International Journal of Control, Automation and Systems 20(8) (2022) 2594-2605
http://dx.doi.org/10.1007/s12555-021-0467-4
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

A New Variational Bayesian-based Kalman Filter with Random Measurement Delay and Non-Gaussian Noises

Chenghao Shan, Weidong Zhou*, Hanyu Shan, and Lu Liu

Abstract: To improve the estimation accuracy of the Kalman filter in the scenario of random measurement delay and non-Gaussian process and measurement noises, a new variational Bayesian (VB)-based Kalman filter is proposed in this paper. First, the state expansion method and Bernoulli random variable (BRV) are utilized to characterize random measurement delay. Second, the one-step predicted probability density function (PDF) and measurement noise vectors are modeled as Student's t (ST) distributions. Third, the likelihood function of two ST distributions is converted from a weighted sum to an exponential product to establish a hierarchical Gaussian state space model (HGSSM). Finally, the system state, BRV and intermediate random variables (IRV) are simultaneously estimated using the variational Bayesian (VB) method. Simulation experiment results indicate that the proposed filter has superior estimation performance to current filters to address the filtering problem of random measurement delay and non-Gaussian process and measurement noises.

Keywords: Kalman filter, non-Gaussian noise, random measurement delay, variational Bayesian (VB).

1. INTRODUCTION

For linear systems, the Kalman filter (KF) is the optimal estimator under the minimum mean square error criterion [1]. The KF assumes that the state-space model of signal and noise is known, uses the estimated value of the previous moment and observed value of the current moment to update the estimation of the state variable, and obtains the estimated value of the current moment. KF is the best and most efficient to solve a large part of the problem [2]. On the issue of state estimation, KF and its derivative algorithms have been widely used in the applications of navigation, control, sensor data fusion, and target tracking [3–12]. However, in actual signal transmission, due to network congestion, transmission channel limitations, and complex environmental factors, the system will generate random measurement delays (RMD). In this case, the accuracy of the traditional Kalman filter will significantly decrease or even diverge [13,14].

To address the filtering problem of random measurement delay, many improved filtering algorithms have been proposed. Wang *et al*. proposed a randomly delayed measurement Kalman filter (RDMKF), which recursively operates by combining analytical calculations with Gaussian weighted integration [15,16]. Wang *et al*. proposed a variational Bayesian (VB)-based improved Kalman filter (VBIKF) that introduced discrete Bernoulli random variables and converted the measurement likelihood function of double Gaussian distributions from a weighted sum to an exponential multiplication form. The state vector and unknown parameters are simultaneously inferred using the VB method [17]. When the noise distributions are Gaussian distributions with known parameters, these algorithms have excellent performances.

Unfortunately, in actual applications such as cooperative localization and target tracking by radar, measurement outliers may occur, which induce non-Gaussian heavy-tailed process and measurement noises (NHPMN) [18–20]. In this case, the estimation accuracy of these algorithms will be significantly reduced. Recently, some algorithms have been proposed to process non-Gaussian heavy-tailed noises, such as maximum correntropy-based filters [21–23], Huber-based filters [24,25], and Student's t-based filters [20,26–28]. However, they are not designed for linear systems with random measurement delay which considered in this paper.

Manuscript received June 5, 2021; revised October 1, 2021; accepted October 18, 2021. Recommended by Associate Editor Le Van Hien under the direction of Editor Jay H. Lee. This work was supported by the China Scholarship Council (CSC, No. 202006680080), and the National Natural Science Foundation of China (61573113).

Chenghao Shan is with the Department of Intelligent Systems Science and Engineering, Harbin Engineering University, and the Department of Electrical and Computer Engineering, National University of Singapore, Singapore (e-mail: shanchenghao123@hrbeu.edu.cn). Weidong Zhou is with the Department of Intelligent Systems Science and Engineering, Harbin Engineering University, China (e-mail: zhouweidong@hrbeu.edu.cn). Hanyu Shan is with the Department of Information and Communication Engineering, Harbin Engineering University, China (e-mail: shy@hrbeu.edu.cn). Lu Liu is with the Beijing Institute of Control and Electronic Technology, China (e-mail: douya@hrbeu.edu.cn).
* Corresponding author.

In this paper, a new VB-based KF is proposed to address random measurement delay and NHPMN. The state expansion method and BRV are utilized to characterize random measurement delay. The one-step predicted PDF and measurement noise vectors are assumed to be Student's t (ST) distributions. The system state, BRV and IRVs are simultaneously estimated using the VB method. The target tracking simulation results illustrate the superiority of the proposed filter. The contributions of this paper are as follows:

1) ST distributions are utilized to model the one-step predicted PDF and non-Gaussian heavy-tailed measurement noise.
2) The likelihood function of two ST distributions is converted from a weighted sum to an exponential product, and the VB method can be used directly.
3) By introducing the state expansion method and Bernoulli random variable, a new hierarchical Gaussian state space model is derived.
4) The system state and unknown variables are simultaneously inferred by introducing a variational Bayesian approach.
5) Target tracking simulation results indicate that the proposed filter has higher estimation accuracy than existing algorithms in the scenarios of random measurement delay and NHPMN. The proposed filter is also more robust than current algorithms under different delay probabilities.

The remainder of the paper is organized as follows: The problem formulation is given in Section 2. The construction of the hierarchical Gaussian state-space model and VB approximations of the joint posterior PDFs are illustrated in Section 3. Simulation results are analyzed in Section 4. Conclusions are drawn in Section 5.

2. PROBLEM FORMULATION

Consider the following linear state-space model with random measurement delay and NHPMN

$$x_t = \mathbf{M}_{t-1}x_{t-1} + \omega_{t-1}, \tag{1}$$
$$y_t = \mathbf{H}_t x_t + v_t, \tag{2}$$
$$y_t^r = (1-\tau_t)y_t + \tau_t y_{t-1},\ t \geq 2,\ y_1^r = y_1, \tag{3}$$

where $x_t \in \mathbb{R}^n$ is the system state vector; $\mathbf{M}_t \in \mathbb{R}^{n\times n}$ is the state transform matrix; $y_t \in \mathbb{R}^m$ is the idealized measurement vector without random delay; $\mathbf{H}_t \in \mathbb{R}^{m\times m}$ is the measurement matrix. ω_{t-1} and v_t are the zero-mean non-Gaussian white noises with the heavy-tailed form. The nominal covariance matrix of process noise ω_{t-1} and measurement noise v_t are defined as $\mathbf{Q}_t$ and $\mathbf{R}_t$, respectively. $y_t^r \in \mathbb{R}^m$ is the actual measurement vector with random delay, and t is the discrete time. The BRV $\tau_t \in \mathbb{R}^m$ is used to capture the random measurement delay, and $p(\tau_t = 1)$ and $p(\tau_t = 0)$ are defined as follows:

$$p(\tau_t = 1) = \mathrm{E}_p[\tau_t] = \phi_t, \tag{4}$$
$$p(\tau_t = 0) = 1 - \mathrm{E}_p[\tau_t] = 1 - \phi_t, \tag{5}$$

where $\phi_t \in [0,\ 1]$ is the fixed probability of random measurement delay, and $\mathrm{E}_p[\cdot]$ is the expectation operation. Additionally, parameters x_t, τ_t, ω_{t-1} and v_t are mutually independent in this paper.

3. MAIN RESULTS

3.1. Construction of the hierarchical Gaussian state space model (HGSSM)

By converting the form of the measurement likelihood PDF and selecting prior PDFs, a new HGSSM will be constructed, and the VB method can be used directly.

3.1.1 Conversion of the likelihood PDF

According to (3)-(5), the following measurement likelihood probability density function (PDF) can be obtained

$$\begin{aligned} p(y_t^r|x_t,x_{t-1}) &= \sum_{\tau_t=0}^{1} p(y_t^r,\tau_t|x_t,x_{t-1}) \\ &= p(\tau_t=0)p(y_t^r|x_t,x_{t-1},\tau_t=0) \\ &\quad + p(\tau_t=1)p(y_t^r|x_t,x_{t-1},\tau_t=1) \\ &= (1-\phi_t)p(y_t^r|x_t,x_{t-1},\tau_t=0) \\ &\quad + \phi_t p(y_t^r|x_t,x_{t-1},\tau_t=1). \end{aligned} \tag{6}$$

Based on (2)-(3), the likelihood PDF can be rewritten in the following form

$$p(y_t^r|x_t,x_{t-1},\tau_t=0) = p_{v_t^r}(y_t^r - \mathbf{H}_t x_t), \tag{7}$$
$$p(y_t^r|x_t,x_{t-1},\tau_t=1) = p_{v_t^r}(y_t^r - \mathbf{H}_{t-1} x_{t-1}), \tag{8}$$

where $p_{v_t^r}(\cdot)$ is the measurement noise PDF.

Bringing (7)-(8) into (6), the following equation is derived

$$\begin{aligned} p(y_t^r|x_t,x_{t-1}) &= \sum_{\tau_t=0}^{1} p(y_t^r,\tau_t|x_t,x_{t-1}) \\ &= (1-\phi_t)p_{v_t^r}(y_t^r - \mathbf{H}_t x_t) \\ &\quad + \phi_t p_{v_t^r}(y_t^r - \mathbf{H}_{t-1} x_{t-1}). \end{aligned} \tag{9}$$

Remark 1: The measurement conditional likelihood PDF in (7) has nonclosed and nonconjugate properties, and the variational inference cannot be directly used. To solve this problem, the probability mass function (PMF) of BRV τ_t is introduced to convert the form of (9) from a weighted sum to an exponential product.

According to (4)-(5), the PMF of BRV is written as

$$p(\tau_t) = (1-\phi_t)^{(1-\tau_t)}\phi_t^{\tau_t}. \tag{10}$$

Based on (7) and (8), the conditional PDF in (9) is

$$\begin{aligned} p(y_t^r|x_t,x_{t-1}) &= \sum_{\tau_t=0}^{1} p(y_t^r,\tau_t|x_t,x_{t-1}) \\ &= \sum_{\tau_t=0}^{1} p(\tau_t) p_{v_t^r}(y_t^r - \mathbf{H}_t x_t)^{(1-\tau_t)} \\ &\quad \times p_{v_t^r}(y_t^r - \mathbf{H}_{t-1} x_{t-1})^{\tau_t}. \end{aligned} \tag{11}$$

Using (11), the conditional likelihood PDF of measurement is formulated as follows:

$$\begin{aligned} p(y_t^r|x_t,x_{t-1},\tau_t) = p_{v_t^r}(y_t^r - \mathbf{H}_t x_t)^{(1-\tau_t)} \\ \times p_{v_t^r}(y_t^r - \mathbf{H}_{t-1} x_{t-1})^{\tau_t}. \end{aligned} \tag{12}$$

3.1.2 Selection of prior PDFs

To more reasonably model the heavy-tailed process noise and improve the estimation accuracy, the one-step predicted PDF can be modeled as the following ST distribution [26]

$$p(x_t^e|y_{1:t-1}^r) = \mathrm{ST}(x_t^e;\hat{x}_t^e,\mathbf{\Sigma}_t,\xi), \tag{13}$$

where ST$(\cdot;\,\mu,\,\epsilon,\,\delta)$ is ST PDF, and μ, ϵ and δ are the mean vector, scale matrix and degrees of freedom (DoF) parameter, respectively. x_t^e is the expanded state vector, and $\hat{x}_t^e$ is the one-step predicted estimation value of the expanded state vector. In this paper, the predicted error covariance matrix $\mathbf{P}_{t|t-1}^{ee}$ is assumed to be the scale matrix $\mathbf{\Sigma}_t$ in (13). The formulas of x_t^e, $\hat{x}_t^e$ and $\mathbf{\Sigma}_t$ are given as follows:

$$x_t^e = \begin{bmatrix} x_t^T & x_{t-1}^T \end{bmatrix}^T, \tag{14}$$

$$\hat{x}_{t|t-1}^e = \begin{bmatrix} \hat{x}_{t|t-1} \\ \hat{x}_{t-1|t-1} \end{bmatrix}, \tag{15}$$

$$\mathbf{\Sigma}_t = \begin{bmatrix} \mathbf{P}_{t|t-1} & \mathbf{P}_{t-1,t|t-1}^T \\ \mathbf{P}_{t-1,t|t-1} & \mathbf{P}_{t-1|t-1} \end{bmatrix} = \mathbf{P}_{t|t-1}^{ee}, \tag{16}$$

where $\hat{x}_{t-1|t-1}$ is the one-step predictive estimation value of the state vector, $\mathbf{P}_{t|t-1}$ is the predicted error covariance, and $\mathbf{P}_{t-1,t|t-1}$ is the mutual covariance matrix. They can be calculated by the time update of the standard Kalman filter algorithm, i.e.,

$$\hat{x}_{t|t-1} = \mathbf{M}_{t-1} x_{t-1|t-1}, \tag{17}$$

$$\mathbf{P}_{t|t-1} = \mathbf{M}_{t-1}\mathbf{P}_{t-1|t-1}\mathbf{M}_{t-1}^T + \mathbf{Q}_{t-1}, \tag{18}$$

$$\mathbf{P}_{t-1,t|t-1} = \mathbf{P}_{t-1|t-1}\mathbf{M}_{t-1}^T. \tag{19}$$

Remark 2: The estimated values of $\mathbf{P}_{t|t-1}^{ee}$ and $\mathbf{\Sigma}_t$ are affected by the outliers of process noise. To improve the performance of the proposed filter, the conjugate prior PDF of $\mathbf{\Sigma}_t$ must be defined. In this paper, the inverse-Wishart (IW) distribution is utilized to model the conjugate prior PDF of $\mathbf{\Sigma}_t$, which is given as

$$p(\mathbf{\Sigma}_t) = \mathrm{IW}(\mathbf{\Sigma}_t;\,u_t,\,\boldsymbol{U}_t), \tag{20}$$

where IW$(\mathbf{A};\,b,\,d)$ is the inverse-Wishart PDF of random matrix $\mathbf{A}$, b is the DoF parameter, and d is the inverse scale matrix. The detailed definition of the IW distribution is

$$\mathrm{IW}(\mathbf{A};b,d) = \frac{|d|^{b/2}|\mathbf{A}|^{-(1+b+j)/2}\exp[-0.5\mathrm{tr}(dA^{-1})]}{2^{jb/2}\Gamma(b/2)}, \tag{21}$$

where j is the dimension of random matrix $\mathbf{A}$, tr$(\cdot)$ represents the trace operation, and $\Gamma(\cdot)$ is the gamma function [29].

To capture the prior properties of $\mathbf{\Sigma}_t$, the mean value of $\mathbf{\Sigma}_t$ is assumed to be the nominal predicted error covariance matrix $\mathbf{P}_{t|t-1}^{ee}$, which can be formulated as follows:

$$\frac{\boldsymbol{U}_t}{u_t - n - 1} = \mathbf{P}_{t|t-1}^{ee}, \tag{22}$$

where u_t is defined as follows with tuning parameter $\gamma \geq 0$, i.e.,

$$u_t = n + \gamma + 1. \tag{23}$$

Using (21)-(22), we have

$$\boldsymbol{U}_t = \gamma \mathbf{P}_{t|t-1}^{ee}. \tag{24}$$

The one-step predicted PDF in (13) can be rewritten as the following hierarchical form

$$p(x_t^e|y_{1:t-1}^r,\beta_t) = \mathrm{N}\left(x_t^e;\hat{x}_{t|t-1}^e,\mathbf{\Sigma}_t/\beta_t\right), \tag{25}$$

$$p(\beta_t) = \mathrm{G}(\beta_t;\xi/2,\xi/2), \tag{26}$$

where G$(\cdot;\,\pi,\,\sigma)$ represents the Gamma PDF, π is the shape parameter, σ is the rate parameter, and β_t is the intermediate random variable (IRV).

In terms of measurement noise processing, the non-Gaussian heavy-tailed measurement noise vectors are assumed to be the following ST distributions

$$p(v_t) = \mathrm{ST}(v_t;0,\mathbf{R}_t,h_t), \tag{27}$$

$$p(v_{t-1}) = \mathrm{ST}(v_{t-1};0,\mathbf{R}_{t-1},h_{t-1}), \tag{28}$$

where $p(v_t)$ and $p(v_{t-1})$ denote the PDF of the current step and last step, respectively. Then, (27) and (28) can be further derived as the following Gaussian-double-Gamma hierarchical form

$$p(v_t) = \iint \mathrm{N}(v_t;0,R_t/\lambda_t)\,p(\lambda_t)\,p(h_t)\,d\lambda_t dh_t, \tag{29}$$

$$p(\lambda_t) = \mathrm{G}(\lambda_t;h_t/2,h_t/2), \tag{30}$$

$$p(h_t) = \mathrm{G}(h_t,e_t,g_t), \tag{31}$$

$$\begin{aligned} p(v_{t-1}) = \iint \mathrm{N}(v_{t-1};0,R_{t-1}/\lambda_{t-1})\,p(\lambda_{t-1}) \\ \times p(h_{t-1})\,d\lambda_{t-1}dh_{t-1}, \end{aligned} \tag{32}$$

$$p(\lambda_{t-1}) = \mathrm{G}(\lambda_{t-1};h_{t-1}/2,h_{t-1}/2), \tag{33}$$

$$p(h_{t-1}) = \mathrm{G}(h_{t-1}, e_{t-1}, g_{t-1}), \tag{34}$$

where λ_t, λ_{t-1}, h_t and h_{t-1} represent the IRVs.

Based on (12) and (27)-(34), the following likelihood PDF can be obtained

$$p(y_t^r|x_t^e, \lambda_t, \lambda_{t-1}, \tau_t) = \mathrm{N}(y_t^r; \mathbf{H}_t x_t, \mathbf{R}_t/\lambda_t)^{(1-\tau_t)} \times \mathrm{N}(y_t^r; \mathbf{H}_{t-1} x_{t-1}, \mathbf{R}_{t-1}/\lambda_{t-1})^{\tau_t}. \tag{35}$$

Obviously, from (35), the expanded state vector x_t^e, BRV τ_t and IRVs λ_t, λ_{t-1}, h_t and h_{t-1} affect the measurement vector y_t^r. The following prior PDF with must be calculated

$$\begin{aligned} &p(x_t^e, \mathbf{\Sigma}_t, \beta_t, \lambda_t, \lambda_{t-1}, h_t, h_{t-1}, \tau_t|y_{1:t-1}^r) \\ &= p(\mathbf{\Sigma}_t)\, p(\beta_t)\, p(x_t^e|y_{1:t-1}^r, \beta_t)\, p(\lambda_t)\, p(h_t)\, p(\lambda_{t-1}) \\ &\quad \times p(h_{t-1})\, p(\tau_t) \\ &= \mathrm{N}\left(x_t^e; \hat{x}_{t|t-1}^e, \mathbf{\Sigma}_t/\beta_t\right) \mathrm{IW}(\mathbf{\Sigma}_t; u_t, U_t)\, \mathrm{G}(\beta_t; \xi/2, \xi/2) \\ &\quad \times \mathrm{G}(\lambda_t; h_t/2, h_t/2)\, \mathrm{G}(h_t, e_t, g_t) \\ &\quad \times \mathrm{G}(\lambda_{t-1}; h_{t-1}/2, h_{t-1}/2)\, \mathrm{G}(h_{t-1}, e_{t-1}, g_{t-1}) \\ &\quad \times (1-\phi_t)^{(1-\tau_t)} \phi_t^{\tau_t}. \end{aligned} \tag{36}$$

Thus far, the HGSSM consisting of (12), (14)-(19), (25)-(26) and (29)-(36) is constructed. Next, the expanded state vector, BRV, scale matrix and IRVs will be inferred by introducing a variational Bayesian approach.

3.1.3 VB approximation of the joint posterior PDFs

The parameters in (36) are mutually coupled. It is difficult to calculate the analytic solution of the posterior PDF $p(\mathbf{\Theta}|y_{1:t}^r)$, $\mathbf{\Theta} = (x_t^e, \mathbf{\Sigma}_t, \beta_t, \lambda_t, \lambda_{t-1}, h_t, h_{t-1}, \tau_t)$. The free form factored approximate PDF for $p\left(\mathbf{\Theta}|y_{1:t}^r\right)$ will be solved using the VB approach as follows:

$$p(\mathbf{\Theta}|y_{1:t}^r) \approx q(x_t^e)\, q(\mathbf{\Sigma}_t)\, q(\beta_t)\, q(\lambda_t) \times q(\lambda_{t-1})\, q(h_t)\, q(h_{t-1})\, q(\tau_t), \tag{37}$$

where $q(\cdot)$ denote the approximate posterior PDFs of the element in $\mathbf{\Theta}$.

Remark 3: In the framework of the VB method, the indicator of the distance between the true joint PDF and the factored approximate PDF is Kullback–Leibler divergence (KLD). By minimizing the KLD, the optimal solution can be calculated. The closed-form solution for the approximate PDF can be obtained by the VB method, which can also guarantee the local convergence of the fixed-point iterations.

According to Remark 3, the approximate posterior PDFs in (37) can be obtained by minimizing the KLD, i.e., [30],

$$\begin{aligned} &\{q(x_t^e)\, q(\mathbf{\Sigma}_t)\, q(\beta_t)\, q(\lambda_t)\, q(\lambda_{t-1})\, q(h_t)\, q(h_{t-1})\, q(\tau_t)\} \\ &= \operatorname{argmin} \mathrm{KLD}\{q(x_t^e)\, q(\mathbf{\Sigma}_t)\, q(\beta_t)\, q(\lambda_t)\, q(\lambda_{t-1}) \\ &\quad \times q(h_t)\, q(h_{t-1})\, q(\tau_t)\, ||\, p(\mathbf{\Theta}|y_{1:t}^r)\}, \end{aligned} \tag{38}$$

where KLD is defined as

$$\mathrm{KLD}[q(x) \parallel p(x)] \triangleq \int q(x) \log[q(x)/p(x)]\, dx. \tag{39}$$

The optimal solution of (39) is calculated as

$$\log q(\theta) = \mathrm{E}_{p\,\Theta^{-\theta}}[\log p(\mathbf{\Theta}|y_{1:t}^r)] + c_\theta, \tag{40}$$

where $\log q(\theta)$ is the natural logarithmic operation of $q(\theta)$. $\mathbf{\Theta}^{-\theta}$ is the collection of all elements in $\mathbf{\Theta}$ apart from θ, and c_θ is the constant with respect to θ. Next, the fixed-point iterations method is employed to calculate the approximate formation of the parameters that are coupled in (28).

Furthermore, the joint PDF $p(\mathbf{\Theta}|y_{1;t}^r)$ is derived as

$$\begin{aligned} &p(\mathbf{\Theta}, y_{1:t}^r) \\ &= p(y_{1:t-1}^r)\, \mathrm{N}(y_t^r; \mathbf{H}_t x_t, \mathbf{R}_t/\lambda_t)^{(1-\tau_t)} \\ &\quad \times \mathrm{N}(y_t^r; \mathbf{H}_{t-1} x_{t-1}, \mathbf{R}_{t-1}/\lambda_{t-1})^{\tau_t} \mathrm{N}\left(x_t^e; \hat{x}_{t|t-1}^e, \mathbf{\Sigma}_t/\beta_t\right) \\ &\quad \times \mathrm{IW}(\mathbf{\Sigma}_t; u_t, U_t)\, \mathrm{G}(\beta_t; \xi/2, \xi/2)\, (1-\phi_t)^{(1-\tau_t)} \phi_t^{\tau_t} \\ &\quad \times \mathrm{G}(\lambda_{t-1}; h_{t-1}/2, h_{t-1}/2)\, \mathrm{G}(h_{t-1}, e_{t-1}, g_{t-1}) \\ &\quad \times \mathrm{G}(\lambda_t; h_t/2, h_t/2)\, \mathrm{G}(h_t, e_t, g_t). \end{aligned} \tag{41}$$

Using (41), $\log p(\mathbf{\Theta}, y_{1;t}^r)$ is derived as follows:

$$\begin{aligned} &\log p(\mathbf{\Theta}, y_{1:t}^r) \\ &= \left(\frac{\xi+n}{2} - 1\right) \log \beta_t - 0.5\xi \beta_t - 0.5\beta_t \\ &\quad \times \left(x_t^e - \hat{x}_{t|t-1}^e\right)^T \mathbf{\Sigma}_t \left(x_t^e - \hat{x}_{t|t-1}^e\right) - 0.5\mathrm{tr}\left(A_t \mathbf{\Sigma}_t^{-1}\right) \\ &\quad - \log|\mathbf{\Sigma}_t|(\xi + 2n + 1) \\ &\quad + 0.5[m(1-\tau_t) + h_t - 2] \log \lambda_t - 0.5 h_t \lambda_t \\ &\quad - 0.5(1-\tau_t)\lambda_t (y_t^r - \mathbf{H}_t x_t)^T \mathbf{R}_t^{-1} (y_t^r - \mathbf{H}_t x_t) \\ &\quad + 0.5(m\tau_t + h_{t-1} - 2) \log \lambda_t - 0.5 h_{t-1} \lambda_{t-1} \\ &\quad - 0.5\tau_t \lambda_{t-1} (y_t^r - \mathbf{H}_{t-1} x_{t-1})^T \mathbf{R}_{t-1}^{-1} (y_t^r - \mathbf{H}_{t-1} x_{t-1}) \\ &\quad + 0.5 h_t \log \frac{h_t}{2} + (e_t - 1) \log h_t - \log \Gamma(0.5 h_t) \\ &\quad - g_t h_t + 0.5 h_{t-1} \log \frac{h_{t-1}}{2} + (e_{t-1} - 1) \log h_{t-1} \\ &\quad - \log \Gamma(0.5 h_{t-1}) - g_{t-1} h_{t-1} + (1-\tau_t) \log(1-\phi_t) \\ &\quad + \tau_t \log \phi_t + c_{\mathbf{\Theta}}. \end{aligned} \tag{42}$$

Proposition 1: Let $\theta = x_t^e$ and using (40) in (42), the following equation can be obtained

$$\begin{aligned} &\log q^{(s+1)}(x_t^e) \\ &= -0.5\mathrm{E}_p^{(s)}[\beta_t] \left(x_t^e - \hat{x}_{t|t-1}^e\right)^T \mathrm{E}_p^{(s)}\left[\mathbf{\Sigma}_t^{-1}\right] \left(x_t^e - \hat{x}_{t|t-1}^e\right) \\ &\quad - 0.5\mathrm{E}_p^{(s)}[(1-\tau_t)]\, \mathrm{E}_p^{(s)}[\lambda_t]\, (y_t^r - \mathbf{H}_t x_t)^T \mathbf{R}_t^{-1} \end{aligned}$$

$$\times (y_t^r - \mathbf{H}_t x_t) - 0.5\mathrm{E}_p^{(s)}[\tau_t]\,\mathrm{E}_p^{(s)}[\lambda_{t-1}]$$
$$\times (y_t^r - \mathbf{H}_{t-1}x_{t-1})^T \mathbf{R}_t^{-1}(y_t^r - \mathbf{H}_{t-1}x_{t-1}) + c_{x_t^e}. \quad (43)$$

where $q^{(s+1)}(\cdot)$ is the approximation of PDF $q(\cdot)$ at the $(s+1)$th iteration, and $\mathrm{E}_p^{(s)}[D]$ is the expectation of variable D at the sth iteration.

Furthermore, (43) can be rewritten as follows:

$$\log q^{(s+1)}(x_t^e) = -0.5\mathrm{E}_p^{(s)}[\beta_t]\left(x_t^e - \hat{x}_{t|t-1}^e\right)^T \mathrm{E}_p^{(s)}[\mathbf{\Sigma}_t^{-1}]\left(x_t^e - \hat{x}_{t|t-1}^e\right) - 0.5(y_t^{er} - \mathbf{H}_t^e x_t^e)^T \tilde{\mathbf{R}}_t^{-1}(y_t^{er} - \mathbf{H}_t^e x_t^e) + c_{x_t^e}. \quad (44)$$

The expanded real measurement vector y_t^{er}, expanded measurement matrix $\mathbf{H}_t^e$, and modified measurement noise covariance matrix $\tilde{\mathbf{R}}_t$ are given as

$$y_t^{er} = \begin{bmatrix} y_t^{rT} & y_t^{rT} \end{bmatrix}^T, \quad (45)$$

$$\mathbf{H}_t^e = \begin{bmatrix} \mathbf{H}_t & 0 \\ 0 & \mathbf{H}_{t-1} \end{bmatrix}, \quad (46)$$

$$\tilde{\mathbf{R}}_t^{(s+1)} = \begin{bmatrix} \frac{\mathbf{R}_t}{\mathrm{E}_p^{(s)}[(1-\tau_t)]\mathrm{E}_p^{(s)}[\lambda_t]} & 0 \\ 0 & \frac{\mathbf{R}_{t-1}}{\mathrm{E}_p^{(s)}[\tau_t]\mathrm{E}_p^{(s)}[\lambda_{t-1}]} \end{bmatrix}. \quad (47)$$

Next, the modified one-step predicted PDF $p(x_t^e|y_{1:t-1}^r)$ and modified likelihood PDF $p(y_t^{er}|x_t^e)$ are defined by the following Gaissian distributions

$$p^{(s+1)}(x_t^e|y_{1:t-1}^r) = \mathrm{N}\left(x_t^e; \hat{x}_{t|t-1}^e, \tilde{\mathbf{\Sigma}}_t^{(s+1)}\right), \quad (48)$$

$$p^{(s+1)}(y_t^{er}|x_t^e) = \mathrm{N}\left(y_t^{er}; \mathbf{H}_t^e x_t^e, \tilde{\mathbf{R}}_t^{(s+1)}\right), \quad (49)$$

where $\tilde{\mathbf{\Sigma}}_t^{(s+1)}$ is the modified predicted error covariance matrix, which is formulated as

$$\tilde{\mathbf{\Sigma}}_t^{(s+1)} = \frac{\left\{\mathrm{E}_p^{(s)}[\mathbf{\Sigma}_t^{-1}]\right\}^{-1}}{\mathrm{E}_p^{(s)}[\beta_t]} = \begin{bmatrix} \mathbf{P}_{t|t-1}^{(s+1)} & \mathbf{P}_{t-1,t|t-1}^{(s+1)} \\ \left(\mathbf{P}_{t-1,t|t-1}^{(s+1)}\right)^T & \mathbf{P}_{t-1|t-1}^{(s+1)} \end{bmatrix}^T. \quad (50)$$

Based on (44)-(50), the approximate PDF $q^{(s+1)}(x_t^e)$ can be updated as follows:

$$q^{(s+1)}(x_t^e) = \mathrm{N}\left(x_t^e; \hat{x}_{t|t}^{e(s+1)}, \mathbf{\Sigma}_{t|t}^{(s+1)}\right), \quad (51)$$

where $\hat{x}_{t|t-1}^{e(s+1)}$ and $\mathbf{\Sigma}_{t|t}^{(s+1)}$ are defined by

$$\hat{x}_{t|t}^{e(s+1)} = \begin{bmatrix} \hat{x}_{t|t}^{(s+1)} & \hat{x}_{t-1|t}^{(s+1)} \end{bmatrix}, \quad (52)$$

$$\mathbf{\Sigma}_{t|t}^{(s+1)} = \begin{bmatrix} \mathbf{P}_{t|t}^{(s+1)} & \mathbf{P}_{t-1,t|t}^{(s+1)} \\ \left(\mathbf{P}_{t-1,t|t}^{(s+1)}\right)^T & \mathbf{P}_{t-1|t}^{(s+1)} \end{bmatrix}^T. \quad (53)$$

The state estimate $\hat{x}_{t|t}^{(s+1)}$ and the corresponding covariance matrix of estimation error $\mathbf{P}_{t|t}^{(s+1)}$ can be calculated by the traditional Kalman filter as follows:

$$\hat{x}_{t|t}^{(s+1)} = \hat{x}_{t|t-1}^{(s+1)} + \mathbf{K}_t^{(s+1)}\left(y_t^{er} - \hat{y}_{t|t-1}^{er}\right), \quad (54)$$

$$\mathbf{P}_{t|t}^{(s+1)} = \mathbf{P}_{t|t-1}^{(s+1)} - \mathbf{K}_t^{(s+1)}\mathbf{P}_{t|t-1}^{yy(s+1)}\left(\mathbf{K}_t^{(s+1)}\right)^T, \quad (55)$$

$$\mathbf{K}_t^{(s+1)} = \mathbf{P}_{t|t-1}^{xy(s+1)}\left(\mathbf{P}_{t|t-1}^{xy(s+1)}\right)^{-1}, \quad (56)$$

$$\mathbf{P}_{t|t-1}^{xy^{er}(s+1)} = \left[\mathbf{P}_{t|t-1}^{(s+1)}\left(\mathbf{P}_{t-1,t|t-1}^{(s+1)}\right)^T\right](\mathbf{H}_t^e)^T, \quad (57)$$

$$\mathbf{P}_{t|t-1}^{y^{er}y^{er}(s+1)} = \mathbf{H}_t^e\tilde{\mathbf{\Sigma}}_t^{(s+1)}(\mathbf{H}_t^e)^T + \tilde{\mathbf{R}}_t^{(s+1)}, \quad (58)$$

where $\mathbf{K}_t^{(s+1)}$ is the gain of the Kalman filter; $\mathbf{P}_{t|t-1}^{xy(s+1)}$ and $\mathbf{P}_{t|t-1}^{y^{er}y^{er}(s+1)}$ are the cross-covariance matrix and innovation covariance matrix, respectively.

The state estimation $\hat{x}_{t-1|t}^{(s+1)}$ of the one-step smoothing and the related estimation error covariance matrix $\mathbf{P}_{t-1|t}^{(s+1)}$ can be obtained as follows:

$$\hat{x}_{t-1|t}^{(s+1)} = \hat{x}_{t-1|t-1} + \mathbf{K}_{t-1}^{o(s+1)}\left(y_t^{er} - \hat{y}_{t|t-1}^{er}\right), \quad (59)$$

$$\mathbf{P}_{t-1|t}^{(s+1)} = \mathbf{P}_{t-1|t-1}^{(s+1)} - \mathbf{K}_{t-1}^{o(s+1)}\mathbf{P}_{t|t-1}^{y^{er}y^{er}(s+1)}\left(\mathbf{K}_{t-1}^{o(s+1)}\right)^T, \quad (60)$$

$$\mathbf{K}_{t-1}^{o(s+1)} = \mathbf{P}_{t-1,t|t-1}^{xy^{er}(s+1)}\left(\mathbf{P}_{t|t-1}^{y^{er}y^{er}(s+1)}\right)^{-1}, \quad (61)$$

$$\mathbf{P}_{t-1,t|t-1}^{xy^{er}(s+1)} = \left[\mathbf{P}_{t-1,t|t-1}^{(s+1)}\left(\mathbf{P}_{t-1|t-1}^{(s+1)}\right)^T\right](\mathbf{M}_t^e)^T, \quad (62)$$

where $\mathbf{K}_{t-1}^{o(s+1)}$ is the gain matrix of the one-step smoothing; $\mathbf{P}_{t-1,t|t-1}^{xy^{er}(s+1)}$ is the cross-covariance matrix.

Additionally, matrix $\mathbf{P}_{t-1,t|t}^{(s+1)}$ of (53) is calculated as

$$\mathbf{P}_{t-1,t|t}^{(s+1)} = \mathbf{P}_{t-1,t|t-1}^{(s+1)} - \mathbf{K}_{t-1}^{o(s+1)}\mathbf{P}_{t|t-1}^{y^{er}y^{er}(s+1)}\left(\mathbf{K}_{t-1}^{o(s+1)}\right)^T. \quad (63)$$

Proposition 2: Let $\theta = \mathbf{\Sigma}_t$ and using (40) in (42), the following equation can be obtained

$$\log q^{(s+1)}(\mathbf{\Sigma}_t) = -0.5(n+\xi+2)\log|\mathbf{\Sigma}_t| - 0.5\mathrm{tr}\left[\left(U_t + \mathrm{E}_p^{(s)}[\beta_t]\mathbf{A}_t^{(s+1)}\right)\mathbf{\Sigma}_t^{-1}\right] + c_{\mathbf{\Sigma}_t}, \quad (64)$$

where the required expectation $\mathbf{A}_t^{(s+1)}$ is formulated as

$$\mathbf{A}_t^{(s+1)} = \mathrm{E}_p^{(s+1)}\left[\left(x_t^e - \hat{x}_{t|t-1}^e\right)\left(x_t^e - \hat{x}_{t|t-1}^e\right)^T\right] = \mathrm{E}_p^{(s+1)}\left[\left(x_t^e - \hat{x}_{t|t}^{e(s+1)} + \hat{x}_{t|t}^{e(s+1)} - \hat{x}_{t|t-1}^e\right) \times \left(x_t^e - \hat{x}_{t|t}^{e(s+1)} + \hat{x}_{t|t}^{e(s+1)} - \hat{x}_{t|t-1}^e\right)^T\right]$$

$$= \left(\hat{x}_{t|t}^{e(s+1)} - \hat{x}_{t|t-1}^{e}\right)\left(\hat{x}_{t|t}^{e(s+1)} - \hat{x}_{t|t-1}^{e}\right)^T + \boldsymbol{\Sigma}_{t|t}^{(s+1)}. \quad (65)$$

According to (64), $q^{(s+1)}(\boldsymbol{\Sigma}_t)$ is uploaded as the following inverse-Wishart distribution, i.e.,

$$q^{(s+1)}(\boldsymbol{\Sigma}_t) = \mathrm{IW}\left(\boldsymbol{\Sigma}_t; \hat{u}_t^{(s+1)}, \hat{\boldsymbol{U}}_t^{(s+1)}\right), \quad (66)$$

$$\hat{u}_t^{(s+1)} = u_t + 1, \quad (67)$$

$$\hat{\boldsymbol{U}}_t^{(s+1)} = \boldsymbol{U}_t + \mathrm{E}_p^{(s)}[\beta_t]\,\mathbf{A}_t^{(s+1)}. \quad (68)$$

And the expectation of $\boldsymbol{\Sigma}_t^{-1}$ is formulated as

$$\mathrm{E}_p^{(s+1)}\left[\boldsymbol{\Sigma}_t^{-1}\right] = \frac{\left(\hat{u}_{t|t}^{(s+1)} - 2n - 1\right)}{\hat{\boldsymbol{U}}_{t|t}^{(s+1)}}. \quad (69)$$

Proposition 3: Let $\theta = \beta_t$ and using (40) in (42), the following equation can be obtained

$$\log q^{(s+1)}(\beta_t) = \left(\frac{\xi + n}{2} - 1\right)\log\beta_t - 0.5\beta_t\left\{\xi + \mathrm{tr}\left(\mathbf{A}_t^{(s+1)}\mathrm{E}_p^{(s+1)}\left[\boldsymbol{\Sigma}_t^{-1}\right]\right)\right\} + c_{\beta_t}. \quad (70)$$

According to (70), $q^{(s+1)}(\beta_t)$ is uploaded as the following Gamma distribution

$$q^{(s+1)}(\beta_t) = \mathrm{G}\left(\beta_t; t_t^{(s+1)}, T_t^{(s+1)}\right), \quad (71)$$

$$t_t^{(s+1)} = n + 0.5\xi, \quad (72)$$

$$\boldsymbol{T}_t^{(s+1)} = 0.5\left\{\xi + tr\left(\mathbf{A}_t^{(s+1)}\mathrm{E}_p^{(s+1)}\left[\boldsymbol{\Sigma}_t^{-1}\right]\right)\right\}. \quad (73)$$

The required expectation of β_t is

$$\mathrm{E}_p^{(s+1)}[\beta_t] = t_t^{(s+1)}\left(\boldsymbol{T}_t^{(s+1)}\right)^{-1}. \quad (74)$$

Proposition 4: Let $\theta = \lambda_t$ and using (40) in (42), the following equation can be obtained

$$\log q^{(s+1)}(\lambda_t) = \left(\frac{m\mathrm{E}_p^{(s)}[1-\tau_t] + \mathrm{E}_p^{(s)}[h_t]}{2} - 1\right)\log\lambda_t - 0.5\lambda_t\mathrm{E}_p^{(s)}[1-\tau_t](y_t^r - \mathbf{H}_t x_t)^T\mathbf{R}_t^{-1}(y_t^r - \mathbf{H}_t x_t) - 0.5\lambda_t\mathrm{E}_p^{(s)}[h_t] + c_{\lambda_t}. \quad (75)$$

According to (75), $q^{(s+1)}(\lambda_t)$ is uploaded as the following Gamma distribution

$$q^{(s+1)}(\lambda_t) = \mathrm{G}\left(\lambda_t; \epsilon_t^{(s+1)}, \varrho_t^{(s+1)}\right), \quad (76)$$

$$\epsilon_t^{(s+1)} = 0.5\left(\mathrm{E}_p^{(s)}[h_t] + m\mathrm{E}_p^{(s)}[1-\tau_t]\right), \quad (77)$$

$$\varrho_t^{(s+1)} = 0.5\left\{\mathrm{E}_p^{(s)}[h_t] + \mathrm{tr}\left(\mathbf{B}_t^{(s+1)}\mathbf{Y}_1\left(\mathbf{F}_t^{(s+1)}\right)^{-1}\right)\right\}, \quad (78)$$

$$\mathbf{B}_t^{(s+1)} = \mathrm{E}_p^{(s+1)}\left[(y_t^r - \mathbf{H}_t x_t)(y_t^r - \mathbf{H}_t x_t)^T\right], \quad (79)$$

$$\mathbf{Y}_1 = \begin{bmatrix} \mathrm{ID} & 0 \\ 0 & 0 \end{bmatrix}, \quad (80)$$

$$\mathbf{F}_t^{(s+1)} = \begin{bmatrix} \frac{\mathbf{R}_t}{\mathrm{E}_p^{(s+1)}[(1-\tau_t)]} & 0 \\ 0 & \frac{\mathbf{R}_{t-1}}{\mathrm{E}_p^{(s+1)}[\tau_t]} \end{bmatrix}, \quad (81)$$

where the element ID in (80) is the m-dimensional identity matrix. The required expectations of λ_t and $\log\lambda_t$ are given as

$$\mathrm{E}_p^{(s+1)}[\lambda_t] = \frac{\epsilon_t^{(s+1)}}{\varrho_t^{(s+1)}}, \quad (82)$$

$$\mathrm{E}_p^{(s+1)}[\log(\lambda_t)] = \psi\left(\epsilon_t^{(s+1)}\right) - \log\left(\varrho_t^{(s+1)}\right), \quad (83)$$

where $\psi(\cdot)$ represents the digamma function.

Proposition 5: Let $\theta = \lambda_{t-1}$ and using (40) in (42), the following equation can be obtained

$$\log q^{(s+1)}(\lambda_{t-1}) = \left(\frac{m\mathrm{E}_p^{(s)}[\tau_t] + \mathrm{E}_p^{(s)}[h_{t-1}]}{2} - 1\right)\log\lambda_{t-1} - 0.5\lambda_{t-1}\left\{\mathrm{E}_p^{(s)}[\tau_t](y_t^r - \mathbf{H}_{t-1}x_{t-1})^T \times \mathbf{R}_{t-1}^{-1}(y_t^r - \mathbf{H}_{t-1}x_{t-1}) + \mathrm{E}_p^{(s)}[h_{t-1}]\right\} + c_{\lambda_{t-1}}. \quad (84)$$

According to (84), $q^{(s+1)}(\lambda_{t-1})$ is also uploaded as the following Gamma distribution

$$q^{(s+1)}(\lambda_{t-1}) = \mathrm{G}\left(\lambda_{t-1}; \epsilon_{t-1}^{(s+1)}, \varrho_{t-1}^{(s+1)}\right), \quad (85)$$

$$\epsilon_{t-1}^{(s+1)} = 0.5\left(\mathrm{E}_p^{(s)}[h_{t-1}] + m\mathrm{E}_p^{(s)}[\tau_t]\right), \quad (86)$$

$$\varrho_{t-1}^{(s+1)} = 0.5\left\{\mathrm{E}_p^{(s)}[h_{t-1}] + \mathrm{tr}\left(\mathbf{B}_t^{(s+1)}\mathbf{Y}_2\left(\mathbf{F}_t^{(s+1)}\right)^{-1}\right)\right\}, \quad (87)$$

$$\mathbf{Y}_2 = \begin{bmatrix} 0 & 0 \\ 0 & \mathrm{ID} \end{bmatrix}. \quad (88)$$

The required expectations of λ_{t-1} and $\log\lambda_{t-1}$ are

$$\mathrm{E}_p^{(s+1)}[\lambda_{t-1}] = \frac{\epsilon_{t-1}^{(s+1)}}{\varrho_{t-1}^{(s+1)}}, \quad (89)$$

$$\mathrm{E}_p^{(s+1)}[\log(\lambda_{t-1})] = \psi\left(\epsilon_{t-1}^{(s+1)}\right) - \log\left(\varrho_{t-1}^{(s+1)}\right). \quad (90)$$

Proposition 6: Let $\theta = h_t$ and using (40) in (42), the following equation can be obtained

$$\log q^{(s+1)}(h_t)$$

$$= \log(h_t)(e_t - 0.5) - h_t g_t - h_t \left\{0.5\mathrm{E}_p^{(s+1)}[\lambda_t] - 0.5\mathrm{E}_p^{(s+1)}[\log(\lambda_t)] + 0.5\right\} + c_{h_t}. \tag{91}$$

According to (91), $q^{(s+1)}(h_t)$ is uploaded as the following Gamma distribution

$$q^{(s+1)}(h_t) = \mathrm{G}\left(h_t; \hat{e}_t^{(s+1)}, \hat{g}_t^{(s+1)}\right), \tag{92}$$

$$\hat{e}_t^{(s+1)} = 0.5 + e_t^{(s+1)}, \tag{93}$$

$$\hat{g}_t^{(s+1)} = 0.5 + 0.5\mathrm{E}_p^{(s+1)}[\lambda_t] - 0.5\mathrm{E}_p^{(s+1)}[\log(\lambda_t)] + g_t^{(s+1)}. \tag{94}$$

The required expectation of h_t is given as

$$\mathrm{E}_p^{(s+1)}[h_t] = \frac{\hat{e}_t^{(s+1)}}{\hat{g}_t^{(s+1)}}. \tag{95}$$

Proposition 7: Let $\theta = h_{t-1}$ and using (40) in (42), the following equation can be obtained

$$\log q^{(s+1)}(h_{t-1}) = \log(h_{t-1})(e_{t-1} - 0.5) - h_{t-1}g_{t-1} - 0.5h_{t-1}\mathrm{E}_p^{(s+1)}[\lambda_{t-1}] + 0.5h_{t-1}\mathrm{E}_p^{(s+1)}[\log(\lambda_{t-1})] - 0.5h_{t-1} + c_{h_{t-1}}. \tag{96}$$

According to (96), $q^{(s+1)}(h_{t-1})$ is also uploaded as the following Gamma distribution

$$q^{(s+1)}(h_{t-1}) = \mathrm{G}\left(h_{t-1}; \hat{e}_{t-1}^{(s+1)}, \hat{g}_{t-1}^{(s+1)}\right), \tag{97}$$

$$\hat{e}_{t-1}^{(s+1)} = 0.5 + e_{t-1}^{(s+1)}, \tag{98}$$

$$\hat{g}_{t-1}^{(s+1)} = 0.5 + 0.5\mathrm{E}_p^{(s+1)}[\lambda_{t-1}] - 0.5\mathrm{E}_p^{(s+1)}[\log(\lambda_{t-1})] + g_{t-1}^{(s+1)}. \tag{99}$$

The required expectation of h_{t-1} is

$$\mathrm{E}_p^{(s+1)}[h_{t-1}] = \frac{\hat{e}_{t-1}^{(s+1)}}{\hat{g}_{t-1}^{(s+1)}}. \tag{100}$$

Proposition 8: Let $\theta = \tau_t$ and using (40) in (42), the following equation can be obtained

$$\begin{aligned} &\log q^{(s+1)}(\tau_t) \\ &= \tau_t \left\{\log\phi_t + 0.5\left[m\mathrm{E}_p^{(s+1)}[\log(\lambda_{t-1})] - \mathrm{E}_p^{(s+1)}[\lambda_{t-1}] \times (y_t^r - \mathbf{H}_{t-1}x_{t-1})^T \mathbf{R}_{t-1}^{-1}(y_t^r - \mathbf{H}_{t-1}x_{t-1})\right]\right\} \\ &+ 0.5\left[m\mathrm{E}_p^{(s+1)}[\log(\lambda_t)] - \mathrm{E}_p^{(s+1)}[\lambda_t](y_t^r - \mathbf{H}_t x_t)^T \times \mathbf{R}_t^{-1}(y_t^r - \mathbf{H}_{t-1}x_{t-1})\right](1-\tau_t) \\ &+ \log(1-\phi_t)(1-\tau_t) + c_{\tau_t}. \end{aligned} \tag{101}$$

According to (101), the probability of BRV is

$$p(\tau_t = 0) = \exp\left\{\log(1-\phi_t) + \mathrm{E}_p^{(s+1)}[\log(\lambda_t)] - 0.5\mathrm{tr}\left(\mathbf{A}_t^{(s+1)}\mathbf{Y}_1\left(\mathbf{L}_t^{(s+1)}\right)^{-1}\right)\right\}\Delta_t^{(s+1)}, \tag{102}$$

$$p(\tau_t = 1) = \exp\left\{\log(\phi_t) + \mathrm{E}_p^{(s+1)}[\log(\lambda_{t-1})] - 0.5\mathrm{tr}\left(\mathbf{A}_t^{(s+1)}\mathbf{Y}_2\left(\mathbf{L}_t^{(s+1)}\right)^{-1}\right)\right\}\Delta_t^{(s+1)}, \tag{103}$$

$$\mathbf{L}_t^{(s+1)} = \begin{bmatrix} \frac{\mathbf{R}_t}{\mathrm{E}_p^{(s+1)}[\lambda_t]} & 0 \\ 0 & \frac{\mathbf{R}_{t-1}}{\mathrm{E}_p^{(s+1)}[\lambda_{t-1}]} \end{bmatrix}, \tag{104}$$

where $\exp(\cdot)$ represents an exponential calculation, and $\Delta_t^{(s+1)}$ is the normalizing constant.

The required expectations of τ_t and $(1-\tau_t)$ are given as

$$\mathrm{E}_p^{(s+1)}[\tau_t] = \frac{p^{(s+1)}(\tau_t = 1)}{p^{(s+1)}(\tau_t = 1) + p^{(s+1)}(\tau_t = 0)}, \tag{105}$$

$$\mathrm{E}_p^{(s+1)}[1-\tau_t] = 1 - \mathrm{E}_p^{(s+1)}[\tau_t]. \tag{106}$$

The proposed filter in this paper consists of (14)-(19), (50), (52)-(63), (65)-(69), (71)-(74), (76)-(83), (85)-(90), (92)-(95), (97)-(100), and (102)-(106). The implementation is listed in Algorithm 1.

4. SIMULATIONS

The proposed filter is contrasted with the current filter algorithms in a simulation of a target tracking environment, which is given as the following linear stochastic system

$$x_t = \begin{bmatrix} \mathrm{ID}_2 & \Delta t\mathrm{ID}_2 \\ 0_2 & \mathrm{ID}_2 \end{bmatrix} x_{t-1} + \omega_{t-1}, \tag{107}$$

$$y_t^r = \begin{bmatrix} \mathrm{ID}_2 & 0_2 \end{bmatrix} + v_t, \tag{108}$$

where Δt denotes the interval of sampling, and 0_2 represents 2-dimensional zero matrix. The initial state vector x_0 and the corresponding error covariance matrix $\mathbf{P}_0$ are set as follows:

$$x_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T, \tag{109}$$

$$\mathbf{P}_0 = diag\begin{bmatrix} 1000 & 1000 & 100 & 100 \end{bmatrix}^T. \tag{110}$$

The true NHPMN are given as

$$\omega_t = \begin{cases} \mathrm{N}(0,\mathbf{Q}) \; 1 - prp = 0.95, \\ \mathrm{N}(0,100\mathbf{Q}) \; prp = 0.05, \end{cases} \tag{111}$$

$$v_t = \begin{cases} \mathrm{N}(0,\mathbf{R}) \; 1 - prm = 0.95, \\ \mathrm{N}(0,100\mathbf{R}) \; prm = 0.05, \end{cases} \tag{112}$$

where prp and prm are the probability of the outliers in process noise and measurement noise, respectively. Nominal process noise $\mathbf{Q}$ and nominal measurement noise $\mathbf{R}$ are set as

Algorithm 1: The proposed VB-based Kalman filter with random measurement delay and non-Gaussian heavy-tailed process and measurement noises.

Inputs: $\hat{x}_{t-1|t-1}$, $\mathbf{P}_{t-1|t-1}$, $\mathbf{M}_{t-1}$, $\mathbf{H}_t$, y_t^r, $\mathbf{Q}_{t-1}$, $\mathbf{R}_t$, m, n, e_t, g_t, ϕ_t, ξ, γ, N, ε.

Time update:

1: Obtain $\hat{x}^e_{t|t-1}$ and $\mathbf{\Sigma}_t$ utilizing (14)-(19).

Variational update:

2: Parameter Initialization: $\hat{x}^{e(0)}_{t|t-1} = x^e_{t|t-1}$, $\mathbf{\Sigma}_t^{(0)} = \mathbf{\Sigma}_t$, $\mathrm{E}_p^{(0)}[\lambda_{t-1}] = 1$, $\mathrm{E}_p^{(0)}[\lambda_t] = 1$, $\mathrm{E}_p^{(0)}[\tau_t] = 0.5$, $\mathrm{E}_p^{(0)}[1-\tau_t] = 0.5$, $\hat{u}^{(0)}_{t|t-1} = n+\gamma+1$, $\hat{\boldsymbol{U}}^{(0)}_{t|t-1} = \gamma \mathbf{P}^{ee}_{t|t-1}$, $\mathrm{E}_p^{(0)}[h_{t-1}] = 5$, $\mathrm{E}_p^{(0)}[h_t] = 5$.

for $s = 0: N-1$

Update $q^{(s+1)}(x_t^e)$ as (51):

3: Obtain $\hat{x}^{(s+1)}_{t|t}$ and $\mathbf{\Sigma}^{(s+1)}_{t|t}$ utilizing (45)-(47), (50), (54)-(63), (69), (74), (82)-(83), (89)-(90), (95), (100) and (105)-(106).

Update $q^{(s+1)}(\mathbf{\Sigma}_t)$ as (66).

4: Obtain $\hat{u}_t^{(s+1)}$ and $\hat{\boldsymbol{U}}_t^{(s+1)}$ utilizing (65) and (67)-(69).

Update $q^{(s+1)}(\beta_t)$ as (71).

5: Obtain $t_t^{(s+1)}$ and $\boldsymbol{T}_t^{(s+1)}$ utilizing (65) and (72)-(74).

Update $q^{(s+1)}(\lambda_t)$ as (76):

6: Obtain $\epsilon_t^{(s+1)}$ and $\varrho_t^{(s+1)}$ utilizing (77)-(83).

Update $q^{(s+1)}(\lambda_{t-1})$ as (85):

7: Obtain $\epsilon_{t-1}^{(s+1)}$ and $\varrho_{t-1}^{(s+1)}$ utilizing (86)-(90).

Update $q^{(s+1)}(h_t)$ as (92):

8: Obtain $\hat{e}_t^{(s+1)}$ and $\hat{g}_t^{(s+1)}$ utilizing (93)-(95).

Update $q^{(s+1)}(h_{t-1})$ as (97):

9: Obtain $\hat{e}_{t-1}^{(s+1)}$ and $\hat{g}_{t-1}^{(s+1)}$ utilizing (98)-(100).

Update $q^{(s+1)}(\tau_t)$ as Bernoulli distribution:

10: Obtain $p(\tau_t = 1)$ and $p(\tau_t = 0)$ utilizing (102)-(106).

11: If $\frac{\left\|\hat{x}^{e(s+1)}_{t|t} - \hat{x}^{e(s)}_{t|t}\right\|}{\left\|\hat{x}^{e(s)}_{t|t}\right\|} \leq \varepsilon$, stop iteration.

end for

12: $\hat{x}^e_{t|t} = \hat{x}^{e(N)}_{t|t}$, $\mathbf{\Sigma}_{t|t} = \mathbf{\Sigma}^{(N)}_{t|t}$.

Outputs: $\hat{x}_{t|t}$ and $\mathbf{P}_{t|t}$.

$$\mathbf{Q} = a \begin{bmatrix} \frac{\Delta t^3}{3}\mathrm{ID}_2 & \frac{\Delta t^2}{2}\mathrm{ID}_2 \\ \frac{\Delta t^2}{2}\mathrm{ID}_2 & \Delta t \mathrm{ID}_2 \end{bmatrix}, \tag{113}$$

$$\mathbf{R} = b \times \mathrm{ID}_2, \tag{114}$$

where the parameters are set as $a = 1$ m^2/s^3 and $b = 100$ m^2.

The proposed filter is compared with the traditional KF, RDMKF, and VBIKF. All algorithms are coded with MALAB 2018a. The simulations are run on a computer with Intel Core i5-6300HQ. The parameters in the proposed filter are set as $\xi = 5$, $\gamma = 3$, $e_t = 5$, $g_t = 1$, $\phi_t = 0.5$, $N = 10$, and $\varepsilon = 10^{-18}$.

To evaluate the estimation performance of each filter, the root mean square error (RMSE) and averaged RMSE (AGRMSE) are utilized as performance indices, and the RMSE and AGRMSE in positions are defined as

$$\mathrm{RMSE_{pos}} = \left(\frac{1}{M_c} \sum_{i=1}^{M_c} \left(x^i_{post} - \hat{x}^i_{post}\right)^2 + \left(y^i_{post} - \hat{y}^i_{post}\right)^2 \right)^{1/2}, \tag{115}$$

$$\mathrm{AGRMSE_{pos}} = \left(\frac{1}{M_c T} \sum_{t=1}^{T} \sum_{i=1}^{M_c} \left(x^i_{post} - \hat{x}^i_{post}\right)^2 + \left(y^i_{post} - \hat{y}^i_{post}\right)^2 \right)^{1/2}, \tag{116}$$

where $\left(x^i_{post}, y^i_{post}\right)$ is the true position of the target at the i-th Monte Carlo run; $\left(\hat{x}^i_{post}, \hat{y}^i_{post}\right)$ represents the corresponding estimated position of each filter at the i-th Monte Carlo (MC) run; $M_c = 250$ is the total number of MC runs; $T = 100$ s is the total simulation time. The RMSE and AGRMSE of velocity can be obtained in a similar form.

Figs. 1 and 2 indicate the performance of different filters with delay probability $\phi_t = 0.5$. The proposed filter has better target position and velocity estimations than the existing algorithms in the scenario of random measurement delay and NHPMN. Table 1 lists the AGRMSE$_{\mathrm{pos}}$, AGRMSE$_{\mathrm{vel}}$ and single-step running times of different filters under the delay probability $\phi_t = 0.5$. Obviously, compared with the existing algorithms, the proposed filter in this paper has smaller AGRMSEs. Compared with VBIKF, which uses the same VB method, the accuracy of AGRMSE in position and velocity is improved by 29.8% and 17.0%, respectively. However, the proposed filter has higher computational complexity than the existing algorithms.

Figs. 3 and 4 show the AGRMSE curves of the proposed filter and existing filters with different probabilities

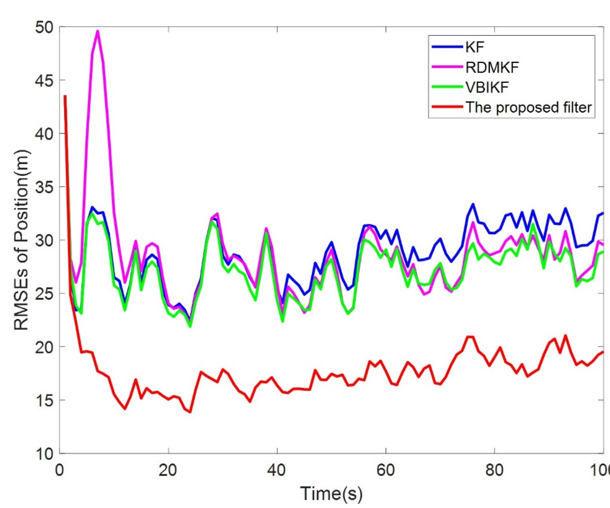

Fig. 1. RMSE$_{\mathrm{pos}}$ with different filters.

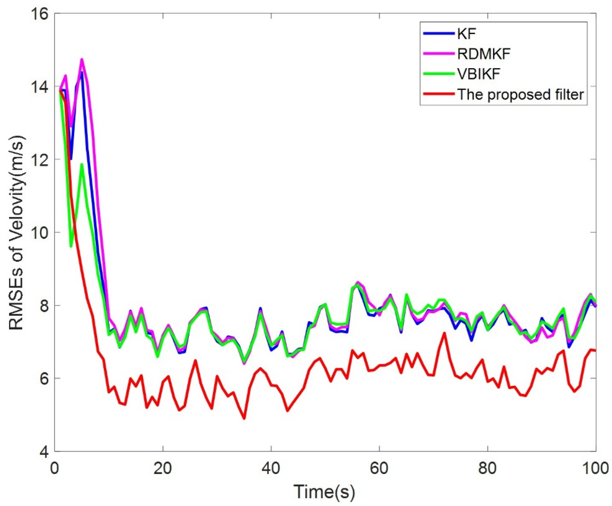

Fig. 2. $RMSE_{vel}$ with different filters.

Table 1. $AGRMSE_{pos}$, $AGRMSE_{vel}$ and single-step running times of different filters.

Filters	KF	RDMKF	VBIKF	Proposed filter
$AGRMSE_{pos}$	28.64	28.25	26.71	18.74
$AGRMSE_{vel}$	8.09	8.24	7.94	6.59
Times (ms)	0.038	0.078	1.009	1.847

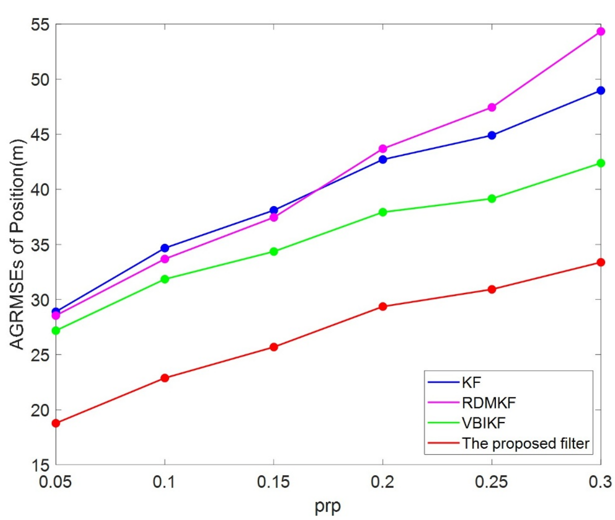

Fig. 3. $AGRMSE_{pos}$ of each filter with different probabilities of outliers in process noise.

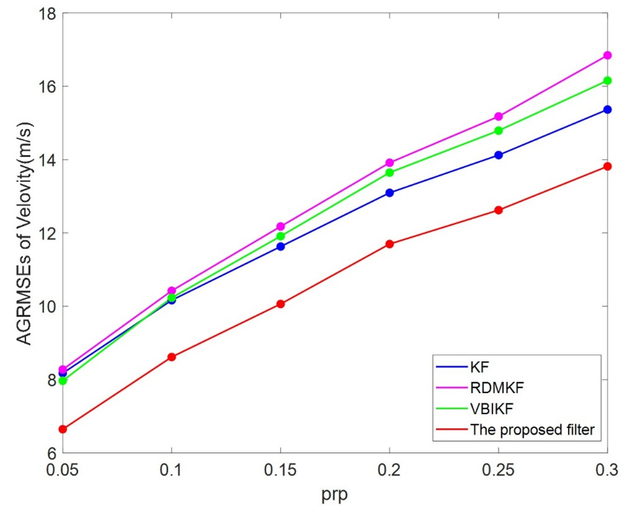

Fig. 4. $AGRMSE_{vel}$ of each filter with different probabilities of outliers in process noise.

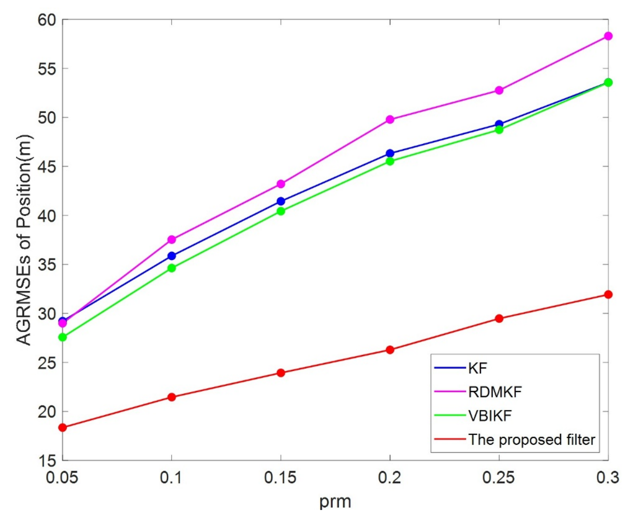

Fig. 5. $AGRMSE_{pos}$ of each filter with different probabilities of outliers in measurement noise.

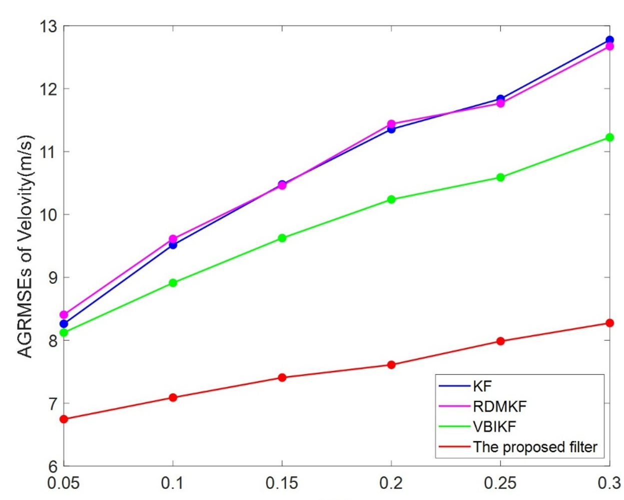

Fig. 6. $AGRMSE_{vel}$ of each filter with different probabilities of outliers in measurement noise.

of outliers in process noise $prp = 0.05, 0.1, ..., 0.3$. The proposed filter performs better than existing filters when $prp = 0.05, 0.1, ..., 0.3$.

The AGRMSE curves of the proposed filter and existing filters with different probabilities of outliers in measurement noise $prm = 0.05, 0.1, ..., 0.3$ are shown in Figs. 5 and 6. The proposed filter performs better than existing filters when $prm = 0.05, 0.1, ..., 0.3$.

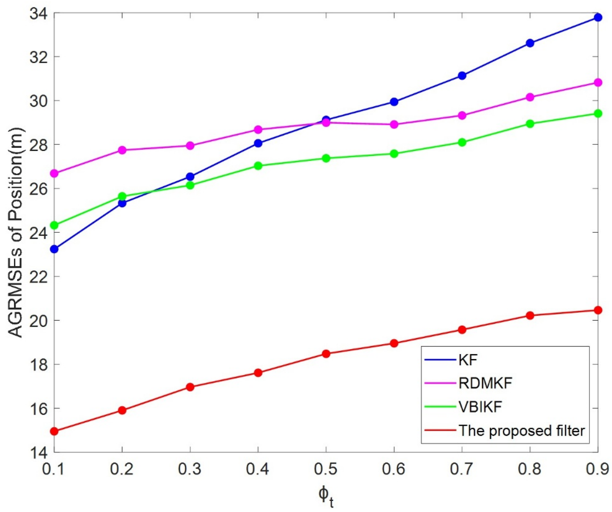

Fig. 7. $AGRMSE_{pos}$ of each filter under different delay probabilities.

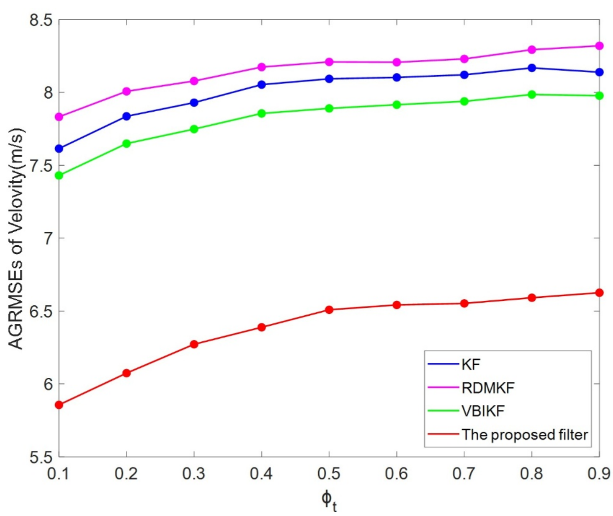

Fig. 8. $AGRMSE_{vel}$ of each filter under different delay probabilities.

In order to compare the estimation performance of the existing algorithms and proposed filter under different delay probabilities, the AGRMSE curves of each filter with different delay probabilities $\phi_t = 0.1, 0.2, ..., 0.9$ are simulated in Figs. 7 and 8. Obviously, in the scenario with NH-PMN, the proposed filter has better position and velocity estimations under different delay probabilities.

5. CONCLUSION

In this paper, a new VB-based Kalman filter is proposed to address the issue of a linear stochastic system with random measurement delay and non-Gaussian process and measurement noises. The system state, BRV and IRVs are simultaneously estimated by utilizing the variational Bayesian method. The target tracking simulation results illustrate that the proposed filter has better estimation performance and robustness than current filters to address the filtering issue for a linear system with random measurement delay and non-Gaussian process and measurement noises.

REFERENCES

[1] D. Simon, *Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches*, John Wiley&Sons, 2006.

[2] A. Ferrero, H. V. Jetti, and S. Salicone, "The possibilistic Kalman filter: Definition and comparison with the available methods," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, July 2021.

[3] C. Hajiyev and D. C. Guler, "Review on gyroless attitude determination methods for small satellites," *Progress in Aerospace Sciences*, vol. 90, pp. 54-66, April 2017.

[4] G. Y. Kulikov and M. V. Kulikova, "The accurate continuous-discrete extended Kalman filter for radar tracking," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 948-958, February 2016.

[5] F. Zhao, C. Chen, W. He, and S. S. Ge, "A filtering approach based on MMAE for a SINS/CNS integrated navigation system," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 6, pp. 1113-1120, November 2018.

[6] T. Cui, F. Ding, X. B. Jin, A. Alsaedi, and T. Hayat, "Joint multi-innovation recursive extended least squares parameter and state estimation for a class of state-space systems," *International Journal of Control, Automation, and Systems*, vol. 18, no. 6, pp. 1412-1424, June 2020.

[7] L. Xu, W. L. Xiong, A. Alsaedi, and T. Hayat, "Hierarchical parameter estimation for the frequency response based on the dynamical window data," *International Journal of Control, Automation, and Systems*, vol. 16, no. 4, pp. 1756-1764, August 2018.

[8] J. Pan, X. Jiang, X. K. Wan, and W. F. Ding, "A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems," *International Journal of Control, Automation, and Systems*, vol. 15, no. 3, pp. 1189-1197, June 2017.

[9] J. Pan, W. Li, and H. P. Zhang, "Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control," *International Journal of Control, Automation, and Systems*, vol. 16, no. 6, pp. 2878-2887, December 2018.

[10] M. H. Li and X. M. Liu, "Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique," *International Journal of Control, Automation, and Systems*, vol. 18, no. 6, pp. 1581-1592, June 2020.

[11] F. Ding, L. Lv, J. Pan, X. K. Wan, and X. B. Jin, "Two-stage gradient-based iterative estimation methods for controlled autoregressive systems using the measurement data," *International Journal of Control, Automation, and Systems*, vol. 18, no. 4, pp. 886-896, April 2020.

[12] L. J. Wan, F. Ding, X. M. Liu, and C. P. Chen, "A new iterative least squares parameter estimation approach for equation-error autoregressive systems," *International Journal of Control, Automation, and Systems*, vol. 18, no. 3, pp. 780-790, March 2020.

[13] D. Wang, Z. D. Wang, B. Shen, and F. E. Alsaadi, "Security-guaranteed filtering for discrete-time stochastic delayed systems with randomly occurring sensor saturations and deception attacks," *International Journal of Robust and Nonlinear Control*, vol. 27, no. 7, pp. 1194-1208, May 2017.

[14] J. Y. Mao, D. R. Ding, Y. Song, Y. R. Liu, and F. E. Alsaadi, "Event-based recursive filtering for time-delayed stochastic nonlinear systems with missing measurements," *Signal Processing*, vol. 134, pp. 158-165, May 2017.

[15] X. X. Wang, Y. Liang, Q. Pan, and C. H. Zhao, "Gaussian filter for nonlinear systems with one-step randomly delayed measurements," *Automatica*, vol. 49, no. 4, pp. 976-986, April 2013.

[16] X. Wang, Q. Pan, Y. Liang, and F. Yang, "Gaussian smoothers for nonlinear systems with one-step randomly delayed measurements," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1828-1835, July 2013.

[17] Z. Wang, Y. Huang, Y. Zhang, G. Jia, and J. Chambers, "An improved Kalman filter with adaptive estimate of latency probability," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 10, pp. 2259-2263, November 2020.

[18] A. Vasilijevic, B. Borovic, and Z. Vukic, "Underwater vehicle localization with complementary filter: Performance analysis in the shallow water environment," *Journal of Intelligent & Robotic Systems*, vol. 68, no. 3, pp. 373-386, December 2012.

[19] M. Roth, E. Özkan, and F. Gustafsson, "A Student's t filter for heavy tailed process and measurement noise," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, pp. 5770-5774, May 2013.

[20] Y. Huang, Y. Zhang, Y. Zhao, and J. A. Chambers, "A novel robust Gaussian–Student's *t* mixture distribution based Kalman filter," *IEEE Transactions on Signal Processing*, vol. 67, no. 13, pp. 3606-3620, May 2019.

[21] B. Chen and J. C. Principe, "Maximum correntropy estimation is a smoothed MAP estimation," *IEEE Signal Processing Letters*, vol. 19, no. 8, pp. 491-494, June 2012.

[22] M. V. Kulikova, "Square-root approach for Chandrasekhar-based maximum correntropy Kalman filtering," *IEEE Signal Processing Letters*, vol. 26, no. 12, pp. 1803-1807, October 2019.

[23] X. Liu, Z. G. Ren, H. Q. Lyu, Z. H. Jiang, P. J. Ren, and B. D. Chen, "Linear and nonlinear regression-based maximum correntropy extended Kalman filtering," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 5, pp. 3093-3102, May 2021.

[24] Z. Li and S. Guan, "Diffusion normalized Huber adaptive filtering algorithm," *Journal of the Franklin Institute-Engineering and Applied Mathematics*, vol. 355, no. 8, pp. 3812-3825, May 2018.

[25] Z. B. Qiu, Y. L. Huang, and H. M. Qian, "Adaptive robust nonlinear filtering for spacecraft attitude estimation based on additive quaternion," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 1, pp. 100-108, January 2020.

[26] Y. Huang, Y. Zhang, N. Li, Z. Wu, and J. A. Chambers, "A novel robust Student's t-based Kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1545-1554, January 2017.

[27] Y. Huang, Y. Zhang, B. Xu, Z. Wu, and J. A. Chambers, "A new adaptive extended Kalman filter for cooperative localization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 353-368, September 2018.

[28] Y. Huang, G. Jia, B. Chen, and Y. Zhang, "A new robust Kalman filter with adaptive estimate of time-varying measurement bias," *IEEE Signal Processing Letters*, vol. 27, pp. 700-704, March 2020.

[29] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Transactions on Automatic Control*, vol. 63, no. 2, pp. 594-601, September 2018.

[30] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131-146, December 2008.

Chenghao Shan received his B.E. degree in automation from Harbin University of Science and Technology, Harbin, China, in 2017. He is currently pursuing the Ph.D. degree in control science and engineering with Harbin Engineering University, Harbin, China. Since 2021, he has also been a joint Ph.D. student funded by the China Scholarship Council, at Department of Electrical and Computer Engineering, National University of Singapore (NUS), His research interests include statistical signal processing, variational Bayesian filtering methods, information fusion and their applications in target tracking and integrated navigation.

Weidong Zhou received his B.E. and M.E. degrees in automatic control from Harbin Institute of Technology, in 1988, and Harbin University of Science and Technology, in 1991, respectively. He received a Ph.D. degree in navigation from Harbin Engineering University, China, in 2006. Since 2005, he has been a full professor at Department of Intelligent Systems Science and Engineering, Harbin Engineering University, China. He is an editorial board member of National Marine Boat Standardization Technical Committee, and a member of the Northeast Regional Branch of the Chinese Society of Inertial Technology. His research interests include automatic control theory, integrated navigation, estimation theory, and multi-sensor data fusion.

Hanyu Shan received her B.E. degree in electronic information engineering from Harbin Engineering University, Harbin, China, in 2020. She is currently pursuing a Ph.D. degree in information and communication engineering with Harbin Engineering University, Harbin, China. Her research interests include digital beamforming, array signal processing, and spatial filtering theory.

Lu Liu received her B.E. degree in automation from Henan University, Henan, China, in 2012, and her M.E. and Ph.D. degrees in control science and engineering from Harbin Engineering University, Harbin, China, in 2015 and 2020, respectively. She currently works at Beijing Institute of Control and Electronic Technology. Her research interests include nonlinear state estimation and information fusion and their applications in target tracking.