Metrika (2016) 79:919–951
DOI 10.1007/s00184-016-0583-y

Nonparametric estimation in a mixed-effect Ornstein–Uhlenbeck model

Charlotte Dion[1,2]

Received: 20 March 2015 / Published online: 23 May 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Two adaptive nonparametric procedures are proposed to estimate the density of the random effects in a mixed-effect Ornstein–Uhlenbeck model. First a kernel estimator is introduced with a new bandwidth selection method developed recently by Goldenshluger and Lepski (Ann Stat 39:1608–1632, 2011). Then, we adapt an estimator from Comte et al. (Stoch Process Appl 7:2522–2551, 2013) in the framework of small time interval of observation. More precisely, we propose an estimator that uses deconvolution tools and depends on two tuning parameters to be chosen in a data-driven way. The selection of these two parameters is achieved through a two-dimensional penalized criterion. For both adaptive estimators, risk bounds are provided in terms of integrated $\mathbb{L}^2$-error. The estimators are evaluated on simulations and show good results. Finally, these nonparametric estimators are applied to neuronal data and are compared with previous parametric estimations.

Keywords Stochastic differential equations · Ornstein–Uhlenbeck process · Mixed-effect model · Nonparametric estimation · Deconvolution method · Kernel estimator · Neuronal data

Mathematics Subject Classification 62G07 · 62M05

✉ Charlotte Dion
charlotte.dion@imag.fr

1 LJK, UMR CNRS 5224, Université Grenoble Alpes, 51 rue des Mathématiques, 38041 Grenoble, France

2 MAP5, UMR CNRS 8145, Université Paris Descartes, 45 rue des Saints Pères, 75006 Paris, France

1 Introduction

Stochastic differential models have been intensively surveyed in the theoretical literature with either continuous observations (e.g. Kutoyants 2004) or discrete observations, both in the parametric field (e.g. Genon-Catalot and Jacod 1993) or in the nonparametric field (e.g. Hoffmann 1999; Comte et al. 2007). More recently, stochastic differential equations with random effects have been introduced with various applications such as neuronal modelling or pharmacokinetics (e.g. Picchini et al. 2008; Delattre and Lavielle 2013; Donnet and Samson 2013). Mixed-effects models are used to analyse repeated measurements with similar functional form but with some variability between experiments (see Davidian and Giltinan 1995; Pinheiro and Bates 2000; Diggle et al. 2002). The advantage is that a single estimation procedure is used to fit the overall data simultaneously.

Estimation methods in stochastic differential models with random effects have been proposed, especially in the parametric framework (e.g. Donnet and Samson 2008, 2014; Donnet et al. 2010; Picchini et al. 2010; Picchini and Ditlevsen 2011; Delattre and Lavielle 2013; Genon-Catalot, and Larédo 2016; Delattre et al. 2015). All these parametric estimation methods of the density of the random effects are developed assuming a known model on the density, which is often Gaussian. However, one can wonder if this assumption is reasonable depending on the application context. We focus here on the nonparametric estimation of the density of the independent identically distributed random effects. To the best of our knowledge, the only references in this context are Comte et al. (2013) and Dion and Genon-Catalot (2015). The first one provides a nonparametric estimator of the density under restrictive assumptions on the drift and diffusion coefficients. The second one studies the more general case of two linear random effects in the drift. It provides a kernel estimator of the bivariate density of the couple of random parameters. Assuming that the process is at its stationary regime, the authors obtain $\mathbb{L}^2$-convergence results.

The present work proposes two nonparametric estimation methods in the simpler model, i.e. an Ornstein–Uhlenbeck stochastic differential model with one additive random effect, X, the time scale parameter being assumed known. More precisely, we consider N real valued stochastic processes $(X_j(t), t \in [0, T])$, $j = 1, \ldots, N$, with dynamics ruled by the following SDEs:

$$\begin{cases} dX_j(t) = \left(\phi_j - \frac{X_j(t)}{\alpha}\right) dt + \sigma dW_j(t) \\ X_j(0) = x_j \end{cases} \tag{1}$$

where $(W_j)_{1 \le j \le N}$ are N independent Wiener processes, and $(\phi_j)_{1 \le j \le N}$ are N unobserved independent and identically distributed (i.i.d.) random variables taking values in $\mathbb{R}$, with a common density f. The sequences $(\phi_j)_{1 \le j \le N}$ and $(W_j)_{1 \le j \le N}$ are independent. Here $(x_1, \ldots, x_N)$ are known values. The positive constants σ and α are supposed to be known; in practice they are estimated from experimental data. The estimation of σ can be done using the quadratic variation of the process. The constant α is a physical quantity. Picchini et al. (2010) give an estimator of α for the Ornstein–Uhlenbeck model (1) where the likelihood function is explicit and one can

compute maximum likelihood estimators. Each process $(X_j(t), 0 \leq t \leq T)$ represents an individual and the variable ϕ_j is the random effect of individual j. Due to the independence of the ϕ_j and the W_j, the $X_j(t)$, for $j = 1, \ldots, N$ are i.i.d. random variables when t is fixed, also the N trajectories $(X_j(t), 0 \leq t \leq T)$, $j = 1, \ldots, N$ are i.i.d. Nevertheless, differences between observations are due to the realization of both W_j and ϕ_j. The Ornstein–Uhlenbeck model is very useful in practice: at first in physics to describe the movement of a particle, then in the econometric field, or for example in neuroscience to describe the membrane potential of a neuron.

The purpose of the present work is to build nonparametric estimators of the random effect density f, considering that only the processes are observed on $[0, T]$ with $T > 0$ given. In practice we consider discrete observations of the X_j's with a very small time step δ. We are able to evaluate the error made by this discretization. The main difficulty is that we do not observe the ϕ_j's but only the $X_j(k\delta)$'s. Thus the first step is to find an estimator of the random effects ϕ_j and then to estimate f, taking into account the approximation introduced by the estimation of the ϕ_j.

In the context of stochastic differential equations with random effects, Comte et al. (2013) propose different nonparametric estimators with good theoretical properties for large T. Here we adopt two different approaches. First we assume that T is large and we propose a direct estimation method of the density from the estimator of the ϕ_j's. This is the kernel estimator. Then, assuming that T may be small (due to the chosen units for example but still with high frequency data) we focus on the deconvolution estimator.

The kernel estimator depends on a bandwidth to be chosen from the data. Several selection methods of the bandwidth of kernel estimators are known. The originality here is that we use a method, proposed by Goldenshluger and Lepski (2011), which provides an adaptive estimator. This kind of non-asymptotic result is new in this context.

Then we study an estimator built by a deconvolution method (see Butucea and Tsybakov 2007; Comte et al. 2013, for example). The novelty lies in the introduction of an additional tuning parameter to control the variance of the noise. The value of T is then allowed to be small but we still need high frequency observation meaning a small time step.

We obtain a collection of estimators depending on two parameters. To select the final estimator among this collection, we extend the Goldenshluger and Lepski method for a two-dimensional model selection (Goldenshluger and Lepski 2011). Finally we have a consistent estimator satisfying an oracle inequality, for any value of T. This estimator is likely to be applied to experimental data with small T.

We illustrate the properties of the proposed estimators with a simulation study. Especially, we compare them with standard bandwidth selection method of cross-validation type. Then, the estimators are applied to neuronal data. They are intracellular measurements of the neuronal membrane potential between two spikes which can be modelled with an Ornstein–Uhlenbeck model with one random effect as in (1). The potential being reset at a fixed initial value after a spike, we consider that the measure between two spikes is an independent experimental unit with a different realization of the random effect. This assumption has already been considered with parametric strategies in Picchini et al. (2008, 2010), where it is assumed that the random effect is

Gaussian and proven that the Ornstein–Uhlenbeck model with one random effect fits better the data than the model without them. Our goal is to estimate nonparametrically the density of the random effect. This estimated density could be used in further works to model this phenomenon (instead of using the Gaussian density systematically).

The paper is organized as follows. Section 2 is dedicated to giving definitions and presenting the estimators investigated in this work. Then in Sect. 3 we set up a method of bandwidth selection for the kernel estimator. In Sect. 4 we define and study the final data-driven estimator built by deconvolution. In Sect. 5 we calibrate the selection methods and illustrate the good performances of both estimators on simulated data. In Sect. 6 we experiment the procedures on real data. We conclude this article with a discussion in Sect. 7. All proofs are gathered in Sect. 8, and the computation of the error made by discretization is done in "Appendix 1".

2 Presentation of the strategies

2.1 Notation and assumptions

Let us introduce some notations. For two functions g_1 and g_2 in $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, the convolution product of g_1 and g_2 for all $x \in \mathbb{R}$, is $g_1 \star g_2(x) = \int_{\mathbb{R}} g_1(x-y)g_2(y)dy$ and the scalar product is: $\langle g_1, g_2 \rangle = \int_{\mathbb{R}} g_1(x)\overline{g_2(x)}dx$. Then the Fourier transform of g_1 is $g_1^*(x) = \int_{\mathbb{R}} e^{iux} g_1(u)du$ for all $x \in \mathbb{R}$ and the $\mathbb{L}^2$-norm is $\|g_1\|^2 = \int_{\mathbb{R}} |g_1(x)|^2 dx$. Finally we recall the Plancherel–Parseval's formula: $2\pi \|g_1\|^2 = \|g_1^*\|^2$.

We assume **(A)** $f \in \mathbb{L}^2(\mathbb{R})$, $f^* \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$.

2.2 Initial idea

As previously mentioned, the first step of the procedure is to estimate the random effect ϕ_j which are not observed, in order to recover their density, in a second time. For this purpose, we introduce the following random variables for $j = 1, \ldots, N$ and $\tau \in]0, T]$,

$$Z_{j,\tau} := \frac{X_j(\tau) - X_j(0) - \int_0^\tau \left(-\frac{X_j(s)}{\alpha} ds\right)}{\tau} = \phi_j + \frac{\sigma}{\tau} W_j(\tau). \tag{2}$$

The $(Z_{j,\tau})_\tau$ are estimators of the ϕ_j based on the trajectory $(X_j(t))$. They correspond to the maximum in φ of the conditional likelihood of (1) given $\phi_j = \varphi$. Moreover random variables satisfy $\mathbb{E}[Z_{j,\tau}] = \mathbb{E}[\phi_j]$ and when τ goes to infinity, the noise $\sigma W_j(\tau)/\tau$ goes to zero. This attests the goodness of the estimator. Notice that the $(Z_{j,\tau})_{j=1,\ldots,N}$ are i.i.d. when τ is fixed, with density f_{Z_τ}, due to the independence of $(\phi_j)_{j=1,\ldots,N}$ and $(W_j)_{j=1,\ldots,N}$. These new random variables are available, depending only on the observations and known parameters. Nevertheless, we only have discrete observations of the process. Thus we discretize the integral: $\int_0^\tau X_s ds \approx \delta \sum_{k=1}^{\lfloor \tau/\delta \rfloor} X_{(k-1)\delta}$. The error due to this approximation is studied in sec-

tion 'Discretization' of "Appendix 2". At this point, two strategies materialize which we explain in the following Section.

2.3 Estimation strategies

Let us present the two investigated methods.

Kernel strategy
The first idea is to reduce the noise which appears in formula (2). Indeed, $\mathrm{Var}(\sigma W_j(\tau)/\tau) = \sigma^2/\tau$ leads to focus on the largest τ: $\tau = T$. Moreover, when T is large $Z_{j,T}$ clearly approximates ϕ_j without needing to remove the noise. Then we are able to build a kernel estimator of the density f of the ϕ_j's based on the $Z_{j,T}$ using directly the $Z_{j,T}$ as an approximation of the non-observed random effects ϕ_j. These N random variables are i.i.d. and the resulting kernel estimator is given for all $x \in \mathbb{R}$, by

$$\widehat{f_h}(x) = \frac{1}{N}\sum_{j=1}^{N} K_h(x - Z_{j,T}) \tag{3}$$

where $h > 0$ is a bandwidth, and $K : \mathbb{R} \to \mathbb{R}$ is a $\mathcal{C}^2$ kernel such that

$$\int K(u)du = 1, \quad \|K\|^2 = \int K^2(u)du < +\infty, \quad \int (K''(u))^2 du < +\infty,$$
$$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right). \tag{4}$$

This natural estimator is studied in detail in Sect. 3.

Deconvolution strategy
The other idea is to build an estimator of f using all variables $Z_{j,\tau}$ for different $\tau \in [0, T]$. Recovering f from the observations $(X_1(t), \ldots, X_N(t))_{t\in[0,T]}$ is called the deconvolution problem because the common density of $(Z_{j,\tau})_{j=1,\ldots,N}$ is a convolution product between two densities. Indeed, the two members of the sum (2) are independent when τ is fixed, which implies for all $j = 1, \ldots, N$,

$$f_{Z_\tau}(u) = f \star f_{\frac{\sigma}{\tau}W_1(\tau)}(u).$$

Then the characteristic function of ϕ_j is recoverable from that of Z_τ. Taking the Fourier transform under assumption **(A)** gives the simple product

$$f^*_{Z_\tau}(u) = f^*(u) f^*_{\frac{\sigma}{\tau}W_1(\tau)}(u),$$

with $f^*_{\frac{\sigma}{\tau}W_1(\tau)}(u) = e^{-\frac{u^2\sigma^2}{2\tau}}$. In this particular case the noise is Gaussian and this convolution problem has been investigated in the literature, see Fan (1991), Butucea and Tsybakov (2007) for example. However it has been proven in Carroll and Hall (1988) that the best rates of convergence obtained in this case are logarithmic. This suggests

to improve the deconvolution procedure and this is the reason why we choose not to use previous estimators but to propose a new method, based on repeated observations and new parameters chosen carefully.

We have $f^*(u) = f^*_{Z_\tau}(u)e^{u^2\sigma^2/2\tau}$. Finally the Fourier inversion gives the closed formula, for all $x \in \mathbb{R}$,

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iux} f^*_{Z_\tau}(u)e^{\frac{u^2\sigma^2}{2\tau}}du. \tag{5}$$

Then, we estimate $f^*_{Z_\tau}(u)$ by its empirical estimator $\widehat{f}^*_{Z_\tau}(u) = (1/N)\sum_{j=1}^N e^{iuZ_{j,\tau}}$. However, plugging this in formula (5) involves integrability problems. Indeed the integrability of $\widehat{f}^*_{Z_\tau}(u)e^{u^2\sigma^2/2\tau}$ is not ensured. Therefore, we have to introduce a cut-off. The nonparametric estimation using a deconvolution method in the Gaussian case commonly yields bad speeds of convergence. To improve the rates, an idea of Comte and Samson (2012), for linear mixed models, was to link this cut-off and the time of the process. Comte et al. (2013) link the time of the process τ and the cut-off as follows:

$$\widehat{f}_\tau(x) = \frac{1}{2\pi}\int_{-\sqrt{\tau}}^{\sqrt{\tau}} e^{-iux}\frac{1}{N}\sum_{j=1}^N e^{iuZ_{j,\tau}}e^{\frac{u^2\sigma^2}{2\tau}}du. \tag{6}$$

Then the time τ is chosen by a Goldenshluger and Lepski's method and the final estimator is denoted $\widetilde{f}_{\widetilde{\tau}}$. Nevertheless, when τ is small (which is the case for the real dataset we investigate), the integration domain is not large enough, and the estimators of f are not satisfactory (*cf* an explicit example in Sect. 6). We adapt Comte et al. (2013) estimator to this small T framework. Indeed, to improve the previous estimator, we introduce a new parameter s in the cut-off:

$$\widehat{f}_{s,\tau}(x) = \frac{1}{2\pi}\int_{-s\sqrt{\tau}}^{s\sqrt{\tau}} e^{-iux}\frac{1}{N}\sum_{j=1}^N e^{iuZ_{j,\tau}}e^{\frac{u^2\sigma^2}{2\tau}}du.$$

Then, in order to simplify the theoretical study, we replace $s\sqrt{\tau}$ in the integral by a new parameter m. The resulting estimator $\widetilde{f}_{m,s}$ is defined when $m^2/s^2 \in]0, T]$, by

$$\widetilde{f}_{m,s}(x) = \frac{1}{2\pi}\int_{-m}^{m} e^{-iux}\frac{1}{N}\sum_{j=1}^N e^{iuZ_{j,m^2/s^2}}e^{\frac{u^2\sigma^2 s^2}{2m^2}}du \tag{7}$$

with m and s in two finite sets $\mathcal{M}$ and $\mathcal{S}$ that we will precise later.
In the following we survey in detail the two strategies.

3 Study of the kernel estimator

The kernel estimator given by (3) has been investigated in Comte et al. (2013). First we recall the MISE bound that the kernel estimator $\widehat{f}_h$ satisfies. Then we develop the bandwidth selection procedure we are interested in in this work.

3.1 Risk bound

Let us define $f_h := K_h \star f$, for $h > 0$. We denote for all $p \in \mathbb{R}$, $\|f\|_p = (\int |f(x)|^p dx)^{1/p}$ and for $p = 2$ we still use $\|f\|_2 = \|f\|$. Notice that $\|K_h\| = \|K\|/\sqrt{h}$ and $\|K_h\|_1 = \|K\|_1$. We recall the result proven in Comte et al. (2013) for the MISE.

Proposition 3.1 *Considering estimator $\widehat{f_h}$ given by (3), we have*

$$\mathbb{E}[\|\widehat{f_h} - f\|^2] \leq 2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + \frac{2\sigma^4\|K''\|^2}{3T^2h^5}. \tag{8}$$

The right-hand side of (8) involves three terms, and the middle one is the integrated variance. The integrated bias is $\|\mathbb{E}[\widehat{f_h}] - f\|^2 \leq 2\|f - f_h\|^2 + 2\|\mathbb{E}[\widehat{f_h}] - f_h\|^2$, with

$$\|\mathbb{E}[\widehat{f_h}] - f_h\|^2 \leq \frac{\sigma^4\|K''\|^2}{3T^2h^5}. \tag{9}$$

Therefore, the first term $\|f - f_h\|^2$ is a bias term, which decreases when h decreases. The second term is the term of variance which increases when h decreases. Finally, the third term, also given in (9), is an unusual error term due to the approximation of the ϕ_j's by the $Z_{j,T}$ also increasing when h decreases. We see on this bound that the rate σ^2/T must be small to obtain a small risk.

3.2 Adaptation of the bandwidth

Now that we have at hand a collection of estimators depending on a bandwidth h, we focus on the crucial matter namely how to choose the bandwidth from the data. The best choice of h is the one which minimizes the sum of these three terms. The selection of the bandwidth can be done for example using cross validation, see e.g. the R-function *density* which is commonly used. However, the only theoretical results known for cross-validation procedure are asymptotic and to the best of our knowledge there is no adaptive result on the final estimator. In the present work, we propose to adapt a selection method due to Goldenshluger and Lepski (2011) mentioned before, which provides a data-driven bandwidth for which we provide non-asymptotic theoretical results.

We denote $\mathcal{H}_{N,T}$ the finite set of bandwidths h, to be defined later. The best theoretical choice of the bandwidth is the h which minimizes the bound on the MISE given by (8). Nevertheless, in practice, the bias term is unknown, and this bound has to be estimated.

To choose h adequately, we use a Goldenshluger and Lepski's criterion introduced in Goldenshluger and Lepski (2011). The idea is to estimate $\|f - f_h\|^2$ by the $\mathbb{L}^2$-distance between two estimators defined in (3). But this induces an error which has to be corrected by the variance term. Then the estimator of the bias term is

$$A(h) = \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|^2 - V(h') \right)_+ \tag{10}$$

where

$$\widehat{f}_{h,h'}(x) := K_{h'} \star \widehat{f}_h(x) = \frac{1}{N} \sum_{j=1}^{N} K_{h'} \star K_h(x - Z_{j,T})$$

and V correspond to the two terms of variance

$$V(h) = \kappa_1 \frac{\|K\|_1^2 \|K\|^2}{Nh} + \kappa_2 \frac{\sigma^4 \|K\|_1^2 \|K''\|^2}{T^2 h^5} \tag{11}$$

with κ_1 and κ_2 two numerical positive constants. We will prove that $A(h)$ has the order of the bias term (see Eq. (24)). Finally the bandwidth is selected as follows:

$$\widehat{h} = \underset{h \in \mathcal{H}_{N,T}}{\operatorname{argmin}} \{A(h) + V(h)\} \tag{12}$$

with $\mathcal{H}_{N,T}$ a finite discrete set of bandwidths h such that $h > 0$, $\frac{1}{Nh} \leq 1$, $\frac{1}{h^5 T^2} \leq 1$ and $\text{Card}(\mathcal{H}_{N,T}) \leq N$. It must be chosen such that when N goes to infinity, for all $c \in \mathcal{C}$ $\sum_{h \in \mathcal{H}_{N,T}} h^{-1/2} e^{-c/\sqrt{h}} \leq S(c)$ with $S(c)$ a positive constant depending on c. For example notice that taking $\mathcal{H}_{N,T} = \{1/k^2, k = 1, \ldots, \sqrt{N}\}$, the sum $\sum_{h \in \mathcal{H}_{N,T}} h^{-1/2} e^{-c/\sqrt{h}} \leq \sum_{k \geq 1} k e^{-ck}$ converges, which is a necessary condition for the proof.
Then we can prove the following Theorem.

Theorem 3.1 *Consider estimator $\widehat{f}_h$ given by (3) with $h \in \mathcal{H}_{N,T}$. Then, there exist two penalty constants κ_1, κ_2 such that*

$$\mathbb{E}[\|\widehat{f}_{\widehat{h}} - f\|^2] \leq C_1 \inf_{h \in \mathcal{H}_{N,T}} \left\{ \|f - f_h\|^2 + V(h) \right\} + \frac{C_2}{N}$$

where C_1, C_2 are two positive constants such that $C_1 = \max(7, 30\|K\|_1^2 + 6)$ and C_2 depends on $\|f\|$, $\|K\|$, $\|K\|_1$, $\|K\|_{4/3}$.

The theoretical study gives $\kappa_1 \geq \max(40/\|K\|_1^2, 40)$ and $\kappa_2 \geq \max(10/3, 10/(3\|K\|_1^2))$. But in practice these two constants are calibrated from a simulation study (and always smaller than the theoretical ones). Theorem 3.1 is an oracle inequality: the bias variance compromise is automatically obtained and in a data-driven and non-asymptotic way.

This strategy requires large T as we assume $1/h^5 \leq T^2$. The error implied by the discrete observations and the use of Riemann sums to compute the $Z_{j,T}$ are detailed in Comte et al. (2013).

4 Study of the deconvolution estimator

4.1 Risk bound

Let us emphasize that the estimator $\widetilde{f}_{m,s}$ given by (7) depends on two parameters which have to be selected from the data. This is not usual in the deconvolution setting, where only one cut-off parameter is often introduced. The selection of these two parameters (m, s) among the finite sets $\mathcal{M}$, $\mathcal{S}$ is thus more difficult. It is even more challenging here because the cut-off m appears both in the integral and in the integrand. But this will induce gains in the rates of the estimators. Before proposing a selection method of (m, s) we start by evaluating the quality of the estimator with the mean integrated squared error (MISE):

$$\mathbb{E}\left[\|\widetilde{f}_{m,s} - f\|^2\right] = \|f - \mathbb{E}[\widetilde{f}_{m,s}]\|^2 + \mathbb{E}\left[\|\widetilde{f}_{m,s} - \mathbb{E}[\widetilde{f}_{m,s}]\|^2\right].$$

In Proposition 4.1 we prove that $\mathbb{E}[\widetilde{f}_{m,s}] = f_m$ where f_m is defined by its Fourier transform

$$f_m^* := f^* \mathbf{1}_{[-m,m]}.$$

It means that the bias does not depend on s. We obtain the following bound on the MISE of $\widetilde{f}_{m,s}$.

Proposition 4.1 *Under (A), the estimator $\widetilde{f}_{m,s}$ given by (7) is an unbiased estimator of f_m and we have*

$$\mathbb{E}\left[\|\widetilde{f}_{m,s} - f\|^2\right] \leq \|f_m - f\|^2 + \frac{m}{\pi N}\int_0^1 e^{\sigma^2 s^2 v^2} dv. \tag{13}$$

The proofs are relegated to Sect. 8. Let us look at the risk bound. The first term of the bound (13) is the bias term. It represents the error resulting from estimating f by f_m and it decreases when m increases, indeed:

$$\|f_m - f\|^2 = \frac{1}{2\pi}\int_{|u|\geq m} |f^*(u)|^2 du.$$

The second term is the variance term, and it increases with m and s. One can notice that it is bounded as soon as s is bounded and $m \leq N$.
We specify the two sets $\mathcal{M}$ and $\mathcal{S}$. We notice that the quality of the estimate in the Fourier domain is good on an interval around zero with length related with σ. The chosen set for s is

$$\mathcal{S} := \left\{ s_l = \frac{1}{2^l}\frac{2}{\sigma},\ l = 0, \ldots, P \right\}.$$

Notice that for $s_l \in \mathcal{S}$, $1/2^{P-1} \leq \sigma s_l \leq 2$. Moreover with this chosen collection $\mathcal{S}$, the order of the variance term is m/N. With the idea that m^2/s^2 is homogeneous to a time, we choose m in the finite collection:

$$\mathcal{M} := \left\{ m = \frac{\sqrt{k\Delta}}{\sigma},\ k \in \mathbb{N}^*,\ 0 < m \leq N \right\}$$

with $0 < \Delta < 1$ a small step to be fixed. The collection of couples of parameters is

$$\mathcal{C} := \{(m, s) \in \mathcal{M} \times \mathcal{S}, \quad m^2/s^2 \leq T\}.$$

The final estimator is the estimator from the collection $\mathcal{C}$ which achieves the bias-variance compromise. Choosing the final estimator is not an easy task except if we know the regularity of f. Indeed, let us assume that f is in the Sobolev ball with regularity parameter b, i.e. f belongs to the set defined by

$$\mathcal{A}_b(L) = \left\{ f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \int_{\mathbb{R}} |f^*(x)|^2 (1 + x^2)^b dx \leq L \right\}$$

with $b > 0$, $L > 0$. For example the standard normal distribution is in a space $\mathcal{A}_b(L)$ for some L and for all $b > 0$, an exponential distribution is in some $\mathcal{A}_b(L)$ for $b < 1/2$ or more generally a Gamma distribution with shape parameter k is in some $\mathcal{A}_b(L)$ for $b < (k - 1/2)$. Thus when $f \in \mathcal{A}_b(L)$, the bias term satisfies:

$$\| f_m - f \|^2 = \frac{1}{2\pi} \int_{|u| \geq m} |f^*(u)|^2 du \leq \frac{L}{2\pi} m^{-2b}.$$

Consequently, the $\mathbb{L}^2$-risk of $\widetilde{f}_{m,s}$ is bounded by

$$\mathbb{E}\left[\| \widetilde{f}_{m,s} - f \|^2 \right] \leq \frac{L}{2\pi} m^{-2b} + \frac{m}{\pi N} e^{\sigma^2 s^2}.$$

Therefore, the best theoretical choice of s is s_P the smallest s in our collection, and

$$m = m^* = K_b N^{\frac{1}{(2b+1)}}$$

with $K_b = (bL \exp(-1/(2^{2(P-1)}))^{1/(2b+1)}$. Then we obtain the following asymptotic result.

Corollary 4.1 *If $f \in \mathcal{A}_b(L)$, and if we choose $s = s_P$ and $m = m^*$, there exists a constant K depending on b, L, P, such that*

$$\mathbb{E}\left[\| \widetilde{f}_{m^*,s_P} - f \|^2 \right] \leq K N^{-\frac{2b}{2b+1}}.$$

The order of the risk in this case is $N^{-2b/(2b+1)}$ for a large N, and it is the nonparametric estimation rate of convergence obtained when the observations are N realizations of the variable of interest. Nevertheless, it is not easy to see that $(m, s) \in \mathcal{C}$ and this choice is theoretical because it depends on the regularity b of f, which is unknown. The next section provides a data-driven method to select (m, s).

4.2 Selection of the final estimator

In this Section we deal with the choice of the best estimator among the available collection of $\widetilde{f}_{m,s}$. In the previous work Comte et al. (2013), s was fixed to $s = 1$ and m was selected. But we saw empirically that it did not work in the setting corresponding to the data. This is why we experimented different values for s. But then we did not find any reliable criterion to select m for any given s. On the contrary, if we look at the bound and try to select s first, we just get $s = 0$, which is not of interest if we are looking to improve the estimator through s in particular. This implies to select the couple (m, s) minimizing the MISE and realizing the compromise between the two terms, in a data-driven way. This is a crucial issue. Indeed, the role of the two parameters is not the same. Thus we propose a new criterion adapted from the Goldenshluger and Lepski (2011) method.

The idea is to select the couple which minimizes the MISE: $\mathbb{E}\left[\|\widetilde{f}_{m,s} - f\|^2\right]$. As it is unknown, we have to find a computable approximation of this quantity. We define the best couple (m, s) as the one minimizing a criterion defined as the sum of a squared bias term and a variance term called penalty. We define the penalty function, which has the same order as the bound on the variance term:

$$\mathrm{pen}(m, s) = \kappa \frac{m}{N} e^{\sigma^2 s^2},$$

where κ is a numerical constant to be calibrated. Note that for $m \in \mathcal{M}$ and $s \in \mathcal{S}$, the penalty function is bounded.

To estimate the bias term, we generalize Goldenshluger and Lepski's criterion for a two-dimensional index. The method is inspired by the ideas developed for kernel estimators by Goldenshluger and Lepski (2011) and adapted to model selection in one dimension in Comte and Johannes (2012) and in two dimensions by Chagny (2013). The idea is to estimate $\|f - f_m\|^2$ by the $\mathbb{L}^2$-distance between two estimators defined in (7). But this induces a bias which has to be corrected by the penalty function. We consider the following estimator of the bias, with $(m', s') \wedge (m, s) := (m' \wedge m, s' \wedge s)$,

$$\Gamma_{m,s} = \max_{(m',s') \in \mathcal{C}} \left(\|\widetilde{f}_{m',s'} - \widetilde{f}_{(m',s') \wedge (m,s)}\|^2 - \mathrm{pen}(m', s') \right)_+ \tag{14}$$

for $(m, s) \in \mathcal{C}$. Finally the selected couple is:

$$(\widetilde{m}, \widetilde{s}) = \arg \min_{(m,s) \in \mathcal{C}} \{\Gamma_{m,s} + \mathrm{pen}(m, s)\}. \tag{15}$$

We are now able to obtain the following result.

Theorem 4.1 *Under (A), consider the estimator $\widetilde{f}_{\widetilde{m},\widetilde{s}}$ given by (7) and (15). There exists κ_0 a numerical constant such that, for all penalty constants $\kappa \geq \kappa_0$,*

$$\mathbb{E}\left[\|\widetilde{f}_{\widetilde{m},\widetilde{s}} - f\|^2\right] \leq C \inf_{(m,s)\in\mathcal{C}} \left\{\|f - f_m\|^2 + pen(m,s)\right\} + \frac{C'(P+1)}{N} \tag{16}$$

where $C > 0$ is a numerical constant and C' is a constant depending on $\|f\|$, σ, Δ, and $P+1$ the cardinality of $\mathcal{S}$.

The key of the proof is to prove that

$$\mathbb{E}[\Gamma_{m,s}] \leq 18\|f - f_m\|^2 + \frac{C'(P+1)}{N}$$

(see the proof in Sect. 8.3.) Inequality (16) means that $\widetilde{f}_{\widetilde{m},\widetilde{s}}$ automatically makes the bias-variance trade-off. Moreover, our result is of non asymptotic nature w.r.t. N.

One should notice that this new parameter s generalizes the results of Comte et al. (2013) even if T is large. We choose the two parameters in an adaptive way, thus this gives more flexibility in the choice of the estimator.

It follows from the proof that $\kappa_0 = 24$ would suit. But in practice, values obtained from the theory are generally too large and the constant is calibrated by simulations. Once chosen, it remains fixed for all simulation experiments. Besides the cardinality P of the set $\mathcal{S}$ is chosen small in practice ($P = 3$ or 10 for example).

In the section 'Discretization' of "Appendix 2' we investigate the error implied by the discrete observations and thus of the discretization of $Z_{j,\tau}$ given by (2).

5 Simulation study

In the following section we compare on simulations the two procedures we compute $\widehat{f}_{\widehat{h}}$ and $\widetilde{f}_{\widetilde{m},\widetilde{s}}$, the estimator of Comte et al. (2013) $\widetilde{f}_{\widetilde{\tau}}$ and we compare our bandwidth selection method with the estimator from the `R`-function `density` with the cross-validation argument `bw = ucv`, on the $Z_{j,T}$, $j = 1, \ldots, N$, we denote it $\widehat{f}_{cv}$.

We simulate data by computing the exact solutions of (1) given by Itô's formula,

$$X_j(t) = X_j(0)e^{-t/\alpha} + \phi_j\alpha(1 - e^{-t/\alpha}) + \sigma e^{-t/\alpha}\int_0^t e^{s/\alpha} dW_j(s) \tag{17}$$

at discrete times $t_k \in \mathcal{T} := \{k\delta, k \in \{0, \ldots, J\}, J\delta = T\}$. For the simulation study, we have to fix $N, \delta, T, \sigma, \alpha$, and the density f. We take $\sigma = 0.0135, 0.05, 1$ and $\sigma = 0.05$, $\alpha = 0.039, 1, 39$. For the time T, we choose $T = 0.3, 10, 50, 100, 300$ with different values of δ the discrete time step at which observations are recorded. The value of J, the number of observations for one trajectory ranges from 150 to 5000 for Table 1 and is fixed to $J = 2000$ for Table 2. All these parameter values are chosen in relation with the parameters of the real dataset. In this study we hope to highlight the influence of each one. For f, we investigate four different distributions:

- Gaussian distribution $\mathcal{N}(0.278, (0.041)^2)$
- Gamma distribution $\Gamma(1.88, 0.148)$
- mixed Gaussian distribution $0.3\mathcal{N}(0, (0.02)^2) + 0.7\mathcal{N}(1, (0.02)^2)$ (denoted M-Gaussian)
- mixed Gamma distribution $0.4\Gamma(3, 0.08)+0.6\Gamma(30, 0.035)$ (denoted M-Gamma) where we write $\Gamma(k, \theta)$ with k the scale parameter and θ the shape.

First, we implement the two collections of estimators: $\widehat{f}_{\widehat{h}}$ and $\widetilde{f}_{\widetilde{m},\widetilde{s}}$. We begin by computing the random variables used by both estimators: $Z_{j,\tau}$ given by (2), with Riemann sums approximations (see section 'Discretization' of "Appendix 2' for details). For the deconvolution estimator given by (7) we also use Riemann sums to compute the integral. For the collection of m, we choose $\Delta = 0.08$ and δ changes. Furthermore, for the kernel estimator given by (3), we choose a Gaussian kernel: $K(u) = (1/\sqrt{2\pi})e^{-u^2/2}$. In this case $\|K\|_1 = 1$, $\|K\|_2^2 = 1/(2\sqrt{\pi})$, $\|K''\|_2^2 = (1 + 1/\sqrt{2})/(\sqrt{2\pi})$.

Then, the selected bandwidth $\widehat{h}$ is given by Eq. (12). Note that for all $(h, h') \in \mathcal{H}^2$,

$$K_{h'} \star K_h(x) = \frac{1}{\sqrt{2\pi}\sqrt{h'^2 + h^2}} e^{-x^2/[2(h'^2+h^2)]}.$$

We use this relation to compute the $\widehat{f}_{h,h'}$.

Secondly, we have to calibrate the penalty constants: κ_1, κ_2 for the kernel estimator and κ for the deconvolution estimator. Classically, the constants are fixed thanks to preliminary simulation experiments. Different functions f have been investigated with different parameter values, and a large number of replications. Comparing the MISE obtained as functions of the constants κ_1, κ_2 and κ yield to select values making a good compromise over all experiments. Finally we choose $\kappa_1 = 1$, $\kappa_2 = 0.0001$ and $\kappa = 0.3$. A recent work Lacour and Massart (2016) proposes to change the calibration constants in the variance term $V(h)$: taking κ in the term $\Gamma_{(m,s)}$ (14) and 2κ for the second $V(h)$ in the selection criterion (12). It has been done in practice for the kernel estimator. We notice that this strategy produces very good results in practice, better than choosing the same κ for the two apparitions of term $V(h)$. On Fig. 1 25 estimators $\widehat{f}_{\widehat{h}}$ are plotted and on Fig. 2 25 estimators $\widetilde{f}_{\widetilde{m},\widetilde{s}}$, for the 4 investigated densities f . The batch of estimators is close to the estimated density.

In order to evaluate the performances of each estimator on the different designs, we compare their empirical MISE computed from 100 simulated data sets.

Table 1 summarises the results for different parameters values. It shows the bad performances of the estimator of Comte et al. (2013) $\widetilde{f}_{\widetilde{\tau}}$ when T is small compared to $\widetilde{f}_{\widetilde{m},\widetilde{s}}$. It performs clearly better when T is increasing. Besides we notice that both kernel estimators have good results. Nevertheless these results are satisfying because it appears that our estimator $\widehat{f}_{\widehat{h}}$ fits slightly better the true density than $\widehat{f}_{cv}$. The computation time is close for both selection method. We show the results for different values of α do not seem to influence the quality of estimators (while the selected h, m, s are very different). During the simulation study, we noticed that the parameter α is important is the sense that when the value of α does not have the same order as the values of ϕ, the estimation is harder. Except when $T = 300$ the ratio signal noise

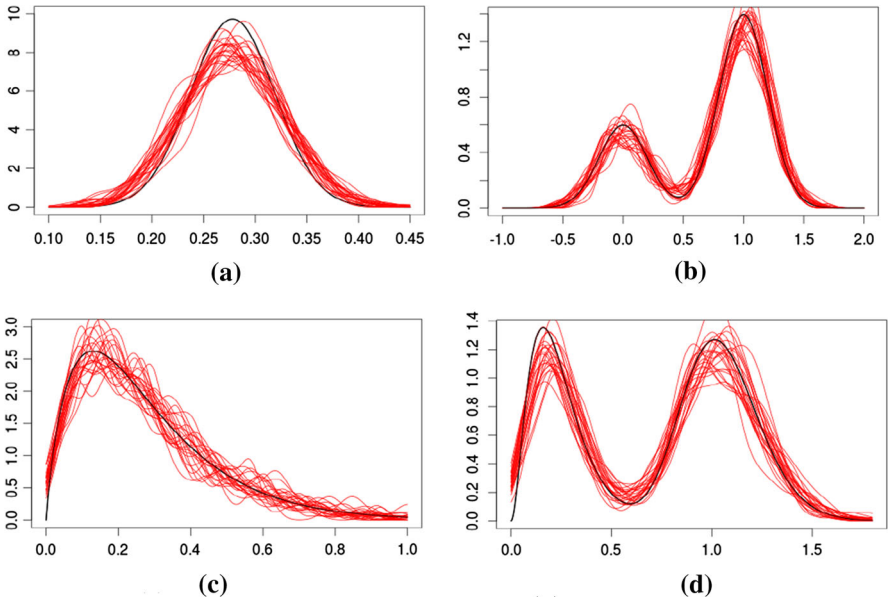

Fig. 1 Simulated data. In *plain red* (*black*) 25 estimators $\widehat{f_{\widehat{h}}}$ with parameters: $N = 240$, $T = 0.3$, $\delta = 0.00015$, $\sigma = 0.0135$, $\alpha = 0.039$ and the true density f in *plain bold black line*. **a** f Gaussian, **b** f mixed Gaussian density, **c** f Gamma and **d** f mixed Gamma density

which is the standard deviation of the random effect divided by σ is larger than one thus the settings are favourables. But for both Gamma and mixed Gamma cases, the standard deviations are respectively 0.2 and 0.15, which is not small compared to σ or to their mean. The mixed Gamma case is difficult for the nonparametric estimation: Fig. 3 illustrates the performances of estimators for this choice. Finally, it is interesting to note that when the standard deviation of the random effect of interest has a larger variance, the density estimation is easier, which is the case of the chosen Gamma density for example compared with the Gaussian case.

In the following as the two kernel estimators seem very close, we only show the results for $\widehat{f_{\widehat{h}}}$ which is of interest here. Besides we no longer investigate the previous estimator $f_{\widetilde{\tau}}$ in light of Table 1 for $T < N$.

We further compare $\widehat{f_{\widehat{h}}}$ and $\widetilde{f}_{\widetilde{m},\widetilde{s}}$. The two estimators seem close to the true density on graphs, see Figs. 1 and 2. In Table 2 on the MISE it is clear that the kernel estimator is the best. Furthermore, we can point out some differences. The first row of the Table corresponds to simulations with the parameters of the real dataset. In the first column, the Gaussian case, the MISE are 10 times larger than the ones for other cases. This can be easily explained: the values of the estimated density are 10 times larger than others. Nevertheless, on lines 3 and 4 for the Gaussian case, the MISE are very large. This is due to the bad estimation of the ϕ_j by the $Z_{j,T}$ with $\sigma = 0.05$ and $T = 0.3$.[1]

[1] We insist that this bad estimation is not due to the fact that the noise is Gaussian. Indeed even if Fan (1991) proves the rates to be logarithmic in that case, the rates are improved and can be polynomial when the density under estimation is of the same type as the noise (see Lacour 2006; Comte et al. 2006).

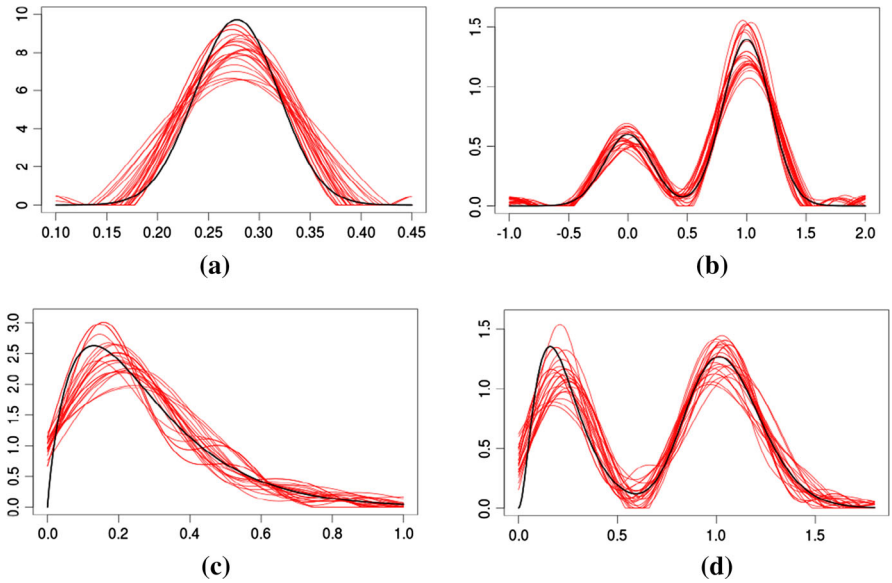

Fig. 2 Simulated data. In *plain red* (*black*) 25 estimators $\widetilde{f}_{\widetilde{m},\widetilde{s}}$ with parameters: $N = 240$, $T = 0.3$, $\delta = 0.00015$, $\sigma = 0.0135$, $\alpha = 0.039$ and the true density f in *bold plain black line*. **a** f Gaussian, **b** f mixed Gaussian, **c** f Gamma and **d** f mixed Gamma

Table 1 Empirical MISE computed from 100 simulated data sets, with $N = 200$, with various T, δ, σ and α for two kernel estimators $\widehat{f}_{cv}$, $\widehat{f}_{\widehat{h}}$ and two deconvolution estimators $\widehat{f}_{\widetilde{\tau}}$ and $\widetilde{f}_{\widetilde{m},\widetilde{s}}$

	$\widehat{f}_{cv}$	$\widehat{f}_{\widehat{h}}$	$\widehat{f}_{\widetilde{\tau}}$	$\widetilde{f}_{\widetilde{m},\widetilde{s}}$
Case f Gamma				
$T = 0.3\,\delta = 0.0005\,\sigma = 0.0135\,\alpha = 0.039$	0.043	0.037	1.547	0.071
$T = 300\,\delta = 0.5\,\sigma = 0.5\,\alpha = 39$	0.048	0.039	0.049	0.055
$T = 50\,\delta = 0.1\,\sigma = 0.05\,\alpha = 1$	0.042	0.039	0.218	0.050
Case f mixed Gamma				
$T = 0.3\,\delta = 0.0005\,\sigma = 0.0135\,\alpha = 0.039$	0.033	0.030	0.712	0.035
$T = 300\,\delta = 0.5\,\sigma = 0.5\,\alpha = 39$	0.032	0.030	0.035	0.043
$T = 50\,\delta = 0.1\,\sigma = 0.05\,\alpha = 1$	0.033	0.031	0.145	0.043

The quality of the estimation is significantly better when we try a $\mathcal{N}(0.278, 0.2)$ (0.2 is the variance of the mixed Gaussian density we implement for example). In general one can notice that when σ is larger than the standard deviation of the density of the random effects f, the estimation is less precise, which is coherent in term of ratio signal noise.

Table 2 shows that if T increases, it improves the results for $\sigma = 0.05$, compare cases 2 and 5 with 4 and 7 for example. If J is large enough, meaning if δ is small enough (which is the case even for $J = 150$ when $T = 0.3$), the deconvolution estimator fits well the density. In practice, when T increases, the selected value of s

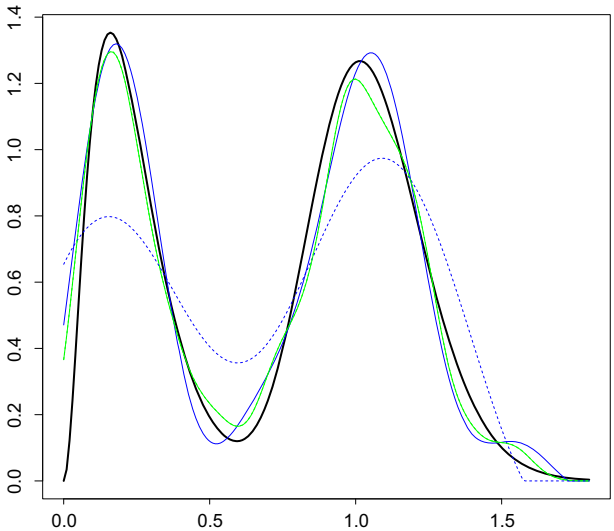

Fig. 3 Simulated data. In *bold plain black curve* is the true density f mixed Gamma, the estimator $\widehat{f}_{cv}$ estimator $\widehat{f}_{\widehat{h}}$ are superposed is in *plain green* (*grey*), estimator $\widetilde{f}_{\widetilde{\tau}}$ is *dotted black* (*blue*) and estimator $\widetilde{f}_{\widetilde{m},\widetilde{s}}$ is *plain blue* (*black*), with $N = 200$, $T = 50$, $\delta = 0.05$, $\sigma = 0.05$, $\alpha = 1$

decreases, which could have been predicted. The results are still satisfying for large T. For the kernel estimator, although the theoretical condition $1/h^5 < T^2$ is not satisfied, the numerical results are good.

Another point is, as expected, that the larger N is, the better the estimators $\widehat{f}_{\widehat{h}}$ and $\widetilde{f}_{\widetilde{m},\widetilde{s}}$. We can refer to Comte et al. (2013) for a study with different values for N. It highlights the influence of N when the estimated density has two modes; for example with $N = 50$ the estimation is clearly less precise than for $N = 200$.

A main difference between our two estimators $\widehat{f}_{\widehat{h}}$ and $\widetilde{f}_{\widetilde{m},\widetilde{s}}$ is the computation time: a few seconds for the first one and ten minutes for the second one. Thus the kernel estimator with the method of bandwidth selection is very efficient, especially in the case of multi-modal densities, and performs often better than the deconvolution one.

6 Application to neuronal data

6.1 Dataset

We describe quickly the data but we refer to Yu et al. (2004), Lansky et al. (2006) for example for details on data acquisition. The data are intracellular measurements of the membrane potential in volts along time, for one single neuron of a pig between the spikes. This is the depolarization phase. In this neuronal context, between the $(j-1)$th and the jth spike, the depolarization of the membrane potential receiving a random input can be described by the Ornstein–Uhlenbeck model with one random effect (1). The spikes are not intrinsic to the model but are generated when the voltage reaches for the first time a certain threshold S, then the process is reset to a fixed

Table 2 Empirical MISE computed from 100 simulated data sets, with $N = 240$, $\alpha = 0.039$ and various T, δ, σ for the kernel estimator $\widehat{f}_{\widehat{h}}$ and the deconvolution estimator $\widetilde{f}_{\widetilde{m},\widetilde{s}}$

Case	T	δ	σ	Estimator	f			
					Gaussian	Gamma	M-Gaussian	M-Gamma
1	0.3	0.0002	0.0135	$\widehat{f}_{\widehat{h}}$	0.254	0.037	0.016	0.024
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.381	0.072	0.021	0.038
2	0.3	0.002	0.0135	$\widehat{f}_{\widehat{h}}$	0.270	0.032	0.016	0.024
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	1.584	0.071	0.019	0.038
3	0.3	0.0002	0.05	$\widehat{f}_{\widehat{h}}$	2.544	0.098	0.026	0.054
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	3.402	0.231	0.078	0.226
4	0.3	0.002	0.05	$\widehat{f}_{\widehat{h}}$	2.524	0.092	0.028	0.053
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	3.269	0.154	0.029	0.232
5	10	0.002	0.0135	$\widehat{f}_{\widehat{h}}$	0.091	0.031	0.017	0.024
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.080	0.037	0.035	0.052
6	10	0.02	0.0135	$\widehat{f}_{\widehat{h}}$	0.085	0.033	0.014	0.023
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.151	0.048	0.033	0.034
7	10	0.002	0.05	$\widehat{f}_{\widehat{h}}$	0.111	0.031	0.016	0.025
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.687	0.061	0.016	0.034
8	10	0.02	0.05	$\widehat{f}_{\widehat{h}}$	0.121	0.035	0.014	0.025
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.181	0.053	0.023	0.040
9	100	0.2	0.0135	$\widehat{f}_{\widehat{h}}$	0.106	0.032	0.014	0.024
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.123	0.062	0.091	0.046

initial voltage. Thus each trajectory is observed on an interval $[0, T_j]$ where $T_j = \inf\{t > 0, X_j(t) \geq S\}$. The initial voltage (the value following a spike) is assumed to be equal to the resting potential. The present dataset has been normalised to obtain N trajectories which begin in zero: $x_j = 0$.

The positive constant parameter α is called the time constant of the neuron (the coefficient of decay in the exponential, when there is no noise), which is intrinsic to the neuron and fixed to $\alpha = 0.039$ [s] (Lansky et al. 2006). The diffusion coefficient σ [V/$\sqrt{\text{s}}$] has been estimated using the estimator $\widehat{\sigma^2} = (1/N)\sum_{j=1}^{N}\left((1/J)\sum_{k=1}^{J}((X_j(\delta(k+1)) - X_j(\delta k))^2/\delta\right)$. We obtain $\hat{\sigma} = 0.0135$, which is the same value as that used in Picchini et al. (2008). The ϕ_j represents the local average input that the neuron receives during the jth inter-spike interval. We assume that ϕ_j changes from one trajectory to another because of other neurons or the influence of the environment, for example. So parameters ϕ and σ characterize the input, while α, x_j (the resting potential) and S (the firing threshold) describe the neuron irrespectively of the incoming signal (Picchini et al. 2008).

Data are composed of $N = 312$ inter-spike trajectories. For each interval $[0, T_j]$ the time step is the same: $\delta = 0.00015$ [s]. We decide to keep only realizations with more than 2000 observations ($T_j/\delta \geq 2000$). Finally we have $N = 240$ realizations with $J = 2000$ observations and for $j = 1, \ldots, N$, $T = T_j = 0.3$ [s]. Also the data

are normalized in order to begin with zero at the initial time. The study of the units of measurement can highlight the collections given in Sect. 4. One can notice that the unit of measurement of v in the integrand must be $[\mathtt{s}/\mathtt{V}]$ (same unit as $1/Z_{j,\tau}$) such that the exponential terms are without unit. The unit of s is $[\sqrt{\mathtt{s}}/\mathtt{V}]$, and the choice of $\mathcal{M}$ with the same unit as v seems natural.

It is interesting to note that the normality of the $Z_{j,T}$ is rejected by Shapiro and Wilk test (p-value 10^{-7}) and Kolmogorov–Smirnov test (p-value 10^{-3}). This suggests that the ϕ_j's are not Gaussian. Thus we want to estimate nonparametrically their density. In the following we compare our results to the estimation obtained in Picchini et al. (2010) under the parametric Gaussian assumption.

6.2 Comparison of estimators

The estimation of the density f obtained by Picchini et al. (2010) under the Gaussian assumption on these data, with $\alpha = 0.039$ are $\mathcal{N}(\mu = 0.278, \eta^2 = (0.041)^2)$. Using a maximum-likelihood estimator on the $(Z_{j,T})$'s we obtain for the mean 0.270 and for the standard deviation 0.046. We notice that these two estimations are close to that of Picchini et al. (2010). We use our two nonparametric estimators to see how close to a Gaussian density they are.

On Fig. 4 we represent both estimators $\widehat{f}_{\widehat{h}}$ and $\widetilde{f}_{\widetilde{m},\widetilde{s}}$ applied on the real data and the density $\mathcal{N}(\mu, \eta^2)$. The two estimations are close, and close to the estimation of Picchini et al. (2010). However, it is also legitimate to think about a Gamma distribution to model the random parameters ϕ_j's because, as the data have been normalised, the estimated random effects are positive. Then it seems reasonable to use a non-negative random variable to model this local average input. Thus, a Gamma distribution may seem more appropriate than a Gaussian distribution, even if the chosen Gaussian has small probability to be negative. We look for the Gamma distribution which has for mean $\mu = 0.278$ and for standard deviation $\eta = 0.041$. This corresponds to a Gamma distribution with the shape parameter 46.3 and the scale parameter 0.006. We notice the similarity between the previous Gaussian curve and the new one. Thus this distribution seems also suitable to fit the distribution of the ϕ_j's, as Fig. 4 shows.

The Gaussian assumption is strong and leads to parametric tractable models. The present work confirms that this approximation is acceptable. However, the nonparametric estimation gives a density for the ϕ_j's that can be used to simulate the random effect and could be closer to the true one.

Notice that, as mentioned in introduction, Comte et al. (2013)'s estimator cannot handle small values of T while our new proposals are successful in such case. Let us precise this point. If the number of points is large enough, the variable $Z_{j,T}$ can be approximate:

$$Z_{j,T} \approx \left(X_j(T) - X_j(0) + \frac{\delta}{\alpha} \sum_{l=1}^{J} X((l-1)\delta) \right) \frac{1}{T}.$$

It is unchanged when we change the units: $\mathtt{V}, \mathtt{s}$ to $\mathtt{mV}, \mathtt{ms}$ thus $T = 0.3$ to $T = 300$. However, the deconvolution estimator of Comte et al. (2013) for $\tau = T$ is changing when the value of T is changing. In fact, the estimator is

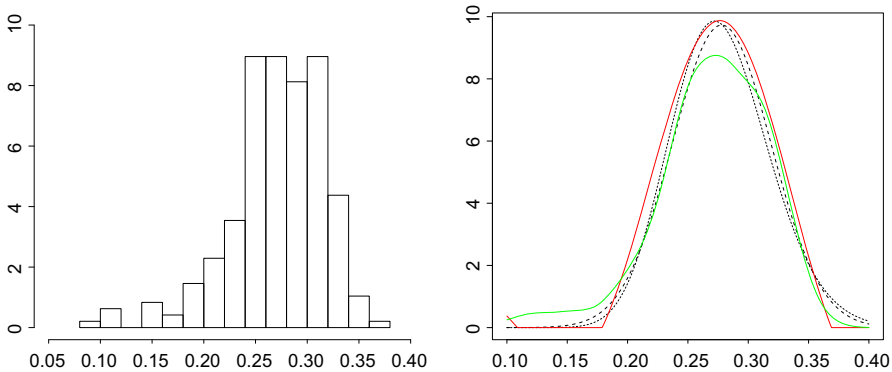

Fig. 4 Real data. In *bold green* (*grey*) estimator $\widehat{f}_{\widehat{h}}$, in *bold red* (*black*) $\widetilde{f}_{\widetilde{m},\widetilde{s}}$, the *black dotted* and *bold line* the density $\mathcal{N}(\mu, \eta^2)$ from Picchini et al. (2010) and the *black dotted thin line* the density $\Gamma(46.3, 0.006)$

$$\widehat{f}_T(x) = \frac{1}{2\pi}\int_{-\sqrt{T}}^{\sqrt{T}} e^{-iux}\frac{1}{N}\sum_{j=1}^{N} e^{iuZ_{j,T}} e^{\frac{u^2\sigma^2}{2T}} du$$
$$\approx \frac{1}{2\pi}\sum_{k=1}^{n_{\text{pas}}}\left((u_{k+1}-u_k)e^{-iu_kx}\frac{1}{N}\sum_{j=1}^{N} e^{iu_kZ_{j,T}} e^{\frac{u_k^2\sigma^2}{2T}}\right)$$

and according to $T = 0.3$ or $T = 300$ the values of u and thus the interval of integration and the three exponential terms are changing. Finally estimator $\widehat{f}_T$ is changing with the units. And the interval of integration is not large enough in the case $T = 0.3$ to give a good estimation.

To solve this problem we have proposed a new estimator $\widetilde{f}_{\widetilde{m},\widetilde{s}}$ to allow the user to deal with data with the units he/she wants and to not oblige him/her to change it.

But one can wonder if the new estimators are robust when increasing T. Indeed, our method works for larger T. Precisely changing volts in millivolts and seconds in milliseconds implies $T = 300$, $\sigma = 0.426$, and on simulated data we adequately reconstruct the shape of the density.

7 Discussion

In this work we study a stochastic differential Ornstein–Uhlenbeck mixed-effects model. We propose two estimators of the density of the random effect. Both estimators are not very sensitive to the effect of the time of observation T. Indeed the kernel strategy corresponds to a context with large T while we built a deconvolution estimator especially for small values of T. Both are data-driven and satisfy an oracle-type inequality. According to the numerical study, the kernel estimator seems to be the efficient one: the numerical results are convincing and close to the ones obtained by cross-validation. Besides we provide non-asymptotic theoretical results. Furthermore we study neuronal data with nonparametric estimation strategy. Instead of making any parametric assumptions for the random effect distribution, we build an estimator of

its density. Future work based on this estimation could be more precise and closer to the real neuronal data. To complete the study, the method for different times of observation T_j could be settled up. Besides, some goodness-of-fit tests could be produced; we refer to Bissantz et al. (2007) who construct confidence bands for an estimator of f in the ordinary smooth deconvolution problem.

The model can be completed by adding another random effect: the time constant of the neuron. Picchini and Ditlevsen (2011) have investigated this model in a parametric way and Dion and Genon-Catalot (2015) in a nonparametric way. A recent work Delattre et al. (2016) assumes that the density of the random effect is a Gaussian mixture and uses data clustering method, which is an interesting approach for the data described in the present paper. The question of a random effect in the diffusion coefficient also is open (see Delattre et al. 2015). Moreover, a model with a drift $b(X_j(t)) + \phi_j$, where b is a known function, can be treated with the same method. However, dealing with a diffusion $\sigma(X_j(t))$ where σ is a known function is a more complex problem.

Acknowledgments The author would like to thank Fabienne Comte and Adeline Samson for very useful discussions and advice.

8 Appendix 1: Proofs

8.1 Proof of Theorem 3.1

Given $h \in \mathcal{H}_{N,T}$, we denote:

$$V(h) = \kappa_1 \frac{\|K\|_1^2 \|K\|^2}{Nh} + \kappa_2 \frac{\sigma^4 \|K\|_1^2 \|K''\|^2}{T^2 h^5} =: V_1(h) + V_2(h).$$

Using the definition of $A(h)$ and of $\widehat{h}$ we obtain

$$\begin{aligned}
\|\widehat{f}_{\widehat{h}} - f\|^2 &\leq 3\|\widehat{f}_{\widehat{h}} - \widehat{f}_{h,\widehat{h}}\|^2 + 3\|\widehat{f}_{h,\widehat{h}} - \widehat{f}_h\|^2 + 3\|\widehat{f}_h - f\|^2 \\
&\leq 3\left(A(h) + V(\widehat{h})\right) + 3\left(A(\widehat{h}) + V(h)\right) + 3\|\widehat{f}_h - f\|^2 \\
&\leq 6A(h) + 6V(h) + 3\|\widehat{f}_h - f\|^2.
\end{aligned}$$

Thus,

$$\mathbb{E}[\|\widehat{f}_{\widehat{h}} - f\|^2] \leq 6\mathbb{E}[A(h)] + 6V(h) + 3\mathbb{E}[\|\widehat{f}_h - f\|^2],$$

hence, we only have to study the term $\mathbb{E}[A(h)]$. We can decompose $\|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|^2$ as follows:

$$\begin{aligned}
\|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|^2 \leq 5\|\widehat{f}_{h,h'} - \mathbb{E}[\widehat{f}_{h,h'}]\|^2 + 5\|\mathbb{E}[\widehat{f}_{h,h'}] - f_{h,h'}\|^2 + 5\|f_{h,h'} \\
- f_{h'}\|^2 + 5\|f_{h'} - \mathbb{E}[\widehat{f}_{h'}]\|^2 + 5\|\mathbb{E}[\widehat{f}_{h'}] - \widehat{f}_{h'}\|^2
\end{aligned}$$

thus

$$A(h) \leq 5(D_1 + D_2 + D_3 + D_4 + D_5)$$

with:

$$\begin{aligned}
D_1 &:= \sup_{h' \in \mathcal{H}_{N,T}} \|f_{h,h'} - f_{h'}\|^2, \\
D_2 &:= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h'} - \mathbb{E}[\widehat{f}_{h'}]\|^2 - \frac{V_1(h')}{10} \right)_+, \\
D_3 &:= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h,h'} - \mathbb{E}[\widehat{f}_{h,h'}]\|^2 - \frac{V_1(h')}{10} \right)_+ \\
D_4 &:= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h'}] - f_{h'}\|^2 - \frac{V_2(h')}{10} \right)_+, \\
D_5 &:= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h,h'}] - f_{h,h'}\|^2 - \frac{V_2(h')}{10} \right)_+.
\end{aligned}$$

According to Young's inequality (see Theorem 9.1), we obtain

$$\|f_{h,h'} - f_{h'}\|^2 = \|K_{h'} \star (f_h - f)\|^2 \leq \|K_{h'}\|_1^2 \|f_h - f\|^2 = \|K\|_1^2 \|f_h - f\|^2$$

thus

$$D_1 \leq \|K\|_1^2 \|f_h - f\|^2. \tag{18}$$

Let us study the term D_2. We denote $\mathcal{B}(1) = \{g \in \mathbb{L}^2(\mathbb{R}), \|g\| = 1\}$. We define

$$\nu_{N,h}(g) := \langle g, \widehat{f}_h - \mathbb{E}[\widehat{f}_h] \rangle$$

then $|\nu_{N,h}(g)| \leq \|g\| \|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|$ thus, the estimator $\widehat{f}_h$ satisfies:

$$\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2 = \sup_{g \in \mathcal{B}(1)} (\nu_{N,h}(g))^2.$$

We can also compute the scalar product which defines $\nu_{N,h}$ and we obtain

$$\nu_{N,h}(g) = \frac{1}{N} \sum_{j=1}^{N} \left(g \star K_h^-(Z_{j,T}) - \mathbb{E}[g \star K_h^-(Z_{j,T})] \right) \tag{19}$$

with $K_h^-(x) := K_h(-x)$. This finally conducts to:

$$\mathbb{E}[D_2] \leq \sum_{h' \in \mathcal{H}_{N,T}} \mathbb{E}\left[\sup_{g \in \mathcal{B}(1)} (\nu_{N,h'}(g))^2 - \frac{V_1(h')}{10} \right]_+.$$

This bound and Eq. (19) leads to apply Talagrand's Theorem (9.2). We have to compute 3 quantities: M, H^2 and v.
First:

$$\begin{aligned}\sup_{g\in\mathcal{B}(1)} \|g\star K_{h'}^-\|_\infty &= \sup_{g\in\mathcal{B}(1)}\sup_{x\in\mathbb{R}} \left|\int g(y)K_{h'}^-(x-y)dy\right| \\ &= \sup_{g\in\mathcal{B}(1)}\sup_{x\in\mathbb{R}} |\langle g, K_{h'}^-(.-x)\rangle| \\ &\le \sup_{g\in\mathcal{B}(1)} \|g\|\|K_{h'}\| = \frac{\|K\|}{\sqrt{h'}} := M. \end{aligned} \tag{20}$$

Secondly, the bound of Proposition 3.1 gives

$$\mathbb{E}\left[\sup_{g\in\mathcal{B}(1)} (\nu_{N,h}(g))^2\right] = \mathbb{E}\left[\|\widehat{f_h} - \mathbb{E}[\widehat{f_h}]\|^2\right] \le \frac{\|K\|^2}{Nh} := H^2. \tag{21}$$

Thirdly:

$$\begin{aligned}\sup_{g\in\mathcal{B}(1)} \left(\mathrm{Var}(g\star K_{h'}^-(Z_{1,T}))\right) &\le \sup_{g\in\mathcal{B}(1)} \mathbb{E}[(g\star K_{h'}^-(Z_{1,T}))^2] \\ &\le 2\sup_{g\in\mathcal{B}(1)} \mathbb{E}[(g\star K_{h'}^-(\phi_1))^2] \\ &\quad + 2\sup_{g\in\mathcal{B}(1)} \mathbb{E}[(g\star(K_{h'}^-(Z_{1,T}) - K_{h'}^-(\phi_1))^2].\end{aligned}$$

Let us investigate the two terms separately. Young's inequality gives:

$$\begin{aligned}\mathbb{E}\left[(g\star K_{h'}^-(\phi_1))^2\right] &= \int \left(g\star K_{h'}^-(x)\right)^2 f(x)dx \le \|f\|\|g\star K_{h'}^-\|_4^2 \\ &= \frac{\|f\|\|K\|_{4/3}^2}{\sqrt{h'}} := v_1.\end{aligned} \tag{22}$$

Then, one can write: $K_{h'}(x - Z_{1,T}) - K_{h'}(x-\phi_1) = (\phi_1 - Z_{1,T})\int_0^1 (K_{h'})'(x-\phi_1 + u(\phi_1 - Z_{1,T}))du$, thus

$$\begin{aligned}&(g\star K_{h'}^-(Z_{1,T}) - g\star K_{h'}(\phi_1))^2 \\ &\quad = (\phi_1 - Z_{1,T})^2\left(\int g(x)\int_0^1 (K_{h'})'(x-\phi_1+u(\phi_1-Z_{1,T}))dudx\right)^2 \\ &\quad \le (\phi_1 - Z_{1,T})^2 \int g^2(x)\left(\int_0^1 (K_{h'})'^2(x-\phi_1+u(\phi_1-Z_{1,T}))du\right)dx \\ &\quad \le (\phi_1 - Z_{1,T})^2\|g\|^2\int (K_{h'})'^2(y)dy = (\phi_1 - Z_{1,T})^2\|(K_{h'})'\|^2.\end{aligned}$$

With $\mathbb{E}[(\phi_1 - Z_{1,T})^2] = \frac{\sigma^2}{T^2}\mathbb{E}[W_1(T)^2] = \frac{\sigma^2}{T}$, the assumption $T^{-1} \leq h^{5/2}$ leads to

$$\mathbb{E}\left[(g\star K_{h'}^{-}(Z_{1,T}) - g\star K_{h'}(\phi_1))^2\right] \leq \frac{\|K'\|^2\sigma^2}{h'^3 T} \leq \frac{\|K'\|^2\sigma^2}{\sqrt{h'}} := v_2. \tag{23}$$

Finally $v = v_1 + v_2 = A_0/\sqrt{h'}$ with $A_0 = \|f\|\|K\|_{4/3}^2 + \|K'\|^2\sigma^2$.
If $\kappa_1\|K\|_1^2 \geq 40$, with the assumption $1/(Nh) \leq 1$, Talagrand's inequality (under the assumptions of the Theorem 3.1) gives

$$\begin{aligned}\mathbb{E}\left(\sup_{g\in\mathcal{B}(1)}(\nu_{N,h'}(g))^2 - \frac{V_1(h')}{10}\right)_+ &\leq \frac{C_1}{N\sqrt{h'}}e^{-C_2/\sqrt{h'}} + C_3\frac{1}{h'N^2}e^{-C_4\sqrt{N}} \\ &\leq \frac{C_5}{N}\sum_{h'\in\mathcal{H}_{N,T}}\frac{1}{\sqrt{h'}}e^{-C_6/\sqrt{h'}} \leq \frac{C_5 S(C_6)}{N}.\end{aligned}$$

One can lead the study of D_3 as we have done for D_2, using the same steps and tools. However $K_h\star K_{h'}$ instead of $K_{h'}$, adds $\|K\|_1$ in M and $\|K\|_1^2$ in H^2 and v. Then, let us study the term D_4. If $\kappa_2 \geq 10/(3\|K\|_1^2)$, the bound (9) leads us to

$$\begin{aligned}D_4 &= \sup_{h'\in\mathcal{H}_{N,T}}\left(\|\mathbb{E}[\widehat{f}_{h'}] - f_{h'}\|^2 - \frac{V_2(h')}{10}\right)_+ \\ &\leq \sup_{h'\in\mathcal{H}_{N,T}}\left(\frac{\|K''\|^2\sigma^4}{3h'^5T^2} - \frac{\kappa_2\|K\|_1^2\|K''\|^2\sigma^4}{10T^2h'^5}\right)_+ = 0\end{aligned}$$

thus $D_4 = 0$. Finally, similarly, if $\kappa_2 \geq 10/3$, we obtain

$$\begin{aligned}D_5 &= \sup_{h'\in\mathcal{H}_{N,T}}\left(\|\mathbb{E}[\widehat{f}_{h,h'}] - f_{h,h'}\|^2 - \frac{V_2(h')}{10}\right)_+ \\ &\leq \sup_{h'\in\mathcal{H}_{N,T}}\left(\frac{\|K''\|^2\|K\|_1^2\sigma^4}{3h^5T^2} - \frac{\kappa_2\|K\|_1^2\|K''\|^2\sigma^4}{10T^2h'^5}\right)_+ = 0.\end{aligned}$$

Thus finally we obtained that:

$$\mathbb{E}[A(h)] \leq 5\left(\|K\|_1^2\|f_h - f\|^2 + \frac{c}{N}\right) \tag{24}$$

with c a constant depending on $\|f\|$, $\|K\|_1$, $\|K\|$, $\|K\|_{4/3}$. Finally we have shown that for all $h \in \mathcal{H}_{N,T}$:

$$\begin{aligned}\mathbb{E}[\|\widehat{f}_{\widehat{h}} - f\|^2] \leq\ & 6\kappa_1\frac{\|K\|_1^2\|K\|^2}{Nh} + 6\kappa_2\frac{\|K\|_1^2\|K''\|^2\sigma^4}{T^2h^5} \\ &+ 3\left(2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + \frac{2\|K''\|^2\sigma^4}{3T^2h^5}\right)\end{aligned}$$

$$
\begin{aligned}
&+30\left(\|K\|_1^2\|f-f_h\|^2+\frac{c}{N}\right)\\
&\le\left(6+\frac{3}{\|K\|_1^2\kappa_1}\right)V_1(h)+\left(6+\frac{9}{2\|K\|_1^2\kappa_2}\right)V_2(h)\\
&\quad+(30\|K\|_1^2+6)\|f_h-f\|^2+\frac{C}{N}\\
&\le C_1\inf_{h\in\mathcal{H}_{N,T}}\{\|f-f_h\|^2+V(h)\}+\frac{C_2}{N}.
\end{aligned}
$$

where $C_1=\max(7,30\|K\|_1^2+6)$ and C_2 depends on $\|f\|$, $\|K\|_1$, $\|K\|$, $\|K\|_{4/3}$. □

8.2 Proof of Proposition 4.1

The bias term is $\|f-\mathbb{E}[\widetilde{f}_{m,s}]\|^2$. Let us compute $\mathbb{E}[\widetilde{f}_{m,s}]$. As the $Z_{j,\tau}$ are i.i.d. when τ is fixed and due to the independence of ϕ_1 and W_1, we obtain:

$$
\begin{aligned}
\mathbb{E}[\widetilde{f}_{m,s}(x)]&=\frac{1}{2\pi}\int_{-m}^{m}e^{-iux}\mathbb{E}\left[e^{iuZ_{1,m^2/s^2}+u^2\sigma^2s^2/(2m^2)}\right]du\\
&=\frac{1}{2\pi}\int_{-m}^{m}e^{-iux}\mathbb{E}\left[e^{iu\phi_1+iu\sigma W_1(m^2/s^2)s^2/m^2+u^2\sigma^2s^2/(2m^2)}\right]du\\
&=\frac{1}{2\pi}\int_{-m}^{m}e^{-iux+u^2\sigma^2s^2/(2m^2)}f^*(u)\mathbb{E}\left[e^{iu\sigma W_1(m^2/s^2)s^2/m^2}\right]du\\
&=\frac{1}{2\pi}\int_{-m}^{m}e^{-iux+u^2\sigma^2s^2/(2m^2)}f^*(u)e^{-u^2\sigma^2s^2/(2m^2)}du\\
&=\frac{1}{2\pi}\int_{-m}^{m}e^{-iux}f^*(u)du=:f_m(x).
\end{aligned}
$$

Therefore this gives $\mathbb{E}[\widetilde{f}_{m,s}(x)]=f_m(x)$, and $\|f-\mathbb{E}[\widetilde{f}_{m,s}]\|^2=\|f-f_m\|^2=\frac{1}{2\pi}\int_{|u|\ge m}|f^*(u)|^2du$.
The variance term is:

$$
\begin{aligned}
\mathbb{E}\left[\|\widetilde{f}_{m,s}-f_m\|^2\right]&=\frac{1}{2\pi}\mathbb{E}\left[\int_{-m}^{m}\left|\frac{1}{N}\sum_{j=1}^{N}e^{iuZ_{j,m^2/s^2}}e^{\frac{u^2\sigma^2s^2}{2m^2}}-f^*(u)\right|^2du\right]\\
&=\frac{1}{2\pi N}\int_{-m}^{m}e^{\frac{u^2\sigma^2s^2}{m^2}}\mathrm{Var}\left(e^{iuZ_{1,m^2/s^2}}\right)du\\
&\le\frac{1}{2\pi N}\int_{-m}^{m}e^{\frac{u^2\sigma^2s^2}{m^2}}du=\frac{m}{\pi N}\int_0^1e^{s^2\sigma^2v^2}du.
\end{aligned}
$$

□

8.3 Proof of Theorem 4.1

Let us study the term $\|\widetilde{f}_{\widetilde{m},\widetilde{s}} - f\|^2$. We decompose it into a sum of three terms and the definition of $(\widetilde{m}, \widetilde{s})$ (15) implies for all $(m, s) \in \mathcal{C}$

$$\begin{aligned}\|\widetilde{f}_{\widetilde{m},\widetilde{s}} - f\|^2 &\leq 3\left(\|\widetilde{f}_{\widetilde{m},\widetilde{s}} - \widetilde{f}_{(\widetilde{m},\widetilde{s})\wedge(m,s)}\|^2 + \|\widetilde{f}_{(\widetilde{m},\widetilde{s})\wedge(m,s)} - \widetilde{f}_{m,s}\|^2 + \|\widetilde{f}_{m,s} - f\|^2\right)\\ &\leq 3\left(\Gamma_{m,s} + \mathrm{pen}(\widetilde{m}, \widetilde{s})\right) + 3\left(\Gamma_{\widetilde{m},\widetilde{s}} + \mathrm{pen}(m, s)\right) + 3\|\widetilde{f}_{m,s} - f\|^2\\ &\leq 6\Gamma_{m,s} + 6\mathrm{pen}(m, s) + 3\|\widetilde{f}_{m,s} - f\|^2 \qquad (25)\end{aligned}$$

Now we study $\Gamma_{m,s}$. First:

$$\begin{aligned}&\|\widetilde{f}_{(m,s)\wedge(m',s')} - \widetilde{f}_{m',s'}\|^2\\ &\quad\leq 3\left(\|\widetilde{f}_{m',s'} - f_{m'}\|^2 + \|f_{m'} - f_{m\wedge m'}\|^2 + \|f_{m\wedge m'} - \widetilde{f}_{(m',s')\wedge(m,s)}\|^2\right).\end{aligned}$$

Thus:

$$\begin{aligned}\Gamma_{m,s} &\leq \max_{(m',s')\in\mathcal{C}}\Big(3\|\widetilde{f}_{m',s'} - f_{m'}\|^2 + 3\|f_{m'} - f_{m\wedge m'}\|^2 + 3\|f_{m\wedge m'}\\ &\qquad - \widetilde{f}_{(m',s')\wedge(m,s)}\|^2 - \mathrm{pen}(m', s')\Big)_+\\ &\leq 3\max_{(m',s')\in\mathcal{C}}\left(\|\widetilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\mathrm{pen}(m', s')\right)_+\\ &\qquad + 3\max_{(m',s')\in\mathcal{C}}\left(\|\widetilde{f}_{(m',s')\wedge(m,s)} - f_{m\wedge m'}\|^2 - \frac{1}{6}\mathrm{pen}(m', s')\right)_+\\ &\qquad + 3\max_{m'\in\mathcal{M}}\|f_{m'} - f_{m\wedge m'}\|^2.\end{aligned}$$

The last maximum can be explicit. If $m' \leq m$, then $\|f_{m'} - f_{m\wedge m'}\|^2 = \|f_{m'} - f_{m'}\|^2 = 0$. Otherwise,

$$\|f_{m'} - f_{m\wedge m'}\|^2 = \|f_{m'} - f_m\|^2 = \int_{m\leq|u|\leq m'} |f^*(u)|^2 du \leq \|f - f_m\|^2.$$

Finally:

$$\max_{m'\in\mathcal{M}}\|f_{m'} - f_{m\wedge m'}\|^2 \leq \|f - f_m\|^2.$$

We get the following bound for $\Gamma_{m,s}$:

$$\begin{aligned}\Gamma_{m,s} &\leq 3\max_{(m',s')\in\mathcal{C}}\left(\|\widetilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\mathrm{pen}(m', s')\right)_+\\ &\qquad + 3\max_{(m',s')\in\mathcal{C}}\left(\|\widetilde{f}_{(m',s')\wedge(m,s)} - f_{m\wedge m'}\|^2 - \frac{1}{6}\mathrm{pen}(m', s')\right)_+\\ &\qquad + 3\|f - f_m\|^2. \qquad (26)\end{aligned}$$

Then we gather Eqs. (25) and (26):

$$\begin{aligned}\|\widetilde{f}_{\widetilde{m},\widetilde{s}} - f\|^2 \leq 6\mathrm{pen}(m,s) + 3\|\widetilde{f}_{m,s} - f\|^2 + 18\|f - f_m\|^2 \\ + \max_{(m',s')\in\mathcal{C}} 18\left(\|\widetilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\mathrm{pen}(m',s')\right)_+ \\ + \max_{(m',s')\in\mathcal{C}} 18\left(\|\widetilde{f}_{(m',s')\wedge(m,s)} - f_{m\wedge m'}\|^2 - \frac{1}{6}\mathrm{pen}(m',s')\right)_+ .\end{aligned}$$

We first notice that our penalty function is increasing in s and m, thus we get the following bound for the last term:

$$\begin{aligned}&\mathbb{E}\left[\max_{(m',s')\in\mathcal{C}}\left(\|\widetilde{f}_{(m',s')\wedge(m,s)} - f_{m\wedge m'}\|^2 - \frac{1}{6}\mathrm{pen}((m',s')\wedge(m,s))\right)_+\right] \\ &\leq \mathbb{E}\left[\max_{m'\leq m, s'\leq s}\left(\|\widetilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\mathrm{pen}(m',s')\right)_+\right] \\ &\quad + \mathbb{E}\left[\max_{m\leq m', s\leq s'}\left(\|\widetilde{f}_{m,s} - f_m\|^2 - \frac{1}{6}\mathrm{pen}(m,s)\right)_+\right] \\ &\quad + \mathbb{E}\left[\max_{m\leq m', s'\leq s}\left(\|\widetilde{f}_{m,s'} - f_m\|^2 - \frac{1}{6}\mathrm{pen}(m,s')\right)_+\right] \\ &\quad + \mathbb{E}\left[\max_{m'\leq m, s\leq s'}\left(\|\widetilde{f}_{m',s} - f_{m'}\|^2 - \frac{1}{6}\mathrm{pen}(m',s)\right)_+\right] \\ &\leq 4\sum_{m'\in\mathcal{M}}\sum_{s'\in\mathcal{S}}\mathbb{E}\left[\|\widetilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\mathrm{pen}(m',s')\right]_+ .\end{aligned}$$

Moreover, according to Proposition 4.1 and using the inequality $\int_0^1 e^{\sigma^2 s^2 v^2} dv \leq e^{\sigma^2 s^2}$, we obtain, for all $(m,s)\in\mathcal{C}$,

$$\begin{aligned}\mathbb{E}\left[\|\widetilde{f}_{\widetilde{m},\widetilde{s}} - f\|^2\right] \leq 5\times 18\sum_{m'\in\mathcal{M}}\sum_{s'\in\mathcal{S}}\mathbb{E}\left[\|\widetilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\mathrm{pen}(m',s')\right]_+ + 6\mathrm{pen}(m,s) \\ + 3\frac{m}{\pi N}e^{\sigma^2 s^2} + 21\|f - f_m\|^2.\end{aligned}$$

Then we obtain the announced result with the following Lemma.

Lemma 8.1 *There exists a constant $C' > 0$ such that for $pen(m,s)$ defined by $pen(m,s) = \kappa\frac{m}{N}e^{\sigma^2 s^2}$,*

$$\sum_{m'\in\mathcal{M}}\sum_{s'\in\mathcal{S}}\mathbb{E}\left[\|\widetilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}pen(m',s')\right]_+ \leq \frac{C'(P+1)}{N}.$$

According to Lemma 8.1, to be proved next, we choose $\text{pen}(m,s) = \kappa \frac{m}{N} e^{\sigma^2 s^2}$, thus, there exist two constants $C = 145,\ C' > 0$ such that,

$$\begin{aligned}\mathbb{E}\left[\|\widetilde{f}_{\widetilde{m},\widetilde{s}} - f\|^2\right] &\leq 5 \times 18 \sum_{m' \in \mathcal{M}} \sum_{s' \in \mathcal{S}} \mathbb{E}\left[\|\widetilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\text{pen}(m',s')\right]_+ \\ &\quad + \left(6\kappa + \frac{3}{\pi}\right)\frac{m}{N} e^{\sigma^2 s^2} + 21\|f - f_m\|^2 \\ &\leq C \inf_{(m,s)\in\mathcal{C}} \left\{\|f - f_m\|^2 + \frac{m}{N} e^{\sigma^2 s^2}\right\} + \frac{C'}{N}.\end{aligned}$$

□

Proof of Lemma 8.1 For a couple $(m,s) \in \mathcal{C}$ fixed, let us consider the subset $S_m := \{t \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \text{supp}(t^*) = [-m,m]\}$. For $t \in S_m$,

$$\nu_N(t) = \frac{1}{N}\sum_{j=1}^{N}\left(\varphi_t(Z_{j,m^2/s^2}) - \mathbb{E}\left[\varphi_t\left(Z_{j,m^2/s^2}\right)\right]\right)$$

with $\varphi_t(x) := \frac{1}{2\pi}\int \overline{t^*(u)} e^{iux + \sigma^2 u^2 s^2/(2m^2)} du$, then $\nu_N(t) = \frac{1}{2\pi}\langle t^*, (\widetilde{f}_{m,s} - f_m)^*\rangle$. This leads to

$$\|\widetilde{f}_{m,s} - f_m\|^2 = \sup_{t \in S_m,\ \|t\|=1} |\nu_N(t)|^2. \tag{27}$$

We also have by Cauchy–Schwarz inequality

$$\begin{aligned}\|\varphi_t\|_\infty &\leq \frac{1}{2\pi}\int |t^*(u)| e^{\sigma^2 u^2 s^2/(2m^2)} du \leq \frac{1}{2\pi}\left(\int_{-m}^{m} |t^*(u)|^2 du\right)^{1/2}\left(\int_{-m}^{m} e^{\sigma^2 u^2 s^2/m^2} du\right)^{1/2} \\ &\leq \frac{\sqrt{2m}}{\sqrt{2\pi}} e^{\sigma^2 s^2/2}\end{aligned}$$

thus

$$\sup_{t \in S_m, \|t\|=1} \|\varphi_t\|_\infty \leq \frac{\sqrt{m}}{\sqrt{\pi}} e^{\sigma^2 s^2/2} := M.$$

Then, by Proposition 4.1,

$$\mathbb{E}\left[\sup_{t \in S_m, \|t\|=1} |\nu_N(t)|^2\right] = \mathbb{E}\left[\|\widetilde{f}_{m,s} - f_m\|^2\right] \leq \frac{m}{\pi N}\int_0^1 e^{\sigma^2 s^2 v^2} dv \leq \frac{m}{\pi N} e^{\sigma^2 s^2} := H^2.$$

Using Fubini and Cauchy–Schwarz inequalities we obtain for all $(m,s) \in \mathcal{C}$:

$$\begin{aligned}4\pi \sup_{t \in S_m,\ \|t\|=1} \text{Var}(\varphi_t(Z_{j,m^2/s^2})) &\leq \sup_{t \in S_m, \|t\|=1} \iint t^*(u) t^*(-v) \mathbb{E}\left[e^{i(u-v)Z_{j,m^2/s^2}}\right] \\ &\quad \times e^{(u^2+v^2)\sigma^2 s^2/(2m^2)} du dv\end{aligned}$$

$$
\begin{aligned}
&\leq 2\pi \left(\iint_{[-m,m]^2} |f^*(u-v)|^2 e^{(u^2+v^2)\sigma^2 s^2/m^2} du dv \right)^{1/2} \\
&\leq 2\pi \left(e^{2\sigma^2 s^2} \iint_{[-m,m]^2} |f^*(u-v)|^2 du dv \right)^{1/2} \\
&\leq 2\pi e^{\sigma^2 s^2} \sqrt{2m} \left(\int_{-2m}^{2m} |f^*(z)|^2 dz \right)^{1/2} \\
&\leq 2\sqrt{2m}\sqrt{2}\pi\sqrt{\pi} e^{\sigma^2 s^2} \|f\| =: 4\pi^2 v, \\
v &:= \frac{\sqrt{m} e^{\sigma^2 s^2} \|f\|}{\sqrt{\pi}}.
\end{aligned}
$$

Finally using that $m \leq N$, $s \leq 2/\sigma$ and $\sum_{s\in\mathcal{S}} s = (4/\sigma)(1-(1/2)^{P+1}) < 4/\sigma$, the Talagrand's inequality with $\alpha = 1/2$ if $4H^2 \leq \mathrm{pen}(m,s)/6$ implies,

$$
\begin{aligned}
&\sum_{s\in\mathcal{S}} \sum_{m\in\mathcal{M}} \mathbb{E}\left[\|\widetilde{f}_{m,s} - f_m\|^2 - \frac{1}{6}\mathrm{pen}(s,m) \right]_+ \\
&\leq \sum_{s\in\mathcal{S}} \sum_{m\in\mathcal{M}} \left(\frac{C_1\|f\|}{N} e^{\sigma^2 s^2} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} + C_3 \frac{m}{N^2} e^{\sigma^2 s^2} e^{-C_4\sqrt{N}} \right) \\
&\leq \sum_{s\in\mathcal{S}} \frac{C_1\|f\|}{N} e^{\sigma^2 s^2} \left(\sum_{m\in\mathcal{M}} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} \right) + \sum_{s\in\mathcal{S}} \sum_{m\in\mathcal{M}} C_3 e^4 \frac{1}{N} e^{-C_4\sqrt{m}} \\
&\leq \frac{C_1\|f\|(P+1)e^4}{N} \left(\sum_{m\in\mathcal{M}} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} \right) + C_3 e^4 \frac{P+1}{N} \sum_{m\in\mathcal{M}} e^{-C_4\sqrt{m}} \\
&\leq \frac{C'(P+1)}{N}
\end{aligned}
$$

because with the definition of $\mathcal{M}$, $\sum_{m\in\mathcal{M}} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} \leq a_1 \sum_{k\in\mathbb{N}} k^{1/4} e^{-a_2 k^{1/4}} < +\infty$, and $\sum_{m\in\mathcal{M}} e^{-C_4 m^{1/2}} \leq \sum_{k\in\mathbb{N}} e^{-a_3 k^{1/4}} < +\infty$, with a_1, a_2, a_3 three positive constants. Notice that $C' > 0$ depends on σ, $\|f\|$, Δ.
We choose $\mathrm{pen}(m,s) = \kappa m e^{\sigma^2 s^2}/N$ with $\kappa \geq 24$. □

9 Appendix 2

9.1 Young inequality

This inequality can be found in Briane and Pagès (2006) for example.

Theorem 9.1 *Let f be a function belonging to $\mathbb{L}^p(\mathbb{R})$ and g belonging to $\mathbb{L}^q(\mathbb{R})$, let p, q, r be real numbers in $[1, +\infty]$ and such that*

$$
\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1.
$$

Then,

$$\|f \star g\|_r \leq \|f\|_p \|g\|_q.$$

9.2 Talagrand's inequality

The following result is a consequence of the Talagrand concentration inequality Talagrand (1996) given in Birgé and Massart (1997).

Theorem 9.2 *Consider $n \in \mathbb{N}^*$, $\mathcal{F}$ a class at most countable of measurable functions, and $(X_i)_{i\in\{1,\dots,N\}}$ a family of real independent random variables. One defines, for all $f \in \mathcal{F}$,*

$$\nu_N(f) = \frac{1}{N}\sum_{i=1}^{N}(f(X_i) - \mathbb{E}[f(X_i)]).$$

Supposing there are three positive constants M, H and v such that $\sup_{f\in\mathcal{F}} \|f\|_\infty \leq M$, $\mathbb{E}[\sup_{f\in\mathcal{F}} |\nu_N f|] \leq H$, and $\sup_{f\in\mathcal{F}} (1/N)\sum_{i=1}^{N} \mathrm{Var}(f(X_i)) \leq v$, then for all $\alpha > 0$,

$$\mathbb{E}\left[\left(\sup_{f\in\mathcal{F}} |\nu_N(f)|^2 - 2(1+2\alpha)H^2\right)_+\right] \leq \frac{4}{a}\left(\frac{v}{N}\exp\left(-a\alpha\frac{NH^2}{v}\right) + \frac{49M^2}{aC^2(\alpha)N^2}\exp\left(-\frac{\sqrt{2}aC(\alpha)\sqrt{\alpha}}{7}\frac{NH}{M}\right)\right)$$

with $C(\alpha) = (\sqrt{1+\alpha} - 1) \wedge 1$, and $a = \frac{1}{6}$.

9.3 Discretization

Indeed, if we assume that the times of observations are the $t_k = k\delta$, $k = 1, \dots, N$ and $0 < \delta < 1$, we must study the error applied by discretization of the $Z_{j,\tau}$. Then, for any $0 < m^2/s^2 \leq T$ we use:

$$\widehat{Z}_{j,m^2/s^2} = \frac{s^2}{m^2}\left[X_j(\delta[m^2/(s^2\delta)]) - X_j(0) + \frac{\delta}{\alpha}\sum_{k=1}^{[m^2/(s^2\delta)]} X_j((k-1)\delta)\right] \tag{28}$$

to approximate $Z_{j,m^2/s^2}$ given by (2). The corresponding estimator of f is

$$\widehat{\widetilde{f}}_{m,s}(x) = \frac{1}{2\pi}\int_{-m}^{m} e^{-iux}\frac{1}{N}\sum_{j=1}^{N} e^{iu\widehat{Z}_{j,m^2/s^2}} e^{\frac{u^2\sigma^2 s^2}{2m^2}} du. \tag{29}$$

We investigate the error:

$$\mathbb{E}[\|\widehat{\widetilde{f}}_{m,s} - f\|^2] \leq 2\mathbb{E}[\|\widehat{\widetilde{f}}_{m,s} - \widetilde{f}_{m,s}\|^2] + 2\mathbb{E}\left[\|\widetilde{f}_{m,s} - f\|^2\right]$$

where the second term of the right hand side is bounded by Proposition 4.1. Then, Plancherel–Parseval's Theorem implies:

$$\begin{aligned}\mathbb{E}[\|\widehat{\widetilde{f}}_{m,s} - \widetilde{f}_{m,s}\|^2] &\leq \frac{1}{2\pi}\mathbb{E}\left[\int_{-m}^{m} \frac{1}{N}\sum_{j=1}^{N} e^{u^2\sigma^2 s^2/m^2}\left|e^{iu\widehat{Z}_{j,m^2/s^2}} - e^{iuZ_{j,m^2/s^2}}\right|^2 du\right] \\ &\leq \frac{1}{2\pi}\int_{-m}^{m} e^{u^2\sigma^2 s^2/m^2}\mathbb{E}\left[\left|e^{iu\widehat{Z}_{1,m^2/s^2}} - e^{iuZ_{1,m^2/s^2}}\right|^2\right] du\end{aligned}$$

and

$$\mathbb{E}\left[\left|e^{iu\widehat{Z}_{1,m^2/s^2}} - e^{iuZ_{1,m^2/s^2}}\right|^2\right] \leq |u|^2\mathbb{E}\left[\left|\widehat{Z}_{1,m^2/s^2} - Z_{1,m^2/s^2}\right|^2\right]$$

thus we study the last term. For all $(m, s) \in \mathcal{C}$, $m^2/s^2 \leq T$,

$$\begin{aligned}Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2} &= \frac{s^2}{m^2}\left(X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)])\right) \\ &\quad + \frac{s^2}{\alpha m^2}\sum_{k=1}^{[m^2/(s^2\delta)]}\int_{(k-1)\delta}^{k\delta}(X_j(s) - X_j((k-1)\delta))ds\end{aligned}$$

then by Cauchy–Schwarz's inequality we obtain

$$\begin{aligned}(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2 &\leq \frac{2s^4}{m^4}\left(X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)])\right)^2 \\ &\quad + \frac{2s^4}{\alpha^2 m^4}\left[\sum_{k=1}^{[m^2/(s^2\delta)]}\int_{(k-1)\delta}^{k\delta}(X_j(s) - X_j((k-1)\delta))ds\right]^2.\end{aligned}$$

Höder's inequality yields

$$\begin{aligned}\left[\sum_{k=1}^{\left[\frac{m^2}{s^2\delta}\right]}\int_{(k-1)\delta}^{k\delta}(X_j(s) - X_j((k-1)\delta))ds\right]^2 &\leq \sum_{k=1}^{\left[\frac{m^2}{s^2\delta}\right]}\left[\int_{(k-1)\delta}^{k\delta}(X_j(s) - X_j((k-1)\delta))ds\right]^2\left[\frac{m^2}{s^2\delta}\right] \\ &\leq \left[\frac{m^2}{s^2\delta}\right]\delta\sum_{k=1}^{\left[\frac{m^2}{s^2\delta}\right]}\int_{(k-1)\delta}^{k\delta}(X_j(s) - X_j((k-1)\delta))^2 ds.\end{aligned}$$

Let us study $\mathbb{E}[(X_j(s) - X_j((k-1)\delta))^2]$, for $(k-1)\delta \leq s \leq k\delta$:

$$X_j(s) - X_j((k-1)\delta) = \int_{(k-1)\delta}^{s} \left(\phi_j - \frac{X_j(u)}{\alpha}\right) du + \int_{(k-1)\delta}^{s} \sigma dW_j(u)$$

and Cauchy–Schwarz's inequality gives

$$\begin{aligned}
\mathbb{E}[(X_j(s) - X_j((k-1)\delta))^2] &\leq 2\mathbb{E}\left[\left(\int_{(k-1)\delta}^{s} \left(\phi_j - \frac{X_j(u)}{\alpha}\right) du\right)^2\right] + 2\mathbb{E}\left[\left(\int_{(k-1)\delta}^{s} \sigma dW_j(u)\right)^2\right] \\
&\leq 2\mathbb{E}\left[\int_{(k-1)\delta}^{s} \left(\phi_j - \frac{X_j(u)}{\alpha}\right)^2 du\right] + 2\delta\sigma^2 \\
&\leq 4\delta^2\left(\mathbb{E}(\phi_j^2) + \frac{1}{\alpha^2}\sup_{s\geq 0}\mathbb{E}[X_j(s)^2]\right) + 2\delta\sigma^2. \qquad (30)
\end{aligned}$$

Finally, after simplification and using for all $x \in \mathbb{R}^+$, $[x] \leq x$,

$$\begin{aligned}
\mathbb{E}\left[(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2\right] \leq \frac{2s^4}{m^4}\mathbb{E}[\left(X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)])\right)^2] \\
+ \frac{2}{\alpha^2}\left(4\delta^2\left(\mathbb{E}(\phi_j^2) + \frac{1}{\alpha^2}\sup_{s\geq 0}\mathbb{E}[X_j(s)^2]\right) + 2\delta\sigma^2\right)
\end{aligned}$$

and we can deal with the term $\mathbb{E}[\left(X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)])\right)^2]$ using formula (30) and $m^2/s^2 - \delta[m^2/(s^2\delta)] \leq \delta$. Thus:

$$\mathbb{E}\left[(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2\right] \leq \left(\frac{2s^4}{m^4} + \frac{2}{\alpha^2}\right)\left(4\delta^2\left(\mathbb{E}(\phi_j^2) + \frac{1}{\alpha^2}\sup_{s\geq 0}\mathbb{E}[X_j(s)^2]\right) + 2\delta\sigma^2\right).$$

Besides, for model (1), Eq. (17) implies $\mathbb{E}[X_j(s)^2] \leq 3x_j^2 + 3\alpha^2\mathbb{E}[\phi_j^2] + 3\sigma^2$, and $0 < \delta < 1$ implies

$$\mathbb{E}\left[(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2\right] \leq C\delta\left(\frac{2s^4}{m^4} + \frac{2}{\alpha^2}\right)$$

with C a positive constant which does not depend on δ or m^2/s^2. Finally,

$$\begin{aligned}
\mathbb{E}[\|\widehat{\widetilde{f}}_{m,s} - \widetilde{f}_{m,s}\|^2] &\leq C\delta\left(\frac{2s^4}{m^4} + \frac{2}{\alpha^2}\right)\frac{1}{2\pi}\int_{-m}^{m} u^2 e^{u^2\sigma^2 s^2/m^2} du \\
&\leq C'\delta\left(\int_0^1 v^2 e^{v^2\sigma^2 s^2} dv\right)\left(\frac{s^4}{m} + \frac{m^3}{\alpha^2}\right).
\end{aligned}$$

But $s \leq 2/\sigma$ and $m = \sqrt{k\Delta}/\sigma$, with $k \in \mathbb{N}^*$ and $0 < \Delta < 1$, thus we obtain

$$\mathbb{E}[\|\widehat{\widetilde{f}}_{m,s} - \widetilde{f}_{m,s}\|^2] \leq \frac{C'}{\sigma^3}\left(\int_0^1 v^2 e^{v^2\sigma^2 s^2} dv\right)\left(2^4\sqrt{k}\left(\frac{\delta}{\sqrt{\Delta}}\right) + \frac{k^{3/2}}{\alpha^2}\left(\delta\Delta^{3/2}\right)\right).$$

Proposition 9.1 *Under (A), assuming* $\mathbb{E}[\phi_j^2] < +\infty$, *the estimator* $\widehat{\widetilde{f}}_{m,s}$ *given by (29) satisfies*

$$\mathbb{E}\left[\|\widetilde{f}_{m,s} - f\|^2\right] \leq \|f_m - f\|^2 + \frac{\sqrt{k\Delta}}{\sigma\pi N}e^{\sigma^2 s^2} + \frac{C'}{\sigma^3}\frac{e^{\sigma^2 s^2}}{2\sigma^2 s^2}\left(2^4\sqrt{k}\left(\frac{\delta}{\sqrt{\Delta}}\right) + \frac{k^{3/2}}{\alpha^2}\left(\delta\Delta^{3/2}\right)\right).$$

Finally if Δ is fixed and δ is small, the error is acceptable. For example if $\delta = \Delta$ the error is of order $\sqrt{\delta}$.
For study on the kernel estimator we refer to Comte et al. (2013).

References

Birgé L, Massart P (1997) From model selection to adaptive estimation. Springer, New York

Birgé L, Massart P (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. Bernoulli 4:329–375

Bissantz N, Dümbgen L, Holzmann H, Munk A (2007) Nonparametric confidence bands in deconvolution density estimation. J R Stat Soc Series B (Stat Methodol) 69:483–506

Briane M, Pagès G (2006) Théorie de l'intégration. Vuibert, Paris

Butucea C, Tsybakov A (2007) Sharp optimality in density deconvolution with dominating bias II. Teor Veroyatnost i Primenen 52:336–349

Carroll R, Hall P (1988) Optimal rates of convergence for deconvolving a density. J Am Stat Assoc 83:1184–1186 ISSN 01621459

Chagny G (2013) Warped bases for conditional density estimation. Math Methods Stat 22:253–282

Comte F, Genon-Catalot V, Rozenholc Y (2007) Penalized nonparametric mean square estimation of the coefficients of diffusion processes. Bernoulli 13:514–543

Comte F, Genon-Catalot V, Samson A (2013) Nonparametric estimation for stochastic differential equation with random effects. Stoch Process Appl 7:2522–2551

Comte F, Johannes J (2012) Adaptive functional linear regression. Ann Stat 40:2765–2797

Comte F, Rozenholc Y, Taupin M-L (2006) Penalized contrast estimator for adaptive density deconvolution. Can J Stat 34:431–452

Comte F, Samson A (2012) Nonparametric estimation of random-effects densities in linear mixed-effects model. J Nonparametr Stat 24:951–975

Davidian M, Giltinan D (1995) Nonlinear models for repeated measurement data. CRC press

Delattre M, Genon-Catalot V, Samson A (2015) Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. ESAIM Probab Stat 19:671–688

Delattre M, Genon-Catalot V, Samson A (2016) Mixtures of stochastic differential equations with random effects: application to data clustering. J Stat Plan Inference 173:109–124

Delattre M, Lavielle M (2013) Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. Stat Interface 6:519–532

Diggle P, Heagerty P, Liang K, Zeger S (2002) Analysis of longitudinal data. Oxford statistical science series

Dion C, Genon-Catalot V (2015) Bidimensional random effect estimation in mixed stochastic differential model. Stoch Inference Stoch Process 18(3):1–28

Donnet S, Foulley J, Samson A (2010) Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. Biometrics 66:733–741

Donnet S, Samson A (2008) Parametric inference for mixed models defined by stochastic differential equations. ESAIM Prob Stat 12:196–218

Donnet S, Samson A (2013) A review on estimation of stochastic differential equations for pharmacokinetic–pharmacodynamic models. Adv Drug Deliv Rev 65:929–939

Donnet S, Samson A (2014) Using PMCMC in EM algorithm for stochastic mixed models: theoretical and practical issues. J Soc Fr Stat 155:49–72

Fan J (1991) On the optimal rates of convergence for nonparametric deconvolution problems. Ann Statist 19:1257–1272

Genon-Catalot V, Jacod J (1993) On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. Ann Inst Henri Poincaré B Probab Stat 29:119–151

Genon-Catalot V, Larédo C (2016) Estimation for stochastic differential equations with mixed effects. Statistics. doi:10.1080/02331888.2016.1141910

Goldenshluger A, Lepski O (2011) Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. Ann Stat 39:1608–1632

Hoffmann M (1999) Adaptive estimation in diffusion processes. Stoch Process Appl 79:135–163

Klein T, Rio E (2005) Concentration around the mean for maxima of empirical processes. Ann Probab 33:1060–1077

Kutoyants Y (2004) Statistical inference for ergodic diffusion processes. Springer, London

Lacour C (2006) Rates of convergence for nonparametric deconvolution. C R Math Acad Sci Paris 342:877–882

Lacour C, Massart P (2016) Minimal penalty for Goldenshluger–Lepski method. Stoch Processes Appl. doi:10.1016/j.spa.2016.04.015

Lansky P, Sanda P, He J (2006) The parameters of the stochastic leaky integrate-and-fire neuronal model. J Comput Neurosci 21:211–223

Picchini U, De Gaetano A, Ditlevsen S (2010) Stochastic differential mixed-effects models. Scand J Stat 37:67–90

Picchini U, Ditlevsen S (2011) Practicle estimation of high dimensional stochastic differential mixed-effects models. Comput Stat Data Anal 55:1426–1444

Picchini U, Ditlevsen S, De Gaetano A, Lansky P (2008) Parameters of the diffusion leaky integrate-and-fire neuronal model for a slowly fluctuating signal. Neural Comput 20:2696–2714

Pinheiro J, Bates D (2000) Mixed-effect models in S and Splus. Springer, New York

Talagrand M (1996) New concentration inequalities in product spaces. Invent Math 126:505–563

Yu Y, Xiong Y, Chan Y, He J (2004) Corticofugal gating of auditory information in the thalamus: an in vivo intracellular recording study. J Neurosci 24:3060–3069