

# Rate of uniform consistency for a class of mode regression on functional stationary ergodic data

Mohamed Chaouch<sup>1</sup> · Naâmane Laïb<sup>2,3</sup> ·  
Djamal Louani<sup>2,4</sup>

Accepted: 27 February 2016 / Published online: 9 March 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** The aim of this paper is to study the asymptotic properties of a class of kernel conditional mode estimates whenever functional stationary ergodic data are considered. To be more precise on the matter, in the ergodic data setting, we consider a random elements  $(X, Z)$  taking values in some semi-metric abstract space  $E \times F$ . For a real function  $\varphi$  defined on the space  $F$  and  $x \in E$ , we consider the conditional mode of the real random variable  $\varphi(Z)$  given the event “ $X = x$ ”. While estimating the conditional mode function, say  $\theta_\varphi(x)$ , using the well-known kernel estimator, we establish the strong consistency with rate of this estimate uniformly over Vapnik–Chervonenkis classes of functions  $\varphi$ . Notice that the ergodic setting offers a more general framework than the usual mixing structure. Two applications to energy data are provided to illustrate some examples of the proposed approach in time series forecasting framework. The first one consists in forecasting the *daily peak* of electricity demand in France (measured in Giga-Watt). Whereas the second one deals with the

---

Chaouch’s and Laïb’s research was supported by the United Arab emirates University Start-up Research Grant No: 31B029.

---

✉ Mohamed Chaouch  
m.chaouch@uaeu.ac.ae  
Naâmane Laïb  
naamane.laib@upmc.fr  
Djamal Louani  
djamal.louani@upmc.fr

<sup>1</sup> Department of Statistics, United Arab Emirates University, Al Ain, UAE

<sup>2</sup> L.S.T.A., Université de Paris 6, Paris, France

<sup>3</sup> EISTI, Avenue du Parc, 95011 Cergy-Pontoise, France

<sup>4</sup> Université de Reims-Champagne-Ardenne, PB 1039, 51687 Reims Cedex 2, France

short-term forecasting of the electrical *energy* (measured in Giga-Watt per Hour) that may be consumed over some time intervals that cover the peak demand.

**Keywords** Conditional mode estimation · Energy data · Entropy · Ergodic processes · Functional data · Martingale difference · Peak load · Strong consistency · Time series forecasting · VC-classes

### 1 Introduction

Let  $(X, Z)$  be a  $E \times F$ -valued random elements, where  $E$  and  $F$  are some semi-metric abstract spaces. Denote by  $d_E$  and  $d_F$  semi-metrics associated to spaces  $E$  and  $F$  respectively. Let  $\mathcal{C}$  be a class of real functions defined upon  $F$ . Obviously, for any  $\varphi \in \mathcal{C}$ ,  $\varphi(Z)$  is a real random variable. Suppose now that we observe a sequence  $(X_i, Z_i)_{i \geq 1}$  of copies of  $(X, Z)$  that we assume to be stationary and ergodic. For any  $x \in E$  and any  $\varphi \in \mathcal{C}$ , let  $g_\varphi(\cdot|x)$  be the conditional density of  $\varphi(Z)$  given  $X = x$ . We assume that  $g_\varphi(\cdot|x)$  is unimodal on some compact  $S_\varphi \subset \mathbb{R}$ . The conditional mode is defined, for any fixed  $x \in E$ , by

$$\Theta_\varphi(x) = \arg \sup_{y \in S_\varphi} g_\varphi(y|x).$$

Note that, if there exists  $\xi > 0$  such that for any  $\varphi \in \mathcal{C}$

$$g_\varphi(\cdot|x) \uparrow \text{ on } (\Theta_\varphi(x) - \xi, \Theta_\varphi(x)) \text{ and } g_\varphi(\cdot|x) \downarrow \text{ on } (\Theta_\varphi(x), \Theta_\varphi(x) + \xi), \tag{1}$$

and if we choose  $S_\varphi = [\Theta_\varphi(x) - \xi, \Theta_\varphi(x) + \xi]$ , then the mode  $\Theta_\varphi(x)$  is uniquely defined for any  $\varphi$ . The kernel estimator, say  $\hat{\Theta}_{\varphi,n}(x)$ , of  $\Theta_\varphi(x)$  may be defined as the value maximizing the kernel estimator  $g_{\varphi,n}(y|x)$  of  $g_\varphi(y|x)$ , that is,

$$g_{\varphi,n}(\hat{\Theta}_{\varphi,n}(x)|x) = \sup_{y \in S_\varphi} g_{\varphi,n}(y|x). \tag{2}$$

Here,

$$g_{\varphi,n}(y|x) = \frac{f_{\varphi,n}(x, y)}{l_n(x)},$$

where

$$f_{\varphi,n}(x, y) = \frac{1}{nh_H \mathbb{E}[\Delta_1(x)]} \sum_{i=1}^n \left[ \Delta_i(x) H \left( \frac{y - \varphi(Z_i)}{h_H} \right) \right],$$

$$l_n(x) = \frac{1}{n \mathbb{E}[\Delta_1(x)]} \sum_{i=1}^n \Delta_i(x)$$

and  $\Delta_i(x) = K \left( \frac{d(x, X_i)}{h_K} \right)$ , with  $K$  and  $H$  are two real valued kernels and  $(h_K, h_H) := (h_{K,n}, h_{H,n})$  a sequence of positive real numbers tending to zero as  $n \rightarrow \infty$ .

The aim of this paper is to establish the uniform consistency, with respect to the function parameter  $\varphi \in \mathcal{C}$ , of the conditional mode estimator  $\hat{\Theta}_{\varphi,n}(x)$  when data are assumed to be sampled from a stationary and ergodic process. More precisely, under suitable conditions upon the entropy of the class  $\mathcal{C}$  and the rate of convergence of the smoothing parameters  $h_K$  and  $h_H$  together with some regularity conditions on the distribution of the random element  $(X, Z)$ , we obtain results of type

$$\sup_{\varphi \in \mathcal{C}} |\hat{\Theta}_{\varphi,n}(x) - \Theta_{\varphi}(x)| = O(\alpha_n), \text{ a.s.}$$

where  $\alpha_n$  is a quantity to be specified later on. Notice that, besides the infinite dimensional character of the data, the ergodic framework avoid the widely used strong mixing condition and its variants to measure the dependency and the very involved probabilistic calculations that it implies [see, for instance, [Masry \(2005\)](#)]. Further motivations to consider ergodic data are discussed in [Laïb \(2005\)](#) and [Laïb and Louani \(2010\)](#) where details defining the ergodic property of processes together with examples of such processes are also given.

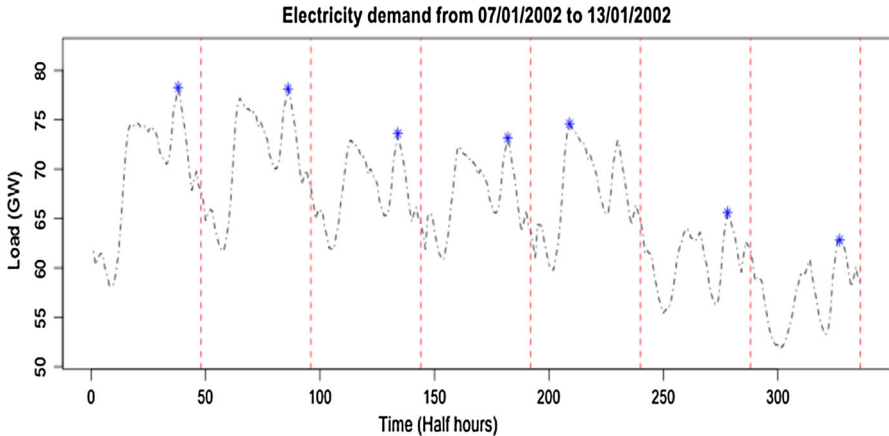
Indexing by a function  $\varphi$  allows to consider simultaneously various situations related to model fitting and time series forecasting. Whenever  $Z := \{Z(t) : t \in T\}$  denotes a process defined on some real set  $T$ , one may consider the following functionals  $\varphi_1(Z) = \sup_{t \in T} Z(t)$  and  $\varphi_2(Z) = \inf_{t \in T} Z(t)$  giving extremes of the process  $Z$  that are of interest in various domains as, for example, the finance, hydraulics and the weather forecasting. For some weight function  $W$  defined on  $T$  and some  $p > 0$ , one may consider the functional  $\varphi_{p,W}$  defined by  $\varphi_{p,W}(Z) = \int_T W(t) Z^p(t) dt$ . Further situation is to consider, for some subset  $A$  of  $T$ , the functional  $Z \rightarrow \varphi_{\rho}(Z) = \inf\{t \in A : Z(t) \geq \rho\}$  for some threshold  $\rho$ . Such a case is very useful in threshold and barrier crossing problems encountered in various domains as finance, physical chemistry and hydraulics. Moreover, indexing by a class of functions  $\mathcal{C}$  is a step towards modelling a functional response random variable. Indeed, the quantity  $Z(\varphi) := \{\varphi(Z) : \varphi \in \mathcal{C}\}$  may be viewed as a functional random variable offering, in this respect, a device for such investigations.

The modelization of the functional variable is becoming more and more popular since the publication of the monograph of [Ramsay and Silverman \(1997\)](#) on functional data analysis. Note however that the first results dealing with nonparametric models (mainly the regression function) were obtained by [Ferraty and Vieu \(2000\)](#). Since then, an increasing number of papers on this topic has been published. One may refer to the monograph by [Ferraty and Vieu \(2006\)](#) for an overview on the subject and the references therein. Extensions to other regression issues as the time series prediction have been carried out in a number of publications, see for instance [Delsol \(2009\)](#). The general framework of ergodic functional data has been considered by [Laïb and Louani \(2010, 2011\)](#) who stated consistencies with rates together with the asymptotic normality of the regression function estimate.

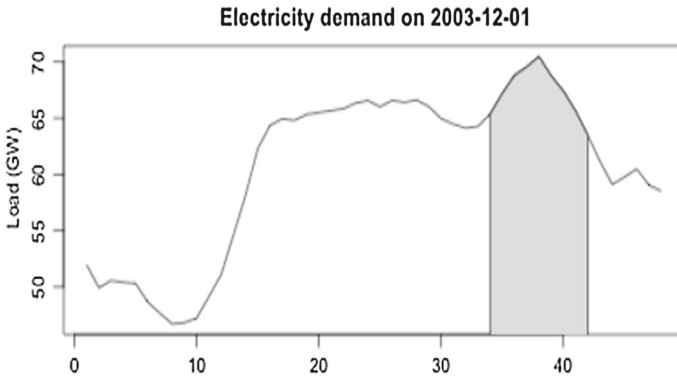
Asymptotic properties of the conditional mode estimator have been investigated in various situations throughout the literature. Ferraty et al. (2006) studied asymptotic properties of kernel-type estimators of some characteristics of the conditional cumulative distribution with particular applications to the conditional mode and conditional quantiles. Ezzahrioui and Ould-Saïd (2008), Ezzahrioui and Ould-Saïd (2010) established the asymptotic normality of the kernel conditional mode estimator in both i.i.d. and strong mixing cases. Dabo-Niang and Laksaci (2007) provided a convergence rate in  $L^p$  norm sense of the kernel conditional mode estimator whenever functional  $\alpha$ -mixing observations are considered. Demongeot et al. (2010) have established the pointwise and uniform almost complete convergences with rates of the local linear estimator of the conditional density. They used their results to deduce some asymptotic properties of the local linear estimator of the conditional mode. Attaoui et al. (2011) have established the pointwise and uniform almost complete convergence, with rates, of the kernel estimate of the conditional density when the observations are linked with a single-index structure. They applied their results to the prediction problem via the conditional mode estimate. Notice also that, considering a scalar response variable  $Y$  with a covariate  $X$  taking values in a semi-metric space, Ferraty et al. (2010) studied, in the i.i.d. case, the nonparametric estimation of some functionals of the conditional distribution including the regression function, the conditional cumulative distribution, the conditional density together with the conditional mode. They established the uniform almost complete convergence, with rates, of kernel estimators of these quantities.

It is well-known that the conditional mode provides an alternative prediction method to the classical approach based on the usual regression function. Since there exist cases where the conditional density is such that the regression function vanishes everywhere, then it makes no sense to use this approach in problems involving prediction. An example in a finite-dimensional space is given in Ould Saïd (1997) to illustrate this situation. Moreover, a simulation study in infinite-dimensional spaces carried out by Ferraty et al. (2006), shows that the conditional mode approach gives slightly better results than the usual regression approach.

In this paper, two applications to energy data are provided to illustrate some examples of the proposed approach in time series forecasting framework. The first real case consists in forecasting the *daily peak* of electricity demand in France (measured in Giga-Watt). Let us denote by  $(Z_i(t))_{t \in [0, T]}$  the curve of electricity demand (called also *load curve*) measured over an interval  $[0, T]$ . If we have hourly (reps. half-hour) measures then  $T = 24$  (resp.  $T = 48$ ). The peak demand observed for any day  $i$  is defined as  $\mathcal{P}_i = \sup_{t \in [0, T]} Z_i(t)$ . In such case  $\varphi(\cdot)$  is fixed to be the supremum function, over  $[0, T]$ , of the function  $Z(t)$ . Accurate prediction of daily peak load demand is very important for decision in the energy sector. In fact, short-term load forecasts enable effective load shifting between transmission substations, scheduling of startup times of peak stations, load flow analysis and power system security studies. Figure 1 provides a sample of seven daily load curves (from 07/01/2002 to 13/01/2002). Vertical dotted lines separate days and the star points correspond to the peak demand for each day.



**Fig. 1** Half hour electricity consumption in France from 07/01/2002 to 13/01/2002 (7 days). The vertical dotted lines separate days and the star points correspond to the peak demand for each day (in Giga-Watt)



**Fig. 2** Half hour daily load curve in solid line and the grey surface corresponds to the electrical energy (in Giga-Watt per Hour) consumed over an interval of 3 h around the peak

It is well-known that, in addition to peak demand, some other characteristics of the load curve may be of interest from an operational point of view. In fact the prediction of the *electrical energy* (measured in Giga-Watt per Hour) consumed over an interval of 3 h around the peak demand may help in the determination of consistent and reliable supply schedules during peak period. Therefore, the second application in this paper deals with the short-term forecasting of the electrical energy that may be consumed between 6 p.m. and 9 p.m. in winter and between 12 a.m. and 3 p.m. in summer. Those time intervals cover the peak demand which happens around 7pm in winter and 2pm in summer. Formally, if we consider  $Z_i(t)$  the load curve of some day  $i$ , then the electrical energy consumed between two instants  $t_1$  and  $t_2$  is defined as  $\mathcal{E}_i = \int_{t_1}^{t_2} Z_i(t) dt$ . In this case,  $\varphi(\cdot)$  is the integral function. An example of half hour daily load curve is plotted in Fig. 2. Solid line is the daily load curve and the grey surface corresponds to the electrical energy consumed over an interval of 3 h around the peak.

## 2 Results

In order to state our results, we introduce some notations. Let  $\mathcal{F}_i$  be the  $\sigma$ -field generated by  $((X_1, Z_1), \dots, (X_i, Z_i))$  and  $\mathcal{G}_i$  the one generated by  $((X_1, Z_1), \dots, (X_i, Z_i), X_{i+1})$ . Let  $B(x, u)$  be a ball centered at  $x \in E$  with radius  $u$ . Let  $D_i(x) := d(x, X_i)$  so that  $D_i(x)$  is a nonnegative real-valued random variable. Working on the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ , let  $F_x(u) = \mathbb{P}(D_i(x) \leq u) := \mathbb{P}(X_i \in B(x, u))$  and  $F_x^{\mathcal{F}_{i-1}}(u) = \mathbb{P}(D_i(x) \leq u | \mathcal{F}_{i-1}) = \mathbb{P}(X_i \in B(x, u) | \mathcal{F}_{i-1})$  be the distribution function and the conditional distribution function, given the  $\sigma$ -field  $\mathcal{F}_{i-1}$ , of  $(D_i(x))_{i \geq 1}$  respectively. Denote by  $o_{a.s.}(u)$  a real random function  $l$  such that  $l(u)/u$  converges to zero almost surely as  $u \rightarrow 0$ . Similarly, define  $O_{a.s.}(u)$  as a real random function  $l$  such that  $l(u)/u$  is almost surely bounded.

Our results are stated under some assumptions we gather hereafter for easy reference.

**A1** For  $x \in E$ , there exist a sequence of nonnegative random functional  $(f_{i,1})_{i \geq 1}$  almost surely bounded by a sequence of deterministic quantities  $(b_i(x))_{i \geq 1}$  accordingly, a sequence of random functions  $(\psi_{i,x})_{i \geq 1}$ , a deterministic nonnegative bounded functional  $f_1$  and a nonnegative nondecreasing real function  $\phi$  tending to zero as its argument goes to zero, such that

(i)  $F_x(u) = \phi(u)f_1(x) + o(\phi(u))$ , as  $u \rightarrow 0$ ,

(ii) For any  $i \in \mathbb{N}$ ,  $F_x^{\mathcal{F}_{i-1}}(u) = \phi(u)f_{i,1}(x) + \psi_{i,x}(u)$  with  $\psi_{i,x}(u) = o_{a.s.}(\phi(u))$  as  $u \rightarrow 0$ ,  $\frac{\psi_{i,x}(u)}{\phi(u)}$  is almost surely bounded and  $\frac{1}{n} \sum_{i=1}^n \psi_{i,x}(u) = o_{a.s.}(\phi(u))$  as  $n \rightarrow \infty, u \rightarrow 0$ .

(iii)  $n^{-1} \sum_{i=1}^n f_{i,1}(x) \rightarrow f_1(x)$  almost surely as  $n \rightarrow \infty$ .

(iv) There exists a nondecreasing bounded function  $\tau_0$  such that, uniformly in  $u \in [0, 1]$ ,

$$\frac{\phi(h_K u)}{\phi(h_K)} = \tau_0(u) + o(1), \text{ as } h \downarrow 0 \text{ and, for } 1 \leq j \leq 2, \int_0^1 (K^j(u))' \tau_0(u) du < \infty.$$

(v)  $n^{-1} \sum_{i=1}^n b_i(x) \rightarrow D(x) < \infty$  as  $n \rightarrow \infty$ .

**A2**  $K$  is a nonnegative bounded kernel of class  $\mathcal{C}^1$  over its support  $[0, 1]$ , with  $K(1) > 0$  and the derivative  $K'$  is such that  $K'(t) < 0$ , for any  $t \in [0, 1]$ .

**A3** (i) For any  $\epsilon > 0$ , there exists  $\eta > 0$  such that for any  $(\varphi_1, \varphi_2) \in \mathcal{C} \times \mathcal{C}$ ,  $d_{\mathcal{C}}(\varphi_1, \varphi_2) < \eta$  implies that  $|\Theta_{\varphi_1}(x) - \Theta_{\varphi_2}(x)| < \epsilon$ .

(ii) Uniformly in  $\varphi \in \mathcal{C}$ ,  $g_{\varphi}(\cdot|x)$  is uniformly continuous on  $S_{\varphi}$ .

(iii)  $g_{\varphi}(\cdot|x)$  is differentiable up to order 2 and  $\lim_{n \rightarrow \infty} \sup_{\varphi \in \mathcal{C}} |g_{\varphi}^{(2)}(\hat{\Theta}_{\varphi,n}(x)|x)| :=$

$$\Phi(x) \neq 0.$$

(iv) For any  $x \in E$ , there exist  $V(x)$  a neighborhood of  $x$ , some constants  $C_x > 0$ ,  $\beta > 0$  and  $\nu \in (0, 1]$ , independent of  $\varphi$ , such that for any  $\varphi \in \mathcal{C}$ , we have  $\forall (y_1, y_2) \in S_{\varphi} \times S_{\varphi}, \forall (x_1, x_2) \in V(x) \times V(x)$ ,

$$|g_\varphi^{(j)}(y_1|x_1) - g_\varphi^{(j)}(y_2|x_2)| \leq C_x(|y_1 - y_2|^v + d(x_1, x_2)^\beta), \quad j = 0, 2.$$

**A4** Consider the space of functions  $\mathcal{D} = \{\psi = t - \varphi \text{ where } t \in \mathbb{R} \text{ and } \varphi \in \mathcal{C}\}$  on which we define the distance  $\rho$  given, for any  $(t_1 - \varphi_1, t_2 - \varphi_2) \in \mathcal{D}^2$  by,  $\rho(t_1 - \varphi_1, t_2 - \varphi_2) = |t_2 - t_1| + d_{\mathcal{C}}(\varphi_1, \varphi_2)$ . The kernel  $H$  is such that

- (i)  $\int_{\mathbb{R}} |t|^v H(t)dt < \infty$  and  $\int_{\mathbb{R}} t H(t)dt = 0$ ,
- (ii) For all  $(t_1 - \varphi_1, t_2 - \varphi_2) \in \mathcal{D}^2$  and  $\forall z \in F$ ,

$$|H(t_1 - \varphi_1(z)) - H(t_2 - \varphi_2(z))| \leq C_H \rho(t_1 - \varphi_1, t_2 - \varphi_2),$$

where  $C_H$  is a positive constant.

**A5** For  $j = 0, 1, 2$  and any  $\varphi \in \mathcal{C}$ ,

$$\mathbb{E} \left[ H^{(j)} \left( \frac{y - \varphi(Z_i)}{h_H} \right) \mid \mathcal{G}_{i-1} \right] = \mathbb{E} \left[ H^{(j)} \left( \frac{y - \varphi(Z_i)}{h_H} \right) \mid X_i \right].$$

*Comments on the hypotheses* As to discuss the conditions **A1**, it is worth noticing that the fundamental hypothesis **A1(ii)** involves the functional nature of the data together with their dependency. As usually in such a framework, small balls techniques are used to handle the probabilities on infinite dimension spaces where the Lebesgue measure does not exist. Several examples of processes fulfilling this condition are given in [Laïb and Louani \(2011\)](#). Note that the hypothesis **A1(i)** stands as a particular case of **A1(ii)** while conditioning by the trivial  $\sigma$ -field. A number of processes satisfying this condition are given through out the literature, see, for instance, [Ferraty and Vieu \(2006\)](#). Conditions **A1(iii)** and **A1(v)** are set basically to meet the ergodic Theorem which may be expressed as the classical law of large numbers. Conditions **A2** and **A4** impose some regularity conditions upon the kernels used in our estimates. When indexing by a class of functions  $\mathcal{C}$ , it is natural to consider regularity conditions as the continuity of the mode with respect to the index function  $\varphi$  assumed in **A3(i)**. Defined as an argmax and, furthermore, indexed by the class  $\mathcal{C}$ , the conditional mode is sensitive to fluctuations. The differentiability of the conditional density  $g_\varphi$  with some kind of smoothness of its derivatives is needed to reach the rates of the convergence obtained in our results. All these conditions are summarised in the assumption **A3**. Hypothesis **A5** is of Markov’s nature.

*Remark 1* In order to check the condition (A4)(ii), let  $T$  be an index set and  $d$  a distance over  $T$ . Suppose that  $\mathcal{C} = \{\varphi_u : u \in T\}$  is a class of functions defined on  $F$ , that are Lipschitz with respect to the index parameter  $u$  in the sense that for any  $z \in F$

$$|\varphi_u(z) - \varphi_v(z)| \leq d(u, v)\kappa(z),$$

where  $\kappa$  is a function defined on  $F$  such that  $\int_F \kappa^2(z)d\mathbb{P}(z) < \infty$ . Let  $d_{\mathcal{C}}$  be the  $L_2$ -distance defined on  $\mathcal{C}$  by

$$d_{\mathcal{C}}(\varphi_u, \varphi_v) := \left[ \int_F (\varphi_u(z) - \varphi_v(z))^2 d\mathbb{P}(z) \right]^{1/2}.$$

Then, taking  $d(\cdot, \cdot)$  as the absolute distance, we have

$$d_{\mathcal{C}}(\varphi_u, \varphi_v) \leq d(u, v) \left[ \int_F \kappa^2(z) d\mathbb{P}(z) \right]^{1/2} := c_0|u - v|.$$

Therefore, taking  $H$  the Epanechnikov kernel given by  $H(u) = \frac{3}{4}(1 - u^2)\mathbb{1}_{[-1,1]}(u)$  and using the condition A4(ii), we obtain

$$|H(t_1 - \varphi_u(z)) - H(t_2 - \varphi_v(z))| \leq \frac{3}{2}(|t_1 - t_2| + c_0|u - v|).$$

Before establishing the uniform convergence with rate, with respect to the class of functions  $\mathcal{C}$ , of the conditional mode estimator, we introduce the following notation. For any  $\epsilon > 0$ , set

$$\mathcal{N}(\epsilon, \mathcal{C}, d_{\mathcal{C}}) = \min\{n : \text{there exist } c_1 \cdots, c_n \text{ in } \mathcal{C} \text{ such that } \forall \varphi \in \mathcal{C} \\ \text{there exists } 1 \leq k \leq n \text{ such that } d_{\mathcal{C}}(\varphi, c_k) < \epsilon\}.$$

This number measures how full is the class  $\mathcal{C}$ . Obviously, conditions upon the number  $\mathcal{N}(\epsilon, \mathcal{C}, d_{\mathcal{C}})$  have to be set to state uniform over the class  $\mathcal{C}$  results.

The following proposition establishes the uniform asymptotic behavior (with rate) of the conditional density estimator  $g(y|x)$  with respect to  $y$  and the function  $\varphi \in \mathcal{C}$ . This proposition which is of interest by itself may be used, as an intermediate result, to state uniform results over the class  $\mathcal{C}$ .

**Proposition 1** *Assume that the conditions A1–A5 hold true and that*

$$\lim_{n \rightarrow \infty} \frac{\log(n)}{n\phi(h_K)} = 0. \tag{3}$$

Furthermore, for a sequence of positive real numbers  $\lambda_n$  tending to zero, as  $n \rightarrow \infty$ , and for  $\eta := \eta_n$ , suppose that

$$(i) \lim_{n \rightarrow \infty} \frac{\log \mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}})}{\eta \lambda_n^2 n h_H \phi(h_K)} = 0 \text{ and } (ii) \sum_{n \geq 1} \exp\{-\lambda_n^2 O(nh_H \phi(h_K))\} < \infty, \tag{4}$$

we have

$$\sup_{\varphi \in \mathcal{C}} \sup_{y \in \mathcal{S}_{\varphi}} |g_{\varphi,n}(y|x) - g_{\varphi}(y|x)| = O_{a.s.}(h_K^{\beta} + h_H^{\nu}) + O_{a.s.}(\eta h_H^{-2}) + O_{a.s.}(\lambda_n) \\ + O_{a.s.}\left(\left(\frac{\log n}{n\phi(h_K)}\right)^{1/2}\right).$$

*Comment* In the statement (7), the deviation between  $g_{\varphi,n}(\cdot|x)$  and  $g_{\varphi}(\cdot|x)$  is decomposed as to introduce the conditional bias  $B_{\varphi,n}(\cdot, \cdot)$  and the pseudo-variances  $R_{\varphi,n}(\cdot, \cdot)$



and  $\mathcal{Q}_{\varphi,n}(\cdot, \cdot)$ . The first element on the right hand side of the result of Proposition 1 corresponds to the rate of convergence of  $B_{\varphi,n}(\cdot, \cdot)$  while the three following ones give the convergence rate of  $\mathcal{Q}_{\varphi,n}(\cdot, \cdot)$ . Notice that the element  $R_{\varphi,n}(\cdot, \cdot)$  is negligible.

Our principal result considers the pointwise in  $x \in \mathcal{E}$  and uniform over the class  $\mathcal{C}$  convergence of the kernel estimate  $\hat{\Theta}_{\varphi,n}(x)$  of the conditional mode  $\Theta_{\varphi}(x)$ .

**Theorem 1** *Under the same conditions of Proposition 1, we have*

$$\sup_{\varphi \in \mathcal{C}} |\hat{\Theta}_{\varphi,n}(x) - \Theta_{\varphi}(x)| = O_{a.s.}(h_K^{\beta/2} + h_H^{v/2}) + O_{a.s.}\left(\lambda_n^{\frac{1}{2}}\right) + O_{a.s.}\left(\eta^{1/2}h_H^{-1}\right) + O_{a.s.}\left(\left(\frac{\log n}{n\phi(h_K)}\right)^{1/4}\right).$$

*Remark 2* Replacing the condition (4)(i) by  $\lim_{n \rightarrow \infty} \frac{\log \mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}})}{\eta \lambda_n^2 n h_H \phi(h_K)} = \delta$ , for some  $\delta > 0$  small enough, with  $\lambda_n = O\left(\left(\frac{\log n}{nh_H \phi(h_K)}\right)^{\frac{1}{2}}\right)$ , the condition (4)(ii) is clearly satisfied inducing the uniform consistency, with respect to  $\varphi \in \mathcal{C}$ , of  $\hat{\Theta}_{\varphi,n}(x)$  with the rate

$$O_{a.s.}\left(h_K^{\beta/2} + h_H^{v/2}\right) + O_{a.s.}\left(\eta^{1/2}h_H^{-1}\right) + O_{a.s.}\left(\left(\frac{\log n}{nh_H \phi(h_K)}\right)^{\frac{1}{4}}\right)$$

*Remark 3* Note, whenever  $\eta = O(h_H^{2+\nu})$  and  $\lambda_n^{-1} = O((nh_H^{5+2\nu}\phi(h_K))^{\frac{1}{2}})$ , that the condition (4)(i) takes the form

$$\lim_{\eta \rightarrow 0} \eta \log \mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}}) = 0. \tag{5}$$

The condition (5) is very usual in defining Vapnik–Chervonenkis classes. Examples of classes fulfilling the condition (5) are given throughout the literature, see, for instance, Laïb and Louani (2011) and Vaart and Wellner (1996).

The main application of Theorem 1 is devoted to prediction of time series when considering the conditional mode estimates.

For  $n \in \mathbb{N}^*$ , let  $Z_i(t)$  and  $X_i(t)$ ,  $i = 1, \dots, n$ , be two functional random variables with  $t \in [0, T)$ . For each curve  $X_i(t)$  (the covariate), we have a real response  $Y_{\varphi,i} = \varphi(Z_i(t))$ , a transformation of some functional variable  $Z_i(t)$ . Given a new curve  $X_{n+1} = x_{\text{new}}$ , our purpose is to predict the corresponding response  $y_{\varphi,\text{new}} := \Theta_{\varphi}(x_{\text{new}})$  using as predictor the conditional mode, say  $\hat{y}_{\varphi,\text{new}} := \hat{\Theta}_{\varphi,n}(x_{\text{new}})$ . The following Corollary based on Theorem 1 gives the asymptotic behavior with rate of the empirical error prediction.

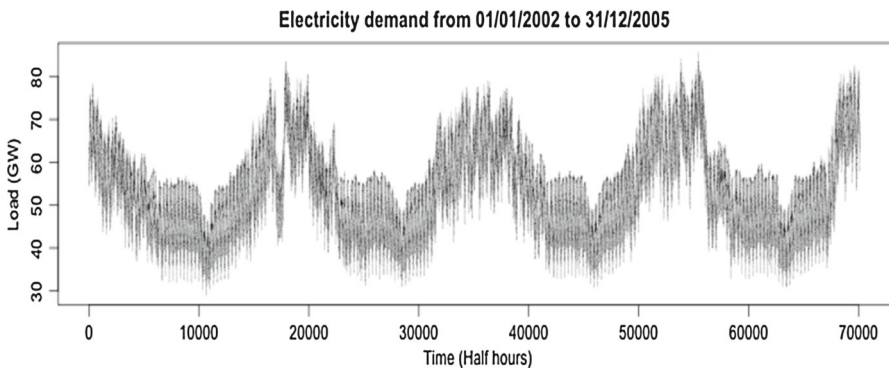
**Corollary 1** Assume that conditions of Theorem 1 hold. Then we have

$$\left| \widehat{y}_{\varphi, new} - y_{\varphi, new} \right| = O_{a.s.} \left( h_K^{\beta/2} + h_H^{v/2} \right) + O_{a.s.} \left( \lambda_n^{\frac{1}{2}} \right) + O_{a.s.} \left( \left( \frac{\log n}{n\phi(h_K)} \right)^{1/4} \right).$$

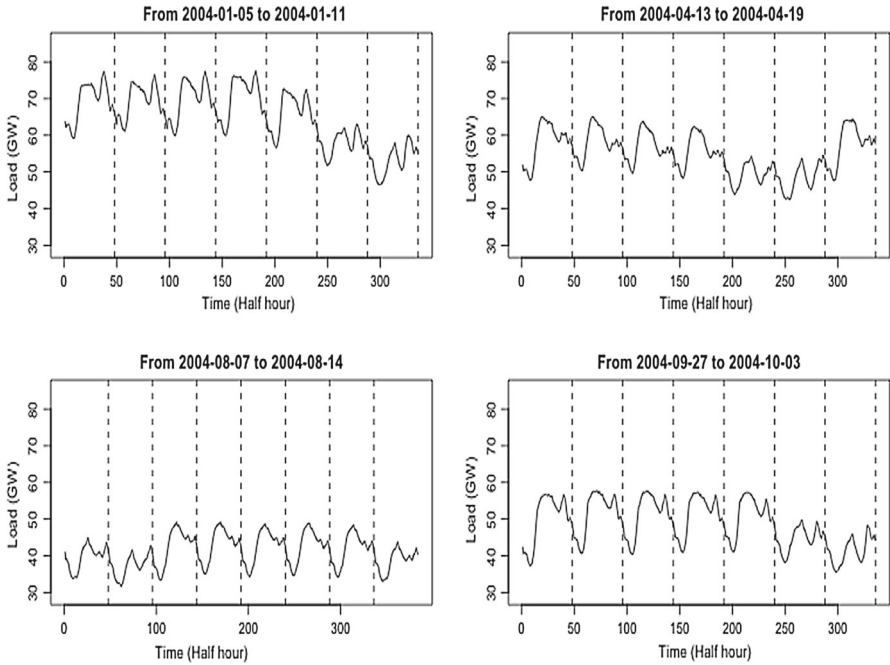
*Proof* The proof of Corollary 1 is a direct consequence of Theorem 1.

### 3 Application to real data

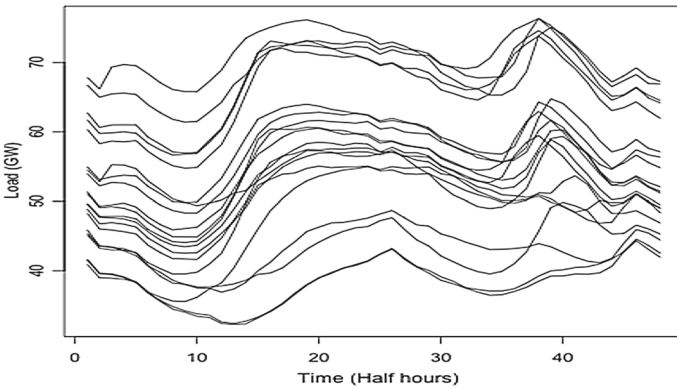
The data-set analyzed in this paper contains half hourly observations of a stochastic process  $\xi(t)$ ,  $t \in \mathbb{R}^+$ . Here  $\xi(t)$  represents the electricity demand at time  $t$  in France. This process has been observed at each half hour from 01 January 2002 to 31 December 2005 (which corresponds to a total of 1461 days). Figure 3 shows the evolution of the process  $\xi(t)$  over time. One can easily see a high seasonality since the variation of the electricity consumption is due to the climatic conditions in France. In fact, the winter and autumn are rather cold, whereas the climate in summer and spring is relatively warm. This remark is confirmed by Fig. 4 which displays the half hourly electricity consumption in France in four selected weeks. We can clearly mark out the intra-daily periodical pattern and can note also the difference in terms of level of consumption from a one season to another. The repetitiveness of the daily shape is due some inertia in the demand that reflects the aggregated behavior of consumers. The evolution of the energy data described by the process  $\xi(t)$ . In order now to construct our functional data  $Z(t)$  and to get its transformation  $\varphi(Z(t))$ , we proceed by slicing the original process  $\xi(t)$  into segments of similar length. Since our target is a day-ahead short term forecasting, we divide the observed original time series ( $\xi(t)$ ) of half hourly electricity consumption into  $n = 1461$  segments ( $Z(t)$ ) of length 48 which correspond to the *functional observations*. Each segment coincide with a specific daily load curve. Formally, let  $[0, T]$  be the time interval on which the process  $\xi(t)$  is observed. We divide this interval into subintervals of length 48, say  $[\ell \times 48, (\ell + 1) \times 48]$ ,  $\ell =$



**Fig. 3** Half hour electricity consumption in France from 07/01/2002 to 31/12/2005 generated according to the process  $\xi(t)$



**Fig. 4** Half hour electricity demand in four selected weeks (the panel contains 1 week data from January, April, August and October 2004)



**Fig. 5** A sample of 20 daily load curves randomly chosen generated according to the process  $Z(t)$

$0, 1, \dots, n - 1$ , with  $n = T/48 = 1461$ . Denoting by  $Z_i(t)$  the functional-valued discrete-time stochastic process defined by

$$Z_i(t) = \xi(t + (i - 1) \times 48); \quad i = 1, \dots, n, \quad \forall t \in [0, 48). \tag{6}$$

Figure 5 shows a randomly chosen sample of 20 realizations of the functional random variable  $(Z(t))_{t \in [0, 48)}$  which corresponds to a daily load curves.

Once the original time series is transformed into functional data-type, as explained in the introduction, we can start to deal with the short term forecasting of the daily peak demand and the electrical energy consumed over an interval using the conditional mode as a predictor.

### 3.1 Short-term daily peak load forecasting

Let us now consider the observed daily peak of the electricity consumption defined, for any day  $i = 1, \dots, n$ , as

$$\mathcal{P}_i = \sup_{t \in [0, 48)} Z_i(t).$$

The goal of this subsection is to forecast the peak  $\mathcal{P}_i$  on the basis of the load curve of the previous day,  $Z_{i-1}(t)$ . Forecasting peak load demand is one of the most relevant issues in electricity companies. In fact, the electricity market is more and more open to competition and companies take care on the quality of their services in order to increase the number of their customers. On the other hand, because of the electrification of appliances (e.g. electric heating, air conditioning, ...) and mobility applications (e.g. electric vehicle, ...), the peak demand is increasing which can lead to a serious issue in the electric network. It is important, therefore, to produce very accurate short-term peak demand forecasts for the day-to-day operation, scheduling and load-shedding plans of power utilities. Forecasting peak load took a lot of interest in the statistical literature. For instance [Goia et al. \(2010\)](#) used a functional linear regression model and [Sigauke and Chikobvu \(2010\)](#) a multivariate adaptive regression splines model. These methods are based on the regression function as a predictor. We propose here the mode regression as an alternative.

In this paper, we compare two predictors based on different choices of the covariable  $X(t)$ . Since the peak electricity demand is highly correlated to the electricity consumption of the previous day and also to the temperature measures, we have then two possibilities to choose the covariable  $X(t)$ :

- (a)  $X_i(t) = Z_{i-1}(t)$ , the curve of electricity consumption of the previous day (Prev. Day) and
- (b)  $X_i(t) = \widehat{T}_i(t)$ , the predicted half hourly temperature curve of the target day (Pred. Temp.).

To evaluate the proposed approach, we split the sample of  $n = 1461$  days into:

- learning sample, say  $\mathcal{L} = \{(Z_{i-1}(t), \mathcal{P}_i)\}_{i=2, \dots, 1096}$ , containing the first 1096 days (corresponding to the period from 01/01/2002 to 31/12/2004) and,
- test sample, say  $\mathcal{T} = \{(Z_{i-1}(t), \mathcal{P}_i)\}_{i=1, \dots, 365}$ , with the last 365 days (corresponding to the period between 01/01/2005 and 31/12/2005).

*Remark 4* When the functional covariate is fixed to be the predicted temperature curve, say  $\widehat{T}_i(t)$ , the notations for the learning and test sample can be changed as follow:  $\mathcal{L} = \{(\widehat{T}_i(t), \mathcal{P}_i)\}_{i=1, \dots, 1096}$  and  $\mathcal{T} = \{(\widehat{T}_i(t), \mathcal{P}_i)\}_{i=1, \dots, 365}$ .

The learning sample is used to build the proposed estimator given by (2) and to find the “optimal” smoothing parameter. To estimate the conditional mode, some tuning parameters should be fixed. For both the covariate and the response variable, the quadratic kernel function defined by  $K(u) = H(u) := 1.5(1 - u^2)\mathbb{1}_{[0,1]}$  is considered. It is proved in the nonparametric estimation literature that the choice of the kernel does not significantly impact the accuracy of estimation.

Another important tuning parameter, that ensures a good behavior of the functional nonparametric estimation, is the semi-metric  $d(\cdot, \cdot)$ . Several possible choices of semi-metric have been discussed in Ferraty and Vieu (2006), p. 28. Usually, the choice of the semi-metric is motivated by the shape of the curves. Here, it is clear that the load curves of the previous day as well as the temperature curves are smooth. Consequently, the  $L_2$ -distance between the second derivative of the curves seems to be the appropriate choice of the semi-metric  $d(\cdot, \cdot)$ .

In contrast to the kernel, an optimal choice of the smoothing parameters  $h_K$  and  $h_H$  is crucial. Here, we adopt the *local cross-validation* method on the  $\kappa$ -nearest neighbors introduced in Ferraty and Vieu (2006), p. 116. Explicitly, for each curve  $X_i$  in the test sample,

$$\widehat{\Theta}_n(h_K, h_H, X_i) = \arg \min_y g_{n, h_K^{i,*}, h_H^{i,*}}(y | X_i^*),$$

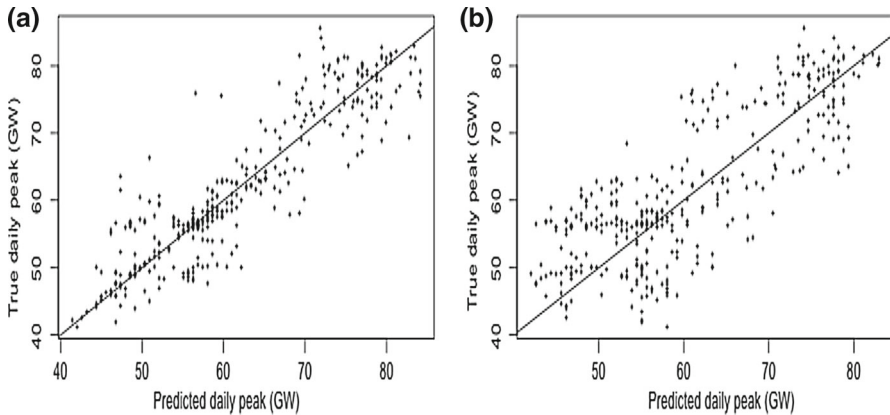
where  $X_i^* = \arg \min_{X_j \in \text{learning sample}} d(X_i, X_j)$  and

$$(h_K^{i,*}, h_H^{i,*}) = \arg \min_{(h_K, h_H) \in H_n(X_i^*, y)} |Y_i^* - \widehat{\Theta}_n(h_K, h_H, X_i^*)|,$$

where  $H_n(x, y)$  is the set of pairs  $(h_K(x), h_H(y))$  such that, for  $h_K(x)$  (respectively, for  $h_H(y)$ ) the ball centred at  $x$  (respectively, the interval centred at  $y$ ) with radius  $h_K(x)$  (respectively, with radius  $h_H(y)$ ) contains exactly  $\kappa$  neighbors of  $x$  (respectively, of  $y$ ). Here, the Ferraty and Vieu’s R-routine called *funopare.mode.lcv*<sup>1</sup> is used to compute the conditional mode.

The test sample will be used to compare our forecasts to the observed daily peak electricity demand for the year 2005. Figure 6a (resp. (b)) displays the observed and the predicted values of the daily peak electricity demand using as covariable the load curve of the previous day (resp. the predicted temperature curve). Since cross-points,  $(\widehat{P}_i, P_i)_{i=1, \dots, 365}$ , represented in Fig. 6a are more concentrated on the diagonal line than those in Fig. 6b, one can deduce that the first approach provides better results than the second one. Moreover, Table 1 provides a numerical summary of the RAE obtained by using as covariate the predicted temperature curve or the last observed daily load curve. One can observe that monthly errors obtained by the second approach are usually less than those given by the first one. Therefore, one can conclude that the peak electricity demand might be better modeled by the previous daily load curve rather than by the predicted temperature curve.

<sup>1</sup> Available on the website: “[www.lsp.ups-tlse.fr/staph/npfda](http://www.lsp.ups-tlse.fr/staph/npfda)”.

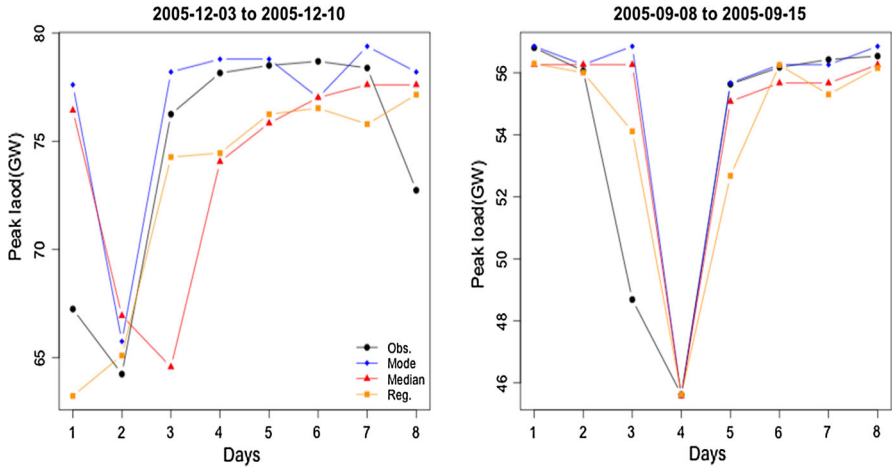


**Fig. 6** Results for 1 year day-ahead forecasting of the daily peak electricity demand using as predictor the mode regression and as functional covariate: **a** the curve of electricity consumption of the previous day (*Prev. Day*), **b** the predicted half hourly temperature curve of the target day (*Pred. Temp.*)

**Table 1** Distribution (by month) of the RAE of the peak load obtained by using as predictor the conditional mode and as covariate the previous day *Prev. Day* and predicted temperature *Pred. Temp.*

	<i>Prev. Day</i> (%)				<i>Pred. Temp.</i> (%)			
	MAPE <sub>m</sub>	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	MAPE <sub>m</sub>	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$
Jan.	4.2	1.4	3.1	5.2	9.5	5.4	9.4	14.2
Feb.	5.4	1.9	4.0	8.3	6.9	1.3	6.3	10.2
Mar.	6.1	2.0	3.7	8.0	9.0	3.5	6.5	13.3
Apr.	6.7	2.6	4.9	9.4	11.9	6.0	11.4	18.2
May	4.7	0.5	1.5	9.5	9.9	3.2	8.0	14.1
Jun.	2.4	0.7	1.2	2.2	9.3	3.5	7.7	15.5
Jul.	3.1	0.5	1.5	3.2	8.3	2.2	7.4	11.2
Aug.	3.0	0.6	1.3	3.6	11.1	4.6	8.0	14.2
Sep.	2.5	0.3	0.8	1.7	10.9	4.0	9.5	18.7
Oct.	4.0	0.9	2.3	4.4	11.0	4.5	9.9	17.5
Nov.	5.4	2.0	3.9	7.1	8.9	3.7	6.8	14.1
Dec.	5.9	2.2	5.5	8.2	7.1	3.2	5.8	9.3

Following the above analysis, the last observed daily load curve will be considered as the suitable covariate to forecast the peak demand in the remainder of this section. Our goal, now, consists in comparing the conditional mode predictor to the conditional median and the regression function (conditional mean) [see [Ferraty and Vieu \(2006\)](#) for more details about the properties of these last two predictors]. Let us mention here that for the nonparametric estimation of the regression operator (respectively, the conditional median), the same semi-metric and the optimal bandwidth,  $h_K$  (respectively,  $h_K$  and  $h_H$ ) are chosen using the same arguments as for the conditional mode estimation (see details above). Moreover, to estimate the regression operator,



**Fig. 7** Examples of daily peak load forecast using conditional mode, conditional median and regression function for eight consecutive days

only the quadratic kernel function  $K(\cdot)$  is used. For the conditional median model, notice that the quadratic kernel is used to smooth the covariate variable  $X$  while the distribution function  $\int_{-\infty}^x \frac{3}{4}(1-t^2)\mathbb{1}_{[-1,1]}(t)dt$  is considered to smooth the response  $Y$ . The computation devoted to the regression operator (respectively, the conditional median) is performed through the R-routine *funopare.knn.lcv*<sup>2</sup> (respectively, *funopare.quantile.lcv*).

For a deeper analysis and evaluation of the accuracy of the proposed approach, we use the Relative Absolute Errors (RAE) and the *monthly* Mean Absolute Prediction Error ( $MAPE_m$ ) as validation criteria. They are defined in the test sample, for any day  $i = 1, \dots, 365$ , respectively by

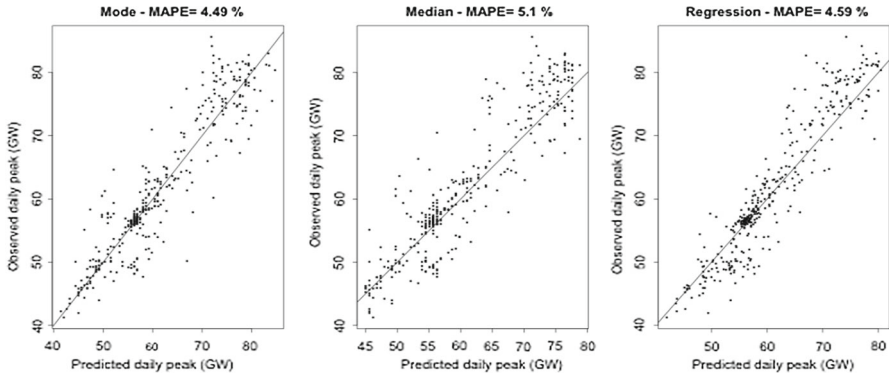
$$RAE_i = \frac{|P_i - \hat{P}_i|}{P_i} \quad \text{and} \quad MAPE_m = \frac{1}{N_m} \sum_{i=1}^{N_m} RAE_i,$$

where  $N_m$  is a number of days for a given month  $m \in \{1, \dots, 12\}$  and  $\hat{P}_i$  is the predicted value of the daily peak obtained by the conditional mode, conditional median or regression function.

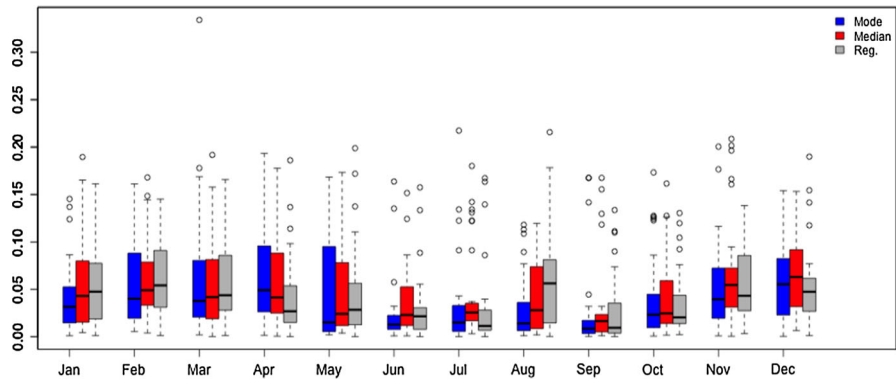
Figure 8 shows examples of peak load forecasts for eight consecutive days. One can see that conditional mode provides more accurate predictions than the two other methods. In Fig. 7, the 365 forecasted daily peak load are plotted against the observed ones. Clearly, one can observe that conditional mode performs well the forecasts while conditional median and regression function under-predict peaks in the cold season and over-predict it in hot season.

Figure 9 provides the distribution of the daily RAE for each month in 2005, obtained using the three prediction methods. One can observe that the conditional mode-based

<sup>2</sup> Available on the website: “[www.lsp.ups-tlse.fr/staph/npfa](http://www.lsp.ups-tlse.fr/staph/npfa)”.



**Fig. 8** Observed daily peak load versus the predicted one obtained by the three forecast methods



**Fig. 9** Distribution (by month) of the daily RAE of the peak load

approach is much more efficient in winter as well as in summer than the other methods. Accurate forecasts in winter are of particular interest since the electricity demand might exceed the supply capacity in this period. Therefore, an efficient energy management in the electrical grid is highly required.

A numerical summary of Fig. 9 is detailed in Table 2 where the monthly  $MAPE_m$ , the first quartile  $Q_{0.25}$ , the median  $Q_{0.5}$  and the third quartile  $Q_{0.75}$ , of the RAE, are provided for the three used methods. One can see that the conditional mode approach performs better the forecasts almost over all the year.

### 3.2 Electrical energy consumption forecasting for battery storage management

The electrical grid in most of the developed countries is expected to be under a large amount of strain in the future due to changes in demand behavior, the electrification of transport, heating and an increased penetration of distributed generation. The current electrical grids infrastructure may not be able to endure these changes and storage of energy produced by solar and wind power generations. One of the most used approaches to solve this technical issue consists in the storage in batteries of the

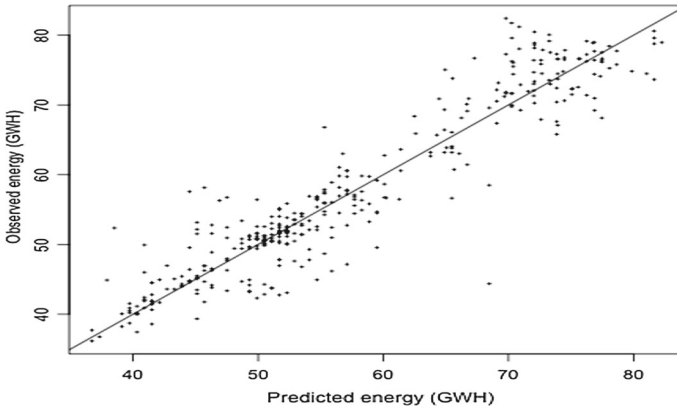


**Table 2** Distribution (by month) of the RAE of the peak load obtained by using as covariate the previous day  $\text{Prev. Day}$  and as prediction methods the conditional mode, conditional median and regression function respectively

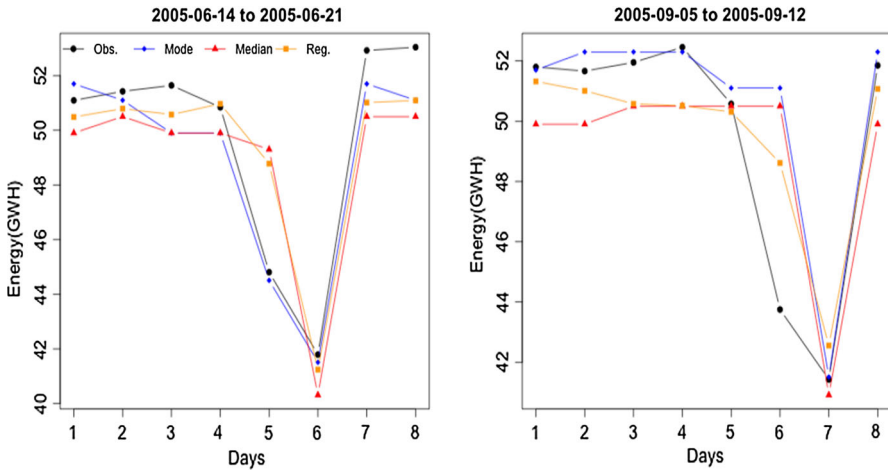
	Mode (%)				Median (%)				Reg. (%)			
	$\text{MAPE}_m$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$\text{MAPE}_m$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$\text{MAPE}_m$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$
Jan.	4.2	1.4	3.1	5.2	5.8	1.5	4.3	7.9	5.1	1.8	4.7	7.7
Feb.	5.4	1.9	4.0	8.3	5.9	3.3	4.9	7.8	6.4	3.1	5.4	9.0
Mar.	6.1	2.0	3.7	8.0	5.6	1.9	4.1	8.1	6.0	2.7	4.3	8.5
Apr.	6.7	2.6	4.9	9.4	5.9	2.5	4.1	8.4	4.1	1.5	2.6	5.2
May	4.7	0.5	1.5	9.5	4.6	1.1	2.4	7.8	4.5	1.2	2.8	5.6
Jun.	2.4	0.7	1.2	2.2	3.7	1.2	2.2	5.2	2.9	0.8	2.1	3.0
Jul.	3.1	0.5	1.5	3.2	4.5	1.6	2.5	3.5	2.9	0.6	1.1	2.8
Aug.	3.0	0.6	1.3	3.6	4.0	0.8	2.8	7.3	5.8	1.4	5.6	8.1
Sep.	2.5	0.3	0.8	1.7	3.0	0.5	1.6	2.3	2.7	0.4	0.9	3.1
Oct.	4.0	0.9	2.3	4.4	4.2	1.4	2.4	5.9	3.4	1.3	2.0	4.3
Nov.	5.4	2.0	3.9	7.1	6.8	3.1	5.4	7.1	5.3	2.8	4.3	8.3
Dec.	5.9	2.2	5.5	8.2	6.6	3.1	6.2	9.1	5.4	2.6	4.7	6.1

energy coming from the traditional energy plants (e.g. nuclear, hydraulic, ...) and from renewable energy resources (e.g. solar and wind) during the day and then use it at the evening and especially over the 3 h around the peak (around 7 p.m. in winter and 2 p.m. in summer). Therefore, an accurate forecast of the energy that will be consumed in the evening allows to optimize the capacity of the storage and consequently to increase the batteries life.

In this subsection, we suggest to solve this forecasting issue by using the mode regression (Fig. 10). Regarding to the discussion made in the previous subsection, we use as covariate the load curve of the previous day. Formally, if we consider  $Z_i(t)$  the load curve of some day  $i$ , then the electrical energy consumed between  $t_1 = 17:30$  and  $t_2 = 20:30$  (respectively  $t_1 = 12:30$  and  $t_2 = 15:30$ ) in winter (respectively, in summer) is defined as  $\mathcal{E}_i = \int_{t_1}^{t_2} Z_i(t) dt$  and measured in Giga-Watt per Hour (GWH). Therefore, here  $\varphi(\cdot)$  is the integral function. As in the previous subsection, the same data sets and the same evaluation procedures used here. As for the peak load forecasting (see details in Sect. 3.1), the same choices for the tuning parameters ( $K$ ,  $H$ ,  $d(\cdot, \cdot)$ ,  $h_K$  and  $h_H$ ) are used. As mentioned before, the functional covariable is supposed to be the last observed daily load curve. Figures 11 and 12 show that conditional mode approach performs well energy forecasts and that conditional median, as well as, regression function, under-predict energy for cold days and over-predict it in hot ones. Figure 13 provides the distribution by month of the daily RAE. We can observe that accurate results are obtained with the conditional mode predictor. Numerical details, namely monthly  $\text{MAPE}$ , first quartile  $Q_{0.25}$ , the median  $Q_{0.5}$  and the third quartile  $Q_{0.75}$ , of the obtained errors are given in Table 3. We can see again that conditional mode performs better the forecasts of the consumed energy.



**Fig. 10** Observed daily consumed energy in 2005 versus its predicted values



**Fig. 11** Examples of energy demand forecast using conditional mode, conditional median and regression function for eight consecutive days

### 4 Proofs

In order to prove our results, we introduce some further notation. Let

$$\tilde{f}_{\varphi,n}(x, y) = \frac{1}{nh_H \mathbb{E}[\Delta_1(x)]} \sum_{i=1}^n \mathbb{E} \left[ \Delta_i(x) H \left( \frac{y - \varphi(Z_i)}{h_H} \right) \mid \mathcal{F}_{i-1} \right].$$

and

$$\bar{l}_n(x) = \frac{1}{n \mathbb{E}(\Delta_1(x))} \sum_{i=1}^n \mathbb{E}(\Delta_i(x) \mid \mathcal{F}_{i-1}).$$

Define the conditional bias of the conditional density estimate of  $\varphi(Z_i)$  given  $X = x$  as

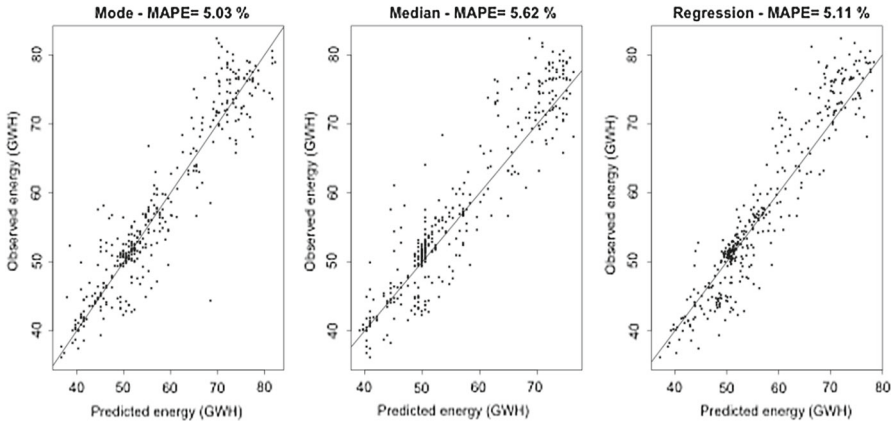


Fig. 12 Observed daily consumed energy in 2005 versus its predicted values

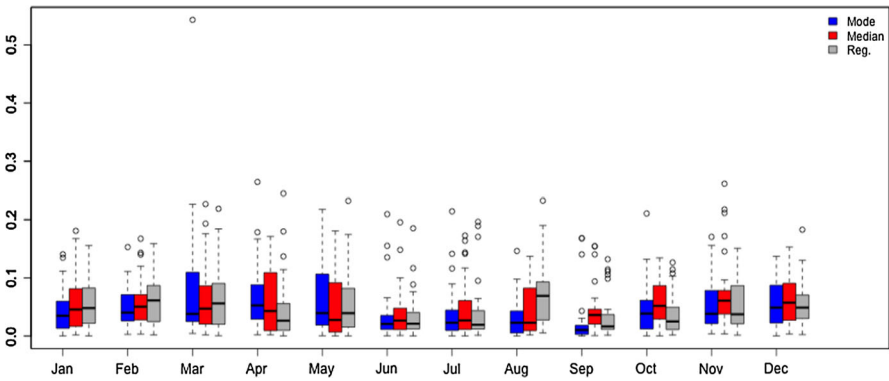


Fig. 13 Distribution (by month) of the daily RAE of the consumed energy

$$B_{\varphi,n}(x, y) = \frac{\bar{f}_{\varphi,n}(x, y)}{\bar{l}_n(x)} - g_{\varphi}(y|x).$$

Consider now the following quantities

$$R_{\varphi,n}(x, y) = -B_{\varphi,n}(x, y)(l_n(x) - \bar{l}_n(x)),$$

and

$$Q_{\varphi,n}(x, y) = (f_{\varphi,n}(x, y) - \bar{f}_{\varphi,n}(x, y)) - g_{\varphi}(y|x)(l_n(x) - \bar{l}_n(x)).$$

It is then clear that the following decomposition holds

$$g_{\varphi,n}(y|x) - g_{\varphi}(y|x) = B_{\varphi,n}(x, y) + \frac{R_{\varphi,n}(x, y) + Q_{\varphi,n}(x, y)}{l_n(x)}. \tag{7}$$

**Table 3** Distribution (by month) of the relative absolute errors of the energy consumed obtained by using as covariate the previous day `Prev.Day` and as prediction methods the conditional mode, conditional median and regression function respectively

	Mode (%)				Median (%)				Reg. (%)			
	MAPE <sub>m</sub>	Q <sub>0.25</sub>	Q <sub>0.5</sub>	Q <sub>0.75</sub>	MAPE <sub>m</sub>	Q <sub>0.25</sub>	Q <sub>0.5</sub>	Q <sub>0.75</sub>	MAPE <sub>m</sub>	Q <sub>0.25</sub>	Q <sub>0.5</sub>	Q <sub>0.75</sub>
Jan.	4.3	1.3	3.4	5.9	5.7	1.6	4.5	8.1	5.4	2.2	4.8	8.2
Feb.	5.0	2.6	4.0	5.0	5.7	2.7	5.0	6.9	6.2	2.6	6.1	8.2
Mar.	8.6	2.4	3.8	10.9	6.3	2.0	4.7	8.6	6.6	2.0	5.6	9.0
Apr.	6.8	3.0	5.2	8.8	6.4	1.3	4.3	10.3	4.7	1.1	2.6	5.5
May	6.4	1.8	3.9	10.6	5.2	0.6	2.7	9.1	5.7	1.5	3.9	8.2
Jun.	3.6	1.1	2.0	3.5	4.0	1.1	2.6	4.7	3.5	1.2	2.1	3.9
Jul.	3.9	0.9	2.2	4.4	4.8	1.1	2.6	6.0	4.0	1.2	1.9	4.3
Aug.	3.2	0.5	2.2	4.3	4.6	0.9	2.2	8.2	6.9	2.7	6.9	9.3
Sep.	2.5	0.3	1.0	1.7	4.4	2.1	3.6	4.5	3.3	1.1	1.6	3.5
Oct.	4.5	1.2	3.8	6.1	5.5	2.9	5.1	8.6	3.7	1.1	2.5	4.9
Nov.	5.2	2.1	3.8	7.7	7.6	3.7	6.0	7.7	5.3	2.2	3.7	8.4
Dec.	5.8	2.2	4.8	8.7	6.5	2.7	5.7	9.0	5.4	3.0	4.8	7.0

The proofs of our results need the following lemmas as tools for which details of their proofs may be found in [Laïb and Louani \(2011\)](#).

**Lemma 1** *Let  $(X_n)_{n \geq 1}$  be a sequence of martingale differences with respect to the sequence of  $\sigma$ -fields  $(\mathcal{F}_n = \sigma(X_1, \dots, X_n))_{n \geq 1}$ , where  $\sigma(X_1, \dots, X_n)$  is the  $\sigma$ -field generated by the random variables  $X_1, \dots, X_n$ . Set  $S_n = \sum_{i=1}^n X_i$ . Suppose that the random variables  $(X_i)_{i \geq 1}$  are bounded by a constant  $M > 0$ , i.e., for any  $i \geq 1$ ,  $|X_i| \leq M$  almost surely, and  $\mathbb{E}(X_i^2 | \mathcal{F}_{i-1}) \leq d_i^2$  almost surely. Then we have, for any  $\lambda > 0$ , that*

$$\mathbb{P}(|S_n| > \lambda) \leq 2 \exp \left\{ -\frac{\lambda^2}{4D_n + 2M\lambda} \right\},$$

where  $D_n = \sum_{i=1}^n d_i^2$ .

**Lemma 2** *Assume that conditions **A1** ((i), (ii), (iv)) and **A2** hold true. For  $1 \leq j \leq 2 + \delta$  for some  $\delta > 0$ , we have*

- (i)  $\frac{1}{\phi(h_K)} \mathbb{E}[\Delta_i^j(x) | \mathcal{F}_{i-1}] = M_j f_{i,1}(x) + O_{a.s.} \left( \frac{\psi_{i,x}(h_K)}{\phi(h_K)} \right)$ ,
- (ii)  $\frac{1}{\phi(h_K)} \mathbb{E}[\Delta_1^j(x)] = M_j f_1(x) + o(1)$ ,

where  $M_j = K^j(1) - \int_0^1 (K^j)' \tau_0(u) du$ .

*Proof of Proposition 1* Considering the decomposition (7), the proof follows from Lemmas 3, 4, 5 and 6 given hereafter, establishing respectively the convergence of  $l_n(x)$  to 1 together with the rate convergence of  $l_n(x) - \bar{l}_n(x)$  to zero and the orders of terms  $B_{\varphi,n}(x, y)$ ,  $R_{\varphi,n}(x, y)$  and  $Q_{\varphi,n}(x, y)$ . Note that, due to the condition (3), the term  $R_{\varphi,n}(x, y)$  is negligible as compared to the term  $B_{\varphi,n}(x, y)$ .  $\square$

**Lemma 3** *Under assumptions A1 and A2, we have*

- (i)  $l_n(x) - \bar{l}_n(x) = O_{a.s} \left( \sqrt{\frac{\log(n)}{n\phi(h_K)}} \right)$ ,
- (ii)  $\lim_{n \rightarrow \infty} l_n(x) = \lim_{n \rightarrow \infty} \bar{l}_n(x) = 1, \quad a.s.$

*Proof of Lemma 3* The results follow by making use of Lemma 1 and Lemma 2 in Laïb and Louani (2011). Details of the proof may be found in Laïb and Louani (2010).  $\square$

**Lemma 4** *Under assumptions A1, A2, A3(iv), A4(i) and A5, we have, as  $n \rightarrow \infty$ ,*

$$\sup_{\varphi \in \mathcal{C}} \sup_{y \in S_\varphi} (\bar{f}_{\varphi,n}(x, y) - \bar{l}_n(x)g_\varphi(y|x)) = O_{a.s.}(h_K^\beta + h_H^\nu).$$

*Proof of Lemma 4* By condition A5 with  $j = 0$ , we have

$$\bar{f}_{\varphi,n}(x, y) = \frac{1}{nh_H \mathbb{E}[\Delta_1(x)]} \sum_{i=1}^n \mathbb{E} \left[ \Delta_i(x) \mathbb{E} \left( H \left( \frac{y - \varphi(Z_i)}{h_H} \right) \mid X_i \right) \mid \mathcal{F}_{i-1} \right].$$

A change of variables and the fact that  $\int_{\mathbb{R}} H(t)dt = 1$  allow us to write

$$\begin{aligned} \mathbb{E} \left[ H \left( \frac{y - \varphi(Z_i)}{h_H} \right) \mid X_i \right] &= h_H \int_{\mathbb{R}} H(t)[g_\varphi(y - th_H | X_i) - g_\varphi(y|x)]dt + h_H g_\varphi(y|x) \\ &=: J_{i,\varphi} + h_H g_\varphi(y|x). \end{aligned}$$

Thus,

$$\begin{aligned} \bar{f}_{\varphi,n}(x, y) &= \frac{g_\varphi(y|x)}{n \mathbb{E}[\Delta_1(x)]} \sum_{i=1}^n \mathbb{E} [\Delta_i(x) | \mathcal{F}_{i-1}] \\ &\quad + \frac{1}{nh_H \mathbb{E}[\Delta_1(x)]} \sum_{i=1}^n \mathbb{E} [\Delta_i(x) J_{i,\varphi} | \mathcal{F}_{i-1}] \\ &=: \bar{l}_n(x)g_\varphi(y|x) + S_2. \end{aligned}$$

Using condition A3(iv), one may write

$$|S_2| \leq \frac{1}{nh_H \mathbb{E}[\Delta_1(x)]} \sum_{i=1}^n \mathbb{E} [\Delta_i(x) | J_{i,\varphi} | \mathcal{F}_{i-1}]$$

$$\begin{aligned} &\leq \frac{1}{n\mathbb{E}[\Delta_1(x)]} \sum_{i=1}^n \mathbb{E} \left[ \Delta_i(x) \left\{ \int_{\mathbb{R}} H(t) (|t|^\nu h_H^\nu + (d(x, X_i))^\beta) dt \right\} | \mathcal{F}_{i-1} \right] \\ &\leq C_x \left\{ h_H^\nu \int_{\mathbb{R}} |t|^\nu H(t) dt + h_K^\beta \right\} \frac{1}{n\mathbb{E}[\Delta_1(x)]} \sum_{i=1}^n \mathbb{E} [\Delta_i(x) | \mathcal{F}_{i-1}]. \end{aligned}$$

Moreover, considering Lemma 2 in [Laïb and Louani \(2011\)](#) combined with the condition **A4(i)** imply that

$$\bar{f}_{\varphi,n}(x, y) - \bar{l}_n(x)g_\varphi(y|x) = O_{a.s.} \left( h_K^\beta + h_H^\nu \right),$$

where  $O_{a.s.}$  does not depend on  $\varphi \in \mathcal{C}$ . □

The following Lemma describes the asymptotic behavior of the conditional bias term  $B_{\varphi,n}(x, y)$  as well as that of  $R_{\varphi,n}(x, y)$  and  $Q_{\varphi,n}(x, y)$ .

**Lemma 5** *Under conditions **A1**, **A2**, **A3(ii)**, **A4(i)** and **A5**, we have*

$$\sup_{\varphi \in \mathcal{C}} \sup_{y \in \mathcal{S}_\varphi} |B_{\varphi,n}(x, y)| = O_{a.s.}(h_K^\beta + h_H^\nu), \tag{8}$$

$$\sup_{\varphi \in \mathcal{C}} \sup_{y \in \mathcal{S}_\varphi} |R_{\varphi,n}(x, y)| = O_{a.s.} \left( (h_K^\beta + h_H^\nu) \left( \sqrt{\frac{\log(n)}{n\phi(h_K)}} \right) \right). \tag{9}$$

Moreover, when hypotheses (3)–(4) are satisfied, we have

$$\sup_{\varphi \in \mathcal{C}} \sup_{y \in \mathcal{S}_\varphi} |Q_{\varphi,n}(x, y)| = O_{a.s.} \left( \eta h_H^{-2} \right) + O_{a.s.}(\lambda_n) + O_{a.s.} \left( \left( \frac{\log n}{n\phi(h_K)} \right)^{1/2} \right). \tag{10}$$

*Proof of Lemma 5* Observe that

$$B_{\varphi,n}(x, y) = \frac{\bar{f}_{\varphi,n}(x, y) - g_\varphi(y|x)\bar{l}_n(x)}{\bar{l}_n(x)} := \frac{\tilde{B}_{\varphi,n}(x, y)}{\bar{l}_n(x)}.$$

Making use of Lemma 4, we obtain  $\sup_{\varphi \in \mathcal{C}} \sup_{y \in \mathcal{S}_\varphi} |\tilde{B}_{\varphi,n}(x, y)| = O_{a.s.}(h_K^\beta + h_H^\nu)$ . The statement (8) follows then from the second part of Lemma 3.

To deal now with the quantity  $R_{\varphi,n}(x, y)$ , write it as

$$R_{\varphi,n}(x, y) = -\frac{\tilde{B}_{\varphi,n}(x, y)}{\bar{l}_n(x)}(l_n(x) - \bar{l}_n(x)).$$

Therefore, the statement (9) follows from the statement (8) combined with Lemma 3 (i).

In order to check the result (10), recall that

$$Q_{\varphi,n}(x, y) = (f_{\varphi,n}(x, y) - \bar{f}_{\varphi,n}(x, y)) - g_{\varphi}(y|x)(l_n(x) - \bar{l}_n(x)).$$

Therefore the statement (10) results from Lemma 3 and the use of Lemma 6 established hereafter. This completes the proof of Lemma 5.  $\square$

The following Lemma is needed as a step in proving Theorem 1

**Lemma 6** *Under assumptions A1, A2, A3, A4(ii), A5 together with hypotheses (3)–(4), for  $n$  large enough, we have*

$$\sup_{\varphi \in \mathcal{C}} \sup_{y \in S_{\varphi}} |f_{\varphi,n}(x, y) - \bar{f}_{\varphi,n}(x, y)| = O_{a.s.}(\eta h_H^{-2}) + O_{a.s.}(\lambda_n).$$

*Proof of Lemma 6* Recall that, for any  $\varphi \in \mathcal{C}$ ,

$$S_{\varphi} = [\theta_{\varphi}(x) - \xi, \theta_{\varphi}(x) + \xi].$$

Let  $\varphi_1, \varphi_2 \in \mathcal{C}$  and, for any  $\epsilon > 0$ , define the set

$$S_{\varphi_2}^{\epsilon} = \{y \in \mathbb{R} : |y - \Theta_{\varphi_2}(x)| \leq \xi + \epsilon\}.$$

It is easily seen, by condition A3(i), that, for any  $\epsilon > 0$ , there exists  $\eta > 0$  for which the fact that  $\varphi_1 \in B(\varphi_2, \eta)$  implies  $S_{\varphi_1} \subset S_{\varphi_2}^{\epsilon}$ . Consider a grid  $(\varphi_j)_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}})}$  on the space  $\mathcal{C}$  such that  $d_{\mathcal{C}}(\varphi_i, \varphi_j)$  for any  $i \neq j$ . Therefore, we have

$$\begin{aligned} & \sup_{\varphi \in \mathcal{C}} \sup_{y \in S_{\varphi}} |f_{\varphi,n}(x, y) - \bar{f}_{\varphi,n}(x, y)| \\ & \leq \max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}})} \sup_{\varphi \in B(\varphi_j, \eta)} \sup_{y \in S_{\varphi}} |f_{\varphi,n}(x, y) - \bar{f}_{\varphi,n}(x, y)| \\ & \leq \max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}})} \sup_{\varphi \in B(\varphi_j, \eta)} \sup_{y \in S_{\varphi_j}^{\epsilon}} |f_{\varphi,n}(x, y) - \bar{f}_{\varphi,n}(x, y)|. \end{aligned} \tag{11}$$

Using now the compactness of  $S_{\varphi_j}^{\epsilon}$  and the fact that its length is  $2(\xi + \epsilon)$  for any  $\varphi_j$ , we can write  $S_{\varphi_j}^{\epsilon} \subset \cup_{k=1}^{d_{\epsilon,n}} S_{\varphi_j,k}^{\epsilon}$  where  $S_{\varphi_j,k}^{\epsilon} = (t_{\varphi_j,k}^{\epsilon} - m_n; t_{\varphi_j,k}^{\epsilon} + m_n)$  and  $m_n$  and  $d_{\epsilon,n}$  are such that  $d_{\epsilon,n} = C_{\epsilon} m_n^{-1}$  for some positive constant  $C_{\epsilon}$ . Moreover, we have

$$\begin{aligned} \sup_{y \in S_{\varphi_j}^{\epsilon}} |f_{\varphi,n}(x, y) - \bar{f}_{\varphi,n}(x, y)| & \leq \max_{1 \leq k \leq d_{\epsilon,n}} \sup_{y \in S_{\varphi_j,k}^{\epsilon}} |f_{\varphi,n}(x, y) - f_{\varphi,n}(x, t_{\varphi_j,k}^{\epsilon})| \\ & \quad + \max_{1 \leq k \leq d_{\epsilon,n}} |f_{\varphi,n}(x, t_{\varphi_j,k}^{\epsilon}) - \bar{f}_{\varphi,n}(x, t_{\varphi_j,k}^{\epsilon})| \\ & \quad + \max_{1 \leq k \leq d_{\epsilon,n}} \sup_{y \in S_{\varphi_j,k}^{\epsilon}} |\bar{f}_{\varphi,n}(x, t_{\varphi_j,k}^{\epsilon}) - \bar{f}_{\varphi,n}(x, y)| \\ & =: J_{\varphi,n,1} + J_{\varphi,n,2} + J_{\varphi,n,3}. \end{aligned} \tag{12}$$

Making use of **A4(ii)**, we obtain

$$\begin{aligned} J_{\varphi,n,1} &\leq \frac{1}{nh_H \mathbb{E}[\Delta_1(x)]} \max_{1 \leq k \leq d_{\epsilon,n}} \\ &\quad \sup_{y \in S_{\varphi_j,k}^\epsilon} \sum_{i=1}^n \Delta_i(x) \left| H\left(\frac{y - \varphi(Z_i)}{h_H}\right) - H\left(\frac{t_{\varphi_j,k}^\epsilon - \varphi(Z_i)}{h_H}\right) \right| \\ &\leq C_H m_n h_H^{-2} l_n(x). \end{aligned} \quad (13)$$

Similarly, we have also

$$J_{\varphi,n,3} \leq C_H m_n h_H^{-2} \bar{l}_n(x). \quad (14)$$

Therefore,

$$\max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_C)} \sup_{\varphi \in B(\varphi_j, \eta)} (J_{\varphi,n,1} + J_{\varphi,n,3}) \leq C_H m_n h_H^{-2} (l_n(x) + \bar{l}_n(x)). \quad (15)$$

Using Lemma 3(ii), it follows that

$$\max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_C)} \sup_{\varphi \in B(\varphi_j, \eta)} (J_{\varphi,n,1} + J_{\varphi,n,3}) = O_{a.s.} \left( m_n h_H^{-2} \right). \quad (16)$$

To identify the convergence rate to zero of the term  $\max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_C)} \sup_{\varphi \in B(\varphi_j, \eta)} J_{\varphi,n,2}$ , observe that

$$\begin{aligned} &\max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_C)} \sup_{\varphi \in B(\varphi_j, \eta)} J_{\varphi,n,2} \\ &\leq \max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_C)} \sup_{\varphi \in B(\varphi_j, \eta)} \max_{1 \leq k \leq d_{\epsilon,n}} |f_{\varphi,n}(x, t_{\varphi_j,k}^\epsilon) - f_{\varphi_j,n}(x, t_{\varphi_j,k}^\epsilon)| \\ &\quad + \max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_C)} \max_{1 \leq k \leq d_{\epsilon,n}} |f_{\varphi_j,n}(x, t_{\varphi_j,k}^\epsilon) - \bar{f}_{\varphi_j,n}(x, t_{\varphi_j,k}^\epsilon)| \\ &\quad + \max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_C)} \sup_{\varphi \in B(\varphi_j, \eta)} \max_{1 \leq k \leq d_{\epsilon,n}} |\bar{f}_{\varphi_j,n}(x, t_{\varphi_j,k}^\epsilon) - \bar{f}_{\varphi,n}(x, t_{\varphi_j,k}^\epsilon)| \\ &=: J_{n,1} + J_{n,2} + J_{n,3}. \end{aligned} \quad (17)$$

By the same arguments as in the statement (15), we can show, under Condition **A4(ii)**, that

$$(J_{n,1} + J_{n,3}) \leq C_H \eta h_H^{-2} (l_n(x) + \bar{l}_n(x)) = O_{a.s.} \left( \eta h_H^{-2} \right). \quad (18)$$

We have to deal now with the middle term  $J_{n,2}$ . Observe that



$$\begin{aligned} \mathbb{P} (J_{n,2} > \lambda) &= \mathbb{P} \left( \max_{1 \leq j \leq \mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}})} \max_{1 \leq k \leq d_{\epsilon, n}} |f_{\varphi_j, n}(x, t_{\varphi_j, k}^\epsilon) - \bar{f}_{\varphi_j, n}(x, t_{\varphi_j, k}^\epsilon)| > \lambda \right) \\ &\leq \sum_{j=1}^{\mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}})} \sum_{k=1}^{d_{\epsilon, n}} \mathbb{P} \left( \frac{1}{nh_H} \left| \sum_{i=1}^n L_{i, \varphi_j}(x, t_{\varphi_j, k}^\epsilon) \right| \geq \lambda \right), \end{aligned}$$

where  $L_{i, \varphi_j}(x, y) = \frac{1}{\mathbb{E}[\Delta_1(x)]} \left[ \Delta_i(x) H\left(\frac{y - \varphi_j(Z_i)}{h_H}\right) - \mathbb{E} \left[ \Delta_i(x) H\left(\frac{y - \varphi_j(Z_i)}{h_H}\right) \mid \mathcal{F}_{i-1} \right] \right]$ . Notice that  $L_{i, \varphi_j}(x, y)$  is a martingale difference bounded by the quantity  $M := \frac{2\bar{K}\bar{H}}{\phi(h_K)[M_1 f_1(x) + o(1)]}$ . In fact, since the kernel  $K$  and the function  $H$  are bounded, it follows easily in view of Lemma 2 (ii) in [Laïb and Louani \(2011\)](#) that

$$|L_{i, \varphi_j}(x, y)| \leq \frac{2\bar{K}\bar{H}}{\mathbb{E}[\Delta_1(x)]} = \frac{2\bar{K}\bar{H}}{\phi(h_K)[M_1 f_1(x) + o(1)]},$$

where  $\bar{H} := \sup_{y \in \mathbb{R}} H(y)$  and  $\bar{K} := \sup_{y \in \mathbb{R}} K(y)$ . Observe now that

$$\begin{aligned} &\mathbb{E}[(L_{i, \varphi_j}(x, t_{\varphi_j, k}^\epsilon))^2 \mid \mathcal{F}_{i-1}] \\ &\leq \frac{1}{(\mathbb{E}[\Delta_1(x)])^2} \mathbb{E} \left[ \left( \Delta_i(x) H \left( \frac{t_{\varphi_j, k}^\epsilon - \varphi_j(Z_i)}{h_H} \right) \right)^2 \mid \mathcal{F}_{i-1} \right]. \end{aligned}$$

Therefore, by condition **A5**, we have

$$\begin{aligned} &\mathbb{E} \left[ \left( \Delta_i(x) H \left( \frac{t_{\varphi_j, k}^\epsilon - \varphi_j(Z_i)}{h_H} \right) \right)^2 \mid \mathcal{F}_{i-1} \right] \\ &= \mathbb{E} \left[ (\Delta_i(x))^2 \mathbb{E} \left[ \left( H \left( \frac{t_{\varphi_j, k}^\epsilon - \varphi_j(Z_i)}{h_H} \right) \right)^2 \mid \mathcal{G}_{i-1} \right] \mid \mathcal{F}_{i-1} \right] \\ &= \mathbb{E} \left[ (\Delta_i(x))^2 \mathbb{E} \left[ \left( H \left( \frac{t_{\varphi_j, k}^\epsilon - \varphi_j(Z_i)}{h_H} \right) \right)^2 \mid X_i \right] \mid \mathcal{F}_{i-1} \right] \\ &= \mathbb{E} \left[ (\Delta_i(x))^2 \int_{\mathbb{R}} \left( H \left( \frac{u}{h_H} \right) \right)^2 g_{\varphi_j}(t_{\varphi_j, k}^\epsilon - u \mid X_i) du \mid \mathcal{F}_{i-1} \right] \\ &:= \mathbb{E} \left[ (\Delta_i(x))^2 \mathcal{T}_{i,1} \mid \mathcal{F}_{i-1} \right] + \mathcal{T}_2 \mathbb{E} \left[ (\Delta_i(x))^2 \mid \mathcal{F}_{i-1} \right], \end{aligned}$$

where we have set  $\mathcal{T}_{i,1} = \int_{\mathbb{R}} \left( H \left( \frac{u}{h_H} \right) \right)^2 \left( g_{\varphi_j}(t_{\varphi_j, k}^\epsilon - u \mid X_i) - g_{\varphi_j}(t_{\varphi_j, k}^\epsilon \mid x) \right) du$  and  $\mathcal{T}_2 = \int_{\mathbb{R}} \left( H \left( \frac{u}{h_H} \right) \right)^2 g_{\varphi_j}(t_{\varphi_j, k}^\epsilon \mid x) du$ . Subsequently, for  $\eta > 0$ , we have

$$\begin{aligned} \mathcal{T}_{i,1} &\leq \int_{|u|\leq\eta} \left( H\left(\frac{u}{h_H}\right) \right)^2 \left| g_{\varphi_j}(t_{\varphi_j,k}^\epsilon - u|X_i) - g_{\varphi_j}(t_{\varphi_j,k}^\epsilon|x) \right| du \\ &\quad + \int_{|u|>\eta} \left( H\left(\frac{u}{h_H}\right) \right)^2 \left| g_{\varphi_j}(t_{\varphi_j,k}^\epsilon - u|X_i) - g_{\varphi_j}(t_{\varphi_j,k}^\epsilon|x) \right| du \\ &\leq h_H \sup_{|u|\leq\eta} |g_{\varphi_j}(t_{\varphi_j,k}^\epsilon - u|X_i) - g_{\varphi_j}(t_{\varphi_j,k}^\epsilon|x)| \int_{|u|\leq\eta/h_H} (H(u))^2 du \\ &\quad + h_H \sup_{|u|>\eta/h_H} (H(u))^2 + h_H g_{\varphi_j}(t_{\varphi_j,k}^\epsilon|x) \int_{|u|>\eta/h_H} (H(u))^2 du; \end{aligned}$$

Condition **A3**(iv) allows us, for any  $\eta > 0$ , to write

$$\begin{aligned} \mathcal{T}_{i,1} &\leq h_H C_x (|\eta|^\nu + d(x, X_i)^\beta) \int_{|u|\leq\eta/h_H} (H(u))^2 du \\ &\quad + h_H \sup_{|u|>\eta/h_H} (H(u))^2 + h_H g_{\varphi_j}(t_{\varphi_j,k}^\epsilon|x) \int_{|u|>\eta/h_H} (H(u))^2 du. \end{aligned}$$

Thus,

$$\begin{aligned} &\mathbb{E} \left[ (\Delta_i(x))^2 \mathcal{T}_{i,1} \mid \mathcal{F}_{i-1} \right] \\ &\leq h_H \mathbb{E} \left[ (\Delta_i(x))^2 \left( C_x (|\eta|^\nu + d(x, X_i)^\beta) \int_{|u|\leq\frac{\eta}{h_H}} H^2(u) du \right. \right. \\ &\quad \left. \left. + \sup_{|u|>\frac{\eta}{h_H}} H^2(u) + g_{\varphi_j}(t_{\varphi_j,k}^\epsilon|x) \int_{|u|>\frac{\eta}{h_H}} H^2(u) du \right) \mid \mathcal{F}_{i-1} \right] \\ &\leq h_H \left( C_x (|\eta|^\nu + h_K^\beta) \int_{|u|\leq\eta/h_H} (H(u))^2 du + \sup_{|u|>\frac{\eta}{h_H}} H^2(u) \right. \\ &\quad \left. + g_{\varphi_j}(t_{\varphi_j,k}^\epsilon|x) \int_{|u|>\frac{\eta}{h_H}} H^2(u) du \right) \mathbb{E} \left[ (\Delta_i(x))^2 \mid \mathcal{F}_{i-1} \right]. \end{aligned}$$

On another hand, we can see easily, for some positive constant  $C_0$ , that

$$\mathcal{T}_2 \mathbb{E} \left[ (\Delta_i(x))^2 \mid \mathcal{F}_{i-1} \right] \leq C_0 h_H \mathbb{E} \left[ (\Delta_i(x))^2 \mid \mathcal{F}_{i-1} \right].$$

Therefore, since  $H$  is bounded and  $h_H \rightarrow 0$ , it follows then that there exists a constant  $C_1 > 0$  such that

$$\mathbb{E} \left[ \left( \Delta_i(x) H \left( \frac{t_{\varphi_j,k}^\epsilon - \varphi(Z_i)}{h_H} \right) \right)^2 \mid \mathcal{F}_{i-1} \right] \leq C_1 h_H \mathbb{E} \left[ (\Delta_i(x))^2 \mid \mathcal{F}_{i-1} \right] \text{ as } n \rightarrow \infty.$$

Furthermore, using Condition **(A1)**, which supposes almost surely that  $f_{i,1}$  is bounded by a deterministic function  $b(x)$  and that  $\psi_{i,x}(h_K) \leq \phi(h_K)$  as  $h_K \rightarrow 0$ , together with Lemma 2 in [Laïb and Louani \(2011\)](#), we have for  $n$  large enough

$$\begin{aligned} \mathbb{E}[(L_{i,\varphi_j}(x, t_{\varphi_j,k}^\epsilon))^2 | \mathcal{F}_{i-1}] &\leq \frac{C_1 h_H}{(\mathbb{E}[\Delta_1(x)])^2} \mathbb{E}[(\Delta_i(x))^2 | \mathcal{F}_{i-1}] \\ &\leq \frac{C_1 h_H}{\phi(h_K)[M_1^2 f_1^2(x) + o(1)]} [M_2 b_i(x) + 1] =: d_i^2. \end{aligned}$$

Moreover, using Conditions **A1**(iii), (v), one may write

$$\frac{4D_n}{n} + 2Mh_H\lambda = \frac{h_H}{\phi(h_K)} \left[ \frac{4C_1[M_2D(x) + 1]}{M_1^2 f_1^2(x) + o(1)} + \frac{4\lambda\bar{K}\bar{H}}{M_1 f_1(x) + o(1)} \right]$$

and  $\frac{nh_H^2\lambda^2}{4D_n/n + 2Mh_H\lambda} = nh_H\phi(h_K)\lambda^2 C_\epsilon(x)$ , where

$$C_\epsilon(x) = \frac{M_1 f_1(x)}{4} \cdot \frac{1}{\frac{C_1(M_2D(x)+1)}{M_1 f_1(x)} + \lambda\bar{H}\bar{K} + o(1)}.$$

Consequently,

$$\mathbb{P}\left(\frac{1}{nh_H} \left| \sum_{i=1}^n L_{i,\varphi_j}(x, t_{\varphi_j,k}^\epsilon) \right| \geq \lambda\right) \leq 2 \exp\left\{-nh_H\phi(h_K)\lambda^2 C_\epsilon(x)\right\}.$$

Choosing  $\lambda = \lambda_n$  and  $m_n = \eta$ , we obtain

$$\begin{aligned} \mathbb{P}(J_{2,n} \geq \lambda_n) &\leq 2\mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}})d_{\epsilon,n} \exp\left\{-nh_H\phi(h_K)\lambda_n^2 C_\epsilon(x)\right\} \\ &\leq 2 \exp\left\{-\lambda_n^2 nh_H\phi(h_K) \left[ C_\epsilon(x) - \frac{c_\epsilon \log \mathcal{N}(\eta, \mathcal{C}, d_{\mathcal{C}})}{\eta\lambda_n^2 nh_H\phi(h_K)} \right]\right\}. \end{aligned}$$

Taking into account the condition (4), it suffices to use the Borel–Cantelli Lemma to conclude the proof. □

*Proof of Theorem 1* Taylor series expansion of the function  $g_\varphi(\hat{\Theta}_{\varphi,n}(x)|x)$  around  $\Theta_\varphi(x)$  together with the definition of  $\Theta_\varphi(x)$  yield

$$g_\varphi\left(\hat{\Theta}_{\varphi,n}(x)|x\right) = g_\varphi\left(\Theta_\varphi(x)|x\right) + \left(\hat{\Theta}_{\varphi,n}(x) - \Theta_\varphi(x)\right)^2 \frac{1}{2} g_\varphi^{(2)}\left(\Theta_{\varphi,n}^*(x)|x\right), \tag{19}$$

where  $\Theta_{\varphi,n}^*$  is between  $\hat{\Theta}_{\varphi,n}(x)$  and  $\Theta_\varphi(x)$ . Subsequently, considering the statement (19) we obtain

$$\left(\hat{\Theta}_{\varphi,n}(x) - \Theta_\varphi(x)\right)^2 |g^{(2)}(\Theta_{\varphi,n}^*(x)|x)| = O\left(\sup_{y \in S_\varphi} |g_n(y|x) - g(y|x)|\right). \tag{20}$$

To end the proof of the theorem, we need the following lemma which deals with the uniform (with respect to  $\varphi \in \mathcal{C}$ ) asymptotic behavior of the conditional mode estimate.

**Lemma 7** *Under assumptions of Proposition 1, we have*

$$\lim_{n \rightarrow \infty} \sup_{\varphi \in \mathcal{C}} |\hat{\Theta}_{\varphi,n}(x) - \Theta_{\varphi}(x)| = 0 \quad a.s.$$

*Proof of Lemma 7* Since by the assumption **A3(ii)**, uniformly in  $\varphi \in \mathcal{C}$ ,  $g_{\varphi}(\cdot|x)$  is uniformly continuous on the compact set  $S_{\varphi}$  on which  $\theta_{\varphi}(x)$  is the unique mode. Then, proceeding as in [Parzen \(1962\)](#), for any  $\varepsilon > 0$ , there exists  $\zeta > 0$  such that, for any  $y \in S_{\varphi}$ ,

$$\sup_{\varphi \in \mathcal{C}} |\Theta_{\varphi}(x) - y| \geq \varepsilon \Rightarrow \sup_{\varphi \in \mathcal{C}} |g_{\varphi}(\Theta_{\varphi}(x)|x) - g_{\varphi}(y|x)| \geq \zeta. \tag{21}$$

On another hand, we have

$$\begin{aligned} \sup_{\varphi \in \mathcal{C}} |g_{\varphi}(\hat{\Theta}_{\varphi,n}(x)|x) - g_{\varphi}(\Theta_{\varphi}(x)|x)| &\leq \sup_{\varphi \in \mathcal{C}} |g_{\varphi,n}(\hat{\Theta}_{\varphi,n}(x)|x) - g_{\varphi}(\hat{\Theta}_{\varphi,n}(x)|x)| \\ &\quad + \sup_{\varphi \in \mathcal{C}} |g_{\varphi,n}(\hat{\Theta}_{\varphi,n}(x)|x) - g_{\varphi}(\Theta_{\varphi}(x)|x)| \\ &\leq \sup_{\varphi \in \mathcal{C}} \sup_{y \in S_{\varphi}} |g_{\varphi,n}(y|x) - g_{\varphi}(y|x)| \\ &\quad + \sup_{\varphi \in \mathcal{C}} \left| \sup_{y \in S_{\varphi}} g_{\varphi,n}(y|x) - \sup_{y \in S_{\varphi}} g_{\varphi}(y|x) \right| \\ &\leq 2 \sup_{\varphi \in \mathcal{C}} \sup_{y \in S_{\varphi}} |g_{\varphi,n}(y|x) - g_{\varphi}(y|x)|. \end{aligned} \tag{22}$$

Using the statements (21) and (22) combined with Proposition 1, we obtain the result. □

We come back now on the proof of the Theorem. Making use of Lemma 7 combined with conditions **A3(iii)–(iv)**, we deduce that

$$\lim_{n \rightarrow \infty} \sup_{\varphi \in \mathcal{C}} |g^{(2)}(\Theta_{\varphi,n}^*(x)|x)| = \sup_{\varphi \in \mathcal{C}} |g^{(2)}(\Theta_{\varphi}(x)|x)| = \Phi(x) \neq 0. \tag{23}$$

Moreover, the statements (20), (23) imply that

$$\sup_{\varphi \in \mathcal{C}} (\hat{\Theta}_{\varphi,n}(x) - \Theta_{\varphi}(x))^2 = O \left( \sup_{\varphi \in \mathcal{C}} \sup_{y \in S_{\varphi}} |g_n(y|x) - g(y|x)| \right), \tag{24}$$

which is enough, while considering Proposition 1, to complete the proof of Theorem 1. □

### References

Attaoui S, Laksaci A, Ould Saïd E (2011) A note on the conditional density estimate in the single functional index model. *Stat Probab Lett* 81:45–53

- Dabo-Niang S, Laksaci A (2007) Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *C R Acad Sci Paris* 344:49–52
- Delsol L (2009) Advances on asymptotic normality in non-parametric functional time series analysis. *Statistics* 43(1):13–33
- Demongeot J, Laksaci A, Madani F, Rachdi M (2010) Local linear estimation of the conditional density for functional data. *C R Acad Sci Paris* 348:931–934
- Ezzahrioui M, Ould-Saïd E (2008) Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J Nonparametric Stat* 20:3–18
- Ezzahrioui M, Ould-Saïd E (2010) Some asymptotic results of a non-parametric conditional mode estimator for functional time-series data. *Stat Neerl* 64:171–201
- Ferraty F, Vieu P (2000) Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *C R Acad Sci Paris Ser I Math* 330:139–142
- Ferraty F, Laksaci A, Vieu P (2006) Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat Inference Stoch Process* 9:47–76
- Ferraty F, Vieu P (2006) *Nonparametric functional data analysis. Theory and practice*. Springer series in statistics. Springer, New York
- Ferraty F, Laksaci A, Tadj A, Vieu P (2010) Rate of uniform consistency for nonparametric estimates with functional variables. *J Stat Plan Inference* 140:335–352
- Goia A, May C, Fusai G (2010) Functional clustering and linear regression for peak load forecasting. *Int J Forecast* 26:700–711
- Laïb N (2005) Kernel estimates of the mean and the volatility functions in a nonlinear autoregressive model with ARCH errors. *J Stat Plan Inference* 134:116–139
- Laïb N, Louani D (2010) Nonparametric kernel regression estimation for functional stationary ergodic data: asymptotic properties. *J Multivar Anal* 101:2266–2281
- Laïb N, Louani D (2011) Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. *J Stat Plan Inference* 141:359–372
- Masry E (2005) Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stoch Process Appl* 115:155–177
- Ould Saïd E (1997) A note on ergodic processes prediction via estimation of the conditional mode function. *Scand J Stat* 24:231–239
- Parzen E (1962) On the estimation of a probability density function and mode. *Ann Math Stat* 33:1065–1076
- Ramsay J, Silverman BW (1997) *Functional data analysis*. Springer, New York
- Sigauke C, Chikobvu D (2010) Daily peak electricity load forecasting in South Africa using a multivariate nonparametric regression approach. *ORiON* 26(2):97–111
- van der Vaart AW, Wellner JA (1996) *Weak convergence and empirical processes. With applications to statistics*. Springer series in statistics. Springer, New York