Electronic Journal of Statistics
Vol. 3 (2009) 239–258
ISSN: 1935-7524
DOI: 10.1214/08-EJS256

Online data processing: Comparison of Bayesian regularized particle filters*

Roberto Casarin[†,‡]

University of Brescia, Dept. of Economics, Via S. Faustino, 74/B, 25122 Brescia, Italy
e-mail: casarin@eco.unibs.it

and

Jean-Michel Marin[§,¶,‖]

CC51, Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier cedex 5, France
e-mail: Jean-Michel.Marin@univ-montp2.fr

Abstract: The aim of this paper is to compare three regularized particle filters in an online data processing context. We carry out the comparison in terms of hidden states filtering and parameter estimation, considering a Bayesian paradigm and a univariate Stochastic Volatility (SV) model. We discuss the use of an improper prior distribution in the initialization of the filtering procedure and show that the regularized Auxiliary Particle Filter (APF) outperforms the regularized Sequential Importance Sampling (SIS) and the regularized Sampling Importance Resampling (SIR).

AMS 2000 subject classifications: Primary 65C60.
Keywords and phrases: Online data processing, Bayesian estimation, regularized particle filters, Stochastic Volatility models.

Received June 2008.

Contents

*This work has been supported by the french Agence Nationale de la Recherche (ANR) through the 2006-2008 project Adap'MC.

[†]Department of Economics, University of Brescia, Italy.
[‡]GRETA Associati, Venice, Italy.
[§]Corresponding author.
[¶]Institut de Mathématiques et Modélisation, Université Montpellier 2, France.
[‖]CREST, INSEE, Paris, France.

1. Introduction

The analysis of phenomena, which evolve over time is a common problem to many fields like engineering, physics, biology, statistics, economics and finance. A time varying system can be represented through a dynamic model, which is constituted by an observable component and an unobservable internal state. The hidden states (or latent variables) represent the information we want to extrapolate from the observations.

In time series analysis, many approaches have been used for the estimation of dynamic models. The seminal works of Kalman (1960) and Kalman and Bucy (1960) introduce filtering techniques (the Kalman-Bucy filter) for continuous valued, linear and Gaussian dynamic systems. Maybeck (1982) motivates the use of stochastic dynamic systems in engineering and examines the estimation problems for state space models, in both a continuous and a discrete time framework. In economics, Harvey (1989) studies the state space representation of dynamic structural models and uses Kalman filter for hidden states filtering. Hamilton (1989) analyzes nonlinear time series models and introduces a filter (Hamilton-Kitagawa filter) for discrete time and discrete valued dynamic systems with a finite number of states.

In this paper, the online data processing problem is considered. In these situations, as pointed out by Liu and Chen (1998), Markov Chain Monte Carlo (MCMC) samplers are much too time demanding. To overcome this difficulty, some sequential Monte Carlo techniques have been recently developed. Doucet et al. (2001) provide the state of the art on these methods. They discuss both applications and theoretical convergence of the algorithms.

The contribution of this work is the comparison of three types of regularized particle filters - the regularized Sequential Importance Sampling (SIS), the regularized Sampling Importance Resampling (SIR) and the regularized Auxiliary Particle Filter (APF) - when the model parameters are unknown. The online estimation of model parameters is a difficult task Kitagawa (1998); Storvik (2002); Berzuini and Gilks (2001); Fearnhead (2002); Djuric et al. (2002); Storvik (2002); Andrieu and Doucet (2003); Doucet and Tadic (2003); Polson et al. (2002). We consider here the Bayesian paradigm and the regularization (see Chen and Haykin (2002)) approach of Oudjane (2000); Liu and West (2001); Musso et al. (2001); Rossi (2004) based on a kernel approximation in the parameter-augmented state space. We also discuss the initialization of the filtering procedure.

This work is structured as follow. Section 2 introduces the general representation of a Bayesian dynamic model and presents the Stochastic Volatility (SV) model. Section 3 reviews some regularized particle filters, presents their application to the stochastic volatility model and discusses some methodological issues. Finally, Section 4 gives the results.

2. Bayesian dynamic models

We introduce the general formulation of a Bayesian dynamic model and show some fundamental relations for Bayesian inference on it. Our definition of dynamic models is general enough to include the models analyzed in Kalman (1960), Hamilton (1994), Carter and Kohn (1994), Harrison and West (1989) and in Doucet et al. (2001). Throughout this work, we use a notation similar to that one commonly used in particle filter literature (see Doucet et al. (2001)).

We denote by $\{\mathbf{x}_t;\, t \in \mathbb{N}\}$, $\mathbf{x}_t \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$, the hidden states of the system, by $\{\mathbf{y}_t;\, t \in \mathbb{N}_0\}$, $\mathbf{y}_t \in \mathcal{Y} \subseteq \mathbb{R}^{n_y}$, the observable variables and by $\{\boldsymbol{\theta}_t;\, t \in \mathbb{N}\}$, $\boldsymbol{\theta}_t \in \Theta \subseteq \mathbb{R}^{n_{\boldsymbol{\theta}}}$, the parameters of the model. We denote by $\mathbf{x}_{0:t} = (\mathbf{x}_0, \ldots, \mathbf{x}_t)$ the collection of hidden states up to time t and with $\mathbf{x}_{-t} = (\mathbf{x}_0, \ldots, \mathbf{x}_{t-1}, \mathbf{x}_{t+1}, \ldots, \mathbf{x}_T)$ the collection of all hidden states without the t-th element. We use the same notations for the observable variables and parameters.

The Bayesian state space representation of a dynamic model is given by:

$$\begin{aligned}
\mathbf{y}_t &\sim p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{\theta}_t, \mathbf{y}_{1:t-1}) && \text{measurement density}\,,\\
(\mathbf{x}_t, \boldsymbol{\theta}_t) &\sim p(\mathbf{x}_t, \boldsymbol{\theta}_t|\mathbf{x}_{0:t-1}, \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t-1}) && \text{transition density}\,,\\
\mathbf{x}_0 &\sim p(\mathbf{x}_0|\boldsymbol{\theta}_0) && \text{initial density}\,,\\
\boldsymbol{\theta}_0 &\sim \pi(\boldsymbol{\theta}_0) && \text{prior density}\,,
\end{aligned}$$

for $t = 1, \ldots, T$.

In the following we suppose that

$$p(\mathbf{x}_t, \boldsymbol{\theta}_t|\mathbf{x}_{0:t-1}, \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t-1}) = p(\mathbf{x}_t, \boldsymbol{\theta}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}_{t-1})\,.$$

We also assume that the parameters are constant over time: the transition density of the parameters is then $\delta_{\boldsymbol{\theta}_{t-1}}(\boldsymbol{\theta}_t)$ with initial value $\boldsymbol{\theta}_0 = \boldsymbol{\theta}$, $\delta_x(y)$ denotes the Dirac's mass centered in x.

In that case, the joint transition of hidden states and parameters is:

$$p(\mathbf{x}_t, \boldsymbol{\theta}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}_{t-1}) = p(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}_t)\delta_{\boldsymbol{\theta}_{t-1}}(\boldsymbol{\theta}_t)\,.$$

Let us denote by $\mathbf{z}_t = (\mathbf{x}_t, \boldsymbol{\theta}_t)$ the parameter-augmented state vector and by $\mathcal{Z}$ the corresponding augmented state space. For such models, we are interested in the prediction and filtering densities which are given by:

$$p(\mathbf{z}_{t+1}|\mathbf{y}_{1:t}) = \int_{\mathcal{Z}} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \boldsymbol{\theta}_{t+1})\delta_{\boldsymbol{\theta}_t}(\boldsymbol{\theta}_{t+1})p(\mathbf{z}_t|\mathbf{y}_{1:t})d\mathbf{z}_t\,, \tag{2.1}$$

$$p(\mathbf{z}_{t+1}|\mathbf{y}_{1:t+1}) = \frac{p(\mathbf{y}_{t+1}|\mathbf{z}_{t+1}, \mathbf{y}_{1:t})p(\mathbf{z}_{t+1}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})}\,. \tag{2.2}$$

Due to the high number of integrals that must be solved, previous densities may be difficult to evaluate with general dynamics. Some Monte Carlo simulation methods, such as particle filters, allow us to overcome these difficulties.

As an example, let us consider the SV model. Two of the main features of the financial time series are time varying volatility and clustering phenomena in volatility. SV models widely used in finance have been introduced, in order to account for these features. Let y_t be the observable variable with time varying volatility and x_t the stochastic log-volatility process. An example of SV model is:

$$\begin{aligned} y_t|x_t &\sim \mathcal{N}\left(0, e^{x_t}\right) \\ x_t|x_{t-1}, \boldsymbol{\theta} &\sim \mathcal{N}\left(\alpha + \phi x_{t-1}, \sigma^2\right) \\ x_0|\boldsymbol{\theta} &\sim \mathcal{N}\left(0, \sigma^2/(1-\phi^2)\right) \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}) \end{aligned}$$

where $\boldsymbol{\theta} = (\alpha, \log((1+\phi)/(1-\phi)), \log(\sigma^2))$. The choice of $\pi(\boldsymbol{\theta})$ will be discussed in Section 3.5. Figure 1 shows two simulated paths of y_t and x_t. The paths correspond to different parameter settings, which can be usually found in financial applications. The first one is: $\alpha = 0$, $\phi = 0.99$ and $\sigma^2 = 0.01$ and corresponds to the case the data are sampled at a daily frequency. The second one is: $\alpha = 0$, $\phi = 0.9$ and $\sigma^2 = 0.1$, and corresponds to a weekly sampling frequency.

In the next section, we deal with the problem of parameter and states joint estimation in a kernel-regularized sequential Monte Carlo framework.

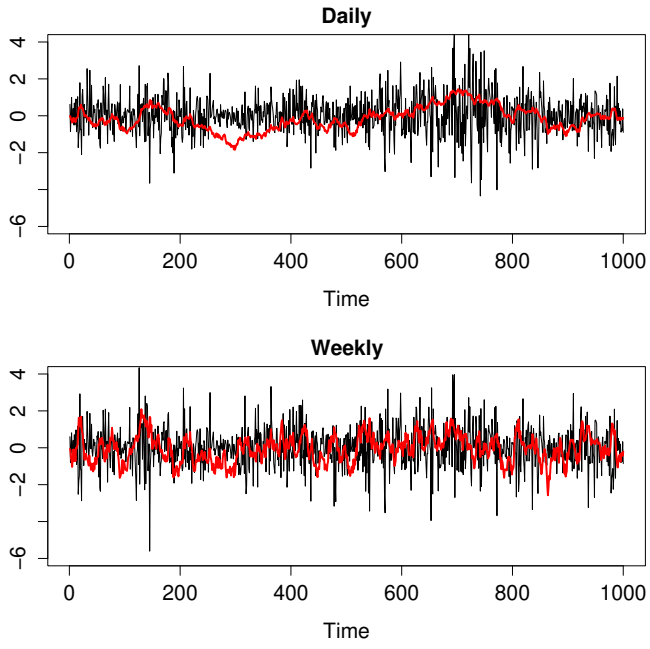

FIG 1. *Simulated paths for x_t (red line) and y_t (black line). Upper plot: daily dataset ($\alpha = 0$, $\phi = 0.99$ and $\sigma^2 = 0.01$). Bottom plot: weekly dataset ($\alpha = 0$, $\phi = 0.9$ and $\sigma^2 = 0.1$).*

3. Regularized particular filters

For making inference on the Bayesian dynamic model given in Section 2 in an online data processing context, MCMC algorithms are much too time demanding. Sequential importance sampling and more advanced sequential Monte Carlo algorithms called Particle Filters (Doucet et al., 2001) represent a promising alternative. The main advantage in using particle filters is that they can deal with nonlinear models and non-Gaussian innovations. In contrast to Hidden Markov Model filters, which work on a state space discretized to a fixed grid, particle filters focus sequentially on the higher density regions of the state space. This feature is common to one of the early sequential methods, the Adaptive Importance Sampling algorithm due to West (1992, 1993).

Different particle filters exist in the literature and different simulation approaches like rejection sampling, MCMC and importance sampling, can be used for the construction of a particle filter. In this work, we present some kernel-regularized particle filters, which combine the importance sampling reasoning with a suitable modification of the importance weights. The regularization approach we use is the same than the one of Liu and West (2001) and Musso et al. (2001). This approach relies upon a kernel-based reconstruction of the empirical filtering densities which produces a systematic modification of the true importance weights.

3.1. Regularized SIS

Let us start from the non-regularized SIS. We assume that at iteration $t > 0$ a properly weighted particle set $\{\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \gamma_t^i\}_{i=1}^N$, approximating the filtering density $p(\mathbf{x}_t, \boldsymbol{\theta}_t|\mathbf{y}_{1:t})$, is available. The empirical distribution corresponding to this approximation is:

$$p_N(\mathbf{x}_t, \boldsymbol{\theta}_t|\mathbf{y}_{1:t}) = \sum_{i=1}^N \gamma_t^i \delta_{(\mathbf{x}_t^i, \boldsymbol{\theta}_t^i)}(\mathbf{x}_t, \boldsymbol{\theta}_t)\,. \tag{3.1}$$

The particles set, $\{\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \gamma_t^i\}_{i=1}^N$, can be viewed as a random discretisation of the state space $\mathcal{X} \times \Theta$ with associated probability weights $\{\gamma_t^i\}_{i=1}^N$. Thanks to this discretisation, it is possible to approximate the prediction and filtering densities given in (2.1) and (2.2):

$$\begin{aligned} p_N(\mathbf{x}_{t+1}, \boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t}) &= \sum_{i=1}^N \gamma_t^i p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}) \delta_{\boldsymbol{\theta}_t^i}(\boldsymbol{\theta}_{t+1})\,, \\ p_N(\mathbf{x}_{t+1}, \boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t+1}) &\propto \sum_{i=1}^N \gamma_t^i p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \boldsymbol{\theta}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}) \delta_{\boldsymbol{\theta}_t^i}(\boldsymbol{\theta}_{t+1})\,. \end{aligned}$$

The goal is now to obtain N particles $\{\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i, \gamma_{t+1}^i\}_{i=1}^N$ from the filtering density in (2.2). It is proposed to sample $(\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i)$ according to the importance density $q(\cdot|\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \mathbf{y}_{1:t+1})$. The importance weight of particle $(\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i)$

is then calculated using the recursive formula:

$$\gamma_{t+1}^i \propto \gamma_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i) p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}^i) \delta_{\boldsymbol{\theta}_t^i}(\boldsymbol{\theta}_{t+1}^i)}{q(\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \mathbf{y}_{t+1})} . \tag{3.2}$$

The choice of an optimal importance density $q(\cdot|\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \mathbf{y}_{t+1})$, that is, a density which minimizes the variance of the importance weights is discussed in Pitt and Shephard (1999) and Crisan and Doucet (2000). In many cases, it is not possible to use this optimal importance density as the weight updating associated to this density does not admit a closed-form expression. In that case, the transition density of the parameter-augmented state vector represents a natural alternative for the importance density. Indeed, the transition density represents a sort of prior at time t for the parameter-augmented state vector $(\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i)$.

In our case, due to the presence of the Dirac point mass in the numerator of the weights it is impossible to modify over the filtering iterations the particle values for the parameters. In practice due to the loss of particle diversity in the parameter space, the weights will tend to zeros and of course stay zero forever, so we are facing a problem of degeneracy of the empirical filtering distribution. This scenario motivates particle filtering methods known as regularized particle filters. In order to avoid the degeneracy problem and to force the exploration of the parameter space toward regions which are not covered by the prior distribution, Liu and West (2001) and Musso et al. (2001) propose to use a regularized version of the filtering density. This approach results in the modification of the weights in (3.2) and the definition of a new set of weights:

$$\omega_{t+1}^i \propto \omega_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i) p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}^i) K_h\left(\boldsymbol{\theta}_{t+1}^i - (a\boldsymbol{\theta}_t^i + (1-a)\bar{\boldsymbol{\theta}}_t)\right)}{q(\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \mathbf{y}_{t+1})}$$

where $a \in [0,1]$ is shrinkage factor, $\bar{\boldsymbol{\theta}}_t$ is the empirical mean over the particle set at the time t and $K_h(y) = h^{-d}K(y/h)$ is a regularization kernel, K being a positive function defined on $\mathbb{R}^{n_{\boldsymbol{\theta}}}$ and h a positive smoothing factor (bandwidth).

The modification of the importance weights defined in (3.2) results from two steps. The first one is the regularization of the empirical density in (3.1) by a kernel estimator:

$$p_N^R(\mathbf{x}_t, \boldsymbol{\theta}_t|\mathbf{y}_{1:t}) = \sum_{i=1}^N \omega_t^i \delta_{\mathbf{x}_t^i}(\mathbf{x}_t) K_h\left(\boldsymbol{\theta}_t - (a\boldsymbol{\theta}_t^i + (1-a)\bar{\boldsymbol{\theta}}_t)\right) .$$

The second one is the application of an importance sampling argument to the approximated filtering density:

$$\begin{aligned} p_N^R(\mathbf{x}_{t+1}, \boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t+1}) = & \sum_{i=1}^N \omega_t^i p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \boldsymbol{\theta}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}) \\ & \times K_h\left(\boldsymbol{\theta}_{t+1} - (a\boldsymbol{\theta}_t^i + (1-a)\bar{\boldsymbol{\theta}}_t)\right) . \end{aligned}$$

The convergence results associated with this type of approximation are recalled in (Doucet et al., 2001), Musso et al. (2001) and Oudjane (2000). Under some regularity conditions on the kernel, when the number of particles increases

to infinity, the regularized empirical density converges to the right one for various criteria. For instance, we have $p_N^R \longrightarrow_{\mathcal{L}^2} p$.

Thanks to this approximation, the regularization kernel becomes the natural choice for the parameters proposal distribution. Thus, we sample $(\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i)$ according to:

$$q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}, \mathbf{y}_{t+1}) K_h\left(\boldsymbol{\theta}_{t+1} - (a\boldsymbol{\theta}_t^i + (1-a)\bar{\boldsymbol{\theta}}_t)\right) .$$

In that case, we have:

$$\omega_{t+1}^i \propto \omega_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i) p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}^i)}{q(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}^i, \mathbf{y}_{t+1})} .$$

In Algorithm 1, we give a pseudo-code representation of this method.

Algorithm 1. - ***Regularized SIS Particle Filter*** -

· *At time* $t = t_0$, *for* $i = 1, \ldots, N$, *simulate* $\mathbf{z}_{t_0}^i \sim p(\mathbf{z}_{t_0})$ *and set* $\omega_{t_0}^i = 1/N$

· *At time* $t_0 < t \leq T-1$, *given* $\{\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \omega_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$:

1. *Simulate* $\boldsymbol{\theta}_{t+1}^i \sim K_h\left(\boldsymbol{\theta}_{t+1} - (a\boldsymbol{\theta}_t^i + (1-a)\bar{\boldsymbol{\theta}}_t)\right)$
2. *Simulate* $\mathbf{x}_{t+1}^i \sim q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}^i, \mathbf{y}_{t+1})$
3. *Update the weights:* $\omega_{t+1}^i \propto \omega_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i) p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}^i)}{q(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}^i, \mathbf{y}_{t+1})}$.

Let us consider the SV model presented in Section 2.

Given the initial weighted random sample $\left\{x_{t_0}^i, \boldsymbol{\theta}_{t_0}^i, \omega_{t_0}^i\right\}_{i=1}^N$, where $\boldsymbol{\theta}_{t_0} = (\alpha_{t_0}, \log((1+\phi_{t_0})/(1-\phi_{t_0})), \log(\sigma_{t_0}^2))$, if we use the transition density as proposal distribution for the hidden states, the regularized SIS performs the following steps, for $t_0 < t \leq T-1$ and for $i = 1, \ldots, N$:

(i) `Simulate` $\boldsymbol{\theta}_{t+1}^i \sim \mathcal{N}\left(a\boldsymbol{\theta}_t^i + (1-a)\bar{\boldsymbol{\theta}}_t, h^2 V_t\right)$ `where` V_t `and` $\bar{\boldsymbol{\theta}}_t$ `are, respectively, the empirical covariance matrix and the empirical mean,` $a \in [0,1]$ `and` $h^2 = (1-a^2)$`,`

(ii) `Simulate` $x_{t+1}^i \sim \mathcal{N}\left(\alpha_{t+1}^i + \phi_{t+1}^i x_t^i, \left(\sigma^2\right)_{t+1}^i\right)$`,`

(iii) `Update the weights as follow`

$$w_{t+1}^i \propto w_t^i \exp\left\{-\frac{1}{2}\left[y_{t+1}^2 \exp\left(-x_{t+1}^i\right) + x_{t+1}^i\right]\right\} .$$

In the following we call SIS the previous scheme.

3.2. Regularized SIR

As it is well known in the literature (see for example Arulampalam et al. (2001)), basic SIS algorithms have a degeneracy problem. After some iterations the empirical distribution degenerates into a Dirac's mass on a single particle. This is due to the fact that the variance of the importance weights is non-decreasing over time (see Doucet et al. (2000)). In order to solve this degeneracy problem,

Gordon et al. (1993) introduce the SIR algorithm. This algorithm belongs to a wider class of bootstrap filters. At each iteration, a resampling step is used to generate a new set of particles. After this resampling step, the weights of the resampled particles are uniformly distributed over the particle set.

In the initial SIR, the resampling step is done at each iteration of the algorithm. This systematic resampling can introduce extra Monte Carlo variations, see Liu and Chen (1998). This can be reduced be doing resampling only when the Effective Sample Size (ESS) is small. The ESS measures the overall efficiency of an importance sampling algorithm. The ESS is a function of the coefficient of variation of the importance weights. At iteration t, the empirical ESS is:

$$\mathrm{ESS}_t = \frac{N}{1 + N\sum_{i=1}^{N}\left(\omega_t^i - N^{-1}\sum_{i=1}^{N}\omega_t^i\right)^2 \Big/ \left(\sum_{i=1}^{N}\omega_t^i\right)^2}.$$

In Algorithm 2, we give a pseudo-code representation of this method.

Algorithm 2. - ***Regularized SIR Particle Filter*** -

· *At time* $t = t_0$, *for* $i = 1, \ldots, N$, *simulate* $\mathbf{z}_{t_0}^i \sim p(\mathbf{z}_{t_0})$ *and set* $\omega_{t_0}^i = 1/N$

· *At time* $t_0 < t \leq T-1$, *given* $\{\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \omega_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$:

1. *Simulate* $\boldsymbol{\theta}_{t+1}^i \sim K_h\left(\boldsymbol{\theta}_{t+1} - (a\boldsymbol{\theta}_t^i + (1-a)\bar{\boldsymbol{\theta}}_t)\right)$
2. *Simulate* $\mathbf{x}_{t+1}^i \sim q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \tilde{\boldsymbol{\theta}}_{t+1}^i, \mathbf{y}_{t+1})$
3. *Update the weights:* $\omega_{t+1}^i \propto \omega_i^t \dfrac{p(\mathbf{y}_{t+1}|\tilde{\mathbf{x}}_{t+1}^i, \tilde{\boldsymbol{\theta}}_{t+1}^i)p(\tilde{\mathbf{x}}_{t+1}^i|\mathbf{x}_t^i, \tilde{\boldsymbol{\theta}}_{t+1}^i)}{q(\tilde{\mathbf{x}}_{t+1}^i|\mathbf{x}_t^i, \tilde{\boldsymbol{\theta}}_{t+1}^i, \mathbf{y}_{t+1})}$
4. *If* $ESS_{t+1} < \kappa$, *simulate* $\{\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i\}_{i=1}^N$ *from* $\{\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i, \omega_{t+1}^i\}_{i=1}^N$ *(Multinomial resampling) and set* $\omega_{t+1}^i = 1/N$.

Let us consider the SV model presented in Section 2. Given the initial weighted random sample $\left\{x_{t_0}^i, \boldsymbol{\theta}_{t_0}^i, \omega_{t_0}^i\right\}_{i=1}^N$, if we use the transition density as proposal distribution for the hidden states, the regularized SIR performs the following step, for $t_0 < t \leq T-1$ and for $i = 1, \ldots, N$:

(i) `Simulate` $\boldsymbol{\theta}_{t+1}^i \sim \mathcal{N}\left(a\boldsymbol{\theta}_t^i + (1-a)\bar{\boldsymbol{\theta}}_t, h^2 V_t\right)$ `where` V_t `and` $\bar{\boldsymbol{\theta}}_t$ `are the empirical covariance matrix and the empirical mean respectively and` $a \in [0,1]$ `and` $h^2 = (1-a^2)$,

(ii) `Simulate` $x_{t+1}^i \sim \mathcal{N}\left(\alpha_{t+1}^i + \phi_{t+1}^i x_t^i, \left(\sigma^2\right)_{t+1}^i\right)$,

(iii) `Update the weights`

$$w_{t+1}^i \propto w_t^i \exp\left\{-\frac{1}{2}\left[y_{t+1}^2 \exp\{-x_{t+1}^i\} + x_{t+1}^i\right]\right\},$$

(v) `If` $\mathrm{ESS}_{t+1} < \kappa$, `simulate` $\mathbf{z}_{t+1}^i \sim \sum_{j=1}^N w_{t+1}^j \delta_{\mathbf{z}_{t+1}^j}(\mathbf{z}_{t+1})$ `and set` $w_{t+1}^i = 1/N$.

If $\kappa = N$, the resampling step is done all the time. In that case, we call SIR the previous scheme. The value of $\kappa < N$ should be calibrated depending on the problem. After some numerical experiments, we have found that a good value for κ is $\kappa = 0.9 \times N$. In that case, the resampling step is done at regular time intervals and we called SIR-r the resulting algorithm.

3.3. Regularized APF

Due to the resampling step, the basic SIR algorithm produces a progressive impoverishment (loss of diversity) of the information contained in the particle set. To overcome this difficulty, many solutions have been proposed in the literature. We refer to the APF due to Pitt and Shephard (1999) and to the regularized APF algorithm due to Liu and West (2001). In order to avoid the resampling step, the APFs use the particle index (auxiliary variable) to select most representative particles in the proposal of the new particles. The regularized joint distribution of parameter-augmented state vector and the particle index is:

$$p_N^R(\mathbf{x}_{t+1}, \boldsymbol{\theta}_{t+1}, i|\mathbf{y}_{1:t+1}) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \boldsymbol{\theta}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \boldsymbol{\theta}_t^i) \times K_h\left(\boldsymbol{\theta}_{t+1} - (a\boldsymbol{\theta}_t^i + (1-a)\bar{\boldsymbol{\theta}}_t)\right) \omega_t^i .$$

A sample approximating that distribution can be obtained by using the proposal:

$$q(\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i, j^i|\mathbf{y}_{t+1}) = p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^{j^i}, \boldsymbol{\theta}_{t+1}^i) \times K_h\big(\boldsymbol{\theta}_{t+1}^i - (a\boldsymbol{\theta}_t^{j^i} + (1-a)\bar{\boldsymbol{\theta}}_t)\big) q(j^i|\mathbf{y}_{t+1}) ,$$

where

$$q(j^i|\mathbf{y}_{t+1}) \propto p(\mathbf{y}_{t+1}|\mu_{t+1}^{j^i}, m_{t+1}^{j^i}) w_t^{j^i} ,$$

$\mu_{t+1}^{j^i}$ and $m_{t+1}^{j^i}$ are evaluated using the initial particle set. Therefore, the importance weight of particle $(\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i, j^i)$ is:

$$\omega_{t+1}^i \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{j^i}, m_{t+1}^{j^i})} .$$

In Algorithm 3 we give a pseudo-code representation of the regularized APF.

Algorithm 3. - ***Regularized Auxiliary Particle Filter*** -

· *At time* $t = t_0$, *for* $i = 1, \ldots, N$, *simulate* $\mathbf{z}_{t_0}^i \sim p(\mathbf{z}_{t_0})$ *and set* $\omega_{t_0}^i = 1/N$

· *At time* $t_0 < t \leq T - 1$, *given* $\{\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \omega_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$:

1. *Simulate* $j^i \sim q(j|\mathbf{y}_{1:t+1})$ *with* $j \in \{1, \ldots, N\}$ *(Multinomial sampling) where* $\mu_{t+1}^j = \mathbb{E}(\mathbf{x}_{t+1}|\mathbf{x}_t^j, \boldsymbol{\theta}_t^j)$ *and* $m_{t+1}^j = \mathbb{E}(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t^j)$
2. *Simulate* $\boldsymbol{\theta}_{t+1}^i \sim K_h\big(\boldsymbol{\theta}_{t+1} - (a\boldsymbol{\theta}_t^{j^i} + (1-a)\bar{\boldsymbol{\theta}}_t)\big)$
3. *Simulate* $\mathbf{x}_{t+1}^i \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^{j^i}, \boldsymbol{\theta}_{t+1}^i)$
4. *Update particles weights:* $\omega_{t+1}^i \propto \dfrac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{j^i}, m_{t+1}^{j^i})}$.

We can say that, in the APF, the selection step is done before simulating the hidden states. This selection depends on the current value of the observable. Therefore:

- the APF is a standard way to construct a proposal distribution for the hidden states that depends on the current value of the particle;
- as we will see after, to use this selection step and the transition distribution as proposal distribution for the hidden states, results in a good proposal distribution.

When applied to the SV model and given the initial weighted random sample $\{x^i_{t_0}, \boldsymbol{\theta}^i_{t_0}, \omega^i_{t_0}\}_{i=1}^N$ the regularized APF performs the following steps, for $t_0 < t \leq T-1$ and for $i = 1, \ldots, N$:

(i) `Simulate` $j^i \sim q(j) \propto \sum_{k=1}^N w_t^k \mathcal{N}(y_{t+1}|\mu^k_{t+1})\delta_k(j)$ `where` $\mu^k_{t+1} = \phi^k_t x^k_t + \alpha^k_t$,

(ii) `Simulate` $\boldsymbol{\theta}^i_{t+1} \sim \mathcal{N}\left(a\boldsymbol{\theta}^{j^i}_t + (1-a)\bar{\boldsymbol{\theta}}_t, h^2 V_t\right)$ `where` V_t `and` $\bar{\boldsymbol{\theta}}_t$ `are the empirical variance matrix and the empirical mean respectively and` $a \in [0,1]$ `and` $h^2 = (1-a^2)$,

(iii) `Simulate` $x^i_{t+1} \sim \mathcal{N}\left(x_{t+1}|\alpha^i_{t+1} + \phi^i_{t+1} x^{j^i}_t, \left(\sigma^2\right)^i_{t+1}\right)$,

(iv) `Update the weights`

$$w^i_{t+1} \propto \exp\left\{-\frac{1}{2}\left[y^2_{t+1}\left(\exp\{-x^i_{t+1}\} - \exp\{-\mu^{j^i}_{t+1}\}\right) + x^i_{t+1} - \mu^{j^i}_{t+1}\right]\right\}.$$

Note that, following Pitt and Shephard (1999), one could alternatively use in the selection step a value of μ^k_{t+1} based on the Taylor expansion of the likelihood at time $t+1$. The parameter a is fixed following the usual optimal criterion.

3.4. Proposal distributions

In the APF algorithm the proposal distribution $q(\mathbf{x}^i_{t+1}, \boldsymbol{\theta}^i_{t+1}, j^i|\mathbf{y}_{t+1})$ depends on the current value of the observable variable. The information coming from $\mathbf{y}_{t+1}$ could help to improve the filtering procedure and is not used in the current version of the SIS and SIR algorithms. Thus, instead of the transition density $p(\mathbf{x}_{t+1}, \boldsymbol{\theta}_{t+1}|\mathbf{x}_t, \boldsymbol{\theta}_t)$, we suggest for SIS and SIR algorithms an alternative proposal distribution which depends on the current value of the observable.

For the SV model at hand we resort to the approach proposed by Shephard and Pitt (1997) for building the proposal distribution of a M.-H. algorithm for the hidden states of a similar SV model.

We consider the quadratic Taylor expansion of the term $\exp(-x_t)$ which appears in the conditional density (see also Celeux et al. (2006)):

$$p(x_t|x_{1:t-1}, y_{1:t}, \alpha, \phi, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}(x_t - \alpha - \phi x_{t-1})^2 - \frac{1}{2}\left(x_t + y_t^2 \exp(-x_t)\right)\right\}. \tag{3.3}$$

Let $m_t = \alpha + \phi x_{t-1}$, then the second-order Taylor approximation about m_t allows us to approximate the second term in Equation 3.3 as:

$$\exp\left\{-\frac{1}{2\sigma^2}(x_t - m_t)^2 - \frac{1}{2}\left(x_t + y_t^2\exp(-m_t)\left(1-(x_t-m_t)+\frac{1}{2}(x_t-m_t)^2\right)\right)\right\}, \tag{3.4}$$

that is proportional to the density of a Gaussian distribution with mean:

$$\left(\frac{1}{\sigma^2}m_t - \frac{1}{2}y_t^2\exp(-m_t)(1+m_t) - \frac{1}{2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{2}y_t^2\exp(-m_t)\right)^{-1} \tag{3.5}$$

and variance:

$$\left(\frac{1}{\sigma^2}+\frac{1}{2}y_t^2\exp(-m_t)\right)^{-1}. \tag{3.6}$$

The use of this Gaussian distribution as proposal defines new SIS and SIR algorithms which will be respectively denoted by SIS-p and SIR-p.

3.5. Initialization of the filters

The initialization represents a crucial step in applying particle filters in presence of unknown parameters. Classically, in the Bayesian paradigm, if the prior distribution of the parameters is proper, this distribution is used to initialize the filters (that corresponds to the case $t_0 = 0$). If the prior distribution is improper, we have to start the filter with an initial set of t_0 observations, such that the posterior distribution is well defined and to chose an importance sampling distribution to create a first properly weighted sample. This choice could be delicate. Indeed, even if the regularized particle filters have some theoretical stability properties with respect to the initial step, in our simulation experiments, we observed that a bad initialization can lead to poor performances in terms of filtering and parameters estimation. The initialization of particle filters in presence of unknown parameters have not yet been discussed in the literature.

To avoid the difficulty presented above, whatever the prior distribution is, we propose to initialize the filters with $t_0 > 0$ observations and to use a long MCMC run to create a sample with uniform weights:

$$\{\mathbf{x}_{t_0}^i, \boldsymbol{\theta}_{t_0}^i, 1/N\}_{i=1}^N.$$

For the Stochastic Volatility model, we consider the case with no prior information on the process of interest, that is a noninformative case. In this setting, we use an improper prior distribution that does not depend on any hyperparameters. For the parameters β^2, ϕ and σ^2, we assume the prior

$$\pi(\beta^2, \phi, \sigma^2) \propto 1/(\sigma\beta)\mathbb{I}_{(-1,1)}(\phi),$$

where $\beta^2 = e^\alpha$.

That gives:

$$\pi(\boldsymbol{\theta}) = \pi(\log(\beta^2), \log((1+\phi)/(1-\phi)), \log(\sigma^2)) = \pi(\theta_1, \theta_2, \theta_3) \propto \exp(\theta_1/2 + \theta_2 + \theta_3/2)/(1-\exp(\theta_2))^2.$$

This noninformative prior distribution has been introduced by Celeux et al. (2006): the parameter ϕ is constrained in the open interval $(-1, 1)$ (in order to impose the usual stationarity condition) and Jeffrey-like priors are used for the variance parameters β^2 and σ^2. For such an improper prior distribution, the posterior distribution is well-defined when the number of observations is greater than 2.

For the SV model, given the initial set of $t_0 \geq 2$ observations and the prior presented above, we obtain the following full conditional distributions:

$$\begin{aligned}
\beta^2|\cdots &\sim \mathcal{IG}\left(\sum_{t=1}^{t_0} y_t^2 \exp(-x_t)/2, (t_0-1)/2\right), \\
\sigma^2|\cdots &\sim \mathcal{IG}\left(\sum_{t=2}^{t_0} (x_t - \phi x_{t-1})^2/2 + x_1^2(1-\phi^2), (t_0-1)/2\right), \\
\pi(\phi|\cdots) &\propto (1-\phi^2)^{1/2} \exp\left(-\phi^2 \sum_{t=2}^{t_0-1} x_t^2 - 2\phi \sum_{t=2}^{t_0} x_t x_{t-1}\right) / 2\sigma^2 \mathbb{I}_{(-1,+1)}(\phi), \\
\pi(x_t|\cdots) &\propto \exp\left\{-\frac{1}{2\sigma^2}\left((x_t - \alpha - \phi x_{t-1})^2 - \right.\right. \\
&\qquad \left.\left.(x_{t+1} - \alpha - \phi x_t)^2\right) - \frac{1}{2}\left(x_t + y_t^2 \exp(-x_t)\right)\right\}.
\end{aligned}$$

The full conditional distributions of ϕ and x_t are not conventional and the standard Gibbs does not apply. We propose to use the Metropolis-Hastings within Gibbs algorithm studied in Celeux et al. (2006). A detailed description of the proposal distributions for ϕ and x_t can be found in Celeux et al. (2006). In that paper, the authors compare this MCMC scheme to an iterated importance sampling one. Note that one could alternatively use this iterated importance sampling algorithm to create a first weighted sample for the filters.

In our experiments we focus on the performances of the filters when there are well initialized. To this aim we use $t_0 = 100$ observations in the initialization step.

4. Simulation results for the SV model

In this section, we compare the regularized particle filters discussed in Section 3. First, in order to introduce some basic facts that characterize the behavior of the filters, we study their performances on the basis of single-run experiments. Secondly, we carried out a wider set of experiments based on multiple runs for a lot of different parameter settings.

4.1. Single-run results

The results of a typical run of the regularized SIS on the synthetic dataset in Figure 1, with $N = 10,000$ particles and $t_0 = 100$ for the Gibbs initialization,

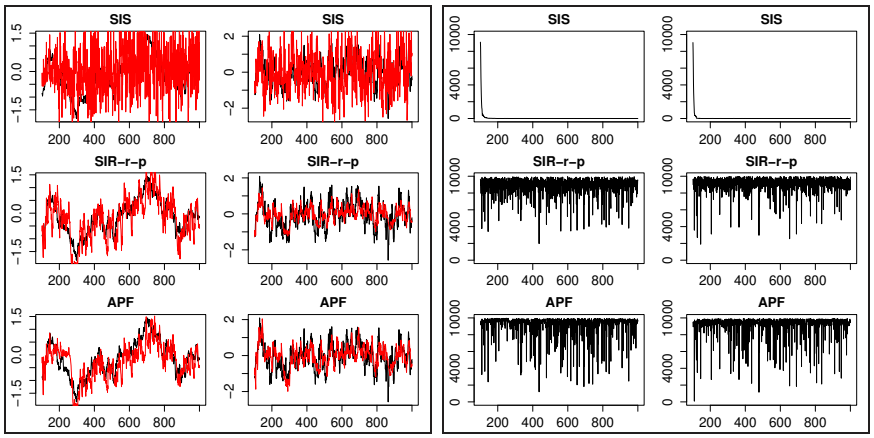

FIG 2. *Daily (left column) and weekly (right column) true (black line) and filtered (red line) log-volatility. Right chart: daily (left column) and weekly (right column) effective sample sizes over iterations.*

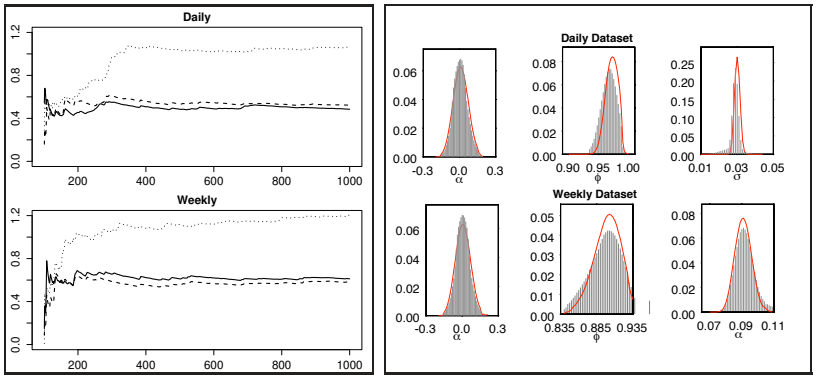

FIG 3. *Left chart: Cumulative root mean square errors over iterations for the APF (solid line), the SIR-r-p (dashed line) and the SIS (dotted line). Right chart: at the last iteration, kernel density estimates of the posteriors based on a long MCMC run (red line) and the regularized APF empirical posteriors (histograms).*

are given in Figures 2–5. We can see (first row of the left chart in Figure 2) that after a few iterations the filtered log-volatility does not fit well to the true log-volatility. The poor performance of the regularized SIS is due to the fact that the empirical posterior of the states and parameters degenerates into a Dirac's mass after a few iterations. The ESSs in the right chart of Figure 3 show that the regularized SIS degenerates after 30 iterations in both the daily and weekly cases.

We measure sequentially the filtering performance of the regularized SIS by evaluating the cumulated root mean square error (RMSE). It measures the distance between the true and the filtered states and is defined as:

$$RMSE_t = \left\{ \frac{1}{t} \sum_{u=1}^{t} (\hat{\mathbf{z}}_u - \mathbf{z}_u)^2 \right\}^{\frac{1}{2}},$$

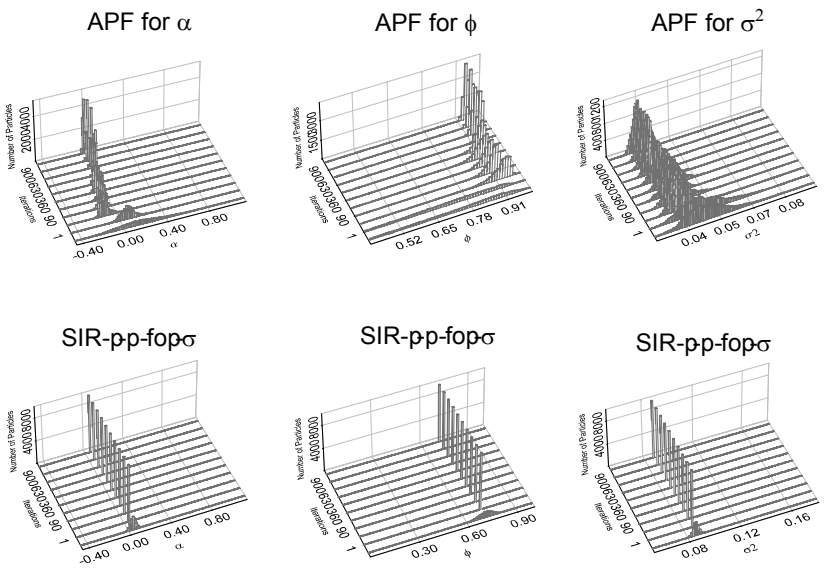

FIG 4. ***Daily dataset:*** *evolution of the empirical posteriors.*

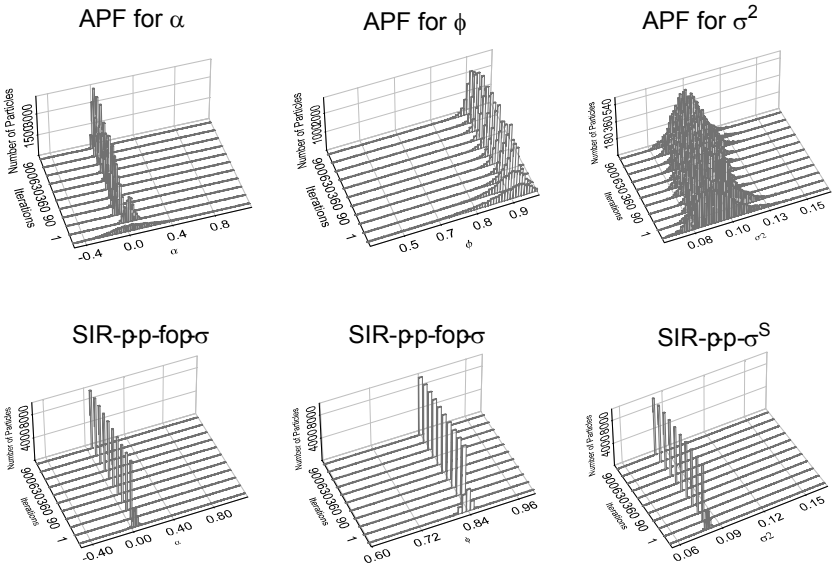

FIG 5. ***Weekly dataset:*** *evolution of the empirical posteriors.*

where $\hat{\mathbf{z}}_t$ is the filtered state, which also includes the parameter sequential estimates. The cumulative RMSEs stabilize rapidly over time in both daily and weekly datasets (see upper and bottom plots in the left chart of Figure 3).

We apply the regularized SIR-r-p and APF with $N = 10,000$ and $t_0 = 100$ to the weekly and daily datasets of Figure 1 and obtain the results given in Figures 2–5. The regularized SIR-r-p and APF outperform the regularized SIS in terms of effective sample sizes and cumulated RMSEs. The effective sample sizes can detect the degeneracy in the particle weights, but are not useful to determine the presence of another form of degeneracy, that is the absence of diversity in the particle values. The histogram of the empirical filtering distribution allows us to detect this second form of degeneracy.

As our work deals with the sequential estimation of the parameters, we focus now on the parameters posteriors. In both the daily and the weekly cases the empirical posterior of the regularized SIS and SIR-r-p degenerates into a Dirac's mass, while the regularized APF posterior does not degenerate.

We report in the right chart of Figure 3 the histograms of the regularized APF posteriors and the approximate posteriors (continuous line) based on a long MCMC run. We use the same Metropolis-Hastings within Gibbs algorithm that we apply in the initialization step of the filters. Note that the regularized APF is not affected by the strong degeneracy which characterizes instead the other filters. In particular we note that in the regularized SIR-r-p algorithm, the degeneracy of the empirical posterior appears after few iterations. See Figures 4 and 5 for a comparison of the regularized SIR-r-p and APF along the iterations.

4.2. Multiple-run results

To confirm the single-run results, we consider a lot of different parameter settings and compare some performances of the seven regularized algorithms: SIS, SIS-p, SIR, SIR-p, SIR-r, SIR-r-p and APF. For each parameter setting we generate 50 independent datasets.

We first compare the algorithms in terms of hidden states filtering. We use the cumulative RMSEs at the last iteration of the filters. Naturally, the cumulative RMSEs are estimated by averaging the RMSEs over the 50 independent runs. The results are summarized in Table 1. Figures 6 and 7 show the evolution of the cumulative RMSEs over the filters iterations for the 50 datasets generated with the two most interesting parameter settings: the weekly setting ($\alpha = 0$, $\phi = 0.9$ and $\sigma^2 = 0.1$) and the daily setting ($\alpha = 0$, $\phi = 0.99$ and $\sigma^2 = 0.01$).

All the results show that the regularized SIR-r-p and APF outperform the other algorithms. Moreover, the regularized APF sensibly outperforms the regularized SIR-r-p for relatively small values of σ^2, i.e. 0.1 and 0.01, over different values of the persistence parameter ϕ. In a context of high volatility persistence (i.e. ϕ equal to 0.9 or 0.99), the difference between the performances of the two algorithms decreases as σ^2 increases.

We then compare the algorithms in terms of parameter estimation. As measure of performances, we use the mean square errors for the parameters α, ϕ

TABLE 1
At the last iteration of the filters, cumulative root mean square errors estimates using the 50 independent runs

Root Mean Square Error (for $\alpha = 0$)							
Grid on ϕ							
(ϕ, σ^2)	SIS	SIS-p	SIR	SIR-p	SIR-r	SIR-r-p	APF
$(0.1, 0.1)$	0.72933	0.51180	0.32873	0.34046	0.35047	0.35129	0.33251
$(0.3, 0.1)$	0.91218	0.58701	0.37109	0.36331	0.34601	0.35111	0.35193
$(0.5, 0.1)$	0.84856	0.66223	0.39764	0.38378	0.36792	0.38284	0.36264
$(0.7, 0.1)$	0.88831	0.68257	0.44713	0.43600	0.48886	0.44521	0.41832
$(0.9, 0.1)$	1.09917	0.84238	0.64995	0.63394	0.61935	0.63096	0.57241
$(0.99, 0.1)$	1.19800	1.11601	1.04905	0.90219	0.84267	0.76512	0.76991
$(0.1, 0.01)$	0.84167	0.56347	0.16680	0.14070	0.19127	0.17083	0.12650
$(0.3, 0.01)$	0.84654	0.56747	0.18419	0.17122	0.19963	0.16366	0.13955
$(0.5, 0.01)$	0.84801	0.61379	0.17655	0.17236	0.17964	0.17130	0.15395
$(0.7, 0.01)$	0.85362	0.72914	0.19182	0.23104	0.25719	0.22406	0.17383
$(0.9, 0.01)$	0.86194	0.77001	0.26563	0.25503	0.27645	0.26479	0.23717
$(0.99, 0.01)$	1.06134	0.81632	0.59902	0.58653	0.57693	0.63302	0.44511
Grid on σ^2							
(ϕ, σ^2)	SIS	SIS-p	SIR	SIR-p	SIR-r	SIR-r-p	APF
$(0.9, 0.01)$	0.86194	0.77001	0.26563	0.25503	0.27645	0.26479	0.23717
$(0.9, 0.1)$	1.09917	0.84238	0.64995	0.63394	0.61935	0.63096	0.57241
$(0.9, 0.3)$	1.31495	1.2636	0.90018	0.88380	0.92874	0.87206	0.86916
$(0.9, 0.5)$	1.53640	1.41958	1.08581	1.08485	1.11529	1.09327	1.09150
$(0.9, 0.7)$	1.73480	1.53763	1.41561	1.39923	1.32798	1.39163	1.31295
$(0.9, 0.9)$	1.91470	1.80927	2.04841	2.00359	2.08913	1.91549	1.55934
$(0.99, 0.01)$	1.06134	0.81632	0.59902	0.58653	0.57693	0.63302	0.44511
$(0.99, 0.1)$	1.19800	1.11601	1.04905	0.90219	0.84267	0.76512	0.76991
$(0.99, 0.3)$	1.52260	1.30912	1.15329	1.08760	1.00661	1.12514	1.05234
$(0.99, 0.5)$	1.70912	1.68504	1.50314	1.43112	1.34881	1.41008	1.49683
$(0.99, 0.7)$	2.01121	1.90713	1.81290	1.75030	1.73903	1.81161	1.68995
$(0.99, 0.9)$	2.30941	2.30941	2.10651	2.09197	2.01498	2.12914	1.97058

TABLE 2
At the last iteration of the filters, mean square errors estimates using 50 independent runs

Daily Data ($\alpha = 0$, $\phi = 0.99$ and $\sigma^2 = 0.01$)							
$\boldsymbol{\theta}$	SIS	SIS-p	SIR	SIR-p	SIR-r	SIR-r-p	APF
α	0.09140	0.00773	0.01154	0.01021	0.00726	0.01234	0.00201
	(0.10115)	(0.00831)	(0.01196)	(0.00974)	(0.00868)	(0.01321)	(0.00247)
ϕ	0.70080	0.71982	0.14037	0.19391	0.14196	0.12992	0.01051
	(0.42131)	(0.43980)	(0.11816)	(0.15843)	(0.12997)	(0.12567)	(0.00721)
σ^2	0.88194	0.87814	0.00668	0.00670	0.00662	0.00677	0.00572
	(0.07652)	(0.08774)	(0.00097)	(0.00061)	(0.00075)	(0.00020)	(0.00158)
Weekly Data ($\alpha = 0$, $\phi = 0.9$ and $\sigma^2 = 0.1$)							
$\boldsymbol{\theta}$	SIS	SIS-p	SIR	SIR-p	SIR-r	SIR-r-p	APF
α	0.00317	0.00401	0.00571	0.00493	0.00555	0.00737	0.00054
	(0.00462)	(0.00493)	(0.00814)	(0.00753)	(0.01027)	(0.01074)	(0.00067)
ϕ	0.52602	0.54642	0.04975	0.07910	0.04520	0.06314	0.00124
	(0.40102)	(0.41621)	(0.07248)	(0.07592)	(0.07975)	(0.08494)	(0.00174)
σ^2	0.70171	0.70212	0.00017	0.00009	0.00010	0.00009	0.00009
	(0.05895)	(0.05756)	(0.00011)	(0.00008)	(0.00004)	(0.00005)	(0.00007)

and σ^2 at the last iteration of the filters. For each parameter setting, the mean square errors has been estimated using 50 independent runs of the filters. Table 2 gives the results. The regularized APF clearly outperforms the other filters.

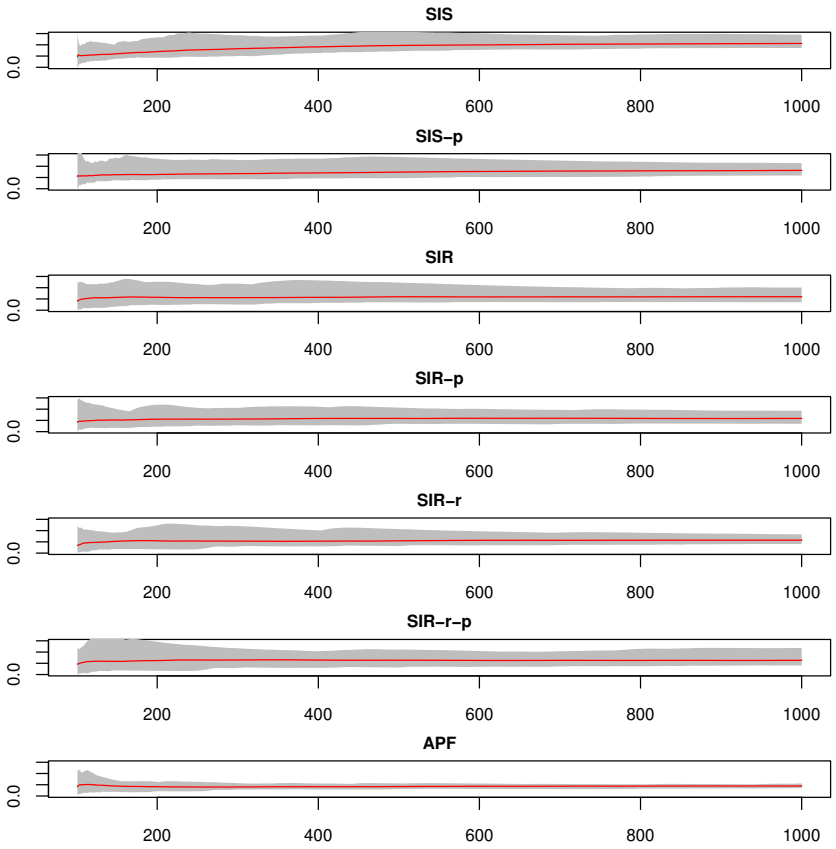

FIG 6. ***Daily datasets:*** *average cumulative root mean square errors between the true and the filtered log-volatility (red line) using 50 independent runs. The grey area represents the area between maximum and minimum cumulative root mean square errors.*

5. Conclusion

In this work we illustrate the kernel regularization technique for particle filters and deal with the online parameter estimation problem. While the regularized APF has been already used for parameter estimation, the regularized versions of SIS and SIR have not been considered to that aim. We focus on the joint estimation of the states and parameters and compare some algorithms on a Bayesian nonlinear model: the Bayesian SV model. As we expected, we find evidence of the degeneracy of two different regularized SIS. Finally, we find that, in terms of parameter estimation, the regularized APF outperforms all the others schemes.

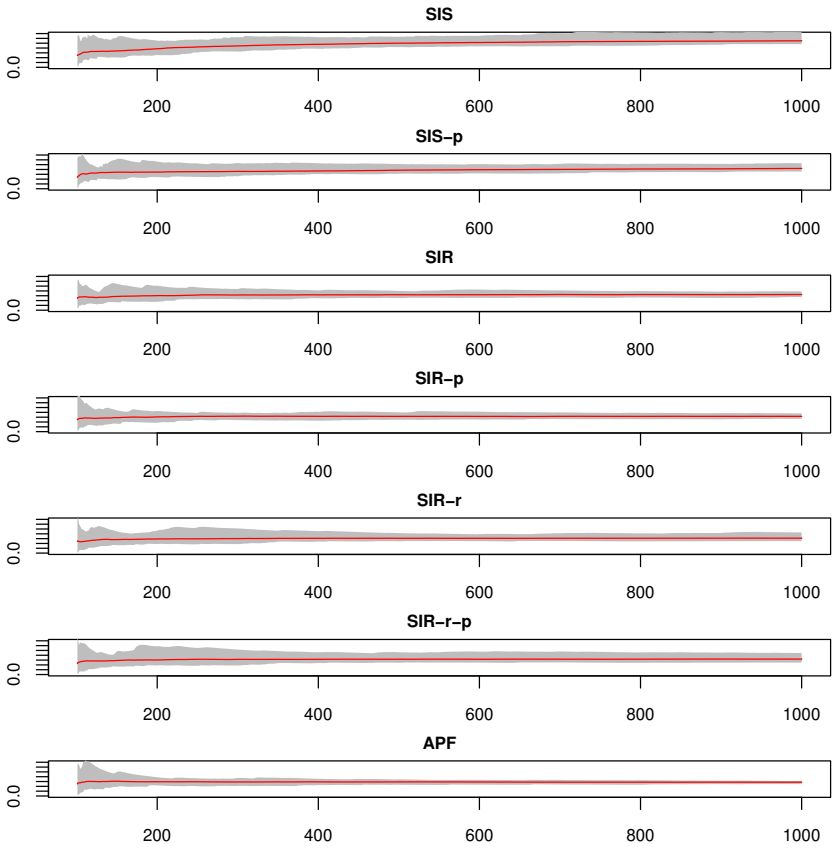

FIG 7. ***Weekly datasets:*** *average cumulative root mean square errors between the true and the filtered log-volatility (red line). The grey area represents the area between maximum and minimum cumulative root mean square errors.*

Acknowledgements

The authors wish to thank the Editor and the Associate Editor for their encouraging comments. Moreover, the authors thank Professor Christian Robert for his helpful suggestions.

References

ANDRIEU, C. and DOUCET, A. (2003). Online expectation–maximization type algorithms for parameter estimation in general state space models. *Proc. IEEE ICASSP*, 6:VI–69–VI–72.

ARULAMPALAM, S., MASKELL, S., GORDON, N., and CLAPP, T. (2001). A tutorial on particle filters for on-line nonlinear/non-gaussian bayesian tracking. Technical Report, QinetiQ Ltd., DSTO, Cambridge.

Berzuini, C. and Gilks, W. (2001). Following a moving average Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society, B*, 63:127–146. MR1811995

Carter, C. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81(3):541–553. MR1311096

Celeux, G., Marin, J.-M., and Robert, C. (2006). Iterated importance sampling in missing data problems. *Computational Statistics and Data Analysis*, 50(12):3386–3404. MR2236856

Chen, Z. and Haykin, S. (2002). On different facets of regularization theory. *Neural Comput.*, 14:2791–2846.

Crisan, D. and Doucet, A. (2000). Convergence of sequential monte carlo methods. Technical Report N. 381, CUED-F-INFENG.

Djuric, P.M., Kotecha, J., Esteve, F., and Perret, E. (2002). Sequential parameter estimation of time-varying non-gaussian autoregressive processes. *EURASIP Journal on Applied Signal Processing*, 8:865–875.

Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag. MR1847783

Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208.

Doucet, A. and Tadic, C. (2003). Parameter estimation in general state-space models using particle methods. *Annals of the Institute of Statistical Mathematics*, 55(2):409–422. MR2001872

Fearnhead, P. (2002). MCMC, sufficient statistics and particle filter. *Journal of Computational and Graphical Statistics*, 11:848–862. MR1951601

Gordon, N., Salmond, D., and Smith, A.F.M. (1993). Novel approach to nonlinear and nongaussian bayesian state estimation. *IEE Proceedings-F*, 140:107–113.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press. MR1278033

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384. MR0996941

Harrison, J. and West, M. (1989). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, 2 edition. MR1020301

Harvey, A. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Transaction of the ASME, Journal of Basic Engineering, Series D*, 82:35–45.

Kalman, R. and Bucy, R. (1960). New results in linear filtering and prediction problems. *Transaction of the ASME, Journal of Basic Engineering, Series D*, 83:95–108.

Kitagawa, G. (1998). A self-organized state-space model. *Journal of the American Statistical Association*, 93(443):1203–1215.

Liu, J. and Chen, R. (1998). Sequential Monte Carlo methods for dynamical system. *Journal of the American Statistical Association*, 93:1032–1044. MR1649198

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation based filtering. In Doucet, A., de Freitas, N., and Gordon, N., editors, *Sequential Monte Carlo Methods in Practice*. Springer-Verlag. MR1847793

Maybeck, P. (1982). *Stochastic Models, Estimation and Control, volume 1-3*. Academic Press.

Musso, C., Oudjane, N., and Legland, F. (2001). Improving regularised particle filters. In Doucet, A., de Freitas, N., and Gordon, N., editors, *Sequential Monte Carlo Methods in Practice*. Springer-Verlag. MR1847795

Oudjane, N. (2000). Stabilité et approximation particulaires en filtrage nonlinéaire. Application au pistage. Thèse du Doctorat en Science, Université de Rennes.

Pitt, M. and Shephard, N. (1999). Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, 94(446):590–599. MR1702328

Polson, N.G., Stroud, J.R., and Müller, P. (2002). Practical Filtering with sequential parameter learning. Tech. report, Graduate School of Business, University of Chicago.

Rossi, V. (2004). Filtrage non linéaire par noyaux de convolution. Application à un procédé de dépollution biologique. Thèse du Doctorat en Science, École Nationale Supérieure Agronomique de Montpellier.

Shephard, N. and Pitt, M. (1997). Likelihood Analysis of Non-Gaussian Measurement Time Series. *Biometrika*, 84:653–667. MR1603940

Storvik, G. (2002). Particle filters for state space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, 50:281–289.

West, M. (1992). Mixture models, Monte Carlo, Bayesian updating and dynamic models. *Computer Science and Statistics*, 24:325–333.

West, M. (1993). Approximating posterior distribution by mixtures. *Journal of the Royal Statistical Society, B*, 55:409–442. MR1224405