Particle Filtering for Partially Observed Gaussian State Space Models

Christophe Andrieu
University of Bristol, Bristol, UK.
Arnaud Doucet†
University of Melbourne, Parkville, Australia.

Summary. Solving Bayesian estimation problems where the posterior distribution evolves over time through the accumulation of data has many applications for dynamic models. A large number of algorithms based on particle filtering methods, also known as sequential Monte Carlo (SMC), have recently been proposed to solve these problems. In this paper, we propose a special particle filtering method which uses random mixtures of normal distributions to represent the posterior distributions of partially observed Gaussian state space models. This algorithm is based on a marginalization idea for improving efficiency and can lead to substantial gains over standard algorithms. It differs from previous algorithms described in Chen and Liu (2000) and Doucet *et al.* (2000), which were only applicable to conditionally linear Gaussian state space models. Computer simulations are carried out to evaluate the performance of the proposed algorithm for dynamic Tobit and Probit models.
Keywords: Bayesian estimation; Filtering; Generalized linear time series; Importance Sampling; Sequential Monte Carlo; State space model.

1. Introduction

1.1. Background

Many data analysis tasks involve estimating the state of a dynamic model when only partial or inaccurate observations are available (West and Harrison, 1997). Except in a few special cases, including linear Gaussian state space models, on-line state estimation is a problem that does not admit a closed form solution. As most real-world models are nonlinear and non-Gaussian, it is of great interest to develop efficient computational methods so as to numerically solve this so-called Bayesian filtering problem.

Many approximation schemes, such as the extended Kalman filter, have been proposed to surmount this problem. However, in many realistic scenarios, these approximating methods are unreliable and faults are difficult to diagnose on-line. Recently there has been a surge of interest in SMC methods for nonlinear/non-Gaussian time series analysis (Doucet *et al.*, 2001). These methods, initiated in Gordon *et al.* (1993) utilize a random sample (or particle) based representation of the posterior probability distributions.

†*Address for correspondence:* Arnaud Doucet, Department of Electrical Engineering, University of Melbourne, VIC 3052, Australia. Email: a.doucet@ee.mu.oz.au

1.2. General Problem

For any sequence l_t, we define $l_{i:j} \triangleq (l_i, l_{i+1}, \ldots, l_j)$. In this paper, we will concentrate on the following class of state space models. Let $t = 1, 2, \ldots$ denote discrete time, then

$$\begin{aligned} x_t &= A_t x_{t-1} + B_t v_t + F_t u_t,\ x_0 \sim \mathcal{N}(\hat{x}_0, P_0), & (1) \\ y_t &= C_t x_t + D_t \varepsilon_t + G_t u_t, & (2) \\ z_t &\sim p(z_t | y_t), & (3) \end{aligned}$$

where $u_t \in \mathbb{R}^{n_u}$ is an exogeneous process, $x_t \in \mathbb{R}^{n_x}$ and $y_t \in \mathbb{R}^{n_y}$ are *unobserved* processes. The sequences $v_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_{n_v}) \in \mathbb{R}^{n_v}$ and $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_{n_\varepsilon}) \in \mathbb{R}^{n_\varepsilon}$ are independent identically distributed (i.i.d.) Gaussian. We assume $P_0 > 0$; x_0, v_t and w_t are mutually independent for all t, and the model parameters $\lambda \triangleq (\hat{x}_0, P_0, A_t, B_t, C_t, D_t, F_t, G_t; t = 1, 2, \ldots)$ are known. The processes (x_t) and (y_t) define a standard linear Gaussian state space model. One does not observe (y_t) in our case, but (z_t). The *observations* (z_t) are conditionally independent given the processes (x_t) and (y_t) and marginally distributed according to $p(z_t | y_t)$; it is assumed that $p(z_t | y_t)$ can be evaluated pointwise up to a normalizing constant. Typically $p(z_t | y_t)$ belongs to the exponential family. Alternatively z_t may be a censored or quantized version of y_t. This class of *partially observed Gaussian state space models* has numerous applications; many examples are discussed for instance in de Jong (1997), Manrique and Shephard (1998) and West and Harrison (1997).

We want to estimate *sequentially in time* some characteristics of the posterior distribution $p(x_{0:t} | z_{1:t})$. Typically, one is interested in computing $E(x_t | z_{1:t})$ (filtering), $E(x_{t+L} | z_{1:t})$ (prediction) and $E(x_{t-L} | z_{1:t})$ (fixed-lag smoothing), where L is a positive integer. These estimates do not in general admit analytical expressions and one has to resort to numerical methods.

1.3. Resolution

SMC methods are, loosely speaking, a combination of importance sampling and resampling methods that allow one to efficiently propagate over time a large set of samples/particles distributed approximately according to $p(x_{0:t} | z_{1:t})$. One could apply standard SMC methods such as the bootstrap filter (Gordon *et al.*, 1993) to estimate $p(x_{0:t}, y_{1:t} | z_{1:t})$ and consequently $p(x_{0:t} | z_{1:t})$. However, in its standard form, this algorithm does not use all the salient structure of the model. Our algorithm is based on a marginalization technique, often called the Rao-Blackwellization method (Gelfand and Smith, 1990), that improves the efficiency of the procedure. It focuses on the estimation of $p(y_{1:t} | z_{1:t})$ rather than the joint density $p(x_{0:t}, y_{1:t} | z_{1:t})$. The process (x_t) is integrated out analytically. Once $p(y_{1:t} | z_{1:t})$ is estimated, one can obtain estimates of $E(x_t | z_{1:t})$, $E(x_{t+L} | z_{1:t})$ and $E(x_{t-L} | z_{1:t})$ through the Kalman filter as discussed further. In a Markov chain Monte Carlo (MCMC) framework, de Jong (1997) has proposed the so-called scan sampler to sample from $p(y_{1:t} | z_{1:t})$ in a similar class of state space models; see Manrique and Shephard (1998) for some applications.

1.4. Plan

The rest of the paper is organized as follows. Section 2 shows how it is possible to restrict ourselves to estimating $p(y_{1:t} | z_{1:t})$ instead of $p(x_{0:t}, y_{1:t} | z_{1:t})$, leading to an improvement of the Monte Carlo efficiency. The particle filtering algorithm is then described in details.

Section 3 demonstrates the performance of the proposed algorithm via computer simulations for dynamic Tobit and Probit models.

2. Rao-Blackwellized Particle Filtering

2.1. Marginalization

Consider the state space model defined by Eq. (1)-(2)-(3). One has

$$p\left(x_{0:t}\middle|\, z_{1:t}\right)=\int p\left(x_{0:t}\middle|\, y_{1:t}\right)p\left(y_{1:t}\middle|\, z_{1:t}\right)dy_{1:t}.$$

Thus if one obtains (through a SMC method described further) an approximation of the probability distribution associated to the density $p\left(y_{1:t}\middle|\, z_{1:t}\right)$ of the form

$$\widehat{p}_N\left(dy_{1:t}\middle|\, z_{1:t}\right)=\sum_{i=1}^{N}w_t^{(i)}\delta_{\widetilde{y}_{1:t}^{(i)}}\left(dy_{1:t}\right),\ w_t^{(i)}\geq 0,\ \sum_{i=1}^{N}w_t^{(i)}=1,$$

then $p\left(x_{0:t}\middle|\, z_{1:t}\right)$ can be approximated using

$$\widehat{p}_N\left(x_{0:t}\middle|\, z_{1:t}\right)=\sum_{i=1}^{N}w_t^{(i)}p\left(x_{0:t}\middle|\, \widetilde{y}_{1:t}^{(i)}\right);$$

that is, a mixture of Gaussians. From such an approximation, one can estimate $E\left(x_t\middle|\, z_{1:t}\right)$ and $E\left(x_{t-L}\middle|\, z_{1:t}\right)$. For example an estimate of $E\left(x_t\middle|\, z_{1:t}\right)$ is given by

$$\widehat{E}_N\left(x_t\middle|\, z_{1:t}\right)=\int x_t\widehat{p}_N\left(x_{0:t}\middle|\, z_{1:t}\right)dx_{0:t}=\sum_{i=1}^{N}w_t^{(i)}E\left(x_t\middle|\, \widetilde{y}_{1:t}^{(i)}\right),$$

where $E\left(x_t\middle|\, \widetilde{y}_{1:t}^{(i)}\right)$ is computed through the Kalman filter associated with the linear Gaussian state space model defined by (1)-(2). Using the variance decomposition formula, it is clear that for any function $h\left(\cdot\right)$

$$var\left\{h\left(x_t\right)\middle|\, z_{1:t}\right\}\geq var\left[E\left\{h\left(x_t\right)\middle|\, y_{1:t},z_{1:t}\right\}\middle|\, z_{1:t}\right],$$

which shows that estimating $p\left(y_{1:t}\middle|\, z_{1:t}\right)$ only is more efficient.

In order to obtain a SMC approximation of the marginal posterior density $p\left(y_{1:t}\middle|\, z_{1:t}\right)$, one needs to be able to estimate this "target" density pointwise up to a normalizing constant. One has

$$p\left(y_{1:t}\middle|\, z_{1:t}\right)\propto\prod_{k=1}^{t}p\left(z_k\middle|\, y_k\right)p\left(y_k\middle|\, y_{1:k-1}\right),\tag{4}$$

where $p\left(y_1\middle|\, y_{1:0}\right)\triangleq p\left(y_1\right)$. As $p\left(z_k\middle|\, y_k\right)$ is assumed known up to a normalizing constant, it is only necessary to estimate $p\left(y_k\middle|\, y_{1:k-1}\right)$ up to a normalizing constant. This predictive density can be computed using the Kalman filter.

The Kalman filter equations are the following. Set $x_{0|0}=\hat{x}_0$ and $P_{0|0}=P_0$, then for $t=1,2,...$ compute

$$\begin{aligned}
&x_{t|t-1}=A_tx_{t-1|t-1}+F_tu_t,\ P_{t|t-1}=A_tP_{t-1|t-1}A_t^{\mathrm{T}}+B_tB_t^{\mathrm{T}},\\
&y_{t|t-1}=C_tx_{t|t-1}+G_tu_t,\ S_t=C_tP_{t|t-1}C_t^{\mathrm{T}}+D_tD_t^{\mathrm{T}},\\
&x_{t|t}=x_{t|t-1}+P_{t|t-1}C_t^{\mathrm{T}}S_t^{-1}\left(y_t-y_{t|t-1}\right),\ P_{t|t}=P_{t|t-1}-P_{t|t-1}C_t^{\mathrm{T}}S_t^{-1}C_tP_{t|t-1},
\end{aligned}\tag{5}$$

where $x_{t|t-1} \triangleq E(x_t|y_{1:t-1})$, $x_{t|t} \triangleq E(x_t|y_{1:t})$, $y_{t|t-1} \triangleq E(y_t|y_{1:t-1})$, $P_{t|t-1} \triangleq cov(x_t|y_{1:t-1})$, $P_{t|t} \triangleq cov(x_t|y_{1:t})$ and $S_t \triangleq cov(y_t|y_{1:t-1})$. One obtains $p(y_k|y_{1:k-1}) = \mathcal{N}(y_k; y_{k|k-1}, S_k)$, where $\mathcal{N}(y_k; y_{k|k-1}, S_k)$ is a Gaussian distribution of argument y_k, mean $y_{k|k-1}$ and covariance S_k.

2.2. *Particle Filtering*

2.2.1. *Sequential Importance Sampling and Resampling*

We describe briefly here how to apply the Sequential Importance Sampling Resampling (SISR) method so as to sample approximately from $p(y_{1:t}|z_{1:t})$; see Doucet *et al.* (2001) for further details.

At time $t-1$, assume we have say N particles $\left\{\widetilde{y}_{1:t-1}^{(i)}\right\}_{i=1}^{N}$ approximately distributed according to $p(y_{1:t-1}|z_{1:t-1})$ and we want to obtain N particles $\left\{y_{1:t}^{(i)}\right\}_{i=1}^{N}$ distributed according to $p(y_{1:t}|z_{1:t})$. At time t, we "extend" each particle $\widetilde{y}_{1:t-1}^{(i)}$ by sampling $\widetilde{y}_{t}^{(i)}$ according to a conditional density $q_t\left(y_t|\widetilde{y}_{1:t-1}^{(i)}, z_{1:t}\right)$. Thus, each particle $\widetilde{y}_{1:t}^{(i)}$ is distributed according to $p(y_{1:t-1}|z_{1:t-1})\, q_t(y_t|y_{1:t-1}, z_{1:t})$. To correct for the discrepancy between $p(y_{1:t-1}|z_{1:t-1})\, q_t(y_t|y_{1:t-1}, z_{1:t})$ and $p(y_{1:t}|z_{1:t})$, we use importance sampling so that the distribution associated to the density $p(y_{1:t}|z_{1:t})$ is approximated by

$$\widehat{p}_N(dy_{1:t}|z_{1:t}) = \frac{\sum_{i=1}^{N} w\left(\widetilde{y}_{1:t}^{(i)}\right)\delta_{\widetilde{y}_{1:t}^{(i)}}(dy_{1:t})}{\sum_{i=1}^{N} w\left(\widetilde{y}_{1:t}^{(i)}\right)} = \sum_{i=1}^{N} w_t^{(i)}\delta_{\widetilde{y}_{1:t}^{(i)}}(dy_{1:t}), \tag{6}$$

where, using Eq. (4), one has for the importance weight

$$w(y_{1:t}) \propto \frac{p(y_{1:t}|z_{1:t})}{p(y_{1:t-1}|z_{1:t-1})\, q_t(y_t|y_{1:t-1}, z_{1:t})} \propto \frac{p(z_t|y_t)\, p(y_t|y_{1:t-1})}{q_t(y_t|y_{1:t-1}, z_{1:t})}.$$

The performance of the algorithm depends on the importance density $q_t(y_t|y_{1:t-1}, z_{1:t})$. One can select $p(y_t|y_{1:t-1})$ since it is a Gaussian density. In this case the associated importance weight is equal to $w(y_{1:t}) \propto p(z_t|y_t)$. Note that the "optimal" importance density, that is the density minimizing the conditional variance of the weight conditional upon $y_{1:t-1}$ (Doucet *et al.*, 2000), is

$$p(y_t|y_{1:t-1}, z_{1:t}) \propto p(z_t|y_t)\, p(y_t|y_{1:t-1}),$$

and the associated importance weight is

$$w(y_{1:t}) \propto p(z_t|y_{1:t-1}) = \int p(z_t|y_t)\, p(y_t|y_{1:t-1})\, dy_t. \tag{7}$$

It might be possible to compute this weight or not, depending on $p(z_t|y_t)$.

Finally, one obtains N particles $\left\{y_{1:t}^{(i)}\right\}_{i=1}^{N}$ approximately distributed according to $p(y_{1:t}|z_{1:t})$ by resampling from the weighted empirical distribution given in Eq. (6). There are several resampling procedures available in the literature. We adopt here the stratified sampling scheme described in Kitagawa (1996).

Alternative SMC methods can be applied to estimate $p(y_{1:t}|\, z_{1:t})$. In particular, the Auxiliary Particle Filtering (APF) technique of Pitt and Shephard (1999) could be used. The idea behind APF is to extend existing particles $\widetilde{y}_{1:t-1}^{(i)}$ that are the most promising, in the sense that the predictive likelihood $p\left(z_t|\, \widetilde{y}_{1:t-1}^{(i)}\right)$ is large. When $p\left(z_t|\, \widetilde{y}_{1:t-1}^{(i)}\right)$ cannot be computed analytically, the APF proposes an analytical approximation. In this case, the APF and SISR differ significantly. However, when $p\left(z_t|\, \widetilde{y}_{1:t-1}^{(i)}\right)$ can be computed analytically, then the APF uses the "optimal" importance density. This is referred to as "perfect adaptation" (Pitt and Shephard, 1999). In this particular case, the APF and SISR are essentially similar except that the APF reverses the order of the sampling and resampling steps; this is possible as the importance weight is independent of y_t and is obviously more efficient.

2.2.2. *Algorithm*

We limit our presentation to standard choices of importance densities where $q_t\left(y_t|\, y_{1:t-1}, z_{1:t}\right)$ only depends on $(y_{1:t-1}, z_{1:t})$ via z_t and the set of low-dimensional sufficient statistics $x_{t|t-1}$ and $P_{t|t-1}$. We will write $q_t\left(y_t|\, x_{t|t-1}, P_{t|t-1}, z_t\right) \triangleq q_t\left(y_t|\, y_{1:t-1}, z_{1:t}\right)$. This class of densities includes

$$p\left(y_t|\, x_{t|t-1}, P_{t|t-1}\right) \triangleq p\left(y_t|\, y_{1:t-1}\right) = \mathcal{N}\left(y_t; y_{t|t-1}, S_t\right),$$

where $y_{t|t-1}$ and S_t are deterministic functions of $x_{t|t-1}$ and $P_{t|t-1}$. As one typically focuses on features of the marginal $p(y_t|\, z_{1:t})$, it is only necessary to store in memory $\left\{y_t^{(i)}, x_{t|t-1}^{(i)}\right\}_{i=1}^N$ and $P_{t|t-1}$ instead of $\left\{y_{1:t}^{(i)}\right\}_{i=1}^N$. Contrary to algorithms presented in Chen and Liu (2000) and Doucet *et al.* (2000), we point out that one does not have to compute N "full" Kalman filter recursions as most of the calculations only need to be done once. More precisely, one notices that $P_{t|t-1}^{(i)} = P_{t|t-1}$ and $S_{t|t}^{(i)} = S_{t|t}$ for any $i \in \{1, \ldots, N\}$.

Given N particles $\left\{y_{t-1}^{(i)}\right\}_{i=1}^N$ at time $t-1$ distributed approximately according to $p(y_{t-1}|\, z_{1:t-1})$ and the associated sufficient statistics $\left\{x_{t|t-1}^{(i)}\right\}_{i=1}^N$ and $P_{t|t-1}$, the particle filter proceeds as follows at time t.

Sequential Importance Sampling step

- For $i = 1, ..., N$, set $\widetilde{x}_{t|t-1}^{(i)} \triangleq x_{t|t-1}^{(i)}$ and sample $\widetilde{y}_t^{(i)} \sim q_t\left(y_t|\, \widetilde{x}_{t|t-1}^{(i)}, P_{t|t-1}, z_t\right)$.
- For $i = 1, ..., N$, evaluate and normalize the importance weights

$$w_t^{(i)} \propto \frac{p\left(z_t|\, \widetilde{y}_t^{(i)}\right) p\left(\widetilde{y}_t^{(i)} \middle|\, \widetilde{x}_{t|t-1}^{(i)}, P_{t|t-1}\right)}{q_t\left(\widetilde{y}_t^{(i)} \middle|\, \widetilde{x}_{t|t-1}^{(i)}, P_{t|t-1}, z_t\right)}, \quad \sum_{i=1}^N w_t^{(i)} = 1. \tag{8}$$

Resampling Step

- Multiply/Discard particles $\left\{\widetilde{y}_t^{(i)}, \widetilde{x}_{t|t-1}^{(i)}\right\}_{i=1}^N$ with respect to high/low importance weights $w_t^{(i)}$ to obtain N particles $\left\{y_t^{(i)}, x_{t|t-1}^{(i)}\right\}_{i=1}^N$.

Updating step

- Compute $P_{t+1|t}$ given $P_{t|t-1}$ using one step of the Kalman recursion (5).
- For $i = 1, ..., N$, use one step of the Kalman recursion (5) to compute $\left\{x_{t+1|t}^{(i)}\right\}$ given $\left\{y_t^{(i)}, x_{t|t-1}^{(i)}\right\}$ and $P_{t|t-1}$.

The computational complexity of this algorithm at each time step is $\mathcal{O}(N)$. If the (unnormalized) importance weights given by Eq. (8) are upper bounded, then asymptotic convergence ($N \to \infty$) of Monte Carlo estimates towards their true values can be ensured (Crisan, 2001).

2.3. Extensions

There are many potential extensions to both the model and the algorithm.

Model. For the linear Gaussian model (1)-(2), one can readily consider the case where v_t and w_t are correlated and/or add a nonlinear term $\varphi(y_{1:t-1})$ to the right hand side of (2). It is also possible to apply the marginalization method presented above to integrate out analytically (x_t) when the model (1)-(2) is not linear Gaussian but conditionally linear Gaussian as described in Shephard (1994); this extension allows one to consider finite or continuous mixtures of Gaussians.

Another interesting extension consists of partially observed hidden Markov model (HMM): (x_t) is modeled as a finite state space Markov chain and

$$p(y_{1:t}, z_{1:t} | x_{1:t}) = \prod_{k=1}^{t} p(y_k | x_k, y_{1:k-1}) p(z_k | y_k).$$

One can integrate out (x_t) and compute $p(y_k | y_{1:k-1})$ using the HMM filter instead of the Kalman filter as part of the method developed in the previous section.

Algorithm. When the distribution of the importance weights $w_t^{(i)}$ is skewed, the particles $\left\{\widetilde{y}_{1:t}^{(i)}\right\}_{i=1}^{N}$ having high importance weights are selected many times; this results in a "depletion" of samples as numerous particles $y_{1:t}^{(i)}$ and $y_{1:t}^{(j)}$ are in fact equal for $i \neq j$. To perform sample "regeneration", it is possible to use a recently proposed approach based on MCMC steps (Gilks and Berzuini, 2001). It consists of applying to each particle $y_{1:t}^{(i)}$ a (possibly non ergodic) transition kernel $K_t\left(y_{1:t} | y_{1:t}^{(i)}\right)$ of invariant density $p(y_{1:t} | z_{1:t})$. There are an infinite number of possible choices for this kernel. One possibility consists of updating at time t the values $y_{t-M+1:t}$ $(M > 0)$ using the efficient scan sampler (de Jong, 1997), whose computational complexity is $\mathcal{O}(M)$. Though this step is not necessary to ensure theoretical convergence of the algorithm, it can improve results.

3. Simulations

3.1. Dynamic Tobit model

Let us consider the following Tobit model (Manrique and Shephard, 1998),

$$\begin{aligned} x_{t+1} &= \phi x_t + \sigma_v v_{t+1},\ x_0 \sim \mathcal{N}\left(0, \sigma_v^2 / \left(1 - \phi^2\right)\right),\ v_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \\ y_t &= x_t + \sigma_\varepsilon \varepsilon_t,\ \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \\ z_t &= \max(y_t, 0). \end{aligned}$$

Table 1. $m\,(SE)$ for standard filter and RB filter.

Algorithm/N	100	250	500	1000	2500	5000	10000	25000
Standard Filter	33.70	33.64	33.90	33.41	33.45	33.55	33.54	33.52
RB Filter	33.52	33.49	33.51	33.50	33.50	33.49	33.51	33.50

Table 2. $10\sigma\,(SE)$ for standard filter and RB filter.

Algorithm/N	100	250	500	1000	2500	5000	10000	25000
Standard Filter	3.76	2.19	1.61	1.20	0.83	0.62	0.37	0.26
RB Filter	2.50	1.30	1.21	0.99	0.51	0.38	0.29	0.14

It is clear that this model is of the form (1)-(2)-(3). We chose as importance density the "optimal" density‡ $p\,(y_t|\,y_{1:t-1}, z_t)$. If $z_t > 0$, then $y_t = z_t$ and if $z_t = 0$, then

$$p\,(y_t|\,y_{1:t-1}, z_t = 0) \propto p\,(y_t|\,y_{1:t-1})\,\mathbb{I}_{(-\infty,0)}\,(y_t)\,.$$

For the importance weight, one obtains using Eq. (7)

$$w\,(y_{1:t}) \propto \begin{cases} \Phi\left(-\frac{y_{t|t-1}}{\sqrt{S_t}}\right) & \text{if } z_t = 0 \\ \mathcal{N}\left(z_t; y_{t|t-1}, S_t\right) & \text{if } z_t > 0 \end{cases},$$

where $\Phi\,(\cdot)$ is the cumulative function of the standard normal distribution.

We simulate $T = 200$ observations with the known hyperparameters $\phi = 0.99$, $\sigma_v^2 = 0.05$ and $\sigma_w^2 = 0.30$. For various number of particles N, we generate $K = 100$ different realizations of our Rao-blackwellized (RB) filter and of the standard algorithm that estimates $p\,(x_{0:t}, y_{1:t}|\,z_{1:t})$ using the importance density $p\,(x_t, y_t|\,x_{t-1}, y_{t-1}, z_t)$ as $p\,(x_t, y_t|\,x_{t-1}, y_{t-1})$ cannot be used. Our comparison is in terms of the mean and variance of the square error (SE) computed as follows:

$$\begin{aligned} SE\,(l) &= \sum_{t=1}^{T}\left\{x_t - \widehat{E}_{N,l}\,(x_t|\,z_{1:t})\right\}^2, \\ m\,(SE) &= \frac{1}{K}\sum_{l=1}^{K} SE\,(l)\,,\ \sigma^2\,(SE) = \frac{1}{K}\sum_{l=1}^{K}\{SE\,(l) - m\,(SE)\}^2\,, \end{aligned}$$

where $\widehat{E}_{N,l}\,(x_t|\,z_{1:t})$ is computed using the lth realization of the particle filter. We present in Tables 1 and 2 the performance of the standard and RB filters. For a fixed number of particles, the RB filter is not more computationally intensive than the standard filter and it performs significantly better.

3.2. Dynamic Probit model

We analyze here a non-stationary binary time series and more specifically the (aggregated) Tokyo rainfall data set (Knorr-Held, 1999). It consists of $T = 366$ observations with $z_t = 1$

‡Note that one could not use $p\,(y_t|\,y_{1:t-1})$ for importance density when $z_t = 0$. In this case, the importance weight (i.e. Radon Nykodym derivative) does not exist.

indicating that it rained on the tth day of the year and $z_t = 0$ otherwise. We model z_t using a dynamic probit model, that is

$$\Pr(z_t = 1 | \alpha_t) = \Phi(\alpha_t),$$

where (α_t) is modeled using a second order random walk

$$\alpha_t = 2\alpha_{t-1} - \alpha_{t-2} + \sigma_v v_t, \ v_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1). \tag{9}$$

Let us introduce an artificial latent process (y_t) such that

$$y_t = \alpha_t + \varepsilon_t, \ \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \tag{10}$$

and define

$$z_t = \mathbb{I}_{[0,+\infty)}(y_t). \tag{11}$$

It is easy to check that one has

$$\Pr(z_t = 1 | \alpha_t) = \Pr(y_t > 0 | \alpha_t) = \Pr(\varepsilon_t > -\alpha_t) = \Phi(\alpha_t).$$

One can easily rewrite (9)-(10)-(11) in a state space model of the form (1)-(2)-(3) by defining $x_t \triangleq (\alpha_t, \alpha_{t-1})$.

In this case, the introduction of (y_t) being artificial, it is necessary to compare our procedure to a particle filtering method applied to the estimation of $p(x_{0:t} | z_{1:t})$ and not $p(x_{0:t}, y_{1:t} | z_{1:t})$. The motivation for introducing (y_t) comes from the fact that it is possible to use the "optimal" density as importance density

$$p(y_t | y_{1:t-1}, z_t) \propto \begin{cases} p(y_t | y_{1:t-1}) \mathbb{I}_{[0,+\infty)}(y_t) & \text{if } z_t = 1 \\ p(y_t | y_{1:t-1}) \mathbb{I}_{(-\infty,0)}(y_t) & \text{if } z_t = 0 \end{cases},$$

which is a truncated Gaussian distribution, and, using Eq. (7), one obtains

$$w(y_{1:t}) \propto p(z_t | y_{1:t-1}) = \left(1 - \Phi\left(-\frac{y_{t|t-1}}{\sqrt{S_t}}\right)\right)^{z_t} \Phi\left(-\frac{y_{t|t-1}}{\sqrt{S_t}}\right)^{1-z_t}.$$

If one considers $p(x_{0:t} | z_{1:t})$, the "optimal" density cannot be used since the associated importance weight $p(z_t | x_{t-1})$ does not admit an analytical expression. Note that the introduction of (y_t) has already been used to develop efficient MCMC samplers (Albert and Chib, 1993).

The hyperparameter σ_v^2 was set to 0.01. In Figure 1, we display the observations z_t, $E(\Phi(\alpha_t) | y_{1:t})$ and $E(\Phi(\alpha_t) | y_{1:t}) \mp \sqrt{var(\Phi(\alpha_t) | y_{1:t})}$; the estimates are obtained using $N = 1000$ particles. To obtain similar results, the bootstrap filter (Gordon *et al.*, 1993) requires as many as $N = 5000$ particles.

4. Conclusion

This paper proposes a method for recursive state estimation of partially observed Gaussian models. Our algorithm is a SMC method based on marginalization. This marginalization is performed through Kalman filtering methods. Our simulations show that our approach can significantly outperform standard SMC methods.

Throughout this article, the model parameters λ are assumed known. It is possible, however, to perform batch and recursive estimation of these parameters, combining the particle filtering method developed here and recent methods proposed in Liu and West (2001).

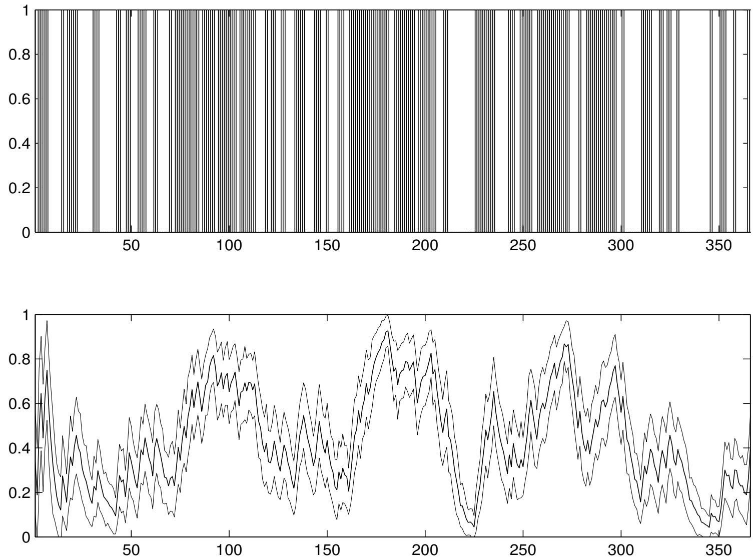

Fig. 1. Top: Binary observations (z_t). Bottom: $E\left(\Phi\left(\alpha_t\right)\middle|\, y_{1:t}\right)$ (thick line) and $E\left(\Phi\left(\alpha_t\right)\middle|\, y_{1:t}\right) \pm \sqrt{var\left(\Phi\left(\alpha_t\right)\middle|\, y_{1:t}\right)}$ (thin lines).

5. Acknowledgments

We are thankful to Leonhard Knorr-Held for providing the Tokyo rainfall dataset and to Sanjeev Arulampalam, Nando de Freitas, Arthur Gretton and Elena Punskaya for valuable comments.

References

[1] Albert J. and Chib S. (1993) Bayesian analysis of binary and polychtomous response data. *J. Am. Statist. Ass.*, **88**, 669-679.

[2] Chen R. and Liu J.S. (2000) Mixture Kalman filters. *J. R. Statist. Soc.* B, **62**, 493-508.

[3] Crisan D. (2001) Particle filters - A theoretical perspective. In *Sequential Monte Carlo Methods in Practice* (eds A. Doucet, J.F.G. de Freitas and N.J. Gordon). New York: Springer-Verlag, 17-41.

[4] de Jong P. (1997) The scan sampler. *Biometrika*, **84**, 929-937.

[5] Doucet A., de Freitas J.F.G. and Gordon N.J. (eds.) (2001) *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

[6] Doucet A., Godsill S.J. and Andrieu C. (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comput.*, **10**, 197-208.

[7] Gelfand A. and Smith, A.F.M. (1990) Sampling based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, **85**, 398-409.

[8] Gilks W.R. and Berzuini C. (2001) Following a moving target - Monte Carlo inference for dynamic Bayesian models. *J. R. Statist. Soc.* B, **63**, pp. 127-146.

[9] Gordon N.J., Salmond D.J. and Smith A.F.M. (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F*, **140**, 107-113.

[10] Kitagawa G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.*, **5**, 1-25.

[11] Knorr-Held L. (1999) Conditional prior proposals in dynamic models. *Scand. J. Statist.*, **26**, 129-144.

[12] Liu J. and West M. (2001) Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice* (eds A. Doucet, J.F.G. de Freitas and N.J. Gordon). New York: Springer-Verlag, 197-223.

[13] Manrique A. and Shephard N. (1998) Simulation-based likelihood inference for limited dependent processes. *Econometrics J.*, **1**, 174-202.

[14] Pitt M.K. and Shephard N. (1999) Filtering via simulation: auxiliary particle filter. *J. Am. Statist. Ass.*, **94**, 590-599.

[15] Shephard N. (1994) Partial non-Gaussian times series models. *Biometrika*, **81**, 115-131.

[16] West M. and Harrison P.J. (1997) *Bayesian Forecasting and Dynamic Models.* 2nd edition, New York: Springer-Verlag.