APPROXIMATE CONDITIONAL MEAN PARTICLE FILTER

Derek Yee, James P. Reilly, Thia Kirubarajan[*]

McMaster University
Department of Electrical and Computer Engineering
Hamilton, ON, Canada
Email: yeed@grads.ece.mcmaster.ca, reillyj@mcmaster.ca, kiruba@mcmaster.ca

ABSTRACT

We consider partially observed non-Gaussian dynamic state space models in which the process equation consists of a combination of linear and nonlinear states and the process noise for the nonlinear state update is distributed according to a mixture of Gaussians. In this paper, we solve a Bayesian filtering problem. The proposed filter is an efficient combination of the particle filter and the approximate conditional mean filter. Simulation results on a time–varying autoregressive signal demonstrate the effectiveness of the proposed algorithm.

1. INTRODUCTION

In many signal processing problems, we encounter a dynamic state space model (DSSM) of the form of (1) – (3). That is, a *partially observed non-Gaussian* DSSM which consists of a combination of linear and nonlinear states, and where the process noise for the nonlinear state update can be modeled as a mixture of Gaussians. Such models are useful e.g., in time–varying autoregressive models. Similar models have been considered in [1]. In this case, optimal filtering is a difficult problem. Arguably, the prescribed solution is the extended mixture Kalman filter (EMKF) [3]. Although it is efficient, it may be possible to outperform the former. In this paper, we propose a novel particle filter that combines the approximate conditional mean (ACM) filter [4] and the particle filter (PF) [2] to render an efficient alternative for solving this problem.

The remainder of this paper is organized as follows. In Section 2, we introduce the considered DSSM. In Sections 3 and 4, we review the PF, and the EMKF, respectively. Section 5 presents the proposed approximate conditional mean particle filter. Section 6 presents some simulation results, and section 7 concludes this paper.

2. DYNAMIC STATE SPACE MODEL

We consider a DSSM of the following form:

$$\boldsymbol{x}_k^1 = \boldsymbol{F}^1(\boldsymbol{x}_{k-n:k-1}^2) + \boldsymbol{A}^1(\boldsymbol{x}_{k-n:k-1}^2)\boldsymbol{x}_{k-1}^1 + \boldsymbol{w}_k^1 \quad (1)$$

$$\boldsymbol{x}_k^2 = \boldsymbol{F}^2(\boldsymbol{x}_{k-n:k-1}^2) + \boldsymbol{A}^2(\boldsymbol{x}_{k-n:k-1}^2)\boldsymbol{x}_k^1 + \boldsymbol{w}_k^2 \quad (2)$$

$$\boldsymbol{y}_k = \boldsymbol{H}(\boldsymbol{x}_k^2) + \boldsymbol{e}_k \quad (3)$$

where $\boldsymbol{x}_k^1 \in \mathbb{R}^{n_1}$, $\boldsymbol{x}_k^2 \in \mathbb{R}^{n_2}$ are the unobserved processes and $\boldsymbol{y}_k \in \mathbb{R}^{n_y}$ the noisy observations. Here, $\boldsymbol{F}^1(\cdot)$, $\boldsymbol{A}^1(\cdot)$, $\boldsymbol{F}^2(\cdot)$, $\boldsymbol{A}^2(\cdot)$, $\boldsymbol{H}(\cdot)$ are known functions with proper dimensions and the notation $(\cdot)_{l:m}$, indicates all the elements from time l to time m. The process noise $\boldsymbol{w}_k^1$, and measurement noise $\boldsymbol{e}_k$ are assumed to be mutually independent zero-mean Gaussian white noise sequences, $\boldsymbol{w}_k^1 \sim \mathcal{N}(\boldsymbol{w}_k^1; \boldsymbol{0}, \boldsymbol{Q}_k^1)$ and $\boldsymbol{e}_k \sim \mathcal{N}(\boldsymbol{e}_k; \boldsymbol{0}, \boldsymbol{R}_k)$. The process noise driving (2) is assumed to be a white noise sequence that is distributed according to a Gaussian Mixture Model (GMM)

$$p(\boldsymbol{w}_k^2) = \sum_{j=1}^{N} p_j \mathcal{N}(\boldsymbol{w}_k^2; \overline{\boldsymbol{w}}_k^{2,(j)}, \boldsymbol{Q}_k^{2,(j)}) \quad (4)$$

where $\sum_{j=1}^{N} p_j = 1$. Furthermore, the initial states are assumed to be mutually independent zero-mean Gaussian random variables, $\boldsymbol{x}_0^1 \sim \mathcal{N}(\boldsymbol{x}_0^1; \hat{\boldsymbol{x}}_0^1, \hat{\boldsymbol{P}}_0^1)$ and $\boldsymbol{x}_0^2 \sim \mathcal{N}(\boldsymbol{x}_0^2; \hat{\boldsymbol{x}}_0^2, \hat{\boldsymbol{P}}_0^2)$.

The main objective is to sequentially in time compute the minimum mean square error estimate (MMSE) of $\boldsymbol{x}_k = [\boldsymbol{x}_k^{1\,T}, \boldsymbol{x}_k^{2\,T}]^T$, and its associated covariance $cov_{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k]$. That is,

$$\mathbb{E}_{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k] = \int \boldsymbol{x}_k p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) d\boldsymbol{x}_k \quad (5)$$

$$cov_{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k] = \int \overline{\boldsymbol{x}}_k \overline{\boldsymbol{x}}_k^T p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) d\boldsymbol{x}_k \quad (6)$$

where $\overline{\boldsymbol{x}}_k = \boldsymbol{x}_k - \mathbb{E}_{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k]$. Unfortunately, the evaluation of (5) and (6) involve complex intractable multidimensional integrals. Thus, we propose to use particle filtering to recursively estimate (5) and (6).

3. PARTICLE FILTER

A standard PF utilizes a weighted set of samples to approximate the *joint* posterior probability density function (pdf) $p(\boldsymbol{x}_k^1, \boldsymbol{x}_k^2|\boldsymbol{y}_{1:k})$. Thus at time k, by drawing N_p samples of $(\boldsymbol{x}_k^1, \boldsymbol{x}_k^2)$ from a *importance function* $q(\boldsymbol{x}_k^1, \boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^1, \boldsymbol{x}_{1:k-1}^2, \boldsymbol{y}_k)$, that is $(\boldsymbol{x}_k^{1,(i)}, \boldsymbol{x}_k^{2,(i)}) \sim q(\boldsymbol{x}_k^1, \boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{1,(i)}, \boldsymbol{x}_{1:k-1}^{2,(i)}, \boldsymbol{y}_k)$ for $i = 1, \ldots, N_p$ and recursively updating the importance weights $\{\boldsymbol{w}_k^{(i)}\}_{i=1}^{N_p}$ with

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{2,(i)}) p(\boldsymbol{x}_k^{1,(i)}, \boldsymbol{x}_k^{2,(i)}|\boldsymbol{x}_{1:k-1}^{1,(i)}, \boldsymbol{x}_{1:k-1}^{2,(i)})}{q(\boldsymbol{x}_k^{1,(i)}, \boldsymbol{x}_k^{2,(i)}|\boldsymbol{x}_{1:k-1}^{1,(i)}, \boldsymbol{x}_{1:k-1}^{2,(i)}, \boldsymbol{y}_k)} . \quad (7)$$

We have for an estimate of (5) and (6)

$$\widehat{\mathbb{E}}_{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k] = \sum_{i=1}^{N_p} \tilde{w}_k^{(i)} \boldsymbol{x}_k \quad (8)$$

[*]This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC).

$$\widehat{cov}_{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k] = \sum_{i=1}^{N_p} \tilde{w}_k^{(i)} \overline{\boldsymbol{x}}_k^{(i)} \overline{\boldsymbol{x}}_k^{(i)\,T} \quad (9)$$

where $\tilde{w}_k^{(i)} = [\sum_{j=1}^{N_p} w_k^{(j)}]^{-1} w_k^{(i)}$ is the normalized importance weight and $\overline{\boldsymbol{x}}_k^{(i)} = \boldsymbol{x}_k^{(i)} - \widehat{\mathbb{E}}_{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k]$.

In (7), the likelihood $p(\boldsymbol{y}_k|\boldsymbol{x}_k^{2,(i)})$ is given by

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k^{2,(i)}) = N(\boldsymbol{y}_k; \boldsymbol{H}(\boldsymbol{x}_k^{2,(i)}), \boldsymbol{R}_k) \quad (10)$$

and the prior $p(\boldsymbol{x}_k^1, \boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{1,(i)}, \boldsymbol{x}_{1:k-1}^{2,(i)})$ by

$$p(\boldsymbol{x}_k^1, \boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{1,(i)}, \boldsymbol{x}_{1:k-1}^{2,(i)}) = p(\boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{1,(i)}, \boldsymbol{x}_k^1, \boldsymbol{x}_{1:k-1}^{2,(i)}) p(\boldsymbol{x}_k^1|\boldsymbol{x}_{1:k-1}^{1,(i)}, \boldsymbol{x}_{1:k-1}^{2,(i)}) \quad (11)$$

where

$$p(\boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{1,(i)}, \boldsymbol{x}_k^1, \boldsymbol{x}_{1:k-1}^{2,(i)}) = \sum_{j=1}^{N} p_j \mathcal{N}(\boldsymbol{x}_k^2, \boldsymbol{F}^2(\boldsymbol{x}_{k-n:k-1}^{2,(i)}) + \overline{\boldsymbol{w}}_k^{2,(j)} + \boldsymbol{A}^2(\boldsymbol{x}_{k-n:k-1}^{2,(i)})\boldsymbol{x}_k^1, \boldsymbol{Q}_k^{2,(j)}) \quad (12)$$

$$p(\boldsymbol{x}_k^1|\boldsymbol{x}_{1:k-1}^{1,(i)}, \boldsymbol{x}_{1:k-1}^{2,(i)}) = \mathcal{N}(\boldsymbol{x}_k^1, \boldsymbol{F}^1(\boldsymbol{x}_{k-n:k-1}^{2,(i)}) + \boldsymbol{A}^1(\boldsymbol{x}_{k-n:k-1}^{2,(i)})\boldsymbol{x}_{k-1}^{1,(i)}, \boldsymbol{Q}_k^1) \quad (13)$$

In practice, particle filtering suffer from *the Degeneracy problem*. That is, after a few iterations, all but a few particles possess insignificant weights. As a result, (8) to (9) form poor approximations of (5) and (6), respectively. Typically, to mitigate this problem, we introduce a resampling step [2]. The basic idea is to discard particles with weak importance weights and to multiply ones with significant importance weights. In this work, we introduce resampling dynamically. That is, we only resample if the effective sample size $\widehat{N}_{eff} = 1/\sum_{i=1}^{N_p}(\tilde{w}_k^{(i)})^2$ is below the threshold $N_{th} = 0.5N_p$.

4. EXTENDED MIXTURE KALMAN FILTER

For the considered DSSM, it is possible to design a better PF that admits estimates with lower variances. The idea is to exploit the linear sub-structure of our given DSSM. Indeed, if we follow the lead of [3], and introduce a indicator random variable $I_k \in \mathcal{I}^N = \{n|n = 1, \dots, N\}$ that satisfies

$$I_k = \begin{cases} 1 & \text{if } \boldsymbol{w}_k^2 \sim \mathcal{N}(\boldsymbol{w}_k^2; \overline{\boldsymbol{w}}_k^{2,(1)}, \boldsymbol{Q}_k^{2,(1)}) \\ \vdots & \\ N & \text{if } \boldsymbol{w}_k^2 \sim \mathcal{N}(\boldsymbol{w}_k^2; \overline{\boldsymbol{w}}_k^{2,(N)}, \boldsymbol{Q}_k^{2,(N)}) \end{cases}$$

where $p(I_k = 1) = p_1$, ..., $p(I_k = N) = p_N$, we can note that (1)-(2) conditional on $\boldsymbol{x}_{1:k}^2$ and $I_{1:k}$ reduces to a linear Gaussian (LG) system in $\boldsymbol{x}_k^1$ for which the Kalman filter (KF) is the optimal estimator. Intuitively, the random variable I_k indicates the *effective* distribution of $\boldsymbol{w}_k^2$ at time index k. Thus if we write $p(\boldsymbol{x}_k^1, \boldsymbol{x}_{1:k}^2, I_{1:k}|\boldsymbol{y}_{1:k})$ as

$$p(\boldsymbol{x}_k^1, \boldsymbol{x}_{1:k}^2, I_{1:k}|\boldsymbol{y}_{1:k}) = p(\boldsymbol{x}_k^1|\boldsymbol{x}_{1:k}^2, I_{1:k}) p(\boldsymbol{x}_{1:k}^2, I_{1:k}|\boldsymbol{y}_{1:k}), \quad (14)$$

it is apparent that we can use the optimal KF to obtain the Gaussian pdf $p(\boldsymbol{x}_k^1|\boldsymbol{x}_{1:k}^2, I_{1:k})$, and use the PF to estimate $p(\boldsymbol{x}_{1:k}^2, I_{1:k}|\boldsymbol{y}_{1:k})$. This approach is also known as the Rao-Blackwellized PF (RBPF) [1]. Thus at time k, if we draw N_p samples of $(\boldsymbol{x}_k^2, I_k)$ from a *importance function* $q(\boldsymbol{x}_k^2, I_k|\boldsymbol{x}_{1:k-1}^2, I_{1:k-1}, \boldsymbol{y}_k)$, that is $(\boldsymbol{x}_k^{2,(i)}, I_k^{(i)}) \sim q(\boldsymbol{x}_k^2, I_k|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)}, \boldsymbol{y}_k)$ for $i = 1, \dots, N_p$ and recursively update the importance weights $\{\boldsymbol{w}_k^{(i)}\}_{i=1}^{N_p}$ as

$$w_k^{(i)} \propto \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{2,(i)}) p(\boldsymbol{x}_k^{2,(i)}, I_k^{(i)}|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)})}{q(\boldsymbol{x}_k^{2,(i)}, I_k^{(i)}|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)}, \boldsymbol{y}_k)} w_{k-1}^{(i)}. \quad (15)$$

Expectations of interest, such as $\mathbb{E}_{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k]$ and its associated covariance $cov_{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k]$ can be approximated by

$$\widehat{\mathbb{E}}_{p(\boldsymbol{x}_k^1|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^1] = \sum_{i=1}^{N_p} \tilde{w}_k^{(i)} \boldsymbol{x}_{k|k}^{1,(i)} \quad (16)$$

$$\widehat{cov}_{p(\boldsymbol{x}_k^1|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^1] = \sum_{i=1}^{N_p} \tilde{w}_k^{(i)} \Big(\boldsymbol{P}_{k|k}^{1,(i)} + (\boldsymbol{x}_{k|k}^{1,(i)} - \widehat{\mathbb{E}}_{p(\boldsymbol{x}_k^1|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^1])(\boldsymbol{x}_{k|k}^{1,(i)} - \widehat{\mathbb{E}}_{p(\boldsymbol{x}_k^1|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^1])^T \Big) \quad (17)$$

$$\widehat{\mathbb{E}}_{p(\boldsymbol{x}_k^2|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^2] = \sum_{i=1}^{N_p} \tilde{w}_k^{(i)} \boldsymbol{x}_k^{2,(i)} \quad (18)$$

$$\widehat{cov}_{p(\boldsymbol{x}_k^2|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^2] = \sum_{i=1}^{N_p} \tilde{w}_k^{(i)} \overline{\boldsymbol{x}}_k^{2,(i)} \overline{\boldsymbol{x}}_k^{2,(i)\,T} \quad (19)$$

where $\boldsymbol{x}_{k|k}^{1,(i)} = \mathbb{E}_{p(\boldsymbol{x}_k^1|\boldsymbol{x}_{1:k}^{2,(i)}, I_{1:k}^{(i)})}[\boldsymbol{x}_k^1]$, $\boldsymbol{P}_{k|k}^{1,(i)} = cov_{p(\boldsymbol{x}_k^1|\boldsymbol{x}_{1:k}^{2,(i)}, I_{1:k}^{(i)})}[\boldsymbol{x}_k^1]$ and $\overline{\boldsymbol{x}}_k^{2,(i)} = \boldsymbol{x}_k^{2,(i)} - \widehat{\mathbb{E}}_{p(\boldsymbol{x}_k^2|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^2]$.

In (15), the likelihood $p(\boldsymbol{y}_k|\boldsymbol{x}_k^{2,(i)})$ is given by (10) and the prior $p(\boldsymbol{x}_k^2, I_k|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)})$ by

$$p(\boldsymbol{x}_k^2, I_k|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)}) = p(\boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)}, I_k) p(I_k) \quad (20)$$

where

$$p(\boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)}, I_k) = N(\boldsymbol{x}_k^2; \boldsymbol{x}_{k|k-1}^{2,(i)}, \boldsymbol{S}_{k|k-1}^{(i)}) \quad (21)$$

is a Gaussian pdf with mean $\boldsymbol{x}_{k|k-1}^{2,(i)} = \mathbb{E}[\boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)}, I_k]$, and covariance $\boldsymbol{S}_{k|k-1}^{(i)} = cov[\boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)}, I_k]$. As mentioned before, we can use the KF to efficiently compute $p(\boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^{2,(i)}, I_{1:k-1}^{(i)}, I_k)$, $\boldsymbol{x}_{k|k}^{1,(i)}$, $\boldsymbol{P}_{k|k}^{1,(i)}$ in (20), (16), (17), respectively, i.e.,

$$\boldsymbol{x}_{k|k}^{1,(i)} = \boldsymbol{x}_{k|k-1}^{1,(i)} + \boldsymbol{P}_{k|k-1}^{1,(i)} \boldsymbol{A}^2(\boldsymbol{x}_{k-n:k-1}^{2,(i)})^T \times \boldsymbol{S}_{k|k-1}^{(i)\,-1}(\boldsymbol{x}_k^2 - \boldsymbol{x}_{k|k-1}^{2,(i)}) \quad (22)$$

$$\boldsymbol{P}_{k|k}^{1,(i)} = \boldsymbol{P}_{k|k-1}^{1,(i)} - \boldsymbol{P}_{k|k-1}^{1,(i)} \boldsymbol{A}^2(\boldsymbol{x}_{k-n:k-1}^{2,(i)})^T \times \boldsymbol{S}_{k|k-1}^{(i)\,-1} \boldsymbol{A}^2(\boldsymbol{x}_{k-n:k-1}^{2,(i)}) \boldsymbol{P}_{k|k-1}^{1,(i)} \quad (23)$$

where

$$\boldsymbol{x}_{k|k-1}^{1,(i)} = \boldsymbol{F}^1(\boldsymbol{x}_{k-n:k-1}^2) + \boldsymbol{A}^1(\boldsymbol{x}_{k-n:k-1}^{2,(i)}) \boldsymbol{x}_{k-1|k-1}^{1,(i)} \quad (24)$$

$$\boldsymbol{P}_{k|k-1}^{1,(i)} = \boldsymbol{A}^1(\boldsymbol{x}_{k-n:k-1}^{2,(i)}) \boldsymbol{P}_{k-1|k-1}^{1,(i)} \boldsymbol{A}^1(\boldsymbol{x}_{k-n:k-1}^{2,(i)})^T + \boldsymbol{Q}_k^1 \quad (25)$$

$$\boldsymbol{x}_{k|k-1}^{2,(i)} = \boldsymbol{F}^2(\boldsymbol{x}_{k-n:k-1}^{2,(i)}) + \boldsymbol{A}^2(\boldsymbol{x}_{k-n:k-1}^{2,(i)}) \boldsymbol{x}_{k|k-1}^{1,(i)} + \overline{\boldsymbol{w}}_k^{2,(I_k)} \quad (26)$$

$$\boldsymbol{S}_{k|k-1}^{(i)} = \boldsymbol{A}^2(\boldsymbol{x}_{k-n:k-1}^{2,(i)}) \boldsymbol{P}_{k|k-1}^{1,(i)} \boldsymbol{A}^2(\boldsymbol{x}_{k-n:k-1}^{2,(i)})^T + \boldsymbol{Q}_k^{2,(I_k)} \quad (27)$$

Thus, the EMKF is more efficient than the standard PF. Yet, it may be possible to further increase efficiency. Indeed, for particle filtering, it is advantageous to reduce the dimensionality of the space in which we draw samples from [2]. Thus, for the considered DSSM, we endeavor to design a novel PF that exploits the structure of the considered DSSM while dispensing of the need to introduce a Indicator random variable I_k. The advantages are clear. By eliminating the need to introduce a Indicator random variable I_k, the task of using a PF to approximate $p(\boldsymbol{x}^2_{1:k}, I_{1:k}|\boldsymbol{y}_{1:k})$ is reduced to one of approximating a lower dimensional pdf $p(\boldsymbol{x}^2_{1:k}|\boldsymbol{y}_{1:k})$. Intuitively, we require a reduced number of particles to achieve a certain level of performance. In the following, we develop these ideas, and proceed with the derivation of the ACM-PF.

5. APPROXIMATE CONDITIONAL MEAN PARTICLE FILTER

We begin by writing the joint posterior pdf $p(\boldsymbol{x}^1_{1:k}, \boldsymbol{x}^2_{1:k}|\boldsymbol{y}_{1:k})$ as

$$p(\boldsymbol{x}^1_k, \boldsymbol{x}^2_{1:k}|\boldsymbol{y}_{1:k}) = p(\boldsymbol{x}^1_k|\boldsymbol{x}^2_{1:k})p(\boldsymbol{x}^2_{1:k}|\boldsymbol{y}_{1:k}). \tag{28}$$

As mentioned before, we aim to exploit the linear substructure of the considered DSSM. Thus, we only need to derive a particle filtering algorithm that approximates $p(\boldsymbol{x}^2_{1:k}|\boldsymbol{y}_{1:k})$. Thus at time k, if we draw N_p samples of $\boldsymbol{x}^2_k$ from $q(\boldsymbol{x}^2_k|\boldsymbol{x}^2_{1:k-1}, \boldsymbol{y}_k)$, that is $\boldsymbol{x}^{2,(i)}_k \sim q(\boldsymbol{x}^2_k|\boldsymbol{x}^{2,(i)}_{1:k-1}, \boldsymbol{y}_k)$ for $i = 1, \ldots, N_p$ and update the importance weights $\{\boldsymbol{w}^{(i)}_k\}^{N_p}_{i=1}$ as

$$w^{(i)}_k \propto \frac{p(\boldsymbol{y}_k|\boldsymbol{x}^{2,(i)}_k)p(\boldsymbol{x}^{2,(i)}_k|\boldsymbol{x}^{2,(i)}_{1:k-1})}{q(\boldsymbol{x}^{2,(i)}_k|\boldsymbol{x}^{2,(i)}_{1:k-1}, \boldsymbol{y}_k)} w^{(i)}_{k-1} \tag{29}$$

where the likelihood $p(\boldsymbol{y}_k|\boldsymbol{x}^{2,(i)}_k)$ is given by (10) and the prior $p(\boldsymbol{x}^2_k|\boldsymbol{x}^2_{1:k-1})$ by (39) which is discussed below. It can be shown that an approximation of $\mathbb{E}_{p(\boldsymbol{x}^1_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}^1_k]$ and its associated conditional covariance $cov_{p(\boldsymbol{x}^1_k|\boldsymbol{y}_{1:k})}[\boldsymbol{x}^1_k]$ are in the form of

$$\widehat{\mathbb{E}}_{p(\boldsymbol{x}^1_k|\boldsymbol{y}_{1:k})}\left[\boldsymbol{x}^1_k\right] = \sum_{i=1}^{N_p} \tilde{w}^{(i)}_k \boldsymbol{x}^{1,(i)}_{k|k} \tag{30}$$

$$\widehat{cov}_{p(\boldsymbol{x}^1_k|\boldsymbol{y}_{1:k})}\left[\boldsymbol{x}^1_k\right] = \sum_{i=1}^{N_p} \tilde{w}^{(i)}_k \Bigg(\boldsymbol{P}^{1,(i)}_{k|k} + (\boldsymbol{x}^{1,(i)}_{k|k} - \widehat{\mathbb{E}}_{p(\boldsymbol{x}^1_k|\boldsymbol{y}_{1:k})}\left[\boldsymbol{x}^1_k\right])(\boldsymbol{x}^{1,(i)}_{k|k} - \widehat{\mathbb{E}}_{p(\boldsymbol{x}^1_k|\boldsymbol{y}_{1:k})}\left[\boldsymbol{x}^1_k\right])^T \Bigg) \tag{31}$$

where $\boldsymbol{x}^{1,(i)}_{k|k} = \mathbb{E}_{p(\boldsymbol{x}^1_k|\boldsymbol{x}^{2,(i)}_{1:k})}\left[\boldsymbol{x}^1_k\right]$ and $\boldsymbol{P}^{1,(i)}_{k|k} = cov_{p(\boldsymbol{x}^1_k|\boldsymbol{x}^{2,(i)}_{1:k})}\left[\boldsymbol{x}^1_k\right]$. For this case, we also use (18) and (19) to compute approximations for $\mathbb{E}_{p(\boldsymbol{x}^2_k|\boldsymbol{y}_{1:k})}\left[\boldsymbol{x}^2_k\right]$ and $cov_{p(\boldsymbol{x}^2_k|\boldsymbol{y}_{1:k})}\left[\boldsymbol{x}^2_k\right]$, respectively[1]. As in the case of the EMKF, we aim to compute $\boldsymbol{x}^{1,(i)}_{k|k}$ and $\boldsymbol{P}^{1,(i)}_{k|k}$ analytically. Recall for the EMKF, the former are computed via the KF. However, in [5] it is shown that $p(\boldsymbol{x}^1_k|\boldsymbol{x}^{2,(i)}_{1:k})$ is a GMM such that the number of mixands increases exponentially with k. Hence, $\boldsymbol{x}^{1,(i)}_{k|k}$ and $\boldsymbol{P}^{1,(i)}_{k|k}$ are obtained through a growing symphony of KF's, each corresponding to one mixand of $p(\boldsymbol{x}^1_k|\boldsymbol{x}^{2,(i)}_{1:k})$. For practical applications, it is infeasible to sequentially in time compute $\boldsymbol{x}^{1,(i)}_{k|k}$ and $\boldsymbol{P}^{1,(i)}_{k|k}$ for all time indexes k. Thus, with the aim of designing a computationally attractive algorithm, we follow [4] and adopt the surprisingly effective assumption that for all time indexes k, $p(\boldsymbol{x}^1_k|\boldsymbol{x}^2_{1:k-1})$ is a Gaussian pdf $\mathcal{N}(\boldsymbol{x}^1_k; \boldsymbol{x}^1_{k|k-1}, \boldsymbol{P}^1_{k|k-1})$, that is

$$p(\boldsymbol{x}^1_k|\boldsymbol{x}^2_{1:k-1}) = \mathcal{N}(\boldsymbol{x}^1_k; \boldsymbol{x}^1_{k|k-1}, \boldsymbol{P}^1_{k|k-1}). \tag{32}$$

Under this assumption, which is known as the *Masreliez approximation*, we can derive a so-called ACM filter [7, 6] for $\boldsymbol{x}^1_k$. In [4], for a linear DSSM with non-Gaussian observation noise distributed in accordance to a GMM, the ACM filter yields near optimal performance. In particular, conditional on $\boldsymbol{x}^2_{1:k}$, (1)-(2) constitutes such a model, that is, a linear substructure with non-Gaussian observation noise distributed in accordance to a GMM. Hence, we propose to merge the ACM filter, and the standard PF into a hybrid algorithm called the ACM-PF. That is, conditional on the i-th particle $\boldsymbol{x}^{2,(i)}_{1:k}$, we compute $\boldsymbol{x}^{1,(i)}_{k|k}$ and $\boldsymbol{P}^{1,(i)}_{k|k}$ with a ACM filter [4]:

$$\boldsymbol{x}^{1,(i)}_{k|k} = \boldsymbol{x}^{1,(i)}_{k|k-1} + \boldsymbol{P}^{1,(i)}_{k|k-1}\boldsymbol{A}^2(\boldsymbol{x}^{2,(i)}_{k-n:k-1})^T g_k(\boldsymbol{x}^{2,(i)}_k) \tag{33}$$

$$\boldsymbol{P}^{1,(i)}_{k|k} = \boldsymbol{P}^{1,(i)}_{k|k-1} - \boldsymbol{P}^{1,(i)}_{k|k-1}\boldsymbol{A}^2(\boldsymbol{x}^{2,(i)}_{k-n:k-1})^T \times G_k(\boldsymbol{x}^{2,(i)}_k)\boldsymbol{A}^2(\boldsymbol{x}^{2,(i)}_{k-n:k-1})\boldsymbol{P}^{1,(i)}_{k|k-1} \tag{34}$$

where

$$\boldsymbol{x}^{1,(i)}_{k|k-1} = \boldsymbol{F}^1(\boldsymbol{x}^{2,(i)}_{k-n:k-1}) + \boldsymbol{A}^1(\boldsymbol{x}^{2,(i)}_{k-n:k-1})\boldsymbol{x}^{1,(i)}_{k-1|k-1} \tag{35}$$

$$\boldsymbol{P}^{1,(i)}_{k|k-1} = \boldsymbol{A}^1(\boldsymbol{x}^{2,(i)}_{k-n:k-1})\boldsymbol{P}^{1,(i)}_{k-1|k-1}\boldsymbol{A}^1(\boldsymbol{x}^{2,(i)}_{k-n:k-1})^T + \boldsymbol{Q}^1_k \tag{36}$$

and

$$g_k(\boldsymbol{x}^{2,(i)}_k) = -\frac{1}{p(\boldsymbol{x}^{2,(i)}_k|\boldsymbol{x}^{2,(i)}_{1:k-1})}\nabla_{\boldsymbol{x}^{2,(i)}_k} p(\boldsymbol{x}^{2,(i)}_k|\boldsymbol{x}^{2,(i)}_{1:k-1}) \tag{37}$$

$$G_k(\boldsymbol{x}^{2,(i)}_k) = \nabla_{\boldsymbol{x}^{2,(i)}_k} g_k(\boldsymbol{x}^{2,(i)}_k)^T. \tag{38}$$

In the above, the notation ∇ denotes the gradient operator. Moreover, since $p(\boldsymbol{x}^1_k|\boldsymbol{x}^2_{1:k-1}) = \mathcal{N}(\boldsymbol{x}^1_k; \boldsymbol{x}^1_{k|k-1}, \boldsymbol{P}^1_{k|k-1})$, it can be shown that the prior density $p(\boldsymbol{x}^2_k|\boldsymbol{x}^2_{1:k-1})$ is in the form of

$$p(\boldsymbol{x}^2_k|\boldsymbol{x}^2_{1:k-1}) = \sum_{j=1}^{N} p_j N(\boldsymbol{x}^2_k; \boldsymbol{x}^{2,(j)}_{k|k-1}, \boldsymbol{P}^{2,(j)}_{k|k-1}) \tag{39}$$

where

$$\boldsymbol{x}^{2,(j)}_{k|k-1} = \boldsymbol{F}^2(\boldsymbol{x}^2_{k-n:k-1}) + \boldsymbol{A}^2(\boldsymbol{x}^2_{k-n:k-1})\boldsymbol{x}^1_{k|k-1} + \overline{\boldsymbol{w}}^{2,(j)}_k \tag{40}$$

$$\boldsymbol{P}^{2,(j)}_{k|k-1} = \boldsymbol{A}^2(\boldsymbol{x}^2_{k-n:k-1})\boldsymbol{P}^1_{k|k-1}\boldsymbol{A}^2(\boldsymbol{x}^2_{k-n:k-1})^T + \boldsymbol{Q}^{2,(j)}_k. \tag{41}$$

Thus, by using (29), (10), (39) and a appropriately chosen importance function $q(\boldsymbol{x}^2_k|\boldsymbol{x}^{2,(i)}_{1:k-1}, \boldsymbol{y}_k)$, we can recursively in time compute the importance weight $w^{(i)}_k$, and all the estimates of interest.

Finally, we point out that the ACM filter (33)–(38) has a structure that is similar to the KF (22)–(27). Indeed, it can be shown that the ACM filter reduces to the KF when $\boldsymbol{w}^2_k$ is Gaussian distributed. Therefore, it follows that the ACM-PF is equivalent to the EMKF if $\boldsymbol{w}^2_k \sim \mathcal{N}(\boldsymbol{w}^2_k; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is some mean and covariance, respectively. The algorithm for the ACM-PF is

[1] For the ACM-PF, we emphasize that the importance weights $w^{(i)}_k$ are given by (29) not (15).

summarized as follows.

Approximate conditional mean Particle filter (ACM-PF)

---

1. Initialization: For $i = 1, \ldots, N_p$, we initialize the particles, $\boldsymbol{x}_0^{2,(i)} \sim p(\boldsymbol{x}_0^2)$, $\boldsymbol{x}_{0|0}^{1,(i)} = \hat{\boldsymbol{x}}_0^1$, $\boldsymbol{P}_{0|0}^{1,(i)} = \hat{\boldsymbol{P}}_0^1$ and set $w_0^{(i)} = \frac{1}{N_p}$.
2. New particles: For $i = 1, \ldots, N_p$, set $\tilde{\boldsymbol{x}}_{k-n:k-1}^{2,(i)} = \boldsymbol{x}_{k-n:k-1}^{2,(i)}$, $\tilde{\boldsymbol{x}}_{k-1|k-1}^{1,(i)} = \boldsymbol{x}_{k-1|k-1}^{1,(i)}$, $\tilde{\boldsymbol{P}}_{k-1|k-1}^{1,(i)} = \boldsymbol{P}_{k-1|k-1}^{1,(i)}$.
   - Proposals: Draw $\tilde{\boldsymbol{x}}_k^{2,(i)} \sim q(\boldsymbol{x}_k^2 | \tilde{\boldsymbol{x}}_{1:k-1}^{2,(i)}, \boldsymbol{y}_k)$.
   - ACM prediction: Compute $\tilde{\boldsymbol{x}}_{k|k-1}^{1,(i)}$, $\tilde{\boldsymbol{P}}_{k|k-1}^{1,(i)}$ using (35), and (36), respectively.
   - ACM update: Compute $\tilde{\boldsymbol{x}}_{k|k}^{1,(i)}$, $\tilde{\boldsymbol{P}}_{k|k}^{1,(i)}$ using (33), and (34), respectively.
3. Calculate Importance Weights: For $i = 1, \ldots, N_p$, evaluate the importance weights up to a normalizing constant
$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(\boldsymbol{y}_k | \tilde{\boldsymbol{x}}_k^{2,(i)}) p(\tilde{\boldsymbol{x}}_k^{2,(i)} | \tilde{\boldsymbol{x}}_{1:k-1}^{2,(i)})}{q(\tilde{\boldsymbol{x}}_k^{2,(i)} | \tilde{\boldsymbol{x}}_{1:k-1}^{2,(i)}, \boldsymbol{y}_k)}$$
and normalize importance weights to $\tilde{w}_k^{(i)}$.
4. If $\widehat{N}_{eff} < N_{th}$,
   - Resample $\{\tilde{\boldsymbol{x}}_{k-n+1:k}^{2,(i)}\}_{i=1}^{N_p}$, $\{\tilde{\boldsymbol{x}}_{k|k}^{1,(i)}\}_{i=1}^{N_p}$, $\{\tilde{\boldsymbol{P}}_{k|k}^{1,(i)}\}_{i=1}^{N_p}$ w.r.t importance weights to obtain $\{\boldsymbol{x}_{k-n+1:k}^{2,(i)}\}_{i=1}^{N_p}$, $\{\boldsymbol{x}_{k|k}^{1,(i)}\}_{i=1}^{N_p}$, $\{\boldsymbol{P}_{k|k}^{1,(i)}\}_{i=1}^{N_p}$ and set $w_k^{(i)} = \frac{1}{N_p}$ for $i = 1, \ldots, N_p$.

   otherwise

   - Set $\tilde{\boldsymbol{x}}_{k-n+1:k}^{2,(i)} = \boldsymbol{x}_{k-n+1:k}^{2,(i)}$, $\tilde{\boldsymbol{x}}_{k|k}^{1,(i)} = \boldsymbol{x}_{k|k}^{1,(i)}$, and $\tilde{\boldsymbol{P}}_{k|k}^{1,(i)} = \boldsymbol{P}_{k|k}^{1,(i)}$ for $i = 1, \ldots, N_p$.
5. Estimates: Compute $\widehat{\mathbb{E}}_{p(\boldsymbol{x}_k^1|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^1]$, $\widehat{cov}_{p(\boldsymbol{x}_k^1|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^1]$, $\widehat{\mathbb{E}}_{p(\boldsymbol{x}_k^2|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^2]$, and $\widehat{cov}_{p(\boldsymbol{x}_k^2|\boldsymbol{y}_{1:k})}[\boldsymbol{x}_k^2]$ using (30), (31), (18), and (19), respectively.
6. Set $k = k + 1$, and go back to step 2.

---

6. SIMULATIONS

The proposed algorithm is applied to a P-th order time-varying autoregressive (TVAR) model that is driven by mixture Gaussian noise. That is,

$$\begin{aligned} \boldsymbol{a}_k &= \boldsymbol{F}\boldsymbol{a}_{k-1} + \boldsymbol{w}_k^1 && (42)\\ z_k &= \boldsymbol{G}(z_{k-P:k-1})\boldsymbol{a}_k + w_k^2 && (43)\\ y_k &= z_k + e_k, \quad k = 1, \ldots, N && (44) \end{aligned}$$

where $\boldsymbol{F} = \beta \boldsymbol{I}_{P\times P}$, $\boldsymbol{a}_k = [a_k^1, \ldots, a_k^P]^T$ are the AR coefficients, $\boldsymbol{G}(z_{k-P:k-1}) = [z_{k-1} \ldots z_{k-P}]$, $\boldsymbol{w}_k^1 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_k^1)$ is the process noise, and $e_k \sim \mathcal{N}(0, R_k)$ is the measurement noise. As mentioned before, the driving noise w_k^2 is distributed according to a GMM

$$p(w_k^2) = 0.8\mathcal{N}(w_k^2; -3, 1) + 0.2\mathcal{N}(w_k^2; 8, 1). \quad (45)$$

Here, we choose $\boldsymbol{Q}_k^1 = (0.01)^2 I_{P\times P}$, and $R_k = 1$. The elements of $\boldsymbol{a}_0$ and $z_{0:P-1}$ are each distributed in accordance with a Gaussian distribution with mean 0, and variance 0.5. We consider a fourth order TVAR model ($P = 4$) with known coefficient $\beta = 0.995$. It should be noted that the considered DSSM is a special case of (1) - (3). That is, $\boldsymbol{a}_k$ plays the role of $\boldsymbol{x}_k^1$, and z_k the role of x_k^2 in (1) - (3).

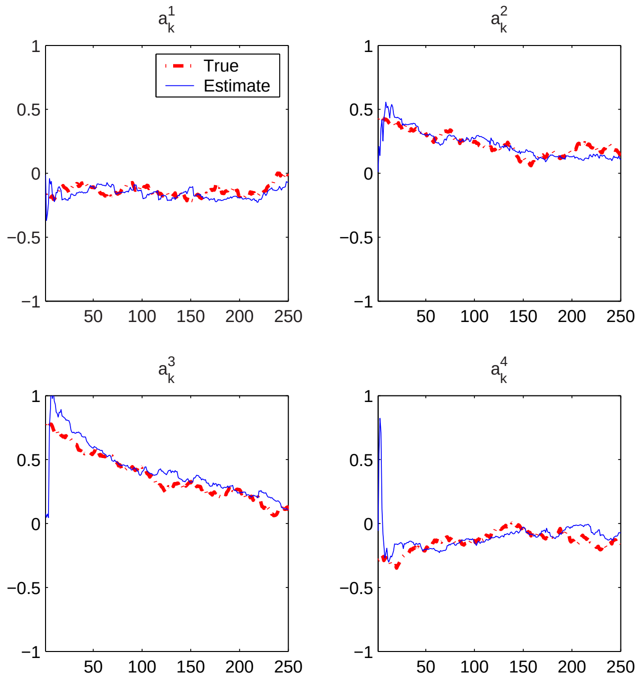

Fig. 1. True and estimated trajectory of $\boldsymbol{a}_k$ for $N = 250$ observations, $N_p = 50$ particles

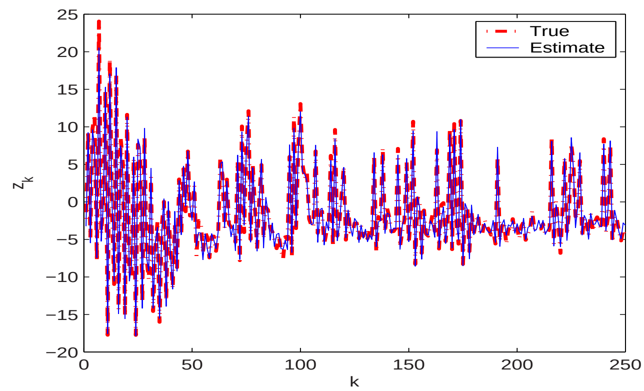

Fig. 2. True and estimated trajectory of z_k for $N = 250$ observations, $N_p = 50$ particles

Figs. 1 and 2, show a typical realization of $\boldsymbol{a}_k$ and z_k, respectively. As shown above, with only $N_p = 50$ particles and the prior

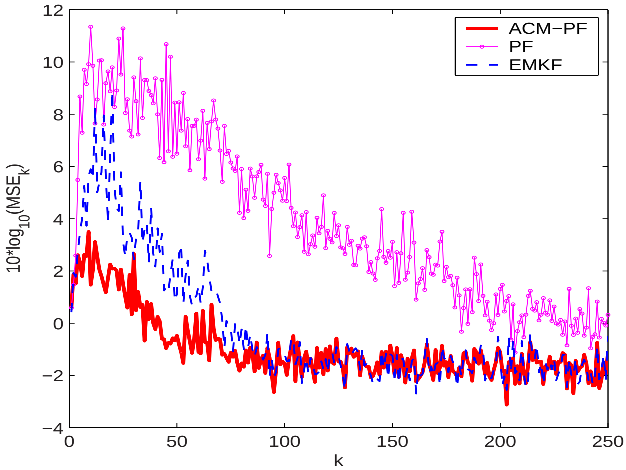

Fig. 3. MSE curves for $N_p = 50$ and $M = 200$ Monte Carlo runs

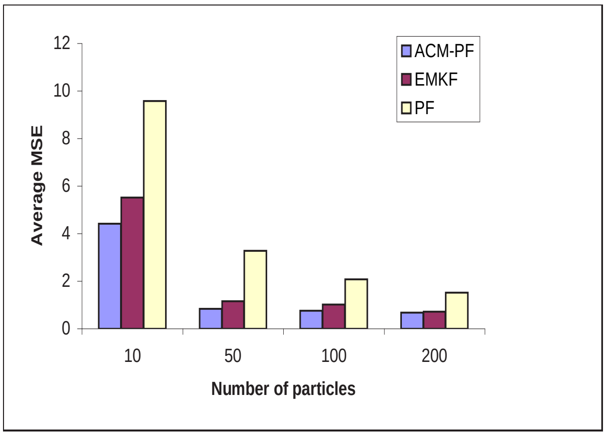

Fig. 4. Average MSE for $N = 250$ observations and $M = 200$ Monte Carlo runs

(39) for the importance function $q(\boldsymbol{x}_k^2|\boldsymbol{x}_{1:k-1}^2, \boldsymbol{y}_k)$, the ACM-PF tracks both the AR coefficients $\boldsymbol{a}_k$ and the AR process z_k remarkably well.

In the following, we show the error performances of the considered algorithms. This includes the ACM-PF, the standard PF, and the EMKF, each using the appropriate prior distribution for the importance function. To this end, we consider two metrics. The first is the mean square error (MSE) at the k-th time step, i.e.,

$$MSE_k = \frac{1}{M}\sum_{i=1}^{M}\left(\|\boldsymbol{a}_k^i - \hat{\boldsymbol{a}}_{k|k}^i\|_2^2 + \|\boldsymbol{z}_k^i - \hat{\boldsymbol{z}}_{k|k}^i\|_2^2\right) \quad (46)$$

where $\|\cdot\|_2$ is the euclidean norm, $\hat{\boldsymbol{a}}_{k|k}^i$ and $\hat{z}_{k|k}^i$ is an estimate of $\boldsymbol{a}_k^i$ and z_k^i for the i-th MC simulation. The other is the average MSE defined as

$$\overline{MSE} = \frac{1}{N}\sum_{k=1}^{N} MSE_k. \quad (47)$$

Note, in calculating (46) and hence (47) , we ran each filter on the same realizations of data and repeated the experiment $M = 200$ times.

Figure 3 depicts the MSE curves for $N = 250$ observations. Each filter was implemented with $N_p = 50$ particles. Clearly, both the ACM-PF and the EMKF significantly outperform the PF. In particular, the steady state MSE of both the ACM-PF and the EMKF are nearly identical. However, of the considered algorithms, the ACM-PF yielded the shortest acquisition time. Thus, for this experiment, the ACM-PF offers the best performance.

Finally, we calculate the average MSE for $N_p = 10, 50, 100$ and 200 particles. The results are shown in Figure 4. The ACM-PF yields much improved performance over the PF. For example, in Figure 4, the ACM-PF merely uses 50 particles to yield a average MSE of 0.7, while in the case of the PF, even 200 particles does not result in a comparable MSE. As compared to the EMKF, the ACM-PF provides better performance. However, one must be careful in making any general conclusions. Indeed, for the majority of the simulations, we have found that the EMKF and the ACM-PF yield comparable steady state MSE as shown in Figure 3. Thus, the performance differences between the ACM-PF and the EMKF are largely attributed to the apparently large acquisition time of the EMKF.

7. CONCLUSION

In this paper, we have proposed a novel filter for a particular class of partially observed non-Gaussian DSSM's. The proposed method called the ACM-PF is a efficient combination of the ACM filter and the PF. The performance of this approach has been demonstrated by simulations.

8. REFERENCES

[1] C. Andrieu and A. Doucet, "Particle filtering for partially observed Gaussian state space models," *J. Royal Statist. Soc. B*, vol. 64, pp. 827–836, 2002.

[2] A. Doucet, N. de Freitas, and N. J. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

[3] J. S. Liu and R. Chen, "Mixture Kalman filters," *J. Roy. Statistic. Soc. B*, vol. 62, pp. 493–508, 2000.

[4] C. J. Masreliez, "Approximate non-Gaussian filtering with linear state and observation relations," *IEEE Trans. Automat. Contr.*, vol. 20, no. 1, pp. 107–110, Jan. 1975.

[5] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.

[6] R. Vijayan and H. V. Poor, "Nonlinear techniques for interference suppression in spread-spectrum systems," *IEEE Trans. Commun.*, vol. 38, no. 7, pp. 1060–1065, 1990.

[7] X. Wang and H. V. Poor, "Joint channel estimation and symbol detection in Rayleigh flat-fading channels with impulsive noise," *IEEE Commun. Lett.*, vol. 1, no. 1, pp. 19–21, Jan. 1997.