

Copyright © 2008 IEEE. Reprinted from IEEE Transactions on  
Aerospace and Electronic Systems, 2007; 43 (1):93-111

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Adelaide's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

# Integrated Track Maintenance for the PMHT via the Hysteresis Model

**SAMUEL J. DAVEY**, Member, IEEE  
Defence Science and Technology Organisation  
Australia

**DOUGLAS A. GRAY**, Member, IEEE  
The University of Adelaide  
Australia

**Unlike other tracking algorithms the probabilistic multi-hypothesis tracker (PMHT) assumes that the true source of each measurement is an independent realisation of a random process. Given knowledge of the prior probability of this assignment variable, data association is performed independently for each measurement. When the assignment prior is unknown, it can be estimated provided that it is either time independent, or fixed over the batch. This paper presents a new extension of the PMHT, which incorporates a randomly evolving Bayesian hyperparameter for the assignment process. This extension is referred to as the PMHT with hysteresis. The state of the hyperparameter reflects each model's contribution to the mixture, and thus can be used to quantify the significance of mixture components. The paper demonstrates how this can be used as a method for automated track maintenance in clutter. The performance benefit gained over the standard PMHT is demonstrated using simulations and real sensor data.**

Manuscript received December 20, 2004; revised October 16, 2005; released for publication June 29, 2006.

IEEE Log No. T-AES/43/1/889312.

Refereeing of this contribution was handled by V. Krishnamurthy.

Authors' addresses: S. J. Davey, Defence Science and Technology Organisation, ISRD 200, PO Box 1500, Edinburgh, South Australia 5111, Australia, E-mail: (samuel.davey@dsto.defense.gov.au); D. A. Gray, The University of Adelaide, Adelaide, Australia.

0018-9251/07/\$25.00 © 2007 IEEE

## I. INTRODUCTION

Multi-target tracking is a problem comprised of two parts: a tracking algorithm must be able to both associate ambiguous observations with dynamic models, and estimate the parameters of these models. In general, the estimation part of the tracking task is performed in much the same way for different tracking algorithms, depending on the specific models chosen. The data association task is then the portion in which competing algorithms differ [1, 2].

The probabilistic multi-hypothesis tracker (PMHT) is a relatively recent tracking approach derived by the application of expectation maximisation (EM) to multi-target tracking [3]. The PMHT offers an attractive alternative to more traditional tracking approaches because it easily accommodates model complexity (such as multiple target dynamics models for manoeuvring targets) [4–6], and because its computation requirements scale linearly with the problem size [3].

The main difference between the PMHT and other tracking algorithms is the assumed measurement model. Under the usual measurement model, there is an assumed prior processing stage that ensures that every target present produces, at most, one measurement. If the target is distributed over several sensor bins, a quasi-sufficient statistic is assumed to be achieved through locating the peak of this distributed response, its centroid, or some other summarising point. The result of this assumption is that there is, at most, one measurement belonging to each track. This makes the track-to-observation association process dependent between measurements because the assignment of one observation may alter the possible assignment options for the next. Joint assignment hypotheses must therefore be formed, and the number of these hypotheses grows more than combinatorially with the number of tracks and measurements, and exponentially with time.

The PMHT uses a different model, which treats the measurements as a collection of independent observations of a mixture. The PMHT assumes that the true assignment of each measurement is an independent random variable with a possibly unknown prior probability mass function. The result of this assumption is that the track-to-observation association is independent for each different observation. This independence leads to linear complexity in the problem size, and allows the PMHT to simply perform batch processing, which is a more complicated task under the standard model, due to the prohibitive growth in computational requirements. The independent assignment assumption conflicts with the more usual assumption of a single measurement per target. However, it should be recognised that the single measurement assumption is usually a choice, and results from the design of data preprocessing.

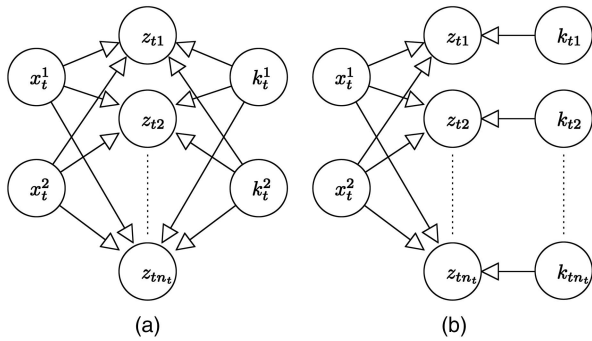


Fig. 1. Measurement model BINs. (a) Standard measurement model. (b) PMHT measurement model.

In problems where the the targets of interest are large compared with the sensor resolution, it may be advantageous to remove this preprocessing (e.g. see [7]).

The difference between the assumed measurement process for the standard tracking paradigm and the PMHT is highlighted by the Bayesian inference networks (BINs) shown in Fig. 1. Each random variable is represented by a circle in the BIN and the directed lines linking the circles show the conditional dependence of the variables. In the figure, there are two target models,  $\mathbf{x}_t^1$  and  $\mathbf{x}_t^2$ , and  $n_t$  measurements,  $\mathbf{z}_{t1} \dots \mathbf{z}_{tn_t}$ . Under the standard tracking measurement model, there is an assignment variable for each target model, i.e.,  $k_t^1$  and  $k_t^2$ , as illustrated in Fig. 1(a). All measurements are dependent on these indices. Unless merged measurements are considered, each measurement can be assigned to only one target. Hence the assignments are also dependent on each other. For the PMHT model, there is different assignment for each measurement, i.e.,  $k_{t1}, k_{t2}, \dots, k_{tn_t}$ , as illustrated in Fig. 1(b). These assignment indices are independent of each other. Under the standard model, all measurements must be used to jointly estimate each assignment. Under the PMHT model, each measurement is used independently to estimate the single assignment variable associated with it.

The assignment variables in the PMHT are treated as independent realisations of a random process. The probability mass function of the assignments is a possibly unknown prior. The standard PMHT restricts the prior to be either constant, or time independent. Existing PMHT research does not address the assignment prior; effort has instead focussed on the target dynamics model [4, 8, 6, 9], more sophisticated measurement models [10–12], and matters of practical significance for realistic implementations [7, 13, 14]. A model is proposed here which allows the assignment prior to be a randomly evolving quantity. This model is referred to as hysteresis and employs a scalar parameter per target to quantify each target’s contribution to the measurement mixture. Each scalar parameter follows a Markov chain whose statistics are known. The set of scalar parameters can be viewed

as a Bayesian hyperparameter for the assignment process.

In earlier work [15], it was shown that the existence model used for track maintenance with probabilistic data association based trackers such as [16, 17, 18, 19] can be incorporated into the PMHT through the assignment prior. A track that corresponds to a real target has a non-zero probability of producing a measurement, whereas a track that does not correspond to a target cannot produce a measurement. The work presented here demonstrates that the target existence model is actually the simplest possible realisation of a hysteresis model, where the scalar parameter for each target is binary.

In the remainder of this paper, the problem of track maintenance and the target existence model used to address it is reviewed. The standard PMHT algorithm is also reviewed to highlight the changes made here. The hysteresis model is introduced, and target existence is used as a special case to demonstrate its application. The PMHT with hysteresis is derived by treating the assignment hyperparameter as missing data and incorporating it into the PMHT assignment weights. The PMHT with hysteresis is then demonstrated on the track initiation problem, and compared with two maintenance schemes based on the standard PMHT (one similar to an  $M$  of  $N$  rule, and one motivated by model order estimation). The discrimination between valid and false tracks for each approach is measured for simulated scenarios and for recorded sensor data.

## II. TRACK MAINTENANCE

An important problem in realistic tracking applications is that of track maintenance. The number of targets in the scene will be a dynamic quantity, requiring an automated tracking system to be able to initiate tracks when new targets appear, and terminate tracks when existing targets disappear. Most tracking algorithms are not able to perform these tasks without incorporating ancillary functions [2, 1]. The typical tracking algorithm addresses the problem of recursively updating the state estimates for various targets under ambiguous measurement conditions. The temporal recursion formulation implicitly assumes that the algorithm is provided with prior information for each target, that each target has a corresponding track, and that there are no false tracks (i.e., tracks which do not correspond to targets).

The track maintenance problem is essentially one of model order estimation. One way to perform model order estimation is to overmodel the system, and then attempt to reject extraneous models using a statistical significance test. In the multi-target tracking framework, this amounts to forming more tracks than there are targets in the scene, and then rejecting the false tracks based on a track significance

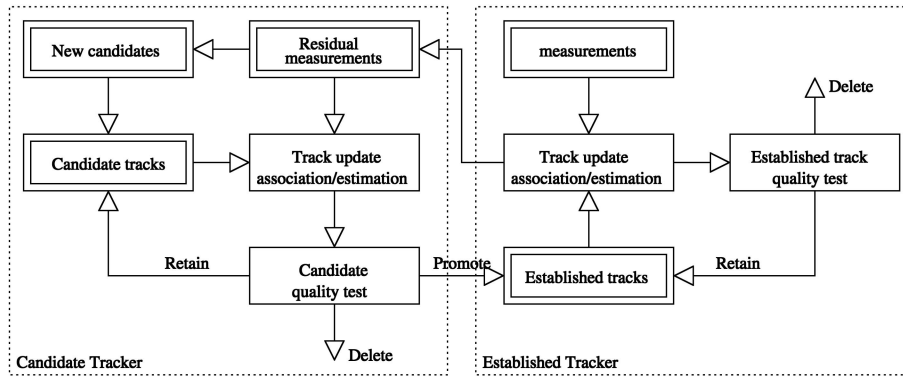


Fig. 2. General track maintenance flow diagram.

test. Under this approach, there will be established tracks which exist from some previous time, and which have previously been accepted as belonging to targets. There will also be candidate tracks, which are new tracks, initialised on whichever measurements are deemed to be likely to be target orientated. The candidate tracks are then accepted or rejected, based on a statistical test, and if accepted they become established tracks. Candidate tracks are not shown to the user, and act as method of vetting false tracks while keeping good detection performance. A general framework for track initiation by this method is illustrated in Fig. 2. In the figure, established tracks, which were initiated at some earlier time, get the first pickings of the sensor measurements. These tracks remove measurements that have a high assignment probability and the remaining measurements (the residual measurements) are used to form candidate tracks. The candidate tracks perform association and state estimation using the residual measurements. This prevents the candidate tracker from running a candidate track on the same target as is represented by an established track.

This is the approach used by the integrated probabilistic data association filter (IPDAF) which augments the target state to include a track quality variable, variously referred to as observability [16], existence [17], perceivability [18], and visibility [19]. The probability of this quality variable (which is binary) is used as a test statistic for track maintenance decisions, e.g. [20]. Under this strategy, track initiation is the same process as track termination, possibly with different thresholds for track significance.

The PMHT, in its standard form, does not provide for track maintenance [3]. It assumes that the number of targets is fixed over the data batch, that this number is known, and that prior information is available for each target. Two of these assumptions are easily relaxed: prior information can be approximated from data, and the fixed number of targets over a batch is unimportant if the PMHT is run using a small sliding window, or in time recursive form. However, the number of targets must be known. Using the

above framework, the assumption of a known number of targets can be relaxed, by over modelling and rejecting extraneous tracks. All that is required is a method for gauging the significance of candidate tracks.

The significance test for candidate tracks is a crucial part of the track initiation method, because it ultimately determines the delay statistics for establishing track on new targets, and the rate of false tracks. The hysteresis model provides an in-built significance test for candidate tracks with the PMHT: the assignment state quantifies the contribution of each model to the total mixture, and hence the significance of candidate tracks. For comparison, two approaches for track initiation with the standard PMHT are also considered.

#### A. Track Termination

Track maintenance requires the ability to start a track on a new target, and the ability to terminate track on an old target that no longer exists. This second problem is very similar to the first: to give good performance for track termination, the filter needs to discriminate between a valid track, and one that was valid but is now false. There are two differences between this and track initiation: firstly, the previous life of the track means that the termination problem starts with good initialisation, and secondly in termination one prefers to keep a track unless the filter is sure that it is gone, whereas in initiation one prefers to discard tracks unless the filter is sure they are valid. Overall this means that the same testing approach can be used, but that a different threshold is applied. Receiver operating characteristic (ROC) curves for track initiation show the discrimination of a particular test statistic, and so also quantify performance for termination.

### III. PMHT PARLANCE

The PMHT is formulated for a batch of  $T$  scans, where scan  $t$  contains  $n_t$  measurements collected at

time  $\tau_r$ . A time recursive algorithm can be obtained by using a batch length of  $T = 1$ . Let  $\mathbf{z}_{tr}$  denote the  $r$ th measurement in scan  $t$ , and define the set  $\mathbf{Z}_t \equiv \{\mathbf{z}_{t1} \dots \mathbf{z}_{tm_r}\}$  for each scan, and the batch set  $\mathbf{Z} \equiv \{\mathbf{Z}_1 \dots \mathbf{Z}_T\}$ .

There are  $M$  state models, which include target models and false measurement (clutter) models. The state of model  $m$  at scan  $t$  is denoted by  $\mathbf{x}_t^m$ . Define the set  $\mathbf{X}_t \equiv \{\mathbf{x}_t^1 \dots \mathbf{x}_t^M\}$  for each scan  $t$ , and the batch set  $\mathbf{X} \equiv \{\mathbf{X}_0 \dots \mathbf{X}_T\}$ . The states are assumed to be first-order Markov processes with known transition probabilities denoted as  $\psi_t^m(\mathbf{x}_t^m | \mathbf{x}_{t-1}^m)$ , and prior densities denoted as  $\psi_0^m(\mathbf{x}_0^m)$ . Each model has a known measurement probability density function, which is denoted by  $\zeta_t^m(\mathbf{z}_{tr} | \mathbf{x}_t^m)$  for the  $r$ th measurement at scan  $t$ .

Let the assignment index  $k_{tr}$  denote the model that gave rise to measurement  $\mathbf{z}_{tr}$  and define the sets  $\mathbf{K}_t \equiv \{k_{t1} \dots k_{tm_r}\}$  and  $\mathbf{K} \equiv \{\mathbf{K}_1 \dots \mathbf{K}_T\}$ . The value of each assignment index is unknown, and the estimation of them, or their probabilities is the data association problem.

The key difference between PMHT and other tracking algorithms is that the assignments  $k_{tr}$  are now treated as identically distributed independent realisations of a prior probability mass. The prior probability that the  $r$ th measurement at scan  $t$  is due to model  $m$  is denoted by  $\pi_t^m$ . Notice that this is independent of  $r$  since the  $k_{tr}$  are identically distributed ( $r$  is an arbitrary ordering of the measurements). Let  $\mathbf{\Pi}_t \equiv (\pi_t^1, \dots, \pi_t^M)$  and  $\mathbf{\Pi} \equiv (\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_T)$  be the per scan and batch sets of the prior values, respectively. Under the standard PMHT, the values of  $\mathbf{\Pi}_t$  are assumed to be independent of those of  $\mathbf{\Pi}_s$  for  $t \neq s$ .

The standard PMHT algorithm proceeds as follows.

- 1) Assume an initial state sequence,  $\hat{\mathbf{x}}_0^m$ , and an initial prior probability mass function,  $\hat{\pi}_0^m$ , for all models at each scan, i.e.,  $t = 0 \dots T$ ,  $m = 1 \dots M$ .
- 2) Calculate assignment probabilities for each model and measurement combination for every scan in the batch.
- 3) Refine the state and prior estimates using the measurement assignments, obtaining new  $\hat{\mathbf{x}}_t^m$  and  $\hat{\pi}_t^m$ .
- 4) Repeat steps 2 and 3 until convergence.

When the state statistics are linear and Gaussian (i.e.,  $\psi_t^m(\mathbf{x}_t^m | \mathbf{x}_{t-1}^m)$ ,  $\psi_0^m(\mathbf{x}_0^m)$ , and  $\zeta_t^m(\mathbf{z}_{tr} | \mathbf{x}_t^m)$  are linear Gaussian functions), then the state estimate can be calculated using a Kalman smoother [3]. When the  $\pi_t^m$  are temporally independent, then their values are estimated using the relative frequency:

$$\hat{\pi}_t^m = \frac{1}{n_t} \sum_{r=1}^{n_t} w_{mtr} \quad (1)$$

where  $w_{mtr}$  is referred to as the assignment weight for measurement  $\mathbf{z}_{tr}$  and model  $m$ . This weight is the probability of the missing data (the assignment  $k_{tr}$ ) given the observations, that is the posterior probability that  $k_{tr} = m$ , whereas  $\pi_t^m$  is the prior. The weight is given by

$$w_{mtr} = \frac{\pi_t^{m(i)} \zeta_t^m(\mathbf{z}_{tr} | \mathbf{x}_t^{m(i)})}{\sum_{s=1}^M \pi_t^{s(i)} \zeta_t^s(\mathbf{z}_{tr} | \mathbf{x}_t^{s(i)})} \quad (2)$$

where  $\pi_t^{m(i)}$  and  $\mathbf{x}_t^{m(i)}$  are the estimated values from the  $i$ th (previous) iteration. The calculation of these weights is the second step in the algorithm description above.

#### IV. ASSIGNMENT STATE MODEL WITH HYSTERESIS

A state model for the assignment prior  $\mathbf{\Pi}$  is now introduced to give the PMHT the capability of tracking correlated dynamic changes in its values. This state model can be viewed as a hyperparameter for the assignment index  $k_{tr}$ . Let the scalar assignment state for model  $m$  at scan  $t$  be  $d_t^m$ . It will be seen to be convenient to choose the assignment state to be discrete (in order to achieve a problem solution). However, this is not a fundamental requirement of the hysteresis model. In the track existence problem,  $d_t^m$  is binary, corresponding to a target that does or does not exist.

Define the per scan set  $\mathbf{D}_t \equiv \{d_t^1 \dots d_t^M\}$  and the batch set  $\mathbf{D} \equiv \{\mathbf{D}_0 \dots \mathbf{D}_T\}$ . The assignment state is assumed to be a first-order Markov random process with known evolution probability density (or mass) function  $\Delta_t^m(d_t^m | d_{t-1}^m)$ . The prior distribution for the assignment state of each model is also assumed known and is denoted as  $\Delta_0^m(d_0^m)$ .

For the target existence problem,  $\Delta_t^m(d_t^m | d_{t-1}^m)$  is a two by two matrix, with elements corresponding to the probability that an existing target continues to exist, the probability that a nonexistent target remains nonexistent (often chosen to be unity) and their complements.  $\Delta_0^m(d_0^m)$  is the prior probability that a new track corresponds to a target that exists.

There are two main approaches which can now be used to relate the  $d_t^m$  to the priors  $\pi_t^m$ . These result from the normalisation requirement of the  $\pi_t^m$  values. Effectively, the introduction of  $M$  independent hyperparameters is incongruous with the  $M - 1$  degrees of freedom in the prior itself. The first option is to ensure proper normalisation by choosing the relation

$$\pi_t^m(\mathbf{D}_t) = \frac{\phi_t^m(d_t^m)}{\sum_{s=1}^M \phi_t^s(d_t^s)} \quad (3)$$

where the functions  $\phi_t^m(d_t^m)$  are assumed to be known. The functions  $\phi_t^s(d_t^s)$  represent scaled probabilities and so must be positive semi-definite functions, namely they may not give negative values for any  $d_t^s$ . The  $\pi_t^m$  values will form a valid probability mass when

$\phi_t^m(d_t^m) \geq 0$ , for all models  $m$ , and  $\sum_{s=1}^M \phi_t^s(d_t^s) \neq 0$ . The problem with this approach is that continuity in the  $d_t^m$  does not necessarily translate to smoothly varying  $\pi_t^m$  values. As an example, consider a mixture with ten models, each of which has discrete  $d_{t-1}^m = 1$  and  $\phi_t^m(d_t^m) = d_t^m / 10$ . All of the prior values at scan  $t-1$  are equal,  $\pi_{t-1}^m = 1/10$ . If the  $\Delta_t^m(d_t^m | d_{t-1}^m)$  are chosen so that steps of more than one have zero probability, then the  $d_t^m$  values can evolve over time only very slowly, a unit step being the smallest change possible for a discrete variable. Yet, if  $d_t^m = 0$  for  $m = 2 \dots M$  and  $d_t^1 = 1$  then the result is  $\pi_t^1 = 1$ , and  $\pi_t^m = 0$  for  $m = 2 \dots M$ . Small transitions in the state may translate to rapid  $\pi_t^m$  variations. The problem is that the value of  $\pi_t^m$  depends on all of the  $\mathbf{D}_t$  through normalisation. This means that the individual element  $d_t^m$  is not indicative of the relative importance of model  $m$  on its own: all of the  $\mathbf{D}_t$  are required. This is an undesirable property.

The second approach is to use additive normalisation. For the track maintenance problem, there are two types of models in the mixture: target models, which produce at most one measurement, and false detection models, which may produce many measurements. Assuming that there are many more false detections than target-originated measurements, the  $\pi_t^m$  values for target models will be low compared with the values for clutter models. Also, the targets will be expected to provide a persistent set of measurements, whereas the number of false measurements may fluctuate, so there is little value in trying to model dynamics in the clutter priors. Consequently, assume that  $M_X$  of the models have correlated prior values (i.e., are target models), and the remaining  $M_Y = M - M_X$  are independent from scan to scan, conditioned on the  $d_t^m$  (corresponding to clutter models). Arbitrarily, index the independent prior models  $m = 1 \dots M_Y$  and the dependent prior models  $m = M_Y + 1 \dots M$ . Assuming that the temporally dependent models have relatively low priors, write

$$\pi_t^m(\mathbf{D}_t) = \begin{cases} \phi_t^m(d_t^m) & M_Y < m \leq M \\ \sigma_t^m \left[ 1 - \sum_{s=M_Y+1}^M \phi_t^s(d_t^s) \right] & 1 \leq m \leq M_Y \end{cases} \quad (4)$$

where the parameters  $\sigma_t^m$  are the relative mixing proportions of the independent models, which are assumed to be unknown. The  $\sigma_t^m$  represent the prior probability that a measurement is due to independent model  $m$  given that it is due to one of the independent models, and are constrained by

$$0 \leq \sigma_t^m \leq 1 \quad (5)$$

$$\sum_{m=1}^{M_Y} \sigma_t^m = 1. \quad (6)$$

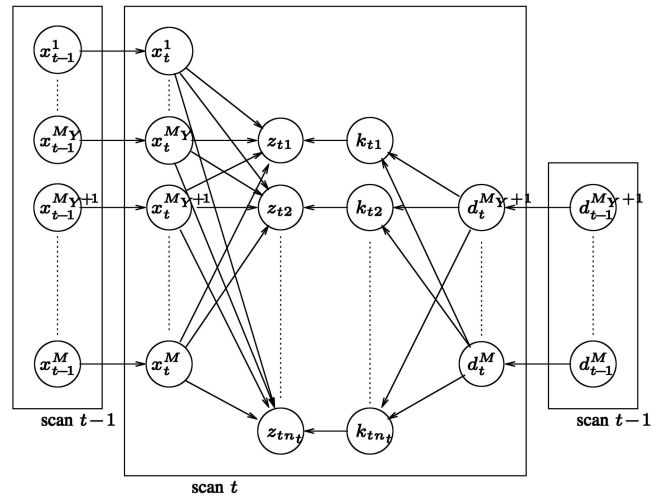


Fig. 3. BIN for PMHT with hysteresis.

For the case of target existence in clutter, the  $\sigma_t^m$  represent the relative strengths of different clutter models. If the clutter is homogenous, then it may be possible to use a single model for false measurements, and the  $\sigma_t^m$  variables are not required.

Define the sets  $\Sigma_t \equiv \{\sigma_t^1 \dots \sigma_t^{M_Y}\}$  and  $\Sigma \equiv \{\Sigma_1 \dots \Sigma_T\}$ .

Equation (4) ensures that the  $\pi_t^m$  sum to unity for any values of  $\phi_t^s(d_t^s)$ , provided that the constraint (6) holds. However, if some of the  $\phi_t^s(d_t^s)$  take values that are too large, then some of the  $\pi_t^m$  will not correspond to legal probabilities. For example, if the sum of the  $\phi_t^s(d_t^s)$  is greater than unity, the independent model priors will all be negative. This is why it is necessary to assume that the priors for the dependent models are small. This assumption then corresponds to constraining the maximum values of each of the  $\phi_t^m(d_t^m)$  to be sufficiently low to ensure that all the  $\pi_t^m$  are positive.

The second approach ensures that dynamic constraints on the  $d_t^m$  translate to dynamic constraints on the  $\pi_t^m$  with  $\phi_t^m(d_t^m)$  acting as a mapping. The assumption of small  $\phi_t^m(d_t^m)$  is acceptable if the dependent models are target models and the independent models are clutter models in a dense clutter multi-target tracking environment. These difficulties in normalisation arise because the priors of the various models are being modelled as independent entities, whereas the prior probability mass is inherently joint. However, using independent models simplifies the prior transition dynamics.

The PMHT with hysteresis uses the second approach derived above. Fig. 3 shows a one scan slice of the BIN for the PMHT with hysteresis. The BIN demonstrates how the assignment state variables provide a dynamic model for the probability mass of  $K$ . This dynamic model is the key contribution of this work.

The assignment state parameter is assumed to occupy a finite discrete space. Without loss of generality, this space is assumed to contain  $M_D$  elements, with the assignment state taking an integer value between 0 and  $M_D - 1$ . The particular forms of the functions  $\phi_t^m(d_t^m)$  and  $\Delta_t^m(d_t^m | d_{t-1}^m)$  are chosen according to the specific problem to be solved. The PMHT algorithms derived here are applicable for any functions, when  $d_t^m$  is discrete. This is because discrete problems are solved in a direct numerical manner by algorithms such as the Viterbi algorithm, or the hidden Markov model smoother, without the problematic analytic integration required for optimal solution of continuous problems.

For the case of target existence, where the  $d_t^m$  is binary, then the functions  $\phi_t^m(d_t^m)$  are given by

$$\begin{aligned} \phi_t^m(0) &= 0 \\ \phi_t^m(1) &= \frac{Pd}{n_t}. \end{aligned} \quad (7)$$

#### A. PMHT with Hysteresis

The PMHT with hysteresis is derived in substantially the same way as the standard PMHT. The full derivation of the standard PMHT is given in [3]. Therefore, the derivation of the PMHT with hysteresis is merely outlined here, although detail is given where it deviates from the derivation in [3].

There are two approaches that can be used to develop an algorithm based on this model. Firstly, the assignment states  $\mathbf{D}$  can be treated as further missing data in an EM context. This means that the auxiliary function will be the expectation over the assignments  $\mathbf{K}$  and the assignment states  $\mathbf{D}$ . In this case, the algorithm calculates the probabilities of the  $d_t^m$ , much as the standard PMHT calculates the probabilities of the assignments,  $k_{tr}$  (these are the weights,  $w_{mtr}$ ).

The second method of solution is to estimate the assignment states. In this case, the auxiliary function becomes a function of the target states and the assignment states, dependent on their values from a previous iteration. Under this approach, the assignment state is treated in a similar way to the target state estimates. An initial assignment state sequence is assumed, and this sequence is iteratively refined based on the measurement assignment weights. One could view this approach as augmenting the state vector to include the  $d_t^m$ .

Each of the above approaches was explored in [21], but only the first is presented here. The problem with the second approach, is that it requires the joint state estimation of  $M_X$  discrete variables, each of which occupies a state space with  $M_D$  elements. The joint state space thus has  $M_D^{M_X}$  elements, which will be impracticably large for all but the smallest  $M_D$ . For the hysteresis model to provide something useful, the  $M_D$  needs to be large in this case, and so the

algorithm is infeasible. In contrast, when the  $d_t^m$  are treated as missing data, then the algorithm calculates a probability mass, whose values are continuous even in the extreme case of  $d_t^m$  being binary.

#### V. HYSTERESIS AS MISSING DATA

The assignment states are treated as additional missing data. This means that the EM auxiliary function [3] for the modified PMHT will be the conditional expectation over the assignment states, and the assignments. Namely,

$$\begin{aligned} Q(\mathbf{X}, \Sigma | \mathbf{X}^{(i)}, \Sigma^{(i)}) \\ = \sum_{\mathbf{D}} \sum_{\mathbf{K}} \log\{P(\mathbf{X}, \mathbf{D}, \mathbf{K}, \mathbf{Z})\} P(\mathbf{D}, \mathbf{K} | \mathbf{X}^{(i)}, \mathbf{Z}) \end{aligned} \quad (8)$$

where the summation denoted  $\sum_{\mathbf{D}}$  is the sum over all possible permutations of the assignment states  $\mathbf{D}$ . Explicitly,

$$\sum_{\mathbf{D}} \{\cdot\} \equiv \sum_{d_0^{M_Y+1}=0}^{M_D-1} \sum_{d_0^{M_Y+2}=0}^{M_D-1} \dots \sum_{d_0^M=0}^{M_D-1} \sum_{d_1^{M_Y+1}=0}^{M_D-1} \dots \sum_{d_1^M=0}^{M_D-1} \{\cdot\}. \quad (9)$$

The BIN in Fig. 3 illustrates the independence assumptions of the filter. These assumptions are that the state models are independent of each other and that the assignment states are independent of each other. Under these assumptions, the complete data likelihood is

$$P(\mathbf{X}, \mathbf{D}, \mathbf{K}, \mathbf{Z}) = P(\mathbf{X})P(\mathbf{D})P(\mathbf{K} | \mathbf{D})P(\mathbf{Z} | \mathbf{K}, \mathbf{X}). \quad (10)$$

The assignments  $k_{tr}$  are conditionally independent given the assignment state vector for scan  $t$ ,  $\mathbf{D}_t$ . Hence

$$P(\mathbf{K} | \mathbf{D}) = \prod_{t=1}^T \prod_{r=1}^{n_t} \pi_t^{k_{tr}}(\mathbf{D}_t) \quad (11)$$

where  $\pi_t^{k_{tr}}(\mathbf{D}_t)$  is defined by (4).

The assignment states are independent of each other and are each first-order Markov processes, so

$$P(\mathbf{D}) = \prod_{m=M_Y+1}^M \left\{ \Delta_0^m(d_0^m) \prod_{t=1}^T \Delta_t^m(d_t^m | d_{t-1}^m) \right\}. \quad (12)$$

The remaining terms in (10) are the same as the standard PMHT and are given by

$$P(\mathbf{Z} | \mathbf{K}, \mathbf{X}) = \prod_{t=1}^T \prod_{r=1}^{n_t} P(\mathbf{z}_{tr} | k_{tr}, \mathbf{x}_t^{k_{tr}}) \quad (13)$$

and

$$P(\mathbf{X}) = \prod_{m=1}^M \left\{ \psi_0^m(\mathbf{x}_0^m) \prod_{t=1}^T \psi_t^m(\mathbf{x}_t^m | \mathbf{x}_{t-1}^m) \right\}. \quad (14)$$

The conditional probability of the assignment states and the assignments in (8) can be expanded via Bayes rule as

$$P(\mathbf{D}, \mathbf{K} | \mathbf{X}^{(i)}, \mathbf{Z}) = P(\mathbf{K} | \mathbf{D}, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D} | \mathbf{X}^{(i)}, \mathbf{Z}). \quad (15)$$

It is important to note that (15) is implicitly dependent on the clutter parameters  $\Sigma^{(i)}$  from the previous EM iteration. It is not a function of the unknown parameters  $\Sigma$ .

Substituting (10) and (15) into (8) gives

$$\begin{aligned} Q(\mathbf{X}, \Sigma | \mathbf{X}^{(i)}, \Sigma^{(i)}) &= \log\{P(\mathbf{X})\} + \sum_{\mathbf{D}} \log\{P(\mathbf{D})\}P(\mathbf{D} | \mathbf{X}^{(i)}, \mathbf{Z}) \\ &+ \sum_{\mathbf{D}} \sum_{\mathbf{K}} \log\{P(\mathbf{K} | \mathbf{D})\}P(\mathbf{K} | \mathbf{D}, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D} | \mathbf{X}^{(i)}, \mathbf{Z}) \\ &+ \sum_{\mathbf{D}} \sum_{\mathbf{K}} \log\{P(\mathbf{Z} | \mathbf{X}, \mathbf{K})\}P(\mathbf{K} | \mathbf{D}, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D} | \mathbf{X}^{(i)}, \mathbf{Z}). \end{aligned} \quad (16)$$

The first term in (16) is the log likelihood of the state sequence. This term is the same as for the standard PMHT. The second term is independent of  $\mathbf{X}$  and  $\Sigma$  and can be ignored. The third and fourth terms in (16) both involve a double sum over  $\mathbf{D}$  and  $\mathbf{K}$ . Note that the dependence upon  $\Sigma$  is implicit in the probability masses of the assignments  $P(\mathbf{K} | \mathbf{D})$ .

$$\begin{aligned} &\sum_{\mathbf{D}} \sum_{\mathbf{K}} \log\{P(\mathbf{K} | \mathbf{D})\}P(\mathbf{K} | \mathbf{D}, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D} | \mathbf{X}^{(i)}, \mathbf{Z}) \\ &= \sum_{\mathbf{D}} \sum_{\mathbf{K}} \left\{ \sum_{t=1}^T \sum_{r=1}^{n_t} \log\{P(k_{tr} | \mathbf{D}_t)\} \right\} P(\mathbf{K} | \mathbf{D}, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D} | \mathbf{X}^{(i)}, \mathbf{Z}) \\ &= \sum_{t=1}^T \left\{ \sum_{\mathbf{D}_t} \sum_{\mathbf{K}_t} \sum_{r=1}^{n_t} \log\{\pi_t^{k_{tr}}(\mathbf{D}_t)\} P(\mathbf{K}_t | \mathbf{D}_t, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z}) \right\} \\ &= \sum_{t=1}^T \left\{ \sum_{\mathbf{D}_t} \sum_{r=1}^{n_t} \sum_{k_{tr}=1}^M \log\{\pi_t^{k_{tr}}(\mathbf{D}_t)\} P(\mathbf{K}_t | \mathbf{D}_t, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z}) \right\} \\ &= \sum_{t=1}^T \sum_{\mathbf{D}_t} \sum_{m=1}^M \log\{\pi_t^m(\mathbf{D}_t)\} \sum_{r=1}^{n_t} P(k_{tr} = m | \mathbf{D}_t, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z}) \\ &= \sum_{t=1}^T \sum_{\mathbf{D}_t} \left\{ \sum_{m=M_Y+1}^M \log\{\phi_t^m(d_t^m)\} \sum_{r=1}^{n_t} w_{mtr}(\mathbf{D}_t) + \sum_{s=1}^{M_Y} \left[ \log\{\sigma_t^s\} + \log\left\{1 - \sum_{m=M_Y+1}^M \phi_t^m(d_t^m)\right\} \right] \sum_{r=1}^{n_t} w_{str}(\mathbf{D}_t) \right\} P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z}) \end{aligned} \quad (17)$$

The  $\log\{P(\mathbf{K} | \mathbf{D})\}$  term is the only part of (16) that is dependent on the clutter parameter  $\Sigma$ . Substituting (4) and (11), the  $\log\{P(\mathbf{K} | \mathbf{D})\}$  term in (16) can be written as shown in (17) where  $w_{mtr}(\mathbf{D}_t)$  is defined as

$$w_{mtr}(\mathbf{D}_t) \equiv P(k_{tr} = m | \mathbf{D}_t, \mathbf{X}^{(i)}, \mathbf{Z}). \quad (18)$$

Note that  $w_{mtr}(\mathbf{D}_t)$  is implicitly dependent on the iteration index  $i$ , but this index is suppressed to somewhat simplify notation.

The terms involving  $\phi_t^m(d_t^m)$  in (17) are constant, since they are summed over  $d_t^m$ . So, the only term in (17) that is significant is the one involving  $\sigma_t^s$ . Let

$$Q_{t\pi} \equiv \sum_{s=1}^{M_Y} \left\{ \log\{\sigma_t^s\} \sum_{r=1}^{n_t} w_{str} \right\} \quad (19)$$

where

$$\begin{aligned} w_{mtr} &= \sum_{\mathbf{D}} w_{mtr}(\mathbf{D}_t)P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z}) \\ &= \sum_{\mathbf{D}} P(k_{tr} = m | \mathbf{D}_t, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z}). \end{aligned} \quad (20)$$

Then, (17) can be written as

$$\begin{aligned} &\sum_{\mathbf{D}} \sum_{\mathbf{K}} \log\{P(\mathbf{K} | \mathbf{D})\}P(\mathbf{K} | \mathbf{D}, \mathbf{X}^{(i)}, \mathbf{Z})P(\mathbf{D} | \mathbf{X}^{(i)}, \mathbf{Z}) \\ &= \sum_{t=1}^T Q_{t\pi} + A_K \end{aligned} \quad (21)$$

where  $A_K$  is an irrelevant constant.



The final term in (21) depends on the measurement likelihood,  $\log\{P(\mathbf{Z} | \mathbf{X}, \mathbf{K})\}$ , and can be written as

$$\sum_{\mathbf{D}} \sum_{\mathbf{K}} \log\{P(\mathbf{Z} | \mathbf{X}, \mathbf{K})\} P(\mathbf{K} | \mathbf{D}, \mathbf{X}^{(i)}, \mathbf{Z}) P(\mathbf{D} | \mathbf{X}^{(i)}, \mathbf{Z})$$

$$= \sum_{m=1}^M \sum_{t=1}^T \sum_{r=1}^{n_t} \log\{\zeta_t^m(z_{tr} | x_t^m)\} w_{mtr}. \quad (22)$$

The auxiliary function can now be written as

$$Q(\mathbf{X}, \Sigma | \mathbf{X}^{(i)}, \Sigma^{(i)}) = \sum_{m=1}^M Q_X^m + \sum_{t=1}^T Q_{t\pi} + Q_D \quad (23)$$

where  $Q_D$  is constant, and the model term  $Q_X^m$  is given by

$$Q_X^m = \log\{\psi_0^m(\mathbf{x}_0^m)\} + \sum_{t=1}^T \log\{\psi_t^m(\mathbf{x}_t^m | \mathbf{x}_{t-1}^m)\}$$

$$+ \sum_{r=1}^{n_t} \log\{\zeta_t^m(\mathbf{z}_{tr} | \mathbf{x}_t^m)\} w_{mtr}. \quad (24)$$

Equation (24) is the same expression as is achieved for the standard PMHT except that the weights are obtained using (20). As is described in [3], (24) can be optimised using a Kalman smoother when the processes are linear and the random elements are Gaussian.

The  $Q_{t\pi}$  term in (23) is a function of the relative clutter probabilities  $\Sigma$  and is essentially the same as the prior auxiliary function for the standard PMHT. This is to be maximised subject to the constraint (6), namely

$$\sum_{m=1}^{M_y} \sigma_t^m = 1.$$

The maximisation can be achieved by using the Lagrangian  $L_{t\pi} = Q_{t\pi} + \lambda_t(1 - \sum_{k=1}^{M_y} \sigma_t^k)$ . Setting the derivative  $dL_{t\pi}/d\sigma_t^k = 0$  gives

$$\sigma_t^k = \frac{1}{\lambda_t} \sum_{r=1}^{n_t} w_{tr}^k \quad (25)$$

and reapplying the constraint gives the estimate for  $\sigma_t^k$  as

$$\sigma_t^k = \frac{\sum_{r=1}^{n_t} w_{ktr}}{\sum_{m=1}^{M_y} \sum_{r=1}^{n_t} w_{mtr}}. \quad (26)$$

This is the same form as the solution for  $\pi_t^m$  in the standard PMHT except that the weights are implicitly dependent on the assignment state probabilities, and the expression only includes the independent models  $m = 1 \dots M_y$ .

#### A. Assignment Weights

To complete the algorithm, an expression for the assignment weight  $w_{mtr}$  is required. The weight, as

defined in (20), is comprised of two terms

$$w_{mtr} = \sum_{\mathbf{D}} P(k_{tr} = m | \mathbf{D}, \mathbf{X}^{(i)}, \mathbf{Z}) P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z}).$$

The first term in (20) is the conditional probability of the assignments, and the second term is the posterior probability of the assignment states.

The conditional probability of the assignments is found via Bayes rule

$$P(k_{tr} = m | \mathbf{X}^{(i)}, \mathbf{D}, \mathbf{Z}) = \frac{P(k_{tr} = m, \mathbf{X}^{(i)}, \mathbf{D}, \mathbf{Z})}{\sum_{s=1}^M P(k_{tr} = s, \mathbf{X}^{(i)}, \mathbf{D}, \mathbf{Z})}$$

$$= \frac{P(\mathbf{X}^{(i)}, \mathbf{D}, \mathbf{Z} \setminus \mathbf{z}_{tr}) \pi_t^m(\mathbf{D}_t) \zeta_t^m(\mathbf{z}_{tr} | \mathbf{x}_t^{m(i)})}{\sum_{s=1}^M P(\mathbf{X}^{(i)}, \mathbf{D}, \mathbf{Z} \setminus \mathbf{z}_{tr}) \pi_t^s(\mathbf{D}_t) \zeta_t^s(\mathbf{z}_{tr} | \mathbf{x}_t^{s(i)})}$$

$$= \frac{\pi_t^m(\mathbf{D}_t) \zeta_t^m(\mathbf{z}_{tr} | \mathbf{x}_t^{m(i)})}{\sum_{s=1}^M \pi_t^s(\mathbf{D}_t) \zeta_t^s(\mathbf{z}_{tr} | \mathbf{x}_t^{s(i)})} \quad (27)$$

where  $\mathbf{Z} \setminus \mathbf{z}_{tr}$  is the set of all measurements in the batch except the measurement  $z_{tr}$ . Equation (27) is the same as the standard PMHT weight equation except that the prior distribution of the assignments  $\pi_t^m(\mathbf{D}_t)$  is dependent on the assignment state  $\mathbf{D}_t$ .

The posterior probability of the assignment state vector  $\mathbf{D}_t$  given all of the batch measurements and the state estimates at the previous iteration, namely  $P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z})$ , can be determined by using the hidden Markov model (HMM) smoother [22].

Using Bayes rule write

$$P(\mathbf{D}_t | \mathbf{X}, \mathbf{Z}) \propto P(\mathbf{D}_t, \mathbf{X}, \mathbf{Z})$$

$$= P(\mathbf{D}_t, \mathcal{X}_1^t, \mathcal{X}_{t+1}^T, \mathcal{Z}_1^t, \mathcal{Z}_{t+1}^T)$$

$$= P(\mathbf{D}_t, \mathcal{X}_1^t, \mathcal{Z}_1^t) P(\mathcal{X}_{t+1}^T, \mathcal{Z}_{t+1}^T | \mathbf{D}_t, \mathcal{X}_1^t, \mathcal{Z}_1^t)$$

$$\equiv \alpha_t(\mathbf{D}_t) \beta_t(\mathbf{D}_t) \equiv \gamma_t(\mathbf{D}_t) \quad (28)$$

where

$$\mathcal{X}_{t_1}^{t_2} \equiv \{X_{t_1} \dots X_{t_2}\} \quad (29)$$

$$\mathcal{Z}_{t_1}^{t_2} \equiv \{Z_{t_1} \dots Z_{t_2}\}. \quad (30)$$

The required probabilities are the  $\gamma_t(\mathbf{D}_t)$  and these are found by deriving recursive relations for  $\alpha_t(\mathbf{D}_t)$  and  $\beta_t(\mathbf{D}_t)$ .

$$\alpha_t(\mathbf{D}_t) = P(\mathbf{D}_t, \mathcal{X}_1^t, \mathcal{Z}_1^t)$$

$$= \sum_{\mathbf{D}_{t-1}} P(\mathbf{D}_t, \mathbf{D}_{t-1}, \mathcal{X}_1^t, \mathcal{Z}_1^t)$$

$$= \sum_{\mathbf{D}_{t-1}} P(\mathbf{D}_t | \mathbf{D}_{t-1}) P(\mathbf{X}_t | \mathbf{X}_{t-1}) P(\mathbf{Z}_t | \mathbf{D}_t, \mathbf{X}_t) \alpha_{t-1}$$

$$\propto \sum_{\mathbf{D}_{t-1}} P(\mathbf{D}_t | \mathbf{D}_{t-1}) P(\mathbf{Z}_t | \mathbf{D}_t, \mathbf{X}_t) \alpha_{t-1}$$

$$= \sum_{\mathbf{D}_{t-1}} \left\{ \prod_{m=M_y+1}^M \Delta_t^m(d_t^m | d_{t-1}^m) \right\} P(\mathbf{Z}_t | \mathbf{D}_t, \mathbf{X}_t) \alpha_{t-1} \quad (31)$$

where the dependence of  $\alpha_{t-1}$  on  $\mathbf{D}_{t-1}$  is suppressed for compactness, and  $\alpha_t(\mathbf{D}_t)$  is normalised by dividing by  $\sum_{\mathbf{D}_t} \alpha_t(\mathbf{D}_t)$ .

$$\begin{aligned}
\beta_{t-1}(\mathbf{D}_{t-1}) &= P(\mathcal{X}_t^T, \mathcal{Z}_t^T | \mathbf{D}_{t-1}, \mathcal{X}_1^{t-1}, \mathcal{Z}_1^{t-1}) \\
&= \sum_{\mathbf{D}_t} P(\mathcal{X}_t^T, \mathcal{Z}_t^T, \mathbf{D}_t | \mathbf{D}_{t-1}, \mathcal{X}_1^{t-1}, \mathcal{Z}_1^{t-1}) \\
&= \sum_{\mathbf{D}_t} P(\mathbf{D}_t | \mathbf{D}_{t-1}) P(\mathbf{X}_t | \mathbf{X}_{t-1}) P(\mathbf{Z}_t | \mathbf{D}_t, \mathbf{X}_t) \beta_t \\
&\propto \sum_{\mathbf{D}_t} P(\mathbf{D}_t | \mathbf{D}_{t-1}) P(\mathbf{Z}_t | \mathbf{D}_t, \mathbf{X}_t) \beta_t \\
&= \sum_{\mathbf{D}_t} \left\{ \prod_{m=M_T+1}^M \Delta_t^m(d_t^m | d_{t-1}^m) \right\} P(\mathbf{Z}_t | \mathbf{D}_t, \mathbf{X}_t) \beta_t
\end{aligned} \tag{32}$$

where the dependence of  $\beta_t$  on  $\mathbf{D}_t$  is again suppressed for compactness, and  $\beta_t(\mathbf{D}_t)$  is normalised by dividing by  $\sum_{\mathbf{D}_t} \beta_t(\mathbf{D}_t)$ .

Both  $\alpha_t(\mathbf{D}_t)$  and  $\beta_t(\mathbf{D}_t)$  depend on the incomplete conditional data likelihood  $P(\mathbf{Z}_t | \mathbf{D}_t, \mathbf{X}_t)$  which is given by

$$\begin{aligned}
P(\mathbf{Z}_t | \mathbf{D}_t, \mathbf{X}_t) &= \prod_{r=1}^{n_t} P(\mathbf{z}_{tr} | \mathbf{D}_t, \mathbf{X}_t) \\
&= \prod_{r=1}^{n_t} \left\{ \sum_{m=1}^M \pi_t^m(\mathbf{D}_t) \zeta_t^m(\mathbf{z}_{tr} | \mathbf{x}_t^m) \right\}.
\end{aligned} \tag{33}$$

## B. Statement of Hysteresis PMHT Algorithm

The hysteresis PMHT algorithm is mostly the same as the standard PMHT. Where it differs, is that the hysteresis PMHT has an additional step where the assignment state probabilities are calculated via the HMM smoother, and the assignment weights are dependent on these probabilities. The algorithm proceeds as follows.

- 1) Initialise the model state estimates  $\mathbf{X}^{(0)}$ , and the mixing proportion estimates  $\Sigma^{(0)}$ .
- 2) Determine the posterior probability of assignment state using

$$P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z}) = \frac{\alpha_t(\mathbf{D}_t) \beta_t(\mathbf{D}_t)}{\sum_{\mathbf{U}} \alpha_t(\mathbf{D}_t = \mathbf{U}) \beta_t(\mathbf{D}_t = \mathbf{U})}$$

where  $\alpha_t(\mathbf{D}_t)$  and  $\beta_t(\mathbf{D}_t)$  are defined using (31) and (32). This probability is implicitly iteration dependent through the conditioning on the state  $\mathbf{X}^{(i)}$ .

- 3) Calculate the assignment weights for each measurement and model,

$$w_{mtr}^{(i)} = \sum_{\mathbf{D}_t} \frac{P(\mathbf{D}_t | \mathbf{X}^{(i)}, \mathbf{Z}) \pi_t^m(\mathbf{D}_t) \zeta_t^m(\mathbf{z}_{tr} | \mathbf{x}_t^{m(i)})}{\sum_{p=1}^M \pi_t^p(\mathbf{D}_t) \zeta_t^p(\mathbf{z}_{tr} | \mathbf{x}_t^{p(i)})}$$

- 4) Update the state estimates using the maximum likelihood estimator to give  $\mathbf{X}^{(i+1)}$ . This part of the algorithm is identical to the standard PMHT. Update the mixing proportion estimates using (26).
- 5) Repeat steps 2...4 until convergence.

## C. Unknown Assignment State Parameters

The derivation of the hysteresis PMHT assumes that the parameters of the assignment state model are known. Namely,  $\Delta_0^m(d_0^m)$ ,  $\Delta_t^m(d_t^m | d_{t-1}^m)$ , and  $\phi_t^m(d_t^m)$  are all known. The validity of this assumption depends on the application. If the assignment state model is adopted simply as a means of smoothing the estimated  $\pi_t^m$  then these parameters are design variables, chosen to ensure the desired behaviour of the estimated  $\pi_t^m$ . However, if it is believed that the true underlying prior follows a Markov chain, then it is not appropriate to choose arbitrary values. Under this circumstance it may be desirable to estimate the true assignment state parameters.

Firstly, consider the family of functions  $\phi_t^m(d_t^m)$ . These functions are part of the measurement function. However, they do not control the measurements directly, but influence the probability of selecting different models. Effectively, this function acts as a discretisation of the  $\pi_t^m$  function. Thus, the designer can always choose the  $\phi_t^m(d_t^m)$ .

In contrast, the  $\Delta_t^m(d_t^m | d_{t-1}^m)$  represent the true underlying dynamic behaviour of the  $\pi_t^m$ . This function can be incorporated in the same way as the unknown confusion matrix in [5]. In a similar way, the optimal estimate for each matrix element is given by

$$\Delta_t^m(d_t^m = i | d_{t-1}^m = j) = \frac{\gamma_t(d_t^m = i, d_{t-1}^m = j)}{\gamma_{t-1}(d_{t-1}^m = j)} \tag{34}$$

where the probability  $\gamma_t(d_t^m = i, d_{t-1}^m = j)$  is determined using an HMM smoother with forwards and backwards recursions similar to those used for  $\gamma_t(d_t^m)$  as given in (31) and (32). The above estimate for  $\Delta$  suffers from a scarcity of data, and it would probably be prudent to assume that the transition function is stationary. This leads to temporal averaging and the estimated  $\Delta$  is then given by

$$\Delta_t^m(d_t^m = i | d_{t-1}^m = j) = \frac{\sum_{t=1}^T \gamma_t(d_t^m = i, d_{t-1}^m = j)}{\sum_{t=1}^T \gamma_{t-1}(d_{t-1}^m = j)} \tag{35}$$

## VI. MAINTENANCE USING THE HYSTERESIS MODEL

Track maintenance has already been identified as a special case of the hysteresis model. The specific models required for maintenance are now presented fully. When a candidate track is valid, then it is detected with probability  $Pd$  and if it is detected, then

it forms one measurement. Thus, for a valid candidate, the true a priori probability that the  $r$ th measurement at scan  $t$  is due to model  $m$ , i.e.,  $\pi_t^m$ , is  $Pd/n_t$ . If the candidate is false, then the true  $\pi_t^m$  is zero. This means that the true  $\pi_t^m$  has two possible values, one corresponding to valid candidates, and one to false candidates. The assignment state space will therefore be chosen to have two values, i.e.,  $d_t^m$  is binary. This is the simplest possible model, and has the benefit of minimising the computational requirements of the algorithm. Under this model,

$$\begin{aligned}\phi_t^m(0) &= 0 \\ \phi_t^m(1) &= \frac{Pd}{n_t}\end{aligned}$$

as presented earlier in (7).

This special case, with a binary assignment state model, is the same as the visibility/existence/perceivability model used for initiation with the PDAF and other filters [16–18, 23].

The hysteresis PMHT calculates the posterior probability mass  $P(d_t^m = 1)$ , which is the probability that the candidate has a prior consistent with a valid track. This then provides a natural quality statistic. Since the probability mass of the assignment state variable can change over the batch, the candidate test statistic is

$$q_y^m = \frac{1}{T} \sum_{t=1}^T P(d_t^m = 1 | \mathbf{X}, \mathbf{Z}). \quad (36)$$

All that remains is the selection of the transition probability matrix and the prior probability mass. Since the model is binary, these are fully specified by three parameters:  $\Delta_0^m(d_0^m)$  has two elements, one of which is constrained by normalisation, and  $\Delta_t^m(d_t^m | d_{t-1}^m)$  has four elements, two of which are constrained by normalisation. These parameters are design parameters which can be chosen either by subjective belief, or by optimising particular criteria.

In [20] a set of heuristic rules is optimised in order to select parameters of the perceivability model. Although other authors use the same model (by other names), [20] is the only work to address the issue of parameter selection. Other authors either present quantities with no explanation, or do not provide these numbers due to propriety issues. The parameters derived in [20] are

$$\begin{aligned}P(d_0^m = 1) &= 0.5 \\ P(d_t^m = 1 | d_{t-1}^m = 1) &= 0.988 \\ P(d_t^m = 1 | d_{t-1}^m = 0) &= 0.0.\end{aligned}$$

Notice that the above parameters make the  $d_t^m = 0$  state absorbing. This is a reflection of the philosophy that a false track cannot become a valid target track, but a valid target track can become false if the target

disappears. The choice of  $P(d_t^m = 1 | d_{t-1}^m = 0) = 0$  turns out to be a poor one for the hysteresis PMHT. Choosing this to be zero means that if a measurement is found at time  $t_k$  then the values of  $d_t^m$  for all  $t < t_k$  must be unity. This means that the contributions of measurements at earlier times are swamped by those of measurements near the decision time. A single clutter measurement falling close to the track will cause the algorithm to be highly confident in the track. To avoid this, the prior

$$P(d_t^m = 1 | d_{t-1}^m = 0) = 0.1$$

is chosen. The behaviour discussed above does not occur with PDAF based algorithms because they use a recursive estimate of the existence probability.

Two candidate tests using the standard PMHT are now presented.

#### A. Sum of Weights Quality Statistic

A common test for promoting candidate tracks is referred to as an  $M$  of  $N$  rule [1]. Under an  $M$  of  $N$  rule, candidates are promoted if they receive  $M$  validated measurements in  $N$  scans. This rule is also used to select potential measurements with which to create candidates. In this case, a candidate is formed if there are  $M$  measurements within a gate volume in  $N$  scans. Rather than simply counting the number of measurements that are within an arbitrary distance of the track, a more intuitively appealing approach is to test the sum of the association probabilities. This has the advantage of not promoting a track simply because it is in a high clutter environment. Also it gives a greater choice of promotion threshold, since the test statistic is no longer integer valued. This leads to a simple ad hoc quality statistic given by the sum of the association weights:

$$q_w^m = \sum_{t=1}^T \sum_{r=1}^{n_t} w_{mtr}. \quad (37)$$

This statistic can be interpreted as the estimated number of measurements caused by model  $m$ . When normalised by the total number of measurements in the batch,  $q_w^m$  is also the maximum likelihood estimator for the prior probability of a measurement being caused by model  $m$  if this prior is time invariant.

This simple quality statistic is referred to as the sum of weights quality statistic (for obvious reasons) and is used as a benchmark for other proposed initiation schemes. The sum of weights test could be implemented with no alteration to the PMHT algorithm and carries very little overheads.

#### B. Cost Function Increment

Track initiation was previously mentioned to be a problem of model order estimation. One method

for solving the model order estimation problem is to find the maximum likelihood model fit where the likelihood is penalised by a model order term. This penalty term is necessary because the likelihood can always be increased by adding further model complexity until the order of the model equals the number of data. The criteria due to Akaike [24], Rissanen [25], and Schwartz [26] can be written (for large data sets) as

$$\tilde{L}(d) = \log\{L(\mathbf{X}(d), \mathbf{Z})\} - d\epsilon \quad (38)$$

where  $\epsilon$  varies depending on the selection criterion, and  $d$  is the order of the model.

This is equivalent to choosing a model of order  $d$  over a model of order  $d - 1$  when the log likelihood under model  $d$  is at least  $\epsilon$  more than the log likelihood under model  $d - 1$ :

$$\begin{aligned} \tilde{L}(d) - \tilde{L}(d - 1) &> 0 \\ \log\{L(\mathbf{X}(d), \mathbf{Z})\} - \log\{L(\mathbf{X}(d - 1), \mathbf{Z})\} &> \epsilon. \end{aligned} \quad (39)$$

The objective function for the PMHT (the EM auxiliary function) is the conditional expectation of the log likelihood, so an obvious analogous candidate test statistic would be

$$q_Q^c = Q(\mathbf{X}, \mathbf{X}^c, \mathbf{Z}; \Pi) - Q(\mathbf{X}, \mathbf{Z}; \Pi). \quad (40)$$

Thus, the candidate test statistic is the increase in the EM auxiliary function by the incorporation of the candidate track. The addition of the candidate model is guaranteed to increase the auxiliary function, and the amount of this increase provides a measure of how much the candidate model adds to the overall data description. This approach was used to estimate the number of components and the parameters of a static Gaussian mixture with an unknown number of components by Vlassis et al. in [27] and [28].

Notice that the statistic implicitly combines three criteria for candidate merit: the scatter of measurements associated with the track, the prior likelihood of the estimated state sequence, and the likelihood of the associated measurements without the candidate. The previous test, sum of weights test, which is based on the assignment weights, only addresses the first of these. This test is also similar to the cumulative log likelihood measure of track merit [29]. Where it differs is that it measures the increase in likelihood due to the new track, not simply the likelihood of the new track. In principle, this should make this test automatically resilient to forming duplicate tracks on the same measurement sequences, and it will tend to penalise tracks formed in heavy clutter.

This candidate test statistic is referred to as the cost increment in the following analysis.

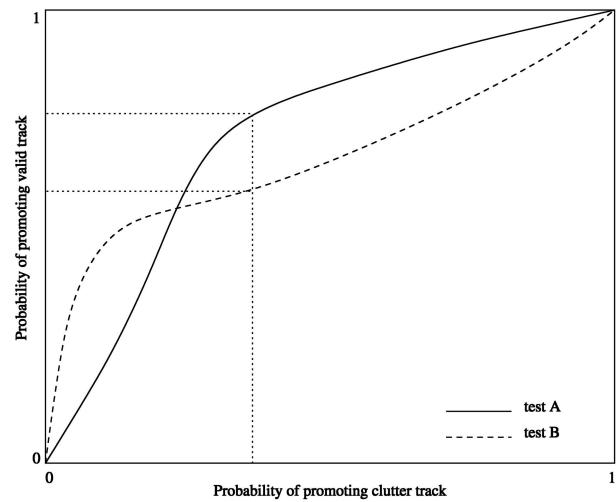


Fig. 4. Example tracker ROC curve.

## VII. QUANTIFYING INITIATION PERFORMANCE

Three methods for track initiation with PMHT have now been presented. The performance of these methods is compared through simulation and with recorded radar data. The performance measure for track initiation should quantify how well the candidate test discriminates between valid and false candidate tracks. This will be done by using an ROC curve. The ROC curve for a decision process plots the true positive probability as a function of the false positive probability. For the case of track initiation, this is the probability of promoting a valid candidate as a function of the probability of promoting a false candidate, i.e., a clutter track. Each point in the locus of the curve represents a particular detector setting (threshold).

Fig. 4 shows an example ROC curve for track initiation. The ROC curves to follow, plot the probability of promoting a valid track on a linear scale and the probability of promoting a clutter track on a logarithmic scale. This is because the probability of promoting a clutter track is required to be very low. Commonly there are many more false detections than valid target detections and consequently the majority of the candidate tracks will be clutter tracks. In Fig. 4, test A gives superior performance at high probability of promoting clutter tracks, but test B gives better performance for low probabilities. In such a case, neither test is universally preferred, and the design choice depends on the operating point at which the system will work.

Although the ROC curve has been used to quantify track initiation performance (e.g. [30–32]), it is not common. Rather, authors tend to choose to fix the false track rate (equivalent to choosing a particular probability of promoting a clutter track) and observe the valid track initiation performance. This may result in misleading conclusions. In Fig. 4 dotted lines show the performance obtained when a mediocre

probability of promoting clutter tracks is specified. In this case, the curve corresponding to test A would give better results than the curve for test B, leading to the conclusion that test A is a superior test. Clearly this is not universally true, and in fact test B is preferable when very low probability of promoting clutter tracks is required (which it usually is). In order to show results representative of the algorithm under test and independent of the particular operating point chosen, the ROC curve is used as the primary ruler for measuring track initiation performance.

The complicated data association process involved makes it impractical to analytically derive the probabilities required for the curve. Instead, the curves presented here have been formed by approximating the probability densities with a finite set of observations. A kernel estimate [33] is used to approximate the probability densities to avoid excessive fluctuations on the resulting curves.

#### A. Other Performance Assessment Approaches

The ROC curve is a good method for assessing track initiation performance, but this is only one facet of the tracking algorithm. Aspects such as manoeuvre handling, estimation error, overshoot, and false track duration are all important factors in overall tracker rating. In [34], 17 metrics for tracker performance are used to compare competing algorithms. This approach is more appropriate to gain an overall indication of tracking performance. However, the main focus of this paper is track initiation performance quantified via the ROC curve. Other studies have already considered the performance of PMHT under established track criterion, such as estimation accuracy (for example, [35–38]).

### VIII. SIMULATED TRACK INITIATION PERFORMANCE

The performance of the three PMHT track initiation approaches presented above is now examined through simulation.

A crucial factor in the production of the initiation ROC curves is a knowledge of the underlying truth. The statistics are conditioned on knowledge of whether each track is valid or false. In order to guarantee this knowledge, two test scenarios are used. To estimate the candidate statistics of false tracks, a scene containing no target is used. The statistics of valid tracks are estimated using a scene with a single target, and a valid target measurement is used to initialise the track. The valid candidate is discarded if the final state estimate deviates from the true target state by more than a prescribed amount. This approach avoids the possibility of coincident tracks, and false tracks being assigned valid target measurements.

The simulated target model is the two-dimensional almost constant velocity model. This is a Cartesian model, with target motion independent in the two coordinate axes. The target state is two-dimensional position and velocity, and the state evolution pdf is given by  $\psi_t^m(\mathbf{x}_t^m | \mathbf{x}_{t-1}^m) \sim N(\mathbf{F}\mathbf{x}_{t-1}^m, \mathbf{G}\mathbf{Q}\mathbf{G}^T)$  where  $N(\mu, \Sigma)$  is a multivariate Gaussian probability density function (pdf) with mean  $\mu$  and covariance  $\Sigma$ . The matrices  $\mathbf{F}$  and  $\mathbf{G}$  are determined by dynamics and are given by

$$\mathbf{F} = \begin{bmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} \frac{1}{2}\tau^2 & 0 \\ \tau & 0 \\ 0 & \frac{1}{2}\tau^2 \\ 0 & \tau \end{bmatrix}$$

where  $\tau$  is the sampling interval, and is assumed constant for convenience (irregular sampling requires a different form of the  $\mathbf{Q}$  matrix for consistency). The model assumes random fluctuations in the velocity components, and the matrix  $\mathbf{Q}$  (often referred to as the process noise covariance) gives the covariance of these fluctuations. It is set to  $\mathbf{Q} = 0.001\mathbf{I}(2)$ , where  $\mathbf{I}(n)$  is the  $n$ -dimensional identity matrix. This low process noise ensures that the target trajectory is approximately straight. When a target is present, it begins in the middle of the sensor footprint with an initial state given by  $\mathbf{x}_0 = [0, 0.35, 50, 0.35]^T$  (corresponding to a velocity vector of magnitude 0.5).

The sensor provides a two-dimensional measurement vector which contains an observation of the target position in each dimension, corrupted with independent Gaussian noise. Thus the measurement pdf for the target is given by

$$\zeta_t^m(\mathbf{z}_t^m | \mathbf{x}_t^m) \sim N(\mathbf{H}\mathbf{x}_t^m, \mathbf{R}) \quad \text{where} \quad \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and the measurement noise covariance is set to  $\mathbf{R} = \mathbf{I}(2)$ . When a target is present, it is detected with probability  $Pd = 0.6$ . The sensor detects measurements over a footprint arbitrarily labelled from  $-50$  to  $50$  in the  $x$  direction and from  $0$  to  $100$  in the  $y$  direction. Measurements are collected over a batch of  $T = 11$  scans.

Different distributions of the clutter measurements are considered, although each consists of only a single model. Thus,  $M_Y = 1$  and  $\sigma_t^1 = 1$ .

Tracks are initialised using a single measurement at zero velocity. The initial state distribution is assumed to be  $\psi_0^m(\mathbf{x}_0^m) \sim N(\bar{\mathbf{x}}_0, \mathbf{P}_0)$  with

$$\bar{\mathbf{x}}_0 = [z_{1r}[x], 0, z_{1r}[y], 0]^T \quad (41)$$

where  $z_{1r}[x]$  and  $z_{1r}[y]$  are the two components of the  $r$ th measurement at the first scan, i.e.,  $\mathbf{z}_{1r} \equiv [z_{1r}[x], z_{1r}[y]]^T$ . The initial covariance assigned to the tracks (i.e., the assumed covariance of the initial state

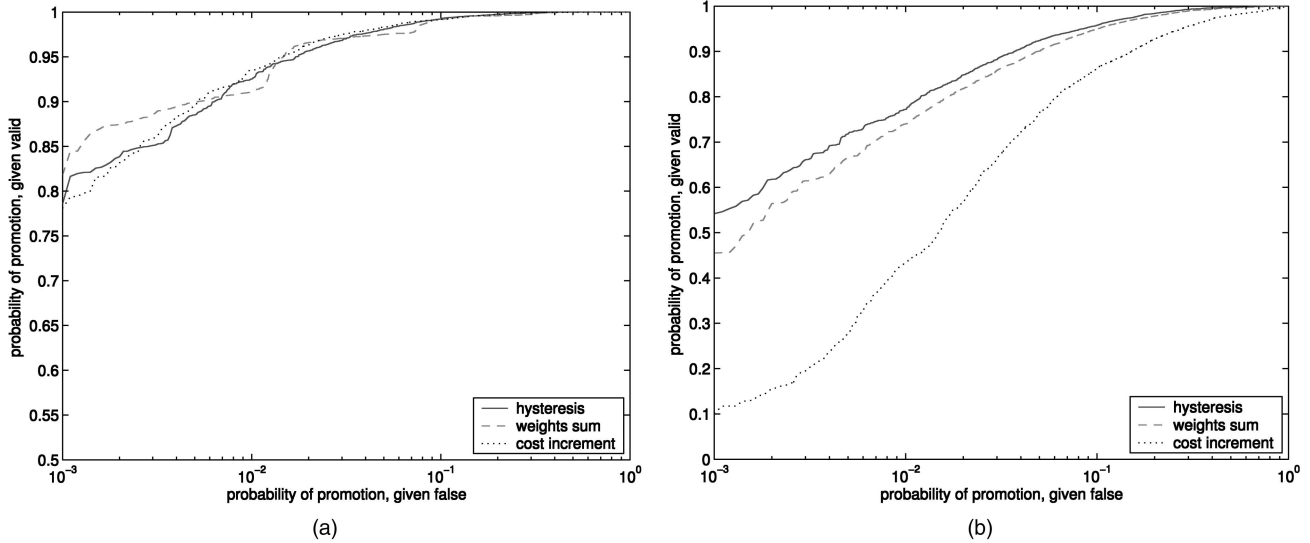


Fig. 5. Track initiation ROC curves with uniform clutter. (a) Low rate of false detections. (b) High rate of false detections.

distribution) is

$$P_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}. \quad (42)$$

The covariance of the position part of the initial state is the covariance of the measurement used to initialise it. The covariance of the velocity part of the initial state is chosen so that the true target velocity is several standard deviations from the initial estimate.

For the challenging scenarios chosen, 10000 Monte Carlo trials are used for each statistic to be tested. The cost increment statistic and the weights sum statistic are both derived from the standard PMHT algorithm. Thus, each plot in the following analysis is generated using 40000 random trials (10000 each of valid and false tracks for the standard PMHT and the hysteresis PMHT).

#### A. Divergent Tracks

On some valid track trials, the state estimate will diverge from the true target trajectory. This may occur because the track is seduced by clutter, or it may be that the track is not able to recover from its poor initialisation (perhaps due to early missed detections). Such tracks are no longer valid. The candidate track is effectively false, and they are not included in the analysis. To test for this condition, the distance between the track position and the true target position is measured through the batch. If this distance exceeds a particular threshold, then the track is deemed divergent, and not included. The distance in each dimension is normalised by the measurement variance, and the total squared distance is compared with a threshold of 9 (approximately a 99 percent

confidence interval). This means a track is divergent if

$$\max_{t=1}^T (\hat{\mathbf{x}}_t - \mathbf{x}_t)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} (\hat{\mathbf{x}}_t - \mathbf{x}_t) > 9. \quad (43)$$

The number of divergent tracks is used independently as a measure of initialisation robustness, and is presented after the ROC curve analysis.

#### B. Uniform Clutter Distribution

The first case considered is the ubiquitous uniform clutter distribution. The performance of the three proposed initiation schemes is examined for two different rates of false detections. The first of these corresponds to 10 clutter measurements per scan. This is quite a low concentration of false measurements, and all of the schemes have a fairly easy time in discriminating between false and valid tracks. In this example there is little to distinguish the different approaches. The estimated ROC curve for each of the initiation schemes on this relatively low false detection rate is shown in Fig. 5(a). The weights sum and cost increment quality measures are shown as a dashed line and a dotted line, respectively. The hysteresis performance is shown as a solid line.

The second clutter false detection rate corresponds to 50 clutter measurements per scan. The estimated ROC curves for this relatively high concentration clutter are shown in Fig. 5(b). As is expected, the performance of all approaches is degraded from that obtained on the lower concentration clutter. Of the three approaches, the hysteresis PMHT gives the best performance, although the difference between it and the weights sum approach is not significant. The cost increment approach gives significantly worse performance than the other two. This is a little surprising, since the cost increment approach

is linked to established model order estimation techniques, whereas the weights sum approach is purely an intuitive, ad hoc method. This performance is expected, however, if the nature of the two statistics is more closely examined.

1) *Relationship Between Cost Increment and Weights Sum for a Uniform Clutter Distribution:*

It may have been intuitively expected that the weights sum (being an ad hoc test) would give the worst initiation results. However, the cost function increment has worse performance. By manipulating the expression for the cost increment quality statistic, this result can be predicted.

In this experiment, there is only one target model, and one clutter model. If the target model is removed, then the cost function is

$$\begin{aligned} Q(\mathbf{X}) &= \sum_{t=1}^T \sum_{r=1}^{n_t} \log\{\zeta_r^1(\mathbf{z}_{tr} | \mathbf{x}_t^1)\} \\ &= \sum_{t=1}^T n_t \log\{\zeta_t^1\} \end{aligned} \quad (44)$$

since all of the measurements have the same probability when the clutter pdf is uniform.

With the target model, the cost function is

$$Q(\mathbf{X}, \mathbf{X}^c) = Q_X^2 + \sum_{t=1}^T Q_{t\pi} + \sum_{t=1}^T \sum_{r=1}^{n_t} w_{1tr} \log\{\zeta_r^1\}. \quad (45)$$

Thus

$$\begin{aligned} q_Q^c &= Q(\mathbf{X}, \mathbf{X}^c) - Q(\mathbf{X}) \\ &= Q_X^2 + \sum_{t=1}^T Q_{t\pi} + \sum_{t=1}^T \sum_{r=1}^{n_t} w_{1tr} \log\{\zeta_r^1\} - \sum_{t=1}^T n_t \log\{\zeta_t^1\} \\ &= Q_X^2 + \sum_{t=1}^T Q_{t\pi} + \sum_{t=1}^T \sum_{r=1}^{n_t} (1 - w_{2tr}) \log\{\zeta_r^1\} - \sum_{t=1}^T n_t \log\{\zeta_t^1\} \\ &= Q_X^2 + \sum_{t=1}^T Q_{t\pi} - \sum_{t=1}^T \sum_{r=1}^{n_t} w_{2tr} \log\{\zeta_r^1\} \\ &= Q_X^2 + \sum_{t=1}^T Q_{t\pi} - \log\{\zeta_t^1\} q_w^c \end{aligned} \quad (46)$$

where  $Q_X^2$  is the model cost for the candidate track.

The statistic clearly consists of a scaled version of the weights sum  $q_w^c$  plus two other terms. The  $Q_{t\pi}$  term is relatively small and has secondary effect. The candidate term  $Q_X^2$  is negative definite since it consists of the sum of the norms of the random errors in the model scaled by their corresponding covariances and by a factor of  $-1/2$ . The assignment weights also scale the measurement orientated error terms. The term proportional to  $q_w^c$  is positive definite since  $\zeta_1 < 1$ . Thus the  $Q_X^2$  acts against the  $q_w^c$  term. When the track corresponds to a valid target, it is likely that

many measurements will be assigned to it, and this makes the  $Q_X^2$  term larger.

The cost increment statistic can be decomposed into two terms: the first term is the reduction in the clutter measurement cost due to assigning measurements to the candidate track. The second term penalises this cost improvement based on the discrepancy between the candidate track and the assumed model. When the clutter is uniformly distributed, then the first term is simply proportional to the sum of the candidate assignment weights. However, when a nonuniform clutter distribution is present, the first term acts to reduce the quality statistic of tracks in highly cluttered regions and increase the quality statistic of those in sparsely cluttered regions. Thus the cost increment approach is more suited to nonuniform clutter distributions.

C. Nonuniform Clutter Distribution

The performance of the algorithms changes if the clutter distribution is nonuniform. To illustrate this, a scenario is now considered where the clutter is exponentially distributed in the  $y$  coordinate direction, and uniform in the  $x$  coordinate direction. The exponential distribution is chosen for simplicity, and because the decaying response loosely emulates the behaviour of the amplitude of false measurements. The clutter distribution is thus given by

$$\zeta_r^1(\mathbf{z}_{tr}) = \frac{1}{2000} \exp\left\{-\frac{z_{tr}[y]}{20}\right\}. \quad (47)$$

As with the uniform distribution, two different rates of false detections are considered. Fig. 6 shows an example trial. All of the measurements for the batch are shown in a spatial plot. The nonuniform distribution of clutter measurements in the vertical axis is clearly seen. The target is present in the trial, and the target detections can be seen in the middle of the plot. The target measurements are shown as circles, and false detections as crosses.

Fig. 7 shows the ROC curves generated for nonuniform distributed clutter. Again, the hysteresis-based algorithm shows better performance than the weights sum. In this case, the performance difference is more significant than in the uniform clutter example. In contrast to the uniform case, the cost increment approach now performs almost as well as the hysteresis approach.

The weights sum approach performs particularly poorly for this example. Also, the ROC curve exhibits a staircase appearance. The staircase effect occurs because the PMHT weights tend to converge to extreme values: close to unity, or close to zero. This means that the weights sum statistic tends to have a pdf with peaks around integer values, for both valid and false tracks. The false track peaks are more broad because the false measurements have a higher

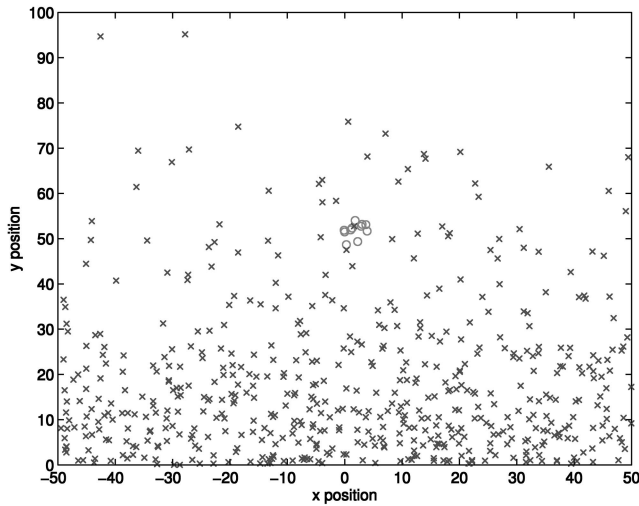
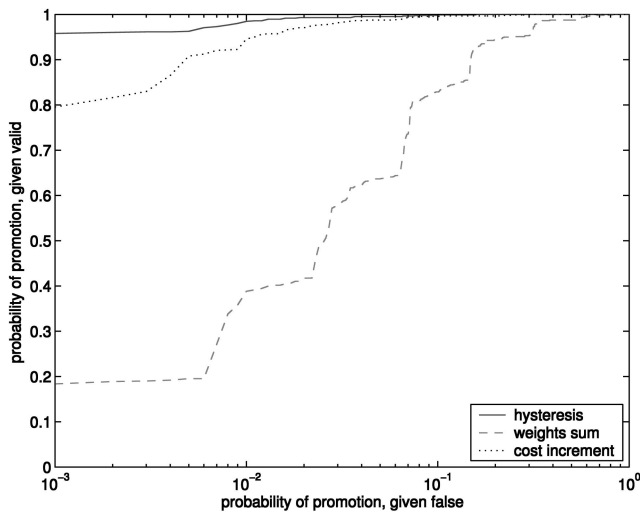


Fig. 6. Example trial with target for nonuniform clutter.

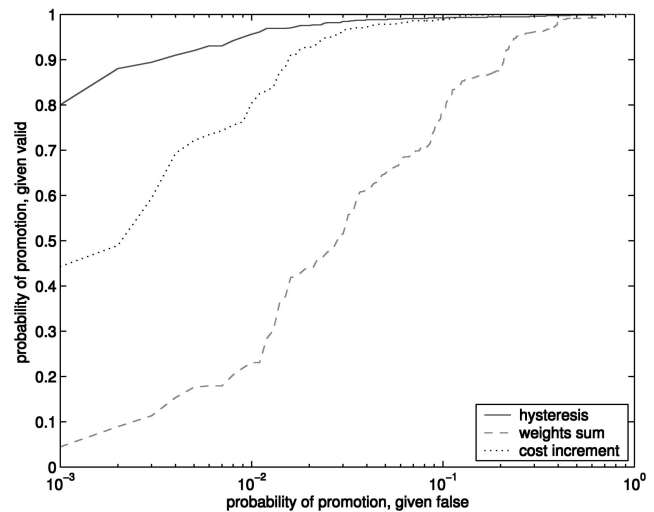
scatter. Each point on the ROC curve corresponds to a possible promotion threshold setting. As this threshold increases, the resulting promotion probabilities follow the ROC curve from the top right corner (zero threshold where all tracks are promoted) to the bottom left corner (an extremely high threshold where all tracks are denied promotion). When this threshold changes from slightly less than a particular integer to slightly more, then a large number of tracks are no longer promoted, since there are peaks in the pdf at integer values. When many valid tracks are thus suppressed, this causes an almost vertical drop in the ROC curve. Conversely, when the threshold is varied between integer values, very few valid tracks are suppressed, but false tracks may be rejected because the false track peaks have more spread. This causes an almost horizontal segment in the curve. The result is a staircase appearance which is somewhat smoothed by the kernel pdf estimator.

The overall poor performance of the weights sum occurs because the approach only considers the number of measurements which are assigned to a track, and does not take account for how well these measurements may have been described by the existing model without the new candidate. It is a track-orientated approach, not a total system one. In contrast, the cost increment statistic is the increase in the log likelihood by the addition of the candidate. Thus if the candidate assigns measurements in a high clutter density region the statistic will be small: those measurements already have a high likelihood under the hypothesis that they are due to clutter. If the candidate assigns measurements in a lower density region, then the statistic will be higher: those measurements had a low likelihood under the clutter hypothesis. Notice that the true target trajectory lies through a lower density part of the clutter distribution. If it were in the highest density part, the tracker would have no hope of following the target. Similarly, the hysteresis algorithm uses an HMM smoother to estimate the probability of the assignment state. This smoother is driven by a gain term which is the conditional measurement likelihood under each assignment state hypothesis,  $P(\mathbf{Z}_t | \mathbf{D}_t, \mathbf{X}_t)$ . When there is only one target, it can be seen that the HMM smoother is driven by the conditional likelihood ratio of the measurements with and without the candidate. If the candidate makes little change to the measurement likelihood (for reasons described above) then this ratio will be close to unity. If the likelihood is greatly increased, then the ratio is large and this drives the HMM smoother to give a high probability of the candidate track corresponding to a valid target.

The cost increment and hysteresis approaches perform better than the weights sum because they both take account of the likelihood that the measurements assigned to the candidate track were due to clutter,



(a)



(b)

Fig. 7. Track initiation ROC curves with nonuniform clutter. (a) Low rate of false detections. (b) High rate of false detections.



TABLE I  
Percentage of Divergent Trials

Cutter Type	Uniform		Nonuniform	
	Low	High	Low	High
standard PMHT	6.0%	31.2%	2.6%	5.0%
Hysteresis PMHT	6.2%	23.5%	2.9%	5.6%

whereas the weights sum effectively counts the number of measurements close to the candidate trajectory. The cost increment and the hysteresis should thus be expected to always outperform the weights sum in nonuniform clutter densities. The hysteresis gives better performance than the cost increment, because it also exploits the temporal history of the track. Candidates which assign measurements in many scans are given higher credence than those who assign a few outlier measurements.

#### D. Initialisation Robustness

The empirical ROC curves are generated using only those valid tracks which remain within an error tolerance of the true target state. This ensures that tracks which have diverged from the correct trajectory do not adversely affect the initiation metric. However, the number of these divergent tracks is also of interest. If the initiation algorithm is able to perfectly discriminate valid tracks from clutter tracks, but diverges from the true trajectory most of the time, then it is much less effective than the ROC will suggest.

The number of divergent trials is therefore listed here to demonstrate that the different schemes have similar performance in this area, and that each only diverges in an acceptably small number of trials. Since the cost increment and weights sum approaches both use information provided by the standard PMHT, they have the same number of divergent trials.

Table I lists the percentage of divergent trials for each of the scenarios.

The uniformly distributed clutter has the highest number of divergent tracks. This is because the nonuniform clutter densities concentrate the false detections away from the simulated target. The higher density clutter in this region makes discrimination between valid and false tracks more difficult because it produces false tracks with more matching measurements. However, the total number of measurements is the same, so the density around the target is lower for the nonuniform clutter. Since most divergent tracks occur when the track is seduced by clutter detections, this lower density around the target leads to a lower rate of divergence.

In general, the rate of divergent tracks is approximately the same for the PMHT with and

without hysteresis, with the hysteresis showing a marginally higher rate. This happens because the parameter choice for the hysteresis model was to use a very low prior probability that the target is visible. In Section VI, the prior probability of a target being valid was chosen to be  $\Delta_0(d_0^m = 1) = 0.1$ . This means that the algorithm may decide that some difficult trials, where early measurements are missed, do not contain a valid track. When the probability that  $d_t^m = 1$  is very low, then the track is not updated by measurements and the track is marked as divergent because it never recovers from initialisation. This effect can be reduced by choosing a higher initial probability, but that degrades the ROC performance. Since the increase in the rate of divergent tracks is only marginal, it was preferred to tolerate it, in preference for superior ROC performance.

The exception to the above trend is the high false detection rate uniform clutter, where the hysteresis model performs much better than the standard PMHT. This is the only case where the two show a significant difference in performance. In this case, the hysteresis model gives better performance because it constrains the allowable values of  $\pi_t^m$ . If more than one measurement is close to the candidate track in one scan, then the standard PMHT will give an estimated  $\hat{\pi}_t^m > 1/n_t$ . The hysteresis model limits the prior to  $Pd/n_t$ . Thus the converged assignment weights for the standard PMHT will be higher than those for the hysteresis PMHT in such a case. Since the true measurement model used for the simulation can only produce one valid measurement per scan, the multiple measurement case only occurs due to false detections. These false detections can seduce the track away from the true path, and by limiting the value of  $\pi_t^m$ , the hysteresis PMHT is made somewhat more resilient to this factor. This behaviour is not observed for any of the other data conditions because the nonuniform clutter pdfs concentrate the false detections in a different area to where the target is present. So, the high rate uniform clutter pdf is the case where the valid track is faced with the highest rate of false detections in its immediate vicinity.

An important point to raise is that the divergent track rate reflects the probability that the candidate track will remain within a tolerance gate of the true target position, given that it starts on a valid target measurement. In the implementation used here, candidate tracks were formed on all unassigned measurements. That is, all measurements for which the largest  $\pi_t^m$  for target models was less than a threshold level of 0.1. Additional tests, such as requiring sequential scans with measurements within a gate distance, can be used to reduce the number of candidates considered, and this reduces the system load (since less tracks are formed) but does not affect the performance in terms of false track rate or target detection ability.

## IX. RADAR DATA PERFORMANCE

The simulation results of the previous section are now verified by running the three track initiation approaches through recorded radar data. The performance of the algorithms is quantified through an empirical ROC curve for track initiation. The difference between the recorded data and simulation is that each of the tracks used in the simulated examples is known to be either valid or false, since this is predetermined by the simulation. When data from a real sensor is used, it is not axiomatic which measurements are valid and which are false. This underlying truth is difficult to obtain, yet it is vital to the analysis. Mislabeled tracks with high or low quality might easily lead to incorrect ROC curves.

The first step required in the production of real data ROC curves is the generation of truth. In a coordinated experiment, this might be done using accurate position logging devices, such as a Global Positioning System (GPS). However, the data sets used here contain targets of opportunity, for which no secondary data source is available. Therefore the validity of measurements is determined by direct examination of the data using displays which are capable of overlaying track reports with the radar image. This approach is used to form a set of valid tracks, and then the candidates are correlated against these, rather than ground truth.

### A. Data Set Features

Two data sets have been selected for use in this analysis. The first data set was collected in early evening. Due to the propagation conditions, and the waveform parameters used, it contains large amounts of interference, in the form of spread Doppler clutter. The radar noise suppression algorithm rejects some of this interference, but not enough to prevent a tracker from producing numerous false tracks. It is known that no targets are present, and the only valid measurements are those due to the calibration track. This data is primarily a test of false track performance. It presents a particularly difficult clutter scenario because the spread Doppler clutter occupies a particular spatial region and false detections produced by it are concentrated in this area—in effect the spatial clutter distribution is nonuniform.

The second data set was recorded under more amiable propagation. The data contains a single target of opportunity which is supported by a single mode of propagation. This target is of relatively low amplitude and is sometimes undetected due to signal fading. This data provides a low detectability valid target. The data sets are combined to produce a single ROC curve for each algorithm tested.

Both data sets also contain high amplitude targets of opportunity which are relatively easy to track. The

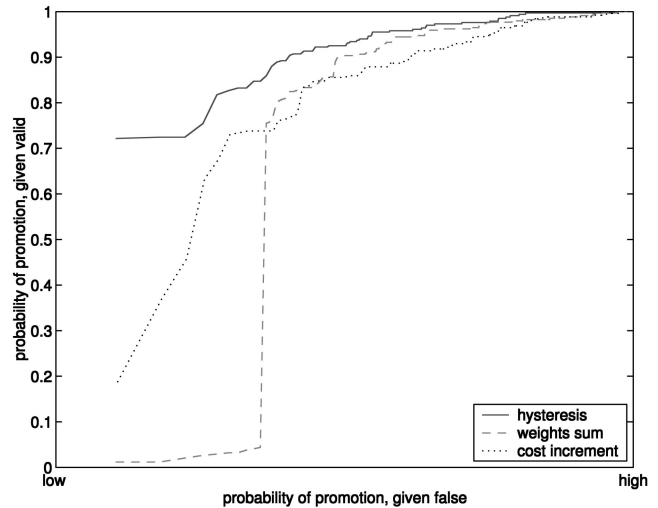


Fig. 8. Track initiation ROC curves for radar data.

effect of these two classes of target can be seen in some of the ROC curves, where an initial decline in promotion probability is followed by a plateau. In this plateau region, the difficult target has been suppressed, but the tracks on the strong targets are still promoted with quality levels much higher than the threshold.

ROC curves generated using radar data for the three approaches are shown in Fig. 8. As was the general trend in the simulations, the hysteresis PMHT gives the best performance.

The clutter pdf for this data is nonuniform, so the weights sum approach is expected to give the worst performance, and this is confirmed by the results. The weights sum also has a step drop off in performance. This occurs because the strong targets are detected in almost all scans. If there are no nearby clutter measurements, the weights sum will thus be approximately equal to the batch length. Once it becomes necessary to raise the promotion threshold above this level, all of those tracks become suppressed.

## X. SUMMARY

This paper has introduced a dynamic model for the assignment prior probability for the PMHT algorithm. This model uses a Bayesian hyperparameter that is assumed to follow a Markov chain, and is referred to as hysteresis. The hysteresis model uses an independent discrete state variable for each target model, and the state variable for a particular model represents its significance in the measurement mixture distribution.

A PMHT algorithm was derived by treating the assignment state variables as missing data in an EM sense, and this algorithm is referred to as the hysteresis PMHT. The hysteresis PMHT determines the probability of the assignment state, in much the same way as the standard PMHT determines the

probability of the assignments. This probability can be calculated using the HMM smoother.

The hysteresis assignment model for PMHT was then applied to the important problem of track initiation. The initiation performance of the hysteresis-based algorithm was compared with two alternatives based on the standard PMHT and common initiation approaches. The hysteresis-based track initiation scheme was found to give the best performance for various simulated clutter conditions, and for data from a real sensor.

## REFERENCES

- [1] Bar-Shalom, Y., and Li, X. R. *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs, CT: YBS, 1995.
- [2] Blackman, S. S., and Popoli, R. *Design and Analysis of Modern Tracking Systems*. Norwood, MA: Artech House, 1999.
- [3] Streit, R. L., and Luginbuhl, T. E. Probabilistic multi-hypothesis tracking. NUWC, Newport, RI, Technical Report 10428, Feb. 1995.
- [4] Willett, P., Ruan, Y., and Streit, R. L. The PMHT for manoeuvring targets. In *Signal and Data Processing of Small Targets*, vol. SPIE 3373, SPIE Annual International Symposium on Aerosense, Orlando, FL, 1998, 416–427.
- [5] Davey, S. J., Gray, D. A., and Streit, R. L. Tracking, association and classification—A combined PMHT approach. *Digital Signal Processing*, **12**, 2 (July 2002), 372–382.
- [6] Willett, P., Ruan, Y., and Streit, R. L. PMHT: Problems and some solutions. *IEEE Transactions on Aerospace and Electronic Systems*, **38**, 3 (July 2002), 738–754.
- [7] Luginbuhl, T. E., Sun, Y., and Willett, P. A track management system for the PMHT algorithm. In *Proceedings of the 4th International Conference on Information Fusion*, Montreal, Canada, Aug. 2001.
- [8] Logothetis, A., Krishnamurthy, V., and Holst, J. On manoeuvring target tracking via the PMHT. In *Thirty-Six Conference on Decision and Control*, San Diego, CA, Dec. 1997.
- [9] Ruan, Y., and Willett, P. Multiple model PMHT and its application to the second benchmark radar tracking problem. *IEEE Transactions on Aerospace and Electronic Systems*, **40**, 4 (Oct. 2004), 1337–1350.
- [10] Streit, R. L. Tracking on intensity-modulated data streams. NUWC, Newport, RI, Technical Report 11221, May 2000.
- [11] Streit, R. L. Tracking targets with specified spectra using the H-PMHT algorithm. NUWC, Newport, RI, Technical Report 11291, June 2001.
- [12] Luginbuhl, T., and Willett, P. Estimating the parameters of general frequency modulated signals. *IEEE Transactions on Signal Processing*, **52**, 1 (Jan. 2004).
- [13] Efe, M., Ruan, Y., and Willett, P. The pedestrian PMHT. In *Proceedings of the 5th International Conference on Information Fusion*, vol. 2, Annapolis, MD, July 2002, 838–845.
- [14] Gauvrit, H., Le Cadre, J.-P., and Jauffret, C. Combinatorial optimization for initialization of probabilistic approaches. In *Workshop commun GdR ISIS (GT 1) and NUWC, approches probabilistes pour l'extraction multipistes*, Nov. 1998 (published in English).
- [15] Davey, S. J., Gray, D. A., and Colegrove, S. B. A hidden Markov model for track initiation with the PMHT. In *Proceedings of the 5th International Conference on Information Fusion*, Annapolis, MD, July 2002.
- [16] Colegrove, S. B., Davis, A. W., and Ayliffe, J. K. Track initiation and nearest neighbours incorporated into probabilistic data association. *Journal of Electrical Electronics Engineering*, Australia, **6**, 3 (1986), 191–198.
- [17] Mušicki, D., Evans, R., and Stankovic, S. Integrated probabilistic data association. *IEEE Transactions on Automatic Control*, **39**, 6 (June 1994), 1237–1241.
- [18] Li, N., and Li, X. R. Target perceivability and its application. *IEEE Transactions on Signal Processing*, **49**, 11 (Nov. 2001), 2588–2604.
- [19] Colegrove, S. B. Advanced Jindalee tracker: Probabilistic data association multiple model initiation filter. Defence Science and Technology Organisation, Australia, Technical Report TR-0659, June 1999.
- [20] Li, N., and Li, X. R. Tracker design based on target perceivability. *IEEE Transactions on Aerospace and Electronic Systems*, **37**, 1 (Jan. 2001), 214–225.
- [21] Davey, S. J. Extensions to the probabilistic multi-hypothesis tracker for improved data association. Ph.D. dissertation, School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, South Australia, 2003.
- [22] Baum, L. E., Petrie, T., Soules, G., and Weiss, N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, **41**, 1 (1970).
- [23] Colegrove, S. B., and Davey, S. J. PDAF with multiple clutter regions and target models. *IEEE Transactions on Aerospace and Electronic Systems*, **39**, 1 (Jan. 2003), 110–124.
- [24] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 6 (Dec. 1974), 716–723.
- [25] Rissanen, J. Modeling by shortest data description. *Automatica*, **14** (1978), 465–471.
- [26] Schwartz, G. Estimating the dimension of a model. *Annals of Statistics*, **6**, 2 (1978), 461–464.
- [27] Vlassis, N., Likas, A., and Krose, B. Multivariate Gaussian mixture modeling with unknown number of components. Computer Science Institute, University of Amsterdam, Intelligent Autonomous Systems, The Netherlands, Technical Report, Apr. 2000.

- [28] Vlassis, N., and Likas, A.  
An EM-VDM algorithm for gaussian mixtures with unknown number of components.  
Computer Science Institute, University of Amsterdam, Intelligent Autonomous Systems, The Netherlands, Technical Report, Apr. 2000.
- [29] Sittler, R. W.  
An optimal data association problem in surveillance theory.  
*IEEE Transactions on Military Electronics*, **MIL-8** (Apr. 1964), 125–139.
- [30] Bar-Shalom, Y., Campo, L. J., and Luh, P. B.  
From receiver operating characteristic to system operating characteristic: Evaluation of a track formation system.  
*IEEE Transactions on Automatic Control*, **35**, 2 (Feb. 1990), 172–179.
- [31] Leung, H., Hu, Z., and Blanchette, M.  
Evaluation of multiple target track initiation techniques in real radar tracking environments.  
*IEE Proceedings on Radar, Sonar Navigation*, **143**, 4 (Aug. 1996), 246–254.
- [32] Bar-Shalom, Y., and Fortman, T. E.  
*Tracking and Data Association*.  
San Diego, CA: Academic Press, 1988.
- [33] Silverman, B. W.  
*Density Estimation for Statistics and Data Analysis*.  
Boca Raton, FL: Chapman and Hall/CRC Press, 1986.
- [34] Colegrove, S. B., Davis, L. M., and Davey, S. J.  
Performance assessment of tracking systems.  
In *International Symposium on Signal Processing and its Applications*, vol. 1, Gold Coast, Australia, Aug. 1996, 188–192.
- [35] Rago, C., Willett, P., and Streit, R. L.  
A comparison of the JPDAF and PMHT tracking algorithms.  
In *International Conference on Acoustics, Speech and Signal Processing*, Detroit, MI, 1995, 3571–3574.
- [36] Ruan, Y., Willett, P., and Streit, R. L.  
A comparison of the PMHT and PDAF tracking algorithms based on their model CRLBs.  
In *SPIE Conference on Acquisition, Tracking and Pointing XIII*, vol. SPIE 3692, Orlando, FL, 1999, 177–188.
- [37] Ruan, Y., and Willett, P.  
An improved PMHT using an idea from coding.  
In *2001 Aerospace Conference*, Big Sky, MT, Mar. 2001.
- [38] Willett, P., Ruan, Y., and Streit, R. L.  
A variety of PMHTs.  
In *Workshop commun GdR ISIS (GT 1) and NUWC, approches probabilistes pour l'extraction multipistes*, Nov. 1998 (published in English).



**Samuel J. Davey** (S'00—M'03) received the Bachelor of Engineering, Master of Mathematical Science and Ph.D. degrees from the University of Adelaide, Australia, in 1966, 1999, and 2003, respectively.

Since 1995 he has worked for the Defence Science and Technology Organisation in target tracking, tracker performance assessment, and multi-sensor fusion.



**Douglas A. Gray** (M'82) received his Ph.D. from the University of Adelaide, Australia, in 1973.

He spent 20 years with the Defence Science and Technology Organisation applying signal processing to sonar and electronic warfare. He is currently a professor of electrical engineering at Adelaide University. His current research interests are in the application of signal processing to radar, GPS, and electricity grids, particularly in adaptive processes, beamforming, signal sorting and classification techniques, data fusion and tracking, estimation, and system identification.