Nonparametric Spatial Prediction

Gérard BIAU [a,*] and Benoît CADRE [b]

[a] Laboratoire de Statistique Théorique et Appliquée,
Université Pierre et Marie Curie – Paris VI,
Boîte 158, 175 rue du Chevaleret, 75013 Paris, France;
[b] Laboratoire de Probabilités et Statistique, Université Montpellier II,
Cc 051, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

Abstract

Let $(\mathbb{N}^*)^N$ be the integer lattice points in the N-dimensional Euclidean space. We define a nonparametric spatial predictor for the values of a random field indexed by $(\mathbb{N}^*)^N$ using a kernel method. We first examine the general problem of the regression estimation for random fields. Then we show the uniform consistency on compact sets of our spatial predictor as well as its asymptotic normality.

***Index Terms* — Kernel regression estimation, Random fields, Spatial prediction, Mixing.**

***AMS 2000 Classification*: 62G08, 62H11.**

1 Introduction

Spatial prediction, in general, is any prediction method that incorporates spatial dependence. In fields such as petroleum exploration, mining, or water pollution analysis, data are available at specific spatial locations (such as experimental stations positioned above the ground or at certain distances in the air), and the goal is to *predict* unsampled locations. The unsampled locations are often mapped on a regular grid, and the predictions are used to produce surface plots or contour maps. One of the simplest and very popular spatial prediction method is *kriging*, which is basically studied by *Geostatistics* (see Wackernagel [17] for an introduction). The origins of Geostatistics are to be found within the mining industry in the early 50s. Since that period

*Corresponding author. Email: biau@ccr.jussieu.fr .

of time, high speed, inexpensive computing methods have made it easy to look at data in ways that were once difficult. One area that has benefited greatly from this new freedom is that of *nonparametric statistics*. However, most of existing theoretical nonparametric results of dependent random variables pertain to time series (see Bosq [2] for a complete survey), where the unidirectional flow of time underlies the construction of the models, and relatively few generalizations to the spatial domain are available. Key references on this latter topic are Tran [15], Tran and Yakowitz [16], Carbon, Hallin and Tran [3], Carbon, Tran and Wu [4], Hallin, Lu and Tran [8], [10] and Biau [1], who have investigated nonparametric density estimation for random fields indexed by $(\mathbb{N}^*)^N$, with $N \geq 1$. Roughly speaking, data sets whose spatial locations are regular lattices in $\mathbb{R}^N$ are the closest analogue to time series observed at equally spaced time points. Such data sets occur frequently in practice, either because the process under study is essentially discrete (many examples of such processes are to be found in forestry) or just because it is the result of a sampling scheme. To illustrate the latter point, think of satellites orbiting the Earth. By various sampling and integration methods, the Earth's surface is divided into small rectangles (*e.g.* 56 m $\times$ 56 m) called *pixels* (short for *picture elements*). An agriculture scene of interest (around, say, 34,000 km^2) has certain proportions devoted to wheat, corn, soybeans, and so forth that need to be estimated. These various crops have their reflectance properties that, together with noise, are remotely sensed. Thus the data are received as a regular lattice in $\mathbb{R}^2$ and are identified with the centers of their respective pixels (Cressie [5]).

Not many theoretical works have been devoted so far to study nonparametric predictors for spatial fields indexed by lattices (for related references, see Lu [12] and Hallin, Lu and Tran [9]). As a first modest step towards this direction, we will be concerned in this paper with the kernel prediction of a strictly stationary $\mathbb{R}$-valued random field $(Z_{\mathbf{i}})_{\mathbf{i}\in(\mathbb{N}^*)^N}$ $(N \geq 1)$. The bold letter $\mathbf{i}$ will always denote an element of $(\mathbb{N}^*)^N$, and we shall use the notation $\mathbf{i} = (i_1, \ldots, i_N)$. For $\mathbf{n}$ in $(\mathbb{N}^*)^N$, we define the rectangular region $\mathcal{I}_{\mathbf{n}}$ by $\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} \in (\mathbb{N}^*)^N : 1 \leq i_k \leq n_k, k = 1, \ldots, N\}$, we set $\hat{\mathbf{n}} = n_1 \ldots n_N$, and we write $\mathbf{n} \to \infty$ if $\min_{k=1,\ldots,N} n_k \to \infty$. All limits are taken as $\mathbf{n} \to \infty$, unless indicated otherwise. Assume now that the field is observed on a set $\mathcal{S}_{\mathbf{n}}$ contained in $\mathcal{I}_{\mathbf{n}}$. We wish to interpolate the value of the field at a given fixed point $\mathbf{j}_0 \in \mathcal{I}_{\mathbf{n}} - \mathcal{S}_{\mathbf{n}}$. To this aim, we assume that a generic value $Z_{\mathbf{j}}$ only depends on the values taken by the field in a vicinity of $\mathbf{j}$ not containing $\mathbf{j}$. More precisely, this vicinity, say $\mathcal{V}_{\mathbf{j}}$, is assumed to be of the form $\mathbf{j} + \mathcal{V}$, where $\mathcal{V}$ is a fixed bounded set that does not contain $\mathbf{0}$. We call $\mathcal{V}_{\mathbf{j}}$ the *prediction vicinity*. Observe that in the time context $(N = 1)$, the above property means that the field is Markovian. A typical spatial example is depicted on Figure 1.

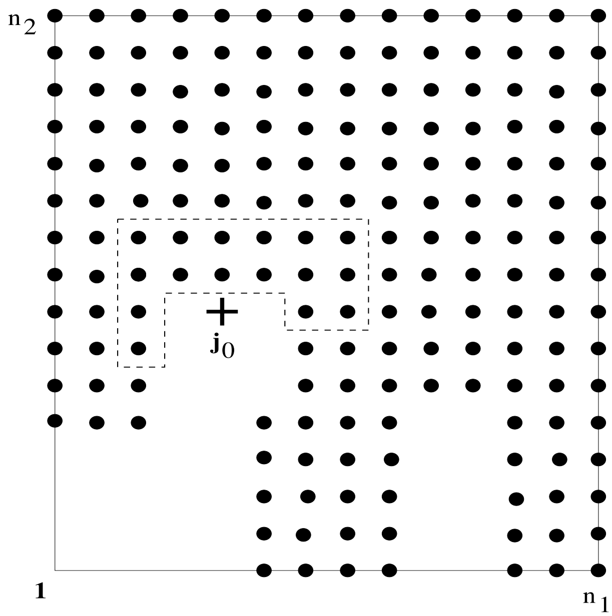

Figure 1: A typical prediction problem in dimension $\mathbf{N} = \mathbf{2}$. Observed data are in bold, the prediction vicinity is contained in the dashed line.

As is well known, the best prediction of $Z_{\mathbf{j}_0}$ given the values taken by the field in $\mathcal{V}_{\mathbf{j}_0}$ is, for the L_2 minimizing criterion,

$$\mathbf{E}[Z_{\mathbf{j}_0}|Z_{\mathbf{i}}, \mathbf{i} \in \mathcal{V}_{\mathbf{j}_0}].$$

Our goal in this paper is to estimate the above quantity using a nonparametric kernel method. Before we define the predictor, we give some notation.

- $\tilde{Z}_{\mathbf{j}}$ will stand for the vector whose components are the $\{Z_{\mathbf{i}}, \mathbf{i} \in \mathcal{V}_{\mathbf{j}}\}$ that have been concatenated and ordered according to an arbitrary order on indices, for example the lexicographic order.

- The cardinal of $\mathcal{V}$ (which is also the cardinal of $\mathcal{V}_{\mathbf{j}}$) will be denoted by d. As a consequence, the random vector $\tilde{Z}_{\mathbf{j}}$ is a $\mathbb{R}^d$-valued random variable.

- We assume that there exists a bounded subset $\mathcal{A} \subset (\mathbb{N}^*)^N$ such that, for $\hat{\mathbf{n}}$ large enough,
$$\mathcal{S}_{\mathbf{n}} = \mathcal{I}_{\mathbf{n}} - \mathcal{A}. \tag{1.1}$$

It is worth pointing out that condition (1.1) is required to hold *for* $\hat{\mathbf{n}}$ *large enough*. Thus, in practice–that is to say at finite distance–(1.1) involves no loss of generality on the shape of the sampling region $\mathcal{S}_{\mathbf{n}}$ that can be, for instance, star-shaped. We just make this assumption in order to get asymptotic results. Large-sample properties of estimators and predictors based on spatial observations have been studied under several different asymptotic structures. All studies appear to arise from different combinations of two basic sampling paradigms. In the present paper, we assume that the sampling sites are separated by a distance greater than some fixed positive number, and the sampling region becomes unbounded as the sample size increases. The resulting structure leads to what is known as *increasing-domain asymptotics*. This is the most common framework used for asymptotics for spatial data. The other form, known as *infill asymptotics*, is inherently different and is more suitable for making inferences about a continuously indexed random field observed on a bounded region. When an increasing number of samples are collected from within a bounded sampling region (that does not grow with the sample size), we obtain the *infill structure*. For references and discussion, we refer the reader to Cressie [5] and Lahiri, Kaiser, Cressie and Hsu [11].

Now, the prediction of the value of the field at the point $\mathbf{j}_0$ reads

$$\hat{r}(\tilde{Z}_{\mathbf{j}_0}) = \frac{\displaystyle\sum_{\substack{\mathbf{i}\in\mathcal{S}_{\mathbf{n}}\\ \mathcal{V}_{\mathbf{i}}\subset\mathcal{S}_{\mathbf{n}}}} Z_{\mathbf{i}}\, K\Big(\frac{\tilde{Z}_{\mathbf{j}_0}-\tilde{Z}_{\mathbf{i}}}{h}\Big)}{\displaystyle\sum_{\substack{\mathbf{i}\in\mathcal{S}_{\mathbf{n}}\\ \mathcal{V}_{\mathbf{i}}\subset\mathcal{S}_{\mathbf{n}}}} K\Big(\frac{\tilde{Z}_{\mathbf{j}_0}-\tilde{Z}_{\mathbf{i}}}{h}\Big)}\,, \tag{1.2}$$

where the *kernel* $K:\mathbb{R}^d\to\mathbb{R}^+$ is a probability density, and the *smoothing parameter* (or *bandwidth*) $h=h(\mathbf{n})$ is a sequence of positive real numbers such that $h\to 0$ as $\mathbf{n}\to\infty$. Observe that in (1.2), we have to impose the condition $\mathcal{V}_{\mathbf{i}}\subset\mathcal{S}_{\mathbf{n}}$ just to ensure that the sums make sense. Anyway, by (1.1), these sums have the same asymptotic behavior as the sums taken over $\mathcal{I}_{\mathbf{n}}$.

The paper is organized as follows. In a second part, we consider as prerequisites the general problem of regression estimation for random fields, *i.e.*, for random pairs $(X_{\mathbf{i}},Y_{\mathbf{i}})$ with values in $\mathbb{R}^d\times\mathbb{R}$. We show the uniform consistency on compact sets of a kernel estimator of the regression function of $Y_{\mathbf{i}}$ given $X_{\mathbf{i}}$, as well as its asymptotic normality. In a third part, we particularize these results to the spatial predictor (1.2). Proofs are gathered at the end of the paper (Section 4).

2 Regression estimation for random fields

2.1 Notation and general hypotheses

Let (X, Y) be a $\mathbb{R}^d \times \mathbb{R}$-valued random variable admitting a density g with respect to the Lebesgue measure. In the remainder of the paper, $\mathbb{R}^d$ is endowed with the Euclidean norm $||.||$. We assume that Y is integrable and we denote by f the density of X. Then the *regression function* $r(.)$ of Y given X is defined by

$$r(x) = \begin{cases} \dfrac{1}{f(x)} \displaystyle\int_{\mathbb{R}} y g(x, y) \mathrm{d}y & \text{if} \quad f(x) \neq 0; \\ \mathbf{E}Y & \text{if} \quad f(x) = 0. \end{cases}$$

In this section, we intend to estimate the function r using $\hat{\mathbf{n}}$ observations $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}}$ drawn from a *strictly stationary* random field $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i} \in (\mathbb{N}^*)^N}$ of random variables with the same law as (X, Y). To this aim, we first introduce the *kernel estimator* $f_{\mathbf{n}}$ of f (Bosq [2]) defined by

$$f_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}}\, h^d} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K\Big(\frac{x - X_{\mathbf{i}}}{h}\Big),$$

where the *kernel* $K : \mathbb{R}^d \to \mathbb{R}^+$ is a Lipschitzian density with compact support, and the *smoothing parameter* (or *bandwidth*) $h = h(\mathbf{n})$ is a sequence of positive real numbers satisfying $h \to 0$ as $\mathbf{n} \to \infty$. In this framework, the kernel estimator $r_{\mathbf{n}}$ of r is defined by

$$r_{\mathbf{n}}(x) = \begin{cases} \dfrac{1}{\hat{\mathbf{n}}\, h^d f_{\mathbf{n}}(x)} \displaystyle\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} K\Big(\frac{x - X_{\mathbf{i}}}{h}\Big) & \text{if} \quad f_{\mathbf{n}}(x) \neq 0; \\ \dfrac{1}{\hat{\mathbf{n}}} \displaystyle\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} & \text{if} \quad f_{\mathbf{n}}(x) = 0. \end{cases}$$

In order to obtain asymptotic results, we will assume throughout the paper, and *unless specified otherwise*, that the field $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i} \in (\mathbb{N}^*)^N}$ satisfies the following *mixing condition*: there exists a function $\chi : \mathbb{R}^+ \to \mathbb{R}^+$ with $\chi(t) \searrow 0$ as $t \to \infty$, such that whenever $E, E' \subset (\mathbb{N}^*)^N$ with finite cardinals,

$$\begin{aligned} \alpha\big(\mathcal{B}(E), \mathcal{B}(E')\big) &:= \sup\big\{|\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)|\,;\, A \in \mathcal{B}(E), B \in \mathcal{B}(E')\big\} \\ &\leq \psi(\text{Card } E, \text{Card } E')\, \chi\big(\text{dist}(E, E')\big), \end{aligned}$$

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

$$\psi(\mathbf{i}, \mathbf{j}) \leq c \min(\mathbf{i}, \mathbf{j}), \quad \forall \mathbf{i}, \mathbf{j} \in \mathbb{N}, \tag{2.1}$$

for some constant $c > 0$. If $\psi \equiv 1$, then the field $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in(\mathbb{N}^*)^N}$ is called *strongly mixing*. Finally, in order to have a large choice of bandwidths, we will require the existence of two positive real numbers a, b such that

$$\mathbf{E} \exp\left(a|Y|^b\right) < \infty\,. \tag{2.2}$$

2.2 Uniform strong consistency of $r_{\mathbf{n}}$

In this paragraph, we fix a compact subset $\mathcal{S}$ of $\mathbb{R}^d$. By a *dilated* of $\mathcal{S}$, we mean a set $\tilde{\mathcal{S}}$ such that

$$\tilde{\mathcal{S}} = \{x \in \mathbb{R}^d : \mathrm{dist}(x, \mathcal{S}) \leq \delta\}$$

for some $\delta > 0$, where

$$\mathrm{dist}(x, \mathcal{S}) = \inf_{y\in\mathcal{S}} ||x - y||\,.$$

In order to establish the uniform strong consistency of $r_{\mathbf{n}}$ on $\mathcal{S}$, we need the following assumptions.

(**H1**) The functions f and r are continuous on some dilated $\tilde{\mathcal{S}}$ of $\mathcal{S}$ and moreover $\inf_{\mathcal{S}} f > 0$.

(**H2**) For all $\mathbf{i} \neq \mathbf{j}$, one has

$$\sup_{x\in\mathcal{S}} \mathbf{E}\left[K\Big(\frac{x - X_{\mathbf{i}}}{h}\Big)K\Big(\frac{x - X_{\mathbf{j}}}{h}\Big)\right] = \mathrm{O}\big(h^{d+1}\big).$$

(**H3**) There exists $D \geq 0$ such that the pairs $(X_{\mathbf{i}}, X_{\mathbf{j}})$ and $\big((X_{\mathbf{i}}, Y_{\mathbf{i}}), (X_{\mathbf{j}}, Y_{\mathbf{j}})\big)$ admit a density, say $f_{\mathbf{i},\mathbf{j}}$ and $g_{\mathbf{i},\mathbf{j}}$, as soon as $\mathrm{dist}(\mathbf{i}, \mathbf{j}) > D$. Moreover, for some constant $c \geq 0$,

$$|f_{\mathbf{i},\mathbf{j}}(u, v) - f(u)f(v)| \leq c, \quad \forall u, v \in \mathbb{R}^d$$

and

$$|g_{\mathbf{i},\mathbf{j}}(s, t) - g(s)g(t)| \leq c, \quad \forall s, t \in \mathbb{R}^{d+1}.$$

Note first that condition (**H3**) is automatically fulfilled when the univariate and bivariate densities are bounded. Observe moreover that when (**H3**) holds with $D = 0$ (as it is usually assumed in the context of spatial regression, see Hallin, Lu and Tran [9]), then (**H2**) holds as soon as (**H1**) holds. Here, we shall work under the more general framework $D \geq 0$ in order to derive results for the prediction problem (we refer the reader to the discussion in Section 3, where a sufficient condition for a random field to satisfy (**H2**) and (**H3**) for $D > 0$ is explicited). From now on, $\lfloor u \rfloor$ stands for the integer part of a real number u.

Theorem 2.1 *Let p be an integer such that $p = \left\lfloor (\hat{\mathbf{n}}\, h^d / (\log \hat{\mathbf{n}})^{1+2/b})^{1/N} \right\rfloor$ and assume that*

$$p \to \infty \quad \textit{and} \quad \sum_{\mathbf{n} \in (\mathbb{N}^*)^N} h^{-d(d+2)} (\log \hat{\mathbf{n}})^{(2d+1)/b} p^N \chi(p) < \infty \,. \tag{2.3}$$

If (**H1**)-(**H3**) *are satisfied we have, with probability 1,*

$$\sup_{x \in \mathcal{S}} |r_{\mathbf{n}}(x) - r(x)| \to 0.$$

If we assume now that the field is *geometrically mixing*, we obtain the following corollary, whose proof is clear.

Corollary 2.1 *Assume that there exists $\xi > 0$ such that $\chi(t) = \mathrm{O}(e^{-\xi t})$ for $t \geq 0$. If* (**H1**)-(**H3**) *are satisfied, and if*

$$\frac{\hat{\mathbf{n}}\, h^d}{(\log \hat{\mathbf{n}})^{1+N+2/b}} \to \infty \,, \tag{2.4}$$

we have, with probability 1,

$$\sup_{x \in \mathcal{S}} |r_{\mathbf{n}}(x) - r(x)| \to 0.$$

In the time context ($N = 1$), the condition (2.4) on the bandwidth h is very similar to the condition that appears in Bosq ([2], Theorem 3.2).

2.3 Asymptotic normality of $r_{\mathbf{n}}$

In this paragraph, we fix $x \in \mathbb{R}^d$. If $f(z) > 0$, we let

$$v(z) = \mathbf{E}[Y^2 | X = z] - \mathbf{E}^2[Y | X = z].$$

In order to derive the asymptotic normality of $r_{\mathbf{n}}$, we need some additional conditions.

(**H4**) The functions f and r are Lipschitzian in a neighborhood of x, the function v is continuous at x and, moreover, $v(x)\, f(x) > 0$.

(**H5**) For all $\mathbf{i} \neq \mathbf{j}$, one has

$$\mathbf{E}\Big[K\Big(\frac{x - X_{\mathbf{i}}}{h}\Big)K\Big(\frac{x - X_{\mathbf{j}}}{h}\Big)\Big] = \mathrm{O}\big(h^{d+1}\big).$$

Observe, as in Paragraph 2.2, that when (**H3**) holds with $D = 0$, then (**H5**) holds as soon as (**H4**) is satisfied. The case $D > 0$ will be more specifically considered in our spatial prediction problem.

Theorem 2.2 *Assume that* (**H3**)-(**H5**) *are satisfied. Assume moreover that*

(*i*) $\hat{\mathbf{n}}\, h^{d+2} \to 0$;

(*ii*) *There exists* $\delta \in]0, 1/2[$ *such that* $\hat{\mathbf{n}}\, h^{d(1+2\delta N)}(\log \hat{\mathbf{n}})^{-8N/b} \to \infty$;

(*iii*) *There exists a sequence of positive integers* $q = q(\mathbf{n}) \to \infty$ *such that*

$$q = \mathrm{o}\big((\hat{\mathbf{n}}\, h^{d(1+2\delta N)}(\log \hat{\mathbf{n}})^{-8N/b})^{1/(2N)}\big)$$

and

$$\hat{\mathbf{n}} \sum_{i=1}^{\infty} i^{N-1}\chi(iq) \to 0\,;$$

(*iv*) *The bandwidth* h *tends to zero in a manner such that*

$$h^{-d\delta}(\log \hat{\mathbf{n}})^{2/b} \sum_{i=q}^{\infty} i^{N-1}\big(\chi(i)\big)^{\delta} \to 0.$$

Then, with the notation

$$\sigma^2(x) = \frac{v(x) \displaystyle\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u}{f(x)}\,,$$

we have

$$\frac{\sqrt{\hat{\mathbf{n}}\, h^d}}{\sigma(x)}\big(r_{\mathbf{n}}(x) - r(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ *denotes the one-dimensional standard Gaussian distribution and* $\xrightarrow{\mathcal{D}}$ *the convergence in distribution.*

Observe that under (iv), one has $\sum_{i=1}^{\infty} i^{N-1}\big(\chi(i)\big)^{\delta} < \infty$. In the geometrically mixing case, the following corollary is easily deduced.

Corollary 2.2 *Assume that there exists* $\xi > 0$ *such that* (i) $\chi(t) = \mathrm{O}(e^{-\xi t})$ *for* $t \geq 0$*. If* (**H3**)-(**H5**) *are satisfied, and*

(ii) $\hat{\mathbf{n}}\, h^{d+2} \to 0$;

(iii) *For some* $\delta \in]0, 1/2[$, $\hat{\mathbf{n}}\, h^{d(1+2\delta N)}(\log \hat{\mathbf{n}})^{-8N/b-2N} \to \infty$,

then, with the notation of Theorem 2.2, we have

$$\frac{\sqrt{\hat{\mathbf{n}}\, h^d}}{\sigma(x)}\big(r_{\mathbf{n}}(x) - r(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.5}$$

If Y is almost surely bounded, one may choose $b = \infty$. In this case, the conditions on the bandwidth that appear in Theorem 2.2 and Corollary 2.2 are the same as those of Tran [15], who considers the problem of estimating the density for a random field.

As for most nonparametric regression techniques, the shape and smoothness of the estimated function depends to a large extent on the specific value chosen for the smoothing parameter h. This choice is very important in practice, and it typically involves a bias-variance trade off. Here, the usual data-driven selection methods used in the time series context (such as cross-validation or plug-in, see *e.g.* Györfi, Härdle, Sarda and Vieu [7]) can be adapted to our spatial framework without further difficulties. We believe however that such results are beyond the scope of the present paper and we will deal with this problem elsewhere. We also refer the reader to Opsomer, Wang and Yang [13] whose comparative study and results would provide valuable information on the choice of the smoothing parameter. We would like as well to shed light on the fact that there are some problems associated with the use of results of this type. As an example, if one is interested in constructing confidence sets, it will be necessary to estimate the limiting variance, which involves not only $f(x)$ but also the conditional variance $v(x)$. A possible answer is to use the weakly consistent estimates $f_{\mathbf{n}}(x)$ and

$$v_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}}\, h^d f_{\mathbf{n}}(x)} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^2\, K\Big(\frac{x - X_{\mathbf{i}}}{h}\Big) - r_{\mathbf{n}}^2(x)$$

of $f(x)$ and $v(x)$, respectively, as well as (2.5) in order to obtain, under suitable assumptions,

$$\frac{\sqrt{\hat{\mathbf{n}}\, h^d}}{\sigma_{\mathbf{n}}(x)}\big(r_{\mathbf{n}}(x) - r(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

with

$$\sigma_{\mathbf{n}}^2(x) = \frac{v_{\mathbf{n}}(x)\displaystyle\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u}{f_{\mathbf{n}}(x)}.$$

3 Spatial prediction

We shall use in this section the notation of the introduction as well as the notation and hypotheses of Section 2, with $X_{\mathbf{i}} = \tilde{Z}_{\mathbf{i}}$ and $Y_{\mathbf{i}} = Z_{\mathbf{i}}$. Moreover, we assume that the kernel K is a *product kernel*, *i.e.*, that it may be written for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ as

$$K(x) = K_1(x_1) \ldots K_d(x_d),$$

where the K_i's are univariate and bounded densities. We now introduce the last hypothesis.

(**H6**) For all $\mathbf{i}, \mathbf{j} \in (\mathbb{N}^*)^N$ with $\mathbf{i} \neq \mathbf{j}$, there exists $\mathbf{l} \in \mathcal{V}_{\mathbf{j}}$ such that the pair $(\tilde{Z}_{\mathbf{i}}, Z_{\mathbf{l}})$ has a locally bounded density on $\mathbb{R}^{d+1}$.

Observe that since $\mathcal{V}$ is bounded, it is possible for a pair $\big((\tilde{Z}_{\mathbf{i}}, Z_{\mathbf{i}}), (\tilde{Z}_{\mathbf{j}}, Z_{\mathbf{j}})\big)$ to have a density as soon as $\mathrm{dist}(\mathbf{i}, \mathbf{j}) > D$, for some $D > 0$ which only depends on the size of $\mathcal{V}$. Therefore, in the context of prediction, the assumption (**H3**) makes sense. Condition (**H6**) holds in the important case where the random field is such that, for any bounded subset $\mathcal{B} \subset (\mathbb{N}^*)^N$, $(Z_{\mathbf{i}})_{\mathbf{i} \in \mathcal{B}}$ admits a bounded density. In this particular case, assumption (**H3**) holds for some $D > 0$ (depending upon the diameter of $\mathcal{V}$), as well as (**H2**) (see the proof of Corollary 3.1) and (**H5**) (see the proof of Corollary 3.2).

Corollary 3.1 *Assume that* (**H1**) *holds for any compact subset $\mathcal{S}$ of $\mathbb{R}^d$ and suppose moreover that* (**H3**) *and* (**H6**) *hold. If (2.3) is satisfied or if (2.4) is satisfied with $\chi(t) = \mathrm{O}(e^{-\xi t})$ for some $\xi > 0$, then*

(*i*) $\hat{r}\big(\tilde{Z}_{\mathbf{j}_0}\big) \to r\big(\tilde{Z}_{\mathbf{j}_0}\big)$ *in probability*;

(*ii*) $\hat{r}\big(\tilde{Z}_{\mathbf{j}_0}\big) \to r\big(\tilde{Z}_{\mathbf{j}_0}\big)$ *with probability 1, if $Z_{\mathbf{i}}$ is almost surely bounded.*

Proof of Corollary 3.1 By Theorem 2.1 or Corollary 2.1, one only needs to prove that the hypothesis (**H2**) holds. Let $\mathbf{i}, \mathbf{j} \in (\mathbb{N}^*)^N$ with $\mathbf{i} \neq \mathbf{j}$. Since K is a product kernel, there exists $c > 0$ such that for all $\mathbf{l} \in \mathcal{V}_{\mathbf{j}}$ and all $x \in \mathcal{S}$,

$$\mathbf{E}\Big[K\Big(\frac{x - \tilde{Z}_{\mathbf{i}}}{h}\Big)K\Big(\frac{x - \tilde{Z}_{\mathbf{j}}}{h}\Big)\Big] \le c\,\mathbf{E}\Big[K\Big(\frac{x - \tilde{Z}_{\mathbf{i}}}{h}\Big)K_s\Big(\frac{x_s - Z_{\mathbf{l}}}{h}\Big)\Big],$$

for some $s \in \{1, \ldots, d\}$. Using (**H6**), we choose $\mathbf{l} \in \mathcal{V}_{\mathbf{j}}$ such that the pair $(\tilde{Z}_{\mathbf{i}}, Z_{\mathbf{l}})$ has a locally bounded density on $\mathbb{R}^{d+1}$, and we easily deduce that (**H2**) holds. ■

In order to obtain the asymptotic normality of $\hat{r}\big(\tilde{Z}_{\mathbf{j}_0}\big)$, we need to introduce another concept of mixing, known as *φ-mixing*. More precisely, we shall assume that the random field $(\tilde{Z}_{\mathbf{i}}, Z_{\mathbf{i}})_{\mathbf{i}\in(\mathbb{N}^*)^N}$ is such that for all finite subsets E, E' of $(\mathbb{N}^*)^N$,

$$\sup\big\{|\mathbf{P}(A|B) - \mathbf{P}(A)|\big\} \leq \varphi\big(\mathrm{dist}(E, E')\big), \tag{3.1}$$

the supremum being taken over all $A \in \mathcal{B}(E)$ and $B \in \mathcal{B}(E')$ with $P(B) > 0$, and where φ is a real valued function with

$$\varphi(t) \to 0 \quad \text{as } t \to \infty\,.$$

Corollary 3.2 *Assume that* (**H3**)*,* (**H4**) *and* (**H6**) *hold for all $x \in \mathbb{R}^d$. Assume moreover that $(\tilde{Z}_{\mathbf{i}}, Z_{\mathbf{i}})_{\mathbf{i}\in(\mathbb{N}^*)^N}$ satisfies (3.1). Then, under conditions (i)-(iv) of Theorem 2.2, or conditions (i)-(iii) of Corollary 2.2, we have*

$$\frac{\sqrt{\hat{\mathbf{n}}\, h^d}}{\sigma(\tilde{Z}_{\mathbf{j}_0})}\big(\hat{r}(\tilde{Z}_{\mathbf{j}_0}) - r(\tilde{Z}_{\mathbf{j}_0})\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Proof of Corollary 3.2 Following the proof of Corollary 3.1, one may prove that (**H5**) holds, for all $x \in \mathbb{R}^d$. Consequently, by Theorem 2.2 or Corollary 2.2, we have, for all $x \in \mathbb{R}^d$,

$$\frac{\sqrt{\hat{\mathbf{n}}\, h^d}}{\sigma(x)}\big(r_{\mathbf{n}}(x) - r(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.2}$$

Let $k(\mathbf{n}) = \lfloor \log\log \hat{\mathbf{n}} \rfloor$. If $\hat{\mathbf{n}}$ is large enough, one can find a subset $\overline{\mathcal{P}}_{\mathbf{n}}$ of $\mathcal{P}_{\mathbf{n}} = \{\mathbf{i} \in \mathcal{S}_{\mathbf{n}} \,:\, \mathcal{V}_{\mathbf{i}} \subset \mathcal{S}_{\mathbf{n}}\}$ such that

$$\mathrm{dist}\big(\overline{\mathcal{P}}_{\mathbf{n}}, \mathcal{V}_{\mathbf{j}_0}\big) \geq k(\mathbf{n}) \quad \text{and} \quad \mathrm{Card}\,\big(\mathcal{P}_{\mathbf{n}} - \overline{\mathcal{P}}_{\mathbf{n}}\big) \leq c\, k(\mathbf{n})^d,$$

for some constant $c > 0$. We then denote, for all $x \in \mathbb{R}^d$,

$$L_{\mathbf{n}}(x) = \frac{\sqrt{\hat{\mathbf{n}}\, h^d}}{\sigma(x)}\left(\frac{\displaystyle\sum_{\mathbf{i}\in\overline{\mathcal{P}}_{\mathbf{n}}} Z_{\mathbf{i}}\, K\Big(\frac{x - \tilde{Z}_{\mathbf{i}}}{h}\Big)}{\displaystyle\sum_{\mathbf{i}\in\overline{\mathcal{P}}_{\mathbf{n}}} K\Big(\frac{x - \tilde{Z}_{\mathbf{i}}}{h}\Big)} - r(x)\right).$$

It is an easy exercise to prove that the result of the corollary will hold if we prove the convergence

$$L_{\mathbf{n}}(\tilde{Z}_{\mathbf{j}_0}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Notice also that a straightforward consequence of (1.1) and (3.2) is that, for all $x \in \mathbb{R}^d$,

$$L_{\mathbf{n}}(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.3}$$

Let now $u \in \mathbb{R}$ and $x \in \mathbb{R}^d$. From Serfling [14] and the definition of φ-mixing (3.1), we can write

$$\Big|\mathbf{E}\big[\exp\big(iuL_{\mathbf{n}}(x)\big)\big|\tilde{Z}_{\mathbf{j}_0}\big] - \mathbf{E}\exp\big(iuL_{\mathbf{n}}(x)\big)\Big| \leq 2\,\varphi\big(k(\mathbf{n})\big).$$

Since $\varphi\big(k(\mathbf{n})\big) \to 0$, we deduce from (3.3) that, with probability 1,

$$\mathbf{E}\big[\exp\big(iuL_{\mathbf{n}}(x)\big)\big|\tilde{Z}_{\mathbf{j}_0}\big] \to \exp(-u^2/2).$$

Finally, according to the Lebesgue convergence theorem,

$$\mathbf{E}\exp\big(iuL_{\mathbf{n}}(\tilde{Z}_{\mathbf{j}_0})\big) = \int_{\mathbb{R}^d} \mathbf{E}\big[\exp\big(iuL_{\mathbf{n}}(x)\big)\big|\tilde{Z}_{\mathbf{j}_0} = x\big] f(x)\mathrm{d}x \to \exp(-u^2/2),$$

and hence the corollary is proved. ■

4 Proofs

4.1 Block decomposition and general notation

In this section, we will make use of the following notation. If $x \in \mathbb{R}^d$, we let

$$\zeta(x) = \int_{\mathbb{R}^d} y g(x,y)\mathrm{d}y\,;$$

$$\zeta_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}}\, h^d} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}\, K\Big(\frac{x - X_{\mathbf{i}}}{h}\Big);$$

$$\zeta^*_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}}\, h^d} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}\, \mathbf{1}_{(|Y_{\mathbf{i}}| \leq \gamma_{\mathbf{n}})} K\Big(\frac{x - X_{\mathbf{i}}}{h}\Big),$$

where $\gamma_{\mathbf{n}} = (3a^{-1}\log \hat{\mathbf{n}})^{1/b}$. Recall that the constants a and b have been defined in (2.2). Moreover, we set

$$\Delta_{\mathbf{i}}(x) = \frac{1}{\hat{\mathbf{n}}\, h^d}\bigg(Y_{\mathbf{i}}\, \mathbf{1}_{(|Y_{\mathbf{i}}| \leq \gamma_{\mathbf{n}})} K\Big(\frac{x - X_{\mathbf{i}}}{h}\Big) - \mathbf{E}\Big[Y\mathbf{1}_{(|Y| \leq \gamma_{\mathbf{n}})} K\Big(\frac{x - X}{h}\Big)\Big]\bigg).$$

Let us now introduce a spatial block decomposition that has been used by Tran [15] and Carbon, Tran and Wu [4]. Let us fix two integers $p, q \geq 1$ (possibly functions of $\mathbf{n}$) and assume that for some integers $t_1, \ldots, t_N$,

$$n_i = t_i(p+q), \quad i = 1, \ldots, N. \tag{4.1}$$

The random variables $\Delta_{\mathbf{i}}(x)$ are now set into blocks of different sizes. Let

$$U(1, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+1 \\ k=1,\ldots,N}}^{j_k(p+q)+p} \Delta_{\mathbf{i}}(x)\,,$$

$$U(2, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+1 \\ k=1,\ldots,N-1}}^{j_k(p+q)+p} \; \sum_{i_N = j_N(p+q)+p+1}^{(j_N+1)(p+q)} \Delta_{\mathbf{i}}(x)\,,$$

$$U(3, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+1 \\ k=1,\ldots,N-2}}^{j_k(p+q)+p} \; \sum_{i_{N-1} = j_{N-1}(p+q)+p+1}^{(j_{N-1}+1)(p+q)} \; \sum_{i_N = j_N(p+q)+1}^{j_N(p+q)+p} \Delta_{\mathbf{i}}(x)\,,$$

$$U(4, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+1 \\ k=1,\ldots,N-2}}^{j_k(p+q)+p} \; \sum_{i_{N-1} = j_{N-1}(p+q)+p+1}^{(j_{N-1}+1)(p+q)} \; \sum_{i_N = j_N(p+q)+p+1}^{(j_N+1)(p+q)} \Delta_{\mathbf{i}}(x)$$

and so on. Note that

$$U(2^{N-1}, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+p+1 \\ k=1,\ldots,N-1}}^{(j_k+1)(p+q)} \; \sum_{i_N = j_N(p+q)+1}^{j_N(p+q)+p} \Delta_{\mathbf{i}}(x)\,.$$

Finally

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

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

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

Then, with this notation, we obtain the decomposition

$$\zeta_{\mathbf{n}}^*(x) - \mathbf{E}\zeta_{\mathbf{n}}^*(x) = \sum_{i=1}^{2^N} T(\mathbf{n}, x, i)\,.$$

If it is not the case that $n_1 = t_1(p+q), \dots, n_N = t_N(p+q)$ for some integers $t_1, \dots, t_N$, then a term, say $T(\mathbf{n}, x, 2^N+1)$, containing all the $\Delta_{\mathbf{i}}(x)$'s at the ends not included in the blocks above can be added.

From now on, the letter c is used to denote constants whose values are unimportant and may vary from line to line.

4.2 General results

In this paragraph, we provide three general propositions that will be useful in the sequel.

Proposition 4.1 *Under (2.2), one has*

$$\sum_{\mathbf{n}\in(\mathbb{N}^*)^N} \mathbf{P}\big(\exists \mathbf{i} \in \mathcal{I}_{\mathbf{n}} \,:\, |Y_{\mathbf{i}}| > \gamma_{\mathbf{n}}\big) < \infty \,.$$

Proof of Proposition 4.1 By stationarity of the random field $(Y_{\mathbf{i}})_{\mathbf{i}\in(\mathbb{N}^*)^N}$ and according to Markov inequality, we have

$$\begin{aligned}
\mathbf{P}\big(\exists \mathbf{i} \in \mathcal{I}_{\mathbf{n}} \,:\, |Y_{\mathbf{i}}| > \gamma_{\mathbf{n}}\big) &\le \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \mathbf{P}\big(|Y_{\mathbf{i}}| > \gamma_{\mathbf{n}}\big) \\
&= \hat{\mathbf{n}}\, \mathbf{P}\big(|Y| > \gamma_{\mathbf{n}}\big) \\
&\le \hat{\mathbf{n}}\, \frac{\mathbf{E}\exp\big(a|Y|^b\big)}{\exp\big(a\gamma_{\mathbf{n}}^b\big)} \\
&= \frac{1}{(\hat{\mathbf{n}})^2}\mathbf{E}\exp\big(a|Y|^b\big),
\end{aligned}$$

hence the proposition, using (2.2). ■

Proposition 4.2 *Let $i \in \{1, 2\}$. Under (2.2), there exists $c > 0$ such that for $\hat{\mathbf{n}}$ large enough and all $x \in \mathbb{R}^d$,*

$$\mathbf{E}\Big[|Y|\mathbf{1}_{(|Y|>\gamma_{\mathbf{n}})} K\Big(\frac{x-X}{h}\Big)\Big]^i \le c\,(\hat{\mathbf{n}})^{-3/4}\mathbf{E}^{1/2} K\Big(\frac{x-X}{h}\Big)^{2i}.$$

Proof of Proposition 4.2 For $\hat{\mathbf{n}}$ large enough, we can write

$$\mathbf{E}\Big[|Y|\mathbf{1}_{(|Y|>\gamma_{\mathbf{n}})} K\Big(\frac{x-X}{h}\Big)\Big]^i \le \mathbf{E}\Big[\exp\big(a|Y|^b/(4i)\big)\mathbf{1}_{(|Y|>\gamma_{\mathbf{n}})} K\Big(\frac{x-X}{h}\Big)\Big]^i.$$

We deduce from the Cauchy-Schwarz inequality that the rightmost term is smaller than

$$\begin{aligned}&\mathbf{E}^{1/2}\Big[\exp\big(a|Y|^b/2\big)\mathbf{1}_{(|Y|>\gamma_{\mathbf{n}})}\Big]\mathbf{E}^{1/2}K\Big(\frac{x-X}{h}\Big)^{2i}\\&\leq \exp(-a\gamma_{\mathbf{n}}^b/4)\mathbf{E}^{1/2}\exp\big(a|Y|^b\big)\mathbf{E}^{1/2}K\Big(\frac{x-X}{h}\Big)^{2i},\end{aligned}$$

hence the result, by (2.2) and the very definition of $\gamma_{\mathbf{n}}$. ■

Proposition 4.3 *(i) Under the hypotheses of Theorem 2.1, one has, for large enough* $\hat{\mathbf{n}}$,

$$\hat{\mathbf{n}}\,h^d\sup_{x\in\mathcal{S}}\Big(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\mathbf{E}\Delta_{\mathbf{i}}^2(x)+\sum_{\substack{\mathbf{i},\mathbf{j}\in\mathcal{I}_{\mathbf{n}}\\ \mathbf{i}\neq\mathbf{j}}}|\mathbf{E}\Delta_{\mathbf{i}}(x)\Delta_{\mathbf{j}}(x)|\Big)\leq c\,\gamma_{\mathbf{n}}^2\,;$$

(ii) Under the hypotheses of Theorem 2.2, one has

$$\hat{\mathbf{n}}\,h^d\mathrm{Var}\,\zeta_{\mathbf{n}}^*(x)\to f(x)\big(v(x)+r^2(x)\big)\int_{\mathbb{R}^d}K^2(u)\mathrm{d}u\,.$$

Proof of Proposition 4.3 The proof of (i) follows the proof of Lemma 2.2 in Tran [15] and the proof of (ii) below. Therefore, we only prove (ii). Let

$$\tilde{I}_{\mathbf{n}}=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\mathbf{E}\Delta_{\mathbf{i}}^2(x)$$

and

$$\tilde{R}_{\mathbf{n}}=\sum_{\substack{\mathbf{i},\mathbf{j}\in\mathcal{I}_{\mathbf{n}}\\ \mathbf{i}\neq\mathbf{j}}}\mathbf{E}\Delta_{\mathbf{i}}(x)\Delta_{\mathbf{j}}(x).$$

Clearly, $\mathrm{Var}\,\zeta_{\mathbf{n}}^*(x)=\tilde{I}_{\mathbf{n}}+\tilde{R}_{\mathbf{n}}$. Observe that, by stationarity,

$$\begin{aligned}\hat{\mathbf{n}}\,h^d\tilde{I}_{\mathbf{n}}&=h^{-d}\bigg(\mathbf{E}\Big[Y\mathbf{1}_{(|Y|\leq\gamma_{\mathbf{n}})}K\Big(\frac{x-X}{h}\Big)\Big]^2-\mathbf{E}^2\Big[Y\mathbf{1}_{(|Y|\leq\gamma_{\mathbf{n}})}K\Big(\frac{x-X}{h}\Big)\Big]\bigg)\\&=h^{-d}\bigg(\mathbf{E}\Big[YK\Big(\frac{x-X}{h}\Big)\Big]^2-\mathbf{E}\Big[Y\mathbf{1}_{(|Y|>\gamma_{\mathbf{n}})}K\Big(\frac{x-X}{h}\Big)\Big]^2\\&\quad-\mathbf{E}^2\Big[Y\mathbf{1}_{(|Y|\leq\gamma_{\mathbf{n}})}K\Big(\frac{x-X}{h}\Big)\Big]\bigg).\end{aligned}$$

By Proposition 4.2 and hypothesis (**H4**), one has

$$\begin{aligned}h^{-d}\mathbf{E}\Big[Y\mathbf{1}_{(|Y|>\gamma_{\mathbf{n}})}K\Big(\frac{x-X}{h}\Big)\Big]^2&\leq c\,h^{-d}(\hat{\mathbf{n}})^{-3/4}\mathbf{E}^{1/2}K\Big(\frac{x-X}{h}\Big)^4\\&\leq c\,h^{-d/2}(\hat{\mathbf{n}})^{-3/4}.\end{aligned}$$

Since $\hat{\mathbf{n}}\, h^d \to \infty$ by Theorem 2.2 (ii), one deduces that

$$h^{-d}\mathbf{E}\Big[Y\mathbf{1}_{(|Y|>\gamma_{\mathbf{n}})}K\Big(\frac{x-X}{h}\Big)\Big]^2 \to 0.$$

In a similar fashion, one may prove that

$$h^{-d}\mathbf{E}^2\Big[Y\mathbf{1}_{(|Y|\leq\gamma_{\mathbf{n}})}K\Big(\frac{x-X}{h}\Big)\Big] \to 0.$$

Finally, by a classical change of variable, we have

$$\begin{aligned} h^{-d}\mathbf{E}\Big[YK\Big(\frac{x-X}{h}\Big)\Big]^2 &= \int_{\mathbb{R}^{d+1}} y^2K^2(u)g(x-hu,y)\mathrm{d}u\mathrm{d}y \\ &= \int_{\mathbb{R}^d} K^2(u)\mathbf{E}[Y^2|X=x-hu]f(x-hu)\mathrm{d}u\,, \end{aligned}$$

so that by (**H4**) and the fact that K is compactly supported,

$$h^{-d}\mathbf{E}\Big[YK\Big(\frac{x-X}{h}\Big)\Big]^2 \to f(x)\big(v(x)+r^2(x)\big)\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u\,.$$

Consequently,

$$\hat{\mathbf{n}}\, h^d\tilde{I}_{\mathbf{n}} \to f(x)\big(v(x)+r^2(x)\big)\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u\,.$$

We have now to show that

$$\hat{\mathbf{n}}\, h^d\tilde{R}_{\mathbf{n}} \to 0.$$

To this aim, we introduce the set S_0 defined by

$$S_0 = \{\mathbf{i},\mathbf{j}\in\mathcal{I}_{\mathbf{n}}\,:\,0<\mathrm{dist}(\mathbf{i},\mathbf{j})\leq D\},$$

where D is the real number of hypothesis (**H3**). Split $\tilde{R}_{\mathbf{n}}$ into two separate sums, *i.e.*, $\tilde{R}_{\mathbf{n}} = J_0 + J_0'$, where

$$J_0 = \sum_{\mathbf{i},\mathbf{j}\in S_0} \mathbf{E}\Delta_{\mathbf{i}}(x)\Delta_{\mathbf{j}}(x) \quad \text{and} \quad J_0' = \sum_{\mathbf{i},\mathbf{j}\in S_0^c} \mathbf{E}\Delta_{\mathbf{i}}(x)\Delta_{\mathbf{j}}(x),$$

where S_0^c stands for the complement of S_0. A straightforward modification of the proof of Lemma 2.2 in Tran [15] gives

$$\hat{\mathbf{n}}\, h^d J_0' \to 0,$$

and it remains to prove that the same result holds for J_0. Since Card $S_0 \leq c\,\hat{\mathbf{n}}$, one obtains by strict stationarity of the field $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in(\mathbb{N}^*)^N}$ that, for $\mathbf{i} \neq \mathbf{j}$,

$$\begin{aligned}
\hat{\mathbf{n}}\, h^d J_0 &\leq c\, h^{-d}\left(\mathbf{E}\left[|Y_{\mathbf{i}}|\mathbf{1}_{(|Y_{\mathbf{i}}|\leq\gamma_{\mathbf{n}})}K\Big(\frac{x-X_{\mathbf{i}}}{h}\Big)|Y_{\mathbf{j}}|\mathbf{1}_{(|Y_{\mathbf{j}}|\leq\gamma_{\mathbf{n}})}K\Big(\frac{x-X_{\mathbf{j}}}{h}\Big)\right]\right.\\
&\qquad \left.+\,\mathbf{E}^2\left[|Y|\mathbf{1}_{(|Y|\leq\gamma_{\mathbf{n}})}K\Big(\frac{x-X}{h}\Big)\right]\right)\\
&\leq c\, h^{-d}\gamma_{\mathbf{n}}^2\left(\mathbf{E}\left[K\Big(\frac{x-X_{\mathbf{i}}}{h}\Big)K\Big(\frac{x-X_{\mathbf{j}}}{h}\Big)\right]+\mathbf{E}^2K\Big(\frac{x-X}{h}\Big)\right)\\
&\leq c\, h(\log\hat{\mathbf{n}})^{2/b}h^{-d-1}\mathbf{E}\left[K\Big(\frac{x-X_{\mathbf{i}}}{h}\Big)K\Big(\frac{x-X_{\mathbf{j}}}{h}\Big)\right]\\
&\qquad +\,c\, h^d(\log\hat{\mathbf{n}})^{2/b}h^{-2d}\mathbf{E}^2K\Big(\frac{x-X}{h}\Big).
\end{aligned}$$

By (i) of Theorem 2.2, we have $h(\log\hat{\mathbf{n}})^{2/b} \to 0$ and hence, the rightmost terms tend to 0 according to (**H4**) and (**H5**). Consequently, $\hat{\mathbf{n}}\, h^d J_0 \to 0$ so that $\hat{\mathbf{n}}\, h^d \tilde{R}_{\mathbf{n}} \to 0$, and the proposition is proved. ■

4.3 Proof of Theorem 2.1

We assume in this paragraph that assumptions (**H1**)-(**H3**) hold as well as (2.3). Let l be the integer defined by $l = \lfloor \gamma_{\mathbf{n}}^2 h^{-(d+1)} \rfloor$. Since $\mathcal{S}$ is compact, one can find a real number $v > 0$ such that $lv \leq c$ and

$$\mathcal{S} \subset \bigcup_{j=1}^{l^d} B(x^j, v),$$

where $B(x,\rho)$ is the closed ball in $\mathbb{R}^d$ with center at x and radius $\rho > 0$, and the x^j 's are elements of $\mathcal{S}$. Theorem 2.1 will be a consequence of the following three lemmas, whose proofs are deferred to the end of the paragraph.

Lemma 4.1 *One has*

$$\sup_{x\in\mathcal{S}} |\mathbf{E}\zeta_{\mathbf{n}}^*(x) - \zeta(x)| \to 0.$$

Lemma 4.2 *One has, almost completely,*

$$\max_{j\leq l^d} |\mathbf{E}\zeta_{\mathbf{n}}^*(x^j) - \zeta_{\mathbf{n}}^*(x^j)| \to 0.$$

Lemma 4.3 *One has, with probability 1,*

$$\sup_{x\in\mathcal{S}} |f_{\mathbf{n}}(x) - f(x)| \to 0.$$

We are now ready to prove Theorem 2.1. Observe first that for any $x \in \mathcal{S}$, provided $f_{\mathbf{n}}(x) \neq 0$,

$$r_{\mathbf{n}}(x) - r(x) = \frac{\zeta_{\mathbf{n}}(x) - \zeta(x)}{f_{\mathbf{n}}(x)} + \zeta(x)\, \frac{f_{\mathbf{n}}(x) - f(x)}{f_{\mathbf{n}}(x) f(x)}\,.$$

According to (**H1**) and Lemma 4.3, it is enough to show that, with probability 1,

$$\sup_{x \in \mathcal{S}} |\zeta_{\mathbf{n}}(x) - \zeta(x)| \to 0.$$

Now, let $\varepsilon > 0$. One has

$$\begin{aligned}
\mathbf{P}\big(\sup_{x\in\mathcal{S}} |\zeta_{\mathbf{n}}(x) - \zeta(x)| \geq \varepsilon\big) &\leq \mathbf{P}\big(\sup_{x\in\mathcal{S}} |\zeta_{\mathbf{n}}(x) - \zeta(x)| \geq \varepsilon, \forall \mathbf{i} \in \mathcal{I}_{\mathbf{n}} \,:\, |Y_{\mathbf{i}}| \leq \gamma_{\mathbf{n}}\big) \\
&\quad + \mathbf{P}\big(\exists \mathbf{i} \in \mathcal{I}_{\mathbf{n}} \,:\, |Y_{\mathbf{i}}| > \gamma_{\mathbf{n}}\big) \\
&\leq \mathbf{P}\big(\sup_{x\in\mathcal{S}} |\zeta^*_{\mathbf{n}}(x) - \zeta(x)| \geq \varepsilon\big) \\
&\quad + \mathbf{P}\big(\exists \mathbf{i} \in \mathcal{I}_{\mathbf{n}} \,:\, |Y_{\mathbf{i}}| > \gamma_{\mathbf{n}}\big).
\end{aligned}$$

According to Proposition 4.1, it will be enough to show that

$$\sum_{\mathbf{n} \in (\mathbb{N}^*)^N} \mathbf{P}\big(\sup_{x\in\mathcal{S}} |\zeta^*_{\mathbf{n}}(x) - \zeta(x)| \geq \varepsilon\big) < \infty\,.$$

But, for each $x \in \mathcal{S}$, one can find $j \leq l^d$ such that $\|x - x^j\| \leq v$ and consequently, as K is Lipschitzian,

$$|\zeta^*_{\mathbf{n}}(x) - \zeta^*_{\mathbf{n}}(x^j)| \leq c\,\gamma_{\mathbf{n}} v h^{-(d+1)}.$$

Therefore we get

$$\begin{aligned}
&\sup_{x\in\mathcal{S}} |\zeta^*_{\mathbf{n}}(x) - \zeta(x)| \\
&\quad \leq \sup_{x\in\mathcal{S}} |\zeta^*_{\mathbf{n}}(x) - \mathbf{E}\zeta^*_{\mathbf{n}}(x)| + \sup_{x\in\mathcal{S}} |\mathbf{E}\zeta^*_{\mathbf{n}}(x) - \zeta(x)| \\
&\quad \leq c\,\gamma_{\mathbf{n}} v h^{-(d+1)} + \max_{j \leq l^d} |\zeta^*_{\mathbf{n}}(x^j) - \mathbf{E}\zeta^*_{\mathbf{n}}(x^j)| + \sup_{x\in\mathcal{S}} |\mathbf{E}\zeta^*_{\mathbf{n}}(x) - \zeta(x)| \\
&\quad \leq c(\gamma_{\mathbf{n}})^{-1} + \max_{j \leq l^d} |\zeta^*_{\mathbf{n}}(x^j) - \mathbf{E}\zeta^*_{\mathbf{n}}(x^j)| + \sup_{x\in\mathcal{S}} |\mathbf{E}\zeta^*_{\mathbf{n}}(x) - \zeta(x)|\,,
\end{aligned}$$

where the last inequality arises from the very definition of l and the fact that $lv \leq c$. Therefore, according to Lemma 4.1, we obtain, for $\hat{\mathbf{n}}$ large enough,

$$\mathbf{P}\big(\sup_{x\in S} |\zeta^*_{\mathbf{n}}(x) - \zeta(x)| \geq \varepsilon\big) \leq \mathbf{P}\big(\max_{j \leq l^d} |\zeta^*_{\mathbf{n}}(x^j) - \mathbf{E}\zeta^*_{\mathbf{n}}(x^j)| \geq \varepsilon/2\big),$$

hence the theorem, thanks to Lemma 4.2. ■

Proof of Lemma 4.1 The proof runs in two steps. First, we observe that by Proposition 4.2, for all $x \in \mathcal{S}$,

$$\begin{aligned}|\mathbf{E}\zeta_{\mathbf{n}}^*(x) - \mathbf{E}\zeta_{\mathbf{n}}(x)| &\le \mathbf{E}\Big[|Y|\mathbf{1}_{(|Y|>\gamma_{\mathbf{n}})} K\Big(\frac{x-X}{h}\Big)\Big] \\ &\le c\,(\hat{\mathbf{n}})^{-3/4}\mathbf{E}^{1/2}K\Big(\frac{x-X}{h}\Big)^2,\end{aligned}$$

and hence, using (**H1**), one has

$$\sup_{x\in\mathcal{S}} |\mathbf{E}\zeta_{\mathbf{n}}^*(x) - \mathbf{E}\zeta_{\mathbf{n}}(x)| \le c\,(\hat{\mathbf{n}})^{-3/4} h^{d/2},$$

which tends to 0. Second, we observe that for all $x \in \mathcal{S}$,

$$\begin{aligned}\mathbf{E}\zeta_{\mathbf{n}}(x) - \zeta(x) &= \int_{\mathbb{R}^{d+1}} yK(u)g(x-hu,y)\mathrm{d}u\mathrm{d}y - r(x)f(x) \\ &= \int_{\mathbb{R}^d} K(u)\big(r(x-hu)f(x-hu) - r(x)f(x)\big)\mathrm{d}u\,. \qquad (4.2)\end{aligned}$$

Now, K is compactly supported and, by (**H1**), the function $r\,f$ is continuous on $\tilde{\mathcal{S}}$. Consequently, according to the Lebesgue theorem,

$$\sup_{x\in\mathcal{S}} |\mathbf{E}\zeta_{\mathbf{n}}(x) - \zeta(x)| \to 0,$$

and the proof is complete. ■

Proof of Lemma 4.2 We follow Carbon, Tran and Wu [4]. Without loss of generality, we assume that $\mathbf{n}$ satisfies (4.1). We adopt the notation of Paragraph 4.1, and we choose $p = q$, where $p = p(\mathbf{n})$ is defined in Theorem 2.1. Using the block decomposition of Paragraph 4.1, we have, for $x \in \mathcal{S}$,

$$\zeta_{\mathbf{n}}^*(x) - \mathbf{E}\zeta_{\mathbf{n}}^*(x) = \sum_{i=1}^{2^N} T(\mathbf{n}, x, i).$$

Thus, to prove the lemma, it is enough to show that for all $i = 1, \dots, 2^N$,

$$\max_{j\le l^d} |T(\mathbf{n}, x^j, i)| \to 0,$$

almost completely. For a sake of simplicity, we will only consider the case $i = 1$. Let us first fix a generic $x \in \mathcal{S}$. We enumerate in an arbitrary way the $\hat{t} = t_1 \dots t_N$ terms $(U(1, \mathbf{n}, x, \mathbf{j}), \mathbf{j} \in \mathcal{T})$ of the sum $T(\mathbf{n}, x, 1)$ that we call $W_1, \dots, W_{\hat{t}}$. Observe that the set of sites associated with each of these random variables contains p^N sites, that two distinct sets of sites are separated by a distance of at least p and finally, that for all $j = 1, \dots, \hat{t}$,

$$|W_j| \le c\, p^N (\hat{\mathbf{n}}\, h^d)^{-1} \gamma_{\mathbf{n}}.$$

According to Lemma 4.5 in Carbon, Tran and Wu [4], one can find independent random variables $W_1^*, \dots, W_{\hat{t}}^*$ such that for all $j = 1, \dots, \hat{t}$, W_j^* has the same law as W_j and

$$\begin{aligned} \mathbf{E}|W_j - W_j^*| &\le c\, p^N (\hat{\mathbf{n}}\, h^d)^{-1} \gamma_{\mathbf{n}}\, \psi(\hat{\mathbf{n}}, p^N)\, \chi(p) \\ &\le c\, p^{2N} (\hat{\mathbf{n}}\, h^d)^{-1} \gamma_{\mathbf{n}}\, \chi(p), \end{aligned}$$

where the last inequality comes from (2.1) and the fact that $p^N \le \hat{\mathbf{n}}$. Then we fix $\varepsilon > 0$. Bernstein and Markov inequalities lead to

$$\begin{aligned} \mathbf{P}\big(|T(\mathbf{n}, x, 1)| \ge 2\varepsilon\big) &\le \mathbf{P}\Big(\Big|\sum_{j=1}^{\hat{t}} W_j^*\Big| \ge \varepsilon\Big) + \mathbf{P}\Big(\sum_{j=1}^{\hat{t}} |W_j - W_j^*| \ge \varepsilon\Big) \\ &\le 2\, \exp\Big(- c \big/ \Big(\sum_{j=1}^{\hat{t}} \mathbf{E}(W_j^*)^2 + p^N (\hat{\mathbf{n}}\, h^d)^{-1} \gamma_{\mathbf{n}} \Big)\Big) \\ &\quad + c\, \hat{t}\, p^{2N} (\hat{\mathbf{n}}\, h^d)^{-1} \gamma_{\mathbf{n}}\, \chi(p). \end{aligned}$$

By Proposition 4.3 (i), one has $\hat{\mathbf{n}}\, h^d \sum_{j=1}^{\hat{t}} \mathbf{E}(W_j^*)^2 \le c\, \gamma_{\mathbf{n}}^2$. Since $\hat{\mathbf{n}} = 2^N p^N \hat{t}$ by (4.1), one deduces that

$$\begin{aligned} \mathbf{P}\big(\max_{j \le l^d} |T(\mathbf{n}, x^j, 1)| \ge 2\varepsilon\big) &\le 2\, l^d \exp\Big(- c\, \hat{\mathbf{n}}\, h^d / \big(\gamma_{\mathbf{n}}^2 + p^N \gamma_{\mathbf{n}}\big)\Big) \\ &\quad + c\, l^d p^N h^{-d} \gamma_{\mathbf{n}}\, \chi(p). \end{aligned}$$

Now, let

$$A_{\mathbf{n}} = \min\Big(\frac{\hat{\mathbf{n}}\, h^d}{(\log \hat{\mathbf{n}})^{1+2/b}}\, ,\ (\log \hat{\mathbf{n}})^{1/b} \Big).$$

By the definitions of l and $\gamma_{\mathbf{n}}$, we have

$$\begin{aligned} \mathbf{P}\big(\max_{j \le l^d} |T(\mathbf{n}, x^j, 1)| \ge 2\varepsilon\big) &\le c\, l^d \big(\exp(-c A_{\mathbf{n}} \log \hat{\mathbf{n}}) + p^N h^{-d} \gamma_{\mathbf{n}} \chi(p)\big) \\ &\le c\, h^{-d(d+1)} (\log \hat{\mathbf{n}})^{2d/b} \big(\exp(-c A_{\mathbf{n}} \log \hat{\mathbf{n}}) \\ &\quad + p^N h^{-d} (\log \hat{\mathbf{n}})^{1/b} \chi(p)\big). \end{aligned}$$

Observe that since $\hat{\mathbf{n}}\, h^d/(\log \hat{\mathbf{n}})^{1+2/b} \to \infty$ by (2.3), one has

$$\sum_{\mathbf{n}\in(\mathbb{N}^*)^N} h^{-d(d+1)} (\log \hat{\mathbf{n}})^{2d/b} \exp(-cA_{\mathbf{n}} \log \hat{\mathbf{n}}) < \infty\,.$$

Consequently, from (2.3),

$$\sum_{\mathbf{n}\in(\mathbb{N}^*)^N} \mathbf{P}\big(\max_{j\le l^d} |T(\mathbf{n}, x^j, 1)| \ge 2\varepsilon\big) < \infty\,.$$

This completes the proof of the lemma. ■

Proof of Lemma 4.3 Just adapt the proof of Carbon, Tran and Wu [4] to the case where f is only continuous on $\mathcal{S}$, or adapt the arguments considered above to the case $Y \equiv 1$. ■

4.4 Proof of Theorem 2.2

4.4.1 Main arguments

In the whole Paragraph 4.4, assumptions (**H3**)-(**H5**) as well as hypotheses (i)-(iv) of Theorem 2.2 are assumed to be satisfied. The proof of Theorem 2.2 will strongly rely on the next two propositions, whose proofs are deferred to the end of the paragraph.

Proposition 4.4 *For any pair* $(\alpha, \beta) \in \mathbb{R}^2$*, one has*

$$\sqrt{\hat{\mathbf{n}}\, h^d}\Big[\alpha\big(\zeta^*_{\mathbf{n}}(x) - \mathbf{E}\zeta^*_{\mathbf{n}}(x)\big) + \beta\big(f_{\mathbf{n}}(x) - \mathbf{E} f_{\mathbf{n}}(x)\big)\Big] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^*),$$

where

$$\sigma^* = \big[\alpha^2 f(x)\big(v(x) + r^2(x)\big) + \beta^2 f(x) + 2\alpha\beta r(x) f(x)\big] \int_{\mathbb{R}^d} K^2(u)\mathrm{d}u\,.$$

Proposition 4.5 *One has*

(i) $\sqrt{\hat{\mathbf{n}}\, h^d}\big(\mathbf{E}\zeta^*_{\mathbf{n}}(x) - \zeta(x)\big) \to 0$;

(ii) $\sqrt{\hat{\mathbf{n}}\, h^d}\big(\mathbf{E} f_{\mathbf{n}}(x) - f(x)\big) \to 0.$

We are now in position to prove Theorem 2.2. Let $(\alpha, \beta) \in \mathbb{R}^2$ and denote by $W_{\mathbf{n}}(\alpha, \beta)$ the random variable

$$W_{\mathbf{n}}(\alpha, \beta) = \sqrt{\hat{\mathbf{n}}\, h^d}\Big[\alpha\big(\zeta_{\mathbf{n}}(x) - \zeta(x)\big) + \beta\big(f_{\mathbf{n}}(x) - f(x)\big)\Big],$$

and by $W^*_{\mathbf{n}}(\alpha,\beta)$ the random variable being defined similarly, with $\zeta^*_{\mathbf{n}}(x)$ instead of $\zeta_{\mathbf{n}}(x)$. Denote also by $\mathcal{G}_{\mathbf{n}}$ the event

$$\mathcal{G}_{\mathbf{n}} = \big[\forall \mathbf{i} \in \mathcal{I}_{\mathbf{n}} \,:\, |Y_{\mathbf{i}}| \le \gamma_{\mathbf{n}}\big].$$

Fix $t \in \mathbb{R}$ and observe that

$$\mathbf{P}\big(W_{\mathbf{n}}(\alpha,\beta) \le t, \mathcal{G}_{\mathbf{n}}\big) = \mathbf{P}\big(W^*_{\mathbf{n}}(\alpha,\beta) \le t, \mathcal{G}_{\mathbf{n}}\big).$$

By Proposition 4.1, $\mathbf{P}(\mathcal{G}_{\mathbf{n}}) \to 1$ and hence

$$\mathbf{P}\big(W_{\mathbf{n}}(\alpha,\beta) \le t\big) - \mathbf{P}\big(W^*_{\mathbf{n}}(\alpha,\beta) \le t\big) \to 0. \tag{4.3}$$

Moreover, according to Proposition 4.5,

$$W^*_{\mathbf{n}}(\alpha,\beta) - \sqrt{\hat{\mathbf{n}}\, h^d}\Big[\alpha\big(\zeta^*_{\mathbf{n}}(x) - \mathbf{E}\zeta^*_{\mathbf{n}}(x)\big) + \beta\big(f_{\mathbf{n}}(x) - f(x)\big)\Big] \to 0. \tag{4.4}$$

We then deduce from Proposition 4.4 and properties (4.3) and (4.4) that

$$W_{\mathbf{n}}(\alpha,\beta) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^*).$$

From this convergence, it follows that $f_{\mathbf{n}}(x) \to f(x)$ and $\zeta_{\mathbf{n}}(x) \to \zeta(x)$ in probability, so that

$$W_{\mathbf{n}}\Big(\frac{1}{f_{\mathbf{n}}(x)}, -\frac{\zeta(x)}{f(x)f_{\mathbf{n}}(x)}\Big) - W_{\mathbf{n}}\Big(\frac{1}{f(x)}, -\frac{\zeta(x)}{f^2(x)}\Big) \to 0 \quad \text{in probability.}$$

Finally, we conclude with the decomposition

$$\begin{aligned} r_{\mathbf{n}}(x) - r(x) &= \frac{\zeta_{\mathbf{n}}(x)}{f_{\mathbf{n}}(x)} - \frac{\zeta(x)}{f(x)} \\ &= \frac{1}{f_{\mathbf{n}}(x)}\big(\zeta_{\mathbf{n}}(x) - \zeta(x)\big) - \frac{\zeta(x)}{f(x)f_{\mathbf{n}}(x)}\big(f_{\mathbf{n}}(x) - f(x)\big) \\ &= W_{\mathbf{n}}\Big(\frac{1}{f_{\mathbf{n}}(x)}, -\frac{\zeta(x)}{f(x)f_{\mathbf{n}}(x)}\Big). \end{aligned}$$

■

4.4.2 Proof of Proposition 4.4

For the sake of simplicity, we only give the proof in the case $\alpha = 1$ and $\beta = 0$. The proof of the general case is similar, putting $\alpha Y_{\mathbf{i}} + \beta$ instead of $Y_{\mathbf{i}}$.

Without loss of generality, we assume that $\mathbf{n}$ satisfies (4.1) and we use the decomposition of Paragraph 4.1, where the integer $p = p(\mathbf{n})$ is defined by

$$p = \left\lfloor \frac{(\hat{\mathbf{n}}\, h^d)^{1/(2N)}}{\gamma_{\mathbf{n}}^2} \right\rfloor,$$

and the integer $q = q(\mathbf{n})$ satisfies the conditions (iii) and (iv) of Theorem 2.2. Note that by assertion (ii) of Theorem 2.2, one has $p \to \infty$. We first prove a technical lemma.

Lemma 4.4 *One has*

(i) *For all* $u \in \mathbb{R}$,

$$\mathbf{E}\exp\big(iu\sqrt{\hat{\mathbf{n}}\, h^d}T(\mathbf{n},x,1)\big) - \prod_{\mathbf{j}\in\mathcal{T}}\mathbf{E}\exp\big(iu\sqrt{\hat{\mathbf{n}}\, h^d}U(1,\mathbf{n},x,\mathbf{j})\big) \to 0;$$

(ii) $\hat{\mathbf{n}}\, h^d \mathbf{E}\Big(\sum_{i=2}^{2^N} T(\mathbf{n},x,i)\Big)^2 \to 0;$

(iii) $\hat{\mathbf{n}}\, h^d \sum_{\mathbf{j}\in\mathcal{T}} \mathbf{E}U(1,\mathbf{n},x,\mathbf{j})^2 \to f(x)\big(v(x)+r^2(x)\big)\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u\,;$

(iv) *For all* $\varepsilon > 0$,

$$\hat{\mathbf{n}}\, h^d \sum_{\mathbf{j}\in\mathcal{T}} \mathbf{E}U(1,\mathbf{n},x,\mathbf{j})^2 \mathbf{1}_{\{|U(1,\mathbf{n},x,\mathbf{j})|>\varepsilon(\hat{\mathbf{n}}\, h^d)^{-1/2}\}} \to 0.$$

Proof of Lemma 4.4 Assertions (i)-(iv) are the strict equivalents of assertions (3.5)-(3.8) of the proof of Lemma 3.2 in Tran [15]. More or less immediate adaptations of the proofs of Tran to our context of regression allow us to prove the lemma. For this reason, we only give the important points of the proofs of assertions (i)-(iv).

(i) The proof is *exactly* the same as the proof of (3.5) in Tran [15].

(ii) One only needs to prove for instance that

$$\hat{\mathbf{n}}\, h^d \mathbf{E}T(\mathbf{n},x,2)^2 \to 0.$$

Enumerate the random variables $\big(U(2,\mathbf{n},x,\mathbf{j}),\mathbf{j}\in\mathcal{T}\big)$ in an arbitrary manner and refer to them as $\hat{U}_1,\ldots,\hat{U}_M$, where $M = \text{Card}\,\mathcal{T} = \prod_{k=1}^N t_k = \hat{\mathbf{n}}(p+$

$q)^{-N}$. Now,

$$\mathbf{E}T(\mathbf{n},x,2)^2 = \sum_{i=1}^{M} \mathrm{Var}\ \hat{U}_i + \sum_{\substack{i,j=1\\ i\neq j}}^{M} \mathrm{Cov}(\hat{U}_i,\hat{U}_j).$$

An easy adaptation of the proof of (3.6) in Tran [15] gives

$$\hat{\mathbf{n}}\, h^d \sum_{i=1}^{M} \mathrm{Var}\ \hat{U}_i \leq c\, M \gamma_{\mathbf{n}}^2\, p^{N-1} q\, (\hat{\mathbf{n}})^{-1} h^{-d\delta} \sum_{i\geq 1} i^{N-1}\chi(i)^{\delta}$$

and

$$\hat{\mathbf{n}}\, h^d \sum_{\substack{i,j=1\\ i\neq j}}^{M} \mathrm{Cov}(\hat{U}_i,\hat{U}_j) \leq c\, \gamma_{\mathbf{n}}^2\, h^{-d\delta} \sum_{i\geq q} i^{N-1}\chi(i)^{\delta}.$$

Consequently,

$$\begin{aligned}\hat{\mathbf{n}}\, h^d \mathbf{E}T(\mathbf{n},x,2)^2 &\leq c\, \gamma_{\mathbf{n}}^2\, q\, p^{-1} h^{-d\delta} \sum_{i\geq 1} i^{N-1}\chi(i)^{\delta} + c\, \gamma_{\mathbf{n}}^2\, h^{-d\delta} \sum_{i\geq q} i^{N-1}\chi(i)^{\delta}\\ &\leq c\, q \big(\hat{\mathbf{n}}\, h^{d(1+2\delta N)} (\log \hat{\mathbf{n}})^{-8N/b}\big)^{-1/(2N)} \sum_{i\geq 1} i^{N-1}\chi(i)^{\delta}\\ &\quad + c\, h^{-d\delta} (\log \hat{\mathbf{n}})^{2/b} \sum_{i\geq q} i^{N-1}\chi(i)^{\delta},\end{aligned}$$

and the last two terms tend to 0 by (iii) and (iv) of Theorem 2.2.

(iii) The proof is a straightforward adaptation of the proof of (ii), using as well Proposition 4.3 (ii). See also the proof of (3.7) in Tran [15].

(iv) We clearly have $|\Delta_{\mathbf{j}}(x)| \leq c\, \gamma_{\mathbf{n}} (\hat{\mathbf{n}}\, h^d)^{-1}$ and hence

$$|U(1,\mathbf{n},x,\mathbf{j})| \leq c\, \gamma_{\mathbf{n}}\, p^N (\hat{\mathbf{n}}\, h^d)^{-1}.$$

Then, by definitions of $\gamma_{\mathbf{n}}$ and p,

$$\sqrt{\hat{\mathbf{n}}\, h^d}\, |U(1,\mathbf{n},x,\mathbf{j})| \leq c\, (\log \hat{\mathbf{n}})^{(1-2N)/b},$$

which goes to 0. Consequently, if $\hat{\mathbf{n}}$ is large enough, we have, for all $\mathbf{j}$,

$$\mathbf{P}\big(|U(1,\mathbf{n},x,\mathbf{j})| > \varepsilon (\hat{\mathbf{n}}\, h^d)^{-1/2}\big) = 0,$$

hence (iv). ■

We are now in position to prove Proposition 4.4. Recall that only the case $\alpha = 1$ and $\beta = 0$ is considered here. Write

$$\sqrt{\hat{\mathbf{n}}\, h^d}\big(\zeta^*_{\mathbf{n}}(x) - \mathbf{E}\zeta^*_{\mathbf{n}}(x)\big) = \sqrt{\hat{\mathbf{n}}\, h^d}\, T(\mathbf{n}, x, 1) + \sqrt{\hat{\mathbf{n}}\, h^d} \sum_{i=2}^{2^N} T(\mathbf{n}, x, i).$$

According to Lemma 4.4 (ii), the rightmost term vanishes in probability. Moreover, the random variables $(U(1, \mathbf{n}, x, \mathbf{j}), \mathbf{j} \in \mathcal{T})$ are asymptotically independent by Lemma 4.4 (i). Thus, the asymptotic normality

$$\sqrt{\hat{\mathbf{n}}\, h^d} T(\mathbf{n}, x, 1) \stackrel{\mathcal{D}}{\to} \mathcal{N}\Big(0, f(x)\big(v(x) + r^2(x)\big) \int_{\mathbb{R}^d} K^2(u)\mathrm{d}u\Big)$$

follows from Lemma 4.4 (iii) and the Lindeberg-Feller condition (iv), by a straightforward adaptation of the proof of the Lindeberg theorem (see for instance Dudley [6]). ■

4.4.3 Proof of Proposition 4.5

We only prove (i), the proof of (ii) being similar. By (**H4**), the function $r\,f$ is Lipschitzian in a neighborhood of x, so that by (4.2), one has for $\hat{\mathbf{n}}$ large enough,

$$\sqrt{\hat{\mathbf{n}}\, h^d} |\mathbf{E}\zeta_{\mathbf{n}}(x) - \zeta(x)| \leq c\, \sqrt{\hat{\mathbf{n}}\, h^{d+2}}\,,$$

which tends to 0 according to condition (i) of Theorem 2.2. Moreover, by (**H4**) and Proposition 4.2, one has

$$\begin{aligned}\sqrt{\hat{\mathbf{n}}\, h^d} |\mathbf{E}\zeta^*_{\mathbf{n}}(x) - \mathbf{E}\zeta_{\mathbf{n}}(x)| &\leq \sqrt{\hat{\mathbf{n}}\, h^d}\, h^{-d}\mathbf{E}\left[|Y| \mathbf{1}_{(|Y|>\gamma_{\mathbf{n}})} K\Big(\frac{x - X}{h}\Big)\right] \\ &\leq c\, \sqrt{\hat{\mathbf{n}}\, h^{-d}}\, (\hat{\mathbf{n}})^{-3/4} \mathbf{E}^{1/2} K\Big(\frac{x - X}{h}\Big)^2 \\ &\leq c\, (\hat{\mathbf{n}})^{-1/4},\end{aligned}$$

hence the proposition. ■

Acknowledgments. The authors greatly thank an Associate Editor and an anonymous referee for a careful reading of the paper.

References

[1] Biau, G. (2002). Optimal asymptotic quadratic errors of density estimators on random fields, *Statistics & Probability Letters*, **Vol. 60**, pp. 297–307.

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

[3] Carbon, M., Hallin, M. and Tran, L.T. (1996). Kernel density estimation for random fields: the L_1 theory, *Journal of Nonparametric Statistics*, **Vol. 6**, pp. 157–170.

[4] Carbon, M., Tran, L.T. and Wu, B. (1997). Kernel density estimation for random fields, *Statistics & Probability Letters*, **Vol. 36**, pp. 115–125.

[5] Cressie, N.A.C. (1993). *Statistics for Spatial Data* (revised edition), Wiley Series in Probability and Mathematical Statistics, New York.

[6] Dudley, R.M. (1989). *Real Analysis and Probability*, Chapman and Hall, New York.

[7] Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*, Springer–Verlag, Berlin.

[8] Hallin, M., Lu, Z. and Tran, L.T. (2001). Density estimation for spatial linear processes, *Bernoulli*, **Vol. 7**, pp. 657–668.

[9] Hallin, M., Lu, Z. and Tran, L.T. (2002). Local linear spatial regression, *ISRO Preprint Series 2002/189*, Université Libre de Bruxelles.

[10] Hallin, M., Lu, Z. and Tran, L.T. (2003). Kernel density estimation for spatial processes: the L_1 theory, *Journal of Multivariate Analysis*, to appear.

[11] Lahiri, S.N., Kaiser, M.S., Cressie, N. and Hsu, N.J. (1999). Prediction of spatial cumulative distribution functions using subsampling (with discussion), *Journal of the American Statistical Association*, **Vol. 94**, pp. 86–110.

[12] Lu, Z. (2000). Spatial nonparametric regression estimation, *Research Report AMSS-2000-061*, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing.

[13] Opsomer, J., Wang, Y. and Yang, Y. (2001). Nonparametric regression with correlated errors, *Statistical Science*, **Vol. 16**, pp. 134–153.

[14] Serfling, R.J. (1968). Contributions to central limit theory for dependent variables, *The Annals of Mathematical Statistics*, **Vol. 39**, pp. 1158–1175.

[15] Tran, L.T. (1990). Kernel density estimation on random fields, *Journal of Multivariate Analysis*, **Vol. 34**, pp. 37–53.

[16] Tran, L.T. and Yakowitz, S. (1993). Nearest neighbor estimators for random fields, *Journal of Multivariate Analysis*, **Vol. 44**, pp. 23–46.

[17] Wackernagel, H. (1995). *Multivariate Geostatistics*, Springer–Verlag, Berlin.