

Marginalized particle PHD Filters for multiple object Bayesian filtering

Yohan Petetin, Mark Morelande, and François Desbouvries, *Senior Member, IEEE*

Abstract—The Probability Hypothesis Density (PHD) filter is a recent solution to the multi-target filtering problem. Because the PHD filter is not computable, several implementations have been proposed including the Gaussian Mixture (GM) approximations and Sequential Monte Carlo (SMC) methods. In this paper, we propose a marginalized particle PHD filter which improves the classical solutions when used in stochastic systems with partially linear substructure.

Index Terms—Multi-object filtering, Probability Hypothesis Density, Gaussian Mixture, Particle Filter, Sequential Monte-Carlo filter, Marginalized Particle Filters, Jump Markov Linear Systems

I. INTRODUCTION

SINGLE object Bayesian filtering has been developed in non linear and / or non Gaussian stochastic models; available solutions include extended (EKF) or unscented (UKF) Kalman filter methods [1] as well as SMC algorithms, including particle filters (PF) and auxiliary particle filters (APF) (see e.g. [2] [3]). However SMC methods can fail when the dimension of the state-space increases, so that one needs to significantly increase the number of particles. This problem can somehow be attenuated when the state can be decomposed into a linear component and a nonlinear one. It is then of interest to use the marginalized particle filter (MPF) solutions, which roughly speaking estimate the nonlinear part of the state by PF, and then given that non linear part estimate the linear one via Kalman filtering (KF) [4] [5].

The (more recent) multi-target filtering problem consists in estimating random states of an unknown number of targets from a set of observations which are either due to detected targets or are false alarms measurements. Early solutions include the Joint Probabilistic Data Association (JPDA) filter [6] and the Multiple Hypothesis Tracking (MHT) filter [7], which both include a matching mechanism that aims at optimizing the association between observations and targets. An alternative class of solutions, based on Random Finite Sets (RFS), has also been proposed [8] [9] [10]. RFS-based solutions consider the set of targets and that of observations as RFS, i.e. as sets of random variables with random and time-varying cardinal. RFS techniques avoid the intensive computational cost of the association mechanism required in the classical solutions. Among RFS-based solutions, the Bayesian multi target filter (MTF) is

a direct generalization of the Bayesian single target filter, in the sense that it propagates over time the so-called posterior multi-target density (mtd), which is a real and positive function that generalizes the probability density function (pdf). Because the MTF is not computable, Mahler proposed to propagate a first order moment of the posterior mtd, the so-called PHD or intensity, a positive density function which operates in the single target state space domain and which enables the number of targets as well as the state of each target to be estimated.

In general, the PHD cannot be computed exactly. In practice, two implementations of the PHD filter, with variants, are popular. The Gaussian mixture (GM) implementation [11] assumes that the PHD is a GM, but requires that each target (when it is detected) and associated measurement follow a linear and Gaussian model; if not, an approximation of the GM implementation based on local linearizations or the Unscented Transformation (UT) [11] can be used provided the non linearities of the model are not too severe; the SMC implementation [12] [13] approximates the PHD by a set of random weighted particles and does require any assumption on the target dynamics.

Now, multiple object filtering SMC-based solutions inherit the same problems as their single object filtering counterparts. In particular, the specific difficulties of high dimensional problems get worse in multi target scenarios since one needs to estimate the number of targets in addition to the states. As in the single object case, this major drawback of SMC based multi object filtering solutions can be attenuated in some specific situations and indeed MPF solutions have recently been adapted to multi-target filtering, with an approach based on data-association [14] in the sense that the SMC part only acts on the data-association step and not on the estimation of states, and another on the Bayesian MTF for Gaussian models with partially linear substructure [15]. On the other hand, even when small dimensional problems are involved, SMC methods may fail, particularly if the particles are sampled without taking into account the available measurements. Among the class of SMC methods, the APF is an alternative technique to the classical Importance Sampling (IS) one. Roughly speaking, in the APF implementation, particles are first selected according to the recent observations before being propagated contrary to the IS one where particles are directly propagated. In the RFS context, a general APF implementation of the PHD filter has been proposed [16].

This paper further adapts the MPF idea, but now to the PHD filter which presents good results in highly cluttered environments [17] [11], avoids a data-association step and has the advantage of being more tractable than the

Yohan Petetin and François Desbouvries are with Mines Telecom Institute, Telecom SudParis, CITI Department, 9 rue Charles Fourier, 91011 Evry, France and with CNRS UMR 5157. Mark Morelande is with the University of Melbourne, Electrical and Electronic Engineering Department, 161 Barry Street, Parkville 3010, VIC Australia. We would like to thank the French MOD DGA/MRIS for financial support of the Ph.D. of Y. Petetin.

Bayesian MTF. The adaptation of marginalization to PHD filter approximation proceeds as follows:

- We first derive the general marginalized PHD (M-PHD) filter, which can be used in Gaussian and partially linear models. We propose two implementations of our M-PHD filter: the first is a “bootstrap one” which holds for a general class of Gaussian and partially linear models; the second is an auxiliary implementation which guides the particles to regions where targets are likely to be present, but which is computationally efficient under certain additional assumptions and approximations. Our approach improves the PHD filter approximation, and so that of the number of targets and of their state, without increasing the number of particles.
- We next consider linear and Gaussian jump Markov state-space systems (JMSS) in the context of multi-object filtering. For such models the PHD is a GM with a number of components which increases exponentially as measurements are acquired. This increase is due uncertainty in the origin of the measurements and in the mode. In [18] the number of mixture components is managed by mixture reduction techniques such as pruning or merging. On the other hand, the SMC implementation proposed in [19] samples particles in augmented dimension (state and mode). We adapt our M-PHD algorithm to linear/Gaussian JMSS so that only the mode governing the jumps is sampled and inference over the kinematic state. Bootstrap and auxiliary implementations are proposed for these models.
- We finally adapt our M-PHD filter to Gaussian JMSS with partially linear and partially non-linear substructures. Following [20], an efficient importance distribution for the auxiliary implementation is proposed for such models.

This paper improves the preliminary work appearing in [20] and [21] by the extension to JMSS and the inclusion of a more detailed performance analysis, including comparisons with the auxiliary PF of [16], consideration of the effect of clutter density and results for maneuvering targets.

The paper is organized as follows. In Section II we first recall the multi-target filtering problem in terms of RFS, then we briefly describe available implementations of the PHD filter. In Section III we derive a new implementation of the PHD filter which takes into account the linearity of some of the components of the state vector of a target. In Section IV we adapt our implementation to linear, or partially linear, Gaussian JMSS models. In Section V we finally compare the performances and complexity of our approach to the classical SMC implementation in a number of range-bearing scenarios.

II. MULTI-TARGET FILTERING

A. RFS description of the model and the MTF

In a standard multi object tracking problem the number of targets and their parameters are unknown at a given time k . It is thus convenient to represent the set of targets and of measurements by RFS (i.e. a set of random vectors with random cardinality) $X_k = \{\mathbf{x}_{k,1}, \dots, \mathbf{x}_{k,n}\}$, $Z_k = \{\mathbf{z}_{k,1}, \dots, \mathbf{z}_{k,q}\}$ where n and q are random integers and $\mathbf{x}_{k,i}$ (resp. $\mathbf{z}_{k,i}$)

$\in \mathbb{R}^m$ (resp. \mathbb{R}^p) for all k and i . The transition pdf of a target between times $k-1$ and k is $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ and the likelihood of a measurement \mathbf{z} with a state \mathbf{x}_k at time k is $g_k(\mathbf{z}|\mathbf{x}_k)$. A direct application of the RFS theory consists in propagating the conditional mtd $p(X_k|Z_{0:k})$. However, the MTF involves equations which are not computable. Therefore, SMC implementations have been proposed (see e.g. [12] [15]), but remain computationally intensive, above all when the number of targets is large.

B. The PHD

To cope with this issue, Mahler introduced the PHD as a first order moment $v(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^m$, of a mtd $f(X)$ or of an RFS X . A PHD is a real and positive function which has the following property: let $S \subset \mathbb{R}^m$, then

$$\int_S v(\mathbf{x}) d\mathbf{x} = \int |X \cap S| f(X) \delta X = \mathbb{E}(|X \cap S|), \quad (1)$$

where $\int \delta X$ denotes the set integral and $|X \cap S|$ is the cardinality of the set of targets which belong to region S . So the integral of the PHD $v(\mathbf{x})$ over region S is the expected number of objects in this region.

In this framework, the multi-target filtering problem can be seen as the propagation of PHD $v_k(\mathbf{x})$, defined as the first order moment of mtd $p(X_k|Z_{0:k})$. The so-called PHD filter, which is a set of equations which propagate recursively the PHD, can be derived under the following hypotheses [10]: the targets evolve and generate measurements independently of one another; the clutter is independent of measurements due to detected targets; and the clutter and the predicted-target processes are Poisson. Also, let us set some notations: $p_{s,k}(\mathbf{x})$ is the probability that a target with state \mathbf{x} at time $k-1$ still exists at time k ; $p_{d,k}(\mathbf{x})$ is the probability that a target with state \mathbf{x} is detected at time k ; κ_k is the intensity of clutter measurements at time k ; and γ_k is the intensity of birth targets at time k . The propagation of the PHD $v_k(\mathbf{x})$ of the posterior mtd $p(X_k|Z_{0:k})$ is then the succession of a prediction step and of an updating step [10] (we assume without loss of generality that there is no spawning):

$$v_{k|k-1}(\mathbf{x}) = \int p_{s,k}(\boldsymbol{\xi}) f_{k|k-1}(\mathbf{x}|\boldsymbol{\xi}) v_{k-1}(\boldsymbol{\xi}) d\boldsymbol{\xi} + \gamma_k(\mathbf{x}), \quad (2)$$

$$v_k(\mathbf{x}) = [1 - p_{d,k}(\mathbf{x})] v_{k|k-1}(\mathbf{x}) + \sum_{\mathbf{z} \in Z_k} \frac{p_{d,k}(\mathbf{x}) g_k(\mathbf{z}|\mathbf{x}) v_{k|k-1}(\mathbf{x})}{\kappa_k(\mathbf{z}) + \int p_{d,k}(\boldsymbol{\xi}) g_k(\mathbf{z}|\boldsymbol{\xi}) v_{k|k-1}(\boldsymbol{\xi}) d\boldsymbol{\xi}}. \quad (3)$$

1) *Implementation of the PHD Filter:* Eqs. (2) and (3) are not computable; several suboptimal algorithms have thus been proposed in the literature.

The GM implementation [11] assumes that the model is linear and Gaussian, that the probabilities of detection $p_{d,k}$ and of survival $p_{s,k}$ are constant, and that $v_{k-1}(\mathbf{x})$ and $\gamma_{k-1}(\mathbf{x})$ are GM. Briefly, since $f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1})$ is a linear and Gaussian pdf, the predicted PHD is a new GM. Further injecting this new mixture in (3) still provides a GM. When the model is Gaussian but non-linear, alternatives based on the UT or on the EKF and other approximations have been proposed in [22].

The SMC implementation [12] [23] approximates PHD v_{k-1} by the discrete and random mixture $\hat{v}_{k-1}(\mathbf{x}) = \sum_{i=1}^{L_{k-1}} w_{k-1}^{(i)} \delta_{\mathbf{x}_{k-1}^{(i)}}(\mathbf{x})$ where δ_x denotes the Dirac mass at point x . Plugging it in (2) we get a continuous approximation $\tilde{v}_{k|k-1}$ of $v_{k|k-1}$ for which we can derive a new discrete approximation of $v_{k|k-1}(\mathbf{x})$ by using IS, $\hat{v}_{k|k-1}(\mathbf{x}) = \sum_{i=1}^{L_k} w_{k|k-1}^{(i)} \delta_{\mathbf{x}_k^{(i)}}(\mathbf{x})$, and finally inject it in (3). We get the discrete approximation \hat{v}_k of v_k given by $\hat{v}_k(\mathbf{x}) = \sum_{i=1}^{L_k} w_k^{(i)} \delta_{\mathbf{x}_k^{(i)}}(\mathbf{x})$ where

$$w_k^{(i)} = (1 - p_{d,k}(\mathbf{x}_k^{(i)}))w_{k|k-1}^{(i)} + \sum_{\mathbf{z} \in \mathcal{Z}_k} \frac{p_{d,k}(\mathbf{x}_k^{(i)})g_k(\mathbf{z}|\mathbf{x}_k^{(i)})w_{k|k-1}^{(i)}}{\kappa(\mathbf{z}) + \sum_{i=1}^{L_{k|k-1}} p_{d,k}(\mathbf{x}_k^{(i)})g_k(\mathbf{z}|\mathbf{x}_k^{(i)})w_{k|k-1}^{(i)}}. \quad (4)$$

Variants of the algorithm based on APF have also been proposed [16], and a Gaussian Particle algorithm which propagates a GM has been proposed in [24]. From now on we only focus on the SMC implementation of the PHD filter.

2) *Analysis of the SMC PHD Filter:* The SMC implementation of the PHD filter suffers from the same drawbacks as those of single object SMC filtering algorithms. The critical term in (4) is $p_{d,k}(\mathbf{x}_k^{(i)})g_k(\mathbf{z}|\mathbf{x}_k^{(i)})w_{k|k-1}^{(i)}$, since it will determine the total mass $\int v_k(\mathbf{x}_k) d\mathbf{x}_k$ which is an estimator of the number of targets. This term depends on the likelihood $g_k(\mathbf{z}|\mathbf{x}_k^{(i)})$. If all particles have a weak likelihood with all measurements, even if some measurements are due to the presence of targets, then this estimator of the number of targets will be close to 0. Such a phenomenon occurs when we do not succeed in concentrating particles in regions of high likelihood, and this task is all the more demanding when the dimension of the particles gets high. Moreover it will be difficult to produce an accurate estimator of the state of a target, if few particles contribute to the mass of a region S where a target is probably present. To cope with this issue without increasing the number of particles, some solutions can be thought of.

First, one can try to sample particles according to an importance distribution which would guide particles toward high likelihood regions, as in single object filtering. However this is a challenging point when we use IS in multi-object filtering, since we do not know which measurement is due to which detected target. An auxiliary implementation can address this problem by first selecting randomly measurements which are probably due to the presence of targets [16].

Next, one can avoid the dispersion of particles by using random samples only when necessary, i.e. for some components of the state vector.

As we show in this paper, these two solutions are not incompatible, and we will actually combine them. More precisely, we examine a class of models for which the GM and SMC approaches can be combined. For this class of models, it is possible to use the classical SMC approach, but by taking into account the specificities of the model one can improve both the estimation of states of the targets and that of the number of targets, without increasing the number of particles. We first address the dimensional reduction problem; next we derive a bootstrap and an APF implementation of our M-PHD filter.

III. AN M-PHD FILTER

A. A brief review of MPF in single object filtering

Assume that the state vector \mathbf{x}_k of a target can be divided into a linear component \mathbf{x}_k^l (with dimension m_l) and a non linear one \mathbf{x}_k^n (with dimension m_n) (i.e., \mathbf{x}_k^l follows a linear dynamics and \mathbf{x}_k^n a non linear one), and that the relations between these components between k and $k+1$ and the associated measurement \mathbf{z}_k are given by:

$$\mathbf{x}_{k+1}^n = \mathbf{f}_k^n(\mathbf{x}_k^n) + \mathbf{F}_k^n(\mathbf{x}_k^n)\mathbf{x}_k^l + \mathbf{u}_k^n, \quad (5)$$

$$\mathbf{x}_{k+1}^l = \mathbf{f}_k^l(\mathbf{x}_k^n) + \mathbf{F}_k^l(\mathbf{x}_k^n)\mathbf{x}_k^l + \mathbf{u}_k^l, \quad (6)$$

$$\mathbf{z}_{k+1} = \mathbf{h}_k(\mathbf{x}_{k+1}^n) + \mathbf{H}_k(\mathbf{x}_{k+1}^n)\mathbf{x}_{k+1}^l + \mathbf{v}_k, \quad (7)$$

in which $\mathbf{f}_k^n(\cdot)$ is a function from \mathbb{R}^{m_n} to \mathbb{R}^{m_n} , $\mathbf{F}_k^n(\cdot)$ is a $m_n \times m_l$ matrix which depends on the components of vector \mathbf{x}_k^n and which acts on vector \mathbf{x}_k^l (functions $\mathbf{f}_k^l(\cdot)$, $\mathbf{h}_k(\cdot)$, $\mathbf{F}_k^l(\cdot)$ and $\mathbf{H}_k(\cdot)$ are defined similarly), and

$$\begin{bmatrix} \mathbf{u}_k^n \\ \mathbf{u}_k^l \\ \mathbf{v}_k \end{bmatrix} \sim \mathcal{N} \left(\cdot; \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; \begin{bmatrix} \mathbf{Q}_k^n & (\mathbf{Q}_k^{nl})^T & \mathbf{0} \\ (\mathbf{Q}_k^{nl}) & \mathbf{Q}_k^l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_k \end{bmatrix} \right), \quad (8)$$

where $[\mathbf{u}_1^n, \mathbf{u}_1^l], \dots, [\mathbf{u}_1^k, \mathbf{u}_1^l]$ and $\mathbf{v}_1, \dots, \mathbf{v}_k$ are independent. A probabilistic representation of this model is given by the transition and the likelihood pdf ($\mathcal{N}(\mathbf{x}; \mathbf{m}; \mathbf{P})$ is the Gaussian pdf with mean \mathbf{m} and covariance matrix \mathbf{P} taken at point \mathbf{x}):

$$f_{k+1|k}(\mathbf{x}_{k+1}|\mathbf{x}_k) = \mathcal{N} \left(\mathbf{x}_{k+1}; \begin{bmatrix} \mathbf{f}_k^n(\mathbf{x}_k^n) \\ \mathbf{f}_k^l(\mathbf{x}_k^n) \end{bmatrix} + \begin{bmatrix} \mathbf{F}_k^n(\mathbf{x}_k^n) \\ \mathbf{F}_k^l(\mathbf{x}_k^n) \end{bmatrix} \mathbf{x}_k^l; \begin{bmatrix} \mathbf{Q}_k^n & (\mathbf{Q}_k^{nl})^T \\ (\mathbf{Q}_k^{nl}) & \mathbf{Q}_k^l \end{bmatrix} \right), \quad (9)$$

$$g_{k+1}(\mathbf{z}|\mathbf{x}_{k+1}) = \mathcal{N}(\mathbf{z}; \mathbf{h}_k(\mathbf{x}_{k+1}^n) + \mathbf{H}_k(\mathbf{x}_{k+1}^n)\mathbf{x}_{k+1}^l; \mathbf{R}_k). \quad (10)$$

1) *MPF implementation:* Let us now recall MPF [5] [4] [25] in model (5)-(7). Remember that the goal consists in approximating the posterior density $p(\mathbf{x}_k|\mathbf{z}_{0:k})$. MPF is based on the factorization

$$p(\mathbf{x}_k^l, \mathbf{x}_{0:k}^n | \mathbf{z}_{0:k}) = \underbrace{p(\mathbf{x}_k^l | \mathbf{x}_{0:k}^n, \mathbf{z}_{0:k})}_{\text{KF}} \times \underbrace{p(\mathbf{x}_{0:k}^n | \mathbf{z}_{0:k})}_{\text{PF}}. \quad (11)$$

The second factor $p(\mathbf{x}_{0:k}^n | \mathbf{z}_{0:k})$ is approximated by PF, but since system (5)-(7) becomes a classical state-space model once \mathbf{x}_k^n is fixed, the first factor $p(\mathbf{x}_k^l | \mathbf{x}_{0:k}^n, \mathbf{z}_{0:k})$ can be computed exactly via KF. This idea is popular and its interest is enlightened by the Rao-Blackwell (RB) theorem applied to sequential filtering: assume that one needs to estimate $\Theta = \int f(x_1, x_2)p(x_1, x_2)dx_1dx_2$, in a context where it is possible to calculate $\int f(x_1, x_2)p(x_2|x_1)dx_2$ exactly. Consider the IS based estimators $\hat{\Theta}_1$ (resp. $\hat{\Theta}_2$) of Θ obtained by sampling $(x_1^i, x_2^i) \stackrel{\text{i.i.d}}{\sim} q(x_1, x_2) = q(x_1)q(x_2|x_1)$ (resp. $x_1^i \stackrel{\text{i.i.d}}{\sim} q(x_1)$):

$$\hat{\Theta}_1 = \frac{1}{N} \sum_{i=1}^N \frac{f(x_1^i, x_2^i)p(x_1^i, x_2^i)}{q(x_1^i, x_2^i)}, \quad (12)$$

$$\hat{\Theta}_2 = \frac{1}{N} \sum_{i=1}^N \int \frac{f(x_1^i, x_2)p(x_1^i)}{q(x_1^i)} p(x_2|x_1^i) dx_2. \quad (13)$$

Then $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased, but the RB equality reads

$$\text{var}(\hat{\Theta}_1) = \text{var}(\hat{\Theta}_2) + \frac{1}{N} \mathbb{E} \left[\text{var} \left(\frac{f(x_1^1, x_2^1)p(x_1^1, x_2^1)}{q(x_1^1, x_2^1)} | x_1^1 \right) \right].$$

So $\text{var}(\widehat{\Theta}_1) \geq \text{var}(\widehat{\Theta}_2)$, and thus $\widehat{\Theta}_2$ is preferable to $\widehat{\Theta}_1$.

Remark 1 In the sequential filtering context we look for calculating a moment of a function of interest f given by $\int f(\mathbf{x}_{1,0:n}, \mathbf{x}_{2,0:n}) p(\mathbf{x}_{1,0:n}, \mathbf{x}_{2,0:n} | \mathbf{z}_{0:n}) d\mathbf{x}_{1,0:n} d\mathbf{x}_{2,0:n}$, but the normalizing constant of $p(\mathbf{x}_{1,0:n}, \mathbf{x}_{2,0:n} | \mathbf{z}_{0:n})$ is unknown. So we have to resort to Bayesian (or normalized) IS [2] and the previous analysis does not hold any longer. Nevertheless, the result still holds when the number of particles tends to infinity [25]. Also, some contributions take into account the resampling step which creates dependency among particles and compare both estimators asymptotically [26].

2) *IS vs APF*: Next, even in a MPF implementation, we need to sample particles. So let us briefly describe two well-known techniques. Let us assume that we need to compute $\Theta = \int f(\mathbf{x}) \pi(\mathbf{x}) d\mathbf{x}$ where $\pi(\mathbf{x}) = \sum_{i=1}^N \tau^i(\mathbf{x}) p_i(\mathbf{x})$. Note that $\pi(\mathbf{x})$ is not necessarily a pdf since $\int \pi(\mathbf{x}) d\mathbf{x}$ can differ from 1. A classical IS technique draws N samples $\mathbf{x}^i \sim q_i(\mathbf{x})$, where $q_i(\mathbf{x})$ is a given pdf, and computes $\widehat{\Theta}^{\text{IS}} = \sum_{i=1}^N w^i(\mathbf{x}^i) f(\mathbf{x}^i)$ where $w^i(\mathbf{x}^i) = \tau^i(\mathbf{x}^i) p_i(\mathbf{x}^i) / q_i(\mathbf{x}^i)$. An alternative consists in considering IS in an augmented dimension which is also known as the auxiliary implementation [27]. More precisely, we consider an importance distribution $q(k, \mathbf{x})$. We first select an index $j^i \sim q(k)$ and we sample $\mathbf{x}^i \sim q(\mathbf{x} | j^i)$. An estimator of Θ is given by $\widehat{\Theta}^{\text{APF}} = \sum_{i=1}^N w^i(\mathbf{x}^i) f(\mathbf{x}^i)$ where $w^i(\mathbf{x}^i) = \tau^{j^i}(\mathbf{x}^i) p_{j^i}(\mathbf{x}^i) / q(j^i, \mathbf{x}^i)$.

One of the advantages of the APF implementation is that it is possible to select a number $N' \neq N$ of samples. Next, the first step is actually a resampling one which can be beneficial or detrimental according to parameters p_i , τ^i and f . A comparison between both methods when the optimal importance distributions are chosen has been studied in several contributions, see e.g. [28] and [29].

B. Extension to multi-object filtering : the M-PHD filter

We now address PHD-based multi-object filtering algorithms. Remember from section II-B that the expected number of targets at time k is given by $\int_{\mathbb{R}^m} v_k(\mathbf{x}) d\mathbf{x}$, where $v_k(\mathbf{x})$ is the a posteriori filtering PHD (i.e., the PHD of the posterior mtd $p(X_k | Z_{0:k})$), and that looking for states boils down to finding regions $S \subset \mathbb{R}^m$ such that $\int_S v_k(\mathbf{x}) d\mathbf{x}$ is high. Estimating $\int_S v_k(\mathbf{x}) d\mathbf{x}$ over any region S (or over \mathbb{R}^m) is thus essential in the PHD-based approach, since in practice the integral is not directly computable.

The aim of this section is to extend the MPF idea from the single to the multi object context, in the case where (5)-(7) now describe the dynamics of one single target between instants k and $k+1$, and the link between the state of a detected target and the associated measurement. We will show that in such a model, under some additional assumptions, the PHD can be integrated with respect first to the non linear part of the state, and next the linear one, and the second integral can be computed exactly from a PF approximation of the first one:

$$\int v_k(\mathbf{x}) d\mathbf{x} = \int \underbrace{\left[\int v_k(\mathbf{x}^n, \mathbf{x}^l) d\mathbf{x}^n \right]}_{\text{PF implementation}} d\mathbf{x}^l = \int \underbrace{\tilde{v}_k(\mathbf{x}^l) d\mathbf{x}^l}_{\text{Computable}}.$$

Let us first give some conditions under which it will be possible to develop an M-PHD filter. We assume that

A.1 The probabilities of survival $p_{s,k}(\mathbf{x}) = p_{s,k}(\mathbf{x}^n)$ and of detection $p_{d,k}(\mathbf{x}) = p_{d,k}(\mathbf{x}^n)$ depend only on the non-linear component of the state vector;

A.2 Let $p_{i,k+1}(\mathbf{x})$ be given pdfs, for $1 \leq i \leq N_{\gamma_{k+1}}$. We assume that $\gamma_k(\mathbf{x}_{k+1})$ is a mixture (subscript ‘‘2’’ is used for birth components)

$$\gamma_{k+1}(\mathbf{x}) = \sum_{i=1}^{N_{\gamma_{k+1}}} w_{k+1|k}^{2,(i)} p_{i,k+1}(\mathbf{x}^n) \times \mathcal{N}(\mathbf{x}^l; \mathbf{m}_{k+1|k}^{2,l,(i)}(\mathbf{x}^n); \mathbf{P}_{k+1|k}^{2,l,(i)}). \quad (14)$$

Note that γ_{k+1} is not necessarily a GM, which by contrast is an hypothesis required in the GM implementation of the PHD filter.

The key idea of our approach is to assume that at time k and for a given value \mathbf{x}^l the PHD is approximated by a discrete measure which support is given by a set $A_k = \{\mathbf{x}_k^{n,(i)}\}_{i=1}^{L_k}$ (one can have $\mathbf{x}_k^{n,(i)} = \mathbf{x}_k^{n,(j)}$ for $i \neq j$) and for a given element of A_k , the PHD is approximated by a GM. In other words, we assume that an approximation of the PHD at time k is a mixture with the following form, hereafter denoted normal discrete (ND):

$$\hat{v}_k(\mathbf{x}) = \sum_{i=1}^{L_k} w_k^{(i)} \mathcal{N}(\mathbf{x}^l; \mathbf{m}_k^{l,(i)}; \mathbf{P}_k^{l,(i)}) \delta_{\mathbf{x}_k^{n,(i)}}(\mathbf{x}^n), \quad (15)$$

with $\mathbf{x} = (\mathbf{x}^l, \mathbf{x}^n)$. So we now look for deriving an ND approximation of the PHD at time $k+1$; starting from an ND approximation $\hat{v}_k(\mathbf{x})$ of $v_k(\mathbf{x})$, we derive a continuous approximation $\tilde{v}_{k+1|k}(\mathbf{x})$ of $v_{k+1|k}(\mathbf{x})$.

1) *Prediction*: Let $\mathbf{F}_k^{l,(i)} \triangleq \mathbf{F}_k^l(\mathbf{x}_k^{n,(i)})$, $\mathbf{F}_k^{n,(i)} \triangleq \mathbf{F}_k^n(\mathbf{x}_k^{n,(i)})$, $\mathbf{f}_k^{l,(i)} \triangleq \mathbf{f}_k^l(\mathbf{x}_k^{n,(i)})$, $\mathbf{f}_k^{n,(i)} \triangleq \mathbf{f}_k^n(\mathbf{x}_k^{n,(i)})$ and $p_{s,k+1}^{(i)} \triangleq p_{s,k+1}(\mathbf{x}_k^{n,(i)})$. Let \hat{v}_k in (15) be an ND approximation of PHD v_k . Plugging (15) into (2) yields the following approximation $\tilde{v}_{k+1|k}$ of the predicted PHD $v_{k+1|k}$:

$$\tilde{v}_{k+1|k}(\mathbf{x}) = \tilde{v}_{k+1|k}^1(\mathbf{x}) + \tilde{v}_{k+1|k}^2(\mathbf{x}), \quad (16)$$

$$\tilde{v}_{k+1|k}^1(\mathbf{x}) = \sum_{i=1}^{L_k} w_{k+1|k}^{1,(i)} \mathcal{N}(\mathbf{x}^l; \mathbf{m}_{k+1|k}^{1,l,(i)}(\mathbf{x}^n); \mathbf{P}_{k+1|k}^{1,l,(i)}) \times \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{1,n,(i)}; \mathbf{P}_{k+1|k}^{1,n,(i)}), \quad (17)$$

$$\tilde{v}_{k+1|k}^2(\mathbf{x}) = \gamma_{k+1}(\mathbf{x}), \quad (18)$$

where

$$w_{k+1|k}^{1,(i)} = p_{s,k+1}^{(i)} w_k^{(i)}, \quad (19)$$

$$\mathbf{m}_{k+1|k}^{1,n,(i)} = \mathbf{f}_k^{n,(i)} + \mathbf{F}_k^{n,(i)} \mathbf{m}_k^{l,(i)}, \quad (20)$$

$$\mathbf{P}_{k+1|k}^{1,n,(i)} = \mathbf{Q}_k^n + \mathbf{F}_k^{n,(i)} \mathbf{P}_k^{l,(i)} \mathbf{F}_k^{n,(i)T}, \quad (21)$$

$$\mathbf{m}_{k+1|k}^{1,l,(i)}(\mathbf{x}^n) = \mathbf{A}_k^{(i)} \tilde{\mathbf{m}}_{k+1|k}^{l,(i)}(\mathbf{x}^n) + f_k^{l,(i)} + \tilde{\mathbf{Q}}_k^n(\mathbf{x}^n - f_k^{n,(i)}), \quad (22)$$

$$\mathbf{P}_{k+1|k}^{1,l,(i)} = \mathbf{Q}_k^l - (\tilde{\mathbf{Q}}_k^n)^{-1} (\mathbf{Q}_k^{nl})^T + \mathbf{A}_k^{(i)} \tilde{\mathbf{P}}_{k+1|k}^{l,(i)} \mathbf{A}_k^{(i)T}, \quad (23)$$

where $\tilde{\mathbf{m}}_{k+1|k}^{l,(i)}(\mathbf{x}^n)$, $\tilde{\mathbf{P}}_{k+1|k}^{l,(i)}$ and $\mathbf{A}_k^{(i)}$ are defined by

$$\tilde{\mathbf{m}}_{k+1|k}^{l,(i)}(\mathbf{x}^n) = \mathbf{m}_k^{l,(i)} + \tilde{\mathbf{K}}_{k+1|k}^{l,(i)} \left[\mathbf{x}^n - f_k^{n,(i)} - \mathbf{F}_k^{n,(i)} \mathbf{m}_k^{l,(i)} \right], \quad (24)$$

$$\tilde{\mathbf{P}}_{k+1|k}^{l,(i)} = \left[\mathbf{I} - \tilde{\mathbf{K}}_{k+1|k}^{l,(i)} \mathbf{F}_k^{n,(i)} \right] \mathbf{P}_k^{l,(i)}, \quad (25)$$

$$\mathbf{A}_k^{(i)} = \mathbf{F}_k^{l,(i)} - (\tilde{\mathbf{Q}}_k^n)^{-1} \mathbf{F}_k^{n,(i)}, \quad (26)$$

$$\tilde{\mathbf{K}}_{k+1|k}^{l,(i)} = \mathbf{P}_k^{l,(i)} (\mathbf{F}_k^{n,(i)})^T (\mathbf{P}_{k+1|k}^{n,(i)})^{-1}, \quad (27)$$

$$\tilde{\mathbf{Q}}_k^n = \mathbf{Q}_k^{nl} (\mathbf{Q}_k^n)^{-1}. \quad (28)$$

Next, we look for an approximation of posterior PHD v_{k+1} .

2) *Update* : Let us set $Z'_{k+1} = \{\theta\} \cup Z_{k+1}$ where θ corresponds to a null measurement. Plugging (16) in (3), we get the following approximation of v_{k+1} :

$$\tilde{v}_{k+1}(\mathbf{x}) = \tilde{v}_{k+1}^1(\mathbf{x}) + \tilde{v}_{k+1}^2(\mathbf{x}), \quad (29)$$

where

$$\tilde{v}_{k+1}^1(\mathbf{x}) = \sum_{i=1}^{L_k} \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{1,n,(i)}; \mathbf{P}_{k+1|k}^{1,n,(i)}) \times \sum_{\mathbf{z} \in Z'_{k+1}} \Phi_{k+1}^{1,(i)}(\mathbf{x}^n, \mathbf{x}^l, \mathbf{z}), \quad (30)$$

$$\tilde{v}_{k+1}^2(\mathbf{x}) = \sum_{i=1}^{N_{\gamma_{k+1}}} p_{i,k+1}(\mathbf{x}^n) \sum_{\mathbf{z} \in Z'_{k+1}} \Phi_{k+1}^{2,(i)}(\mathbf{x}^n, \mathbf{x}^l, \mathbf{z}), \quad (31)$$

$$\Phi_{k+1}^{j,(i)}(\mathbf{x}^n, \mathbf{x}^l, \mathbf{z}) = \mathcal{N}(\mathbf{x}^l; \mathbf{m}_{k+1}^{j,l,(i)}(\mathbf{z}, \mathbf{x}^n); \mathbf{P}_{k+1}^{j,l,(i)}(\mathbf{z}, \mathbf{x}^n)) \times w_{k+1}^{j,(i)}(\mathbf{z}, \mathbf{x}^n) \text{ for } j = 1, 2, \quad (32)$$

with, for $j = 1, 2$, and $\mathbf{z} \in Z'_{k+1}$,

$$w_{k+1}^{j,(i)}(\mathbf{z}, \mathbf{x}^n) = \begin{cases} (1 - p_{d,k+1}(\mathbf{x}^n)) w_{k+1|k}^{j,(i)}, & \mathbf{z} = \theta, \\ \frac{p_{d,k+1}(\mathbf{x}^n) q_{k+1}^{j,(i)}(\mathbf{z}, \mathbf{x}^n) w_{k+1|k}^{j,(i)}}{\kappa(\mathbf{z}) + \tilde{B}^1(\mathbf{z}) + \tilde{B}^2(\mathbf{z})}, & \mathbf{z} \in Z_{k+1}, \end{cases} \quad (33)$$

$$\mathbf{m}_{k+1}^{j,l,(i)}(\mathbf{z}, \mathbf{x}^n) = \begin{cases} \mathbf{m}_{k+1|k}^{j,l,(i)}(\mathbf{x}^n), & \mathbf{z} = \theta, \\ \mathbf{m}_{k+1|k}^{j,l,(i)}(\mathbf{x}^n) + \mathbf{K}_{k+1}^{j,(i)}(\mathbf{x}^n) \tilde{\mathbf{z}}_{k+1}^{(i)}(\mathbf{x}^n), & \mathbf{z} \in Z_{k+1} \end{cases} \quad (34)$$

$$\mathbf{P}_{k+1}^{j,l,(i)}(\mathbf{z}, \mathbf{x}^n) = \begin{cases} \mathbf{P}_{k+1|k}^{j,l,(i)}, & \mathbf{z} = \theta, \\ \left[\mathbf{I} - \mathbf{K}_{k+1}^{j,(i)}(\mathbf{x}^n) \mathbf{H}_k(\mathbf{x}^n) \right] \mathbf{P}_{k+1|k}^{j,l,(i)}, & \mathbf{z} \in Z_{k+1}, \end{cases} \quad (35)$$

where

$$q_{k+1}^{j,(i)}(\mathbf{z}, \mathbf{x}^n) = \mathcal{N}(\mathbf{z}; h_k(\mathbf{x}^n) + \mathbf{H}_k(\mathbf{x}^n) \mathbf{m}_{k+1|k}^{j,l,(i)}(\mathbf{x}^n); \mathbf{S}_{k+1}^{j,(i)}(\mathbf{x}^n)), \quad (36)$$

$$\mathbf{K}_{k+1}^{j,(i)}(\mathbf{x}^n) = \mathbf{P}_{k+1|k}^{j,l,(i)} \mathbf{H}_k(\mathbf{x}^n)^T (\mathbf{S}_{k+1}^{j,(i)}(\mathbf{x}^n))^{-1}, \quad (37)$$

$$\tilde{\mathbf{z}}_{k+1}^{j,(i)}(\mathbf{x}^n) = \mathbf{z} - \mathbf{H}_k(\mathbf{x}^n) \mathbf{m}_{k+1|k}^{j,l,(i)}(\mathbf{x}^n) - h_k(\mathbf{x}^n), \quad (38)$$

$$\mathbf{S}_{k+1}^{j,(i)}(\mathbf{x}^n) = \mathbf{H}_k(\mathbf{x}^n) \mathbf{P}_{k+1|k}^{j,l,(i)} (\mathbf{H}_k(\mathbf{x}^n))^T + \mathbf{R}_k, \quad (39)$$

$$\tilde{B}^1(\mathbf{z}) = \sum_{i=1}^{L_k} \int w_{k+1|k}^{1,(i)} p_{d,k+1}(\mathbf{x}^n) q_{k+1}^{1,(i)}(\mathbf{z}, \mathbf{x}^n) d\mathbf{x}^n \times \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{1,n,(i)}; \mathbf{P}_{k+1|k}^{1,n,(i)}), \quad (40)$$

$$\tilde{B}^2(\mathbf{z}) = \sum_{i=1}^{N_{\gamma_{k+1}}} \int w_{k+1|k}^{2,(i)} p_{d,k+1}(\mathbf{x}^n) q_{k+1}^{2,(i)}(\mathbf{z}, \mathbf{x}^n) p_{i,k+1}(\mathbf{x}^n) d\mathbf{x}^n. \quad (41)$$

Remark 2 The derivation of the predicted and posterior approximations of the PHD holds under assumption **A.1** above. However, these results can be extended easily to the case where $p_{s,k}(\mathbf{x})$ and $p_{d,k}(\mathbf{x})$ are now mixtures (see e.g. [11]):

$$p_{s,k}(\mathbf{x}) = w_{s,k}^{(0)} p_{s,0,k}(\mathbf{x}^n) + \sum_{i=1}^{N_{s,k}} w_{s,k}^{(i)} p_{s,i,k}(\mathbf{x}^n) \mathcal{N}(\mathbf{x}^l; \mathbf{m}_{s,k}^{l,(i)}(\mathbf{x}^n); \mathbf{P}_{s,k}^{l,(i)}), \quad (42)$$

$$p_{d,k}(\mathbf{x}) = w_{d,k}^{(0)} p_{d,0,k}(\mathbf{x}^n) + \sum_{i=1}^{N_{d,k}} w_{d,k}^{(i)} p_{d,i,k}(\mathbf{x}^n) \mathcal{N}(\mathbf{x}^l; \mathbf{m}_{d,k}^{l,(i)}(\mathbf{x}^n); \mathbf{P}_{d,k}^{l,(i)}) \quad (43)$$

such that $0 \leq p_{s,k}(\mathbf{x}) \leq 1$ and $0 \leq p_{d,k}(\mathbf{x}) \leq 1$.

Now, let us focus on the computation of $\int \tilde{v}_{k+1}(\mathbf{x}) d\mathbf{x}$. Remark that $\int \tilde{v}_{k+1}(\mathbf{x}) d\mathbf{x}^l$ is computable since $\int \Phi_{k+1}^{j,(i)}(\mathbf{x}^n, \mathbf{x}^l, \mathbf{z}) d\mathbf{x}^l$ reduces to $w_{k+1}^{j,(i)}(\mathbf{z}, \mathbf{x}^n)$. So the challenge is now to compute $\int [\int \tilde{v}_{k+1}(\mathbf{x}) d\mathbf{x}^l] d\mathbf{x}^n$; two implementations are proposed.

C. A marginal Bootstrap Implementation

This implementation relies on the observation that it is possible to derive respectively a discrete approximation of $\tilde{v}_{k+1}^1(\mathbf{x})$ (this term is due to persistent targets) and of $\tilde{v}_{k+1}^2(\mathbf{x})$ (this term is due to birth targets) as a function of \mathbf{x}^n by using the IS mechanism recalled in paragraph III-A2. In addition, pdf $\mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{1,n,(i)}; \mathbf{P}_{k+1|k}^{1,n,(i)})$ and $p_{i,k+1}$ are available and can be used for sampling a new set of particles.

1) *ND approximation of \tilde{v}_{k+1}^1* : We first draw samples for persistent targets. An ND approximation of \tilde{v}_{k+1}^1 with L_k components is given by

$$\hat{v}_{k+1}^1(\mathbf{x}^n, \mathbf{x}^l) = \sum_{i=1}^{L_k} \sum_{\mathbf{z} \in Z'_{k+1}} \Phi_{k+1}^{1,(i)}(\mathbf{x}_{k+1}^{1,n,(i)}, \mathbf{x}^l, \mathbf{z}) \delta_{\mathbf{x}_{k+1}^{1,n,(i)}}(\mathbf{x}^n), \quad (44)$$

where $\mathbf{x}_{k+1}^{1,n,(i)} \sim \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{1,n,(i)}; \mathbf{P}_{k+1|k}^{1,n,(i)})$ for $1 \leq i \leq L_k$.

2) *ND approximation of \tilde{v}_{k+1}^2* : Next, we draw samples for birth targets. It may be unreliable to draw only $N_{\gamma_{k+1}}$ particles if $N_{\gamma_{k+1}}$ is small. So let us derive an ND approximation of \tilde{v}_{k+1}^2 by using a set of $L_{k+1}^2 = K \times N_{\gamma_{k+1}}$ particles. For $i = 1, \dots, L_{k+1}^2$, let $j^i = \lceil i/K \rceil$. A ND approximation of \tilde{v}_{k+1}^2 is given by

$$\hat{v}_{k+1}^2(\mathbf{x}^n, \mathbf{x}^l) = \frac{1}{K} \sum_{i=1}^{L_{k+1}^2} \sum_{\mathbf{z} \in Z_{k+1}} \Phi_{k+1}^{2,(j^i)}(\mathbf{x}_{k+1}^{2,n,(i)}, \mathbf{x}^l, \mathbf{z}) \delta_{\mathbf{x}_{k+1}^{2,n,(i)}}(\mathbf{x}^n), \quad (45)$$

where $\mathbf{x}_{n+1}^{2,n,(i)} \sim p_{j^i,k+1}(\mathbf{x}^n)$.

3) *Implementation issues:* Let $L_{k+1|k} = L_k + L_{k+1}^2$ denote the total number of samples drawn for persistent and birth targets. By gathering the ND approximations of \hat{v}_{k+1}^1 and of \hat{v}_{k+1}^2 , an ND approximation of the PHD at time $k+1$ is given by

$$\hat{v}_{k+1}(\mathbf{x}^n, \mathbf{x}^l) = \sum_{i=1}^{L_{k+1|k}} \sum_{\mathbf{z} \in Z'_{k+1}} \Phi_{k+1}^{(i)}(\mathbf{x}_{k+1}^{n,(i)}, \mathbf{x}^l, \mathbf{z}) \delta_{\mathbf{x}_{k+1}^{n,(i)}}(\mathbf{x}^n). \quad (46)$$

Following (32), we can write $\Phi_{k+1}^{(i)}(\mathbf{x}_{k+1}^{n,(i)}, \mathbf{x}^l, \mathbf{z}) = w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{n,(i)}) \mathcal{N}(\mathbf{x}^l; \mathbf{m}_{k+1}^{l,(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{n,(i)}); \mathbf{P}_{k+1}^{l,(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{n,(i)}))$, where $w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{n,(i)})$, $i = 1, \dots, L_{k+1|k}$, $\mathbf{z} \in Z'_{k+1}$ are the sample weights.

Our final approximation $\hat{v}_{k+1}(\mathbf{x}^n, \mathbf{x}^l)$ is a ND mixture with $L_{k+1|k}(1 + |Z_{k+1}|)$ components. This indicates an exponential increase in the number of components with time. Following [12], resampling according to the weights $w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{n,(i)})$ is performed to reduce the number of components to a number L_{k+1} which is proportional to the estimated number of targets. If the measurements do not depend on the linear part \mathbf{x}^l of the state the ND approximation of the PHD can be written as

$$\hat{v}_{k+1}(\mathbf{x}^n, \mathbf{x}^l) = \sum_{i=1}^{L_{k+1|k}} w_{k+1}^{(i)}(\mathbf{x}^n) \times \mathcal{N}(\mathbf{x}^l; \mathbf{m}_{k+1}^{l,(i)}(\mathbf{x}_{k+1}^{n,(i)}); \mathbf{P}_{k+1}^{l,(i)}(\mathbf{x}_{k+1}^{n,(i)})) \delta_{\mathbf{x}_{k+1}^{n,(i)}}(\mathbf{x}^n). \quad (47)$$

where $\mathbf{m}_{k+1}^{l,(i)}(\mathbf{x}_{k+1}^{n,(i)}) = \mathbf{m}_{k+1}^{l,(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{n,(i)})$, $\mathbf{P}_{k+1}^{l,(i)}(\mathbf{x}_{k+1}^{n,(i)}) = \mathbf{P}_{k+1}^{l,(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{n,(i)})$ and $w_{k+1}^{(i)}(\mathbf{x}^n, \mathbf{x}^l) = \sum_{\mathbf{z} \in Z'_{k+1}} w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{n,(i)})$. In this case the number of components in the ND mixture is given by the number $L_{k+1|k}$ of sampled particles. Resampling may still be required to remove unlikely particles and to produce a sample size proportional to the estimated number of targets.

It can be seen from (33) that computation of the sample weights $w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{n,(i)})$ requires $\tilde{B}^j(\mathbf{z})$ of (40) and (41). These quantities can be approximated by using the sampled particles:

$$\hat{B}^1(\mathbf{z}) = \sum_{i=1}^{L_k} w_{k+1|k}^{1,(i)} p_{d,k+1}(\mathbf{x}_{k+1}^{1,n,(i)}) q_{k+1}^{1,(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{1,n,(i)}), \quad (48)$$

$$\hat{B}^2(\mathbf{z}) = \frac{1}{K} \sum_{i=1}^{L_{k+1}^2} w_{k+1|k}^{2,(i)} p_{d,k+1}(\mathbf{x}_{k+1}^{2,n,(i)}) q_{k+1}^{2,(i)}(\mathbf{z}, \mathbf{x}_{k+1}^{2,n,(i)}). \quad (49)$$

D. A marginal APF Implementation

Even if our previous implementation relies on particles only for the nonlinear part of the state vector, it may be inefficient because the importance distributions with which we sampled the particles did not take into account available measurements Z_{k+1} . The main risk is that new particles are in regions where targets are not present. To address this problem, we derive an APF implementation which guides particles into promising regions. To achieve this, an efficient importance distribution should be chosen to sample the new set of particles. So in order to propose a practical and efficient implementation, we now assume that the following hypotheses are satisfied:

A'1 probability of detection $p_{d,k+1}(\mathbf{x}^n) = p_{d,k+1}$ is constant; **A'2** birth intensity γ_{k+1} in (14) is a GM, i.e. $p_{i,k+1}(\mathbf{x}^n) = \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{2,n,(i)}; \mathbf{P}_{k+1|k}^{2,n,(i)})$ in (14).

It is possible to rewrite $\tilde{v}_{k+1}(\mathbf{x})$ in a more direct form in which we do not distinguish persistent and birth targets. Setting $L_{k+1|k} = L_k + N_{\gamma_{k+1}}$, \tilde{v}_{k+1} reads

$$\tilde{v}_{k+1}(\mathbf{x}) = \sum_{i=1}^{L_{k+1|k}} \sum_{\mathbf{z} \in Z'_{k+1}} \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i)}; \mathbf{P}_{k+1|k}^{n,(i)}) \Phi_{k+1}^{(i)}(\mathbf{x}^n, \mathbf{x}^l, \mathbf{z}),$$

where $\mathbf{m}_{k+1|k}^{n,(i)}$, $\mathbf{P}_{k+1|k}^{n,(i)}$, $\Phi_{k+1}^{(i)}(\mathbf{x}^n, \mathbf{x}^l, \mathbf{z})$ are respectively equal to $\mathbf{m}_{k+1|k}^{j,n,(i)}$, $\mathbf{P}_{k+1|k}^{j,n,(i)}$, $\Phi_{k+1}^{j,(i)}(\mathbf{x}^n, \mathbf{x}^l, \mathbf{z})$ where $j = 1$ if $1 \leq i \leq L_k$ and $j = 2$ if $L_k + 1 \leq i \leq L_{k+1|k}$. Marginalizing out the linear part \mathbf{x}^l of the state gives

$$\tilde{v}_{k+1}(\mathbf{x}^n) = \sum_{i=1}^{L_{k+1|k}} \sum_{\mathbf{z} \in Z'_{k+1}} \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i)}; \mathbf{P}_{k+1|k}^{n,(i)}) w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n);$$

where $w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n) = w_{k+1}^{1,(i)}(\mathbf{z}, \mathbf{x}^n)$ for $i = 1, \dots, L_k$ and $w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n) = w_{k+1}^{2,(i-L_k)}(\mathbf{z}, \mathbf{x}^n)$ for $i = L_k + 1, \dots, L_k + N_{\gamma_{k+1}}$.

The goal is to sample L_{k+1} samples according to $\tilde{v}_{k+1}(\mathbf{x}^n)$. As recalled in paragraph III-A2 one can now use an APF implementation to that end. Due to the form of $\tilde{v}_{k+1}(\mathbf{x}^n)$ which can be seen as a mixture of mixtures, an APF implementation relies on two auxiliary variables: one for a component due to a persistent or a birth target (see point III-D1 below) and the other for measurement \mathbf{z} (see point III-D2 below). So for $1 \leq j \leq L_{k+1}$, we will randomly choose a component index $j^j \in \{1, \dots, L_{k+1|k}\}$, a measurement $\mathbf{z}^j \in Z'_{k+1}$, and finally a new particle $\mathbf{x}_{k+1}^{n,(j)}$. The auxiliary variables and the final samples are independently and identically drawn according to an importance distribution $q(\mathbf{z}, i, \mathbf{x}^n)$. The weight associated to a particle is computed from

$$w(\mathbf{z}, i, \mathbf{x}^n) = \frac{\mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i)}; \mathbf{P}_{k+1|k}^{n,(i)}) w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n)}{q(\mathbf{z}, i, \mathbf{x}^n)}. \quad (50)$$

Next, it is well-known that the importance distribution $q(\mathbf{z}, i, \mathbf{x}^n)$ has to be chosen carefully. The optimal importance distribution (i.e. that which minimizes the conditional variance of weights) is

$$q^{\text{opt}}(\mathbf{z}, i, \mathbf{x}^n) \propto \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i)}; \mathbf{P}_{k+1|k}^{n,(i)}) w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n). \quad (51)$$

However, it is not computable here. Indeed, remember that the dependence of $w_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n)$ in \mathbf{x}^n is due to the term $q_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n)$ in (36); because $\mathbf{h}(\mathbf{x}^n)$ and $\mathbf{H}(\mathbf{x}^n)$ are not linear in \mathbf{x}^n , $\int \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i)}; \mathbf{P}_{k+1|k}^{n,(i)}) q_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n) d\mathbf{x}^n$ is not computable so the optimal distribution is not either. So we need to derive an importance distribution which approximates the optimal one; it relies on the approximations

$$p^{(i)}(\mathbf{z}) \approx \int q_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n) \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i)}; \mathbf{P}_{k+1|k}^{n,(i)}) d\mathbf{x}^n \quad (52)$$

$$p^{(i)}(\mathbf{x}^n | \mathbf{z}) \tilde{\propto} q_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n) \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i)}; \mathbf{P}_{k+1|k}^{n,(i)}) \quad (53)$$

where $\tilde{\propto}$ means ‘‘is approximately proportional to’’. It is well known that Gaussian approximations, $p^{(i)}(\mathbf{z})$ and $p^{(i)}(\mathbf{x}^n | \mathbf{z})$,

to the predictive likelihood and the optimal sampling density can be obtained by linearizing $\mathbf{h}(\cdot)$ and $\mathbf{H}(\cdot)$ or by applying the UT [16] [30] [31]. These approximations are accurate if the supports of the prior and the likelihood overlap. This occurs, for instance, if the prior is informative as in the case of a persisting target. However, if there is little overlap between the prior and the likelihood then both the UT and linearization perform poorly. This generally happens for birth targets which initially have a large amount of uncertainty in their position and velocity. To avoid this problem we use progressive correction (PC) to find the Gaussian approximations in (52) and (53). PC was originally proposed for Monte Carlo approximation [32] but can also be used more generally [33]. A brief review is given in the appendix. We next explain step by step how to draw a sample $(i^j, \mathbf{z}^j, \mathbf{x}_{k+1}^{n,(j)}) \sim q(i) \times q(\mathbf{z}|i) \times q(\mathbf{x}^n|i, \mathbf{z})$.

1) *Selection of old components:* The first step is to select an index $i \in \{1, \dots, L_{k+1|k}\}$ which governs how the state vector is sampled. If $i \in \{1, \dots, L_k\}$, then we have selected a persisting target with the state to be drawn conditional on the i th sample $\mathbf{x}_k^{n,(i)}$ at time k . If $i \in \{L_k + 1, \dots, L_k + N_{\gamma_{k+1}}\}$ then the state is drawn from the $(i - L_k)$ th birth density. For $\mathbf{z} \in Z'_{k+1}$, define

$$\tilde{q}(\mathbf{z}|i) = \begin{cases} 1 - p_{d,k+1}, & \mathbf{z} = \theta, \\ p_{d,k+1} p^{(i)}(\mathbf{z}) / [\kappa(\mathbf{z}) + \hat{B}(\mathbf{z})], & \mathbf{z} \in Z_{k+1}, \end{cases} \quad (54)$$

where $\hat{B}(\mathbf{z})$ is an approximation of $\tilde{B}_1(\mathbf{z}) + \tilde{B}_2(\mathbf{z})$. The computation of $\hat{B}(\mathbf{z})$ is discussed below. The sampling density for the component index is

$$q(i) = w_{k+1|k}^{(i)} \sum_{\mathbf{z} \in Z'_{k+1}} \tilde{q}(\mathbf{z}|i) / C_{k+1}, \quad (55)$$

$$C_{k+1} = \sum_{i=1}^{L_{k+1|k}} w_{k+1|k}^{(i)} \sum_{\mathbf{z} \in Z'_{k+1}} \tilde{q}(\mathbf{z}|i) \quad (56)$$

If approximation (70) is accurate, then sampling from (55) will produce components corresponding to persisting or birth particles as suggested by the measurements. This has the potential to greatly improve efficiency compared to blind sampling using the prior, particularly in the initiation of new targets.

2) *Selection of a measurement:* The next step is to select a measurement with which to associate the selected component index. This can be achieved by selecting $\mathbf{z} \in Z'_{k+1}$ from

$$q(\mathbf{z}|i) = \tilde{q}(\mathbf{z}|i) \Big/ \sum_{\mathbf{y} \in Z'_{k+1}} \tilde{q}(\mathbf{y}|i) \quad (57)$$

3) *Sampling the state:* Finally, we extend, or initiate for the case of a birth component, the i -th trajectory, which is associated to measurement \mathbf{z} by sampling from

$$q(\mathbf{x}^n|i, \mathbf{z}) = \begin{cases} \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i^j)}, \mathbf{P}_{k+1|k}^{n,(i^j)}), & \mathbf{z} = \theta, \\ p^{(i)}(\mathbf{x}^n|\mathbf{z}), & \mathbf{z} \in Z_{k+1}. \end{cases} \quad (58)$$

4) *Computation of weights:* Let $i^j, \mathbf{z}^j, \mathbf{x}_{k+1}^{n,(j)}$, $j = 1, \dots, L_{k+1}$ denote the j th sampled component index, measurement and particle. According to (50), the expression of

weights is given by

$$w_{k+1}^{(j)} = \begin{cases} C_{k+1}/L_{k+1}, & \mathbf{z}^j = \theta, \\ \frac{C_{k+1} q_{k+1}^{(i^j)}(\mathbf{z}^j, \mathbf{x}_{k+1}^{n,(j)}) \mathcal{N}(\mathbf{x}_{k+1}^{n,(j)}; \mathbf{m}_{k+1|k}^{n,(i)}; \mathbf{P}_{k+1|k}^{n,(i)})}{L_{k+1} p^{(i^j)}(\mathbf{z}^j) p^{(i^j)}(\mathbf{x}_{k+1}^{n,(j)}|\mathbf{z}^j)}, & \mathbf{z}^j \in Z_{k+1}. \end{cases} \quad (59)$$

It can be seen from (59) that if (52) and (53) are accurate then the weights will be approximately uniform with $w_{k+1}^{(j)} \approx C_{k+1}/L_{k+1}$. Finally, a ND approximation of posterior PHD v_{k+1} can be found as

$$\hat{v}_{k+1}(\mathbf{x}) = \sum_{j=1}^{L_{k+1}} w_{k+1}^{(j)} \times \mathcal{N}(\mathbf{x}; \mathbf{m}_{k+1}^{l,(i^j)}(\mathbf{z}^j, \mathbf{x}_{k+1}^{n,(j)}); \mathbf{P}_{k+1}^{l,(i^j)}(\mathbf{z}^j, \mathbf{x}_{k+1}^{n,(j)}) \delta_{\mathbf{x}_{k+1}^{n,(j)}}(\mathbf{x}^n). \quad (60)$$

In a bootstrap filter the quantities $\tilde{B}^j(\mathbf{z})$ of (40)-(41) can be computed using the particles drawn from the prior and then used in the weight calculation. In the APF proposed here these quantities are required *before* sampling is performed. We propose a computationally efficient approximation using the same sampling distributions used for the APF. In particular, for $i = 1, \dots, L'$, we draw an index j^i such that $\Pr(j^i = b) \propto w_{k+1|k}^{(b)} p^{(b)}(\mathbf{z})$ and then sample $\tilde{\mathbf{x}}^{(i)} \sim p^{j^i}(\mathbf{x}^n|\mathbf{z})$. We then obtain the approximation

$$\hat{B}(\mathbf{z}) = \frac{\hat{B}_0(\mathbf{z})}{L'} \sum_{i=1}^{L'} \frac{\mathcal{N}(\tilde{\mathbf{x}}^{(i)}; \mathbf{m}_{k+1|k}^{n,(j^i)}; \mathbf{P}_{k+1|k}^{n,(j^i)}) q_{k+1}^{(j^i)}(\mathbf{z}, \tilde{\mathbf{x}}^{(i)})}{p^{j^i}(\mathbf{z}) p^{j^i}(\tilde{\mathbf{x}}^{(i)}|\mathbf{z})}, \quad (61)$$

where $\hat{B}_0(\mathbf{z}) = p_{d,k+1} \sum_{i=1}^{L_{k+1|k}} w_{k+1|k}^{(i)} p^{(i)}(\mathbf{z})$. If the approximations (52) and (53) is good then the terms in the sum (61) will be close to unity and a small value can be chosen for the sample size L' . Remark that in this case $\hat{B}(\mathbf{z}) \approx \hat{B}_0(\mathbf{z})$.

IV. OTHER APPLICATIONS OF THE M-PHD FILTER

We next focus on models where dependence on a discrete MC $\{r_k\}$ is introduced: we now to consider the augmented state (\mathbf{x}, r) and look for computing the integral of PHD $\sum_r \int v_k(\mathbf{x}, r) d\mathbf{x}$. We explain how the M-PHD algorithm of the previous section can be adapted to Gaussian and linear (Section IV-A) or partially linear (Section IV-B) JMSS.

A. Linear and Gaussian JMSS

Let

$$f_{k+1|k}(\mathbf{x}^l, r|\mathbf{x}_k^l, r_k) = p(r|r_k) \mathcal{N}(\mathbf{x}^l; \mathbf{F}_k(r) \mathbf{x}_k^l; \mathbf{Q}_k(r)), \quad (62)$$

$$g_k(\mathbf{z}|\mathbf{x}_{k+1}^l, r_{k+1}) = \mathcal{N}(\mathbf{z}; \mathbf{H}_k(r_{k+1}) \mathbf{x}_{k+1}^l; \mathbf{R}_k(r_{k+1})). \quad (63)$$

Let us also assume the following hypotheses:

A.3 The probabilities of survival $p_{s,k+1}(r)$ and of detection $p_{d,k+1}(r)$ depend only on the mode;

A.4 The birth intensity is a mixture which reads

$$\gamma_{k+1}(\mathbf{x}^l, r) = \sum_{i=1}^{N_{\gamma_{k+1}}} w_{k+1|k}^{2,(i)} p_{i,k+1}(r) \times \mathcal{N}(\mathbf{x}^l; \mathbf{m}_{k+1|k}^{2,l,(i)}(r); \mathbf{P}_{k+1|k}^{2,l,(i)}(r)). \quad (64)$$

The mode $r_k \in \{1, \dots, d\}$, where d is the number of modes, is a discrete MC characterized by the probability transition $p(r_k|r_{k-1})$. In such models, the PHD $v_k(\mathbf{x}, r)$ is a GM which grows exponentially due to the measurements and the number of modes. A GM implementation [18] has been proposed to propagate $v_k(\mathbf{x}, r)$ and relies on numerical approximations such as pruning and merging to avoid the exponential growth. Here, we propose an alternative which involves adapting the M-PHD in order to avoid the growth of the mixture due to the number d of different modes.

In such models, (2) and (3) still hold if we consider the augmented state $\mathbf{x} = (\mathbf{x}^l, r)$. The role of the non linear part of state vector \mathbf{x} is now played by the MC $\{r_k\}$; so a marginalized implementation for these models appears as a particular case of our previous M-PHD where we replace \mathbf{x}^n by r . Therefore, we now look for propagating samples which act on the discrete part r of the state vector and so for propagating an ND implementation of the PHD,

$$\hat{v}_k(\mathbf{x}, r) = \sum_{i=1}^{L_k} w_k^{(i)} \mathcal{N}(\mathbf{x}; \mathbf{m}_k^{(i)}; \mathbf{P}_k^{(i)}) \delta_{r_k^{(i)}}(r). \quad (65)$$

Starting from (65), the predicted and updated approximations (16) and (29) hold up to some adaptations: dependence on $\mathbf{x}_k^{n,(i)}$ and \mathbf{x}^n have to be replaced by $r_k^{(i)}$ and r ; so pdf $\mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{1,n,(i)}; \mathbf{P}_{k+1|k}^{1,n,(i)})$ in (16) is replaced by $p(r|r_k^{(i)})$; $\mathbf{m}_{k+1|k}^{1,l,(i)}$ and $\mathbf{P}_{k+1|k}^{1,l,(i)}$ in (22) and (23) reduce to $\mathbf{m}_{k+1|k}^{1,l,(i)} = \mathbf{F}_k^{l,(i)} \mathbf{m}_k^{l,(i)}$, and $\mathbf{P}_{k+1|k}^{1,l,(i)} = \mathbf{F}_k^{l,(i)} \mathbf{P}_k^{l,(i)} \mathbf{F}_k^{l,(i)T} + \mathbf{Q}_k^{l,(i)}$.

Based on this observation, our two implementations still hold except that we draw samples $\{r_{k+1}^{(i)}\}$:

- the bootstrap implementation of Section III-C now relies on distributions $p(r|r_k^{(i)})$ to sample particles for persistent targets and on $p_{i,k+1}(r)$ to sample particles for birth targets;
- the APF implementation of Section III-D relies on an importance distribution $q(\mathbf{z}, i, r)$; the optimal importance distribution (51) is however computable contrary to the previous case (since integrals on \mathbf{x}^n which were not computable are now replaced by a sum on r).

B. Non-Linear and Gaussian JMSS, with linear substructure

Finally we adapt the marginalized PHD algorithm proposed in Section III-B to a more general class of models. Let $\xi_k = (\mathbf{x}_k^n, r_{k+1})$ and $\psi_{k+1} = (\mathbf{x}_{k+1}^n, r_{k+1})$. Then,

$$f_{k+1|k}(\mathbf{x}|\mathbf{x}_k) = p(r_{k+1}|r_k) \times \mathcal{N}\left(\mathbf{x}; \begin{bmatrix} \mathbf{F}_k^n(\xi_k) \\ \mathbf{f}_k^l(\xi_k) \end{bmatrix} + \begin{bmatrix} \mathbf{F}_k^n(\xi_k) \\ \mathbf{F}_k^l(\xi_k) \end{bmatrix} \mathbf{x}_k; \begin{bmatrix} \mathbf{Q}_k^n(r_{k+1}) & (\mathbf{Q}_k^{nl}(r_{k+1}))^T \\ (\mathbf{Q}_k^{nl}(r_{k+1})) & \mathbf{Q}_k^l(r_{k+1}) \end{bmatrix}\right), \quad (66)$$

$$g_{k+1}(\mathbf{z}|\mathbf{x}_{k+1}) = \mathcal{N}(\mathbf{z}; \mathbf{h}_k(\psi_{k+1}) + \mathbf{H}_k(\psi_{k+1})\mathbf{x}_{k+1}^l; \mathbf{R}_k(r_{k+1})). \quad (67)$$

A.5 The probabilities of survival $p_{s,k+1}(\mathbf{x}^n, r)$ and of detection $p_{d,k+1}(\mathbf{x}^n, r)$ depend only on the mode r and nonlinear part of the kinematic state;

A.6 The birth intensity γ_{k+1} reads

$$\gamma_{k+1}(\mathbf{x}, r) = \sum_{i=1}^{N_{\gamma_{k+1}}} w_{k+1|k}^{2,l,(i)} p_{i,k+1}(r) p_{i,k+1}(\mathbf{x}^n|r) \times \mathcal{N}(\mathbf{x}; \mathbf{m}_{k+1|k}^{2,l,(i)}(r); \mathbf{P}_{k+1|k}^{2,l,(i)}(r)). \quad (68)$$

Again, the technique of Section III can be applied to derive an ND approximation of the PHD in such models. The state vector is now $\mathbf{x} = (\mathbf{x}^l, \mathbf{x}^n, r)$, with non linear part (\mathbf{x}^n, r) ; so the goal now is to propagate samples $\{\mathbf{x}_k^{n,(i)}, r_k^{(i)}\}$.

Starting from an ND approximation at time k

$$\hat{v}_k(\mathbf{x}, r) = \sum_{i=1}^{L_k} w_k^{(i)} \mathcal{N}(\mathbf{x}^l; \mathbf{m}_k^{l,(i)}; \mathbf{P}_k^{l,(i)}) \delta_{\mathbf{x}_k^{n,(i)}, r_k^{(i)}}(\mathbf{x}^n, r), \quad (69)$$

approximation (16)-(29) can be used with some modifications: dependence on $\mathbf{x}_k^{n,(i)}$ and \mathbf{x}^n are respectively replaced by $(\mathbf{x}_k^{n,(i)}, r_k^{(i)})$ and by (\mathbf{x}^n, r) ; the pdf $\mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{1,n,(i)}; \mathbf{P}_{k+1|k}^{1,n,(i)})$ is replaced by $p(r|r_k^{(i)}) \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{1,n,(i)}(r); \mathbf{P}_{k+1|k}^{1,n,(i)}(r))$. So the bootstrap and the APF implementations can be used to propagate samples $\{\mathbf{x}_k^{n,(i)}, r_k^{(i)}\}$:

- In the bootstrap implementation of Sections III-C1 and III-C2, samples $(\mathbf{x}_{k+1}^{n,(i)}, r_{k+1}^{(i)})$ are now obtained from $p(r|r_k^{(i)}) \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{1,n,(i)}(r); \mathbf{P}_{k+1|k}^{1,n,(i)}(r))$ for persistent targets and from $p_{i,k+1}(r) p_{i,k+1}(\mathbf{x}^n|r)$ for birth targets.
- In the APF implementation, samples are obtained from an importance density $q(\mathbf{z}, i, (\mathbf{x}^n, r))$ which extends that developed for non-maneuvring targets in Section III-D. For $r = 1, \dots, d$, we assume the availability of the approximation

$$\mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i)}(r); \mathbf{P}_{k+1|k}^{n,(i)}(r)) q_{k+1}^{(i)}(\mathbf{z}, \mathbf{x}^n) \approx p^{(i)}(\mathbf{x}^n|r, \mathbf{z}) p^{(i)}(\mathbf{z}|r) \quad (70)$$

where $\mathbf{m}_{k+1|k}^{n,(i)}(r)$ and $\mathbf{P}_{k+1|k}^{n,(i)}(r)$ are the predicted mean and covariance matrix under mode r , as shown in (20) and (21). The approximation (70) can be found using progressive correction as described in the appendix. We define

$$\tilde{q}(\mathbf{z}|i) = \begin{cases} 1 - p_{d,k+1}, & \mathbf{z} = \theta, \\ p_{d,k+1} / [\kappa(\mathbf{z}) + \hat{B}(\mathbf{z})] \sum_{r=1}^d p^{(i)}(\mathbf{z}|r), & \mathbf{z} \in Z_{k+1}, \end{cases} \quad (71)$$

The sampling densities for the component index and for the measurements are given by $q(i)$ and $q(\mathbf{z}|i)$ defined in (55) and (57) except that we use (71) for the definition of $\tilde{q}(\mathbf{z}|i)$. The mode is sampled from

$$q(r|\mathbf{z}, i) \propto \begin{cases} p(r|r_{k-1}^{(i)}), & \mathbf{z} = \theta, \\ p(r|r_{k-1}^{(i)}) p^{(i)}(\mathbf{z}|r), & \mathbf{z} \in Z_{k+1}. \end{cases} \quad (72)$$

Finally, the state is sampled from

$$q(\mathbf{x}^n|i, \mathbf{z}) = \begin{cases} \mathcal{N}(\mathbf{x}^n; \mathbf{m}_{k+1|k}^{n,(i)}(r); \mathbf{P}_{k+1|k}^{n,(i)}(r)), & \mathbf{z} = \theta, \\ p^{(i)}(\mathbf{x}^n|r, \mathbf{z}), & \mathbf{z} \in Z_{k+1}. \end{cases} \quad (73)$$

The sample weights are calculated as in (59) with obvious adjustments for the mode.

V. SIMULATIONS

Four algorithms are considered in the performance analysis. The algorithms are as follows:

- 1) Bootstrap filter (BF). The BF is implemented as in [12]. In this implementation the number of particles used for persisting targets depends on the estimated number of existing targets. In particular, if the total sample size is L_k and \hat{n}_{k-1} is the estimated number of targets at time $k-1$, then $L_k^1 = \min\{L_k, \hat{n}_{k-1} \min\{500, L_k/20\}\}$ particles are used for persisting targets and $L_k^2 = L_k - L_k^1$ particles are used for birth targets. This ensures that the total number of likelihood evaluations for each update is n and allows comparison with the algorithms which use a fixed sample size.
- 2) Marginal bootstrap filter (M-BF). This is a marginalized bootstrap implementation developed in Sections III-C and IV-B.
- 3) APF PHD filter. This is the algorithm of [16] with some slight differences. First, birth targets are drawn from a Gaussian mixture rather than a single Gaussian. Second, progressive correction, rather than the less accurate UT, is used to approximate the optimal importance distribution. Third, the algorithm is extended to include sampling of a manoeuvring mode for the manoeuvring target example of Section V-C.
- 4) Marginalized APF (M-APF) implementation. This is the implementation proposed in Sections III-D and IV-B. Our approximation of $\hat{B}(z)$ relies on (61) with $L' = 100$.

Both APF implementations use a fixed sample size at all times.

A. The Optimal Subpattern Assignment (OSPA) metric

The algorithms obtain target state estimates by applying the k -means clustering algorithm to the collection of particles. The number of clusters sought by k -means is the estimated number of targets, i.e., the sum of the sample weights rounded to the nearest integer. Errors arise in both the individual target state estimates and in the estimated number of targets. These errors can be measured by the OSPA metric, derived in [34]. Let $X = \{x_1, \dots, x_m\}$ and $Y = \{y_1, \dots, y_n\}$ be two finite sets. Here, X represents the estimated finite set of targets and Y represents the true finite set of targets. For $1 \leq p < +\infty$ and $c > 0$, we denote $d^{(c)}(x, y) = \min(c, \|x - y\|)$ ($\|\cdot\|$ is the Euclidean norm) and Π_n the set of permutations on $\{1, 2, \dots, n\}$. The OSPA metric is defined by:

$$\bar{d}_p^c(X, Y) \triangleq \left(\frac{1}{n} \left(\min_{\pi \in \Pi_n} \sum_{i=1}^m d^{(c)}(x_i, y_{\pi(i)})^p + c^p(n-m) \right) \right)^{1/p}$$

if $m \leq n$ and $\bar{d}_p^c(X, Y) \triangleq \bar{d}_p^c(Y, X)$ if $m > n$. The parameter c determines the weighting of the penalties assigned to cardinality and localization errors and p determines the sensitivity of the localization error and cardinality error. In our simulations, we take $c = 100$ and $p = 2$.

B. Range bearing tracking of constant velocity (CV) targets

In this scenario targets move with CV in the region $[-500, 500]^2$. The model used in the PHD filters is as follows. The state vector is $\mathbf{x}_k = [x_k, y_k, \dot{x}_k, \dot{y}_k]'$ where (x_k, y_k) is target position in Cartesian coordinates and the dot notation indicates differentiation with respect to time. The transition pdf at time k due to a target with state \mathbf{x}_{k-1} at time $k-1$ is

$$f(\mathbf{x}_k | \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}\mathbf{x}_{k-1}, \mathbf{Q}), \quad (74)$$

$$\mathbf{F} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \otimes \mathbf{I}_2, \quad \mathbf{Q} = \sigma \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix} \otimes \mathbf{I}_2, \quad (75)$$

where \otimes is the Kronecker product. The particular parameter values are $p_{s,k} = 0.95$, $T = 1$ and $\sigma = 1/625$. The birth intensity is

$$\gamma(\mathbf{x}_k) = \lambda_b / c^2 \sum_{b=1}^{c^2} \mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_b, \mathbf{\Omega}_b), \quad (76)$$

where λ_b is the mean arrival rate of new targets, $\mathbf{\Omega}_b = \text{diag}(3086, 3086, 36, 36)$ for $b = 1, \dots, c^2$ and, for $i, j = 1, \dots, c$,

$$\bar{\mathbf{x}}_{i+(j-1)c} = \begin{bmatrix} -500 + 1000(i-1/2)/(c-1) \\ -500 + 1000(j-1/2)/(c-1) \\ 0 \\ 0 \end{bmatrix} \quad (77)$$

Note that the means form a regular $c \times c$ grid over $[-500, 500]^2$ so that the birth intensity covers the region in which the targets move. In the implementation $\lambda_b = 1/4$ and $c = 6$.

The targets are observed by a sensor placed at the origin. The sensor produces measurements of position in polar coordinates. The measurement function is then $\mathbf{h}(\mathbf{x}) = [\sqrt{x^2 + y^2}, \arctan(y/x)]'$. The measurement noise has covariance matrix $\mathbf{R} = \text{diag}(16, (\pi/180)^2)$. Targets are detected with probability $p_{d,k} = 0.9$. Clutter is uniformly distributed with density λ_c . Simulations are performed for clutter densities of $\lambda_c = 10^{-5}$ and $\lambda_c = 4 \times 10^{-5}$. The former clutter density produces an average of 10 clutter points in the surveillance volume at each scan while the latter produces 40 clutter points. The dynamical and measurement models satisfy (9) and (10) with $\mathbf{x}_k^n = [x_k, y_k]'$ and $\mathbf{x}_k^l = [\dot{x}_k, \dot{y}_k]'$.

The scenario considered here is shown in Figure 1. In this scenario the number of targets varies as shown in Figure 1(a). The target trajectories are shown in Figure 1(b). In this figure trajectories are plotted in red at the start of the surveillance interval and gradually transform to blue at the end of the surveillance interval. This colour coding allows times at which targets are in proximity to be identified.

The algorithms are implemented with sample sizes between 100 and 10 000. Algorithm performance is measured using the mean OSPA distance averaged over 200 realisations.

The time-averaged mean OSPA for each algorithm is plotted against sample size in Figure 2. The results show that marginalization produces significant benefits, particularly when combined with measurement-directed sampling. These benefits are evident for both clutter densities. It should be kept in mind that in our implementations, for a given sample size, the APF and M-APF have roughly three times the

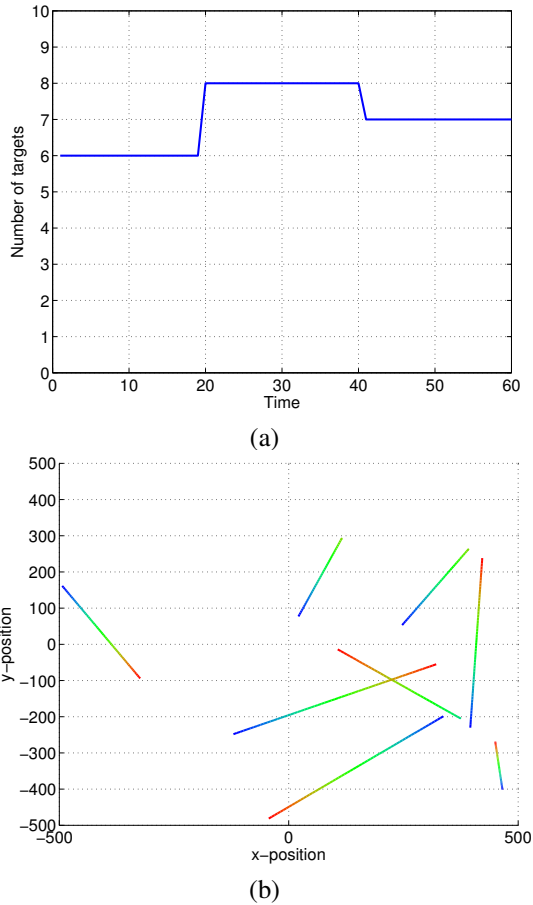


Fig. 1. Simulation scenario for CV targets: (a) Number of existing targets plotted against time. (b) Target trajectories.

computational expense of the BF and M-BF. This increased expense is clearly worthwhile as the BF and M-BF would require far more than three times the sample size to produce similar performance to the APF and M-APF.

C. Range bearing tracking of manoeuvring targets

We now consider a scenario in which targets perform coordinated turn manoeuvres. The target state is $\mathbf{x}_k = [x_k, y_k, \dot{x}_k, \dot{y}_k, \omega_k]'$ where ω_k is the turn rate. Target motion is modelled by a jump Markov model with two motion models. The first motion model ($r = 1$) is the CV model described in Section V-B and given by (74) with

$$\mathbf{F} = \text{diag} \left(\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \otimes \mathbf{I}_2, 0 \right), \quad (78)$$

$$\mathbf{Q} = \text{diag} \left(\sigma \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix} \otimes \mathbf{I}_2, 0 \right). \quad (79)$$

The second motion model ($r = 2$) is a coordinated turn model for which the transition matrix and process noise covariance matrix are

$$\mathbf{F}(\mathbf{x}) =$$

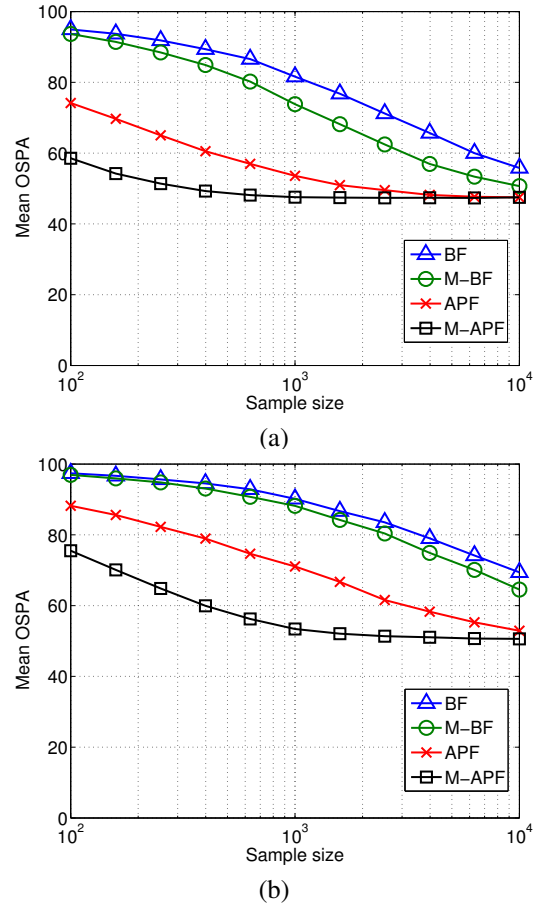


Fig. 2. Mean OSPA plotted against sample size for CV target tracking with the BF, M-BF, APF and M-APF. Results are shown for (a) $\lambda_c = 10^{-5}$ (b) $\lambda_c = 4 \times 10^{-5}$.

$$\text{diag} \left(\begin{bmatrix} 1 & 0 & \sin(\omega T)\omega & -[1 - \cos(\omega T)]/\omega \\ 0 & 1 & [1 - \cos(\omega T)]/\omega & \sin(\omega T)\omega \\ 0 & 0 & \cos(\omega T) & -\sin(\omega T) \\ 0 & 0 & \sin(\omega T) & \cos(\omega T) \end{bmatrix}, 1 \right), \quad (80)$$

$$\mathbf{Q} = \text{diag} \left(\sigma \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix} \otimes \mathbf{I}_2, \beta \right). \quad (81)$$

The coordinated turn motion model satisfies (66)-(67) with $\mathbf{x}_k^n = [x_k, y_k, \omega_k]'$ and $\mathbf{x}_k^l = [\dot{x}_k, \dot{y}_k]'$. The particular parameter values are $T = 1$, $\sigma = 1/625$ and $\beta = (5\pi/180)^2$. The persistence probability is $p_{s,k} = 0.95$. The transition probabilities for the discrete mode parameter are $p(r_k = 1|r_{k-1} = 1) = 0.93$ and $p(r_k = 2|r_{k-1} = 2) = 0.8$.

The birth intensity is as shown in (76) with the turn rate having all its mass concentrated at zero. The measurement model is also identical to the CV scenario.

Target trajectories are constructed by simulating the mode parameter using the same transition model assumed by the filter. When $r_k = 1$ CV motion is performed while a coordinated turn is performed if $r_k = 2$. The turn rate is held constant throughout a coordinated turn manoeuvre at a random value drawn from the mixture $[\mathcal{N}(5\pi/180, (2\pi/180)^2) + \mathcal{N}(-5\pi/180, (2\pi/180)^2)]/2$. Only randomly generated trajectories which remain in the region $[-500, 500]^2$ are accepted. The resulting scenario is shown in Figure 3. The number of

targets varies as shown in Figure 3(a) and the target trajectories are shown in Figure 3(b).

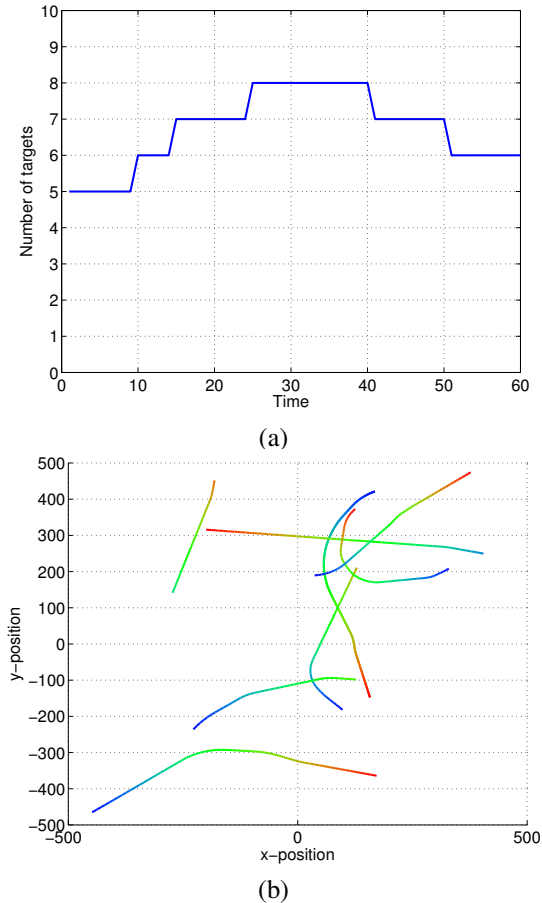


Fig. 3. Simulation scenario for manoeuvring targets: (a) Number of existing targets plotted against time. (b) Target trajectories.

The time-averaged mean OSPA for each algorithm is plotted against sample size in Figure 4. The results are similar in nature to those obtained for CV motion. Thus the same comments regarding the impact of marginalization and measurement-directed sampling apply for manoeuvring targets. As expected, the sample size required to reach an optimal level of performance, i.e., the sample size for which further increases in sample size produce no performance benefits, is larger for manoeuvring targets than for CV targets. Thus, for $\lambda = 10^{-5}$, about 4000 particles are required for the M-APF to achieve approximately optimal performance for the manoeuvring target scenario while roughly 1500 particles are required for CV target tracking. This increase in required sample size is probably also partly due to the increased variability in target number for the manoeuvring target scenario, as seen by comparing Figures 1(a) and 3(a). It may be expected that tracking the target number is more difficult if it changes frequently.

VI. CONCLUSION

In this paper, we proposed a family of M-PHD filters for multi-target filtering. We first focused on models which present both a linear and a non linear substructure. Two implementations of the M-PHD were developed: a bootstrap filter, the

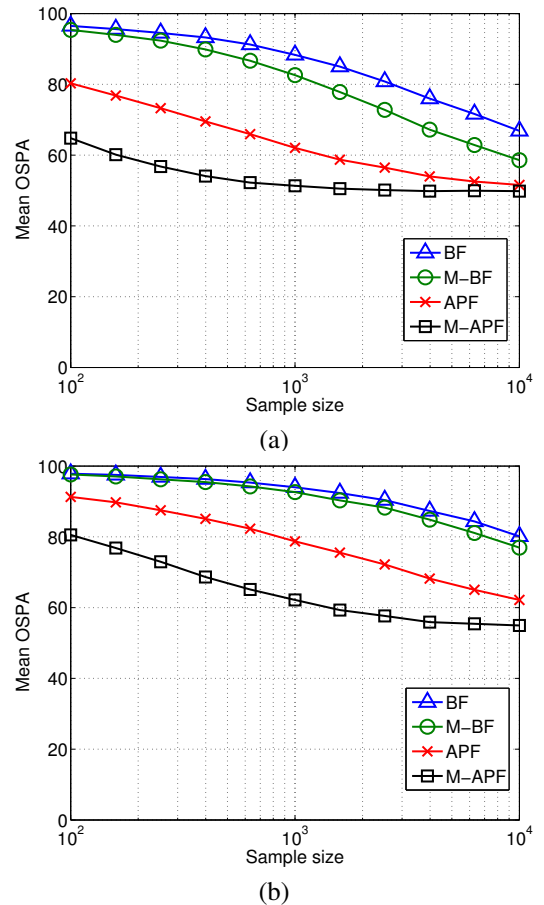


Fig. 4. Mean OSPA plotted against sample size for manoeuvring target tracking with the BF, M-BF, APF and M-APF. Results are shown for (a) $\lambda_c = 10^{-5}$ (b) $\lambda_c = 4 \times 10^{-5}$.

M-BF, which draws samples from the prior and an auxiliary particle filter, the M-APF, which uses a measurement-directed sampling density. It was then shown how these filters can be used with Gaussian JMSS with linear or partially linear substructures.

The M-BF and M-APF were applied to the problems of tracking constant velocity targets and manoeuvring targets using range-bearings measurements. The marginalised particle filters exhibited significant improvement compared to their non-marginalized counterparts, the BF and APF. Of the two M-PHD implementations, the M-APF provided better performance than the M-BF. Although the expense of drawing a sample is greater for the M-APF, it offers better performance for a given computational expense through the use of a much smaller sample size.

APPENDIX

A BRIEF REVIEW OF PROGRESSIVE CORRECTION

As seen in (52) and (53), the measurement-directed proposals used in the APF require approximation of quantities of the forms

$$\bar{\ell}(\mathbf{z}) = \int g(\mathbf{z}|\mathbf{x})\pi_0(\mathbf{x}) d\mathbf{x}, \quad (82)$$

$$\pi(\mathbf{x}) \propto g(\mathbf{z}|\mathbf{x})\pi_0(\mathbf{x}), \quad (83)$$

where $g(\cdot|\mathbf{x})$ is the conditional measurement PDF and $\pi_0(\cdot)$ is a prior PDF. We are interested in cases where (82) and (83) are not available in closed-form. The idea of PC is to define correction factors $\gamma_1, \dots, \gamma_s$ such that $0 < \gamma_i \leq 1$ and $\sum_{i=1}^s \gamma_i = 1$. For $i = 1, \dots, s$, let

$$\pi_i(\mathbf{x}) = g(\mathbf{z}|\mathbf{x})^{\Gamma_i} \pi_0(\mathbf{x}) / \bar{\ell}(\mathbf{z}), \quad (84)$$

$$\bar{\ell}_i(\mathbf{z}) = \int g(\mathbf{z}|\mathbf{x})^{\Gamma_i} \pi_0(\mathbf{x}) \, d\mathbf{x}, \quad (85)$$

where $\Gamma_i = \sum_{a=1}^i \gamma_a$. We then have, for $i \geq 1$,

$$\bar{\ell}_i(\mathbf{z}) = \int g(\mathbf{z}|\mathbf{x})^{\gamma_i} g(\mathbf{z}|\mathbf{x})^{\Gamma_{i-1}} \pi_0(\mathbf{x}) \, d\mathbf{x}, \quad (86)$$

$$= \bar{\ell}_{i-1}(\mathbf{z}) \int g(\mathbf{y}|\mathbf{x})^{\gamma_i} \pi_{i-1}(\mathbf{x}) \, d\mathbf{x}, \quad (87)$$

with $\bar{\ell}_0(\mathbf{z}) = 1$. It follows that

$$\bar{\ell}(\mathbf{z}) = \bar{\ell}_s(\mathbf{z}) = \prod_{i=1}^s \nu_i(\mathbf{z}), \quad (88)$$

$$\nu_i(\mathbf{z}) = \int g(\mathbf{z}|\mathbf{x})^{\gamma_i} \pi_{i-1}(\mathbf{x}) \, d\mathbf{x}. \quad (89)$$

The densities $\pi_i(\cdot)$ are found recursively for $i = 1, \dots, s$ using

$$\pi_i(\mathbf{x}) \propto g(\mathbf{z}|\mathbf{x})^{\gamma_i} \pi_{i-1}(\mathbf{x}). \quad (90)$$

Although (88) and (90) cannot be evaluated in closed-form they can be accurately approximated using standard techniques such as linearisation or the UT. This can be seen by considering (90) for $i = 1$:

$$\pi_1(\mathbf{x}) \propto g(\mathbf{z}|\mathbf{x})^{\gamma_1} \pi_0(\mathbf{x}). \quad (91)$$

Compared with the complete correction (83), the partial correction (91) uses a broad likelihood which ensures large overlap between the prior and the likelihood provided γ_1 is sufficiently small. This allows good approximations to $\pi_1(\cdot)$ and $\nu_1(\mathbf{z})$ to be obtained using conventional methods such as linearisation or the UT. The process can then be continued for $\pi_2(\cdot)$, $\nu_2(\mathbf{z})$ and so on.

REFERENCES

- [1] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.
- [2] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. New York: Springer Verlag, 2001.
- [3] P. Fearnhead, "Computational methods for complex stochastic systems: A review of some alternatives to MCMC," *Statistics and Computing*, vol. 18, no. 2, pp. 151–171, 2008.
- [4] R. Chen and J. S. Liu, "Mixture Kalman Filters," *J. R. Statist. Soc. B*, vol. 62, pp. 493–508, 2000.
- [5] T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear nonlinear state-space models," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2279–2289, 2005.
- [6] Y. Bar-Shalom, *Tracking and data association*. San Diego, CA, USA: Academic Press Professional, Inc., 1987.
- [7] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 2009.
- [8] J. Goutsias, R. P. Mahler, and H. T. Nguyen, *Random Sets Theory and Applications*. Springer-Verlag New York, 1997.
- [9] I. R. Goodman, R. P. Mahler, and H. T. Nguyen, *Mathematics of Data Fusion*. Norwell, MA, USA: Kluwer Academic Publishers, 1997.
- [10] R. Mahler, "Multitarget Bayes Filtering via First-Order Multitarget Moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, October 2003.
- [11] B.-N. Vo and W.-K. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," *IEEE transactions Signal Processing*, vol. 54, pp. 4091–4104, November 2006.
- [12] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with Random Finite Sets," *IEEE Trans. Aerospace and Electronic Systems*, vol. 41, 2005.
- [13] T. Zajic and R. Mahler, "A particle-systems implementation of the PHD multitarget tracking filter," in *Proceedings of SPIE*, vol. 5096, 2003.
- [14] S. Särkkä, A. Vehtari, and J. Lampinen, "Rao-blackwellized particle filter for multiple target tracking," *Information Fusion Journal*, vol. 8, 2007.
- [15] M. Vihola, "Rao-blackwellised particle filtering in random set multitarget tracking," *IEEE Transactions on Aerospace and Electronic Systems*, 2006.
- [16] N. Whiteley, S. Singh, and S. Godsill, "Auxiliary particle implementation of the probability hypothesis density filter," *IEEE Transactions on Aerospace and Electronic Systems*, July 2010.
- [17] R. Mahler, *Statistical Multisource Multitarget Information Fusion*. Artech House, 2007.
- [18] A. Pasha, B.-N. Vo, H. Tuan, and W.-K. Ma, "A Gaussian Mixture PHD Filter for Jump Markov System Models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 919–936, 2009.
- [19] K. Punithakumar, T. Kirubarajan, and A. Sinha, "Multiple-Model Probability Hypothesis Density Filter for Tracking Maneuvering Targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 87–98, 2008.
- [20] M. Morelande, "A sequential Monte Carlo method for PHD approximation with conditionally linear/Gaussian models," in *Proceedings of the International Conference on Information Fusion*, Edinburgh, Great Britain, 2010.
- [21] Y. Petetin and F. Desbouvries, "Marginalized PHD filters for Multi-Target Filtering," in *International Conference on Information Sciences, Signal Processing and their Applications*, Montreal, Canada, 2012.
- [22] D. Clark, B.-T. Vo, B.-N. Vo, and S. Godsill, "Gaussian Mixture implementations of Probability Hypothesis Density Filters for non-linear dynamical models," in *IET seminar on Target Tracking and Data Fusion*, April 2008, pp. 21–28.
- [23] D. Clark and J. Bell, "Convergence results for the Particle-PHD Filter," in *IEEE Trans. Signal Processing*, vol. 54, 2006.
- [24] D. Clark, B.-T. Vo, and B.-N. Vo, "Gaussian Particle Implementations of Probability Hypothesis Density Filters," in *IEEE Aerospace Conference*, 2007.
- [25] A. Doucet, N. J. Gordon, and V. Krishnamurthy, "Particle Filters for State Estimation of Jump Markov Linear Systems," *IEEE Transactions on Signal Processing*, vol. 49, pp. 613–624, 1999.
- [26] F. Lindsten, T. Schön, and J. Olsson, "An explicit variance reduction expression for the Rao-Blackwellized particle filter," in *18th World Congress of the International Federation of Automatic Control (IFAC)*, August 2011.
- [27] M. K. Pitt and N. Shephard, "Filtering via simulation : Auxiliary particle filter," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 550–599, June 1999.
- [28] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputation," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 567–76, June 1995.
- [29] Y. Petetin and F. Desbouvries, "Optimal SIR algorithm vs. fully adapted auxiliary particle filter: a non asymptotic analysis," *Accepted for publication, Statistics and computing*, July 2012.
- [30] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.
- [31] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," in *Proceedings of the IEEE*, vol. 92, March 2004, pp. 401–422.
- [32] N. Oudjane and C. Musso, "Progressive correction for regularized particle filters," in *Proceedings of the International Conference on Information Fusion*, 2000.
- [33] M. Morelande and A. Skvortsov, "Radiation field estimation using a Gaussian mixture," in *Proceedings of the International Conference on Information Fusion*, Seattle, USA, 2009.
- [34] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, August 2008.