Stat Infer Stoch Process
DOI 10.1007/s11203-007-9020-1

L^1-convergence of smoothing densities in non-parametric state space models

Valérie Monbet · Pierre Ailliot · Pierre-François Marteau

Received: 15 January 2005 / Accepted: 15 September 2007

Abstract This paper addresses the problem of reconstructing partially observed stochastic processes. The L^1 convergence of the filtering and smoothing densities in state space models is studied, when the transition and emission densities are estimated using non parametric kernel estimates. An application to real data is proposed, in which a wave time series is forecasted given a wind time series.

Keywords Filtering · Non parametric density estimation · Smoothing · Stability · State space models · Wind-wave models

AMS Classifications (2000) 62M20 · 60J25 · 60M35 · 93E14

1 Introduction

A state-space model is basically a Markov process with two components. One of them, the state process, is assumed to be hidden and its evolution is characterized by its initial distribution and its transition kernel. The other one, the observed process, is assumed to be observable and is related to the state process by the emission probabilities.

These models were first introduced in the 60s in the fields of control and speech recognition (Baum and Petrie 1966; Kalman 1960). They appear in the statistical literature only in the seventies (Akaike 1974; Harrison and Stevens 1976). Finally, during the last decade they became a focus of interest due to their wide range of applications. Recently, particular attention has been given to the properties of filtering (Douc and Matias 2001; Le Gland and Mevel 2000) and smoothing (Godsill et al. 2004) recursions, which permit to forecast the hidden state given the observed process, and in particular to the stability of the Markovian

Valérie Monbet—supported by IFREMER, Brest, France.

V. Monbet (✉) · P. Ailliot · P. F. Marteau
CERYC, Université de Bretagne Sud, Campus de Tohannic, 56000 Vannes, France
e-mail: valerie.monbet@univ-ubs.fr

operator used in the filtering recursions. These results have been used, for example, to study the convergence properties of particle filters and of maximum likelihood estimates.

In the present paper, we assume that both components of the Markov process are simultaneously observed on a period of time and this learning sequence is used to estimate the transition kernel of the state process and the emission probabilities. The originality of this work is to use non parametric estimates for these conditional densities. Then, we consider another period of time in which only the observation is available. The corresponding state sequence is forecasted by computing filtering and smoothing recursions with the "true" transition kernel and emission probabilities replaced by their non-parametric estimates.

Such a situation can occur, for example, in meteorology. For instance, let us assume that wind and wave follow a state-space model, with the wave corresponding to the state process and the wind to the observed process. Generally, wind and wave time series are available only on short periods of time, but this can be sufficient to learn the state space model. For these meteorological time series, it is well-known that it is hard to find appropriate parametric models for the multivariate joint distributions (strong non-linearities, positive and asymmetric marginal distributions, etc...), and, as a consequence, for the transition kernel of the state process and the emission probabilities. In this context, it is natural to use non parametric estimates in order to have enough flexibility to restore the complexity of the phenomena.

In this paper, we study the asymptotic convergence of the proposed estimates. For that purpose, we combine convergence results for non parametric kernel density estimates for stochastic processes (Bosq 1996; Liebscher 1999, 2001) with the contractivity properties of the filtering and smoothing recursions. We show that the convergence rates of the filtering and smoothing densities essentially depend on the convergence rates of the non parametric estimates and on the contractivity properties of the filtering and smoothing operators.

In Sect. 2, we introduce the methodology and the notations. Then, the asymptotic properties of the filtering and smoothing densities are studied, respectively, in Sects. 3 and 4. In Sect. 5, the results are illustrated through simulations. Finally, in Sect. 6, the method is applied to meteorological data.

2 Problem statement

2.1 State space model

We consider the following model, with $\{S_t\}$ being the state process and $\{Y_t\}$ being the observed process:

(M1) $\{S_t\}$ is a homogeneous Markov chain on $\mathcal{S} \subset \mathbb{R}^d$, with $d \geq 1$, equipped with the Borel σ-field $\mathcal{B}(\mathcal{S})$. For $s \in \mathcal{S}$ and $B \in \mathcal{B}(\mathcal{S})$, we denote by $A(s, B)$ the transition kernel of this Markov chain. Furthermore, we assume that for all $s \in \mathcal{S}$ the measure $A(s, .)$ has a probability density function (pdf) $a(.|s)$ with respect to a common finite dominating measure μ on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ and that the initial distribution of the Markov chain, which is the distribution of S_0, is absolutely continuous with respect to μ. ξ_0 denotes the corresponding pdf.

(M2) $\{Y_t\}$ takes its values in $\mathcal{Y} \subset \mathbb{R}^{d'}$ with $d' \geq 1$, equipped with the Borel σ-field $\mathcal{B}(\mathcal{Y})$. For each $r \geq 1$, Y_r is conditionally independent of $\{Y_t\}_{t=1,\ldots,r-1}$ given S_r. We also assume that for each $s \in \mathcal{S}$, the conditional distribution $P(Y_r \in B | S_r = s)$ has a density $b(.|s)$ with respect to a common finite dominating measure ν on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$.

This model is a special case of a graphical model on an acyclic directed graph.

$$\begin{array}{lccccc} \text{State process} & \ldots \to & S_{t-1} & \to & S_t & \to & S_{t+1} \to \ldots \\ & & \downarrow & & \downarrow & & \downarrow \\ \text{Observed process} & & Y_{t-1} & & Y_t & & Y_{t+1} \end{array}$$

One usually refers to this model as *Hidden Markov Model* (HMM) when the hidden state space $\mathcal{S}$ is finite and as *State Space Model* when it is infinite.

Throughout this paper, we make the following assumptions:

$[\mathcal{A}]$
1. There exist κ_a^- and κ_a^+ such that for all $s, s' \in \mathcal{S}$, $0 < \kappa_a^- \leq a(s|s') \leq \kappa_a^+ < +\infty$
2. There exist κ_b^- and κ_b^+ such that for all $y \in \mathcal{Y}$, $0 < \kappa_b^- \leq \int b(y|s)\mu(ds) \leq \kappa_b^+ < +\infty$

Hypothesis $[\mathcal{A}]1$ is fundamental here. It implies that the state space of the Markov chain $\{S_t\}$ is 1-small (see Meyn and Tweedie 1993). Thus, the chain is uniformly ergodic and has a unique invariant distribution. Since the transition kernel A admits a density, it is easy to check, thanks to Radon-Nicodym's theorem, that the invariant distribution also admits a density which is denoted by f. Condition $\kappa_a^+ < +\infty$ is reasonable for many applications, but condition $\kappa_a^- > 0$ may be restrictive. It is an usual assumption to get the contractivity of the forward operator (see Atar and Zeitouni 1997; Del Moral and Guionnet 2001), but several authors have recently proposed to refine it (see Chigansky and Liptser 2004 and references therein). In this paper, we keep this assumption because it is also convenient to establish both the convergence of the non parametric estimates (see Sects. 3.2 and 4.1) and the contractivity of forward and backward operators (see Sects. 3.3 and 4.2).

2.2 Filtering and smoothing recursions

For any sequence $\{x_1, \ldots, x_t\}$, we denote for $t' \leq t$, $x_{t'}^t = \{x_{t'}, \ldots, x_t\}$. Assumptions **(M1)**, **(M2)** and $[\mathcal{A}]$ imply that the conditional probability $P(S_t \in B | Y_1^r = y_1^r)$ admits a density $f_{t|r}(.)$ with respect to μ. We distinguish between prediction ($r < t$), filtering ($r = t$) and smoothing ($r > t$). It is well known that the filtering and smoothing densities verify, respectively, the recursions (1) and (3) below. More precisely, let $f_{0|0} = \xi_0$ denotes the density of the initial distribution on S (the choice of this distribution is discussed later on). For $t \geq 1$, we have

$$f_{t|t} = \Pi_F^{(t)} f_{t-1|t-1} \tag{1}$$

where $\Pi_F^{(t)}$ denotes the *forward operator* at time t. This operator acts on the probability density functions on $\mathcal{S}$. If ξ denotes an arbitrary pdf on $\mathcal{S}$, $\Pi_F^{(t)}\xi$ is the pdf given by

$$\Pi_F^{(t)}\xi(s_t) = \frac{b(y_t|s_t) \int a(s_t|s_{t-1})\xi(s_{t-1})\mu(ds_{t-1})}{\iint a(s|s_{t-1})b(y_t|s)\xi(s_{t-1})\mu(ds_{t-1})\mu(ds)} \tag{2}$$

We can check that, under hypothesis $[\mathcal{A}]$,

$$\iint a(s|s_{t-1})b(y_t|s)\xi(s_{t-1})\mu(ds)\mu(ds_{t-1}) \geq \kappa_a^-\kappa_b^-$$

so that $\Pi_F^{(t)}$ is well defined. The proof of relation (2) is straightforward using the conditional independence properties which characterize the state space model together with the law of

total probability and Bayes' theorem. The smoothing recursion is written in the same way starting with $f_{T|T}$. For $t < T$

$$f_{t|T} = \Pi_B^{(t)} f_{t+1|T}$$

where $\Pi_B^{(t)}$ denotes the *backward operator* at time t. For any pdf ξ on $\mathcal{S}$, $\Pi_B^{(t)}\xi$ is the pdf given by

$$\Pi_B^{(t)}\xi(s_t) = f_{t|t}(s_t) \int \frac{a(s_{t+1}|s_t)}{f_{t+1|t}(s_{t+1})}\xi(s_{t+1})\mu(ds_{t+1}) \tag{3}$$

with one step ahead forecast density

$$f_{t+1|t}(s_{t+1}) = \int f_{t|t}(s)a(s_{t+1}|s)\mu(ds)$$

The operator $\Pi_B^{(t)}$ is well defined under assumption $[\mathcal{A}]$ since it implies that $f_{t+1|t}(s) \geq \kappa_a^-$ for any $s \in \mathcal{S}$.

One of the most common problems in state space models consists in forecasting the hidden sequence s_1^T corresponding to an observed time series y_1^T and several algorithms can be found in the literature to compute the smoothing densities $f_{t|T}$. When the state space $\mathcal{S}$ is finite, all the integrals in filtering and smoothing recursions (2) and (3) are simply sums and the calculation is straightforward. The algorithm is then referred to as a forward–backward algorithm. Another case where the general recursions simplify considerably is the linear state space model with Gaussian innovations. In this case the algorithm is the well-known Kalman filter. Monte Carlo methods such as particular filtering can be used to approximate general state space models in other cases (see for example Godsill et al. 2004).

2.3 Non parametric estimates for the state space model

Let us now describe the method that is used to estimate conditional pdf a and b as well as the filtering and smoothing densities.

Assume that $\{S_t, Y_t\}$ is a state space process satisfying **(M1)**−**(M2)**. This process is a first order Markov chain. We denote Q_{μ_0} the law of $\{S_t, Y_t\}$ when the initial distribution of the Markov chain is μ_0. Suppose also that we have:

- a realization $\hat{s}_1^n, \hat{y}_1^n$ of $\{S_t, Y_t\}$. At this stage, the state process $\{S_t\}$ is not hidden. In the sequel $(\hat{s}_1^n, \hat{y}_1^n)$ is referred to as the *learning sequence* and it will be used to estimate the pdf a of the transition kernel and b of the emission probabilities.
- a realization y_1^T of $\{Y_t\}$ while $\{S_t\}$ is hidden. y_1^T is referred to as the *observed sequence*. Our goal consists in forecasting $\{S_t\}$ given y_1^T.

When the smoothing density $f_{t|T}$ is known, the state s_t can be forecasted using $s_t^* = E(S_t|Y_1^T = y_1^T) = \int s f_{t|T}(s)\mu(ds)$ for each time $t \in \{1, \ldots, T\}$. Since, it is considered that $\{S_t\}$ and $\{Y_t\}$ are observed simultaneously during a learning period, a and b can be estimated using kernel density estimates $\hat{a}$ and $\hat{b}$. Estimates $\hat{\Pi}_F^{(t)}$ and $\hat{\Pi}_B^{(t)}$ of $\Pi_F^{(t)}$ and $\Pi_B^{(t)}$ can be deduced for all t and hence the estimates $\hat{f}_{t|t}$ and $\hat{f}_{t|T}$ of the filtering and smoothing densities $f_{t|t}$ and $f_{t|T}$ too.

More precisely, let us denote f the stationary pdf of the Markov chain $\{S_t\}$, q_1 the stationary joint pdf of (S_{t-1}, S_t) and q_2 the joint stationary pdf of (Y_t, S_t). We define the estimates of

f, q_1 and q_2 by

$$\hat{f}(s) = n^{-1}h_1(n)^{-d}\sum_{k=1}^{n} K_1((s-\hat{s}_k)h_1(n)^{-1})$$

$$\hat{q}_1(s',s) = n^{-1}h_2(n)^{-2d}\sum_{k=2}^{n} K_2((s'-\hat{s}_{k-1}, s-\hat{s}_k)h_2(n)^{-1})$$

$$\hat{q}_2(y,s) = n^{-1}h_3(n)^{-(d+d')}\sum_{k=1}^{n} K_3((y-\hat{y}_k, s-\hat{s}_k)h_3(n)^{-1})$$

where K_1, K_2 and K_3 are kernel functions with some properties specified later on (see hypothesis $[\mathcal{K}(\zeta)]$) and where $h_1(n)$, $h_2(n)$ and $h_3(n)$ are bandwidth parameters.

We deduce estimates of transition and emission pdf as follows:

$$\hat{a}(s|s') = \begin{cases} \frac{\hat{q}_1(s',s)}{\hat{f}(s)} & \text{if } \hat{f}(s) > \gamma n^{-1} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$\hat{b}(y|s) = \begin{cases} \frac{\hat{q}_2(y,s)}{\hat{f}(s)} & \text{if } \hat{f}(s) > \gamma n^{-1} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

with $\gamma > 0$.

Now, given $\hat{a}$ and $\hat{b}$, one can deduce a non parametric estimate $\hat{\Pi}_F^{(t)}$ of the forward operator $\Pi_F^{(t)}$, for all observed time series y_1^t, by substituting $\hat{a}$ and $\hat{b}$ to a and b in Eq. 2,

$$\hat{\Pi}_F^{(t)}\xi(s_t) = \frac{\int \hat{a}(s_t|s_{t-1})\hat{b}(y_t|s_t)\xi(s_{t-1})\mu(ds_{t-1})}{\int\int \hat{a}(s|s_{t-1})\hat{b}(y_t|s)\xi(s_{t-1})\mu(ds_{t-1})\mu(ds)} \tag{6}$$

and then define the following recursive estimate of the filtering density, for $t \geq 1$

$$\hat{f}_{t|t} = \hat{\Pi}_F^{(t)}\hat{f}_{t-1|t-1}$$

For $t = 0$, $\hat{f}_{0|0}$ is chosen arbitrarily (cf Sect. 3.4).

In the same way, a non parametric estimate of the backward operator can be defined by

$$\hat{\Pi}_B^{(t)}\xi(s_t) = \hat{f}_{t|t}(s_t)\int \frac{\hat{a}(s_{t+1}|s_t)}{\hat{f}_{t+1|t}(s_{t+1})}\xi(s_{t+1})\mu(ds_{t+1}) \tag{7}$$

with

$$\hat{f}_{t+1|t}(s_{t+1}) = \int \hat{a}(s_{t+1}|s_t)\hat{f}_{t|t}(s_t)\mu(ds_t)$$

And $\hat{f}_{t|T}$ is defined recursively, starting with $\hat{f}_{T|T}$, by

$$\hat{f}_{t|T} = \hat{\Pi}_B^{(t)}\hat{f}_{t+1|T}$$

The convergence results, demonstrated later on, induce that $\hat{a}$ and $\hat{b}$ inherit properties $[\mathcal{A}]$ for n large enough. Hence, there exists an integer n_0 such that operators $\hat{\Pi}_F^{(t)}$ and $\hat{\Pi}_B^{(t)}$ are well-defined for $n > n_0$, Q_{μ_0} almost surely. We suppose in the sequel that this last condition is verified.

3 L^1 convergence of filtering densities

In this section, the L^1 convergence of the non parametric estimates of the filtering densities is studied when length n of the learning sequence tends to infinity.

The L^1 convergence of $\hat{f}_{t|t}$ to $f_{t|t}$ is obtained by using two main arguments, i.e. the convergence of $\hat{\Pi}_F^{(t)}$ to $\Pi_F^{(t)}$ (Proposition 1) and the exponential forgetting of the initial distribution for the forward operator Π_F (Proposition 2) which enable us to control the growth of the error $||\hat{f}_{t|t} - f_{t|t}||_1$ when t increases.

3.1 Hypothesis

Let us now list the assumptions which are made throughout the paper to get the uniform convergence of the kernel estimates.

We make the following assumptions on the regularity of the pdf and on the mixing properties of the Markov chain:

[$\mathcal{B}$]
1. The stationary pdf f, the two-dimensional stationary pdf q_1 and the joint density q_2 admits bounded derivatives (or partial derivatives) of order $\zeta \geq 2$.
2. $\inf_{s\in\mathcal{S}} f(s) \geq \kappa_f^- > 0$
3. The Markov Chain $\{S_t\}$ is geometrically ergodic.

Hypothesis [$\mathcal{A}$] and [$\mathcal{B}$] are partly redundant. However, they are presented in this form for the sake of simplicity. Assumption [$\mathcal{B}$]1 controls the regularity of the pdf f, q_1 and q_2 and it is required in order to get the uniform convergence of the non parametric estimates $\hat{a}$ and $\hat{b}$. [$\mathcal{B}$]1 and [$\mathcal{B}$]2 imply that a and b are bounded. We can also remark that, as $f(s) = \int a(s|s')f(s')\mu(ds')$, [$\mathcal{A}$]1 implies [$\mathcal{B}$]2 with $\kappa_f^- \geq \kappa_a^-$. And assumption [$\mathcal{A}$]1 implies the uniform ergodicity of $\{S_t\}$ and thus [$\mathcal{B}$]3.

Concerning the kernel functions, we assume that for some $\zeta \geq 2$

[$\mathcal{K}(\zeta)$]
1. K_1, K_2, K_3 are Lipschitz-continuous functions
2. $K_i(t) = 0$ for $t \notin [-1,1]^{d_i}$ and $\int_{[-1,1]^{d_i}} K_i(t)dt = 1$, for $i = 1, 2, 3$.
3. $\int_{[-1,1]^{d_i}} \prod_{j=1}^{l} z_{i_j} K_i(z_1, \ldots, z_{d_i})dz_1 \ldots dz_{d_i} = 0$ for every choice $i_1, \ldots, i_l \in \{1, \ldots, d_i\}, l = 1, \ldots, \zeta - 1$, for $i = 1, 2, 3$ with $d_1 = d, d_2 = 2d$ and $d_3 = d + d'$.

Assumptions [$\mathcal{K}(\zeta)$]1 and [$\mathcal{K}(\zeta)$]2 are common in kernel estimation (see Liebscher 2001) and [$\mathcal{K}(\zeta)$]3 allows us to obtain appropriate convergence rate for the bias and to improve slightly former results, such as those reported in Bosq (1996).

Finally, the bandwidth parameters $h_1(n)$, $h_2(n)$ and $h_3(n)$ are supposed to satisfy

[$\mathcal{H}$] $h_i(n) = const. \left(\frac{\log(n)}{n}\right)^{\frac{1}{2\zeta+d_i}}$ for $i = 1, 2, 3$.

3.2 Convergence of the forward operator

In order to demonstrate Proposition 1, we need to establish the uniform convergence of $\hat{a}$ and $\hat{b}$ to the conditional pdf a and b.

Lemma 1 *Under assumption* $[\mathcal{B}]$ *and if hypothesis* $[\mathcal{H}]$ *and* $[\mathcal{K}(\zeta)]$ *are verified with* $\zeta \geq 2$ *then,*

$$r_f(n) = \sup_{s\in\mathcal{S}} |\hat{f}(s) - f(s)| = O\left(\left(\frac{\log(n)}{n}\right)^{\zeta/(2\zeta+d)}\right) \ Q_{\mu_0} - a.s. \tag{8}$$

$$r_a(n) = \sup_{s,s'\in\mathcal{S}} |\hat{a}(s'|s) - a(s'|s)| = O\left(\left(\frac{\log(n)}{n}\right)^{\zeta/(2\zeta+2d)}\right) \ Q_{\mu_0} - a.s. \tag{9}$$

$$r_b(n) = \sup_{s\in\mathcal{S},y\in\mathcal{Y}} |\hat{b}(y|s) - b(y|s)| = O\left(\left(\frac{\log(n)}{n}\right)^{\zeta/(2\zeta+d+d')}\right) \ Q_{\mu_0} - a.s. \tag{10}$$

Proof of Lemma 1 The proof of Lemma 1 follows the same scheme as the one described in Liebscher (1999, 2001), therefore it is not developed in details here. One can notice that assumptions $[\mathcal{B}]$ induce the condition $\mathcal{I}$ given in Liebscher (2001). Convergence of $\hat{f}$ is proved in Liebscher (2001) and convergence of $\hat{a}$ to a is a direct application of Theorem 3.2 of Liebscher (1999). The convergence of $\hat{b}$ can be demonstrated in the same way.

Firstly, we remark that for any $\alpha, \hat{\alpha} \in \mathbb{R}$ and $\beta, \hat{\beta} \in \mathbb{R}^*$,

$$\left|\frac{\alpha}{\beta} - \frac{\hat{\alpha}}{\hat{\beta}}\right| \leq \frac{1}{|\beta\hat{\beta}|}(|\hat{\alpha}||\hat{\beta} - \beta| + |\hat{\beta}||\hat{\alpha} - \alpha|) \tag{11}$$

Then, we easily deduce that

$$\begin{aligned} &|\hat{b}(y|s) - b(y|s)| \\ &\leq \frac{1}{f(s)}\left(\frac{\hat{q}_2(y,s)|\hat{f}(s) - f(s)|}{\hat{f}(s)} + |\hat{q}_2(y,s) - q_2(y,s)|\right) \\ &\leq \frac{1}{\kappa_f^-}\left(\frac{\hat{q}_2(y,s)|\hat{f}(s) - f(s)|}{\hat{f}(s)} + |\hat{q}_2(y,s) - q_2(y,s)|\right) \end{aligned} \tag{12}$$

Secondly, assumptions $[\mathcal{B}]$ imply that $\{S, Y\}$ is a β-mixing process and thus the convergence properties of non parametric estimates of densities given in Liebscher (2001) apply here. Hence, we obtain the uniform almost sure convergence of $\hat{q}_2$ to q_2. These results also permit to bound $\hat{q}_2(y,s)\hat{f}(s)^{-1}$ and the conclusion follows from (12). □

Proposition 1 *Under assumptions* $[\mathcal{A}]$, $[\mathcal{B}]$, $[\mathcal{H}]$ *and* $[\mathcal{K}(\zeta)]$, *there is a constant* κ_F *such that for all* $t \in \{1, \ldots, T\}$, $y_1^t \in \mathcal{Y}^t$ *and all pdf* ξ *on* $\mathcal{S}$ *we have*

$$||(\hat{\Pi}_F^{(t)} - \Pi_F^{(t)})\xi||_1 \leq \kappa_F(r_a(n) + r_b(n)) \ \ Q_{\mu_0} - a.s.$$

Proof of Proposition 1 Let ξ be a pdf on $\mathcal{S}$ and $t \in \{1, \ldots, T\}$. By definition, we have

$$||(\Pi_F^{(t)} - \hat{\Pi}_F^{(t)})\xi||_1 = \int \left| \frac{\int b(y_t|s_t)a(s_t|s)\xi(s)\mu(ds)}{\int\int b(y_t|s')a(s'|s)\xi(s)\mu(ds)\mu(ds')} - \frac{\int \hat{b}(y_t|s_t)\hat{a}(s_t|s)\xi(s)\mu(ds)}{\int\int \hat{b}(y_t|s')\hat{a}(s'|s)\xi(s)\mu(ds)\mu(ds')} \right| \mu(ds_t) \tag{13}$$

Then, using inequality (11) and the relation between the numerators and the denominators in (13), we get

$$||(\Pi_F^{(t)} - \hat{\Pi}_F^{(t)})\xi||_1 \leq \frac{|\hat{\beta} - \beta|}{|\beta|} + \frac{1}{|\beta|} \int\int \left| b(y_t|s')a(s'|s) - \hat{b}(y_t|s')\hat{a}(s'|s) \right| \xi(s)\mu(ds)\mu(ds')$$

with $\beta = \int\int b(y_t|s')a(s'|s)\xi(s)\mu(ds)\mu(ds')$ and $\hat{\beta} = \int\int \hat{b}(y_t|s')\hat{a}(s'|s)\xi(s)\mu(ds)\mu(ds')$. Then

$$\begin{aligned}
&\int\int \left| b(y_t|s')a(s'|s) - \hat{b}(y_t|s')\hat{a}(s'|s) \right| \xi(s)\mu(ds)\mu(ds') \\
&\leq \int\int \hat{a}(s'|s) \left| b(y_t|s') - \hat{b}(y_t|s') \right| \xi(s)\mu(ds)\mu(ds') + \int\int b(y_t|s') \left| a(s'|s) \right. \\
&\quad \left. - \hat{a}(s'|s) \right| \xi(s)\mu(ds)\mu(ds') \\
&\leq r_b(n) \int\int \hat{a}(s'|s)\xi(s)\mu(ds)\mu(ds') + r_a(n) \int\int b(y_t|s')\xi(s)\mu(ds)\mu(ds') \\
&\leq r_b(n) + \kappa_b^+ r_a(n)
\end{aligned}$$

And, as a consequence,

$$\begin{aligned}
|\hat{\beta} - \beta| &\leq \int\int \left| b(y_t|s')a(s'|s) - \hat{b}(y_t|s')\hat{a}(s'|s) \right| \xi(s)\mu(ds)\mu(ds') \\
&\leq r_b(n) + \kappa_b^+ r_a(n)
\end{aligned}$$

Then, using assumptions $[\mathcal{A}]1$ and $[\mathcal{A}]2$, we can show that

$$\beta > \kappa_a^- \kappa_b^-$$

and finally, we get

$$||(\Pi_F^{(t)} - \hat{\Pi}_F^{(t)})\xi||_1 \leq 2 \frac{r_b(n) + \kappa_b^+ r_a(n)}{\kappa_a^- \kappa_b^-}$$

□

3.3 Exponential forgetting of initial conditions

The forward operator $\Pi_F^{(t)}$ is a composition of a contractant Markov operator and a Bayes operator, which is not necessarily contractant, and as a consequence $\Pi_F^{(t)}$ is not contractant in general. But, in order to get $f_{t|t}$ from $f_{r|r}$ for $t > r$, it is equivalent to apply $\Pi_F^{(t)} \ldots \Pi_F^{(r+1)}$ or to apply the Bayes operator once followed by $t - r$ Markov operators. The contractivity of the Markov operator can then beat the expansion of the Bayes operator. Such an idea was first proposed in Araposthatis and Marcus (1990) and was then extended in Douc and Matias (2001), Künsch (2001) and Le Gland and Mevel (2000).

Proposition 2 *Under assumptions* $[\mathcal{A}]$, *for all* $r \leq t$, *for any pdf* ξ *and* ξ' *on* $\mathcal{S}$

$$||\Pi_F^t \Pi_F^{t-1} \cdots \Pi_F^r (\xi - \xi')||_1 \leq C_a \left(1 - \frac{1}{C_a^2}\right)^{t-r+1} ||\xi - \xi'||_1$$

with $C_a = \frac{\kappa_a^+}{\kappa_a^-}$.

The proof of Proposition 2 can be found in Künsch (2001) (see Lemma 8 and Theorem 1).

3.4 Convergence of the filtering densities

The main result concerning the convergence of the filtering densities is given in the theorem below.

Theorem 1 *Under hypothesis* $[\mathcal{A}]$, $[\mathcal{B}]$, $[\mathcal{H}]$ *and* $[\mathcal{K}(\zeta)]$ *with* $\zeta \geq 2$ *there exists a constant* K_F, *such that for all* $T \in \mathcal{N}^*$, $t \in \{1, \ldots, T\}$ *and* $y_1^T \in \mathcal{Y}^T$, *we have*

$$||\hat{f}_{t|t} - f_{t|t}||_1 \leq K_F(r_a(n) + r_b(n)) + C_a \left(1 - \frac{1}{C_a^2}\right)^t ||f_{0|0} - \hat{f}_{0|0}||_1 \quad Q_{\mu_0} - a.s.$$

This theorem states that the rate of convergence of the filtering densities is the same as the slowest convergence rate obtained in Lemma 1 for the non parametric estimates, with a constant K_F which does not depend on T.

In practice, $f_{0|0}$ and $\hat{f}_{0|0}$ can be chosen arbitrarily. For instance, we can choose the stationary distributions, i.e. $f_{0|0} = f$, and its kernel density estimate $\hat{f}_{0|0} = \hat{f}$. In this case, $||f_{0|0} - \hat{f}_{0|0}||_1$ is controlled by $r_f(n)$. Another natural choice is $\hat{f}_{0|0} = f_{0|0}$ with $f_{0|0}$ some arbitrary pdf on $\mathcal{S}$. In this case, the term $||f_{0|0} - \hat{f}_{0|0}||_1$ vanishes. In all cases, the error made on the initial condition is forgotten at an exponential rate when t tends to infinity. In Sect. 4, we will assume, for the sake of simplicity, that $\hat{f}_{0|0} = f_{0|0}$.

Proof of Theorem 1 The L^1 norm of error on the filtering density is given by

$$||f_{t|t} - \hat{f}_{t|t}||_1 = ||\Pi_F^{(t)} f_{t-1|t-1} - \hat{\Pi}_F^{(t)} \hat{f}_{t-1|t-1}||_1$$

It can be bounded as follows:

$$||f_{t|t} - \hat{f}_{t|t}||_1 \leq ||(\Pi_F^{(t)} - \hat{\Pi}_F^{(t)}) \hat{f}_{t-1|t-1}||_1 + ||\Pi_F^{(t)} (f_{t-1|t-1} - \hat{f}_{t-1|t-1})||_1$$

By iterating, we obtain

$$\begin{aligned}
&||f_{t|t} - \hat{f}_{t|t}||_1 \\
&\leq ||(\Pi_F^{(t)} - \hat{\Pi}_F^{(t)}) \hat{f}_{t-1|t-1}||_1 + ||\Pi_F^{(t)} (\Pi_F^{(t-1)} - \hat{\Pi}_F^{(t-1)}) \hat{f}_{t-2|t-2}||_1 \\
&\quad + \cdots + ||\Pi_F^{(t)} \cdots \Pi_F^{(2)} (\Pi_F^{(1)} - \hat{\Pi}_F^{(1)}) \hat{f}_{0|0}||_1 + ||\Pi_F^{(t)} \cdots \Pi_F^{(1)} (f_{0|0} - \hat{f}_{0|0})||_1
\end{aligned}$$

Then, using Propositions 1 and 2, we get

$$\begin{aligned}
||f_{t|t} - \hat{f}_{t|t}||_1 &\leq \kappa_F(r_a(n) + r_b(n)) \left(1 + C_a \sum_{r=1}^{t-1} \left(1 - \frac{1}{C_a^2}\right)^r\right) \\
&\quad + C_a \left(1 - \frac{1}{C_a^2}\right)^t ||f_{0|0} - \hat{f}_{0|0}||_1 \\
&\leq K_F(r_a(n) + r_b(n)) + C_a \left(1 - \frac{1}{C_a^2}\right)^t ||f_{0|0} - \hat{f}_{0|0}||_1
\end{aligned}$$

with

$$K_F = \kappa_F \left(1 + C_a \left(1 - C_a^2\right)\right)$$

□

4 Convergence of the smoothing densities

The L^1-convergence of the smoothing densities $\hat{f}_{t|T}$ to $f_{t|T}$ is obtained using similar arguments as those used to get the convergence of the filtering densities. More precisely, we combine the facts that $\hat{\Pi}_B^{(t)}$ tends to $\Pi_B^{(t)}$ when n tends to infinity (Proposition 3) and that during the backward task of smoothing, the "future" is forgotten at an exponential rate (Proposition 4).

4.1 Convergence of the backward operator

Let us first consider the convergence of the backward operator.

Proposition 3 *Under assumptions* [$\mathcal{A}$], [$\mathcal{B}$], [$\mathcal{H}$] *and* [$\mathcal{K}(\zeta)$] *with* $\zeta > 2$, *there is a constant* κ_B *such that for all* $t \in \{1, \ldots, T\}$, $y_1^t \in \mathcal{Y}^t$ *and pdf* ξ *on* $\mathcal{S}$, *we have*

$$||(\hat{\Pi}_B^{(t)} - \Pi_B^{(t)})\xi||_1 \leq \kappa_B(r_a(n) + r_b(n))$$

Proof of Proposition 3 Let ξ be a pdf on $\mathcal{S}$ and $t \in \{1, \ldots, T\}$. By definition, we have

$$\begin{aligned}
&||(\hat{\Pi}_B^{(t)} - \Pi_B^{(t)})\xi||_1 \\
&\quad = \int \left| \int \left(\frac{f_{t|t}(s_t) a(s_{t+1}|s_t)}{f_{t+1|t}(s_{t+1})} - \frac{\hat{f}_{t|t}(s_t)\hat{a}(s_{t+1}|s_t)}{\hat{f}_{t+1|t}(s_{t+1})} \right) \xi(s_{t+1})\mu(ds_{t+1}) \right| \mu(ds_t)
\end{aligned}$$

Then, using inequality (11), we get

$$\begin{aligned}
&||(\hat{\Pi}_B^{(t)} - \Pi_B^{(t)})\xi||_1 \\
&\quad \leq \int \int \frac{f_{t|t}(s_t) a(s_{t+1}|s_t) \left| f_{t+1|t}(s_{t+1}) - \hat{f}_{t+1|t}(s_{t+1}) \right|}{f_{t+1|t}(s_{t+1}) \hat{f}_{t+1|t}(s_{t+1})} \xi(s_{t+1})\mu(ds_t)\mu(ds_{t+1}) \\
&\qquad + \int \int \frac{\left| f_{t|t}(s_t) a(s_{t+1}|s_t) - \hat{f}_{t|t}(s_t)\hat{a}(s_{t+1}|s_t) \right|}{\hat{f}_{t+1|t}(s_{t+1})} \xi(s_{t+1})\mu(ds_t)\mu(ds_{t+1}) \qquad (14)
\end{aligned}$$

Using Theorem 1, with $f_{0|0} = \hat{f}_{0|0}$, and Lemma 1 we can show that for all $s_{t+1} \in \mathcal{S}$,

$$\begin{aligned}
&\int \left| a(s_{t+1}|s_t) f_{t|t}(s_t)\mu(ds_t) - \int \hat{a}(s_{t+1}|s_t)\hat{f}_{t|t}(s_t) \right| \mu(ds_t) \\
&\quad \leq \int a(s_{t+1}|s_t) \left| f_{t|t}(s_t) - \hat{f}_{t|t}(s_t) \right| \mu(ds_t) + \int \hat{f}_{t|t}(s_t) \left| a(s_{t+1}|s_t) - \hat{a}(s_{t+1}|s_t) \right| \mu(ds_t) \\
&\quad \leq \kappa_a^+ K_F(r_a(n) + r_b(n)) + r_a(n)
\end{aligned}$$

As a consequence of the previous inequality,

$$\begin{aligned}
\left| f_{t+1|t}(s_{t+1}) - \hat{f}_{t+1|t}(s_{t+1}) \right| &= \left| \int a(s_{t+1}|s_t) f_{t|t}(s_t)\mu(ds_t) - \int \hat{a}(s_{t+1}|s_t)\hat{f}_{t|t}(s_t)\mu(ds_t) \right| \\
&\leq \kappa_a^+ K_F(r_a(n) + r_b(n)) + r_a(n) \qquad (15)
\end{aligned}$$

We have $f_{t+1|t}(s_{t+1}) = \int f_{t|t}(s_t)a(s_t|s_{t+1})\mu(ds_t) > \kappa_a^-$ for all s_{t+1}. Then Eq. 15 implies that we can find n_0 such that for $n \geq n_0$, $\hat{f}_{t+1|t}(s_{t+1}) > \kappa_a^-/2$.

Finally, we have, for $n \geq n_0$,

$$
\begin{aligned}
&||(\hat{\Pi}_B^{(t)} - \Pi_B^{(t)})\xi||_1 \\
&\leq \frac{\kappa_a^+}{\kappa_a^-\kappa_a^-/2} \int\int f_{t|t}(s_t) \left| f_{t+1|t}(s_{t+1}) - \hat{f}_{t+1|t}(s_{t+1}) \right| \xi(s_{t+1})\mu(ds_t)\mu(ds_{t+1}) \\
&\quad + \frac{1}{\kappa_a^-/2} \int\int \left| f_{t|t}(s_t)a(s_{t+1}|s_t) - \hat{f}_{t|t}(s_t)\hat{a}(s_{t+1}|s_t) \right| \xi(s_{t+1})\mu(ds_t)\mu(ds_{t+1}) \\
&\leq \frac{\kappa_a^+ \left(\kappa_a^+ K_F(r_a(n) + r_b(n)) + r_a(n)\right)}{\kappa_a^-\kappa_a^-/2} \int\int f_{t|t}(s_t)\xi(s_{t+1})\mu(ds_t)\mu(ds_{t+1}) \\
&\quad + \frac{\left(\kappa_a^+ K_F(r_a(n) + r_b(n)) + r_a(n)\right)}{\kappa_a^-/2} \int \xi(s_{t+1})\mu(ds_{t+1}) \\
&\leq \frac{\kappa_a^+ K_F(r_a(n) + r_b(n)) + r_a(n)}{\kappa_a^-/2} \left(\frac{\kappa_a^+}{\kappa_a^-} + 1\right)
\end{aligned}
$$

□

4.2 Exponential forgetting for the backward operator

Proposition 4 states that the backward operator $\Pi_B^{(t)}$ is contractant. As far as we know, there is no former result on this in the literature.

Proposition 4 *Under hypothesis* [$\mathcal{A}$], *for any pdf* ξ *and* ξ' *on* $\mathcal{S}$,

$$||\Pi_B^{(t)}(\xi - \xi')|| \leq \rho_B ||\xi - \xi'||_1$$

with $\rho_B = 1 - \left(\frac{\kappa_a^-}{\kappa_a^+}\right)^2$.

Proof of Proposition 4 Let us first remark that for any pdf ξ on $\mathcal{S}$, we have

$$\Pi_B^{(t)}\xi(s_t) = \int_S q_B^{(t)}(s_t|s_{t+1})\xi(s_{t+1})\mu(ds_{t+1}) \tag{16}$$

with

$$q_B^{(t)}(s|s') = \frac{f_{t|t}(s)}{f_{t+1|t}(s')}a(s'|s)$$

Then for all $s, s', s'' \in \mathcal{S}$

$$
\begin{aligned}
\frac{q_B^{(t)}(s|s')}{q_B^{(t)}(s|s'')} &= \frac{a(s'|s)\, f_{t+1|t}(s'')}{a(s''|s)\, f_{t+1|t}(s')} \\
&\leq \left(\frac{\kappa_a^+}{\kappa_a^-}\right)^2
\end{aligned}
$$

The proof of the proposition follows using Lemma 2 below. □

Lemma 2 *If there exists* $C_b > 0$ *such that for all* $s, s', s'' \in \mathcal{S}$

$$\frac{q_B^{(t)}(s|s')}{q_B^{(t)}(s|s'')} \leq C_b$$

then for all pdf ξ *and* $\xi\prime$ *on* $\mathcal{S}$ *we have*

$$||\Pi_B^{(t)}\xi - \Pi_B^{(t)}\xi\prime||_1 \leq \left(1 - \frac{1}{C_b}\right)||\xi - \xi\prime||_1$$

A similar lemma is demonstrated in Douc and Matias (2001) and Künsch (2001) .

4.3 Convergence of the smoothing densities

We can now state the theorem on the convergence of the smoothing densities.

Theorem 2 *Under hypothesis* [$\mathcal{A}$], [$\mathcal{B}$], [$\mathcal{H}$] *and* [$\mathcal{K}(\zeta)$], *for* $\zeta \geq 2$, *there exists a constant* K_B *such that for all* $T \in \mathbb{N}^*$, $t \in \{1, \ldots, T\}$ *and* $y_1^T \in \mathcal{Y}^T$

$$||f_{t|T} - \hat{f}_{t|T}||_1 \leq K_B(r_n(a) + r_n(b))\ Q_{\mu_0} - a.s. \tag{17}$$

Let us give some direct consequences of this theorem. We abusively denote, for $k \geq 1$, $\hat{E}[(S_t)^k|Y_1^T = y_1^T]$ the conditional moments of order k corresponding to $\hat{f}_{t|T}$. Under assumptions of Theorem 2, we have for all $T \in \mathbb{N}^*$, $t \in \{1, \ldots, T\}$ and y_1^T

$$|E[(S_t)^k|Y_1^T = y_1^T] - \hat{E}[(S_t)^k|Y_1^T = y_1^T]| \leq K_B \sup_{s\in\mathcal{S}}(|s|^k)(r_a(n) + r_b(n))\ Q_{\mu_0} - a.s.$$

It is also easy to check that for all $\alpha \in]0, 1[$, $\hat{F}_{t|T}^{-1}(\alpha) \to F_{t|T}^{-1}(\alpha)\ Q_{\mu_0} - a.s.$ when $n \to \infty$. Here $F_{t|T}$ and $\hat{F}_{t|T}$ denote the cumulative distribution functions associated with $f_{t|T}$ and $\hat{f}_{t|T}$ respectively. It justifies the use of $\hat{F}_{t|T}^{-1}$ to compute prediction intervals.

Proof of Theorem 2 Using Propositions 3 and 4, we can show that

$$\begin{aligned}
||f_{t|T} - \hat{f}_{t|T}||_1 &= ||\Pi_B^{(t)} f_{t+1|T} - \hat{\Pi}_B^{(t)} \hat{f}_{t+1|T}||_1 \\
&\leq ||(\Pi_B^{(t)} - \hat{\Pi}_B^{(t)})\hat{f}_{t+1|T}||_1 + ||\Pi_B^{(t)}(f_{t+1|T} - \hat{f}_{t+1|T})||_1 \\
&\leq \kappa_B(r_a(n) + r_b(n)) + \rho_B||f_{t+1|T} - \hat{f}_{t+1|T}||_1
\end{aligned}$$

By iteration, we deduce that

$$||f_{t|T} - \hat{f}_{t|T}||_1 \leq \kappa_B(r_a(n) + r_b(n))\Sigma_{k=0}^{T-t+1}\rho_B^k + \rho_B^{T-t}||f_{T|T} - \hat{f}_{T|T}||_1$$

Finally, applying Theorem 1 with $f_{0|0} = \hat{f}_{0|0}$, we get

$$||f_{t|T} - \hat{f}_{t|T}||_1 \leq \frac{\kappa_B}{1 - \rho_B}(r_a(n) + r_b(n)) + \rho_B^{T-t} K_F(r_a(n) + r_b(n))$$

□

5 Simulation results

In this section, we illustrate the results demonstrated in the previous sections with simulations.

Let us first define the state space model that has been used in this section. The hidden state space is chosen as the unit circle $\mathcal{S} = \mathbb{R}/2\pi\mathbb{Z}$, and we assume that $\{S_t\}$ is a von Mises process with transition kernel

$$a(s_t|s_{t-1}) = \frac{1}{I(\rho)} e^{\rho \cos(s_t - s_{t-1})}$$

with $\rho > 0$ and I the modified Bessel function of order 0.

We also assume that the observed process $\{Y_t\}$ takes its values in $\mathcal{Y} = \mathbb{R}/2\pi\mathbb{Z}$ and that the emission probabilities are parameterized using the von Mises distribution

$$b(y_t|s_t) = \frac{1}{I(\rho')} e^{\rho' \cos(y_t - s_t)}$$

with $\rho' > 0$.

It is easy to check that conditions $[\mathcal{A}]$ and $[\mathcal{B}]$ are verified by this state space model.

Then, we have carried out the following numerical experiment:

- We have simulated a learning sequence $\hat{s}_1^n$, $\hat{y}_1^n$ from the state space model, with $\rho = \rho' = 1$ and n increasing from 500 to 20,000 with a step equal to 500.
- We have computed the corresponding non-parametric estimates $\hat{a}$ and $\hat{b}$ of a and b. We have used Epanechnikov kernels (see Bosq 1996) for K_1,K_2 and K_3, and these kernels satisfy $[\mathcal{K}(\zeta)]$ with $\zeta = 2$. The bandwidth parameters $h_1(n)$, $h_2(n)$ and $h_3(n)$ have been chosen according to $[\mathcal{H}]$. Then, we have computed the quantities $r_a(n) = \sup_{s,s' \in \mathcal{S}} |\hat{a}(s'|s) - a(s'|s)|$ and $r_b(n) = \sup_{s \in \mathcal{S}, y \in \mathcal{Y}} |\hat{b}(y|s) - b(y|s)|$. The results are shown in Fig. 1. According to Lemma 1, we have $r_a(n) = O\left(\left(\frac{\log(n)}{n}\right)^{1/3}\right)$ and $r_b(n) = O\left(\left(\frac{\log(n)}{n}\right)^{1/3}\right)$. The function $\left(\frac{\log(n)}{n}\right)^{1/3}$ is also plotted in Fig. 1 for comparison purpose, and the agreement with $r_a(n)$ and $r_b(n)$ is good.
- We have simulated an observed sequence y_1^T with the "true" state space model and $T = 1000$. Then, we have computed the corresponding "true" and "estimated" filtering and smoothing probabilities. In practice, we have used the forward–backward algorithm to compute these probabilities, after having discretized the hidden state space $\mathcal{S}$. According to Theorems 1 and 2, the speed of convergence of these estimates should be close to that of the non parametric estimates and Fig. 1 illustrates that the L^1 norms of the errors $||\hat{f}_{t|t} - f_{t|t}||_1$ and $||\hat{f}_{t|T} - f_{t|T}||_1$ are of the same order than $r_a(n)$ and $r_b(n)$.

6 Application to meteorological data

The significant wave height is an important parameter for coastal and offshore engineering (reliability, fatigue, ...). When direct measurements of this parameter are not available for some specific location, a wind-wave model is generally used. Most of the time, numerical models based on physical considerations are used (see Liu et al. 2002), but it is well-known that they are often inaccurate in coastal areas, because of the lack of complete physical process modeling. The computational cost of these methods is also very high. An alternative consists in using stochastic models as explained below.

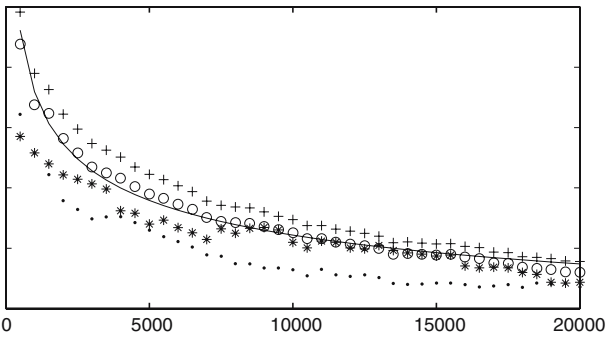

Fig. 1 Comparison of the convergence rate for the non parametric estimates ($r_a(n)$ (***) and $r_b(n)$ (...)) and the filtering (ooo) and smoothing (+++) densities. The different curves have been scaled to help the visual comparison. The solid line is proportional to $(\log(n)/n)^{\frac{1}{3}}$

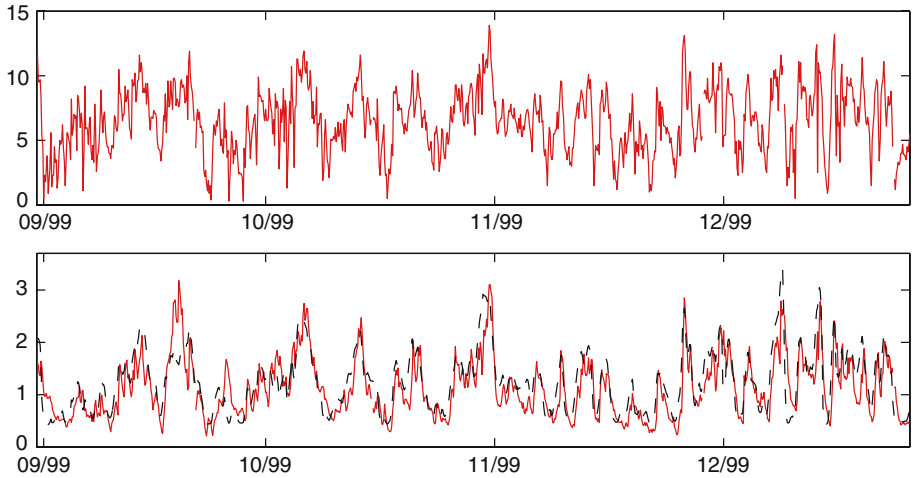

Fig. 2 Wind intensity (top) and significant wave height (bottom). *Solid line*: observed time series, *dotted line*: forecasted time series

We will denote $S_t \geq 0$ the significant wave height at time t at some specific location and $\{Y_t\} = \{u_t, v_t\}$ where u_t and v_t denote, respectively, the zonal and meridional component of the wind at time t at the same location. We will assume that $\{S_t, Y_t\}$ follows a state space model. In practice, we have considered data from the buoy 42039 (Pensacola, Gulf of Mexico), with geographical coordinates (28.80 N, 86.06 W). These data are available through the NOAA ftp server (ftp://polar.wwb.noaa.gov/). The data set is split in two parts: the first one (3 years) is used as a learning sequence and the second one (4 months) is used for validation. The state sequence is forecasted by the conditional expectation of S_t given the observed sequence y_1^t.

Figure 2 illustrates how the forecasted time series matches the observed one. The agreement is generally good and the forecasted time series restores the most important features for the applications (peak occurrences, peak amplitude, calm weather durations, ...).

These results have been compared with those obtained with linear regression models and neural networks, and we have obtained better results with the non parametric methodology

proposed in this paper. The results obtained with these different models will be discussed in a forthcoming paper.

References

Akaike H (1974) Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. Ann Inst Statist Math 26:363–387

Araposthatis A, Marcus SI (1990) Analysis of an identification algorithm arising in the adaptative estimation of Markov chains. Math Control Signal Syst 3:1–19

Atar R, Zeitouni O (1997) Exponential stability for nonlinear filtering. Annales de l'institut Henri Poincaré (B) 33(6):697–725

Baum LE, Petrie T (1966) Statistical inference for probabilistic functions of finite state Markov chains. Ann Math Stat 37:1554–1563

Bosq D (1996) Non parametric statistics for stochastic processes. Springer Verlag

Chigansky P, Liptser R (2004) Stability of nonlinear filters in nonmixing case. Ann Appl Prob 14(4):2038–2056

Del Moral P, Guionnet A (2001) On the stability of interacting processes with applications to filtering and genetic algorithms. Annales de l'Institut Henri Poincaré (B) 37(2):155–194

Douc R, Matias C (2001) Asymptotics of the maximum likelihood estimator for general hidden Markov models. Bernouilli 7(3):381–420

Godsill SJ, Doucet A, West M (2004) Monte Carlo smoothing for non-linear time series. J Am Stat Assoc 50:438–449

Harrison BJ, Stevens CF (1976) Bayesian forecasting (with discussion). J Roy Stat Soc Ser B 38(3):205–229

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82(Ser D):35–45

Künsch HR (2001) State space and hidden Markov models. Chapman & Hall

Le Gland F, Mevel L (2000) Exponential forgetting and geometric ergodicity in hidden Markov models. Math Control Signal Syst 13(1):66–93

Liebscher E (1999) Asymptotic properties of the kernel estimator for the transition density of a Markov chain. Preprint M 25/99 TU Ilmenau

Liebscher E (2001) Estimation of the density and the regression function under mixing conditions. Stat Decision 19(1):9–26

Liu PC, Schwab DJ, Bennett JR (2002) Has wind-wave modeling reached its limit? Ocean Eng 29:81–98

Meyn SP, Tweedie RL (1993) Markov chains and stochastic stability. Springer-Verlag