18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

# Multiple-Model Hypothesis Testing Using Adaptive Representative Model

**Bao Liu** **Jian Lan**

Center for Information Engineering Science Research (CIESR)
The School of Electronic and Information Engineering
Xi'an Jiaotong University, Xi'an, Shaanxi 710049, P. R. China
xiaobei0077@163.com, lanjian@mail.xjtu.edu.cn

**X. Rong Li**

Department of Electrical Engineering
University of New Orleans
New Orleans, LA 70148, U.S.A
xli@uno.edu

***Abstract*—This paper presents a multiple-model hypothesis testing (MMHT) approach using a representative model (RM) for detecting unknown events that may have multiple distributions. It addresses various difficulties of MMHT for composite, multivariate, nondisjoint, and mis-specified hypothesis sets with correlated observations, and decides which region of the mode space covered by the model set is better. The model-set likelihood (MSL) based MMHT method (MMHT-MSL) is promising because of its efficiency and theoretical validity. The MSL is dominated by the likelihood of the closest-to-truth model in the model set as the sample size increases. However, the multiple-model approach usually intends to deal with all possible modes in the convex hull of the model set rather than only the models in the model set. Consequently, when mis-specification exists, this dominating model is not necessarily representative; that is, it is inappropriate for the model set rather than the region of the mode space covered by the model set. Our approach utilizes model-set adaptation (e.g., expected-mode augmentation and best model augmentation) to improve coverage ability of the model set, and then searches for the model which is closest to the truth under some criterion in the model-set-covered region as the RM. The RM based MMHT method (MMHT-RM) can be expected to provide a more efficient detection in the sense of minimizing the expected sample size subject to the error probability constraints. Moreover, in contrast to the MMHT-MSL, MMHT-RM is highly computationally efficient and easy to implement. Performance of MMHT-RM is evaluated for model-set selection problems in several scenarios. Simulation results demonstrate the effectiveness of the proposed MMHT-RM compared with MMHT-MSL.***

## I. Introduction

In statistics, hypothesis testing plays a fundamental role and has been studied extensively. Traditional hypothesis testing aims at determining one or more hypotheses possibly with uncertain parameters, which is thought to specify a distribution of a known structure. We can consider the traditional hypothesis testing as a single model hypothesis testing because it handles a single distribution case that may involve parametric uncertainties. In some cases, however, the distribution (including structure and parameters) is uncertain, which is a complex problem. In practice, the distribution may be partially known rather than completely unknown, and it may have multiple possible forms. This is referred to as the multi-distribution detection (MDD) problem, where existing parametric or nonparametric tests (see, e.g., [1], [2]) would not work well. Take the maneuver detection problem as an example. The measurements of "nonmaneuver" (straight and level motion at a constant velocity) follow a corresponding distribution, and "maneuver" essentially includes all other motion patterns. Some patterns may be known. For instance, different maneuver models correspond to different patterns whose measurements may have different distributions (including structures and parameters). However, the multiple-model (MM) approach (e.g., autonomous MM (AMM), cooperating MM, and variable-structure MM) [3] can utilize the partially known distributions (e.g., possible motion patterns) available in some cases. Accordingly, multiple-model hypothesis testing (MMHT) techniques (see, e.g., [4], [5], [6]) have been developed to deal with MDD problems. Through the MM approach, MMHT techniques formulate the MDD problem as one of special binary hypothesis testing. Contrast to the traditional hypothesis which usually specifies a distribution of a known structure with one or more unknown parameters, the hypothesis in MMHT may consist of different distributions, and we can consider it as a special composite hypothesis.

Now that the MDD problem is formulated as one of binary hypothesis testing by MMHT, sequential testing is preferred to the non-sequential testing primarily for the following reasons: the measurements are obtained sequentially; sequential tests are usually substantially more efficient in terms of the use of information in the measurements than non-sequential tests, and thus lead to a quicker decision at the same level of decision errors (see, e.g., [7], [8]). Wald's Sequential Probability Ratio Test (SPRT) [9] or 2-SPRT [10] as an (asymptotically) optimal decision rule comes to mind naturally. However, SPRT or 2-SPRT is only for simple hypothesis testing, but MMHT involves composite hypotheses. Fortunately, the ideas of MMHT using SPRT (MMSPRT) [4] and 2-SPRT (2-MMSPRT) [11] have been proposed, in which the model-set likelihood (MSL) [4] can convert each composite hypothesis to a simple one. So, MMSPRT and 2-MMSPRT belong to the class of MSL

Research supported in part by grant for State Key Program for Basic Research of China (973) (2013CB329405), the National Natural Science Foundation of China (61203120), NASA/LEQSF(2013-15)-Phase3-06 through grant NNX13AD29A, the Research Fund for the Doctoral Program of Higher Education of China (20110201120007), and the Fundamental Research Funds for the Central Universities of China.

based MMHT methods (MMHT-MSL). MMSPRT has a high efficiency of detection when the truth is included in the hypothesis sets, and meanwhile 2-MMSPRT can expect the asymptotic optimality in the sense of minimizing the maximum expected sample size subject to the error probability constraints when the mis-specified problem exists. The ideas of MMSPRT and 2-MMSPRT seem promising because of their efficiency and theoretical validity. These superiorities stem from the merits of the MSL: (a) theoretically, the MSL considers the model set as a whole, and this argument fits the description of the MDD problem well; (b) the MSL can convert each composite hypothesis in MMHT to a simple one, and then the optimal (SPRT) or asymptotically optimal (2-SPRT) test can be applied.

The likelihoods of the model sets from MM algorithms are naturally of major interest [4] in the MMHT-MSL methods. The MSL as the expected value of the model likelihood calculated through MM algorithms can represent the whole model set to some extent. However, it would be dominated by the closest-to-truth model (e.g., the one with the smallest Kullback-Leibler "distance" [12]) in the model set rather than the region of the mode space covered by the model set (i.e., the set of model "candidates") as the number of measurements increases, because the MM algorithms in existing MMHT-MSL methods are based on the basic assumption that one (and only one) model in the model set is true at each time instant.

In practice, the set of models used in the MM approach is usually used to cover the region of the mode space that is not deemed unlikely to be in effect. In other words, the model set is a set that represents this region of the mode space. We refer to this region as the "model-set-covered region". It is not unlikely for any point (i.e., mode or model) in this region to be in effect. In mathematics, we can formulate this region as the convex hull of the model set. Moreover, we consider that "mis-specification" exists when the assumption does not match the truth—in other words, the true mode is not in the model set. Consequently, the above behavior of the MM algorithm is restrictive once it privileges only one model in the model set, which is not the best one in the model-set-covered region, so the convergent model from the MSL may be unrepresentative when the model set is mis-specified. For example in Figure 1, with an increase in the sample size, the model-set likelihoods of $M_1$ and $M_2$ would be dominated by model likelihoods of $m_1$ and $m_2$, respectively. As pointed out in [5], the MMHT formulation is superior to the traditional M-ary hypothesis testing [13] formulation since the problem is detection rather than classification or identification. However, as shown above, the results of the MMHT-MSL methods in Figure 1 might be the same as the traditional M-ary hypothesis testing because of the inappropriate dominating problem as the sample size increases. On the other hand, $m_i$ $(i = 1, 2)$ as the best model in the original model set $M_i$ rather than $S_i$ (the region covered by $M_i$) is barely adequate to represent $S_i$ as a whole. In other words, the MMHT approach is to decide which model-set-covered region (rather than the original model set) is better. From this point of view, the MMHT-MSL methods may be inappropriate when mis-specification exists.

Since the convergent model of a model set in an MM algorithm is usually unrepresentative when there is a mis-specification, we should find a model that can represent the model-set-covered region and is as close to the true mode as possible at each time instant. This representative model (RM) is not necessarily in the original model set, but it should be related with the original model set under a certain criterion. This model relates to the data and original model set, so it maybe adaptive for a better accuracy to approximate the true mode as the sample size increases. This idea is also illustrated in Figure 1. As observed, $R_1$ and $R_2$ as the closest-to-truth model in the model-set-covered region under some criterion can represent well the model set as a whole for MMHT. The model set is designed to cover the truth; however, the convergence to an unrepresentative model implies that the truth is not included in the original model set. To improve the coverage ability of the model set, two model-set adaptation methods (expected-mode augmentation [14] and best model augmentation [15]) are applied in this paper. A model in the model-set-covered region that may be closest to the truth under some criterion is augmented to the original model set, and then an MM algorithm is run, and the model having the maximum probability is selected as the RM for the model-set-covered region. As the name implies, MMHT-RM aims to find an RM to represent the model-set-covered region under some criterion, such as $R_1$ and $R_2$ for $S_1$ and $S_2$, respectively, in Figure 1, and then the composite hypothesis in MMHT is converted to a simple one. Finally, an (asymptotically) optimal test can be applied.

Compared with the existing approaches for MDD problems, this paper makes the following main contributions: (a) A new MMHT method using a representative model is proposed to handle the MDD problem in dynamic systems. All existing MMHT methods (see, e.g., [4], [5], [6], [11]) are based on the model-set likelihood, and none of them consider the inappropriate dominating problem of the MSL described before when mis-specification exists. So, our method is more general and suitable for MDD problems. The RM as the closet-to-truth model in the model-set-covered region under some criterion can help MMHT to make a quicker decision than the MSL. Specifically, 2-MMSPRT [11] needs some approximations, especially for calculating the model-set likelihood of dynamic

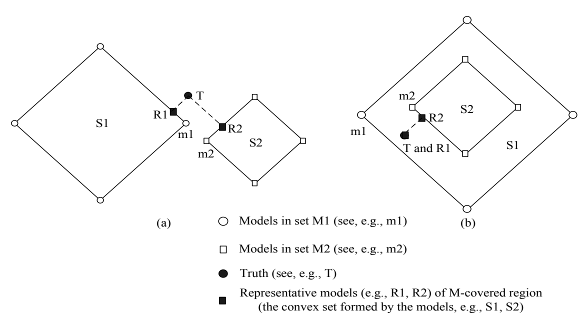



Figure 1: Some examples of MDD problems

systems and the nominal middle model between two model sets; however, this is usually not needed in the RM approach. (b) The proposed method provides a basic idea to derive an RM to replace the whole model set in order to make a quickest decision for MDD problems. Two versions of the RM are presented for MMHT under different criteria describing different relationships between the model set and the RM. Since the existing approaches do not consider the inappropriate dominating problem of the MSL in the presence of mis-specification, the RM approach provides a new solution to this important problem.

This paper is organized as follows. The MMHT techniques and the problem formulation are presented in Section II. Section III presents the RM approach with two specific model-set adaptation methods. Section IV provides simulation results of some illustrative examples for the MMHT-RM compared with the MMHT-MSL. The conclusions are presented in Section V.

## II. Multiple-model hypothesis testing and problem formulation

As an emerging approach, MMHT has received much attention in recent years due to its unique power and success in handling MDD problems with structural (and parametric) uncertainties, in converting each composite hypothesis testing to a simple one, and in applying to the fields of maneuver detection, fault detection and isolation, signal processing, etc. (see, e.g., [4], [5], [6], [11]). To our knowledge, however, they are all based on model-set likelihood calculated by MM algorithms without considering the inappropriate dominating problem. This section briefly describes MMHT techniques, and then presents problems to be studied in this paper.

### A. MMHT Formulation of MDD Problems

In an MDD problem, such as detection of a signal [16], fault [17], and maneuver [18] [19], what we want to decide is whether there is a thing (e.g., a signal) or not, rather than what kind of thing is present. The distribution of the thing to be detected might be partially known, and it usually depends on some parameters related to the event class and special behavior in typical situations. The MMHT approach formulates this uncertainty as a special binary hypothesis test with two model families in the MM framework [11]:

$$H_0 \quad : \quad \begin{cases} H_{01} : z_k \sim f\left(z_k | m_0^1, z^{k-1}\right) \\ \qquad \vdots \\ H_{0n} : z_k \sim f\left(z_k | m_0^n, z^{k-1}\right) \end{cases} \tag{1}$$

$$H_1 \quad : \quad \begin{cases} H_{11} : z_k \sim f\left(z_k | m_1^1, z^{k-1}\right) \\ \qquad \vdots \\ H_{1r} : z_k \sim f\left(z_k | m_1^r, z^{k-1}\right) \end{cases} \tag{2}$$

where $z^k \triangleq (z_1, \ldots, z_k)$ is the measurements up to time $k$, $M_0 = \left\{m_0^1, \ldots, m_0^n\right\}$ and $M_1 = \left\{m_1^1, \ldots, m_1^r\right\}$ are sets of possible models without and with the thing to be detected, respectively, and $f\left(z_k | m, z^{k-1}\right)$ is the likelihood function of model $m$ conditioned on $z^{k-1}$. Note that the true model is not necessarily in the hypothesis sets, so this is a typical mis-specified problem.

Apparently, we should solve this hypothesis testing problem in two phases: First, convert this composite hypothesis in (1)–(2) to a simple one (e.g., using model-set likelihoods). Second, choose a good test (e.g., MMSPRT [4] and 2-MMSPRT [11]) according to the prior knowledge (e.g., mis-specification is possible or not) to make a quickest decision. This paper focuses on the first task.

### B. Existing Methods

In the MMHT formulation of MDD problems, the MM approach gets around the distribution/model uncertainty and formulates this uncertainty as part of a composite hypothesis. In order to make a good decision, the MMHT-MSL methods utilize the model-set likelihood to convert the above composite hypothesis to a simple one. Therefore, the marginal and joint likelihoods of a model set involved are naturally of major interest in the MMHT-MSL methods.

The marginal likelihood of a model set $M_j$ ($j = 0, 1$) is defined as [4]

$$\begin{aligned} L_k^{M_j} &\triangleq f\left[\tilde{z}_k | s \in M_j, z^{k-1}\right] \\ &= \sum_{m_i \in M_j} f\left[\tilde{z}_k | s = m_i, z^{k-1}\right] \\ &\qquad \cdot P\left\{s = m_i | s \in M_j, z^{k-1}\right\}, \end{aligned} \tag{3}$$

where $\tilde{z}_k$ is the measurement residual, $s$ is the true mode, $m_i$ is the $i$th model, and $s = m_i$ denotes the event that model $m_i$ matches the system mode $s$. The predicted probability $P\left\{s = m_i | s \in M_j, z^{k-1}\right\}$ for each model $m_i$ in the set $M_j$ is available from an MM algorithm. The joint likelihood of the model set $M_j$ is defined as [4]

$$L_{M_j}^k \triangleq f\left[\tilde{z}^k | s \in M_j\right]. \tag{4}$$

A subscript $k$ and a superscript $k$ are used for quantities at time $k$ and through time $k$, respectively. Finally, we need the joint likelihood ratio

$$\Lambda^k \triangleq \frac{L_{M_1}^k}{L_{M_0}^k} = \prod_{k_0 \leq \kappa \leq k} \frac{L_\kappa^{M_1}}{L_\kappa^{M_0}} \tag{5}$$

in MMSPRT (see, e.g., [4], [5], [6]) and

$$\hat{\Lambda}_j^k \triangleq \frac{L_{m_\kappa^M}^k}{L_{M_j}^k} = \prod_{k_0 \leq \kappa \leq k} \frac{L_\kappa^{m_\kappa^M}}{L_\kappa^{M_j}}, j = 0, 1 \tag{6}$$

in 2-MMSPRT [11] if the residual sequence $\langle \tilde{z}_k \rangle$ is white, where $k_0$ is the test start time, and $m_\kappa^M$ is the nominal middle model between model sets $M_0$ and $M_1$ at time $k$. $m_\kappa^M$'s likelihood function $L_k^{m_k^M}$ satisfies the following equation:

$$\frac{\log \alpha_0^{-1}}{\mathrm{KL}(L_k^{m_k^M}, L_k^{M_0})} = \frac{\log \alpha_1^{-1}}{\mathrm{KL}(L_k^{m_k^M}, L_k^{M_1})} \tag{7}$$

where $\alpha_0 = \alpha$ and $\alpha_1 = \beta$ are the error probability constraints [20]

$$P\left\{\text{“}H_1\text{”} | H_0\right\} \leq \alpha \text{ and } P\left\{\text{“}H_0\text{”} | H_1\right\} \leq \beta, \tag{8}$$

and $\mathrm{KL}(f_1(x), f_0(x)) = \int [\log \frac{f_1(x)}{f_0(x)}] f_1(x) dx$ is the Kullback-Leibler information [12].

The use of model-set likelihoods satisfies the SPRT's and 2-SPRT's simple hypothesis and (approximate) independence requirements (through the forced identical independent distribution method [11]). For the MM detection problem (1)–(2) (with consideration of model mis-specification) subject to the error probability constraints (8), the following MMHT-MSL method is of (asymptotically) optimal efficiency (the proof can be found in [4] and [11]):

1) choose $M_1$ if $\Lambda^k \geq A$ in MMSPRT or $\hat{\Lambda}_0^k \geq A_0$ in 2-MMSPRT, where $A = \frac{1-\beta}{\alpha}$ and $A_0 = \alpha^{-1}$ (acceptance region);

2) choose $M_0$ if $\Lambda^k \leq B$ in MMSPRT or $\hat{\Lambda}_1^k \geq A_1$ in 2-MMSPRT, where $B = \frac{\beta}{1-\alpha}$ and $A_1 = \beta^{-1}$ (rejection region);

3) otherwise, continue to test using more observations (continuation region).

### C. Problem Formulation

This section presents the following challenges (see, e.g., [4], [11]) of the MMHT formulation for the MDD problem. (a) $H_0$ and $H_1$ are not necessarily disjoint: $M_0 \cap M_1 \neq \emptyset$. The traditional hypothesis testing does not consider such a nondisjoint case. (b) Both $H_0$ and $H_1$ are composite, which can be seen from (1) and (2). Few optimal tests are available for the composite case. (c) Sequential observations are usually non-i.i.d. The sequence of likelihood ratios is not identically distributed [21]. (d) The parameters that characterize each model are usually multidimensional, especially in the dynamic systems. Most tests are for the one-dimensional case, and the efficiency of the multidimensional hypothesis testing needs to be verified. (e) The quickest decision and desired error probability constraints should be guaranteed no matter whether the candidate models are specified or not.

Fortunately, the MSL based methods (see, e.g., [4], [11]) gave some solutions to overcome these difficulties and can be used as a starting point. The 2-MMSPRT approach [11] considered the mis-specified problem; however, it does not address the inappropriate dominating problem of the model-set likelihood when mis-specification exists. This paper presents a more general approach based on a representative model for MMHT problems in meeting these challenges. Our final goals in this paper are: (a) Find an RM to represent the model-set-covered region at every time instant in order to convert each composite hypothesis to a simple one. (b) Make a decision as quickly as possible subject to the error probability constraints. (c) Decide on the hypothesis that is closest to the truth. It is appealing that the more likely hypothesis set has a higher probability to be accepted.

## III. The RM approach

Consider a discrete time linear system. The $i$th model of which in the MM method is:

$$x_{k+1} = F_k^{(i)} x_k + G_k^{(i)} w_k^{(i)} \tag{9}$$

$$z_k = H_k^{(i)} x_k + v_k^{(i)} \tag{10}$$

where $E(w_k^{(i)}) = \bar{w}_k^{(i)}$, $\mathrm{cov}(w_k^{(i)}) = Q_k^{(i)}$, $E(v_k^{(i)}) = \bar{v}_k^{(i)}$, and $\mathrm{cov}(v_k^{(i)}) = R_k^{(i)}$. Superscript $(i)$ denotes quantities pertinent to model $m^{(i)}$ in the model-set $M$, and $m_k^{(i)}$ denotes the event that model $m^{(i)}$ matches the true mode $s$ at time $k$:

$$m_k^{(i)} \triangleq \{s_k = m^{(i)}\}.$$

For simplicity and to capture the model close to the true mode as soon as possible, we do not consider mode jumps in this paper, and so the AMM algorithm is considered. The AMM algorithm (see, e.g., [3], [22]) is one of the most widely used MM methods, which suits the case of a time-invariant true mode, because of its algorithmic simplicity, computational efficiency, and good performance. One cycle of the AMM algorithm is given in Table I.

Table I One cycle of AMM algorithm [3]

| | |
|---|---|
| 1. Model-conditioned filtering (for $i = 1, 2, \ldots, l$): | |
| Predicted state: | $\hat{x}_{k\|k-1}^{(i)} = F_{k-1}^{(i)} \hat{x}_{k-1\|k-1}^{(i)} + G_{k-1}^{(i)} \bar{w}_{k-1}^{(i)}$ |
| Predicted covariance: | $P_{k\|k-1}^{(i)} = F_{k-1}^{(i)} P_{k-1\|k-1}^{(i)} \left(F_{k-1}^{(i)}\right)' + G_{k-1}^{(i)} Q_{k-1}^{(i)} \left(G_{k-1}^{(i)}\right)'$ |
| Measurement residual: | $\tilde{z}_k^{(i)} = z_k - H_k^{(i)} \hat{x}_{k\|k-1}^{(i)} - \bar{v}_k^{(i)}$ |
| Residual covariance: | $S_k^{(i)} = H_k^{(i)} P_{k\|k-1}^{(i)} \left(H_k^{(i)}\right)' + R_k^{(i)}$ |
| Filter gain: | $K_k^{(i)} = P_{k\|k-1}^{(i)} \left(H_k^{(i)}\right)' \left(S_k^{(i)}\right)^{-1}$ |
| Updated state: | $\hat{x}_{k\|k}^{(i)} = \hat{x}_{k\|k-1}^{(i)} + K_k^{(i)} \tilde{z}_k^{(i)}$ |
| Updated covariance: | $P_{k\|k}^{(i)} = P_{k\|k-1}^{(i)} - K_k^{(i)} S_k^{(i)} \left(K_k^{(i)}\right)'$ |
| 2. Mode probability update (for $i = 1, 2, \ldots, l$): | |
| Model likelihood: | $L_k^{(i)} \triangleq p\left[\tilde{z}_k^{(i)} \| m_{(i)}^k, z^{k-1}\right] = \mathcal{N}\left(\tilde{z}_k^{(i)}; 0, S_k^{(i)}\right)$ |
| Mode probability: | $\mu_k^{(i)} = \frac{\mu_{k-1}^{(i)} L_k^{(i)}}{\sum_j \mu_{k-1}^{(j)} L_k^{(j)}}$ |
| 3. Estimate fusion: | |
| Overall estimate: | $\hat{x}_{k\|k} = \sum_i \hat{x}_{k\|k}^{(i)} \mu_k^{(i)}$ |
| Overall covariance: | $P_{k\|k} = \sum_i [P_{k\|k}^{(i)} + \left(\hat{x}_{k\|k} - \hat{x}_{k\|k}^{(i)}\right)\left(\hat{x}_{k\|k} - \hat{x}_{k\|k}^{(i)}\right)'] \mu_k^{(i)}$ |

For the AMM algorithm, [23] pointed out that the probability of the model in $M$ with the smallest "distance" (which was given a Kullback-type information theoretic interpretation in [23]) to the truth tends to unity almost surely as time increases if this smallest "distance" is unique. Consequently, the model-set likelihood would converge to the likelihood of the (unique) model closest to the truth in the original model set (e.g., $m_1$ and $m_2$ in Figure 1).

To present a more general formulation and achieve a quickest detection, this section will address the challenges stated above using the representative model approach.

Before considering how to find a single model to represent a model-set-covered region, we should answer a question first: what kind of model is representative? In reality, the one has the maximum probability value is probably most widely used for representing a model set. However, to overcome the inappropriate dominating problem of the MSL, the model-set adaptation (MSA) should be used. The expected-mode augmentation (EMA) [14] and the best model augmentation (BMA) [15] are popular for MSA. The augmenting model at each time instant is closest to the truth under some criterion. For example, EMA augments the basic model set using the minimum mean-square error estimation with fixed models, while BMA uses the best model that minimizes the KL criterion for the augmentation [15]. So, this section will focus on these two directions to find the RM from the model-set-covered region.

### A. RM Based on EMA (RM-EMA)

An example of the good representative model is the expected value of the true mode since it is statistically closest to the true mode. This expected mode can be approximated by a sum of mode estimates weighted by their probabilities, readily available from the MM algorithm [14]. The theoretical foundation for the EMA approach has been provided in [14], and this paper will focus on its application to the derivation of the RM of a model-set-covered region.

Given a model set $M$, let $M^+ = M \cup \bar{m}$ denote the model set $M$ augmented by its expected mode $\bar{m}$. So, this expected mode is based on the model set $M_i$ in $H_i (i = 0, 1)$ conditioned on data through time $k$:

$$\begin{aligned}\bar{m}_{k|k}^{M_i} &:= E[s|s \in M_i, z^k] \\ &= \sum_{m_j \in M_i} m_j \mu_{k|k}^{(j)}\end{aligned} \tag{11}$$

or through time $k-1$:

$$\begin{aligned}\bar{m}_{k|k-1}^{M_i} &:= E[s|s \in M_i, z^{k-1}] \\ &= \sum_{m_j \in M_i} m_j \mu_{k|k-1}^{(j)},\end{aligned} \tag{12}$$

where $z^k \triangleq (z_1, \ldots, z_k)$ is the sequence of measurements, $\mu_{k|k-1}^{(j)} = P\{s = m_j | s \in M_i, z^{k-1}\}$ and $\mu_{k|k}^{(j)} = P\{s = m_j | s \in M_i, z^k\}$ denote the predicted and updated probabilities of model $j$ being the correct one, and $m_j$ is the parameter value that characterizes model $j$ $(j = 1, \ldots, l)$. The RM-EMA algorithm is given in Table II.

Table II RM-EMA ALGORITHM

| |
|---|
| 1. Initialization: |
| - For $i = 1, 2, \ldots, l$, $m^{(i)} \in M$ and $[\hat{x}_0^{(i)}, P_0^{(i)}, \mu_0^{(i)}] = [x_0, P_0, \mu_0]$ are given |
| - For $i = l+1$, $m^{(i)} = \bar{m}_0^M$ and $\mu_0^{(i)} = 1$ |
| 2. Model probability reinitialization: |
| - For $i = 1, 2, \ldots, l+1$, $\mu_0^{(i)} = \frac{\mu_0^{(i)}}{l+1}$ |
| 3. For $k = 1, 2, \ldots$, $M^+ = M \cup m^{(l+1)}$, and $i = 1, 2, \ldots, l+1$ |
| - Run AMM algorithm [3] |
| $S_k^{(i)} = H_k^{(i)} P_{k\|k-1}^{(i)} \left(H_k^{(i)}\right)' + R_k^{(i)}$ |
| $L_k^{(i)} = \mathcal{N}\left(\tilde{z}_k^{(i)}; 0, S_k^{(i)}\right)$ |
| $\mu_k^{(i)} = \frac{\mu_{k-1}^{(i)} L_k^{(i)}}{\sum_j \mu_{k-1}^{(j)} L_k^{(j)}}$ |
| - $\bar{m}_k^{M^+} = \sum_i \mu_k^{(i)} m_k^{(i)}$ |
| - Augmenting model: $m_{k+1}^{(l+1)} = \bar{m}_k^{M^+}$ |

In general, the expected-mode $\bar{m}$ is time varying and a generic cycle is $M_{k-1}^+ \to \bar{m}_{k|k}$ or $\bar{m}_{k|k-1}$, and the evolution of the augmented model set is $M_k^+ = M \cup \bar{m}_{k|k}$ or $M \cup \bar{m}_{k|k-1}$. Finally, the probability of this expected-mode $\bar{m}$ in $M^+$ tends to unity almost surely as time increases in the AMM algorithm. $\bar{m}$ as the model statistically closest to the true mode derived from the RM-EMA approach will give us the model in the model-set-covered region that is "closest" to the truth.

As pointed out in [14], this approach is quite general—it is valid for all problems where the above $\bar{m}$ is meaningful—and is simple to implement because $\bar{m}$ is readily available from the AMM estimator with little extra computation. However, the EMA approach can only handle the set of models which differ only in the parameters with the same physical meaning (additivity) [15]. The RM derived from the candidate models with different structures or parameters will be presented in the next section.

### B. RM Based on BMA (RM-BMA)

In the RM-EMA approach, the RM is derived as the expected value of the true mode, which is generated adaptively in real time as a probabilistically weighted sum of mode estimates over the augmented model set. This section will focus on the RM as the model which is closest to the truth under some criterion in the model-set-covered region. First, we should provide a criterion to serve as a general measure of the closeness between the true mode and the candidate models (with different structures or parameters) in the model-set-covered region. This section follows the idea of the BMA approach in [15].

Given a model set $M$ in which the models have different structures or parameters. We consider that the RM of the $M$-covered region $S$ can be selected as the one with the minimum Kullback-Leiber (KL) "distance" [12], [24], [15]:

$$\hat{m} = \underset{m^{(i)} \in S}{\arg\min} D(s, m^{(i)}), \tag{13}$$

where

$$\begin{aligned}D(s, m^{(i)}) &\triangleq D(p(z_k|s), p(z_k|m^{(i)})) \\ &= \int [\ln \frac{p(z_k|s)}{p(z_k|m^{(i)})}] p(z_k|s) dz_k,\end{aligned} \tag{14}$$

and $p(z_k|s)$ and $p(z_k|m^{(i)})$ are the probability density functions (pdfs) of $z_k$ conditioned on $s$ and $m^{(i)}$, respectively. $D(s, m^{(i)})$ quantifies the closeness of model $m^{(i)}$ to the true mode $s$ [15]. The realtime information provided by $M$, $S$, and $z^{k-1}$ should be considered, which leads to

$$\begin{aligned}&D(s, m^{(i)}) \\ &\triangleq D(p(z_k|M, S, s, z^{k-1}), p(z_k|M, S, m^{(i)}, z^{k-1})) \\ &= \int [\ln \frac{p(z_k|s, z^{k-1})}{p(z_k|m^{(i)}, z^{k-1})}] p(z_k|s, z^{k-1}) dz_k.\end{aligned} \tag{15}$$

For the modes (models) with Gaussian assumptions as in (9) and (10), $z_k$ can be assumed a Gaussian vector with the distribution $\mathcal{N}(z_k; \bar{z}_k, \Sigma)$. Assume

$$\begin{aligned}p(z_k|s, z^{k-1}) &= \mathcal{N}(z_k; \bar{z}_k^s, \Sigma_k^s) \\ p(z_k|m^{(i)}, z^{k-1}) &= \mathcal{N}(z_k; \bar{z}_k^i, \Sigma_k^i),\end{aligned} \tag{16}$$

where

$$
\begin{aligned}
\bar{z}_k^s &= E[z_k|s, z^{k-1}] \\
\Sigma_k^s &= E[(z_k - \bar{z}_k^s)(\cdot)'|s, z^{k-1}] \\
\bar{z}_k^i &= E[z_k|m_k^{(i)}, z^{k-1}] \\
&\approx E[z_k|m_k^{(i)}, \hat{x}_{k-1}, P_{k-1}] \\
\Sigma_k^i &= E[(z_k - \bar{z}_k^i)(\cdot)'|m_k^{(i)}, z^{k-1}] \\
&\approx E[(z_k - \bar{z}_k^i)(\cdot)'|m_k^{(i)}, \hat{x}_{k-1}, P_{k-1}],
\end{aligned} \tag{17}
$$

and $(\cdot)$ denotes the same term right before it. Then (15) becomes [15]

$$
\begin{aligned}
&D(s, m^{(i)}) \\
=&\frac{1}{2}\{\ln \frac{|\Sigma_k^i|}{|\Sigma_k^s|} - n + \text{tr}[(\Sigma_k^i)^{-1}(\Sigma_k^s + (\bar{z}_k^s - \bar{z}_k^i)(\cdot)')]\},
\end{aligned} \tag{18}
$$

where $n$ is the dimension of $z_k$, tr$[\cdot]$ means the trace of $[\cdot]$. Actually, for a practical process, the true mode $s$ is not known to us. As pointed out in [15], without given $s$, we can assume

$$
p(z_k|s, z^{k-1}) \approx \hat{p}[z_k] = p[z_k|M_{k|k-1}^+, z^{k-1}], \tag{19}
$$

where $M_{k|k-1}^+ = M \cup \hat{m}_{k-1}$. Under this assumption, the KL "distance" in (18) can be calculated by letting

$$
\begin{aligned}
\bar{z}_k^s &\approx E[z_k|M_{k|k-1}^+, z^{k-1}] \\
&= \sum_{m^{(j)} \in M_{k|k-1}^+} \hat{z}_{k|k-1}^j \mu_{k|k-1}^{(j)} \\
\Sigma_k^s &\approx E[(z_k - \bar{z}_k^s)(\cdot)'|M_{k|k-1}^+, z^{k-1}] \\
&= \sum_{m^{(j)} \in M_{k|k-1}^+} [P_{k|k-1}^j + (\hat{z}_{k|k-1}^j - \bar{z}_k^s)(\cdot)']\mu_{k|k-1}^{(j)},
\end{aligned} \tag{20}
$$

where

$$
\begin{aligned}
\hat{z}_{k|k-1}^j &= E[z_k|m_k^{(j)}, z^{k-1}] \\
P_{k|k-1}^j &= E[(z_k - \hat{z}_{k|k-1}^j)(\cdot)'|m_k^{(j)}, z^{k-1}] \\
\mu_{k|k-1}^{(j)} &\triangleq P\{m_k^{(j)}|M_{k|k-1}^+, z^{k-1}\}.
\end{aligned} \tag{21}
$$

The RM-BMA algorithm is given in Table III.

Table III RM-BMA ALGORITHM

| |
|---|
| 1. Initialization: |
| - For $i = 1, 2, \ldots, l$, $m^{(i)} \in M$, $[\hat{x}_0^{(i)}, P_0^{(i)}, \mu_0^{(i)}] = [x_0, P_0, \mu_0]$, and $S$ are given |
| - For $i = l+1$, $m^{(i)} = \hat{m}_0^M$ and $\mu_0^{(i)} = 1$ |
| 2. Model probability reinitialization: |
| - For $i = 1, 2, \ldots, l+1$, $\mu_0^{(i)} = \frac{\mu_0^{(i)}}{l+1}$ |
| 3. For $k = 1, 2, \ldots$, $M^+ = M \cup m^{(l+1)}$, and $i = 1, 2, \ldots, l+1$ |
| - Run AMM algorithm [3] |
| $\hat{z}_{k\|k-1}^{(i)} = H_k^{(i)} \hat{x}_{k\|k-1}^{(i)} + \bar{v}_k^{(i)}$ |
| $S_k^{(i)} = H_k^{(i)} P_{k\|k-1}^{(i)} \left(H_k^{(i)}\right)' + R_k^{(i)}$ |
| $L_k^{(i)} = \mathcal{N}\left(\tilde{z}_k^{(i)}; 0, S_k^{(i)}\right)$ |
| $\mu_k^{(i)} = \frac{\mu_{k-1}^{(i)} L_k^{(i)}}{\sum_j \mu_{k-1}^{(j)} L_k^{(j)}}$ |
| - $\bar{z}_k^s = \hat{z}_{k\|k-1} = \sum_i \hat{z}_{k\|k-1}^{(i)} \mu_{k-1}^{(i)}$ |
| $\Sigma_k^s = S_k$ |
| $= \sum_i [S_k^{(i)} + (\hat{z}_{k\|k-1} - \hat{z}_{k\|k-1}^{(i)})(\cdot)']\mu_{k-1}^{(i)}$ |
| - $\hat{m}_k^{M^+} = \underset{m^{(l+1)} \in S}{\arg\min} D(s, m^{(l+1)})$ |
| - Augmenting model: $m_{k+1}^{(l+1)} = \hat{m}_k^{M^+}$ |

*Remark* 1. Assumption (19) is the key in the RM-BMA algorithm, because it provides an approximation to the truth through the model set $M$ and the augmenting model from $S$ with the realtime information.

The probability of the augmenting model calculated through the above EMA or BMA based algorithm will tend to unity almost surely as time increases in the AMM algorithm [14] [15], so the augmenting model should be selected as the RM at this time. After the RM-EMA or RM-BMA, each composite hypothesis in (1)–(2) is converted to a simple one (an MM system is convert to an adaptive single model), and then the difficulties (a) and (b) in Section II.C are addressed naturally. For the last difficulty in Section II.C related to the decision rule, we can borrow the idea of MMSPRT [4] or 2-MMSPRT [11], and it is omitted in this paper due to space limitation.

## IV. Illustrative Examples

The model-set selection problem as a primary difficulty in MM algorithms has received much attention in recent years. A better model set will lead to better performance of an MM algorithm for a given problem. On the other hand, the MMHT problems can also be formulated as particular model-set selection problems from another standpoint. For example, consider CA (nearly constant acceleration) models with different inputs and CT (nearly constant turn) models with different turn rates (see, e.g., [14], [15]). A system has the following equations:

$$
x_{k+1} = F_k^j x_k + G_k^j \left(a_k^j + w_k^j\right) \tag{22}
$$

$$
z_k = H_k^j x_k + v_k^j, \; k = 0, 1, 2 \ldots \tag{23}
$$

where $x = [\mathsf{x}, \dot{\mathsf{x}}, \mathsf{y}, \dot{\mathsf{y}}]'$ denotes the target state, $z = [z_x, z_y]'$ the measurement, $v_k^j \sim \mathcal{N}(0, R)$ the measurement noise, and $H_k^j = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$. For the CA models, $a_k^j = [a_x, a_y]'$ $\left(m/s^2\right)$ are the inputs, $w_k^j \sim \mathcal{N}(0, Q)$ with $Q = 0.003^2 I$ $\left(m^2/s^4\right)$ the process noise, $F_k^j = \text{diag}[F_2, F_2]$ with $F_2 = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$, and $G_k^j = \text{diag}[G_2, G_2]$ with $G_2 = \begin{bmatrix} T^2/2 \\ T \end{bmatrix}$. For the CT models, $a_k^j = 0$, $G_k^j = I$, $w_k^j \sim \mathcal{N}(0, Q_j)$, and $F_k^j = F_{\text{CT}}^j$, where $F_{\text{CT}}^j$ and $Q_j$ are given on the next page with $j = 1, 2, 3$, $\omega_1 = -5/30$, $\omega_2 = -3/30$, and $\omega_3 = 3/30$. The parameters are $T = 1s$ and $R = 1250I$ $m^2$. All results are over 1000 Monte Carlo (MC) runs and are expressed in terms of the discrete time $k$ (i.e., multiples of $T$).

Consider two cases of the model-set selection problem with parameters given in Table IV. The true acceleration $a$ for Case 1 changes from $[-6, -6]'$ to $[6, 6]'$ with step length $[0.5, 0.5]'$, and the true turn rate $\omega$ for Case 2 changes from $-6/30$ to $6/30$ with step length $0.5/30$. The filter is initialized by $\hat{x}_0 \sim \mathcal{N}(x_0, P_0)$. We set the error probability constraints $\alpha = \beta = 0.01$, $x_0 = [8000\ m, 25\ m/s, 8000\ m, 200\ m/s]'$ for Case 1 [14], and $x_0 = [8000\ m, 600\ m/s, 8000\ m, 600\ m/s]'$ for Case 2 [15]. In Case 2 the radius of the turn $r_t = 18000\sqrt{2}$

m and its turn rate $\omega_t = 1/30$ rad/s.

Table IV PARAMETERS OF ALL EXAMPLES

| | Set for hypothesis $H_0$ | Set for hypothesis $H_1$ | Range of true $a$ and $\omega$ |
|---|---|---|---|
| Case 1 | $M_0 = \{\text{CA } (a = [-5,-5]'), \text{CA } (a = [-3,-3]')\}$ | $M_1 = \{\text{CA } (a = [5,5]'), \text{CA } (a = [3,3]')\}$ | $a$ from $[-6,-6]'$ to $[6,6]'$ |
| Case 2 | $M_0 = \{\text{CT } (\omega = -5/30), \text{CT } (\omega = -3/30)\}$ | $M_1 = \{\text{CT } (\omega = 3/30), \text{CA } (a = [-100,100]')\}$ | $\omega$ from $-6/30$ to $6/30$ |

As pointed out in [15], BMA can be viewed as a generalization of EMA, and they have practical equivalence. The models in Case 1 differ only in parameters and their weighted sum is meaningful, so only the RM-EMA is considered for Case 1. In Case 2 the models have different structures, and the RM-EMA can not handle it, so only the RM-BMA is considered.

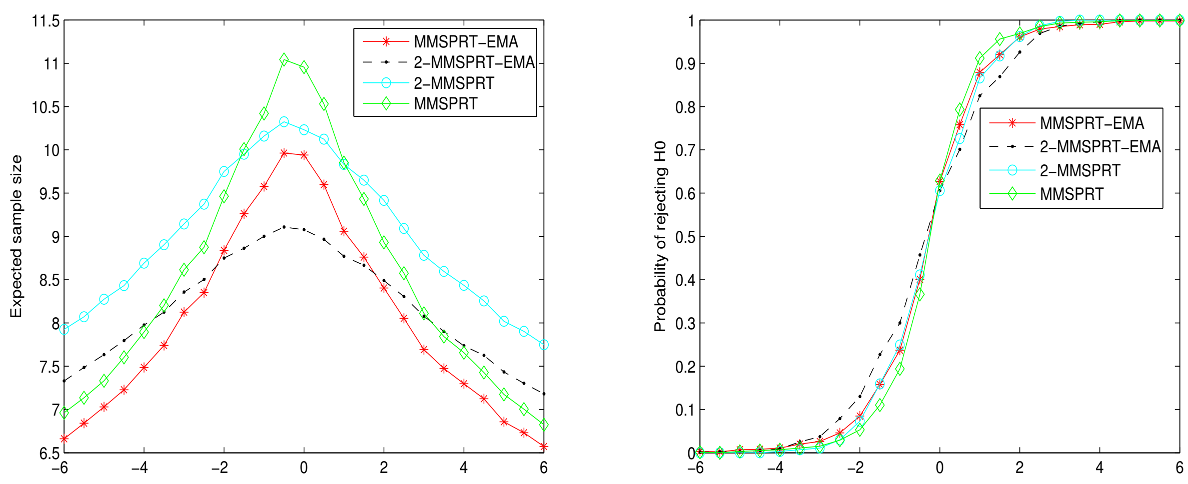
As is clear from Figure 2 and Figure 3, and not surprisingly, under the error probability requirements, the expected sample size of MMHT-RM (MMHT-EMA or MMHT-BMA) is smaller than MMHT-MSL's. Second, the simulation results confirm the (asymptotic) efficiency of the SPRT based and 2-SPRT based MMHT methods whether using the MSL or RM. Third, the RM approach lead to the decision which is the closest-to-truth one from the probability of rejecting $H_0$ (which is defined in Table IV). We can see that the more likely model set has a higher probability to be accepted. Finally, the proposed RM approach addresses the challenges stated in Section II, and thus it is an effective method for multi-distribution detection problems, especially for the mis-specified case.

## V. CONCLUSION

This paper makes contributions to the multi-distribution detection problems when mis-specification exists. The MMHT approach formulates this problem as one of special binary composite hypothesis testing. To overcome the inappropriate dominating problem of MSL based MMHT methods, a novel and integrated approach based on the adaptive representative model has been proposed in a general setting. Two specific RM methods with model-set adaptation have been derived to convert each composite hypothesis to a simple one, i.e., RM-EMA (for the models differently only in parameters) and RM-BMA (for the models with different parameters or structures). Through the proposed RM approach, an MM system is convert to a single model (although adaptive), and thus some challenges (e.g., composite, nondisjoint, and mis-specified hypothesis sets) in MMHT techniques for the MDD problem can be circumvented.

Several scenarios for the model-set selection problem involving the models with different parameters and structures have been simulated to compare MMHT-RM with MMHT-MSL. Simulation results demonstrate that MMHT-RM can address the inappropriate dominating problem in the MMHT-MSL methods and it can be expected to make a good decision.

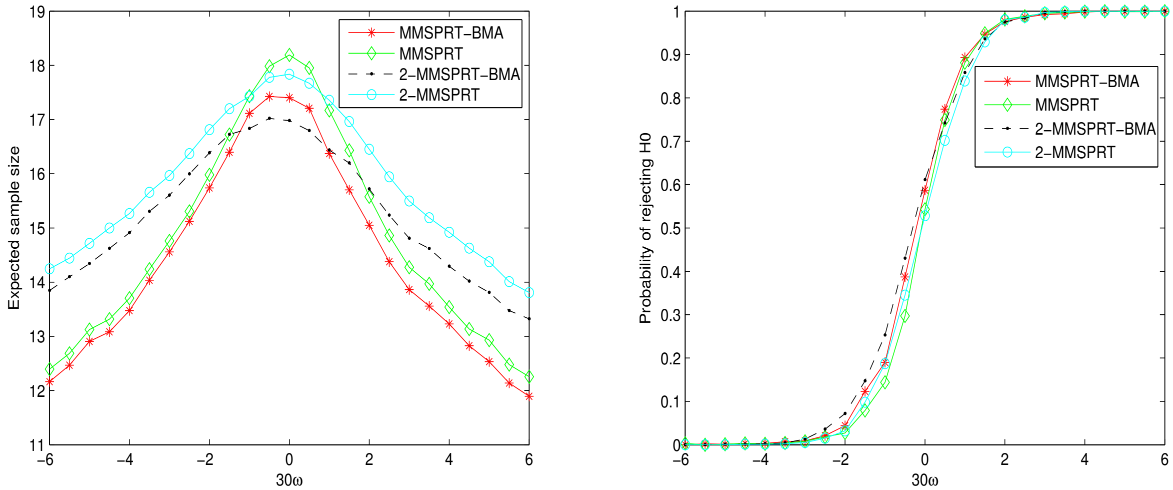
$$F_{CT}^j = \begin{bmatrix} 1 & \frac{\sin\omega_j T}{\omega_j} & 0 & -\frac{1-\cos\omega_j T}{\omega_j} \\ 0 & \cos\omega_j T & 0 & -\sin\omega_j T \\ 0 & \frac{1-\cos\omega_j T}{\omega_j} & 1 & \frac{\sin\omega_j T}{\omega_j} \\ 0 & \sin\omega_j T & 0 & \cos\omega_j T \end{bmatrix}, Q_j = 0.001 \begin{bmatrix} \frac{2(\omega_j-\sin\omega_j T)}{\omega_j^3} & \frac{1-\cos\omega_j T}{\omega_j^2} & 0 & \frac{\omega_j T-\sin\omega_j T}{\omega_j^2} \\ \frac{1-\cos\omega_j T}{\omega_j^2} & T & -\frac{\omega_j T-\sin\omega_j T}{\omega_j^2} & 0 \\ 0 & -\frac{\omega_j T-\sin\omega_j T}{\omega_j^2} & \frac{2(\omega_j-\sin\omega_j T)}{\omega_j^3} & \frac{1-\cos\omega_j T}{\omega_j^2} \\ \frac{\omega_j T-\sin\omega_j T}{\omega_j^2} & 0 & \frac{1-\cos\omega_j T}{\omega_j^2} & T \end{bmatrix}$$




(a) Expected sample size

(b) Probability of rejecting $H_0$ (see Table IV)

Figure 2: Performance comparison between MMHT-MSL and MMHT-EMA for Case 1




(a) Expected sample size

(b) Probability of rejecting $H_0$ (see Table IV)

Figure 3: Performance comparison between MMHT-MSL and MMHT-BMA for Case 2

In summary, the proposed RM based MMHT method is general for MDD problems and also computationally feasible. Moreover, it is conceptually simple, generally applicable, easily implementable, and significantly superior to the existing methods.

## REFERENCES


[1] A. Stuart, J. K. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics: Volume 2A—Classical Inference and the Linear Model*, 6th ed. London: Hodder Arnold, 1999.

[2] J. D. Gibbons and S. Chakraborty, *Nonparametric Statistical Inference*, 4th ed. New York: Marcel Dekker, 2003.

[3] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking—Part V: Multiple-model methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1255–1321, 2005.

[4] X. R. Li, "Multiple-model estimation with variable structure—Part II: Model-set adaptation," *IEEE Trans. Automat. Control*, vol. 45, pp. 2047–2060, 2000.

[5] J. F. Ru, X. R. Li, and V. P. Jilkov, "Multiple model detection of target maneuvers," in *Proc. Signal Data Process. Small Target*, vol. 5913, 2005, pp. 100–108.

[6] Y. Liu and X. R. Li, "Sequential multiple-model detection of target maneuver termination," in *Proc. 14th Int. Conf. Information Fusion*, Chicago, Illinois, USA, 2011, pp. 873–880.

[7] X. R. Li and L. Wang, "Fault detection using sequential probability ratio test," in *Proc. Power Engineering Society 1999 Winter Meeting*, vol. 2, 1999, pp. 938–943.

[8] R. Chandramouli and N. D. Memon, "On sequential watermark detection," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1034–1044, April 2003.

[9] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.

[10] G. Lorden, "2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size," *Ann. Statist.*, vol. 4, pp. 281–291, 1976.

[11] B. Liu, J. Lan, and X. R. Li, "Multiple-model hypothesis testing based on 2-SPRT," in *Proc. 2015 Amer. Control Conf.*, 2015.

[12] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 55, pp. 79–86, 1951.

[13] S. Zacks, *The Theory of Statistical Inference*. New York: Wiley, 1971.

[14] X. R. Li, V. P. Jilkov, and J. Ru, "Multiple-model estimation with variable structure—Part VI: Expected-mode augmentation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, pp. 853–867, 2005.

[15] J. Lan, X. R. Li, and C. Mu, "Best-model augmentation for variable-structure multiple-model estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, pp. 2008–2025, 2011.

[16] X. Wang and D. Musicki, "Target tracking using energy based detections," in *Proc. 10th Int. Conf. Information Fusion*, Quebec, Canada, 2007.

[17] M. Efe and D. P. Atherton, "The IMM approach to the fault detection problem," in *Proc. 11th IFAC Symposium on System Identification*, Fukuoka, Japan, 1997.

[18] I. Hwang, H. Balakrishnan, and C. Tomlin, "Flight-mode-based aircraft conflict detection using a residual-mean Interacting Multiple Model algorithm." in *Proc. AIAA Guidance, Navigation, and Control Conf.*, Austin, TX, 2003.

[19] C. Yang, Y. Bar-Shalom, and C. F. Lin, "Discrete-time point process filter for mode estimation," *IEEE Trans. Automat. Control*, vol. 37, no. 11, pp. 1812–1816, 1992.

[20] A. Wald and J. Wolfowitz, "Optimum character of the Sequential Probability Ratio Test," *Ann. Math. Statist.*, vol. 19, pp. 326–339, 1948.

[21] Y. Liu and X. R. Li, "Performance analysis of Wald's SPRT with independent but non-stationary log-likelihood ratios," in *Proc. 14th Int. Conf. Information Fusion*, Chicago, Illinois, USA, 2011, pp. 145–152.

[22] P. S. Maybeck, *Stochastic Models, Estimation and Control*. New York: Academic Press, 1982.

[23] Y. Baram and N. R. Sandell, "An information theoretic approach to dynamical systems modeling and identification." *IEEE Trans. Automat. Control*, vol. 23, no. 1, pp. 61–66, 1978.

[24] S. Kullback, *Information Theory and Statistics*. Mineola, NY: Dover, 1968.