Density estimation for spatial-temporal models

Liliana Forzani, Ricardo Fraiman & Pamela Llop

TEST
An Official Journal of the Spanish Society of Statistics and Operations Research

ISSN 1133-0686
Volume 22
Number 2

TEST (2013) 22:321-342
DOI 10.1007/s11749-012-0313-3

Test (2013) 22:321–342
DOI 10.1007/s11749-012-0313-3

ORIGINAL PAPER

Density estimation for spatial-temporal models

Liliana Forzani · Ricardo Fraiman · Pamela Llop

Received: 18 December 2011 / Accepted: 30 September 2012 / Published online: 14 December 2012

Abstract In this paper a k-nearest neighbor type estimator of the marginal density function for a random field which evolves with time is considered. Considering dependence, the consistency and asymptotic distribution are studied for the stationary and nonstationary cases. In particular, the parametric rate of convergence $\sqrt{T}$ is proven when the random field is stationary. The performance of the estimator is shown by applying our procedure to a real data example.

Keywords Spatio-temporal data · Density estimation · Local times

Mathematics Subject Classification 91B72 · 62G07 · 60J55

1 Introduction

In the last decade there has been a significant growth on research of functional data as well as spatial-temporal data (see Tang et al. 2008). However, not all the theory developed for curves has been extended to random fields as is the case of nonparametric marginal density estimation which so far has only been performed for $\mathbb{R}^N$-valued random fields. For instance, for mixing stationary random fields, Tran and Yakowitz (1993) proved the asymptotic normality of the k-nearest neighbor estimator. Tran

L. Forzani · P. Llop (✉)
Facultad de Ingeniería Química and Instituto de Matemática Aplicada del Litoral, UNL-CONICET, Santa Fe, Argentina
e-mail: lloppamela@gmail.com

R. Fraiman
Departamento de Matemática and Ciencias, Universidad de San Andrés, Buenos Aires, Argentina

R. Fraiman
CMAT, Universidad de la República, Montevideo, Uruguay

(1990) obtained the asymptotic normality of kernel type estimator, while Carbon et al. (1996) studied its L_1 convergence. The uniform consistency of this kind of estimator was shown by Carbon et al. (1997) and further extensions were studied by Hallin et al. (2001, 2004).

For functional data, the problem of nonparametric marginal density estimation has been considered by several authors in different setups. The case when a single sample path is observed over an increasing interval $[0, T]$ as T grows to infinity has been studied by Rosenblatt (1970), Nguyen (1979), and Castellana and Leadbetter (1986). In particular, the latter showed that for continuous time processes a parametric speed of convergence is attained by kernel type estimators. More recently, some extensions have been obtained by Blanke and Bosq (1997), Blanke (2004), Kutoyants (2004) among others. In particular, Labrador (2008) proposed a k-nearest neighbor type estimator using local time ideas. Later, Llop et al. (2011) redefined the k-nearest neighbor estimator for the case when an independent sample is observed, obtaining parametric rates of convergence and its asymptotic normality.

This work addresses the problem of nonparametric marginal density estimation for random fields which evolve in time. More precisely, given the following random field:

$$\mathcal{X}(\mathbf{s}) = \mu(\mathbf{s}) + e(\mathbf{s}), \quad \mathbf{s} \in \mathbf{S} \subset \mathbb{R}^d, \tag{1}$$

we estimate its marginal density function using a k-nearest neighbor type estimator defined via local time when a dependent sample of identically distributed random fields $\mathcal{X}_1, \ldots, \mathcal{X}_T$ is given. For this estimator we study its asymptotic properties. The functional nature of the data (a random surface) plays a fundamental role which allows us to obtain parametric rates of convergence of this density estimator in the stationary case, contrary to what generally happens in nonparametric problems. This kind of data appears naturally when analyzing the evolution of some measurements in a geographical area (such as the Amazon), or when recording responses from the brain during a time interval, among other interesting practical problems.

This paper is organized as follows: in Sect. 2 we recall some well-known dependence notions and give a new one which will be used in this work. Section 3 is dedicated to theoretical results. More precisely, in Sect. 3.2 we introduce the estimator for the stationary case, prove its consistency obtaining strong rates of convergence, and show its asymptotic normality. In Sect. 3.3, we extend the definition given in Sect. 3.2 to nonstationary random fields, and also obtain its rates of convergence. Section 4 is devoted to numerically show the performance of our estimation methods. In Sect. 4.1 a simulated example is presented for $d = 2$, and in Sect. 4.2 a real data example of fMRI images corresponding to the brain in the *resting state* is considered. Main and auxiliary proofs are given in Appendices B and C, respectively.

2 Dependence notions

In this section we will review some known dependence notions and introduce a new one which will be used in this work to find convergence rates of our density estimators. We will start with the classical α-mixing condition, introduced by Rosenblatt (1956) whose definition is the following:

Definition 1 We say that a sequence of random variables $\{X_t\}_{t\in\mathbb{N}}$ is *α-mixing* if there exists a sequence $\{\alpha(r), r \in \mathbb{N}\}$ decreasing to zero at infinity such that for each r,

$$\big|P(A \cap B) - P(A)P(B)\big| \le \alpha(r),$$

for all $A \in \mathcal{M}_{t_1}^{t_u}$ and $B \in \mathcal{M}_{l_1}^{l_v}$ with $\mathcal{M}_a^b$ the σ-field generated by the random variables $\{X_t\}_{t=a}^b$, where $1 \le t_1 \le t_u < t_u + r = l_1 \le l_v \le \infty$. If in addition there exist $0 < \rho < 1$ and $a > 0$ such that $\alpha(r) \le a\rho^r$ then, we say that the sequence $\{X_t\}_{t\in\mathbb{N}}$ is *geometrically α-mixing*.

Although this is one of the weakest notions of dependence, Doukhan and Louhichi (1999) observed that certain processes which are of interest in statistics are not α-mixing. This is the case of the autoregressive process of order one (AR(1)) given by $\mathcal{X}_t = \theta\mathcal{X}_{t-1} + \epsilon_t, t \in \mathbb{Z}$, where ϵ_t are independent and $P(\epsilon_t = 1) = P(\epsilon_t = -1) = \frac{1}{2}$, $0 < |\theta| \le \frac{1}{2}$. Inspired by this problem, the authors introduced a new notion of dependence which was called *weak dependence* (Definition 2), and using the AR(1) process they showed that this definition is even weaker than the α-mixing dependence (see Lemma 1).

Definition 2 (Doukhan and Louhichi 1999) We say that a sequence of random variables $\{X_t\}_{t\in\mathbb{N}}$ is *$(\mathcal{G}, \alpha, \psi)$-weakly dependent* if there exist a class $\mathcal{G}$ of real-valued functions, a sequence $\{\alpha(r), r \in \mathbb{N}\}$ of positive numbers decreasing to zero at infinity, and a function $\psi : \mathcal{G}^2 \times \mathbb{N}^2 \to \mathbb{R}$ such that for any u-tuple $(t_1, \ldots, t_u)$ and for any v-tuple $(l_1, \ldots, l_v)$ with $1 \le t_1 \le t_u < t_u + r = l_1 \le l_v \le \infty$,

$$\Big|\mathrm{cov}\big(f(X_{t_1}, \ldots, X_{t_u}), g(X_{l_1}, \ldots, X_{l_v})\big)\Big| \le \psi(f, g, u, v)\alpha(r),$$

for all functions $f, g \in \mathcal{G}$ defined on $\mathbb{R}^u$ and $\mathbb{R}^v$, respectively.

Lemma 1 (Doukhan and Louhichi 1999) *Let us consider the process AR*(1), $\mathcal{X}_t = \theta\mathcal{X}_{t-1} + \epsilon_t$ *with* ϵ_t *independent and* $P(\epsilon_t = 1) = P(\epsilon_t = -1) = \frac{1}{2}, 0 < |\theta| \le \frac{1}{2}$. *Then,* $\mathcal{X}_t$ *is* $(\mathcal{G}, \alpha, \psi)$*-weakly dependent with* $\mathcal{G}$ *the set of the bounded Lipschitz functions.*

Here we will use the $(\mathcal{G}, \alpha, \psi)$-weak dependence for $\mathcal{G} = \{\mathbb{I}_E : E \in \mathcal{A}\}$ the class of the indicator functions over measurable sets and $\psi(f, g, u, v) = \phi(u, v)$ with $\phi : \mathbb{N} \times \mathbb{N} \to \mathbb{R}^+$ a function greater or equal to one. This weak dependence will be called *(α, ϕ)-weak dependence* as specified in the following definition:

Definition 3 We say that a sequence of random variables $\{X_t\}_{t\in\mathbb{N}}$ is *(α, ϕ)-weakly dependent* if there exist a sequence $\{\alpha(r), r \in \mathbb{N}\}$ decreasing to zero at infinity and a function $\phi : \mathbb{N} \times \mathbb{N} \to \mathbb{R}^+$ with $\phi(u, v) \ge 1$ such that

$$\big|P(A \cap B) - P(A)P(B)\big| \le \alpha(r)\phi(u, v),$$

for all $A \in \mathcal{M}_{t_1}^{t_u}$ and $B \in \mathcal{M}_{l_1}^{l_v}$ with $\mathcal{M}_a^b$ the σ-field generated by the random variables $\{X_t\}_{t=a}^b$, where $1 \le t_1 \le t_u < t_u + r = l_1 \le l_v \le \infty$. If, in addition, there exist constants $L_1, L_2 < \infty$ and $\mu \geq 0$ such that

$$\sum_{j=0}^{\infty}(j+1)^k \alpha(j) \leq L_1 L_2^k (k!)^{\mu}, \quad \forall k \geq 0,$$

then, we say that the sequence $\{X_t\}_{t\in\mathbb{N}}$ is *geometrically* (α, ϕ)*-weakly dependent.*

The following lemma, whose proof is given in Appendix C, shows the relationship between the (geometric) α-mixing and the (geometric) (α, ϕ)-weak dependence.

Lemma 2 *If the sequence* $\{X_t\}_{t\in\mathbb{N}}$ *is* (*geometric*) α*-mixing then it is* (*geometric*) (α, ϕ)*-weakly dependent with* $L_1 = \frac{a}{1-\rho}$, $L_2 = \frac{1}{1-\rho}$, $\mu = 1$ *and any* $\phi \geq 1$.

Observe that we need to prove only the geometric case since Rosenblatt (1956) showed the explicit weak dependence structure of the α-mixing process for $\mathcal{G} = L^{\infty}$ and $\psi(f, g, u, v) = 4\|f\|_{\infty}\|g\|_{\infty}$.

3 Density estimation

In this section we will introduce the density estimators for the stationary and non-stationary cases and present some asymptotic results. We will start the section by introducing the setting of our work and some useful notation.

3.1 Setting and notation

Let $S_1, S_2, \ldots, S_d \subset \mathbb{R}$ be finite intervals and let $\mathbf{S} \doteq S_1 \times S_2 \times \cdots \times S_d \in \mathbb{R}^d$ be a rectangle of measure $|\mathbf{S}|$. Let us consider the first-order stationary random field $\{e(\mathbf{s}, \omega) \in \mathbb{R} : \mathbf{s} \in \mathbf{S},\ \omega \in \Omega\}$ defined in a probability space $(\Omega, \mathcal{A}, P)$ with unknown marginal density function f_e, and which admits a local time. More precisely, if λ is the occupation measure of $e(\mathbf{s}, \omega)$,

$$\lambda(A, e) \doteq \lambda\big(A, e(\mathbf{s}, \omega)\big) = \int_{\mathbf{S}} \mathbb{I}_A\big(e(\mathbf{s}, \omega)\big)\, d\mathbf{s}, \quad A \in \mathcal{B}(\mathbb{R}), \quad \omega \in \Omega,$$

where $\mathcal{B}(\mathbb{R})$ stands for the Borel sigma–algebra on $\mathbb{R}$, and if λ is absolutely continuous with respect to the Lebesgue measure, then the local time is defined as a regular version of the Radon–Nikodym derivative $l_T(\cdot, e) \doteq l_T(\cdot, e(\mathbf{s}, \omega))$ for almost all ω (from now on $e(\mathbf{s})$). In this case we can write

$$\lambda(A, e) = \int_A l_T(u, e)\, du.$$

Given a sequence of random fields $\{e_t(\mathbf{s}, \omega)\}_{t=1}^{T}$ of $e(\mathbf{s}, \omega)$ (from now on $e(\mathbf{s})$ and $e_t(\mathbf{s})$), $I_{(x,r)} = [x - r, x + r]$ the interval of center x and radius r, and $\{k_T\}$ being a sequence of real numbers such that $k_T/T < |\mathbf{S}|$, we define the random variable $h_T^e(x)$ such that the time spend by $\{e_t(\mathbf{s})\}_{t=1}^T$ at $I_{(x,h_T^e(x))}$ be k_T. That is,

$$k_T = \sum_{t=1}^{T} \int_{\mathbf{S}} \mathbb{I}_{I_{(x,h_T^e(x))}}\big(e_t(\mathbf{s})\big)\, d\mathbf{s}. \tag{2}$$

Lemma 3 *If $\mathcal{X}(\mathbf{s})$ admits a local time, then the random variable h_T^e exists and it is unique.*

Next we state some assumptions which will be used in this work.

H1 $\{e_t(\mathbf{s}), \mathbf{s} \in \mathbf{S}\}_{t=1}^T$ is a sequence of random fields with the same distribution as $e(\mathbf{s})$, where $e(\mathbf{s})$ is a first-order stationary random field with unknown strictly positive density function f_e which admits a local time;
H2 the density f_e is a strictly positive Lipschitz function with constant K;
H2′ the density f_e has two bounded derivatives;
H3 for each $\mathbf{s}$ fixed, $\{e_t(\mathbf{s})\}_{t=1}^T$ is geometrically (α, ϕ)-weakly dependent with ϕ some of the following functions:

- $\phi(u, v) = 2v$,
- $\phi(u, v) = u + v$,
- $\phi(u, v) = uv$,
- $\phi(u, v) = \rho(u + v) + (1 - \rho)uv$, for some $\rho \in (0, 1)$;

H3′ for each $\mathbf{s}$ fixed, $\{e_t(\mathbf{s})\}_{t=1}^T$ is geometrically (α, ϕ)-mixing;
H4 $\{k_T\}$ and $\{v_T\}$ are sequences of positive integers which converge to infinity such that $v_T(\frac{k_T}{T}) \to 0$ and $\frac{1}{T}(\frac{k_T}{v_T})^2 \to \infty$.
H4′ $\{k_T\}$ and $\{v_T\}$ are sequences of positive integers which converge to infinity such that $v_T(\frac{k_T}{T}) \to 0$ and $\frac{k_T}{v_T} \to \infty$.
H5 for each $\mathbf{s}$ fixed, $\{e_t(\mathbf{s})\}_{t=1}^T$ is α-mixing with the mixing coefficients $\alpha(r)$ verifying

$$N \sum_{r=N}^{\infty} \alpha(r) \to 0 \quad \text{as } N \to \infty.$$

H6 For each $c > 0$,

$$c^{-2} c_T^{-2} \int_{\mathbf{S}} \int_{\mathbf{S}} \int_{\{x:|u-x|\le cc_T\}} \int_{\{x:|v-x|\le cc_T\}} \big(f_{\mathbf{sr}}(u, v) - f_{\mathcal{X}}(u) f_{\mathcal{X}}(v)\big)\, du\, dv\, d\mathbf{s}\, d\mathbf{r}$$
$$\to \int_{\mathbf{S}\times\mathbf{S}} \big(f_{\mathbf{sr}}(x, x) - f_e^2(x)\big) d\mathbf{s}\, d\mathbf{r} \doteq c_0^2(x) > 0,$$

where $f_{\mathbf{sr}}$ is the joint density of $(e(\mathbf{s}), e(\mathbf{r}))$ and $c_T = \frac{k_T}{T}$ with $\{k_T\}$ is a sequence of positive integers going to infinity such that $\frac{k_T}{T} \to 0$.
H7 For some $c > 0$,

$$\left| \int_{\mathbf{S}} \int_{\mathbf{S}} \int_{\{x:|u-x|\le cc_T\}} \int_{\{x:|v-x|\le cc_T\}} \big(f_{\mathbf{sr}}(u, v) - f_e(u) f_e(v)\big)\, du\, dv\, d\mathbf{s}\, d\mathbf{r} \right| \le cc_T^4,$$

where again $f_{\mathbf{sr}}$ is the joint density of $(e(\mathbf{s}), e(\mathbf{r}))$ and $c_T = \frac{k_T}{T}$ with $\{k_T\}$ is a sequence of positive integers going to infinity such that $\frac{k_T}{T} \to 0$.

Remark 1 The choice of the function ϕ in H3 is related with the exponential inequality (Theorem 4) used in the proofs (see Doukhan and Neumann 2006).

The symbol C will denote generic constants whose value may be different in each occurrence.

3.2 The stationary case: $\mu(\mathbf{s}) = \mu$ constant

If in model (1) we assume that the deterministic unknown mean function $\mu(\mathbf{s})$ is constant with respect to the space (that is, $\mu(\mathbf{s}) = \mu$) then the process $\mathcal{X}(\mathbf{s})$ (with density $f_{\mathcal{X}}$) inherits the stationarity of $e(\mathbf{s})$. In this case, given a sample sequence of random fields $\{\mathcal{X}_t(\mathbf{s})\}_{t=1}^T$ and a sequence $\{k_T\}$ of real numbers such that $k_T/T < |\mathbf{S}|$ we define the estimator of the density function $f_{\mathcal{X}}$ as

$$\widehat{f}_{\mathcal{X}}(x) \doteq \frac{k_T}{2T|\mathbf{S}|h_T^{\mathcal{X}}(x)}, \tag{3}$$

where the random variable $h_T^{\mathcal{X}}(x)$ is defined as in (2) but with $\mathcal{X}$ instead of e.

Remark 2 From Lemma 3, if $\mathcal{X}(\mathbf{s})$ admits a local time then $\widehat{f}_{\mathcal{X}}(x)$ is well defined.

If $\{\mathcal{X}_t(\mathbf{s})\}_{t=1}^T$ is an iid sample of $\mathcal{X}(\mathbf{s})$, a direct extension to $\mathbb{R}^d$ of the results given in Llop et al. (2011) will allow us obtain the consistency of the estimator (3). However, in this work we will not assume independence of the data but the processes will be (α, ϕ)-weakly dependent and therefore we will need another exponential inequalities to obtain rates of convergence.

Theorem 1 (Rates of convergence) *For each $x \in \mathbb{R}$ we have*

$$\lim_{T\to\infty} v_T\big(\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\big) = 0 \quad a.co.$$

in the following two cases:

(a) *weakly dependent case*: *for $\mathcal{X}$ fulfilling* H1, H2, H3 *and* H4;
(b) *α-mixing case*: *for $\mathcal{X}$ fulfilling* H1, H2, H3′ *and* H4′.

Remark 3

(a) Weakly dependent case: Our assumptions imply that we can choose k_T such that $v_T = T^\gamma$ for any $\gamma < \frac{1}{4}$. More precisely, if $k_T = T^\beta$ and $v_T = T^\gamma$, in order that conditions

$$v_T\left(\frac{k_T}{T}\right) \to 0 \quad \text{and} \quad \frac{1}{T}\left(\frac{k_T}{v_T}\right)^2 \to \infty$$

are met, it enough that $\beta - 1 + \gamma < 0$ and $\beta - \gamma - \frac{1}{2} > 0$ or equivalently, $\beta < 1 - \gamma$ and $\beta > \gamma + \frac{1}{2}$ from which follows that given $\gamma < \frac{1}{4}$ we can choose β such that the conditions hold.

(b) α-mixing case: Our assumptions imply that we can choose k_T such that $\upsilon_T = T^{\gamma}$ for any $\gamma < \frac{1}{2}$. More precisely, if $k_T = T^{\beta}$ and $\upsilon_T = T^{\gamma}$, in order that conditions

$$\upsilon_T \left(\frac{k_T}{T} \right) \to 0 \quad \text{and} \quad \frac{k_T}{\upsilon_T} \to \infty$$

are met, it enough that $\beta - 1 + \gamma < 0$ and $\beta - \gamma > 0$ or equivalently, $\beta < 1 - \gamma$ and $\beta > \gamma$ from which follows that given $\gamma < \frac{1}{2}$ we can choose β such that the conditions hold.

Next we show the asymptotic normality of our estimator assuming the classical α-mixing dependence condition. The result can still be shown for $(\mathcal{G}, \alpha, \psi)$-weak dependent sequences (using for instance Theorem 6.1 in Neumann and Paparoditis 2008), although it requires some technical hypotheses which we would like to avoid. On the other hand, for (α, ϕ)-weak dependent sequences there is not yet a Central Limit Theorem (CLT from now on) available. As in Castellana and Leadbetter (1986) and Llop et al. (2011), the functional nature of the data makes possible to attain a parametric rate of convergence $\sqrt{T}$ for our density estimator as stated in the following result:

Theorem 2 (Asymptotic normality) *Assume that* H1, H2$'$ *and* H5 *hold for* $\mathcal{X}(\mathbf{s})$. *Choose a sequence* $\{k_T\}$ *of positive real numbers going to infinite such that*

$$k_T^2/T^{3/2} \to 0 \quad \textit{and} \quad k_T^2/T \to \infty. \tag{4}$$

For that k_T *assume that* H6 *and* H7 *hold. Then, for all* $x \in \mathbb{R}$

$$\sqrt{T}\big(\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\big) \to \mathcal{N}\left(0, \frac{2|\mathbf{S}|}{c_o}\right) \quad \textit{in distribution.}$$

Remark 4 Observe that for this result we ask α to verify H5 which is weaker than the condition for the (α, ϕ)-weak dependence required in Theorem 1. However, for this result we require H7, which is another condition on the decay of the covariances.

3.3 The nonstationary case: $\mu(\mathbf{s})$ any function

If in model (1) we assume that the mean function $\mu(\mathbf{s})$ is any function then the process $\mathcal{X}(\mathbf{s})$ (with density $f_{\mathcal{X}_\mathrm{s}}$) is not stationary any more. In this case, given a sample sequence of random fields $\{\mathcal{X}_t(\mathbf{s})\}_{t=1}^{T}$ and a sequence $\{k_T\}$ of real numbers such that $k_T/T < |\mathbf{S}|$ we define the estimator of the density function $f_{\mathcal{X}_\mathrm{s}}$ as

$$\widehat{f}_{\mathcal{X}_\mathrm{s}}(x) = \widehat{f}_u\big(x - \bar{\mathcal{X}}_T(\mathbf{s})\big),$$

being

$$\widehat{f}_u(x) \doteq \frac{k_T}{2T|\mathbf{S}|h_T^u(x)} \tag{5}$$

with $u = \{\mathcal{U}_{T1}, \ldots, \mathcal{U}_{TT}\}$ given by $\mathcal{U}_{Tt}(\mathbf{s}) = \mathcal{X}_t(\mathbf{s}) - \bar{\mathcal{X}}_T(\mathbf{s}) = e_t(\mathbf{s}) - \bar{e}_T(\mathbf{s})$. Here $\{e_1(\mathbf{s}), \ldots, e_T(\mathbf{s})\}$ is a sequence with the same distribution as $e(\mathbf{s})$, $\bar{e}_T(\mathbf{s}) = \frac{1}{T}\sum_{t=1}^{T} e_t(\mathbf{s})$ and h_T^u is defined as in (2) replacing $\{e_t(\mathbf{s})\}_{t=1}^{T}$ by u.

Remark 5 From Lemma 3, if $\mathcal{X}(\mathbf{s})$ admits a local time then $\widehat{f}_{\mathcal{X}_{\mathbf{s}}}(x)$ is well defined.

For every $\mathbf{s}$ fixed, the random variables $\{\mathcal{U}_{T1}(\mathbf{s}), \ldots, \mathcal{U}_{TT}(\mathbf{s})\}$ with $\mathbb{E}(\mathcal{U}_{Tt}(\mathbf{s})) = 0$ are identically distributed but not necessarily (α, ϕ)-weakly dependent. Therefore, we cannot use directly the results given in Sect. 3.2. However, we can still prove the complete convergence of the estimator of $f_{\mathcal{X}_{\mathbf{s}}}$ and obtain rates of convergence.

Theorem 3 (Rates of convergence) *Assume* H1–H3 *and choose two sequences* $\{k_T\}$ *and* $\{v_T\}$ *of positive real numbers converging to infinity such that, for each fixed* $\mathbf{s}$, $v_T(T/k_T)|\bar{e}_T(\mathbf{s})| \to 0$ *a.co. For that sequences, suppose that* H4 *holds. Then for each* $x \in \mathbb{R}$

$$\lim_{T\to\infty} v_T\left(\widehat{f}_{\mathcal{X}_{\mathbf{s}}}(x) - f_{\mathcal{X}_{\mathbf{s}}}(x)\right) = 0 \quad a.co.$$

Remark 6 If for $\mathbf{s}$ fixed the sequence $\{e_t(\mathbf{s})\}_{t=1}^{T}$ is geometric α-mixing with $|e(\mathbf{s})| < M$ for some constant $M > 0$, we can choose k_T such that $v_T = T^\gamma$ for any $\gamma < \frac{1}{4}$. More precisely, let $k_T = T^\beta$ and $v_T = T^\gamma$. In order that conditions

$$v_T\left(\frac{k_T}{T}\right) \to 0 \quad \text{and} \quad \frac{1}{T}\left(\frac{k_T}{v_T}\right)^2 \to \infty$$

are met, it is sufficient that $\beta < 1 - \gamma$ and $\beta > \gamma + \frac{1}{2}$. In addition, in order that $v_T(T/k_T)|\bar{e}_T(\mathbf{s})| \to 0$ a.co., as $\bar{e}_T(\mathbf{s}) = o(T^{-\alpha})$ with $\alpha < \frac{1}{2}$ it would be $\beta > \gamma + \frac{1}{2}$ from which follows that given $\gamma < \frac{1}{4}$ we can choose β such that the conditions hold.

Remark 7 In the stationary case, we got a faster rate of convergence in the geometric α-mixing case than in the geometric (α, ϕ)-weak case. However, we are not able to do the same in the nonstationary case since here we need the condition $v_T(T/k_T)|\bar{e}_T(\mathbf{s})| \to 0$ to hold and therefore $v_T = T^\gamma$ for $\gamma < \frac{1}{4}$.

4 Numerical examples

4.1 A simulated example

In this section we illustrate the performance of our method by estimating the marginal density function of a nonstationary random field process $\mathcal{X}_t(\mathbf{s}) : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\mathcal{X}_t(\mathbf{s}) = \mu(\mathbf{s}) + e_t(\mathbf{s}) + \theta e_{t-1}(\mathbf{s}), \qquad \theta \in [0, 1],$$
$$\mathbf{s} = (s_1, s_2) \in \mathbf{S} = (0, 1] \times (0, 1], \qquad t \geq 1,$$

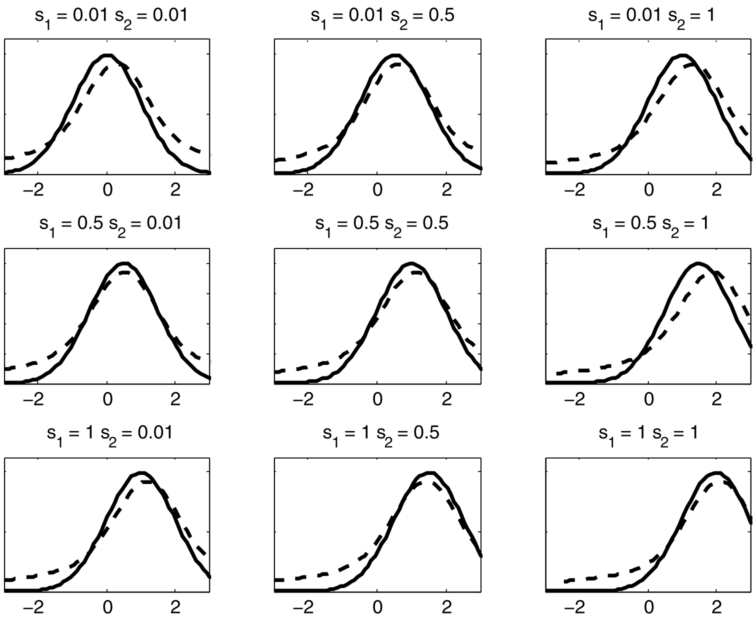

Fig. 1 Estimated (*dashed curve*) and theoretical (*solid curve*) density functions of $\mathcal{X}_t(\mathbf{s})$ for $\mu(s_1, s_2) = s_1 + s_2$, $(s_1, s_2) \in (0, 1] \times (0, 1]$ for $\theta = 1/2$

where $\mu(\mathbf{s}) = \mu(s_1, s_2) = s_1 + s_2$ and, for each t, $e_t(\mathbf{s})$ is a Gaussian stationary random field with $\mathbb{E}(e_t(\mathbf{s})) = 0$, $\mathrm{var}(e_t(\mathbf{s})) = 1$, $e_t(\mathbf{s})$ independent of $e_l(\mathbf{s})$, for $t \neq l$ and $\mathrm{cov}(e_t(\mathbf{s}), e_t(\mathbf{r})) = \exp\{-d(\mathbf{s}, \mathbf{r})/0.25\}$ with $d(\mathbf{s}, \mathbf{r})$ the Euclidean distance between $\mathbf{s}$ and $\mathbf{r}$.

We use the estimator defined in (5) for with 100 equispaced points in time and 100×100 in space and compute the L_1-norm between the estimated (computed with cross-validated parameter) and the true density for different for different values of θ:

θ	L_1-norm	k_T
0	0.21	19.1
0.5	0.32	26.4
0.99	0.57	29.2

As expected, the good behavior of the estimator decreases as θ increases. In Fig. 1 we plot the estimator (dashed line) and the theoretical density (solid line) for some points of $\mathbf{S}$ for $\theta = 0.5$. As we can see, the estimator fits very well the true density except in the tails where, due to the nature of the field and classical border effects, we are not able to perform a good estimation.

4.2 A real data example: brain activity

In neuroscience, it is well known that BOLD (blood-oxygen-level-dependence) activity in gray matter and white matter of the brain are different. The activity in areas

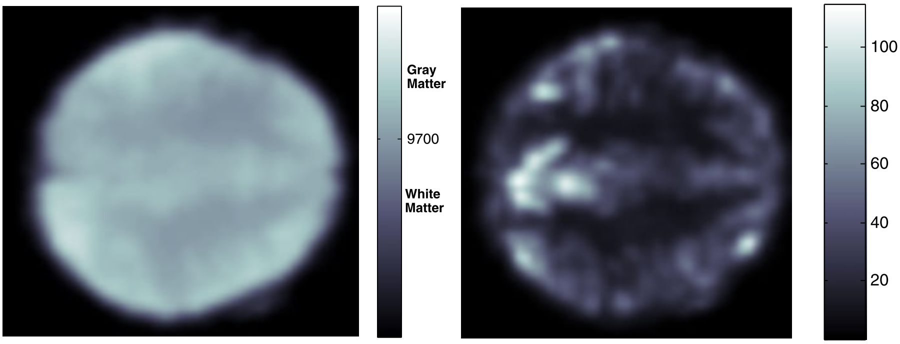

Fig. 2 (*Left*) Estimated mean. (*Right*) Estimated scale function

of gray matter is higher than in the white matter, which is reflected in the difference between the mean $\mu_{\text{gray}}(s,t)$ and $\mu_{\text{white}}(s,t)$, the first being greater than the second one. One question that arises is whether the fluctuations around the mean may be detectable at the level of the distribution, i.e. if the density of the centered data is different from the BOLD activity at the gray and the white matter. To do this type of analysis, functional magnetic resonance imaging (from now fMRI) is used with two types of experimental paradigm: the common stimulus response paradigm and the "nonstimulus experiment". When the brain is in absence of stimulus, we say that it is in a "resting state". The brain at this condition has rich dynamics (see Fox and Raichle 2007), so this state is the most useful to study brain activity. Here we analyze a two dimensional (slice) axial cut of the brain at a resting state condition. The data studied correspond to one subject from the 1000 Functional Connectomes Project (see Chao-Gan and Yu-Feng 2010a, 2010b). For the same axial cut $T = 225$ images of the brain were measured. Each image contains the same $\mathbf{s} = 1532$ spatial points (voxels) at different times. Among the 1532 voxels 617 of them corresponds to the brain white matter and 884 to brain gray matter. The plots of the mean and scale functions are given in the next two figures (Fig. 2). We first standardize the data subtracting the mean value image and dividing by the scale image. Next we estimate the stationary density functions of the white and gray matter separately using the corresponding cross-validated parameters k_T. The estimates obtained for the brain gray matter and the white gray matter (for the standardized data) are very close (see Fig. 3), which suggests that the main differences are just explained by the mean and the scale functions.

5 Conclusions

In this paper the problem of estimating the marginal density function of a random field that evolves in time was studied. For this purpose, a nearest-neighbor type estimator based on the occupation measure was defined for stationary and nonstationary random fields. Under dependence, strong rates of convergence were obtained and, for the stationary case, the asymptotic normality was also proven. The functional structure of the data allowed us to obtain parametric rates of convergence $\sqrt{T}$ contrary

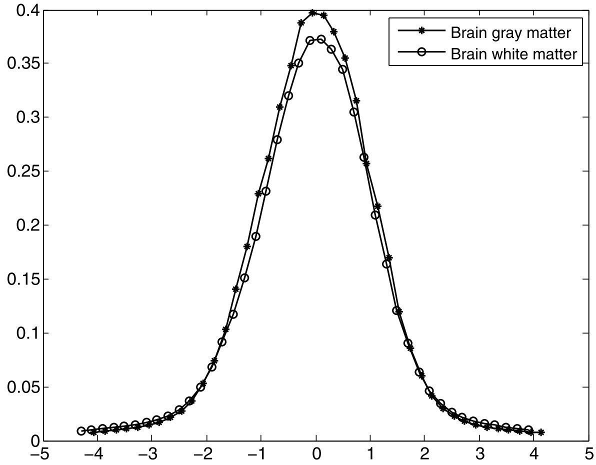

Fig. 3 Estimated stationary density functions of brain gray ($k_T = 63.6$) and white matter ($k_T = 46.7$)

to what generally happens in nonparametric problems. A small simulation study was performed and a real data example was analyzed to check the performance of the estimator.

Acknowledgements We are most grateful to Daniel Fraiman for his very helpful insights to analyze the brain fMRI data. We would also like to thank Roberto Scotto for helpful discussions. The authors were supported by PICT2008-0921, PICT2008-0622, PI 62-309 and PIP 112-200801-0218.

Appendix A: Auxiliary results

Theorem 4 (Doukhan and Neumann 2006) *Suppose that $X_1, X_2, \ldots, X_T$ are real-valued random variables defined on a probability space $(\Omega, \mathcal{A}, P)$ with $\mathbb{E}(X_t) = 0$ and $P(|X_t| \leq M) = 1$, for all $t = 1, \ldots, T$ and some $M < \infty$. Let $S_T = \sum_{t=1}^{T} X_t$ and $\phi : \mathbb{N}^2 \to \mathbb{R}^+$ be one of the following functions*:

- $\phi(u, v) = 2v$;
- $\phi(u, v) = u + v$;
- $\phi(u, v) = uv$;
- $\phi(u, v) = \rho(u + v) + (1 - \rho)uv$, *for some* $\rho \in (0, 1)$.

We assume that there exist constants $K, L_1, L_2 < \infty$, $\mu \geq 0$ and a nonincreasing sequence of real coefficients $\{\alpha(n)\}_{n\geq 1}$ such that, for all u-tuples $(t_1, \ldots, t_u)$ and all v-tuples $(l_1, \ldots, l_v)$ with $1 \leq t_1 \leq t_u < t_u + r = l_1 \leq l_v \leq \infty$ the following inequality is fulfilled:

$$\left|\operatorname{cov}(X_{t_1} \cdots X_{t_u}, X_{l_1} \cdots X_{l_v})\right| \leq K^2 M^{u+v-2} \phi(u, v) \alpha(r),$$

where

$$\sum_{j=0}^{\infty} (j+1)^k \alpha(j) \leq L_1 L_2^k (k!)^{\mu}, \quad \forall k \geq 0.$$

Then,

$$P(S_T \geq \epsilon) \leq \exp\left(-\frac{\epsilon^2/2}{\Sigma_T + \Gamma_T^{1/(\mu+2)}\epsilon^{(2\mu+3)/(\mu+2)}}\right),$$

where Σ_T can be chosen as any number greater than or equal to $\sigma_T^2 = \operatorname{var}(S_T)$ and

$$\Gamma_T = 2(K \vee M)L_2\left(\left(\frac{2^{4+\mu}TK^2L_1}{\Sigma_T}\right) \vee 1\right).$$

Theorem 5 (Doukhan et al. 1984) *Let $X_1, X_2, \ldots, X_T$ be a sequence of α-mixing random variables verifying $\mathbb{E}(X_t) = 0$ and $|X_t| \leq 1$, for all $t = 1, \ldots, T$. Let $S_T = \sum_{t=1}^{T} X_t$ and denote $\gamma = 2/(1-\theta)$ and $\sigma = \sup_{1\leq t\leq T}\{\mathbb{E}(|X_t|^\gamma)^{1/\gamma}\}$. Then, there exist constants C_1 and C_2 which depend only on the mixing coefficients, such that for $0 < \theta < 1$,*

$$P(S_T \geq \epsilon) \leq C_1 \frac{1}{\theta} \exp\left(-\frac{C_{2,T}\,\epsilon^{1/2}}{T^{1/4}\sigma^{1/2}}\right),$$

where $C_{2,T} = C_2$ if $T^{1/2}\sigma \leq 1$ and $C_{2,T} = C_2 T^{1/4}\sigma^{1/2}$ if $T^{1/2}\sigma > 1$.

Theorem 6 (Robinson 1983) *Let $\{V_{tT}\}_{t=1}^{T}$ be a triangular array of zero mean random variables and $\{b_T\}_{T\geq 1}$ a sequence of positive constants such that:*

- (i) *for each T, V_{tT}, $t = 1, \ldots, T$ are identically distributed and α-mixing with the mixing coefficients $\alpha(r)$ verifying*

$$N \sum_{r=N}^{\infty} \alpha(r) \to 0 \quad \text{as } N \to \infty;$$

- (ii) *there exists $M > 0$ such that $P(|V_{tT}| \leq M) = 1$ for all $t = 1, \ldots, T$;*
- (iii) *$b_T \to 0$ and $Tb_T \to \infty$ as $T \to \infty$;*
- (iv) *there exists $\sigma^2 > 0$ such that $\mathbb{E}(V_{tT}^2)/b_T \to \sigma^2$ as $T \to \infty$;*
- (v) *$\mathbb{E}(|V_{tT}V_{(t+s)T}|) \leq Cb_T^2$ for $s \geq 1$, $t = 1, \ldots, T$ where C is independent of T.*

Then

$$\frac{1}{\sqrt{Tb_T}} \sum_{t=1}^{T} V_{tT} \to \mathcal{N}(0, \sigma^2) \quad \text{in distribution.}$$

Appendix B: Proof of main results

Proof of Theorem 1 Let $x \in \mathbb{R}$ be fixed. By definition of complete convergence we need to show that for all $\epsilon > 0$,

$$\sum_{T=1}^{\infty} P\left(v_T\left|\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\right| > \epsilon\right) < \infty.$$

Using the definition of the estimator $\widehat{f}_{\mathcal{X}}(x)$, it is enough to prove that

$$\sum_{T=1}^{\infty} P(A_T) < \infty \quad \text{and} \quad \sum_{T=1}^{\infty} P(B_T) < \infty, \tag{6}$$

where for $\epsilon_T \doteq \frac{\epsilon}{v_T}$, the sets A_T and B_T are defined by

$$A_T = A_T(x) \doteq \left\{ h_T^{\mathcal{X}}(x) < \frac{k_T}{2T|\mathbf{S}|(f_{\mathcal{X}}(x) + \epsilon_T)} \right\}$$

and

$$B_T = B_T(x) \doteq \begin{cases} \{h_T^{\mathcal{X}}(x) > \frac{k_T}{2T|\mathbf{S}|(f_{\mathcal{X}}(x) - \epsilon_T)}\} & \text{si } f_{\mathcal{X}}(x) > \epsilon_T \\ \emptyset & \text{si } f_{\mathcal{X}}(x) \le \epsilon_T. \end{cases}$$

To prove the left-side inequality of (6) (the proof of the right-side inequality is identical and it will be omitted), let us define $a_T = a_T(x) \doteq \frac{k_T}{2T|\mathbf{S}|(f_{\mathcal{X}}(x)+\epsilon_T)}$ so that $A_T = \{h_T^{\mathcal{X}} < a_T\}$. From the equivalence

$$h_T^{\mathcal{X}} < a_T \Leftrightarrow \sum_{t=1}^{T} \underbrace{\int_{\mathbf{S}} \mathbb{I}_{(x-a_T, x+a_T)}\big(\mathcal{X}_t(\mathbf{s})\big)\, d\mathbf{s}}_{\doteq Y_{Tt}} > k_T,$$

we have

$$P(A_T) = P\left(\sum_{t=1}^{T} Y_{Tt} > k_T \right). \tag{7}$$

Next we define $\overline{Y}_{Tt} \doteq Y_{Tt} - \mathbb{E}(Y_{Tt})$ and the probability $p_T \doteq P(\mathcal{X}_t(\mathbf{s}) \in (x - a_T, x + a_T))$. Since $\mathbb{E}(Y_{Tt}) = |\mathbf{S}| p_T$, using the definition of a_T we get

$$P(A_T) = P\left(\sum_{t=1}^{T} \overline{Y}_{Tt} > k_T \left(1 - \frac{1}{f_{\mathcal{X}}(x) + \epsilon_T} \frac{p_T}{2a_T} \right) \right). \tag{8}$$

By the Mean Value Theorem there exists $x_T \in (x - a_T, x + a_T)$ for which $\frac{p_T}{2a_T} = f_{\mathcal{X}}(x_T)$. In addition, by condition H2, the definition of a_T and the fact that $\epsilon_T \to 0$ we have

$$\left| \frac{p_T}{2a_T} - f_{\mathcal{X}}(x) \right| = \left| f_{\mathcal{X}}(x_T) - f_{\mathcal{X}}(x) \right| \le K|x_T - x| \le K a_T \le \frac{\epsilon_T}{2},$$

from which it follows that $\frac{p_T}{2a_T} \le f_{\mathcal{X}}(x) + \frac{\epsilon_T}{2}$ and then

$$1 - \frac{1}{f_{\mathcal{X}}(x) + \epsilon_T} \frac{p_T}{2a_T} \ge \frac{\epsilon_T/2}{f_{\mathcal{X}}(x) + \epsilon_T} \ge C(x, \epsilon) \frac{1}{v_T}.$$

Therefore, with this inequality in (8) we get

$$P(A_T) \leq P\left(\sum_{t=1}^{T} \overline{Y}_{Tt} > C\frac{k_T}{v_T}\right). \tag{9}$$

(a) Weakly dependent case: in order to use a Bernstein type inequality in (9), we need the following Lemma which will be proved in Appendix B.

Lemma 4 *Under* H3 *for* $\mathcal{X}(\mathbf{s})$ *we have*

(i) *for any* u*-tuple* $(t_1, \dots, t_u)$ *and any* v*-tuple* $(l_1, \dots, l_v)$, $1 \leq t_1 \leq t_u < t_u + r = l_1 \leq l_v \leq T$ *we have*

$$\left|\operatorname{cov}(\overline{Y}_{Tt_1} \cdots \overline{Y}_{Tt_u}, \overline{Y}_{Tl_1} \cdots \overline{Y}_{Tl_v})\right| \leq \left(2|\mathbf{S}|\right)^{u+v} \phi(u, v)\alpha(r),$$

where $2|\mathbf{S}|$ *is such that* $|\overline{Y}_{Tt}| \leq 2|\mathbf{S}|$ *for all* t, $\alpha(r) \to 0$ *with* ϕ *any of the functions given in* H3;

(ii) *for some constant* C,

$$\operatorname{var}\left(\sum_{t=1}^{T} \overline{Y}_{Tt}\right) \leq CT.$$

Therefore, Lemma 4(i) implies that the sequence $\{\overline{Y}_{Tt}\}_{t=1}^{T}$ is $(\mathcal{G}, \alpha, \psi)$-weakly with $f : \mathbb{R}^u \to \mathbb{R}$ and $g : \mathbb{R}^v \to \mathbb{R}$ given by $f(\overline{Y}_{Tt_1}, \dots, \overline{Y}_{Tt_u}) = \overline{Y}_{Tt_1} \cdots \overline{Y}_{Tt_u}$ and $g(\overline{Y}_{Tl_1}, \dots, \overline{Y}_{Tl_v}) = \overline{Y}_{Tl_1} \dots \overline{Y}_{Tl_v}$, respectively, and $\psi : \mathcal{G}^2 \times \mathbb{N}^2 \to \mathbb{R}^+$ defined by $\psi(f, g, u, v) = (2|\mathbf{S}|)^{u+v}\phi(u, v)$ with ϕ any of the four functions given in H3. Therefore, applying Theorem 4 with $K = (2|\mathbf{S}|)^2$, $M = 2|\mathbf{S}|$, $\Gamma_T = C$ and $\Sigma_T = CT$ in (9) we have

$$\begin{aligned}\sum_{T=1}^{\infty} P(A_T) &\leq T_0 + \sum_{T=T_0}^{\infty} \exp\left(-\frac{Ck_T^2}{Cv_T^2 T + Cv_T^2(\frac{k_T}{v_T})^{(2\mu+3)/(\mu+2)}}\right)\\ &\leq T_0 + \sum_{T=T_0}^{\infty} \exp\left(-\frac{C}{C(\frac{v_T}{k_T})^2 T + C(\frac{v_T}{k_T})^{1/(\mu+2)}}\right),\end{aligned}$$

and the result follows since $(2\mu + 3)/(\mu + 2) < 2$ and H4 $(\frac{v_T}{k_T})^2 T \to 0$ and, as a consequence, $\frac{v_T}{k_T} \to 0$.

(b) α-mixing case: under H1, H2, H3′ and H4′ for $\mathcal{X}(\mathbf{s})$. Since $\mathcal{X}_t$ verifies H5, $\overline{Y}_{Tt}$ inherits the same condition and then we apply the Bernstein inequality given in Theorem 5 to the random variables $Z_{Tt} \doteq \frac{\overline{Y}_{tT}}{2|\mathbf{S}|}$ with $|Z_{Tt}| \leq 1$ and $\mathbb{E}(Z_{Tt}) = 0$. For $\gamma = \frac{2}{1-\theta}$ with $0 < \theta < 1$, it is easy to verify that $\mathbb{E}(|Z_{Tt}|^\gamma)^{1/\gamma} \leq (2|\mathbf{S}|)^{1-\gamma}$ so that $\sigma \leq (2|\mathbf{S}|)^{1-\gamma}$ and then, since $\frac{C_{2,T}}{T^{1/4}\sigma^{1/2}} \geq C$ in (9) we have

$$\sum_{T=1}^{\infty} P(A_T) \leq T_0 + C\frac{1}{\theta} \sum_{T=T_0}^{\infty} \exp\left(-C\sqrt{\frac{k_T}{v_T}}\right),$$

which is finite by H4′. □

Proof of Theorem 2 For a fixed $x \in \mathbb{R}$ we define

$$S_T = \sum_{t=1}^{T} Y_{Tt}, \quad s_T^2 = \mathrm{var}(S_T) \quad \text{and} \quad a_T = \frac{k_T}{2T|\mathbf{S}|(f_{\mathcal{X}}(x) + \frac{z}{\sqrt{T}})}.$$

From the definition of $\widehat{f_{\mathcal{X}}}(x)$ and analogously to (7) we have

$$P\big(\sqrt{T}\big(\widehat{f_{\mathcal{X}}}(x) - f_{\mathcal{X}}(x)\big) \le z\big) = P\big(h_T^{\mathcal{X}} \ge a_T\big)$$
$$= P\left(\frac{S_T - \mathbb{E}(S_T)}{s_T} \le \frac{k_T - T|\mathbf{S}|p_T}{s_T}\right).$$

Then the proof will be completed if we prove:

(a) $\frac{S_T - \mathbb{E}(S_T)}{s_T} \to \mathcal{N}(0, 1)$ in distribution;

(b) $\frac{k_T - T|\mathbf{S}|p_T}{s_T} \to \frac{2|\mathbf{S}|}{c_o} z$.

(a) It will a consequence of $\frac{S_T - \mathbb{E}(S_T)}{\sqrt{T}a_T} = \frac{1}{\sqrt{T}a_T} \sum_{t=1}^{T} \overline{Y}_{tT} \to \mathcal{N}(0, c_0^2)$ and $\frac{s_T^2}{Ta_T^2} \to c_0^2$ where c_0^2 is given in H6. To prove the second part observe that

$$s_T^2 = \mathrm{var}(S_T) = \sum_{t=1}^{T} \mathrm{var}(Y_{Tt}) + 2\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \mathrm{cov}(Y_{Tt}, Y_{Ts})$$
$$= \sum_{t=1}^{T} \mathrm{var}(Y_{Tt}) + 2\sum_{t=1}^{T-1}\sum_{l=1}^{T-t} \mathrm{cov}(Y_{Tt}, Y_{Tt+l}) \quad (\text{taking } l = s - t).$$

Then,

$$\frac{s_T^2}{Ta_T^2} = \frac{1}{Ta_T^2}\sum_{t=1}^{T} \mathrm{var}(Y_{Tt}) + 2\frac{1}{Ta_T^2}\sum_{t=1}^{T-1}\sum_{l=1}^{T-t} \mathrm{cov}(Y_{Tt}, Y_{Tl+l}) \doteq I + II. \tag{10}$$

Since $a_T \sim k_T/T$, from H6 we get

$$\frac{1}{a_T^2}\mathrm{var}(Y_{Tt})$$
$$= \frac{1}{a_T^2}\int_{\mathbf{S}}\int_{\mathbf{S}}\int_{\{u:|u-x|\le a_T\}}\int_{\{v:|v-x|\le a_T\}} \big(f_{st}(u, v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)\big)\,du\,dv\,dt\,ds \to c_0^2 > 0, \tag{11}$$

from which follows that $I \to c_0^2$ as $T \to \infty$. On the other hand, for N integer we write,

$$|II| \le 2\frac{1}{Ta_T^2}\sum_{t=1}^{T-1}\sum_{l=1}^{N-1}\big|\mathrm{cov}(Y_{Tt}, Y_{Tt+l})\big| + 2\frac{1}{Ta_T^2}\sum_{t=1}^{T-1}\sum_{l=N}^{T-t}\big|\mathrm{cov}(Y_{Tt}, Y_{Tt+l})\big| \doteq III + IV,$$

where the term *IV* is considered zero if $T - t < N$. Since

$$\operatorname{cov}(Y_{Tt}, Y_{Tl+l}) = \int_{\mathbf{S}}\int_{\mathbf{S}} \big(P\big(A_t(\mathbf{s}) \cap A_{t+l}(\mathbf{r})\big) - P\big(A_t(\mathbf{r})\big)P\big(A_{t+l}(\mathbf{r})\big)\big)\, d\mathbf{s}\, d\mathbf{r},$$

and $A_t(\mathbf{s}) \doteq \{\mathcal{X}_t(\mathbf{s}) \in (x - a_T, x + a_T)\} \in \mathcal{M}_t^t$ and $A_{t+l}(\mathbf{r}) \doteq \{\mathcal{X}_{t+l}(\mathbf{s}) \in (x - a_T, x + a_T)\} \in \mathcal{M}_{t+l}^{t+l}$ for each fixed $\mathbf{s}$ and $\mathbf{r}$, hypothesis H5 implies that

$$\begin{aligned}\big|\operatorname{cov}(Y_{Tt}, Y_{Tl+l})\big| &\le \int_{\mathbf{S}}\int_{\mathbf{S}} \big|P\big(A_t(\mathbf{s}) \cap A_{t+l}(\mathbf{r})\big) - P\big(A_t(\mathbf{r})\big)P\big(A_{t+l}(\mathbf{r})\big)\big|\, d\mathbf{s}\, d\mathbf{r} \\ &\le |\mathbf{S}|^2 \alpha(l),\end{aligned}$$

from which follows

$$IV \le \frac{C}{Ta_T^2}\sum_{t=1}^{T-1}\sum_{l=N}^{T-t}\alpha(l) \le \frac{C}{a_T^2}\sum_{l=N}^{\infty}\alpha(l) = \frac{C}{Na_T^2}N\sum_{l=N}^{\infty}\alpha(l).$$

On the other hand, from H7 we have

$$\begin{aligned}&\big|\operatorname{cov}(Y_{Tt}, Y_{Tl+l})\big| \\ &= \left|\int_{\mathbf{S}}\int_{\mathbf{S}} \big(P\big(A_t(\mathbf{s}) \cap A_{t+l}(\mathbf{r})\big) - P\big(A_t(\mathbf{r})\big)P\big(A_{t+l}(\mathbf{r})\big)\big)\, d\mathbf{s}\, d\mathbf{r}\right| \\ &= \left|\int_{\mathbf{S}}\int_{\mathbf{S}}\int_{\{x:|u-x|\le cc_T\}}\int_{\{x:|v-x|\le cc_T\}} \big(f_{\mathbf{sr}}(u, v) - f_{\mathcal{X}}(u) f_{\mathcal{X}}(v)\big)\, du\, dv\, d\mathbf{s}\, d\mathbf{r}\right| \\ &\le C a_T^4 \end{aligned} \tag{12}$$

for some constant C which implies that $III \le 2CNa_T^2$ and hence

$$|II| \le 2CNa_T^2 + \frac{C}{Na_T^2}N\sum_{l=N}^{\infty}\alpha(l).$$

Let $\epsilon > 0$ fixed and $N = \lfloor \frac{\epsilon}{a_T^2} \rfloor$. For this choice we get $|II| \le C\epsilon + \frac{C}{\epsilon}N\sum_{l=N}^{\infty}\alpha(l)$ for T large enough (since $a_T \sim \frac{k_T}{T} \to 0$). From hypothesis H5 there exists N_0 such that if $M \ge N_0$, $M\sum_{l=M}^{\infty}\alpha(l) < \epsilon^2$. On the other hand, again since $a_T \to 0$, there exists T_0 such that if $T \ge T_0$, $\lfloor \frac{\epsilon}{a_T^2} \rfloor \ge N_0$ and then $\lfloor \frac{\epsilon}{a_T^2} \rfloor \sum_{l=\lfloor \frac{\epsilon}{a_T^2} \rfloor}^{\infty}\alpha(l) < \frac{\epsilon^2}{2}$ which implies that $II \le C\epsilon$ which implies

$$\frac{s_T^2}{Ta_T^2} \to c_0^2. \tag{13}$$

To prove that $\frac{S_T - \mathbb{E}(S_T)}{\sqrt{T}a_T} = \frac{1}{\sqrt{T}a_T}\sum_{t=1}^{T}\overline{Y}_{tT} \to \mathcal{N}(0, c_0^2)$ we will show that the variables $V_{Tt} \doteq \overline{Y}_{tT}$ verify the assumptions (i)–(v) of the Lindeberg version of the CLT given in Theorem 6 with $b_T = a_T^2$. As before, since $\mathcal{X}_t$ verifies H5, $\overline{Y}_{Tt}$ inherits the same condition from which (i) follows. Condition (ii) holds for $M = 2|\mathbf{S}|$. (iii) follows from (4) (since $k_T \to \infty$ and $a_T \sim \frac{k_T}{T} \to 0$), H5 implies (iv) (see (11)) and H7 implies (v) (see (12)). Therefore, from Theorem 6 we get the result since $\sigma^2 = c_0^2$.

To prove (b) we use the definition of a_T to get

$$\frac{k_T - T|\mathbf{S}|p_T}{s_T} = \frac{2Ta_T|\mathbf{S}|(f_{\mathcal{X}}(x) + z/\sqrt{T}) - T|\mathbf{S}|\int_{x-a_T}^{x+a_T} f_{\mathcal{X}}(u)\,du}{s_T}$$
$$= s_T^{-1}T|\mathbf{S}|\int_{x-a_T}^{x+a_T}\big(f_{\mathcal{X}}(x) - f_{\mathcal{X}}(u)\big)\,du + \frac{2Ta_T|\mathbf{S}|z}{s_T\sqrt{T}}. \qquad (14)$$

By Taylor Theorem and since f has two derivatives bounded, for x^* between x and u we have

$$\left|\int_{x-a_T}^{x+a_T}\big(f_{\mathcal{X}}(x) - f_{\mathcal{X}}(u)\big)\,du\right| = \left|-\frac{1}{2}\int_{x-a_T}^{x+a_T} f_{\mathcal{X}}''(x^*)(u-x)^2\,du\right| \le Cs_T^{-1}Ta_T^3.$$

Therefore, in (14) we have

$$\frac{k_T - T|\mathbf{S}|p_T}{s_T} = O\big(s_T^{-1}Ta_T^3\big) + \frac{2|\mathbf{S}|z}{s_T/\sqrt{T}a_T}.$$

Finally, by hypothesis (4) and (13) $s_T^{-1}Ta_T^3 \to 0$ then from (11) we get (b). □

Proof of Theorem 3 This proof will be an immediate consequence of Theorem 1 and of the following lemma which was proved in Llop et al. (2011, Lemma 5, p. 85). □

Lemma 5 *Assume* H1–H3 *and choose two sequences* $\{k_T\}$ *and* $\{v_T\}$ *of positive real numbers converging to infinity such that, for each fixed* $\mathbf{s}$, $v_T(T/k_T)|\bar{e}_T(\mathbf{s})| \to 0$ *a.co. For this sequences, suppose that* H4 *hold. Then for each* $x \in \mathbb{R}$

$$\lim_{n\to\infty} v_T\big(\widehat{f}_u\big(x - \bar{\mathcal{X}}_T(\mathbf{s})\big) - \widehat{f}_e\big(x - \mu(\mathbf{s})\big)\big) = 0, \quad a.co.$$

where $\widehat{f}_e$ *is the estimator of* f_e.

Appendix C: Proof of the auxiliary lemmas

Proof of Lemma 2 We need to prove that if $\alpha(r) \le a\rho^r$ with $0 < \rho < 1$ and $a > 0$ then, $\sum_{j=0}^{\infty}(j+1)^k\alpha(j) \le L_1L_2^k(k!)^\mu, \forall k \ge 0$. For that, let us suppose that $\alpha(r) \le a\rho^r$ for some $0 < \rho < 1$ and $a > 0$. If $f^{(k)}$ denotes the kth derivative of f then

$$\sum_{j=0}^{\infty}(j+1)^k\alpha(j) \le a\sum_{j=0}^{\infty}(j+1)^k\rho^j$$
$$\le a\sum_{j=0}^{\infty}(j+k)(j+k-1)\cdots(j+1)\rho^j$$
$$= a\left(\sum_{j=0}^{\infty}\rho^{j+k}\right)^{(k)} = a\left(\rho^k\frac{1}{1-\rho}\right)^{(k)}. \qquad (15)$$

Let $C(k,j)$ be denote the combinatorial number $\binom{n}{k}$. Since

$$(fg)^{(k)} = \sum_{j=0}^{k} C(k,j) f^{(k-j)} g^{(j)},$$

$$\left(\rho^k\right)^{(j)} = \frac{k!}{(k-j)!}\rho^{k-j} \quad \text{and} \quad \left(\frac{1}{1-\rho}\right)^{(j)} = \frac{j!}{(1-\rho)^{j+1}}$$

we have

$$\begin{aligned} a\left(\rho^k \frac{1}{1-\rho}\right)^{(k)} &= a\sum_{j=0}^{k} C(k,j)\left(\rho^k\right)^{(k-j)}\left(\frac{1}{1-\rho}\right)^{(j)} \\ &= a\sum_{j=0}^{k} C(k,j)\frac{k!}{j!}\rho^j \frac{j!}{(1-\rho)^{j+1}} \\ &= a\frac{k!}{1-\rho}\sum_{j=0}^{k} C(k,j)\left(\frac{\rho}{1-\rho}\right)^j 1^{k-j} \\ &= \frac{a}{1-\rho}\left(\frac{1}{1-\rho}\right)^k k! \end{aligned}$$

where in the last equality we have used the Binomial Theorem. Then, with this inequality in (15), we get

$$\sum_{j=0}^{\infty}(j+1)^k \alpha(j) \le \frac{a}{1-\rho}\left(\frac{1}{1-\rho}\right)^k k!$$

Therefore, taking $L_1 = \frac{a}{1-\rho}$, $L_2 = \frac{1}{1-\rho}$ and $\mu = 1$ we get the result. □

Proof of Lemma 3 For T and x fixed, the function

$$G(r) \doteq \frac{1}{T}\sum_{t=1}^{T}\int_{I_{(x,r)}} l_T(u,\mathcal{X}_t)\,du,$$

is strictly increasing with $G(0)=0$. On the other hand, due to the existence of local time we can write

$$G(r) = \frac{1}{T}\sum_{t=1}^{T}\lambda(I_{(x,r)},\mathcal{X}_t),$$

then, $G(r)\to|\mathbf{S}|$ when $r\to\infty$ and therefore, the existence and uniqueness of $h_T^{\mathcal{X}}(x)$ is ensured. For a further reading on local times see Geman and Horowitz (1980). □

Proof of Lemma 4 To prove part (i) let us consider the u-tuple $(t_1, \ldots, t_u)$ and the v-tuple $(l_1, \ldots, l_v)$ for $1 \le t_1 \le t_u < t_u + r = l_1 \le l_v \le T$ and let $C(k, j)$ be denote the combinatorial number $\binom{n}{k}$. Since $\mathbb{E}(Y_{Tt}) = |\mathbf{S}| p_T$ then,

$$\prod_{k=1}^{u} \overline{Y}_{Tt_k} = \prod_{k=1}^{u} \big(Y_{Tt_k} - p_T|\mathbf{S}|\big) \doteq \sum_{k=0}^{u} C(u,k)\big(-p_T|\mathbf{S}|\big)^{u-k} \prod_{i=1}^{k} Y_{Tt_i}.$$

Therefore, taking absolute value and considering that $|p_T| \le 1$ we get

$$\begin{aligned}
&\big|\mathrm{cov}(\overline{Y}_{Tt_1} \cdots \overline{Y}_{Tt_u}, \overline{Y}_{Tl_1} \cdots \overline{Y}_{Tl_v})\big| \\
&\quad = \left|\mathrm{cov}\left(\prod_{k=1}^{u} \overline{Y}_{Tt_k}, \prod_{m=1}^{v} \overline{Y}_{Tl_m}\right)\right| \\
&\quad = \left|\mathrm{cov}\left(\sum_{k=1}^{u} C(u,k)\big(-p_T|\mathbf{S}|\big)^{u-k} \prod_{i=1}^{k} Y_{Tt_i}, \sum_{m=1}^{v} C(v,m)\big(-p_T|\mathbf{S}|\big)^{v-m} \prod_{j=1}^{m} Y_{Tl_j}\right)\right| \\
&\quad \le \sum_{k=1}^{u} C(u,k) \sum_{m=1}^{v} C(v,m) |\mathbf{S}|^{u+v-(k+m)} \left|\mathrm{cov}\left(\prod_{i=1}^{k} Y_{Tt_i}, \prod_{j=1}^{m} Y_{Tl_j}\right)\right|. \qquad (16)
\end{aligned}$$

To compute

$$\left|\mathrm{cov}\left(\prod_{i=1}^{k} Y_{Tt_i}, \prod_{j=1}^{m} Y_{Tl_j}\right)\right| = \left|\mathbb{E}\left(\prod_{i=1}^{k} Y_{Tt_i} \prod_{j=1}^{m} Y_{Tl_j}\right) - \mathbb{E}\left(\prod_{i=1}^{k} Y_{Tt_i}\right)\mathbb{E}\left(\prod_{j=1}^{m} Y_{Tl_j}\right)\right|,$$

we will consider the following events:

$$A_{t_i}(\mathbf{s}) \doteq \big\{\omega \in \Omega : \mathcal{X}_{t_i}(\mathbf{s}, \omega) \in (x - a_T, x + a_T)\big\}, \quad i = 1, \ldots, u$$

and

$$A_{l_j}(\mathbf{r}) \doteq \big\{\omega \in \Omega : \mathcal{X}_{l_j}(\mathbf{r}, \omega) \in (x - a_T, x + a_T)\big\}, \quad j = 1, \ldots, v.$$

In addition, we will denote

$$\int_{\mathbf{S}^u} d\mathbf{s}^u \doteq \int_{\mathbf{S}} \cdots \int_{\mathbf{S}} d\mathbf{s}_1 \cdots d\mathbf{s}_u.$$

Then,

$$\begin{aligned}
&\mathbb{E}\left(\prod_{i=1}^{k} Y_{Tt_i} \prod_{j=1}^{m} Y_{Tl_j}\right) \\
&\quad = \mathbb{E}\left(\prod_{i=1}^{k}\left(\int_{\mathbf{S}} \mathbb{I}_{(x-a_T, x+a_T)}\big(\mathcal{X}_{t_i}(\mathbf{s})\big)\, d\mathbf{s}\right) \prod_{j=1}^{m}\left(\int_{\mathbf{S}} \mathbb{I}_{(x-a_T, x+a_T)}\big(\mathcal{X}_{l_j}(\mathbf{r})\big)\, d\mathbf{r}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbf{S}^u}\int_{\mathbf{S}^v} \mathbb{E}\left(\prod_{i=1}^{k} \mathbb{I}_{(x-a_T,x+a_T)}\big(\mathcal{X}_{t_i}(\mathbf{s})\big)\prod_{j=1}^{m} \mathbb{I}_{(x-a_T,x+a_T)}\big(\mathcal{X}_{l_j}(\mathbf{s})\big)\right) d\mathbf{s}^u\, d\mathbf{r}^v \\
&= \int_{\mathbf{S}^u}\int_{\mathbf{S}^v} P\left(\bigcap_{i=1}^{k} A_{t_i}(\mathbf{s}) \cap \bigcap_{j=1}^{m} A_{l_j}(\mathbf{r})\right) d\mathbf{s}^u\, d\mathbf{r}^v, \qquad (17)
\end{aligned}
$$

and similarly

$$
\mathbb{E}\left(\prod_{i=1}^{k} Y_{Tt_i}\right)\mathbb{E}\left(\prod_{j=1}^{m} Y_{Tl_j}\right) = \int_{\mathbf{S}^u}\int_{\mathbf{S}^v} P\left(\bigcap_{i=1}^{k} A_{t_i}(\mathbf{s})\right) P\left(\bigcap_{j=1}^{m} A_{l_j}(\mathbf{r})\right) d\mathbf{s}^u\, d\mathbf{r}^v.
$$

Hence,

$$
\begin{aligned}
&\left|\operatorname{cov}\left(\prod_{i=1}^{k} Y_{Tt_i}, \prod_{j=1}^{m} Y_{Tl_j}\right)\right| \\
&\quad \le \int_{\mathbf{S}^u}\int_{\mathbf{S}^v} \left| P\left(\bigcap_{i=1}^{k} A_{t_i}(\mathbf{s}) \cap \bigcap_{j=1}^{m} A_{l_j}(\mathbf{r})\right) - P\left(\bigcap_{i=1}^{k} A_{t_i}(\mathbf{s})\right) P\left(\bigcap_{j=1}^{m} A_{l_j}(\mathbf{r})\right)\right| d\mathbf{s}^u\, d\mathbf{r}^v \\
&\quad \le \int_{\mathbf{S}^u}\int_{\mathbf{S}^v} \alpha(r)\phi(u,v)\, d\mathbf{s}^u\, d\mathbf{r}^v = |\mathbf{S}|^{u+v}\alpha(r)\phi(u,v), \qquad (18)
\end{aligned}
$$

where in the last inequality we have used hypothesis H4 since, for each fixed $\mathbf{s}, \mathbf{r}$, since $k \le u$ and $m \le v$,

$$
\bigcap_{i=1}^{k} A_{t_i}(\mathbf{s}) \in \mathcal{M}_{t_1}^{t_u} \quad \text{and} \quad \bigcap_{j=1}^{m} A_{l_j}(\mathbf{r}) \in \mathcal{M}_{l_1}^{l_v},
$$

with $1 \le t_1 \le t_u < t_u + r = l_1 \le l_v \le T$. Finally, with inequality (18) in (16) we get

$$
\begin{aligned}
&\left|\operatorname{cov}\left(\prod_{i=1}^{k} \overline{Y}_{Tt_i}, \prod_{j=1}^{m} \overline{Y}_{Tl_j}\right)\right| \\
&\quad \le \sum_{k=1}^{u} C(u,k) \sum_{m=1}^{v} C(v,m)|\mathbf{S}|^{u+v-(k+m)}\left|\operatorname{cov}\left(\prod_{i=1}^{k} Y_{Tt_i}, \prod_{j=1}^{m} Y_{Tl_j}\right)\right| \\
&\quad \le \sum_{k=1}^{u} C(u,k) \sum_{m=1}^{v} C(v,m)|\mathbf{S}|^{u+v-(k+m)}|\mathbf{S}|^{k+m}\alpha(l_1 - t_i)\phi(u,v) \\
&\quad \le |\mathbf{S}|^{u+v} \sum_{k=1}^{u} C(u,k) \sum_{m=1}^{v} C(v,m)\alpha(r)\phi(u,v) \\
&\quad \le \big(2|\mathbf{S}|\big)^{u+v}\alpha(r)\phi(u,v).
\end{aligned}
$$

To prove part (ii) of this lemma, first observe that since $\mathrm{var}(Y_{Tt}) \le \mathbb{E}(Y_{Tt}^2) \le |\mathbf{S}|^2 p_T^2 \le |\mathbf{S}|^2$ we may write

$$\begin{aligned}\mathrm{var}\left(\sum_{t=1}^{T} \overline{Y}_{Tt}\right) &\le \sum_{t=1}^{T} \mathrm{var}(Y_{Tt}) + 2 \sum_{t=1}^{T-1} \sum_{l=t+1}^{T} \left|\mathrm{cov}(\overline{Y}_{Tt}, \overline{Y}_{Tl})\right| \\ &\le T|\mathbf{S}|^2 + 2^3 |\mathbf{S}|^2 \phi(1,1) \sum_{t=1}^{T-1} \sum_{l=t+1}^{T} \alpha(l-t) \\ &\le CT. \qquad \square\end{aligned}$$

References

Blanke D (2004) Adaptive sampling schemes for density estimation. J Stat Plan Inference 136(9):2898–2917

Blanke D, Bosq D (1997) Accurate rates of density estimators for continuous-time processes. Stat Probab Lett 33(2):185–191

Carbon M, Hallin M, Tran L (1996) Kernel density estimation for random fields: the L_1 theory. J Nonparametr Stat 6(2–3):157–170

Carbon M, Hallin M, Tran L (1997) Kernel density estimation for random fields (density estimation for random fields). Stat Probab Lett 36(2):115–125

Castellana JV, Leadbetter MR (1986) On smoothed probability density estimation for stationary processes. Stoch Process Appl 21(2):179–193

Chao–Gan Y, Yu-Feng Z (2010a) DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting–state fMRI. Frontiers Syst Neurosci 4:1–7

Chao–Gan Y, Yu-Feng Z (2010b). http://www.nitrc.org

Doukhan P, Leon J, Portal F (1984) Vitesse de convergence dans le théoréme central limite pour des variables aleatoires mtlangeantes a valeurs dans un espace de Hilbert. C R Acad Sci Paris 289:305–308

Doukhan P, Louhichi S (1999) A new dependence condition and applications to moment inequalities. Stoch Process Appl 84:313–342

Doukhan P, Neumann M (2006) Probability and moment inequalities for sums of weakly dependent random variables, with applications. Stoch Process Appl 117:878–903

Fox MD, Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional networks. Nat Rev Neurosci 8:700–711

Geman D, Horowitz J (1980) Occupation densities. Ann Probab 8(1):1–67

Hallin M, Lu Z, Tran L (2001) Density estimation for spatial linear processes. Bernoulli 7(4):657–668

Hallin M, Lu Z, Tran L (2004) Kernel density estimation for spatial processes: the L_1 theory. Ann Stat 88:61–75

Kutoyants Y (2004) On invariant density estimation for ergodic diffusion processes. SORT 28(2):111–124

Labrador B (2008) Strong pointwise consistency of the k_T-occupation time density estimator. Stat Probab Lett 78(9):1128–1137

Llop P, Forzani L, Fraiman R (2011) On local times, density estimation and supervised classification from functional data. J Multivar Anal 102(1):73–86

Nguyen H (1979) Density estimation in a continuous-time stationary Markov process. Ann Stat 7(2):341–348

Neumann M, Paparoditis E (2008) Goodness-of-t tests for Markovian time series models: central limit theory and bootstrap approximations. Bernoulli 14(1):14–46

Robinson PM (1983) Nonparametric estimators for time series. J Time Ser Anal 4:185–206

Rosenblatt M (1956) A central limit theorem and a strong mixing condition. Proc Natl Acad Sci USA 42:43–47

Rosenblatt M (1970) Density estimates and Markov sequences. In: Nonparametric techniques in statistical inference. Cambridge University Press, Cambridge, pp 199–210

Tang X, Liu Y, Zhang J, Kainz W (2008) In: Advances in spatio-temporal analysis. ISPRS, vol 5

Tran LT (1990) Kernel density estimation on random fields. J Multivar Anal 34(1):37–53

Tran L, Yakowitz S (1993) Nearest neighbour estimators for random fields. J Multivar Anal 44(1):23–46