilon) \le 2^{N+1} exp\left(-\frac{\hat{\mathbf{q}}\varepsilon^2}{4(\,M_2^2 + 2^N c\varepsilon)}\right) + 2^N \times \hat{\mathbf{q}} \times 11 \left(1 + \frac{4c\,p^N M_2^{2/r}}{\varepsilon}\right)^{\frac{r}{2r+1}} \alpha([p])^{2r/(2r+1)}$$

where $M_2^2 = \sup_{\mathbf{v}\in\mathcal{I}_{\mathbf{n}}} \mathbf{E}\zeta_{\mathbf{v}}^2$.

Remark 6.2. Actually, this result is an extension of Lemma 3.2 of Dabo-Niang and Yao [12] for bounded processes. This extension is necessary since in the problem of our interessed, assuming the boundness of the processes amounts to assume that the $X_{\mathbf{i}}$'s are bounded. It is a restrictive condition which (generally) is incompatible with the cornerstone condition of the *inverse regression* (if X is elliptically distributed for example).

We will use the following lemma to get the weak consistency and a law of iterated of the logarithm as well as for the matrix Σ (as we will see immediately) than for the matrix Σ_e (see the proofs of results of Section 3).

Lemma 6.3. *Let* $\{X_{\mathbf{n}}, \mathbf{n} \in \mathbb{N}^N\}$ *be a zero-mean stationary spatial process sequence, of strong mixing random variables.*

(1) If $\mathbf{E}||X||^{2+\delta} < +\infty$ and $\sum \alpha(\hat{\mathbf{n}})^{\frac{\delta}{2+\delta}} < \infty$, for some $\delta > 0$. Then,
$$\frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} X_{\mathbf{i}} = \mathcal{O}_p\left(\frac{1}{\hat{\mathbf{n}}}\right).$$
(2) If $\mathbf{E}||X||^{2+\delta} < +\infty$ and $\sum \alpha(\hat{\mathbf{n}})^{\frac{\delta}{2+\delta}} < \infty$, for some $\delta > 0$. Then,
$$\sqrt{\hat{\mathbf{n}}}(\frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} X_{\mathbf{i}})/\sigma \to \mathcal{N}(0,1)$$
with $\sigma^2 = \sum_{i \in \mathbb{Z}^N} \mathrm{cov}(X_k, X_{\mathbf{i}})$

(3) If $\mathbf{E} \exp \gamma ||X|| < \infty$ for some constant $\gamma > 0$, if for all $u > 0$, $\alpha(u) \leq a\rho^{-u}$, $0 < \rho < 1$ or $\alpha(u) = C.u^{-\theta}$, $\theta > N$ then,
$$\frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} X_{\mathbf{i}} = o_{a.s}\left(\sqrt{\frac{\log \log \hat{\mathbf{n}}}{\hat{\mathbf{n}}}}\right).$$

Remark 6.4.

- *The first result is obtained by using covariance inequality for strong mixing processes (see Bosq* [7]*). Actually, it suffices to enumerate the* X_i *'s into an arbitrary order and sketch the proof in Theorem 1.5 of Bosq* [7]*.*
- *The law of the iterated of the logarithm holds by applying the previous Lemma 6.1 with* $\varepsilon = \eta \sqrt{\frac{\log \log \hat{\mathbf{n}}}{\hat{\mathbf{n}}}}$, $\eta > 0$ *and* $\hat{q} = \left[\frac{\hat{\mathbf{n}}}{\log \log \hat{\mathbf{n}}}\right] + 1$.

6.2. **Consistency of the inverse regression.** In Section 3, we have seen that the results are based on consistency results of the function $r(.)$ which are presented now under some regularity conditions on the functions: $K(.)$, $f(.)$ and $r(.)$.

- The kernel function $K(.) : \mathbb{R} \to \mathbb{R}^+$ is a $k-$order kernel with compact support and satisfying a Lipschitz condition $|K(x) - K(y)| \leq C|x - y|$

- $f(.)$ and $r(.)$ are functions of $C^k(\mathbb{R})$ $(k \geq 2)$ such that $\sup_y |f^{(k)}(y)| < C_1$ and $\sup_y ||\varphi^{(k)}(y)|| < C_2$ for some constants C_1 and C_2,

we have convergence result:

Lemma 6.5. *Suppose $\alpha(t) \leq Ct^{-\theta}$, $t > 0$, $\theta > 2N$ and $C > 0$. If $\widehat{\mathbf{n}}h_{\mathbf{n}}^3(\log \widehat{\mathbf{n}})^{-1} \to 0$, $\widehat{\mathbf{n}}h_{\mathbf{n}}^{\theta_1}(\log \widehat{\mathbf{n}})^{-1} \to \infty$ with $\theta_1 = \frac{4N+\theta}{\theta-2N}$, then*

(1) (see, [10])

$$\sup_{y\in\mathbb{R}}|f_{\mathbf{n}}(y) - f(y)| = \mathcal{O}_p\left(\Psi_{\mathbf{n}}\right). \tag{6.3}$$

(2) Furthermore, if $E(||X||) < \infty$ and $\psi(.) = \mathbf{E}(||X||^2|Y = .)$ is continuous, then

$$\sup_{y\in\mathbb{R}}||\varphi_{\mathbf{n}}(y) - \varphi(y)|| = \mathcal{O}_p\left(\Psi_{\mathbf{n}}\right). \tag{6.4}$$

Remark 6.6. Actually, only the result (6.3) is shown in Carbon et al [10] but the result (6.4) is easily obtained by noting that for all $\varepsilon > 0$,

$$\mathbf{P}(\sup_{y\in\mathbb{R}}||\varphi_{\mathbf{n}}(y) - \mathbf{E}\varphi_{\mathbf{n}}(y)|| > \varepsilon) \leq \frac{\mathbf{E}||X||}{a_{\mathbf{n}}} + \mathbf{P}(\sup_{y\in\mathbb{R}}||\varphi_{\mathbf{n}}(y) - \mathbf{E}\varphi_{\mathbf{n}}(y)|| > \varepsilon,\ \forall i,\ ||X_i|| \leq a_{\mathbf{n}})$$

with $a_{\mathbf{n}} = \eta\left(\log \hat{\mathbf{n}}\right)^{1/4}$, $\eta > 0$.

Lemma 6.7. *If $(Z_{\mathbf{u}})$ is GSM, $\widehat{\mathbf{n}}h_{\mathbf{n}}^3(\log \widehat{\mathbf{n}})^{-1} \to 0$ and $\hat{\mathbf{n}}\, h_{\mathbf{n}}(\log \hat{\mathbf{n}})^{-2N-1} \to \infty$, then*

$$sup_{y\in\mathbb{R}}|f_{\mathbf{n}}(y) - f(y)| = \mathcal{O}_{a.s}\left(\Psi_{\mathbf{n}}\right). \tag{6.5}$$

Furthermore, if $\mathbf{E}\left(\exp \gamma\, ||X||\right) < \infty$ for some $\gamma > 0$ and $\psi(.) = \mathbf{E}(||X||^2|Y = .)$ is continuous, then

$$sup_{y\in\mathbb{R}}||\varphi_{\mathbf{n}}(y) - \varphi(y)|| = \mathcal{O}_{a.s}\left(\Psi_{\mathbf{n}}\right). \tag{6.6}$$

Remark. The equality (6.5) is due to Carbon et al [10]. The proof of the equality (6.6) is obtained applying Lemma 6.1 and sketching the proofs of Theorem 3.1 and 3.3 of Carbon et al [10]. Then it is omitted.

We will need the following lemma and the spatial block decomposition:

Lemma 6.8. *(Bradley's Lemma in Bosq* [6]*)*

Let (X, Y) be an $\mathbb{R}^d \times \mathbb{R}-$valued random vector such that $Y \in \mathbf{L}^r(P)$ for some $r \in [1, +\infty]$. Let c be a real number such that $||Y + c||_r > 0$ and $\xi \in (0,\ ||Y + c||_r]$. Then there exists a random variable Y^* such that:

(1) *$P_{Y^*} = P_Y$ and Y^* is independent of X,*
(2) *$P(|Y^* - Y| > \xi) \leq 11\left(\xi^{-1}||Y + c||_r\right)^{r/(2r+1)} \times \left[\alpha\left(\sigma(X),\, \sigma(Y)\right)\right]^{2r/(2r+1)}$.*

Spatial block decomposition.

Let $Y_{\mathbf{u}} = \zeta_{\mathbf{v}=([u_i]+1,\, 1\leq i\leq N)}$, $\mathbf{u} \in \mathbb{R}^N$. The following spatial blocking idea here is that of Tran [31] and Politis and Romano [26].

Let $\Delta_{\mathbf{i}} = \int_{(i_1-1)}^{i_1} \dots \int_{(i_N-1)}^{i_N} Y_{\mathbf{u}} d\mathbf{u}$. Then,

$$S_{\mathbf{n}} = \int_0^{n_1} \dots \int_0^{n_N} Y_{\mathbf{u}} d\mathbf{u} = \sum_{\substack{1 \leq i_k \leq n_k \\ k=1,\dots N}} \Delta_{\mathbf{i}}.$$

So, $S_{\mathbf{n}}$ is the sum of $2^N P^N \; q_1 \times q_2 \times \cdots \times q_N$ terms $\Delta_{\mathbf{i}}$. And each of them is an integral of $Y_{\mathbf{u}}$ over a cubic block of side p. Let consider the classical block decomposition:

$$U(1,\mathbf{n},\mathbf{j}) = \sum_{k_i=2j_ip+1,\; 1\leq i\leq N}^{(2j_i+1)p} \Delta_{\mathbf{k}},$$

$$U(2,\mathbf{n},\mathbf{j}) = \sum_{k_i=2j_ip+1,\; 1\leq i\leq N-1}^{(2j_i+1)p} \; \sum_{k_N=(2j_N+1)p+1}^{2(j_N+1)p} \Delta_{\mathbf{k}},$$

$$U(3,\mathbf{n},x,\mathbf{j}) = \sum_{k_i=2j_ip+1,\; 1\leq i\leq N-2}^{(2j_i+1)p} \; \sum_{k_{N-1}=(2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \; \sum_{k_N=2j_Np+1}^{(2j_N+1)p} \Delta_{\mathbf{k}},$$

$$U(4,\mathbf{n},\mathbf{j}) = \sum_{k_i=2j_ip+1,\; 1\leq i\leq N-2}^{(2j_i+1)p} \; \sum_{k_{N-1}=(2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \; \sum_{k_N=(2j_N+1)p+1}^{2(j_N+1)p} \Delta_{\mathbf{k}},$$

and so on. Note that

$$U(2^{N-1},\mathbf{n},\mathbf{j}) = \sum_{k_i=(2j_i+1)p+1,\; 1\leq i\leq N-1}^{2(j_i+1)p} \; \sum_{k_N=2j_Np+1}^{(2j_N+1)p} \Delta_{\mathbf{k}}.$$

Finally,

$$U(2^N,\mathbf{n},\mathbf{j}) = \sum_{k_i=(2j_i+1)p+1,\; 1\leq i\leq N}^{2(j_i+1)p} \Delta_{\mathbf{k}}.$$

So,

$$S_{\mathbf{n}} = \sum_{i=1}^{2^N} T(\mathbf{n}, i), \tag{6.7}$$

with $T(\mathbf{n}, i) = \sum_{j_l=0,\, l=1,\dots,N}^{q_l-1} U(i,\mathbf{n},\mathbf{j})$.

If $n_i \neq 2pt_i$, $i = 1, \dots, N$, for all set of integers $t_1, \dots, t_N$, then a term, say $T\left(\mathbf{n},\, 2^N+1\right)$ containing all the $\Delta_{\mathbf{k}}$'s at the end, and not included in the blocks above, can be added (see Tran [31] or Biau and Cadre [4]). This extra term does not change the result of previous proof.

Proof of Lemma 6.1.

Using (6.7) it suffices to show that

(6.8)
$$\mathbf{P}\left(|T(\mathbf{n},i)| > \frac{\hat{\mathbf{n}}\varepsilon}{2^N}\right) \leq 2\exp\left(-\frac{\varepsilon^2}{4v^2(\mathbf{q})}\hat{\mathbf{q}}\right)+\hat{\mathbf{q}}\times 11\left(1+\frac{4C\,p^N M_2^{2/r}}{\varepsilon}\right)^{r/(2r+1)}\alpha([p])^{2r/(2r+1)}$$

for each $1 \leq i \leq 2^N$.

Without loss of generality we will show (6.8) for $i = 1$. Now, we enumerate (as it is often done in this case) in arbitrary way the $\hat{\mathbf{q}} = q_1 \times q_2 \times \cdots \times q_N$ terms $U(1, \mathbf{n}, \mathbf{j})$ of sum of $T(\mathbf{n}, 1)$ that we call $W_1, ..., W_{\hat{\mathbf{q}}}$. Note that the $U(1, \mathbf{n}, \mathbf{j})$ are measurable with respect to the $\sigma-$field generated by $Y_{\mathbf{u}}$ with $\mathbf{u}$ such that $2j_i p \leq u_i \leq (2j_i + 1)p$, $i = 1, ..., N$.

These sets of sites are separated by a distance at least p and since for all $m = 1, ..., \hat{\mathbf{q}}$ there exists a $\mathbf{j}(m)$ such that $W_m = U(1, \mathbf{n}, \mathbf{j}(m))$ which have the same distribution as W_m^* ,

$$\mathbf{E}|W_m|^r = \mathbf{E}|W_m^*|^r = \mathbf{E}\left|\int_{2j_1(m)p}^{(2j_1(m)+1)p} \cdots \int_{2j_N(m)p}^{(2j_N(m)+1)p} Y_{\mathbf{u}} d\mathbf{u}\right|^r, \ r \in [1, +\infty].$$

Noting that

$$\int_{2j_k(m)p}^{(2j_k(m)+1)p} Y_{\mathbf{u}}\, d\mathbf{u} = \int_{2j_k(m)p}^{[2j_k(m)p]+1} Y_{\mathbf{u}}\, d\mathbf{u} + \sum_{v_k=[2j_k(m)p]+2}^{[(2j_k(m)+1)p]} \zeta_{\mathbf{v}} + \int_{[(2j_k(m)+1)p]}^{2j_k(m)+1)p} Y_{\mathbf{u}}\, d\mathbf{u}$$

$$= ([2j_k(m)p] + 1 - 2j_k(m)p)\,\zeta_{(\mathbf{v},\, v_k=[2j_k(m)p]+1)} + \sum_{v_k=[2j_k(m)p]+2}^{[(2j_k(m)+1)p]} \zeta_{\mathbf{v}}$$
$$+ ((2j_k(m) + 1)p - [(2j_k(m) + 1)p])\,\zeta_{(\mathbf{v},\, v_k=[(2j_k(m)+1)p]+1)}$$
$$= \sum_{v_k=[2j_k(m)p]+1}^{[(2j_k(m)+1)p]+1} w(\mathbf{j}, \mathbf{v})_k\, \zeta_{\mathbf{v}}$$

and $|w(\mathbf{j}, \mathbf{v})_k| \leq 1$ $\forall k = 1, ..., N$, we have by using Minkovski's inequality and 6.1 one get

$$\mathbf{E}\left|\frac{W_m}{p^N}\right|^r \leq c^{r-2} r!\, M_2^2\,, \forall r \geq 2. \tag{6.9}$$

Then, using recursively the version of Bradley's lemma gives in Lemma 6.8 we define independent random variables $W_1^*, ..., W_{\hat{\mathbf{q}}}^*$ such that for all $r \in [1, +\infty]$ and for all $m = 1, ..., \hat{\mathbf{q}}$, W_m^* has the same distribution with W_m and setting $\omega_r^r = p^{rN} c^{r-2} M_2^2$, we have:

$$P(|W_m - W_m^*| > \xi) \leq 11 \left(\frac{||W_m + \omega_r||_r}{\xi} \right)^{r/(2r+1)} \alpha([p])^{2r/(2r+1)},$$

where, $c = \delta \omega_r p$ and $\xi = \min\left(\frac{\hat{\mathbf{n}}\varepsilon}{2^{N+1}\hat{\mathbf{q}}}, (\delta - 1)\omega_r p^N\right) = \min\left(\frac{\varepsilon p^N}{2}, (\delta - 1)\omega_r p^N\right)$ for some $\delta > 1$ specified below. Note that for each m,

$$||W_m + c||_r \geq c - ||W_m||_r \geq (\delta - 1)\omega_r p^N > 0$$

so that $0 < \xi < ||W_m + c||_r$ as required in Lemma 6.8.

Then, if $\delta = 1 + \frac{\varepsilon}{2\omega_r}$,

$$P(|W_m - W_m^*| > \xi) \leq 11 \left(1 + \frac{4\omega_r}{\varepsilon} \right)^{r/(2r+1)} \alpha([p])^{2r/(2r+1)}$$

and

$$P\left(\sum_{m=1}^{\hat{\mathbf{q}}} |W_m - W_m^*| > \frac{\hat{\mathbf{n}}\varepsilon}{2^{N+1}} \right) \leq \hat{\mathbf{q}} \times 11 \left(1 + \frac{4\omega_r}{\varepsilon} \right)^{r/(2r+1)} \alpha([p])^{2r/(2r+1)}.$$

Now, note that Inequality (6.9) also leads (by Bernstein's inequality) to :

$$P\left(\left| \sum_{m=1}^{\hat{\mathbf{q}}} W_m^* \right| > \frac{\hat{\mathbf{n}}\varepsilon}{2^{N+1}} \right) \leq 2\exp\left(- \frac{\left(\frac{\hat{\mathbf{n}}\varepsilon}{2^{N+1}}\right)^2}{4\sum_{m=1}^{\hat{\mathbf{q}}} \mathbf{E}\, W_m^2 + \frac{c\hat{\mathbf{n}}p^N}{2^{N+1}}\varepsilon} \right)$$

Thus

$$P(|T(\mathbf{n}, 1)| > \tfrac{\hat{\mathbf{n}}\varepsilon}{2^N}) \;\leq\; 2\exp\left(- \frac{\hat{\mathbf{q}}\varepsilon^2}{4(M_2^2 + 2^N c\varepsilon)} \right) + \hat{\mathbf{q}} \times 11 \left(1 + \frac{4c\, p^N M_2^{2/r}}{\varepsilon} \right)^{r/(2r+1)} \alpha([p])^{2r/(2r+1)}$$

Then, since $\hat{\mathbf{q}} = q_1 \times ... \times q_N$ and $\hat{\mathbf{n}} = 2^N p^N \hat{\mathbf{q}}$, we get inequality (6.8) the proof is completed by noting that $P(|S_{\mathbf{n}}| > \hat{\mathbf{n}}\varepsilon) \leq 2^N P(|T(\mathbf{n}, i)| > \frac{\hat{\mathbf{n}}\varepsilon}{2^N})$. □

6.3. **Proof of the Theorem 3.1.** We will prove the desired result on $\Sigma_{e,\mathbf{n}} - \Sigma_e$ using an intermediate matrix

$$\overline{\Sigma}_{e,\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} r(Y_{\mathbf{i}}) r(Y_{\mathbf{i}})^T.$$

Start with the following decomposition

$$\Sigma_{e,\mathbf{n}} - \Sigma_e = \Sigma_{e,\mathbf{n}} - \overline{\Sigma}_{e,\mathbf{n}} + \overline{\Sigma}_{e,\mathbf{n}} - \Sigma_e.$$

We first show that:

$$\Sigma_{e,\mathbf{n}} - \overline{\Sigma}_{e,\mathbf{n}} = \mathcal{O}_p\left(\frac{1}{\hat{\mathbf{n}}^{\frac{1}{2}+\delta}} + \frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2} \right). \tag{6.10}$$

To this aim, we set :

$$\Sigma_{e,\mathbf{n}} - \overline{\Sigma}_{e,\mathbf{n}} = S_{\mathbf{n},1} + S_{\mathbf{n},2} + S_{\mathbf{n},3} \tag{6.11}$$

with

$$S_{\mathbf{n},1} = \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_\mathbf{n}} \left(\hat{r}_{e_\mathbf{n}}(Y_\mathbf{i}) - r(Y_\mathbf{i})\right)\left(\hat{r}_{e_\mathbf{n}}(Y_\mathbf{i}) - r(Y_\mathbf{i})\right)^T,$$

$$S_{\mathbf{n},2} = \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_\mathbf{n}} r(Y_\mathbf{i})\ \left(\hat{r}_{e_n}(Y_i) - r(Y_\mathbf{i})\right)^T$$

and

$$S_{\mathbf{n},3} = \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_\mathbf{n}} \left(\hat{r}_{e_n}(Y_i) - r(Y_\mathbf{i})\right)\ r(Y_\mathbf{i})^T.$$

Note that $S^T_{\mathbf{n},3} = S_{\mathbf{n},2}$, hence we only need to control the rate of convergence of the first two terms $S_{\mathbf{n},1}$ and $S_{\mathbf{n},2}$

We will successively prove that

$$S_{\mathbf{n},1} = \mathcal{O}_p\left(\frac{\Psi^2_\mathbf{n}}{e^2_\mathbf{n}}\right),$$

and

$$S_{\mathbf{n},2} = \mathcal{O}_p\left(\frac{\Psi^2_\mathbf{n}}{e^2_\mathbf{n}} + h^k_\mathbf{n}\right)$$

this latter will immediately implies that

$$S_{\mathbf{n},3} = \mathcal{O}_p\left(\frac{\Psi^2_\mathbf{n}}{e^2_\mathbf{n}} + h^k_\mathbf{n}\right).$$

- Control on $S_{\mathbf{n},1}$

Since for each $y \in \mathbb{R}$:

$$\hat{r}_{e_\mathbf{n}}(y) - r(y) = \frac{r(y)}{f_{e_\mathbf{n}}(y)}\left(f(y) - f_{e_\mathbf{n}}(y)\right) + \frac{1}{\hat{f}_{e_\mathbf{n}}(y)}\left(\varphi_\mathbf{n}(y) - \varphi(y)\right) \tag{6.12}$$

and

$$f(y) - f_{e_\mathbf{n}}(y) = f(y) - f_\mathbf{n}(y) + (f_\mathbf{n}(y) - e_\mathbf{n})\mathbf{1}_{\{f_\mathbf{n}(y)<e_\mathbf{n}\}}, \tag{6.13}$$

for each $\mathbf{i} \in (\mathbb{N}^*)^N$

$$\|r_{e_n}(Y_\mathbf{i}) - r(Y_\mathbf{i})\| \leq \frac{\|r(Y_\mathbf{i})\|}{e_\mathbf{n}}\,||f_\mathbf{n} - f||_\infty + 2\,\|r(Y_\mathbf{i})\|\,\mathbf{1}_{\{f_\mathbf{n}(Y_\mathbf{i})<e_\mathbf{n}\}} + \frac{\|\varphi_\mathbf{n} - \varphi\|_\infty}{e_\mathbf{n}}.$$

and

$$\|r_{e_\mathbf{n}}(Y_\mathbf{i}) - r(Y_\mathbf{i})\|^2 \leq 3\left[\|r(Y_\mathbf{i})\|^2\,\frac{||f_\mathbf{n} - f||^2_\infty}{e^2_n} + 4\,\|r(Y_\mathbf{i})\|^2\,\mathbf{1}_{\{f_\mathbf{n}(Y_\mathbf{i})<e_\mathbf{n}\}} + \frac{||\varphi_\mathbf{n} - \varphi||^2_\infty}{e^2_\mathbf{n}}\right].$$

Using the following inequality (see Ferré and Yao [17] for details):

$$\mathbf{1}_{\{f_{\mathbf{n}}(Y_{\mathbf{i}})<e_n\}} \le \mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}} + \frac{||f_{\mathbf{n}} - f||_\infty^2}{e_n^2}, \tag{6.14}$$

and by results on Lemmas 6.3 and 6.5, we have:

$$S_{\mathbf{n},1} \le \frac{C}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \|r(Y_{\mathbf{i}})\|^2 \mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}} + \mathcal{O}_p\left(\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right), \, C>0.$$

Now, noting that

$$\frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \|r(Y_{\mathbf{i}})\|^2 \mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}} \le e_{\mathbf{n}}^2 \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \frac{\|r(Y_{\mathbf{i}})\|^2}{f(Y_{\mathbf{i}})^2} \mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}},$$

we have (since $\mathbf{E}\left(\frac{||r(Y_{\mathbf{i}})||^2}{f(Y_{\mathbf{i}})^2}\mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}}\right) = \mathcal{O}\left(\frac{1}{\hat{\mathbf{n}}^{1+\delta}}\right)$ by assumption):

$$\frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \|r(Y_{\mathbf{i}})\|^2 \mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}} = \mathcal{O}_p\left(\frac{e_{\mathbf{n}}^2}{\hat{\mathbf{n}}^{1+\delta}}\right) \tag{6.15}$$

and

$$S_{\mathbf{n},1} = \mathcal{O}_p\left(\frac{e_{\mathbf{n}}^2}{\hat{\mathbf{n}}^{1+\delta}} + \frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right)$$

because of Assumption $\mathbf{E}\left(\frac{\|r(Y)\|^2}{f(Y)^2}\mathbf{1}_{\{f(Y)<e_{\mathbf{n}}\}}\right) = \mathcal{O}\left(\frac{1}{\hat{\mathbf{n}}^{1+\delta}}\right)$.

Now, since $\Psi_{\mathbf{n}} = h_{\mathbf{n}}^k + \sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}}\, h_{\mathbf{n}}}}$ and $\frac{e_{\mathbf{n}}}{\hat{\mathbf{n}}^{\frac{1+\delta}{2}}} \le C\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}}\, h_{\mathbf{n}}}}$ (for $\hat{\mathbf{n}}$ large and $C>0$ an arbitrary constante), we have:

$$S_{\mathbf{n},1} = \mathcal{O}_p\left(\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right). \tag{6.16}$$

- Control on $S_{\mathbf{n},2}$.

Noting that : $\frac{1}{f_{e_n}} = \frac{1}{f} + \frac{f-\tilde{f}_{e_{\mathbf{n}}}}{\tilde{f}_{e_{\mathbf{n}}} f} + \frac{\tilde{f}_{e_{\mathbf{n}}}-f_{e_{\mathbf{n}}}}{\tilde{f}_{e_{\mathbf{n}}} f_{e_{\mathbf{n}}}} = \frac{1}{f} + \frac{f-e_{\mathbf{n}}}{\tilde{f}_{e_{\mathbf{n}}} f}\mathbf{1}_{\{f<e_{\mathbf{n}}\}} + \frac{\tilde{f}_{e_{\mathbf{n}}}-f_{e_{\mathbf{n}}}}{\tilde{f}_{e_{\mathbf{n}}} f_{e_{\mathbf{n}}}}$, with $\tilde{f}_{e_{\mathbf{n}}} = \max\{f, e_{\mathbf{n}}\}$, we have:

$$\begin{aligned} S_{\mathbf{n},2} &= \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \frac{r(Y_{\mathbf{i}})r(Y_{\mathbf{i}})^T}{f_{e_{\mathbf{n}}}(Y_{\mathbf{i}})} \left(f(Y_{\mathbf{i}}) - f_{e_{\mathbf{n}}}(Y_{\mathbf{i}})\right) + \frac{r(Y_{\mathbf{i}})}{f_{e_{\mathbf{n}}}(Y_{\mathbf{i}})} \left(\varphi_{\mathbf{n}}(Y_{\mathbf{i}}) - \varphi(Y_{\mathbf{i}})\right)^T \\ &= \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \frac{r(Y_{\mathbf{i}})r(Y_{\mathbf{i}})^T}{f(Y_{\mathbf{i}})} \left(f(Y_{\mathbf{i}}) - f_{\mathbf{n}}(Y_{\mathbf{i}})\right) + \frac{r(Y_{\mathbf{i}})}{f(Y_{\mathbf{i}})} \left(\varphi_{\mathbf{n}}(Y_{\mathbf{i}}) - \varphi(Y_{\mathbf{i}})\right)^T + R_{\mathbf{n}_1} + R_{\mathbf{n}_2}. \end{aligned}$$

where

$$\begin{aligned} R_{\mathbf{n}_1}(Y_{\mathbf{i}}) &= r(Y_{\mathbf{i}}) \left[r(Y_{\mathbf{i}})^T \left(f(Y_{\mathbf{i}}) - f_{\mathbf{n}}(Y_{\mathbf{i}})\right) + \left(\varphi_{\mathbf{n}}(Y_{\mathbf{i}}) - \varphi(Y_{\mathbf{i}})\right)^T\right] \\ & \quad \left(\frac{1}{f(Y_{\mathbf{i}})}\mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}} + \frac{\tilde{f}_{e_{\mathbf{n}}}(Y_{\mathbf{i}}) - f_{e_{\mathbf{n}}}(Y_{\mathbf{i}})}{\tilde{f}_{e_{\mathbf{n}}}(Y_{\mathbf{i}}) f_{e_{\mathbf{n}}}(Y_{\mathbf{i}})}\right) \end{aligned}$$

and

$$R_{\mathbf{n}_2}=\frac{r(Y_{\mathbf{i}})r(Y_{\mathbf{i}})^T}{f_{e_{\mathbf{n}}}(Y_{\mathbf{i}})}\left(f_{\mathbf{n}}(Y_i)-f_{e_{\mathbf{n}}}(Y_{\mathbf{i}})\right).$$

Futhermore :

- since for all $y\in\mathbb{R}$ we have $\frac{1}{\tilde{f}_{e_{\mathbf{n}}}(y)f_{e_{\mathbf{n}}}(y)}\leq\frac{1}{e_{\mathbf{n}}^2}$ and by several calculus we also have $\left|\tilde{f}_{e_{\mathbf{n}}}(y)-f_{e_{\mathbf{n}}}(y)\right|\leq|f(y)-f_{\mathbf{n}}(y)|$ and then $||\tilde{f}_{e_{\mathbf{n}}}-f_{e_{\mathbf{n}}}||_\infty\leq||f_{\mathbf{n}}-f||_\infty$, we also have one hand:

$$R_{\mathbf{n}_1}\leq\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\left(||r(Y_{\mathbf{i}})||\,||\varphi_{\mathbf{n}}-\varphi||_\infty+||r(Y_{\mathbf{i}})||^2\,||f_{\mathbf{n}}-f||_\infty\right)\left(\frac{1}{f(Y_{\mathbf{i}})}\mathbf{1}_{\{f<e_{\mathbf{n}}\}}+\frac{||f_{\mathbf{n}}-f||_\infty}{e_{\mathbf{n}}^2}\right)\quad(6.17)$$

- on the other hand we have

$$\begin{aligned}R_{\mathbf{n}_2}&\leq\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}||r(Y_{\mathbf{i}})||^2\,\frac{|f_{\mathbf{n}}(Y_{\mathbf{i}})-e_{\mathbf{n}}|}{f_{e_{\mathbf{n}}}(Y_{\mathbf{i}})}\mathbf{1}_{\{f_{\mathbf{n}}(Y_i)<e_{\mathbf{n}}\}}\\&\leq\frac{2}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}||r(Y_{\mathbf{i}})||^2\,\mathbf{1}_{\{f_{\mathbf{n}}(Y_i)<e_{\mathbf{n}}\}}.\end{aligned}$$

  because for all $y\in\mathbb{R}$, $|f_{\mathbf{n}}(y)-f_{e_{\mathbf{n}}}(y)|=|f_{\mathbf{n}}(y)-e_{\mathbf{n}}|\,\mathbf{1}_{\{f_{\mathbf{n}}(y)<e_{\mathbf{n}}\}}\leq 2e_{\mathbf{n}}\mathbf{1}_{\{f_{\mathbf{n}}(y)<e_{\mathbf{n}}\}}$.

Then, it follows from (6.14 and 6.15) that:

$$R_{\mathbf{n}_2}=\mathcal{O}_p\left(\frac{e_{\mathbf{n}}^2}{\hat{\mathbf{n}}^{1+\delta}}+\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right)$$

as for $S_{\mathbf{n}_1}$, we deduce:

$$R_{\mathbf{n}_2}=\mathcal{O}_p\left(\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right)$$

Now, observious that,

$$\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\frac{||r(Y_{\mathbf{i}})||^2}{f(Y_{\mathbf{i}})}\mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}}\leq e_{\mathbf{n}}\,\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\frac{||r(Y_{\mathbf{i}})||^2}{f(Y_{\mathbf{i}})^2}\mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}},$$

we have (as previously):

$$\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\frac{||r(Y_{\mathbf{i}})||^2}{f(Y_{\mathbf{i}})}\mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}}=\mathcal{O}_p\left(\frac{e_{\mathbf{n}}}{\hat{\mathbf{n}}^{1+\delta}}\right).\quad(6.18)$$

Moreover, since $\mathbf{E}\left(\frac{||r(Y)||^2}{f(Y)^2}\mathbf{1}_{\{f(Y)<e_{\mathbf{n}}\}}\right)=\mathcal{O}\left(\frac{1}{\hat{\mathbf{n}}^{1+\delta}}\right)$, we also have:

$$\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\frac{||r(Y_{\mathbf{i}})||}{f(Y_{\mathbf{i}})}\mathbf{1}_{\{f(Y_{\mathbf{i}})<e_{\mathbf{n}}\}}=\mathcal{O}_p\left(\frac{1}{\hat{\mathbf{n}}^{\frac{1+\delta}{2}}}\right)\quad(6.19)$$

So combining (6.17), (6.18) and (6.19), we get:

$$R_{\mathbf{n}_1}=\mathcal{O}_p\left(\frac{e_{\mathbf{n}}\Psi_{\mathbf{n}}}{\hat{\mathbf{n}}^{1+\delta}}+\frac{\Psi_{\mathbf{n}}}{\hat{\mathbf{n}}^{\frac{1+\delta}{2}}}+\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right)=\mathcal{O}_p\left(\frac{\Psi_{\mathbf{n}}}{\hat{\mathbf{n}}^{\frac{1+\delta}{2}}}+\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right);$$

and since $\frac{e_{\mathbf{n}}}{\hat{\mathbf{n}}^{\frac{1+\delta}{2}}} \leq C\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}}\, h_{\mathbf{n}}}}$ (for $\hat{\mathbf{n}}$ large) we have:

$$R_{\mathbf{n}_1} = \mathcal{O}_p\left(\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right).$$

Then,

$$S_{\mathbf{n},2} = S_{\mathbf{n},2}^{(1)} + S_{\mathbf{n},2}^{(2)} + \mathcal{O}_p\left(\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right);$$

with

$$S_{\mathbf{n},2}^{(1)} = \frac{1}{\hat{\mathbf{n}}}\sum_{i=1} \frac{r(Y_{\mathbf{i}})r(Y_{\mathbf{i}})^T}{f(Y_{\mathbf{i}})}\left(f_{\mathbf{n}}(Y_{\mathbf{i}}) - f(Y_{\mathbf{i}})\right)$$

and

$$S_{\mathbf{n},2}^{(2)} = \frac{1}{\hat{\mathbf{n}}}\sum_{i=1} \frac{r(Y_{\mathbf{i}})}{f(Y_{\mathbf{i}})}\left(\varphi_{\mathbf{n}}(Y_{\mathbf{i}}) - \varphi(Y_{\mathbf{i}})\right)^T.$$

To finish, we are going to show that

$$S_{\mathbf{n},2}^{(1)} = \mathcal{O}_p(h_{\mathbf{n}}^k + \frac{1}{\hat{\mathbf{n}}h_{\mathbf{n}}})$$

$$S_{\mathbf{n},2}^{(2)} = \mathcal{O}_p(h_{\mathbf{n}}^k + \frac{1}{\hat{\mathbf{n}}h_{\mathbf{n}}})$$

Note that:

$$S_{\mathbf{n},2}^{(1)} = \frac{1}{\hat{\mathbf{n}}}\sum_{i=1} \tau(Y_{\mathbf{i}})\, f(Y_{\mathbf{i}}) - \frac{1}{h_{\mathbf{n}}} V_{\mathbf{n}}$$

where $\tau(.)$ is a function defined by $\tau(y) = \frac{r(y)\, r(y)^T}{f(y)}$ for $y \in \mathbb{R}$ and

$$V_{\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}^2}\sum_{\mathbf{i},\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} \tau(Y_{\mathbf{i}})\, K_{h_{\mathbf{n}}}(Y_{\mathbf{i}} - Y_{\mathbf{j}})$$

is a second-order Von Mises functional statistic which associated U-statistic is:

$$U_{\mathbf{n}} = \frac{1}{2\hat{\mathbf{n}}(\hat{\mathbf{n}}-1)}\sum_{i=1}\sum_{j\neq i}[\tau(Y_{\mathbf{i}}) + \tau(Y_{\mathbf{j}})]K_{h_{\mathbf{n}}}(Y_{\mathbf{i}} - Y_{\mathbf{j}}).$$

Since: $V_{\mathbf{n}} = U_{\mathbf{n}} + \mathcal{O}_p(\frac{1}{\hat{\mathbf{n}}})$,

$$S_{\mathbf{n},2}^{(1)} = \frac{1}{\hat{\mathbf{n}}}\sum_{i=1} \tau(Y_{\mathbf{i}})\, f(Y_{\mathbf{i}}) - \frac{1}{h_{\mathbf{n}}} U_{\mathbf{n}} + \mathcal{O}_p\left(\frac{1}{\hat{\mathbf{n}}h_{\mathbf{n}}}\right).$$

We apply Lemma 2.1 with, $m = 2$, $h(y_1, y_2) = [\tau(y_1) + \tau(y_2)]\, K_{h_{\mathbf{n}}}(y_1 - y_2)$

$$h_1(y) = \frac{1}{2}\left[\tau(y).f * K_{h_{\mathbf{n}}}(y) + (\tau.f) * K_{h_n}(y)\right],$$

and

$$\Theta(F) = \mathbf{E}\left(h_1(Y)\right) = \mathbf{E}\ \left(\tau(y).f * K_{h_\mathbf{n}}(y)\right).$$

Since

$$||h(Y_1, Y_2)||_{4+\delta} \leq C.||\tau(Y)||_{4+\delta} < \infty,$$

by assumption (3.1) then,

$$U_\mathbf{n} = \Theta(F) + \frac{2}{\hat{\mathbf{n}}}\sum_{\mathbf{i}}\ (h_1(Y_\mathbf{i}) - \Theta(F)) + \mathcal{O}_p(\frac{1}{\hat{\mathbf{n}}}).$$

and

$$\begin{aligned} S^{(1)}_{\mathbf{n},2} &= \frac{1}{\hat{\mathbf{n}}}\sum_{i=1}\tau(Y_\mathbf{i})\, f(Y_\mathbf{i}) - \frac{\Theta(F)}{h_\mathbf{n}} - \frac{2}{\hat{\mathbf{n}}}\sum_{\mathbf{i}}\left(\frac{h_1(Y_\mathbf{i})}{h_\mathbf{n}} - \frac{\Theta(F)}{h_\mathbf{n}}\right) + \mathcal{O}_p\left(\frac{1}{\hat{\mathbf{n}}h_\mathbf{n}}\right) \\ &= \frac{1}{\hat{\mathbf{n}}}\sum_{i=1}\tau(Y_\mathbf{i})\left(f(Y_\mathbf{i}) - \frac{f * K_{h_\mathbf{n}}(Y_\mathbf{i})}{h_\mathbf{n}}\right) + \frac{\Theta(F) - (\tau.f) * K_{h_\mathbf{n}}(y_\mathbf{i})}{h_\mathbf{n}} + \mathcal{O}_p\left(\frac{1}{\hat{\mathbf{n}}h_\mathbf{n}}\right). \end{aligned}$$

Since f and $r(.)$ belongs to $C^k(\mathbb{R})$, we get,

$$\left\|\frac{f * K_{h_\mathbf{n}}(y)}{h_\mathbf{n}} - f(y)\right\|_\infty = \mathcal{O}(h^k_\mathbf{n})$$

and

$$\left\|\frac{(\tau.f) * K_{h_\mathbf{n}}(y)}{h_\mathbf{n}} - (\tau.f)(y)\right\|_\infty = \mathcal{O}(h^k_\mathbf{n}).$$

Then, we have

$$\frac{1}{\hat{\mathbf{n}}}\sum_{i\in\mathcal{I}_\mathbf{n}}\frac{\tau(Y_\mathbf{i})\, f * K_{h_\mathbf{n}}(Y_\mathbf{i})}{h_\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}}\sum_{i\in\mathcal{I}_\mathbf{n}}(\tau.f)(Y_\mathbf{i}) + \mathcal{O}_p(h^k_\mathbf{n}),$$

$$\frac{1}{\hat{\mathbf{n}}}\sum_{i\in\mathcal{I}_\mathbf{n}}\frac{(\tau.f) * K_{h_\mathbf{n}}(Y_\mathbf{i})}{h_\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}}\sum_{i\in\mathcal{I}_\mathbf{n}}(\tau.f)(Y_\mathbf{i}) + \mathcal{O}(h^k_\mathbf{n})$$

$$\frac{\Theta(F)}{h_\mathbf{n}} = \mathbf{E}((\tau.f)(Y)) + \mathcal{O}(h^k_\mathbf{n}).$$

Finally:

$$S^{(1)}_{\mathbf{n},2} = \mathcal{O}_p(h^k_\mathbf{n} + \frac{1}{\hat{\mathbf{n}}} + \frac{1}{\hat{\mathbf{n}}h_\mathbf{n}}) = \mathcal{O}_p(h^k_\mathbf{n} + \frac{1}{\hat{\mathbf{n}}h_\mathbf{n}}).$$

By using similar arguments and applying Lemma 2.1 with $m = 3$, one also gets

$$S^{(2)}_{\mathbf{n},2} = \mathcal{O}_p(h^k_\mathbf{n} + \frac{1}{\hat{\mathbf{n}}h_\mathbf{n}}).$$

So,

$$S_{\mathbf{n},2} = \mathcal{O}_p\left(\frac{\Psi^2_\mathbf{n}}{e^2_\mathbf{n}} + h^k_\mathbf{n} + \frac{1}{\hat{\mathbf{n}}h_\mathbf{n}}\right)$$

Then, equality, (6.16), and (6.20) lead to (6.10).

Recall that $\Psi_{\mathbf{n}} = h_{\mathbf{n}}^k + \sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} h_{\mathbf{n}}}}$. Then, the fact that there exist a real $A > 0$ such that $\forall \hat{\mathbf{n}} > A$, $\frac{1}{\hat{\mathbf{n}} h_{\mathbf{n}}} < \frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} h_{\mathbf{n}} e_{\mathbf{n}}^2}$ and :

$$S_{\mathbf{n},2} = \mathcal{O}_p\left(\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2} + h_{\mathbf{n}}^k\right) \tag{6.20}$$

Finally, using equality (6.10) one has;

$$\Sigma_{e,\mathbf{n}} - \Sigma_e = \overline{\Sigma}_{e,\mathbf{n}} - \Sigma_e + \mathcal{O}_p\left(\frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2} + h_{\mathbf{n}}^k\right). \tag{6.21}$$

To complete the proof, we will use Lemma 6.3. To this aim, it suffices to choose $\theta = \delta$ with $\delta > 2N$ then $\mathbf{E}||X||^{4+\delta} < \infty$ and $\sum_k \alpha(k)^{\frac{\delta}{\delta+4}} < \infty$; hence we have:

$$\overline{\Sigma}_{e,\mathbf{n}} - \Sigma_e = \mathcal{O}_p(\frac{1}{\hat{\mathbf{n}}}).$$

which ends the proof. $\square$

6.4. **Proof of corollary 3.2.**

The proof is achieved by replacing $h_{\mathbf{n}} \simeq \hat{\mathbf{n}}^{-c_1}$ and $e_{\mathbf{n}} \simeq \hat{\mathbf{n}}^{-c_2}$ with $\frac{c_2}{k} + \frac{1}{4k} < c_1 < \frac{1}{2} - 2c_2$ on equality (6.21) $\square$

6.5. **Proof of corrollary 3.3.**

Chosing $h_{\mathbf{n}} \simeq \hat{\mathbf{n}}^{-c_1}$ and $e_{\mathbf{n}} \simeq \hat{\mathbf{n}}^{-c_2}$ where $\frac{c_2}{k} + \frac{1}{4k} < c_1 < \frac{1}{2} - 2c_2$ on equality (6.21), one gets $\Sigma_{e,\mathbf{n}} - \Sigma_e = \overline{\Sigma}_{e,\mathbf{n}} - \Sigma_e + o_p(\frac{1}{\sqrt{\hat{\mathbf{n}}}})$ and the central limit theorem for spatial data and Slusky's theorem completes the proof.

Proof of Theorem 3.4. Let $v_{\mathbf{n}} = \left(\frac{\log\log \hat{\mathbf{n}}}{\hat{\mathbf{n}}}\right)^{\frac{1}{2}}$ note that since Y take place on a compact set, $\frac{1}{f}$ is bounded and replace the assumption $\left\|\frac{r(Y)}{f(Y)}\mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}\right\|_2 = \mathcal{O}\left(\frac{1}{\hat{\mathbf{n}}^{\frac{1}{2}+\delta}}\right)$ by $\mathbf{E}\left(\exp\left(\|r(Y)\|\ \mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}\right)\right) = \mathcal{O}\left(\hat{\mathbf{n}}^{-\xi}\right)$ for some $\xi > 0$. Then,

$$P\left(\left\|\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\frac{r(Y_{\mathbf{i}})}{f(Y_{\mathbf{i}})}\mathbf{1}_{\{f(Y_{\mathbf{i}})\le e_{\mathbf{n}}\}}\right\| > \frac{\varepsilon}{v_{\mathbf{n}}}\right) \le P\left(\frac{C}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\|r(Y_{\mathbf{i}})\|\ \mathbf{1}_{\{f(Y_{\mathbf{i}})\le e_{\mathbf{n}}\}} > \frac{\varepsilon}{v_{\mathbf{n}}}\right)$$

and because of Minskovski's inequality: for all $k \in \mathbb{N}^*$, $\mathbf{E}\left(\left(\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\|r(Y_{\mathbf{i}})\|\ \mathbf{1}_{\{f(Y_{\mathbf{i}})\le e_{\mathbf{n}}\}}\right)^k\right) \le \left\|r(Y)\,\mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}\right\|_k^k$, we can say that with using the argument $\mathbf{E}\left(\exp\left(\|r(Y)\|\ \mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}\right)\right) =$

$\mathcal{O}\left(\hat{\mathbf{n}}^{-\xi}\right)$:

$$
\begin{aligned}
P\left(\left\|\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\frac{r(Y_{\mathbf{i}})}{f(Y_{\mathbf{i}})}\mathbf{1}_{\{f(Y_{\mathbf{i}})\le e_{\mathbf{n}}\}}\right\| > \frac{\varepsilon}{v_{\mathbf{n}}}\right) &\le \mathbf{E}\left[\exp\left(\|r(Y)\|\ \mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}\right)\right]\exp\left(-\frac{\varepsilon}{v_{\mathbf{n}}}\right)\\
&\le C_1\,\hat{\mathbf{n}}^{-\xi}.\exp\left(-\varepsilon\left(\frac{\log\log\ \hat{\mathbf{n}}}{\hat{\mathbf{n}}}\right)^{-\frac{1}{2}}\right) \text{ for some } C_1 > 0.\\
&\le C_1\ \exp\left(-\xi\ \log\hat{\mathbf{n}} - \varepsilon\left(\frac{\hat{\mathbf{n}}}{\log\log\ \hat{\mathbf{n}}}\right)^{\frac{1}{2}}\right)\\
&\le C_1\ \exp\left(-\min(\xi\,,\varepsilon)\left(\log\hat{\mathbf{n}} + \frac{\sqrt{\log\hat{\mathbf{n}}}}{\sqrt{\log\log\ \hat{\mathbf{n}}}}\right)\right)\\
&\le C_1\ \exp\left(-\min(\xi\,,\varepsilon)\ \log\hat{\mathbf{n}}\ \left(1+\frac{1}{\sqrt{(\log\hat{\mathbf{n}})\log\log\ \hat{\mathbf{n}}}}\right)\right)
\end{aligned}
$$

as $\hat{\mathbf{n}} \to +\infty$, $\exp\left(-\min(\xi\,,\varepsilon)\ \log\hat{\mathbf{n}}\ \left(1+\frac{1}{\sqrt{(\log\hat{\mathbf{n}})\log\log\ \hat{\mathbf{n}}}}\right)\right) \simeq \hat{\mathbf{n}}^{-C_2}$ where c_2 is positive constant. So, $\frac{1}{\hat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\frac{r(Y_{\mathbf{i}})}{f(Y_{\mathbf{i}})}\mathbf{1}_{\{f(Y_{\mathbf{i}})\le e_{\mathbf{n}}\}} = o_{a.s}\left(\left(\frac{\log\log\ \hat{\mathbf{n}}}{\hat{\mathbf{n}}}\right)^{\frac{1}{2}}\right)$ and the proof is complet by using Lemma 6.7 and sketching the proof of Theorem 3.1 □

Proof of Corollary 3.5. If moreover we chose $h_{\mathbf{n}} \simeq \hat{\mathbf{n}}^{-c_1}$ and $e_{\mathbf{n}} \simeq \hat{\mathbf{n}}^{-c_2}$ where $\frac{c_2}{k} + \frac{1}{4k} < c_1 < \frac{1}{2} - 2c_2$, then,

$$
\sqrt{\frac{\hat{\mathbf{n}}}{\log\log\hat{\mathbf{n}}}} \times \frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2} = \sqrt{\frac{\hat{\mathbf{n}}}{\log\log\hat{\mathbf{n}}}} \times \left(\hat{\mathbf{n}}^{-2kc_1+2c_2} + \hat{\mathbf{n}}^{-1+c_1+2c_2}\log\hat{\mathbf{n}}\right)
$$

$\sqrt{\frac{\hat{\mathbf{n}}}{\log\log\hat{\mathbf{n}}}} \times \frac{\Psi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2} = \frac{\hat{\mathbf{n}}^{\frac{1}{2}-2kc_1+2c_2}}{\sqrt{\log\log\hat{\mathbf{n}}}} + \hat{\mathbf{n}}^{-\frac{1}{2}+c_1+2c_2}\log\hat{\mathbf{n}}$ this latter tend to zero as soon as $\frac{c_2}{k} + \frac{1}{4k} \le c_1 < \frac{1}{2} - 2c_2$

The proof is obtained by sketching the proof of Corolary 3.2 and using the law of the iterated logarithm recalled in Lemma 6.3.

References

[1] L. Anselin and A. Bera. *Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics*. In A. Ullah and D.E.A. Giles (eds.), Handbook of Applied Economic Statistics., New York: Marcel Dekker, 1998.

[2] L. Anselin and R. J.G. M. Florax. *New Directions in Spatial Econometrics*. Springer, Berlin, 1995.

[3] G. Arbia and G. Lafratta. Exploring nonlinear spatial dependency in the tails. *Geographical Analysis*, 37(4):423–437, 2005.

[4] G. Biau. Spatial kernel density estimation. *Mathematical Methods of Statistics*, 12:371–390, 2003.

[5] G. Biau and B. Cadre. Nonparametric spatial prediction. *Statistical Inference for Stochastic Processes*, 7:327–349, 2004.

[6] D. Bosq. Parametric rates of nonparametric estimators and predictors for continuous time processes. *The Annals of Statistics*, 25:982–1000, 1997.

[7] D. Bosq. *Nonparametric Statistics for Stochastic Processes - Estimation and Prediction - 2nd Edition*. Lecture Notes in Statistics, Springer-Verlag, New York, 1998.

[8] D. R. Brillinger. *A Generalized Linear Model with "Gaussian" regressor variables. In A Festschrift for Eric L. Lehmann (P. J. Bickel, K. A. Doksum and J. A. Hodges, eds.* Wadsworth, Belmont, C, 1983.

[9] M. Carbon, C. Francq, and L. T. Tran. Kernel regression estimation for random fields. *Journal of Statistical Planning and Inference*, 137(Issue 3):778–798, 2007.

[10] M. Carbon, L. T. Tran, and B. Wu. Kernel density estimation for random fields: The l_1 theory. *Nonparametric Statistics*, 6:157–170, 1997.

[11] N. A.C. Cressie. *Statistics for Spatial Data*. New-York, 1991.

[12] S. Dabo-Niang and A. F. Yao. Kernel regression estimation for continuous spatial processes. *To appear on Mathematical Methods for statistics*, 2007.

[13] P. Diggle. *Statistical Analysis of Spatial Point Patterns*. Oxford, 2003.

[14] P. Diggle, J.A. Tawn, and R.A. Moyeed. Model-based geostatistics (with discussion). *J. of Royal Stat. Soc., Ser. C*, 47:299–350, 1998.

[15] P. Doukhan. *Mixing - Properties and Examples*. Lecture Notes in Statistics, Springer-Verlag, New York, 1994.

[16] L. Ferré. Determination of the dimension in sir and related methods. *Journal of the American Statistical Association*, 3(441):132–140, 1998.

[17] L. Ferré and A.F. Yao. Smoothed functional inverse regression. *Statistica Sinica*, 15(3):665–683, 2005.

[18] X. Guyon. *Random Fields on a Network - Modeling, Statistics, and Applications*. Springer, New-York, 1995.

[19] M. Hallin, Z. Lu, and L. T. Tran. Local linear spatial regression. *Ann. of Stat.*, 32(6):2469–2500, 2004.

[20] W. Härdle and T. M. Stoker. Investigating smooth multiple regression by method of average derivatives. *J. Am. Statist. Ass.*, 84:986–995, 1989.

[21] T. J. Hastie and R. Tibshirani. Generalized additive models (with discussion). *Statist. Sci.*, 1:297–318, 1986.

[22] M. Hirstache, A. Judisky, J. Polzehl, and V. Spokoiny. Structure adaptative for dimension reduction. *Ann of Stat*, 29(6):1537–1566, 2001.

[23] T. Hsing. Nearest-neighbor inverse regression. *Ann. of Stat*, 27(2):697–731, 1999.

[24] K. C. Li. Sliced inverse regression for dimension reduction, with discussions. *Journal of the American Statistical Association*, 86:316–342, 1991.

[25] Z. Lu and X. Chen. Spatial kernel regression estimation: Weak concistency. *Stat and Probab. Lett*, 68:125–136, 2004.

[26] D. N. Politis and J. P. Romano. Nonparametric resampling for homogeneous strong mixing random fields. *J. of Mult. Anal*, 47(2):301–328, 1993.

[27] B. Ripley. *Spatial Statistics*. Wiley, New-York, 1981.

[28] M. Rosenblatt. *Stationnary Sequences and Random Fields*. Birkhauser, Boston, 1985.

[29] S.H Song and J. Lee. A note on s^2 in a spatially correlated error component regression for panel data. *Economics Letters*, 101:41–43, 2008.

[30] S. Sun and C. Chiang. Limiting behavior of the perturbed empirical distribution functions evaluated at u-statistics for strongly mixing sequences of random variables. *J. of Applied Mathematics and Stochastic Analylsis*, 10(1):3–20, 1997.

[31] L. T. Tran. Kernel density estimation on random fields. *Journal of Multivariate Analysis*, 34:37–53, 1990.

[32] L. T. Tran and S. Yakowitz. Nearest neighbor estimators for random fields. *Journal of Multivariate Analysis*, 44:23–46, 1993.

[33] H. Wackernagel. *Multivariate Geostatistics.* Springer-Verlag, Berlin, Heidelberg, New York, 1995. An Introduction with Applications.

[34] H-M. Wu and H. H-S. Lu. Supervised motion segmention by spatial-frequential analysis and dynamic sliced inverse regression. *Statistica Sinica*, 14:413–430, 2004.

[35] Y. Xia, H. Tong, W. Li, and L. et Zhu. An adaptive estimation of dimension reduction space. *JR Stat. Soc., Ser. B*, 64(3):363–410, 2002.

[36] H Zhang. On estimation and prediction for spatial generalized linear mixed models. *Trans. Amer. Math. Soc.*, 58(1):129–136, 2002.

[37] L.Z. Zhu and TK. Fang. Asymptotics for kernel estimate of sliced inverse regression. *Annals of Statistics*, 24:1053–1068, 1996.