Online data processing: comparison of Bayesian regularized particle filters

Roberto Casarin, Jean-Michel Marin

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Online data processing: comparison of Bayesian regularized particle filters

Roberto Casarin — Jean-Michel Marin

N° 6153

Mars 2007

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--6153--FR+ENG

Online data processing: comparison of Bayesian regularized particle filters

Roberto Casarin * , Jean-Michel Marin†

Thème COG — Systèmes cognitifs
Projets SELECT

Rapport de recherche n° 6153 — Mars 2007 — 20 pages

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 parameters estimation, considering a Bayesian paradigm and a univariate stochastic volatility 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 (R-APF) outperforms the Regularized Sequential Importance Sampling (R-SIS) and the Regularized Sampling Importance Resampling (R-SIR).

Key-words: Online data processing, Bayesian estimation, regularized particle filters, stochastic volatility model

* Department of Economics, University of Brescia
† INRIA Futurs, Projet SELECT, Université Paris-Sud

Unité de recherche INRIA Futurs
Parc Club Orsay Université, ZAC des Vignes,
4, rue Jacques Monod, 91893 ORSAY Cedex (France)
Téléphone : +33 1 72 92 59 00 — Télécopie : +33 1 60 19 66 08

Traitement de données en temps réel : comparaison de filtres particulaires bayésiens régularisés

Résumé : L'objectif de ce travail est de comparer trois filtres particulaires régularisés pour le traitement de données en temps réel. Les trois filtres sont évalués pour leurs capacités à reconstituer les états latents du système et à estimer les paramètres du modèle. Nous considérons le paradigme bayésien et le modèle à volatilité stochastique. Nous montrons que les performances du filtre particulaire auxiliaire sont meilleures que celles des filtres particulaires classiques d'échantillonnage préférentiel séquentiel.

Mots-clés : Traitement de données en temps réel, estimation bayésienne, filtres particulaires régularisés, modèle à volatilité stochastique

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 informations we want to extrapolate from the observations.

In time series analysis, many approaches have been used for the estimation of dynamics 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 too much 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 (R-SIS), the Regularized Sampling Importance Resampling (R-SIR) and the Regularized Auxiliary Particle Filter (R-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 regularisation (see Chen and Haykin (2002)) approach of (Musso et al., 2001; Oudjane, 2000; 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 present the stochastic volatility model. Section 3 reviews some regularized particle filters. Finally, Section 4 presents 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 model 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 filters 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 $\{\mathbf{y}_t;\, t \in \mathbb{N}^*\}$, $\mathbf{y}_t \in \mathcal{Y} \subseteq \mathbb{R}^{n_y}$, 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:

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

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

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

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

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, filtering and smoothing densities which are given by:

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

$$p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \int_{\mathcal{Z}} p(\mathbf{y}_{t+1}|\mathbf{z}_{t+1}, \mathbf{y}_{1:t})p(\mathbf{z}_t|\mathbf{y}_{1:t})d\mathbf{z}_{t+1}\,,$$

$$p(\mathbf{z}_{t+1}|\mathbf{y}_{1:t+1}) = \frac{p(\mathbf{y}_{t+1}|\mathbf{z}_{t+1}, \mathbf{y}_{1:t})p(\mathbf{z}_{t+1}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})}\,, \tag{2}$$

$$p(\mathbf{z}_s|\mathbf{y}_{1:t}) = p(\mathbf{z}_s|\mathbf{y}_{1:s}) \int_{\mathcal{Z}} \frac{p(\mathbf{z}_{s+1}|\mathbf{z}_s, \mathbf{y}_{1:s})p(\mathbf{z}_{s+1}|\mathbf{y}_{1:t})}{p(\mathbf{z}_{s+1}|\mathbf{y}_{1:t})}d\mathbf{z}_{s+1}, \quad s < t\,.$$

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 stochastic volatility model. Two of the main features of the financial time series are time varying volatility and clustering phenomena in volatility. Stochastic volatility models widely used in finance have been introduced, in order to account

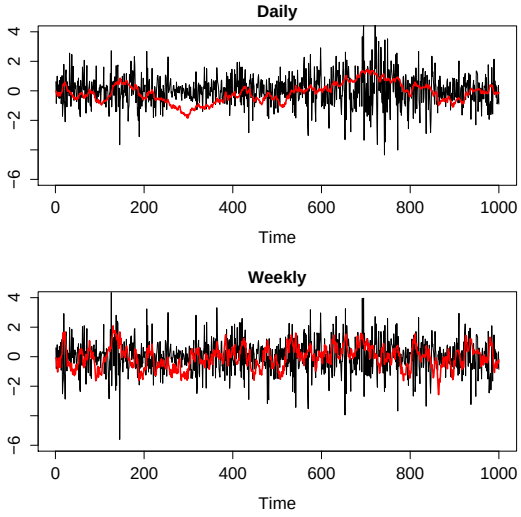

Figure 1: Simulation paths for x_t (*grey line*) and y_t (*black line*). Upper plot: daily dataset ($\alpha = 0$, $\phi = 0.99$ and $\sigma^2 = 0.01$). Bottom plot: weekly dataset ($\alpha = 0$, $\phi = 0.9$ and $\sigma^2 = 0.1$).

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 stochastic volatility model is:

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

where $\boldsymbol{\theta} = (\alpha, \log((1+\phi)/(1-\phi)), \log(\sigma^2))$. The choice of $\pi(\boldsymbol{\theta})$ will be discussed in Section 4. Fig. 1 shows two simulated paths of y_t and x_t.

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

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 too much 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-regularised particle filters, which combine the importance sampling reasoning with a suitable modification of the importance weights. This approach relies upon a kernel-based reconstruction of the empirical filtering densities which produces a systematic modification of the true importance weights. Some convergence results for this kind of particle filters are given by Musso et al. (2001).

3.1 Regularized SIS

Let us start from the non-regularized SIS. We assume that at iteration $t > 0$ a properly weighted particle set $\{\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \alpha_t^i\}_{i=1}^N$, 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:

$$p_N(\mathbf{x}_t, \boldsymbol{\theta}_t|\mathbf{y}_{1:t}) = \sum_{i=1}^N \alpha_t^i \delta_{(\mathbf{x}_t^i, \boldsymbol{\theta}_t^i)}(\mathbf{x}_t, \boldsymbol{\theta}_t)\,. \tag{3}$$

The particles set, $\{\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \alpha_t^i\}_{i=1}^N$, can be viewed as a random discretisation of the state space $\mathcal{X} \times \Theta$ with associated probability weights $\{\alpha_t^i\}_{i=1}^N$. Thanks to this discretisation, it is possible to approximate the prediction and filtering densities given in (1) and (2):

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

The goal is now to obtain N particles $\{\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i, \alpha_{t+1}^i\}_{i=1}^N$ from the filtering density in (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:

$$\alpha_{t+1}^i \propto \alpha_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i, \mathbf{y}_{1:t}) p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}^i, \mathbf{y}_{1:t}) \delta_{\boldsymbol{\theta}_t^i}(\boldsymbol{\theta}_{t+1}^i)}{q(\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i|\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \mathbf{y}_{1:t+1})}\,. \tag{4}$$

The choice of an optimal importance density $q(\cdot|\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \mathbf{y}_{1:t+1})$, 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 the 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 at 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 for ever, so we are facing a problem of degeneracy of the empirical filtering distribution. This scenario motivates particle filtering methods known as regularised 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, we modify the weights in (4) and define a new set of weights:

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

where $K_h(y) = h^{-d}K(y/h)$ is a regularization kernel, K being a positive function defined on $\mathbb{R}^{n_{\boldsymbol{\theta}}}$ and h a positive smoothing factor (bandwidth). These modified importance weights correspond to a regularized version of the filtering density. Indeed, the modification of the importance weights defined in (4) results from two steps. The first one is the regularization of the empirical density in (3) by a kernel estimator:

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

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

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

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:

$$p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \boldsymbol{\theta}_{t+1}, \mathbf{y}_{1:t}) K_h(\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t^i)\,.$$

In that case, we have:

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

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

Algorithm 1 - *Regularized SIS Particle Filter* -
· *At time* $t = 0$, *for* $i = 1, \ldots, N$, *simulate* $\mathbf{z}_0^i \sim p(\mathbf{z}_0)$ *and set* $\omega_0^i = 1/N$
· *At time* $t > 0$, *given* $\{\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \omega_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$:

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

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 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, which use a resampling step to generate a new set of particles with uniform weights. As for the SIS, we assumed that the proposal distribution for $\tilde{\mathbf{x}}_{t+1}^i$ is $p(\cdot|\mathbf{x}_t^i, \tilde{\boldsymbol{\theta}}_{t+1}^i, \mathbf{y}_{1:t})$, $\{\tilde{\mathbf{x}}_{t+1}^i, \tilde{\boldsymbol{\theta}}_{t+1}^i\}_{i=1}^N$ being the particle set before resampling. Due to the resampling step, the weights of the resampled particles are uniformly distributed over the particle set, $\omega_t^i = 1/N$, and the weight of particle $(\tilde{\mathbf{x}}_{t+1}^i, \tilde{\boldsymbol{\theta}}_{t+1}^i)$ is such that:

$$\tilde{\omega}_{t+1}^i \propto \omega_t^i \, p(\mathbf{y}_{t+1}|\tilde{\mathbf{x}}_{t+1}^i, \tilde{\boldsymbol{\theta}}_{t+1}^i, \mathbf{y}_{1:t}) \propto p(\mathbf{y}_{t+1}|\tilde{\mathbf{x}}_{t+1}^i, \tilde{\boldsymbol{\theta}}_{t+1}^i, \mathbf{y}_{1:t}) \, .$$

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

Algorithm 2 - *Regularized SIR Particle Filter* -
· *At time* $t = 0$, *for* $i = 1, \ldots, N$, *simulate* $\mathbf{z}_0^i \sim p(\mathbf{z}_0)$ *and set* $\omega_0^i = 1/N$
· *At time* $t > 0$, *given* $\{\mathbf{x}_t^i, \boldsymbol{\theta}_t^i, \omega_t^i = 1/N\}_{i=1}^N$, *for* $i = 1, \ldots, N$:

1. *Simulate* $\tilde{\boldsymbol{\theta}}_{t+1}^i \sim K_h(\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t^i)$
2. *Simulate* $\tilde{\mathbf{x}}_{t+1}^i \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \tilde{\boldsymbol{\theta}}_{t+1}^i, \mathbf{y}_{1:t})$
3. *Update the weights:* $\tilde{\omega}_{t+1}^i \propto p(\mathbf{y}_{t+1}|\tilde{\mathbf{x}}_{t+1}^i, \tilde{\boldsymbol{\theta}}_{t+1}^i, \mathbf{y}_{1:t})$.
4. *Simulate* $\{\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i\}_{i=1}^N$ *from* $\{\tilde{\mathbf{x}}_t^i, \tilde{\boldsymbol{\theta}}_t^i, \tilde{\omega}_t^i\}_{i=1}^N$ *(Multinomial resampling) and set* $\omega_{t+1}^i = 1/N$.

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 R-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:

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

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

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

where

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

$\mu_{t+1}^{j^i}$ 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:

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

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

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

· *At time* $t = 0$*, for* $i = 1, \ldots, N$*, simulate* $\mathbf{z}_0^i \sim p(\mathbf{z}_0)$ *and set* $\omega_0^i = 1/N$

· *At time* $t > 0$*, 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* $\boldsymbol{\theta}_{t+1}^i \sim K_h(\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t^{j^i})$
3. *Simulate* $\mathbf{x}_{t+1}^i \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^{j^i}, \boldsymbol{\theta}_{t+1}^i, \mathbf{y}_{1:t})$
4. *Update particles weights:* $\omega_{t+1}^i \propto \dfrac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \boldsymbol{\theta}_{t+1}^i, \mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{j^i}, m_{t+1}^{j^i}, \mathbf{y}_{1:t})}$.

In the next section apply R-SIS, R-SIR and R-APF to the stochastic volatility model and make a comparison between them, in terms of root mean square filtering errors and of variance of the particle weights.

4 Application to the stochastic volatility model and comparison

In this section, we apply the three regularized particle filters to the stochastic volatility model presented in Section 2. We assume that the initial value of the SV process follows the stationary distribution:

$$x_0 \sim \mathcal{N}(0, \sigma^2/(1-\phi^2))\,.$$

For the parameter β, σ and ϕ, we assume the prior

$$p(\beta^2, \phi, \sigma^2) = 1/(\sigma\beta)\mathbb{I}_{]-1,1[}(\phi)\,,$$

where $\beta = e^{\alpha}$. We constrain the parameter ϕ to take values in $]-1,1[$ in order to impose the usual stationarity condition. As we use an improper prior, it is not possible to use the prior distribution for initializing the three particle filters. We need to start with a proper weighted sample $\{x_0^i, \boldsymbol{\theta}_0^i, \omega_0^i\}_{i=1}^N$. We propose to:

1) start the sequential filtering procedure at least at the value n of t, such that the posterior distribution of the parameters given all the observations up to time n is well defined: for the considered SV model, this corresponds to set $n \geq 2$;

2) use a Metropolis within Gibbs algorithm to create a sample with uniform weights.

We will use the Metropolis within Gibbs algorithm studied 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. We recall here just some full conditional distributions:

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

We refer to Celeux et al. (2006) for a detailed description of the proposal distributions for ϕ and x_t.

Given the initial weighted random sample $\left\{x_t^i, \boldsymbol{\theta}_t^i, \omega_t^i\right\}_{i=1}^N$, where $\boldsymbol{\theta}_t = (\alpha_t,\ \log((1+\phi_t)/(1-\phi_t)), \log(\sigma_t^2))$, the R-SIS performs the following steps:

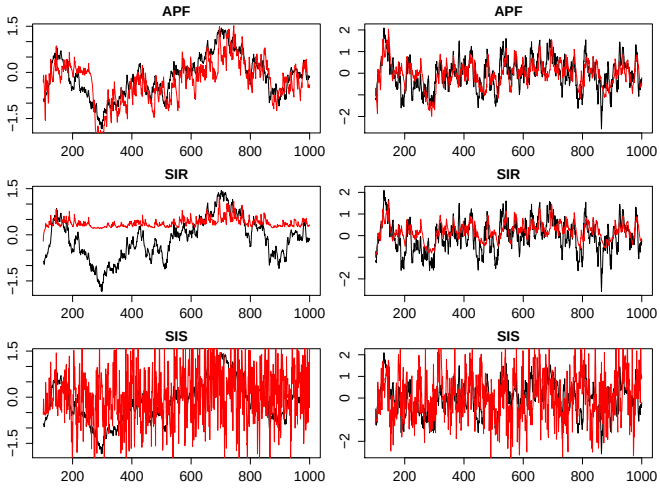

Figure 2: Daily (*left column*) and weekly (*right column*) true (*black line*) and filtered (*grey line*) log-volatility.

For $n \leq t \leq T-1$ and for $i = 1, \ldots, N$:

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

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

(iii) Update the weights as follow

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

The results of a typical run of the R-SIS on the synthetic dataset in Fig. 1, with $N = 10,000$ particles and $n = 100$ for the Gibbs initialization, is represented in Fig. from 2 to 6. We can see (last row in Fig. 2) that after a few iterations the filtered log-volatility does not fit well to the true log-volatility. We measure sequentially the filtering performance of the R-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: $RMSE_t = \{\frac{1}{t}\sum_{u=1}^{t}(\hat{\mathbf{z}}_u - \mathbf{z}_u)^2\}^{\frac{1}{2}}$, where $\hat{\mathbf{z}}_t$ is the filtered state, which includes also the parameter sequential estimate. The

RMSEs cumulate rapidly over time in both daily and weekly datasets (see upper and bottom plots in Fig. 3). The poor performance of the R-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 Fig. 4 show that the R-SIS degenerates after 30 iterations in both the daily and weekly datasets.

The R-SIR performs the following step:

For $n \leq t \leq T-1$ and for $i = 1, \ldots, N$:

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

(ii) Simulate $\tilde{x}_{t+1}^i \sim \mathcal{N}\left(x_{t+1} | \tilde{\alpha}_{t+1}^i + \tilde{\phi}_{t+1}^i x_t^i, \left(\tilde{\sigma}^2\right)_{t+1}^i\right)$,

(iii) Update the weights

$$\tilde{w}_{t+1}^i \propto w_t^i \exp\left\{-\frac{1}{2}\left[y_{t+1}^2 \exp\{-\tilde{x}_{t+1}^i\} + \tilde{x}_{t+1}^i\right]\right\},$$

(v) Simulate $\mathbf{z}_{t+1}^i \sim \sum_{j=1}^N \tilde{w}_{t+1}^j \delta_{\tilde{\mathbf{z}}_{t+1}^j}(\mathbf{z}_{t+1})$ and set $w_{t+1}^i = 1/N$.

The R-APF performs the following steps:

For $n \leq t \leq T-1$ and for $i = 1, \ldots, N$:

(i) Simulate $j^i \sim q(j) \propto \sum_{k=1}^N w_t^k \mathcal{N}(y_{t+1} | \mu_{t+1}^k) \delta_k(j)$ where $\mu_{t+1}^k = \phi_t^k x_t^k + \alpha_t^k$,

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

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

(iv) Update the weights

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

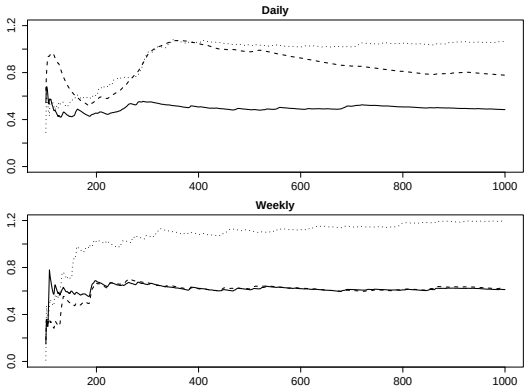

Figure 3: Daily (*upper plot*) and weekly (*bottom plot*) Root Mean Square Errors for the regularized APF (*solid line*), SIR (*dashed line*) and SIS (*dotted line*) over iterations.

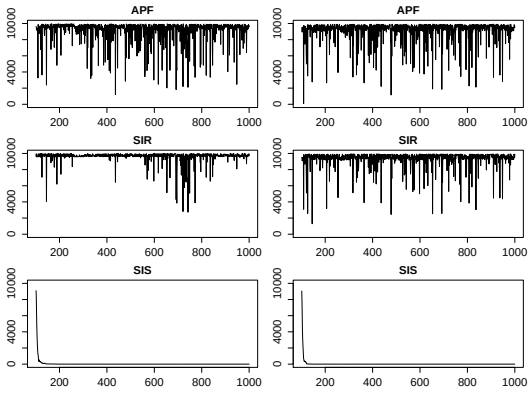

Figure 4: Daily (*left column*) and weekly (*right column*) Effective Sample Sizes over iterations.

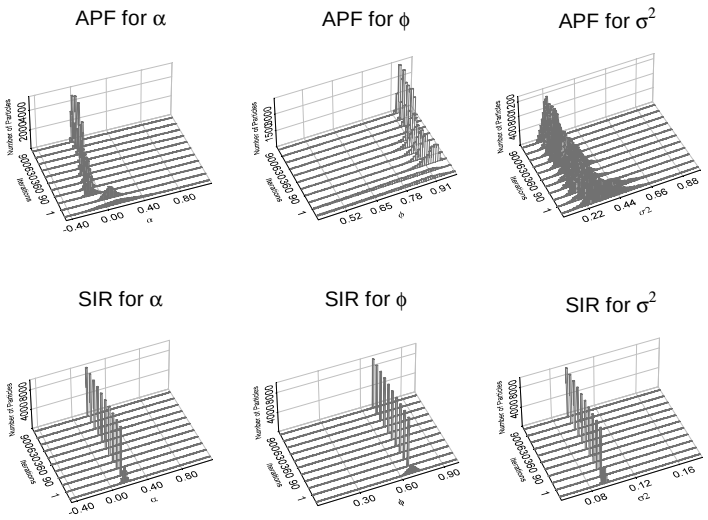

Figure 5: Evolution on daily dataset of the empirical posterior distributions of α, ϕ and σ^2.

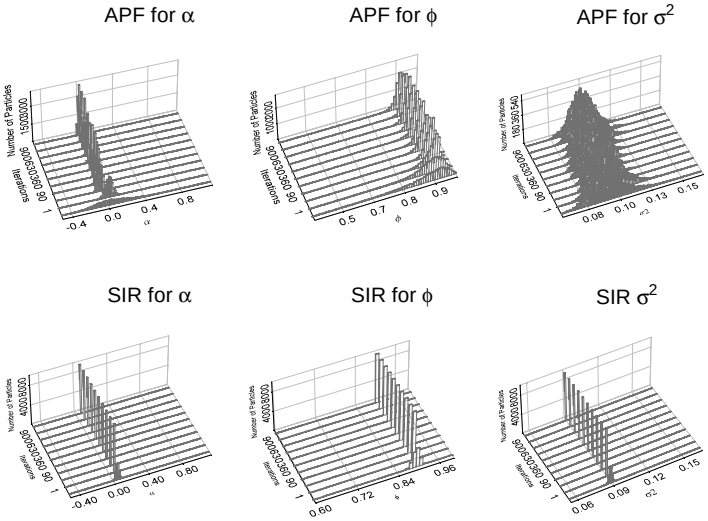

Figure 6: Evolution on weekly dataset of the empirical posterior distributions of α, ϕ and σ^2.

Daily Data				Weekly Data			
$\boldsymbol{\theta}$	R-APF	R-SIR	R-SIS	$\boldsymbol{\theta}$	R-APF	R-SIR	R-SIS
α	0.00065	0.00885	0.00719	α	0.00016	0.00318	0.00534
ϕ	0.00855	0.12433	0.66767	ϕ	0.00029	0.18422	0.51290
σ^2	0.00506	0.00676	0.89327	σ^2	8e-05	0.73326	0.7054

Table 1: Mean Square Errors of the estimators of α, ϕ and σ^2. The Mean Square Errors are estimated using the last iteration of the 10 independent runs of the filters.

Note that, following Pitt and Shephard (1999), one could alternatively use in the selection step a value of μ_{t+1}^k based on the Taylor expansion of the likelihood at time $t+1$. For the three regularized particle filters, we have used a Gaussian kernel where the parameter a is fixed following the usual optimal criterion.

We apply the R-SIR and R-APF with $N = 10,000$ and $n = 100$ to the weekly and daily datasets of Fig. 1 and obtain the results given in Fig. from 2 to 6. The R-SIR and R-APF outperform the R-SIS in terms of ESSs and cumulated RMSEs. The ESSs can detect the degeneracy in the particle weights but is not useful to reveal 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 choose to show the histogram of the parameters posterior. Fig. 5 and 6 exhibit the evolution over iterations of the posterior of the parameters α, ϕ and σ^2. In both the daily and the weekly datasets, after a few iterations the empirical posterior of the R-SIR degenerates into a Dirac's mass.

To confirm the previous results, we have done ten independent runs of the three algorithms. The weekly and daily datasets vary across the 10 experiments. Our simulation study confirms the results of the single-run experiment. Fig. 7 and 8 show a comparison between R-SIS, R-SIR and R-APF in terms of RMSEs. The RMSEs are estimated over the 10 independent runs of the algorithms. The R-SIR and the R-APF outperform the R-SIS in both the daily and the weekly datasets. The estimated Mean Square Errors for the parameters α ϕ and σ^2 (see Table 1), based on 10 independent runs of the filters, show that in term of parameters estimation the R-APF outperforms the R-SIR and the R-SIS.

Fig. 9 and 10 show a comparison of the three filter in terms of ESS. As one could expect, in all the independent runs the R-SIS weights degenerate after a few iterations.

We also observe that the ESSs of the R-SIR and R-APF are substantially equivalent when filtering daily data, which are characterized by high persistence and low variance in the volatility process (i.e. $\phi = 0.99$ and $\sigma^2 = 0.01$). The parameters and hidden states joint estimation is more challenging on weekly data, which are characterized by low persistence and high variance in the volatility process (i.e. $\phi = 0.9$ and $\sigma^2 = 0.1$). In that context, the R-APF outperforms the R-SIR (see Fig. 8).

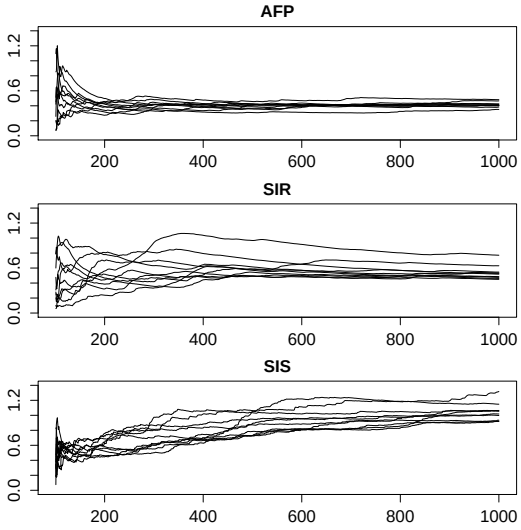

Figure 7: Comparison on daily datasets of cumulative Root Mean Square Errors between the true and the filtered log-volatility.

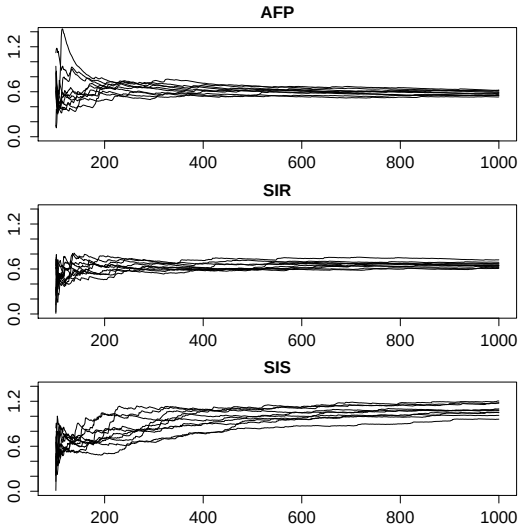

Figure 8: Comparison on weekly datasets of cumulative RMSEs between the true and the filtered log-volatility.

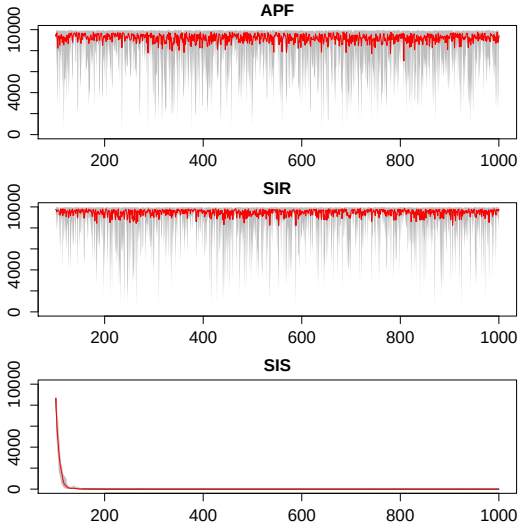

Figure 9: Comparison on daily datasets of average Effective Sample Sizes (*black line*). We represent the area between maximum and minimum Effective Sample Sizes (*grey area*).

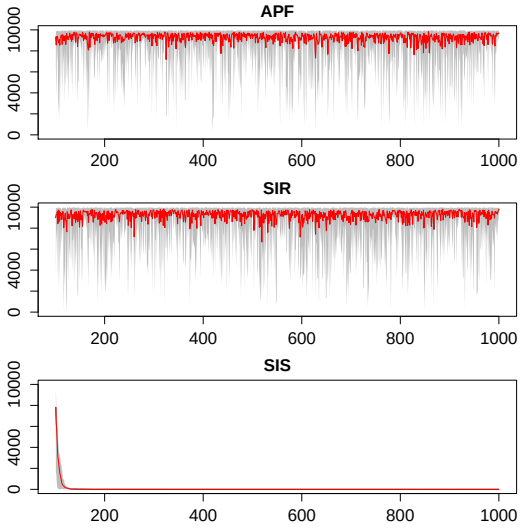

Figure 10: Comparison on weekly datasets of average Effective Sample Sizes (*black line*). We represent the area between maximum and minimum Effective Sample Size (*grey area*).

5 Conclusion

In this work we bring into action the kernel regularization technique for particle filters and deal with the online parameter estimation problem. While the regularized auxiliary particle filter (R-APF) has been already used for the parameter estimation, the regularized version of the Sequential Importance Sampling (R-SIS) and the Sampling Importance Resampling (R-SIR) have not been considered to that aim. We focus on the joint estimation of the states and parameters and compare the three algorithms on a Bayesian nonlinear model: the Bayesian stochastic volatility model. As we expected, we find evidence of the degeneracy of the R-SIS. We find that when the volatility is persistent, both R-SIR and R-APF perform well, while the R-APF is superior in estimating both hidden states and parameters in a context of low persistence in the volatility. In terms of parameters estimation, the R-APF outperforms the R-SIR in all cases.

Acknowledgments

We thank Professor Christian Robert for its helpful comments and a careful reading of a preliminary version of the paper.

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 Royal Statistical Society, B*, 63:127–146.

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

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.

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.

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.

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

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.

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

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

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.

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.

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.

Oudjane, N. (2000). Stabilité et approximation particulaires en filtrage non-linéaire. 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.

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.

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 Royal Statistical Society, B*, 55:409–442.

Unité de recherche INRIA Futurs
Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399