


Article

A New Variational Bayesian-Based Kalman Filter with Unknown Time-Varying Measurement Loss Probability and Non-Stationary Heavy-Tailed Measurement Noise

Chenghao Shan ¹ , Weidong Zhou ^{1,*}, Yefeng Yang ² and Hanyu Shan ³

¹ Department of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China; shanchenghao123@hrbeu.edu.cn

² Center for Control Theory and Guidance Technology, Harbin Institute of Technology, Harbin 150001, China; 18B904013@stu.hit.edu.cn

³ Department of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China; shy@hrbeu.edu.cn

* Correspondence: zhouweidong@hrbeu.edu.cn; Tel.: +86-0451-8256-8107

Abstract: In this paper, a new variational Bayesian-based Kalman filter (KF) is presented to solve the filtering problem for a linear system with unknown time-varying measurement loss probability (UTVMLP) and non-stationary heavy-tailed measurement noise (NSHTMN). Firstly, the NSHTMN was modelled as a Gaussian-Student's t -mixture distribution via employing a Bernoulli random variable (BM). Secondly, by utilizing another Bernoulli random variable (BL), the form of the likelihood function consisting of two mixture distributions was converted from a weight sum to an exponential product and a new hierarchical Gaussian state-space model was therefore established. Finally, the system state vector, BM, BL, the intermediate random variables, the mixing probability, and the UTVMLP were jointly inferred by employing the variational Bayesian technique. Simulation results revealed that in the scenario of NSHTMN, the proposed filter had a better performance than current algorithms and further improved the estimation accuracy of UTVMLP.

Keywords: variational Bayesian; Kalman filter; measurement loss probability; mixture distribution; non-stationary heavy-tailed measurement noise



Citation: Shan, C.; Zhou, W.; Yang, Y.; Shan, H. A New Variational Bayesian-Based Kalman Filter with Unknown Time-Varying Measurement Loss Probability and Non-Stationary Heavy-Tailed Measurement Noise. *Entropy* **2021**, *23*, 1351. <https://doi.org/10.3390/e23101351>

Academic Editors: Jiayi Ma, Yu Liu, Junjun Jiang, Zheng Wang and Han Xu

Received: 7 September 2021

Accepted: 14 October 2021

Published: 16 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Under the minimal mean square error criteria, the KF is the optimal estimator for the linear Gaussian state-space model [1,2]. KF has been widely employed in a variety of applications [3–5]. Unfortunately, in many practical applications, when the sensor produces intermittent faults, the actual measurement of the sensors may not be accurately represented by the KF measurement model [6,7]. If the random measurement loss occurs, the measurement of the sensors contains only pure noise. In this situation, the estimation accuracy of a typical KF will drop significantly or even diverge. Various filtering methods have been developed to address the measurement loss filtering issue, such as the intermittent KF (IKF) [8,9]. However, IKF has an important assumption: the measurement loss probability is known. In practical applications, the measurement loss probability is usually unknown and the IKF is no longer applicable in this case [7].

In order to address the filtering issues of the unknown measurement loss probability of the linear system, the first Bayesian Kalman filter and the second Bayesian Kalman filter were designed by estimating the process and posterior distribution of the measurement loss, respectively [10]. The above two filters, however, are no longer valid if the unknown measurement loss probability is time-varying. Recently, the variational Bayesian-based adaptive KF (VBAKF) was derived for a linear system with unknown time-varying measurement loss probability (UTVMLP) and both the system vector and UTVMLP are jointly estimated by introducing the variational Bayesian technique [7]. Additionally,

VBAKF shows excellent performance in the context of white Gaussian measurement noise with known statistical characteristics. Unfortunately, in realistic engineering applications, measurement outliers may occur at various periods due to environmental changes and unreliable sensors, resulting in NSHTMN, i.e., when the system runs healthily, the measurement noise is the Gaussian-distributed, and when the time-varying measurement outliers erode the system, the measurement noise is heavy-tail-distributed [11,12]. In the scenario of NSHTMN, the estimation accuracy of VBAKF will drop sharply.

Recently, some mixture distribution-based algorithms have been presented to address NSHTMN, such as the Gaussian-Student's t -mixture distribution-based KF (GSTKF) [13,14]. However, the filtering problem with UTVMLP and NSHTMN cannot directly solved by employing a mixture distribution, that is, under the scenario of UTVMLP and NSHTMN, the current likelihood function is a weighted sum of double-mixture distributions, which is an unclosed and unconjugated distribution that makes the Bayesian inference difficult to employ directly.

In this paper, a new variational Bayesian-based KF is presented to settle the filtering issue for linear discrete-time systems with UTVMLP and NSHTMN. Firstly, the Gaussian-Student's t -mixture distribution with BM is employed to model the NSHTMN. Secondly, the form of the likelihood function is converted to an exponential product and constructs a new hierarchical Gaussian state-space model by utilizing BL. Thirdly, the variational Bayesian method is introduced to simultaneously estimate the system state vector, BM, BL, the intermediate random variables, the mixing probability, and the UTVMLP. Finally, a numerical simulation experiment reveals that the proposed filter has better estimation accuracy but is more time-consuming than existing filtering algorithms in the scenarios of NSHTMN and UTVMLP.

The contributions of this paper are as follows:

- (a) By employing a Bernoulli-distributed variable, the NSHTMN is modelled as a Gaussian-Student's t -mixture distribution;
- (b) The measurement likelihood function is converted from the weight sum of two mixture distributions to an exponential product and a new hierarchical Gaussian state-space model is therefore derived;
- (c) The system state vector, UTVMLP, and the unknown variables are simultaneously estimated by utilizing the variational Bayesian technique;
- (d) Numerical simulation results indicate that the proposed filter has better performance than that of existing algorithms in the scenarios of NSHTMN and UTVMLP

2. Problem Formulation

Consider the linear stochastic system with the following state and measurement equations:

$$x_t = \mathbf{F}_{t-1}x_{t-1} + e_{t-1} \quad (1)$$

$$y_t = b_t \mathbf{H}_t x_t + g_t \quad (2)$$

where $x_t \in \mathbb{R}^m$ denotes the system state vector; $\mathbf{F}_{t-1} \in \mathbb{R}^{m \times m}$ denotes the state transition matrix; $e_t \in \mathbb{R}^m$ represents the Gaussian-distributed white process noise vector with a zero mean value and covariance matrix \mathbf{Q}_t ; $y_t \in \mathbb{R}^n$ represents the measurement vector; $\mathbf{H}_t \in \mathbb{R}^{n \times m}$ is the measurement matrix; $g_t \in \mathbb{R}^n$ is the white NSHTMN vector; and t represents the index of discrete time. The phenomenon of measurement loss is described by introducing the identically distributed and mutually uncorrelated measurement loss defined as the Bernoulli random variable (BL) b_t , which is expressed by the following equations.

$$p(b_t = 1) = \mathbb{E}[b_t] = \gamma_t \quad (3)$$

$$p(b_t = 0) = 1 - \mathbb{E}[b_t] = 1 - \gamma_t \quad (4)$$

where $\gamma_t \in [0, 1]$ denotes the time-varying measurement loss probability. Note that the value of γ_t is unknown in this paper. The initial Gaussian-distributed system state vector

x_0 is the random vector with mean $\hat{x}_{0|0} = 0$ and covariance matrix $\mathbf{P}_{0|0}$. Additionally, it is assumed that the initial system state vector x_0 , the noise vectors e_{t-1} and g_t , and the Bernoulli random variable b_t are mutually independent.

It can be seen from Equations (1)–(4) that the ideal measurement was received by the sensor when $b_t = 0$ and the measurement loss with UTVMLP occurred when $b_t = 1$. Meanwhile, the measurement noise is NSHTMN due to measurement outliers, that is, when the system runs healthily, the measurement noise is Gaussian-distributed, and when measurement outliers erode the system, the measurement noise is heavy-tail-distributed. The NSHTMN and UTVMLP can result in estimation errors or even in filtering divergence. Therefore, a new variational Bayesian-based Kalman filter with NSHTMN and UTVMLP will be proposed.

3. Proposed Variational Bayesian-Based Kalman Filter

In this section, a new variational Bayesian-based Kalman filter is proposed to address the filtering issue for a linear system with NSHTMN and UTVMLP. Firstly, the Gaussian-Student's t -mixture distribution is utilized to model the NSHTMN and the hierarchical form is derived. Secondly, by converting the measurement likelihood function into an exponential multiplication, a new hierarchical Gaussian state-space model is established. Thirdly, by using the variational Bayesian method, the system state and unknown variables are simultaneously estimated. Finally, the required mathematical expectations are given.

3.1. Gaussian-Student's t -Mixture Distribution

The NSHTMN vector can be modeled as the Gaussian-Student's t -mixture distribution by employing another mixing-defined Bernoulli random variable (BM), ζ_t , and the probability density function (PDF), $p(g_t)$ is given as

$$p(g_t) = \sum_{\zeta_t=0}^1 \int [N(g_t; 0, \mathbf{R}_t)]^{\zeta_t} [ST(g_t; 0, \mathbf{R}_t, \mu)]^{(1-\zeta_t)} p(\zeta_t | \varphi_t) p(\varphi_t) d\varphi_t \quad \zeta_t \in \{0, 1\} \quad (5)$$

where $N(x; 0, \Sigma)$ represents the Gaussian PDF with a zero mean vector and covariance matrix Σ , and $ST(x; 0, Y, \omega)$ represents the student's t -PDF with a zero mean vector, covariance matrix Y , and degree of freedom (dof) parameter ω . \mathbf{R}_t represents the covariance matrix of the nominal measurement noise. The PDF of the mixing probability φ_t and the probability mass function (PMF) of ζ_t are defined as follows, respectively.

$$p(\varphi_t) = \text{Be}(\varphi_t; h_0, 1 - h_0) \quad (6)$$

$$p(\zeta_t | \varphi_t) = \varphi_t^{\zeta_t} (1 - \varphi_t)^{(1-\zeta_t)} \quad (7)$$

where $\text{Be}(x; \sigma, \kappa)$ represents the Beta PDF with shape parameters σ and κ .

Due to the hierarchical properties of the student's t -distribution, Equation (5) can be rewritten as such:

$$p(g_t) = \sum_{\zeta_t=0}^1 \int_0^{+\infty} [N(g_t; 0, \mathbf{R}_t)]^{\zeta_t} [N(g_t; 0, \mathbf{R}_t / \beta_t)]^{(1-\zeta_t)} p(\beta_t) p(\zeta_t | \varphi_t) p(\varphi_t) d\varphi_t d\beta_t \quad (8)$$

$$p(\beta_t) = G\left(\beta_t, \frac{\mu}{2}, \frac{\mu}{2}\right) \quad (9)$$

where $G(x, a, b)$ represents the Gamma PDF with shape parameter a and rate parameter b , and β_t represents the intermediate random variable.

3.2. New Hierarchical Gaussian State-Space Model (HGSSM)

According to Equations (2)–(4), the measurement likelihood PDF is derived as Based on Equation (2), the following equation can be obtained.

$$\begin{aligned} p(y_t|x_t, \gamma_t) &= \sum_{b_t=0}^1 p(y_t, b_t|x_t, \gamma_t) \\ &= p(y_t|x_t, b_t=1)p(b_t=1) + p(y_t|x_t, b_t=0)p(b_t=0) \\ &= (1-\gamma_t)p(y_t|x_t, b_t=1) + \gamma_t p(y_t|x_t, b_t=0) \end{aligned} \quad (10)$$

$$p(y_t|x_t, b_t=1) = p_{g_t}(y_t - \mathbf{H}_t x_t) \quad (11)$$

$$p(y_t|x_t, b_t=0) = p_{g_t}(y_t) \quad (12)$$

where $p_{g_t}(\cdot)$ represents the measurement noise PD. Substituting Equations (11) and (12) in Equation (10) results in

$$p(y_t|x_t, \gamma_t) = (1-\gamma_t)p_{g_t}(y_t - \mathbf{H}_t x_t) + \gamma_t p_{g_t}(y_t) \quad (13)$$

Remark 1. The measurement likelihood PDF in Equation (13) is an unclosed and un-conjugated weighted sum form, and it is impossible to infer the system state vector and unknown parameters directly by utilizing the variational Bayesian. The weighted sum will then be converted into an exponential multiplication form to address this problem.

The PMF of BL b_t is given as

$$p(b_t|\gamma_t) = (1-\gamma_t)^{b_t} \gamma_t^{(1-b_t)} \quad (14)$$

Exploiting Equations (13) and (14), the measurement likelihood PDF is reformulated as

$$\begin{aligned} p(y_t|x_t, \gamma_t) &= \sum_{b_t=0}^1 p(y_t|x_t, b_t)p(b_t|\gamma_t) \\ &= \sum_{b_t=0}^1 \left[(1-\gamma_t)^{b_t} [p_{g_t}(y_t - \mathbf{H}_t x_t)]^{b_t} \gamma_t^{(1-b_t)} [p_{g_t}(y_t)]^{(1-b_t)} \right] \\ &= \sum_{b_t=0}^1 \left[[p_{g_t}(y_t - \mathbf{H}_t x_t)]^{b_t} [p_{g_t}(y_t)]^{(1-b_t)} \right] p(b_t|\gamma_t) \end{aligned} \quad (15)$$

According Equation (15), the exponential multiplication-formed likelihood PDF $p(y_t|x_t, b_t)$ is given as follows.

$$p(y_t|x_t, b_t) = [p_{g_t}(y_t - \mathbf{H}_t x_t)]^{b_t} [p_{g_t}(y_t)]^{(1-b_t)} \quad (16)$$

Remark 2. The variational Bayesian method must select the suitable conjugate-prior distributions for unknown variables. Therefore, the appropriate prior PDFs to construct a new HGSSM are selected.

The one-step predicted PDF $p(x_t|y_{1:t-1})$ of system state vector x_t is assumed as being Gaussian distributed as follows.

$$p(x_t|y_{1:t-1}) = \mathbf{N}(x_t; \hat{x}_{t|t-1}, \mathbf{P}_{t|t-1}) \quad (17)$$

where $\hat{x}_{t|t-1}$ represents the mean vector and $\mathbf{P}_{t|t-1}$ represents the covariance matrix. Both $\hat{x}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ can be updated by the typical Kalman filter, which is given as

$$\hat{x}_{t|t-1} = \mathbf{F}_{t-1} \hat{x}_{t-1|t-1} \quad (18)$$

$$\mathbf{P}_{t|t-1} = \mathbf{H}_{t-1} \mathbf{P}_{t-1|t-1} \mathbf{H}_{t-1}^T + \mathbf{Q}_{t-1} \quad (19)$$

In employing Equations (8), (9) and (16), the conditional likelihood PDF $p(y_t|x_t, b_t, \beta, \zeta_t)$ is derived as

$$\begin{aligned} p(y_t|x_t, b_t, \beta_t, \zeta_t) &= \left[N(y_t; \mathbf{H}_t x_t, \mathbf{R}_t)^{\zeta_t} N(y_t; \mathbf{H}_t x_t, \mathbf{R}_t / \beta_t)^{(1-\zeta_t)} \right]^{b_t} \\ &\times \left[N(y_t; 0, \mathbf{R}_t)^{\zeta_t} N(y_t; 0, \mathbf{R}_t / \beta_t)^{(1-\zeta_t)} \right]^{(1-b_t)} \\ &= N(y_t; \mathbf{H}_t x_t, \mathbf{R}_t)^{\zeta_t b_t} N(y_t; \mathbf{H}_t x_t, \mathbf{R}_t / \beta_t)^{(1-\zeta_t) b_t} \\ &\times N(y_t; 0, \mathbf{R}_t)^{\zeta_t (1-b_t)} N(y_t; 0, \mathbf{R}_t / \beta_t)^{(1-\zeta_t)(1-b_t)} \end{aligned} \quad (20)$$

It can be seen from Equations (6)–(9), (13) and (20) that the measurement vector y_t depends on system state vector x_t , intermediate random variable β_t , BM ζ_t , BL b_t , mixing probability φ_t , and measurement loss probability γ_t . The following joint-prior PDF must be calculated, i.e.,

$$\begin{aligned} p(\Xi|y_{1:t-1}) &= p(x_t|y_{1:t-1})p(\beta_t)p(b_t|\gamma_t)p(\zeta_t|\varphi_t)p(\varphi_t)p(\gamma_t|y_{1:t-1}), \\ \Xi &\triangleq \{x_t, \beta_t, b_t, \zeta_t, \gamma_t, \varphi\} \end{aligned} \quad (21)$$

where the definitions of $p(\gamma_t)$, $p(\zeta_t|\varphi)$, $p(\beta)$, $p(b_t|\gamma_t)$, and $p(x_t|y_{1:t-1})$ are given in Equations (6), (7), (9), (14) and (17), respectively. Additionally, $p(\gamma_t|y_{1:t-1})$ denotes the prior PDF of the time-varying measurement loss probability, which can be assumed as the following Beta distribution.

$$\begin{aligned} p(\gamma_t|y_{1:t-1}) &= \text{Be}(\gamma_t; \hat{\eta}_{t|t-1}, \hat{\delta}_{t|t-1}) \\ &= \int p(\gamma_{t-1}|y_{1:t-1})p(\gamma_t|\gamma_{t-1})d\gamma_{t-1} \text{ (Bayes' theorem)} \end{aligned} \quad (22)$$

where the shape parameters $\hat{\eta}_{t|t-1}$ and $\hat{\delta}_{t|t-1}$ can be calculated by introducing the forgetting factor $\rho \in (0, 1]$ as follows.

$$\hat{\eta}_{t|t-1} = \rho \hat{\eta}_{t-1|t-1} \quad (23)$$

$$\hat{\delta}_{t|t-1} = \rho \hat{\delta}_{t-1|t-1} \quad (24)$$

where $\hat{\eta}_{t-1|t-1}$ and $\hat{\delta}_{t-1|t-1}$ represent posterior shape parameters.

The proposed new HGSSM is comprised of Equations (14) and (17)–(24). System state vector x_t , intermediate random variable β_t , BM ζ_t , BL b_t , mixing probability φ_t , and measurement loss probability γ_t will be simultaneously estimated by utilizing the variational Bayesian method.

3.3. Variational Bayesian Approximation of the Joint Posterior PDFs

Aiming at the estimation of the unknown variables of the new HGSSM, the joint posterior PDF $p(\Xi|y_t)$ with $\Xi \triangleq \{x_t, \beta_t, b_t, \zeta_t, \varphi_t, \gamma_t\}$ is required to be solved. However, the analytical solution of $p(\Xi|y_t)$ is not accessible. The variational Bayesian approach is therefore employed to determine an approximate PDF for $p(\Xi|y_t)$ and to compute an approximate solution [15–17], i.e.,

$$p(\Xi|y_{1:t}) \approx q_a(x_t)q_a(\beta_t)q_a(b_t)q_a(\zeta_t)q_a(\varphi_t)q_a(\gamma_t) \quad (25)$$

where θ represents an arbitrary element of Ξ and $q_a(\theta)$ denotes the approximate PDF or PMF. By minimizing the Kullback–Leibler divergence (KLD) between $p(\Xi|y_{1:t})$ and $q_a(x_t)q_a(\beta_t)q_a(b_t)q_a(\zeta_t)q_a(\varphi_t)q_a(\gamma_t)$, $q_a(\theta)$ can be obtained as follows.

$$\{q_a(x_t), q_a(\beta_t), q_a(b_t), q_a(\zeta_t), q_a(\varphi_t), q_a(\gamma_t)\} = \underset{(q_a(x_t)q_a(\beta_t)q_a(b_t)q_a(\zeta_t)q_a(\varphi_t)q_a(\gamma_t))}{\text{argminKLD}} \quad (26)$$

$$\text{KLD}(q_a(A)||p(A)) \triangleq \int q_a(A) \log \frac{q_a(A)}{p(A)} dA \quad (27)$$

where $\text{KLD}(q_a(A)||p(A))$ represents the KLD between $q_a(A)$ and $p(A)$, and the optimal solution of Equation (26) can be calculated via the following formula [15,17].

$$\log q_a(\theta) = E_{\Xi-\theta}[\log p(\Xi, y_{1:t})] + c_\theta \quad (28)$$

where $E(\theta)$ denotes the mathematical expectation operation, $\Xi^{-\theta}$ signifies a grouping of all the components in Ξ apart from θ , and the constant with regard to θ is denoted by c_θ . Additionally, the fixed-point iteration technique is utilized to derive the approximate formation of $q_a(\theta)$ due to the fact that estimated parameters are mutually coupled.

The joint PDF $p(\Xi, y_{1:t})$ in Equation (26) can be derived as

$$\begin{aligned} p(\Xi, y_{1:t}) &= N(y_t; \mathbf{H}_t x_t, \mathbf{R}_t)^{\zeta_t b_t} N(y_t; \mathbf{H}_t x_t, \mathbf{R}_t / \beta_t)^{(1-\zeta_t)b_t} \\ &\times N(y_t; 0, \mathbf{R}_t)^{\zeta_t(1-b_t)} N(y_t; 0, \mathbf{R}_t / \beta_t)^{(1-\zeta_t)(1-b_t)} N(x_t; \hat{x}_{t|t-1}, \mathbf{P}_{t|t-1}) \\ &\times G(\beta_t, \frac{\mu_t}{2}, \frac{\mu_t}{2}) \varphi_t^{\zeta_t} (1 - \varphi_t)^{(1-\zeta_t)} (1 - \gamma_t)^{b_t} \gamma_t^{(1-b_t)} \\ &\times \text{Be}(\varphi_t; h_0, 1 - h_0) \text{Be}(\gamma_t; \hat{\eta}_{t|t-1}, \hat{\delta}_{t|t-1}) p(y_{1:t-1}) \end{aligned} \quad (29)$$

Proposition 1. Let $\theta = x_t$ and by using Equation (29) in (28), $q_a^{(s+1)}(x_t)$ can be updated as Gaussian, i.e.,

$$q_a^{(s+1)}(x_t) = N(x_t; \hat{x}_{t|t}^{(s+1)}, \mathbf{P}_{t|t}^{(s+1)}) \quad (30)$$

where $q_a^{(s+1)}(\cdot)$ represents the approximate PDF in the $(s+1)$ th iteration, while the mean vector $\hat{x}_{t|t}^{(s+1)}$ and the covariance matrix $\mathbf{P}_{t|t}^{(s+1)}$ are assumed to be updated in accordance with the traditional Kalman filter as follows.

$$\hat{x}_{t|t}^{(s+1)} = \hat{x}_{t|t-1} + \mathbf{K}_t^{(s+1)} (y_t - \mathbf{H}_t \hat{x}_{t|t-1}) \quad (31)$$

$$\mathbf{P}_{t|t}^{(s+1)} = \hat{\mathbf{P}}_{t|t-1}^{(s+1)} - \mathbf{K}_t^{(s+1)} \mathbf{H}_t \hat{\mathbf{P}}_{t|t-1}^{(s+1)} \quad (32)$$

$$\mathbf{K}_t^{(s+1)} = \hat{\mathbf{P}}_{t|t-1}^{(s+1)} \mathbf{H}_t^T \left(\mathbf{H}_t \hat{\mathbf{P}}_{t|t-1}^{(s+1)} \mathbf{H}_t^T + \tilde{\mathbf{R}}_t^{(s+1)} \right)^{-1} \quad (33)$$

where $\mathbf{K}_t^{(s+1)}$ represents the Kalman gain matrix. The modified measurement noise covariance matrix at $(s+1)$ th iteration $\tilde{\mathbf{R}}_t^{(s+1)}$ is formulated as

$$\tilde{\mathbf{R}}_t^{(s+1)} = \frac{\mathbf{R}_t}{E^{(s)}[\zeta_t] E^{(s)}[b_t] + E^{(s)}[1 - \zeta_t] E^{(s)}[b_t] E^{(s)}[\beta_t]} \quad (34)$$

where $E^{(s)}[\cdot]$ represents the mathematical expectation of variables in the s th iteration.

Proof: see Appendix A.

Proposition 2. Let $\theta = \beta_t$ and by using Equation (29) in Equation (28), $q_a^{(s+1)}(\beta_t)$ can be updated as Gamma, i.e.,

$$q_a^{(s+1)}(\beta_t) = G(\beta_t; \pi_t^{(s+1)}, \nu_t^{(s+1)}) \quad (35)$$

where the shape parameter $\pi_t^{(s+1)}$ and rate parameter $\nu_t^{(s+1)}$ are formulated as

$$\pi_t^{(s+1)} = 0.5 \left(n E^{(s+1)}[1 - \zeta_t] + \mu_t \right) \quad (36)$$

$$\nu_t^{(s+1)} = 0.5 \left[\text{tr} \left(\mathbf{G}_t^{(s+1)} \mathbf{R}_t^{-1} \right) + \mu_t \right] \quad (37)$$

where n represents the dimension of the measurement vector, $\text{tr}(\cdot)$ represents the trace operation, and $\mathbf{G}_t^{(s+1)}$ is defined as

$$\begin{aligned} \mathbf{G}_t^{(s+1)} &= \mathbb{E}^{(s)}[1 - \zeta_t] \left\{ \mathbb{E}^{(s)}[b_t] \left[\mathbf{H}_t \mathbf{P}_{t|t}^{(s+1)} \mathbf{H}_t^T + \left(y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t}^{(s+1)} \right) \left(y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t}^{(s+1)} \right)^T \right] \right. \\ &\quad \left. + \mathbb{E}^{(s)}[1 - \zeta_t] \mathbb{E}^{(s)}[1 - b_t] y_t (y_t)^T \right\} \end{aligned} \quad (38)$$

Proof: see Appendix B.

Proposition 3. Let $\theta = b_t$ and by using Equation (29) in Equation (28), $q_a^{(s+1)}(b_t)$ can be updated as the Bernoulli distribution. The posterior probabilities $p(b_t = 1)$ and $p(b_t = 0)$ of BL b_t are given as

$$p^{(s+1)}(b_t = 1) \Delta^{(s+1)} \exp \left\{ \mathbb{E}^{(s)}[\log(1 - \gamma_t)] - 0.5 \text{tr} \left(\mathbf{C}_t^{(s+1)} \mathbf{R}_t^{-1} \right) + 0.5 n \mathbb{E}^{(s)}[1 - \zeta_t] \mathbb{E}^{(s)}[\log(\beta_t)] \right\} \quad (39)$$

$$p^{(s+1)}(b_t = 0) = \Delta^{(s+1)} \exp \left\{ \mathbb{E}^{(s)}[\log(\gamma_t)] - 0.5 \text{tr} \left(\mathbf{D}_t^{(s+1)} \mathbf{R}_t^{-1} \right) + 0.5 n \mathbb{E}^{(s)}[1 - \zeta_t] \mathbb{E}^{(s)}[\log(\beta_t)] \right\} \quad (40)$$

where $\Delta^{(s+1)}$ represents the normalizing constant and the parameters $\mathbf{C}_t^{(s+1)}$ and $\mathbf{D}_t^{(s+1)}$ are, respectively, defined as

$$\mathbf{C}_t^{(s+1)} = \left\{ \mathbb{E}^{(s)}[\zeta_t] + \mathbb{E}^{(s)}[\beta_t] \mathbb{E}^{(s)}[1 - \zeta_t] \right\} \left[\mathbf{H}_t \mathbf{P}_{t|t}^{(s+1)} \mathbf{H}_t^T + \left(y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t}^{(s+1)} \right) \left(y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t}^{(s+1)} \right)^T \right] \quad (41)$$

$$\mathbf{D}_t^{(s+1)} = \left\{ \mathbb{E}^{(s)}[\zeta_t] + \mathbb{E}^{(s)}[\beta_t] \mathbb{E}^{(s)}[1 - \zeta_t] \right\} y_t y_t^T \quad (42)$$

Proof: see Appendix C.

Proposition 4. Let $\theta = \zeta_t$ and by using Equation (29) in (28), $q_a^{(s+1)}(\zeta_t)$ can be also updated as the Bernoulli distribution. The posterior probabilities $p(\zeta_t = 1)$ and $p(\zeta_t = 0)$ of BM ζ_t are given as

$$p^{(s+1)}(\zeta_t = 1) = \nabla^{(s+1)} \exp \left\{ \mathbb{E}^{(s)}[\log(\varphi_t)] - 0.5 \text{tr} \left(\mathbf{V}_t^{(s+1)} \mathbf{R}_t^{-1} \right) \right\} \quad (43)$$

$$p^{(s+1)}(\zeta_t = 0) = \nabla^{(s+1)} \exp \left\{ \mathbb{E}^{(s)}[\log(1 - \varphi_t)] - 0.5 \text{tr} \left(\mathbf{W}_t^{(s+1)} \mathbf{R}_t^{-1} \right) + 0.5 n \mathbb{E}^{(s)}[\log(\beta_t)] \right\} \quad (44)$$

where $\nabla^{(s+1)}$ also represents the normalizing constant and the definitions of parameters $\mathbf{V}_t^{(s+1)}$ and $\mathbf{W}_t^{(s+1)}$ are, respectively, given as

$$\begin{aligned} \mathbf{V}_t^{(s+1)} &= \mathbb{E}^{(s)}[b_t] \left[\mathbf{H}_t \mathbf{P}_{t|t}^{(s+1)} \mathbf{H}_t^T + \left(y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t}^{(s+1)} \right) \left(y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t}^{(s+1)} \right)^T \right] \\ &\quad + \mathbb{E}^{(s)}[1 - b_t] y_t y_t^T \end{aligned} \quad (45)$$

$$\begin{aligned} \mathbf{W}_t^{(s+1)} &= \mathbb{E}^{(s)}[b_t] \mathbb{E}^{(s)}[\beta_t] \left[\mathbf{H}_t \mathbf{P}_{t|t}^{(s+1)} \mathbf{H}_t^T + \left(y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t}^{(s+1)} \right) \left(y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t}^{(s+1)} \right)^T \right] \\ &\quad + \mathbb{E}^{(s)}[1 - b_t] \mathbb{E}^{(s)}[\beta_t] y_t y_t^T \end{aligned} \quad (46)$$

Proof: see Appendix D.

Proposition 5. Let $\theta = \varphi_t$ and by using Equation (29) in (28), $q_a^{(s+1)}(\varphi_t)$ can be updated as the Beta distribution, i.e.,

$$q_a^{(s+1)}(\varphi_t) = \text{Be} \left(\varphi_t; h_t^{(s+1)}, d_t^{(s+1)} \right) \quad (47)$$

where the shape parameters $h_t^{(s+1)}$ and $d_t^{(s+1)}$ are formulated as

$$h_t^{(s+1)} = E^{(s+1)}[\zeta_t] + h_0 \quad (48)$$

$$d_t^{(s+1)} = E^{(s+1)}[1 - \zeta_t] + 1 - h_0 \quad (49)$$

Proof: see Appendix E.

Proposition 6. Let $\theta = \gamma_t$ and by using Equation (29) in Equation (28), $q_a^{(s+1)}(\gamma_t)$ can be also updated as the Beta distribution, i.e.,

$$q_a^{(s+1)}(\gamma_t) = \text{Be}\left(\gamma_t; \hat{\eta}_t^{(s+1)}, \hat{\delta}_t^{(s+1)}\right) \quad (50)$$

where the definitions of shape parameters $\hat{\eta}_t^{(s+1)}$ and $\hat{\delta}_t^{(s+1)}$ are given as

$$\hat{\eta}_t^{(s+1)} = E^{(s+1)}[1 - b_t] + \hat{\eta}_{t|t-1} \quad (51)$$

$$\hat{\delta}_t^{(s+1)} = E^{(s+1)}[b_t] + \hat{\delta}_{t|t-1} \quad (52)$$

Proof: see Appendix F.

3.4. Calculation of the Required Mathematical Expectations

The required mathematical expectations $E^{(s+1)}[\beta_t]$, $E^{(s+1)}[\log(\beta_t)]$, $E^{(s+1)}[b_t]$, $E^{(s+1)}[1 - b_t]$, $E^{(s+1)}[\zeta_t]$, $E^{(s+1)}[1 - \zeta_t]$, $E^{(s+1)}[\log(\varphi_t)]$, $E^{(s+1)}[\log(1 - \varphi_t)]$, $E^{(s+1)}[\log(\gamma_t)]$, and $E^{(s+1)}[\log(1 - \gamma_t)]$ in Section 3.3 are defined, respectively, as follows:

$$E^{(s+1)}[\beta_t] = \frac{\pi_t^{(s+1)}}{\nu_t^{(s+1)}} \quad (53)$$

$$E^{(s+1)}[\log(\beta_t)] = -\Psi\left(\pi_t^{(s+1)}\right) - \log\left(\nu_t^{(s+1)}\right) \quad (54)$$

$$E^{(s+1)}[b_t] = \frac{p^{(s+1)}(b_t = 1)}{p^{(s+1)}(b_t = 1) + p^{(s+1)}(b_t = 0)} \quad (55)$$

$$E^{(s+1)}[1 - b_t] = 1 - E^{(s+1)}[b_t] \quad (56)$$

$$E^{(s+1)}[\zeta_t] = \frac{p^{(s+1)}(\zeta_t = 1)}{p^{(s+1)}(\zeta_t = 1) + p^{(s+1)}(\zeta_t = 0)} \quad (57)$$

$$E^{(s+1)}[1 - \zeta_t] = 1 - E^{(s+1)}[\zeta_t] \quad (58)$$

$$E^{(s+1)}[\log(\varphi_t)] = \Psi\left(h_t^{(s+1)}\right) - \Psi\left(h_t^{(s+1)} + d_t^{(s+1)}\right) \quad (59)$$

$$E^{(s+1)}[\log(1 - \varphi_t)] = \Psi\left(d_t^{(s+1)}\right) - \Psi\left(h_t^{(s+1)} + d_t^{(s+1)}\right) \quad (60)$$

$$E^{(s+1)}[\log(\gamma_t)] = \Psi\left(\eta_t^{(s+1)}\right) - \Psi\left(\eta_t^{(s+1)} + \delta_t^{(s+1)}\right) \quad (61)$$

$$E^{(s+1)}[\log(1 - \gamma_t)] = \Psi\left(\delta_t^{(s+1)}\right) - \Psi\left(\eta_t^{(s+1)} + \delta_t^{(s+1)}\right) \quad (62)$$

where $\Psi(\cdot)$ represents the digamma function [18].

The presented variational Bayesian-based Kalman filter with UTVMLP and NSHTMN consists of Equations (18), (19) and (30)–(62). Table 1 describes the implementation of the proposed new KF.

Table 1. The proposed variational Bayesian-based Kalman filter with UTVMLP and NSHTMN (one-time step).

Inputs: $\hat{x}_{t-1 t-1}, \mathbf{P}_{t-1 t-1}, \mathbf{Q}_{t-1 t-1}, \mathbf{R}_{t-1 t-1}, y_t, \mathbf{F}_{t-1}, \mathbf{H}_t, n, m, \mu_t, h_0, \hat{\eta}_{t-1}, \hat{\delta}_{t-1}, N_I, \varsigma$
Time update:
1. Obtain $\hat{x}_{t t-1}$ and $\mathbf{P}_{t t-1}$ utilizing Equations (18) and (19) (time update of typical Kalman filter).
Variational measurement update:
2. Initialization: $\hat{x}_{t t}^{(0)} = x_{t t-1}, \mathbf{P}_{t t}^{(0)} = \mathbf{P}_{t t-1}, E^{(0)}[\beta_t] = 1, E^{(0)}[\log(\beta_t)] = 0, E^{(0)}[b_t] = \hat{\eta}_{t-1}/\hat{\delta}_{t-1},$ $E^{(0)}[1 - b_t] = 1 - E^{(0)}[b_t], \hat{\eta}_t^{(0)} = \hat{\eta}_{t-1}, \hat{\delta}_t^{(0)} = \hat{\delta}_{t-1}, E^{(0)}[\zeta_t] = 1, E^{(0)}[\log(\zeta_t)] = 1 - E^{(0)}[\zeta_t],$ $E^{(0)}[\log(\varphi_t)] = \Psi(h_0) - \Psi(1), E^{(s+1)}[\log(1 - \varphi_t)] = \Psi(1 - h_0) - \Psi(1),$ $E^{(s+1)}[\log(\gamma_t)] = \Psi(\hat{\eta}_t^{(0)}) - \Psi(\hat{\eta}_t^{(0)} + \hat{\delta}_t^{(0)}), E^{(s+1)}[\log(1 - \gamma_t)] = \Psi(\hat{\delta}_t^{(0)}) - \Psi(\hat{\eta}_t^{(0)} + \hat{\delta}_t^{(0)})$ for $s = 0 : N_I - 1.$
3. Update $q_a^{(s+1)}(x_t)$ by Equation (30).
4. Obtain $\hat{x}_{t t}^{(s+1)}, \mathbf{P}_{t t}^{(s+1)},$ and $\tilde{\mathbf{R}}_t^{(s+1)}$ by utilizing Equations (31)–(34) (typical Kalman filter).
5. Update the Gamma-distributed $q_a^{(s+1)}(\beta_t)$ by Equation (35).
6. Obtain $\pi_t^{(s+1)}, \nu_t^{(s+1)},$ and $\mathbf{G}_t^{(s+1)}$ by utilizing Equations (36)–(38).
7. Update the Bernoulli-distributed $q_a^{(s+1)}(b_t)$.
8. Obtain $p^{(s+1)}(b_t = 1), p^{(s+1)}(b_t = 0), \mathbf{C}_t^{(s+1)},$ and $\mathbf{D}_t^{(s+1)}$ by utilizing Equations (39)–(42).
9. Obtain $E^{(s+1)}[b_t]$ and $E^{(s+1)}[1 - b_t]$ by utilizing Equations (55) and (56).
10. Update the Bernoulli-distributed $q_a^{(s+1)}(\zeta_t)$.
11. Obtain $p^{(s+1)}(\zeta_t = 1), p^{(s+1)}(\zeta_t = 0), \mathbf{V}_t^{(s+1)},$ and $\mathbf{W}_t^{(s+1)}$ by utilizing Equations (43)–(46).
12. Obtain $E^{(s+1)}[\zeta_t]$ and $E^{(s+1)}[1 - \zeta_t]$ by utilizing Equations (57) and (58).
13. Update the Beta-distributed $q_a^{(s+1)}(\varphi_t)$ by Equation (47).
14. Obtain $h_t^{(s+1)}$ and $d_t^{(s+1)}$ by utilizing Equations (48) and (49).
15. Obtain $E^{(s+1)}[\log(\varphi_t)]$ and $E^{(s+1)}[\log(1 - \varphi_t)]$ by utilizing Equations (59) and (60).
16. Update the Beta-distributed $q_a^{(s+1)}(\gamma_t)$ by Equation (50).
17. Obtain $\hat{\eta}_t^{(s+1)}$ and $\hat{\delta}_t^{(s+1)}$ by utilizing Equations (51) and (52).
18. Obtain $E^{(s+1)}[\log(\gamma_t)]$ and $E^{(s+1)}[\log(1 - \gamma_t)]$ by utilizing Equations (61) and (62).
19. If $(\ \hat{x}_{t t}^{(s+1)} - \hat{x}_{t t}^{(s)}\ / \ \hat{x}_{t t}^{(s)}\) \leq \varsigma$, the iteration stopped.
End for:
20. $\hat{x}_{t t} = \hat{x}_{t t}^{(s)}, \mathbf{P}_{t t} = \mathbf{P}_{t t}^{(s)}, h_t = h_t^{(s)}, d_t = d_t^{(s)}, \eta_t = \eta_t^{(s)}, \delta_t = \delta_t^{(s)}$
Outputs: $\hat{x}_{t t}, \mathbf{P}_{t t}, h_t, d_t, \eta_t, \delta_t, h_t/(h_t + d_t), \eta_t/(\eta_t + \delta_t)$

4. Simulations

Aimed at demonstrating the superiority of the presented filter in the scenario with UTVMLP and NSHTMN, a numerical example is simulated. The process and measurement equations of the stochastic system are, respectively, given as [7]

$$x_t = \begin{bmatrix} 0.6 & 0.4 \\ 0.1 & 0.9 \end{bmatrix} x_{t-1} + e_{t-1} \quad (63)$$

$$y_t = b_t \begin{bmatrix} 1 & -2 \end{bmatrix} x_t + g_t \quad (64)$$

where the Gaussian process noise e_{t-1} and the NSHTMN g_t are given as [12]

$$e_{t-1} \sim N(0, \mathbf{Q}_t) \quad (65)$$

$$\left\{ \begin{array}{l} g_t \sim N(0, \mathbf{R}_t) \quad t \in [1, 100] \text{ (Gaussian)} \\ g_t \sim \begin{cases} N(0, \mathbf{R}_t) \text{ w.p.} = 0.98 \\ N(0, 500\mathbf{R}_t) \text{ w.p.} = 0.02 \end{cases} \quad t \in [101, 200] \text{ (slightly heavy-tailed)} \\ g_t \sim \begin{cases} N(0, \mathbf{R}_t) \text{ w.p.} = 0.95 \\ N(0, 500\mathbf{R}_t) \text{ w.p.} = 0.05 \end{cases} \quad t \in [201, 300] \text{ (moderately heavy-tailed)} \\ g_t \sim N(0, \mathbf{R}_t) \quad t \in [301, 400] \text{ (Gaussian)} \end{array} \right. \quad (66)$$

where w.p. represents “with probability”. The true process noise covariance matrix \mathbf{Q}_t with parameter $M = 1$ and the nominal measurement noise covariance matrix \mathbf{R}_t with parameter $N = 150 \text{ m}^2$ are set as

$$\mathbf{Q}_t = M \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad (67)$$

$$\mathbf{R}_t = N \quad (68)$$

The real UTVMLP is set as

$$p(\gamma_t) = \begin{cases} 0.1 & t \in [1, 100] \\ 0.15 & t \in [101, 200] \\ 0.3 & t \in [201, 300] \\ 0.1 & t \in [301, 400] \end{cases} \quad (69)$$

From Equations (66)–(69), it can be seen that the measurement noise and UTVMLP are divided into four stages, as shown in Table 2. The remaining system parameters are as follows: the sampling interval $\Delta k = 0.01 \text{ s}$ and the total simulation time $T = 400 \text{ s}$. The proposed filter is compared with the typical Kalman filter (KF) [2]; the existing variational Bayesian-based adaptive KF with UTVMLP (VBAKF) [7]; the existing Gaussian-Student's t -mixture distribution-based KF (GSTKF) with Gaussian process noise [14]; and the existing IKF with known real measurement loss probability [8]. The parameters of VBAKF are selected as $p_0 = 0.5$, $\alpha_0 = 5$, $\beta_0 = 5$, $\rho = 1 - \exp(-5)$, and $N_m = 10$. The parameters of GSTKF are selected as $v_t = 5$ and $e_0 = 0.85$. The parameters of the proposed filter are given as $\rho = 0.99$, $\mu = 5$, $h_0 = 0.85$, $N_I = 10$, $\eta_0 = 5$, and $\delta_t = 5$, $\zeta = 10^{-16}$. All filters are programmed with MATLAB R2018a and run on a computer with Intel Core i5-6300HQ CPU at 2.30 GHz and 8 GB of RAM.

Table 2. The stages of measurement and the corresponding measurement noise and UTVMLP.

Measurement Stage	Measurement Noise	UTVMLP
Stage 1, time 1 s~100s	$g_t \sim N(0, \mathbf{R}_t)$ (Gaussian)	0.1 (slight loss)
Stage 2, time 101 s~200 s	$g_t \sim \begin{cases} N(0, \mathbf{R}_t) \text{ w.p.} = 0.98 \\ N(0, 500\mathbf{R}_t) \text{ w.p.} = 0.02 \end{cases}$ (slightly heavy-tailed)	0.15 (slight loss)
Stage 3, time 201 s~300 s	$g_t \sim \begin{cases} N(0, \mathbf{R}_t) \text{ w.p.} = 0.95 \\ N(0, 500\mathbf{R}_t) \text{ w.p.} = 0.05 \end{cases}$ (moderately heavy-tailed)	0.3 (moderate loss)
Stage 4, time 301 s~400 s	$g_t \sim N(0, \mathbf{R}_t)$ (Gaussian)	0.1 (slight loss)

Aimed at evaluating the performances in the estimation of the system state vector of all the algorithms, the root-mean square error (RMSE) and the averaged root-mean square error (AGRMSE) are utilized as performance indicators. The definitions of RMSE and AGRMSE of the system state are given as

$$\text{RMSE}_x = \sqrt{\frac{1}{M_c} \sum_{r=1}^{M_r} \left((x_t^r - \hat{x}_t^r)^2 + (y_t^r - \hat{y}_t^r)^2 \right)} \quad (70)$$

$$\text{AGRMSE}_x = \sqrt{\frac{1}{M_c T} \sum_{t=1}^T \sum_{r=1}^{M_r} \left((x_t^r - \hat{x}_t^r)^2 + (y_t^r - \hat{y}_t^r)^2 \right)} \quad (71)$$

where (x_t^r, y_t^r) and $(\hat{x}_t^r, \hat{y}_t^r)$ denote the actual and estimated system state at the j th Monte Carlo run and discrete-time t , respectively. $M_c = 500$ represents the total Monte Carlo run time.

Different from the proposed algorithm and VBAKF, the KF, IKF and GSTKF do not estimate UTVMLP. Although IKF can also address the filtering problem with measurement loss, IKF is based on the assumption that the measurement loss probability is known. Therefore, only VBAKF and the proposed algorithm participate in the comparison of the UTVMLP estimation performance.

Figure 1 shows the $RMSE_{x_s}$ of the proposed filter and the existing filters over 500 times of the Monte Carlo run. Additionally, the $AGRMSE_{x_s}$ and SSRTs of different filters are listed in Table 2. It can be seen from Figure 1 and Table 3 that in the contexts of UTVMLP and NSHTMN, when the measurement is the Gaussian measurement noise and there is slight loss probability, as shown in stages 1 and 4, the estimation accuracy of the proposed filter is close to the IKF with true loss probability and the performance of the proposed algorithm is better than the other algorithms. We can also find that the proposed algorithm still has better performance than the existing algorithms when the measurement has heavy-tailed measurement noise and larger measurement loss probability, as shown in stages 2 and 4. In addition, the proposed algorithm has longer SSRT and higher computational complexity than the existing filters, which can be observed from Table 3.

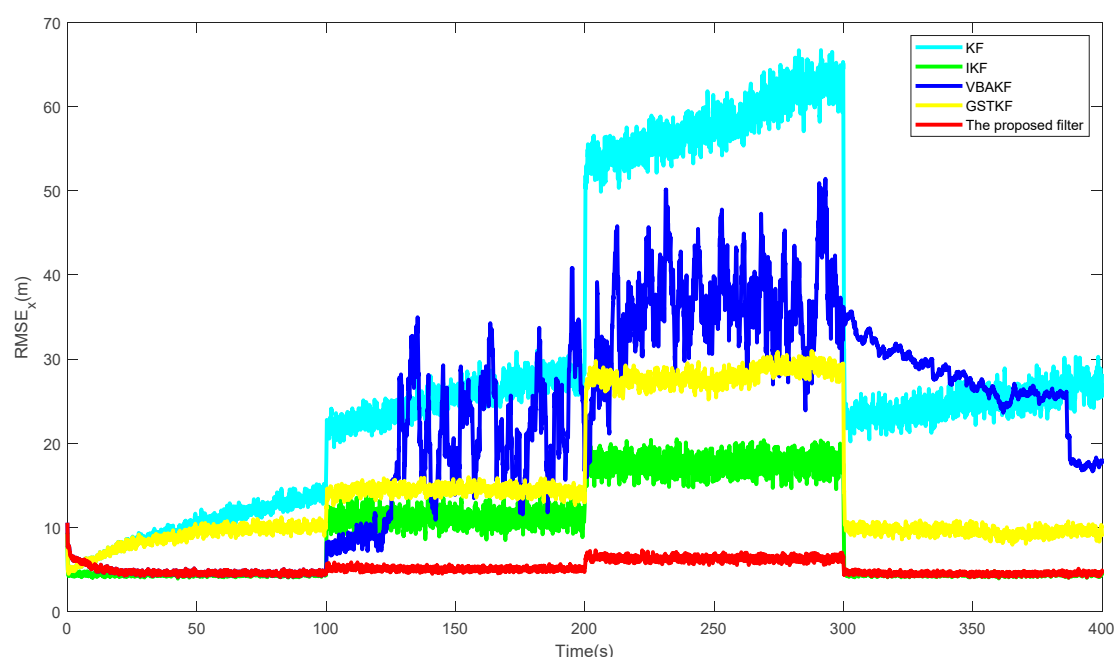


Figure 1. $RMSE_x$ of different filters.

Table 3. $AGRMSE_{x_s}$ and single-step running times (SSRT) of different filters.

Filters	KF	IKF	VBAKF	GSTKF	The Proposed Filter
$AGRMSE_x$ in Stage 1	10.1336	4.4326	4.7705	8.7683	4.7704
$AGRMSE_x$ in Stage 2	25.4009	11.1049	18.7603	14.4343	5.0660
$AGRMSE_x$ in Stage 3	58.0561	17.5274	35.8194	28.0354	6.3044
$AGRMSE_x$ in Stage 4	24.9129	4.4448	27.0252	9.5892	4.5081
$AGRMSE_x$ in all stages	29.6259	9.3774	21.5944	15.2068	5.1623
SSRT (ms)	0.0276	0.0925	0.1166	0.1491	0.2989

Figure 2 shows the curves of the true and estimated UTVMLPs of VBAKF and the proposed filter over 500 times of the Monte Carlo run. Obviously, the NSHTMN has a great influence on the filtering performance of VBAKF and the proposed filter has better UTVMLP estimation accuracy than VBAKF in the scenario of NSHTMN.

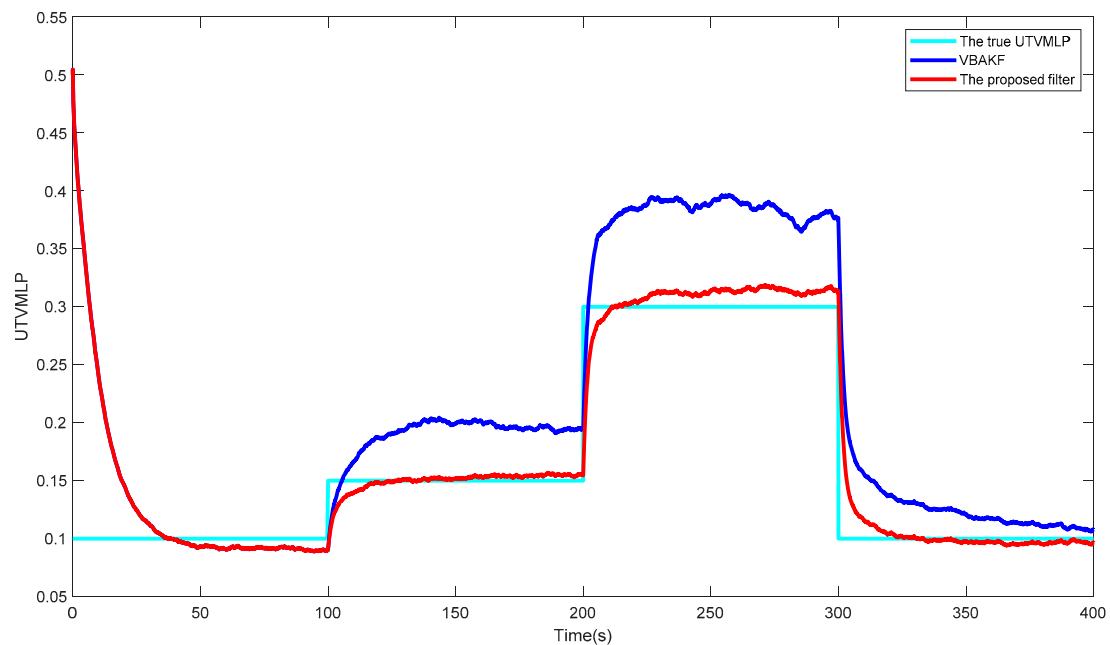


Figure 2. The true and estimated UTVMLPs.

Figures 3 and 4 show the $RMSE_x$ s and the estimated UTVMLPs of the proposed filter with shape parameter $\mu = 3, 4, 5, 6$ over the 500 Monte Carlo run, respectively. The corresponding SSRTs of the proposed filter with $\mu = 3, 4, 5, 6$ are 0.2991, 0.2983, 0.2993, and 0.2989. It can be seen that the proposed filter with different shape parameters has better performance than current algorithms in the system state and UTVMLP estimations. Moreover, the degree of freedom parameter μ has little influence on the estimation accuracy and time complexity of the proposed algorithm, and the recommended value of μ is therefore set as 5.

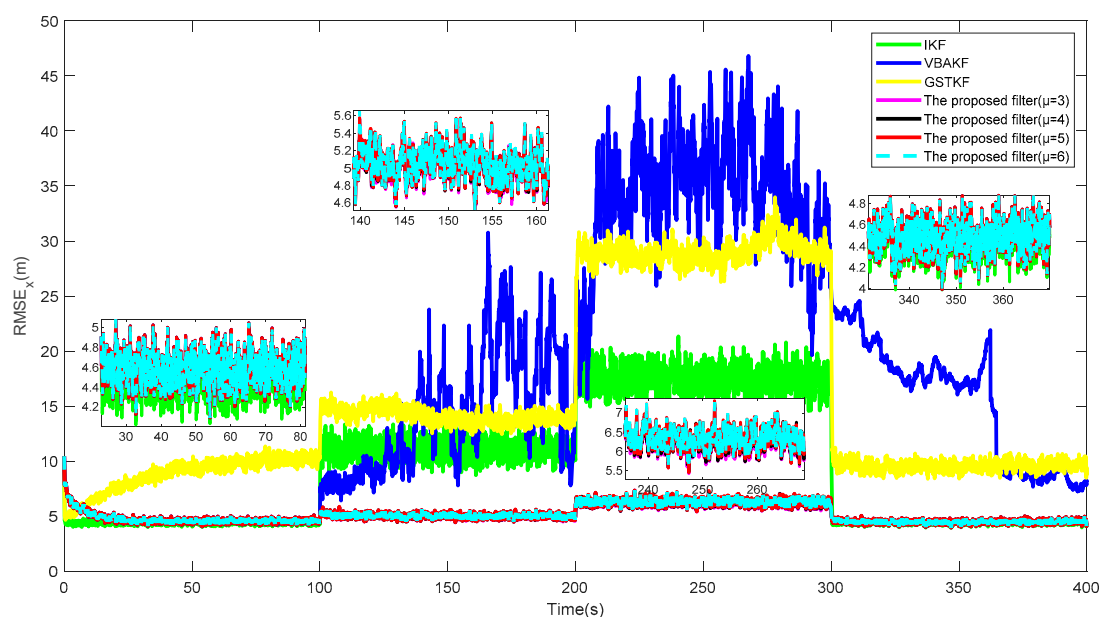


Figure 3. $RMSE_x$ of the proposed filter with $\mu = 3, 4, 5, 6$.

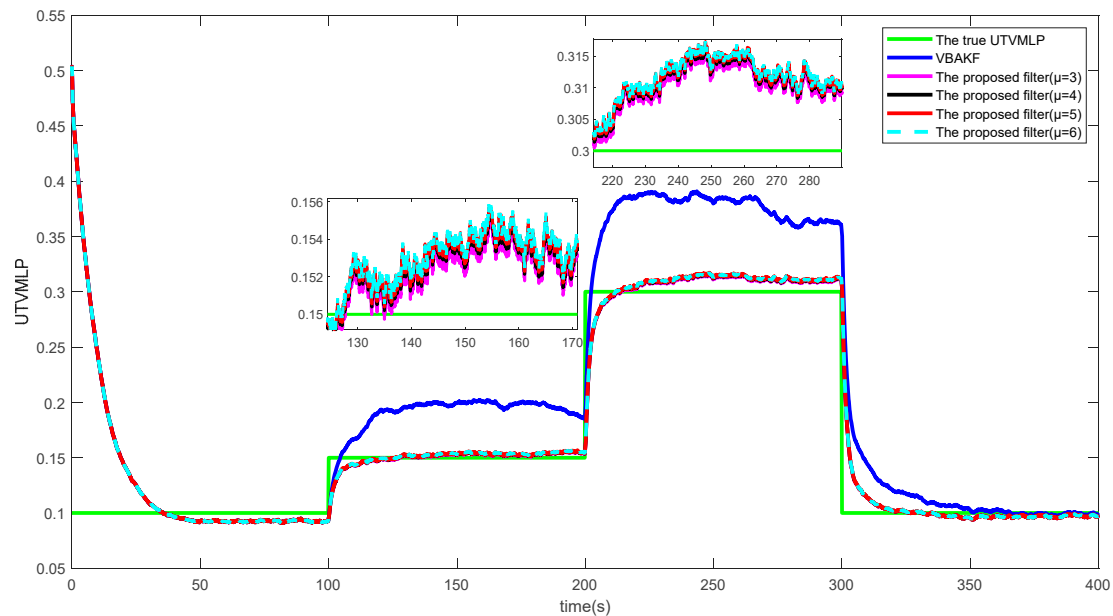


Figure 4. The estimated UTVMLPs of the proposed filter with $\mu = 3, 4, 5, 6$.

Figures 5 and 6 show the $RMSE_x$ s and the estimated UTVMLPs of the proposed filter with forgetting factor $\rho = 0.93, 0.95, 0.97, 0.99$ over the 500 Monte Carlo run, respectively. The corresponding SSRT of the proposed filter with $\rho = 0.93, 0.95, 0.97, 0.99$ is approximately equal to 0.2990. We can find that the proposed filter with $\rho = 0.99$ has the best performance in the system state and UTVMLP estimations, and the value of ρ has little effect on calculation complexity. Therefore, the recommended value of ρ is 0.99.

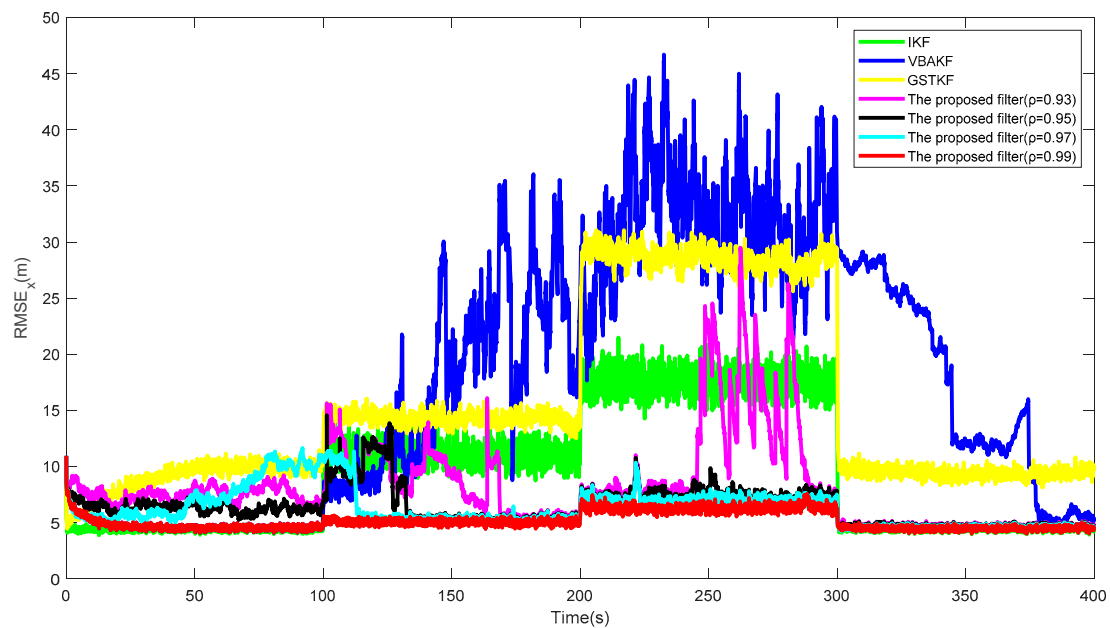


Figure 5. $RMSE_x$ of the proposed filter with forgetting factor $\rho = 0.93, 0.95, 0.97$, and 0.99.

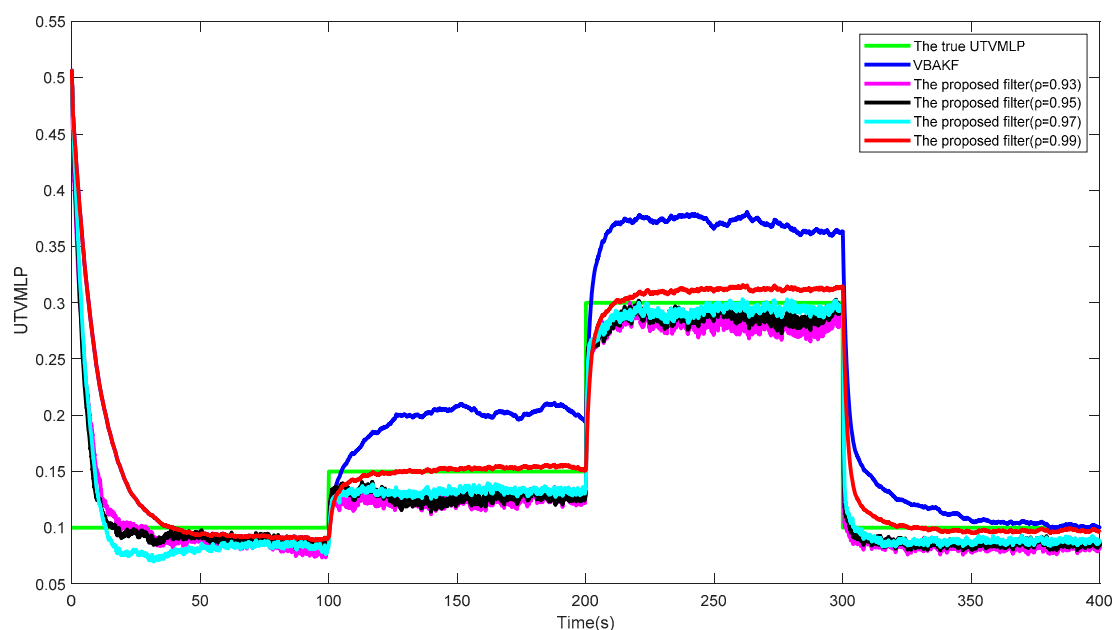


Figure 6. The estimated UTVMLPs of the proposed filter with $\rho = 0.93, 0.95, 0.97$, and 0.99 .

Figure 7 shows the AGRMSE_x s of the proposed filter and the current algorithms with the iteration number $N_I = 1, 2, \dots, 10$. We can see from Figure 7 that the proposed filter has a smaller AGRMSE_x than the existing filters when $N_I \geq 3$ and the proposed filter converges faster than the existing filters. However, Table 4 shows that the setting of N_I has a huge impact on the time consumption of the proposed filter and the SSRT increases with the increase of N_I . Therefore, considering time consumption and estimation accuracy, the recommended value of N_I ranges from 4 to 10.

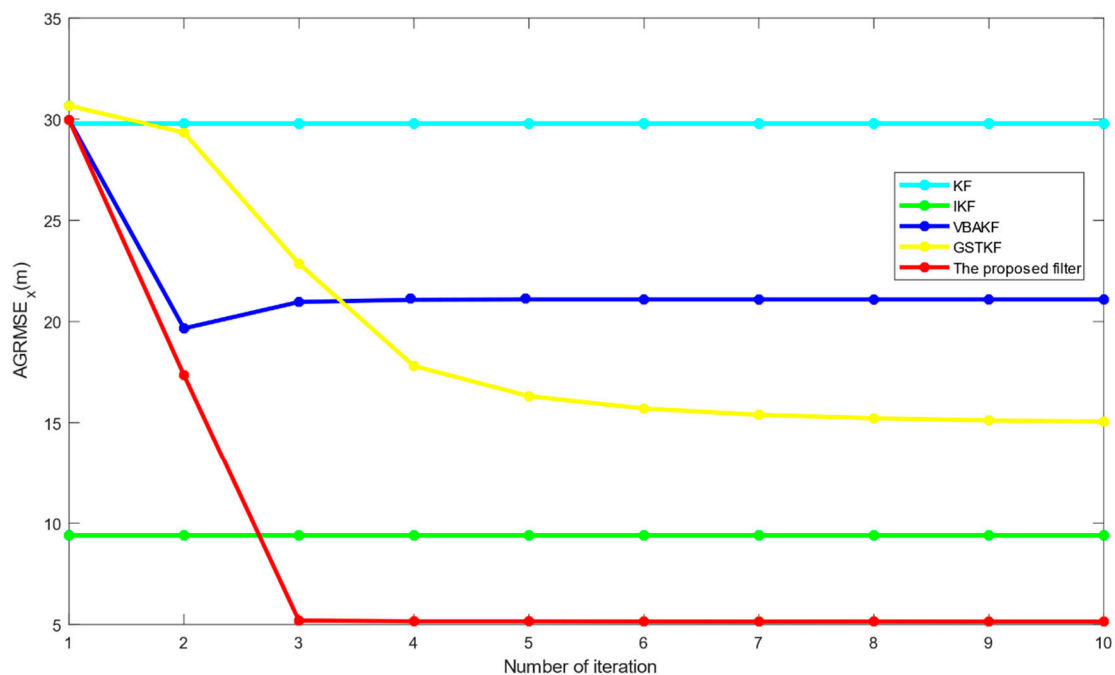


Figure 7. AGRMSE_x of the proposed filter when the iteration number $N_I = 1, 2, \dots, 10$.

Table 4. SSRT of the proposed filter with the iteration number $N_I = 1, 2, \dots, 10$.

Iteration Number N_I	1	2	3	4	5
SSRT (ms)	0.0327	0.0550	0.0894	0.1307	0.1570
Iteration Number N_I	6	7	8	9	10
SSRT (ms)	0.1883	0.2243	0.2451	0.2835	0.2989

5. Conclusions

In this paper, a new VB-based KF is presented to address the filtering issue with UTVMLP and NSHTMN. The system state vector, BM, BL, the intermediate random variables, the mixing probability, and the UTVMLP are simultaneously inferred by introducing the variational Bayesian technique. Simulation results illustrated that the proposed filter has a better performance than existing filters in the estimations of the system state vector and UTVMLP.

6. Future Work

The environmental factors in practical applications may be more complicated than this paper illustrated. Apart from non-stationary heavy-tailed measurement noise and unknown loss probability, the system process noise may also present non-stationary heavy-tailed distribution. In terms of measurement, random delays in measurement will also appear. Therefore, in our future research, we will further consider the factors of non-stationary heavy-tailed process noise and random measurement delay based on the theoretical content of this paper. Additionally, we will design a non-linear filtering method with lower computational complexity to verify the effectiveness in the real world.

Author Contributions: Conceptualization, C.S.; methodology, C.S.; software, C.S. and H.S.; data curation, W.Z.; writing—original draft preparation, C.S.; writing—review and editing, Y.Y.; visualization, C.S. and Y.Y.; supervision, W.Z.; project administration, W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research study was funded by the National Natural Science Foundation of China, grant number 61573113.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Proposition 1

Utilizing Equation (29), $\log p(\Xi, y_{1:t})$ can be derived as

$$\begin{aligned}
 \log p(\Xi, y_{1:t}) &= -0.5\zeta_t b_t (y_t - \mathbf{H}_t x_t)^T \mathbf{R}_t^{-1} (y_t - \mathbf{H}_t x_t) - 0.5(1 - \zeta_t) b_t \beta_t (y_t - \mathbf{H}_t x_t)^T \mathbf{R}_t^{-1} (y_t - \mathbf{H}_t x_t) \\
 &\quad - 0.5\zeta_t (1 - b_t) y_t^T \mathbf{R}_t^{-1} y_t - 0.5(1 - \zeta_t) (1 - b_t) \beta_t y_t^T \mathbf{R}_t^{-1} y_t - 0.5 \left(x_t - \hat{x}_{t|t-1} \right)^T \mathbf{P}_{t|t-1}^{-1} \left(x_t - \hat{x}_{t|t-1} \right) \\
 &\quad \times \left(x_t - \hat{x}_{t|t-1} \right) - \log G(\beta_t, \frac{\mu_t}{2}, \frac{\mu_t}{2}) + \zeta_t \log \varphi_t + (1 - \varphi_t) \log(1 - \zeta_t) + (1 - \gamma_t) \log b_t \\
 &\quad + \gamma_t \log(1 - b_t) + \log \text{Be}(\varphi_t; h_0, 1 - h_0) + \log \text{Be}(\gamma_t; \hat{\eta}_{t|t-1}, \hat{\delta}_{t|t-1}) + c_\Xi
 \end{aligned} \tag{A1}$$

Exploiting $\theta = x_t$ in Equation (28) and utilizing (a) yields

$$\begin{aligned}
 \log q_a^{(s+1)}(x_t) &= -0.5\mathbf{E}^{(s)}[\zeta_t] \mathbf{E}^{(s)}[b_t] (y_t - \mathbf{H}_t x_t)^T \mathbf{R}_t^{-1} (y_t - \mathbf{H}_t x_t) - 0.5\mathbf{E}^{(s)}[1 - \zeta_t] \mathbf{E}^{(s)}[b_t] \\
 &\quad \times \mathbf{E}^{(s)}[\beta_t] (y_t - \mathbf{H}_t x_t)^T \mathbf{R}_t^{-1} (y_t - \mathbf{H}_t x_t) - 0.5 \left(x_t - \hat{x}_{t|t-1} \right)^T \mathbf{P}_{t|t-1}^{-1} \left(x_t - \hat{x}_{t|t-1} \right) + c_{x_t}
 \end{aligned} \tag{A2}$$

By exploiting Equation (34) in Equation (A2), the posterior PDF $q_a^{(s+1)}(x_t)$ is defined as

$$q_a^{(s+1)}(x_t) \propto N\left(x_t; \hat{x}_{t|t}^{(s+1)}, \mathbf{P}_{t|t}^{(s+1)}\right) N\left(y_t; \mathbf{H}_t x_t, \tilde{\mathbf{R}}_t^{(s+1)}\right) \quad (\text{A3})$$

Based on Equation (A3), Equation (30) can be calculated and $q_a^{(s+1)}(x_t)$ is updated by utilizing the measurement update of the traditional Kalman filter.

Appendix B. Proof of Proposition 2

Exploiting $\theta = \beta_t$ in Equation (28) and utilizing Equation (A1) yields

$$\begin{aligned} \log q_a^{(s+1)}(\beta_t) &= 0.5 \left(nE^{(s)}[1 - \zeta_t] + \mu_t - 2 \right) \log(\beta_t) - 0.5\beta_t E^{(s)}[1 - \zeta_t] E^{(s)}[b_t] \\ &\quad \times \text{tr} \left(\mathbf{A}_t^{(s+1)} \mathbf{R}_t^{-1} \right) - 0.5\beta_t E^{(s)}[1 - \zeta_t] E^{(s)}[1 - b_t] \text{tr} \left(\mathbf{B}_t^{(s+1)} \mathbf{R}_t^{-1} \right) - 0.5\mu_t \beta_t + c_{\beta_t} \\ &= 0.5 \left(nE^{(s)}[1 - \zeta_t] + \mu_t - 2 \right) \log(\beta_t) - 0.5\beta_t \left[\text{tr} \left(\mathbf{G}_t^{(s+1)} \mathbf{R}_t^{-1} \right) + \mu_t \right] + c_{\beta_t} \end{aligned} \quad (\text{A4})$$

where $\mathbf{A}_t^{(s+1)}$ and $\mathbf{B}_t^{(s+1)}$ are defined as

$$\begin{aligned} \mathbf{A}_t^{(s+1)} &= E^{(s+1)} \left[(y_t - \mathbf{H}_t x_t)(y_t - \mathbf{H}_t x_t)^T \right] \\ &= \mathbf{H}_t \mathbf{P}_{t|t}^{(s+1)} \mathbf{H}_t^T + \left(y_t - \mathbf{H}_t \hat{x}_{t|t}^{(s+1)} \right) \left(y_t - \mathbf{H}_t \hat{x}_{t|t}^{(s+1)} \right)^T \end{aligned} \quad (\text{A5})$$

$$\mathbf{B}_t^{(s+1)} = E^{(s+1)} \left[y_t (y_t)^T \right] = y_t (y_t)^T \quad (\text{A6})$$

By exploiting Equations (36) and (37) in Equation (A5), we have

$$\log q_a^{(s+1)}(\beta_t) = \left(\pi_t^{(s+1)} - 1 \right) \log(\beta_t) - \nu_t^{(s+1)} \beta_t + c_{\beta_t} \quad (\text{A7})$$

According to Equation (A7), we can obtain

$$q_a^{(s+1)}(\beta_t) \propto \beta_t^{(\pi_t^{(s+1)} - 1)} \exp \left(-\nu_t^{(s+1)} \beta_t \right) \quad (\text{A8})$$

Based on Equation (A8), Equation (41) can be obtained.

Appendix C. Proof of Proposition 3

Exploiting $\theta = b_t$ in Equation (28) and utilizing Equation (A1) yields

$$\log q_a^{(s+1)}(b_t) = b_t \Pi_{ta}^{(s+1)} + (1 - b_t) \Pi_{tb}^{(s+1)} + c_{b_t} \quad (\text{A9})$$

where the intermediate variables $\Pi_{ta}^{(s+1)}$ and $\Pi_{tb}^{(s+1)}$ are defined as

$$\Pi_{ta}^{(s+1)} = E^{(s)}[\log(1 - \gamma_t)] - 0.5 \text{tr} \left(\mathbf{C}_t^{(s+1)} \mathbf{R}_t^{-1} \right) + 0.5nE^{(s)}[1 - \zeta_t] E^{(s)}[\log(\beta_t)] \quad (\text{A10})$$

$$\Pi_{tb}^{(s+1)} = E^{(s)} E^{(s)}[\log(\gamma_t)] - 0.5 \text{tr} \left(\mathbf{D}_t^{(s+1)} \mathbf{R}_t^{-1} \right) + 0.5nE^{(s)}[1 - \zeta_t] E^{(s)}[\log(\beta_t)] \quad (\text{A11})$$

By utilizing (A9)–(A11), the posterior PMF $q_a^{(s+1)}(b_t)$ can be formulated as

$$q_a^{(s+1)}(b_t) = \left[\Delta^{(s+1)} \exp \left(\Pi_{ta}^{(s+1)} \right) \right]^{b_t} \left[\Delta^{(s+1)} \exp \left(\Pi_{tb}^{(s+1)} \right) \right]^{(1-b_t)} \quad (\text{A12})$$

Based on Equation (A12), $q_a^{(s+1)}(b_t)$ can be updated as the Bernoulli distribution.

Appendix D. Proof of Proposition 4

Exploiting $\theta = \zeta_t$ in Equation (28) and utilizing Equation (A1) yields

$$\log q_a^{(s+1)}(\zeta_t) = \zeta_t \Sigma_{ta}^{(s+1)} + (1 - \zeta_t) \Sigma_{tb}^{(s+1)} + c_{b_t} \quad (\text{A13})$$

where the intermediate variables $\Sigma_{ta}^{(s+1)}$ and $\Sigma_{tb}^{(s+1)}$ are defined as

$$\Sigma_{ta}^{(s+1)} = \left\{ E^{(s)}[\log(\varphi_t)] - 0.5 \text{tr}(\mathbf{V}_t^{(s+1)} \mathbf{R}_t^{-1}) \right\} \quad (\text{A14})$$

$$\Sigma_{tb}^{(s+1)} = E^{(s)}[\log(1 - \varphi_t)] - 0.5 \text{tr}(\mathbf{W}_t^{(s+1)} \mathbf{R}_t^{-1}) + 0.5 n E^{(s)}[\log(\beta_t)] \quad (\text{A15})$$

By utilizing Equations (A13)–(A15), the posterior PMF $q_a^{(s+1)}(\zeta_t)$ can be formulated as

$$q_a^{(s+1)}(\zeta_t) = \left[\nabla^{(s+1)} \exp(\Sigma_{ta}^{(s+1)}) \right]^{\zeta_t} \left[\nabla^{(s+1)} \exp(\Sigma_{tb}^{(s+1)}) \right]^{(1-\zeta_t)} \quad (\text{A16})$$

Based on Equation (A16), $q_a^{(s+1)}(\zeta_t)$ can be updated as the Bernoulli distribution.

Appendix E. Proof of Proposition 5

Exploiting $\theta = \varphi_t$ in Equation (28) and utilizing Equation (A1) yields

$$\log q_a^{(s+1)}(\varphi_t) = \log(\varphi_t) \left\{ h_0 + E^{(s)}[\zeta_t] - 1 \right\} + \log(1 - \varphi_t) \left\{ E^{(s)}[1 - \zeta_t] - h_0 \right\} + c_{\varphi_t} \quad (\text{A17})$$

Substituting Equations (48) and (49) in Equation (A17) yields

$$\log q_a^{(s+1)}(\varphi_t) = \log(\varphi_t) \left(h_t^{(s+1)} - 1 \right) + \log(1 - \varphi_t) \left(d_t^{(s+1)} + 1 \right) + c_{\varphi_t} \quad (\text{A18})$$

According to Equation (A18), we obtain

$$q_a^{(s+1)}(\varphi_t) \propto \varphi_t^{(h_t^{(s+1)}-1)} (1 - \varphi_t)^{(d_t^{(s+1)}+1)} \quad (\text{A19})$$

Based on Equation (A19), Equation (47) is obtained.

Appendix F. Proof of Proposition 5

Exploiting $\theta = \gamma_t$ in Equation (28) and utilizing Equation (A1) yields

$$\log q_a^{(s+1)}(\gamma_t) = \log(\gamma_t) \left(\hat{\eta}_{t|t-1} - E^{(s+1)}[b_t] \right) + \log(1 - \gamma_t) \left(\hat{\delta}_{t|t-1} + E^{(s+1)}[1 - b_t] \right) \quad (\text{A20})$$

Substituting Equations (51) and (52) in Equation (A20) yields

$$\log q_a^{(s+1)}(\gamma_t) = \log(\gamma_t) \left(\hat{\eta}_t^{(s+1)} + 1 \right) + \log(1 - \gamma_t) \left(\hat{\delta}_t^{(s+1)} - 1 \right) + c_{\varphi_t} \quad (\text{A21})$$

According to Equation (A21), we obtain

$$q_a^{(s+1)}(\gamma_t) \propto \gamma_t^{(\hat{\eta}_t^{(s+1)}+1)} (1 - \gamma_t)^{(\hat{\delta}_t^{(s+1)}-1)} \quad (\text{A22})$$

Based on Equation (A22), Equation (50) is obtained.

References

1. Mohinder, S.G.; Angus, P.A. Linear Optimal Filters and Predictors. In *Kalman Filtering: Theory and Practice Using MATLAB*; IEEE: Manhattan, NY, USA, 2008; pp. 131–181.
2. Simon, D. The discrete-time Kalman filter. In *Optimal State Estimation*; Wiley-Interscience: New York, NY, USA, 2006; pp. 121–148.
3. Li, Q.; Li, R.; Ji, K.; Dai, W. Kalman Filter and Its Application. In Proceedings of the 2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS), Tianjin, China, 1–3 November 2015; pp. 74–77.

4. Yang, Y.J.; Fan, X.G.; Zhuo, Z.F.; Wang, S.D.; Nan, J.G.; Huang, J.K. AFAKF for manoeuvring target tracking based on current statistical model. *Iet Sci. Meas. Technol.* **2016**, *10*, 637–643. [[CrossRef](#)]
5. Hou, B.W.; He, Z.M.; Zhou, X.Y.; Zhou, H.Y.; Li, D.; Wang, J.Q. Maximum Correntropy Criterion Kalman Filter for alpha-Jerk Tracking Model with Non-Gaussian Noise. *Entropy* **2017**, *19*, 648. [[CrossRef](#)]
6. Joerger, M.; Pervan, B. Kalman Filter-Based Integrity Monitoring Against Sensor Faults. *J. Guid. Control. Dyn.* **2013**, *36*, 349–361. [[CrossRef](#)]
7. Jia, G.L.; Huang, Y.L.; Zhang, Y.G.; Chambers, J. A Novel Adaptive Kalman Filter With Unknown Probability of Measurement Loss. *IEEE Signal Process. Lett.* **2019**, *26*, 1862–1866. [[CrossRef](#)]
8. Sinopoli, B.; Schenato, L.; Franceschetti, M.; Poolla, K.; Jordan, M.I.; Sastry, S.S. Kalman filtering with intermittent observations. *IEEE Trans. Autom. Control* **2004**, *49*, 1453–1464. [[CrossRef](#)]
9. Mo, Y.L.; Sinopoli, B. Kalman Filtering With Intermittent Observations: Tail Distribution and Critical Value. *IEEE Trans. Autom. Control* **2012**, *57*, 677–689.
10. Zhang, J.Q.; You, K.Y.; Xie, L.H. Bayesian Filtering With Unknown Sensor Measurement Losses. *Ieee Trans. Control Netw. Syst.* **2019**, *6*, 163–175. [[CrossRef](#)]
11. Bai, M.M.; Huang, Y.L.; Jia, G.L.; Zhang, Y.G. A robust fixed-interval smoother for nonlinear systems with non-stationary heavy-tailed state and measurement noises. *Signal Process.* **2021**, *180*, 107898. [[CrossRef](#)]
12. Zhu, H.; Zhang, G.R.; Li, Y.F.; Leung, H. A novel robust Kalman filter with unknown non-stationary heavy-tailed noise. *Automatica* **2021**, *127*, 109511. [[CrossRef](#)]
13. Jia, G.; Huang, Y.; Bai, M.B.; Zhang, Y. A Novel Robust Kalman Filter With Non-stationary Heavy-tailed Measurement Noise. *IFAC-PapersOnLine* **2020**, *53*, 368–373. [[CrossRef](#)]
14. Huang, Y.; Zhang, Y.; Zhao, Y.; Chambers, J.A. A Novel Robust Gaussian–Student’s t Mixture Distribution Based Kalman Filter. *IEEE Trans. Signal Process.* **2019**, *67*, 3606–3620. [[CrossRef](#)]
15. Huang, Y.; Zhang, Y.; Wu, Z.; Li, N.; Chambers, J. A Novel Adaptive Kalman Filter With Inaccurate Process and Measurement Noise Covariance Matrices. *IEEE Trans. Autom. Control* **2018**, *63*, 594–601. [[CrossRef](#)]
16. Bishop, C.; Ligne, S. *Pattern Recognition and Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2006; Volume 1.
17. Tzikas, D.G.; Likas, A.C.; Galatsanos, N.P. The variational approximation for Bayesian inference. *IEEE Signal Process. Mag.* **2008**, *25*, 131–146. [[CrossRef](#)]
18. Huang, Y.; Zhang, Y.; Li, N.; Chambers, J. A Robust Gaussian Approximate Fixed-Interval Smoother for Nonlinear Systems With Heavy-Tailed Process and Measurement Noises. *IEEE Signal Process. Lett.* **2016**, *23*, 468–472. [[CrossRef](#)]