STATE AND IMPULSIVE TIME-VARYING MEASUREMENT NOISE DENSITY ESTIMATION IN NONLINEAR DYNAMIC SYSTEMS USING DIRICHLET PROCESS MIXTURES

Nouha Jaoua[1,2], *François Septier*[1,3], *Emmanuel Duflos*[1,2], *Philippe Vanheeghe*[1,2]

[1] LAGIS UMR CNRS 8219, 59651 Villeneuve d'Ascq, France.
[2] Ecole Centrale de Lille, 59651 Villeneuve d'Ascq, France.
[3] Institut Mines-Télécom / Télécom Lille 1, 59658 Villeneuve d'Ascq, France.

ABSTRACT

In this paper, we focus on the challenging task of the online estimation of the state and the unknown measurement noise density in nonlinear dynamic state-space models. We are especially interested in making inference in the presence of impulsive and time-varying noise. A flexible Bayesian nonparametric noise model based on an extension of Dirichlet Process Mixtures, namely the Time Varying Dirichlet process Mixtures, is introduced. A novel method based on sequential Monte Carlo methods is proposed to perform the optimal online estimation. Simulation results demonstrate the efficiency and the robustness of this method.

***Index Terms*—** Bayesian nonparametric, Time-Varying Dirichlet Process Mixture, impulsive noise, α-stable process, particle filter

1. INTRODUCTION

In signal processing literature, noise sources are often assumed to be Gaussian. However, in many fields the conventional Gaussian noise assumption is inadequate and can lead to the loss of resolution and/or accuracy. This is particularly the case of noise that exhibits multimodality and impulsivity. For example, the latter is found in various areas [1–4]. In fact, impulsive noise tends to produce large amplitude excursions from the average value more frequently than Gaussian signals. It contains sharp spikes or occasional bursts. As a result, its probability density function (pdf) decays in the tails less rapidly than Gaussian pdf [5]. Moreover, in most practical applications, the distribution of the errors cannot be perfectly known and may also vary through time. Therefore, it is important to have efficient estimation procedure being able to deal with unknown measurement noise density. In this paper, we address the problem of the optimal state estimation when the time-varying probability density function of the measurement noise sequence is unknown and need to be estimated on-line from the data.

Several algorithms have been developed to estimate noise statistics in dynamical systems. Most of them are based on a given prior parametric shape of the unknown density, i.e. student-t, α-stable [6] or finite mixture of Gaussians [7]. Thus, the algorithm consists in jointly estimating the hidden state and the (possibly time-varying) parameters associated to the parametric shape of the unknown density. However, the main difficulty is the choice of the parametric shape and/or the number of component if a mixture is considered. Indeed, it is theoretically desirable to consider models that are not limited to finite parametrizations. This can be overcome by addressing the uncertainty about the parametric form of the unknown density with a nonparametric prior. The Dirichlet process (DP) is one of the most prominent random probability measures due to its richness, computational ease, and interpretability. It can be used to model the uncertainty about the functional form of the distribution for parameters in a model. The hierarchical models in which the DP is used as a prior over the distribution of the parameters are referred to as the Dirichlet process mixture (DPM) models. DPM can easily be defined as an extension of a parametric mixture model without the need to do model selection for determining the number of components to be used.

Several approaches for filtering with nonparametric density estimation using DPMs have been recently proposed [8–10]. In [8], the authors derive a sequential Monte-Carlo (SMC) algorithm in order to jointly estimate the hidden state and the unknown (static) density in linear dynamical systems by assuming the unknown density to be an infinite mixtures of Gaussians. In [9], the authors extend the algorithm to nonlinear dynamical systems. Finally in [10], we propose to use a DPM of Cauchy distributions, unlike the Gaussian choice in the two previous works, in order to be more robust to impulsive noise. Furthermore, we propose a novel sampling step within the SMC algorithm in order to take into account the observations in the sampling step of the parameters of the DPM. In this paper, we propose to extend this methodology in order to deal with time-varying measurement noise density.

This paper is organized as follows: in section 2 we briefly review Bayesian nonparametric density estimation using DPMs and we describe the time-varying DPM (TVDPM) model. In section 3, we introduce the dynamic model as well as the measurement noise modeling. Section 4 is devoted to the description of the proposed particle filter (PF). Simulation results are presented in section 5 and conclusions are drawn through section 6.

2. BAYESIAN NONPARAMETRIC DENSITY ESTIMATION USING TVDPMS

Consider a set of observations $\{z_t\}_{t=1}^T$ statistically distributed according to an unknown pdf F_t such as $z_t \sim F_t(.)$, $t = 1, ..., T$. We are interested in estimating the pdf $F_t(.)$ based on the sequence of observations $\{z_t\}_{t=1}^T$. To this purpose, we consider the following nonparametric model

$$F_t(.) = \int_\Theta f(.|\theta_t) d\mathbb{G}_t(\theta_t), \quad t = 1, ..., T \tag{1}$$

where $\theta_t \in \Theta$ is called the latent variable or cluster, $f(.|\theta_t)$ is the mixed pdf and $\mathbb{G}_t$ is the mixing distribution. Under a Bayesian framework, $\mathbb{G}_t$ is assumed to be a Random Probability Measure (RPM) distributed according to a prior distribution. The DPM model

Part of this work has been supported by the BNPSI ANR project no ANR-13-BS-03-0006-01.

is based on DP prior for the mixing distribution. Such a model assumes that the RPM $\mathbb{G}_t$ is distributed according to a DP

$$\mathbb{G}_t \sim DP(\mathbb{G}_0, \alpha) \tag{2}$$

where $\mathbb{G}_0$ is a base probability measure and $\alpha > 0$ is the concentration parameter. For a review of DP, see [11]. The DPM defines the following hierarchical Bayesian structure:

$$\begin{array}{l} \mathbb{G}_t|\mathbb{G}_0, \alpha \sim DP(\mathbb{G}_0, \alpha), \quad t = 1, ..., T \\ \theta_t|\mathbb{G}_t \sim \mathbb{G}_t, \quad t = 1, ..., T \\ z_t|\theta_t \sim f(.|\theta_t), \quad t = 1, ..., T \end{array} \tag{3}$$

It should be noted that, due to the discreteness property of the DP [12], there is a strictly positive probability of multiple latent variables θ_t taking identical values. Let $U_1, ..., U_r$ be the unique values or clusters among $\theta_1, ..., \theta_T$.
We can also obtain a DPM model as the limit of the following finite Dirichlet-multinomial model as $p \longmapsto \infty$ [13]:

$$\begin{array}{l} U_1, ..., U_p|\mathbb{G}_0 \sim \mathbb{G}_0, \quad t = 1, ..., T \\ \pi_1, ..., \pi_p|\alpha \sim \mathcal{D}(\frac{\alpha}{p}, ..., \frac{\alpha}{p}), \quad t = 1, ..., T \\ c_t|\pi_{1:p} \sim Multinomial(\pi_{1:p}), \quad t = 1, ..., T \\ z_t|c_t, U_{1:p} \sim f(.|U_{c_t}), \quad t = 1, ..., T \end{array} \tag{4}$$

where $c_t \in \{1, ..., p\}$ is the allocation variable that picks out for each θ_t the unique value U_k such that $\theta_t = U_{c_t}$. As $p \longmapsto \infty$, the predictive distribution of the allocation variables, computed by integrating over the vector $\pi_{1:p}$, follows a Polya urn scheme [14]:

$$\left\{ \begin{array}{l} p(c_t = k|c_{1:t-1}) = \frac{m_k}{\alpha+t-1}, \quad \forall k \in X(c_{1:t-1}) \\ p(c_t \notin X(c_{1:t-1})|c_{1:t-1}) = \frac{\alpha}{\alpha+t-1} \end{array} \right. \tag{5}$$

where $X(c_{1:t-1})$ denotes the set of unique values in $c_{1:t-1}$ and $m_k(c_{1:t-1}) = \sum_{j=1}^{t-1} \delta_{c_j}(k)$ is the number of allocation variables in $c_{1:t-1}$ taking the value k.

The DPM is a widely used model for density estimation and is among the most successful ways of modeling multimodal distributions in a nonparametric Bayesian framework. However, when the available data have a time-varying distribution, such model becomes inadequate. In this paper, we will use a class of TVDPMs, proposed by Caron *et al* [15], which ensures that at each time step the RPM $\mathbb{G}_t$ follows a DPM model. This model allows us to move both the cluster parameters and their weights. Furthermore, it relies on a simple and intuitive birth/death procedure. The main idea behind this model consists at each time step t to delete randomly a subset of the allocations variables sampled at time $k < t$ and which had survived the previous $t-1$ deletion steps. In fact, each allocation variable is deleted with probability $1-\rho$, where $0 \leq \rho \leq 1$. Hence, an allocation variable survives according to a Bernoulli distribution and the sum $m_{k,t}$ is distributed according to a Binomial distribution:

$$m_{k,t} \sim \mathcal{B}(m_{k,t-1} + \delta_{c_{t-1}}(k), \rho) \forall k \in X(\mathbf{m}_{t-1}) \cup \{c_{t-1}\} \tag{6}$$

where $\mathbf{m}_{t-1} = \{m_{k,t-1}|k \in \mathbb{N}\}$.
After the deletion step, the new allocation variable c_t is sampled according to a standard Polya urn scheme based on the surviving allocation variables $p(c_t|\mathbf{m}_t)$:

$$\left\{ \begin{array}{l} p(c_t = k|\mathbf{m}_t) = \frac{m_{k,t}}{\alpha+\sum_{j\in X(\mathbf{m}_t)} m_{j,t}}, \quad \forall k \in X(\mathbf{m}_t) \\ p(c_t \notin X(\mathbf{m}_t)|\mathbf{m}_t) = \frac{\alpha}{\alpha+\sum_{j\in X(\mathbf{m}_t)} m_{j,t}} \end{array} \right. \tag{7}$$

To obtain a first order stationary DPM process, we also need to ensure that at each time t the clusters $U_{k,t}$, $\forall k \in X(\mathbf{m}_t)$ are i.i.d from the base distribution $\mathbb{G}_0$. This can be easily achieved if $\forall k \in X(\mathbf{m}_t) \cup \{c_t\}$

$$U_{k,t} \sim \left\{ \begin{array}{ll} p(U_{k,t}|U_{k,t-1}) & \text{if} \quad k \in X(\mathbf{m}_t) \\ \mathbb{G}_0 & \text{otherwise} \end{array} \right. \tag{8}$$

where $\mathbb{G}_0$ is the invariant distribution of the transition kernel

$$\int \mathbb{G}_0(U_{k,t-1}) p(U_{k,t}|U_{k,t-1}) dU_{k,t-1} = \mathbb{G}_0(U_{k,t}) \tag{9}$$

Such transition kernel can be constructed using standard approaches from the time series literature [16].
Finally, we can summarize the TVDPM by the following hierarchical Bayesian structure for $t = 1, ..., T$:

$$\begin{array}{l} m_{k,t} \sim \mathcal{B}(m_{k,t-1} + \delta_{c_{t-1}}(k), \rho) \quad \forall k \in X(\mathbf{m}_{t-1}) \cup \{c_{t-1}\} \\ c_t|\mathbf{m}_t \sim p(c_t|\mathbf{m}_t) \\ U_{k,t} \sim \left\{ \begin{array}{ll} p(U_{k,t}|U_{k,t-1}) & \text{if } k \in X(\mathbf{m}_t) \\ \mathbb{G}_0 & \text{otherwise} \end{array} \right. \quad \forall k \in X(\mathbf{m}_t) \cup \{c_t\} \\ z_t|U_{c_t,t} \sim f(.|U_{c_t,t}) \end{array} \tag{10}$$

3. TVDPM NOISE MODEL

Consider the following generic nonlinear dynamic system given in state-space form:

$$\left\{ \begin{array}{l} x_{t+1} = g_t(x_t, w_t) \\ y_t = h_t(x_t, v_t) \end{array} \right. \tag{11}$$

where t is the time index, x_t is the state variable, y_t is the measurement, g_t and h_t are respectively the state and the observation functions, and w_t and v_t are mutually independent i.i.d noise processes. Such nonlinear dynamic systems are widely used to model systems across many areas in signal processing such as target tracking, communications, etc. Here, we assume that the distribution of the process noise is known. The measurement noise is assumed to be impulsive, skewed, multimodal and time-varying with an unknown distribution $v_t \sim F_t$.

In order to introduce temporal dependencies between the distributions F_t and F_{t-1}, we assume that the measurement noise v_t is distributed according to the TVDPM model defined by (10) with a heavy-tailed kernel that is the Cauchy distribution. The mixed pdf $f(.|U_t)$ is thus assumed to be a Cauchy distribution with location parameter m_t and scale parameter e_t denoted $\mathcal{C}(m_t, e_t)$. We denote $U_t = \{m_t, e_t\}$ the cluster giving at each time index t the location and the scale of the mixed pdf. The base distribution $\mathbb{G}_0$ is assumed to be a normal inverse Wishart distribution $\mathcal{NIW}(\mu_0, \kappa_0, \nu_0, \Lambda_0)$. Instead of fixing ρ, we assume that it is time-varying with $p(\rho_t|\rho_{t-1}) = Beta(a_\rho, a_\rho \frac{1-\rho_{t-1}}{\rho_{t-1}})$, where $a_\rho \in]0, \infty]$. We denote $\Phi = \{\alpha, \mu_0, \kappa_0, \nu_0, \Lambda_0, a_\rho\}$ the set of hyperparameters which are assumed to be pre-specified and fixed.

4. PARTICLE FILTER FOR SEQUENTIAL STATE AND NOISE DENSITY ESTIMATION

In this paper, our main goal is to jointly estimate the state x_t as well as the time-varying measurement noise distribution F_t at each time t conditional on the observations $y_{1:t}$. The variables of interest are the hidden state x_t, the allocation variables c_t, the vector $\mathbf{m}_t$, the clusters $U_{k\in X(\mathbf{m}_t)\cup\{c_t\},t}$ and the hyperparameter ρ_t. These variables may be written as a single vector

$\mathbf{z}_t = [x_t, c_t, \mathbf{m}_t, U_{k\in X(\mathbf{m}_t)\cup\{c_t\},t}, \rho_t]$. Within a Bayesian framework, we need to compute the joint posterior pdf $p(\mathbf{z}_t|y_{1:t}, \Phi)$. This pdf is analytically intractable. Therefore, we propose to use SMC methods in order to find an estimate of the required posterior pdf. The set of hyperparameters Φ is assumed to be known, therefore it is omitted in the following. The posterior pdf $p(\mathbf{z}_t|y_{1:t})$ will be approximated by a PF:

$$p(\mathbf{z}_t|y_{1:t}) \simeq \sum_{i=1}^{N} \omega_t^{(i)} \delta_{\mathbf{z}_t^{(i)}}(\mathbf{z}_t) \tag{12}$$

where δ is the Dirac delta function, $\mathbf{z}_t^{(i)}$ is the vector of the different variables of interest particles drawn from the importance density $q(\mathbf{z}_t|\mathbf{z}_{0:t-1}, y_{1:t})$ and $\omega_t^{(i)}$ is the normalized importance weight asociated to the ith particle.

Once the posterior density function of interest is identified, the remaining task is the simulation of the differents particles from the importance density. The choice of the importance density is crucial because it determines the efficiency as well as the complexity of the PF. The considered importance density can be decomposed as:

$$\begin{aligned} q(\mathbf{z}_t|\mathbf{z}_{0:t-1}, y_{1:t}) &= q(x_t, c_t, U_{k=c_t,t}|x_{0:t-1}, \mathbf{m}_{1:t}, U_{1:t-1}, y_{1:t}) \\ &\times q(U_{k\in X(\mathbf{m}_t),t}|U_{1:t-1}, \mathbf{m}_{1:t}) q(\rho_t|\rho_{1:t-1}, \mathbf{m}_{1:t}) \\ &\times q(\mathbf{m}_t|\mathbf{m}_{1:t-1}, c_{1:t-1}, \rho_{1:t-1}) \end{aligned} \tag{13}$$

The particles of the vector $\mathbf{m}_t$ are sampled using the transition density $p(\mathbf{m}_t|\mathbf{m}_{t-1}, c_{t-1}, \rho_{t-1})$:

$$\widetilde{m}_{k,t}^{(i)} \sim \mathcal{B}(m_{k,t-1}^{(i)} + \delta_{c_{t-1}^{(i)}}(k), \rho_{t-1}^{(i)}), \forall k \in X(\mathbf{m}_{t-1}^{(i)}) \cup \{c_{t-1}^{(i)}\} \tag{14}$$

For the hyperparameter ρ_t, we use the optimal importance density $p(\rho_t|\rho_{t-1}, N_t, N_{t-1})$ where $N_t = \sum_{k\in X(\mathbf{m}_t)} m_{k,t}$. We have

$$N_t \sim \mathcal{B}(N_{t-1} + 1, \rho_t) \tag{15}$$

Since the beta distribution is a conjugate prior for a binomial likelihood, the optimal importance density is given:

$$\begin{aligned} &p(\rho_t|\rho_{t-1}, N_t, N_{t-1}) \\ &\quad \propto Beta(\rho_t; a_\rho, a_\rho \tfrac{1-\rho_{t-1}}{\rho_{t-1}}) \mathcal{B}(N_{t-1} + 1, \rho_t) \\ &\quad = Beta(\rho_t; a_\rho + N_t, a_\rho \tfrac{1-\rho_{t-1}}{\rho_{t-1}} + N_{t-1} + 1 - N_t) \end{aligned} \tag{16}$$

We note that the current observation y_t provides only information on the cluster $U_{k=c_t,t}$. This is why, the set of surviving clusters such as $k \neq c_t$ are sampled using the transition density $p(U_{k,t}|U_{k,t-1})$. However, for the cluster $U_{k=c_t,t}$ as well as the state x_t and the allocation variable c_t, we consider the optimal importance density in the sense of minimizing the variance of the importance weights [17]. In our context, it is expressed as

$$\begin{aligned} &q(x_t, c_t, U_{k,t}|x_{0:t-1}, \mathbf{m}_{1:t}, U_{1:t-1}, y_{1:t}) \\ &\quad = p(x_t, c_t, U_{k,t}|x_{0:t-1}, \mathbf{m}_t, U_{1:t-1}, y_t) \end{aligned} \tag{17}$$

This importance density is interesting because it incorporates information on the current observation. Consequently, the particles tend to cluster in regions of high probability mass of the posterior pdf. The sampling of $\widetilde{x}_t^{(i)}$, $\widetilde{c}_t^{(i)}$ and $\widetilde{U}_{k,t}^{(i)}$ from (17) requires the analytical expression of the optimal importance density. However, this pdf is analytically intractable. Using Bayes'theorem, the considered importance density can be written as

$$\begin{aligned} &p(x_t, c_t, U_{k,t}|x_{0:t-1}, \mathbf{m}_t, U_{1:t-1}, y_t) \\ &= \frac{p(y_t|x_t,c_t,U_{k,t})p(x_t|c_t,U_{k,t},x_{t-1})p(U_{k,t}|U_{k,t-1},c_t)p(c_t|\mathbf{m}_t)}{p(y_t|x_{0:t-1}, \mathbf{m}_t, U_{1:t-1})} \\ &\propto p(y_t|x_t,c_t,U_{k,t})p(x_t|c_t,U_{k,t},x_{t-1})p(U_{k,t}|U_{k,t-1},c_t)p(c_t|\mathbf{m}_t) \end{aligned} \tag{18}$$

Thus, an approximation of the optimal importance density can be obtained using Monte Carlo method and importance sampling. For this purpose, we consider a set of N_{IS} auxiliary particles $\left\{(\breve{x}_{t,i}^{(j)}, \breve{c}_{t,i}^{(j)}, \breve{U}_{k,t,i}^{(j)})\right\}_{j=1}^{N_{IS}}$ where state particles $\breve{x}_{t,i}^{(j)}$ are sampled using the transition density $p(x_t|x_{0:t-1}^{(i)})$:

$$\breve{x}_{t,i}^{(j)} \sim p(x_t|x_{t-1}^{(i)}) \tag{19}$$

and the allocation variable particles $\breve{c}_{t,i}^{(j)}$ are sampled using a standard Polya urn scheme based on the surviving allocation variables

$$\begin{cases} p(\breve{c}_{t,i}^{(j)} = k|\widetilde{\mathbf{m}}_t^{(i)}) = \frac{\widetilde{m}_{k,t}^{(i)}}{\alpha + \sum_{l\in X(\widetilde{\mathbf{m}}_t^{(i)})} \widetilde{m}_{l,t}^{(i)}}, \quad \forall k \in X(\widetilde{\mathbf{m}}_t^{(i)}) \\ p(\breve{c}_{t,i}^{(j)} \notin X(\widetilde{\mathbf{m}}_t^{(i)})|\widetilde{\mathbf{m}}_t^{(i)}) = \frac{\alpha}{\alpha + \sum_{l\in X(\widetilde{\mathbf{m}}_t^{(i)})} \widetilde{m}_{l,t}^{(i)}} \end{cases} \tag{20}$$

The cluster particles $\breve{U}_{k,t,i}^{(j)}$ are sampled as follows

$$\begin{cases} p(U_{k,t}|U_{k,t-1}), \quad \text{if} \quad \breve{c}_{t,i}^{(j)} \in X(\widetilde{m}_t^{(i)}) \\ \mathbb{G}_0, \quad \text{otherwise} \end{cases} \tag{21}$$

Using this set of particles, the optimal importance density can be approximated by the following empirical distribution:

$$\begin{aligned} &p(x_t, c_t, U_{k,t}|x_{0:t-1}^{(i)}, \widetilde{\mathbf{m}}_t^{(i)}, U_{1:t-1}^{(i)}, y_t) \\ &\quad \simeq \sum_{j=1}^{N_{IS}} \frac{\breve{\omega}_{t,i}^{(j)}}{S_{\breve{\omega}}} \delta_{\breve{x}_{t,i}^{(j)}, \breve{c}_{t,i}^{(j)}, \breve{U}_{k,t,i}^{(j)}}(x_t, c_t, U_{k,t}) \end{aligned} \tag{22}$$

where $\breve{\omega}_{t,i}^{(j)}$ is the unnormalized weight associated to the jth group of particles $(\breve{x}_{t,i}^{(j)}, \breve{c}_{t,i}^{(j)}, \breve{U}_{k,t,i}^{(j)})$ defined as

$$\breve{\omega}_{t,i}^{(j)} = p(y_t|\breve{x}_{t,i}^{(j)}, \breve{c}_{t,i}^{(j)}, \breve{U}_{k,t,i}^{(j)}) \tag{23}$$

and $S_{\breve{\omega}}$ is the sum of unnormalized weights $S_{\breve{\omega}} = \sum_{j=1}^{N_{IS}} \breve{\omega}_{t,i}^{(j)}$. In order to sample the ith group of particles $(\widetilde{x}_t^{(i)}, \widetilde{c}_t^{(i)}, \widetilde{U}_{k,t}^{(i)})$ from the approximate optimal importance density given by (22), we just need to pick one particles group from the set $\left\{(\breve{x}_{t,i}^{(j)}, \breve{c}_{t,i}^{(j)}, \breve{U}_{k,t,i}^{(j)})\right\}_{j=1}^{N_{IS}}$ using weights $\left\{\breve{\omega}_{t,i}^{(j)}\right\}_{j=1}^{N_{IS}}$ as probabilities of selection. This can be done as follows:

$$J \sim Multinomial\left(\frac{\breve{\omega}_{t,i}^{(1)}}{S_{\breve{\omega}}}, \frac{\breve{\omega}_{t,i}^{(2)}}{S_{\breve{\omega}}}, ..., \frac{\breve{\omega}_{t,i}^{(N_{IS})}}{S_{\breve{\omega}}}\right) \tag{24}$$

Thus, the ith group of particles $(\widetilde{x}_t^{(i)}, \widetilde{c}_t^{(i)}, \widetilde{U}_{k,t}^{(i)})$ is given by

$$\left(\widetilde{x}_t^{(i)}, \widetilde{c}_t^{(i)}, \widetilde{U}_{k,t}^{(i)}\right) = \left(\breve{x}_{t,i}^{(j=J)}, \breve{c}_{t,i}^{(j=J)}, \breve{U}_{k,t,i}^{(j=J)}\right) \tag{25}$$

Using these importance densities, the importance weights of the main PF are updated according to the following relation:

$$\omega_t^{(i)} \propto \omega_{t-1}^{(i)} \frac{p(y_t|x_{0:t-1}^{(i)}, \widetilde{\mathbf{m}}_t^{(i)}, U_{k,t-1}^{(i)}) p(\widetilde{\rho}_t^{(i)}|\rho_{t-1}^{(i)})}{q(\widetilde{\rho}_t^{(i)}|\rho_{t-1}^{(i)}, \widetilde{\mathbf{m}}_t^{(i)}, \mathbf{m}_{t-1}^{(i)}, c_{t-1}^{(i)})} \tag{26}$$

where $p(y_t|x_{0:t-1}^{(i)}, \widetilde{\mathbf{m}}_t^{(i)}, U_{k,t-1}^{(i)})$ can be approximated by Monte Carlo method using the weighted set of particles of the importance sampling strategy. In doing so, this pdf is given by the sum of unnormalized weights $S_{\breve{\omega}}$.

The proposed PF for joint state and time-varying noise density estimation denoted by PF-JSTVNDE is summarized in algorithm 1.

Initialization
for $t = 1$ **to** T **do**
  for $i = 1$ **to** N **do**
    $\forall k \in X(\mathbf{m}_{t-1}^{(i)}) \cup \{c_{t-1}^{(i)}\}$, sample $\widetilde{m}_{k,t}^{(i)}$ using (14) ;
    Sample $\widetilde{\rho}_t^{(i)}$ using (16);
    for $j = 1$ **to** N_{IS} **do**
      Sample $\breve{x}_{t,i}^{(j)}$ using (19), $\breve{c}_{t,i}^{(j)}$ using (20) and $\breve{U}_{k,t,i}^{(j)}$ using (21);
      Compute weights $\breve{\omega}_{t,i}^{(j)}$ using (23) ;
    end
    Compute: $S_{\breve{\omega}} = \sum_{j=1}^{N_{IS}} \breve{\omega}_{t,i}^{(j)}$;
    Normalize weights: $\breve{\omega}_{t,i}^{(j)} = \breve{\omega}_{t,i}^{(j)} / S_{\breve{\omega}}$;
    Select a particle indice $J \in \{1, 2, ..., N_{IS}\}$ according to weights $\left\{\breve{\omega}_{t,i}^{(j)}\right\}_{j=1}^{N_{IS}}$ using (24);
    Set $\widetilde{x}_t^{(i)} = \breve{x}_{t,i}^{(J)}$, $\widetilde{c}_t^{(i)} = \breve{c}_{t,i}^{(J)}$ and $\widetilde{U}_{k,t}^{(i)} = \breve{U}_{k,t,i}^{(J)}$;
    For $k \in \overline{X(\widetilde{\mathbf{m}}_t^{(i)}) \cap \widetilde{c}_t^{(i)}}$ sample $\widetilde{U}_{k,t}^{(i)} \sim p(U_{k,t} | U_{k,t-1}^{(i)})$;
    Update importance weights $\omega_t^{(i)}$ using (26) ;
  end
  Normalize importance weights $\omega_t^{(i)} = \omega_t^{(i)} / \sum_{j=1}^{N} \omega_t^{(i)}, \quad i = 1, ..., N$;
  if $N_{eff} < \frac{N}{2}$ **then** Resampling step;
end

Algorithm 1: PF-JSTVNDE algorithm

5. SIMULATIONS

The performance of the proposed method is studied considering the following benchmark scalar nonlinear time series model [18–20]:

$$\begin{cases} x_{t+1} = 0.5x_t + 25\dfrac{x_t}{1+x_t^2} + 8\cos(1.2(t+1)) + w_t \\ y_t = \frac{x_t^2}{20} + v_t \end{cases}$$

This model has been simulated with the following parameters: $x_0 \sim \mathcal{N}(0, 10)$ and $w_t \sim \mathcal{N}(0, 1)$. The measurement noise v_t is assumed to be time-varying and generated for t=1,...,1000 from a sequence of mixtures:

$$\begin{cases} v_t \sim 0.4\mathcal{S}_{1.2}(0.5, 0.7, -4) + 0.6\mathcal{S}_{1.5}(0, 0.5, 0) & t = 1, ..., 300 \\ v_t \sim \mathcal{S}_{1.5}(0, 0.5, 0) & t = 301, ..., 600 \end{cases}$$

where $S_\alpha(\beta, \gamma, \mu)$ denotes the α-stable distribution with characteristic exponent $0 < \alpha < 2$, dispersion parameter $\gamma > 0$, location parameter μ and skewness parameter $\beta \in [-1; 1]$. One difficulty is that they have no closed-form expressions for their pdf. They can be most conveniently described by their characteristic function [5]:

$$\varphi(t) = \begin{cases} \exp\left(i\mu t - \gamma^\alpha |t|^\alpha \left[1 - i\beta \operatorname{sgn}(t) \tan \frac{\alpha\pi}{2}\right]\right), & \alpha \neq 1 \\ \exp\left(i\mu t - \gamma |t|^\alpha \left[1 + i\beta \operatorname{sgn}(t) \frac{2}{\pi} \log|t|\right]\right), & \alpha = 1 \end{cases}$$

The hyperparameters of the base distribution μ_0,κ_0,ν_0 and Λ_0 are respectively set to 0, 0.01, 5 and 6. We fixed the scale parameter of the DPM α to 3 and a_ρ to 1000. The proposed PF has been implemented with $N = 200$ particles and $N_{IS} = 100$ auxiliary particles. Results are illustrated in the different plots of Fig. 1 and Fig. 2. Fig. 1 shows the estimated signal as well as the true and the observed ones. We also plot the measurement noise signal and the estimation error between the true and the estimated signals. From these plots, it can be seen that despite the fact that the noise is important, the state x_t is correctly estimated. Fig. 2 depicts the estimated measurement noise density as well as the true one at different time steps. We can observe that our algorithm is able to capture the evolution of the noise density over time.

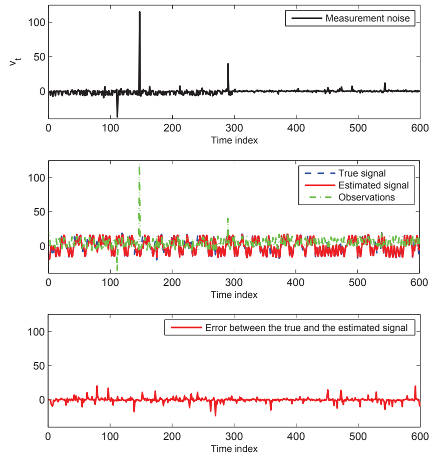

Fig. 1. Top picture: Measurement noise signal. Middle picture: True, estimated and observed signals. Bottom picture: Error between the true and the estimated states.

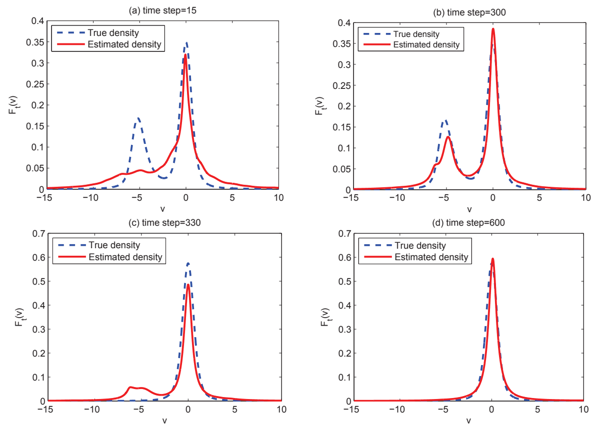

Fig. 2. True and estimated noise densities at different time steps.

6. CONCLUSIONS

In this paper, a novel method for the online estimation of the state and the time-varying measurement noise density is presented. The measurement noise considered here is impulsive and multimodal. The proposed approach relies on the introduction of a flexible Bayesian nonparametric model based on Dirichlet Process Mixture to model the measurement noise density as an infinite mixture of Cauchy distributions. A particle filter based on efficient importance densities is then implemented to perform the joint estimation of the state and the unknown noise density. The efficiency and the robustness of the proposed scheme are illustrated through several simulations. In future works, we plan to make inference on the hyperparameters of the base distribution.

7. REFERENCES

[1] M. Shinde and S. Gupta, "Signal detection in the presence of atmospheric noise in tropics," *IEEE Transactions on Communications*, vol. 22, no. 8, pp. 1055–1063, aug 1974.

[2] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-19, no. 3, pp. 106 –127, aug. 1977.

[3] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *IEEE Transactions on Electromagnetic Compatibility*, vol. 44, no. 1, pp. 249–258, feb 2002.

[4] J. Ilow and D. Hatzinakos, "Impulsive noise modeling with stable distributions in fading environments," in *8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, jun 1996, pp. 140 –143.

[5] C.L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, John Wiley and Sons, 1995.

[6] M. Lombardi and S.J. Godsill, "On-line Bayesian estimation of signals in symmetric alpha-stable noise," *IEEE Transactions on Signal Processing*, vol. 54, pp. 775–779, February 2006.

[7] C. Carter and R. Kohn, "Markov chain monte carlo in conditionally gaussian state space models," *Biometrika*, vol. 83, no. 3, pp. 589?601, 1996.

[8] F. Caron, M. Davy, A. Doucet, E. Duflos, and P. Vanheeghe, "Bayesian inference for linear dynamic models with dirichlet process mixtures," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 71 –84, jan. 2008.

[9] E. Ozkan, S. Saha, F. Gustafsson, and V. Smidl, "Non-parametric bayesian measurement noise density estimation in non-linear filtering," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, may 2011, pp. 5924 –5927.

[10] N. Jaoua, E. Duflos, P. Vanheeghe, and F. Septier, "Bayesian nonparametric state and impulsive measurement noise density estimation in nonlinear dynamic systems," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 5755–5759.

[11] T.S. Ferguson, "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, vol. 1, no. 2, pp. 209–230, 1973.

[12] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[13] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 249–265, 2000.

[14] D. Blackwell and J.B. Macqueen, "Ferguson distributions via Pólya urn schemes," *The Annals of Statistics*, vol. 1, pp. 353–355, 1973.

[15] F. Caron, M. Davy, and A. Doucet, "Generalized polya urn for time-varying dirichlet process mixtures," in *Proceedings of the Twenty-Third Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-07)*, Corvallis, Oregon, 2007, pp. 33–40, AUAI Press.

[16] M. K. Pitt, C. Chatfield, and S. G. Walker, "Constructing First Order Stationary Autoregressive Models via Latent Processes," *Scandinavian Journal of Statistics*, vol. 29, pp. 657–663, 2002.

[17] V. Zaritskii, V. Svetnik, and L. Shimelevich, "Monte carlo technique in problems of optimal data processing," *Automation and Remote Control*, vol. 12, pp. 95 –103, 1975.

[18] G. Kitagawa, "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.

[19] M. West, "Mixture models, Monte Carlo, Bayesian updating and dynamic models.," *Computing Science and Statistics*, vol. 24, pp. 325–333.

[20] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings F Radar and Signal Processing*, Apr. 1993, vol. 140, pp. 107–113.