RESEARCH

Model set adaptive filtering algorithm using variational Bayesian approximations and Rényi information divergence

Tianli Ma[1,2], ChaoBo Chen[2,3] and Song Gao[2,3]*

Abstract

The paper presents a model set adaptive filtering algorithm based on variational Bayesian approximation (MSA-VB) for the target tracking system with the model and noise uncertainties. The Rényi information divergence, as a criterion, is to choose the best match model that has the minimum divergence between candidate models and true mode. Subsequently, the model-conditioned estimation based on variational Bayesian approximation is proposed to estimate system state and measurement noise variances. To deal with the coupled noise intractability, the moments matching technique is used to obtain the mixed statistics of measurement noise at the fusion stage. The proposed algorithm is compared with the interacting multiple models (IMM) algorithm and the variational Bayesian-interacting multiple models (IMM-VB) algorithm via two scenarios for maneuvering target tracking, and simulation results show that the MSA-VB has improved estimation and tracking performance.

Keywords: Target tracking, Variational Bayesian, Model set adaptive, System model uncertainty, Rényi information divergence

1 Introduction

Target tracking plays an important role in a variety of practical applications, such as underwater sonar tracking [1], aircraft surveillance [2], and visual tracking [3, 4]. The objective is to accurately estimate the target state for a sequence of observation sets in presence of noise. When the systems are linear and noise are Gaussian, Kalman filter provides an optimal filtering technique to estimate the target state [5]. Its variant have been studied under numerous relaxed assumptions [6–9]. Of course, this limitation in these filters is that it assumes a complete prior knowledge of the dynamic and measurement model parameters, including the noise statistics. In many practical situations, model uncertainty is caused by unknown dynamic model parameters, and noise uncertainty is caused by

unknown noise statistics; both violate the abovementioned assumption.

The model uncertainty is due to the fact that the target dynamics cannot be properly modeled by a single state space model (SSM) [10]. For example, in the maneuvering target tracking, the target has different maneuvering behaviors, such as constant velocity, acceleration, and turning with different angular velocity. Thus, the interacting multiple models (IMM) algorithm was proposed in [11], where the multiple models are used to match the different maneuvering behaviors and the transition among different models is subject to a Markov process. Since the state estimation of each model is parallel and independent, the computational complexity of the algorithm increases gradually as the number of models increases and the conflict among the models are more significant. In order to deal with the problem, Li and BarShalom [12] proposed the variable structure multiple model methods (VSMM), which adjusts their model sets in real time. In VSMM, the core idea is the model set adaptation (MSA)

*Correspondence: gaos@xatu.edu.cn
[2]Autonomous Systems and Intelligent Control International Joint Research Center, Xi'an Technological University, Xi'an, China
[3]School of Electronic Information Engineering, Xi'an Technological University, Xi'an, China
Full list of author information is available at the end of the article

that aims at finding the best model set for the state estimation. Different implementations for MSA approaches have been proposed. For instance, the expected mode augmentation (EMA) [13], the likely model set (LMS) [14], the minimal sub model set (MSS) [15], and the best model augmentation (BMA) [16].

In addition, the noise uncertainty also affects the performance of the tracking system. Practically, the measurement noise usually varies with interferences. To cope with the noise uncertainty, the adaptive filtering method is used to address the issue of state estimation in the case of unknown noise statistics in [17]. The adaptive filtering algorithms are classified into four categories: maximum likelihood, correlation, covariance matching, and Bayesian method. Among them, the Bayesian method can be seen as a more general case of the other three algorithms. However, most of the Bayesian algorithms are difficult to get the analytical solution—because of the complexity of the probability density function and the high dimensional integral. These adaptive filters considered that the sequence of state variable follows a first-order Markov process. For the stationary random sequence (e.g., image and video signal [18]), the denoising technique based on the intersection of confidence intervals (ICI) rule was presented to provide a noise-free image or its best possible estimate [19, 20].

Recently, the variational Bayesian approximation adaptive Kalman filter (VBAKF) proposed in [17] has been introduced to estimate the target state under the case without knowledge of measurement noise variances. Its main idea is that the joint posterior of the target state and measurement noise variances can be approximated by a factored free-form distribution (for models in the conjugate-exponential class). Unfortunately, linear system and Gaussian distribution assumption do not really exist in actual applications. The VBAKF cannot achieve the demanding filtering performance. The nonlinear estimation methods, such as unscented Kalman filter (UKF) and cubature Kalman filter (CKF), were combined with VB approximation method [21, 22]. Here, the measurement noise variances are approximated by the variational Bayesian approximation (VB) approach; thereafter, system states are updated by these nonlinear estimation methods. Hu [23] proposed the robust version of VBAKF, which models the measurement noise by using the Student t distribution, and [23] was extended in [24] and [25].

However, the abovementioned algorithms considered only one of these uncertainties. In real environment, the model and noise uncertainties have to be considered simultaneously. Several suggestions for dealing with this problem can be found in literature. In [26], a novel estimator was presented for the jump Markov linear systems with unknown measurement noise variance parameters. A merging scheme is adopted for the system noises in the fusion stage of the IMM approach and a fix recursive form is used to estimate the noise variance parameters. Based on the literature [26], Hong [27] presented a robust variational Bayesian-interacting multiple model (IMM-VB), which models the glint noise by using Gaussian mixture distribution. Gao [28] proposed an interacting multiple model estimation-based adaptive robust UKF, which establishes an adaptive fading UKF for the case of process model uncertainty and a robust UKF for the case of measurement model uncertainty. These approaches have obtained better performance for the problem of the absence of model and noise uncertainties, but they have a very high burden of time complexity. That is because more models have been designed in IMM algorithm for demanding results.

In this paper, we present a model set adaptive filtering algorithm based on variational Bayesian approximation to address the state estimation problem under the situation with dual uncertainties. The Rényi information divergence, as a criterion, is used to computing the divergence between the true mode and the candidate models. Subsequently, it develops a model-conditioned estimation based on variational Bayesian approximation to fuse the state and measurement noise variances. Two simulation experiments are provided to illustrate the effectiveness of the proposed algorithm.

The rest of paper is organized as follows. In Section 2, the state estimation problem with unknown noise statistics is formulated and the variational Bayesian method is briefly reviewed. The proposed algorithm is described in Section 3. In Section 4, the simulation results are presented to prove the effectiveness of the proposed algorithms. Finally, the conclusions are given in Section 5.

2   Methods/experimental
The main drawbacks of the interacting multiple model method for target tracking system are the high computational complexity and poor performance. A model set adaptive filtering algorithm based on variational Bayesian approximation is proposed in this paper. The proposed method is designed based on the idea of the VSMM and VB methods. The Rényi information divergence measures the "closeness" of two probability density functions. It has additional flexibility in that in allows for emphasis to be placed on specific portions of the support of the densities to be compared. Hence, the Rényi information divergence is as a criterion to choose the best match model that has the minimum divergence between candidate models and true mode. And the moments matching technique is used to obtain the mixed statistics of measurement noise and system state at the fusion stage.

The paper performs Monte Carlo simulation using MATLAB software to examine the behavior of the proposed method. The root mean square errors (RMSEs)

are used to evaluate the estimation accuracy. We compare our method with IMM-VB and IMM in two different scenarios. The parameters in the experiments are introduced in Section 5.

This paper does not contain any studies with human participants or animals performed by any of the authors.

3 Variational Bayesian approximation

3.1 Problem formulation
Consider the following state space model:

$$\begin{cases} x_k = F_k^{r_k} x_{k-1} + w_{k-1}^{r_k} \\ z_k = H_k^{r_k} x_k + v_k \end{cases} \tag{1}$$

where $x_k \in R^n$ and $z_k \in R^d$ are the target state and the measurement vectors, respectively. r_k denotes the system mode which is described by a discrete-time homogenous Markov chain. The process noise $w_{k-1}^{r_k}$ corresponding to mode r_k and the measurement noise v_k are assumed to be mutually independent zero-mean Gaussian random processes with the covariance matrices $Q_{k-1}^{r_k}$ and Σ_k, respectively. Here, we denote the diagonal covariance matrix comprising of these variances by $\Sigma_k = \text{diag}\{\sigma_{k,1}^2, \sigma_{k,2}^2 \ldots \sigma_{k,d}^2\}$. Due to the fact that the inverse Wishart distribution is the conjugate prior distribution for the variance of the Gaussian distribution [29]. For this reason, a product of inverse Wishart models is adopted to approximate the posterior distribution Σ_k. That is

$$p(\Sigma_k) = IW(\Sigma_k; \kappa_k, \Lambda_k) \tag{2}$$

where the notation $IW(\Sigma_k; \kappa_k, \Lambda_k)$ represents the inverse Wishart distribution for the variable Σ_k with the degree of freedom κ_k, and the symmetric positive definite matrix Λ_k.

Remark 1 *Since the inverse gamma distribution is a special case of the inverse wishart distribution in one-dimensional space, the discussed model in this paper is much more general and thus more information can be utilized for filter design. Interested readers are referred to [30] and [31] for a detailed introduction.*

However, Σ_k is unknown in most cases, which requires to joint estimate the posterior distribution of the target state and the measurement noise covariance. Assume that the dynamic model of the state and the covariance matrix are independent for any mode r_k, that is

$$p(x_k, \Sigma_k | x_{k-1}, \Sigma_{k-1}, r_k) = p(x_k | x_{k-1}, r_k) p(\Sigma_k | \Sigma_{k-1}, r_k) \tag{3}$$

The predicted joint distribution of the state and measurement noise are calculated by the Chapman-Kolmogorov equation.

$$\begin{aligned} p(x_k, \Sigma_k | r_k, Z_{k-1}) = \int p(x_k | x_{k-1}, r_k) p(\Sigma_k | \Sigma_{k-1}, r_k) \\ \times p(x_{k-1}, \Sigma_{k-1} | r_k, Z_{k-1}) \, dx_{k-1} \, d\Sigma_{k-1} \end{aligned} \tag{4}$$

When the measurement z_k is available, the joint posterior distribution is given by the Bays rule.

$$p(x_k, \Sigma_k | r_k, Z_k) = \frac{p(Z_k | x_k, \Sigma_k, r_k, Z_{k-1}) p(x_k, \Sigma_k | r_k, Z_{k-1})}{\int p(Z_k | x_k, \Sigma_k, r_k) p(x_k, \Sigma_k | r_k) \, dx_{k-1} \, d\Sigma_{k-1}} \tag{5}$$

where $p(Z_k | x_k, \Sigma_k, r_k, Z_{k-1})$ denotes the likelihood function which is related with Σ_k.

Notice that the two main problems need to be solved. One is the dynamic model of the measurement noise covariance $p(\Sigma_k | \Sigma_{k-1}, r_k)$ is unknown. The other is that the posterior density is difficult to achieve due to the involved intractable integrals. To calculate the posterior density with the unknown noise covariance, the variational Bayesian approximation method, which uses a simple free-from distribution to approximate the joint posterior density, is proposed.

3.2 Variational Bayesian approximation
Assume that the state vector and measurement noise covariance are independent, and the joint posterior density can be approximated by a free-form factored distribution as follows

$$p(x_k, \Sigma_k | r_k, Z_k) \approx Q_x(x_k) Q_\Sigma(\Sigma_k) \tag{6}$$

where the probability densities $Q_x(x_k)$ and $Q_\Sigma(\Sigma_k)$ are Gaussian distribution and inverse wishart distribution, respectively. The non-negative KL divergence represents the measure of the dissimilarity of the approximation and the true posterior, that can be expressed as

$$\begin{aligned} KL\left[Q_x(x_k) Q_\Sigma(\Sigma_k) \| p(x_k, \Sigma_k | r_k, Z_k)\right] \\ = \int Q_x(x_k) Q_\Sigma(\Sigma_k) \log \frac{Q_x(x_k) Q_\Sigma(\Sigma_k)}{p(x_k, \Sigma_k | r_k, Z_k)} \, dx_k \, d\Sigma_k \end{aligned} \tag{7}$$

The optimal approximation of the joint posterior density can be obtained by minimizing the KL divergence, and the mean field approximation is used to solve the calculation problem of multiple hidden variables [29]. The results are given as

$$Q_x(x_k) = N(x_k; \hat{x}_k, P_k) \tag{8}$$

$$Q_\Sigma(\Sigma_k) = \prod_{u=1}^{d} IW\left(\sigma_{k-1,u}^2; \kappa_{k-1,u}^i, \Lambda_{k-1,u}^i\right) \tag{9}$$

4 Model set adaptive diltering algorithm based on variational Bayesian approximation

In this section, a model set adaptive filtering algorithm is proposed. The model set adaptive approach is used to select the best model set for multiple model estimation. Moreover, the noise statistics and state of each model are estimated by the model set conditioned estimation based on variational Bayesian approximation.

4.1 Model-set adaptation

To address the problem of the model uncertainty, IMM-VB needs a lot of models to improve the algorithm performance. This has two obvious defects [13]. First, the computation load grows with the increase of the number of the models. Second, the competition among these models lead to performance decrease greatly.

To overcome this problem, a MSA algorithm based on Rényi divergence is proposed. Rényi information divergence is a distance measure between two densities (the test density f and the reference density f_0) [32]. The order α Rényi information divergence of f and f_0 is defined as

$$D_\alpha(f\|f_0) = -\frac{1}{1-\alpha} \ln \int f^\alpha(x) f_0^{1-\alpha}(x)\, dx \qquad (10)$$

For any order α, the Rényi information divergence takes on its minimum value if and only $f = f_0$. In our application, we wish to compute the divergence between the true mode s_k and candidate model $r_k^j \in \mathcal{M}^c$ at the time k. That is

$$
\begin{aligned}
D_z\left(s_k, r_k^j\right) &\triangleq D\left(p\left(z_k|s_k\right) \| p\left(z_k|r_k^j\right)\right) \\
&= -\frac{1}{1-\alpha} \ln \int p^\alpha(z_k|s_k) p^{1-\alpha}\left(z_k|r_k^j\right)\, dz_k
\end{aligned}
\qquad (11)
$$

where $p(z_k|s_k)$ and $p\left(z_k|r_k^j\right)$ are the probability density functions of z_k conditioned on s_k and r_k^j, respectively. Due to the true mode s_k of the system is unknown at the time k [16], it is assumed that the best online estimates of the probability density function of z_k for the true mode s_k can be approximated as $p(z_k|\mathcal{M}_{k-1}, s_k, \Sigma_{s_k}, Z_{k-1}) \approx p(z_k|\mathcal{M}_{k-1}, \mathcal{M}_k, \Sigma_k, Z_{k-1})$.

Let $p(z_k|s_k) = \mathcal{N}\left(z_k; \bar{z}_{s_k}, \Phi_{s_k}\right)$ and $p\left(z_k|r_k^j\right) = \mathcal{N}\left(z_k; \bar{z}_k^j, \Phi_k^j\right)$ be the Gaussian densities with vector means $\bar{z}_{s_k}$, $\bar{z}_k^j$ and positive definite covariance matrices Φ_{s_k}, Φ_k^j. The Rényi information divergence between $p(z_k|s_k)$ and $p\left(z_k|r_k^j\right)$ is

$$
\begin{aligned}
D_z\left(s_k, r_k^j\right) = &-\frac{1/2}{1-\alpha} \ln \frac{\left|\Phi_k^j\right|^\alpha \left|\Phi_{s_k}\right|^{1-\alpha}}{\left|\alpha\Phi_k^j + (1-\alpha)\Phi_{s_k}\right|} \\
&+ \frac{\alpha}{2}\triangle^T \left(\alpha\Phi_k^j + (1-\alpha)\Phi_{s_k}\right)^{-1}\triangle
\end{aligned}
\qquad (12)
$$

where $\triangle = \bar{z}_{s_k} - \bar{z}_k^j$. These means and covariance matrices of the Gaussian densities can be calculated as

$$
\begin{cases}
\bar{z}_{s_k} = E\left[z_k|\mathcal{M}_{k-1}, \mathcal{M}_k, Z_{k-1}\right] \\
\Phi_{s_k} = E\left[\left(z_k - \bar{z}_{s_k}\right)\left(z_k - \bar{z}_{s_k}\right)^T | \mathcal{M}_{k-1}, \mathcal{M}_k, Z_{k-1}\right] \\
\bar{z}_k^j = E\left[z_k|r_k^j, x_{k|k-1}^j, P_{k-1}^j\right] \\
\Phi_k^j = E\left[\left(z_k - \bar{z}_k^j\right)\left(z_k - \bar{z}_k^j\right)^T | r_k^j, x_{k|k-1}^j, P_{k-1}^j\right]
\end{cases}
\qquad (13)
$$

Remark 2 *Different selections of the parameter α allow for different parts of these distributions to be emphasized. In the limiting case of $\alpha \to 1$ the Rényi information divergence becomes the Kullback-Liebler divergence. The effect of the order to the Rényi information divergence is detailed and analyzed in [33–35]. The results show that $\alpha = 0.5$ emphasizes the tails of the distribution and allows for the maximum discrimination between two similar distributions. Therefore, we can obtain a better performance by choosing $\alpha = 0.5$ for tracking applications [36, 37].*

The optimal model $\hat{r}_k$ in the candidate model set $\mathcal{M}^c$ can be selected as the one with the minimum Rényi information divergence.

$$\hat{r}_k = \arg \min_{r_k^j \in \mathcal{M}^c} D_z\left(s_k, r_k^j\right) \qquad (14)$$

Thus, the adapted model set is the basic model set $\mathcal{M}^b$ combine with the model $\hat{r}_k$ at the time k.

$$\mathcal{M}_k = \mathcal{M}^b \bigcup \hat{r}_k \qquad (15)$$

4.2 Model-set conditioned estimation based variational Bayesian approximation

Suppose that the posterior probability density function of model i at the time k-1 is described as below

$$
\begin{aligned}
p\left(x_{k-1}, \Sigma_{k-1}|r_{k-1}^i\right) = &\mathcal{N}\left(x_{k-1}; \hat{x}_{k-1|k-1}^i, P_{k-1|k-1}^i\right) \\
&\times \prod_{u=1}^d IW\left(\sigma_{k-1,u}^2; \kappa_{k-1,u}^i, \Lambda_{k-1,u}^i\right)
\end{aligned}
\qquad (16)
$$

Note that Eq. (16) can be seen as a product of a Gaussian distribution and an inverse Wishart distribution. r_{k-1}^i means the event that model i matches the system model in effect at time $k - 1$, $r_{k-1}^i \in$

Ma *et al. EURASIP Journal on Advances in Signal Processing* (2020) 2020:17

Page 5 of 15

$\mathcal{M}_{k-1}$. The notation $\mathcal{N}\left(x_{k-1}; \hat{x}_{k-1|k-1}^i, P_{k-1|k-1}^i\right)$ represents the Gaussian probability density function with mean $\hat{x}_{k-1|k-1}^i$ and covariance $P_{k-1|k-1}^i$. The notation $IW\left(\sigma_{k-1,u}^2; \kappa_{k-1,u}^i, \Lambda_{k-1,u}^i\right)$ represents an inverse Wishart distribution with parameters $\kappa_{k-1,u}^i$ and $\Lambda_{k-1,u}^i$. By using the total probability theorem, one has

$$p(x_k, \Sigma_k|Z_k) = \sum_{j=1}^{\mathcal{M}_{k-1}} p\left(x_k, \Sigma_k|r_k^j, Z_k\right) p\left(r_k^j|Z_k\right) \quad (17)$$

In the model set conditional reinitialization stage, the joint probability density function is described as

$$
\begin{aligned}
&p\left(x_{k-1}, \Sigma_{k-1}|r_k^j, Z_{k-1}\right) \\
&= \sum_{i=1}^{\mathcal{M}_{k-1}} p\left(x_{k-1}, \Sigma_{k-1}|r_{k-1}^i, Z_{k-1}\right) p\left(r_{k-1}^i|r_k^j, \mathcal{M}_{k-1}\right) \\
&= \sum_{i=1}^{\mathcal{M}_{k-1}} \mu_{k-1}^{i|j} \mathcal{N}\left(x_{k-1}; \hat{x}_{k-1|k-1}^i, P_{k-1|k-1}^i\right) \\
&\quad \prod_{u=1}^{d} IW\left(\sigma_{k-1,u}^2; \kappa_{k-1,u}^i, \Lambda_{k-1,u}^i\right)
\end{aligned}
$$

$$(18)$$

where $\mu_{k-1}^{i|j}$ denotes the conditional probability that model j transfers to model i at the time k. It is calculated by

$$\mu_{k-1}^{i|j} = \frac{p\left(r_k^j|r_{k-1}^i\right) p\left(r_{k-1}^i|Z_{k-1}\right)}{p\left(r_k^j|Z_{k-1}\right)} = \frac{\pi_{ij}\mu_{k-1}^i}{\hat{\mu}_{k|k-1}^j} \quad (19)$$

where $\hat{\mu}_{k|k-1}^j$ is the normalization coefficient. Based on the model set $\mathcal{M}_k$, the output of each filter is merged in the fusion stage [13]. Therefore, we aim to approximate the sum term in (18) by a single one, that is

$$
\begin{aligned}
p\left(x_{k-1}, \Sigma_{k-1}|r_k^j\right) &= \mathcal{N}\left(x_{k-1}; \hat{x}_{k-1|k-1}^{0j}, P_{k-1|k-1}^{0j}\right) \\
&\quad \times \prod_{u=1}^{d} IW\left(\sigma_{k-1,u}^{0j}; \kappa_{k-1,u}^{0j}, \Lambda_{k-1,u}^{0j}\right)
\end{aligned}
$$

$$(20)$$

where $\hat{x}_{k-1|k-1}^{0j}$ and $P_{k-1|k-1}^{0j}$ are mixed state and covariance matrix of the model j, respectively.

$$\hat{x}_{k-1|k-1}^{0j} = \sum_{i=1}^{\mathcal{M}_{k-1}} \mu_{k-1}^{i|j} \hat{x}_{k-1|k-1}^i \quad (21)$$

$$
\begin{aligned}
P_{k-1|k-1}^{0j} &= \sum_{i=1}^{\mathcal{M}_{k-1}} \mu_{k-1}^{i|j} P_{k-1|k-1}^i \\
&\quad + \sum_{i=1}^{\mathcal{M}_{k-1}} \mu_{k-1}^{i|j} \left(\hat{x}_{k-1|k-1}^i - \hat{x}_{k-1|k-1}^{0j}\right)(\cdot)^T
\end{aligned}
$$

$$(22)$$

On the basis of the moment matching theory [26], the first and second moments of the variable $\sigma_{k-1,u}^{0j}$ in (20) can be obtained as follows

$$E\left[\sigma_{k-1,u}^{0j}\right] = \frac{\Lambda_{k-1,u}^{0j}}{\kappa_{k-1,u}^{0j} - d - 1} \quad (23)$$

$$Var\left[\sigma_{k-1,u}^{0j}\right] = \frac{2\left(\Lambda_{k-1,u}^{0j}\right)^2}{\left(\kappa_{k-1,u}^{0j} - d - 1\right)^2 \left(\kappa_{k-1,u}^{0j} - d - 3\right)} \quad (24)$$

The mean and variance of the inverse Wishart sum distribution in (18) are given by

$$E\left[\sigma_{k-1,u}^i\right] = \sum_{i=1}^{\mathcal{M}_{k-1}} \mu_{k-1}^{i|j} \frac{\Lambda_{k-1,u}^i}{\kappa_{k-1,u}^i - d - 1} \quad (25)$$

$$
\begin{aligned}
Var\left[\sigma_{k-1,u}^i\right] &= \sum_{i=1}^{\mathcal{M}_{k-1}} \mu_{k-1}^{i|j} \frac{2\left(\Lambda_{k-1,u}^i\right)^2}{\left(\kappa_{k-1,u}^i - d - 1\right)^2 \left(\kappa_{k-1,u}^i - d - 3\right)} \\
&\quad + \sum_{i=1}^{\mathcal{M}_{k-1}} \mu_{k-1}^{i|j} \frac{\Lambda_{k-1,u}^i}{\kappa_{k-1,u}^i - d - 1} - E\left[\sigma_{k-1,u}^i\right]^2
\end{aligned}
$$

$$(26)$$

By solving Eqs. (23) and (24), the parameters $\kappa_{k-1,u}^{0j}$ and $\Lambda_{k-1,u}^{0j}$ are calculated as follows

$$\kappa_{k-1,u}^{0j} = \frac{2\left(E\left[\sigma_{k-1,u}^i\right]\right)^2}{Var\left[\sigma_{k-1,u}^i\right]} + d + 3 \quad (27)$$

$$\Lambda_{k-1,u}^{0j} = \left(\frac{2\left(E\left[\sigma_{k-1,u}^i\right]\right)^2}{Var\left[\sigma_{k-1,u}^i\right]} + 2\right) E\left[\sigma_{k-1,u}^i\right] \quad (28)$$

Remark 3 *In the proposed algorithm, the key feature is that the Gaussian sum distribution for the estimated state is approximated by a single Gaussian distribution. Similarly, the inverse Wishart sum distribution for the measurement noise covariance matrix Σ_k is approximated by a single inverse Wishart distribution by matching the first and second moments.*

Then, the mixed target state and measurement noise covariance are taken as the filter input. The predict density is computed from Eq. (3).

$$
\begin{aligned}
p\left(x_k, \Sigma_k | r_k^j, Z_{k-1}\right) \\
= \int p\left(x_k | x_{k-1}, r_k^j\right) p\left(\Sigma_k | \Sigma_{k-1}, r_k^j\right) p\left(x_{k-1}, \Sigma_{k-1} | r_k^j, Z_{k-1}\right) dx_{k-1} d\Sigma_{k-1} \\
= \mathcal{N}\left(x_k; \hat{x}_{k|k-1}^j, P_{k|k-1}^j\right) \prod_{u=1}^{d} IW\left(\sigma_{k-1,u}^2; \kappa_{k-1,u}^{j-}, \Lambda_{k-1,u}^{j-}\right)
\end{aligned}
$$

(29)

where, taking account into the time variation of the parameters, a forgetting factor ρ is introduced, $\rho \in (0,1]$. Note that the closer ρ is to 0, the more instability the parameters will be in terms of time-fluctuations [38]. The system state, covariance matrix, and the parameters of the inverse wishart distribution are predicted by

$$
\hat{x}_{k|k-1}^j = F_{k-1}^j x_{k-1|k-1}^{0j} \tag{30}
$$

$$
P_{k|k-1}^j = F_{k-1}^j P_{k-1|k-1}^{0j} (F_{k-1}^j)^T + Q_{k-1}^j \tag{31}
$$

$$
\kappa_{k-1,u}^{j-} = \rho \kappa_{k-1,u}^{0j} \tag{32}
$$

$$
\Lambda_{k-1,u}^{j-} = \rho \Lambda_{k-1,u}^{0j} \tag{33}
$$

After receiving the measurement z_k, VB approximation method can be used to obtain a free-form factored approximate distribution for $p\left(x_k, \Sigma_k | r_k^j, Z_k\right)$ [38], the analytical expression of the joint posterior probability density function is given by

$$
\begin{aligned}
p\left(x_k, \Sigma_k | r_k^j, Z_k\right) &\propto p\left(z_k | x_k, \Sigma_k, r_k^j\right) p\left(x_{k-1}, \Sigma_{k-1} | r_k^j, Z_k\right) \\
&= \mathcal{N}\left(z_k; H_k^j x_{k|k-1}^j, \Sigma_k\right) \mathcal{N}\left(x_k^j; \hat{x}_{k|k-1}^j, P_{k|k-1}^j\right) \\
&\quad \prod_{u=1}^{d} IW\left(\sigma_{k-1,u}^2; \kappa_{k-1,u}^{j-}, \Lambda_{k-1,u}^{j-}\right)
\end{aligned}
$$

(34)

To calculate the distributions of state and measurement noise covariance of the ith model, VB assumes that the joint posterior distribution in (34) can be factorized as the product of $q\left(x_k^j\right)$ and $q\left(\Sigma_k^j\right)$. The logarithm of $q\left(x_k^j\right)$ can be computed by fixing the $q\left(\Sigma_k^j\right)$.

$$
\begin{aligned}
\log\left(q\left(x_k^j\right)\right) &\propto -\frac{1}{2}\left(z_k - H_k^j \hat{x}_{k|k-1}^j\right)^T \left(\Sigma_k^j\right)^{-1} \left(z_k - H_k^j \hat{x}_{k|k-1}^j\right) \\
&\quad -\frac{1}{2}\left(x_k^j - \hat{x}_{k|k-1}^j\right)^T \left(P_{k|k-1}^j\right)^{-1} \left(x_k^j - \hat{x}_{k|k-1}^j\right)
\end{aligned}
$$

(35)

Through the simplified formula, $p\left(x_{k-1}, \Sigma_{k-1} | r_k^j, Z_k\right)$ is approximated by

$$
q\left(x_k^j\right) = \mathcal{N}\left(x_k^j; \hat{x}_{k|k-1}^j, P_k^j\right) \tag{36}
$$

The mean and covariance of the Gaussian distribution are derived by the VB approximation as follows

$$
\begin{aligned}
\hat{x}_{k|k}^j &= \hat{x}_{k|k-1}^j \\
&\quad + P_{k|k-1}^j H_k^j \left(\Sigma_k^j + \left(H_k^j\right)^T \left(P_{k|k-1}^j\right)^{-1} H_k^j\right)^{-1} \left(z_k - H_k^j \hat{x}_{k|k-1}^j\right)
\end{aligned}
$$

(37)

$$
P_k^j = \left(\left(P_{k|k-1}^j\right)^{-1} + \left(H_k^j\right)^T \left(\Sigma_k^j\right)^{-1} H_k^j\right)^{-1} \tag{38}
$$

Similarly, the logarithm of $q\left(\Sigma_k^j\right)$ can be calculated by keeping $q\left(x_k^j\right)$ fixed such as

$$
\begin{aligned}
\log\left(q\left(\Sigma_k^j\right)\right) &\propto -\frac{\kappa + d + 2}{2} \log\left(\Sigma_k^j\right) + \left(H_k^j\right)^T P_k^j H_k^j \Sigma_k^{-1} \\
&\quad -\frac{1}{2} tr\left(\Lambda_k^j + \left(z_k - H_k^j \hat{x}_{k|k-1}^j\right)^T \left(z_k - H_k^j \hat{x}_{k|k-1}^j\right)\right)
\end{aligned}
$$

(39)

Here, $q\left(\Sigma_k^j\right)$ is approximated as

$$
q\left(\Sigma_k^j\right) = IW\left(\sigma_{k,u}^2; \kappa_{k,u}^j, \Lambda_{k,u}^j\right) \tag{40}
$$

where

$$
\kappa_{k,u}^j = \kappa_{k-1,u}^{j-} + \frac{d}{2} \tag{41}
$$

$$
\Lambda_{k,u}^j = \Lambda_{k-1,u}^{j-} + \left(z_k - H_k^j \hat{x}_{k|k}^j\right)_u^2 + \left(\left(H_k^j\right)^T P_k^j H_k^j\right)_{uu} \tag{42}
$$

where $u = 1, 2 \dots d$, and

$$
\hat{\Sigma}_k^j = \mathrm{diag}\left(\frac{\Lambda_{k,1}^j}{\kappa_{k,1}^j - d - 1}, \dots, \frac{\Lambda_{k,d}^j}{\kappa_{k,d}^j - d - 1}\right)
$$

Note that Eqs. (37), (38), (41), and (42) are coupled. The fixed-point algorithm is used to proceeded alternatively the state and noise parameters until the convergence is reached. It has been proved that VB converge very fast and most of the time, only a few iterations in [17].

By using the Bayes rule, the probability of each model is updated

$$
\begin{aligned}
\mu_k^j &= p\left(r_k^j | \mathcal{M}_k, Z_k\right) = \frac{p\left(r_k^j | z_k\right) p\left(z_k | r_k^i, z_{k-1}\right)}{p(z_k | z_{k-1})} \\
&= \frac{\hat{\mu}_{k-1}^j L_k^j}{\sum_{i=1}^{\mathcal{M}_k} \hat{\mu}_{k-1}^i L_k^i}
\end{aligned}
$$

(43)

Here, the likelihood function L_k^j of model j is calculated

$$
\begin{aligned}
L_k^j &= p\left(z_k | r_k^j, \mathcal{M}_{k-1}, Z_{k-1}\right) \\
&= \mathcal{N}\left(z_k; H_k^j \hat{x}_{k|k-1}^j, \left(H_k^j\right)^T P_k^j H_k^j + \hat{\Sigma}_k^j\right)
\end{aligned}
\tag{44}
$$

Finally, based on the mixed equation, the state and the covariance in the fusion stage are shown as

$$
\hat{x}_{k|k} = E[x_k | \mathcal{M}_k, z_k] = \sum_{j=1}^{\mathcal{M}_k} \mu_k^j \hat{x}_k^j
\tag{45}
$$

$$
\begin{aligned}
P_{k|k} &= E\left[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T | \mathcal{M}_k, z_k\right] \\
&= \sum_{j=1}^{\mathcal{M}_k} \mu_k^j \left(P_k^j + \left(\hat{x}_{k|k} - \hat{x}_k^j\right)\left(\hat{x}_{k|k} - \hat{x}_k^j\right)^T\right)
\end{aligned}
\tag{46}
$$

The computational complexity of the model set adaptive filter is $O\left(ln^3\right)$, where n is the dimension of the state and l is the number of the model set adaptive filter. The estimation of the measurement noise in VB step involves the following parameters: $O(ld + l)$ for the posteriors of the means, and $O\left(ld^2 + l\right)$ for the posteriors of precision, where d is the dimension of the measurement. So the computational complexity of the proposed algorithm is $O\left(N_t\left[ln^3 + l + ld + l + ld^2\right]\right) = O\left(N_t\left[ln^3 + ld^2\right]\right)$, where N_t is the total number of VB iterations. A description of the proposed algorithm is summarized in the following:

Algorithm 1 The MSA-VB algorithm

Input: Measurement data z_k, the basic model set $\mathcal{M}^b$, the candidate model set for activation $\mathcal{M}^c$, process noise covariance Q, prior variational parameters $\kappa_{0,u}$ and $\Lambda_{0,u}$, state prior parameters x_0, P_0, and the number of VB fixed-point iterations N_t

Output: System state x_k, measurement noise covariance parameters $\kappa_{k,u}$, $\Lambda_{k,u}$
1: Initialize the posterior variational parameters
2: **for** $k = 1 : N$ **do**
3:     **if** $k = 1$ **then**
4:         $\mathcal{M}_k = \mathcal{M}^b$, the system state x_k is estimated by the MSA-VB based on $\mathcal{M}_k$.
5:     **else**
6:         Calculate the Rényi information divergence Eq. (12) for $\hat{r}_k \in \mathcal{M}^c$. The optimal model $\hat{r}_{k-1}$ is replaced with the optimal model at the time k. The system state x_k is estimated by using the MSA-VB based on the new model set $\mathcal{M}_k$.
7:     **end if**
8: **end for**

---

5 Simulation results
In this section, numerical simulations are carried out in order to compare the performance of the five algorithms: IMM3 (3 basic models), IMM11 (3 basic models and 8 CT models), IMM3-VB (3 basic models), IMM11-VB (3 basic models and 8 CT models), and MSA-VB (3 basic models and 8 CT models as the candidate models) with unknown measurement noise and system model. The reference system dynamic can be described by the following state space model

$$
x_{k+1} = \begin{bmatrix}
1 & \frac{\sin(\omega t)}{\omega} & 0 & \frac{\cos(\omega t) - 1}{\omega} \\
0 & \cos(\omega t) & 0 & -\sin(\omega t) \\
0 & \frac{1 - \cos(\omega t)}{\omega} & 1 & \frac{\sin(\omega t)}{\omega} \\
0 & \sin(\omega t) & 0 & \cos(\omega t)
\end{bmatrix} x_k + w_k
\tag{47}
$$

where $x_k = \begin{bmatrix} \phi_k & \dot{\phi}_k & \varphi_k & \dot{\varphi}_k \end{bmatrix}^T$, ϕ_k and $\dot{\phi}_k$ are the target position and velocity in the X-directions, respectively. φ_k and $\dot{\varphi}_k$ are the target position and velocity in the Y-directions, respectively. ω denotes the turn rate. t is the sampling time, and $t = 1s$. The root mean square error (RMSE) of position and velocity are used to evaluate the performance of the proposed algorithm.

For the proposed algorithm, the system state is estimated based on the model set with the basic model set $\mathcal{M}^b$ and the candidate model set $\mathcal{M}^c$. The number of $\mathcal{M}^b$ and $\mathcal{M}^c$ are 3 and 8, respectively. Those models in the model set are constant turn (CT) models [1]. These models differ only in the turn rate ω_i, which belongs to the basic model set $\mathcal{M}^b$ that consists of the initial models of the IMM algorithm and the candidate model set $\mathcal{M}^c$ that consists of these models with different structures compared with the basic models.

5.1 Experiment 1
In this case, a single target moves in a 2-D scenario with $[-100, 100]\,m \times [-100, 100]\,m$ surveillance region. The turn rate in different time periods is shown in Table 1.

The measurement equation is expressed as

$$
z_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x_k + v_k
\tag{48}
$$

Table 1 The turn rate in different time periods

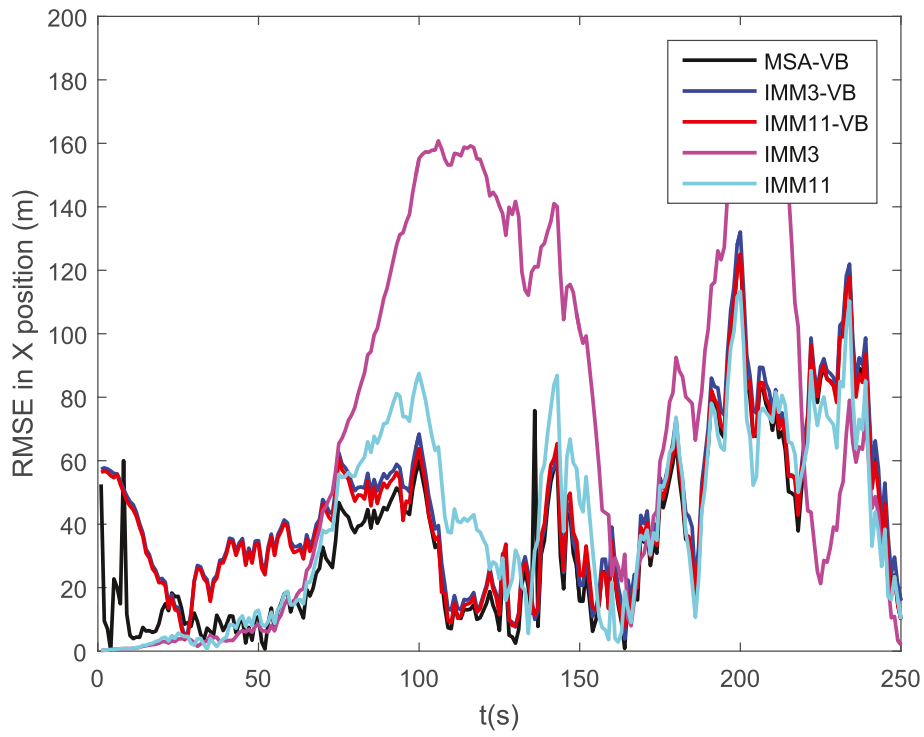| Time periods (s)	ω (rad/s)
1–50	1e− 5
51–100	0.0116
101–140	1e− 5
141–200	− 0.0116
201–225	0.0314
226–250	− 0.0314

Fig. 1 RMSEs in *X* position of the proposed algorithm, IMM3, IMM11, IMM3-VB, and IMM11-VB in scenario 1
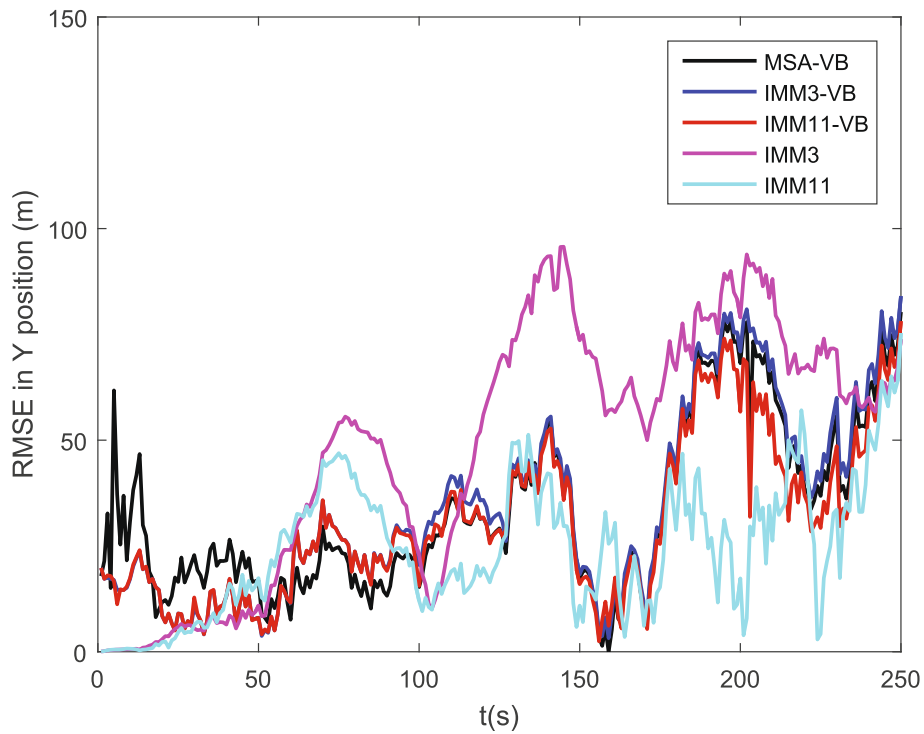

Fig. 2 RMSEs in *Y* position of the proposed algorithm, IMM3, IMM11, IMM3-VB, and IMM11-VB in scenario 1
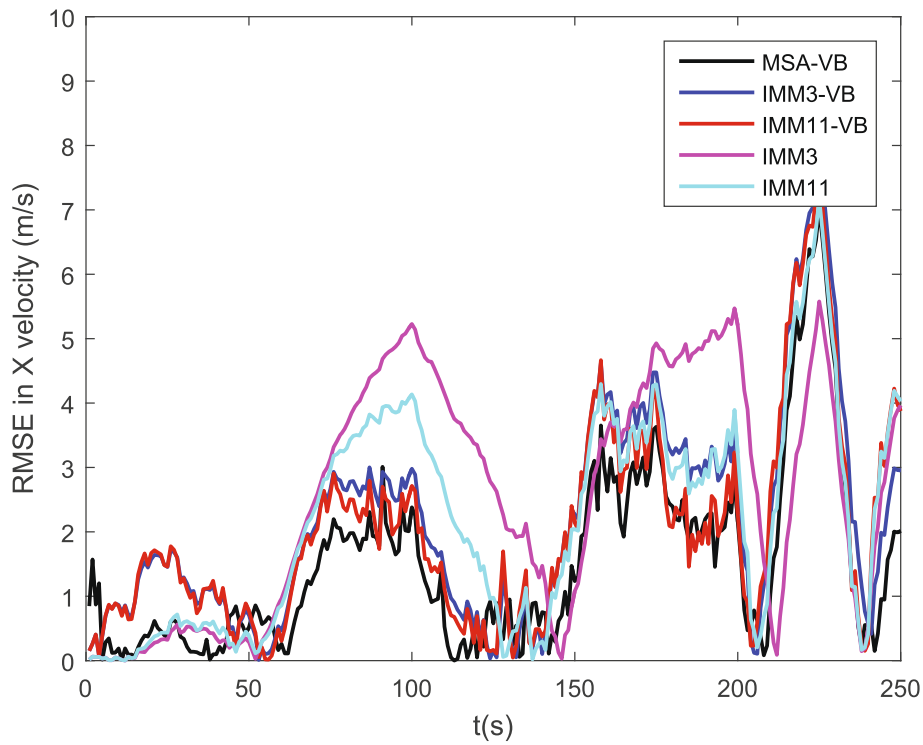
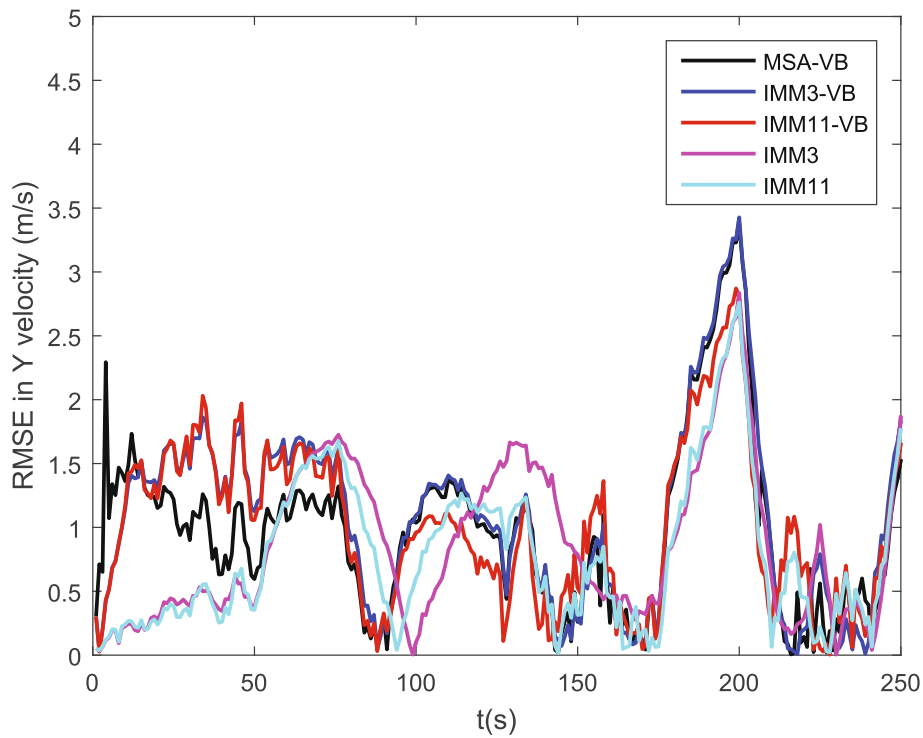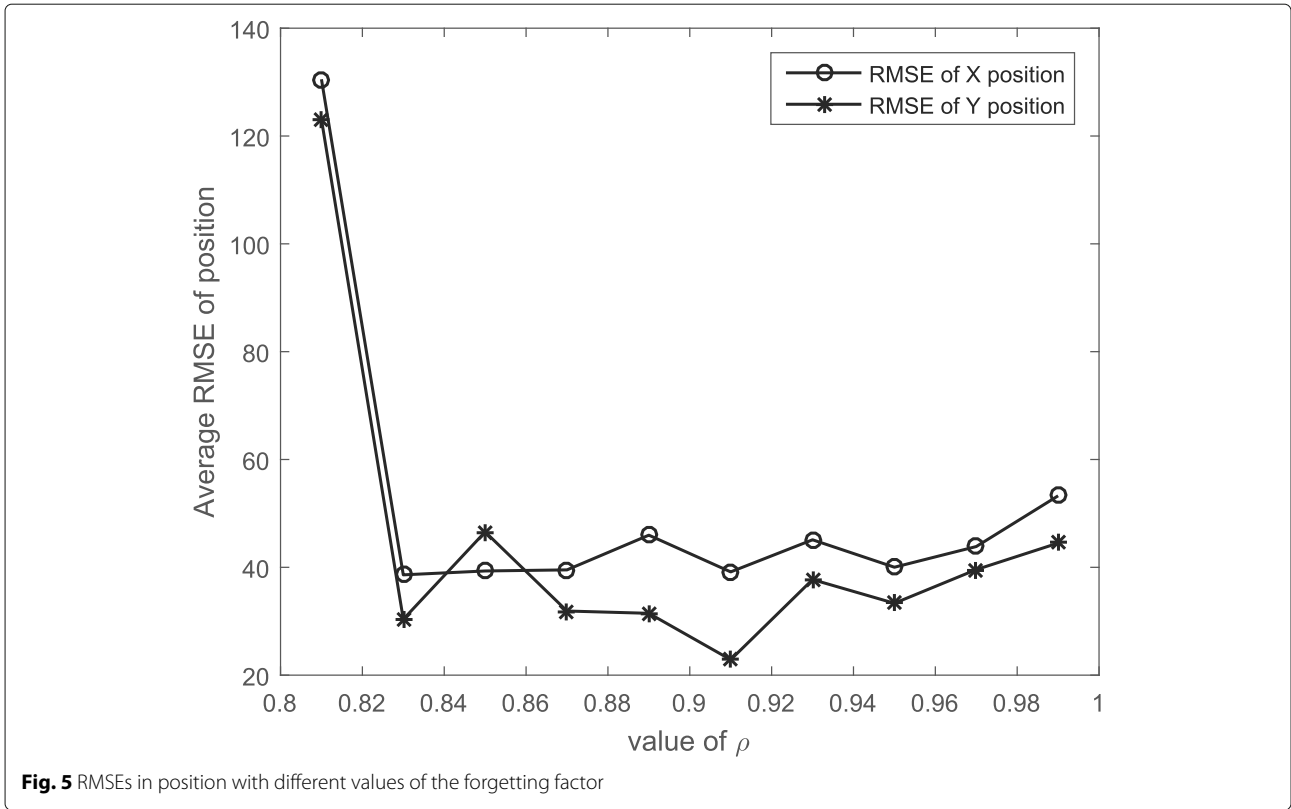**Fig. 3** RMSEs in *X* Velocity of the proposed algorithm, IMM3, IMM11, IMM3-VB, and IMM11-VB in scenario 1

Fig. 4 RMSEs in *Y* Velocity of the proposed algorithm, IMM3, IMM11, IMM3-VB, and IMM11-VB in scenario 1

Fig. 5 RMSEs in position with different values of the forgetting factor

where measurement noise ν_k is the zero-mean Gaussian distribution with the unknown covariance matrix Σ_k. The initial degree of freedom and symmetric positive definite matrix are $\kappa_{0,i} = 5(i = 1, 2)$, $\Lambda_0 = \text{diag}\{20, 20\}$. The turn rates ω_i, ω_j belong to the following set

$$\omega_i \in \left(-6.67 \times 10^{-4}, 0, 6.67 \times 10^{-4}\right) \quad \omega_i \in \mathcal{M}^b$$

$$\omega_j \in \pi \times 10^{-3} \times (\pm 4.2, \pm 5.6, \pm 8.3, \pm 16.7) \quad \omega_j \in \mathcal{M}^c$$

The transition probability matrix (TPM) of the IMM11 $\Pi^{11} = (\pi_{i,j}^{11})_{11 \times 11}$ is extended from the TPM of the basic models set. The TPM $\Pi_b^3 = (\pi_{i,j}^3)_{3 \times 3}$ is shown as

$$\Pi_b^3 = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \tag{49}$$

$$\pi_{i,j}^{11} = \begin{cases} \begin{cases} \pi_{i,j}^3 - a & i = j \\ \pi_{i,j}^3 & i \neq j \end{cases} & i \leq 3, j \leq 3 \\ a/8 & i \leq 3, j > 3 \\ 1 - 10b & i = j & i > 3 \\ b & i \neq j & i > 3 \end{cases} \tag{50}$$

where $i, j = 1, \ldots, 11$, and $a = b = 0.01$.

As can be seen from the RMSEs of position in Figs. 1 and 2. The RMSE of MSA-VB is lower than that of IMM3-VB. Note that the IMM3 and IMM11 perform a bit better than the IMM3-VB, IMM11-VB, and MSA-VB at the beginning ($t < 25$). The reason for this is VB eliminates

the error between initial noise variance and true noise variance by using the iterative computation. The RMSEs of velocity in Figs. 3 and 4. From these figures it can be observed that the RMSEs of velocity of the IMM11-VB, MSA-VB outperform that of the IMM3, IMM11, and IMM3-VB. During the maneuver, MSA also outperforms the other four algorithms.

In [29] and [23], the forgetting factor ρ is chosen empirically. The average RMSEs of the state versus different value of ρ are shown in Fig. 5. Simulation results show that the proposed algorithm performs better when $\rho = 0.92$.

The RMSEs of the five algorithms are given in Table 2 with different measurement noise variance parameters. The variance parameters are chosen as 5, 10, 20, and 50, respectively. From Table 2, with increasing levels of noise variance parameters, the proposed algorithm and the

Table 2 Average RMSEs in position with different measurement noise variance parameters for the Experiment 1

Methods	Measurement noise variance parameters			
	5	10	20	50
MSA-VB	12.5276	21.0012	34.4979	59.5699
IMM3-VB	14.7016	22.8626	39.6366	70.3219
IMM11-VB	13.3429	19.2559	36.3398	63.2457
IMM3	95.3203	96.4174	98.8328	107.2918
IMM11	47.6672	48.2740	52.5462	63.7677

Table 3 Relative computational times for the Experiment 1

IMM3	IMM11	IMM3-VB	IMM11-VB	MSA-VB
1	4.544	4.507	16.992	5.558

IMM11-VB have smaller RMSEs among these algorithms for all measurement noise variance parameters. Compared with the IMM3, IMM3-VB, IMM11, and IMM11-VB, the average RMSEs of the IMM3 and IMM3-VB are larger than that of the IMM11 and IMM11-VB. That is because more models are adapted for maneuvering target tracking. Due to VB effectively estimating the noise variance parameters, the average RMSEs of the IMM3-VB and IMM11-VB is smaller than that of the IMM3 and IMM11, respectively.

The comparison of relative computational times are shown in Table 3. The CPU time needed for IMM11-VB is three times the CPU time of MSA-VB.

5.2 Experiment 2

In this scenario, a bearings-only tracking (BOT) problem is considered. The target located at coordinate $[0, 0]$ *m*. The initial position of observation platform is [500 m, 20 m/s, 800 m, 10 m/s]. First, the observation platform moves in constant velocity (CV) model for 100 s, then moves in CT model with duration 50 s and the turn rate $\omega = -0.0232$(rad/s), and finally moves in CV model for 100 s. The measurement model at the time k is

$$\vartheta_k = \tan^{-1}\left(\frac{\phi_k^o - \phi_k^t}{\varphi_k^o - \varphi_k^t}\right) + v_k \tag{51}$$

The turn rates of the basic model set $\mathcal{M}^b$ and the candidate model set $\mathcal{M}^c$ are

$$\omega_i \in \pi \times 10^{-3} \times (-1, 0, 1) \quad \omega_i \in \mathcal{M}^b$$
$$\omega_j \in \pi \times 10^{-4} \times (\pm 0.5, \pm 0.667, \pm 1, \pm 2) \quad \omega_j \in \mathcal{M}^c$$

The TPMs are shown in Eqs. (49) and (50).

The RMSEs of X and Y position and velocity over 200 Monte Carlo runs are shown in Figs. 6, 7, 8 and 9, respectively. It can be observed that the proposed algorithm performs a bit better than the other four algorithms, and the RMSEs of IMM11 and IMM11-VB become larger around the time steps where the true measurement noise variances are jump changes. As can be seen from Fig. 7, the RMSE of IMM11 is shocked in the noise varying process. The reason is that the number of dimensions of measurement vector is lower than that of state vector. The result of variance estimation is shown in the Fig. 10. We can see that the proposed algorithm is effective in the estimation of the measurement noise statistics with some penalty of time delay. This is because the old noise value will be contained in part from the prior time step to the next time step. The RMSEs of the five algorithms are given in Table 4 with different measurement noise
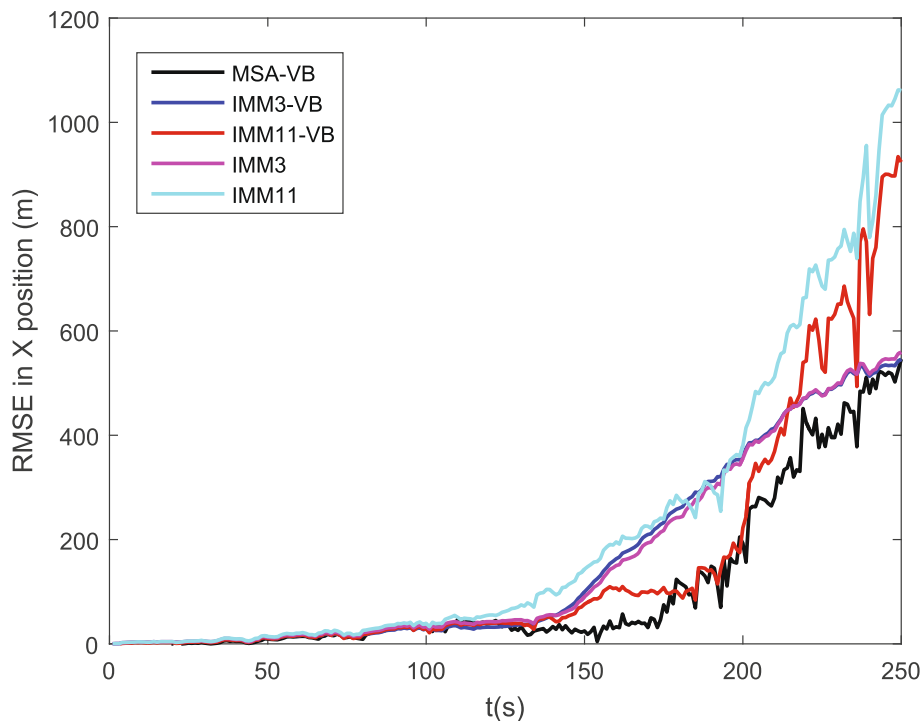

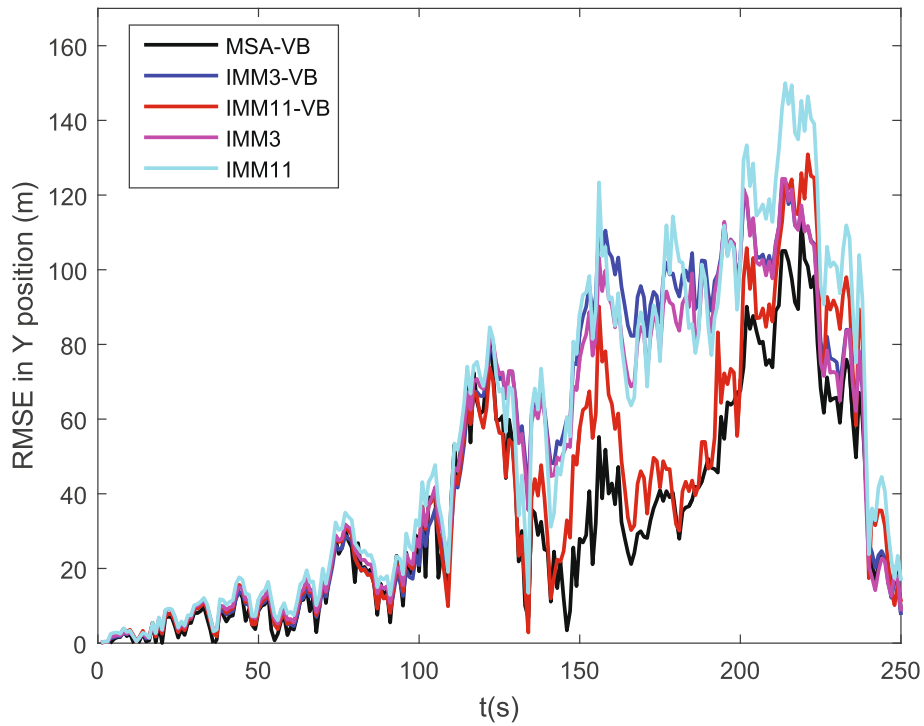**Fig. 6** RMSEs in *X* position of the proposed algorithm, IMM3, IMM11, IMM3-VB, and IMM11-VB in scenario 2

Fig. 7 RMSEs in *Y* position of the proposed algorithm, IMM3, IMM11, IMM3-VB, and IMM11-VB in scenario 2
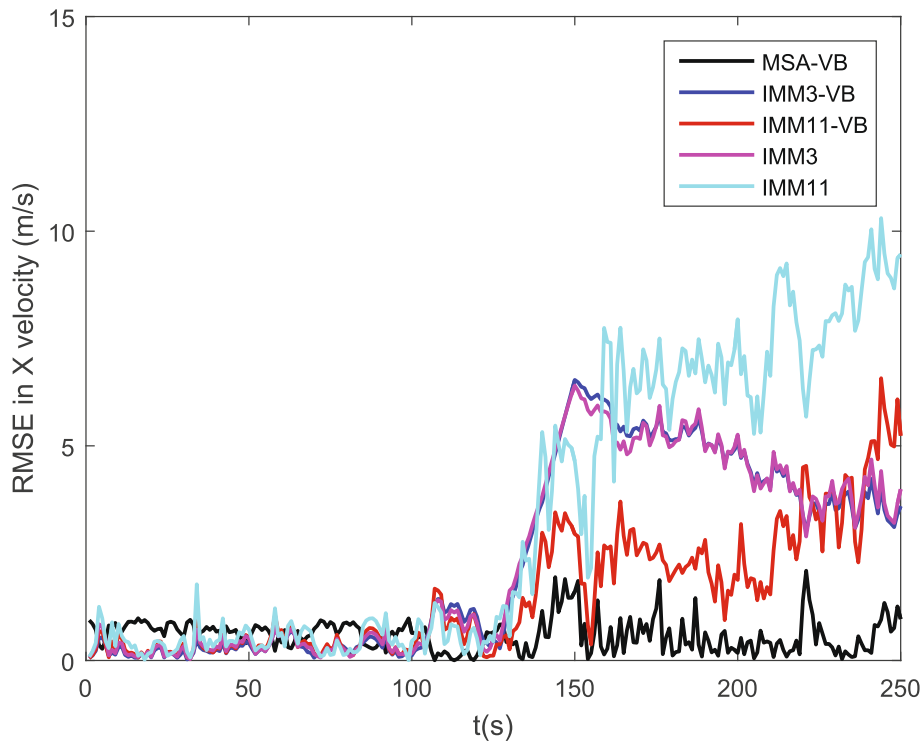

Fig. 8 RMSEs in *X* velocity of the proposed algorithm, IMM3, IMM11, IMM3-VB, and IMM11-VB in scenario 2
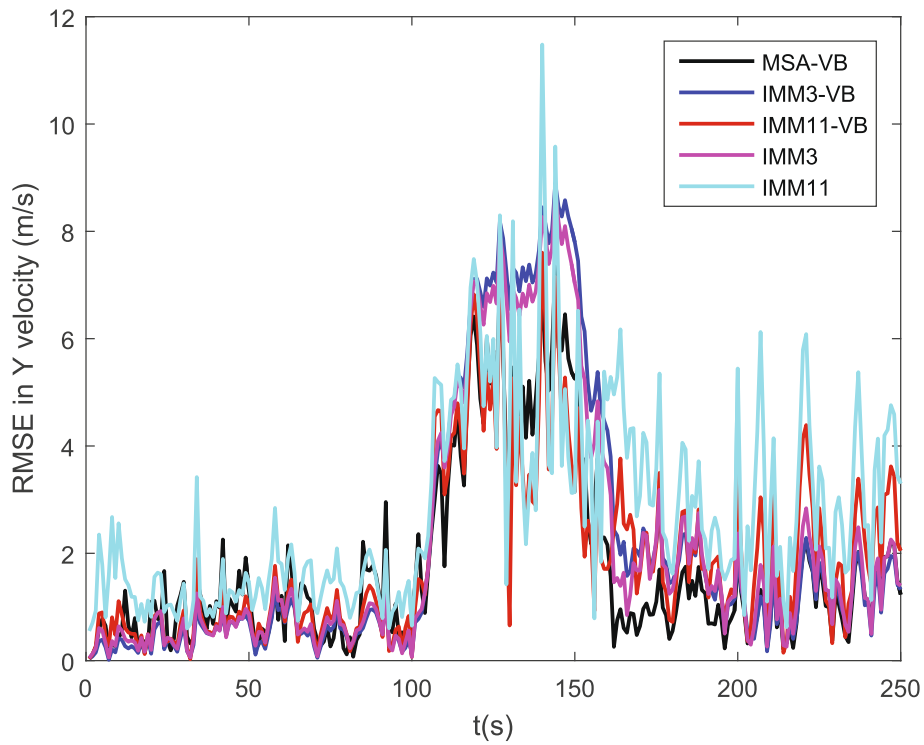
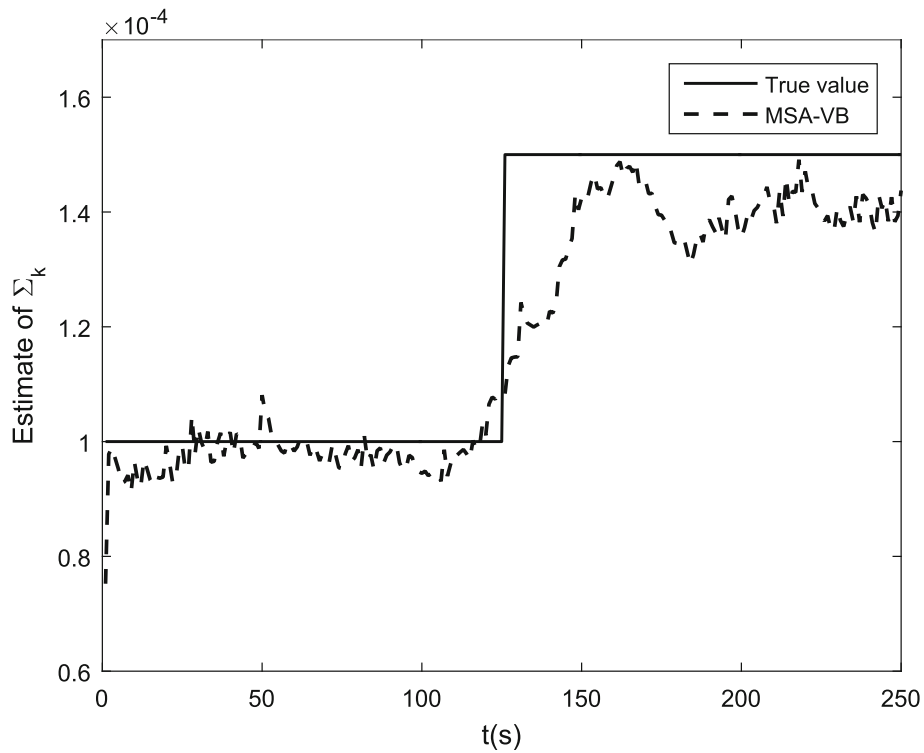**Fig. 9** RMSEs in *Y* velocity of the proposed algorithm, IMM3, IMM11, IMM3-VB, and IMM11-VB in scenario 2

Fig. 10 Estimation result of Σ_k

Table 4 Average RMSEs in position with different measurement noise variance parameters for the experiment 2

Methods	Measurement noise variance parameters			
	0.0001	0.0005	0.001	0.01
MSA-VB	77.8138	108.6454	236.6716	341.4275
IMM3-VB	135.4686	241.3172	288.0869	391.6982
IMM11-VB	126.4245	192.4053	272.2413	369.2875
IMM3	288.3541	290.9831	353.0082	421.0007
IMM11	350.9934	371.9373	424.6710	465.0060

variance parameters. The variance parameters are chosen as 0.0001, 0.0005, 0.001, and 0.01, respectively. It can be seen from Table 4 that the average RMSEs increase as the noise variance parameters increase. Compared to the IMM3, IMM3-VB, IMM11, and IMM11-VB, the proposed algorithm has higher tracking accuracy.

The comparison of relative computational times are shown in Table 5. The CPU time of the IMM3 denotes a unit. It is obvious that MSA-VB needs less computational time than IMM11-VB.

6 Discussion

We have shown that the proposed algorithm can mitigates the effects of system model and noise statistics uncertainties in the target track system. We found that the proposed algorithm can effectively estimate the target state, as demonstrated in the numerical simulations. The estimation accuracy was examined by comparing the proposed algorithm, the IMM-VB and IMM methods. The average RMSEs of the positions and velocities of the proposed algorithm are smaller than that of the other algorithms. Our results show an improved estimation and tracking performance compared to the IMM-VB and IMM methods. In addition, we found that the computational complexity of the proposed method is relatively higher than the IMM3-VB and IMM methods. The reason is that most of its computational time is spent on reconstructing the adapted model set and calculating the noise parameters by using the VB method. It should be noted that this study concentrates on only the single sensor target tracking with measurement noise uncertainty. Hence, we currently focus on extending it to multi-sensor target tracking with unknown measurement and process noise.

7 Conclusions

In this paper, we present an adaptively robust filter to address the performance degradation of the IMM-VB in

Table 5 Relative computational times for the experiment 2

IMM3	IMM11	IMM3-VB	IMM11-VB	MSA-VB
1	4.312	2.062	7.562	5.375

the presence of system model and noise statistics uncertainties. The main contribution of this paper is that a MSA method is designed to choose the best match model by calculating the divergence between the candidate models and true mode. Based on the chosen model, the model-conditioned estimation based on variational Bayesian approximation is proposed to estimate the system state and noise parameters. The performance of the MSA-VB is evaluated over the different target tracking scenes. The RMSE for the positions and velocities are presented which shows higher accuracy compared with the IMM-VB and IMM methods.

Author details
[1] School of Mechatronic Engineering, Xi'an Technological University, Xi'an, China. [2] Autonomous Systems and Intelligent Control International Joint Research Center, Xi'an Technological University, Xi'an, China. [3] School of Electronic Information Engineering, Xi'an Technological University, Xi'an, China.

References

1. R. Dehghannasiri, X. Qian, E. R. Dougherty, A Bayesian robust Kalman smoothing framework for state-space models with uncertain noise statistics. EURASIP J. Adv. Signal Process. **1**, 55 (2018). https://doi.org/10.1186/s13634-018-0577-1

2. L. W. Jochumsen, J. Østergaard, S. H. Jensen, A recursive kinematic random forest and alpha beta filter classifier for 2D radar tracks. EURASIP J. Adv. Signal Process. **1**, 82 (2015). https://doi.org/10.1186/s13634-016-0378-3

3. X. Lan, S. Zhang, P. C. Yuen, Learning common and feature-specific patterns: A novel multiple-sparse-representation-based tracker. IEEE Trans. Image Process. **27**(4), 2022–2037 (2018). https://doi.org/10.1109/TIP.2017.2777183

4. X. Lan, M. Ye, R. Shao, Learning modality-consistency feature templates: A robust RGB-infrared tracking system. IEEE Trans. Ind. Electron. **66**(12), 9887–9897 (2019). https://doi.org/0.1109/TIE.2019.2898618

5. J. Zhao, L. Mili, A framework for robust hybrid state estimation with unknown measurement noise statistics. IEEE Trans. Ind. Inform. **99**, 1–1 (2017). https://doi.org/10.1109/TII.2017.2764800

6. G. Y. Kulikov, M. V. Kulikov, The accurate continuous-discrete extended Kalman filter for radar tracking. IEEE Trans. Signal Process. **64**(4), 948–958 (2016). https://doi.org/10.1109/TSP.2015.2493985

7. X. Li, P. Willett, M. Baum, PMHT approach for underwater bearing only multisensor multitarget tracking in clutter. IEEE J. Ocean. Eng. **41**(4), 831–839 (2016). https://doi.org/10.1109/joe.2015.2506220

8. A. Roy, D. Mitra, Multi-target trackers using cubature Kalman filter for Doppler radar tracking in clutter. IET Signal Process. **10**(8), 888–901 (2016). https://doi.org/10.1049/iet-spr.2015.0540

9. J. Xiao, R. Stolkin, M. Oussalah, Continuously adaptive data fusion and model relearning for particle filter tracking with multiple features. IEEE Sensors J. **16**(8), 2639–2649 (2016). https://doi.org/10.1109/JSEN.2016.2514704

10. M. A. Alirezapouri, H. Khaloozadeh, A. Vali, Set value-based dynamic model development for non-linear manoeuvring target tracking problem in the presence of unknown but bounded disturbances. IET Radar Sonar Navig. **12**(2), 186–194 (2018). https://doi.org/10.1049/iet-rsn.2017.0293

11. Y. Liu, Y. Liang, Z. Liu, A novel robust MM filter against outliers. Asian J. Control. **18**(1), 340–349 (2016). https://doi.org/10.1002/asjc.1042

12. X. R. Li, Y. BarShalom, Multiple-model estimation with variable structure. IEEE Trans. Autom. Control. **41**(4), 478–493 (1996). https://doi.org/10.1109/9.489270

13. X. R. Li, V. P. Jilkov, R. Jifeng, Multiple-model estimation with variable structure-part VI: Expected-mode augmentation. IEEE Trans. Aerosp. Electron. Syst. **41**(3), 853–867 (2005). https://doi.org/10.1109/taes.2005.1541435

14. X. R. Li, Y. M. Zhang, Multiple-model estimation with variable structure. V. Likely-model set algorithm. IEEE Trans. Aerosp. Electron. Syst. **36**(2), 448–466 (2000). https://doi.org/0.1109/7.845222

15. X. Wang, S. Challa, R. Evans, X. R. Li, Minimal submodel-set algorithm for maneuvering target tracking. IEEE Trans. Aerosp. Electron. Syst. **39**(4), 1218–1231 (2003). https://doi.org/10.1109/TAES.2003.1261123

16. J. Lan, X. R. Li, Best model augmentation for variable-structure multiple-model estimation. IEEE Trans. Aerosp. Electron. Syst. **47**(3), 2008–2025 (2011). https://doi.org/10.1109/TAES.2011.5937279

17. S. Sarkka, A. Nummenmaa, Recursive noise adaptive Kalman filtering by variational Bayesian approximations. IEEE Trans. Autom. Control. **54**(3), 596–600 (2009). https://doi.org/10.1109/TAC.2008.2008348

18. J. Lerga, V. Sucic, D. Sersic, in *Proc. International International Symposium on Image and Signal Processing and Analysis*. Performance analysis of the LPA-RICI denoising method, (2009), pp. 28–33. https://doi.org/10.1109/ISPA.2009.5297758

19. G. Segon, J. Lerga, V. Sucic, Improved LPA-ICI-based estimators embedded in a signal denoising virtual instrument. Signal Image Video Process. **11**(2), 211–218 (2017). https://doi.org/10.1007/s11760-016-0921-6

20. I. Volaric, J. Lerga, V. Sucic, A fast signal denoising algorithm based on the LPA-ICI method for real-time applications. Circ. Syst. Signal Process. **36**(11), 4653–4669 (2017). https://doi.org/10.1007/s00034-017-0538-1

21. K. Li, L. Chang, B. Hu, A variational Bayesian-based unscented Kalman filter with both adaptivity and robustness. IEEE Sensors Journal. **16**(18), 6966–6976 (2016). https://doi.org/10.1109/JSEN.2016.2591260

22. Z. Miao, H. Shi, Y. Zhang, Neural network-aided variational Bayesian adaptive cubature Kalman filtering for nonlinear state estimation. Measurement Science & Technology. **28**(10), 6–18 (2017). https://doi.org/10.1088/1361-6501/aa7d70

23. H. Zhu, H. Leung, Z. S. He, A variational Bayesian approach to robust sensor fusion based on Student-*t* distribution. Inf. Sci. **221**(1), 201–214 (2013). https://doi.org/10.1016/j.ins.2012.09.017

24. H. Nurminen, T. Ardeshiri, R. Piché, Skew t filter and smoother with improved covariance matrix approximation. Inf. Sci. **221**(1), 201–214 (2016). https://doi.org/10.1109/TSP.2018.2865434

25. H. Zhu, H. Leung, Z. S. He, State estimation in unknown non-Gaussian measurement noise using variational Bayesian technique. IEEE Trans. Aerosp. Electron. Syst. **49**(4), 2601–2614 (2013). https://doi.org/10.1109/TAES.2013.6621839

26. W. Li, Y. Jia, State estimation for jump Markov linear systems by variational Bayesian approximation. IET Control Theory Appl. **6**(6), 319–326 (2012). https://doi.org/10.1049/iet-cta.2011.0167

27. X. Hong, H. Yuan, W. C. Xie, Variational Bayesian-interacting multiple model tracking filter with angle glint noise. J. Electron. Inf. Technol. **40**(7), 1583–1590 (2018)

28. B. Gao, S. S. Gao, Y. M. Zhong, Interacting multiple model estimation-based adaptive robust unscented Kalman filter. J. Control Autom. Syst. **13**(1), 1–13 (2013). https://doi.org/10.1007/s12555-016-0589-2

29. M. J. Beal, Variational algorithms for approximate Bayesian inference. Ph.D. dissertation, The Gatsby Computational Neuroscience Unit, University College London, London, England (2003)

30. P. Dong, Z. Leung, H. Leung, Variational Bayesian adaptive cubature information filter based on Wishart distribution. IEEE Trans. Autom. Control. **62**(11), 6051–6057 (2017). https://doi.org/10.1109/TAC.2017.2704442

31. C. Shen, D. J. Xu, F. Shen, Generalized noises adaptive Kalman filtering based on variational inference. XI Tong Gong Cheng Yu Dian Zi Ji Shu/Syst. Eng. Electron. **36**(8), 1466–1472 (2014). https://doi.org/10.3969/j.issn.1001-506X.2014.08.03

32. B. Ma, Parametric and nonparametric approaches for multisensor data fusion. Ph.D. dissertation, Dept. Elect. Eng., UMich., Detroit, MI, USA (2003). university of London London

33. J. M. Aughenbaugh, B. R. La Cour, Metric selection for information theoretic sensor management. Proc. Int. Conf. Inf. Fusion. **13**(1), 1–13 (2015). https://doi.org/10.1109/ICIF.2008.4632451

34. V. Sucic, N. Saulig, B. Boashash, Estimating the number of components of a multicomponent nonstationary signal using the short-term time-frequency Rényi entropy. EURASIP J. Adv. Signal Process. **2011**(1), 125–125 (2011). https://doi.org/10.1186/1687-6180-2011-125

35. J. Lerga, N. Saulig, V. Mozetic, in *in Proc. International Multidisciplinary Conference on Computer and Energy Science*. Number of EEG signal components estimated using the short-term Rényi entropy, (Croatia, 2016), pp. 1–6. https://doi.org/10.1109/SpliTech.2016.7555940

36. C. Kreucher, K. Kastella, H. O. Alfred, Sensor management using an active sensing approach. Signal Process. **85**(3), 607–624 (2005). https://doi.org/10.1016/j.sigpro.2004.11.004

37. C. Shen, D. Xu, W. Huang, F. Shen, An interacting multiple model approach for state estimation with non-Gaussian noise using a variational Bayesian method. Asian J. Control. **17**(4), 1424–1434 (2014). https://doi.org/10.1002/asjc.1055

38. J. Zhao, L. Mili, A framework for robust hybrid state estimation with unknown measurement noise statistics. IEEE Trans. Ind. Inform. **99**, 1–1 (2017). https://doi.org/10.1109/TII.2017.2764800

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.