



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/169026/>

Version: Published Version

---

**Article:**

Shen, C. and Mihaylova, L. (2020) A flexible robust student's t-based multimodel approach with maximum Versoria criterion. *Signal Processing*, 182. 107941. ISSN: 0165-1684

<https://doi.org/10.1016/j.sigpro.2020.107941>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



# A flexible robust Student's t-based multimodel approach with maximum Versoria criterion

Chen Shen<sup>a,b,\*</sup>, Lyudmila Mihaylova<sup>b</sup>

<sup>a</sup> School of Information and Electronic Engineering (Sussex Artificial Intelligence Institute), Zhejiang Gongshang University, China

<sup>b</sup> Department of Automatic Control and Systems Engineering, University of Sheffield, UK

## ARTICLE INFO

### Article history:

Received 17 May 2020

Revised 5 October 2020

Accepted 11 December 2020

Available online 16 December 2020

### Keywords:

Robustness

Flexibility

Versoria function

Degrees of freedom

## ABSTRACT

The performance of the state estimation for Gaussian state space models can be degraded if the models are affected by the non-Gaussian process and measurement noises with uncertain degree of non-Gaussianity. In this paper, we propose a flexible robust Student's t-based multimodel approach. More specifically, the degrees of freedom parameter from the Student's t-distribution is assumed unknown and modelled by a Markov chain of state values. In order to capture more information of the Student's t-distributions propagated through multiple models, we establish a model-based Versoria cost function in the form of a weighted mixture rather than the original form, and maximize the function to interact and fuse the multiple models. Simulated results prove the flexibility of the robustness of the proposed Student's t-based multimodel approach when the existence probability of the outliers is uncertain.

© 2020 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

As a classical state estimation method, Kalman filter [1] has been thoroughly applied in a variety of applications such as positioning, navigation, brain imaging, and traffic control, together with its numerous nonlinear or off-line extensions [2–6]. These estimators assume that the model parameters and the noise statistics are exactly known, which is too perfect for most real applications and usually cannot be satisfied, hence resulting in reduced accuracy of the state estimation.

Known and correct information of the process noise and measurement noise is a key factor affecting the state estimation performance. However, uncertainties are practically inevitable to most state estimation problems. Many sources of uncertainties are due to the abnormal changes in the process of state evolution or sensor observation, which makes the Gaussian assumption on the process noise and measurement noise invalid. For the state estimation with measurement outliers, the Masreliez–Martin method has been used to develop a robust extended Kalman filter [7] and a robust unscented Kalman filter together with the adaptivity to process and measurement covariances [8]. In [9], a nonlinear unscented filter that is robust to measurement outliers and adaptive to a

time-variant measurement covariance matrix has been designed by using Huber's M-estimation and the variational Bayesian approach. The variational Bayesian approach has proven its efficiency in suppressing the measurement outliers when the measurement noise is represented by the Student's t-distribution of a hierarchical form [10–12]. Aforementioned Huber's M-estimation approach and the hierarchical Student's t-distribution have also been used for the state estimation with non-Gaussian process noise, in the light of inertial and Global Positioning System navigation [13–14].

In most real-time systems, simultaneous consideration of non-Gaussian process and measurement noises is necessary. By using the similar modelling idea in [14], a robust Student's t-based Kalman filter using variational Bayesian approach is presented in [15] to cope with the unknown non-Gaussian process and measurement noises. In [16], a generalized Gaussian scale mixture model has been proposed to model the heavy-tailed and skewed non-Gaussian process and measurement noises for the Kalman filter. Instead of the minimum mean-square error criterion, the maximum correntropy criterion has been introduced to establish a new state estimation framework [17–18] to deal with outliers both in the process and measurement noises. During recent years, the interests in the Student's t-filter methodology [19] have been significantly increasing, especially thanks to its ability to represent heavy-tailed distributions in many applications, with a light computational cost. Multiple advanced nonlinear Student's t-filters

\* Corresponding author.

E-mail addresses: [write2shen@sina.com](mailto:write2shen@sina.com) (C. Shen), [l.s.mihaylova@sheffield.ac.uk](mailto:l.s.mihaylova@sheffield.ac.uk) (L. Mihaylova).

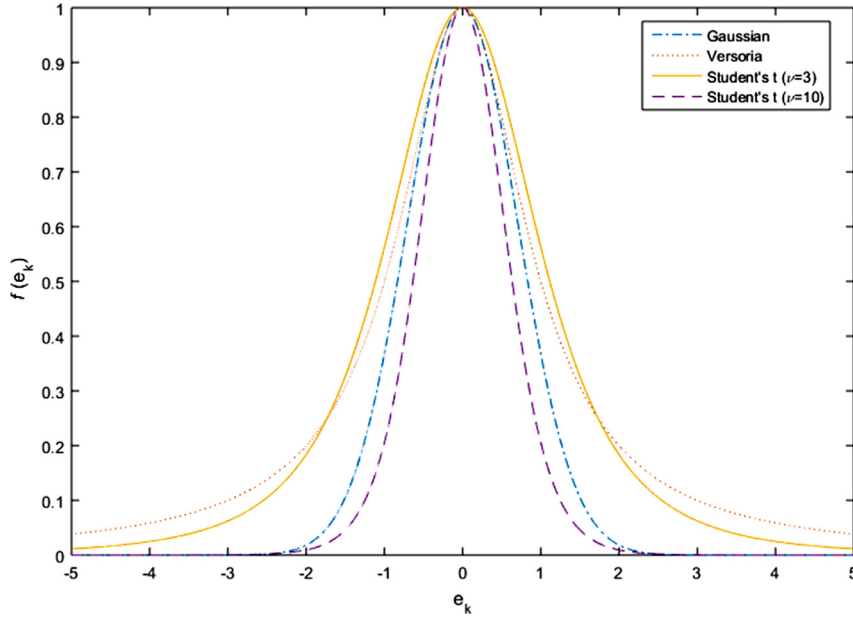


Fig. 1. Comparisons with popular probability density functions.

with efficient computations of integrals involving nonlinear functions have been developed in [20–23].

Although many efforts have been devoted to developing robust state estimation filters, the research on state estimation with non-Gaussian noises with time-variant nature still receives much less attention than its Gaussian counterpart. Especially, there is lack of works targeting on the flexibility of the robustness, i.e. most current available filters are only suitable for the non-Gaussian noise with fixed or specific degree of non-Gaussianity, e.g. heavy-tailedness. These estimators are vulnerable to the outliers with an uncertain or time-variant existence probability in the system and measurement processes. Although an adaptive approach to changeable noise covariances has been developed in [9] and approaches using parameters learning have been proposed in [10–12] for the non-Gaussian noises, these works do not take into account the uncertain degree of heavy-tailedness, which makes the robustness of these filters lack of flexibility. Pioneering works on similar topics have been proposed in [24–26] by constructing several noise models. Nevertheless, the developed multimodel particle filters [24–25] are usually computationally expensive, and these works mainly focus on the measurement noise. In addition, the flexibility of the robustness has not been validated by the noise with different levels of heavy-tailedness. The Bayesian model averaging (BMA) approach has been successful with handling multiple noise models in [24–26], but the model fusion only involves second-order information and high order information that is crucial to the non-Gaussian data has not been considered.

In this paper, we aim at enhancing the robustness of the Student's t-filter against noises with uncertain degree of heavy-tailedness, and propose a flexible robust Student's t-based multimodel approach. The main contribution of our work is two-fold. Firstly, the unknown degrees of freedom (dof) parameter is considered to follow a Markov process with interactive dof modes. Thus the flexibility problem of robustness can be naturally translated into an interacting multiple models (IMM) framework [27]. This modelling method is different from the one in [24] and therein the noise models are only combined directly at the final stage. Secondly, we introduce the maximum Versoria criterion, an effective optimization criterion recently used in kernel adaptive filtering area [28–29], within the multiple Student's t-based models, and

design a novel and effective model fusion strategy for the established dof multiple models. Using the synthetic data, we demonstrate that the proposed fusion strategy offers better model fusion for state estimation in the presence of non-Gaussian noises with uncertain degree of heavy-tailedness.

The outline of this paper is summarized as follows. The maximum Versoria criterion is briefly introduced in Section 2. Section 3 presents the proposed flexible robust Student's t-based multimodel algorithm. The computational complexity and stability of convergence are briefly discussed in Section 4. Simulated results are presented in Section 5 followed by concluding remarks in Section 6.

## 2. Maximum Versoria criterion

The original Versoria function, also named as Agnesi function, is defined as follows [30]

$$f(x) = \frac{A^3}{A^2 + x^2} = \frac{2a}{1 + \tau x^2} \quad (1)$$

where  $x$  denotes the variable,  $A = 2a$  is the diameter of the generating adjoined circle of the Versoria function, meaning that the centroid of the circle  $x^2 + (f(x) - a)^2 = a^2$  is located at  $(0, a)$ , and  $\tau = (1/2a)^2$  is the Versoria shape parameter. In Fig. 1, the Versoria function with  $a = 0.5$  is plotted and compared with the appropriate Gaussian and Student's t probability density functions. It can be seen that the Versoria function has the heaviest tail among all and its tail is even thicker than the tail of a Student's t density function ( $\nu = 3$ ) which is acknowledged as suitable for characterizing data with a heavy-tailed non-Gaussian distribution.

In the literature of kernel adaptive filtering, the objective of using the Versoria function is to estimate the unknown weight vector  $\omega$  from the output signal as defined in [28]

$$D_k = \mathbf{Y}_k^T \omega + \gamma_k \quad (2)$$

where  $\gamma_k$  is the noise with impulsive interferences and  $\mathbf{Y}_k$  is the input vector of the unknown system. The output error  $e_k$  is defined as

$$e_k = D_k - \mathbf{Y}_k^T \omega_{k-1} \quad (3)$$

where  $\boldsymbol{\omega}_{k-1}$  represents the estimate of  $\boldsymbol{\omega}$  of at time  $k-1$ . The optimization problem under maximum Versoria criterion (MVC) is to compute  $f(e_k)$  using (1) or other relevant forms as the cost function and maximize it to achieve the minimum error.

MVC has been shown with excellent robustness against non-Gaussian interferences such as impulsive noise in kernel algorithms [28–32]. It provides superior performance than algorithms using Gaussian-based kernel. Actually, not only have the kernel-based functions been used in the community of adaptive signal processing and communication, but also been extended to develop optimal state estimation rules to capture high order information existing in the stochastic dynamical system [17–18]. Given that the heavy-tailedness of the specific Versoria function shown in Fig. 1, the general Versoria function is believed to have the potential to fully exploit the non-Gaussian characteristics and offer better handling of high order information than the Gaussian function for the stochastic dynamical system. Inspired by the above, we introduce the MVC to the Student's t-filter based model fusion where second-order moments are not accurate enough and high order information in multiple models is crucial to the estimation accuracy. In the next section, the novel MVC based model fusion will be derived.

### 3. Maximum Versoria criterion based model fusion

The following stochastic dynamical system is considered for the Student's t-filter

$$\begin{aligned} \mathbf{x}_k &= \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{z}_k &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \end{aligned} \quad (4)$$

where  $\mathbf{x}_k$  and  $\mathbf{z}_k$  denote the  $m \times 1$  state variable and  $d \times 1$  measurement, respectively,  $\mathbf{F}_k$  is the state evolution matrix and  $\mathbf{H}_k$  is the measurement matrix. The process noise  $\mathbf{w}_{k-1}$  and the measurement noise  $\mathbf{v}_k$  are the non-Gaussian noises with nominal covariance matrices  $\mathbf{Q}_{k-1}$  and  $\mathbf{R}_k$ , respectively. In the standard Student's t-filter framework, the state  $\mathbf{x}_k$  and the noises  $\mathbf{w}_{k-1}$  and  $\mathbf{v}_k$  are assumed to be marginally Student's t-distributed [19].

In order to develop a Student's t-filter able to estimate the system state with non-Gaussian noises of time-variant heavy-tailedness,  $M$  appropriate models are introduced and each of them uses a different value for the dof  $\nu_k^{s_k} \in \{\nu_k^1, \nu_k^2, \dots, \nu_k^M\}$ . These values for the dof could be chosen according to the prior knowledge about the non-Gaussianity, more specifically, the intensity of the heavy-tailedness. Here  $s_k$  indicates the system mode which is represented by a  $M$ -state Markov chain and the state transition probability matrix is  $\Pi = \{\pi_{ij}\}$  satisfying the condition  $\sum_{j=1}^M \pi_{ij} = 1$  for any  $i \in \{1, 2, \dots, M\}$ . Each element  $\pi_{ij}$  denotes the transition probability from mode  $i$  to mode  $j$ .

In fact, this modelling idea is originally taken from the works in the area of wireless localization [33–34] where the condition of the communication channel in the cellular networks switches between LOS and NLOS randomly due to the obstructed radio transmissions. In their works, a two-state Markov process is adopted to represent the mode state for the NLOS and LOS conditions. In our work, for each model  $i \in \{1, 2, \dots, M\}$ , suppose we have obtained the Student's t-distributed filtering density at the end of discrete time instance  $k-1$  such that

$$\begin{aligned} p(\mathbf{x}_{k-1} | s_k = i, \mathbf{z}_{1:k-1}) &= \text{Std}(\mathbf{x}_{k-1} | \hat{\mathbf{x}}_{k-1|k-1}^i, \mathbf{P}_{k-1|k-1}^i, \nu_{k-1}^i) \\ &= \frac{\Gamma\left(\frac{\nu_{k-1}^i + m}{2}\right)}{\Gamma\left(\frac{\nu_{k-1}^i}{2}\right)} |\mathbf{P}_{k-1|k-1}^i|^{-1/2} (\nu_{k-1}^i \pi)^{d/2} \\ &\quad \times \left(1 + \frac{(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}^i)^T (\mathbf{P}_{k-1|k-1}^i)^{-1} (\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}^i)}{\nu_{k-1}^i}\right)^{-(\nu_{k-1}^i + d)/2} \end{aligned} \quad (5)$$

where  $\hat{\mathbf{x}}_{k-1|k-1}^i$  is the estimated state,  $\mathbf{P}_{k-1|k-1}^i$  is the scale matrix,  $\nu_{k-1}^i$  is the dof parameter, and  $\Gamma(\cdot)$  is the gamma function.

#### Step 1) Mixing under maximum Versoria criterion

At the mixing stage, we are interested in calculating the density by using marginalization

$$\begin{aligned} p(\mathbf{x}_{k-1} | s_k = j, \mathbf{z}_{1:k-1}) &= \sum_{i=1}^M p(\mathbf{x}_{k-1} | s_k = j, s_{k-1} = i, \mathbf{z}_{1:k-1}) \Pr(s_{k-1} = i | s_k = j) \\ &= \sum_{i=1}^M \mu_{k-1}^{ij} \text{Std}(\mathbf{x}_{k-1} | \hat{\mathbf{x}}_{k-1|k-1}^i, \mathbf{P}_{k-1|k-1}^i, \nu_{k-1}^i) \end{aligned} \quad (6)$$

where  $\mu_{k-1}^{ij}$  is known as the mixing probability. As can be seen from (6), we use a single Student's t-distribution  $\text{Std}(\mathbf{x}_{k-1} | \hat{\mathbf{x}}_{k-1|k-1}^{0j}, \mathbf{P}_{k-1|k-1}^{0j}, \nu_{k-1}^{0j})$  that will be obtained via the proposed MVC-based optimization to approximate  $p(\mathbf{x}_{k-1} | s_k = j, \mathbf{z}_{1:k-1})$ .

Recall that for the conventional IMM framework [27] where the above Student's t-distribution is replaced by the Gaussian, the moments matching technique has been widely used in most Gaussian mixture based methods and a single approximate Gaussian distribution is found by matching the first two moments of the mixture of Gaussians. However, this assumption is not valid often since a mixture of probability distributions does not necessarily have a Gaussian distribution, particularly for the jump Markov systems with non-Gaussian noises. In our formulated problem, each component of the mixture is assumed to be a Student's t-distribution. For such a heavy-tailed distribution, the moments matching technique faces challenges at the information fusion step due to the loss of high order information.

In order to cope with this problem, we introduce the MVC as the information measure for model interaction of the multimodel state estimates. Instead of using the error (3) defined for adaptive filtering, we define the error function for the model fusion as follows

$$e_{i,1k} = \|\mathbf{x}_{k-1|k-1} - \hat{\mathbf{x}}_{k-1|k-1}^i\| \left(\mathbf{P}_{k-1|k-1}^i\right)^{-1} \quad (7)$$

where the weighted  $l_2$  norm  $\|\cdot\|$  is used,  $\mathbf{x}_{k-1|k-1}$  is the unknown state variable. Since our focus is on fusing  $M$  models, the Versoria cost function here is re-defined in the form

$$J_1(\mathbf{x}_{k-1|k-1}) = 2a \sum_{i=1}^M \mu_{k-1}^{ij} \frac{1}{1 + \tau e_{i,1k}^2} \quad (8)$$

which is of a weighted mixture rather than the original form of the Versoria function. In (8) the mixing probability  $\mu_{k-1}^{ij}$  is chosen as the corresponding weight.

To maximize the Versoria cost function (8), we need to compute its first derivative. After some mathematical arrangements, we obtain

$$\begin{aligned} \mathbf{x}_{k-1|k-1} &= \left( \sum_{i=1}^M \mu_{k-1}^{ij} \frac{(\mathbf{P}_{k-1|k-1}^i)^{-1}}{\left(1 + \tau \|\mathbf{x}_{k-1|k-1} - \hat{\mathbf{x}}_{k-1|k-1}^i\|^2 (\mathbf{P}_{k-1|k-1}^i)^{-1}\right)^2} \right)^{-1} \\ &\quad \times \sum_{i=1}^M \mu_{k-1}^{ij} \frac{(\mathbf{P}_{k-1|k-1}^i)^{-1} \hat{\mathbf{x}}_{k-1|k-1}^i}{\left(1 + \tau \|\mathbf{x}_{k-1|k-1} - \hat{\mathbf{x}}_{k-1|k-1}^i\|^2 (\mathbf{P}_{k-1|k-1}^i)^{-1}\right)^2} \end{aligned} \quad (9)$$

Note that there is no closed-form solution to (9) and a fixed-point iteration algorithm is required to solve it. Let  $L(L \geq 1)$  be the

maximum iteration number and the value from the  $L$ th iteration is assigned to the mixing state vector  $\hat{\mathbf{x}}_{k-1|k-1}^{0j}$  such that

$$\hat{\mathbf{x}}_{k-1|k-1}^{0j} = \mathbf{x}_{k-1|k-1}^{(L)} = \left( \sum_{i=1}^M \mu_{k-1}^{ij} \frac{(\mathbf{P}_{k-1|k-1}^i)^{-1}}{\left(1 + \tau \|\mathbf{x}_{k-1|k-1}^{(L-1)} - \hat{\mathbf{x}}_{k-1|k-1}^i\|^2 (\mathbf{P}_{k-1|k-1}^i)^{-1}\right)^2} \right)^{-1} \times \sum_{i=1}^M \mu_{k-1}^{ij} \frac{(\mathbf{P}_{k-1|k-1}^i)^{-1} \hat{\mathbf{x}}_{k-1|k-1}^i}{\left(1 + \tau \|\mathbf{x}_{k-1|k-1}^{(L-1)} - \hat{\mathbf{x}}_{k-1|k-1}^i\|^2 (\mathbf{P}_{k-1|k-1}^i)^{-1}\right)^2} \quad (10)$$

Although a few iterations are involved, the algorithm converges very fast to one local optimum due to the property of the Versoria function, especially when the error is large. Thus a small number of  $L$  iterations is enough to find a local optimal solution and also provides a satisfactory performance in practice. In the simulation section, the fast convergence of this fixed-point iteration will be validated.

After obtaining the mixing state vector  $\hat{\mathbf{x}}_{k-1|k-1}^{0j}$ , the mixing scale matrix  $\mathbf{P}_{k-1|k-1}^{0j}$  is required as an essential part of the whole cycle of the IMM framework. Hence we merge the scale matrices from multiple dof models using the efficient Kullback-Leibler based method [35] as follows

$$\left(\mathbf{P}_{k-1|k-1}^{0j}\right)^{-1} = \sum_{i=1}^M \mu_{k-1}^{ij} \left(\mathbf{P}_{k-1|k-1}^i\right)^{-1} \quad (11)$$

### Step 2) Mode-conditioned Student's t-filter

At this stage, the standard Student's t-filter is run for each dof model. The prediction stage consists of

$$\hat{\mathbf{x}}_{k|k-1}^j = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}^{0j} \quad (12)$$

$$\mathbf{P}_{k|k-1}^j = \mathbf{F}_k \mathbf{P}_{k-1|k-1}^{0j} \mathbf{F}_k^T + \mathbf{Q}_{k-1} \quad (13)$$

The objective of the subsequent update is to compute the filtering density  $p(\mathbf{x}_k^j | \mathbf{z}_{1:k})$  after receiving the measurement vector  $\mathbf{z}_k$ . The joint density of the state and measurement is Student's t-distributed as follows

$$p(\mathbf{x}_k^j, \mathbf{z}_k | \mathbf{z}_{1:k-1}) = \text{Std} \left( \begin{bmatrix} \mathbf{x}_k^j \\ \mathbf{z}_k \end{bmatrix} \middle| \begin{bmatrix} \hat{\mathbf{x}}_{k|k-1}^j \\ \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}^j \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{k|k-1}^j & \mathbf{P}_{k|k-1}^j \mathbf{H}_k^T \\ \mathbf{H}_k \mathbf{P}_{k|k-1}^j & \mathbf{S}_k^j \end{bmatrix}, \nu_{k-1}^j \right) \quad (14)$$

where

$$\mathbf{S}_k^j = \mathbf{H}_k \mathbf{P}_{k|k-1}^j \mathbf{H}_k^T + \mathbf{R}_k \quad (15)$$

Thus the conditional density of the state can be updated by using the standard Student's t-filter [19]

$$\hat{\mathbf{x}}_{k|k}^{*j} = \hat{\mathbf{x}}_{k|k-1}^j + \mathbf{P}_{k|k-1}^j \mathbf{H}_k^T (\mathbf{S}_k^j)^{-1} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}^j) \quad (16)$$

$$(\Delta_k^j)^2 = (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}^j)^T (\mathbf{S}_k^j)^{-1} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}^j) \quad (17)$$

$$\mathbf{P}_{k|k}^{*j} = \frac{\nu_{k-1}^j + (\Delta_k^j)^2}{\nu_{k-1}^j + d} \left( \mathbf{P}_{k|k-1}^j - \mathbf{P}_{k|k-1}^j \mathbf{H}_k^T (\mathbf{S}_k^j)^{-1} \mathbf{H}_k \mathbf{P}_{k|k-1}^j \right) \quad (18)$$

$$\nu_k^{*j} = \nu_{k-1}^j + d \quad (19)$$

$$\hat{\mathbf{x}}_{k|k}^j = \hat{\mathbf{x}}_{k|k}^{*j} \quad (20)$$

$$\mu_{k|k}^j = \frac{\nu_k^{j*}}{\nu_k^{j*} - 2} \frac{\nu_k^j - 2}{\nu_k^j} \mu_{k|k}^{j*} \quad (21)$$

where  $\nu_k^j = \nu_{k-1}^j$ . As stated in Student's t-filter related literature [19–20], the moments matching technique (not the moments matching in the IMM framework) is often used to obtain (20) and (21) to prevent the loss of robustness after a few time steps such that  $\hat{\mathbf{x}}_{k|k}^{j*}$  and  $\mathbf{P}_{k|k}^{j*}$  are substituted by  $\hat{\mathbf{x}}_{k|k}^j$  and  $\mathbf{P}_{k|k}^j$ , respectively. The results (20)-(21) are still kept within our algorithm so that information from all the dof models using different values of dof can be distinguished and preserved for dof model interaction and fusion by using the MVC. Otherwise, all the Student's t-filters reduce to Kalman filters resulting in the loss of robustness, not to mention the flexibility of the robustness.

### Step 3) Mode probabilities update

The mode probability is updated by the Bayes' rule as

$$\mu_k^j = \frac{L_k^j \sum_{l=1}^M \pi_{lj} \mu_{k-1}^l}{\sum_{i=1}^M \sum_{l=1}^M \pi_{li} \mu_{k-1}^l L_k^i} \quad (22)$$

where the measurement likelihood function is in the form of the standard Student's t-distribution (see (5)) instead of the Gaussian

$$L_k^j = \text{Std} \left( \mathbf{z}_k | \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}^j, \mathbf{H}_k \mathbf{P}_{k|k-1}^j \mathbf{H}_k^T + \mathbf{R}_k, \nu_k^j \right) \quad (23)$$

### Step 4) Estimation fusion under maximum Versoria criterion

Similar to the mixing stage, the MVC is chosen to replace the moments matching technique to find the combination of all the models with updated mode probabilities. This time, the error function is defined as

$$e_{j,2k} = \left\| \mathbf{x}_{k|k} - \hat{\mathbf{x}}_{k|k}^j \right\| \left( \frac{\nu_k^j}{\nu_k^j - 2} \mathbf{P}_{k|k}^j \right)^{-1} \quad (24)$$

from which the covariance matrix  $\frac{\nu_k^j}{\nu_k^j - 2} \mathbf{P}_{k|k}^j$  is selected instead of

the scale matrix  $\mathbf{P}_{k|k}^j$  as the weighting matrix. Unlike the cost function established in (7), the definition of (24) is due to the consideration that this fusion stage immediately follows the mode-conditioned Student's t-filter, in which the state and scale matrix are all updated using different values of the dof. Thus the covariance matrix appears to be more accurate than the scale matrix alone as the dof parameter is included.

The Versoria cost function at this stage can be written as

$$J_2(\mathbf{x}_{k|k}) = 2a \sum_{j=1}^M \mu_k^j \frac{1}{1 + \tau e_{j,2k}^2} \quad (25)$$

Taking the first derivate of (25) and setting it to zero, gives the fused state vector  $\hat{\mathbf{x}}_{k|k}$  in the iterative form as follows

$$\hat{\mathbf{x}}_{k|k} = \mathbf{x}_{k|k}^{(\bar{l})} = \left[ \sum_{j=1}^M \mu_k^j \frac{(\nu_k^j - 2) (\mathbf{P}_{k|k}^j)^{-1}}{\nu_k^j \left(1 + \tau \|\mathbf{x}_{k|k}^{(\bar{l}-1)} - \hat{\mathbf{x}}_{k|k}^j\|^2 (\mathbf{P}_{k|k}^j)^{-1}\right)^2} \right]^{-1} \times \sum_{j=1}^M \mu_k^j \frac{(\nu_k^j - 2) (\mathbf{P}_{k|k}^j)^{-1} \hat{\mathbf{x}}_{k|k}^j}{\nu_k^j \left(1 + \tau \|\mathbf{x}_{k|k}^{(\bar{l}-1)} - \hat{\mathbf{x}}_{k|k}^j\|^2 (\mathbf{P}_{k|k}^j)^{-1}\right)^2} \quad (26)$$

where  $\bar{L}$  denotes the maximum iteration number. The fused scale matrix can be obtained by applying Kullback-Leibler divergence based method

$$(\mathbf{P}_{k|k})^{-1} = \sum_{j=1}^M \mu_k^j (\mathbf{P}_{k|k}^j)^{-1} \quad (27)$$

It is known that a fixed single value chosen for the dof parameter limits the flexibility of the robustness of the Student's t-filter if the degree of heavy-tailedness is uncertain. Our method can cope with such an issue from two perspectives. Establishing multiple models for the dof parameter with possibly switched values makes the robustness of the Student's t-filter adjustable. With the Markovian transition of dof modes, the problem is straightforwardly translated into a multimodel estimation under the IMM framework. In addition, to incorporate the robust Student's t-filter effectively, the celebrated IMM framework is altered by introducing the Versoria function to combine all models with high order moments being processed for better robust fused results. The improvement will be demonstrated shortly in the simulation.

#### 4. Further discussion

In this section, we briefly discuss the proposed algorithm in terms of computational complexity and stability of convergence, respectively.

##### 4.1. Computational complexity

The proposed algorithm is within the IMM framework. It is known that the IMM framework has the best computational complexity among those popular multimodel approaches that have acceptable estimation errors [36]. As can be seen in the previous section, Step 2 and Step 4 are the main parts of the proposed algorithm under the IMM framework. Hence we give the computation complexity of (10) in terms of matrix inversion (using  $O(\cdot)$  notation) and matrix/vector multiplication with respect to the state dimension  $m$  and the number of models  $M$  in the following.

$$C_{Eq.(10)}(m) = \bar{L}[(M+1)O(m^3) + M(m^2 + m) + Mm^2 + m^2] \\ = \bar{L}[(M+1)O(m^3) + (2M+1)m^2 + Mm] \quad (28)$$

which is related to the iteration number  $\bar{L}$ . Since (10) shares the same structures with (26), the above computation complexity also applies to (26).

For a clear comparison, we may introduce moments matching to our Student's t-based problem for model fusion also within the IMM framework. The consequent algorithm would have a similar form to ours except (10), (11), (26) and (27). As a consequence, the mixing mean and scale matrix after replacing MVC with moments matching are approximated respectively as

$$\hat{\mathbf{x}}_{k-1|k-1}^{0j} = \sum_{i=1}^M \mu_{k-1}^{ij} \hat{\mathbf{x}}_{k-1|k-1}^i \quad (29)$$

$$\mathbf{P}_{k-1|k-1}^{0j} = \frac{\nu_{k-1}^j - 2}{\nu_{k-1}^j} \sum_{i=1}^M \left( \frac{\nu_{k-1}^i}{\nu_{k-1}^i - 2} \mu_{k-1}^{ij} \mathbf{P}_{k-1|k-1}^i \right. \\ \left. + \mu_{k-1}^{ij} [\hat{\mathbf{x}}_{k-1|k-1}^i - \hat{\mathbf{x}}_{k-1|k-1}^{0j}] [\hat{\mathbf{x}}_{k-1|k-1}^i - \hat{\mathbf{x}}_{k-1|k-1}^{0j}]^T \right) \quad (30)$$

Therefore the major computational complexity is from (30) and computed in terms of matrix/vector multiplication as

$$C_{Eq.(30)}(m) = Mm^2 \quad (31)$$

One may conclude that the whole computation cost of the proposed algorithm is evidently much more expensive than that of the moments matching based algorithm as the system dimension  $m$  grows. The primary cost arises from the multiple operations of matrix inversion/multiplication and depends on the iteration number  $\bar{L}$  as well. Despite showing predictable higher computational cost, our method is expected to give better robustness performance than the moments matching based approach and other related competing estimation approaches. Fortunately, a large iteration number is usually unnecessary in practice owing to the MVC convergence property. Then the heavy computation loads can be avoided. In the simulation, we will compare our proposed method with other competing methods in terms of computational costs and estimation accuracy.

##### 4.2. Stability of convergence

In this section, we briefly discuss the stability of the MVC-based iterative algorithm. The original MVC arises in the area of information theoretic learning, and its stability of convergence has been proved via the least square mean algorithm. In contrast, we use the MVC to fuse multiple models for the state estimation, which is a completely different subject from those in existing literature. Hence, unlike what has been done in [28], it is hard to obtain recursive update equations for the state estimation error from (10) and (26). This results in difficulty in analyzing the convergence directly. Fortunately, (10) and (26) are iterative fixed-point solutions under MVC, and [37] has offered a complete convergence analysis of the fixed-point algorithm with a similar form. For example, we rewrite (26) in the following form to match the general iterative form shown in [37] such that

$$\mathbf{x}_{k|k}^{(l+1)} = g(\mathbf{x}_{k|k}^{(l)}) = [\mathbf{M}]^{-1} \mathbf{V} \quad (32)$$

where  $l$  denotes the  $l$ th iteration, and we have made the following transformation

$$\mathbf{M} = \sum_{j=1}^M \mu_k^j \frac{f\left(\|\mathbf{x}_{k|k}^{(l)} - \hat{\mathbf{x}}_{k|k}^j\| (\mathbf{P}_{k|k}^j)^{-1}\right)}{2a} (\mathbf{P}_{k|k}^j)^{-1} \quad (33)$$

$$\mathbf{V} = \sum_{j=1}^M \mu_k^j \frac{f\left(\|\mathbf{x}_{k|k}^{(l)} - \hat{\mathbf{x}}_{k|k}^j\| (\mathbf{P}_{k|k}^j)^{-1}\right)}{2a} (\mathbf{P}_{k|k}^j)^{-1} \hat{\mathbf{x}}_{k|k}^j \quad (34)$$

where  $f(\cdot)$  is the function with respect to the norm. Therefore the proof in [37] can be similarly applied into our iterative fixed-point equations like (10) and (26), and a sufficient condition therein guarantees that the iterative solutions are stable to converge to a unique fixed point. However, the above may not be extremely strict because a Gaussian kernel which is always less than one is used in their work, while the Versoria function dependent on the radius  $a$  is considered in ours. In addition, the relationship between the radius and the convergence has not been revealed. Thus a rigorous and in-depth theoretical exploration on these issues remain as our future work. Nevertheless, we will still demonstrate the convergence property of the proposed algorithm as a complement, and test how the radius influences the convergence in the following simulation.

#### 5. Simulated results

In this section, we adopt an example that estimates the location of a single target to assess the performance of the proposed method, namely, maximum Versoria Student's t-based multimodel

**Table 1**  
Effects of the number of iteration.

		$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$
Case1	ARMSE (m)	11.045	10.753	10.751	10.751	10.751
	SD (m)	2.143	1.967	1.967	1.967	1.967
Case2	ARMSE (m)	12.417	12.100	12.099	12.099	12.099
	SD (m)	2.255	2.062	2.056	2.056	2.056

(MVMMStd) method. The constant velocity model and the linear position measurement model in [12] are used.

The total tracking time is 200s and the sampling time is set as 2s. The true initial state vector of the target is  $\mathbf{x}_0 = [50\mathbf{m} \ 10\mathbf{m/s}]^T$ . The nominal covariance matrices of the process and measurement noises are  $\mathbf{Q}_{k-1} = \text{diag}(1\mathbf{m}^2, 1\mathbf{m}^2/\mathbf{s}^2)$  and  $\mathbf{R}_k = 100\mathbf{m}^2$ , respectively. The model-based parameters for MVMMStd are given as follows:  $M = 2$ ,  $\nu^1 = 100$  and  $\nu^2 = 3$ ,  $\Pi = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$ ,  $\mu_0^1 = \mu_0^2 = 0.5$ .

It is worth pointing out that  $M$  is unnecessary to be very large in practice and a large  $M$  does not ensure the improvement of the final estimates but would absolutely cause extra computational burden. Too many dof models would not yield satisfactory fusion results in consequence of the excessive model competition. Similar remarks have been drawn in the area of maneuvering target tracking [38] where multiple models are established for the uncertain target dynamics. Thus  $M$  dof models should be chosen to have distinguishable estimation performance. If there is no accurate prior knowledge about the heavy-tailedness of the noises, it is advised to select at least two values, for instance, a very high value and a low value for the dof models.

In this simulation, the following two setups for the uncertain process noise and measurement noise are considered.

**Case 1.** Noises with abrupt changes: The process noise is purely Gaussian in the first 50s and only contaminated by 5% outliers  $\mathbf{N}(\mathbf{w}_{k-1}|0, 25\mathbf{Q}_{k-1})$  in the rest 150s. The measurement outliers  $\mathbf{N}(\mathbf{v}_k|0, 50\mathbf{R}_k)$  occur with the probabilities of 0%, 5%, 15% and 15% at time intervals [1, 50]s, [51, 100]s, [101, 150]s, and [151, 200]s, respectively, which indicates the Gaussian noise, the moderate outlier-contaminated non-Gaussian noise, the mild one, and the severe one.

**Case 2.** Noises with progressive changes: The process and measurement noises are polluted by outliers  $\mathbf{N}(\mathbf{w}_{k-1}|0, 25\mathbf{Q}_{k-1})$  and  $\mathbf{N}(\mathbf{v}_k|0, 50\mathbf{R}_k)$ , respectively, both with a constantly changing occurrence probability  $0.15(k-1)/200$ , which rises from 0% at the beginning to 14.9% at  $k = 200$ s.

First of all, we study the convergence of the proposed algorithm and test the influence of the radius on the algorithm. Given five predefined values for the number of iteration  $L$  ( $\bar{L} = L$  in this simulation) and the assumption that  $a = 1$ , Table 1 gives the average root mean square errors (ARMSEs) and standard deviations (SDs) of the estimated position after 2000 Monte Carlo repetitions for both two cases. A notable decrease in the ARMSEs and SDs is observed when the iteration number changes from 1 to 2 both in Case 1 and Case 2. It can be seen as well that the ARMSEs and corresponding SDs change slightly in both cases when the number increases from 2 to 3, and remain unchanged if  $L \geq 3$ , which means the algorithm almost converges at  $L = 3$ . This accords with the statement from literature that the Versoria function converges very fast to reach the local optimum.

Next, we move our focus on to the impacts of the radius  $a$  on the estimation accuracy of MVMMStd. The iteration number  $L = \bar{L}$  is chosen as 2 in this part. In Fig. 2, letting the radius range from 0.001 to 100 with selective values, we show the relationship between the ARMSEs of the estimated position varying with the radius  $a$  for both cases. The standard errors of the ARMSEs, repre-

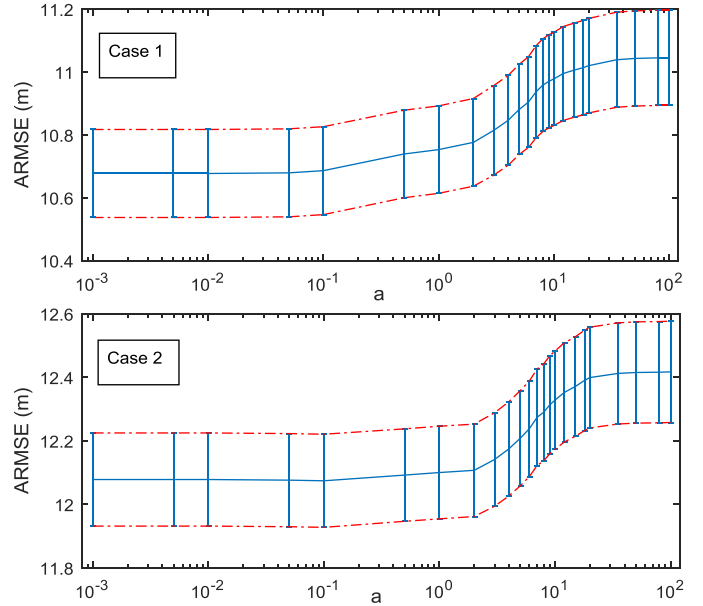


Fig. 2. ARMSEs of the position versus the radius 0.001–100.

sented by upper and lower red chain-dotted lines, are provided in the figure as well. Note that a nonlinear log scale is employed for the horizontal axis to achieve good presentation. From Fig. 2, it is observed that the ARMSE varies remarkably when  $a$  is chosen in the interval [2, 20], while ARMSE curves for both cases tend to be flat out of this interval, which means the algorithm is quite insensitive to  $a < 0.1$  and  $a > 20$ . The estimation error can be diminished when  $0.1 < a < 2$ , but not as greatly as  $2 < a < 20$ . To further prove it, we select first 50 Monte Carlo runs from the whole 2000 runs and compute the ARMSEs of the position over 200s for each run. As displayed in Figs. 3 and 4, the MVMMStd with  $a = 0.1$  has an apparent estimation advantage over the one with  $a = 20$  but performs only slightly better than the one with  $a = 2$ .

Even though the influence of the radius on estimation errors have been studied, it is still difficult to investigate the impact of the radius on the convergence directly. Instead, we combine with the iteration number and study how the ARMSEs change versus different iteration numbers as the radius varies. We continue to do the experiment as Fig. 2 by using different iteration numbers, and obtain the results shown in Fig. 5. Note that only the results in the interval [1, 10] are presented. This is because [1, 10] is the most active interval to affect our algorithm, and the difference between each curve is quite small for the intervals except [1, 10]. In addition to previous results given in Table 1, Fig. 5 visually describes the convergence behavior of the proposed algorithm. A larger iteration number generally leads to a lower estimation error, and the performance of the MVMMStd using  $L \geq 3$  is quite close, especially when the radius is either in [1, 2] or [5, 10]. Thus values of the radius from these intervals do not show significant influence on the convergence. On the other hand, these four curves present the largest gaps between each other when the radius falls into the interval [2, 5]. Moreover, it can be seen from Fig. 5 that to achieve the same estimation error the MVMMStd with a larger radius requires more iteration steps. It is equivalent to say a smaller radius in [2, 5] enables the algorithm to converge faster. Therefore, if the convergence speed and estimation precision are simultaneously considered, a small value in [2, 5] is suggested for the radius.

Secondly, we compare the MVMMStd with several competing robust estimation methods in terms of estimation errors. These methods include the Student's t-filters [19] such as Std3 and Std100 with fixed dof values as 3 and 100, respectively, the robust

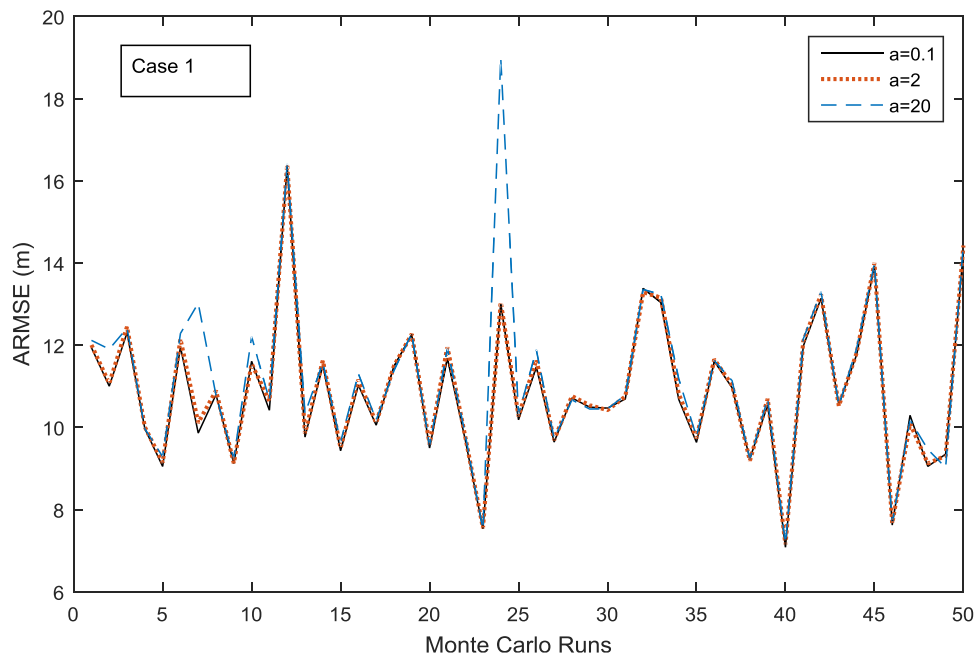


Fig. 3. ARMSEs versus first 50 repetitions from Case 1.

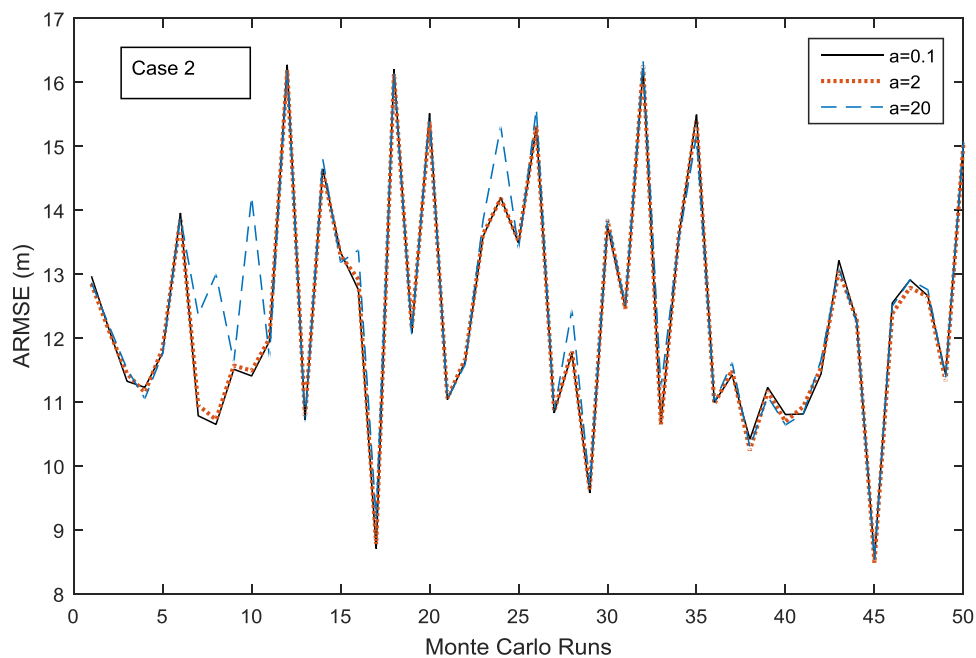


Fig. 4. ARMSEs versus first 50 repetitions from Case 2.

Bayesian filter using Bayesian model averaging approach (RFBMA) [24] which is essentially a multimodel approach combining BMA and restricted variational Bayes, the Student's t-based multimodel approach that shares the same Student's t-based modelling method with ours but uses moments matching (MMMMStd) [27] rather than the MVC, the Student's t-based robust Kalman filter using variational Bayesian (RStdKF) without knowing process and measurement noises [15], the generalized hyperbolic skew Student's t Kalman filter (GHSStdKF) [16], the generalized hyperbolic variance Gamma Kalman filter (GHVGKF) [16], the maximum correntropy Kalman filter (MCKF) [17], and the robust Kalman filter using Huber's M-estimation [13,39] (HMKF). All the multimodel algorithms are initialized by using the same predefined model parameters for fair comparisons. For the MVMMStd,  $L = 2$  and  $a = 1$  are used.

The number of iteration in GHSStdKF, GHVGKF, and RStdKF is chosen as 10 according to literature to ensure convergence. MCKF uses threshold detection to end iterative computations instead of using fixed iteration steps, and hence the number of iteration is dynamic. The actual average iteration steps of MCKF in Case 1 and Case 2 are 3.01 and 3.07, respectively.

Plotted in Fig. 6 are the RMSEs of the position of all the compared methods in Case 1. In first 50s, since only Gaussian process and measurement noises exist, Std100, MCKF and RStdKF have the best estimation precision. The performance of MVMMStd is still comparable to that of Std100, MCKF and RStdKF. MMMStd, HMKF and RFBMA have an advantage over Std3. MMMStd even outperforms MVMMStd in this period, which means the loss of high order information from the dof models is insignificant and

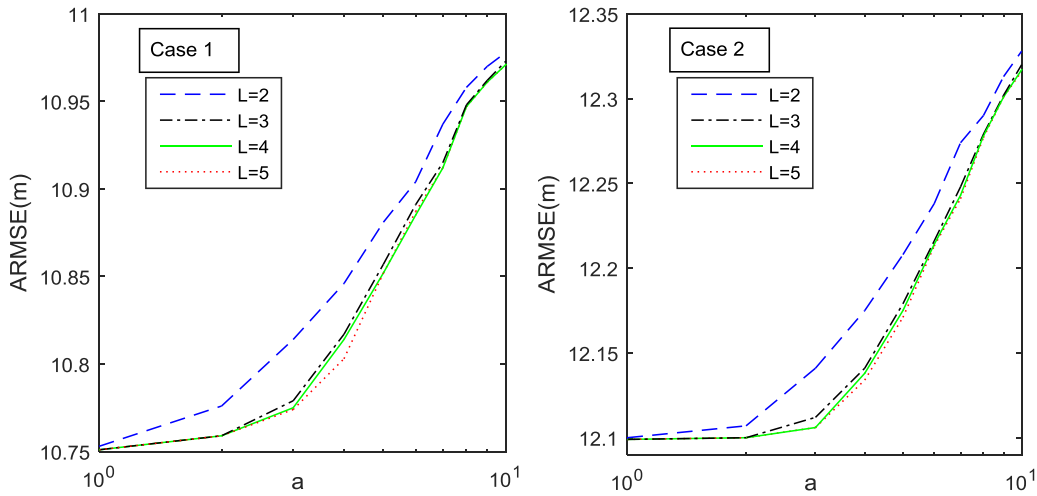


Fig. 5. ARMSEs of the position using different iteration numbers.

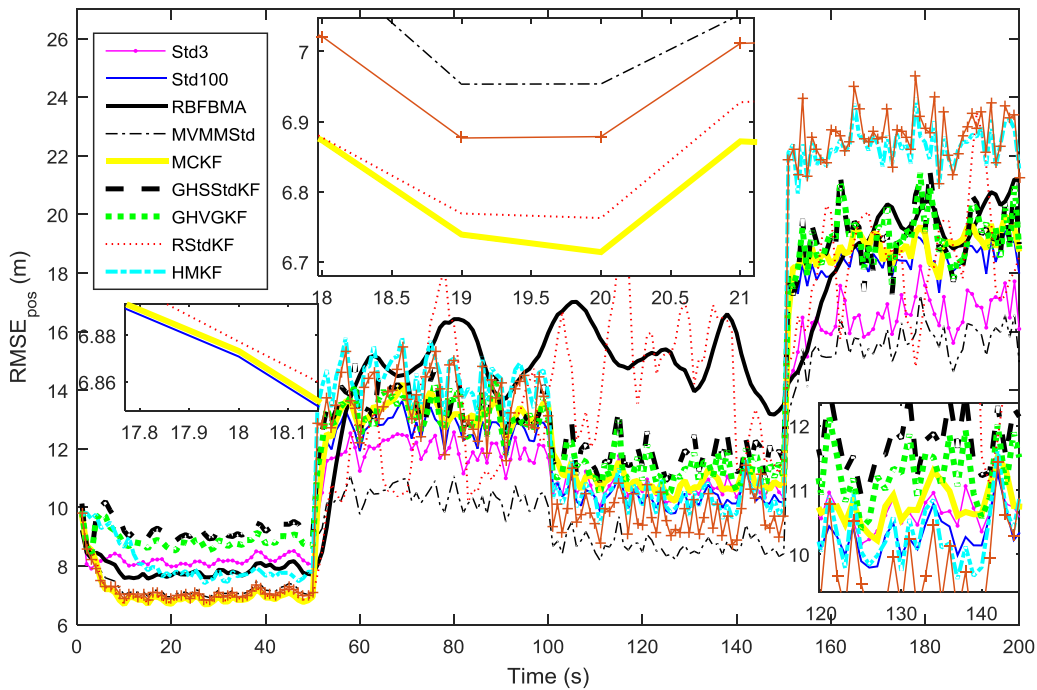


Fig. 6. RMSEs of the position in Case 1.

the moments matching technique is better compared to the MVC under the pure Gaussian environment. In the next 150s with the existence of the moderate, mild and severe outliers in noises in order, MVMMStd always performs the best among all because of the ability to find most probable combination of multiple models. MMMStd and RFBFMA offer worse performance than MVMMStd, although they both use the multimodel approach. This is because MMMStd uses moments matching at the mixing and combination stages, which ignores the high order information contained in Student's t-filter based models, and RFBFMA is unable to handle non-Gaussian process noise apart from using moments matching at the stage of overall approximation. Though RStdKF gives the lowest estimation errors occasionally and is able to learn unknown noise parameters, it is sensitive to the change of heavy-tailedness and hence its estimation results have large fluctuations. GHStdKF and GHVGKF have shown the effectiveness when the noise is heavy-tailed and skewed in literature, however, they exhibit worse than the proposed MVMMStd in this case. It is interesting to note

that in the last 150s, Student's t-filters with fixed dof (3 or 100) present better performance for some time than the multimodel approaches (MMMMStd and RFBFMA) and other robust methods. Nonetheless, it is difficult to select a correct dof value for the Student's t-filter in practice, particularly in the case where the heavy-tailedness is time-variant. It also implies that solutions to the non-Gaussian noise with uncertain heavy-tailedness are necessary, and appropriate processing of the model fusion is required if the multimodel approach is considered.

Likewise, RMSEs of the position in Case 2 are given in Fig. 7. All the algorithms demonstrate a trend of performance deterioration as a result of the increasing contamination probability  $0.15(k-1)/200$  for both the process and measurement noises. Although all of them exhibit close estimation performance at the beginning when the contamination of outliers is very low, the disparity among the algorithms keeps progressively growing along with the increasing outlier occurrence probability. The major reason that the estimation error of RFBFMA significantly

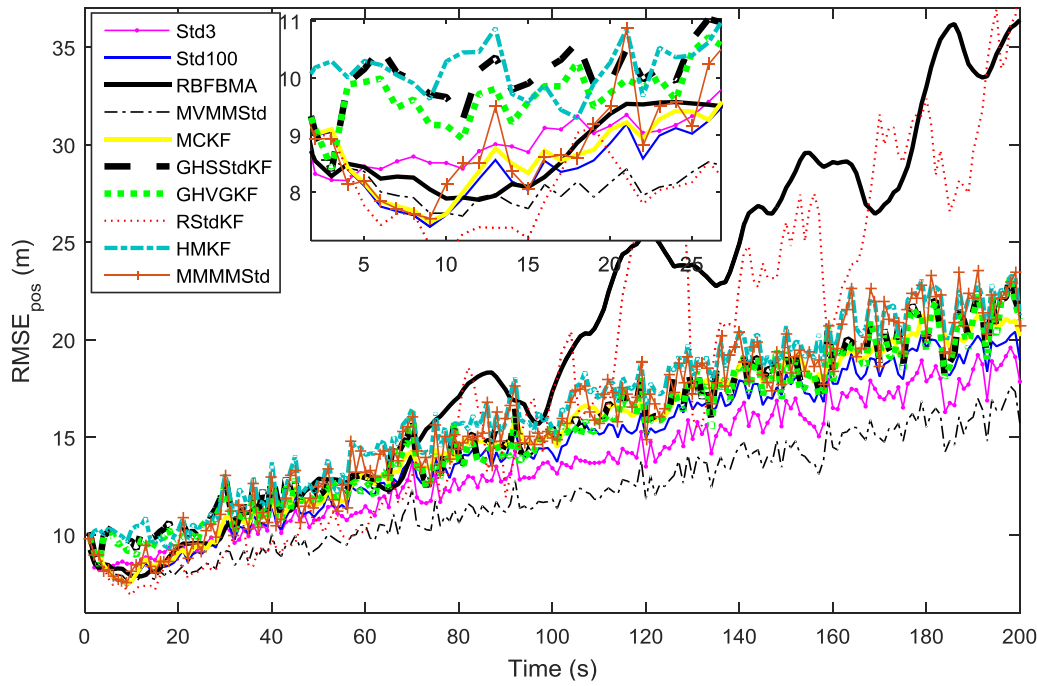


Fig. 7. RMSEs of the position in Case 2.

Table 2

Average root mean square errors of competing robust methods.

		Std3	MMMMStd	MCKF	RFBMA			
ARMSE (m)	Case 1	12.039	14.191	13.038	12.872			
	Case 2	13.728	15.970	15.432	18.221			
ARMSE (m)	Case 1		HMKF	GHSStdKF-10	GHVGKF-10	RStdKF-10		
	Case 2	14.590	13.691	13.453	11.130	16.657	15.696	15.449

grows after 100s is its inability to handle the time-variant non-Gaussian process noise. Due to the same reason explained for Case 1, RStdKF has much larger errors and fluctuations in the last 100s than it has in the first 100s. It shows that the proposed MVMMStd provides the best estimation results throughout the whole time except the first 10s, i.e. when the outlier percentage is quite low and the actual noise distributions approximate to the Gaussian.

As the ARMSEs of MVMMStd using different iteration numbers have been displayed in Table 1, Table 2 gives the ARMSEs of all the rest competing algorithms for Case 1 and Case 2 for comparison in addition to Figs. 6 and 7. Note that the numbers following GHSStdKF, GHVGKF and RStdKF represent the iteration steps they use in the algorithms. As seen from Tables 1 and 2, the proposed MVMMStd has the best estimation results, which is in line with what we observe from Figs. 6 and 7.

Furthermore, as MVMMStd and MMMMStd are structurally similar and only differ from each other in terms of the fusion method under the IMM framework, we show their mode probabilities in Figs. 8 and 9 for Case 1 and Case 2, respectively. As expected, for the MVMMStd in both cases, the probabilities of the model using a large dof value (Mode 1) are prone to decline as the outliers probability rises (at 50s and 150s in Case 1, the whole time in Case 2) and increase as the outliers probability decreases (at 100s in Case 1), while the ones using a smaller value (Mode 2) have an opposite trend. In Case 1, it shows that the mode probabilities estimation of MMMMStd is less sensitive to the sudden changes of outliers percentage at 50s and 100s, and obviously gives incorrect estimates when outliers percentage is high in the last 50s. Even if MMMMStd provides relatively close estimates of mode proba-

bilities to MVMMStd in [50, 150]s, the position estimation errors in these periods are still larger than that of MVMMStd as shown in Fig. 6 because high order information is not appropriately processed. In Case 2, combining the results from Fig. 7, we may find that MMMMStd underestimates the probability of Mode 1 when the degree of heavy-tailedness is mild and moderate except for the first 10s, whereas overestimates it when the degree escalates, and vice versa for Mode 2.

Finally, the computational complexity of MVMMStd is further compared to that of previous competing robust methods. The computational complexity of each algorithm is calculated by using the cputime function in Matlab. Let Std3 have a computational complexity of one, and the complexities of the rest are normalized relative to Std3. Normalized computational complexities obtained from Case 1 are given in Table 3.

As can be observed from Table 3, Std3 has the lowest computational complexity among all the competitors. RFBMA, HMKF, MMMMStd and MCKF are more computationally efficient than the rest competing methods apart from Std3. As expected, the computational complexity of MVMMStd increases when the number of iterations grows from 2 to 5 (Numbers following MVMMStd are iteration steps). MVMMStd-2 has almost an identical computational complexity to RStdKF-10, and is computationally much cheaper than GHSStdKF-10 and GHVGKF-10. Among all the iteration-based algorithms, MVMMStd has a very high computational complexity of a single iteration step because each iteration step of MVMMStd contains not only weighted models aggregation but repeated operations of matrix inversion and multiplication. However, if we take into account the estimation errors given previously in Table 1, Figs. 6, and 7, we may select appropriate iteration steps for MVMMStd and avoid spending unnecessary computational resources. The MVMMStd with 2 iteration steps already provides satisfactory results in both cases, thus  $L=2$  and  $L=3$  are preferable choices in our simulation. Using  $L=4$ ,  $L=5$  or larger iteration number does not provide distinct improvement but leads to 3~4 times the computational complexities of RFBMA, HMKF, MMMMStd and MCKF. Even though RFBMA, HMKF, MMMMStd and MCKF have notably lighter computational complexities than MVMMStd, the robustness of these algorithms are undermined

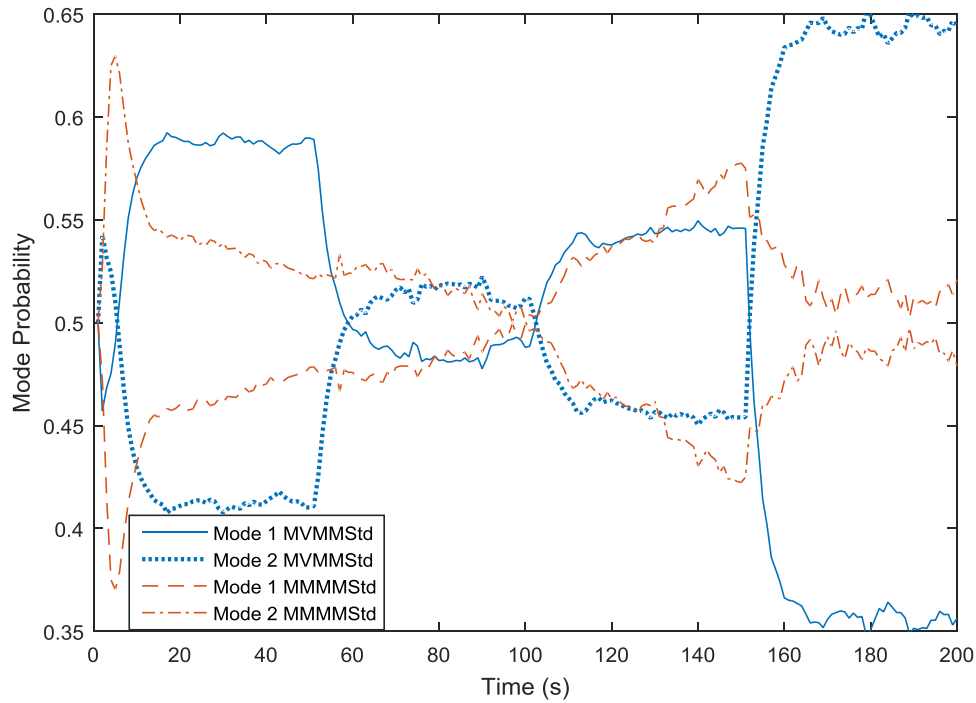


Fig. 8. Mode Probabilities in Case 1.

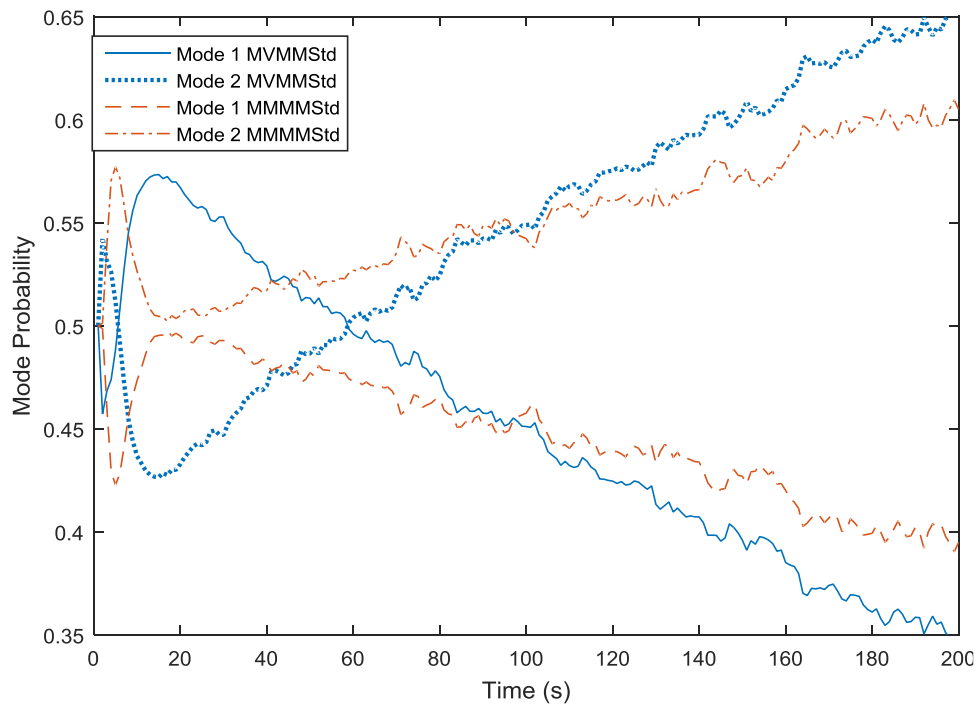


Fig. 9. Mode Probabilities in Case 2.

**Table 3**  
Normalized computational complexities.

Std3	MMMMStd	MVMMStd-2	MVMMStd-3	MVMMStd-4	MVMMStd-5
1	4.12	7.41	9.11	11.32	12.33
MCKF	RFBFMA	RStdKF-10	GHSSdKF-10	GHVGKF-10	HMKF
4.17	1.69	7.20	15.05	17.10	3.33

when the process noise and measurement noise are of time-variant heavy-tailedness.

## 5. Conclusion

In this paper, a flexible robust Student's t-based multimodel approach is proposed for the state estimation. The key feature of the proposed approach is that the unknown degrees of freedom is modelled by a Markov process and the MVC has been used to fuse multiple Student's t-based models. Results from the simulated data show the superiority of the proposed approach when dealing with the non-Gaussian noises with uncertain degree of non-Gaussianity, ranging from mild to severe heavy-tailedness, in spite of higher computational complexity. This approach is expected to offer better estimation performance in practice by either optimizing transition probabilities or selecting more accurate dof models if the prior knowledge about the noises is obtained. The proposed approach can be also used to develop the nonlinear counterparts with light computational loads using existing Student's t-based numerical methods appropriately.

## Funding

This work was supported by Zhejiang Provincial Natural Science Foundation of China (grant number [LQ18F030003](#)), [China Scholarship Council](#) (grant number [201908330102](#)), and the UK's [Engineering and Physical Sciences Research Council](#) (EPSRC) Programme Grant [EP/S016813/1](#).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Chen Shen:** Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review & editing. **Lyudmila Mihaylova:** Methodology, Writing - review & editing.

## Acknowledgment

The authors would like to thank the reviewers for the constructive suggestions for improving the paper and Dr. Yulong Huang for the discussion and advice during the stage of revision.

## References

- [1] R.E. Kalman, A new approach to linear filtering and prediction problems, *Trans. the ASME, J. Basic Eng.* 82 (March (1)) (1960) 35–45.
- [2] K. Ito, K. Xiong, Gaussian filters for nonlinear filtering problems, *IEEE Trans. Autom. Control* 45 (May (5)) (2000) 910–927.
- [3] I. Arasaratnam, S. Haykin, Cubature Kalman smoothers, *Automatica* 47 (Oct. (10)) (2011) 2245–2250.
- [4] A. Aravkin, J.V. Burke, L. Ljung, A. Lozano, G. Pillonetto, Generalized Kalman smoothing: Modeling and algorithms, *Automatica* 86 (Dec.) (2017) 63–86.
- [5] Z. Chen, Bayesian filtering: from Kalman filters to particle filters, and beyond, *Statistics* 181 (Jan. (1)) (2003) 1–69.
- [6] F. Daum, Nonlinear filters: beyond the Kalman filter, *IEEE Trans. Aerosp. Electron. Syst.* 20 (Aug. (8)) (2005) 57–69.
- [7] V. Stojanovic, N. Nedic, Robust Kalman filtering for nonlinear multivariable stochastic systems in the presence of non-Gaussian noise, *Int. J. Robust Nonlin.* 26 (Feb. (3)) (2016) 445–460.
- [8] W. Li, S. Sun, Y. Jia, J. Du, Robust unscented Kalman filter with adaptation of process and measurement noise covariances, *Digit. Signal Process.* 48 (Jan.) (2016) 93–103.
- [9] K. Li, L. Chang, B. Hu, A variational Bayesian-based unscented Kalman filter with both adaptivity and robustness, *IEEE Sens. J.* 16 (Sept. (18)) (2016) 6966–6976.
- [10] R. Piche, S. Sarkka, J. Hartikainen, Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution, in: *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, 2012, pp. 1–6.
- [11] Y. Huang, Y. Zhang, N. Li, J. Chambers, A robust Gaussian approximate filter for nonlinear systems with heavy tailed measurement noises, in: *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, 2016, pp. 4209–4213.
- [12] H. Zhu, H. Leung, Z. He, A variational Bayesian approach to robust sensor fusion based on Student-t distribution, *Inf. Sci.* 221 (Feb.) (2013) 201–214.
- [13] L. Chang, K. Li, B. Hu, Huber's M-estimation-based process uncertainty robust filter for integrated INS/GPS, *IEEE Sens. J.* 15 (June (6)) (2015) 3367–3374.
- [14] Y. Huang, Y. Zhang, A new process uncertainty robust Student's t based Kalman filter for SINS/GPS integration, *IEEE Access* 5 (July (7)) (2017) 14391–14404.
- [15] Y. Huang, Y. Zhang, Z. Wu, N. Li, J. Chambers, A novel robust Student's t-based Kalman filter, *IEEE Trans. Aerosp. Electron. Syst.* 53 (Feb. (1)) (2017) 1545–1554.
- [16] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian, J.A. Chambers, Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking, *IEEE Trans. Syst., Man, Cybern. (Oct. (10))* (2019) 2082–2096 no. 49.
- [17] B. Chen, X. Liu, H. Zhao, J.C. Principe, Maximum correntropy Kalman filter, *Automatica* 76 (Feb.) (2017) 70–77.
- [18] X. Liu, B. Chen, B. Xu, Z. Wu, P. Honeine, Maximum correntropy unscented filter, *Int. J. Syst. Sci.* 48 (Jan. (8)) (2017) 1607–1615.
- [19] M. Roth, E. Ozkan, F. Gustafsson, A Student's t filter for heavy tailed process and measurement noise, in: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 5770–5774.
- [20] Y. Huang, Y. Zhang, N. Li, S.M. Naqvi, J. Chambers, A robust student's t based cubature filter, in: *Proc. Int. Conf. Inf. Fusion*, 2016, pp. 9–16.
- [21] F. Tronarp, R. Hostettler, S. Sarkka, Sigma-point filtering for nonlinear systems with non-additive heavy-tailed noise, in: *Proc. 19th Int. Conf. Inf. Fusion*, Heidelberg, Germany, 2016, pp. 1859–1866.
- [22] O. Straka, J. Dunik, Stochastic integration Student's-t filter, in: *Proc. 20th Int. Conf. Inf. Fusion*, Xi'an, China, 2017, pp. 1–8.
- [23] J. Pruher, F. Tronarp, T. Karvonen, S. Sarkka, O. Straka, Student-t process quadratures for filtering of non-linear systems with heavy-tailed noise, in: *Proc. 20th Int. Conf. Inf. Fusion*, Xi'an, China, 2017, pp. 1–8.
- [24] S.S. Khalid, N.U. Rehman, S. Abrar, L. Mihaylova, Robust Bayesian filtering using Bayesian model averaging and restricted variational Bayes, in: *Proc. 21st Int. Conf. Inf. Fusion*, Cambridge, UK, 2018, pp. 361–368.
- [25] A. Kiring, Shrinkage based particle filters for tracking in wireless sensor networks with correlated sparse measurements," Ph.D. dissertation, Dept. Autom. Control Syst. Eng., Univ. Sheffield, Sheffield, UK, 2018.
- [26] B. Liu, Robust particle filter by dynamic averaging of multiple noise models, in: *Proc. 42nd IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 4034–4038.
- [27] H.A.P. Blom, Y. Bar-Shalom, The interacting multiple model algorithm for systems with Markovian switching coefficients, *IEEE Trans. Autom. Control* 33 (Aug. (8)) (1988) 780–783.
- [28] F. Huang, J. Zhang, S. Zhang, Maximum Versoria criterion-based robust adaptive filtering algorithm, *IEEE Trans. Circuits Syst. II: Exp. Briefs* 64 (Oct. (10)) (2017) 1252–1256.
- [29] S. Jain, R. Mitra, V. Bhatia, Kernel adaptive filtering based on maximum Versoria criterion, in: *Proc. 2018 IEEE Int. Conf. Adv. Netw. Telecommun. Syst.*, Indore, India, 2018.
- [30] X. Yu, J. Liu, H. Li, An adaptive inertia weight particle swarm optimization algorithm for IIR digital filter, in: *Proc. Int. Conf. Artif. Intell. Comput. Intell.*, Shanghai, China, 2009, pp. 114–118.
- [31] Z. Zhang, S. Zhang, J. Zhang, Robust weight-constraint decorrelation normalized maximum Versoria algorithm, 2019 Ninth Int. Workshop Signal Design Appl. Commun., 2019.
- [32] Q. Wu, Y. Li, W. Xue, A kernel recursive maximum Versoria-like criterion algorithm for nonlinear channel equalization, *Symmetry* 11 (Aug. (9)) (2019) 1067.
- [33] J. Liao, B. Chen, Robust mobile location estimator with NLOS mitigation using interacting multiple model algorithm, *IEEE Trans. Wireless Commun.* 5 (Nov. (11)) (2006) 3002–3006.
- [34] W. Li, Y. Jia, Location of mobile station with maneuvers using an IM-M-based cubature Kalman filter, *IEEE Trans. Ind. Electron.* 59 (Nov. (11)) (2012) 4338–4348.
- [35] G. Battistelli, L. Chisci, Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability, *Automatica* 50 (March) (2014) 707–718.
- [36] R.R. Pitre, V.P. Jilkov, X.R. Li, A comparative study of multiple-model algorithms for maneuvering target tracking, in: *Proc. SPIE 5809, Signal Process., Sensor Fusion, Target Recogn. XIV*, 2005, doi:10.1117/12.609681.
- [37] B. Chen, J. Wang, H. Zhao, N. Zheng, J.C. Principe, Convergence of a fixed-point algorithm under maximum correntropy criterion, *IEEE Signal Process. Lett.* 22 (Oct. (10)) (2015) 1723–1727.
- [38] X.R. Li, Multiple-model estimation with variable structure. Part II: model-set adaptation, *IEEE Trans. Autom. Control* 45 (Nov. (11)) (2000) 2047–2060.
- [39] L. Chang, B. Hu, G. Chang, A. Li, Robust derivative-free Kalman filter based on Huber's M-estimation methodology, *J. Process Control* 23 (Nov. (2013)) (2013) 1555–1561.