IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 19, 2018, accepted July 3, 2018, date of publication July 5, 2018, date of current version July 25, 2018.

*Digital Object Identifier 10.1109/ACCESS.2018.2853572*

# Similarity-Based Multiple Model Adaptive Estimation

**AKBAR ASSA, (Member, IEEE), AND KONSTANTINOS N. PLATANIOTIS, (Fellow, IEEE)**
Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada

Corresponding author: Akbar Assa (akbar.assa@utoronto.ca)

The work of A. ASSA was supported by the Natural Sciences and Engineering Research Council of Canada under Grant PDF-487746-2016.

**ABSTRACT** Multiple model adaptive estimation (MMAE) methods are frequently used to overcome the parametric uncertainty of the system's model. Most MMAE methods approximate the state posteriori (posterior probability) by a weighted arithmetic average of model posteriories using a Bayesian weighting scheme. Despite its effectiveness, arguably arithmetic averaging is not the most proper type of averaging for probability densities. Besides, the exploited Bayesian weighting scheme eventually reduces the MMAE to the single best candidate model, which is problematic in many scenarios. Motivated by such shortcomings, this paper proposes a similarity-based approach for MMAE which enhances the estimation accuracy by generalizing the model averaging scheme and providing realistic weights for each model. The proposed approach provides a posteriori which on average is closest to all posteriories and assigns weights to each model based on their similarity to the true model. The choice of similarity measure leads to various schemes. The simulation results confirm the superiority of the proposed MMAE methods as compared to the conventional method.

**INDEX TERMS** Multiple model estimation, generalized averaging, probabilistic similarity measures.

## I. INTRODUCTION

The problem of multiple model (MM) estimation is considered in this paper, where new weighting and fusion schemes are provided by exploiting a similarity-based approach. Known by many names such as ''filter bank'' [1] and ''Gaussian sum filter'' [2], [3], multiple model systems have found vast popularity in numerous applications including control [4], [5], Fault Detection and Isolation (FDI) [6], [7], pattern recognition [8], and target tracking [9], [10]. The MMAE systems, initially introduced by [11] and later enhanced in a series of works [12]–[14], exploit a bank of filters each of which follows a candidate behaviour mode of the system. The estimation of the filters are then combined to provide the overall estimate of the states. The switching between the models was considered later by introducing the generalized pseudo-Bayes algorithm (GPBA) [15] which was enhanced in [16] and promoted further by developing the Interacting Multiple Model (IMM) estimation scheme [17]. The design of the model candidates set was also pursued by a group of works such as [18]. Nonetheless these advancements in MMAE systems follow the same fusion rule of the initial work of [11] and [13].

Traditionally, the fusion of filters in a MMAE system is provided by the Weighted Arithmetic Mean (WAM) operator the wights of which are calculated through a Bayesian scheme. Despite its practical significance, arithmetic mean is not the only optimal way to fuse the information gained from the filters. In fact, there are more suitable options for averaging when dealing with probability densities [19]. Besides, the underlying Bayesian weighting scheme converges to the best candidate model exponentially fast [20]–[22], reducing the system to the best single available model which is not favorable as the mode of the system may change during operation. This paper proposes a novel similarity-based fusion approach to address these shortcomings.

Despite their significance, only few works have considered the aformentioned limitations of the MMAE systems. To alleviate the problem caused by Bayesian weighting scheme of [11], lower bounds were introduced which prevented the weights of other models from reducing to zero [23]. Yet, it is not clear how such lower bounds should be selected [22]. A non-Bayesian weighting scheme was proposed by [24] through which the measurement likelihood of each model is averaged over all time instances to update the model weights.

Similarly the weights are updated in [25], considering only the exponent of the likelihoods. The negative affect of precomputed components in calculation of weights was identified in [26] and eliminated to enhance the performance of the system. However, the weighting problem was not approached in a systematic way. Recently, the Kullback-Liebler Average (KLA) [27] was used to obtain the optimal estimation in an IMM estimation scheme [28]. It was shown by [27] that KLA is equivalent to the weighted geometric average and coincides with the fusion rule provided by the Generalized Covariance Intersection (GCI) method [29]. However, such an averaging is ineffective when the system is left with a single model (as in the traditional Bayesian weighting scheme).

The motivation of this work is to provide novel fusion schemes for MMAE systems to enhance the estimation accuracy. For that matter, probabilistic similarity measures are exploited to achieve suitable averaging and weighting schemes. The proposed approach generalizes the traditional fusion method by providing more optimal options for fusion which are superior to the traditional method in terms of performance. Moreover, the provided fusion scheme can be easily employed for other similar applications such as sensor fusion [30]. It is noteworthy that the convergence of the MMAE systems is closely connected to the convergence of individual models [30], which itself depends on conditions such stabilizability and detectability of the candidate models [31], [32] and also closeness of the model parameters to the true values (e.g., initial conditions of the filter). As it was discussed earlier, the traditional MMAE systems converge to the closest model available to true model [20]–[22]. Therefore, their convergence depends mainly on the convergence of the closest model to reality. In case of the proposed averaging methods, the adverse effect of possibly divergent models is negated by the proposed weighting scheme since the weights of these models diminish exponentially as they diverge from the true model. In that sense, the proposed MMAE system will converge to the average of converging systems, weighted based on their correctness (i.e., distance from the true model). In order to provide a fair comparison with the classical MMAE solution and to demonstrate the versatility and novelty of the proposed contributions, the simplest possible structure for the elemental units (state space models and corresponding estimators) is considered in this work. To that end, it is assumed that each state space model is linear, time-invariant, with additive white Gaussian noise and with known initial conditions. Following the traditional approach [11]–[14] it is assumed that one model is active for the duration of the experiments. Extensions of the MMAE approach, such as generalized Bayesian filters (including the IMM variant), or MMAE solutions operating on nonlinear and/or non-Gaussian models employing filers such as Extended Kalman Filter (EKF), particle or cubature filters are not part of this work since applying such extensions by mirroring the past developments in the MMAE literature is straight-forward.

The key contributions of this work are summarized as follows,

- First, a novel averaging method for fusing the posteriories of the MMAE system is proposed through exploiting a weighted average distance. The fusion is performed by seeking a posteriori which minimizes the proposed average distance. It is shown that the first two moments of such a posteriori are retrievable for some popular similarity measures. It should be noted that even though the exploited similarity measures are not new, the proposed averaging methods based on them are.
- It was shown that the traditional WAM fusion scheme is a special case of the proposed fusion method and is re-derived using the proposed approach. In addition, other optimal fusion rules such as Weighted Geometric Mean (WGM) and Weighted Square Mean Root (WSMR) are also derived using other similarity measures. Simulation results suggest superior performance of these new fusion schemes as compared to WAM.
- The derived averaging methods are generalized as the weighted generalized mean, which can provide various fusion schemes.
- The weighting scheme is revised to weight each filter by the closeness of its predicted measurements to the observed measurements exploiting the similarity measure of choice. Through multiple simulations, it is shown that this approach enhances the estimation accuracy by preventing the models from being neglected.

The rest of the paper is structured as follows. The model of the system and the problems of the traditional MM methods are discussed in Sec. II. Section III comprises the details of the proposed similarity-based averaging method where various options for fusion of all models towards an optimal estimation is considered. Similarity-based weighting of the models is the focus of Sec. IV, where a novel non-Bayesian weighting scheme is presented and various options for weighting of the models are introduced. The simulations of the proposed methods are included in Sec. V, where the performance of the proposed methods is compared to that of the traditional MM. The paper is concluded in Sec. VI.

## II. PROBLEM FORMULATION AND BACKGROUND

Throughout this work, the following notation is followed: the scalar variables are denoted by non-bold letters (e.g., $a$), the vectors are shown by small bold letter (e.g., $\boldsymbol{a}$ ), and the matrices and sets of vectors are denoted by capital bold letters (e.g., $\boldsymbol{A}$). This work is focused on MM estimation of the states in linear dynamical systems. The states in such systems evolve through a linear model,

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}^{*}\boldsymbol{x}_k + \boldsymbol{w}_k \tag{1}$$

where $\boldsymbol{F}^{*}$ is the transition matrix and assumed to be time-invariant, $\boldsymbol{x}_k \in \mathbb{R}^n$ and $\boldsymbol{w}_k \in \mathbb{R}^n$ are the vectors of states and process noise at time step $k$, respectively. The process noise is modeled with an additive zero-mean white Gaussian noise

($\boldsymbol{w}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}_k^*)$). The initial state of the system is assumed to follow a normal distribution, i.e. $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0^*, \boldsymbol{\Sigma}_0^*)$. The states of the system are measured indirectly at each time step,

$$\boldsymbol{z}_k = \boldsymbol{H}^* \boldsymbol{x}_k + \boldsymbol{v}_k \tag{2}$$

where $\boldsymbol{z}_k \in \mathbb{R}^m$, $\boldsymbol{v}_k \in \mathbb{R}^m$ and $\boldsymbol{H}^*$ are the measurement vector, the measurement noise vector, and the observation matrix at time step $k$, respectively. The measurement noise is assumed to be an additive zero-mean white Gaussian noise ($\boldsymbol{v}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_k^*)$). The process noise, measurement noise, and initial states are also assumed to be independent.

In a traditional MMAE scheme, the state posteriori is calculated through marginalizing over all models [11]. If the models probability is discrete,

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k) = \sum_{i=1}^{M} w_k^i \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i) \tag{3}$$

where $p(\boldsymbol{x}_k|\boldsymbol{Z}_k, \mathcal{M}^i) = \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)$ is the posteriori of model $i$ the mean ($\boldsymbol{\mu}_k^i$) and covariance ($\boldsymbol{\Sigma}_k^i$) of which are calculated thought Kalman filtering, $w_k^i = p(\mathcal{M}^i|\boldsymbol{Z}_k)$ is the associated weight of model $i$ and $M$ is the number of models. The posteriori is approximated by a Gaussian ($p(\boldsymbol{x}_k|\boldsymbol{Z}_k) \approx \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$) which preserves the first two moments of (3) as follows [14] (derivation details are available in the appendix),

$$\boldsymbol{\mu}_k = \sum_{i=1}^{M} w_k^i \boldsymbol{\mu}_k^i \tag{4}$$

$$\boldsymbol{\Sigma}_k = \sum_{i=1}^{M} w_k^i \left[\boldsymbol{\Sigma}_k^i + (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)(\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)^T\right] \tag{5}$$

Conventionally, the weight of a model is calculated recursively using the Bayesian theorem [11],

$$\begin{aligned} w_k^i &= \frac{p(\boldsymbol{z}_k|\mathcal{M}^i, \boldsymbol{Z}_{k-1}) p(\mathcal{M}^i|\boldsymbol{Z}_{k-1})}{\sum_{i=1}^{M} p(\boldsymbol{z}_k|\mathcal{M}^i, \boldsymbol{Z}_{k-1}) p(\mathcal{M}^i|\boldsymbol{Z}_{k-1})} \\ &= c_1 \prod_{j=1}^{k} p(\boldsymbol{z}_j|\mathcal{M}^i, \boldsymbol{Z}_{j-1}) \\ &= c_2 e^{\frac{-1}{2}\left[\sum_{j=1}^{k} ln|\boldsymbol{\Omega}_j^i| + tr\{\sum_{j=1}^{k} \tilde{\boldsymbol{z}}_j^i \tilde{\boldsymbol{z}}_j^{iT} \boldsymbol{\Omega}_j^{i^{-1}}\}\right]} \end{aligned} \tag{6}$$

where $c_1$ and $c_2$ are normalizing factors, $p(\boldsymbol{z}_k|\mathcal{M}^i, \boldsymbol{Z}_{k-1}) = \mathcal{N}(\hat{\boldsymbol{z}}_{k|k-1}^i, \boldsymbol{\Omega}_k^i)$ is the measurement likelihood, $p(\mathcal{M}^i|\boldsymbol{Z}_{k-1}) = w_{k-1}^i$ is the likelihood of the model calculated at previous time step, $\tilde{\boldsymbol{z}}_j^i = \boldsymbol{z}_k - \hat{\boldsymbol{z}}_{k|k-1}^i$ is the innovation term and $\boldsymbol{\Omega}_k^i$ is its covariance.

The assumption of discrete probability for models is critical in formation of (3), which may be violated when none of the approximate models are the correct model of the system. As it was discussed in [19], arithmetic averaging is not well-suited for averaging exponential probability densities such as normal distributions. Moreover, the Bayesian weighting scheme (6) causes the weight of the closest model to reality to converge to one, while all other model weights will be reduced to zero as $k$ increases [20]. In other words, the MM estimator will be eventually reduced to the single best model estimator and model averaging is abandoned, which is not favorable if the remaining best model is not close enough to the true model of the system or the mode of the system changes later.

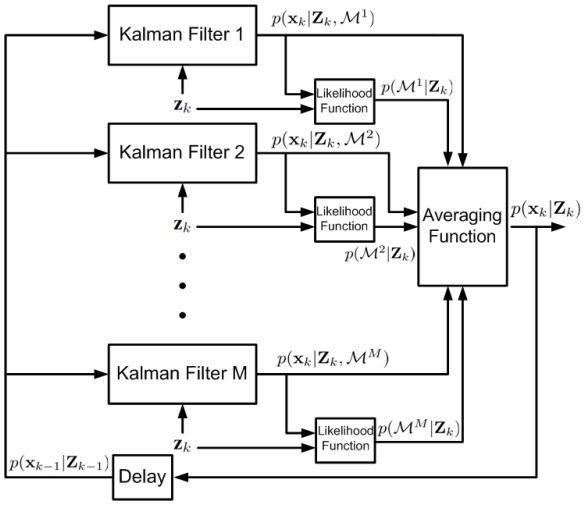


**FIGURE 1.** The structure of a general MMAE system.

## III. SIMILARITY-BASED AVERAGING

In general, MMAE systems use a bank of filters to estimate the posteriori of the states. The structure of a general MMAE system is depicted in Fig. 1. The role of the averaging function shown in this figure is critical as it determines how the estimations of multiple models should lead to a better estimation. In the previous section, the traditional averaging scheme proposed by [11] and [14] was discussed. Here we propose a novel general method to find the optimal posteriori based on multiple posterior distributions of system states. In particular, a normal distribution is sought which on average is closest to all estimated posteriories. For that matter, an average distance between the desired posteriori and posterior distributions of system gained through various models is formulated and then minimized. Such an average distance is defined as follows,

$$L = \sum_{i=1}^{M} w_k^i D(p_k, p_k^i) \tag{7}$$

where $p_k = p(\boldsymbol{x}_k|\boldsymbol{Z}_k) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the sought normal posterior distribution, $p_k^i = p(\boldsymbol{x}_k|\boldsymbol{Z}_k, \mathcal{M}^i) = \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)$ is the state posteriori obtained through model $i$ and $D(p, p^i)$ is a similarity measure of choice between the two probability densities of $p_k$ and $p_k^i$. The goal of the optimization is to determine a normal distribution such that this average distance is minimized, i.e.,

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \arg\min_{p_k} L \tag{8}$$

Since the sought posteriori is normal, it is sufficient to determine its mean and covariance through (8). It is noteworthy that various similarity measures may be exploited in the formulation of (7), each of which leads to distinctive outcomes. Different minimization approaches may be followed to find the desired normal density which minimizes the average distance shown in (7). One way to find the desired normal density is to find the closed-form of the average distance, then differentiate it with respect to the first two moments of the sought normal density and set them to zero. Luckily, the closed-form of the distance between two normal densities are available for many of the similarity measures (such as KLD) and therefore the proposed average distance may be derived in a closed-form. This approach is practiced for the KL-based average distance which yields two averaging methods (due to the asymmetry of KLD), namely the traditional arithmetic average and the less practiced geometric average. However, this approach is not applicable to the Bhattacharyya distance. Instead, an upper bound is derived for the Bhattacharyya-based average distance and the normal density minimizing this upper bound is identified which results in another novel averaging method. Needless to say, even though the exploited similarity measures are not new, the proposed averaging methods based on them are.

### A. KULLBACK-LIEBLER DIVERGENCE

Kullback-Liebler Divergence is one of the most popular similarity measures which was originally introduced in information theory as the relative entropy [33]. Using this similarity measure, the average distance is formulated as follows,

$$L_1 = \sum_{i=1}^{M} w_k^i D_{KL}(\mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i) \parallel p_k(\boldsymbol{x}_k)) \tag{9}$$

It can be shown that,

$$\begin{aligned} L_1 &= \sum_{i=1}^{M} w_k^i \int_{-\infty}^{+\infty} \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i) ln \frac{\mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)}{p_k(\boldsymbol{x}_k)} d\boldsymbol{x}_k \\ &= \int_{-\infty}^{+\infty} \sum_{i=1}^{M} w_k^i \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i) ln \frac{\sum_{i=1}^{M} w_k^i \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)}{p_k(\boldsymbol{x}_k)} d\boldsymbol{x}_k \\ &\quad - \int_{-\infty}^{+\infty} \sum_{i=1}^{M} w_k^i \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i) ln \frac{\sum_{i=1}^{M} w_k^i \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)}{\mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)} d\boldsymbol{x}_k \\ &= D_{KL}\left(\sum_{i=1}^{M} w_k^i \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i) \parallel p_k(\boldsymbol{x}_k)\right) + c_3 \end{aligned} \tag{10}$$

where $c_3$ is a constant which does not depend on the choice of $p_k$. It is apparent that minimizing this average distance is only possible by selecting the desired posteriori to be the WAM of the posterior distributions of all models, similar to the traditional choice of averaging first by [11]. In other words, the traditional formulation for computing the posteriori in MM systems is a special case of the proposed similarity-based averaging method.

It is desirable to have a single normal distribution for the posteriori. Exploiting the KLD formulation between two normal distributions, the average distance shown in (7) may be written as follows,

$$\begin{aligned} L_1 = \sum_{i=1}^{M} w_k^i \left[ tr\{{\boldsymbol{\Sigma}_k}^{-1} \boldsymbol{\Sigma}_k^i\} - n + ln \frac{|\boldsymbol{\Sigma}_k|}{|\boldsymbol{\Sigma}_k^i|} \right] \\ + \sum_{i=1}^{M} w^i \left[ (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)^T {\boldsymbol{\Sigma}_k}^{-1} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i) \right] \end{aligned} \tag{11}$$

The mean and covariance of the desired normal distribution may be found by solving the minimization problem shown in (8). For that matter, the average distance is differentiated with respect to the mean and set equal to zero,

$$\frac{\partial L_1}{\partial \boldsymbol{\mu}_k} = \sum_{i=1}^{M} w_k^i {\boldsymbol{\Sigma}_k}^{-1} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i) = 0 \tag{12}$$

Assuming the covariance matrix to be of full rank, it can be concluded that,

$$\boldsymbol{\mu}_k = \sum_{i=1}^{M} w_k^i \boldsymbol{\mu}_k^i \tag{13}$$

Now differentiation with respect to the covariance and setting it to zero results in,

$$\begin{aligned} \frac{\partial L_1}{\partial \boldsymbol{\Sigma}_k} = -\sum_{i=1}^{M} w_k^i {\boldsymbol{\Sigma}_k}^{-T} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)(\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)^T {\boldsymbol{\Sigma}_k}^{-T} \\ - \sum_{i=1}^{M} w_k^i {\boldsymbol{\Sigma}_k}^{-T} {\boldsymbol{\Sigma}_k^i}^T {\boldsymbol{\Sigma}_k}^{-T} + \boldsymbol{\Sigma}_k^{-T} = 0 \end{aligned} \tag{14}$$

which can be rewritten as follows,

$$\sum_{i=1}^{M} w^i \left[ {\boldsymbol{\Sigma}_k^i}^T + (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)(\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)^T \right] {\boldsymbol{\Sigma}_k}^{-T} = \boldsymbol{I} \tag{15}$$

where $\boldsymbol{I}$ is the identity matrix. The covariance of the posteriori is easily obtainable from (15) as follows,

$$\boldsymbol{\Sigma}_k = \sum_{i=1}^{M} w^i \left[ \boldsymbol{\Sigma}_k^i + (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)(\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)^T \right] \tag{16}$$

It is not surprising to see that the mean and covariance of the posteriori derived in (13) and (16) are the same as those obtained by [11] and [14] in (4) and (5) since as it was shown in (10), the distribution which minimizes such an average is a weighted arithmetic mean of posteriories.

Since KLD is not symmetric with respect to the probability densities, it is also possible to use it differently to form the

average distance as follows,

$$\begin{aligned} L_2 &= \sum_{i=1}^{M} w_k^i D_{KL}(\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \parallel \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)) \\ &= \sum_{i=1}^{M} w_k^i \left[ tr\{{\boldsymbol{\Sigma}_k^i}^{-1} \boldsymbol{\Sigma}_k\} - n + ln \frac{|\boldsymbol{\Sigma}_k^i|}{|\boldsymbol{\Sigma}_k|} \right] \\ &\quad + \sum_{i=1}^{M} w_k^i (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)^T {\boldsymbol{\Sigma}_k^i}^{-1} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i) \end{aligned} \tag{17}$$

To minimize this average distance as shown in (8), the distance is differentiated with respect to the mean of the desired posteriori and set equal to zero as follows,

$$\frac{\partial L_2}{\partial \boldsymbol{\mu}_k} = \sum_{i=1}^{M} w^i {\boldsymbol{\Sigma}_k^i}^{-1} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i) = 0 \tag{18}$$

which yields,

$$\sum_{i=1}^{M} w^i {\boldsymbol{\Sigma}_k^i}^{-1} \boldsymbol{\mu}_k = \sum_{i=1}^{M} w^i {\boldsymbol{\Sigma}_k^i}^{-1} \boldsymbol{\mu}_k^i \tag{19}$$

Similarly, setting the differentiation of $L_2$ with respect to the covariance to zero as follows,

$$\frac{\partial L_2}{\partial \boldsymbol{\Sigma}_k} = \sum_{i=1}^{M} w^i \left[ {\boldsymbol{\Sigma}_k^i}^{-T} - \boldsymbol{\Sigma}_k^{-T} \right] = 0 \tag{20}$$

results in,

$$\boldsymbol{\Sigma}_k^{-1} = \sum_{i=1}^{M} w^i {\boldsymbol{\Sigma}_k^i}^{-1} \tag{21}$$

Using (21) in (19) yields,

$$\boldsymbol{\mu}_k = \boldsymbol{\Sigma}_k \sum_{i=1}^{M} w^i {\boldsymbol{\Sigma}_k^i}^{-1} \boldsymbol{\mu}_k^i \tag{22}$$

The results gained in (21) and (22) are identical to those gained through using the weighted geometric mean (WGM) since the average distance shown in (17) is the same as KLA, which was shown to be equivalent to WGM [27].

A closer look at the average distances obtained by KLD in (11) and (17) reveals that both of these schemes lead to estimations with minimum MSE. However, the mean is weighted by the model covariances in (17), while the weights of all samples are equal in (11). In addition, the second form of KLD-based average distance causes the inverse of the posteriori covariance (i.e., the posteriori information) to be equal to weighted arithmetic average of inverse of posteriori covariance of all models. In that sense, it can be inferred that using this averaging method, the gained information is equal to weighted sum of information from all models. As a result, in calculation of the overall mean, the mean of each model is weighted not only by the likelihood of the model but also by the information it contains, similar to distributed data fusion applications such as [34].

### B. BHATTACHARYYA DISTANCE

One of the widely applied similarity measures in statistics is the Bhattacharyya distance (BD) [35] which is easily calculable for normal distributions. Unfortunately, the minimization of the BD-based average distance would not lead to closed-form solutions for the mean and covariance of the posteriori. However, it is possible to approximate these values by minimizing an upper bound of this average distance. Using the Jensen's inequality, the upper bound of the BD-based average distance is found as follows,

$$\begin{aligned} L_3 &= \sum_{i=1}^{M} w_k^i D_{BC}(p_k(\boldsymbol{x}_k), \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)) \\ &= -\sum_{i=1}^{M} w_k^i ln \int_{-\infty}^{+\infty} \sqrt{p_k(\boldsymbol{x}_k) \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)} d\boldsymbol{x}_k \\ &\leq -ln \int_{-\infty}^{+\infty} \sqrt{p_k(\boldsymbol{x}_k)} \sum_{i=1}^{M} w_k^i \sqrt{\mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)} d\boldsymbol{x}_k \\ &= D_{BC}\left( p_k(\boldsymbol{x}_k), \left( \sum_{i=1}^{M} w_k^i \sqrt{\mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)} \right)^2 \right) \end{aligned} \tag{23}$$

It is apparent that this upper bound is minimized when $p_k(\boldsymbol{x}_k) = \left( \sum_{i=1}^{M} w^i \sqrt{\mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)} \right)^2$. In other words, minimizing the BD-based average distance requires the desired posteriori to be the weighted square mean root (WSMR) of all model posterior distributions. The desired posteriori may be expanded further as follows,

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{Z}_k) &= \left( \sum_{i=1}^{M} w^i \sqrt{\mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)} \right)^2 \\ &= \sum_{i=1}^{M} \sum_{j=1}^{M} w^i w^j \sqrt{\mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i) \mathcal{N}(\boldsymbol{\mu}_k^j, \boldsymbol{\Sigma}_k^j)} \\ &= \sum_{i=1}^{M} \sum_{j=1}^{M} w^{ij} \mathcal{N}(\boldsymbol{\mu}_k^{ij}, \boldsymbol{\Sigma}_k^{ij}) \end{aligned} \tag{24}$$

where,

$$\boldsymbol{\Sigma}_k^{ij} = 2 \left( {\boldsymbol{\Sigma}_k^i}^{-1} + {\boldsymbol{\Sigma}_k^j}^{-1} \right)^{-1}, \tag{25}$$

$$\boldsymbol{\mu}_k^{ij} = 0.5 \boldsymbol{\Sigma}_k^{ij} ({\boldsymbol{\Sigma}_k^i}^{-1} \boldsymbol{\mu}_k^i + {\boldsymbol{\Sigma}_k^j}^{-1} \boldsymbol{\mu}_k^j), \tag{26}$$

$$w^{ij} = \frac{w^i w^j \sqrt{\frac{\sqrt{|\boldsymbol{\Sigma}_k^i||\boldsymbol{\Sigma}_k^j|}}{|\frac{\boldsymbol{\Sigma}_k^i + \boldsymbol{\Sigma}_k^j}{2}|}}}{e^{\frac{1}{4}(\boldsymbol{\mu}_k^i - \boldsymbol{\mu}_k^j)^T (\boldsymbol{\Sigma}_k^i + \boldsymbol{\Sigma}_k^j)^{-1} (\boldsymbol{\mu}_k^i - \boldsymbol{\mu}_k^j)}} \tag{27}$$

The MoG shown in (24) may be approximated by a single normal distribution as follows,

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{28}$$

$$\boldsymbol{\mu}_k = \sum_{i=1}^{M}\sum_{j=1}^{M} w^{ij}\boldsymbol{\mu}_k^{ij} \tag{29}$$

$$\boldsymbol{\Sigma}_k = \sum_{i=1}^{M}\sum_{j=1}^{M} w^{ij}(\boldsymbol{\mu}_k^{ij}-\boldsymbol{\mu}_k)(\boldsymbol{\mu}_k^{ij}-\boldsymbol{\mu}_k)^T + \sum_{i=1}^{M}\sum_{j=1}^{M} w^{ij}\boldsymbol{\Sigma}_k^{ij} \tag{30}$$

The posterior distribution calculated in (24) is a MoG, the elements of which are the geometric mean of a pair of model posteriories. The weights of the mixture are related to the Bhattacharyya coefficient of the related model posteriori pair. If the posterioies in a pair are far from each other, their weight will be relatively reduced and therefore will contribute less to the overall estimation. Therefore, this type of averaging is more robust to model outliers.

### C. GENERALIZED AVERAGING

The averaging methods obtained from the KLD and BD measured are special cases of the generalized mean, also known as the power mean. If the generalized mean is used for averaging of the posteriories, the overall state posteriori is calculated as follows,

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_k) = \left(\sum_{i=1}^{M} w^i \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)^l\right)^{\frac{1}{l}} \tag{31}$$

where $l$ is the power parameter. If $l = 1$, the posteriori will be the WAM, while the resultant posteriori will be the same as WGM and WSMR if $l \to 0$ and $l = 1/2$, respectively. It can be shown that for a set of posteriories, WAM always results in the largest outcome, while WGM provides the smallest average. Since it is not straight-forward to calculate the mean and covariance of such an average without explicitly knowing the value of $l$, this form of averaging is not explored further. Yet, other important averaging schemes may be resulted by exploring other values for $l$ in this general form.

## IV. SIMILARITY-BASED WEIGHTING SCHEME

The Bayesian weighting scheme employed in the traditional MM reduces the weights to zero for all models but the closest one to reality. To remedy this problem, in this section we propose a novel similarity-based method to determine the weights of the models based on their closeness to the true model. Since the true model is not available, we provide a way to find a fair approximation of true measurement likelihood under some assumptions. This approach is discussed in details in the sequel.

The recursive calculation of the Bayesian weights (shown in (6)) resembles a geometric averaging of the measurement likelihood, except the root of the multiplication is not considered. Now assuming the system has reached its steady state and the weights to be the geometric mean of measurement likelihood, they may be calculated as follows,

$$\begin{aligned} w_k^i &= c_1' \sqrt[k]{\prod_{j=1}^{k} p(\boldsymbol{z}_j|\mathcal{M}^i, \boldsymbol{Z}_{j-1})} \\ &= c_2' e^{\frac{-1}{2}\left[ln|\boldsymbol{\Omega}^i| + tr\{\boldsymbol{\Omega}_k^{i,*}\boldsymbol{\Omega}_j^{i^{-1}}\}\right]} \end{aligned} \tag{32}$$

where $c_1'$ and $c_2'$ are normalizing factors, $\boldsymbol{\Omega}^i$ is the covariance of innovation at the steady state and $\boldsymbol{\Omega}_k^{i,*} = \frac{1}{k}\sum_{j=1}^{k} \tilde{\boldsymbol{z}}_j^i \tilde{\boldsymbol{z}}_j^{iT}$. Such weights are similar to those introduced by [24]. Considering the normal distribution of the measurement likelihood, it is easy to see that the ratio of weights with respect the weight of the true model is related to the KLD between the measurement likelihood of model $i$ and that of the true model, i.e.,

$$\frac{w_k^i}{w_k^*} = e^{-D_{KL}} \tag{33}$$

where,

$$\begin{aligned} D_{KL} &= D_{KL}(p(\boldsymbol{z}_k|\boldsymbol{Z}_{k-1}, \mathcal{M}^*) \parallel p(\boldsymbol{z}_k|\boldsymbol{Z}_{k-1}, \mathcal{M}^j)) \\ &= D_{KL}(\mathcal{N}(\hat{\boldsymbol{z}}_{k|k-1}^*, \boldsymbol{\Omega}^*) \parallel \mathcal{N}(\hat{\boldsymbol{z}}_{k|k-1}^i, \boldsymbol{\Omega}^i)) \\ &= \frac{1}{2}\left[tr\{E_*\{\tilde{\boldsymbol{z}}_k^i \tilde{\boldsymbol{z}}_k^{i\,T}\}\boldsymbol{\Omega}^{i^{-1}}\} + ln|\boldsymbol{\Omega}^i|\right] \\ &\quad - \frac{1}{2}\left[ln|\boldsymbol{\Omega}^*| + m\right] \end{aligned} \tag{34}$$

and $E_*\{\cdot\}$ is the expectation function with respect to $p(\boldsymbol{z}_k|\boldsymbol{Z}_{k-1}, \mathcal{M}^*)$. While [24] assumed $E_*\{\tilde{\boldsymbol{z}}_k^i \tilde{\boldsymbol{z}}_k^{i\,T}\} = \boldsymbol{\Omega}_k^{i,*}$ and $\boldsymbol{\Omega}^*$ to be common between all models, we calculate these values differently, which helps us foster other similarity measures for the same purpose. It can be shown that,

$$E_*\{\tilde{\boldsymbol{z}}_k^i \tilde{\boldsymbol{z}}_k^{i\,T}\} = \boldsymbol{\Omega}^* + (\hat{\boldsymbol{z}}_k^* - \hat{\boldsymbol{z}}_k^i)(\hat{\boldsymbol{z}}_k^* - \hat{\boldsymbol{z}}_k^i)^T \tag{35}$$

Assuming the innovation sequences are zero-mean for all models, then (35) is reduced to $\boldsymbol{\Omega}^*$, i.e.,

$$E_*\{\tilde{\boldsymbol{z}}_k^i \tilde{\boldsymbol{z}}_k^{i\,T}\} \approx \boldsymbol{\Omega}^* \tag{36}$$

Assuming the innovation sequence for model $i$ to be ergodic, then the covariance in (36) may be approximated based on previous samples of innovation, i.e.,

$$\boldsymbol{\Omega}^* \approx \boldsymbol{\Omega}_{N,k}^{*,i} = \frac{1}{N}\sum_{j=k-N+1}^{k} \tilde{\boldsymbol{z}}_j^i \tilde{\boldsymbol{z}}_j^{iT} \tag{37}$$

where $N$ is the size of averaging window, which is selected relatively large in the adaptive estimation scenarios and very small for the FDI applications. Therefore, the weights after normalization will be calculated as follows,

$$w_k^i = c_4 e^{\frac{-1}{2}\left[tr\{\boldsymbol{\Omega}_{N,k}^{*,i}\boldsymbol{\Omega}^{i^{-1}}\} + ln|\boldsymbol{\Omega}^i| - ln|\boldsymbol{\Omega}_{N,k}^{*,i}|\right]} \tag{38}$$

where $c_4$ is the normalizing factor. Comparing (38) with the weights introduced by [24] in (32), it can be seen that they are almost identical except for the last term in the exponent of (38), since this term was considered common in all weights

in calculation of (32) and therefore omitted in the normalization process. It is noteworthy that due to the presence of this last term, the weights in (38) contain another comparison between the observed ($\mathbf{\Omega}^{*,i}_{N,k}$) and precomputed ($\mathbf{\Omega}^{i}$) covariance matrices, therefore unlike (32) the precomputed values in (38) will not adversely affect the weights as was discussed in [26].

Since the measurement likelihood of the true model is approximated, it is also possible to exploit other similarity measures to calculate the weights of the models. For instance, if the Bhattacharyya distance is used instead of KLD, the following weighting scheme will be gained,

$$\frac{w^i_k}{w^*_k} = e^{-D_{BC}} \tag{39}$$

where,

$$\begin{aligned} D_{BC} &= D_{BC}(p(\boldsymbol{z}_k|\boldsymbol{Z}_{k-1}, \mathcal{M}^*) \parallel p(\boldsymbol{z}_k|\boldsymbol{Z}_{k-1}, \mathcal{M}^j) \\ &= D_{BC}(\mathcal{N}(\hat{\boldsymbol{z}}^*_{k|k-1}, \mathbf{\Omega}^*) \parallel \mathcal{N}(\hat{\boldsymbol{z}}^i_{k|k-1}, \mathbf{\Omega}^i) \\ &= \frac{1}{2}\left[ ln|\frac{\mathbf{\Omega}^* + \mathbf{\Omega}^i}{2}| - \frac{1}{2}ln|\mathbf{\Omega}^i| - \frac{1}{2}ln|\mathbf{\Omega}^*| \right] \end{aligned} \tag{40}$$

Under the same assumptions, (37) may be used to calculate the weights, which are formulated as follows after normalization,

$$w^i_k = c_5 \sqrt{\frac{\sqrt{|\mathbf{\Omega}^i||\mathbf{\Omega}^{*,i}_{N,k}|}}{|\mathbf{\Omega}^{*,i}_{N,k} + \mathbf{\Omega}^i|}} \tag{41}$$

where $c_5$ is the normalizing factor. Another interesting similarity measure which can be exploited is the Wasserstein (norm 2) similarity measure which is also known as the Frechet distance [36]. If this measure is used for calculation of weights, the following result is achieved,

$$\frac{w^i_k}{w^*_k} = e^{-D_{W_2}} \tag{42}$$

This distance is calculated as follows,

$$\begin{aligned} D_{W_2} &= D_{W_2}(p(\boldsymbol{z}_k|\boldsymbol{Z}_{k-1}, \mathcal{M}^*) \parallel p(\boldsymbol{z}_k|\boldsymbol{Z}_{k-1}, \mathcal{M}^j) \\ &= D_{W_2}(\mathcal{N}(\hat{\boldsymbol{z}}^*_{k|k-1}, \mathbf{\Omega}^*) \parallel \mathcal{N}(\hat{\boldsymbol{z}}^i_{k|k-1}, \mathbf{\Omega}^i) \\ &= \left[ tr\{\mathbf{\Omega}^* + \mathbf{\Omega}^i - 2\sqrt{\mathbf{\Omega}^i\mathbf{\Omega}^*}\} \right] \end{aligned} \tag{43}$$

where $\mathbf{\Omega}^*$ is approximated with $\mathbf{\Omega}^{*,i}_{N,k}$ under the discussed assumptions. A comparison between (43) and (40) reveals that these two measures offer similar structures, except the function "$ln|\cdot|$" in (40) has been replaced by "$tr\{\cdot\}$" in (43). The latter function offers the sum of eigenvalues for the enclosed matrix, while the former gives the log sum of the eigenvalues. In that sense, the Wasserstein distance discriminates more heavily between the weights of the models as compared to the Bhattacharyya distance.

**TABLE 1. System parameters of the model used in scenario A.**

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\beta$ | 0 | 0 | 0.02 | 0.02 | 0.02 | 0.02 |
| $g_n$ | 0 | 0 | 4 | 4 | 8 | 8 |
| $r$ | 1 | 2 | 1 | 2 | 1 | 2 |

## V. SIMULATION RESULTS

The effectiveness of the proposed similarity-based techniques as compared to traditional MM scheme is verified in this section through multiple Monte-Carlo simulations. Various scenarios are considered including the adaptive estimation and FDI applications. The details of these simulations are discussed in the rest of this section.

### A. SCENARIO A

The goal of the first simulation is to compare the performance of the proposed averaging schemes with that of the traditional method in terms of accuracy. For that purpose, the first simulation model from [37] is exploited. The model parameters are as follows,

$$\boldsymbol{F} = \begin{bmatrix} e^{-0.2} & 0 \\ 0 & e^{-\beta} \end{bmatrix} \quad \boldsymbol{H} = \begin{bmatrix} 1 & 1 \end{bmatrix}$$
$$\boldsymbol{Q} = \begin{bmatrix} 2(1-e^{-0.4}) & 0 \\ 0 & g_n(1-e^{-2\beta}) \end{bmatrix} \quad \boldsymbol{R} = r$$
$$\boldsymbol{\mu}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mathbf{\Sigma}_0 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$$

where $\beta$, $g_n$, and $r$ are not known exactly. Five models with possible values for these parameters are considered, the values of which are shown in Table 1 ($i = 2$ to 6) along with the true parameters, which are presented in the first column ($i = 1$). The estimations of the MM systems with the proposed averaging methods are compared to that of the true model using the KLD-based weighting scheme to ensure none of the models is eleminated from the averaging. The performance of each MM method at each time step $k$ is depicted by calculating their Root Mean Square Relative Error (RMSRE) [38] over 1000 rounds of Monte Carlo Simulations as follows,

$$RMSRE_k = \frac{\sqrt{\sum_{j=1}^{N_{mc}} \|\boldsymbol{\mu}_k(j) - \boldsymbol{x}_k(j)\|_2}}{\sqrt{\sum_{j=1}^{N_{mc}} \|\boldsymbol{x}_k(j)\|_2}} \tag{44}$$

and their overall performance is measured by their absolute RSMRE calculated as follows,

$$ARMSRE = \frac{1}{N_S}\sum_{j=1}^{N_S} RMSRE_j \tag{45}$$

where $N_S$ is the number of the time samples, $N_{mc}$ is the number of Monte Carlo simulations, $\|\boldsymbol{a}\|_2 = \boldsymbol{a}^T\boldsymbol{a}$, and $\boldsymbol{a}(j)$ is the vector $\boldsymbol{a}$ from Monte Carlo simulation $j$. It is noteworthy that both RMSRE and ARMSRE are unitless as they represent

the relative errors of the estimations. The RMSRE of the proposed averaging schemes are depicted in Fig. 2. Even though the performance of these methods are close, the estimations provided by WGM is slightly more accurate, while traditional WAM offers less accuracy. This can also be seen from the ARMSRE of the methods briefed in Table 2. The enhancement of the estimation accuracy comes as a result of considering the error covariance of models in the estimations.

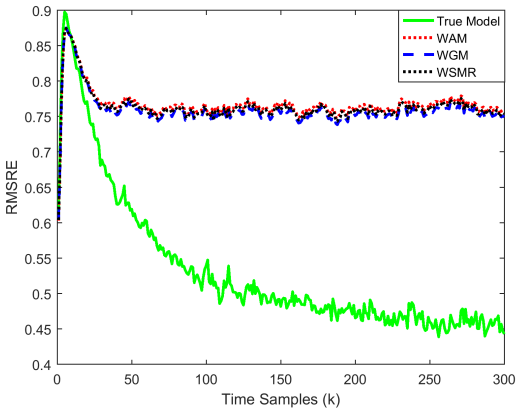


**FIGURE 2.** The RMSRE of the proposed averaging methods and the true model in scenario A.

**TABLE 2.** ARMSRE of the proposed averaging methods and the true model in scenario A.

| Methods | True Model | WAM | WGM | WSMR |
|---|---|---|---|---|
| ARMSRE | 0.5329 | 0.7674 | 0.7596 | 0.7637 |

### B. SCENARIO B

A simple model from [37] is considered in the second simulation to highlight the difference between the proposed weighting schemes. The parameters of this model are as follows,

$$\boldsymbol{F} = \theta \quad \boldsymbol{H} = 1 \quad \boldsymbol{Q} = 1$$
$$\boldsymbol{R} = 1 \quad \boldsymbol{\mu}_0 = 100 \quad \boldsymbol{\Sigma}_0 = 10$$

where $\theta$ of the true model is 0.4. Four models with parameter values of $\theta_1 = 0.1$, $\theta_2 = 0.2$, $\theta_6 = 0.6$, and $\theta_7 = 0.7$ are selected in MM estimation systems, using the proposed weighting schemes. The traditional WAM is used in all of these MM systems. The evolution of weights are shown in Fig.3. As expected, the Bayesian weight of the closest model to reality (model with $\theta_6$) quickly converges to 1, while other Bayesian weights are reduced to zero. This means that the MM estimator with the Bayesian weights is actually reduced to a system with $\mathbf{F} = 0.6$. On the other hand, other similarity-based weighting schemes assign weights to the models based on their closeness to the true mode. Among these methods, Wasserstein-based weighting scheme (shown by $W_2$-Based) discriminate the most between weights, while Bhattacharyya-based (denoted by BC-based) distributes the weights between models almost evenly. It is also noticeable that the discrimination between the models become less significant as the time goes by. This behaviour justifies why the

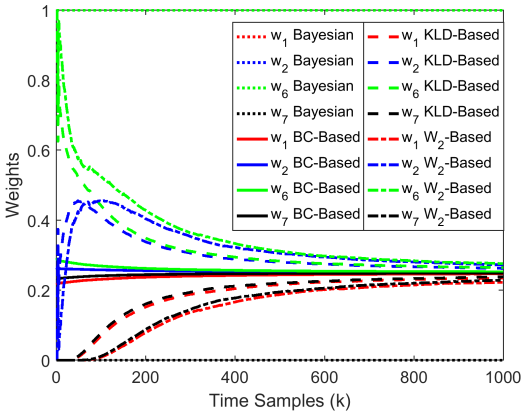


**FIGURE 3.** Model weights using the proposed weighting schemes as compared to the Bayesian weighting method in scenario B.

**TABLE 3.** ARMSRE of the proposed weighting methods and that of the true model in scenario B.

| Methods | ARMSRE |
|---|---|
| True Model | 0.6590 |
| Bayesian | 0.6662 |
| KLD-based | 0.6599 |
| $W_2$-based | 0.6600 |
| BC-based | 0.6593 |

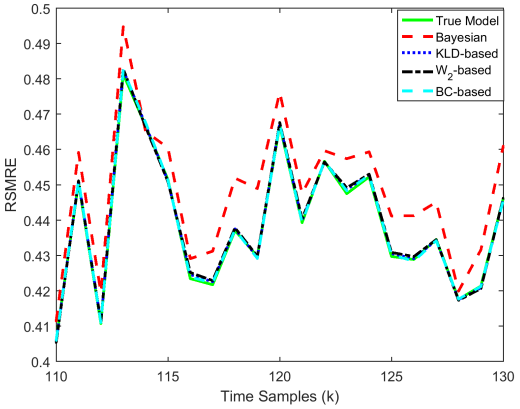


**FIGURE 4.** The estimation RMSRE of the proposed weighting schemes as compared to that of the Bayesian weighting method in scenario B.

sampling window in (37) should be small in FDI applications and large for adaptive estimation ones. Moreover, it can be predicted that the BC-based weights may not be suitable for FDI applications.

The performance of the MMAE systems with these weighting schemes are compared by means of their RMSRE, which is depicted in Fig. 4 for a short period of time to enhance the figure visibility. As it was expected, the accuracy provided by the proposed weighting methods is higher than that of the traditional Bayesian weighting scheme. For this example, the BC-based weighting scheme is slightly more accurate than the other two similarity-based schemes. This fact may also be inferred from the ARMSRE of these methods, briefed in Table 3.

**TABLE 4.** Parameters of the system used in scenario C.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $f_1(i)$ | -0.29 | -0.20 | -0.65 | -0.15 | -0.979 | -0.25 | -0.75 |
| $f_2(i)$ | 0.4 | 0.8 | 0.2 | 0.5 | 1.874 | -0.75 | -0.15 |

**TABLE 5.** ARMSRE of the proposed averaging methods and that of the true model in scenario C.

| Methods | True Model | WAM | WGM | WSMR |
|---|---|---|---|---|
| Set 1 | 0.6174 | 0.6249 | 0.6182 | 0.6192 |
| Set 2 | 0.6174 | 0.6923 | 0.6481 | 0.6681 |

### C. SCENARIO C

In the third scenario, the adaptive estimation example used in [24] and [39] is considered. The system is composed of seven models that are different in their transition matrix (i.e., $\boldsymbol{F}$). The parameters of the system models are as follows,

$$\boldsymbol{F} = \begin{bmatrix} 0 & 1 \\ f_1(i) & f_2(i) \end{bmatrix} \quad \boldsymbol{Q} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.3 \end{bmatrix}$$
$$\boldsymbol{H} = \begin{bmatrix} 1 & 2 \end{bmatrix} \quad \boldsymbol{R} = 0.2$$
$$\boldsymbol{\mu}_0 = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \quad \boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix}$$

where $f_1(i)$ and $f_2(i)$ are found from Table 4. The first column of the table ($i = 1$) is assumed to be the true parameters of the system model. Two simulations are conducted in this scenario. In each simulation, two sets of models are used to estimate the state of the system. The first set (Set 1) includes models 2, 3, and 4, which are relatively close to the true model. The second set (Set 2) is chosen to comprise models 5, 6, and 7, which are comparatively far from the true model. Using each model, Kalman filtering is employed to find the mean and covariance of the posteriori for each model. The temporal performance of the schemes are reflected by their sample-based innovation covariance ($\boldsymbol{\Omega}_k = \frac{1}{k}\sum_{j=1}^{k} \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T$), similar to [24].

In the first simulation of this scenario, the proposed averaging methods are used and compared with the arithmetic averaging of the traditional MM. For a fair competition, the weights of the model are kept equal and constant. For that matter, Set 1 and Set 2 are tested for state estimation. The simulation length is 200 samples ($N_S = 200$). The resultant innovation covariance of these schemes are shown in Fig. 5, and 6 for sample time 100 to 200 for more clarity. The ARMSRE of the methods are also presented in Table 5 for $N_{mc} = 1000$. The performance of all schemes are close, however the performance of WGM is better than that of the other two. Moreover, WSMR provides higher accuracy as compared to the traditional arithmetic mean. The superiority of the proposed averaging methods become more significant when the models are far from the true model (i.e. using Set 2). These performance enhancements come as a consequence of considering the error covariance of each model estimations.

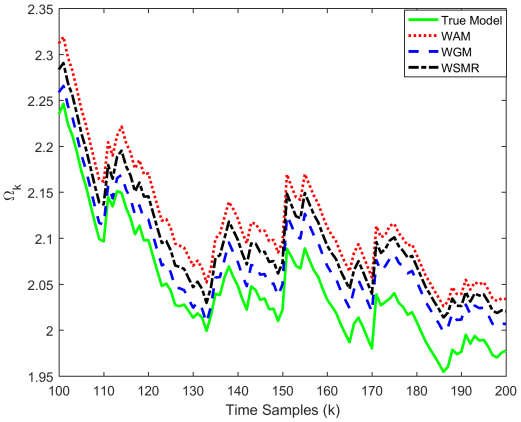


**FIGURE 5.** The innovation covariance of the proposed averaging methods in comparison with that of the true model using Set 1 models in scenario C.

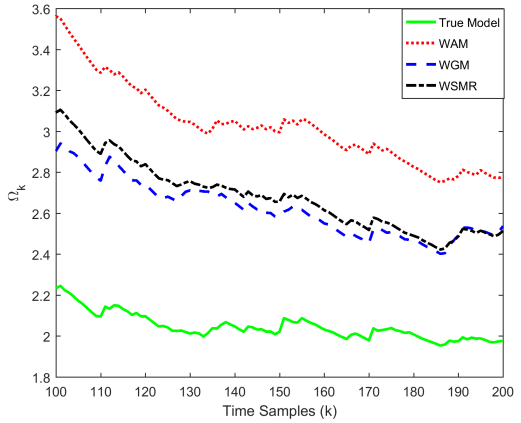


**FIGURE 6.** The innovation covariance of the proposed averaging methods in comparison with that of the true model using Set 2 models in scenario C.

The goal of the second simulation in Scenario C is to demonstrate the effect of the proposed weighting schemes on the performance of the estimations. For that matter, all weighting schemes employed the traditional arithmetic averaging for a fair comparison. Other conditions of the simulation are the same as the previous simulation. The performance of these weighting schemes are depicted in Fig. 7, and 8, while the average performance of the methods are briefed in Table 6. When the models are close to the true model (Set 1), the performance of all weighting schemes are close since the best model match is still a good approximation of the true model. However, when the models are further from the true model (Set 2), the proposed weighting schemes provide higher levels of accuracy. In general, the proposed weighting methods perform equal or better than the Bayesian weighting scheme used in the traditional MM.

### D. SCENARIO D

To test the proposed methods in another MM application, a FDI application is considered in this scenario using the proposed averaging and weighting schemes. For that purpose,

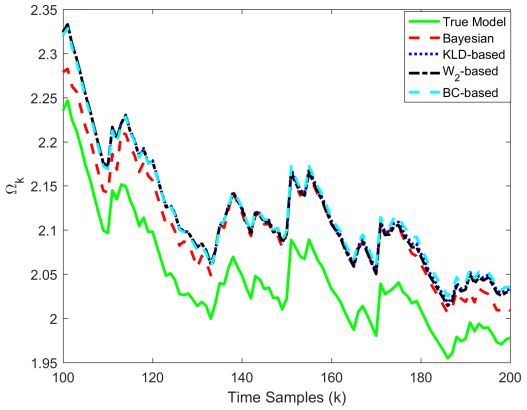


**FIGURE 7.** The innovation covariance of the proposed weighting methods in comparison with that of the true model using Set 1 models in scenario C.

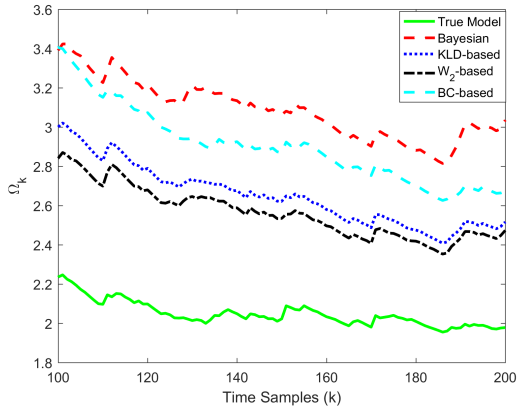


**FIGURE 8.** The innovation covariance of the proposed weighting methods in comparison with that of the true model using Set 2 models in scenario C.

**TABLE 6.** ARMSRE of the proposed weighting methods and that of the true model in scenario C.

| Methods | Set 1 | Set 2 |
|---|---|---|
| True Model | 0.6174 | 0.6174 |
| Bayesian | 0.6249 | 0.6923 |
| KLD-based | 0.6211 | 0.6865 |
| $W_2$-based | 0.6208 | 0.6876 |
| BC-based | 0.6212 | 0.6871 |

the linear system introduced in [40] is exploited and slightly modified to reach a stable system. There are two measurements each of which is assumed to be obtained from a sensor. Five models are available for the measurement matrix ($\boldsymbol{H}$) which model the failure of only one of the two sensors (single failure). The parameters of the system are as follows,

$$\boldsymbol{F} = 0.99 \quad \boldsymbol{Q} = 0.01$$
$$\boldsymbol{H} = \begin{bmatrix} h_1(i) \\ h_2(i) \end{bmatrix} \quad \boldsymbol{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\boldsymbol{\mu}_0 = 100 \quad \boldsymbol{\Sigma}_0 = 10$$

where $h_1(i)$ and $h_2(i)$ are found from table 7. The correct model of the system is shown in the first column of the table ($i = 1$), while other models shown in the rest of the columns use half of the correct value or 0 to represent a

**TABLE 7.** Parameters of models in scenario D.

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $h_1(i)$ | 0.5 | 0.25 | 0.5 | 0 | 0.5 |
| $h_2(i)$ | 1.2 | 1.2 | 0.6 | 1.2 | 0 |

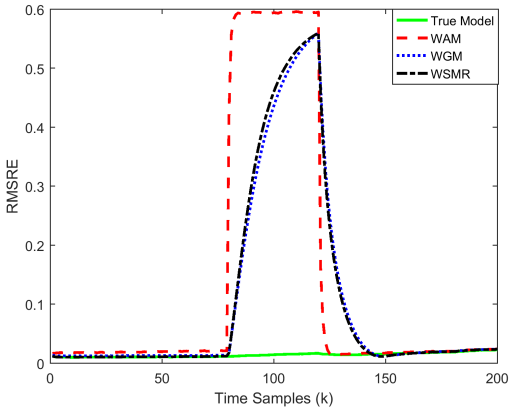


**FIGURE 9.** The RMSRE of the proposed averaging methods in comparison with that of the true model using Set 1 models of Scenario D.

partial or a total failure, respectively. An intermittent total failure in the second sensor is considered in the simulations, where the measurement matrix of the system is changed from the first column to the last column ($i = 5$) between time steps 80 and 120. The system is back to normal after time step 120. Similar to the previous scenario, two sets of models each containing three models are considered. In the first set (Set 1), the first three models are considered, which can only reflect no failure or partial failures. The second set (Set 2) is composed of models 1, 3, and 5, which cover all modes of operation for the second sensor. The temporal efficiency of the methods are measured by RMSRE shown in (44), while the overall performance of each method is evaluated using the ARMSRE over a thousand rounds of Monte Carlo simulations ($N_{mc} = 1000$). Each simulation takes 200 samples to complete ($N_S = 200$).

In the first simulation of this scenario, the averaging methods are tested for FDI. For that matter, the weights of all methods are selected by the traditional Bayesian scheme. However, to avoid having zero weights, a lower bound is selected for all weights as prescribed by [23] the value of which is set to 0.01. Figures 9 and 10 show the results of this comparison for Set 1 and Set 2, respectively, while the overall performance of these methods is summarized in Table 8. As expected, when the fault model is not available (Set 1), the estimation error increases. The WGM and WSMR perform better than the traditional WAM as they consider the uncertainty of each model. In particular, when the fault model is available (Set 2), WSMR outperforms the other two averaging methods.

Next, the effectiveness of the proposed weighting schemes is evaluated for this scenario. All methods are chosen to use the WAM, while their weighting schemes are selected as discussed in Sec. IV. The RMSRE of these methods using

**TABLE 8.** ARMSRE of the proposed averaging methods and that of the true model in scenario D.

| Methods | True Model | WAM | WGM | WSMR |
|---|---|---|---|---|
| Set 1 | 0.0144 | 0.1373 | 0.1056 | 0.1066 |
| Set 2 | 0.0144 | 0.0247 | 0.0174 | 0.0147 |

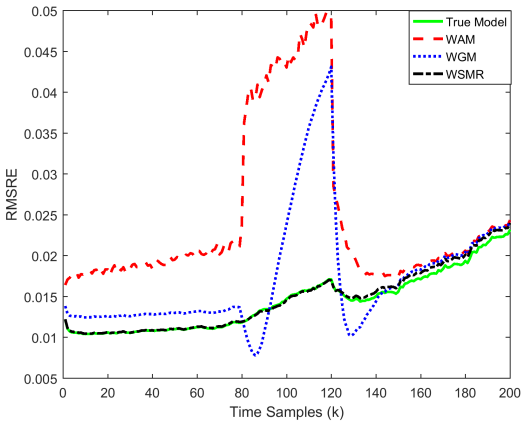


**FIGURE 10.** The RMSRE of the proposed averaging methods in comparison with that of the true model using Set 2 models of Scenario D.

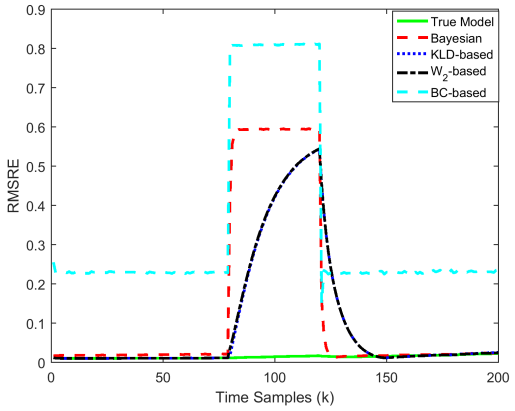


**FIGURE 11.** The RMSRE of the proposed weighting schemes in comparison with that of the true model using Set 1 models of Scenario D.

Set 1 and Set 2 models are depicted in Fig. 11 and 12, while the average performance of the methods are summarized in Table 9. The performance of the KLD-based and Wasserstein ($W_2$)-based weights are quite similar in this simulation. These weights can select the correct model without the interference from other models and therefore their performance is very close to the true model when the fault model is available. On the other hand, the BC-based weighting method cannot exclude the other models and assigns considerable weights to the wrong models, which cause the system to perform poorly. As it was discussed earlier, BC distance discriminates the weights less harshly as compared to the Wasserstein distance and therefore cannot eliminate the wrong models from the averaging. The Bayesian weights perform close to the best model available, yet their performance is suboptimal due to the use of the lower bounds.

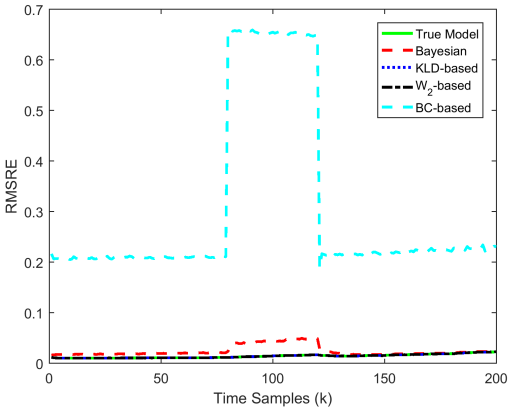


**FIGURE 12.** The RMSRE of the proposed weighting schemes in comparison with that of the true model using Set 2 models of Scenario D.

**TABLE 9.** The ARMSRE of the proposed weighting methods and the true model in Scenario D.

| Methods | Set 1 | Set 2 |
|---|---|---|
| True Model | 0.0144 | 0.0144 |
| Bayesian | 0.1373 | 0.0247 |
| KLD-based | 0.1016 | 0.0144 |
| $W_2$-based | 0.1023 | 0.0144 |
| BC-based | 0.3483 | 0.3050 |

## VI. CONCLUSION

Motivated by the potential problems of the traditional MM estimation, a similarity-based approach was proposed which generalized the traditional averaging scheme and overcome the problem of converging into a single model. In particular, both traditional and novel averaging methods were derived and discussed in terms of their results. Moreover, multiple novel weighting schemes were suggested to solve the problem of the Bayesian weighting scheme. The simulation results confirms the applicability of the proposed MM approach for adaptive estimation and FDI applications. Careful analysis of the simulation results reveals that when the models are quite close to the true model (and one another), the proposed averaging scheme perform only slightly better than the traditional WAM. However, when the models are diverse and far from the true model, the significance of the proposed methods become more apparent (e.g., scenario C using model set 2). This fact can also be seen in the FDI application, where the model sets are quite diverse and far from the true system. In FDI applications, the most discriminative distances (Wasserstein distance and KLD divergence) has the best performance, while the performance of the Bhattacharyya-based weighting method is much better in the adaptive estimation applications (e.g., Scenario B).

It is noteworthy that the proposed averaging methods consider the error covariance of each method in calculation of the mean, while the traditional averaging method does not offer such an advantage. In that sense, the proposed WGM and WSMR perform better than the traditional WAM. This superiority becomes more significant when the models are more diverse (not close to each other). On the other hand,

the advantage of the proposed weighting scheme lies within allocating weights based on the distance of models from the true model. Therefore, the effectiveness of the proposed weighting schemes become more efficient as compared to that of the traditional weighting scheme when the models are diverse and are not very close to the true model.

The proposed approach opens new doors for enhanced estimations in MM systems. Various probabilistic similarity measures may be exploited to achieve new averaging and weighting schemes which best suit the application of interest. In addition, the proposed methodology may be fostered for more complex systems including nonlinear/time-varying system with non-Gaussian noise.

## APPENDIX

The mean of the MoG shown in (5) is calculated as follows,

$$\begin{aligned}\boldsymbol{\mu}_k &= E\{\boldsymbol{x}_k|\boldsymbol{Z}_k\} = \int_{-\infty}^{+\infty}\sum_{i=1}^{M} w^i \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)\boldsymbol{x}_k d\boldsymbol{x}_k \\ &= \sum_{i=1}^{M} w^i \int_{-\infty}^{+\infty} \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)\boldsymbol{x}_k d\boldsymbol{x}_k = \sum_{i=1}^{M} w^i \boldsymbol{\mu}_k^i \end{aligned} \quad (46)$$

The covariance of this MoG is derived as follows,

$$\begin{aligned}\boldsymbol{\Sigma}_k &= E_{p(\boldsymbol{x}_k|\boldsymbol{Z}_k)}\left\{(\boldsymbol{x}_k - \boldsymbol{\mu}_k)(\boldsymbol{x}_k - \boldsymbol{\mu}_k)^T\right\} \\ &= \int_{-\infty}^{+\infty} p(\boldsymbol{x}_k|\boldsymbol{Z}_k)(\boldsymbol{x}_k - \boldsymbol{\mu}_k)(\boldsymbol{x}_k - \boldsymbol{\mu}_k)^T d\boldsymbol{x}_k \\ &= \int_{-\infty}^{+\infty}\sum_{i=1}^{M} w^i \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)(\boldsymbol{x}_k\boldsymbol{x}_k^T - \boldsymbol{x}_k\boldsymbol{\mu}_k^T \\ &\quad - \boldsymbol{\mu}_k\boldsymbol{x}_k^T + \boldsymbol{\mu}_k\boldsymbol{\mu}_k^T)d\boldsymbol{x}_k \\ &= \sum_{i=1}^{M} w^i \left[\int_{-\infty}^{+\infty} \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)\boldsymbol{x}_k\boldsymbol{x}_k^T d\boldsymbol{x}_k \right. \\ &\quad \left. - \boldsymbol{\mu}_k^i\boldsymbol{\mu}_k^T - \boldsymbol{\mu}_k{\boldsymbol{\mu}_k^i}^T + \boldsymbol{\mu}_k\boldsymbol{\mu}_k^T\right] \end{aligned} \quad (47)$$

It can be shown that,

$$\int_{-\infty}^{+\infty} \mathcal{N}(\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i)\boldsymbol{x}_k\boldsymbol{x}_k^T d\boldsymbol{x}_k = \boldsymbol{\Sigma}_k^i + \boldsymbol{\mu}_k^i{\boldsymbol{\mu}_k^i}^T \quad (48)$$

Using (54) in (53) results in the following formulation for the covariance of the mixture,

$$\begin{aligned}\boldsymbol{\Sigma}_k &= \sum_{i=1}^{M} w^i \left[\boldsymbol{\Sigma}_k^i + \boldsymbol{\mu}_k^i{\boldsymbol{\mu}_k^i}^T \right. \\ &\quad \left. - \boldsymbol{\mu}_k^i\boldsymbol{\mu}_k^T - \boldsymbol{\mu}_k{\boldsymbol{\mu}_k^i}^T + \boldsymbol{\mu}_k\boldsymbol{\mu}_k^T\right] \\ &= \sum_{i=1}^{M} w^i \left[\boldsymbol{\Sigma}_k^i + (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)(\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^i)^T\right] \end{aligned} \quad (49)$$

## REFERENCES

[1] P. D. Hanlon and P. S. Maybeck, "Multiple-model adaptive estimation using a residual correlation Kalman filter bank," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 393–406, Apr. 2000.
[2] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Autom. Control*, vol. AC-17, no. 4, pp. 439–448, Aug. 1972.
[3] W. I. Tam, K. N. Plataniotis, and D. Hatzinakos, "An adaptive Gaussian sum algorithm for radar tracking," *Signal Process.*, vol. 77, no. 1, pp. 85–104, 1999.
[4] K. S. Narendra and J. Balakrishnan, "Adaptive control using multiple models," *IEEE Trans. Autom. Control*, vol. AC-42, no. 2, pp. 171–187, Feb. 1997.
[5] M. Athans *et al.*, "The stochastic control of the F-8C aircraft using a multiple model adaptive control (MMAC) method—Part I: Equilibrium flight," *IEEE Trans. Autom. Control*, vol. AC-22, no. 5, pp. 768–780, Oct. 1977.
[6] B. Pourbabaee, N. Meskin, and K. Khorasani, "Sensor fault detection, isolation, and identification using multiple-model-based hybrid Kalman filter for gas turbine engines," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 4, pp. 1184–1200, Jul. 2016.
[7] S. Zhao, B. Huang, and F. Liu, "Fault detection and diagnosis of multiple-model systems with mismodeled transition probabilities," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5063–5071, Aug. 2015.
[8] D. Lainiotis, "Sequential structure and parameter-adaptive pattern recognition—I: Supervised learning," *IEEE Trans. Inf. Theory*, vol. AC-16, no. 5, pp. 548–556, Sep. 1970.
[9] T. Kirubarajan, Y. Bar-Shalom, K. R. Pattipati, and I. Kadar, "Ground target tracking with variable structure IMM estimator," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 1, pp. 26–46, Jan. 2000.
[10] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AC-34, no. 1, pp. 103–123, Jan. 1998.
[11] D. T. Magill, "Optimal adaptive estimation of sampled stochastic processes," *IEEE Trans. Autom. Control*, vol. AC-10, no. 4, pp. 434–439, Oct. 1965.
[12] F. Sims, D. Lainiotis, and D. Magill, "Recursive algorithm for the calculation of the adaptive Kalman filter weighting coefficients," *IEEE Trans. Autom. Control*, vol. AC-14, no. 2, pp. 215–218, Apr. 1969.
[13] D. Lainiotis, "Optimal adaptive estimation: Structure and parameter adaption," *IEEE Trans. Autom. Control*, vol. AC-16, no. 2, pp. 160–170, Apr. 1971.
[14] D. G. Lainiotis, "Partitioning: A unifying framework for adaptive systems, I: Estimation," *Proc. IEEE*, vol. 64, no. 8, pp. 1126–1143, Aug. 1976.
[15] G. Ackerson and K. Fu, "On state estimation in switching environments," *IEEE Trans Autom. Control*, vol. AC-15, no. 1, pp. 10–17, Feb. 1970.
[16] J. K. Tugnait and A. H. Haddad, "A detection-estimation scheme for state estimation in switching environments," *Automatica*, vol. 15, no. 4, pp. 477–481, Jul. 1979.
[17] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Autom. Control*, vol. AC-33, no. 8, pp. 780–783, Aug. 1988.
[18] X.-R. Li and Y. Bar-Shalom, "Multiple-model estimation with variable structure," *IEEE Trans. Autom. Control*, vol. 41, no. 4, pp. 478–493, Apr. 1996.
[19] K. P. Nelson, "Assessing probabilistic inference by comparing the generalized mean of the model and source probabilities," *Entropy*, vol. 19, no. 6, p. 286, 2017.
[20] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, vol. 21. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979, pp. 22–95.
[21] Y. Baram and N. R. Sandell, "Consistent estimation on finite parameter sets with application to linear systems identification," *IEEE Trans. Autom. Control*, vol. AC-23, no. 3, pp. 451–454, Jun. 1978.
[22] Z. Zhao and X. R. Li, "The behavior of model probability in multiple model algorithms," in *Proc. 8th IEEE Int. Conf. Inf. Fusion*, vol. 1, Jul. 2005, pp. 331–336.
[23] P. S. Maybeck and P. D. Hanlon, "Performance enhancement of a multiple model adaptive estimator," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 4, pp. 1240–1254, Oct. 1995.
[24] M. Magalhães and Z. Binder, "A true multimodel estimation algorithm," *IFAC Proc. Vol.*, vol. 20, no. 5, pp. 263–267, 1987.
[25] X. R. Li and Y. Zhang, "Numerically robust implementation of multiple-model algorithms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 1, pp. 266–278, Jan. 2000.

[26] P. S. Maybeck and R. D. Stevens, "Reconfigurable flight control via multiple model adaptive control methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, no. 3, pp. 470–480, May 1991.
[27] G. Battistelli and L. Chisci, "Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, 2014.
[28] W. Li and Y. Jia, "An information theoretic approach to interacting multiple model estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 1811–1825, Jul. 2015.
[29] S. J. Julier, T. Bailey, and J. K. Uhlmann, "Using exponential mixture models for suboptimal distributed data fusion," in *Proc. IEEE Nonlinear Statist. Signal Process. Workshop*, Sep. 2006, pp. 160–163.
[30] V. Shin, Y. Lee, and T.-S. Choi, "Generalized Millman's formula and its application for estimation problems," *Signal Process.*, vol. 86, no. 2, pp. 257–266, Feb. 2006.
[31] B. D. O. Anderson and J. B. Moore, "Detectability and stabilizability of time-varying discrete-time linear systems," *SIAM J. Control Optim.*, vol. 19, no. 1, pp. 20–32, 1981.
[32] W. W. Hager and L. L. Horowitz, "Convergence and stability properties of the discrete Riccati operator equation and the associated optimal control and filtering problems," *SIAM J. Control Optim.*, vol. 14, no. 2, pp. 295–312, 1976.
[33] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
[34] J. Ajgl, M. Šimandl, and J. Dunik, "Millman's formula in data fusion," in *Proc. 10th Int. PhD Workshop Syst. Control*, 2009, pp. 1–6.
[35] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. AC-15, no. 1, pp. 52–60, Feb. 1967.
[36] D. C. Dowson and B. V. Landau, "The Fréchet distance between multivariate normal distributions," *J. Multivariate Anal.*, vol. 12, no. 3, pp. 450–455, Sep. 1982.
[37] D. G. Lainiotis, "Adaptive dynamic modeling, risk assessment, failure detection and correction," in *Proc. Adv. Seminar Struct. Rel. (JRC-Ispra)*, 1984, pp. 165–193.
[38] X. R. Li and Z. Zhao, "Evaluation of estimation algorithms part I: Incomprehensive measures of performance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 4, pp. 1340–1358, Oct. 2006.
[39] J. G. Deshpande, T. N. Upadhyay, and D. G. Lainiotis, "Adaptive control of linear stochastic systems," *Automatica*, vol. 9, no. 1, pp. 107–115, Jan. 1973.
[40] XR. Li, X. Zwi, and Y. Zwang, "Multiple-model estimation with variable structure. III. Model-group switching algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 225–241, Jan. 1999.

**AKBAR ASSA** (S'08–M'15) received the Ph.D. degree in mechanical engineering from Ryerson University in 2015. He was a Post-Doctoral Fellow with Ryerson University in 2015 and served as a Professor with the Sheridan College from 2015 to 2016. He joined the Multimedia Laboratory, ECE Department, University of Toronto, in 2017, as a NSERC Post-Doctoral Fellow. His research interests span from robotics, visual servoing, and control systems to image processing, optimal estimation, and sensor fusion.

**KONSTANTINOS N. (KOSTAS) PLATANIOTIS** (S'90–M'92–SM'03–F'12) served as the Director of the Knowledge Media Design Institute, University of Toronto, from 2010 to 2012. He is currently a Professor and the Bell Canada Chair in multimedia with the Electrical and Computer Engineering Department, University of Toronto, Toronto, ON, Canada. He is also the Co-Founder and the inaugural Director-Research for the Identity, Privacy and Security Institute, University of Toronto. His research interests are knowledge and digital media design, multimedia systems, biometrics, image and signal processing, communications systems, and machine learning. He served as the Technical Co-Chair of the IEEE 2013 International Conference in Acoustics, Speech and Signal Processing, and the General Co-Chair for the 2017 IEEE GlobalSIP Conference. He is the General Co-Chair for the 2018 International Conference on Image Processing and the General Co-Chair for the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing. He was the IEEE Signal Processing Society Vice President for Membership from 2014 to 2016 and he has served as the Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS. He is a Registered Professional Engineer in the Province of Ontario and a fellow of the Engineering Institute of Canada.