2

MMSE-Based Filtering for Linear and Nonlinear Systems in the Presence of Non-Gaussian System and Measurement Noise

I. Bilik[1] and J. Tabrikian[2]
[1]*Dept. of Electrical and Computer Engineering, University of Massachusetts, Dartmouth, MA,*
[2]*Dept. of Electrical and Computer Engineering, Ben-Gurion University of the Negev,*
[1]*USA*
[2]*Israel*

1. Introduction

This chapter addresses the problems of minimum mean square error (MMSE) estimation in non-Gaussian linear and nonlinear systems. In many scientific and practical problems (such as control, astronomy, economic data analysis, communication and radar surveillance), estimation of time-varying system state using a sequence of noisy measurements is performed using the dynamic state-space (DSS) modeling approach. In the DSS approach, the time-varying dynamics of an unobserved state are characterized by the state vector. In most problems, the Bayesian approach can be efficiently used for system state estimation. The posterior probability density function (PDF), which contains the complete statistical information for the system state estimation, can be used for optimal (in any sense) state estimation [1]. Unfortunately, many practical applications, such as target tracking in radar systems are nonlinear and non-Gaussian. Thus, in maneuvering target tracking applications, a heavy-tailed distribution is usually used to model the abrupt changes of the system state due to target maneuver [2]. In addition, changes in the target aspect toward the radar may cause irregular electromagnetic wave reections, resulting significant variations of radar reections [3]. This phenomenon gives rise to outliers in angle tracking, and it is referred to as target glint [4]. It was found that glint has a long-tailed PDF [3], [5], and its distribution can be modeled by mixture of a zero-mean, small-variance Gaussian and a heavy-tailed Laplacian [6]. The Gaussian mixture model (GMM) with two mixture components is widely used in the literature for abrupt changes of the system state and glint noise modeling [3], [7]. This model consists of one small variance Gaussian with high probability and one large variance Gaussian with low probability of occurrence. The nonlinearity behavior in target tracking systems is due to the fact that the target dynamics are usually modeled in Cartesian

Reprinted, with permission, from IEEE Trans. Aerospace and Electronic Systems, references [39] and [45]

Source: Kalman Filter: Recent Advances and Applications, Book edited by: Victor M. Moreno and Alberto Pigazo, ISBN 978-953-307-000-1, pp. 584, April 2009, I-Tech, Vienna, Austria

coordinates, while the observation model is in polar coordinates. There is no general analytic expression for the posterior PDF in nonlinear problems and only suboptimal estimation algorithms have been studied [1]. The extended Kalman filter (EKF) is the most popular approach for recursive nonlinear estimation [8], [9]. The main idea of the EKF is based on a first-order linearization of the model where the posterior PDF and the system and measurement noises are assumed to be Gaussian. The nonlinearity of the measurement model leads to non-Gaussian, multi-modal PDF of the system state, even when the system and the measurement noises are Gaussian. The Gaussian approximation of this multi-modal distribution leads to poor tracking performance. The unscented Kalman filter (UKF) approximates the PDF at the output of the nonlinear transformation using deterministic sampling [10]-[11]. The advantage of the UKF over the EKF stems from the fact that it does not involve approximation of the nonlinear model per se [12], [13]. The UKF provides an unbiased estimate, however its convergence is slow [13]. Many researchers addressed the problem of filtering in non-Gaussian models. One of the effective algorithms in the non-Gaussian problems is the Masreliez filter [14], [15] that employs a nonlinear "score-function", calculated from known *a-priori* noise statistics. The score-function is customized for the noise statistics and has to be redesigned for each application. The main disadvantage of this approach is that it involves a computationally expensive score function calculation [6]. In [16], the Masreliez filter was used in the target tracking problem with glint noise. Recently, a few new filtering approaches have been proposed for the problem of target tracking. One of them is the multiple modeling (MM) approach, in which the time-varying motion of the maneuvering target is described by multiple models [17]. In this approach, the non-Gaussian system is represented by a mixture of parallel Gaussian-distributed modes [8]. Using the Bayesian framework, the posterior PDF of the system state is obtained as a mixture of conditional estimates with *a-priori* probabilities of each mode [18]. Various filters are used for mode-conditioned state estimation. For example, the Gaussian sum filter (GSF), was implemented in [8], [19] using a bank of KFs. The EKF and Masreliez filters were used as mode-conditioned filters for the nonlinear problems of target tracking in [6], [16], [20]. The main drawback of the MM approach is the exponential growth of the number of the modes, and exponentially increasing number of mode-conditioned filters [18], [21]. Therefore, optimal algorithms, such as the GSF, are impractical. The direct approximation methods for target tracking in the presence of clutter with GMM distribution approximation were proposed in [22]-[28]. The joint probabilistic data association (JPDA) [18] and global nearest neighbor (GNN) [25] approximate the entire GMM by a single Gaussian, loosing important information contained in other mixture components. The multiple hypothesis tracking (MHT) [26] and mixture reduction (MR) methods [22], employ *ad-hoc* joining and clustering preserving mean and covariance of the original distribution. The direct approximation algorithms are generally computationally efficient, however, they are suboptimal due to the *ad-hoc* mixture approximation methods and lead to degraded target tracking performance. A suboptimal, computationally-efficient interacting MM (IMM) algorithm was successfully applied to the maneuvering target tracking problem [18], [29], [30]. In [7], [16], [20] the IMM algorithm with EKFs and Masreliez filters were implemented for maneuvering target tracking in the presence of glint noise. The IMM algorithm with greater number of modes was proposed in [31] for non-Gaussian system and measurement

noise. In the recent decade, a new class of filtering methods has been proposed based on the sequential Monte Carlo (MC) approach. The sequential importance sampling technique forms the basis for most MC techniques [32]. In these techniques, the filtering is performed recursively generating MC samples of the state variables. These methods are often very exible in non-Gaussian problems due to the nature of the MC simulations [33]. One of the popular techniques of this approach is the PF, which is a suboptimal estimator that approximates the posterior distribution by a set of random samples with associated weights. The PF models the posterior distribution using discrete random samples rather than using an analytic model [34]. The Gaussian sum particle filter (GSPF) [35] implements the PF assuming Gaussian mixture distributions for the system and measurement noises. The GSPF generalizes the GSF introducing a new model order reduction method. Thus, the model order of the system state PDF remains constant over iterations, discarding mixands with small weights. The PF has been extensively used for maneuvering target tracking (e.g. [2]). In [36], the PF was applied to the problem of tracking in glint noise environment. As it was shown in [37] and [38], the PF outperforms the IMM algorithm when the likelihood function is multi-modal. Different application-driven PFs are presented in the literature, but there is no precise rule, which type of PF should be used in each application. This implies that no rigorous PF exists, which is one of the disadvantages of the PF approach. In this chapter, two recursive methods, based on the MMSE estimator of the GMM distributed state vector, are presented. The first is Gaussian mixture Kalman filter (GMKF), derived in [39], [40], for a linear model with non-Gaussian system and measurement noise. This algorithm relies on the fact that any PDF can be closely approximated by a mixture of finite number of Gaussians [41]. A greedy EM-based model order reduction method for the problem of exponential model order growth, discussed in [18], [21] and [19], is derived. The greedy learning algorithm controls the GMM order of the system state PDF, which might vary over the iterations, but remains finite. The EM-based model order reduction method is optimal with respect to the Kullback-Leibler divergence (KLD), that is, it minimizes the KLD of the reduced-order estimated PDF from the "true" PDF of the system state. The GMKF addresses a general estimation problem with non-Gaussian system and measurement noise, modeled by the GMM. This problem is of a great practical interest, for example, in maneuvering target tracking in the presence of glint noise [31]. The second recursive algorithm, named as nonlinear GMKF (NL-GMKF), extends the GMKF to nonlinear models [44]-[46]. The NL-GMKF considers the case of non-Gaussian system and measurement noises as well as non-Gaussian posterior PDF of the state vector. The expected significance of the NL-GMKF is in practical applications of low to moderate maneuvering target tracking, when maneuver detection is difficult. The advantage of the NL-GMKF over other tracking algorithms is significant especially in the presence of glint measurement noise with small probability of detection and high significance. The correlation between the statistics of glint noise and maneuver (that characterizes a maneuvering target consisting of multiple scattering centers) makes the problem of maneuvering target tracking in the presence of glint noise extremely challenging, due to the difficulty of maneuver and glint detection and filtering simultaneously. The NL-GMKF does not require prior knowledge of the target dynamics such as coordinated turn model, therefore, it might be useful when tracking targets with a complicated maneuvering profile that cannot be modeled by a finite set of simple dynamic models.

2. DSS model

A. Linear model

Consider a state sequence {**s**[n], n = 0, 1, 2, ...} and observations {**x**[n], n = 0, 1, 2, ...} whose time evolution and observation equations are described by the following linear non-Gaussian model

$$\mathbf{s}[n] = \mathbf{A}[n]\mathbf{s}[n-1] + \mathbf{u}[n] , \tag{1}$$

$$\mathbf{x}[n] = \mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] , \tag{2}$$

where the state transition matrices, **A**[n], and the observation matrices, **H**[n], are known. The initial state **s**[–1], the zero-mean driving noise **u**[n], and the zero-mean measurement noise **w**[n] are independent with the following distributions

$$\mathbf{s}[-1] \sim GMM(\alpha_{\mathbf{s}l}[-1], \boldsymbol{\mu}_{\mathbf{s}l}[-1], \boldsymbol{\Gamma}_{\mathbf{s}l}[-1]; l = 1, \ldots, L) , \tag{3}$$

$$\mathbf{u}[n] \sim GMM(\alpha_{\mathbf{u}k}[n], \boldsymbol{\mu}_{\mathbf{u}k}[n], \boldsymbol{\Gamma}_{\mathbf{u}k}[n]; k = 1, \ldots, K) , \tag{4}$$

$$\mathbf{w}[n] \sim GMM(\alpha_{\mathbf{w}m}[n], \boldsymbol{\mu}_{\mathbf{w}m}[n], \boldsymbol{\Gamma}_{\mathbf{w}m}[n]; m = 1, \ldots, M) , \tag{5}$$

where $GMM(\alpha_j, \boldsymbol{\mu}_j, \boldsymbol{\Gamma}_j, j = 1,..., J)$ denotes a Jth-order proper complex Gaussian mixture distribution with weights, $\{\alpha_j\}_{j=1}^{J}$, mean vectors, $\{\boldsymbol{\mu}_j\}_{j=1}^{J}$, and covariance matrices, $\{\boldsymbol{\Gamma}_j\}_{j=1}^{J}$. The driving noise **u**[n] and the measurement noise **w**[n] are temporally independent, i.e. **u**[n] and **u**[n'], and **w**[n] and **w**[n'] are mutually independent for any time instances n = 0, 1, 2,..., n' = 0, 1, 2, ...; $n \neq n'$. The PDF of a GMM-distributed random vector $\mathbf{y} \sim GMM(\alpha_{\mathbf{y}j}, \boldsymbol{\mu}_{\mathbf{y}j}, \boldsymbol{\Gamma}_{\mathbf{y}j}; j = 1,..., J)$ is given by

$$f_{\mathbf{y}}(\mathbf{y}) = \sum_{j=1}^{J} \alpha_{\mathbf{y}j} \Phi(\mathbf{y}; \boldsymbol{\theta}_j) , \tag{6}$$

where $\Phi(\mathbf{y}; \boldsymbol{\theta}_j)$ is a Gaussian PDF and $\boldsymbol{\theta}_j$ contains the mean vector, $\boldsymbol{\mu}_{\mathbf{y}j}$ and the covariance matrix, $\boldsymbol{\Gamma}_{\mathbf{y}j}$. In the following, we will use the term Gaussian for proper Gaussian distributions.

B. Non-linear model

The nonlinear and non-Gaussian DSS model is:

$$\mathbf{s}[n] = \mathbf{a}(\mathbf{s}[n-1], \mathbf{u}[n]) , \tag{7}$$

$$\mathbf{x}[n] = \mathbf{h}(\mathbf{s}[n], \mathbf{w}[n]) , \tag{8}$$

where the nonlinear transition function, $\mathbf{a}(\cdot,\cdot)$, and the observation function, $\mathbf{h}(\cdot,\cdot)$, are assumed to be known. The system and measurement noise are non-Gaussian with known PDFs. The driving noise, **u**[n], and the measurement noise, **w**[n], are temporally independent, i.e. **u**[n] and **u**[n'], and **w**[n] and **w**[n'] are mutually independent for any time instances n = 0, 1, 2,...; n' = 0, 1, 2,... ; $n \neq n'$. The initial state, **s**[–1], the driving noise, **u**[n], and the measurement noise, **w**[n], are independent. The initial state distribution is modeled in (3).

3. MMSE-based filters

Let $\hat{\mathbf{s}}[n|p]$ denote the MMSE estimator of $\mathbf{s}[n]$ from $\mathcal{X}[p]$ where $\mathcal{X}[p] \triangleq (\mathbf{x}^T[0], \mathbf{x}^T[1], ..., \mathbf{x}^T[p])^T$. The notation $\hat{\mathbf{s}}[n|n-1]$ stands for one-step prediction of the state vector $\mathbf{s}[n]$ from data $\mathcal{X}[n-1]$. The objective of this section is to derive recursive methods for estimation of $\mathbf{s}[n]$ from the observed data $\mathcal{X}[n]$ for the linear and nonlinear non-Gaussian models. To this end, the MMSE criterion resulting in the conditional mean estimator

$$\hat{\mathbf{s}}[n|n] \stackrel{\triangle}{=} E[\mathbf{s}[n]|\mathcal{X}[n]] \,, \tag{9}$$

is employed.

A. GMKF

For the linear DSS model, stated in Section IIA, the MMSE estimator of the state vector $\mathbf{s}[n]$ from the measurements $\mathcal{X}[n]$ can be implemented by the following recursive algorithm, named as GSF.

GSF Theorem

1. Initialization:

Initialize the GMM parameters of the state vector at time instance $n = -1$ for $l = 1,..., L$.

$$\alpha_{\mathbf{s}}[-1|-1, \eta_l[-1]] = \alpha_{\mathbf{s}l}[-1] \,, \tag{10}$$

$$\boldsymbol{\mu}_{\mathbf{s}}[-1|-1, \eta_l[-1]] = \boldsymbol{\mu}_{\mathbf{s}l}[-1] \,, \tag{11}$$

$$\boldsymbol{\Gamma}_{\mathbf{s}}[-1|-1, \eta_l[-1]] = \boldsymbol{\Gamma}_{\mathbf{s}l}[-1] \,, \tag{12}$$

where $\eta_{sl}[n]$, $l = 1,..., L$ is the random mixture indicator [47]-[48].

Set $n = 0$.

2. Prediction

2a. Predicted state PDF parameters:

$$\boldsymbol{\theta}[n|n-1] = \{\boldsymbol{\alpha}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]], \boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]], \boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]]\}_{(l,k)=(1,1)}^{(L,K)}$$

in which

$$\alpha_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]] = \alpha_{\mathbf{s}}[n-1|n-1, \eta_l[n-1]]\alpha_{\mathbf{u}k}[n] \,, \tag{13}$$

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]] = \mathbf{A}[n]\boldsymbol{\mu}_{\mathbf{s}}[n-1|n-1, \eta_l[n-1]] + \boldsymbol{\mu}_{\mathbf{u}k}[n] \,, \tag{14}$$

$$\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]] = \mathbf{A}[n]\boldsymbol{\Gamma}_{\mathbf{s}}[n-1|n-1, \eta_l[n-1]]\mathbf{A}^H[n] + \boldsymbol{\Gamma}_{\mathbf{u}k}[n] \,. \tag{15}$$

The PDFs of the state and system noise are modeled by GMM of order L and K, respectively, and therefore in the prediction stage, the number of mixture components grows to LK. The random indicator for this mixture is denoted as $\tilde{\eta}_{lk}[n]$, $l = 1,..., L$; $k = 1,...,K$.

2b. Prediction of the state vector:

$$\hat{\mathbf{s}}[n|n-1] = \sum_{l=1}^{L}\sum_{k=1}^{K} \alpha_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]]\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]] \,. \tag{16}$$

2c. Prediction of the measurements vector and innovation calculation:
The MMSE prediction of $\mathbf{x}[n]$ from $\mathcal{X}[n-1]$ is

$$\hat{\mathbf{x}}[n|n-1] \triangleq E(\mathbf{x}[n]|\mathcal{X}[n-1]) = \mathbf{A}[n]\hat{\mathbf{s}}[n|n-1] \ . \tag{17}$$

Define the innovation process $\tilde{\mathbf{x}}[n]$ as

$$\tilde{\mathbf{x}}[n] \triangleq \mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1] \ .$$

The mixture parameters of the conditional PDF $\tilde{\mathbf{x}}[n] \,|\, \mathcal{X}[n-1]$ are defined as follows: $\boldsymbol{\theta}_{lkm}[n] = \{\boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n \,|\, n-1,\ \bar{\eta}_{lkm}[n]],\ \boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n \,|\, n-1,\ \bar{\eta}_{lkm}[n]]\}$ and

$$\alpha_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]] \ = \ \alpha_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]] \cdot \alpha_{\mathbf{w}m}[n] \ , \tag{18}$$

$$\boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]] \ = \ \mathbf{H}[n]\left(\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]] - \hat{\mathbf{s}}[n|n-1]\right) + \boldsymbol{\mu}_{\mathbf{w}m}[n] \ , \tag{19}$$

$$\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]] \ = \ \mathbf{H}[n]\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]]\mathbf{H}^H[n] + \boldsymbol{\Gamma}_{\mathbf{w}m}[n] \ , \tag{20}$$

3. Kalman gain:

$$\begin{aligned}\mathbf{K}_{lkm}[n] \ = \ & \left(\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1, \bar{\eta}_{lkm}[n]]\mathbf{H}^H[n] + \left(\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \bar{\eta}_{lkm}[n]] - \hat{\mathbf{s}}[n|n-1]\right)\boldsymbol{\mu}_{\mathbf{w}m}[n]\right) \\ & \cdot \left(\mathbf{H}[n]\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1, \bar{\eta}_{lkm}[n]]\mathbf{H}^H[n] + \boldsymbol{\Gamma}_{\mathbf{w}m}[n]\right)^{-1} \ .\end{aligned} \tag{21}$$

The measurement noise PDF is modeled by GMM of order M, and therefore at the estimation stage the number of mixture components grows to LKM. The random indicator for this mixture is denoted as $\bar{\eta}_{lkm}[n]$, $l = 1,..., L$; $k = 1,..., K$; $m = 1,..., M$.

4. Estimation

4a. Estimated state mixture parameters:

$$\bar{\boldsymbol{\theta}}[n|n] \ \triangleq \ \{\alpha_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]], \boldsymbol{\mu}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]], \boldsymbol{\Gamma}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]]\}_{(l,k,m)=(1,1,1)}^{(L,K,M)} \tag{22}$$

where

$$\alpha_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]] \ = \ \frac{\alpha_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]]\Phi(\tilde{\mathbf{x}}[n]; \boldsymbol{\theta}_{lkm}[n])}{\sum_{l'=1}^{L}\sum_{k'=1}^{K}\sum_{m'=1}^{M}\alpha_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{l'k'm'}[n]]\Phi(\tilde{\mathbf{x}}[n]; \boldsymbol{\theta}_{l'k'm'}[n])} \ , \tag{23}$$

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]] \ = \ \boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \bar{\eta}_{lkm}[n]] + \mathbf{K}_{lkm}[n]\left(\tilde{\mathbf{x}}[n] - \boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]]\right) \ , \tag{24}$$

$$\begin{aligned}\boldsymbol{\Gamma}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]] \ = \ & \left(\mathbf{I} - \mathbf{K}_{lkm}[n]\mathbf{H}[n]\right)\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]] + \mathbf{K}_{lkm}[n]\left(\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \tilde{\eta}_{lk}[n]] - \hat{\mathbf{s}}[n|n-1]\right)\boldsymbol{\mu}_{\mathbf{w}m}[n] \ , \\ & \forall l = 1, \ldots, L; \ k = 1, \ldots, K; \ m = 1, \ldots, M \ ,\end{aligned} \tag{25}$$

4b. Estimation of the state vector:

$$\hat{\mathbf{s}}[n|n] = \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{m=1}^{M}\alpha_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]]\boldsymbol{\mu}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]] \ , \tag{26}$$

5. Set $n \rightarrow n+1$, go to step 2.

Note that according to this theorem, the model order grows during the prediction (stage 2) due to the non-Gaussian system noise, and during the estimation (stage 4) due to the non-Gaussian measurement noise. The GMKF algorithm consists of the GSF followed by a model order reduction algorithm implemented after the estimation stage:

$$\begin{aligned}\boldsymbol{\theta}[n|n] &\triangleq \{\alpha_{\mathbf{s}}[n|n,\eta_l[n]],\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]],\boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]]\}_{l=1}^{L} \\ &= EM_{LKM\rightarrow L}\left(\{\alpha_{\mathbf{s}}[n|n,\bar{\eta}_{lkm}[n]],\boldsymbol{\mu}_{\mathbf{s}}[n|n,\bar{\eta}_{lkm}[n]],\boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\bar{\eta}_{lkm}[n]]\}_{(l,k,m)=(1,1,1)}^{(L,K,M)}\right) .\end{aligned} \tag{27}$$

The operator $EM_{LKM\rightarrow L}$ means that the parameters of an L-order GMM are estimated via synthetic data generation according to the GMM PDF with LKM components. Note that in general, the GMM order of the posterior state PDF can vary from iteration to iteration. The GMKF stages that are similar to the KF (see e.g. [49]) are schematically presented in Fig. 1.

Proof: In the following, an algorithm for obtaining $\hat{\mathbf{s}}[n|n]$ is developed from $\hat{\mathbf{s}}[n-1|n-1]$. Let the MMSE prediction of $\mathbf{x}[n]$ from $\mathcal{X}[n-1]$ be defined as

$$\hat{\mathbf{x}}[n|n-1] \triangleq E(\mathbf{x}[n]|\mathcal{X}[n-1]) , \tag{28}$$

and the innovation process be

$$\tilde{\mathbf{x}}[n] \triangleq \mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1] . \tag{29}$$

If the transformation $\mathcal{X}[n] \leftrightarrow [\mathcal{X}^T[n-1], \tilde{\mathbf{x}}^T[n]]^T$ is one-to-one, then the conditional distribution of $\mathbf{s}[n]|\mathcal{X}[n]$ is identical to the conditional distribution of $\mathbf{s}[n]|\mathcal{X}[n-1], \tilde{\mathbf{x}}[n]$. The vector $\hat{\mathbf{x}}[n|n-1]$ is calculated using (2) and the statistical independence between $\mathcal{X}[n-1]$ and the zero-mean measurement noise $\mathbf{w}[n]$, as follows:

$$\hat{\mathbf{x}}[n|n-1] = E(\mathbf{x}[n]|\mathcal{X}[n-1]) = E(\mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n]|\mathcal{X}[n-1]) = \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1] ,$$

where $\hat{\mathbf{s}}[n|n-1]$ is the prediction of the state vector at time instance n, from data $\mathcal{X}[n-1]$, obtained by using (1):

$$\hat{\mathbf{s}}[n|n-1] = E(\mathbf{s}[n]|\mathcal{X}[n-1]) = \mathbf{A}[n]\hat{\mathbf{s}}[n-1|n-1] , \tag{30}$$

where $\hat{\mathbf{s}}[n-1|n-1]$ is the conditional mean estimator of the state vector $\mathbf{s}[n-1]$ from data $\mathcal{X}[n-1]$. The innovation process in (29) can be expressed using (1), (2) and (30), as

$$\begin{aligned}\tilde{\mathbf{x}}[n] &= \mathbf{H}[n](\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1]) + \mathbf{w}[n] \\ &= \mathbf{H}[n]\left(\mathbf{A}[n]\left(\mathbf{s}[n-1] - \hat{\mathbf{s}}[n-1|n-1]\right) + \mathbf{u}[n]\right) + \mathbf{w}[n] .\end{aligned} \tag{31}$$

Using (1), (30) and (31) it can be obtained that

$$\mathbf{y}[n] \triangleq \begin{bmatrix} \mathbf{s}[n] \\ \tilde{\mathbf{x}}[n] \end{bmatrix} = \begin{bmatrix} \mathbf{A}[n] & \mathbf{I} & \mathbf{0} \\ \mathbf{H}[n]\mathbf{A}[n] & \mathbf{H}[n] & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{s}[n-1] \\ \mathbf{u}[n] \\ \mathbf{w}[n] \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1] \end{bmatrix} . \tag{32}$$

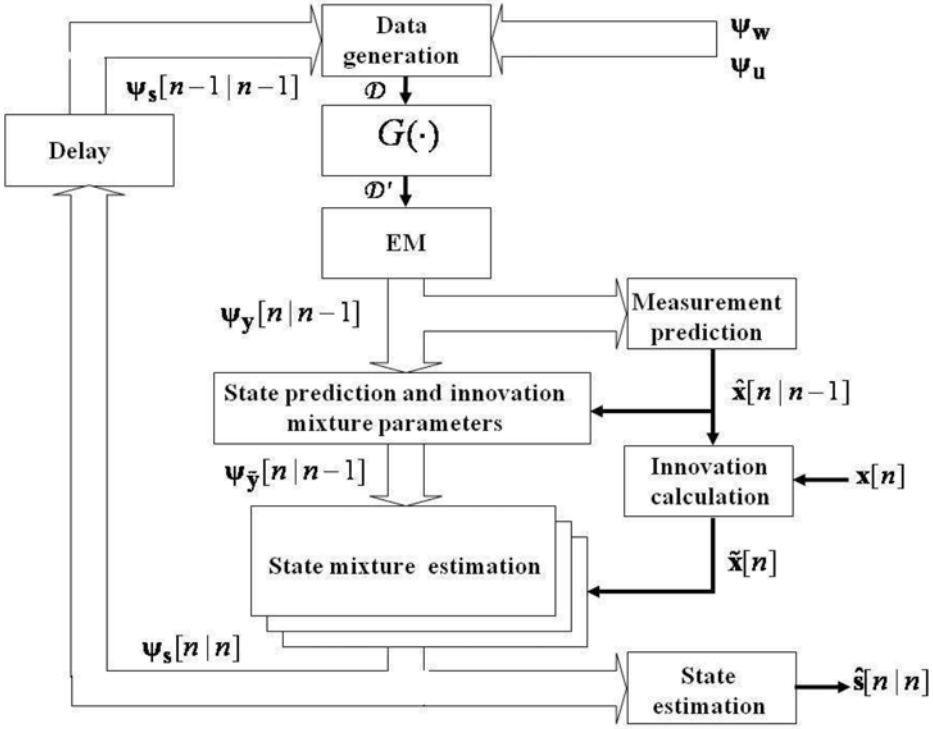

Fig. 1. GMKF schematic diagram.

The conditional distribution of $\begin{bmatrix} \mathbf{s}[n-1] \\ \mathbf{u}[n] \\ \mathbf{w}[n] \end{bmatrix}$ given $\mathcal{X}[n-1]$ is GMM of order *LKM*, because $\mathbf{s}[n-1]$, $\mathbf{u}[n]$ and $\mathbf{w}[n]$ are independent, $\mathbf{s}[n-1]\,|\,\mathcal{X}[n-1]$ is GMM distributed of order L, $\mathbf{u}[n]$ is GMM distributed of order K, and $\mathbf{w}[n]$ is GMM distributed of order M. According to (32), $\mathbf{y}[n]$ is a linear transformation of $\begin{bmatrix} \mathbf{s}[n-1] \\ \mathbf{u}[n] \\ \mathbf{w}[n] \end{bmatrix}$ and therefore, using Proposition 2 in the appendix, the conditional distribution of $\mathbf{y}[n]$ given $\mathcal{X}[n-1]$ is also GMM of order *LKM*:

$$\mathbf{y}[n]|\mathcal{X}[n-1] \sim GMM\left(\alpha_{\mathbf{y}}[n|n-1,\bar{\eta}_{lkm}[n]], \boldsymbol{\mu}_{\mathbf{y}}[n|n-1,\bar{\eta}_{lkm}[n]], \boldsymbol{\Gamma}_{\mathbf{y}}[n|n-1,\bar{\eta}_{lkm}[n]], \right. \\ \left. l=1,\ldots,L;\ k=1,\ldots,K;\ m=1,\ldots,M\right), \tag{33}$$

where

$$\boldsymbol{\mu}_{\mathbf{y}}[n|n-1,\bar{\eta}_{lkm}[n]] \triangleq \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\bar{\eta}_{lkm}[n]] \\ \boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] \end{bmatrix}, \tag{34}$$

$$\boldsymbol{\Gamma}_{\mathbf{y}}[n|n-1,\bar{\eta}_{lkm}[n]] \triangleq \begin{bmatrix} \boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\bar{\eta}_{lkm}[n]] & \boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] \\ \boldsymbol{\Gamma}_{\tilde{\mathbf{x}}\mathbf{s}}[n|n-1,\bar{\eta}_{lkm}[n]] & \boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] \end{bmatrix}, \tag{35}$$

and $\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \bar{\eta}_{lkm}[n]]$ is the MMSE estimator of $\mathbf{s}[n]$ from $\mathcal{X}[n-1]$ given the *lkm*th mixture indicator $\bar{\eta}_{lkm}[n]$, defined as

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \bar{\eta}_{lkm}[n]] = E[\mathbf{s}[n]|\mathcal{X}[n-1], \bar{\eta}_{lkm}[n]] \ . \tag{36}$$

As mentioned above, the distribution of $\mathbf{s}[n]|\mathcal{X}[n]$ is identical to the distribution of $\mathbf{s}[n]|\mathcal{X}[n-1]$, $\tilde{\mathbf{x}}[n]$. Since $\mathbf{s}[n]$ and $\tilde{\mathbf{x}}[n]$ given $\mathcal{X}[n-1]$ are jointly GMM of order *LKM*, then the conditional distribution of $\mathbf{s}[n]|\mathcal{X}[n]$ is GMM:

$$\begin{aligned}\mathbf{s}[n]|\mathcal{X}[n] \sim GMM\ (&\alpha_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]], \boldsymbol{\mu}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]], \mathbf{M}[n|n, \bar{\eta}_{lkm}[n]], \\ &l = 1, \ldots, L; k = 1, \ldots, K; m = 1, \ldots, M)\ ,\end{aligned} \tag{37}$$

where the parameters of this conditional distribution are given by

$$\tilde{\boldsymbol{\theta}}[n|n] \triangleq \{\alpha_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]], \boldsymbol{\mu}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]], \boldsymbol{\Gamma}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]]\}_{(l,k,m)=(1,1,1)}^{(L,K,M)} \ . \tag{38}$$

In the following, expressions for the parameters in $\tilde{\boldsymbol{\theta}}[n|n]$ are derived. Using the Bayesian rule, the estimated mixture weights are given by

$$\alpha_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]] = \frac{\alpha_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]]\Phi(\tilde{\mathbf{x}}[n]; \boldsymbol{\theta}_{lkm}[n])}{\sum_{l'=1}^{L}\sum_{k'=1}^{K}\sum_{m'=1}^{M}\alpha_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{l'k'm'}[n]]\Phi(\tilde{\mathbf{x}}[n]; \boldsymbol{\theta}_{l'k'm'}[n])} \ . \tag{39}$$

The relation between the parameters of the *lkm*th mixture component in the conditional distribution of $\mathbf{s}[n]$, $\tilde{\mathbf{x}}[n]|\mathcal{X}[n-1]$ and in the conditional distribution of $\mathbf{s}[n]|\mathcal{X}[n-1]$, $\tilde{\mathbf{x}}[n]$ is obtained below. For any random mixture indicator $\bar{\eta}_{lkm}[n]$, the vectors $\mathbf{s}[n]$ and $\tilde{\mathbf{x}}[n]$ given $\mathcal{X}[n-1]$, are jointly Gaussian. Therefore, the MMSE estimator of $\mathbf{s}[n]$ from $\mathcal{X}[n]$, given the *lkm*th mixture indicator $\eta_{lkm}[n]$ is obtained using (36) as:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]] = E\ &[\mathbf{s}[n]|\mathcal{X}[n-1], \tilde{\mathbf{x}}[n], \bar{\eta}_{lkm}[n]] = E[\mathbf{s}[n]|\mathcal{X}[n-1], \tilde{\eta}_{lk}[n]] \\ &+ \boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]]\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}^{-1}[n|n-1, \bar{\eta}_{lkm}[n]]\ (\tilde{\mathbf{x}}[n] - \boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]])\ ,\end{aligned} \tag{40}$$

where $\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]]$ and $\boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1, \bar{\eta}_{lkm}[n]]$ are the *lkm*th covariance and cross-covariance matrices of the conditional distribution of $\tilde{\mathbf{x}}[n]|\mathcal{X}[n-1]$, respectively. The covariance matrix $\boldsymbol{\Gamma}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]]$ can be obtained as

$$\begin{aligned}\boldsymbol{\Gamma}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]] = \ &\mathrm{cov}(\mathbf{s}[n]|\tilde{\mathbf{x}}[n], \mathcal{X}[n-1], \bar{\eta}_{lkm}[n]) = \mathrm{cov}(\mathbf{s}[n]|\mathcal{X}[n-1], \tilde{\eta}_{lk}[n]) \\ &-\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}\mathbf{s}}^{H}[n|n-1, \bar{\eta}_{lkm}[n]]\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}^{-1}[n|n-1, \bar{\eta}_{lkm}[n]]\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}\mathbf{s}}[n|n-1, \bar{\eta}_{lkm}[n]] \ .\end{aligned} \tag{41}$$

Following the conventions of the KF, the *lkm*th Kalman gain notation corresponding to the *lkm*th mixture component, is defined as

$$\mathbf{K}_{lkm}[n] \triangleq \boldsymbol{\Gamma}_{\tilde{\mathbf{x}}\mathbf{s}}^{H}[n|n-1, \bar{\eta}_{lkm}[n]]\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}^{-1}[n|n-1, \bar{\eta}_{lkm}[n]] \ . \tag{42}$$

Using (42), expressions (40) and (41) can be rewritten as

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n,\bar{\eta}_{lkm}[n]] = \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]] + \mathbf{K}_{lkm}[n]\left(\tilde{\mathbf{x}}[n] - \boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]]\right) , \quad (43)$$

$$\boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\bar{\eta}_{lkm}[n]] = \boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]] - \mathbf{K}_{lkm}[n]\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}\mathbf{s}}[n|n-1,\bar{\eta}_{lkm}[n]] . \quad (44)$$

The first terms in the RHS of (43) and (44) are the statistics of the predicted state vector. These terms are calculated using Lemma 1.

Lemma 1: The conditional distribution of the state vector $\mathbf{s}[n]$ given $\mathcal{X}[n-1]$ is

$$\mathbf{s}[n]|\mathcal{X}[n-1] \sim GMM\left(\alpha_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]], \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]], \boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]], l=1,\ldots,L\right) ,$$

where the predicted state mixture weights, mean vector and covariance matrix of the *lk*th mixture component are

$$\alpha_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]] = \alpha_{\mathbf{s}}[n-1|n-1,\eta_l[n-1]]\alpha_{\mathbf{u}k}[n] , \quad (45)$$

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]] = E(\mathbf{s}[n]|\mathcal{X}[n-1],\tilde{\eta}_{lk}[n]) = \mathbf{A}[n]\boldsymbol{\mu}_{\mathbf{s}}[n-1|n-1,\eta_l[n-1]] + \boldsymbol{\mu}_{\mathbf{u}k}[n] , \quad (46)$$

$$\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]] = \mathrm{cov}(\mathbf{s}[n]|\mathcal{X}[n-1],\tilde{\eta}_{lk}[n]) = \mathbf{A}[n]\boldsymbol{\Gamma}_{\mathbf{s}}[n-1|n-1,\eta_l[n-1]]\mathbf{A}^H[n] + \boldsymbol{\Gamma}_{\mathbf{u}k}[n] . \quad (47)$$

Proof: According to (1), $\mathbf{s}[n]$ is the sum of two statistically independent GMM-distributed random variables $\mathbf{A}[n]\mathbf{s}[n-1]$ and $\mathbf{u}[n]$. The system noise, $\mathbf{u}[n]$ is independent of $\mathcal{X}[n-1]$, because it is an i.i.d. sequence independent of the state vector at the previous time instance. Therefore, (45)-(47) can be directly obtained from (1) and Proposition 2 in the appendix. □

Lemma 1 proves the state mixture prediction in (13)-(15). Using Proposition 1 in the appendix and the parameters of the conditional state distribution of $\mathbf{s}[n]$ given $\mathcal{X}[n-1]$ obtained in Lemma 1, the state prediction is

$$\hat{\mathbf{s}}[n|n-1] = \sum_{l=1}^{L}\sum_{k=1}^{K} \alpha_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]]\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]] ,$$

which proves the state vector prediction in (16). In order to calculate the second terms in the RHS of (43) and (44), the conditional statistics of the innovation process $\tilde{\mathbf{x}}[n]\,|\,\mathcal{X}[n-1]$ are required.

Lemma 2: The conditional distribution of the innovation process $\tilde{\mathbf{x}}[n]$, defined in (29), given $\mathcal{X}[n-1]$, is GMM of order *LKM*:

$$\tilde{\mathbf{x}}[n]|\mathcal{X}[n-1] \sim GMM\left(\alpha_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]], \boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]], \boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]]; l=1,\ldots,L; k=1,\ldots,K; m=1,\ldots,M\right) , \quad (48)$$

where the mixture parameters are

$$\begin{aligned}
\alpha_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] &= \alpha_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]] \cdot \alpha_{\mathbf{w}m}[n] ,\\
\boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] &= \mathbf{H}[n]\left(\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]] - \hat{\mathbf{s}}[n|n-1]\right) + \boldsymbol{\mu}_{\mathbf{w}m}[n] ,\\
\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] &= \mathbf{H}[n]\mathbf{M}[n|n-1,\tilde{\eta}_{lk}[n]]\mathbf{H}^H[n] + \boldsymbol{\Gamma}_{\mathbf{w}m}[n] .
\end{aligned}$$

Proof: According to (31), the innovation process is a linear transformation of $\mathbf{s}[n-1]$, $\mathbf{u}[n]$ and $\mathbf{w}[n]$. Using the statistical properties of the system state, and system and measurement

noises, the conditional distribution of the innovation process $\tilde{\mathbf{x}}[n]$ given $\mathcal{X}[n-1]$ is GMM of order LKM. The mixture weights, mean vectors and covariance matrices of the lkmth mixture component in the conditional PDF of the innovation process are calculated using (31) and Proposition 2 in the appendix, as follows:

$$\alpha_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] = \alpha_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]]\cdot\alpha_{\mathbf{w}m}[n]\ , \tag{49}$$

$$\begin{aligned}\boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] &= E(\tilde{\mathbf{x}}[n]|\mathcal{X}[n-1],\bar{\eta}_{lkm}[n]) \\ &= \mathbf{H}[n]\left(E[\mathbf{s}[n]|\mathcal{X}[n-1],\tilde{\eta}_{lk}[n]]-\hat{\mathbf{s}}[n|n-1]\right)+E[\mathbf{w}[n]|\mathcal{X}[n-1],\bar{\eta}_{lkm}[n]]\end{aligned} \tag{50}$$

$$= \mathbf{H}[n](\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]]-\hat{\mathbf{s}}[n|n-1])+\boldsymbol{\mu}_{\mathbf{w}m}[n], \tag{51}$$

$$\begin{aligned}\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] &= \mathrm{cov}(\tilde{\mathbf{x}}[n]|\mathcal{X}[n-1],\bar{\eta}_{lkm}[n]) \\ &= \mathbf{H}[n]\mathrm{cov}\left(\mathbf{s}[n]|\mathcal{X}[n-1],\bar{\eta}_{lkm}[n]\right)\mathbf{H}^{H}[n]+\mathrm{cov}\left(\mathbf{w}[n]|\mathcal{X}[n-1],\tilde{\eta}_{lk}[n]\right) \\ &= \mathbf{H}[n]\mathbf{M}[n|n-1,\tilde{\eta}_{lk}[n]]\mathbf{H}^{H}[n]+\boldsymbol{\Gamma}_{\mathbf{w}m}[n]\ ,\end{aligned} \tag{52}$$

where $\alpha_{\mathbf{s}}[n\,|\,n-1,\ \tilde{\eta}_{lk}[n]]$, $\boldsymbol{\mu}_{\mathbf{s}}[n\,|\,n-1,\ \tilde{\eta}_{lk}[n]]$ and $\mathbf{M}[n\,|\,n-1,\ \tilde{\eta}_{lk}[n]]$ are obtained using Lemma 1. □

In the definition of the Kalman gain in (42), the cross-covariance matrix $\boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n\,|\,n-1,\ \bar{\eta}_{lkm}[n]]$ is required. According to (32) and Proposition 2 in the appendix, the joint distribution of $\mathbf{s}[n]$ and $\tilde{\mathbf{x}}[n]$ given $\mathcal{X}[n-1]$ is GMM of order LKM whose lkmth cross-covariance matrix $\boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n\,|\,n-1,\ \bar{\eta}_{lkm}[n]]$, defined in (35), is

$$\boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]]=\mathrm{cov}\left(\mathbf{s}[n],\tilde{\mathbf{x}}[n]|\mathcal{X}[n-1],\bar{\eta}_{lkm}[n]\right)\ . \tag{53}$$

By substitution of (31) into (53) we obtain

$$\begin{aligned}\boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1,\bar{\eta}_{lkm}[n]] &= \mathrm{cov}\left(\mathbf{s}[n],(\mathbf{H}[n](\mathbf{s}[n]-\hat{\mathbf{s}}[n|n-1])+\mathbf{w}[n])\,|\mathcal{X}[n-1],\bar{\eta}_{lkm}[n]\right) \\ &= \mathrm{cov}\left(\mathbf{s}[n]|\mathcal{X}[n-1],\tilde{\eta}_{lk}[n]\right)\mathbf{H}^{H}[n]+E\left[(\mathbf{s}[n]-\hat{\mathbf{s}}[n|n-1])\mathbf{w}^{H}[n]|\mathcal{X}[n-1],\bar{\eta}_{lkm}[n]\right]\end{aligned} \tag{54}$$

$$=\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]]\mathbf{H}^{H}[n]+(\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]]-\hat{\mathbf{s}}[n|n-1])\,\boldsymbol{\mu}_{\mathbf{w}m}^{H}[n]\ . \tag{55}$$

where (54) is obtained under the assumption that $\mathbf{w}[n]$ is independent of $\mathbf{s}[n]$, and since $\hat{\mathbf{s}}[n\,|\,n-1]$ given $\mathcal{X}[n-1]$ is deterministic. Using (52) and (55), the Kalman gain defined in (42), can be rewritten as

$$\begin{aligned}\mathbf{K}_{lkm}[n] &= \left(\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]]\mathbf{H}^{H}[n]+(\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]]-\hat{\mathbf{s}}[n|n-1])\,\boldsymbol{\mu}_{\mathbf{w}m}^{H}[n]\right) \\ &\quad\cdot\left(\mathbf{H}[n]\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]]\mathbf{H}^{H}[n]+\boldsymbol{\Gamma}_{\mathbf{w}m}[n]\right)^{-1}\ .\end{aligned} \tag{56}$$

Equation (56) proves the Kalman gain equation in (21). Finally, using (44) and (55), $\boldsymbol{\Gamma}_{\mathbf{s}}[n\,|\,n,\ \bar{\eta}_{lkm}[n]]$ can be expressed as

$$\boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\bar{\eta}_{lkm}[n]]=(\mathbf{I}-\mathbf{K}_{lkm}[n]\mathbf{H}[n])\,\boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\tilde{\eta}_{lk}[n]]\ . \tag{57}$$

Equations (39), (43), and (44) prove the estimated mixture statistics in (23), (24) and (25). Using the estimated parameters of the distribution in (37), the MMSE state estimation is

$$\hat{\mathbf{s}}[n|n] = \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{m=1}^{M} \alpha_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}]\boldsymbol{\mu}_{\mathbf{s}}[n|n, \bar{\eta}_{lkm}[n]] \ ,$$

which proves the state vector estimate in (26). This result is identical to the GSF presented in [19]. The GSF derived here suffers from exponential model order growth over the iterations. The system state PDF order grows twice: during prediction stage, derived in Lemma 1, and during the innovation process in Lemma 2. Section C.A presents a new model order reduction algorithm, which is optimal in the KLD sense. □

The proof for this theorem appears also in [42]-[43], for the case of non-stationary Gaussian noise processes with Markov noise statistics and Gaussian initial conditions. The assumption of mutually-independent i.i.d. noises enables to provide a simpler closed-form solution.

Model order reduction

In [19] it was shown that the use of GMM distributions for the system state, measurement and system noises results in exponential growth of the number of the mixture components over iterations, and therefore, the GSF is impractical [18],[21]. Several suboptimal techniques for model order reduction have been presented in [21]. In this section, an optimal model order reduction algorithm based on the minimization of the KLD is proposed. The model order reduction stage is implemented once at the end of each iteration. The main idea of the proposed model order reduction scheme is to approximate the posterior *LKM* order system state distribution $\bar{f}_{\mathbf{s}[n]|\mathcal{X}[n]}(\cdot)$ by a reduced order GMM $f_{\mathbf{s}[n]|\mathcal{X}[n]}(\cdot)$. Estimation of the reduced order distribution involves minimization of the distance between the estimated and the "true" PDFs. The KLD is widely used to measure the distance between distributions:

$$KLD\left(\bar{f}_{\mathbf{s}[n]|\mathcal{X}[n]}(\mathbf{s}[n], \bar{\boldsymbol{\theta}}[n|n]) || f_{\mathbf{s}[n]|\mathcal{X}[n]}(\mathbf{s}[n], \boldsymbol{\theta}[n|n])\right) = \int_{\boldsymbol{\xi}} \bar{f}_{\mathbf{s}[n]|\mathcal{X}[n]}(\boldsymbol{\xi}, \bar{\boldsymbol{\theta}}[n|n]) \log \frac{\bar{f}_{\mathbf{s}[n]|\mathcal{X}[n]}(\boldsymbol{\xi}, \bar{\boldsymbol{\theta}}[n|n])}{f_{\mathbf{s}[n]|\mathcal{X}[n]}(\boldsymbol{\xi}, \boldsymbol{\theta}[n|n])} d\boldsymbol{\xi} \ .$$

This induces that our main goal is to obtain:

$$\boldsymbol{\theta}[n|n] = \arg\min_{\boldsymbol{\theta}} KLD\left(\bar{f}_{\mathbf{s}[n]|\mathcal{X}[n]}(\mathbf{s}[n], \bar{\boldsymbol{\theta}}) || f_{\mathbf{s}[n]|\mathcal{X}[n]}(\mathbf{s}[n], \boldsymbol{\theta}[n|n])\right) \ . \tag{58}$$

In [50], [51], it is shown that this minimization problem can equivalently be solved using Monte-Carlo realizations by the maximization of the likelihood function:

$$\boldsymbol{\theta}[n|n] = \arg\max_{\boldsymbol{\theta}} \lim_{J\to\infty} \prod_{j=1}^{J} f_{\mathbf{s}[n]|\mathcal{X}[n]}(\mathbf{s}_j, \boldsymbol{\theta}) \ , \tag{59}$$

where $\mathbf{s}_j$, j = 1,..., J are artificial samples generated from the distribution $\bar{f}_{\mathbf{s}[n]|\mathcal{X}[n]}(\mathbf{s}[n], \bar{\boldsymbol{\theta}}[n|n])$. Note that the size of the artificial data is theoretically unlimited and it is practically set according to the processor computational resources. There is no closed form solution for this maximization problem and the maximization can be carried out by the EM algorithm [50], using the following iterative expressions for mixture parameters estimation:

$$\begin{aligned} P(l|\mathbf{s}_j) &= \frac{\alpha_{\mathbf{s}}[n-1|n-1, \eta_l[n-1]]}{\Phi(\mathbf{s}_j; \boldsymbol{\theta}_l)} \ , \ \forall l = 1, \ldots, L \ , j = 1, \ldots, J \\ \alpha_{\mathbf{s}}[n|n, \eta_l[n]] &= \frac{1}{J}\sum_{j=1}^{J} P(l|\mathbf{s}_j) \ , \end{aligned}$$

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]] = \frac{\sum_{j=1}^{J} P(l|\mathbf{s}_j)\mathbf{s}_j}{\sum_{j=1}^{J} P(l|\mathbf{s}_j)},$$

$$\boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]] = \frac{\sum_{j=1}^{J} P(l|\mathbf{s}_j)(\mathbf{s}_j-\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]])(\mathbf{s}_j-\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]])^H}{\sum_{j=1}^{J} P(l|\mathbf{s}_j)},$$

where $\boldsymbol{\theta}_l = \{\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]],\ \boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]]\}$ and $\Phi(\mathbf{s}_j;\boldsymbol{\theta}_l)$ was defined in (6). One of the computationally efficient variations of the EM algorithm is the greedy EM algorithm, which iteratively minimizes the KLD of the estimated from the "true" distributions. The resulting reduced order mixture estimated using the greedy EM algorithm [51] is:

$$\mathbf{s}[n]|\mathcal{X}[n] \sim GMM\left(\alpha_{\mathbf{s}}[n|n,\eta_l[n]],\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]],\boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]]\ ;\ l=1,\ldots,L\right).$$

Note that in general, the number of the mixture components may vary with n. Thus, the GMM order might be obtained using model order selection algorithms, such as the minimum description length (MDL) [52]. Alternatively, L can be set as an upper bound on the number of mixture components in the conditional PDF of $\mathbf{s}[n]\,|\,\mathcal{X}[n]$. Note that the EM role in the proposed order reduction scheme differs from the off-line EM algorithm presented in [21]. The proposed KLD-based model order reduction stage can be summarized as follows:

1. Generate J samples $\{\mathbf{s}_j\}_{j=1}^{J}$ from the distribution of $\mathbf{s}[n]\,|\,\mathcal{X}[n]$, obtained by the GSF at the end of the estimated stage:

$$\mathbf{s}[n]|\mathcal{X}[n] \sim GMM\left(\alpha_{\mathbf{s}}[n|n,\bar{\eta}_{lkm}[n]],\boldsymbol{\mu}_{\mathbf{s}}[n|n,\bar{\eta}_{lkm}[n]],\boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\bar{\eta}_{lkm}[n]],\right.$$
$$\left. l=1,\ldots,L;k=1,\ldots,K;m=1,\ldots,M\right)$$

2. Estimate a reduced order GMM with L mixture components using the greedy EM algorithm as follows [50], [51]:
   a. Initialization: calculate the ML estimation of the first Gaussian parameters, $\hat{\boldsymbol{\theta}}_0$, using the training data set, $\{\mathbf{s}_j\}_{j=1}^{J}$ in order to obtain $f_{\mathbf{s}}^{(\tilde{K}=0)}(\mathbf{s})$. Set $\tilde{K}=1$.
   b. Generate a new mixture by

$$f_{\mathbf{s}}^{(\tilde{K})}(\mathbf{s}) = (1-\hat{\alpha}_{\mathbf{s}}^{(\tilde{K})})f_{\mathbf{s}}^{(\tilde{K}-1)}(\mathbf{s}) + \hat{\alpha}_{\mathbf{s}}^{(\tilde{K})}\Phi(\mathbf{s},\hat{\boldsymbol{\theta}}_{\tilde{K}}), \tag{60}$$

   where $\hat{\boldsymbol{\theta}}_{\tilde{K}}$ and $\hat{\alpha}_{\mathbf{s}}^{(\tilde{K})}$ are numerically obtained by:

$$\{\hat{\boldsymbol{\theta}}_{\tilde{K}},\hat{\alpha}_{\mathbf{s}}^{(\tilde{K})}\} = \arg\max_{\boldsymbol{\theta},\alpha}\sum_{j=1}^{J}\log\left[(1-\alpha)f_{\mathbf{s}}^{(\tilde{K}-1)}(\mathbf{s}_j)+\alpha\Phi(\mathbf{s}_j,\boldsymbol{\theta})\right]. \tag{61}$$

   See [51] for further details.
   c. Apply the EM algorithm on the mixture $f_{\mathbf{s}}^{(\tilde{K})}(\mathbf{s})$, initialized by the mixture parameters, obtained in the previous step.

d. If the log-likelihood function in (60) does not significantly increase, set $f_{\mathbf{s}[n]|\mathcal{X}[n]}(\cdot|\mathcal{X}[n]) = f_{\mathbf{s}}^{(\tilde{K})}(\cdot)$; otherwise, set $\tilde{K} = \tilde{K} + 1$ and return to step (b).

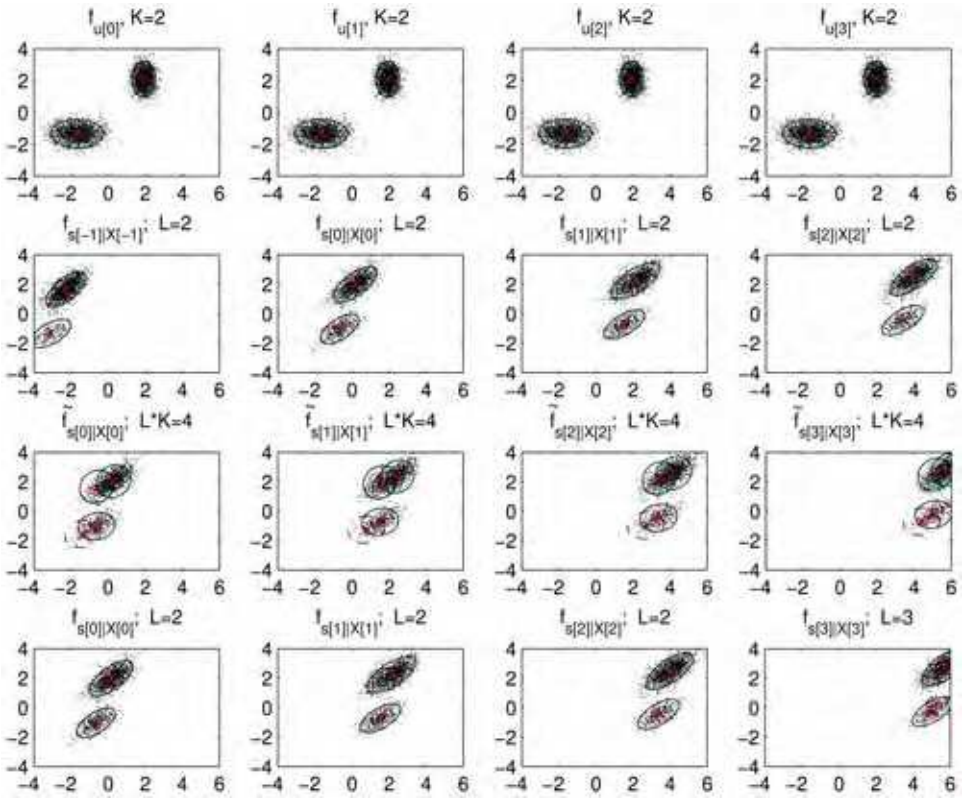

Fig. 2. The model order reduction procedure over 4 time instances.

The KLD-based model order reduction stage is illustrated via a two-dimensional example in Fig. 2. The figure shows the mixture order evaluation over iterations. The measurement noise was assumed to be zero-mean Gaussian. Fig. 2 shows a sequence of four time instances from $n = 0$ to $n = 3$. The corresponding two-dimensional data is represented by dots. The distribution of the system noise $\mathbf{u}[n]$ with two mixture components is shown in the first row of Fig. 2. The estimated distribution of the system state $\mathbf{s}[n-1]|\mathcal{X}[n-1]$ at the previous time instance is shown in the second row. The distribution of the estimated system state with increased model order $\mathbf{s}[n]|\mathcal{X}[n]$ is shown in the third row. The mixture components of this distribution are obtained from the convolution of the mixture components in the first two rows. The figure shows that the mixture order grows at this convolution stage. The data are generated from the distributions of the system state prediction on the third row, and new distributions with reduced order are estimated from the generated data. The reduced-order conditional distributions of the system state given the data at time instance n, $\mathbf{s}[n]|\mathcal{X}[n]$, are shown in the last row. The figure shows that no

significant information was lost during the model order reduction process, and the reduced order distribution adequately represents the data distribution. Next, the KLD was used to evaluate the distortion caused by the mixture order reduction procedure. A Gaussian measurement noise was assumed also in this example, i.e. $M = 1$. The system noise was modeled as GMM of order two. The proposed KLD-based model order reduction method was compared to the MR clustering method proposed in [22]. The first row in Fig. 3 shows the PDF of the system state in sequential steps. The uncertainty ellipses represent Gaussians with means marked by 'x' and their widths correspond to 2σ (2 standard deviations). Ellipses with solid, and dashed-dot lines correspond to mixture component weights satisfying: $\alpha_s[n \mid n, \eta_l[n]] > 0.01$, and $\alpha_s[n \mid n, \eta_l[n]] \le 0.01$, respectively. Thus, the system state order grows from 4 to 32 over four iterations. The second line in Fig. 3 shows the reduced order GMM at the output of the KLD based model order reduction stage. The third line on Fig. 3 shows the reduced order GMM at the output of the MR algorithm. The figure shows significant difference between the mixtures obtained by the two methods. The KLD of the reduced-order GMM from the exponentially growing order GMM was calculated according to the following procedure:

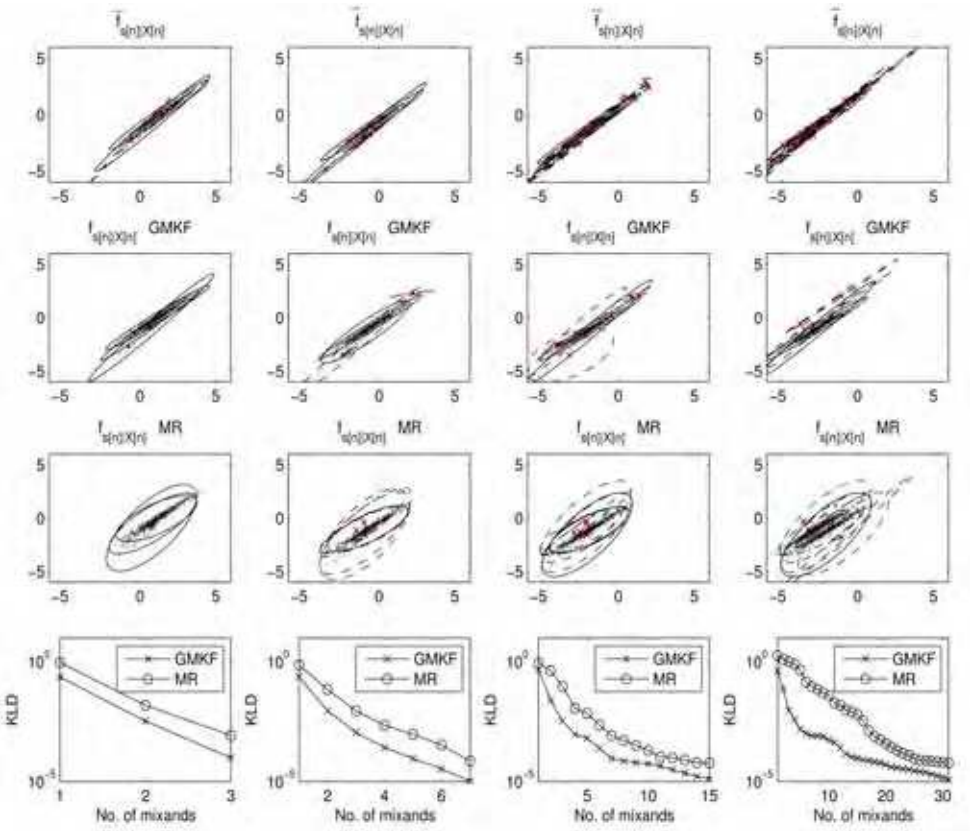

Fig. 3. KLD between system state distributions, obtained with and without model order reduction stage over 4 time instances.

- Sample artificial data vectors $\mathbf{s}_j$, $\forall j = 1,..., J$ from the increased order GMM, representing the posterior state PDF.
- Evaluate the KLD using Monte-Carlo simulations:

$$KLD(\bar{f}_{\mathbf{s}[n]|\mathcal{X}[n]}|f_{\mathbf{s}[n]|\mathcal{X}[n]}) \simeq \frac{1}{J}\sum_{j=1}^{J}\log\frac{\bar{f}_{\mathbf{s}[n]|\mathcal{X}[n]}(\mathbf{s}_j)}{f_{\mathbf{s}[n]|\mathcal{X}[n]}(\mathbf{s}_j)} .$$

In this example, the artificial data size was chosen to be J = 5000. The KLD between these distributions is shown at the last row as a function of the mixture order of $f_{\mathbf{s}[n]|\mathcal{X}[n]}$ for the two tested model order reduction methods. It is shown that for the proposed model order reduction method (a) the KLD decreases when the mixture order grows, (b) the proposed model order reduction method outperforms the *ad-hoc* MR method in the KLD sense. Fig. 3 shows that the distribution obtained by the proposed model order reduction method, represents the data fairly good.

B. NL-GMKF

In this section, the recursive NL-GMKF for the nonlinear and non-Gaussian model described in Section IIB, is derived based on the MMSE criterion.

Summary

The following summarizes the NL-GMKF for recursive estimation of $\mathbf{s}[n]$ from $\mathcal{X}[n]$.

1. Initialization:

Initialize the L-order GMM parameters of the state vector at time instance n = –1.

$$\begin{aligned}\alpha_{\mathbf{s}}[-1|-1,\eta_{\mathbf{s}l}[-1]] &= \alpha_{\mathbf{s}l}[-1] ,\\ \boldsymbol{\mu}_{\mathbf{s}}[-1|-1,\eta_{\mathbf{s}l}[-1]] &= \boldsymbol{\mu}_{\mathbf{s}l}[-1] ,\\ \boldsymbol{\Gamma}_{\mathbf{s}}[-1|-1,\eta_{\mathbf{s}l}[-1]] &= \boldsymbol{\Gamma}_{\mathbf{s}l}[-1] .\end{aligned}$$

Set n = 0.

2. Mixture parameters of the state and measurement prediction:

- Generate an artificial data set $\mathcal{D}$ from the conditional distribution of $\begin{bmatrix}\mathbf{s}[n-1]\\ \mathbf{u}[n]\\ \mathbf{w}[n]\end{bmatrix}$ given $\mathcal{X}[n-1]$, according to the PDF of $\mathbf{s}[n-1]\,|\,\mathcal{X}[n-1]$ from the previous step and PDFs of $\mathbf{u}[n]$ and $\mathbf{w}[n]$.
- Apply the nonlinear function $\mathbf{G}(\cdot) = \begin{bmatrix}\mathbf{a}(\cdot,\cdot)\\ \mathbf{h}(\mathbf{a}(\cdot,\cdot),\cdot)\end{bmatrix}$ on $\mathcal{D}$ and obtain a new artificial data set $\mathcal{D}'$.
- Model the conditional distribution of $\mathbf{y}[n]$ given $\mathcal{X}[n-1]$ using the new artificial data $\mathcal{D}'$ by GMM of order L to obtain the parameters of $\psi_{\mathbf{y}}[n\,|\,n-1]$:

$$\psi_{\mathbf{y}}[n|n-1] \triangleq \left\{\alpha_{\mathbf{y}}[n|n-1,\eta_l[n]], \boldsymbol{\mu}_{\mathbf{y}}[n|n-1,\eta_l[n]], \boldsymbol{\Gamma}_{\mathbf{y}}[n|n-1,\eta_l[n]]\right\}_{l=1}^{L} , \tag{62}$$

where

$$\boldsymbol{\mu}_{\mathbf{y}}[n|n-1,\eta_l[n]] \triangleq \begin{bmatrix}\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_l[n]]\\ \boldsymbol{\mu}_{\mathbf{x}}[n|n-1,\eta_l[n]]\end{bmatrix} , \tag{63}$$

and

$$\mathbf{\Gamma}_{\mathbf{y}}[n|n-1,\eta_l[n]] = \begin{bmatrix} \mathbf{\Gamma}_{\mathbf{s}}[n|n-1,\eta_l[n]] & \mathbf{\Gamma}_{\mathbf{sx}}[n|n-1,\eta_l[n]] \\ \mathbf{\Gamma}_{\mathbf{sx}}^H[n|n-1,\eta_l[n]] & \mathbf{\Gamma}_{\mathbf{x}}[n|n-1,\eta_l[n]] \end{bmatrix} . \tag{64}$$

3. Innovation:
Calculate the innovation vector:

$$\tilde{\mathbf{x}}[n] = \mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1] , \tag{65}$$

where

$$\hat{\mathbf{x}}[n|n-1] = \sum_{l=1}^{L} \alpha_{\mathbf{y}}[n|n-1,\eta_l[n]] \boldsymbol{\mu}_{\mathbf{x}}[n|n-1,\eta_l[n]] . \tag{66}$$

4. PDF parameters of the state prediction and the innovation:
Obtain the parameters

$$\boldsymbol{\psi}_{\tilde{\mathbf{y}}}[n|n-1] = \left\{ \alpha_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]], \boldsymbol{\mu}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]], \mathbf{\Gamma}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] \right\}_{l=1}^{L} , \tag{67}$$

where

$$\boldsymbol{\mu}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_l[n]] \\ \boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] \end{bmatrix} = \boldsymbol{\mu}_{\mathbf{y}}[n|n-1,\eta_l[n]] - \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{x}}[n|n-1] \end{bmatrix} , \tag{68}$$

$$\mathbf{\Gamma}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] = \begin{bmatrix} \mathbf{\Gamma}_{\mathbf{s}}[n|n-1,\eta_l[n]] & \mathbf{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] \\ \mathbf{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}^H[n|n-1,\eta_l[n]] & \mathbf{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] \end{bmatrix} = \mathbf{\Gamma}_{\mathbf{y}}[n|n-1,\eta_l[n]] , \tag{69}$$

$$\alpha_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] = \alpha_{\mathbf{y}}[n|n-1,\eta_l[n]] , \tag{70}$$

5. Kalman gains:
Calculate the Kalman gains

$$\mathbf{K}_l[n] \triangleq \mathbf{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] \mathbf{\Gamma}_{\tilde{\mathbf{x}}}^{-1}[n|n-1,\eta_l[n]] . \tag{71}$$

6. Estimation
6a. Estimated state mixture parameters:
Obtain the estimated state mixture parameters:

$$\boldsymbol{\psi}_{\mathbf{s}}[n|n] \triangleq \{\alpha_{\mathbf{s}}[n|n,\eta_l[n]], \boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]], \mathbf{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]]\}_{l=1}^{L} , \tag{72}$$

where

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]] = \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_l[n]] + \mathbf{K}_l[n] \left(\tilde{\mathbf{x}}[n] - \boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] \right) , \tag{73}$$

$$\mathbf{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]] = \mathbf{\Gamma}_{\mathbf{s}}[n|n-1,\eta_l[n]] - \mathbf{K}_l[n] \mathbf{\Gamma}_{\tilde{\mathbf{x}}\mathbf{s}}[n|n-1,\eta_l[n]] , \tag{74}$$

$$\alpha_{\mathbf{s}}[n|n,\eta_l[n]] \;=\; \frac{\alpha_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]]\Phi(\tilde{\mathbf{x}}[n];\boldsymbol{\theta}_{\tilde{\mathbf{x}}l}[n|n-1])}{\sum_{l'=1}^{L}\alpha_{\tilde{\mathbf{y}}}[n|n-1,\eta_{l'}[n]]\Phi(\tilde{\mathbf{x}}[n];\boldsymbol{\theta}_{\tilde{\mathbf{x}}l'}[n|n-1])}\,, \tag{75}$$

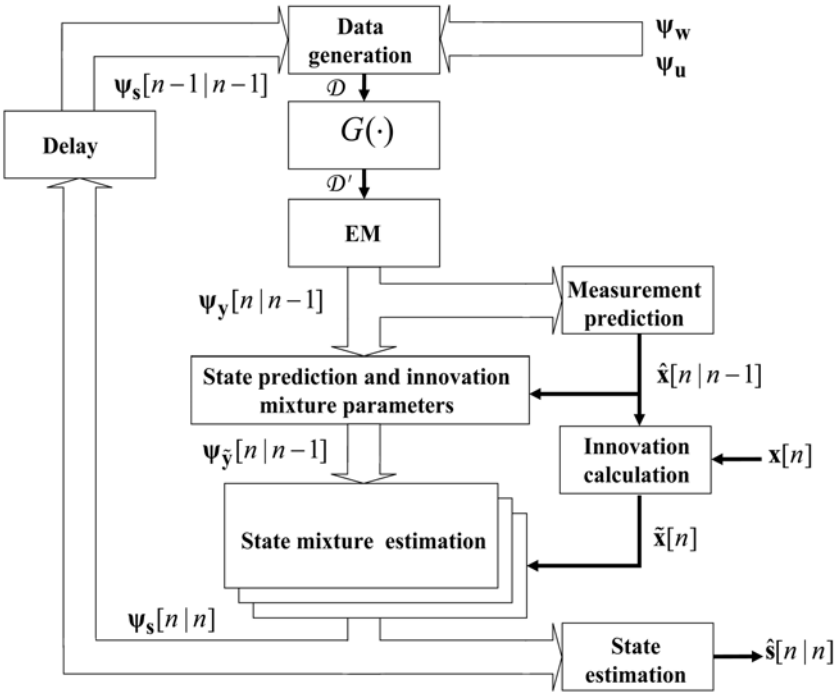

Fig. 4. NL-GMKF schematic diagram.

6b. Estimation of the state vector:
Obtain the system state estimation:

$$\hat{\mathbf{s}}[n|n]=\sum_{l=1}^{L}\alpha_{\mathbf{s}}[n|n,\eta_l[n]]\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]]\;. \tag{76}$$

7. Set $n \rightarrow n+1$, go to step 2.
The NL-GMKF algorithm is schematically presented in Fig. 4.
Derivation
Let $\hat{\mathbf{x}}[n|n-1]$ denote the MMSE estimator of $\mathbf{x}[n]$ from $\mathcal{X}[n-1]$ using (8), $\hat{\mathbf{x}}[n|n-1]$ is given by

$$\hat{\mathbf{x}}[n|n-1] = E(\mathbf{x}[n]|\mathcal{X}[n-1]) = E(\mathbf{h}(\mathbf{s}[n],\mathbf{w}[n])|\mathcal{X}[n-1])\;, \tag{77}$$

and the innovation process defined as $\tilde{\mathbf{x}}[n]$, is given by

$$\tilde{\mathbf{x}}[n] = \mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1] = \mathbf{h}(\mathbf{s}[n],\mathbf{w}[n]) - \hat{\mathbf{x}}[n|n-1]\;. \tag{78}$$

If the transformation $\mathcal{X}[n] \leftrightarrow [\mathcal{X}^T[n-1],\ \tilde{\mathbf{x}}^T[n]]^T$ is one-to-one, then the conditional distribution of $\mathbf{s}[n] \mid \mathcal{X}[n]$ is identical to the conditional distribution of $\mathbf{s}[n] \mid \mathrm{X}[n-1],\ \tilde{\mathbf{x}}[n]$. Since $\mathbf{s}[n]$ and $\tilde{\mathbf{x}}[n]$ given $\mathcal{X}[n-1]$ are assumed to be jointly GMM of order L, the conditional distribution of $\mathbf{s}[n] \mid \tilde{\mathbf{x}}[n], \mathcal{X}[n-1]$ given the random mixture indicator $\eta_l[n]$ [47], is Gaussian. Therefore, the conditional distribution of $\mathbf{s}[n] \mid \mathcal{X}[n]$ is GMM of order L:

$$\mathbf{s}[n]|\mathcal{X}[n] \sim GMM\left(\alpha_{\mathbf{s}}[n|n,\eta_l[n]], \boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]], \boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]]; l=1,\ldots,L\right) . \tag{79}$$

In the following, the parameters of this conditional distribution, $\boldsymbol{\psi}_{\mathbf{s}}[n \mid n]$, where

$$\boldsymbol{\psi}_{\mathbf{s}}[n|n] \triangleq \{\alpha_{\mathbf{s}}[n|n,\eta_l[n]], \boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]], \boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]]\}_{l=1}^{L} , \tag{80}$$

are derived. Since the conditional distribution of $(\mathbf{s}[n],\ \tilde{\mathbf{x}}[n])$ given the random mixture indicator $\eta_l[n]$, is jointly Gaussian, then the mean vector $\boldsymbol{\mu}_{\mathbf{s}}[n \mid n, \eta_l[n]]$ and covariance matrix $\boldsymbol{\Gamma}_{\mathbf{s}}[n \mid n, \eta_l[n]]$ can be obtained as

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]] &= E[\mathbf{s}[n]|\tilde{\mathbf{x}}[n],\mathcal{X}[n-1],\eta_l[n]] \\ &= E[\mathbf{s}[n]|\mathcal{X}[n-1],\eta_l[n]] + \boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]]\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}^{-1}[n|n-1,\eta_l[n]]\left(\tilde{\mathbf{x}}[n]-\boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]]\right) .\end{aligned} \tag{81}$$

$$\begin{aligned}\boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]] &= \mathrm{cov}(\mathbf{s}[n]|\tilde{\mathbf{x}}[n],\mathcal{X}[n-1],\eta_l[n]) \\ &= \mathrm{cov}(\mathbf{s}[n]|\mathcal{X}[n-1],\eta_l[n]) - \boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]]\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}^{-1}[n|n-1,\eta_l[n]]\boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}^{H}[n|n-1,\eta_l[n]] ,\end{aligned} \tag{82}$$

where

$$\begin{aligned}\boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] &= E[\tilde{\mathbf{x}}[n]|\mathcal{X}[n-1],\eta_l[n]] \\ \boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] &= \mathrm{cov}(\tilde{\mathbf{x}}[n]|\mathcal{X}[n-1],\eta_l[n]) \\ \boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] &= \mathrm{cov}(\mathbf{s}[n],\tilde{\mathbf{x}}[n]|\mathcal{X}[n-1],\eta_l[n]) .\end{aligned} \tag{83}$$

Following the conventions of the KF, the lth Kalman gain corresponding to the lth mixture component, is defined as

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]] = \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_l[n]] + \mathbf{K}_l[n]\left(\tilde{\mathbf{x}}[n]-\boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]]\right) , \tag{84}$$

Using (84), Eqs. (81) and (82) can be rewritten as

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]] = \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_l[n]] + \mathbf{K}_l[n]\left(\tilde{\mathbf{x}}[n]-\boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]]\right) , \tag{85}$$

$$\boldsymbol{\Gamma}_{\mathbf{s}}[n|n,\eta_l[n]] = \boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\eta_l[n]] - \mathbf{K}_l[n]\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}\mathbf{s}}[n|n-1,\eta_l[n]] . \tag{86}$$

In the following, the parameters required in (85) and (86), are obtained. Let $\mathbf{y}[n] \triangleq \begin{bmatrix}\mathbf{s}[n]\\ \mathbf{x}[n]\end{bmatrix}$ and $\tilde{\mathbf{y}}[n] \triangleq \begin{bmatrix}\mathbf{s}[n]\\ \tilde{\mathbf{x}}[n]\end{bmatrix}$. Then, by using (7), (8) and (78) one obtains

$$\mathbf{y}[n] = \begin{bmatrix}\mathbf{a}(\mathbf{s}[n-1],\mathbf{u}[n])\\ \mathbf{h}(\mathbf{a}(\mathbf{s}[n-1],\mathbf{u}[n]),\mathbf{w}[n])\end{bmatrix} = \mathbf{G}\left(\begin{bmatrix}\mathbf{s}[n-1]\\ \mathbf{u}[n]\\ \mathbf{w}[n]\end{bmatrix}\right) . \tag{87}$$

and

$$\tilde{\mathbf{y}}[n] \quad = \quad \mathbf{y}[n] - \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{x}}[n|n-1] \end{bmatrix} = \mathbf{G}\left(\begin{bmatrix} \mathbf{s}[n-1] \\ \mathbf{u}[n] \\ \mathbf{w}[n] \end{bmatrix} \right) - \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{x}}[n|n-1] \end{bmatrix}. \tag{88}$$

Since $\tilde{\mathbf{y}}[n]$, is a linear transformation of $\mathbf{y}[n]$, then the vectors $\mathbf{s}[n]$ and $\tilde{\mathbf{x}}[n]$ given $\mathcal{X}[n-1]$ are jointly GMM of order L, that is, the conditional distribution of $\tilde{\mathbf{y}}[n]$ given $\mathcal{X}[n-1]$ can be modeled by an L-order GMM with parameters:

$$\boldsymbol{\psi}_{\tilde{\mathbf{y}}}[n|n-1] = \left\{ \alpha_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]], \boldsymbol{\mu}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]], \boldsymbol{\Gamma}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] \right\}_{l=1}^{L}, \tag{89}$$

where

$$\boldsymbol{\mu}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] \quad \triangleq \quad \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_l[n]] \\ \boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] \end{bmatrix}, \tag{90}$$

$$\boldsymbol{\Gamma}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] \quad \triangleq \quad \begin{bmatrix} \boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\eta_l[n]] & \boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] \\ \boldsymbol{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}}^{H}[n|n-1,\eta_l[n]] & \boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] \end{bmatrix}. \tag{91}$$

Using the properties of the jointly GMM-distributed random processes, $\mathbf{s}[n] \,|\, \mathcal{X}[n-1]$ and $\tilde{\mathbf{x}}[n] \,|\, \mathcal{X}[n-1]$, the mixture weights can be obtained as:

$$\alpha_{\mathbf{s}}[n|n,\eta_l[n]] \quad = \quad \frac{\alpha_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]]\Phi(\tilde{\mathbf{x}}[n];\boldsymbol{\theta}_{\tilde{\mathbf{x}}l}[n|n-1])}{\sum_{l'=1}^{L} \alpha_{\tilde{\mathbf{y}}}[n|n-1,\eta_{l'}[n]]\Phi(\tilde{\mathbf{x}}[n];\boldsymbol{\theta}_{\tilde{\mathbf{x}}l'}[n|n-1])}, \tag{92}$$

where

$$\boldsymbol{\theta}_{\tilde{\mathbf{x}}l}[n|n-1] = \{ \boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]], \boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] \}. \tag{93}$$

Therefore, one can calculate the PDF parameters of $\mathbf{s}[n] \,|\, \mathcal{X}[n]$, given in (85), (86) and (92), using the parameters of the distribution $\tilde{\mathbf{y}}[n] \,|\, \mathcal{X}[n-1]$ obtained in the following. The conditional PDF of $\mathbf{y}[n] \,|\, \mathcal{X}[n-1]$ is modeled by GMM of order L with parameters:

$$\boldsymbol{\psi}_{\mathbf{y}}[n|n-1] \triangleq \left\{ \alpha_{\mathbf{y}}[n|n-1,\eta_l[n]], \boldsymbol{\mu}_{\mathbf{y}}[n|n-1,\eta_l[n]], \boldsymbol{\Gamma}_{\mathbf{y}}[n|n-1,\eta_l[n]] \right\}_{l=1}^{L}, \tag{94}$$

where

$$\boldsymbol{\mu}_{\mathbf{y}}[n|n-1,\eta_l[n]] \quad \triangleq \quad \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_l[n]] \\ \boldsymbol{\mu}_{\mathbf{x}}[n|n-1,\eta_l[n]] \end{bmatrix}, \tag{95}$$

and

$$\boldsymbol{\Gamma}_{\mathbf{y}}[n|n-1,\eta_l[n]] = \begin{bmatrix} \boldsymbol{\Gamma}_{\mathbf{s}}[n|n-1,\eta_l[n]] & \boldsymbol{\Gamma}_{\mathbf{sx}}[n|n-1,\eta_l[n]] \\ \boldsymbol{\Gamma}_{\mathbf{sx}}^{H}[n|n-1,\eta_l[n]] & \boldsymbol{\Gamma}_{\mathbf{x}}[n|n-1,\eta_l[n]] \end{bmatrix}. \tag{96}$$

Eqs. (94), (95), and (96) provides (62), (63), and (64). Since $\hat{\mathbf{x}}[n \,|\, n-1]$ depends on $\mathcal{X}[n-1]$ only, then from (88) we conclude that the conditional PDF of $\tilde{\mathbf{y}}[n] \,|\, \mathcal{X}[n-1]$ is shifted by $\begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{x}}[n|n-1] \end{bmatrix}$ compared to the conditional PDF of $\mathbf{y}[n] \,|\, \mathcal{X}[n-1]$:

$$\boldsymbol{\mu}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] = \boldsymbol{\mu}_{\mathbf{y}}[n|n-1,\eta_l[n]] - \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{x}}[n|n-1] \end{bmatrix}, \quad (97)$$

$$\boldsymbol{\Gamma}_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] = \boldsymbol{\Gamma}_{\mathbf{y}}[n|n-1,\eta_l[n]], \quad (98)$$

$$\alpha_{\tilde{\mathbf{y}}}[n|n-1,\eta_l[n]] = \alpha_{\mathbf{y}}[n|n-1,\eta_l[n]], \quad (99)$$

and

$$\boldsymbol{\mu}_{\tilde{\mathbf{x}}}[n|n-1,\eta_l[n]] = \boldsymbol{\mu}_{\mathbf{x}}[n|n-1,\eta_l[n]] - \hat{\mathbf{x}}[n|n-1]. \quad (100)$$

Hence the mixture weights and covariance matrices in $\psi_{\mathbf{y}}[n \mid n-1]$ and $\psi_{\tilde{\mathbf{y}}}[n \mid n-1]$ are identical except the means as described in (100). Since the function $\mathbf{G}(\cdot)$ is nonlinear, the parameters of the conditional distribution of $\tilde{\mathbf{y}}[n] \mid \mathcal{X}[n-1]$ cannot be obtained analytically. Alternatively, the MC approach can be implemented. Thus, an artificial data set $\mathcal{D}$ is obtained from the conditional distribution of $\begin{bmatrix} \mathbf{s}[n-1] \\ \mathbf{u}[n] \\ \mathbf{w}[n] \end{bmatrix}$, given $\mathcal{X}[n-1]$. Next, the nonlinear function $\mathbf{G}(\cdot)$ is applied on the data set $\mathcal{D}$ to obtain a new artificial data set $\mathcal{D}'$, which is used to obtain the PDF parameters of $\tilde{\mathbf{y}}[n] \mid \mathcal{X}[n-1]$, i.e. $\psi_{\tilde{\mathbf{y}}}[n \mid n-1]$. The statistical parameters required for calculation of (84), (85) and (86) can be obtained from the parameters of $\psi_{\tilde{\mathbf{y}}}[n \mid n-1]$ in (89). The measurement prediction is calculated as a conditional mean estimator of $\mathbf{x}[n]$ given $\mathcal{X}[n-1]$, using parameters obtained in (95) and (96) as follows

$$\hat{\mathbf{x}}[n|n-1] = \sum_{l=1}^{L} \alpha_{\mathbf{y}}[n|n-1,\eta_l[n]]\boldsymbol{\mu}_{\mathbf{x}}[n|n-1,\eta_l[n]]. \quad (101)$$

The MMSE estimator in (9) is obtained using (85) and (92) as follows:

$$\hat{\mathbf{s}}[n|n] = \sum_{l=1}^{L} \alpha_{\mathbf{s}}[n|n,\eta_l[n]]\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_l[n]]. \quad (102)$$

This completes the derivation of the NL-GMKF. Note that the GMM order of the conditional distribution $\mathbf{y}[n] \mid \mathcal{X}[n-1]$ might be obtained using model order selection algorithms, such as the minimum description length (MDL) [52]. Alternatively, L can be set as an upper bound on the number of mixture components in the conditional PDF. In this work, the vector parameter $\psi_{\tilde{\mathbf{y}}}[n \mid n-1]$ is obtained from the data $\mathcal{D}'$using the greedy EM algorithm [50]. The greedy learning algorithm controls the GMM order of the estimated PDF, which varies over the iterations. In [51] it was shown that the greedy EM algorithm is insensitive to the initialization. The PDF estimation using the greedy EM algorithm appears in [51], [50] and is summarized in the appendix.

4. Simulation results

A. GMKF

In the following scenarios, the estimation performance of the GMKF is evaluated for different linear DSS models, and compared to the IMM, PF and GSPF in terms of root-mean-

square error (RMSE) of the estimate of the first element in the state vector. Additionally, the performance of the GMKF is compared to the GSF with clustering MR algorithm [22]. In these scenarios, the standard sampling importance resampling (SIR) PF was used [1] with 10000 particles. In the scenarios with real-valued Gaussian measurement noise, a common IMM algorithm was used [18]. The number of models corresponds to the number of mixture components in the PDF of the system noise. The GMKF performance is evaluated for various system noise PDFs and it is tested for N = 100 time instances. For estimation performance evaluation, each test was performed over 100 trials. In the first three scenarios, a practical application of maneuvering radar target tracking is addressed. In these scenarios, the state vector is composed of the target range and radial velocity, where the target range is observed by the radar. The DSS and measurement models are given by:

$$\begin{aligned} \begin{bmatrix} r[n] \\ \dot{r}[n] \end{bmatrix} &= \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r[n-1] \\ \dot{r}[n-1] \end{bmatrix} + \mathbf{u}[n] \\ x[n] &= \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} r[n] \\ \dot{r}[n] \end{bmatrix} + w[n] \,, \end{aligned} \tag{103}$$

where T = 0.1 *sec* is the sampling interval. Thus, the state transition matrix and the measurement matrix are given by $\mathbf{A}[n] = \begin{bmatrix} 1 & 0.1 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{H}[n]$ = [1 0], respectively. The conditional distribution of $\mathbf{s}[n] \,|\, \mathcal{X}[n]$ was assumed to be real-valued GMM of order L = 16. The GMKF is initialized at time instance n = –1 with $\alpha_{\mathbf{s}}[-1 \,|\, -1, \eta_l[-1]] = \frac{1}{L}$, $\boldsymbol{\mu}_{\mathbf{s}}[-1 \,|\, -1, \eta_l[-1]] = \mathbf{0}$, $\boldsymbol{\Gamma}_{\mathbf{s}}[-1 \,|\, -1, \eta_l[-1]] = \rho\mathbf{I}$ for l = 1,..., L, where ρ is a large number. For the GSPF, the conditional distribution of the state vector $\mathbf{s}[n]$, given $\mathcal{X}[n]$ was assumed to be GMM of order L = 16, and the number of particles for each Gaussian in the mixture was chosen to be 2000. For the GSF with MR algorithm, the number of mixture components at the output of the MR was selected to be the same as in the GMKF.

A.1 Maneuvering target tracking

The measurement noise, $w[n]$, was assumed to be zero-mean Gaussian $\boldsymbol{\theta}_w = \{\mu_w = 0,\ \sigma_w^2 = 0.1\}$. The target maneuvers (timevarying acceleration) are simulated by white-noise acceleration [61]. This model is commonly used for small, random maneuvers which are usually modeled by zero-mean Gaussian system noise with time-varying variance: $\mathbf{u}[n] \sim \mathcal{N}(\mathbf{0},\ \beta[n]\mathbf{I})$, $\forall n$ = 0, 1, 2,..., where $\beta[n]$ controls the target maneuvering. In the first part of the simulation (n = 0,..., 21) this parameter was $\beta[n]$ = 0.01 to represent approximately constant-acceleration. In the second part of the simulation n = 22,..., 82, the target's time-varying acceleration was simulated by a random vector with "wider" Gaussian, defined by $\beta[n]$ = 1. For the GMKF, the GSPF and the GSF with MR, the system noise was modeled by GMM with two mixture components:

$$f_{\mathbf{u}}(\mathbf{u}) = (1-\alpha_{\mathbf{u}})\Phi(\mathbf{u}, \boldsymbol{\theta}_{\mathbf{u}1}) + \alpha_{\mathbf{u}}\Phi(\mathbf{u}, \boldsymbol{\theta}_{\mathbf{u}2}) \,, \quad \forall n = 0, 1, 2, \ldots$$

where $\alpha_{\mathbf{u}}$ = 0.2. For the PF, the system noise is modeled by a single Gaussian with the following first and second order statistics:

$$\begin{aligned} \bar{\boldsymbol{\mu}}_{\mathbf{u}} &= (1-\alpha_{\mathbf{u}})\boldsymbol{\mu}_{\mathbf{u}1} + \alpha_{\mathbf{u}}\boldsymbol{\mu}_{\mathbf{u}2} \,, \\ \bar{\boldsymbol{\Gamma}}_{\mathbf{u}} &= (1-\alpha_{\mathbf{u}})\boldsymbol{\Gamma}_{\mathbf{u}1} + \alpha_{\mathbf{u}}\boldsymbol{\Gamma}_{\mathbf{u}2} + (1-\alpha_{\mathbf{u}})\boldsymbol{\mu}_{\mathbf{u}1}\boldsymbol{\mu}_{\mathbf{u}1}^T + \alpha_{\mathbf{u}}\boldsymbol{\mu}_{\mathbf{u}2}\boldsymbol{\mu}_{\mathbf{u}2}^T - \bar{\boldsymbol{\mu}}_{\mathbf{u}}\bar{\boldsymbol{\mu}}_{\mathbf{u}}^T \,. \end{aligned} \tag{104}$$

The IMM consisted of two KFs, which correspond to the two modes characterized by the following system and measurement noise parameters:

- Mode 1: non-maneuvering - $\boldsymbol{\theta}_{\mathbf{u}1}$, $\boldsymbol{\theta}_w$ = {0, 0.1},
- Mode 2: maneuvering - $\boldsymbol{\theta}_{\mathbf{u}2}$, $\boldsymbol{\theta}_w$ = {0, 0.1}.

The transition probability matrix between the two models was selected to match the transition probabilities to the mean sojourn time in each mode [18]:

$$\pi_1 = \begin{bmatrix} 0.95 & 0.05 \\ 0.02 & 0.98 \end{bmatrix}, \tag{105}$$

which corresponds to 20 samples in first and 60 in a second mode. Note that the estimation performances of the IMM were weakly sensitive to small changes of the transition probabilities. The range and velocity estimate RMSE of GMKF, IMM, PF, GSPF, and MR are presented in Fig. 5. This figure shows that in the first part of the simulation (non-maneuvering target), the estimation RMSEs of all the tested methods are close. However, beyond the switching point (maneuvering target), the GMKF outperforms the other tested algorithms.

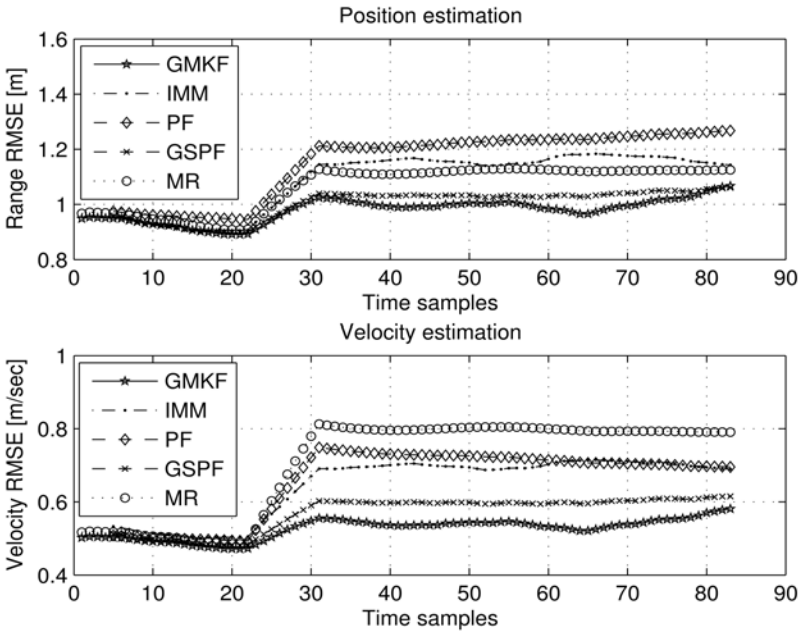

Fig. 5. GMKF vs. IMM, PF, GSPF, and MR for maneuvering target tracking.

A.2 Comparison with the GSF

The estimation performance of the GMKF is compared to the GSF (GMKF without the order reduction stage), which is optimal (but computationally impractical) in the MMSE sense. The measurement noise, $w[n]$, was assumed to be real-valued zero-mean Gaussian with variance σ_w^2 = 0.1. The target maneuvers are modeled by GMM with two mixture components:

$$f_{\mathbf{u}}(\mathbf{u}) = (1-\alpha_{\mathbf{u}})\Phi(\mathbf{u},\boldsymbol{\theta}_{\mathbf{u}1}) + \alpha_{\mathbf{u}}\Phi(\mathbf{u},\boldsymbol{\theta}_{\mathbf{u}2}) \ , \ \ \forall n = 0,1,2,\ldots$$

where

$$\boldsymbol{\theta}_{\mathbf{u}1} = \{\boldsymbol{\mu}_{\mathbf{u}1} = \mathbf{0}, \boldsymbol{\Gamma}_{\mathbf{u}1} = diag\,([0 \ \ 0.01])\} \ , \ \boldsymbol{\theta}_{\mathbf{u}2} = \{\boldsymbol{\mu}_{\mathbf{u}2} = \mathbf{0}, \boldsymbol{\Gamma}_{\mathbf{u}2} = diag\,([0 \ \ 1])\} \ .$$

and $\alpha_{\mathbf{u}}$ = 0.2. As it was shown in Section III-A, the GSF is optimal with respect to the MMSE, but it is impractical due to the model order growth. Therefore, the estimation RMSE of the GSF can be interpreted as a lower bound for the RMSE of GMKF. The range estimation performance of the GMKF and the GSF are shown in Fig. 6 for maneuvering target tracking. The GMKF performance is evaluated for N = 13 only, because longer runs are computationally impractical due to exponential model order growth. This figure shows that the GMKF performance is close to the lower bound provided by the GSF which is the exact MMSE.

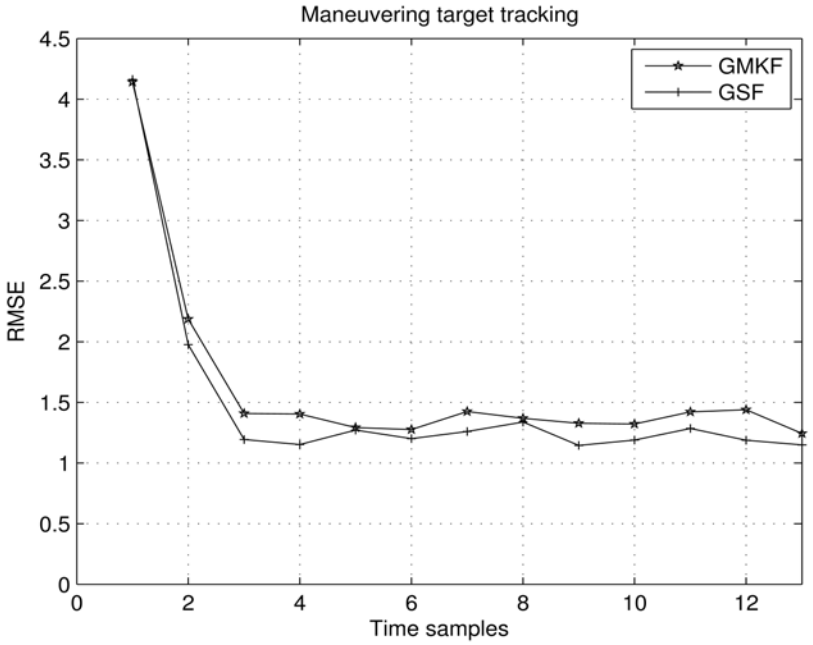

Fig. 6. GMKF vs. GSF for maneuvering target tracking.

A.3 Non-Gaussian glint measurement noise

The Middleton class-A model is widely used for glint measurement noise [3], [62]. This model consists of one Gaussian with high probability of occurrence and small variance and another with small probability of occurrence and very high variance:

$$f_{w[n]}(w) = \alpha_w \Phi(w,\boldsymbol{\theta}_{w_1}) + (1-\alpha_w)\Phi(w,\boldsymbol{\theta}_{w_2}) \ , \ n = 0,\ldots,N-1 \ ,$$

where α_w = 0.9, $\boldsymbol{\theta}_{w1}$ = {μ_{w1} = 0, σ^2_{w1} = 0.01}, $\boldsymbol{\theta}_{w2}$= {μ_{w2} = 0, σ^2_{w2} = 1}. For the PF, the measurement noise is modeled by a single Gaussian with the following first and second order statistics:

$$\begin{aligned} \bar{\mu}_w &= (1-\alpha_w)\mu_{w1} + \alpha_w \mu_{w2} \ , \\ \bar{\sigma}_w^2 &= (1-\alpha_w)\sigma_{w1}^2 + \alpha_w \sigma_{w2}^2 + (1-\alpha_w)\mu_{w1}\mu_{w1}^T + \alpha_w \mu_{w2}\mu_{w2}^T - \bar{\mu}_w \bar{\mu}_w^T \ , \end{aligned} \tag{106}$$

The IMM [20] with 2 modes was used in this scenario. These modes are characterized by the following measurement and system noises parameters:

- Mode 1: no glint: $\boldsymbol{\theta}_{\mathbf{u}} = \{\mathbf{0}, diag\,([1\ 1])\}\,, \boldsymbol{\theta}_{w_1} = \left\{\mu_{w1} = 0, \sigma_{w1}^2 = 0.01\right\}$,
- Mode 2: glint : $\boldsymbol{\theta}_{\mathbf{u}} = \{\mathbf{0}, diag\,([1\ 1])\}\,, \boldsymbol{\theta}_{w_2} = \left\{\mu_{w2} = 0\ , \sigma_{w2}^2 = 1\right\}$.

In this scenario, it was assumed that the glint is time-independent. Therefore, the transition probability matrix between the two models is [20]:

$$\pi_2 = \begin{bmatrix} (1-\alpha_w) & \alpha_w \\ (1-\alpha_w) & \alpha_w \end{bmatrix} . \tag{107}$$

The estimation performances of GMKF, IMM, PF, GSPF, and MR are presented in Fig. 7. This figure shows that the GMKF outperforms the other tested algorithms.

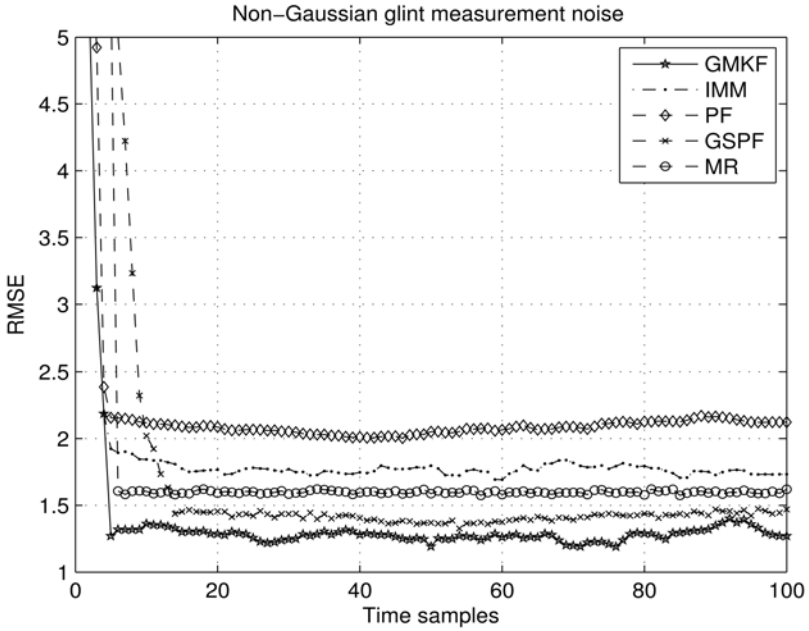

Fig. 7. GMKF vs. IMM, PF, GSPF, and MR for non-Gaussian glint measurement noise.

A.4 GMKF sensitivity

The main motivation of this test is to show that the estimation performance of the GMKF is weakly sensitive to the assumed GMM order of the posterior distribution. In addition, the sensitivity of the GMKF performance to the size of the artificial data used in the model order reduction algorithm, is tested. The DSS model, used in the previous scenario was adopted here. It was assumed that the elements of the system noise $\mathbf{u}[n] = [u_1[n]\ u_2[n]]^T$ are independent where $u_1[n]$ and $u_2[n]$ are Laplacian-distributed with marginal PDFs

$f_{u_i[n]}(u) = \frac{1}{2b_i} e^{-b_i|u|}$, n = 0,...,N –1, i = 1, 2, where b_1 = 0.3 and b_2 = 0.4. The measurement noise, $w[n]$, was assumed to be zero-mean Gaussian with variance σ_w^2 = 0.1. The Laplacian PDF of the system noise vector is approximated by a GMM of order K = 16. Fig. 8 shows that for any number of Gaussians in the mixture, the GMKF estimation performance improves as the size of the artificial data in the EM stage increases. It can be observed that the estimation performance improves to a turning point, from which the estimation performance degrades. The reason for the estimation performance degradation is the overfitting problem [63]. The threshold for the number of the components in the mixture increases with the size of the artificial data used in the EM stage. This figure shows that only a minor improvement in the estimation performance is achieved with increase of the GMM order.

B. NL-GMKF

In this section, the NL-GMKF performance is evaluated using the following nonlinear DSS model with non-Gaussian driving and measurement noise distributions.

Maneuvering target tracking model

Maneuvering target tracking involves modeling of a maneuver control signal, which is unknown to the tracker. Typically, the control signal is modeled as a random process [18], [64]. Most maneuver models assume that target maneuvering is uncoupled in Cartesian coordinates. The well-known Singer model [65] assumes that the target acceleration is a zero-mean first-order Markov process. In this example, a two-dimensional target tracking problem is addressed. The two-dimensional time-varying target dynamics can be described by the system state vector, which consists of the target position $[r_x\, r_y]^T$, velocity $[\dot{r}_x\ \dot{r}_y]^T$, and acceleration $[\ddot{r}_x\ \ddot{r}_y]^T$:

$$\mathbf{s}[n] = [r_x[n],\ r_y[n],\ \dot{r}_x[n],\ \dot{r}_y[n],\ \ddot{r}_x[n],\ \ddot{r}_y[n]]^T \ .$$

The discrete-time Singer model is described by the linear transition function:

$$\mathbf{a}(\mathbf{s}[n-1], \mathbf{u}[n]) = \begin{bmatrix} 1 & 0 & T & 0 & c_1 & 0 \\ 0 & 1 & 0 & T & 0 & c_1 \\ 0 & 0 & 1 & 0 & c_2 & 0 \\ 0 & 0 & 0 & 1 & 0 & c_2 \\ 0 & 0 & 0 & 0 & e^{-\delta T} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-\delta T} \end{bmatrix} \mathbf{s}[n-1] + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{u}[n] \ ,$$

where

$$c_1 = \frac{T}{\delta} - \frac{1 - e^{-\delta T}}{\delta^2}$$
$$c_2 = \frac{1 - e^{-\delta T}}{\delta} \ ,$$

T is the sampling interval, and $\delta = \frac{1}{\tau}$ is reciprocal of the maneuver time constant τ , that describes the maneuver duration. Note that this model can describe both maneuvering and nonmaneuvering motions. Thus, as maneuver time decreases, the Singer model reduces to the constant velocity (CV) model, and as the maneuver time increases, the Singer model reduces to the constant acceleration (CA) model. According to the Singer model, the target position change is determined by its velocity, the target velocity change is determined by its

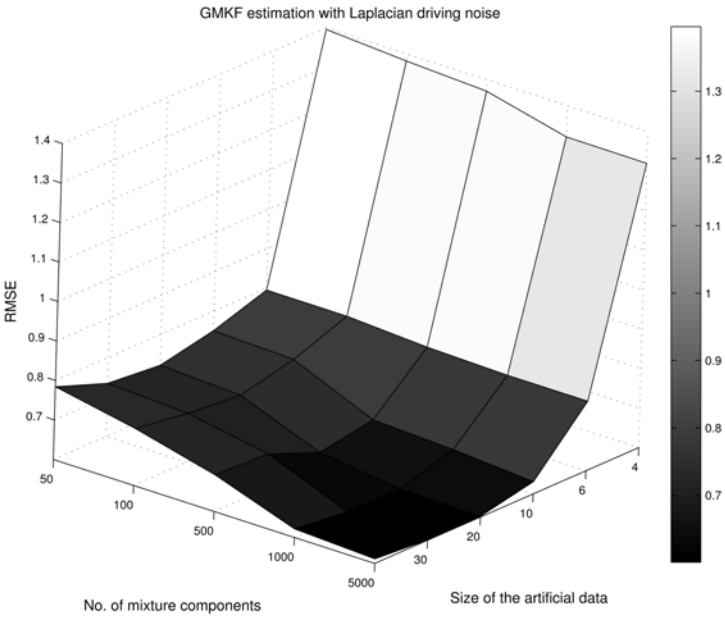

Fig. 8. GMKF sensitivity testing.

acceleration, and the acceleration change is driven by the system noise. Therefore, an alternative way to model target maneuver is to model abrupt changes of target acceleration that corresponds to increased variance of the system noise.

Observation model

The measurements from a typical two-dimensional radar consist of the target range and bearing. Therefore, the measurement equation is nonlinear. Assuming that the radar is placed at the origin (x, y) = (0, 0), the radar measurements: range, $r[n]$, and bearing, $\beta[n]$, of the target are described by the measurement function

$$\mathbf{h}\left(\mathbf{s}[n],\ \mathbf{w}[n]\right) = \begin{bmatrix} \left(r_x^2[n] + r_y^2[n]\right)^{\frac{1}{2}} \\ \arctan\left(\frac{r_y[n]}{r_x[n]}\right) \end{bmatrix} + \mathbf{w}[n]\ ,$$

where $\mathbf{w}[n]$ is a zero-mean additive noise.

Glint noise model

The mixture approach is widely used in modeling the non-Gaussian glint noise. The glint is Gaussian-like around the mean and has a non-Gaussian, long-tailed nature in the tail region [3]. The data at the tail region represent outliers caused by the glint spikes. The outliers with low occurrence probability have a significant inuence on the conventional target tracking algorithms, such as the KF. In [3], [16] and [36], the glint noise was modeled as a mixture of two zero-mean Gaussians, where the outliers were represented by a zero-mean Gaussian with large (comparing to the thermal noise) covariance matrix. In [6] and [20], it was proposed to model the heavy-tailed distribution of the glint noise for each measurement component as a mixture of a zero-mean Gaussian noise with high occurrence probability and a Laplacian noise with low probability of occurrence:

$$f_w(w) = (1-\alpha_w)\Phi(w, \boldsymbol{\theta}_{w1}) + \alpha_w \mathcal{L}(w, \boldsymbol{\theta}_{w2}) \,, \tag{108}$$

where α_w is the glint probability, and $\mathcal{L}(w, \boldsymbol{\theta}_{w2})$ is the Laplacian PDF in which $\boldsymbol{\theta}_{w2}$ contains the mean and the variance. The estimation performance of the NL-GMKF was compared to PF, IMM-EKF, UKF, and GSPF. The performances were tested for N = 150 time instances with sampling interval T = 1 sec. The following target tracking cases were tested:

1. Nonmaneuvering target tracking in the presence of glint measurement noise
2. Maneuvering target tracking with Gaussian measurement noise
3. Maneuvering target tracking in the presence of glint measurement noise
4. Maneuvering target tracking in the presence of glint measurement noise, whose statistics depend on the target maneuver
5. Coordinated turn (CT) maneuvering target tracking in the presence of Gaussian or glint measurement noise.

Table 1 shows time-varying statistical parameters of the system and measurement noises for the first four scenarios. In these scenarios, the target maneuver was modeled by Gaussian-distributed system noise with time-varying variance and the following parameters:

- Stages 1 and 3: $\boldsymbol{\theta}_{\mathbf{u}1} = \{\boldsymbol{\mu}_{\mathbf{u}1}, \Gamma_{\mathbf{u}1}\}$, where $\boldsymbol{\mu}_{\mathbf{u}1} = \mathbf{0}$, $\Gamma_{\mathbf{u}1}^{1/2} = diag\left([1m/sec^2, 1m/sec^2]\right)$.
- Stage 2: $\boldsymbol{\theta}_{\mathbf{u}2} = \{\boldsymbol{\mu}_{\mathbf{u}2}, \Gamma_{\mathbf{u}2}\}$, where $\boldsymbol{\mu}_{\mathbf{u}2} = \mathbf{0}$, $\Gamma_{\mathbf{u}2} = \xi^2 \Gamma_{\mathbf{u}1}$.

Scenario	Noises	Stage 1 $1 \le n \le 50$	Stage 2 $51 \le n \le 100$	Stage 3 $101 \le n \le 150$
1	System Measurement	$\boldsymbol{\theta}_{\mathbf{u}1}$ $\boldsymbol{\theta}_{\mathbf{w}1}, \boldsymbol{\theta}_{\mathbf{w}2}$ $\psi = 5, \alpha_{\mathbf{w}} = 0.3$	$\boldsymbol{\theta}_{\mathbf{u}1},$ $\boldsymbol{\theta}_{\mathbf{w}1}, \boldsymbol{\theta}_{\mathbf{w}2},$ $\psi = 5, \alpha_{\mathbf{w}} = 0.3$	$\boldsymbol{\theta}_{\mathbf{u}1}$ $\boldsymbol{\theta}_{\mathbf{w}1}, \boldsymbol{\theta}_{\mathbf{w}2}$ $\psi = 5, \alpha_{\mathbf{w}} = 0.3$
2	System Measurement	$\boldsymbol{\theta}_{\mathbf{u}1}$ $\boldsymbol{\theta}_{\mathbf{w}1}, \alpha_{\mathbf{w}} = 0$	$\boldsymbol{\theta}_{\mathbf{u}2}, \xi = 10$ $\boldsymbol{\theta}_{\mathbf{w}1}, \alpha_{\mathbf{w}} = 0$	$\boldsymbol{\theta}_{\mathbf{u}1}$ $\boldsymbol{\theta}_{\mathbf{w}1}, \alpha_{\mathbf{w}} = 0$
3	System Measurement	$\boldsymbol{\theta}_{\mathbf{u}1}$ $\boldsymbol{\theta}_{\mathbf{w}1}, \boldsymbol{\theta}_{\mathbf{w}2}$ $\psi = 5, \alpha_{\mathbf{w}} = 0.3$	$\boldsymbol{\theta}_{\mathbf{u}2}, \xi = 10$ $\boldsymbol{\theta}_{\mathbf{w}1}, \boldsymbol{\theta}_{\mathbf{w}2}$ $\psi = 5, \alpha_{\mathbf{w}} = 0.3$	$\boldsymbol{\theta}_{\mathbf{u}1}$ $\boldsymbol{\theta}_{\mathbf{w}1}, \boldsymbol{\theta}_{\mathbf{w}2}$ $\psi = 5, \alpha_{\mathbf{w}} = 0.3$
4	System Measurement	$\boldsymbol{\theta}_{\mathbf{u}1}$ $\boldsymbol{\theta}_{\mathbf{w}1}, \alpha_{\mathbf{w}} = 0$	$\boldsymbol{\theta}_{\mathbf{u}2}, \xi = 10$ $\boldsymbol{\theta}_{\mathbf{w}1}, \boldsymbol{\theta}_{\mathbf{w}2},$ $\psi = 5, \alpha_{\mathbf{w}} = 0.3$	$\boldsymbol{\theta}_{\mathbf{u}1}$ $\boldsymbol{\theta}_{\mathbf{w}1}, \alpha_{\mathbf{w}} = 0$

Table 1.Parameters of the system and measurement noise for scenarios 1-4

The following parameters for the measurement noise in (108) were selected in all the scenarios: $\boldsymbol{\theta}_{\mathbf{w}i} = \{\boldsymbol{\mu}_{\mathbf{w}i}, \boldsymbol{\Gamma}_{\mathbf{w}i}\}$, i = 1, 2, where $\boldsymbol{\mu}_{\mathbf{w}1} = \boldsymbol{\mu}_{\mathbf{w}2} = 0$, $\Gamma_{\mathbf{w}1}^{1/2} = diag\,([10m, 0.5mrad])$, and $\boldsymbol{\Gamma}_{\mathbf{w}2} = \psi^2 \boldsymbol{\Gamma}_{\mathbf{w}1}$. In the first two scenarios, target tracking performances were evaluated for a variety of glint noise statistics and maneuvering levels, respectively. In the first scenario, tracking performances of the different methods are evaluated for different probabilities of glint, $\alpha_{\mathbf{w}}$, and different glint noise levels, ψ in the range of [1, 10]. In the second scenario, maneuver cases [18], [31], starting from low maneuvering target (commercial aircraft) to extremely highly maneuvering target, modeled by different values of ξ in the range [1, 100], were tested. Thus, ξ = 1 models extremely low acceleration standard deviation (STD) of about 0.1g, and ξ = 100 models extremely high acceleration STD of about 10g. In all the scenarios, the initial target position, velocity and acceleration were as follows:

$$\mathbf{s}[-1] \quad = \quad \left[5km,\ 5km,\ -100m/sec,\ 0m/sec,\ 0m/sec^2,\ 0m/sec^2\right]^T \ .$$

For performance evaluation, the RMSE of the two-dimensional target position $\mathbf{r}[n] = [r_x[n]\ r_y[n]]^T$ and velocity $\dot{\mathbf{r}}[n] = [\dot{r}_x[n]\ \dot{\mathbf{r}}_y[n]]^T$ estimates were evaluated. The mean RMSE of the range and velocity estimation, defined as

$$||\hat{\mathbf{r}}[n] - \mathbf{r}[n]|| = \left((\hat{r}_x[n] - r_x[n])^2 + (\hat{r}_y[n] - r_y[n])^2\right)^{\frac{1}{2}} ,$$
$$||\hat{\dot{\mathbf{r}}}[n] - \dot{\mathbf{r}}[n]|| = \left((\hat{\dot{r}}_x[n] - \dot{r}_x[n])^2 + (\hat{\dot{r}}_y[n] - \dot{r}_y[n])^2\right)^{\frac{1}{2}} ,$$

are evaluated in the following scenarios using 100 trials. In all the tests, the NL-GMKF was initialized at time instance $n = -1$ with

$$\begin{aligned} \alpha_{\mathbf{s}l}[-1] &= \frac{1}{L} , \\ \boldsymbol{\mu}_{\mathbf{s}l}[-1] &= \mathbf{0} , \\ \boldsymbol{\Gamma}_{\mathbf{s}l}[-1] &= \rho\mathbf{I} , \forall l = 1, \ldots, L \end{aligned}$$

where ρ is a large number, and $\mathbf{I}$ is an identity matrix. For the NL-GMKF and the GSPF, the conditional distribution of the state vector $\mathbf{s}[n]$, given $\mathcal{X}[n]$ was assumed to be GMM of order $L = 16$. For the GSPF, the number of particles for each Gaussian in the mixture was chosen to be 400 (totally 6400 particles). Addressing the nonlinear problem, the GSPF was implemented using first-order linearization in the time-update stage [35]. In the following tests, the standard sampling importance resampling (SIR) PF with 10000 particles was implemented [1]. For the PF, UKF and the GSPF, the first and second order statistics of the state vector $\mathbf{s}[n]$, given $\mathcal{X}[n]$ were initialized with

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{s}}[-1|-1] &= \mathbf{0} , \\ \boldsymbol{\Gamma}_{\mathbf{s}}[-1|-1] &= \rho\mathbf{I} . \end{aligned}$$

In the UKF [55], the parameter that determines the spread of the sigma points was set to $\alpha = 0.95$. The scaling parameter, k, was set to 0, and the parameter β, used to incorporate prior knowledge of the distribution [55], was set to 2.

B.1 Nonmaneuvering target tracking in the presence of glint measurement noise

The tracking performance of the NL-GMKF in the presence of non-Gaussian glint noise is tested in this scenario. A nonmaneuvering target was considered in order to evaluate the inuence of the non-Gaussian measurement noise on the tracking performance. In practice, this situation might occur when glint noise arises due to environmental factors such as turbulence and vibration and not due to target maneuver [53], [54]. Tracking performances of the tested algorithms are evaluated for various levels of glint noise covariance and its probabilities. In the NL-GMKF, the system noise was assumed to be Gaussian with parameters $\boldsymbol{\theta}_{\mathbf{u}1}$. The measurement noise was modeled by the mixture in (108), with parameters defined in Table 1. The IMM consisted of two EKFs, which correspond to the two modes characterized by the following system and measurement noise parameters:

- Mode 1: no glint: $\boldsymbol{\theta}_{\mathbf{u}1}$, $\boldsymbol{\theta}_{\mathbf{w}1}$,
- Mode 2: glint: $\boldsymbol{\theta}_{\mathbf{u}1}$, $\boldsymbol{\theta}_{\mathbf{w}2}$.

In this scenario, it assumed that the glint is time-independent. Therefore, the transition probability matrix between the two models is [20]

$$\pi_2 = \begin{bmatrix} (1-\alpha_{\mathbf{w}}) & \alpha_{\mathbf{w}} \\ (1-\alpha_{\mathbf{w}}) & \alpha_{\mathbf{w}} \end{bmatrix} . \tag{109}$$

The tracking performances of the NL-GMKF, PF, IMM-EKF, UKF and GSPF in terms of mean RMSE, are shown in Fig. 9 for the two-dimensional position and velocity. It can be observed that the NL-GMKF outperforms the other tested algorithms. The tracking performances as a function of the glint probability, $\alpha_{\mathbf{w}}$, and glint level, ψ, at time instance n = 75 are shown in Figs. 10 and 11, respectively. It can be observed that the NL-GMKF outperforms the other tested algorithms for the tested probabilities of glint and the tested glint levels. Note that the estimation performance of the NL-GMKF remains almost constant across the entire range of the tested probability of glint and glint level.

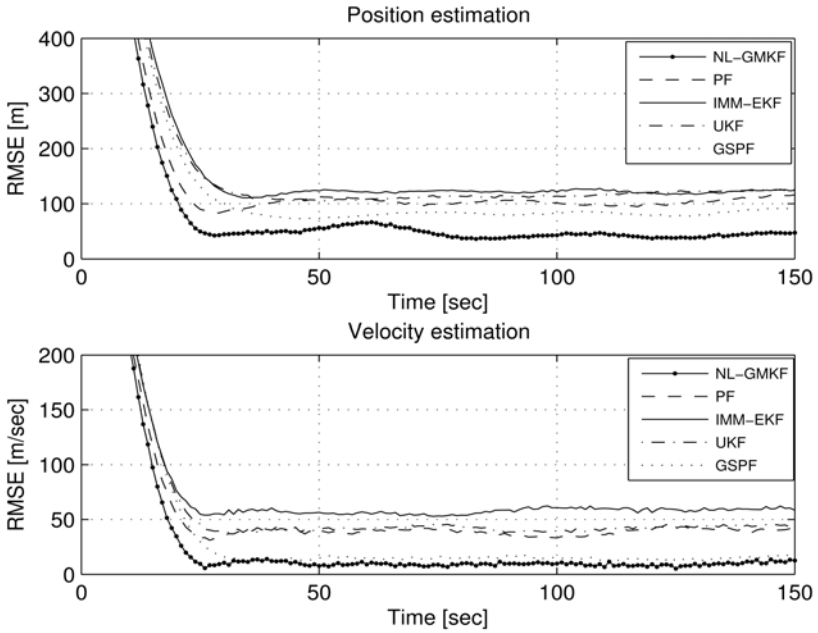

Fig. 9. Tracking performance of NL-GMKF, PF, IMM-EKF, UKF and GSPF in the presence of glint noise with probability of $\alpha_{\mathbf{w}}$ = 0.3, and glint noise level of ψ= 5.

B.2 Maneuvering target tracking

In this scenario, target tracking performances of the NL-GMKF without glint noise are evaluated for a wide range of maneuvers, from low, hardly detectable, to very high. The glint produced by small targets at long distances may be negligible and target tracking errors arises solely due to target maneuver. The tracking performance of the NL-GMKF is tested in a wide range of maneuvers in order to test its ability to track highly maneuverable targets as well as slow maneuverable targets with low probability of maneuver detection. In this example, the maneuvering target was simulated during the second interval of the simulation. In the NL-GMKF, the system noise statistics were modeled by the mixture of two components:

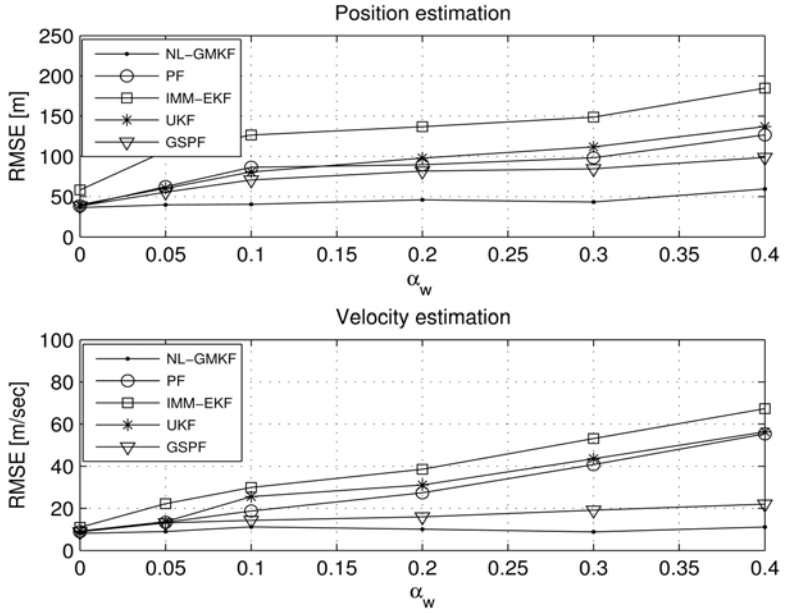

Fig. 10. Tracking performance of NL-GMKF, PF, IMM-EKF, UKF and GSPF vs. glint noise probability with glint noise level $\psi = 5$ at time instance $n = 75$.

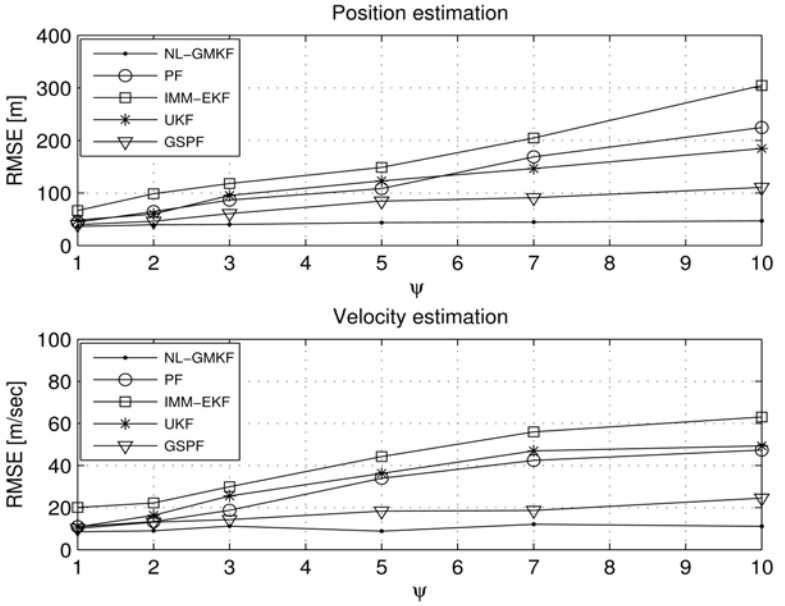

Fig. 11. Tracking performance of NL-GMKF, PF, IMM-EKF, UKF and GSPF as a function of glint noise level with probability of glint $\alpha_w = 0.3$ at time instance $n = 75$.

$$f_{\mathbf{u}}(\mathbf{u}) = (1-\alpha_{\mathbf{u}})\Phi(\mathbf{u},\boldsymbol{\theta}_{\mathbf{u}1}) + \alpha_{\mathbf{u}}\Phi(\mathbf{u},\boldsymbol{\theta}_{\mathbf{u}2}) \ , \quad \forall n = 0,1,2,\dots \ , \tag{110}$$

and the probability of target maneuver was selected to be $\alpha_{\mathbf{u}} = 0.2$. The measurement noise model was assumed to be Gaussian with parameters $\boldsymbol{\theta}_{\mathbf{w}1}$. The IMM consisted of two EKFs, which correspond to the two modes characterized by the following system and measurement noise parameters:

- Mode 1: nonmaneuvering - $\boldsymbol{\theta}_{\mathbf{u}1}$, $\boldsymbol{\theta}_{\mathbf{w}1}$,
- Mode 2: maneuvering - $\boldsymbol{\theta}_{\mathbf{u}2}$, $\boldsymbol{\theta}_{\mathbf{w}1}$.

The first model represents the target motion in the first and third parts of the simulation, and the second represents the target motion in the second part. The transition probability matrix between the two models was selected to match the transition probabilities to the mean sojourn time in each mode [18]:

$$\pi_1 = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix} , \tag{111}$$

which corresponds to 50 samples at each mode. The estimation performance of the IMM was found to be weakly sensitive to small changes of the transition probabilities. The tracking performances of the NL-GMKF, PF, IMM-EKF, UKF, and GSPF in terms of mean RMSE are presented in Fig. 12. This figure shows the mean RMSE of the target position and velocity estimation. It can be observed that at the first and the third parts of the simulation (nonmaneuvering target), the estimation performances of the tested algorithms are close. However, at the second part of the simulation (maneuvering target), the NL-GMKF outperforms the other tested algorithms. It is expected that the estimation errors in tracking highly maneuvering targets would be larger. However, high maneuvers can be easily detected and the estimation errors can be significantly reduced using a proper model. A more challenging scenario arises when tracking slow-maneuvering targets with low probability of maneuver detection. In this case, large errors may be due to mismatch in the model. These situations can be modeled by small covariance matrices of the system noise representing the maneuver, $\boldsymbol{\Gamma}_{\mathbf{u}2}$, which is determined by ξ. The tracking performance of the proposed NL-GMKF at time instance $n = 75$ as a function of the maneuvering level, ξ, is tested and presented in Fig. 13 for accelerations in the range $0.1g$ – $10g$, which covers the entire range of maneuvering aircrafts [7]. This figure shows that the NL-GMKF outperforms the other tested algorithms for all tested maneuvering levels. This figure also shows that the performance of the NL-GMKF is almost constant for all tested maneuvering levels while the tracking performances of other tested algorithms degrade for higher maneuvering levels.

B.3 Maneuvering target tracking in the presence of glint measurement noise

In this example, maneuvering target tracking in the presence of glint noise was tested. It is assumed that the glint noise arises due to environmental factors such as turbulence and vibration and not due to target maneuver per se, and thus it is present during the entire simulation and its statistics are independent of the target maneuver. This scenario occurs in tracking close targets (such as a group of aircrafts) [20], [53], [54] and tests the effect of the glint noise on tracking the maneuvering target with various levels of maneuverability. The target maneuvering during the second interval, modeled by zero-mean Gaussian system noise with time-varying statistics described in scenario 2 with $\xi = 10$, was considered in this example. In the NL-GMKF, the statistics of the system noise used in scenario 2 was adopted.

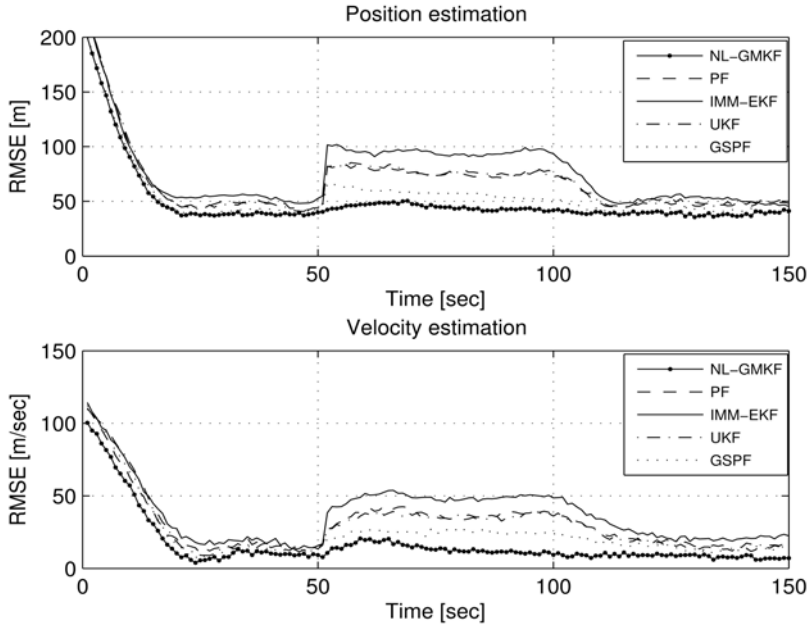

Fig. 12. Tracking performance of NL-GMKF, PF, IMM-EKF, UKF and GSPF in the presence of maneuvering during $n \in [50, 100]$ with $\xi = 10$.

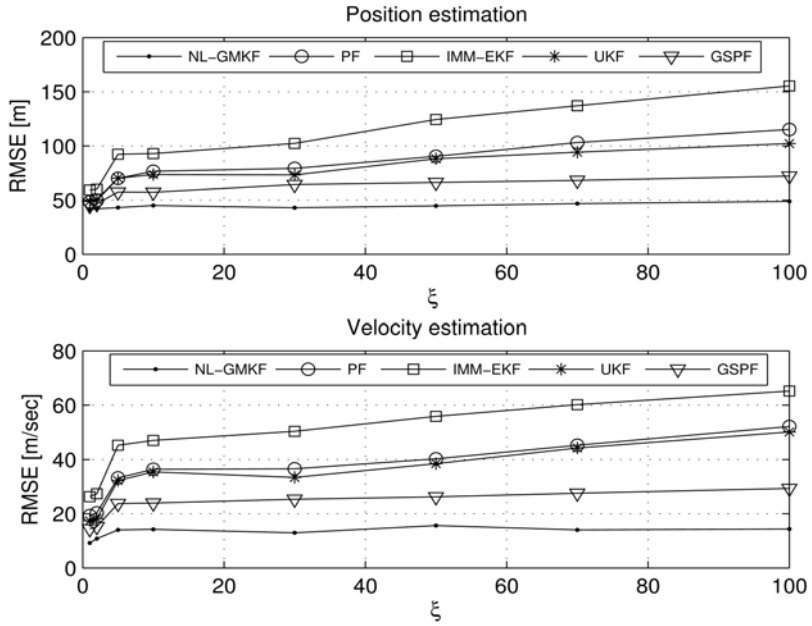

Fig. 13. Tracking performance of NL-GMKF, PF, IMM-EKF, UKF, and GSPF in the presence of maneuvering target for various levels of maneuver at time instance $n = 75$.

The measurement noise was modeled by the mixture in (108), with parameters defined in Table 1. The IMM-EKF algorithm for maneuvering target in the presence of glint noise was implemented using an efficient layered implementation [31]. According to this scheme, two sets of modes are used. One corresponds to the presence or absence of target maneuvering and the other to the presence or absence of glint. Therefore, the IMM is implemented with four modes with transition matrices, defined in (107), (111) according to [31]. The parameters of the measurement and system noise were identical to those defined in scenarios 1 and 2, respectively. The IMM consisted of four EKFs, which correspond to the four modes characterized by the following system and measurement noise parameters:

- Mode 1: nonmaneuvering and no glint - θ_{u1}, θ_{w1},
- Mode 2: maneuvering and no glint - θ_{u2}, θ_{w1},
- Mode 3: nonmaneuvering and glint - θ_{u1}, θ_{w2},
- Mode 4: maneuvering and glint - θ_{u2}, θ_{w2}.

The tracking performances of the NL-GMKF, PF, IMM-EKF, UKF and GSPF in terms of mean RMSE for a maneuvering target with glint are shown in Fig. 14. It can be observed that the NL-GMKF outperforms the other tested algorithms during the entire simulation interval.

B.4 Maneuvering target tracking in the presence of correlated statistics of glint noise

In this example, a scenario of great practical interest in which the statistics of the glint noise and the target maneuver are correlated, was tested. In target tracking, changes in the target aspect with respect to the radar due to maneuver dramatically increases the radar cross section uctuations resulting in significant glint noise [20], [53]-[57]. Therefore, glint noise increases dramatically during the maneuver. This scenario is modeled by correlated time-variations of the statistics of the measurement noise and the system noise. In this example, it was assumed that there is no glint noise during nonmaneuvering parts of the simulation (first and third parts), while it is present during the maneuvering part of the simulation (second part). The models for the tested tracking algorithms presented in the previous scenario, are adopted here too. The tracking performances of NLGMKF, PF, IMM-EKF, UKF, and GSPF of the maneuvering target with glint, are shown in Fig. 15 in terms of mean RMSE of the two-dimensional position and velocity. It can be observed that the NL-GMKF outperforms the other tested algorithms during the entire simulation interval. Note that in contrast to other tested scenarios, the performance of the NL-GMKF degrades at the switching point (comparing to almost constant performance obtained in the previous examples in Figs. 9, 12, and 14). This behavior occurs due to simultaneous changes in the system and measurement noise statistics and the difficulty to associate those changes, expressed by the innovation vector statistics, with glint noise or target maneuver. Fig. 16 shows the tracking performances for various combinations of maneuvers and glint noise levels. One can notice that the tracking performance of the NL-GMKF does not monotonically decrease as a function of ξ or ψ. This phenomenon can be explained by the difficulties of the NL-GMKF to associate the increase in the innovation process variance to glint or to target maneuver. Note that Fig. 16 is in conformance with results in scenarios 1 and 2. The first column in Fig. 16 (ξ = 1) corresponds to the NL-GMKF performance in the non-maneuvering scenario shown in Fig. 11, and the first row in Fig. 16 (ψ= 1) corresponds to the NLGMKF performance in the scenario without glint noise shown in Fig. 13. The NL-GMKF provides higher performance than the other tested algorithms, since it employs prior knowledge on the non-Gaussian PDF of system and measurement noise by approximating

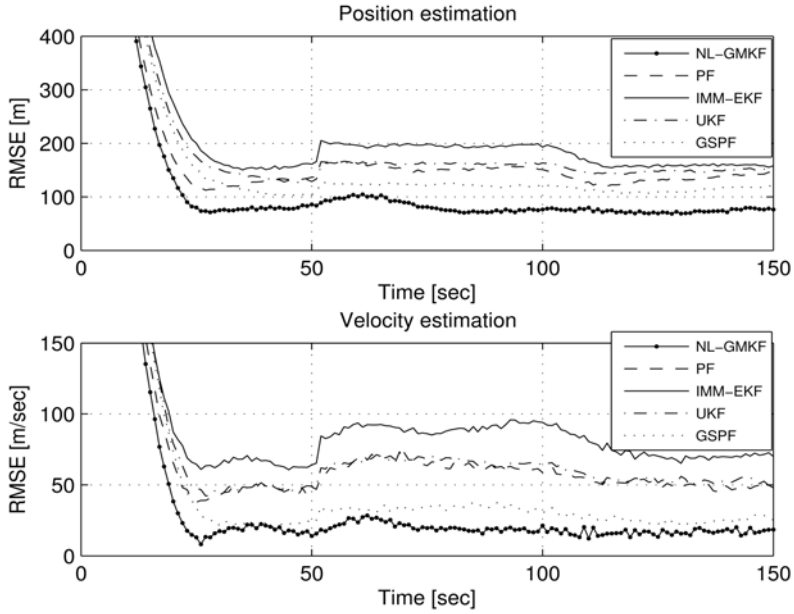

Fig. 14. Tracking performance of NL-GMKF, PF, IMM-EKF, UKF and GSPF in the presence of maneuvering target during $n \in [50, 100]$ with $\xi = 10$, and glint noise probability $\alpha_w = 0.3$ with glint noise level $\psi = 5$.

the non-Gaussian system state PDF by GMM and it does not attempt to estimate their parameters from the data. Therefore, performance degradation of the NL-GMKF in maneuvering or glint noise scenarios is due to reduction of information carried by the system model or measurements during maneuvering or glint samples, while in the other tested algorithms, it is due to modeling mismatch.

B.5 Coordinated turn

In this example, we consider a scenario in which the target maneuver is modeled by motion dynamics and not by increased system noise level. In air traffic control applications, the motion of the civil aircraft can be modeled by combination of intervals of constant velocity (CV) motion and intervals of CT with constant turning speed and a constant angular rate [18], [58]-[60]. In the considered scenario, the radar is positioned at [0*m*, 0*m*], and the target initial position at time $t = 0$ is [5*km*, 5*km*]. During the first 10 seconds, the target approaches the radar along the y-axis with constant speed of 100*m/sec*. Next, the target executes a coordinated turn during 10 seconds with angular velocity of $\frac{3\pi}{20}$ *rad/sec*, which corresponds to acceleration of about 4.5*g*. Finally, the target continues at constant speed of 100*m/sec* motion along x-axis for the next 10 seconds. The target trajectory is shown in Fig. 17. Two scenarios with Gaussian and glint measurement noise correlated to the maneuver statistics, were considered. In the first scenario, a zero-mean Gaussian measurement noise, defined in scenario 1, was considered during the entire simulation. In the second scenario, it was assumed that there is no glint noise during the CV intervals of the target motion and the glint noise increases during the CT interval. The measurement noise PDF in the 3 stages of

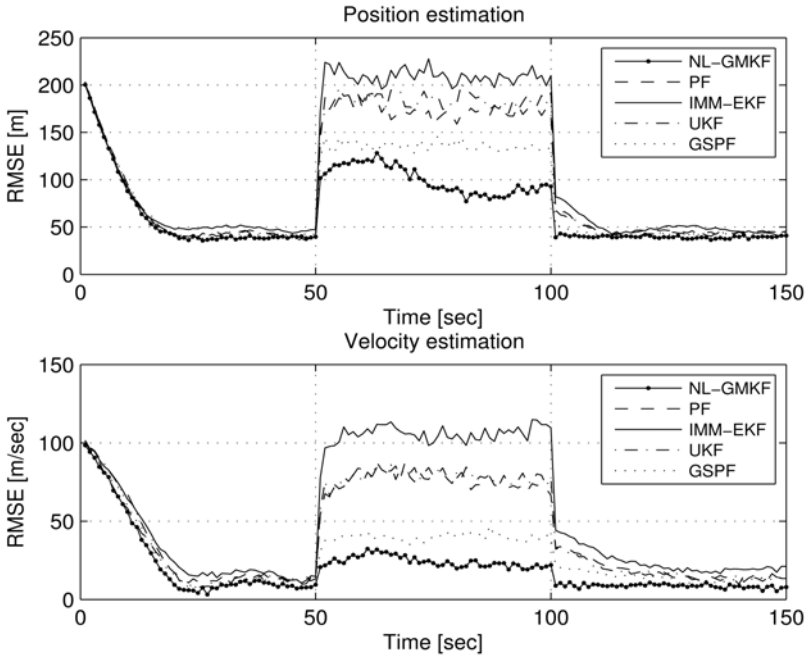

Fig. 15. Tracking performance of NL-GMKF, PF, IMM-EKF, UKF and GSPF in the presence of maneuvering target during $n \in [50, 100]$ with $\xi = 10$, and glint noise with probability $\alpha_{\mathbf{w}} = 0.3$ with glint noise level $\psi = 5$ during $n \in [50, 100]$ with correlated statistics.

target maneuver was taken according to the 3 stages of scenario 4. In the NL-GMKF, the statistics of the measurement noise, considered in scenario 3, was adopted here. The statistics of the system noise were modeled by a mixture of two components, as modeled in scenario 2. The covariance matrix of the second mixture component was selected to be $\mathbf{\Gamma}_{\mathbf{u2}}^{\frac{1}{2}} =$ *diag* ([50*m/sec*^{2}, 50*m/sec*^{2}]). Two types of IMM-EKF algorithms were tested in this scenario: IMM-EKF$^{(1)}$ consists of two CV kinematics models with different system noise levels (similar to the IMM-EKF that was used in scenario 4), and IMM-EKF$^{(2)}$ consists of the CV and the CT models [18]. The IMM-EKF$^{(1)}$ parameters were set to be similar to the IMM-EKF in scenarios 2 and 4, respectively. The covariance matrix of the system noise in the second mode of the IMM-EKF$^{(1)}$ was set to: $\mathbf{\Gamma}_{\mathbf{u2}}^{\frac{1}{2}} =$ *diag* ([50*m/sec*^{2}, 50*m/sec*^{2}]). The first mode of the IMM-EKF$^{(2)}$ is similar to mode 1 of the IMM-EKF in scenario 4. In the second mode of the IMM-EKF$^{(2)}$, the CT model [18] was incorporated by including the turn rate as part of the state vector. The system noise standard deviations of the CT model were 0.5*m/sec*^{2} and 0.03*rad/sec*. The transition probability matrix, defined in (111) was used for the IMM-EKF$^{(2)}$. The tracking performance of the NL-GMKF and IMM-EKFs algorithms for non-glint and glint scenarios are shown in the first and the second columns of Fig. 17, respectively. The two-dimensional CV-CT-CV trajectories estimated by the NL-GMKF and the IMM-EKFs, are shown in the first to third rows. The estimation performances of the tested algorithms as a function of time are shown in Fig. 18. This figure shows that performances of the NL-GMKF and the IMM-EKF$^{(2)}$ are similar and that both outperform the IMM-EKF$^{(1)}$ in the first scenario.

However, the NL-GMKF outperforms both IMM-EKF algorithms in the second scenario, in the presence of glint noise during the CT interval. It should be noted that the model of IMM-EKF(2) is tailored to specific scenarios of CV and CT, while the NL-GMKF does not use such prior information. The use of CV and CT models can enable using smaller system noise variance and could be simply incorporated within the NL-GMKF. However, such a prior information increases the algorithm sensitivity to other types of target maneuvering. The target position and velocity estimation RMSEs in this scenario are slightly higher than the corresponding RMSEs presented in Fig. 16 due to modeling mismatch. Finally, the tracking performance of the NL-GMKF was evaluated for a scenario with longer maneuvering duration with the same high acceleration. This scenario models higher maneuvering intensity. In this scenario, the target executes a coordinated turn during 50 seconds with acceleration magnitude of about 4.5g. As in the previous case, the acceleration vector varies, and this variation is not modeled in the NL-GMKF. The estimation performances of the tested algorithms as a function of time are shown in Fig. 19. This figure shows that performances of the NL-GMKF and the IMM-EKF(2) are similar and both of them outperform the IMM-EKF(1) in the first scenario, while the NL-GMKF outperforms both IMM-EKF algorithms in the second scenario, in the presence of glint noise during the CT interval.

B.6 NL-GMKF sensitivity

The main motivation of this test is to show that the performance of the NL-GMKF is weakly sensitive to the assumed GMM order of the posterior distribution. In addition, the sensitivity of the NL-GMKF performance to the size of the artificial data used in greedy EM learning, is tested. Note that the size of the artificial data is theoretically unlimited and it is practically set according to the processor computational resources [40]. The scenario 4 is considered here and the position estimation performance at time instance n = 75 is used for sensitivity testing. Fig. 20 shows that for the tested number of Gaussians in the mixture, the NL-GMKF estimation performance improves as the size of the artificial data in the EM stage increases. It can be observed that the estimation performance improves to a turning point, from which the estimation performance degrades. The reason for the estimation performance degradation is the overfitting problem [63]. The threshold for the number of the components in the mixture increases with the size of the artificial data used in the EM stage. This figure shows that only a minor improvement in the estimation performance is achieved with increase of the GMM order.

5. Conclusions

Two new recursive filters, named as GMKF and NL-GMKF, for linear and nonlinear, non-Gaussian problems were presented in this chapter. The GMKF algorithm consists of the GSF followed by an efficient model order reduction method. The GSF provides a rigorous solution for state vector estimation in a linear DSS model with GMM-distributed system and measurement noises and it generalizes the original KF to GMMs. Practical implementation of the optimal GSF is limited due to the exponential model order growth. The GMKF solves this problem via an efficient model order reduction method. The problem of exponential growth of the model order was solved via the mixture PDF estimation at each step using the greedy EM algorithm. It was shown that greedy EM-based order reduction scheme does not

signi_cantly reduce the GMKF estimation performance. The estimation performance of the GMKF was tested for non-Gaussian cases using simulations, and it was shown that the GMKF outperforms the PF, IMM and the KF. However this superiority comes at the cost of the extra computational complexity caused by the use of EM for model order reduction procedure. The NL-GMKF was also derived based on the MMSE criterion. It assumes a nonlinear DSS model with general non-Gaussian distributions for the system and measurement noise. The posterior distribution of the state vector is modeled by GMM whose parameters are determined to minimize its estimated KLD from the true distribution. The NL-GMKF was applied to a nonlinear problem of maneuvering radar target tracking in the presence of glint noise. The performance of the derived NL-GMKF was evaluated via simulations and compared to the PF, the IMM-EKF and the EKF. The simulations showed that the NL-GMKF outperforms the PF, the IMM-EKF and the EKF in the considered cases.

APPENDIX

Proposition 1: Consider two M-order jointly GMM-distributed random variables $\mathbf{x}$ and $\mathbf{y}$ whose joint distribution can be written as

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim GMM\left(\alpha_m, \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}m} \\ \boldsymbol{\mu}_{\mathbf{y}m} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Gamma}_{\mathbf{x}m} & \boldsymbol{\Gamma}_{\mathbf{x}\mathbf{y}m} \\ \boldsymbol{\Gamma}_{\mathbf{y}\mathbf{x}m} & \boldsymbol{\Gamma}_{\mathbf{y}m} \end{bmatrix}; m = 1, \ldots, M\right) . \tag{112}$$

The MMSE estimator of $\mathbf{y}$ from $\mathbf{x}$ is

$$E(\mathbf{y}|\mathbf{x}) = \frac{1}{\sum_{m'=1}^{M} \alpha_{m'} \Phi(\mathbf{x}; \boldsymbol{\theta}_{m'})} \sum_{m=1}^{M} \alpha_m \Phi(\mathbf{x}; \boldsymbol{\theta}_m) \left[\boldsymbol{\mu}_{\mathbf{y}m} + \boldsymbol{\Gamma}_{\mathbf{y}\mathbf{x}m} \boldsymbol{\Gamma}_{\mathbf{x}m}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}m})\right] , \tag{113}$$

where $\boldsymbol{\theta}_m$ consists of the mean and covariance matrix of the mth mixture component of the PDF of $\mathbf{x}$.

The proof of this proposition can be obtained by noting that $E(\mathbf{y}|\mathbf{x}) = E_{\eta_m} E_{\mathbf{y}|\mathbf{x},\eta_m}(\mathbf{y}|\mathbf{x}, \eta_m)$ where η_m is the random mixture indicator of the mth Gaussian [47]. Therefore, the conditional distribution of $\mathbf{y}|\mathbf{x}, \eta_m$ is Gaussian, and this implies that the MMSE estimator of a GMM-distributed random vector is a weighted sum of linear MMSE (LMMSE) estimators. Note that the MMSE in (113) is nonlinear for $M > 1$, as expected in non-Gaussian problems.

Proposition 2: Consider two statistically independent GMM-distributed random vectors $\mathbf{x} \sim GMM(\alpha_{\mathbf{x}k}, \boldsymbol{\mu}_{\mathbf{x}k}, \boldsymbol{\Gamma}_{\mathbf{x}k}; k = 1,..., K)$ and $\mathbf{y} \sim GMM(\alpha_{\mathbf{y}l}, \boldsymbol{\mu}_{\mathbf{y}l}, \boldsymbol{\Gamma}_{\mathbf{y}l}; l = 1,..., L)$. The random vector $\mathbf{z}$, defined as:

$$\mathbf{z} = \mathbf{A}_{\mathbf{x}}\mathbf{x} + \mathbf{A}_{\mathbf{y}}\mathbf{y} + \mathbf{b} , \tag{114}$$

is GMM-distributed of order $J = KL$: $\mathbf{z} \sim GMM(\alpha_{\mathbf{z}j}, \boldsymbol{\mu}_{\mathbf{z}j}, \boldsymbol{\Gamma}_{\mathbf{z}j}; j = 1,..., J)$, with

$$\alpha_{\mathbf{z}j} = \alpha_{\mathbf{x}k}\alpha_{\mathbf{y}l} , \tag{115}$$

$$\boldsymbol{\mu}_{\mathbf{z}j} = \mathbf{A}_{\mathbf{x}}\boldsymbol{\mu}_{\mathbf{x}k} + \mathbf{A}_{\mathbf{y}}\boldsymbol{\mu}_{\mathbf{y}l} + \mathbf{b} , \tag{116}$$

$$\boldsymbol{\Gamma}_{\mathbf{z}j} = \mathbf{A}_{\mathbf{x}}\boldsymbol{\Gamma}_{\mathbf{x}k}\mathbf{A}_{\mathbf{x}}^{H} + \mathbf{A}_{\mathbf{y}}\boldsymbol{\Gamma}_{\mathbf{y}l}\mathbf{A}_{\mathbf{y}}^{H} , \tag{117}$$

where the matrices $\mathbf{A_x}$, $\mathbf{A_y}$ and the vector $\mathbf{b}$ are deterministic. The index j represents the combination of (k, l), through the relation: $j = (k-1)L + l$.

Proof of Proposition 2

Equation (114) can be rewritten as

$$\mathbf{z} = \begin{bmatrix} \mathbf{A_x} & \mathbf{A_y} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \mathbf{b} \ . \tag{118}$$

The vectors $\mathbf{x}$ and $\mathbf{y}$ are GMM-distributed of order K and L, respectively, and thus $\mathbf{v} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ is GMM-distributed of order LK. Using GMM definition in (6), the PDF of $\mathbf{v}$ can be rewritten as

$$f_{\mathbf{v}}(\mathbf{v}) = \sum_{l=1}^{L} \sum_{k=1}^{K} \alpha_{\mathbf{x}k} \alpha_{\mathbf{y}l} \Phi(\mathbf{x}, \boldsymbol{\theta}_{\mathbf{x}k}) \Phi(\mathbf{y}, \boldsymbol{\theta}_{\mathbf{y}l}) \ . \tag{119}$$

From (119) it can be easily obtained that the weight of the lkth mixture component is $\alpha_{\mathbf{v}lk} = \alpha_{\mathbf{x}k}\alpha_{\mathbf{y}l}$. Since $\mathbf{z}$ is a linear transformation of the GMM-distributed random vector $\mathbf{v}$, therefore $\mathbf{z}$ is GMM-distributed of order LK. Hence, the lkth mixture weight in the distribution of $\mathbf{z}$ is

$$\alpha_{\mathbf{z}lk} = \alpha_{\mathbf{v}lk} = \alpha_{\mathbf{x}k}\alpha_{\mathbf{y}l} \ . \tag{120}$$

Let $\tilde{\eta}_{lk}$ denote the random mixture indicator in the PDF of $\mathbf{z}$. Using (118) and (119), the mean vector of the lkth mixture component is

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{z}lk} &= E\left[\left(\begin{bmatrix} \mathbf{A_x} & \mathbf{A_y} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \mathbf{b} \right) \middle| \tilde{\eta}_{lk} \right] = \begin{bmatrix} \mathbf{A_x} & \mathbf{A_y} \end{bmatrix} \begin{bmatrix} E[\mathbf{x}|\tilde{\eta}_{lk}] \\ E[\mathbf{y}|\tilde{\eta}_{lk}] \end{bmatrix} + \mathbf{b} \\ &= \mathbf{A_x}\boldsymbol{\mu}_{\mathbf{x}k} + \mathbf{A_y}\boldsymbol{\mu}_{\mathbf{y}l} + \mathbf{b} \ . \end{aligned} \tag{121}$$

Similarly, the covariance matrix of the lkth mixture component is

$$\begin{aligned} \boldsymbol{\Gamma}_{\mathbf{z}lk} &= \operatorname{cov}\left(\left(\begin{bmatrix} \mathbf{A_x} & \mathbf{A_y} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \mathbf{b} \right) |\tilde{\eta}_{lk} \right) = \operatorname{cov}\left(\mathbf{A_x}\mathbf{x} + \mathbf{A_y}\mathbf{y} | \tilde{\eta}_{lk} \right) \\ &= \mathbf{A_x}\boldsymbol{\Gamma}_{\mathbf{x}k}\mathbf{A}_{\mathbf{x}}^{H} + \mathbf{A_y}\boldsymbol{\Gamma}_{\mathbf{y}k}\mathbf{A}_{\mathbf{y}}^{H} \ , \end{aligned} \tag{122}$$

where the cross-covariance of $\mathbf{x}$ and $\mathbf{y}$ vanishes, because $\mathbf{x}$ and $\mathbf{y}$, conditioned on the Gaussian indicator, $\tilde{\eta}_{lk}$, are statistically independent.

6. References

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, pp. 174-188, 2002.

[2] N. Ikoma, N. Ichimura, T. Higuchi, and H. Maeda, "Particle filter based method for maneuvering target tracking," *IEEE International Workshop on Intelligent Signal Processing,* Hungary, pp. 3-8, 2001.
[3] G. Hewer, R. Martin, and J. Zeh, "Robust preprocessing for Kalman filtering of glint noise," *IEEE Trans. AES,* vol. 23, pp. 120-128, 1987.
[4] D. Howard, "Radar target glint in tracking and guidance system based on echo signal phase distortion," *Proc. on National Electronics Conference,* pp. 840-849, 1959.
[5] B. Borden and M. Mumford, "A statistical glint/radar cross section target model," *IEEE Trans. AES,* vol. 19, no. 1, pp. 781-785, 1983.
[6] W. Wu and P. Cheng, "Nonlinear IMM algorithm for maneuvering target tracking," *IEEE Trans. AES,* vol. 30, pp. 875-884, 1994.
[7] X. Li and V. Jilkov, "Survey of maneuvering target tracking. Part V: Multiple-model methods," *IEEE Trans. AES,* vol. 41, no. 4, pp. 1255-1320, 2005.
[8] B. Anderson and J. Moore, *Optimal filtering.* Prentice-Hall 1979.
[9] A. Jazwinski, "Stochastic processes and filtering theory," *Academic Press,* NY, 1970.
[10] S. Julier and J. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," *Proc. The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls.,* 1997.
[11] S. Haykin, *Kalman Filtering and Neural Networks.* pp. 221-280, Wiley Publishing, 2001.
[12] M. Vemula, M. Bugallo, and P. Djuric, "Performance comparison of Gaussian-based filters using information measures," *IEEE Signal Processing Letters,* vol. 14, no. 12, pp. 1-4, 2007.
[13] A. Farina, B. Ristic, and D. Benvenuti, "Tracking a ballistic target: comparison of several nonlinear filters," *IEEE Trans. AES,* vol. 38, no. 3, pp. 854-867, 2002.
[14] C. Masreliez, "Approximate non-Gaussian filtering with linear state and observation relations," *IEEE Trans. Automatic Control*, AC-20, pp. 107-110, 1975.
[15] C. Masreliez and R. Martin, "Robust Bayesian estimation for the linear model and robustifying the Kalman filter," *IEEE Trans. Automatic Control*, AC-22, pp. 361-371, 1975.
[16] W. Wu, "Target tracking with glint noise," *IEEE Trans. AES,* vol. 29, pp. 174-185, 1993.
[17] D. Magill, "Optimal adaptive estimation of sampled stochastic processes," *IEEE Trans. Automat. Control,* vol. 10, no. 1, pp. 434-439, 1965.
[18] Y. Bar-Shalom and X. Li, Estimation with Applications to Tracking and Navigation. Artech House, NY, 2001.
[19] D. Alspach and H. Sorenson, "nonlinear Bayesian estimation using Gaussian sum approximation," *IEEE Trans. Automat. Control,* vol. AC-17, pp. 439-448, 1972.
[20] E. Daeipour and Y. Bar-Shalom, "An interacting multiple model approach for target tracking with glint noise," *IEEE Trans. AES,* vol. 31, no. 2, pp. 706-715, 1995.
[21] F. Gustafsson, *Adaptive filtering and change detection.* John Wiley & Sons, NY, 2000.
[22] D. Salmond, "Mixture reduction algorithms for target tracking in clutter," *SPIE Signal and Data Processing of Small Targets, 1305,* pp. 434-445, 1990.
[23] L. Pao, "Multisensor multitarget mixture reduction algorithms for tracking," *Jour. Guidamce, Control, and Dynamics,* vol. 17, no. 6, pp. 1205-1211, 1994.
[24] Y. Bar-Shalom and L. Xiao-Rong, *Multitarget Multisensor Tracking: Principles and Techniques.* Storrs. CT: YBS Publishing, 1995.
[25] S. Blackman and R. Popoli, *Design and Analysis of Modern tracking Systems.* Norwood. MA: Artech House, 1999.

[26] S. Blackman, *Multiple-Target Tracking with Radar Applications.* Norwood, MA: Artech House, 1986.

[27] J. Williams and P. Maybeck, "Cost-function-based hypothesis control techniques for multiple hypothesis tracking," *Mathematical and Computer Modelling,* vol. 43, Iss. 9-10, pp. 976-989, 2006.

[28] P. Maybeck and B. Smith, "Multiple model tracker based on Gaussian mixture reduction for maneuvering targets in clutter," *Proc. 7th Int. Conference on Information Fusion,* pp. 40-47, 2005.

[29] H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm with markovian switching coefficients," *IEEE Trans. Automat. Control,* vol. 33, no. 8, pp. 780-783, 1988.

[30] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: a survey," *IEEE Trans. AES,* vol. 34, no. 1, pp. 103-123, 1998.

[31] E. Daeipour and Y. Bar-Shalom, "IMM tracking of maneuvering targets in the presence of glint," *IEEE Trans. AES,* vol. 34, no. 3, pp. 996-1002, 1998.

[32] A. Doucet, N. de Freitas, and N. Gordon, *An introduction to sequential Monte Carlo methods.* New York: Springer-Verlag 2001.

[33] R. Chen and N. Liu, "Mixture Kalman filter," *Journal of Royal Statistical Society,* vol. 62, part 3, pp. 493-508, 2000.

[34] P. Djurić, N. Kotecha, N. Zhang, Y. Huang, T. Chirmai, M. Bugallo, and N. Miguez, "Particle filtering," *IEEE Signal Processing Magazine,* pp. 19-38, 2003.

[35] N. Kotecha and P. Djurić, "Gaussian sum particle filtering,"*IEEE Trans. Signal Processing,* vol. 51, no. 10, pp. 2602-2612, 2003.

[36] H. Hu, Z. Jing, A. Li, S. Hu, and H. Tian, "An MCMC-based particle filter for tracking target in glint noise environment," *Journal of Systems Engineering and Electronics*, vol. 16, no. 2, pp. 305-309, 2005.

[37] T. Ho and B. Chen, "Novel extended Viterbi-based multiple model algorithms for state estimation of discrete-time systems with Markov jump parameters," *IEEE Trans. Signal Proc.,* vol. 54, no. 2, pp. 393-404, 2006.

[38] N. Bergman, A. Doucet and N. Gordon, "Optimal estimation and Cram_er-Rao bounds for partial non-Gaussian state space models," *Ann. Inst. Statist. Math.,* vol. 53, no. 1, pp. 97-112, 2001.

[39] I.Bilik and J. Tabrikian, "MMSE-based tracking filter with non-Gaussian process and measurement noises," *IEEE Trans. on Aerospace and Electronic Systems, in press*.

[40] I. Bilik and J. Tabrikian, "Optimal recursive filtering using Gaussian mixture model," *Proc. IEEE Workshop on Statistical Signal Processing,* pp. 399-403, 2005.

[41] N. Li and A. Barron, "Mixture density estimation," In Advances in Neural Information Processing Systems 12. The MIT Press, 2002.

[42] G. Ackerson and K. Fu, "On State Estimation in Switching Environments," *IEEE Trans. AC,* vol.15, no. 1, pp. 10-17, 1970.

[43] J. Tugnait and A. Haddad, "Adaptive estimation in linear systems with unknown markovian noise statistics," *IEEE Trans. on Information Theory,* vol. 26, no. 1, pp. 66-78, 1980.

[44] I. Bilik and J. Tabrikian, "Maneuvering target tracking using the nonlinear non-Gaussian Kalman filter," *Proc. of ICASSP*, vol. III, pp. 724 - 727 May 2006.

[45] I. Bilik and J. Tabrikian, "Target tracking with nonlinear Gaussian mixture Kalman filter," *Trans. on Aerospace and Electronic Systems, in press*.

[46] I. Bilik and J. Tabrikian, "Target tracking in glint noise environment using nonlinear non-Gaussian Kalman filter," *Proc. IEEE International Radar Conference,* pp 24-27, April 2006.

[47] K. Todros and J. Tabrikian, "Application of Gaussian mixture model for blind separation of mixture models," *Proc. of ICA 2004, Granada,* pp. 392-395, 2004.

[48] K. Todros and J. Tabrikian, "Blind Separation of Independent Sources Using Gaussian Mixture Model," *IEEE Trans. Sig. Proc.,* vol. 55, no. 7, pp. 3645-3658, 2007.

[49] S. Kay, *Fundamentals of Statistical Signal Processing.* vol. I, Prentice Hall, Upper Saddle River, NJ, 1998.

[50] D. A. Reynolds, *A Gaussian mixture modeling approach to text-independent speaker identification,* Ph.D. Dissertation, Georgia Institute of Technology, 1992.

[51] J. Verbeek, N. Vlassis, and B. Kröse, "Efficient greedy learning of Gaussian mixture models," *Neural Computations*, MIT Press, 2002.

[52] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Mag.,* 36-47, 2004.

[53] F. Xiaobin, H. Peikang, and F. Zhengfang, "Analysis spectral-power-density of glint of radar-extended target," *Proc. on CIE Radar Conf.*, pp. 1128-1331, 2001.

[54] S. Xiaoli, L. Yongcai, W. Xinmin, and W. Changqing, "Angular glint of aircraft formation and its applications," *Proc. ICMA 2007,* pp. 1334-1339, 2007.

[55] E. Wan and R. van der Merwe, "The unscented Kalman filter for nonlinear estimation," *Proc. Adaptive Systems for Signal Processing, Communications, and Control Symposium,* pp. 153-158, 2000.

[56] J. Lindsay, "Angular glint and the moving, rotating, complex radar target," *IEEE Trans. AES,* vol. 4, no. 2, pp. 164-173, 1968.

[57] G. Sandhu and A. Sauylor, "A real-time statistical radar target model," *IEEE Trans. AES,* vol. 21, no. 4, pp. 490-507, 1985.

[58] X. Yuan, C. Han, Z. Duan, and M. Lei, "Comparison and choice of models in tracking target with coordinated turn," *Proc. FUSION07,* pp. 1497-1502, 2005.

[59] X. Li and Y. Bar-Shalom, "Design of interacting multiple model algorithm for tracking in air traffic control systems," *Proc. 32 Conf. on Design and Control,* pp. 906-911, 1993.

[60] M. Morelande and N. Gordon, "Target tracking through a coordinated turn," *Proc. ICASSP 2005*, vol. IV, pp. 21-24, 2005.

[61] X. Rong and V. Jilkov, "Survey of maneuvering target tracking-Part I: Dynamic models," *IEEE Trans. Aerospace and Electronic Systems,* vol. 39, no. 4, 2003.

[62] D. Middleton, "Statistical-physical model of electromagnetic interference," *IEEE Trans. Eletromagn. Compat.* vol. EMC-19, no. 3, pp. 106-126, 1977.

[63] R. Duda, P. Hart, and D. Stork, *Pattern Classification.* A Wiley-Intersience Publication, 2001.

[64] X. Li and V. Jilkov, "Survey of maneuvering target tracking. Part I: Dynamic models," *IEEE Trans. AES,* vol. 39, no. 4, 2003.

[65] R. Singer, "Estimating optimal tracking filter performance for manned maneuvering target," *IEEE Trans. AES,* vol. 6, pp. 473-483, 1970.

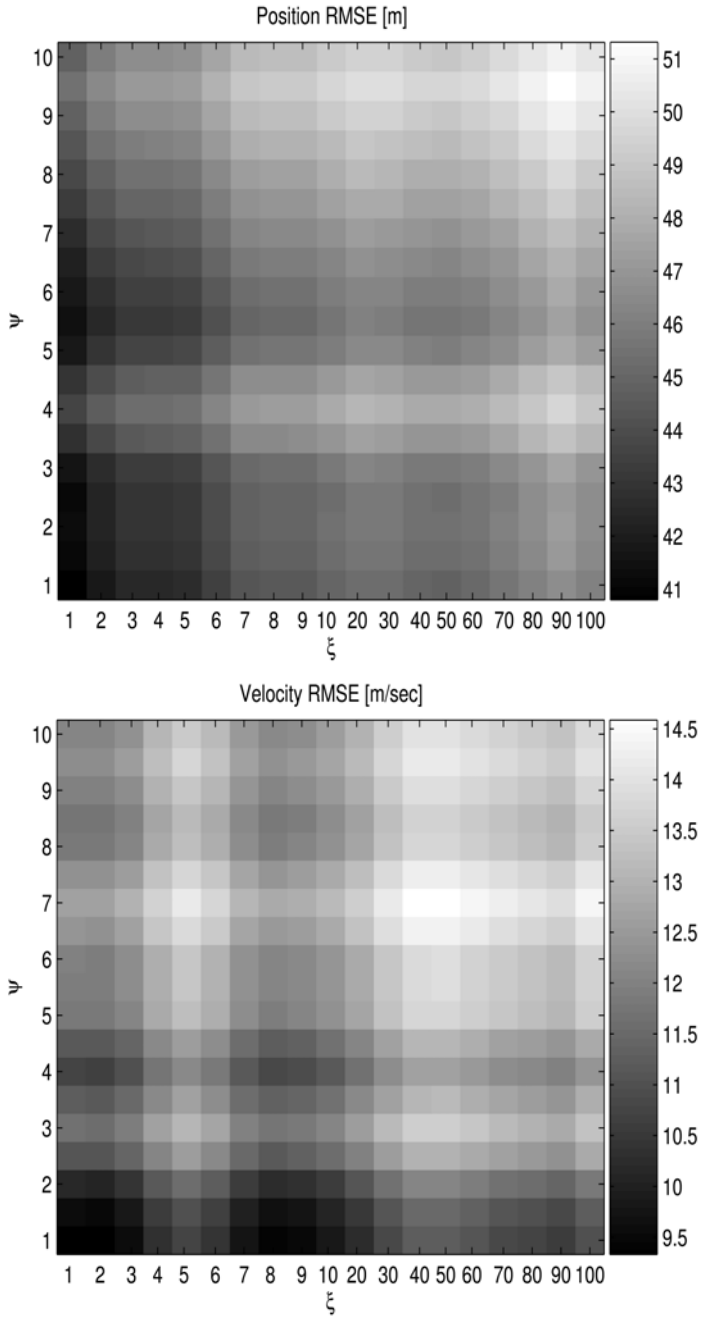

Fig. 16. Tracking performance of NL-GMKF in the presence of maneuvering target with glint noise during the second interval at time instance $n = 75$ and $\alpha_w = 0.3$.

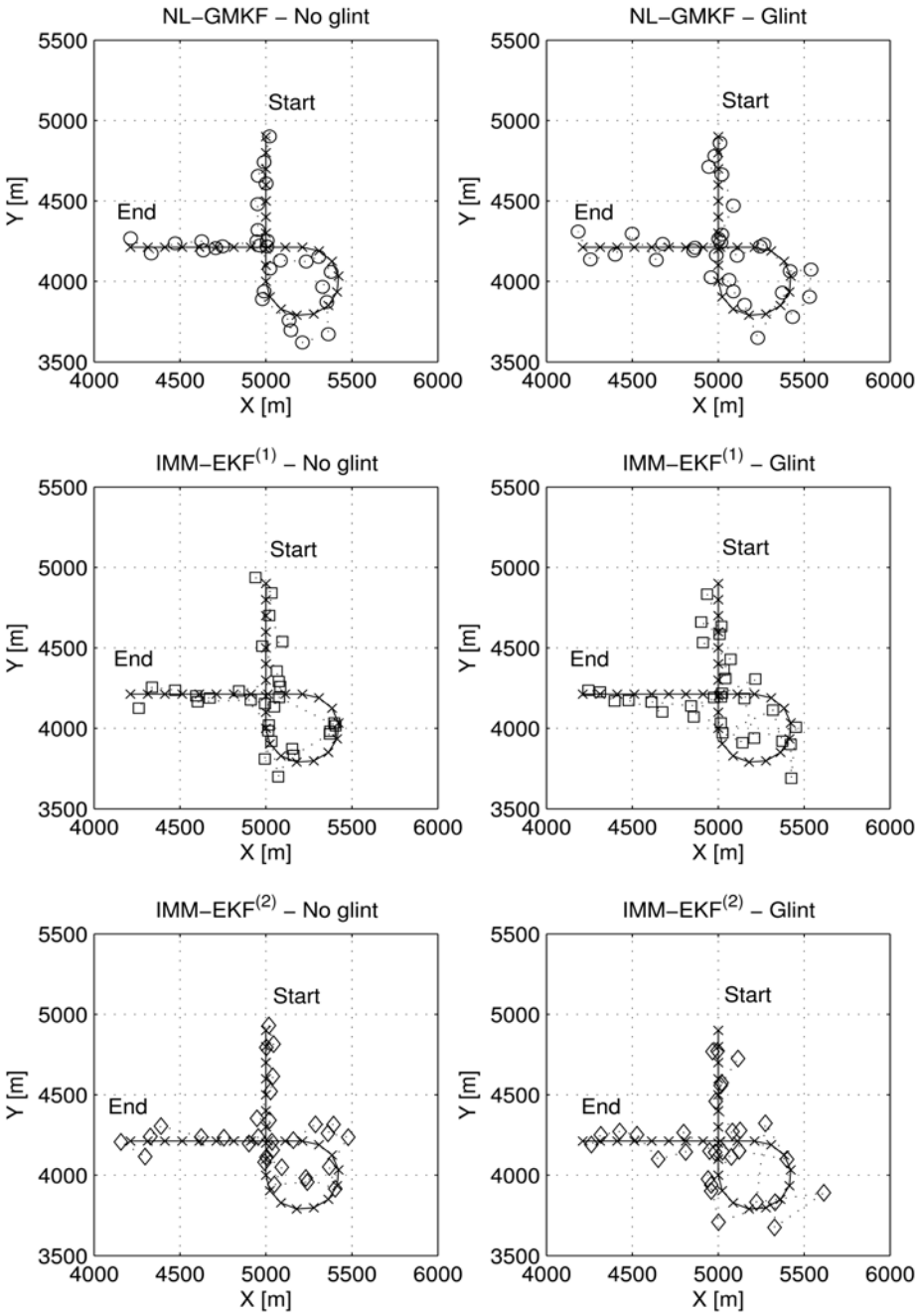

Fig. 17. Tracking performance of NL-GMKF compared to IMM-EKF(1) and IMM-EKF(2) in a CV-CT-CV scenario, with and without glint during the CT period.

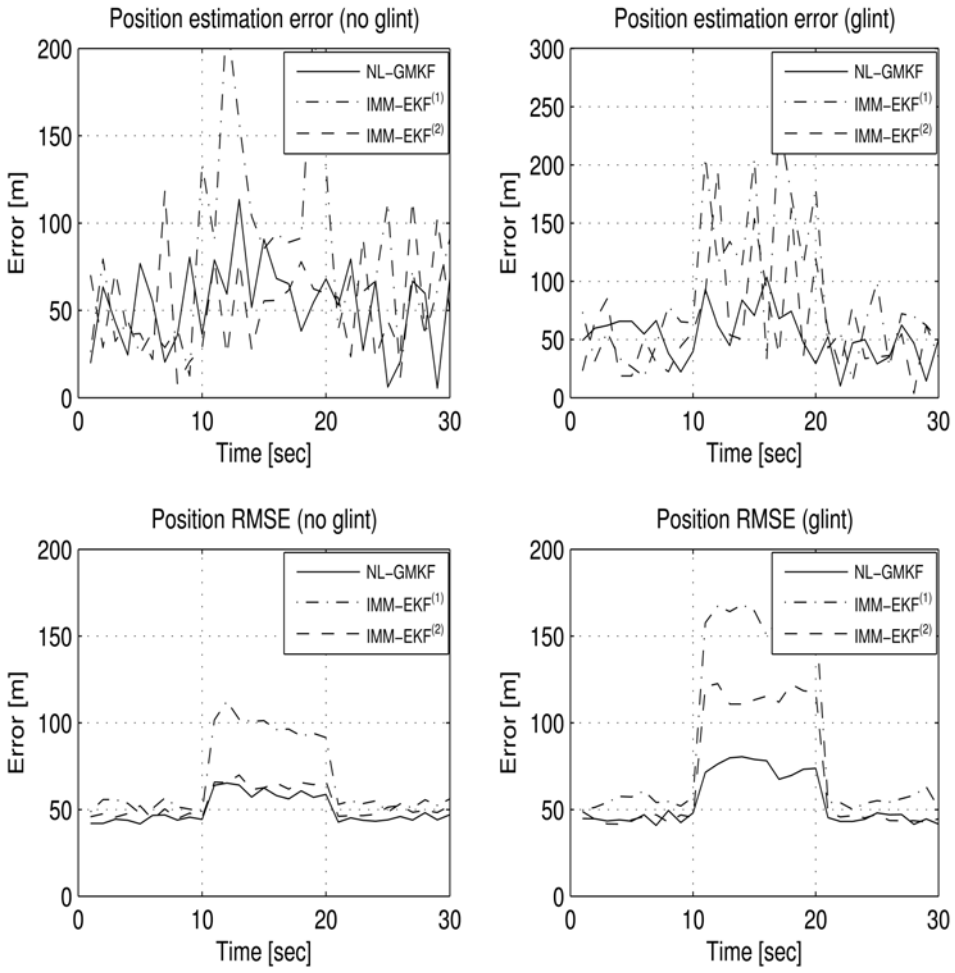

Fig. 18. Tracking performance of NL-GMKF, IMM-EKF$^{(1)}$ and IMM-EKF$^{(2)}$ in a CV-CT-CV scenario, without and with glint noise with probability $\alpha_{\mathbf{w}} = 0.3$ and glint noise level $\psi = 5$ during the CT period.

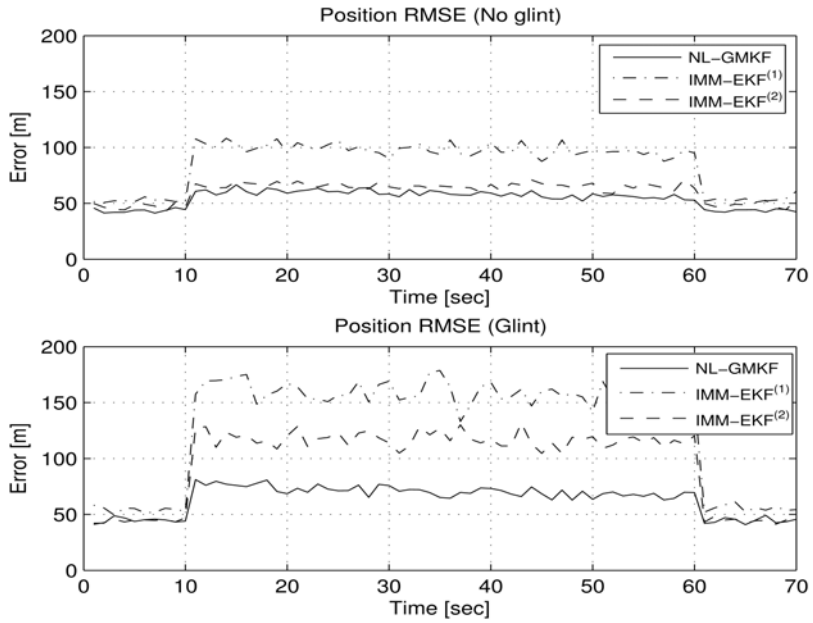

Fig. 19. Tracking performance of NL-GMKF, IMM-EKF$^{(1)}$ and IMM-EKF$^{(2)}$ in a CV-CT-CV high maneuver intensity scenario, without and with glint noise with probability $\alpha_w = 0.3$ and glint noise level $\psi = 5$ during the CT period.

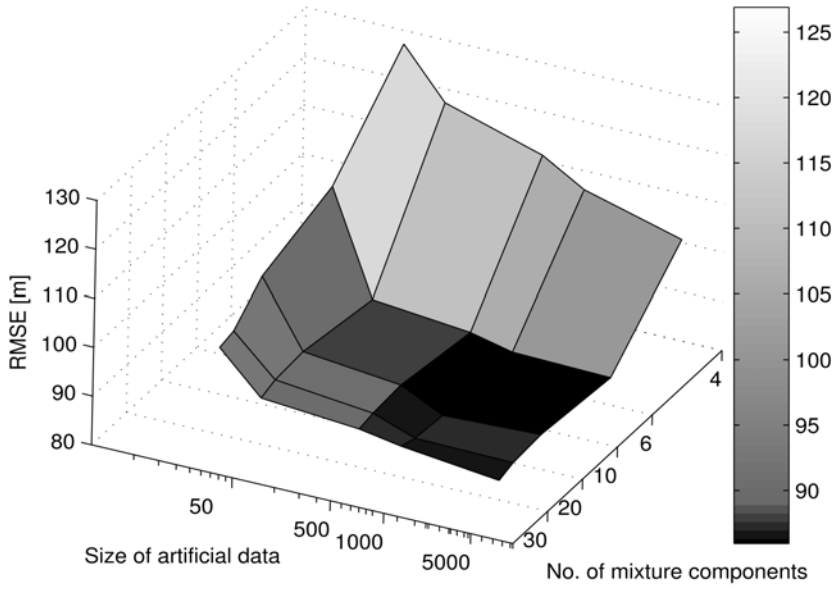

Fig. 20. NL-GMKF sensitivity testing.

Kalman Filter Recent Advances and Applications

Edited by Victor M. Moreno and Alberto Pigazo

ISBN 978-953-307-000-1
Hard cover, 584 pages
Publisher InTech
Published online 01, April, 2009
Published in print edition April, 2009

The aim of this book is to provide an overview of recent developments in Kalman filter theory and their applications in engineering and scientific fields. The book is divided into 24 chapters and organized in five blocks corresponding to recent advances in Kalman filtering theory, applications in medical and biological sciences, tracking and positioning systems, electrical engineering and, finally, industrial processes and communication networks.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

I. Bilik and J. Tabrikian (2009). MMSE-Based Filtering for Linear and Nonlinear Systems in the Presence of Non-Gaussian System and Measurement Noise, Kalman Filter Recent Advances and Applications, Victor M. Moreno and Alberto Pigazo (Ed.), ISBN: 978-953-307-000-1, InTech, Available from: http://www.intechopen.com/books/kalman_filter_recent_adavnces_and_applications/mmse-based_filtering_for_linear_and_nonlinear_systems_in_the_presence_of_non-gaussian_system_and_mea

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821