

Article

Variational Bayesian-Based Improved Maximum Mixture Correntropy Kalman Filter for Non-Gaussian Noise

Xuyou Li, Yanda Guo * and Qingwen Meng

Department of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China; lixuyou@hrbeu.edu.cn (X.L.); mengqingwen@hrbeu.edu.cn (Q.M.)

* Correspondence: 2012042206@hrbeu.edu.cn; Tel.: +86-18845632572

Abstract: The maximum correntropy Kalman filter (MCKF) is an effective algorithm that was proposed to solve the non-Gaussian filtering problem for linear systems. Compared with the original Kalman filter (KF), the MCKF is a sub-optimal filter with Gaussian correntropy objective function, which has been demonstrated to have excellent robustness to non-Gaussian noise. However, the performance of MCKF is affected by its kernel bandwidth parameter, and a constant kernel bandwidth may lead to severe accuracy degradation in non-stationary noises. In order to solve this problem, the mixture correntropy method is further explored in this work, and an improved maximum mixture correntropy KF (IMMCKF) is proposed. By derivation, the random variables that obey Beta-Bernoulli distribution are taken as intermediate parameters, and a new hierarchical Gaussian state-space model was established. Finally, the unknown mixing probability and state estimation vector at each moment are inferred via a variational Bayesian approach, which provides an effective solution to improve the applicability of MCKFs in non-stationary noises. Performance evaluations demonstrate that the proposed filter significantly improves the existing MCKFs in non-stationary noises.

Keywords: Kalman filter; maximum correntropy criterion; mixture correntropy; variational Bayesian inference



Citation: Li, X.; Guo, Y.; Meng, Q.

Variational Bayesian-Based Improved Maximum Mixture Correntropy Kalman Filter for Non-Gaussian Noise. *Entropy* **2022**, *24*, 117. <https://doi.org/10.3390/e24010117>

Academic Editor: Ercan Kuruoglu

Received: 29 November 2021

Accepted: 10 January 2022

Published: 12 January 2022

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



Copyright: © 2022 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

The state estimation problem in dynamic systems is an important research topic in engineering applications and scientific research. As an excellent optimal state-space estimator, the Kalman filter (KF) is commonly applied in various fields like control systems and signal processing. Unfortunately, the optimality of KF requires exact system models and ideal noise conditions as summarized in [1].

The widely used KF, which usually refers to the Kalman filter based on the Hidden Markov Models (HMM), has rigorous requirements for the noise models. Both process and measurement noise are assumed as ideal independent Gaussian noise sequences. However, in practical applications, ideal noise conditions are not likely, and model uncertainties such as system structure changes and environmental disturbances are generally inevitable. Moreover, unexpected noise interference, such as colored noise and non-Gaussian noise, widely exist, and the performance of the KF is likely to worsen when applied to such situations. For example, when the independent noise assumption no longer holds, and the colored noise needs to be considered in the system model, as one of the efficient solutions, the pairwise Markov models (PMM), which can be deemed a general form of HMM, can be taken as an efficient improvement scheme for KF. In [2], the framework of the KF based on PMM is derived, which allows the cross-dependence between observations conditionally on hidden variables. As an extension, the KF based on triplet Markov chains is provided in [3]. These methods provided efficient solutions for the colored noise filtering from the perspective of the state-space models. On this basis, to solve the model parameters' uncertain problems in practical applications. The robust parameter estimation problem for

the PMM model is further explored in [4,5], which provided several representative schemes for the state estimation problems in colored noise or model parameter uncertain conditions.

On the other hand, the research on filtering algorithms under non-Gaussian noise has also attracted extensive attention. Generally, non-Gaussian distribution noises are often caused by impulsive noise, which gives the overall noise distributions obvious heavy-tailed features (such as some special Gaussian-mixture distributions). Extensive research has been carried out regarding the filtering problems in non-Gaussian noises. H-Infinity filtering and m-estimators are typical robust methods used for this purpose [6]. The Huber Kalman filter (HKF) is one of the most representative m-estimators, which retains a consistent form with KF and provides reliable robustness to external impulsive noise interference. In the existing relevant studies, various improvements based on the HKF have also been proposed, such as a batch mode [7] and several typical nonlinear schemes [8,9]. Besides m-estimators, student-t distribution filters have also been designed to deal with state estimate problems under heavy-tailed noise conditions. The filters attenuate the interference of outliers by their heavy-tailed distribution statistical characteristics [10,11]. However, student-t filters suffer from high-order statistics' loss due to moment matching and inaccurate parameter selection, such as the degrees of freedom limiting their applicability. To solve the problem, another novel student-t filter based on the hierarchical Gaussian state-space model and variational Bayes (VB) inference approach is proposed, which is compactly called the robust student-t Kalman filter (RSTKF) [12]. Compared with the student-t filter, several unknown parameters, such as the degrees of freedom of the student-t distribution and the scale coefficient, are updated accordingly in the RSTKF [13–15].

Recently, a new robust maximum correntropy Kalman filter (MCKF) was proposed for the filtering problem in non-Gaussian noises [16,17]. The MCKF is derived on the maximum correntropy criterion (MCC), which introduces the Gaussian correntropy as a novel local similarity measure, and a novel object function based on correntropy is utilized to overcome the non-Gaussian noise interference [18]. Compared with other existing robust filters, the Gaussian correntropy-based MCKF has better statistical characteristics [19]. Meanwhile, the MCKF can be formulated in the same form as the KF. Similarly, the states constrained [20] and nonlinear forms [21,22] have been proposed and validated in various applications [23]. Moreover, an adaptive MCKF was proposed to improve the filtering performance by updating the measurement noise covariance via VB approach [24].

Due to the excellent robustness to outliers, in many cases, the filtering problem of a linear system corrupted by impulsive noise can be solved effectively by the MCKF. However, just like other robust filters, the filtering performance of the MCKF is closely related to its initial parameters, which are usually obtained by experience or trial and error method. Similar to many existing robust filtering algorithms with fixed parameters, it might result in performance degradation in non-stationary noises. Several studies have focused on the parameter problem of correntropy function. In prior works, heuristic solutions were proposed to adjust the kernel bandwidth during filtering [25,26]. As it is difficult to directly find an optimal kernel bandwidth during filtering, the mixture correntropy concept was proposed in [27,28], the method takes mixture Gaussian correntropy with different kernel parameters instead of the Gaussian correntropy, and a new maximum mixture correntropy criterion (MMCC) is derived from replacing MCC. However, the mixing probability of the mixture correntropy needs to be configured manually and fixed, which results in performance degradation similar to that of the MCKF.

To improve the filtering performance when applied to non-stationary noise conditions, it may be an efficient improved solution that considers the model's switching scenarios. For instance, in [29], the optimal recursive filtering method is studied for non-Gaussian Markov switching models, in which a semi-supervised parameter estimation method is used. In addition, for the non-stationary noise conditions concerned in this paper, the variable Bayesian approach is an implementation scheme worthy of consideration. Based on the triplet Markov chain model, the general form of variable Bayesian inference is deduced in [30], and a structured variable Bayesian inference frame with regime switching

is obtained. For typical linear filtering applications, the RSTKF proposed in [12] also adopts a similar operation scheme, where VB approximation is operated by rescaling the covariance and inferring parameters to deal with the non-stationary non-Gaussian noises. In [31], a similar problem is further studied, and a VB-based robust student's t KF is applied to the linear PMM systems, which extends the filter's applicability to more general conditions, and independent noise assumption is no longer needed. Therefore, inspired by the related references, the model switch concept is adopted in this work, and the variational Bayesian approach is also operated to infer estimation results. In view of the filtering accuracy degradation problem of MCKF in the non-stationary and non-Gaussian noise, a series of studies are carried out.

In this work, an improved maximum mixture Kalman filter (IMMCKF) is therefore proposed, the intermediate random variables are used to represent the mixing probability of mixture correntropy. The state variables and mixing probability are approximated by the derived variational Bayesian approach. Compared with the existing MCKF algorithms, the numerical test results demonstrate that the proposed filter could deal with the filtering problem well in a non-stationary non-Gaussian noises environment. The contributions of this paper are listed as follows:

- (a) The accuracy degradation problem of existing fixed-parameter robust filtering algorithms in non-stationary noise environment is considered. By analyzing, it is inferred that the mixture correntropy function can be improved as the breakthrough point and applied to the filtering problem in such noise conditions. On this basis, a novel improved robust filtering algorithm is then further derived.
- (b) By employing Beta-Bernoulli distributed intermediate random variables, a new hierarchical Gaussian state-space model is therefore derived, the system state vector, and the unknown variables are simultaneously estimated by utilizing the variational Bayesian technique.
- (c) Through analyses and derivations, the necessary selection strategy of initial parameters is derived, and the numerical test results show that the filtering performance is enhanced obviously after several iterations. On the one hand, it achieved desired robust filtering performance when applied to non-stationary noise conditions. On the other hand, the proposed algorithm effectively avoids the possible filtering divergence difficulties of MCKF in practical application.

This paper is organized as follows: In Section 2, we review the concept of correntropy and existing MCKF. In Section 3, a variational Bayesian-based improved maximum mixture correntropy KF is derived to solve the filtering problem in non-stationary non-Gaussian noises, in which the variational Bayesian approach is applied in the proposed filter. Section 4 provides performance evaluations and analysis, demonstrating the advantages of the proposed filter in different noise conditions. Conclusions are given in Section 5.

2. Materials and Methods

2.1. Definition of Correntropy and Properties

Correntropy is a useful local similarity measure tool for state estimation in heavy-tailed noise environments. Given two random variables $X, Y \in R$ with a joint distribution function $F_{XY}(x, y)$, correntropy can be defined as

$$V(X, Y) = E[\kappa_\sigma(X, Y)] = \int \kappa_\sigma(x, y) dF_{XY}(x, y) \quad (1)$$

where $\kappa_\sigma(\cdot)$ is a shift-invariant Mercer kernel. In this work, the Gaussian kernel function is given by

$$\kappa_\sigma(X, Y) = G_\sigma(e) = \exp\left(-\frac{e^2}{2\sigma^2}\right) \quad (2)$$

where $e = X - Y$, and $\sigma > 0$ denotes the kernel bandwidth. In order to make correntropy applicable to complex noise conditions, the default Gaussian kernel function can be extended into a mixture correntropy form as follows:

$$M(X, Y) = E[\rho G_{\sigma_1}(e) + (1 - \rho) G_{\sigma_2}(e)] \tag{3}$$

where $G_{\sigma_1}, G_{\sigma_2}$ represent the Gaussian correntropy with two different kernel bandwidth parameters, and $\rho \in [0, 1]$ represents the mixing probability. For convenience of application, (3) can be approximately expressed as

$$M(X, Y) = \frac{1}{N} \left(\sum_{i=1}^N \rho \exp\left(-\frac{e_i^2}{2\sigma_1^2}\right) + (1 - \rho) \exp\left(-\frac{e_i^2}{2\sigma_2^2}\right) \right) \tag{4}$$

where N represents the number of samples. The mixture correntropy can be taken as a generalization of the original correntropy, if $\rho = 1$ or $\rho = 0$, it reduces to the Gaussian correntropy with single kernel parameter and $M(X, Y) = 1$ if $X = Y$.

2.2. Robust Kalman Filter Based on Maximum Correntropy Criterion

Consider the linear state-space system based on HMM, the equations to be

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_k \tag{5}$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \tag{6}$$

where k is a discrete time index, $\mathbf{x}_k \in \mathbb{R}^n$ is the system state vector at discrete-time index k , \mathbf{F}_k is the state transition matrix, $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement vector, \mathbf{H}_k is the measurement matrix; \mathbf{w}_k and \mathbf{v}_k are zero-mean process and measurement noise vector with nominal covariance \mathbf{Q}_k and \mathbf{R}_k , respectively. It is assumed that both process and measurement noise are statistically independent and time uncorrelated. When the process and measurement noises are assumed ideal Gaussian distributions, and initial state \mathbf{x}_0 is random Gaussian variables, the state can be inferred by the KF, which is an optimal filter in the minimum mean square error (MMSE) criterion. Additionally, the quadratic objective function can be formulated as follows:

$$\hat{\mathbf{x}}_{k|k} = \arg \min_{\mathbf{x}_k} J(\mathbf{x}_k) = \arg \min_{\mathbf{x}_k} \left(\frac{1}{2} \|\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\|_{\mathbf{P}_{k|k-1}^{-1}}^2 + \frac{1}{2} \|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|_{\mathbf{R}_k^{-1}}^2 \right) \tag{7}$$

where $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$. To minimize Equation (7), the KF is implemented in two steps as below, the prior estimation is

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} \tag{8}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_{k-1} \tag{9}$$

and the posterior measurement update is

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{10}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \tag{11}$$

$$\mathbf{P}_{k|k} = (\mathbf{I}_n - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \tag{12}$$

where $\mathbf{P}_{k|k-1}$ and $\mathbf{P}_{k|k}$ represents the prior and posterior covariance matrix, respectively. The KF is an optimal estimator in an ideal white Gaussian noise environment. In Equation (7), only the second-order statistics are used during the state update, and the KF is susceptible to non-Gaussian noise interference.

To solve the filtering problem in non-Gaussian noise conditions, recently, MCC is often considered an efficient potential solution. As described in Section 2, correntropy has several necessary properties that make it capable of dealing with the non-Gaussian estimation

problem. Different from the global similarity measure-mean square error (MSE), which only contains the second-order statistics, Gaussian correntropy incorporates all even order moments [17,18]. In geometric meaning, MSE gives the L2 norm distance while correntropy offers a hybrid norm distance, where the correntropy behaves like the L2→L0 norm with the increasing difference between two points.

As proved by the research in [17], maximizing the correntropy of two different random variables can be used as a criterion in dealing with the non-Gaussian noise problem, especially the filtering problem in heavy-tailed noises, which leads to MCC. Therefore, to enhance the robustness of the Kalman filter, the objective function based on MCC is introduced to replace the original quadratic cost function, and the new objective function therefore can be formulated as:

$$\hat{\mathbf{x}}_k = \underset{\mathbf{x}_k}{\operatorname{argmax}} J_{\text{MCC}}(\mathbf{x}_k) = \underset{\mathbf{x}_k}{\operatorname{argmax}} \left(\sum_{i=1}^n \mathbf{a}_i G_{\sigma_m}(\mathbf{e}_i) + \sum_{j=1}^m \mathbf{b}_j G_{\sigma_p}(\mathbf{f}_j) \right) \tag{13}$$

where $\mathbf{e} = \left\| \mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \right\|_{\mathbf{R}_k^{-1}}$, $\mathbf{f} = \left\| \hat{\mathbf{x}}_k - \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} \right\|_{\mathbf{P}_{k|k-1}^{-1}}$ the subscript i represents the i th element of the vector; \mathbf{a} and \mathbf{b} are tuning parameter vectors; and σ_m, σ_p denotes the kernel bandwidths corresponding to the \mathbf{R}_k and $\mathbf{P}_{k|k-1}$, respectively. For simplicity, it is assumed that $\sigma_m = \sigma_p = \sigma$, to ensure the filter converges to the KF when the kernel bandwidth goes to infinity, assuming that $\mathbf{a}_i = \sigma$, $\mathbf{b}_j = \sigma$ [21]. The prior estimation steps of the MCKF are the same as in Equations (8) and (9) and the posterior state estimation of \mathbf{x}_k is obtained by below KF like equations with the updated filter gain as

$$\bar{\mathbf{K}}_k = \bar{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \bar{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T + \bar{\mathbf{R}}_k)^{-1} \tag{14}$$

$$\bar{\mathbf{P}}_{k|k-1} = \mathbf{S}_{\mathbf{P}_{k|k-1}} \boldsymbol{\Lambda}_p^{-1} \mathbf{S}_{\mathbf{P}_{k|k-1}}^T \tag{15}$$

$$\bar{\mathbf{R}}_k = \mathbf{S}_{\mathbf{R}_k} \boldsymbol{\Lambda}_m^{-1} \mathbf{S}_{\mathbf{R}_k}^T \tag{16}$$

where $\boldsymbol{\Lambda}_p = \operatorname{diag}(G_\sigma(\mathbf{f}_1)G_\sigma(\mathbf{f}_2)\dots G_\sigma(\mathbf{f}_n))$, $\boldsymbol{\Lambda}_m = \operatorname{diag}(G_\sigma(\mathbf{e}_1)G_\sigma(\mathbf{e}_2)\dots G_\sigma(\mathbf{e}_m))$, and $\mathbf{P}_{k|k-1} = \mathbf{S}_{\mathbf{P}_{k|k-1}} \mathbf{S}_{\mathbf{P}_{k|k-1}}^T$, $\mathbf{R}_k = \mathbf{S}_{\mathbf{R}_k} \mathbf{S}_{\mathbf{R}_k}^T$. It is worth noting that the MCC solution cannot be obtained in a closed form and usually solves it using an iterative update algorithm such as the fixed-point iterative algorithm, which involves no step size and may converge fast, the condition that guarantees the convergence of the fixed point MCC was given in [32].

3. Main Results

3.1. Robust Kalman Filter Based on Mixture Correntropy Criterion

The kernel parameters of Gaussian correntropy determine the filtering performance of the MCKF, and an improper kernel bandwidth might lead to filtering performance degradation or even diverge. For this reason, the mixture correntropy is an alternative solution to improve the solution of problem by reducing the filter’s sensitivity to the kernel parameters.

The maximum mixture correntropy Kalman filter (MMCKF) is derived in this section. According to the mixture correntropy Function (4), the objective function of the maximum mixture correntropy criterion can be formulated as

$$\hat{\mathbf{x}}_k = \underset{\mathbf{x}_k}{\operatorname{argmax}} J_{\text{MMCC}}(\mathbf{x}_k) = \underset{\mathbf{x}_k}{\operatorname{argmax}} \left(\sum_{i=1}^n \mathbf{a}_i M(\mathbf{e}_i) + \sum_{j=1}^m \mathbf{b}_j M(\mathbf{f}_j) \right) \tag{17}$$

where $M(\mathbf{e}_i)$ represents the mixture correntropy function as (3). Similar to Equation (13), it is assumed that $\mathbf{a}_i = \mathbf{b}_j = \lambda$, and mixture correntropy is the convex combination of the two

Gaussian correntropy functions. To maximize this objective function, the solution can be obtained by solving

$$\frac{\partial J_{\text{MMCC}}(\mathbf{x}_k)}{\partial \mathbf{x}_k} = \lambda \left(\sum_{i=1}^n \boldsymbol{\theta}_i M(\mathbf{e}_i) + \sum_{j=1}^m \boldsymbol{\theta}_j M(\mathbf{f}_j) \right) = 0 \tag{18}$$

where

$$\boldsymbol{\theta}_i = \left(\frac{\rho G_{\sigma_1}(\mathbf{e}_i)}{\sigma_1^2} + \frac{(1-\rho)G_{\sigma_2}(\mathbf{e}_i)}{\sigma_2^2} \right) \tag{19}$$

represents the derivation of the correntropy function.

To maintain consistency to the KF, a tuning factor λ should be properly assigned as

$$\lambda = \frac{\sigma_1^2 \sigma_2^2}{\rho \sigma_2^2 + (1-\rho)\sigma_1^2} \tag{20}$$

and the MMCKF converges to the optimal KF while the process and measurement noises obey ideal Gaussian distributions. A modified mixture correntropy function $C(\mathbf{e}_i)$ is then formulated as

$$C(\mathbf{e}_i) = \lambda \boldsymbol{\theta}_i = \mu G_{\sigma_1}(\mathbf{e}_i) + (1-\mu)G_{\sigma_2}(\mathbf{e}_i) \tag{21}$$

with $\mu = \rho \sigma_2^2 / (\rho \sigma_2^2 + (1-\rho)\sigma_1^2)$, and the optimal estimate of \mathbf{x}_k is obtained by KF like equations, where $\mathbf{\Lambda}_p = \text{diag}(C(\mathbf{f}_1)C(\mathbf{f}_2)...C(\mathbf{f}_n))$, $\mathbf{\Lambda}_m = \text{diag}(C(\mathbf{e}_1)C(\mathbf{e}_2)...C(\mathbf{e}_m))$, and $\mu \in [0, 1]$. Here, $C(\mathbf{e}_i)$ can be regarded as a linear transformation of $M(\mathbf{e}_i)$, they are positive correlated, and $C(\mathbf{e}_i) = M(\mathbf{e}_i)$ if $\rho, \mu \in \{0, 1\}$.

3.2. Improved MMCKF via Variational Bayesian Approximation

Similar to the Gaussian correntropy of the MCKF, in Equation (21), the mixing probability parameters are constant, which inevitably results in the filtering performance degradation in non-stationary noises. To improve the filtering performance in such complex noise environments, an improved algorithm is derived in this section. The mixing probability of the mixture correntropy is reassigned as an unknown random variable that can be further approximated during filtering.

According to the Bayesian theorem, the posterior probability density function (PDF) $p(\mathbf{x}_k | \mathbf{y}_{1:k})$ is formulated as

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \propto p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \tag{22}$$

in which $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ is determined using the Chapman–Kolmogorov equation. Therefore,

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \propto \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} \tag{23}$$

where $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ is the posterior density at time $k - 1$ and $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$ is the transition density.

In this work, we assumed that μ is an unknown random variable, and $\mu \in [0, 1]$. In order to approximate the unknown state and reasonable mixing probability via VB approach, we first need to determine the assumed distribution of unknown variable. Then, it is necessary to introduce a prior probability distribution $p(\mu)$. According to the desired scenario, μ is assumed to be an Beta distributed variable. Since the likelihood distribution of the mixing probability μ can be formulated as a Bernoulli distribution by introducing Bernoulli random variables ξ , according to the Bayesian probability theory, we have

$$p(\mu | \xi) = \frac{p(\xi | \mu) \cdot p(\mu)}{\int p(\xi | \mu) \cdot p(\mu) d\xi} \tag{24}$$

and then, $p(\mu | \xi) \propto p(\xi | \mu) \cdot p(\mu)$. According to the Bayesian formula, the posterior probability distribution is directly proportional to the product of the prior probability distribution

and the likelihood function. The form of the posterior probability distribution will be the same as that of the prior. Therefore, the conjugate prior distribution of μ is selected as a Beta distribution according to Bayesian probability theory, have $p(\mu) = Be(\mu; a, b)$, where $E(\mu) = a / (a + b)$.

Now that we have determined the distribution of the required intermediate variables, we can further derive the following specific formula based on the system model. Using two random variables s_k and t_k obey Bernoulli distribution to generate the likelihood distribution. In order to ensure that the prior and posterior probability distributions retain the same form, the Beta distribution is taken as the prior conjugation distribution for inference. Therefore, the conditional probability $p(t_k|\alpha_k)$ and $p(s_k|\beta_k)$ can be formulated as

$$p(t_k|\alpha_k) = \alpha_k^{t_k}(1 - \alpha_k)^{(1-t_k)}, \text{ s.t. } t_k \in \{0,1\} \tag{25}$$

$$p(s_k|\beta_k) = \beta_k^{s_k}(1 - \beta_k)^{(1-s_k)}, \text{ s.t. } s_k \in \{0,1\} \tag{26}$$

in which α_k and β_k obey the Beta distribution as

$$p(\alpha_k) = Be(\alpha_k; a_0, b_0) \tag{27}$$

$$p(\beta_k) = Be(\beta_k; c_0, d_0) \tag{28}$$

where a_0, b_0, c_0 and d_0 represent the prior Beta parameters for the mixing probabilities. Then, the conditional PDF $p(y_k|x_k, t_k)$ and $p(x_k|y_{1:k-1}, s_k)$ can be rewritten as the following hierarchical Gaussian form by

$$p(\mathbf{y}_k|\mathbf{x}_k, t_k) = N(\mathbf{y}_k; \mathbf{H}_k\mathbf{x}_k, \tilde{\mathbf{R}}_k)^{t_k} N(\mathbf{y}_k; \mathbf{H}_k\mathbf{x}_k, \bar{\mathbf{R}}_k)^{(1-t_k)} \tag{29}$$

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}, s_k) = N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \tilde{\mathbf{P}}_{k|k-1})^{s_k} N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \bar{\mathbf{P}}_{k|k-1})^{(1-s_k)} \tag{30}$$

where $\tilde{\mathbf{R}}_k, \tilde{\mathbf{P}}_{k|k-1}$ and $\bar{\mathbf{R}}_k, \bar{\mathbf{P}}_{k|k-1}$ represent the updated covariance matrix corresponding to the Gaussian correntropy functions G_{σ_1} and G_{σ_2} , respectively. The modified prior error covariance is formulated as $\tilde{\mathbf{P}}_{k|k-1} = \mathbf{S}_{P_{k|k-1}} \Lambda_{p,\sigma_1}^{-1} \mathbf{S}_{P_{k|k-1}}^T$ and $\bar{\mathbf{P}}_{k|k-1} = \mathbf{S}_{P_{k|k-1}} \Lambda_{p,\sigma_2}^{-1} \mathbf{S}_{P_{k|k-1}}^T$, and the measurement noise covariance is formulated as $\tilde{\mathbf{R}}_k = \mathbf{S}_{R_k} \Lambda_{m,\sigma_1}^{-1} \mathbf{S}_{R_k}^T$ and $\bar{\mathbf{R}}_k = \mathbf{S}_{R_k} \Lambda_{m,\sigma_2}^{-1} \mathbf{S}_{R_k}^T$.

Therefore, the conditional probability density distribution of $p(\mathbf{y}_k|\mathbf{x}_k, \alpha_k, t_k)$ and $p(\mathbf{x}_k|\mathbf{y}_{1:k-1}, \beta_k, s_k)$ can be formulated as

$$p(\mathbf{y}_k|\mathbf{x}_k, \alpha_k, t_k) = N(\mathbf{y}_k; \mathbf{H}_k\mathbf{x}_k, \tilde{\mathbf{R}}_k)^{t_k} N(\mathbf{y}_k; \mathbf{H}_k\mathbf{x}_k, \bar{\mathbf{R}}_k)^{(1-t_k)} p(t_k|\alpha_k) p(\alpha_k) \text{ s.t. } t_k \in \{0, 1\} \tag{31}$$

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}, \beta_k, s_k) = N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \tilde{\mathbf{P}}_{k|k-1})^{s_k} N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \bar{\mathbf{P}}_{k|k-1})^{(1-s_k)} p(s_k|\beta_k) p(\beta_k) \text{ s.t. } s_k \in \{0, 1\} \tag{32}$$

According to the likelihood PDF derived above and the Bayesian theorem, the joint PDF can be given as follows:

$$p(\Theta_k, \mathbf{y}_{1:k}) = p(\mathbf{y}_{1:k-1}) N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \tilde{\mathbf{P}}_{k|k-1})^{s_k} N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \bar{\mathbf{P}}_{k|k-1})^{(1-s_k)} N(\mathbf{y}_k; \mathbf{H}_k\mathbf{x}_k, \tilde{\mathbf{R}}_k)^{t_k} N(\mathbf{y}_k; \mathbf{H}_k\mathbf{x}_k, \bar{\mathbf{R}}_k)^{(1-t_k)} \alpha_k^{t_k} (1 - \alpha_k)^{(1-t_k)} \beta_k^{s_k} (1 - \beta_k)^{(1-s_k)} Be(\alpha_k; a_0, b_0) Be(\beta_k; c_0, d_0) \tag{33}$$

where $\Theta_k \triangleq \{\mathbf{x}_k, s_k, t_k, \alpha_k, \beta_k\}$ contains the state variable \mathbf{x}_k that needs to be estimated, and the Beta-Bernoulli variables $\{\alpha_k, \beta_k\}, \{s_k, t_k\}$ are used for the inference of mixing probability parameters. Therefore, to obtain the estimated value of Θ using the VB inference, according to prior work [12,33], the approximate posterior PDF of the element in Θ needs to satisfy the following equation:

$$\log q(\theta) = E_{\Theta_k^{-\theta}}[\log p(\Theta_k, \mathbf{y}_{1:k})] + C_\theta \tag{34}$$

where θ is an element of Θ_k , $\Theta^{-\theta}$ means the remain elements of Θ except θ , and C_θ is a constant with respect to θ . As (34) cannot be solved analytically, fixed-point iteration is required to achieve an approximate solution. By expanding (34), the following can be obtained:

$$\begin{aligned} \log p(\Theta_k, \mathbf{y}_{1:k}) &= (1 - s_k) \log N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) + s_k \log N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \tilde{\mathbf{P}}_{k|k-1}) \\ &+ (1 - t_k) \log N(\mathbf{y}_k; \mathbf{H}_k \mathbf{x}_k, \hat{\mathbf{R}}_k) + t_k \log N(\mathbf{y}_k; \mathbf{H}_k \mathbf{x}_k, \tilde{\mathbf{R}}_k) + t_k \log \alpha_k + (1 - t_k) \log(1 - \alpha_k) \\ &+ s_k \log \beta_k + (1 - s_k) \log(1 - \beta_k) + \log \text{Be}(\alpha_k; a_0, b_0) + \log \text{Be}(\beta_k; c_0, d_0) + C_{\Theta_k} \end{aligned} \tag{35}$$

Initialize the parameters a_0, b_0, c_0, d_0 and calculate $E^{(i+1)}[t_k] = a_0 / (a_0 + b_0)$, $E^{(i+1)}[s_k] = c_0 / (c_0 + d_0)$, note that $E^{(0)}[\log(\alpha_0)] = \psi(a_0) - \psi(a_0 + b_0)$, $E^{(0)}[\log(\beta_0)] = \psi(c_0) - \psi(c_0 + d_0)$, where ψ represents the digamma function.

By exploiting (35), the estimation of the unknown parameters in Θ_k is implemented by fixed-point iteration loop in (a)–(c).

(a) Let $\theta = \mathbf{x}_k$, utilizing (34) in (35), and $\log q^{(i+1)}(\mathbf{x}_k)$ can be rewritten as

$$\begin{aligned} \log q^{(i+1)}(\mathbf{x}_k) &= E^{(i)}[s_k] \log N(\mathbf{x}_k; \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)}, \hat{\mathbf{P}}_{k|k}^{(i)}) + (1 - E^{(i)}[s_k]) \log N(\mathbf{x}_k; \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)}, \tilde{\mathbf{P}}_{k|k}^{(i)}) \\ &E^{(i)}[t_k] \log N(\mathbf{y}_k; \mathbf{H}_k \mathbf{x}_k, \hat{\mathbf{R}}_k^{(i)}) + (1 - E^{(i)}[t_k]) \log N(\mathbf{y}_k; \mathbf{H}_k \mathbf{x}_k, \tilde{\mathbf{R}}_k^{(i)}) + C_{\mathbf{x}_k} \end{aligned} \tag{36}$$

where i represents the i th fixed-point iterations. Using $E^{(i+1)}[s_k]$ and $E^{(i+1)}[t_k]$ replace the mixing probability ρ in (21) and then calculate the conditional PDF $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ and $p(\mathbf{y}_k | \mathbf{x}_k)$ by proposed mixture correntropy as

$$q^{(i+1)}(\mathbf{x}_k | \mathbf{y}_{1:k-1}) q^{(i+1)}(\mathbf{y}_k | \mathbf{x}_k) = N(\mathbf{x}_k; \hat{\mathbf{x}}_{k-1}, \hat{\mathbf{P}}_{k|k-1}) N(\mathbf{y}_k; \mathbf{H}_k \mathbf{x}_k, \hat{\mathbf{R}}_k) \tag{37}$$

where the modified prediction error covariance matrix $\hat{\mathbf{P}}_{k|k-1}^{(i+1)}$ and the modified measurement noise covariance matrix $\hat{\mathbf{R}}_k^{(i+1)}$ are used. After deformation, the posterior PDF $q^{(i+1)}(\mathbf{x}_k)$ can be formulated as

$$q^{(i+1)}(\mathbf{x}_k) \propto N(\mathbf{x}_k; \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}^{(i)}, \hat{\mathbf{P}}_{k|k-1}^{(i)}) N(\mathbf{y}_k; \mathbf{H}_k \mathbf{x}_k, \hat{\mathbf{R}}_k^{(i)}) \tag{38}$$

According to (38), $q^{(i+1)}(\mathbf{x}_k)$ is updated using a nominal Gaussian distribution as $q^{(i+1)}(\mathbf{x}_k) = N(\mathbf{x}_k; \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i+1)}, \hat{\mathbf{P}}_{k|k}^{(i+1)})$. Then $\hat{\mathbf{x}}_{k|k}^{(i+1)}$ and the corresponding estimation error covariance matrix $\hat{\mathbf{P}}_{k|k}^{(i+1)}$ are obtained by the measurement update of the KF as shown in (10)–(12) with $\hat{\mathbf{P}}_{k|k-1}^{(i+1)}$ and $\hat{\mathbf{R}}_k^{(i+1)}$.

(b) As for the expectation of the Bernoulli distribution, $E^{(i+1)}[s_k]$ and $E^{(i+1)}[t_k]$ can be updated as follows:

$$E^{(i+1)}[s_k] = \frac{\Pr^{(i+1)}(s_k = 1)}{\Pr^{(i+1)}(s_k = 1) + \Pr^{(i+1)}(s_k = 0)} \tag{39}$$

$$E^{(i+1)}[t_k] = \frac{\Pr^{(i+1)}(t_k = 1)}{\Pr^{(i+1)}(t_k = 1) + \Pr^{(i+1)}(t_k = 0)} \tag{40}$$

By setting $\theta = s_k$ and $\theta = t_k$, and then updating $q^{(i+1)}(s_k)$ and $q^{(i+1)}(t_k)$ using the Bernoulli distribution, the probabilities of s_k and t_k being 1 are given by:

$$\Pr(t_k = 1) = \Delta_1^{i+1} \exp\{E^{(i)}[\log \alpha_k] + 0.5 \text{tr}(\log \mathbf{A}_{m, \sigma_1}) - 0.5 \text{tr}(\mathbf{A}_k^{i+1} \hat{\mathbf{R}}_k^{-1})\} \tag{41}$$

$$\Pr(s_k = 1) = \Delta_2^{i+1} \exp\{E^{(i)}[\log \beta_k] + 0.5 \text{tr}(\log \mathbf{A}_{p, \sigma_1}) - 0.5 \text{tr}(\mathbf{B}_k^{i+1} \hat{\mathbf{P}}_{k|k-1}^{-1})\} \tag{42}$$

where Δ is a normalizing constant, and two auxiliary variables \mathbf{A}, \mathbf{B} are used to evaluate the prior probability covariance and measurement covariance as

$$\mathbf{A}_k^{(i+1)} = \mathbf{E}^{(i+1)}[(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k)(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k)^T] \tag{43}$$

$$\mathbf{B}_k^{(i+1)} = \mathbf{E}^{(i+1)}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T] \tag{44}$$

To implement the fixed-point iteration, (43) and (44) is approximated as

$$\mathbf{A}_k^{(i+1)} = (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i+1)})(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i+1)})^T + \mathbf{H}_k \mathbf{P}_{k|k}^{i+1} \mathbf{H}_k^T \tag{45}$$

$$\mathbf{B}_k^{(i+1)} = \mathbf{P}_{k|k}^{(i+1)} + (\hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k-1})^T \tag{46}$$

(c) Next, $q^{(i+1)}(\alpha_k)$ and $q^{(i+1)}(\beta_k)$ are updated using the Beta distribution as shown in (27) and (28). Take the PDF of the Beta distribution into derivation, and, by setting $\theta = \alpha_k$ and $\theta = \beta_k$, then

$$q^{(i+1)}(\alpha_k) = (a_0 - \mathbf{E}^{(i+1)}[t_k]) \log(\alpha_k) + (b_0 + \mathbf{E}^{(i+1)}[t_k] - 1) \log(1 - \alpha_k) + C_\alpha \tag{47}$$

$$q^{(i+1)}(\beta_k) = (c_0 - \mathbf{E}^{(i+1)}[s_k]) \log(\beta_k) + (d_0 + \mathbf{E}^{(i+1)}[s_k] - 1) \log(1 - \beta_k) + C_\beta \tag{48}$$

where $q^{(i+1)}(\alpha_k)$ and $q^{(i+1)}(\beta_k)$ can be updated with new shape parameters, therefore, the shape parameters of the Beta prior distribution can be updated as follows:

$$a_k^{(i+1)} = a_0 + \mathbf{E}^{(i+1)}[t_k] \tag{49}$$

$$b_k^{(i+1)} = b_0 - \mathbf{E}^{(i+1)}[t_k] + 1 \tag{50}$$

$$c_k^{(i+1)} = c_0 + \mathbf{E}^{(i+1)}[s_k] \tag{51}$$

$$d_k^{(i+1)} = d_0 - \mathbf{E}^{(i+1)}[s_k] + 1 \tag{52}$$

After completing c), checking the convergence of the iteration according to the preset error threshold, if $\frac{\|\hat{\mathbf{x}}_k^{(i+1)} - \hat{\mathbf{x}}_k^{(i)}\|}{\|\hat{\mathbf{x}}_k^{(i)}\|} < \varepsilon$ then stop the iteration and output the result. Otherwise, return to (a) for the next iteration.

In addition to the previous state and measurement vector, to begin the algorithm, the proposed filter only needs two kernel parameters and the initial prior beta shape parameters for the mixture of Gaussian correntropy. Among them, the kernel parameters of the Gaussian correntropy subfunction are more likely to be chosen by experience. In most cases, they have minimal impact on the filtering performance. This conclusion will be confirmed in the subsequent simulations.

On the other hand, the chosen of prior Beta distribution parameters needs a simple discussion. By combining the characteristics of the Beta distribution, the initial mixing probability ρ is determined by its prior shape parameters. If $a_0 = c_0 = 0$ or $b_0 = d_0 = 0$, the proposed filter converges to the original MCKF. In order to reduce input parameters and simplify the algorithm, it is generally assumed that $a_0 = c_0$ and $b_0 = d_0$ in most cases.

In this work, two Gaussian correntropies with different kernel parameters are used to cope with the stable filtering process and the process corrupted by dynamic abnormal errors, such as impulsive noise disturbance. In order to ensure filtering accuracy and stability, the prior parameters are not arbitrarily set, but obey a certain regularity. For example, in an ideal Gaussian noise environment, we have $\mathbf{E}^{(j+1)}[t_k] \rightarrow 1$ and then the performance of the mixture correntropy converge to the Gaussian correntropy with a large kernel parameter, so that it has basic robustness and retains convergence to KF as much as possible, expanding (41) and we can find that

$$\begin{aligned}
 E^{(i+1)}[t_k] &= \frac{p^{(i+1)}(t_k = 1)}{p^{(i+1)}(t_k = 1) + p^{(i+1)}(t_k = 0)} \\
 &= \frac{1}{1 + p^{(i+1)}(t_k = 0)/p^{(i+1)}(t_k = 1)} \\
 &= \frac{1}{1 + \exp(E^{(i+1)}(1 - \alpha_0) - E^{(i+1)}(\alpha_0) + C_k)} \\
 &= \frac{1}{1 + \exp(\psi(b_0) - \psi(a_0) + C_k)} \rightarrow 1
 \end{aligned}
 \tag{53}$$

where C_k represents the terms independent of a_0 and b_0 . According to (41) and (42), if the distributions of measurement and state noise nearly meet the ideal Gaussian assumption is needed.

In contrast, for the filtering process corrupted by severe impulsive noise, the mixing probability is likely to be redistributed appropriately to enhance the robustness to abnormal noise, and then

$$\begin{aligned}
 E^{(i+1)}[t_k] &= \frac{1}{1 + p^{(i+1)}(t_k = 0)/p^{(i+1)}(t_k = 1)} \rightarrow 0 \\
 &\Rightarrow \exp(\psi(b_0) - \psi(a_0) + C_k) \gg 0
 \end{aligned}
 \tag{54}$$

As the definition of Gaussian correntropy Equation (2), if $\sigma_1 > \sigma_2$, for the residual term \mathbf{e}_i with significant abnormality, there is a great of difference between $G_{\sigma_1}(\mathbf{e}_i)$ and $G_{\sigma_2}(\mathbf{e}_i)$, and $G_{\sigma_1}(\mathbf{e}_i) \gg G_{\sigma_2}(\mathbf{e}_i)$. Therefore, it is further inferred that C_k increases significantly and then $\psi(b_0) - \psi(a_0) + C_k \gg 0$ at this time, where $\psi(b_0) - \psi(a_0)$ can be taken as a constant term to regulate the transition of $E^{(i+1)}[t_k]$, and $a_0 > b_0$ still applies to this case.

Therefore, in this work, the parameters can be chosen within the range $0.95 \leq a_0 \leq 0.85$ and $b_0 = 1 - a_0$, which can be used as a typical parameter configuration. On this basis, it may sometimes be useful to make a minor adjustment on the initial parameter configurations according to the specified application. This will contributes to further improving the algorithm’s performance in complex environments. The related results are shown in the later simulation.

4. Performance Evaluations and Analysis

4.1. Example I: 2-D Moving Target Tracking Model

Consider a two-dimensional (2D) moving target tracking system as

$$\mathbf{F} = \begin{bmatrix} 1 & \frac{\sin(wT)}{w} & 0 & -\frac{(1-\cos(wT))}{w} \\ 0 & \cos(wT) & 0 & -\sin(wT) \\ 0 & -\frac{(1-\cos(wT))}{w} & 1 & \frac{\sin(wT)}{w} \\ 0 & \sin(wT) & 0 & \cos(wT) \end{bmatrix}, \mathbf{\Gamma} = \begin{bmatrix} 0.5T^2 & 0 \\ T & 0 \\ 0 & 0.5T^2 \\ T & 0 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{55}$$

in which $\mathbf{\Gamma}$ represents the process noise gain matrix, $T = 0.2$, $w = 0.2$, and the total simulation steps $N = 200/T$. It is assumed that $\mathbf{Q}_k = 0.1\mathbf{I}_2$ and $\mathbf{R}_k = 10\mathbf{I}_2$. The true initial state $\mathbf{x}_0 = [1 \ 1 \ 1 \ 1]^T$, the initial state estimation $\hat{\mathbf{x}}_0 = [0 \ 0 \ 0 \ 0]^T$ and the error covariance matrix $\mathbf{P}_{0|0} = \mathbf{I}_4$. In addition to the KF, the HKF with loss function parameter $r = 1.345$; the MCKF with typical kernel parameters $\sigma = 2, \sigma = 3, \sigma = 5$ and $\sigma = 9$, are used for comparison. For the algorithm proposed in this paper, we chose $\sigma_1 = 9, \sigma_2 = 3$ and $a_0 = 0.9$ as the default initial parameters. The maximum iteration times $N_m = 10$ for the robust filters. The numerical test was coded with MATLAB and executed on a computer with Intel Core i7-9700 CPU @3.0 GHz.

In order to evaluate the filtering performance, the root mean square errors (RMSE) and averaged RMSE (ARMSE) are chosen as evaluation indicators of position and velocity estimation, which are defined as follows:

$$\begin{cases} \text{RMSE}(k) \triangleq \sqrt{\frac{1}{M} \sum_{s=1}^M ((x_k^s - \hat{x}_k^s)^2 + (y_k^s - \hat{y}_k^s)^2)} \\ \text{ARMSE} \triangleq \sqrt{\frac{1}{MN} \sum_{k=1}^N \sum_{s=1}^M ((x_k^s - \hat{x}_k^s)^2 + (y_k^s - \hat{y}_k^s)^2)} \end{cases} \quad (56)$$

where (x_k^s, y_k^s) and $(\hat{x}_k^s, \hat{y}_k^s)$ are the true and estimated positions(or velocity) at the k th step of the s th Monte Carlo run, and $\text{RMSE}_{pos}(k)$ is the RMSE of position at k step. The ARMSEs of position and velocity are denoted as ARMSE_{pos} and ARMSE_{vel} , respectively [12,21].

In order to verify whether the proposed filtering algorithm improve the solution of accuracy degradation problem of original MCKF in no-stationary non-Gaussian noise conditions, the simulation was divided into two stages, and the noises with different distributions were used to test the performance of the proposed filter. To simulate the non-Gaussian noise sequence contaminated by impulsive noise interference, the Gaussian mixture distribution model with specific parameters are used to generate the noise.

The number of Monte Carlo runs was $M = 1000$ and the specific noise parameters were set as follows:

$$\begin{cases} \mathbf{w}_k \sim 0.95\text{N}(0, \mathbf{Q}_k) + 0.05\text{N}(0, 100\mathbf{Q}_k), \text{stage1} : 1 < t \leq N/2 \\ \mathbf{w}_k \sim 0.90\text{N}(0, \mathbf{Q}_k) + 0.10\text{N}(0, 100\mathbf{Q}_k), \text{stage2} : N/2 < t \leq N \\ \mathbf{v}_k \sim 0.90\text{N}(0, \mathbf{R}_k) + 0.10\text{N}(0, 100\mathbf{R}_k), \text{stage1} : 1 < t \leq N/2 \\ \mathbf{v}_k \sim 0.95\text{N}(0, \mathbf{R}_k) + 0.05\text{N}(0, 100\mathbf{R}_k), \text{stage2} : N/2 < t \leq N \end{cases}$$

The RMSEs of the position and velocity from different filters are shown in Figures 1 and 2. In addition, the ARMSEs from different filters at all stages are listed in Table 1 for comparison. In the intervals $(0, N/2]$ and $(N/2, N]$, the process and measurement vectors were contaminated by non-Gaussian heavy-tailed noise with different distributions. The plots show that the RMSEs from different filters have obvious differences, and the best kernel size for the MCKF at each stage varied accordingly. For example, MCKF2 ($\sigma = 3$) and MCKF3 ($\sigma = 5$) achieved higher accuracy, respectively, at these two stages; nevertheless, a too large or small kernel size also resulted in the filtering performance to degrade or even diverge, as shown by MCKF1 ($\sigma = 2$) and MCKF4 ($\sigma = 9$).

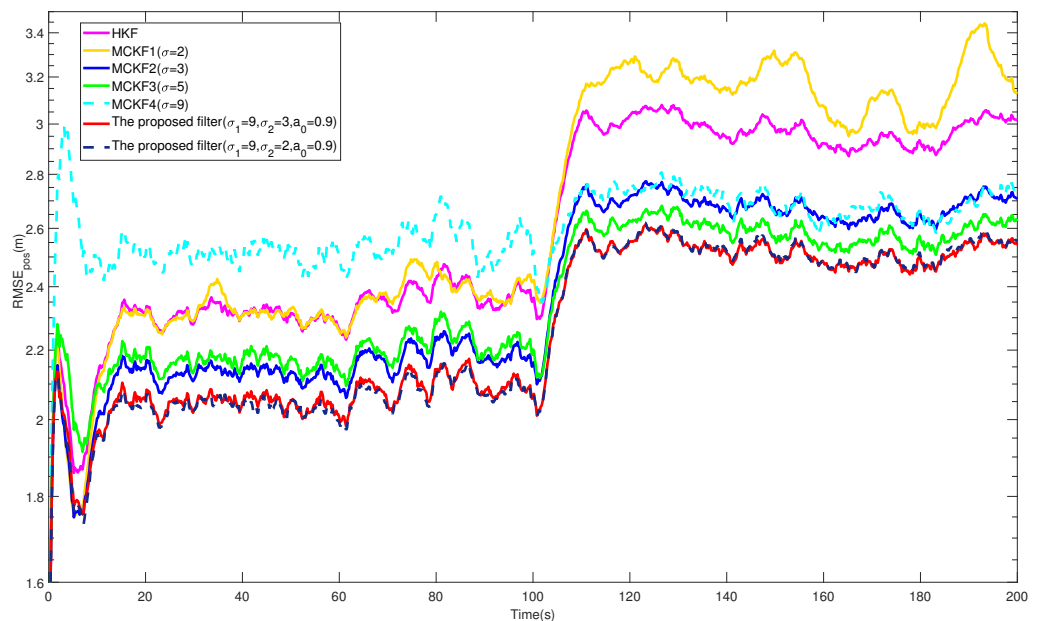


Figure 1. RMSEs of the position from different filters.

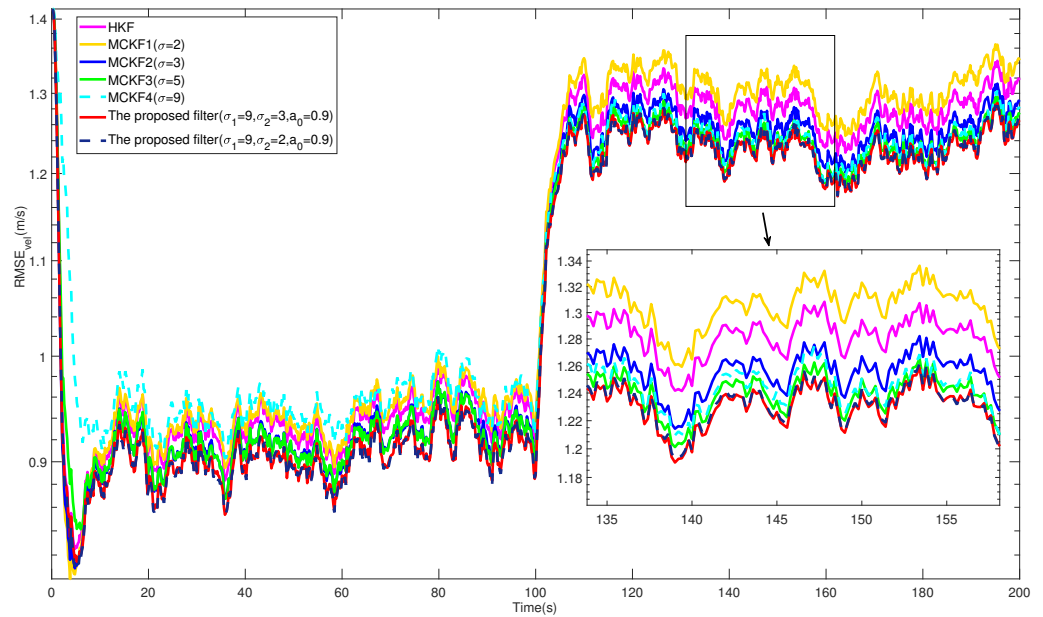


Figure 2. RMSEs of the velocity from different filters.

Table 1. ARMSEs of different filters at each stage.

Filters	Stage 1		Stage 2		All Stages	
	Position (m)	Velocity (m/s)	Position (m)	Velocity (m/s)	Position (m)	Velocity (m/s)
KF	4.014	1.207	3.420	1.302	3.729	1.255
HKF	2.347	0.970	2.945	1.268	2.663	1.129
MCKF1 ($\sigma = 2$)	2.346	0.977	3.081	1.290	2.738	1.144
MCKF2 ($\sigma = 3$)	2.167	0.953	2.659	1.242	2.426	1.107
MCKF3 ($\sigma = 5$)	2.201	0.953	2.572	1.226	2.394	1.098
MCKF4 ($\sigma = 9$)	2.552	0.993	2.683	1.231	2.618	1.118
IMMCKF1 ($\sigma_1 = 9, \sigma_2 = 3$)	2.084	0.939	2.503	1.218	2.303	1.087
IMMCKF2 ($\sigma_1 = 9, \sigma_2 = 2$)	2.078	0.939	2.508	1.218	2.303	1.088

In comparison, the proposed filter had the lowest estimation error at each stage, especially in the case of noise distribution change, the results show that the proposed VB interference method played a positive role in the filtering process, which takes into account both stability performance and robustness to non-Gaussian noise. Therefore, by comparing the filtering performance of all stages, the superiority of the proposed algorithm is preliminarily verified.

For the existing robust filters with fixed parameters, such as the classic MCKF or HKF, the filtering parameters can be obtained by experience or trial and error methods, which are more applicable to stationary noise conditions. However, as mentioned before, the classic MCKF does not always achieve satisfactory estimation accuracy in non-stationary noise environment. In order to show this change more concretely, the process and measurement noise distribution can be expressed as $\mathbf{w}_k \sim (1 - p_1)N(0, \mathbf{Q}_k) + p_1N(0, 100\mathbf{Q}_k)$ and $\mathbf{v}_k \sim (1 - p_2)N(0, \mathbf{R}_k) + p_2N(0, 100\mathbf{R}_k)$, where p_1 and p_2 represent the outlier percentage of noise.

In Figures 3 and 4, the outlier percentage of process noise was fixed as $p_1 = 0.05$, and the ARMSEs of the different filters varying with $p_2: 0 \sim 0.15$ are shown. The overall performance of the ARMSEs showed an increasing trend, which means that as the proportion of impulsive noise increased, the estimation accuracy of the filters also decreased. In this test, the MCKFs with different kernel parameters obtained significant filtering accuracy differences. This means that, under the interference of different non-Gaussian noise, the MCKFs with a fixed kernel parameter could not ensure reliable estimation results, and the MCKFs lack enough self-adaptive ability.

In addition, the fixed-parameters MMCKF without variational Bayesian iteration also taken for comparison. The comparison shows that the results of MMCKF always converges

to the MCKF of a specific fixed parameter. Therefore, although it achieves better accuracy than MCKF in some cases, it does not avoid similar accuracy degradation problem as MCKF. Since the filtering results of MMCKF are similar to the typical MCKF with specific parameters, therefore, the comparison and discussion of the simulation mainly focus on the classic MCKF.

Compared with other filtering results, it can be concluded that the proposed filter has better performance than other existing algorithms. On the one hand, the algorithm does not lose much optimality of estimation caused by changing the objective function of KF, and on the other hand, it retains robustness to the increased impulsive noise probability. Therefore, the proposed filter further demonstrates its performance advantage in various non-Gaussian noise environments.

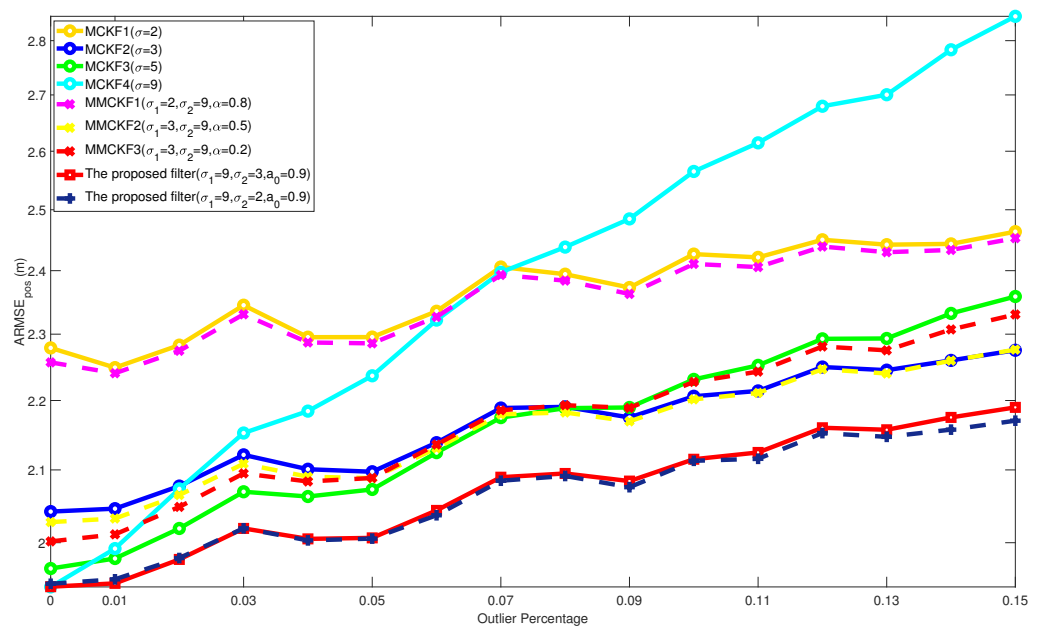


Figure 3. ARMSEs of position from different filters with varying p_2 .

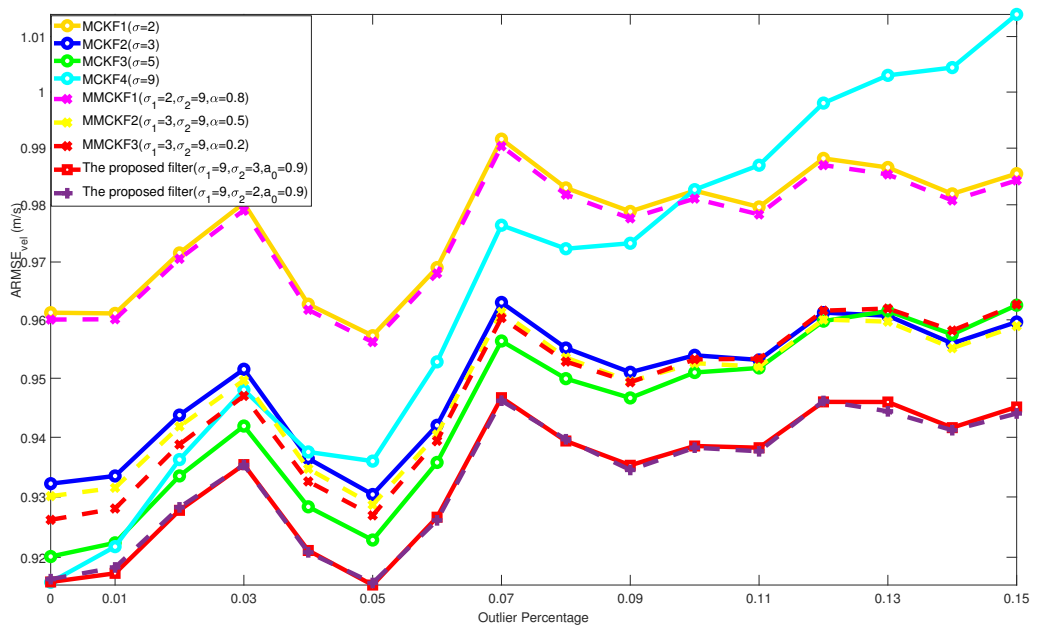


Figure 4. ARMSEs of velocity from different filters with varying p_2 .

In addition, the simulation tests were performed to compare the filtering performance with different initial parameter configurations. For the proposed IMMCKF, the Gaussian

correntropy $G_{\sigma_1}(\mathbf{e})$, $G_{\sigma_2}(\mathbf{e})$ were mixed to generate the mixture correntropy, where σ_1 and σ_2 represent the specific kernel parameters, and $\sigma_1 > \sigma_2$. It has been proved that Gaussian correntropy with a smaller kernel bandwidth is more sensitive to impulsive noise. However, this also degrades the filtering stability and might lead to divergence.

Table 2 lists the RMSEs of the proposed filter with different σ_1 and σ_2 . In general, the filtering results of the proposed filter with different σ_1 and σ_2 were not significant. Within a certain range, the filtering accuracy improved with the increase of σ_1 , as it provided better compatibility to Gaussian noises for the filtering algorithm. As too large of a kernel bandwidth may reduce Gaussian correntropy's robustness to impulsive noise, σ_1 much greater than σ_2 would not be considered as the parameter set. Therefore, generally speaking, it is easy to choose appropriate σ_1 and σ_2 for the proposed algorithm.

Table 2. ARMSEs of the proposed filter with different σ_1 and σ_2 .

Filters	Stage 1		Stage 2		All Stages	
	Position (m)	Velocity (m/s)	Position (m)	Velocity (m/s)	Position (m)	Velocity (m/s)
KF	4.014	1.207	3.420	1.302	3.729	1.255
IMMCKF1 ($\sigma_1 = 6, \sigma_2 = 3$)	2.117	0.943	2.530	1.221	2.332	1.091
IMMCKF2 ($\sigma_1 = 6, \sigma_2 = 2$)	2.108	0.942	2.532	1.222	2.330	1.091
IMMCKF3 ($\sigma_1 = 9, \sigma_2 = 3$)	2.084	0.939	2.503	1.218	2.303	1.087
IMMCKF4 ($\sigma_1 = 9, \sigma_2 = 2$)	2.078	0.939	2.508	1.218	2.303	1.088
IMMCKF5 ($\sigma_1 = 12, \sigma_2 = 3$)	2.074	0.938	2.494	1.217	2.294	1.086
IMMCKF6 ($\sigma_1 = 12, \sigma_2 = 2$)	2.054	0.936	2.502	1.218	2.289	1.086

In this case, several potential shape parameter options were used for comparison. Table 3 lists the ARMSEs of position and velocity from the proposed filter with different initial shape parameters. As demonstrated, the estimation accuracy can be improved somewhat by fine-tuning the parameter within the given range, but in general, it does not significantly affect the overall filtering performance. In summary, for the state estimation problem in non-stationary noises, the filter proposed in this work has good compatibility with the initial parameter sets.

Table 3. ARMSEs of the proposed filter with different a_0 .

Filters	Stage 1		Stage 2		All Stages	
	Position(m)	Velocity (m/s)	Position (m)	Velocity (m/s)	Position (m)	Velocity (m/s)
KF	4.014	1.207	3.420	1.302	3.729	1.255
IMMCKF1 ($a_0 = 0.94$)	2.118	0.943	2.514	1.218	2.324	1.089
IMMCKF2 ($a_0 = 0.92$)	2.095	0.940	2.505	1.218	2.309	1.088
IMMCKF3 ($a_0 = 0.90$)	2.084	0.939	2.503	1.218	2.303	1.087
IMMCKF4 ($a_0 = 0.88$)	2.079	0.938	2.504	1.218	2.301	1.087
IMMCKF5 ($a_0 = 0.86$)	2.077	0.938	2.510	1.219	2.304	1.088

In order to balance the computational efficiency and the filtering performance, it is necessary to choose a reasonable number of iterations for the proposed filter. Figure 5 shows the ARMSEs of filters with different numbers of iterations. For comparison, several filtering results with similar performance in the simulation are also taken into accounts. It can be concluded that the accuracy of the proposed algorithm can be greatly improved after several iterations, and the filtering result also gradually converged. In practical applications, the increase of iteration times brings more computational burden. As shown from Figure 5, when the number of iterations N_m is 3~5, the filtering algorithm has obtained satisfactory estimation accuracy. Therefore, when considering the accuracy and calculation burden factors, the number of iterations N_m set as 3~5 is reasonable.

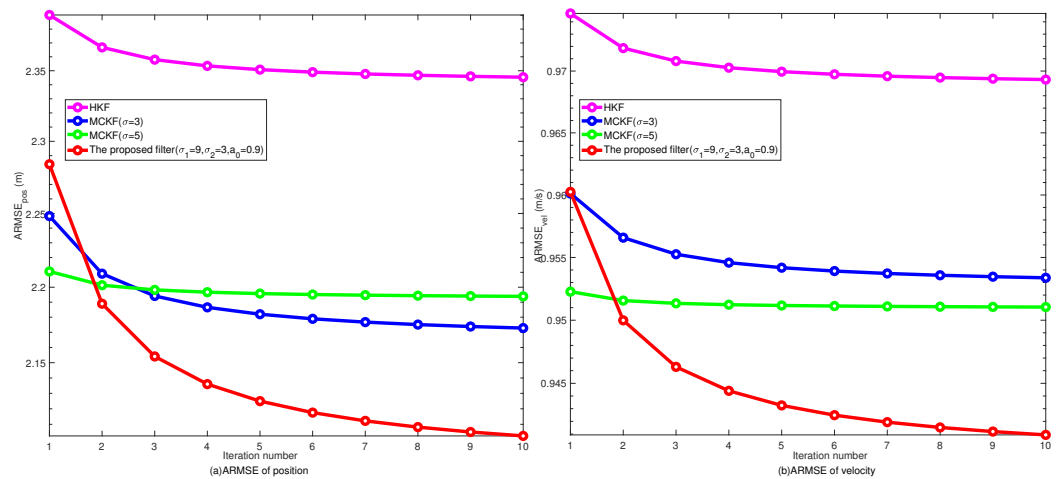


Figure 5. ARMSEs versus iteration number from different filters.

The implementation time of the proposed filter and existing filters in a single-step run with $N_m = 1$ is evaluated. It shows that KF (0.0054 ms) is the fastest, MCKF (0.0253 ms), HKF (0.0251 ms) and MMCKF (0.0212 ms) has similar computational burden and the proposed IMMCKF (0.0490 ms). Compared with the MCKF and MMCKF, the computation complexity of the proposed filter increased simultaneously due to the additional variational Bayesian iterations. In view of this, the additional computation cost of the IMMCKF can be compensated by adjusting the number of iterations. Therefore, it is still feasible to apply this algorithm in real-time applications.

4.2. Example II: INS/GPS Integrated Navigation System

To validate the effectiveness and superiority of the proposed algorithm in this paper, the experimental data collected in a vehicle-mounted INS/GPS integrated navigation experiment was used for the test. The experiment was carried out in the campus of Harbin Engineering University, the test trajectory is shown in Figure 6. A low-cost MEMS-IMU based INS/GPS integrated navigation system is used to provide the navigation data.

An INS/GNSS integration navigation system includes a self-made navigation-grade fiber optic strap-down INS and a double-antenna GPS receiver is used for reference. The initial velocity and position of INS/GPS are obtained directly from GPS measurement, the initial level attitude information is acquired from the alignment results of high accuracy SINS. In the experimental test, the car moves along a bumpy road, and the GPS might work abnormally due to the occlusion of trees and buildings.

The sampling frequencies of the low-cost IMU and GPS were 100 Hz and 1 Hz, respectively. The loosely coupled configuration is used in the integration, and a linear closed-loop feedback scheme as [15,23] is used. The state variables vector is defined as $\mathbf{x}_k = [\varphi^n \ \delta\mathbf{v}^n \ \delta\mathbf{p}^n \ \varepsilon^b \ \nabla^b]^T$, where φ^n , $\delta\mathbf{v}^n$, $\delta\mathbf{p}^n$ denote the attitude error, velocity error and position error expressed in the n-frame, respectively. The bias of the gyro and accelerometer are $\varepsilon^b = 10^\circ/\text{h}$ and $\nabla^b = 100 \mu\text{g}$, respectively. The initial estimation was $\hat{\mathbf{x}}_{0|0} = \mathbf{0}_{15 \times 1}$. The process noise vector is expressed as $\mathbf{w} = [w_{gx}^b \ w_{gy}^b \ w_{gz}^b \ w_{ax}^b \ w_{ay}^b \ w_{az}^b]^T$, where $w_g^b = 1^\circ/\sqrt{\text{h}}$ and $w_a^b = 1500 \mu\text{g}/\sqrt{\text{Hz}}$, respectively, and the measurement noise covariance matrix $\mathbf{R} = \text{diag}(0.5 \text{ m/s} \cdot \mathbf{I}_3 \ 1.5 \text{ m} \cdot \mathbf{I}_3)$. The initial position and velocity errors were set as $\delta\mathbf{p} = [1.5 \text{ m} \ 1.5 \text{ m} \ 1.5 \text{ m}]^T$, $\delta\mathbf{v}^n = [0.1 \text{ m/s} \ 0.1 \text{ m/s} \ 0.1 \text{ m/s}]^T$, and attitude errors $\phi = [0.1^\circ \ 0.1^\circ \ 0.1^\circ]^T$.

The filter only performs time update while there are no GPS outputs. To compare the robustness of different filters to non-Gaussian measurement noise interference, inspired by the scheme in [34,35], the gross errors were added artificially to the measurement data. In the integrated navigation system, the velocity and position state variables are susceptible to external interference. Therefore, the velocity and position errors are used for comparison.



Figure 6. The test trajectory of the vehicle.

For comparison, several typical filtering results are shown in Figures 7 and 8. It can be seen that most filters obtain similar results in stable periods, as the measurement output is reliable in these periods. For the existing robust filters with fixed-parameter, due to the uncertain interference factors in practical application. In this example, both MCKF1 ($\sigma = 2$) and MCKF2 ($\sigma = 3$) have a divergence trend in the filtering process. In contrast, MCKFs with larger kernel parameters and the proposed filter have more stable performance.

In this example, the measurement sequence is disturbed by impulsive noise with a certain probability. It is evident from the Figures that the estimation accuracy of KF is seriously corrupted during filtering. Generally, a small kernel size can be more effective in attenuating the measurement outliers, but at the same time, it inevitably decreases the filter's stability, e.g., MCKF1 ($\sigma = 2$) failed to obtain reliable filtering results due to divergence. By comparing the robustness of MCKF3 ($\sigma = 5$), MCKF4 ($\sigma = 9$), and other algorithms to non-Gaussian noise, it can be concluded that, for existing MCKFs, it is difficult to obtain both stable and robustness estimation results.

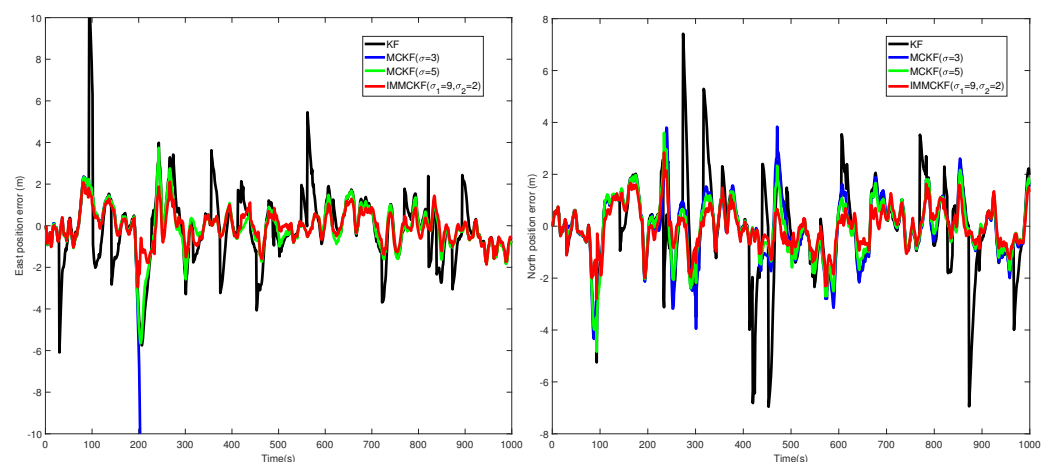


Figure 7. The position errors from different filters in non-Gaussian noises.

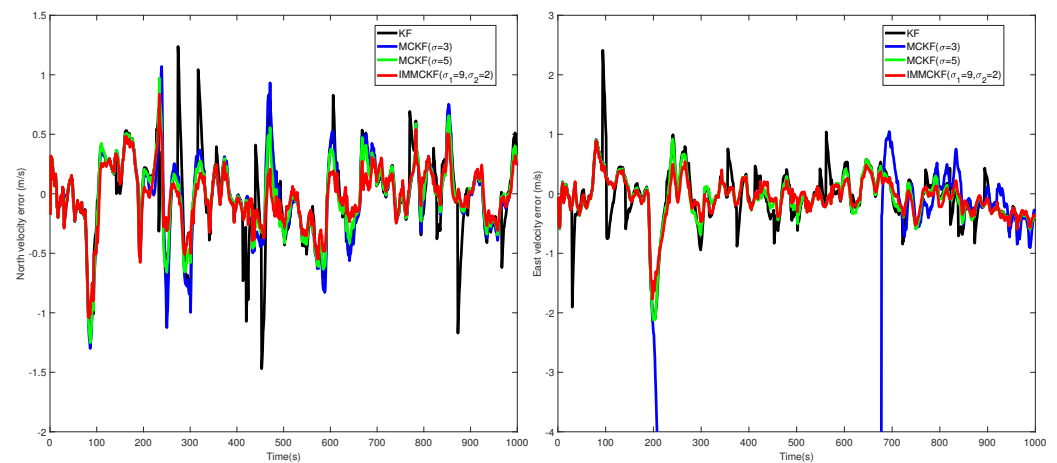


Figure 8. The velocity errors from different filters in non-Gaussian noises.

The results listed in Table 4 confirmed that the filter proposed in this work solves the problem well. NaN represents invalid output due to filter divergence. Compared with other robust filters, the proposed filter achieves better filtering results by taking into account both the robustness and stability of the filter, which demonstrated the conclusions of the previous simulation example.

Table 4. The RMSEs of position, velocity from different filters.

Filtering Algorithms	PosE (m)	PosN (m)	PosU (m)	VelE (m/s)	VelN (m/s)	VelU (m/s)
KF	1.736	1.726	0.799	0.466	0.363	0.118
HKF	1.817	1.428	0.681	0.488	0.320	0.112
MCKF1 ($\sigma = 2$)	NaN	NaN	0.675	NaN	NaN	0.117
MCKF2 ($\sigma = 3$)	NaN	1.187	0.671	NaN	0.333	0.112
MCKF3 ($\sigma = 5$)	1.069	1.061	0.671	0.379	0.290	0.111
MCKF4 ($\sigma = 9$)	1.121	1.111	0.674	0.385	0.300	0.111
MMCKF1 ($\sigma_1 = 2, \sigma_2 = 9, \alpha = 0.8$)	NaN	NaN	0.711	NaN	NaN	0.120
MMCKF2 ($\sigma_1 = 3, \sigma_2 = 9, \alpha = 0.5$)	1.221	0.995	0.674	0.400	0.275	0.112
MMCKF3 ($\sigma_1 = 3, \sigma_2 = 9, \alpha = 0.2$)	1.026	1.007	0.673	0.374	0.278	0.111
IMMCKF1 ($\sigma_1 = 9, \sigma_2 = 3$)	0.888	0.903	0.657	0.345	0.261	0.110
IMMCKF2 ($\sigma_1 = 9, \sigma_2 = 2$)	0.751	0.775	0.640	0.319	0.238	0.110

5. Conclusions

In this work, the performance degradation problem of the existing MCKF in non-stationary noises is explored, and a new improved mixture correntropy filtering algorithm is proposed as an effective solution. To cope with the dynamic process where both Gaussian and non-Gaussian noises may occur, the intermediate random variables is used to construct the mixture correntropy. By derivation, the state variables and intermediate parameters are approximated via a variational Bayesian approach. The theoretical derivation and numerical test results show that the proposed method significantly improves upon the existing MCKF algorithms in different conditions, which offers a promising improvement for the robust filtering problem in complex noise environments.

Author Contributions: Conceptualization, X.L.; methodology, Y.G.; software, Y.G.; data curation, Q.M.; writing—original draft preparation, Y.G.; writing—review and editing, Q.M.; supervision, X.L.; project administration, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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.

References

1. Simon, D. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
2. Pieczynski, W.; Desbouvries, F. Kalman filtering using pairwise Gaussian models. In Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings (ICASSP'03), Hong Kong, China, 6–10 April 2003.
3. Ait-El-Fquih, B.; Desbouvries, F. Kalman Filtering in Triplet Markov Chains. *IEEE Trans. Signal Process.* **2006**, *54*, 2957–2963. [[CrossRef](#)]
4. Kulikova, M.V. Gradient-Based Parameter Estimation in Pairwise Linear Gaussian System. *IEEE Trans. Autom. Control* **2017**, *62*, 1511–1517. [[CrossRef](#)]
5. Nemesin, V.; Derrode, S. Robust Blind Pairwise Kalman Algorithms Using QR Decompositions. *IEEE Trans. Signal Process.* **2013**, *61*, 5–9. [[CrossRef](#)]
6. Huber, P.J. *Robust Statistics*; John Wiley & Sons: Hoboken, NJ, USA, 2004; Volume 523.
7. Gandhi, M.A.; Mili, L. Robust Kalman filter based on a generalized maximum-likelihood-type estimator. *IEEE Trans. Signal Process.* **2009**, *58*, 2509–2520. [[CrossRef](#)]
8. Chang, L.; Hu, B.; Chang, G.; Li, A. Robust derivative-free Kalman filter based on Huber's M-estimation methodology. *J. Process. Control* **2013**, *23*, 1555–1561. [[CrossRef](#)]
9. Karlgaard, C.D.; Schaub, H. Huber-based divided difference filtering. *J. Guid. Control Dyn.* **2007**, *30*, 885–891. [[CrossRef](#)]
10. Roth, M.; Özkan, E.; Gustafsson, F. A Student's t filter for heavy tailed process and measurement noise. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 5770–5774.
11. Huang, Y.; Zhang, Y.; Li, N.; Chambers, J. Robust Student-t based nonlinear filter and smoother. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 2586–2596. [[CrossRef](#)]
12. Huang, Y.; Zhang, Y.; Li, N.; Wu, Z.; Chambers, J.A. A novel robust student's t-based kalman filter. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 1545–1554. [[CrossRef](#)]
13. Agamennoni, G.; Nieto, J.I.; Nebot, E.M. Approximate inference in state-space models with heavy-tailed noise. *IEEE Trans. Signal Process.* **2012**, *60*, 5024–5037. [[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. A New Process Uncertainty Robust Student's t Based Kalman Filter for SINS/GPS Integration. *IEEE Access* **2017**, *5*, 14391–14404. [[CrossRef](#)]
16. Izanloo, R.; Fakoorian, S.A.; Yazdi, H.S.; Simon, D. Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise. In Proceedings of the 2016 Annual Conference on Information Science and Systems (CISS), Princeton, NJ, USA, 16–18 March 2016; pp. 500–505.
17. Chen, B.; Liu, X.; Zhao, H.; Principe, J.C. Maximum correntropy Kalman filter. *Automatica* **2017**, *76*, 70–77. [[CrossRef](#)]
18. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298. [[CrossRef](#)]
19. Chen, B.; Principe, J.C. Maximum correntropy estimation is a smoothed MAP estimation. *IEEE Signal Process. Lett.* **2012**, *19*, 491–494. [[CrossRef](#)]
20. Liu, X.; Chen, B.; Zhao, H.; Qin, J.; Cao, J. Maximum correntropy Kalman filter with state constraints. *IEEE Access* **2017**, *5*, 25846–25853. [[CrossRef](#)]
21. Wang, G.; Li, N.; Zhang, Y. Maximum correntropy unscented Kalman and information filters for non-Gaussian measurement noise. *J. Frankl. Inst.* **2017**, *354*, 8659–8677. [[CrossRef](#)]
22. Wang, H.; Li, H.; Zhang, W.; Zuo, J.; Wang, H. Maximum correntropy derivative-free robust Kalman filter and smoother. *IEEE Access* **2018**, *6*, 70794–70807. [[CrossRef](#)]
23. Liu, X.; Qu, H.; Zhao, J.; Yue, P. Maximum correntropy square-root cubature Kalman filter with application to SINS/GPS integrated systems. *ISA Trans.* **2018**, *80*, 195–202. [[CrossRef](#)]
24. Wang, G.; Gao, Z.; Zhang, Y.; Ma, B. Adaptive maximum correntropy Gaussian filter based on variational Bayes. *Sensors* **2018**, *18*, 1960. [[CrossRef](#)]
25. Huang, F.; Zhang, J.; Zhang, S. Adaptive filtering under a variable kernel width maximum correntropy criterion. *IEEE Trans. Circuits Syst. II Express Briefs* **2017**, *64*, 1247–1251. [[CrossRef](#)]
26. Fakoorian, S.; Izanloo, R.; Shamshirgaran, A.; Simon, D. Maximum correntropy criterion Kalman filter with adaptive kernel size. In Proceedings of the 2019 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 15–19 July 2019; pp. 581–584.
27. Chen, B.; Wang, X.; Lu, N.; Wang, S.; Cao, J.; Qin, J. Mixture correntropy for robust learning. *Pattern Recognit.* **2018**, *79*, 318–327. [[CrossRef](#)]
28. Wang, H.; Zhang, W.; Zuo, J.; Wang, H. Outlier-robust Kalman filters with mixture correntropy. *J. Frankl. Inst.* **2020**, *357*, 5058–5072. [[CrossRef](#)]

29. Zheng, F.; Derrode, S.; Pieczynski, W. Semi-supervised optimal recursive filtering and smoothing in non-Gaussian Markov switching models. *Signal Process.* **2020**, *171*, 107511. [[CrossRef](#)]
30. Petetin, Y.; Janati, Y.; Desbouvries, F. Structured Variational Bayesian Inference for Gaussian State-Space Models with Regime Switching. *IEEE Signal Process. Lett.* **2021**, *28*, 1953–1957. [[CrossRef](#)]
31. Zhang, G.; Lan, J.; Zhang, L.; He, F.; Li, S. Filtering in Pairwise Markov Model With Student's t Non-Stationary Noise with Application to Target Tracking. *IEEE Trans. Signal Process.* **2021**, *69*, 1627–1641. [[CrossRef](#)]
32. Chen, B.; Wang, J.; Zhao, H.; Zheng, N.; Principe, J.C. Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Process. Lett.* **2015**, *22*, 1723–1727. [[CrossRef](#)]
33. 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* **2017**, *63*, 594–601. [[CrossRef](#)]
34. Li, S.; Xu, B.; Wang, L.; Razzaqi, A.A. Improved Maximum Correntropy Cubature Kalman Filter for Cooperative Localization. *IEEE Sens. J.* **2020**, *20*, 13585–13595. [[CrossRef](#)]
35. Yang, C.; Shi, W.; Chen, W. Robust M–M unscented Kalman filtering for GPS/IMU navigation. *J. Geod.* **2019**, *93*, 1093–1104. [[CrossRef](#)]