KERNEL SPATIAL DENSITY ESTIMATION IN INFINITE DIMENSION SPACE

Sophie Dabo-Niang, Anne-Françoise Yao

EQUIPPE

Lille 1 I Lille 2 I Lille 3 I

Document de travail

- [2011-3]

"KERNEL SPATIAL DENSITY ESTIMATION IN INFINITE DIMENSION"

Sophie Dabo-Niang, Anne-Françoise Yao

Université Lille Nord de France
Pôle de Recherche et d'Enseignement Supérieur

Université Lille1 Sciences et Technologies

Université Lille 2 Droit et Santé

Lille 3 Université Sciences humaines et sociales Charles-de-Gaulle

KERNEL SPATIAL DENSITY ESTIMATION IN INFINITE DIMENSION SPACE

SOPHIE DABO-NIANG [♯] AND ANNE-FRANÇOISE YAO [§]

Abstract. In this paper, we propose a nonparametric estimation of the spatial density of a functional stationary random field. This later is with values in some infinite dimensional space and admitted a density with respect to some reference measure. The weak and strong consistencies of the estimator are shown and rates of convergence are given. Special attention is paid to the links between the probabilities of small balls in the concerned infinite dimensional space and the rates of convergence. The practical use and the behavior of the estimator are illustrated through some simulations and a real data application.

Key Words: Density estimation; Random fields; Functional variables; Infinite dimensional space; Small balls probabilities; mixing conditions.

1. Introduction

In many studies, the observations can be summarized as spatially dependent curves. It is the case for example with hyper-spectral images analysis, growth curves of trees distribution in a forester parcel or annual meteorological curves (temperature, precipitation, wind,...) of some region. Furthermore, in some studies it can be interesting to see spatio-temporal data as spatially dependent curves (see the example treated in Dabo-Niang et al. (2010)). For example the space-time evolution of concentration of oil (in soil science) or a pollutant on earth's or aquatic parcel (in environmental science), the observations can be summarized as curves (of evolution) located at different sites. In fact, such data occur in many fields such as geology, econometrics, epidemiology, and so on.

Looking at these data as functional spatially dependent data can provide some information as spatial distribution of the concerned curves, spatial correlation between curves or in a general way some statistics for functional spatial data.

Despite the many possible applications of such tools, there is a few literature dedicated to models that take into account both the functional and spatial dependence features of the problem that concerns us. We refer for example to Bar-Hen et al. (2008), Basse et al. (2008), Delicado et al. (141), Nerini et al. (2010) or Dabo-Niang et al. (2010).

♯ PRES Université Lille Nord de France
Université Charles De Gaulle, Lille 3
maison de la recherche, domaine du pont de bois, BP 60149, 59653 Villeneuve d'ascq cedex, France.
Laboratoire EQUIPPE EA 4018
sophie.dabo@univ-lille3.fr.
§ Corresponding author: Laboratoire LMGEM, Centre d'Océanologie de Marseille. University Aix-Marseille 2, Campus de Luminy, case 901, 13288 Marseille cedex 09, France. anne-francoise.yao@univmed.fr.

In fact, on one hand, the recent abundant literature on functional data analysis concerns either descriptive statistic for functional data (see *e.g.* Ramsay & Silverman (1997), Ramsay & Silverman (2002)) or models for *independent identically distributed (i.i.d.)* or time dependent data (Bosq (2000), Dabo-Niang et al. (2006), Ferraty & Vieu (2006),...).

On the other hand, most of spatial models goes to finite dimensional observations, as one can see in the relatively abundant parametric models literature (see for example Chilès & Delfiner (1999), Guyon (1995), Anselin & Florax (1995), Cressie (1993) or Ripley (1981)), while the nonparametric treatment of finite dimensional spatial data is limited. Only some studies on nonparametric estimation of spatial probability densities (see the key references: Tran (1990), Tran & Yakowitz (1993), Carbon et al. (1997b), **?**, Hallin et al. (2004a), Lu & Chen (1997), Carbon (2006)) or regression models (see Biau & Cadre (2004), Lu & Chen (2002), Lu & Chen (2004), Hallin et al. (2004a), Carbon et al. (2007), Dabo-Niang & Yao (2007)) have been tackled.

In this paper, we are interested in *density estimation for spatially dependent functional data.* Nevertheless, contrary to the finite dimensional setting, when dealing with *probability density,* the functional aspect of the problem requires to specify the reference measure.

Indeed, in finite dimensional setting, when talking about density of probability without specify the reference measure, μ, one implicitly deals with the Lebesgue measure. Unfortunately, Lebesgue measure does not exist for infinite dimensional spaces. Thus, defining μ is a crucial problem in density estimation for functional data (particularly for functional spatially dependent data).

In this work, we present some ideas for choosing μ and illustrate our purpose by some applications (Section 5) where μ is specified. But, in the theoretical study, a part from some hypothesis, we will not specify μ because the behavior of our estimator is the same for all measure, μ, σ-finite and such that $0 < \mu(A) < \infty$, for any open ball $A \subset \mathcal{E}$ and we suppose that all the reference measures used in this paper belongs to this class of measures.

We note that, a similar work has been recently done in Dabo-Niang et al. (2010) which also treats some applications where specification of μ is not necessary.

About the spatial aspect of the problem, it requires the definition of some notions such as the set-index and the dependence measure.

In finite dimensional setting, nonparametric spatial density estimation models are constructed on the rectangular region $\mathcal{I}_{\mathbf{n}} = \left\{\mathbf{i} = (i_1, ..., i_N) \in (\mathbb{N}^*)^N \ , \ 1 \leq i_k \leq n_k, \, k = 1, ..., N\right\}$, $\mathbf{n} = (n_1, ..., n_N) \in (\mathbb{N}^*)^N$ (or more generally, a lattice in $\mathbb{R}^N$). These models appear as natural extensions of the nonparametric estimation of the probability density for time dependent random variables, using various nonparametric approaches. We refer to Robinson (1983, 1987), Masry & Györfi (1987), Yakowitz (1987), Roussas (1988), Bosq (1998), Györfi et al. (1990), Truong & Stone (1992), Tran (1990), Lu (1996) and Lu & Chen (1997) among others.

Our aim is to extend the results of Bosq (1998) to infinite dimensional spatial dependent variables (we recall that Bosq (1998) deals with kernel density estimation for finite dimensional time dependent processes). But, this generalization is far from being trivial.

Indeed, such a generalization deals with many theoretical and practical difficulties essentially due to the lack of canonical ordering for $N > 1$. As raised by Hallin et al. (2004b), this gap leads to re-defined two important notions: *asymptotic notion* and, since we deal with index-dependent data, the notion of *neighborhood* and the related notion of *dependent measure*.

In fact, when $N > 1$ the simple idea of a sample size going to infinity has to be defined. To keep the analogy with the case $N = 1$, in nonparametric spatial modeling, one often suppose; without the loss of generality; that the "sample size" is the rectangular region $\mathcal{I}_{\mathbf{n}}$. But, as we will see later on, the below theoretical results remain valid if $\mathcal{I}_{\mathbf{n}}$ is replaced by any subset of a lattice of $\mathbb{R}^N$.

We will measure the dependency in *strong mixing* meaning; this notion will be recalled on Section 2.

The rest of the paper is organized as follows. In Section 2, devoted to the theoretical framework, we define the spatial kernel density estimator and give the assumptions required to study the asymptotic behavior of our estimator, developed in Section 4. In Section 3, we present our estimator in the special setting of *local weighting* and *small ball probabilities*.

Section 4 is devoted to convergence in probability and strong convergence of the kernel density estimate, under various types of asymptotic and mixing assumptions.

In Section 5, we deal with the practical use of our estimator; we suggest a new method to visualize the density estimator even the random field is with values in an infinite dimension space. We illustrate the behavior of our estimator through some simulations and also a real data application. The dataset is composed of spatial grain size curves observed in a French lagoon called: *Berre Lagoon*. Section 6 is devoted to discussions and conclusions. Proofs and technical preliminary lemmas are given in the Appendix.

2. Theoretical framework

We deal with a measurable stationary spatial process $\left(X_{\mathbf{i}},\ \mathbf{i} \in (\mathbb{N}^*)^N\right),\ N \geq 1$, defined on a probability space $(\Omega,\ \mathcal{A}, \mathbf{P})$ such that the $X_{\mathbf{i}}'$s have the same distribution as a variable X valued in an infinite dimensional separable normed space $(\mathcal{E},\ \|.\|)$ ($\|.\|$ is the norm). We assume that X has an unknown density f with respect to some given measure μ which is σ-finite such that $0 < \mu(A) < \infty$, for any open ball $A \subset \mathcal{E}$. All the theoretical properties studied in this paper concern only this class of measure. So, we will give our results without specifying the reference measure μ (as it is done for example in Ferraty & Vieu (2006) and references therein for $N = 1$). One can choose as reference measure a Wiener measure for example in the case of diffusion process as in Dabo-Niang (2004) or another Gaussian measure, see Section 3.2.

In practice, we suggest the choice of an appropriate reference measure according to the problem considered. For example, if the study concerns spatial distribution of curves (temperature, or grain size curves,...) observed for K ($K > 1$) populations, on can choose as reference measure, the distribution of one of the populations. Of course, if necessary, μ can be Gaussian measure or any known measure over infinite dimensional spaces.

In the following, for any $x \in \mathcal{E}$, $B(x,\, t)$ will denote the opened ball centered at x with radius $t > 0$, the capital letter C (resp. C_x) will denote an arbitrary constant (resp. constant depending on x) that may vary from line to line.

Let us now define our estimator.

2.1. **The spatial kernel estimator.** We aim to estimate the spatial density from data, $X_{\mathbf{i}}$'s, observed on a region $\mathcal{O}_{\mathbf{n}}$. In the following, without the loss of generality, we will often suppose that $\mathcal{O}_{\mathbf{n}}$ is the rectangular region, $\mathcal{I}_{\mathbf{n}}$, as previously defined. Before going further, let us consider some notations.

We will denote by: $\mathbf{i} = (i_1, ..., i_N)$ (in bold) a point in $\mathbb{N}^N$, that will be referred to as a site. We will write $\mathbf{n} \to +\infty$ if $\min_{k=1,...,N} n_k \to +\infty$. This means that the rectangular region does not expand to infinity at the same rate along all directions. Such an expansion is called *non-isotropic* divergence (see *e.g.* Lu & Chen (2004), Hallin et al. (2004b)). However, in some problems the assumption that this rate is the same in all directions is called *isotropic* divergence: that is $\mathbf{n} \to +\infty$ and $\left|\frac{n_j}{n_k}\right| \leq C$ for some constant $0 < C < \infty$ and $\forall\, j, k \in \{1, ..., N\}$ (see **?** or Hallin et al. (2004b)).

In the following, we consider the less restrictive non-isotropic case. But the proofs of results obtained here are similar in the isotropic case.

We will set $\widehat{\mathbf{n}} = n_1 \times ... \times n_N$ the sample size and denote by $\hat{\mathbf{n}}_{\mathcal{V}} = \mathbf{Card}(\mathcal{V})$ the cardinality of any set $\mathcal{V}$.

The *kernel spatial density* estimator for *functional data* based on $(X_{\mathbf{i}},\ \mathbf{i} \in \mathcal{I}_{\mathbf{n}})$ is defined by:

$$f_{\mathbf{n}}(x) = \frac{1}{\widehat{\mathbf{n}} a_{\mathbf{n}}^x} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K_{\mathbf{n}}\left(\|X_{\mathbf{i}} - x\|\right), \qquad x \in \mathcal{E},$$

where $(a_{\mathbf{n}}^x)$ is a sequence of normalization with $\lim_{\mathbf{n} \to +\infty} a_{\mathbf{n}}^x = 0\,(+)$ and $K_{\mathbf{n}}(.)$ is a function defined from $\mathbb{R}^+$ to $\mathbb{R}$ which depends on a kernel. Discussion on some particular cases of $a_{\mathbf{n}}^x$ and $K_{\mathbf{n}}$ will be done later (see Section 3).

Remark 2.1. Note that, since we are in a spatial setting, even the density f or its estimator $f_{\mathbf{n}}$ must be defined at any $x_{\mathbf{j}} \in \mathcal{E}$ and $\mathbf{j} \in \mathbb{Z}^N$. But, for ease of reading the paper, we will often use, without the loss of generality, the generic notations $f(x)$ (*resp.* $f_{\mathbf{n}}(x)$) instead of $f(x_{\mathbf{j}})$ (*resp.* $f_{\mathbf{n}}(x_{\mathbf{j}})$).

In order to shorten the paper, we only consider the uniform consistency of $f_{\mathbf{n}}$ over a set $\mathcal{G}$. The pointwise convergence of $f_{\mathbf{n}}$ to f is easily deduced with weaker assumptions.

Uniform consistency results are particularly useful to study the kernel estimation of the mode of the density f over $\mathcal{G}$ defined by:

$$\mathbf{x_{mode}} = \arg\sup_{\mathcal{G}} f \tag{2.1}$$

which is an interesting centrality statistics that allows to analyze a dataset when the observations are valued in a finite or infinite dimension space. This modal curve can be estimated by the following:

$$\widehat{\mathbf{x}}_{\mathbf{mode}} = \arg\sup_{\mathcal{G}} f_{\mathbf{n}}. \tag{2.2}$$

As in the finite-dimensional case, the density can have several spatial modal curves (In the case $N = 2$, this can be seen through a graphical representation of $f_{\mathbf{n}}$). Then, for a given spatial modal curve, it is interesting to control the size of the set in which one looks for the concerned spatial modal curve (this remains interesting even if there is only one modal curve).
This being, we are interested in some set $\mathcal{G}$ such that $\mathcal{G} \subset \mathcal{G}_{\mathbf{n}} := \bigcup_{j=1}^{d_{\mathbf{n}}} B(x_{\mathbf{i}_j}, r_{\mathbf{n}})$, where $d_{\mathbf{n}}$ is some integer and for $j = 1, ..., d_{\mathbf{n}}$, $B(x_{\mathbf{i}_j}, r_{\mathbf{n}})$ is the opened ball of center $x_{\mathbf{i}_j} \in \mathcal{E}$ and radius $r_{\mathbf{n}} > 0$. For sake of simplicity, in the following we will set $x_{\mathbf{i}_j} = x_j$. Note that such a set, $\mathcal{G}$, can always be built: it suffices to take a finite number of observed curves: $x_1, ..., x_{d_{\mathbf{n}}}$ and construct a set of open balls $B(x_j, r_{\mathbf{n}})$, $j = 1, ..., d_{\mathbf{n}}$ with r_n (verifying Condition **H5** below) such that $\mathcal{G}_{\mathbf{n}} = \bigcup_{j=1}^{d_{\mathbf{n}}} B(x_j, r_{\mathbf{n}})$ covers the whole set of observations of interest.

Similarly to the finite dimensional case (see for example Carbon et al. (1997b)), the study of the asymptotic behavior of $f_{\mathbf{n}}$ requires some assumptions on the kernel and some mixing conditions which are respectively given in Sections 2.2 and 2.4 and the following regularity conditions on f:

$$\mathbf{H_f} - f \text{ is uniformly continuous on } \mathcal{G} \text{ and } \sup_{x \in \mathcal{G}} |f(x)| < \infty.$$

Note that this later condition is the infinite dimensional counterpart of the condition used in the finite dimensional setting (see for example Bosq (1998), Carbon et al. (1997b)).

2.2. **Assumptions on the kernel.** In order to get convergence results (without rates), we assume that the function $K_{\mathbf{n}}$ verifies the following conditions. Particular cases of these conditions will be discussed in Section 3.

$$\mathbf{H1} \text{ - } \forall \delta, 0 < \delta \leq +\infty, \ \lim_{\mathbf{n} \to \infty} \sup_{x \in \mathcal{G}} \left| \frac{1}{a_{\mathbf{n}}^x} \int_{\|y-x\|<\delta} K_{\mathbf{n}}(\|y - x\|) \, d\mu(y) - 1 \right| = 0.$$

H2 - For some constant $C > 0$, we have:

$$\sup_{x\in\mathcal{G},\, y\in\mathcal{E}} \frac{K_{\mathbf{n}}\left(\|y-x\|\right)}{a_{\mathbf{n}}^{x}} \leq CS_{\mathbf{n}} < \infty,$$

where $S_{\mathbf{n}}$ is a sequence of positive numbers satisfying

$$\lim_{\mathbf{n}\to+\infty} S_{\mathbf{n}} = +\infty, \quad \text{and} \quad \lim_{\mathbf{n}\to+\infty} \frac{\widehat{\mathbf{n}}}{S_{\mathbf{n}}\log\widehat{\mathbf{n}}} = +\infty.$$

H3 - $\exists\beta_1 > 0,\ \exists\beta_2 > 0, \forall x_1,\, x_2 \in \overset{o}{\mathcal{G}},\, y \in \mathcal{E}$ ($\overset{o}{\mathcal{G}}$ is the interior of $\mathcal{G}$).

$$\left|\frac{1}{a_{\mathbf{n}}^{x_1}}K_{\mathbf{n}}\left(\|y-x_1\|\right) - \frac{1}{a_{\mathbf{n}}^{x_2}}K_{\mathbf{n}}\left(\|y-x_2\|\right)\right| \leq CS_{\mathbf{n}}^{\beta_2}\left\|x_1-x_2\right\|^{\beta_1}.$$

H4 - For any $\delta > 0$:

$$\lim_{\mathbf{n}\to\infty} \sup_{(x,u)\in\mathcal{G}\times\{u/\|u\|>\delta\}} \frac{1}{a_{\mathbf{n}}^{x}}\left\|u\right\| K_{\mathbf{n}}\left(\|u\|\right) = 0.$$

These conditions are spatial counterpart of those used for the *i.i.d.* functional variables, see Dabo-Niang et al. (2006) but as we will see later, the following hypothesis is a generalization of the one used in the finite dimensional case.

H5 - $d_{\mathbf{n}} = \widehat{\mathbf{n}}^{\beta}$ and $r_{\mathbf{n}} \leq \left((S_{\mathbf{n}})^{\kappa}\log\widehat{\mathbf{n}}/\widehat{\mathbf{n}}\right)^{1/2\beta_1}$ with β an integer and $\kappa \leq -2\beta_2 + 1$.

2.3. **Assumptions to get rates of convergence.** To get rates of convergence, we will need a more restrictive condition on the regularity of f:

$$\mathbf{H}'_{\mathbf{f}} - \quad f \text{ satisfies the Lipshitz condition } : \forall x, y \in \mathcal{G}\ \ |f(x) - f(y)| \leq \|x-y\|.$$

We will replace Assumption **H1** by:

$$\mathbf{H6} - \forall\delta, 0 < \delta \leq +\infty,\ \sup_{x\in\mathcal{G}}\left|\frac{1}{a_n^x}\int_{\|y-x\|<\delta} K_n\left(\|y-x\|\right)d\mu(y) - 1\right| = o\left(D_{\mathbf{n}}\right) \text{ where}$$

$D_{\mathbf{n}} = \sup_{(x,y)\in supp(K_{\mathbf{n}})^2} \|x-y\|$ is the diameter of the support of $K_{\mathbf{n}}$, which is such that $D_{\mathbf{n}} = o(1)$.

2.4. **Dependency conditions.** As it often occurs in spatial dependent data analysis, we need to define the type of dependence. Here, we consider the following two dependence measures:

2.4.1. *Local dependence conditions.* We will assume that the joint probability density $f_{\mathbf{i},\mathbf{j}}\left(x,y\right)$ of $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$ with respect to $\mu\times\mu$ exits and satisfies

$$\left|f_{\mathbf{i},\mathbf{j}}\left(x,y\right) - f\left(x\right)f\left(y\right)\right| \leq C, \tag{2.3}$$

for some constant C and for all $x,\, y \in \mathcal{E}$ and $\mathbf{i},\ \mathbf{j} \in \mathbb{N}^N$, $\mathbf{i} \neq \mathbf{j}$.

In fact, this local dependency condition can be replaced by the condition: for all $\mathbf{i}$, $\mathbf{j} \in \mathbb{N}^N$ the joint probability distribution $\nu_{\mathbf{i},\mathbf{j}}$ of $(X_{\mathbf{i}},\ X_{\mathbf{j}})$ satisfies

$$\exists \epsilon_1 \in (0,1],\ \nu_{\mathbf{i},\mathbf{j}}\left(B(x,h_{\mathbf{n}}) \times B(x,h_{\mathbf{n}})\right) = \left(F^x(h_{\mathbf{n}})\right)^{1+\epsilon_1}, \tag{2.4}$$

where $F^x(h_{\mathbf{n}}) = P(X \in B(x,h_{\mathbf{n}}))$.

Such local dependency condition is necessary to reach the same rate of convergence as in the *i.i.d.* case.

2.4.2. *Mixing conditions.* Another complementary dependency condition concerns the *mixing condition* which measures the dependency by means of α-mixing. We assume that $\left(X_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N\right)$ satisfies the following mixing condition: there exists a function $\varphi(t) \downarrow 0$ as $t \to \infty$, such that for E, $E^{'}$ subsets of $\mathbb{N}^N$ with finite cardinals,

$$\alpha\left(\mathcal{B}(E), \mathcal{B}\left(E^{'}\right)\right) = \sup_{B \in \mathcal{B}(E), C \in \mathcal{B}\left(E^{'}\right)} \left|\mathbf{P}(B \cap C) - \mathbf{P}(B)\,\mathbf{P}(C)\right|$$

$$\leq \chi\left(\mathrm{Card}(E), \mathrm{Card}\left(E^{'}\right)\right) \varphi\left(\mathrm{dist}\left(E, E^{'}\right)\right), \tag{2.5}$$

where $\mathcal{B}(E)(resp.\ \mathcal{B}\left(E^{'}\right))$ denotes the Borel σ-field generated by $(X_{\mathbf{i}},\ \mathbf{i} \in E)$ $(resp.\ \left(X_{\mathbf{i}},\ \mathbf{i} \in E^{'}\right))$, $\mathrm{Card}(E)$ $(resp.\ \mathrm{Card}\left(E^{'}\right))$ the cardinality of E $(resp.\ E^{'})$, $\mathrm{dist}\left(E,\ E^{'}\right)$ the Euclidean distance between E and $E^{'}$ and $\chi : \mathbb{N}^2 \to \mathbb{R}^+$ is a nondecreasing symmetric positive function in each variable. Throughout the paper, it will be assumed that χ satisfies either

$$\chi(n,m) \leq C \min(n,m), \quad \forall n, m \in \mathbb{N} \tag{2.6}$$

or

$$\chi(n,m) \leq C\,(n+m+1)^{\widetilde{\beta}}, \quad \forall n, m \in \mathbb{N} \tag{2.7}$$

for some $\widetilde{\beta} \geq 1$ and some $C > 0$. If $\chi \equiv 1$, then $(X_{\mathbf{i}})$ is said to strongly mixing. Many stochastic processes (among the various useful time series models) satisfy strong mixing properties, which are relatively easy to check. Conditions (2.6)-(2.7) are weaker than the strong mixing condition and have been used for finite dimensional variables see for example Tran (1990), Carbon et al. (1997a,b) and Biau & Cadre (2004). We refer to Doukhan (1994) and Rio (2000) for discussion on mixing and examples.

Concerning the function $\varphi(.)$, as it is often done, two kind of conditions will be assumed: the case where $\varphi(i)$ tends to zero at a polynomial rate, ie.

$$\varphi(i) \leq C i^{-\theta}, \text{ for some } \theta > 0 \tag{2.8}$$

or the case where $\varphi(i)$ tends to zero at an exponential rate: i.e $\varphi(i) = C \exp(-si)$ for some $s > 0$. There is a link between these two dependence conditions (local and mixing), see for example Bosq (1998) for details.

3. Estimation in local weighting and small ball probabilities setting.

To fix idea, we consider in this Section, some cases that illustrate that our estimator and the assumptions on the function $K_{\mathbf{n}}$ (previously mentioned) are less restrictive than they seem.

3.1. **The kernel density estimator and concentration conditions.**

This section deals with the well-known situation where the density estimator is based both on a kernel function and on a smoothing parameter. Namely, we look at some special case where the probability distribution of X satisfies some concentration condition as in the *i.i.d.* case suggested for example by Dabo-Niang et al. (2006) and Ferraty & Vieu (2006). Then, the density estimate is of the simple usual Parzen-Rosenblatt form, studied for example by Carbon et al. (1997a,b) in the finite dimensional spatial setting. This special case allows clearer interpretation of our asymptotic results (see Section 4.3) by linking them directly with the small balls probabilities. It will be also shown that the technical assumptions introduced in Section 2.2 are not really restrictive.

In this setting, the previous estimator can be rewritten as follows:

$$\forall x \in \mathcal{E}, \tilde{f}_{\mathbf{n}}(x) = \frac{1}{\widehat{\mathbf{n}}\, a_{\mathbf{n}}^{x}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K\left(\|X_{\mathbf{i}} - x\| / h_{\mathbf{n}}\right), \tag{3.1}$$

where $(h_{\mathbf{n}})$ is a sequence of positive reals (the bandwidths), K is an integrable kernel and we set $K_{\mathbf{n}}(u) = K(u/h_{\mathbf{n}})$.

Concerning the sequence $(a_{\mathbf{n}}^{x})$, we will consider the case where $a_{\mathbf{n}}^{x}$ is a function of $F^{x}(h_{\mathbf{n}})$ or $\mathbf{E}\left[K\left(\frac{\|Y-x\|}{h_{\mathbf{n}}}\right)\right]$ if the reference measure is the distribution of a stationary random field $(Y_{\mathbf{i}})$. These cases are similar to those proposed for example by respectively Dabo-Niang (2004) and Ferraty & Vieu (2006) on one hand and by Ferraty et al. (2007) on the other hand. The two later references concern regression estimation with functional variables. Other different sequences $(a_{\mathbf{n}}^{x})$ have been considered in the literature: for example Dabo-Niang et al. (2006) considered $a_{\mathbf{n}}^{x} = \mathbf{E}\left[K\left(\frac{\|X-x\|}{h_{\mathbf{n}}}\right)\right] - o\left(g(h_n)\right)$ where g is a positive valued function and Basse et al. (2008) suggested $a_{\mathbf{n}}^{x} = \mu(B(x,\, h_{\mathbf{n}}))$ in some particular kernel case.

As we will see in Section 4.2, the asymptotic results obtained for the more general setting (of Section 2.1) in the first part of Section 4 can be reformulated for small balls probabilities under the following concentration conditions.

Let $\psi(.)$ be some increasing function taking values in $]0, +\infty[$ such that $\lim_{t\to 0} \psi(t) = 0$. We consider the following concentration hypothesis:

H7 - $\limsup_{t\to 0} \sup_{x\in\mathcal{G}} \left| \frac{P(X \in B(x,t))}{\psi(t)} - f(x) \right| = 0$

and

$$\exists \epsilon_1 \in (0,1], \lim_{t\to 0}\sup_{x\in\mathcal{G}}\left|\frac{\max_{\mathbf{i}\neq\mathbf{j}} P((X_{\mathbf{i}},\, X_{\mathbf{j}})\in B(x,t)\times B(x,t))}{\psi(t)^{1+\epsilon_1}} - \sup_{\mathbf{i},\mathbf{j}} f_{\mathbf{i},\mathbf{j}}(x,x)\right| = 0.$$

Remark 3.1.

Note that in particular, $\psi(t)$ can be seen as $\psi(t) = \mu(B(0,t))$ if for all $x \in \mathcal{G}$, $\lim_{t\to 0}\mu(B(0,t)) = \lim_{t\to 0}\mu(B(x,t))$. Then, $f(x) = \lim_{t\to 0}\frac{P(X\in B(x,t))}{\psi(t)}$, see for example Ferraty & Vieu (2006) (p. 139) for more details which merge the definition of the classical density.

For sake of simplicity, we consider in this particular case, only kernels satisfying the following classical assumptions (even if Lemma 7.4 in the Appendix takes into account a large family of kernels):

H8- K is such that: $supp(K) = (0,1)$, $K(1) = 0$, its derivative function K' exists and $-\infty < \tau_1 \leq K' \leq \tau_2 < 0$.

The two following assumptions on the bandwidth and the concentration function will be needed:

H9 - $\exists c > 0$, $\exists \varepsilon_0 > 0$, $\forall \varepsilon < \varepsilon_0$, $\displaystyle\int_0^{\varepsilon} \psi(z)\, dz > c\varepsilon\psi(\varepsilon)$.

H10 - The smoothing parameter $h_{\mathbf{n}}$ satisfies

$$\lim_{\mathbf{n}\to+\infty} h_{\mathbf{n}} = 0 \qquad \text{and} \qquad \lim_{\mathbf{n}\to+\infty}\frac{\widehat{\mathbf{n}}\psi(h_{\mathbf{n}})}{\log\widehat{\mathbf{n}}} = +\infty.$$

The first part of Assumption **H7** and **H10** allow us to choose $a^x_{\mathbf{n}} = \psi(h_{\mathbf{n}}) \simeq \frac{F^x(h_{\mathbf{n}})}{f(x)}$. Under the conditions $\mathbf{H7}-\mathbf{H8}$, **H10** and using the same arguments as Dabo-Niang et al. (2006) and Ferraty & Vieu (2006) (p. 44 and p. 139), one gets $\mathbf{E}\left[K\left(\frac{\|X-x\|}{h_{\mathbf{n}}}\right)\right] = -\displaystyle\int_0^1 K'(t)F^x(h_{\mathbf{n}}t)\, dt \simeq -f(x).\int_0^1 K'(t)\psi(h_{\mathbf{n}}t)\, dt$. This motivates us to choose $a^x_{\mathbf{n}} = -\displaystyle\int_0^1 K'(t)\psi(h_{\mathbf{n}}t) \simeq \frac{\mathbf{E}\left[K\left(\frac{\|X-x\|}{h_{\mathbf{n}}}\right)\right]}{f(x)}$. The choice $\psi(h_{\mathbf{n}})$ or $-\displaystyle\int_0^1 K'(t)\psi(h_{\mathbf{n}}t)$ for $a^x_{\mathbf{n}}$ does not depend on x and is considered in what follows for the particular estimate $\tilde{f}_{\mathbf{n}}$.

It is easy to see that **H8** and **H9** allow to satisfy **H3** with $S_{\mathbf{n}} = 1/\psi(h_{\mathbf{n}})$, $\beta_1 = 1$ and $\beta_2 > 1$ and **H6** with $D_{\mathbf{n}} = h_{\mathbf{n}}$. Assumption **H5** is then satisfied by the following:

H11 - $d_{\mathbf{n}} = \widehat{\mathbf{n}}^{\beta}$ and $r_{\mathbf{n}} \leq \left(\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}(\psi(h_{\mathbf{n}}))^{\kappa}}\right)^{\frac{1}{2}}$, where β is an integer $\kappa \leq -2\beta_2 + 1$.

To get a rate of convergence, we use the following more restrictive assumption about the concentration of the probability of X:

H12 - $\lim_{t\to 0}\sup_{x\in\mathcal{G}}\left|\frac{P(X\in B(x,t))}{\psi(t)}-f(x)\right|=o(t)\,.$

and

$$\exists\epsilon_1\in(0,1],\ \lim_{t\to 0}\sup_{x\in\mathcal{G}}\left|\frac{\max_{\mathbf{i}\neq\mathbf{j}}P((X_{\mathbf{i}},\,X_{\mathbf{j}})\in B(x,t)\times B(x,t))}{\psi(t)^{1+\epsilon_1}}-\sup_{\mathbf{i},\mathbf{j}}f_{\mathbf{i},\mathbf{j}}(x,x)\right|=o(t)\,.$$

Remark 3.2.

- Note that Conditions **H7-H10** imply **H1-H4** (the proof of this implication is easily obtained by sketching the proof of Theorem 3 of Dabo-Niang et al. (2006)).
- Assumptions **H1-H5** are verified in finite dimensional setting, see the next section.

Before going further, let us give some particular examples of random variables that are interested in the setting of small ball probabilities that are generally difficult to evaluate.

3.2. **Examples.**

3.2.1. *The class of fractal (or geometric) processes.*

A special case of particular importance concerns the fractal processes for which, for all positive real t,

$$F^x(t)\simeq C_x t^{\gamma} \tag{3.2}$$

where γ (called the fractal dimension) and C_x are positive constants. In this case, the value of $a_{\mathbf{n}}^x$ can be precise: $a_{\mathbf{n}}^x\simeq C_x h_{\mathbf{n}}^{\gamma}$.

3.2.2. *Case of a process with density respectively to Wiener measure.*

Let μ be the standard Wiener measure. That is, X is a process whose probability distribution is absolutely continuous with respect to the standard Wiener measure (for example diffusion processes, fractional Brownian processes, ..., see for example Bogachev (1999), Li & man Shao (2001)) and $\mathcal{E}=C[0,1]$ is the space of all continuous functions on $[0,1]$ that vanish at 0. If for simplicity, K is the indicator function and $a_{\mathbf{n}}^x=\mu(B(x,h_{\mathbf{n}}))$, the consistency of the density estimator $\tilde{f}_{\mathbf{n}}$, is related to Wiener measure of the ball $B(x,h_{\mathbf{n}})$. The Wiener measure on a ball can be evaluated if the center of the ball lies in the subspace (see Csaki (1980))

$$\mathcal{S}=\{x\in C[0,1]: x(0)=0, x \text{ is absolutely continuous and } \int_0^1 x'^2(t)dt<\infty\}.$$

That is, for $x\in\mathcal{S}$ and small $h_{\mathbf{n}}$

$$\mu(B(x,h_{\mathbf{n}}))\simeq C_x\exp(-\pi^2/(8h_{\mathbf{n}}^2)).$$

We refer to Basse et al. (2008), Dabo-Niang & Rhomari (2009), Ferraty & Vieu (2006), for more discussions on these two types of processes.

3.2.3. *The finite dimensional case.*
In the case where $\mathcal{E} = \mathbb{R}^d$, as raised in Section 4 of Ferraty et al. (2007), any random variable X on $\mathcal{E}$ which has a finite and non zero density function at a point x satisfies (3.2) with $\gamma = d$. Then, $a_{\mathbf{n}}^x = h_{\mathbf{n}}^d \simeq \frac{C_x}{f(x)}.h_{\mathbf{n}}^d = V(d)h_{\mathbf{n}}^d$ and does not depend on x (where $C_x = V(d).f(x)$, $V(d)$ is the volume of the unit ball on $\mathbb{R}^d$).

Since our results appear as generalization of those of Carbon et al. (1997b), let us show that our assumptions are satisfied under that of Carbon et al. (1997b) if $\mathcal{E} = \mathbb{R}^d$.

Clearly $\mathbf{H_f}$ and $\mathbf{H'_f}$ are infinite dimension versions of the classical regularity Assumption 2 (f is a Lipschitz function) of Carbon et al. (1997b).
Note that in the multivariate setting, the kernel is defined as $K(\|u\|/h_{\mathbf{n}}) = K^*(u/h_{\mathbf{n}})$ (where K^* is a multivariate kernel), $\psi(h_{\mathbf{n}}) = h_{\mathbf{n}}^d$ and $\mathcal{G}$ is a compact set, see Carbon et al. (1997b).
Under Assumption 1 (K^* is a Lipschitz density then bounded) of Carbon et al. (1997b) we have:

- K^* verifies **H1** since it is a density.
- Assumption **H2** is satisfied since $\sup_{x\in\mathbb{R}^d} K^*(x) < C$, for some constant $C > 0$.

  In fact, we have $\frac{K_{\mathbf{n}}(\|y-x\|)}{a_{\mathbf{n}}^x} = \frac{K^*\left(\frac{y-x}{h_{\mathbf{n}}}\right)}{h_{\mathbf{n}}^d} \leq \frac{C}{h_{\mathbf{n}}^d}$. Furthermore, Carbon et al. (1997b) supposed that:

$$\lim_{\mathbf{n}\to+\infty} h_{\mathbf{n}} = 0 \quad \text{and} \quad \lim_{\mathbf{n}\to+\infty} \frac{\widehat{\mathbf{n}} h_{\mathbf{n}}^d}{\log \widehat{\mathbf{n}}} = +\infty$$

- One easily gets **H3** by setting $\beta_1 = 1, \quad \beta_2 = 1 + \frac{1}{d}$, since K^* is such that:

$$\left| \frac{1}{a_{\mathbf{n}}^{x_1}} K_{\mathbf{n}}(\|y - x_1\|) - \frac{1}{a_{\mathbf{n}}^{x_2}} K_{\mathbf{n}}(\|y - x_2\|) \right| \leq C\, h_{\mathbf{n}}^{-(d+1)} \|x_1 - x_2\|.$$

Assumption **H4** is satisfies by a kernel on $\mathbb{R}^d$ such that: $\lim_{\|x\|\to\infty} \|x\| K^*(x) = 0$.
Assumption **H5** corresponds to condition (5.1) of Carbon et al. (1997a): with $r_{\mathbf{n}} = Ch_{\mathbf{n}}^{(d+1)} \left(\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} h_{\mathbf{n}}^d}\right)^{1/2} = \left(\left(\frac{1}{h_{\mathbf{n}}^d}\right)^{-(1+2/d)} \frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}\right)^{1/2}$ (since $\beta_2 = 1 + \frac{1}{d}$) .

4. Asymptotic Results

This Section is devoted to consistency results of the density estimator under the general assumptions and also for the particular case of Section 3.1. We are particularly interested in the uniform consistency of f over $\mathcal{G}$.

4.1. **Convergence under polynomial mixing condition.** Let us first consider the case where $\varphi(i)$ tends to zero at a polynomial rate i.e. (2.8) is satisfied. Then, we get the following consistency results for different values of θ (depending on the mixing conditions (2.6) and (2.7)).

4.1.1. *Weak consistency.*

If condition (2.6) is satisfied, θ is such that $\theta > 2N(\beta+1)$ (β is the one of assumption **H5**) and $\widehat{\mathbf{n}} S_{\mathbf{n}}^{\theta_1} (\log \widehat{\mathbf{n}})^{\theta_2} \to \infty$ with,

$$\theta_1 = \frac{\theta}{2N(\beta+1)-\theta}, \theta_2 = \frac{\theta - 2N}{2N(\beta+1)-\theta},$$

then, we have:

Theorem 4.1. *Under the conditions* $\mathbf{H_f}$ *and* **H1-H5**, *(2.6) and (2.8)*

$$\sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - f(x)| \text{ converges in probability to } 0. \tag{4.1}$$

On the other hand, if the mixing coefficient satisfies condition (2.7) and if θ is such that $\theta > N(1+2\beta+2\widetilde{\beta})$ and if $\widehat{\mathbf{n}} S_{\mathbf{n}}^{\theta_3} (\log \widehat{\mathbf{n}})^{\theta_4} \to \infty$ with

$$\theta_3 = \frac{\theta + N}{N(1+2\beta+2\widetilde{\beta})-\theta}, \theta_4 = \frac{\theta - N}{N(1+2\beta+2\widetilde{\beta})-\theta},$$

then we have:

Theorem 4.2. *Under the conditions* $\mathbf{H_f}$ *and* **H1-H5**, *(2.7) and (2.8)*

$$\sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - f(x)| \text{ converges in probability to } 0. \tag{4.2}$$

To get rates of convergence, we replace Assumption **H1** by **H6** and $\mathbf{H_f}$ by $\mathbf{H'_f}$ and get the following result:

Theorem 4.3. *Under the conditions* $\mathbf{H'_f}$, **H2-H6**, *if the mixing coefficient is such that (2.6) is satisfied,* $\theta > 2N(\beta+1)$ *and* $\widehat{\mathbf{n}} S_{\mathbf{n}}^{\theta_1} (\log \widehat{\mathbf{n}})^{\theta_2} \to \infty$, *and (2.8), then we have:*

$$\sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - f(x)| = \mathcal{O}_p\left(D_{\mathbf{n}} + \sqrt{\frac{S_{\mathbf{n}} \log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}} \right). \tag{4.3}$$

The following result which proof is similar to that of Theorem 4.3 and then omitted, concerns the mixing condition (2.7).

Theorem 4.4. *Under the conditions* $\mathbf{H'_f}$, **H2-H6**, *if the mixing coefficient is such that (2.7) is satisfied and* $\theta > N(1+2\beta+2\widetilde{\beta})$ *and if* $\widehat{\mathbf{n}} S_{\mathbf{n}}^{\theta_3} (\log \widehat{\mathbf{n}})^{\theta_4} \to \infty$, *(2.8), we have*

$$\sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - f(x)| = \mathcal{O}_p\left(D_{\mathbf{n}} + \sqrt{\frac{S_{\mathbf{n}} \log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}} \right). \tag{4.4}$$

4.1.2. *Strong consistency.* The next results give the strong convergence of $f_{\mathbf{n}}$ under additional conditions.

Let $g(\mathbf{n}) = \prod_{i=1}^{N}(\log n_i)(\log\log n_i)^{1+\epsilon}$ (then $\sum_{\mathbf{n}\in\mathbb{N}^N} 1/\left(\widehat{\mathbf{n}}g(\mathbf{n})\right) < \infty$). Let θ be such that $\theta > 2N(\beta+1)$ and $\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_1^*}(\log\widehat{\mathbf{n}})^{\theta_2^*}\left(g(\mathbf{n})\right)^{\frac{-2N}{\theta-2N(\beta+2)}} \to \infty$ with

$$\theta_1^* = \frac{\theta}{2N(\beta+2)-\theta},\ \theta_2^* = \frac{\theta-2N}{2N(\beta+2)-\theta},$$

$$\theta_3^* = \frac{\theta+N}{N(3+2\beta+2\widetilde{\beta})-\theta},\ \theta_4^* = \frac{\theta-N}{N(3+2\beta+2\widetilde{\beta})-\theta},$$

then, we have the following strong consistency result:

Theorem 4.5. *Under the conditions* $\mathbf{H_f}$, **H1-H5**, *and if (2.6) and (2.8) are satisfied, then we have*

$$\sup_{x\in\mathcal{G}}|f_{\mathbf{n}}(x)-f(x)| \text{ converges almost surely to } 0. \tag{4.5}$$

The following theorem gives an almost sure rate of convergence and is stated in the case where the mixing coefficient satisfies (2.7) and if $\theta > N(3+2\beta+2\widetilde{\beta})$ and $\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_3^*}(\log\widehat{\mathbf{n}})^{\theta_4^*}\left(g(\mathbf{n})\right)^{\frac{-2N}{\theta-N(2\beta+2\widetilde{\beta}+3)}} \to \infty$. The proof of this Theorem is similar to that of Theorem 4.5 and is omitted.

Theorem 4.6. *Under the conditions* $\mathbf{H_f}$, **H1-H5**, *(2.7) and (2.8), we have*

$$\sup_{x\in\mathcal{G}}|f_{\mathbf{n}}(x)-f(x)| \text{ converges almost surely to } 0. \tag{4.6}$$

The strong rate of convergence of $f_{\mathbf{n}}$ follows in the two cases of mixing as in the weak rate above. Suppose that the mixing coefficient satisfies (2.6), $\theta > 2N(\beta+2)$ and $\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_1^*}(\log\widehat{\mathbf{n}})^{\theta_2^*}\left(g(\mathbf{n})\right)^{\frac{-2N}{\theta-2N(\beta+2)}} \to \infty$, then:

Theorem 4.7. *Under the conditions* $\mathbf{H'_f}$, **H2-H6**, *(2.6) and (2.8), we have*

$$\sup_{x\in\mathcal{G}}|f_{\mathbf{n}}(x)-f(x)| = \mathcal{O}_{a.s}\left(D_{\mathbf{n}} + \sqrt{\frac{S_{\mathbf{n}}\,\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}\right). \tag{4.7}$$

Proof. It follows from Theorems 4.3 and 4.6. □

If (2.7) is satisfied, $\theta > N(3+2\beta+2\widetilde{\beta})$ and if $\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_3^*}(\log\widehat{\mathbf{n}})^{\theta_4^*}\left(g(\mathbf{n})\right)^{\frac{-2N}{\theta-N(2\beta+2\widetilde{\beta}+3)}} \to \infty$, then we have:

Theorem 4.8. *Under the conditions* $\mathbf{H'_f}$, **H2-H6**, *(2.7) and (2.8), we have*

$$\sup_{x\in\mathcal{G}}|f_{\mathbf{n}}(x)-f(x)| = \mathcal{O}_{a.s}\left(D_n + \sqrt{\frac{S_{\mathbf{n}}\,\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}\right). \tag{4.8}$$

Proof. It is similar to that of Theorem 4.3. □

4.2. **Convergence under exponential mixing condition.** It is worth to study the exponential mixing case where

$$\varphi(i) = C\,\exp(-si) \tag{4.9}$$

for some $s > 0$ since it includes the Geometrically Strong Mixing (GSM) case (with $\chi \equiv 1$) which is easier to check in practice.

The proof of the following Theorem is obtained by sketching the proof of Theorem 4.3 and by using similar arguments as in the finite dimensional case (Carbon, *et al.* Carbon et al. (1997b)). It is then omitted.

Theorem 4.9. *Under the conditions* $\mathbf{H}'_{\mathbf{f}}$, **H2-H6**, *(2.6) or 2.7), (4.9) and* $\widehat{\mathbf{n}}\, S_{\mathbf{n}}^{-1}(\log \widehat{\mathbf{n}})^{-2N-1} \to \infty$, *then*

$$\sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - f(x)| = O\left(D_{\mathbf{n}} + \sqrt{\frac{S_{\mathbf{n}} \log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}\right) \; a.s. \tag{4.10}$$

4.3. **The small balls probabilities effects.**
In this section we look at consistency results of the density estimate of the simple usual Parzen-Rosenblatt form of Section 3.1 when the probability distribution of X satisfies some concentration condition. In the following, we are only interested in strong rates of convergence. But under the same conditions as Theorems 4.1, 4.3, 4.5 (respectively Theorems 4.2, 4.4, 4.6) except that **H1-H6** are replace by **H7-H12**, we get the weak (with rates) and strong consistencies under the mixing condition (2.6) (respectively mixing condition (2.7)). So, the following theorems give the strong rates of convergence.

Theorem 4.10. *Under the conditions of Theorem 4.7 except that* **H2-H6** *are replaced by* **H8-H12**, *we have*

$$\sup_{x\in\mathcal{G}} |\tilde{f}_{\mathbf{n}}(x) - f(x)| = \mathcal{O}\left(h_{\mathbf{n}} + \sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\psi(h_{\mathbf{n}})}}\right), \; a.s. \tag{4.11}$$

Proof. The proof follows the same steps as that of Theorem 4.7, by using **H12** instead of **H6** and noting that $D_{\mathbf{n}} = h_{\mathbf{n}}$. □

Theorem 4.11. *Under the same conditions as Theorem 4.8 except that* **H2-H6** *are replaced by* **H8-H12**, *we have*

$$\sup_{x\in\mathcal{G}} |\tilde{f}_{\mathbf{n}}(x) - f(x)| = O\left(h_{\mathbf{n}}\right) + O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\psi(h_{\mathbf{n}})}}\right), \; a.s. \tag{4.12}$$

Proof. It is similar to that of Theorem 4.10. □

Similar arguments used to prove Theorem 4.9 lead to:

Theorem 4.12. *(Exponential mixing case) Under the conditions* $\mathbf{H}'_{\mathbf{f}}$, **H8-H12**, *(2.6) or 2.7), (4.9) and* $\widehat{\mathbf{n}}\, S_{\mathbf{n}}^{-1}(\log \widehat{\mathbf{n}})^{-2N-1} \to \infty$, *we have*

$$\sup_{x\in\mathcal{G}} |\tilde{f}_{\mathbf{n}}(x) - f(x)| = O\left(h_{\mathbf{n}}\right) + O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\psi(h_{\mathbf{n}})}}\right), \ a.s. \tag{4.13}$$

Remark 4.13.

- In this section, we obtain a rate of convergence of the form:

$$O\left(h_{\mathbf{n}}\right) + O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\psi(h_{\mathbf{n}})}}\right)$$

  where $O\left(h_{\mathbf{n}}\right)$ is the rate of the bias of the estimator which only depends on the regularity of f. Note that, by using the same approach as in this paper, one can easily shows that if f satisfies a Hölder condition: $\forall x, y \in \mathcal{G}$, $|f(x) - f(y)| \leq C \ \|x - y\|^{\kappa}$ for some C, $\kappa > 0$, then the rate of convergence is the following:

$$O\left(h_{\mathbf{n}}^{\kappa}\right) + O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\psi(h_{\mathbf{n}})}}\right).$$

- In the case of pointwise convergence at a given point $x \in \mathcal{E}$, we do not need Assumption **H12** to get a rate of convergence. It suffices in this particular case, to sketch the proof of the previous results and get rates of the form:

$$O\left(h_{\mathbf{n}}\right) + O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}F^{x}(h_{\mathbf{n}})}}\right).$$

- The bounds obtained here permit to derive the same rates of convergence as in the i.i.d case only if the space $\mathcal{E}$ is of finite dimension (for example $\mathbb{R}^d$). If the case of infinite dimension, the obtained rates (derived from the bounds in the theorems above) are the same for dependent and i.i.d cases (thanks to dependency condition like (2.3) or (2.4)) but these rates (present in the literature, see the references on non-parametric estimation for functional variables therein) are far from being proved to be the optimal ones.

5. Applications

Before going further, let us first discuss briefly in what follows, the practical use of our estimator.

5.1. **The functional spaces considered.** In the following applications problem, we deal with density estimation with data taking place on the normed space, $\mathcal{E} = C(a,b)$ (space of continuous functions over interval $[a,\, b]$) endowed with the norm defined by $||x|| = \int_a^b |x(t)|dt$, $x \in C(a,b)$. (In fact, a discretization version over a sequence of 101 equi-spaced points of $[a,\, b]$ is considered: $\int_b^b |x(t)|dt \simeq \frac{1}{b-a}\sum_{t_i=a}^b |x(t_i)|$).

5.2. **Our estimator in practice.** As previously mentioned, our kernel density estimator is defined over a rectangular region $\mathcal{I}_{\mathbf{n}}$. Suppose that the process $(X_{\mathbf{i}})$ is observed over a set $\mathcal{O}_{\mathbf{n}} \supseteq \mathcal{I}_{\mathbf{n}}$. Let $(x_{\mathbf{j}},\, \mathbf{j} \in \mathcal{O}_{\mathbf{n}})$ be the observations.

Note that similar rates of convergence as above can be obtained if one replace $\mathcal{I}_{\mathbf{n}}$ by any lattice of $\mathbb{R}^N$. Namely, this rate is for example as follows:

(1) In the case of pointwise estimation, for each site $\mathbf{j}$, if, as it will be done later on, one computes $\tilde{f}_{\mathbf{n}}(x_{\mathbf{j}})$ with observations on $V_{\mathbf{n}}^{\mathbf{j}} \subseteq \mathcal{O}_{\mathbf{n}}$ instead of $\mathcal{I}_{\mathbf{n}}$, where $V_{\mathbf{n}}^{\mathbf{j}}$ is a vicinity of $\mathbf{j}$ (for example the set of the $k_{\mathbf{n}}$ nearest neighbors). Then, if we set $\hat{\mathbf{n}}_{V_{\mathbf{n}}^{\mathbf{j}}} = \hat{\mathbf{n}}\,\lambda_{V_{\mathbf{n}}^{\mathbf{j}}}$ with $0 < \lambda_{V_{\mathbf{n}}^{\mathbf{j}}} \leq 1$, one can establish that $|\tilde{f}_{\mathbf{n}}(x_{\mathbf{j}}) - f(x_{\mathbf{j}})| = O\left(h_{\mathbf{n}} + \sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\,\lambda_{V_{\mathbf{n}}^{\mathbf{j}}}\psi(h_{\mathbf{n}})}}\right)$ since $\log \widehat{\mathbf{n}}_{\mathbf{j}}/\widehat{\mathbf{n}}_{\mathbf{j}} \leq \log \widehat{\mathbf{n}}/\widehat{\mathbf{n}}\,\lambda_{V_{\mathbf{n}}^{\mathbf{j}}}$, where $\hat{\mathbf{n}}_{\mathbf{j}} = \hat{\mathbf{n}}_{V_{\mathbf{n}}^{\mathbf{j}}}$.

(2) In the case of uniform control of the estimator over a set $\mathcal{G}$, we propose to replace $\mathcal{I}_{\mathbf{n}}$ by $\mathcal{V}_{\mathbf{n}}^{\mathcal{G}} \subseteq \mathcal{O}_{\mathbf{n}}$, where $\mathcal{V}_{\mathbf{n}}^{\mathcal{G}}$ is the set of all sites $\mathbf{j}$ such that $x_{\mathbf{j}} \in \mathcal{G}$. For the same reasons as before, one can prove that $\sup_{\mathbf{j}\in\mathcal{V}_{\mathbf{n}}^{\mathcal{G}}} |\tilde{f}_{\mathbf{n}}(x_{\mathbf{j}}) - f(x_{\mathbf{j}})| = O\left(h_{\mathbf{n}} + \sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\,\lambda_{\mathcal{V}_{\mathbf{n}}^{\mathcal{G}}}\psi(h_{\mathbf{n}})}}\right)$.

Moreover, formally, our estimator looks like its *i.i.d.* counterpart. However, as in any spatial modeling, the spatial dependency must be taken into account for applications. Here, we deal with the mixing coefficient (2.5) that tends to zero at polynomial rate (Condition (2.6)) or at exponential rate (Condition (2.7)). That is, without the loss of generality, let us consider the strong mixing case that corresponds to $\chi \equiv 1$ (and then $\alpha \equiv \varphi$) in expression (2.5). Thus, for any sites $\mathbf{i}$ and $\mathbf{j}$, the closeness they are, the strongly is the dependence between $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$.

In order to take into account the spatial dependency at any set $\mathbf{j} \in \mathcal{O}_{\mathbf{n}}$, we suggest to compute $\tilde{f}_{\mathbf{n}}(x_{\mathbf{j}})$ based on observations on $V_{\mathbf{n},\,\rho_{\mathbf{n}}}^{\mathbf{j}} = \{\mathbf{i} \in \mathcal{O}_{\mathbf{n}},\, ||\mathbf{i} - \mathbf{j}|| < \rho_{\mathbf{n}}\}$, where $(\rho_{\mathbf{n}})$ is an increasing sequel of positive reals, tending to infinity. Then, to estimate the spatial density $f(x_{\mathbf{j}})$ at $x_{\mathbf{j}}$, we propose the following procedure.

5.2.1. *Procedure of estimation of* $f(x_{\mathbf{j}})$, $\mathbf{j} \in \mathcal{O}_{\mathbf{n}}$.

(1) Specify the sets of radius and bandwidths: $S(\rho)$ and $S(h)$.

(2) For each $h_{\mathbf{n}} \in S(h)$ and $\rho_{\mathbf{n}} \in S(\rho)$ and each $\mathbf{j} \in \mathcal{O}_{\mathbf{n}}$, compute:

$$\tilde{f}_{\mathbf{n}}(x_{\mathbf{j}}) = \frac{1}{\widehat{\mathbf{n}}_{\mathbf{j}}\, a_{\mathbf{n}}^{x_{\mathbf{j}}}} \sum_{\mathbf{i}\in V_{\mathbf{n},\,\rho_{\mathbf{n}}}^{\mathbf{j}}} K\left(\frac{\|x_{\mathbf{i}} - x_{\mathbf{j}}\|}{h_{\mathbf{n}}}\right)$$

where $a_{\mathbf{n}}^{x_{\mathbf{j}}}$ has been computed as below;

(3) Compute $h_{\mathbf{n},opt}$ and $\rho_{\mathbf{n},opt}$ obtained by optimizing the Entropy (Ferraty & Vieu 2006) (or any other criterion such as *Cross-validation*) over the sets $S(h)$ and $S(\rho)$;

(4) For each $\mathbf{j}$, compute $\tilde{f}_{\mathbf{n},opt}(x_{\mathbf{j}})$ which corresponds to $h_{\mathbf{n},opt}$ and $\rho_{\mathbf{n},opt}$.

5.2.2. *The sequences* $\left(a_{\mathbf{n}}^{x_{\mathbf{j}}}\right)$.

Recall that each $a_{\mathbf{n}}^{x_{\mathbf{j}}}$ depends on the reference distribution μ. In finite dimensional setting, μ is often, the Lebesgue measure that does not exist in infinite dimensional spaces. So, μ must be specified according the considered problem.

In the following applications, we apply our estimation method to detect if two spatial (stationary) distributions $F^{(1)}$ and $F^{(2)}$ are equal or not. Let respectively $(X_{\mathbf{i}}^{(1)})$ and $(X_{\mathbf{i}}^{(2)})$ the processes corresponding respectively to $F^{(1)}$ and $F^{(2)}$. To solve the detection problem, we consider one of the two distributions as a reference measure. That is, we set for example $\mu = F^{(2)}$. Then, for each $x_{\mathbf{j}}$, $\mathbf{j} \in \mathcal{O}_{\mathbf{n}}$, observation of $(F_{\mathbf{i}}^{(1)})$, $a_{\mathbf{n}}^{x_{\mathbf{j}}}$ is defined by:

$$a_{\mathbf{n}}^{x_{\mathbf{j}}} = \frac{1}{\hat{\mathbf{n}}_{\mathbf{j}}} \sum_{\mathbf{i} \in V_{\mathbf{n}}^{\mathbf{j}}} K\left(\frac{\|X_{\mathbf{i}}^{(2)} - x_{\mathbf{j}}\|}{h_{\mathbf{n}}}\right)$$

where $V_{\mathbf{n}}^{\mathbf{j}}$ is a vicinity of $\mathbf{j}$. Then we have

$$\tilde{f}_{\mathbf{n}}(x_{\mathbf{j}}) = \frac{\sum_{\mathbf{i} \in V_{\mathbf{n}}^{\mathbf{j}}} K\left(\frac{\left\|X_{\mathbf{i}}^{(1)} - x_{\mathbf{j}}\right\|}{h_{\mathbf{n}}}\right)}{\sum_{\mathbf{i} \in V_{\mathbf{n}}^{\mathbf{j}}} K\left(\frac{\|X_{\mathbf{i}}^{(2)} - x_{\mathbf{j}}\|}{h_{\mathbf{n}}}\right)}. \tag{5.1}$$

Remark 5.1.

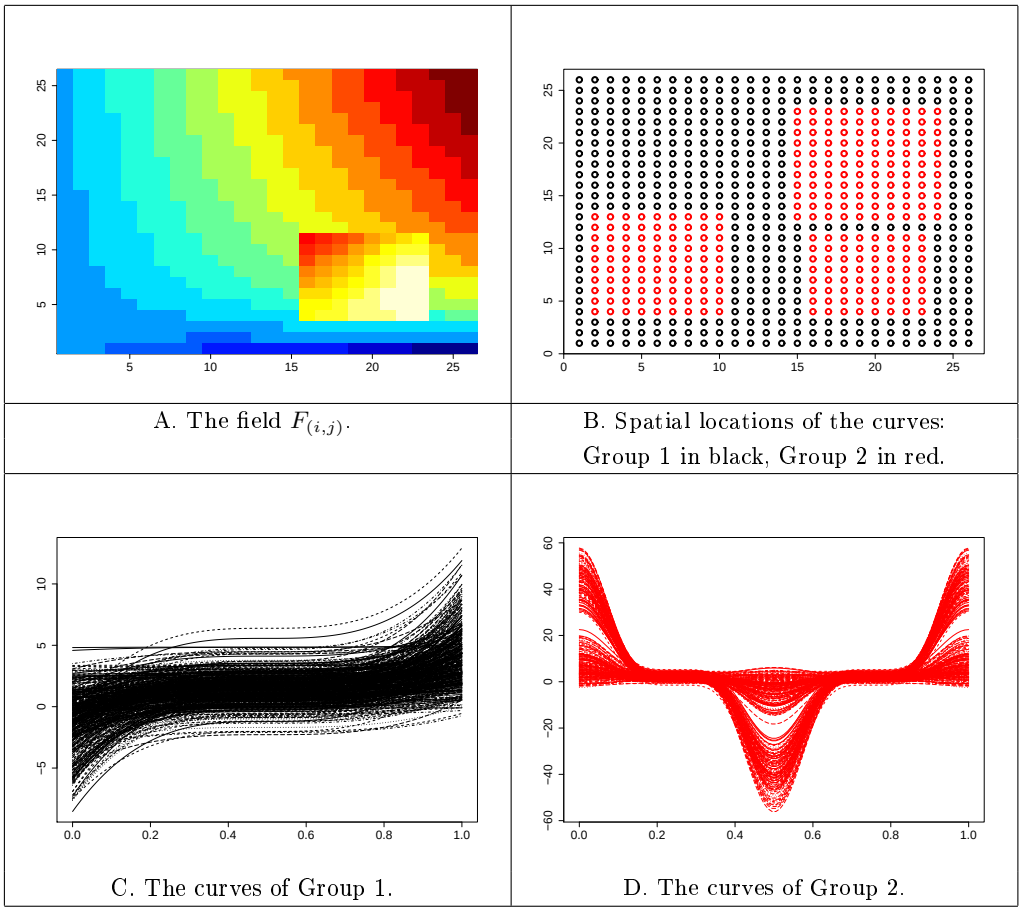
(1) *Visualization of the functional spatial feature.*

The functions f, $\tilde{f}_{\mathbf{n}}$ or $f_{\mathbf{n}}$ are defined from $\mathcal{E}$ to $\mathbb{R}^+$ where $\mathcal{E}$ is a functional space and a graphical representation of its graph is impossible. However, since f (or its estimator) is a spatial density, it is possible to get a graphical representation of the underline spatial feature. Note that this later shows the spatial feature of the process in a given domain (see Figures 5.2 and 5.4). Indeed, such graphics have been obtained by identifying the function f (or its estimator) with the function H defined by:

$$H: \begin{array}{ccccc} \mathcal{O}_{\mathbf{n}} & \to & \mathcal{E} & \to & \mathbb{R}^+ \\ \mathbf{j} & \mapsto & x_{\mathbf{j}} & \mapsto & f(x_{\mathbf{j}}) \,(\text{or } f_{\mathbf{n}}(x_{\mathbf{j}}) \,\text{or } \tilde{f}_{\mathbf{n}}(x_{\mathbf{j}})). \end{array}$$

(2) *Spatial modal curves estimation*

Estimation of a spatial modal curve over observations on a set $\mathcal{V}_{\mathbf{n}}^{\mathcal{G}}$ can be obtained by taking $\rho_{\mathbf{n}}$ such that $\mathcal{V}_{\mathbf{n}}^{\mathcal{G}} \subset \cup_{\mathbf{j} \in \mathcal{V}_{\mathbf{n}}^{\mathcal{G}}} \mathbb{B}(\mathbf{j}, \rho_{\mathbf{n}})$ where $\mathbb{B}(\mathbf{j}, \rho_{\mathbf{n}})$ is the opened ball of center $\mathbf{j}$ and radius $\rho_{\mathbf{n}}$; with $\mathcal{G} = \{x_{\mathbf{j}}, \mathbf{j} \in \mathcal{V}_{\mathbf{n}}^{\mathcal{G}}\}$.



A. The field $F_{(i,j)}$.

B. Spatial locations of the curves: Group 1 in black, Group 2 in red.

C. The curves of Group 1.

D. The curves of Group 2.

FIGURE 5.1. The simulated dataset

5.3. **Simulations.** We now illustrate our purpose through some simulations. In the following, we will deal with $N = 2$, the simulated curves are located on subsets of the area $\mathcal{I}_{(26,26)} = \{(i,j), 1 \leq i \leq 26; 1 \leq j \leq 26\}$ and we denote by $GRF(m, \sigma^2, s)$ a stationary Gaussian random field with mean m and covariance function defined by $C(h) = \sigma^2 \exp(-(\frac{\|h\|}{s})^2)$, $h \in \mathbb{R}^2$ and $s > 0$. Then, we have simulated two datasets.

The first dataset $(X^{(1)}_{(i,j)}, (i,j) \in \mathcal{I}_{(26,26)})$ is built from two different groups of curves generated, for $t \in T = [0,1]$, by:

$$X^{(1)}_{(i,j)}(t) = \begin{cases} F_{(i,j)}.(t-0.5)^3 + \varepsilon_{(i,j)} & \text{for } (i,\ j) \in R_1(\text{Group1}) \\ F_{(i,j)} \cos(2\pi t)^5 + \varepsilon_{(i,j)} & \text{for } (i,\ j) \in R_2(\text{Group2}) \end{cases}$$

with $\varepsilon = GRF(0, 20, 5)$, R_1, R_2 are two disjoint sets of sites. The curves located at $(i,\ j)$ in R_1 are in black on Figure 5.1 and the ones located $(i,\ j)$ in R_2 are in red on Figure 5.1. The field $(F_{(i,j)}\ (i,j) \in \mathcal{I}_{(26,26)})$ is the one presented in Figure 5.1 (on top of the left).

Concerning the dataset $(X^{(2)}_{(i,j)}, (i,j) \in \mathcal{I}_{(26,26)})$, we have considered the two following cases:

Case 1: The second dataset is generated as follows
$\forall (i,j) \in F_{(i,j)}$, $X^{(2)}_{(i,j)} = F_{(i,j)}.(t-0.5)^3 + \varepsilon_{(i,j)} + \epsilon_{(i,j)}$
with $\epsilon_{(i,j)} \sim N(0,.01)$, the $\epsilon_{(i,j)}$ are *i.i.d.*

Then, we have

$$X^{(2)}_{(i,j)} = X^{(1)}_{(i,j)} + \epsilon_{(i,j)} \ \text{ for } (i,j) \in R_1$$
$$X^{(1)}_{(i,j)} \neq X^{(2)}_{(i,j)} \ \text{ otherwise}$$

Case 2: $X^{(2)}_{(i,j)} = H_{(i,j)} \cos(2\pi t)^5 + \varepsilon_{(i,j)} + \epsilon_{(i,j)}$ where $(H_{(i,j)})$ is the field of Figure 5.1-A. that coincides with $F_{(i,j)}$ on $\mathcal{I}_{(26,26)} \backslash \{[16,\, 23] \times [4,\, 11]\}$. Then, we have

$$X^{(2)}_{(i,j)} = X^{(1)}_{(i,j)} + \epsilon_{(i,j)} \ \text{ for } (i,j) \in \{[2,\, 10] \times [4,13]\} \cup \{[15,\, 24] \times [13,23]\}$$
$$X^{(2)}_{(i,j)} \neq X^{(1)}_{(i,j)} \ \text{ otherwise.}$$

We aim to test if our procedure detect the difference or the similarity between the two distributions of $(X^{(1)}_{(i,j)})$ and $(X^{(2)}_{(i,j)})$ in the two cases.

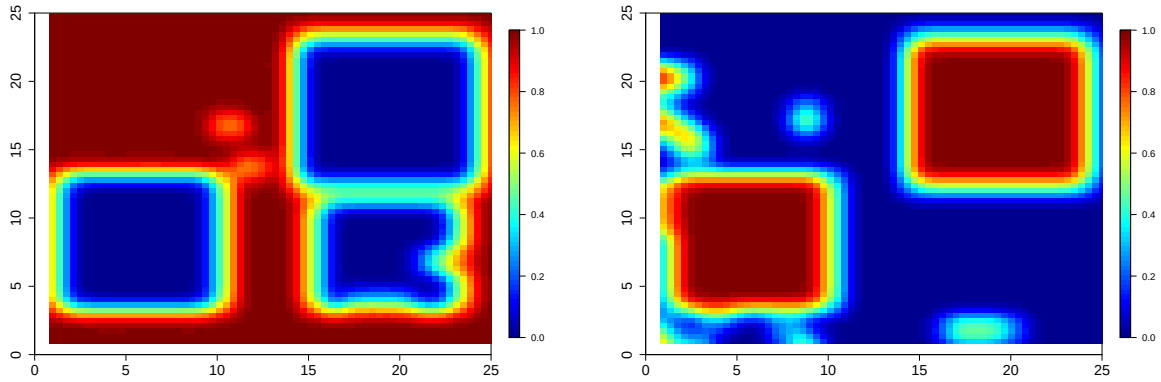
Then, we have computed the spatial density estimation (5.1) based on the simulated dataset.

For the two cases, we have get $\rho_{\mathbf{n},\,opt} = 5$; and, $h_{\mathbf{n},opt} \simeq .05$ in Case 1 and $h_{\mathbf{n},opt} \simeq .15$ in Case 2. The results are given on graphics of Figure 5.2.This later shows clearly that, a part from a few sites, our procedure is able to detect the difference in the two cases.

In fact, Figure 5.2 (left) shows that in Case 1, the kernel density of $(X^{(1)}_{(i,j)})$ with respect to the distribution of $(X^{(2)}_{(i,j)})$ is near 1 in the region R_1 when it is near by 0 in R_2. That means that our procedure detects $X^{(1)}_{(i,j)} \simeq X^{(2)}_{(i,j)}$ on R_1 and $X^{(1)}_{(i,j)}$ different from $X^{(2)}_{(i,j)}$ on the region R_2. Similarly, in case 2 (Figure 5.2 (right)), the result detects that $X^{(1)}_{(i,j)} \simeq X^{(2)}_{(i,j)}$ if $i \in [2,\, 10] \cup [15,\, 24]$ and $j \in [4,13] \cup [13,23]$ and $X^{(1)}_{(i,j)} \neq X^{(2)}_{(i,j)}$ otherwise.

5.4. **Application to spatial grain size curves.**

The dataset is a sample of *grain size curves* collected in the *Berre Lagoon*. This latter is situated in the southeast of *France*, near *Marseille*. It can be divided into two areas separated by a sandy zone: the *main lagoon* with *NW-SE* extension and the *Vaines Lagoon* in the *SE*. There are three natural inputs: the *Caronte pass* in relation with the *Gulf of Fos* located in the *North East* of the main lagoon and the *Vallat river* in the East of Vaines lagoon. Since 1962, an hydroelectric plant is located in the north of the lagoon. The volume of the fresh water used to produce electricity is very large and its injection caused major environmental modifications (Figure 5.4); thus it has been considerably reduced since 1994 (new discharge policy). The *Berre* area is under the



Case 1 Case 2

FIGURE 5.2. Representation of the kernel density of the spatial functional random $(X^{(1)}_{(i,j)})$ with respect to the distribution of $(X^{(2)}_{(i,j)})$.

influences of the wind in two main directions: N 340° *i.e.* Mistral and N 135° *i.e.* east wind; its depth does not exceeds 10 meters.

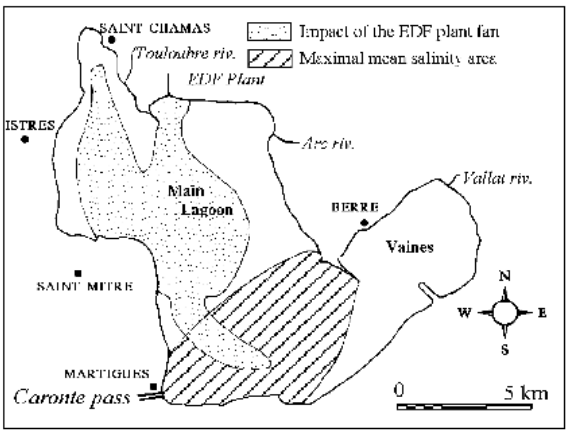
The environmental modifications are handled by the geologist by comparing sediment transport through *grain size* curves distributions.

5.4.1. *Definition of grain size curves.* Grain-size curves are routinely used by geologists to identify sedimentary facies (Buller & McManus 1972), to classify pyroclastic deposits (Lirer & Vinci 1991), or to investigate the patterns of sediment transport (Gao & Collins 1994). Their particularity is that with some transformations, they can see as distribution functions. We deal here with *grain size* curves.

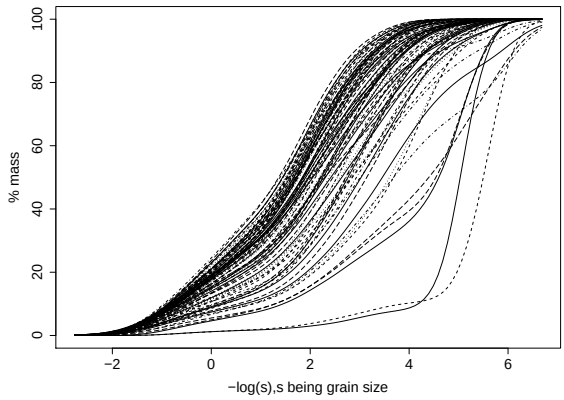
Data sampling. Samples of sediment were collected in the lagoon according to a grid of half a mile in each direction. Grain-size measurements were provided by a Malvern Instrument Ltd. Laser Microsizer; the size detection ranges from 0.063 to 900 micrometer (μm). Some classes, namely 42, were detected, associated with a geometrical progression scale subdivisions. These classes cover a large sedimentologic spectrum: colloids, i.e. organic matter, ($0.063\mu m \leq$ size $\leq 1\mu m$), clay ($1\mu m \leq$ size $< 10\mu m$), silt ($10\mu m \leq$ size $\leq 63\mu m$) and sand ($63\mu m \leq$ size $\leq 900\mu m$). All grain-size curves will be displayed according to the natural logarithmic scale. There have been two campaigns to collect these grain size curves: one in year 1992 and another in year 1997. Figure 5.3 displays the dataset.

The problem. Since, the environmental modifications are handled by the geologist by comparing sediment transport through *grain size* curves distributions. We aim to provide tools that both measure and allow to visualize the change of spatial distribution of the *grain size curve*.

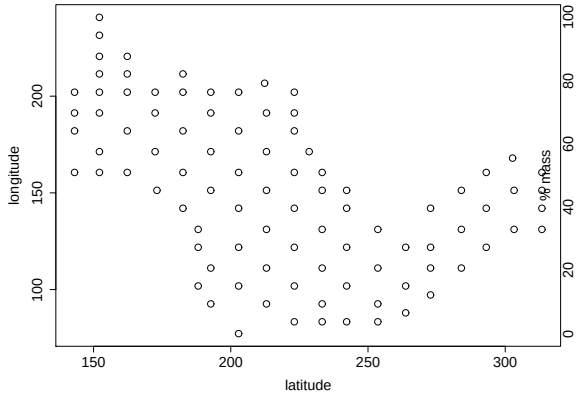
To do that, we consider the two spatial processes: $(X^{(1992)}_{(i,j)}(t))$ and $(X^{(1997)}_{(i,j)}(t))$, $t \in [0.063, 900]$, where the sites (i, j) of the observations are in the region given in Figure 5.3.C.



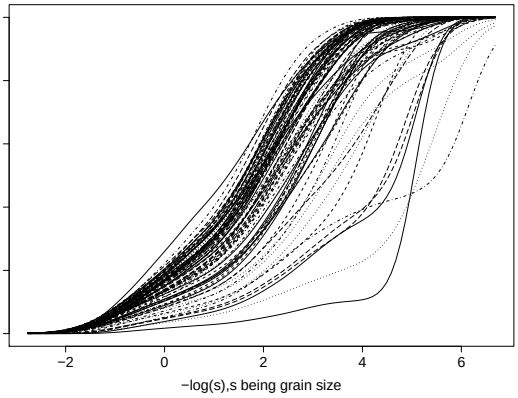
A. The Berre lagoon and its main features.



B. The grain size curves collected in 1992



C. Spatial locations of the grain size curves



D. The grain size curves collected in 1997

FIGURE 5.3. The dataset.

To solve the problem of change in spatial distribution, we propose to estimate the density of $(X_{(i,j)}^{(1992)}(.))$ (or $(X_{(i,j)}^{(1997)}(.))$) with respect to the distribution of $(X_{(i,j)}^{(1997)}(.))$ (or $(X_{(i,j)}^{(1992)}(.))$) using the spatial density estimator, $\tilde{f}_{\mathbf{n}}(.)$ previously defined.
In section 2, we assume that the data are strictly stationary, for this real data application we check throw some density computations if this stationary assumption is satisfied. For this, we consider the density of $(X_{(i,j)}^{(1992)}(.))$ (or $(X_{(i,j)}^{(1997)}(.))$) with respect to the distribution of $(X_{(i,j)}^{(1997)}(.))$ (or $(X_{(i,j)}^{(1992)}(.))$) and compute its estimator, $\tilde{f}_{\mathbf{n}}(.)$ using different sets of sample observation sites (i,j). The densities estimates $\tilde{f}_{\mathbf{n}}(.)$ obtained look very similar. That permits to think that our real data seems stationary. For seek of simplicity and shortness of the paper, we do not present these graphical density stationary results obtained. Other future investigations can be done to confirm the stationary behavior of this real data.

The spatial density estimation.

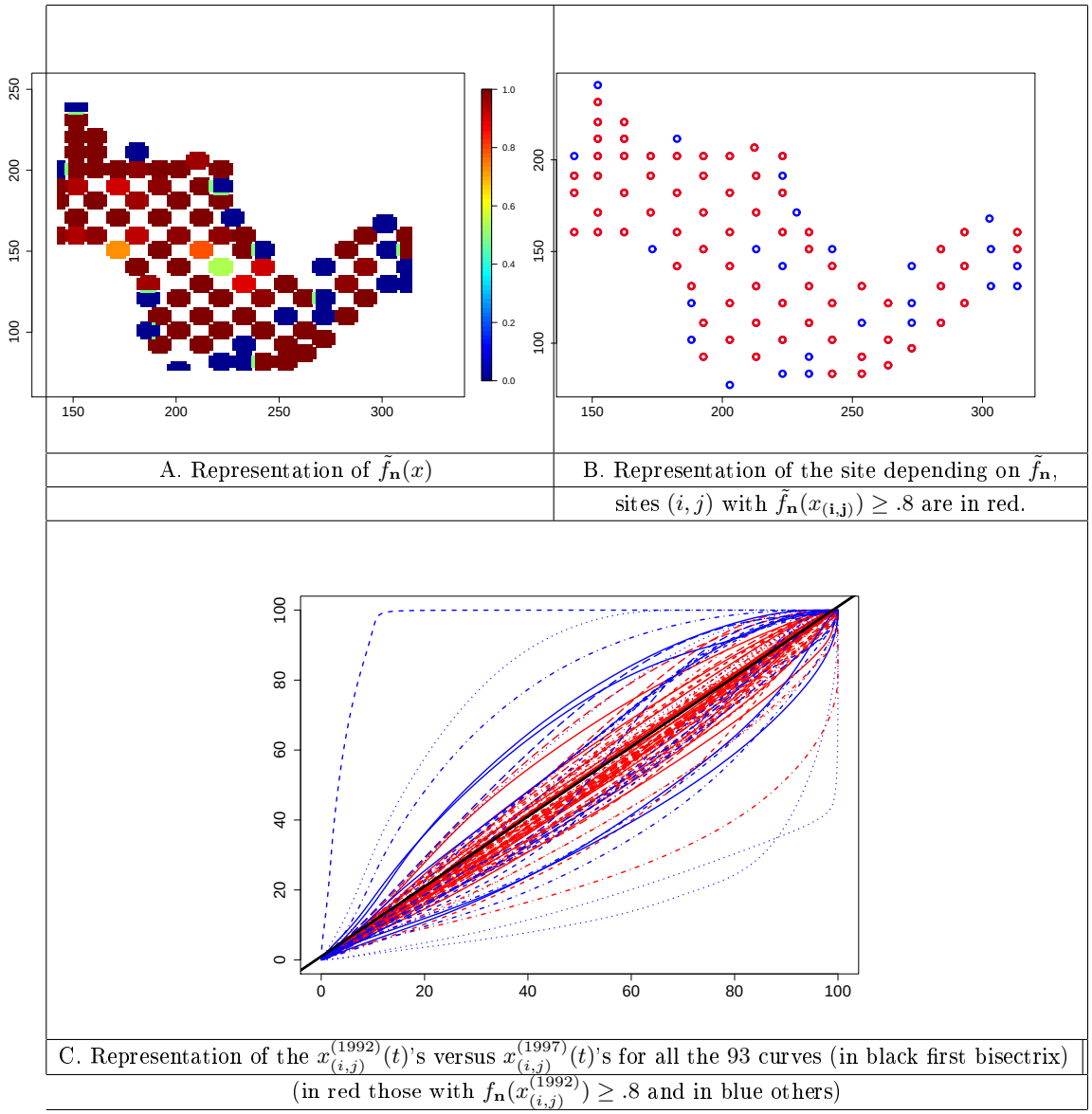
We have computed the spatial density estimation (5.1) based on the simulated dataset. This later was computed using $\rho_{\mathbf{n},opt} = 35$ and the bandwidth, obtained by minimizing the entropy of $\tilde{f}_{\mathbf{n}}$ over a set of bandwidths is $h_{\mathbf{n},opt} \simeq 1.4$. The results are displayed in Figure 5.4.

Figure 5.4.A. shows that the spatial distribution has not changes except at a few locations. Note that, these sites (where the distribution has change) are located near the main cities: *Martigues*, *Vitrolles* and *Berre*. These changes are due to the vanishing of the silt in these zones which could be due to the rainwater network. Other changes can be observed near the natural inputs of the Lagoon: *Arc river*, *Caronte pass* and *Vallat river*. The later could be due to the fact that Sediment become *coarser* in these natural inputs.

To confirm or deny the change in distribution, we have displayed Figures 5.4.B. and 5.4.C. where one can see that each site for which $f_{\mathbf{n}}(x_{(i,j)}^{(1992)}) \geq .8$ (*i.e.* $f_{\mathbf{n}}(x_{(i,j)}^{(1992)})$ closes to 1) on red in Figure 5.4.B. corresponds to $x_{(i,j)}^{(1992)}(t) \simeq x_{(i,j)}^{(1997)}(t)$, $t \in [0.063, 900]$. Indeed, except some cases, the graph of $x_{(i,j)}^{(1992)}(t)$ versus $x_{(i,j)}^{(1997)}(t)$ are closed to the first bisector for the curves on red.

6. Conclusion and discussions

In this paper, we have proposed a new method to estimate the *spatial density* of a *stationary functional random field*. This work combines asymptotic properties and practical use. The theory is directly linked with the applications through the concentration of the probability measure of the underlying functional variable. This concentration on small balls controls the rate of convergence of the nonparametric estimator and allows to select in practice the smoothing parameter. A rule for selecting this later parameter is obtained by minimizing the entropy of $\tilde{f}_{\mathbf{n}}$ over a set of bandwidths that is the spatial counter part of that suggested in the i.i.d case for example in Ferraty & Vieu (2006).



A. Representation of $\tilde{f}_{\mathbf{n}}(x)$

B. Representation of the site depending on $\tilde{f}_{\mathbf{n}}$, sites (i,j) with $\tilde{f}_{\mathbf{n}}(x_{(\mathbf{i},\mathbf{j})}) \geq .8$ are in red.

C. Representation of the $x_{(i,j)}^{(1992)}(t)$'s versus $x_{(i,j)}^{(1997)}(t)$'s for all the 93 curves (in black first bisectrix) (in red those with $f_{\mathbf{n}}(x_{(i,j)}^{(1992)}) \geq .8$ and in blue others)

FIGURE 5.4. Representation of the kernel density of the spatial functional random $(X_i^{(1992)}(.))$ with respect to the distribution of $(X_i^{(1997)}(.))$.

However, from the previous section, a question remains: do the two estimation procedures (*i.i.d.* and spatial dependence cases) lead to so different results in practice? To answer to this question, we refer to Dabo-Niang et al. (2010), where, through some simulations, a first and relevant work shows the effect of taking into account or not the spatial dependency on the unsupervised clustering rule based on kernel modal curves estimation. Namely, Dabo-Niang et al. (2010) shows that these two procedures (*i.i.d* and spatial procedures) do not lead to the same results: the *i.i.d* procedure classifies according to the closeness between curves (functional distance), the spatial procedure classifies according to both closeness between locations and proximity between

curves. That is, Dabo-Niang et al. (2010) provided an heterogeneity measure based on functional spatial modal curve estimation for classification problem. Contrary to the problem treated in this paper, Dabo-Niang et al. (2010) raises some situations where the specification of the reference measure μ is not necessary.

Other interesting fields of real applications are:

- Space-time data problem: In this setting, let $(X_{\mathbf{i}}(t))$ a space-time process. Let $[a_1, b_1]$ and $[a_2, b_2]$ $(a_1 \leq b_1 < a_2 \leq b_2)$ two periods of observations of the process. Our procedure can be used to detect spatial feature that has changes at $[a_2, b_2]$ since $[a_1, b_1]$. The fields of applications are very large: meteorology (spatio-temporal evolution of the wind), ecology (spatio-temporal dynamic of a population), archeology,...
- Unsupervised clustering methods of hyper-spectral images which are spatially distributed spectra: such data are collected in a large field of real-life applications: geology, agriculture, surveillance, identify surface materials,....
- One can compare the distribution of a spatial process with a Wiener measure or any other known functional spatial process distribution.

Now, since our estimator is nonparametric, it is subject to the well-known curse of dimensionality problem. Let us discuss this problem.

The curse of dimensionality problem in functional spatial setting.

This well-known problem for high-dimensionality studies, appears in nonparametric statistical setting as: "the number of observations required for a good estimation (of the density or the regression function for example) increases exponentially with the dimension of the space (of study)". Then, if such a defect is observed in multidimensional setting, naturally, that should be more crucial in infinite dimensional space.

Really, as raised by Donoho (2000) and Ferraty & Vieu (2006) in the *i.i.d.* case, this problem depends on the "concentration of the distribution measure" of the underlying variable.

Let us go back to our estimator and to fix ideas consider the *small ball probabilities* case where the rate of convergence of $\tilde{f}_{\mathbf{n}}(x)$ at a given point $x \in \mathcal{E}$ is of the form:

$$O\left(h_{\mathbf{n}}\right) + O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} F^x(h_{\mathbf{n}})}}\right).$$

As we said before, the term $O\left(h_{\mathbf{n}}\right)$ controls the bias and $O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} F^x(h_{\mathbf{n}})}}\right)$ comes from the variability of $\tilde{f}_{\mathbf{n}}(x) - \mathbf{E}\left(\tilde{f}_{\mathbf{n}}(x)\right)$. So the rate of convergence of our estimator crucially depends on the *small ball probabilities* $F^x(h_{\mathbf{n}})$. The role of the *small ball probabilities* have been discussed in Dabo-Niang & Rhomari (2009), Ferraty & Vieu (2006) (p. 206) and their arguments remain true in the spatial case. In fact, since $F^x(h_{\mathbf{n}}) = P(X \in B(x, h_{\mathbf{n}}))$, the rate of convergence of $\tilde{f}_{\mathbf{n}}(x)$ depends on the concentration of the spatial process around x. Thus, the more the observations are closed to x, the faster $\tilde{f}_{\mathbf{n}}(x)$ will converge (since in such situation, $\mathbf{P}\left(X \in B(x, h_{\mathbf{n}})\right)$ is higher).

Now, in $\mathbb{R}^d$, as remarked in Section 3, $\mathbf{P}\left(X \in B(x, h_{\mathbf{n}})\right) \simeq f(x)\, h_{\mathbf{n}}^d$. So, we retrieve both the *curse of dimensionality* and the following property: in multidimensional setting, the rate of convergence of $f(x)$ depends on the value of $f(x)$. If $f(x)$ is small, the rate of convergence depends crucially on d and high values of $f(x)$ are better estimated with much less data than small values of $f(x)$ for a fixed d.

In short, we can say that actually, the curse of dimensionality is a "curse of *small ball probabilities*" problem. Thus, the rate of convergence of our estimator will crucially suffer from a "curse of *small ball probabilities*" as soon as the data are very sparse and the challenge is to look for solutions (as in multidimensional case) that "increase" the *small ball probabilities*. Such solutions can be seen as the functional counterparts of the well-known *reduction dimension* methods.

This motivate some authors such as Ferraty & Vieu (2006), Dabo-Niang et al. (2006, 2010), Dabo-Niang & Rhomari (2009), Ferraty & Vieu (2006), to endow the space $\mathcal{E}$ with a suitable topological structure. They choose a semi-metric according to a given statistical problem. Note that the most known semi-metrics are those based on projection on finite dimensional subspaces such as first directions of the PCA, B-splines basis, Fourier basis, wavelets basis,... which are well-known as dimension reduction methods.

Acknowledgments. The authors are grateful to two anonymous referees and an Associate Editor for their careful reading of the first version of this paper, and for their insightful and constructive comments. Thanks to the GIPREB (Groupement d'intérêt public pour la réhabilitation de l'étang de Berre) which allows the use of the dataset. The authors wish to thank Claude Manté of Laboratoire LMGEM (Centre d'Océanologie de Marseille) for discussions about the grain size curves and his helpful comments for the applications.

7. Appendix

This section is devoted to prove the asymptotic properties of the kernel density estimate.

Before the proofs of the results, we need to present some preliminary results. The first three lemmas are given in Carbon et al. (1997b).

Lemma 7.1. *Suppose $E_1, ..., E_r$ be sets containing m sites each with $dist(E_i,\ E_j) \geq \gamma$ for all $i \neq j$ where $1 \leq i \leq r$ and $1 \leq j \leq r$. Suppose $Z_1, ..., Z_r$ is a sequence of real-valued r.v.'s measurable with respect to $\mathcal{B}(E_1), ..., \mathcal{B}(E_r)$ respectively, and Z_i takes values in $[a, b]$. Then there exists a sequence of independent r.v.'s $Z_1^*, ..., Z_r^*$ independent of $Z_1, ..., Z_r$ such that Z_i^* has the same distribution as Z_i and satisfies*

$$\sum_{i=1}^{r} E|Z_i - Z_i^*| \leq 2r(b-a)\psi((r-1)m, m)\varphi(\gamma). \tag{7.1}$$

Lemma 7.2.

(i) Suppose that (2.5) holds. Denote by $\mathcal{L}_r(\mathcal{F})$ the class of $\mathcal{F}-$measurable r.v.'s X satisfying

$\|X\|_r = (E|X|^r)^{1/r} < \infty$. *Suppose* $X \in \mathcal{L}_r(\mathcal{B}(E))$ *and* $Y \in \mathcal{L}_s(\mathcal{B}(E'))$. *Assume also that* $1 \leq r,\ s,\ t < \infty$ *and* $r^{-1} + s^{-1} + t^{-1} = 1$. *Then*

$$|EXY - EXEY| \leq C\|X\|_r\|Y\|_s\{\chi(Card(E), Card(E'))\varphi(dist(E, E'))\}^{1/t}. \tag{7.2}$$

(ii) For r.v.'s bounded with probability 1, the right-hand side of (7.2) can be replaced by $C\chi(Card(E), Card(E'))\varphi(dist(E, E'))$.

Lemma 7.3.
If (2.8) holds for $\theta > 2N$, *then*

$$\sum_{i=1}^{\infty} i^{N-1}(\varphi(i))^a < \infty \tag{7.3}$$

for some $0 < a < 1/2$.

The following lemma is due to Dabo-Niang et al. (2006) and will be useful in the case of small balls probabilities effects.

Lemma 7.4.
Let K *be a continuous kernel on* $(0,1)$ *with* $supp(K) = (0,1)$ *and* K^{-1} *integrable where* $K^{-1}(u) = \inf_t\{t \in (0,1),\ K(t) = u\}$. *Then we have:*

$$\int_{\mathcal{E}} K\left(\frac{\|y - x\|}{h_n}\right) d\mu(y) \ = \ K(1)\,\mu\left(B(x,h)\right) + \int_{K(1)}^{K(0)} \mu\left(B(x, K^{-1}(u)h)\right) du.$$

To state the convergence results stated in Section 4, it suffices to study the consistency of the bias $\sup_{x\in\mathcal{G}} |E\left(f_{\mathbf{n}}(x)\right) - f(x)|$ on one hand and that of $\sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - E\left(f_{\mathbf{n}}(x)\right)|$ on the other hand, since $\sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - f(x)| \leq \sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - E\left(f_{\mathbf{n}}(x)\right)| + \sup_{x\in\mathcal{G}} |E\left(f_{\mathbf{n}}(x)\right) - f(x)|$.

Concerning the bias, because of the properties of the expectation, the uniform convergence result of the bias to zero proved by Dabo-Niang et al. (2006) remains valid here.

Concerning the term $\sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - E\left(f_{\mathbf{n}}(x)\right)|$ the following approach is the same for the weak and the strong consistencies.

We set:

$$Q_{\mathbf{n}}(x) = f_{\mathbf{n}}(x) - E\left(f_{\mathbf{n}}(x)\right) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Z_{\mathbf{i},\mathbf{n},x}, \qquad x \in \mathcal{G},$$

where

$$Z_{\mathbf{i},\mathbf{n},x} = \frac{1}{\widehat{\mathbf{n}}a_{\mathbf{n}}^x}\left(K_{\mathbf{n}}\left(\|X_i - x\|\right) - E\left(K_{\mathbf{n}}\left(\|X_i - x\|\right)\right)\right).$$

Define

$$S_{1\mathbf{n}} = \max_{1\leq j\leq d_{\mathbf{n}}} \sup_{x\in B_j} |f_{\mathbf{n}}(x) - f_{\mathbf{n}}(x_j)|,$$

$$S_{2\mathbf{n}} = \max_{1\leq j\leq d_{\mathbf{n}}} \sup_{x\in B_j} |Ef_{\mathbf{n}}(x_j) - Ef_{\mathbf{n}}(x)|,$$

$$S_{3\mathbf{n}} = \max_{1\leq j\leq d_{\mathbf{n}}} |f_{\mathbf{n}}(x_j) - Ef_{\mathbf{n}}(x_j)|.$$

Since $\mathcal{G}$ is covered by $d_{\mathbf{n}}$ balls $B_j = B(x_j, r_{\mathbf{n}})$ of radius $r_{\mathbf{n}}$ and center at x_j, then,

$$\sup_{x\in\mathcal{G}} |f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x)| \leq S_{1\mathbf{n}} + S_{2\mathbf{n}} + S_{3\mathbf{n}}.$$

Using assumptions **H3** and **H5**, one can easily show that $S_{1\mathbf{n}}$ and $S_{2\mathbf{n}}$ are equal to $o\left(\sqrt{\frac{S_{\mathbf{n}}\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}\right)$ a.s. (or in probability for the weak consistency study). To study the consistency of $S_{3\mathbf{n}} = \max_{1\leq j\leq d_{\mathbf{n}}} |Q_{\mathbf{n}}(x_j)|$, we use the following classical spatial block decomposition (due to **?**) .

For that, without loss of generality we assume that each $n_i,\ i = 1, ..., N$ is of the form:

$$n_i = 2pt_i, \tag{7.4}$$

for some integers $p \geq 1$ and $t_1, ..., t_N$.

Remark 7.5. As raised by **?**, if one does not have the equalities (7.4), the term say $T(\mathbf{n}, x, 2^N+1)$ (which contains the $Z_{\mathbf{i},\mathbf{n},x}$'s at the ends not included in the blocks above) can be added.

Then, the random variables $Z_{\mathbf{i},\mathbf{n},x}$ are grouped into blocks of different sizes such that, we can set:

$$U(1,\mathbf{n},x,\mathbf{j}) = \sum_{i_k=2j_kp+1,\ 1\leq k\leq N}^{(2j_k+1)p} Z_{\mathbf{i},\mathbf{n},x},$$

$$U(2,\mathbf{n},x,\mathbf{j}) = \sum_{i_k=2j_kp+1,\ 1\leq k\leq N-1}^{(2j_k+1)p} \sum_{i_N=(2j_N+1)p+1}^{2(j_N+1)p} Z_{\mathbf{i},\mathbf{n},x},$$

$$U(3,\mathbf{n},x,\mathbf{j}) = \sum_{i_k=2j_kp+1,\ 1\leq k\leq N-2}^{(2j_k+1)p} \sum_{i_{N-1}=(2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \sum_{i_N=2j_Np+1}^{(2j_N+1)p} Z_{\mathbf{i},\mathbf{n},x},$$

$$U(4,\mathbf{n},x,\mathbf{j}) = \sum_{i_k=2j_kp+1,\ 1\leq k\leq N-2}^{(2j_k+1)p} \sum_{i_{N-1}=(2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \sum_{i_N=(2j_N+1)p+1}^{2(j_N+1)p} Z_{\mathbf{i},\mathbf{n},x},$$

and so on. Note that

$$U(2^{N-1},\mathbf{n},x,\mathbf{j}) = \sum_{i_k=(2j_k+1)p+1,\ 1\leq k\leq N-1}^{2(j_k+1)p} \sum_{i_N=2j_Np+1}^{(2j_N+1)p} Z_{\mathbf{i},\mathbf{n},x}.$$

Finally,

$$U(2^N,\mathbf{n},x,\mathbf{j}) = \sum_{i_k=(2j_k+1)p+1,\ 1\leq k\leq N}^{2(j_k+1)p} Z_{\mathbf{i},\mathbf{n},x}.$$

Setting $\mathcal{T} = \{0, ..., t_1 - 1\} \times ... \times \{0, ..., t_N - 1\}$, we define for each integer $l = 1, ..., 2^N$,

$$T(\mathbf{n}, x, l) = \sum_{\mathbf{j} \in \mathcal{T}} U(l, \mathbf{n}, x, \mathbf{j}).$$

Then, we obtain the following decomposition

$$Q_{\mathbf{n}}(x) = f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x) = \sum_{l=1}^{2^N} T(\mathbf{n}, x, l).$$

To prove that $\max_{1 \le j \le d_{\mathbf{n}}} \|Q_{\mathbf{n}}(x_j)\| = O\left(\sqrt{\frac{S_{\mathbf{n}} \log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}\right)$ a.s. (resp. in probability), it is sufficient to show that

$$\max_{1 \le j \le d_{\mathbf{n}}} |T(\mathbf{n}, x_j, l)| = O\left(\sqrt{\frac{S_{\mathbf{n}} \log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}\right) \text{ a.s. (resp. in probability)} \tag{7.5}$$

for each $l = 1, ..., N$. Without loss of generality we will show (7.5) for $l = 1$.

For that, we set $\epsilon_{\mathbf{n}} = \eta \sqrt{\frac{S_{\mathbf{n}} \log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}$ (where $\eta > 0$ is a constant to be chosen later) and $\beta_{1\widehat{\mathbf{n}}} = S_{\mathbf{n}} \chi(\widehat{\mathbf{n}}, p^N) \varphi(p) \epsilon_{\mathbf{n}}^{-1}$.

Lemma 7.6. *Given an arbitrary large positive constant c, there exists a positive constant C such that for any $\eta > 0$*

$$P\left[\max_{1 \le j \le d_{\mathbf{n}}} |T(\mathbf{n}, x_j, 1)| > \eta \sqrt{\frac{S_{\mathbf{n}} \log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}\right] \le C d_{\mathbf{n}} (\widehat{\mathbf{n}}^{-c} + \beta_{1\widehat{\mathbf{n}}}).$$

Proof. Let

$$T(\mathbf{n}, x, 1) = \sum_{\mathbf{j} \in \mathcal{T}} U(1, \mathbf{n}, x, \mathbf{j}).$$

be the sum of $\widehat{\mathbf{t}} = t_1 \times ... \times t_N$ of the $U(1, \mathbf{n}, x, \mathbf{j})$'s. Note that each $U(1, \mathbf{n}, x, \mathbf{j})$'s is measurable with respect to the $\sigma-$field generated by $X_{\mathbf{i}}$ with $\mathbf{i}$ belonging to the set of sites

$$\mathcal{I}_{\mathbf{i},\, \mathbf{j}} = \{\mathbf{i} : \ 2j_k p + 1 \le i_k \le (2j_k + 1)p, \ k = 1, ..., N\}.$$

These sets of sites are separated by a distance greater than p. Enumerate the random variables's $U(1, \mathbf{n}, x, \mathbf{j})$ and the corresponding $\sigma-$field with which they are measurable in an arbitrary manner and refer to them respectively as $V_1, ..., V_{\widehat{\mathbf{t}}}$ and $\mathcal{B}_1, ..., \mathcal{B}_{\widehat{\mathbf{t}}}$. Then, $T(\mathbf{n}, x, 1) = \sum_{i=1}^{\widehat{\mathbf{t}}} V_i$, with

$$|V_i| = |U(1, \mathbf{n}, x, \mathbf{j})| < C p^N \widehat{\mathbf{n}}^{-1} S_{\mathbf{n}}, \tag{7.6}$$

Lemma 7.1 allows us to approximate $V_1, ..., V_{\widehat{\mathbf{t}}}$ by $V_1^*, ..., V_{\widehat{\mathbf{t}}}^*$ such that:

$$P\left[|T(\mathbf{n}, x, 1)| > \epsilon_{\mathbf{n}}\right] \le P\left[\left|\sum_{i=1}^{\widehat{\mathbf{t}}} V_i^*\right| > \epsilon_{\mathbf{n}}/2\right] + P\left[\sum_{i=1}^{\widehat{\mathbf{t}}} |V_i - V_i^*| > \epsilon_{\mathbf{n}}/2\right]. \tag{7.7}$$

Now, using: Markov's inequality, (7.6), (7.1) and the fact that the sets of sites (with respect to which V_i's are measurable) are separated by a distance greater than p, we get:

$$P\left[\sum_{i=1}^{\widehat{\mathbf{t}}}|V_i - V_i^*| > \epsilon_{\mathbf{n}}\right] \leq C\widehat{\mathbf{t}}p^N\widehat{\mathbf{n}}^{-1}S_{\mathbf{n}}\chi(\widehat{\mathbf{n}},p^N)\varphi(p)\epsilon_{\mathbf{n}}^{-1} \sim \beta_{1\widehat{\mathbf{n}}}\ . \tag{7.8}$$

Let

$$\lambda_{\mathbf{n}} = \left(\widehat{\mathbf{n}}S_{\mathbf{n}}^{-1}\log\widehat{\mathbf{n}}\right)^{1/2}, \tag{7.9}$$

then,

$$p = \left[\left(\frac{\widehat{\mathbf{n}}}{4S_{\mathbf{n}}\lambda_{\mathbf{n}}}\right)^{1/N}\right] \sim \left(\frac{\widehat{\mathbf{n}}}{S_{\mathbf{n}}\log\widehat{\mathbf{n}}}\right)^{1/2N}, \tag{7.10}$$

and $\lambda_{\mathbf{n}}\epsilon_{\mathbf{n}} = \eta\log\widehat{\mathbf{n}}$.

If (2.8) holds for $\theta > 2N$, then we prove such as in **?** (follow the same steps as those of the proof of Lemma 2.2 of **?**) by using our hypotheses $\mathbf{H_f}$, **H1-H4** and Lemma 7.3, that

$$\lim_{\mathbf{n}\to\infty}\widehat{\mathbf{n}}S_{\mathbf{n}}^{-1}(R_{\mathbf{n}(x)} + U_{\mathbf{n}(x)}) < C \tag{7.11}$$

where C is a constant independent of $x \in \mathcal{G}$ and

$$U_{\mathbf{n}(x)} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} E(Z_{\mathbf{i},\mathbf{n},x})^2$$

$$R_{\mathbf{n}(x)} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\ \sum_{\mathbf{l}\in\mathcal{I}_{\mathbf{n}},\ i_k\neq l_k \text{ for some } k} |Cov(Z_{\mathbf{i},\mathbf{n},x},\ Z_{\mathbf{l},\mathbf{n},x})|.$$

In the case of small balls probabilities effects (Section 3.1) $\mathbf{H7}-\mathbf{H9}$ help us to have this result. By (7.3) and (7.11), we have

$$\lambda_{\mathbf{n}}^2\sum_{i=1}^{\widehat{\mathbf{t}}}E(V_i^*)^2 \leq C\widehat{\mathbf{n}}S_{\mathbf{n}}^{-1}(U_{\mathbf{n}(x)} + R_{\mathbf{n}(x)})\log\widehat{\mathbf{n}} < C\log\widehat{\mathbf{n}}.$$

Using (7.6), we get $|\lambda_{\mathbf{n}}V_i^*| < 1/2$ for large $\widehat{\mathbf{n}}$. We deduce from Bernstein's inequality (see Theorem 1.2 of Bosq (1998), page 24) that

$$P\left[\left|\sum_{i=1}^{\widehat{\mathbf{t}}}V_i^*\right| > \epsilon_{\mathbf{n}}\right] \leq 2\exp(-\lambda_{\mathbf{n}}\epsilon_{\mathbf{n}} + \lambda_{\mathbf{n}}^2\sum_{i=1}^{\widehat{\mathbf{t}}}E(V_i^*)^2) \leq 2\exp((-\eta + C)\log\widehat{\mathbf{n}}) \leq \widehat{\mathbf{n}}^{-c} \tag{7.12}$$

for sufficiently large $\widehat{\mathbf{n}}$ and $\eta > C$. We get from (7.7), (7.8) and (7.12) that

$$P[\max_{1\leq j\leq d_{\mathbf{n}}}|T(\mathbf{n},x_j,1)| > \epsilon_{\mathbf{n}}] \leq Cd_{\mathbf{n}}(\widehat{\mathbf{n}}^{-c} + \beta_{1\widehat{\mathbf{n}}}).$$

□

Proof of Theorem 4.1.

To prove this result it suffices to show that $d_{\mathbf{n}}\widehat{\mathbf{n}}^{-c} \to 0$ and $d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \to 0$ (see Lemma 7.6 with the specified assumptions). In fact, we have

$$d_{\mathbf{n}}\widehat{\mathbf{n}}^{-c} \leq C\widehat{\mathbf{n}}^{\beta-c}$$

which goes to zero as soon as $c > \beta$, note that this inequality is possible since c is chosen as a large positive constant as stated in Lemma 7.6.

Moreover we have:

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} = d_{\mathbf{n}}S_{\mathbf{n}}\chi(\widehat{\mathbf{n}}, p^N)\varphi(p)\epsilon_{\mathbf{n}}^{-1}. \tag{7.13}$$

We have also that $\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_1}(\log\widehat{\mathbf{n}})^{\theta_2} \to \infty$ is equivalent to $(d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}})^{-1} \to \infty$ by assumption (2.6). Then, $d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \to 0$ since

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \leq C\widehat{\mathbf{n}}^{\beta}S_{\mathbf{n}}\chi(\widehat{\mathbf{n}}, p^N)p^{-\theta}\epsilon_{\mathbf{n}}^{-1}$$

$$\leq C\widehat{\mathbf{n}}^{\beta}S_{\mathbf{n}}\left(\frac{\widehat{\mathbf{n}}}{S_{\mathbf{n}}\log\widehat{\mathbf{n}}}\right)^{(1/2)-(\theta/2N)}((\widehat{\mathbf{n}}/S_{\mathbf{n}}\log\widehat{\mathbf{n}}))^{\frac{1}{2}}$$

$$= C(\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_1}(\log\widehat{\mathbf{n}})^{\theta_2})^{\frac{-\theta+2N(\beta+1)}{2N}}.$$

So

$$\lim_{\mathbf{n}\to+\infty}\sup_{x\in\mathcal{G}}|f_{\mathbf{n}}(x) - E(f_{\mathbf{n}}(x))| = 0, \text{ in probability} \tag{7.14}$$

by the way. □

Proof of Theorem 4.2. Analogously to Theorem 4.1, (2.7) allows to show that $\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_3}(\log\widehat{\mathbf{n}})^{\theta_4} \to \infty$ is equivalent to $d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \to 0$. In effect, we have

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \leq C\widehat{\mathbf{n}}^{\beta}S_{\mathbf{n}}\chi(\widehat{\mathbf{n}}, p^N)p^{-\theta}\epsilon_{\mathbf{n}}^{-1}$$

$$\leq C\widehat{\mathbf{n}}^{\beta+\widetilde{\beta}}S_{\mathbf{n}}\left(\frac{\widehat{\mathbf{n}}}{S_{\mathbf{n}}\log\widehat{\mathbf{n}}}\right)^{-(\theta/2N)}((\widehat{\mathbf{n}}/S_{\mathbf{n}}\log\widehat{\mathbf{n}}))^{\frac{1}{2}}$$

$$= C(\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_3}(\log\widehat{\mathbf{n}})^{\theta_4})^{\frac{-\theta+N(2\beta+2\widetilde{\beta}+1)}{2N}}. \qquad \square$$

Proof of Theorem 4.3.

Since **H6** implies **H1**, we can apply Lemma 7.6 and get

$$\sup_{x\in\mathcal{G}}|f_{\mathbf{n}}(x) - E(f_{\mathbf{n}}(x))| = O\left(\sqrt{\frac{S_{\mathbf{n}}\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}\right), a.s.$$

Dabo-Niang et al. (2006) showed that

$$\sup_{x\in\mathcal{G}}|Ef_{\mathbf{n}}(x) - f(x)| = O(D_{\mathbf{n}}).$$

This finishes the proof of the theorem. □

Proof of Theorem 4.5.
We have:

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \leq C(\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_1}(\log\widehat{\mathbf{n}})^{\theta_2})^{\frac{-\theta+2N(\beta+1)}{2N}} = \varphi_\eta(\mathbf{n}).$$

Then,

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}}\widehat{\mathbf{n}}g(\mathbf{n}) \leq C(\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_1}(\log\widehat{\mathbf{n}})^{\theta_2})^{\frac{-\theta+2N(\beta+1)}{2N}}\widehat{\mathbf{n}}g(\mathbf{n})$$

$$\leq C(\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_1^*}(\log\widehat{\mathbf{n}})^{\theta_2^*}g(\mathbf{n})^{-2N/(\theta-(\beta+2)2N)})^{\frac{-\theta+(\beta+2)2N)}{2N}}.$$

Thus, the condition $\widehat{\mathbf{n}}S_{\mathbf{n}}^{\theta_1^*}(\log\widehat{\mathbf{n}})^{\theta_2^*}\,(g(\mathbf{n}))^{\frac{-2N}{\theta-2N(\beta+2)}} \to \infty$ is equivalent to
$\widehat{\mathbf{n}}g(\mathbf{n})\varphi_\eta(\mathbf{n}) \to 0$ which implies that $\sum_{\mathbf{n}\in\mathbb{N}^N}\varphi_\eta(\mathbf{n}) < \infty$. Then, the theorem follows by Borel-Cantelli Lemma. □

References

Anselin, L. & Florax, R. J. M. (1995). *New Directions in Spatial Econometrics*. Berlin: Springer.

Bar-Hen, A., Bel, L., & Cheddadi, R. (2008). *Spatio-temporal functional regression on paleoecological data In Functional and Operatorial Statistics*. Springer Verlag: Edited by S. Dabo-Niang and F. Ferraty.

Basse, M., Dabo-Niang, S., & Diop, S. (2008). Kernel density estimation for functional random fields. *Pub. Inst. Stat. Univ.Paris*, LII(1-2), 91–108.

Biau, G. & Cadre, B. (2004). Nonparametric spatial prediction. *Stat. Inference Stoch. Process.*, 7, 327–349.

Bogachev, V. (1999). *Gaussian measures(Math. surveys and monographs)*. American Mathematical Society.

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

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and applications*. New York: Lecture Notes in Statistics, Springer-Verlag.

Buller, A. & McManus, J. (1972). Simple metric sedimentary statistics used to recognize different environments. *Sedimentology*, 18, 1–21.

Carbon, M. (2006). Polygone des fréquences pour des champs aléatoires. *C. R. Math. Acad. Sci. Paris*, 342(9), 693–696.

Carbon, M., Francq, C., & Tran, L. T. (2007). Kernel regression estimation for random fields. *J. Statist. Plann. Inference*, 137(Issue 3), 778–798.

Carbon, M., Hallin, M., & Tran, L. T. (1997a). Kernel density estimation for random fields: The l_1 theory. *Nonparametric Statistics*, 6, 157–170.

Carbon, M., Tran, L. T., & Wu, B. (1997b). Kernel density estimation for random fields. *Stat. and Probab. Lett.*, 36, 115–125.

Chilès, J. & Delfiner, P. (1999). *Geostatistics. Modeling spatial uncertainty.* New-York: Wiley Series in Probability and Mathematical Statistics.

Cressie, N. A. (1993). *Statistics for Spatial Data.* New-York: Wiley Series in Probability and Mathematical Statistics.

Csaki, E. (1980). A relation between chung's and strassen laws of the iterated logarithm. *Zeit. Wahrs. Ver. Geb.*, 19, 287–301.

Dabo-Niang, S. (2004). Kernel density estimator in an infinite dimensional space with a rate of convergence in the case of diffusion process. *App. Math. Lett.*, 17, 381–386.

Dabo-Niang, S., Ferraty, F., & Vieu, P. (2006). Mode estimation for functional random variable and its application for curves classification. *Far East J. of Theoretical Stat.*, 18(1), 93 – 119.

Dabo-Niang, S. & Rhomari, N. (2009). Kernel regression estimate in a banach space. *J. Statist. Plann. Inf.*, 139, 1421–1434.

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

Dabo-Niang, S. & Yao, A. F. (2007). Kernel regression estimation for continuous spatial processes. *Math. Method. Stat.*, 16(4), 1–20.

Delicado, P., Giraldo, R., & Mateu, J. (2008, pp 135-141). *Point-wise kriging for spatial prediction of functional data.* Functional and Operatorial Statistics, edited by S. Dabo-Niang and F. Ferraty: Springer Verlag.

Donoho, D. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. Lecture delivered at the conference "Math Challenges of the 21st Century" held by the American Math. Society organised in Los Angeles, August 6-11.

Doukhan, P. (1994). *Mixing - Properties and Examples.* New York: Lecture Notes in Statistics, Springer-Verlag.

Ferraty, F., Mas, A., & Vieu, P. (2007). Advances in nonparametric regression for functional variables. *Australian and New-Zealand J. of Statist.*, 49(3), 1–20.

Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis.* berlin: Springer-Verlag.

Gao, S. & Collins, M. (1994). Analysis of grain size trends, for defining sediment transport pathways in marine environments. *J. Coastal Res.*, 10(1), 70–78.

Guyon, X. (1995). *Random Fields on a Network - Modeling, Statistics, and Applications.* New-York: Springer.

Györfi, L., Härdle., Sarda, P., & Vieu, P. (1990). *Nonparametric Curve Estimation from Time Series.* Number 60. New-York.: Lecture Notes in Statistics, Springer-Verlag.

Hallin, M., Lu, Z., & Tran, L. T. (2004a). Kernel density estimation for spatial processes: The l_1 theory. *J. Multivariate Anal.*, 88(1), 61–75.

Hallin, M., Lu, Z., & Tran, L. T. (2004b). Local linear spatial regression. *Ann. Statist.*, 32(6), 2469–2500.

Li, W. V. & man Shao, Q. (2001). Gaussian processes: inequalities, small ball probabilities and applications. stochastic processes: theory and methods. *Handbook of Statist*, (pp. 1861734).

Lirer, L. & Vinci, A. (1991.). Grain-size distributions of pyroclastic deposits. *Sedimentology*, 38, 1075–1083.

Lu, Z. (1996). Weak consistency of nonparametric kernel regression under alpha-mixing dependence. *Chinese Sci. Bull.*, 41, 2219–2221.

Lu, Z. & Chen, P. (1997). Distribution-free strong conistency for nonparametric kernel regression involving nonlinear time series. *J. Statist. Plann. Inference*, 65, 67–86.

Lu, Z. & Chen, X. (2002). Spatial nonparametric regression estimation: Non-isotropic case. *Acta Math. Appl. Sin. Engl. Ser.*, 18(4), 641–656.

Lu, Z. & Chen, X. (2004). Spatial kernel regression estimation: Weak concistency. *Stat. Proba. Lett*, 68, 125–136.

Masry, E. & Györfi, L. (1987). Strong consistency and rates for recursive density estimators for stationary mixing processes. *J. Multivariate Anal.*, 22, 79–93.

Nerini, D., Monestiez, P., & Manté, C. (2010). A cokriging method for spatial functional. *Journal of Multivariate Analysis*, 101(2), 409–418.

Ramsay, J. & Silverman, B. (1997). *Functional data analysis.* New York: Springer Series in Statistics.

Ramsay, J. & Silverman, B. (2002). *Applied functional data analysis; Methods and case studies.* New York: Springer-Verlag.

Rio, E. (2000). *Théorie Asymptotic des Processus Aléatoires Faiblement Dépendants.* Berlin: Springer-Verlag.

Ripley, B. (1981). *Spatial Statistics.* New-York: Wiley.

Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.*, 4, 185–207.

Robinson, P. M. (1987). Time series residuals with applocation to probability density estimation. *J. Time Ser. Anal.*, 8, 329–344.

Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables. *J. Statist. Plann. Inference*, 18, 135–149.

Tran, L. T. (1990). Kernel density estimation on random fields. *J. Multivariate Anal.*, 34, 37–53.

Tran, L. T. & Yakowitz, S. (1993). Nearest neighbor estimators for random fields. *J. Multivariate Anal.*, 44, 23–46.

Truong, Y. M. & Stone, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.*, 20, 77–97.

Yakowitz, S. (1987). Nearest-neighbor methods for times series analysis. *J. Time Ser. Anal.*, 8, 235–247.

Documents de travail récents

- Xavier Chojnicki, Lionel Ragot: “Impacts of Immigration on Aging Welfare-State - An Applied General Equilibrium Model for France” [2011-2]

- Kirill Borissov, Stephane Lambrecht : “Education, Wage Inequality and Growth“ [2011-1]