Bayesian Multi-object Filtering for Pairwise Markov Chains

Yohan Petetin*, François Desbouvries, *Senior Member, IEEE*

Abstract—**Random Finite Sets (RFS) are recent tools for addressing the multi-object filtering problem. The Probability Hypothesis Density (PHD) Filter is an approximation of the multi-object Bayesian filter which results from the RFS formulation of the problem and has been used in many applications. In the RFS framework, it is assumed that each target and associated observation follow a Hidden Markov Chain (HMC) model. HMCs conveniently describe some physical properties of practical interest for practitioners, but they also implicitly imply restrictive independence properties which, in practice, may not be satisfied by data. In this paper, we show that these structural limitations of HMC models can somehow be relaxed by embedding them into the more general class of Pairwise Markov Chain (PMC) models. We thus focus on the computation of the PHD filter in a PMC framework, and we propose a practical implementation of the PHD filter for a particular class of PMC models.**

Index Terms—**Random Finite Sets, Hidden Markov Chains, Pairwise Markov Chains, Multi-object filtering, Probability Hypothesis Density.**

I. Introduction

THE 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 [1] [2] and the Multiple Hypothesis Tracking (MHT) filter [3] [4], which both include a matching mechanism that aims at optimizing the association between observations and targets. An alternative class of solutions, based on RFS, i.e. of sets of random variables (r.v.) with random cardinal, has also been proposed [5] [6] [7]. RFS-based solutions no longer try to associate a target to a given observation, and so avoid the intensive computational cost of the association mechanism required in the classical solutions. Among RFS based solutions, the PHD filter propagates the first order moment of the posterior multi-target density (mtd), which is a real and positive function associated to a RFS that generalizes the probability density function (pdf) [7] [8]. This technique has now become popular and many multi-object filtering problems have been addressed with a PHD approach, see e.g. [9] [10]. So from now, we focus on RFS and PHD based solutions.

Following these pioneering works, a number of contributions which aim at improving the performances of PHD based solutions have been proposed, mainly in two directions.

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. We would like to thank the French MOD DGA/MRIS for financial support of the Ph.D. of Y. Petetin.

On the one hand the PHD filter cannot be computed directly, so some works have focused on implementation issues. Implementations based on numerical approximations, such as the Gaussian Mixture (GM) PHD filter [11] [12] [13], or on Sequential Monte-Carlo (SMC) methods [14] [15] [16] have been proposed and studied theoretically.

On the other hand, other contributions try to relax some of the assumptions underlying the derivation of the PHD filter in order to derive more realistic algorithms. For example, the PHD filter was originally derived under Poisson hypothesis for the predicted-target and clutter processes. Mahler relaxed these hypotheses and derived, at the price of an increasing complexity, the so-called Cardinalized PHD (CPHD) filter, which in addition to propagating the PHD, also propagates the cardinality distribution of the number of targets [17]; later on, practical implementations of the CPHD filter were further derived [18]. Other improvements of the PHD filter have also been developed, see e.g. [19].

In this paper, we focus on the computation of the PHD filter when the usual independence assumptions followed by each target are violated. More precisely, most multi-object filters, including the PHD filter, rely on the assumption that the targets and the observations they produce follow the well known HMC model which is also widely used in single object filtering, see e.g. [20] [21]. In the HMC model, it is assumed that for a given target, the states are a Markov Chain (MC); and that given all the states, the observations are independent and the observation at time k only depends on the state at time k. These models are of practical interest because it is possible to derive filters of interest (Single-object Bayes filter, Multi-object Bayes Filter, PHD filter). Such filters actually rely on the given transition pdf of the MC, and on the given likelihood pdf of a measurement with a given state which characterizes how the measurements are produced from the states. These pdfs are chosen according to the physical tracking problem at hand, and to the sensors which are used to track the targets. However, the Markovian and independence assumptions which are implicit in HMC modeling may not be satisfied in practice, i.e., the data do not necessarily fit the strong assumptions underlying these models [22].

Our motivation in this paper is thus to propagate the PHD, while relaxing some of the assumptions which underly the classical HMC modeling. To that end, we resort to PMC models [23] [24] which are more general than HMC ones. So we first discuss on the statistical properties of PMC as compared to HMC, and next compute a PMC PHD filter adapted to an underlying PMC model. Since any HMC is also a particular PMC, our class of PMC PHD filters encompasses the classical HMC based PHD solutions, which are a particular case. We finally consider several practical aspects of PMC models, and in particular show that it is possible to build PMC models which preserve HMC physical properties of interest.

The rest of this paper is organized as follows. In section II we first motivate our contribution by recalling the main limitations of HMC models and the advantage of PMC ones, and illustrate them in the classical linear and Gaussian model framework. In section III we recall the RFS approach and the HMC based PHD filter, and next derive a PMC based PHD filter, as well as a GM implementation for linear and Gaussian PMC. In section IV we discuss on the practical aspects of PMC models and we validate our solution via simulations. We finally end the paper with a Conclusion.

II. Motivation

Among the assumptions underlying RFS based solutions and, in particular, the PHD filter [6, Ch. 6], it is assumed that each target and the associated observation follow an HMC model (other assumptions will be recalled later since they are not decisive in this section). Here we recall the statistical properties of HMC models and we show their main limitations.

A. HMC models

Let a state $\mathbf{x}_k \in \mathbb{R}^m$ and its associated observation $\mathbf{y}_k \in \mathbb{R}^q$ (as far as notations are concerned, we do not differ r.v. and their realizations). Then $(\mathbf{x}_k, \mathbf{y}_k)_{k \geq 0}$ is an HMC if the joint pdf of $(\mathbf{x}_{0:k}, \mathbf{y}_{0:k}) = (\mathbf{x}_0, \cdots, \mathbf{x}_k, \mathbf{y}_0, \cdots, \mathbf{y}_k)$ can be factorized as follows:

$$p(\mathbf{x}_{0:k}, \mathbf{y}_{0:k}) = \underbrace{p(\mathbf{x}_0) \prod_{i=1}^{k} f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1})}_{p(\mathbf{x}_{0:k})} \underbrace{\prod_{i=0}^{k} g_i(\mathbf{y}_i|\mathbf{x}_i)}_{p(\mathbf{y}_{0:k}|\mathbf{x}_{0:k})}, \quad (1)$$

where $f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1})$ is the transition of state $\mathbf{x}_i$ given $\mathbf{x}_{i-1}$ at time i, and $g_i(\mathbf{y}_i|\mathbf{x}_i)$ is the likelihood pdf of a measurement $\mathbf{y}_i$ given state $\mathbf{x}_i$ at time i. Let $\mathcal{N}(\mathbf{x}; \mathbf{m}; \mathbf{P})$ denote the Gaussian pdf with mean $\mathbf{m}$ and covariance matrix $\mathbf{P}$ taken at point $\mathbf{x}$. A classical example of HMC models is the linear and Gaussian model:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{u}_k, \quad (2)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \quad (3)$$

where $\mathbf{F}_k$ is an $m \times m$ matrix, $\mathbf{H}_k$ is a $q \times m$ matrix, $\mathbf{x}_0$, $\mathbf{u}_1$, $\cdots$, $\mathbf{u}_k$, $\mathbf{v}_0$, $\cdots$, $\mathbf{v}_k$ are independent, and $\mathbf{x}_0 \sim \mathcal{N}(.; \mathbf{m}_0; \mathbf{P}_0)$, $\mathbf{u}_k \sim \mathcal{N}(.; \mathbf{0}; \mathbf{Q}_k)$ and $\mathbf{v}_k \sim \mathcal{N}(.; \mathbf{0}; \mathbf{R}_k)$. We assume that $\mathbf{Q}_k$ and $\mathbf{R}_k$ are positive definite for all k. The hypotheses on noises $\{\mathbf{u}_k\}_{k \geq 1}$ and $\{\mathbf{v}_k\}_{k \geq 0}$ imply that $(\mathbf{x}_k, \mathbf{y}_k)$ is an HMC where $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}_k \mathbf{x}_{k-1}; \mathbf{Q}_k)$ and $g_k(\mathbf{y}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \mathbf{H}_k \mathbf{x}_k; \mathbf{R}_k)$. Model (2)-(3) can be written under a block matrix form (the usefulness of this reformulation will appear clearer later); plugging (2) in (3), and setting $\mathbf{w}_k = [\mathbf{u}_k^T, (\mathbf{H}_k \mathbf{u}_k + \mathbf{v}_k)^T]^T$, (2)-(3) read

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{F}_k & \mathbf{0} \\ \mathbf{H}_k \mathbf{F}_k & \mathbf{0} \end{bmatrix}}_{\mathbf{B}_k} \begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{y}_{k-1} \end{bmatrix} + \mathbf{w}_k, \quad (4)$$

where $\mathbf{w}_1, \cdots, \mathbf{w}_k$ are independent zero-mean Gaussian noises, independent of $(\mathbf{x}_0, \mathbf{y}_0)$, with

$$\mathrm{E}(\mathbf{w}_k \mathbf{w}_k^T) = \begin{bmatrix} \mathbf{Q}_k & (\mathbf{H}_k \mathbf{Q}_k)^T \\ \mathbf{H}_k \mathbf{Q}_k & \mathbf{R}_k + \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^T \end{bmatrix}. \quad (5)$$

Most of single- and multi-target filters have been developed for HMC models (1), and in particular for linear and Gaussian HMC (2)-(3). These models are popular for both computational and physical modeling reasons, as we now recall.

On the one hand, model (1) leads to the recursive computation of pdfs of interest such as the filtering pdf $p(\mathbf{x}_k|\mathbf{y}_{0:k})$, used in single target filters, via the Kalman filter or via numerical and stochastic approximations [25] [21] [26].

On the other hand, HMCs enable practitioners to turn a real problem into a probabilistic model. More precisely, pdf $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ describes the dynamical evolution of the state over time, whereas pdf $g_k(\mathbf{y}_k|\mathbf{x}_k)$ measures the likelihood of observation $\mathbf{y}_k$ with state $\mathbf{x}_k$ and so characterizes the sensor used to detect the targets.

B. Statistical limitations of HMC models

Although HMC models are popular, factorization (1) implies some constraints among the different variables. From a global point of view, HMC models are characterized by the two following properties:

D.1 $\{\mathbf{x}_k\}_{k \geq 0}$ is an MC with transitions $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$;

D.2 Given states $\mathbf{x}_{0:k}$, observations $\mathbf{y}_{0:k}$ are independent, and observation $\mathbf{y}_i$ depends only on state $\mathbf{x}_i$.

Next, from a local point of view, (1) leads to two other properties, which this time are not sufficient to characterize an HMC but which, as we shall see in section IV, can be useful to identify models which are not HMC:

P.1 given $\mathbf{x}_{k-1}$ and $\mathbf{y}_{k-1}$, $\mathbf{x}_k$ depends on $\mathbf{x}_{k-1}$ only, i.e. $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) = f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ (note that **P.1** is a direct consequence of **D.2**);

P.2 given $\mathbf{x}_k$, $\mathbf{x}_{k-1}$ and $\mathbf{y}_{k-1}$, $\mathbf{y}_k$ depends on $\mathbf{x}_k$ only, i.e. $p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_{k-1}) = g_k(\mathbf{y}_k|\mathbf{x}_k)$.

In other words, observation $\mathbf{y}_{k-1}$ at time $k-1$ does not affect the pdf of state $\mathbf{x}_k$ when $\mathbf{x}_{k-1}$ is known, and the law of observation $\mathbf{y}_k$ at time k only depends on state $\mathbf{x}_k$ when $\mathbf{x}_k$, $\mathbf{x}_{k-1}$ and $\mathbf{y}_{k-1}$ are also known.

C. PMC models

We start with the following observation. Let $(\mathbf{x}_k, \mathbf{y}_k)$ be an HMC. In addition to **P.1** and **P.2**, one can check that $(\mathbf{x}_k, \mathbf{y}_k)$ satisfies the following property:

P.3 couple $(\mathbf{x}_k, \mathbf{y}_k)$ is an MC, i.e. $p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{x}_{0:k-1}, \mathbf{y}_{0:k-1}) = p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{x}_{k-1}, \mathbf{y}_{k-1})$.

Next, a model in which **P.3** holds is called a PMC model. So $\boldsymbol{\xi}_k \triangleq (\mathbf{x}_k, \mathbf{y}_k)$ is a PMC if and only if the joint pdf of $(\mathbf{x}_{0:k}, \mathbf{y}_{0:k})$ factorizes as follows:

$$\underbrace{p(\mathbf{x}_{0:k}, \mathbf{y}_{0:k})}_{p(\boldsymbol{\xi}_{0:k})} = \underbrace{p(\mathbf{x}_0, \mathbf{y}_0)}_{p(\boldsymbol{\xi}_0)} \prod_{i=1}^{k} \underbrace{p_{i|i-1}(\mathbf{x}_i, \mathbf{y}_i|\mathbf{x}_{i-1}, \mathbf{y}_{i-1})}_{p_{i|i-1}(\boldsymbol{\xi}_i|\boldsymbol{\xi}_{i-1})}. \quad (6)$$

In other words, in a PMC model only the *pair* $\{\mathbf{x}_k, \mathbf{y}_k\}_{k\geq 0} = \{\boldsymbol{\xi}_k\}_{k\geq 0}$ is assumed to be an MC. Note that an HMC model is a PMC model in which, in addition, **P.1** and **P.2** also hold; in this particular case, $p(\mathbf{y}_0|\mathbf{x}_0)$ and transitions $p_{i|i-1}$ reduce to

$$p(\mathbf{y}_0|\mathbf{x}_0) \overset{\text{HMC}}{=} g_0(\mathbf{y}_0|\mathbf{x}_0), \quad (7)$$

$$p_{i|i-1}(\mathbf{x}_i, \mathbf{y}_i|\mathbf{x}_{i-1}, \mathbf{y}_{i-1}) \overset{\text{HMC}}{=} f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1})g_i(\mathbf{y}_i|\mathbf{x}_i). \quad (8)$$

Particularly interesting, in a PMC, $\{\mathbf{x}_k\}_{k\geq 0}$ is not necessarily an MC [27], which is an hypothesis used in most filtering algorithms; and given $(\mathbf{x}_{0:k}, \mathbf{y}_{0:i-1})$ with $0 \leq i \leq k$, the law of $\mathbf{y}_i$ does not necessarily depend on $\mathbf{x}_i$ only, but also on $\mathbf{x}_{i-1}$, $\mathbf{y}_{i-1}$ and $\mathbf{x}_{i+1:k}$. Note that in a stationary reversible PMC, $\{\mathbf{x}_k\}_{k\geq 0}$ is an MC if and only if $p(\mathbf{y}_i|\mathbf{x}_{0:k}) = p(\mathbf{y}_i|\mathbf{x}_i)$ [28]. So more complex scenarios can be considered via PMC models (as we will see e.g. in section IV), including scenarios where $\{\mathbf{x}_k\}_{k\geq 0}$ is not Markovian.

As an illustrative example, the classical state-space model (4)-(5) is a particular case of

$$\begin{bmatrix}\mathbf{x}_k\\ \mathbf{y}_k\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{F}_k^1 & \mathbf{F}_k^2\\ \mathbf{H}_k^1 & \mathbf{H}_k^2\end{bmatrix}}_{\mathbf{B}_k} \begin{bmatrix}\mathbf{x}_{k-1}\\ \mathbf{y}_{k-1}\end{bmatrix} + \mathbf{w}_k, \quad (9)$$

where $\mathbf{w}_1, \cdots, \mathbf{w}_k$ are independent zero-mean Gaussian noises, independent of $(\mathbf{x}_0, \mathbf{y}_0) \sim \mathcal{N}(.; \mathbf{m}_0; \mathbf{P}_0)$, with

$$\mathrm{E}(\mathbf{w}_k\mathbf{w}_k^T) = \boldsymbol{\Sigma}_k = \begin{bmatrix}\boldsymbol{\Sigma}_k^{11} & {\boldsymbol{\Sigma}_k^{21}}^T\\ \boldsymbol{\Sigma}_k^{21} & \boldsymbol{\Sigma}_k^{22}\end{bmatrix}. \quad (10)$$

The transitions of this PMC model are given by

$$p_{k|k-1}(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) = \mathcal{N}(\boldsymbol{\xi}_k; \mathbf{B}_k\boldsymbol{\xi}_{k-1}; \boldsymbol{\Sigma}_k). \quad (11)$$

From a computational point of view PMC models can also be used in practice because as in HMC models, the filtering pdf $p(\mathbf{x}_k|\mathbf{y}_{0:k})$ can be computed recursively via

$$p(\mathbf{x}_k|\mathbf{y}_{0:k}) \propto \int p_{k|k-1}(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{0:k-1})\mathrm{d}\mathbf{x}_{k-1}, \quad (12)$$

which reduces to the classical Bayesian filter

$$p(\mathbf{x}_k|\mathbf{y}_{0:k}) \propto g_k(\mathbf{y}_k|\mathbf{x}_k) \int f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{0:k-1})\mathrm{d}\mathbf{x}_{k-1}$$

when (7)-(8) are satisfied. Note that in (12), the new state $\mathbf{x}_k$ is no longer introduced via $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ but via $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_{k-1})$, which depends on $\mathbf{y}_{k-1}$; also, the new measurement $\mathbf{y}_k$ is injected via pdf $p(\mathbf{y}_k|\mathbf{x}_{k-1}, \mathbf{x}_k, \mathbf{y}_{k-1})$ which depends on the former measurement $\mathbf{y}_{k-1}$, contrary to the HMC Bayesian filter where the new measurement is injected via $g_k(\mathbf{y}_k|\mathbf{x}_k)$. Finally, (12) can either be computed exactly [29, eqs. (13.56) and (13.57)] [30] (in the linear and Gaussian case) or approximated via Monte Carlo approximations [27].

III. Multi-object Filters for dynamical PMC models

In this section we shall develop multi-object filters based on RFS when the targets and the associated observations can be modeled by a PMC model, and we propose a GM implementation for the particular linear and Gaussian case (9)-(10).

A. A review of the PHD filter based on RFS

1) The classical RFS approach: RFS are well suited to the multi-object filtering problem, since the number of targets and measurements can indeed be considered as random and time varying. More precisely, the RFS approach considers that at a given time k the targets and measurements are two RFS:

$$X_k = \{\mathbf{x}_{1,k}, \cdots, \mathbf{x}_{n',k}\},$$
$$Z_k = \{\mathbf{z}_{1,k}, \cdots, \mathbf{z}_{m',k}\},$$

where n' and m' are random integers. Remember that we do not differ r.v. and their realizations, and so we do not differ RFS and their realizations either. The measurement associated to a given state $\mathbf{x}$ is still noted $\mathbf{y}$, as in section II; in addition, notation $\mathbf{z}$ is used when we do not know if $\mathbf{z}$ is a measurement associated to a given target or is a false alarm measurement, and indeed RFS Z_k contains measurements due to detected targets but also to false alarms. If a given target $\mathbf{x}_{k,i}$ is not detected, then there is no measurement $\mathbf{z}$ associated to $\mathbf{x}_{k,i}$ in Z_k.

In that framework, the multi-target filtering problem consists in computing the mtd $p(X_k|Z_{0:k})$ of RFS X_k from the past measurements up to time k, i.e. from $Z_{0:k} = (Z_0, \cdots, Z_k)$. However, this filter is difficult to compute, above all when the number of targets is large, because it involves the computation of set integrals [8]. We now turn to the PHD filter.

2) The PHD filter: The so-called PHD (or intensity) $v(\mathbf{x})$ of RFS X is defined as the first order moment of mtd $p(X)$ [7]. This function satisfies the following property: let $S \in \mathbb{R}^m$ be a region and let $|X \cap S|$ be the number of objects of RFS X which belong to region S (remember that it is an r.v.); then $v(\mathbf{x})$ satisfies

$$\int_S v(\mathbf{x})\ \mathrm{d}\mathbf{x} = \mathrm{E}(|X \cap S|). \quad (13)$$

In other words, $v(\mathbf{x})$ is the spatial density of the expected number of targets. Using PHDs has several advantages. First, $v(\mathbf{x})$ enables both to compute the expected number of objects in X, and to estimate the states of X by looking for local maxima of $v(\mathbf{x})$. Next $v(\mathbf{x})$ is a function of $\mathbf{x}$, and not of RFS X, which becomes of practical interest when we turn to propagation issues. In this context, the approximated multi-object filtering problem consists in propagating PHD $v_k(\mathbf{x})$ of mtd $p(X_k|Z_{0:k})$ over time k. Let us assume that:

A.1 each target evolves and generates observations independently of one another according to an HMC model [6];
A.2 the clutter process is Poisson and clutter measurements are independent of target-originated measurements;
A.3 the predicted-target process, which governs the birth and surviving targets, is Poisson (Poisson hypothesis can be relaxed, which leads to the CPHD filter, where we also propagate the cardinality distribution of RFS X_k [17] [18]).

Let us introduce or recall some notations:

- $p_{s,k}(\mathbf{x})$ (resp. $p_{d,k}(\mathbf{x})$) is the probability that a target with state $\mathbf{x}$ at time $k-1$ (resp. k) still exists (resp. is detected) at time k;

- $\kappa_k(\mathbf{z})$ is the PHD of the clutter measurements RFS at time k;
- $\gamma_k(\mathbf{x})$ is the PHD of the birth targets RFS at time k;
- $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ (resp. $g_k(\mathbf{z}_k|\mathbf{x}_k)$) is the transition (resp. likelihood) pdf of the HMC model followed by the targets.

For simplicity we consider in this paper that there is no spawning (if there is spawning the extension is immediate). The PHD filter (derived by Mahler [7]) propagates the a posteriori PHD $v_k(\mathbf{x})$, and consists of a prediction step followed by an updating step:

$$v_{k|k-1}(\mathbf{x}) = \int p_{s,k}(\mathbf{x}_{k-1}) f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}) \times v_{k-1}(\mathbf{x}_{k-1}) \mathrm{d}\mathbf{x}_{k-1} + \gamma_k(\mathbf{x}), \quad (14)$$

$$v_k(\mathbf{x}) = [1 - p_{d,k}(\mathbf{x})]\, v_{k|k-1}(\mathbf{x}) + \sum_{\mathbf{z} \in Z_k} \frac{p_{d,k}(\mathbf{x}) g_k(\mathbf{z}|\mathbf{x}) v_{k|k-1}(\mathbf{x})}{\kappa_k(\mathbf{z}) + \int p_{d,k}(\mathbf{x}) g_k(\mathbf{z}|\mathbf{x}) v_{k|k-1}(\mathbf{x})\, \mathrm{d}\mathbf{x}}. \quad (15)$$

Although equations (14) and (15) now involve classical integrals, it is still impossible to compute them exactly, even in the linear and Gaussian case. So many contributions have been proposed [11] [14] [16] which aim at deriving relevant implementations of the PHD filter.

B. A PMC approach

In section II we raised the statistical interest of PMC models. So from now on we assume that the targets and their associated measurements follow a PMC model and we extend the PHD filter (14)-(15) to these new conditions.

1) General considerations - taking into account observations via a 2-dimensional joint intensity: First, note that for a given pair (target, observation) following a PMC model, the local transition pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and likelihood pdf $p(\mathbf{y}_k|\mathbf{x}_k)$ still exist but should not simply be used instead of $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $g_k(\mathbf{z}|\mathbf{x}_k)$ in (14)-(15). Doing so would implicitely mean that properties **P.1** and **P.2** hold which, as we have seen already, is not necessarily true in a PMC model.

Next, directly computing $v_k(\mathbf{x})$ in a PMC seems complicated because $\mathbf{x}_k$ is not necessarily Markovian, and the dynamics of each target now depends on its associated observation (see (12)). However, as we shall see in section III-B2 it will indeed be easier to propagate a joint intensity $v_k(\mathbf{x}, \mathbf{y})$, and eventually obtain the PHD $v_k(\mathbf{x})$ of interest as $v_k(\mathbf{x}) = \int v_k(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{y}$ (note that by a slight abuse of notation, $v_k(.)$ denotes either a function of $\mathbf{x}$ or of $(\mathbf{x}, \mathbf{y})$).

This difficulty can be highlighted with the following single object tracking scenario where misdetections may arise. So let us assume that $p(\mathbf{x}_{k-1}|\mathbf{y}_{0:k-1})$ is available at time $k-1$. At time k, $\mathbf{y}_k$ is not observed (hidden state $\mathbf{x}_k$ is not detected) and we want to compute, at time $k+1$, $p(\mathbf{x}_{k+1}|\mathbf{y}_{0:k-1,k+1})$ ($\mathbf{x}_{k+1}$ is detected), where $\mathbf{y}_{0:k-1,k+1} = (\mathbf{y}_0, \cdots, \mathbf{y}_{k-1}, \mathbf{y}_{k+1})$. Remember that $\boldsymbol{\xi}_k = (\mathbf{x}_k, \mathbf{y}_k)$. We have

$$p(\mathbf{x}_{k+1}|\mathbf{y}_{0:k-1,k+1}) \propto \int p_{k+1|k}(\boldsymbol{\xi}_{k+1}|\boldsymbol{\xi}_k) p(\boldsymbol{\xi}_k|\mathbf{y}_{0:k-1}) \mathrm{d}\boldsymbol{\xi}_k, \quad (16)$$

$$p(\boldsymbol{\xi}_k|\mathbf{y}_{0:k-1}) = \int p_{k|k-1}(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{0:k-1}) \mathrm{d}\mathbf{x}_k. \quad (17)$$

So $p(\mathbf{x}_{k+1}|\mathbf{y}_{0:k-1,k+1})$ can be computed from $p(\mathbf{x}_{k-1}|\mathbf{y}_{0:k-1})$, but via the *joint* pdf $p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{y}_{0:k-1})$. The reason why is that $p_{k+1|k}(\boldsymbol{\xi}_{k+1}|\boldsymbol{\xi}_k)$ (in general) depends on $\mathbf{y}_k$: even though $\mathbf{y}_k$ has not been observed, averaging on all $\mathbf{y}_k$ is necessary to build $p(\mathbf{x}_{k+1}|\mathbf{y}_{0:k-1,k+1})$, whence (16). By contrast, in an HMC model (8) is satisfied, so $p_{k+1|k}(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}|\mathbf{x}_k, \mathbf{y}_k)$ no longer depends on $\mathbf{y}_k$; averaging on $\mathbf{y}_k$ becomes useless in this case, and indeed (16)-(17) reduce to

$$p(\mathbf{x}_{k+1}|\mathbf{y}_{0:k-1,k+1}) \propto g_{k+1}(\mathbf{y}_{k+1}|\mathbf{x}_{k+1}) \times \int f_{k+1|k}(\mathbf{x}_{k+1}|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{0:k-1}) \mathrm{d}\mathbf{x}_k,$$

$$p(\mathbf{x}_k|\mathbf{y}_{0:k-1}) = \int f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{0:k-1}) \mathrm{d}\mathbf{x}_{k-1}.$$

So in an HMC $p(\mathbf{x}_{k+1}|\mathbf{y}_{0:k-1,k+1})$ can be computed from $p(\mathbf{x}_{k-1}|\mathbf{y}_{0:k-1})$ via $p(\mathbf{x}_k|\mathbf{y}_{0:k-1})$, and finally computing $p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{y}_{0:k-1})$ is no longer necessary.

Of course, similar difficulties occur in the multi-target framework, since at each time instant k, each target $\mathbf{x}_{i,k}$ may be detected (which means that among all observations there exists a realization of r.v. $\mathbf{y}_{i,k}$ associated to target $\mathbf{x}_{i,k}$) or undetected, and some of the observations may be false alarms. Consequently in the next section, we shall look for propagating a joint (target, observation) PHD filter.

2) A PHD filter for PMC models: So we now look for computing the PHD of RFS

$$\dot{X}_k = \{\boldsymbol{\xi}_{1,k}, \cdots, \boldsymbol{\xi}_{n',k}\}$$

given a realization of all past measurements $Z_{0:k}$, where

$$Z_k = \{\mathbf{z}_{1,k}, \cdots, \mathbf{z}_{m',k}\}.$$

By construction, the dynamical evolution of a random couple in $\dot{X}_k$, and the connection of a measurement in Z_k with some couple in $\dot{X}_k$, are described as follows:

1) Since each individual pair (target, observation) follows a PMC model, $(\mathbf{x}_k, \mathbf{y}_k)$ is Markovian with transition pdf $p_{k|k-1}(\mathbf{x}_k, \mathbf{y}_k|\mathbf{x}_{k-1}, \mathbf{y}_{k-1})$. Observe that $p_{k|k-1}$ includes the dynamical model of the target via $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_{k-1})$, but also the link to the sensor via $p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_{k-1})$. For physical reasons, we will assume that a couple $\boldsymbol{\xi}_k$ survives with a probability $p_{s,k+1}(\mathbf{x}_k)$ which depends only on state $\mathbf{x}_k$;
2) given couple $\boldsymbol{\xi}_k = (\mathbf{x}_k, \mathbf{y}_k)$, we define the likelihood of a measurement $\mathbf{z}$ in Z_k as $\delta_{\mathbf{z}}(\mathbf{y}_k)$ where $\delta_{\mathbf{z}}(\mathbf{y}_k)$ is the Dirac delta function which satisfies, for $S_q \subset \mathbb{R}^q$, $\int_{S_q} \delta_{\mathbf{z}}(\mathbf{y}_k) \mathrm{d}\mathbf{y}_k = 1$ if $\mathbf{z} \in S_q$ and 0 otherwise (see e.g. [8, p. 693]) Indeed, remember that $\mathbf{y}_k$ is the r.v. observation associated to the target with state $\mathbf{x}_k$; so for a given realization $\mathbf{y}_k$, if $\mathbf{z} \neq \mathbf{y}_k$, one can claim that measurement $\mathbf{z}$ is not produced by $\mathbf{x}_k$ and that (given $\boldsymbol{\xi}_k$) it does not depend on past measurements. For physical reasons, we will assume that a couple $\boldsymbol{\xi}_k$ is detected with a probability $p_{d,k}(\mathbf{x}_k)$ which depends only on state $\mathbf{x}_k$.

From the previous statistical considerations, the multi-object Bayes Filter reads

$$p(\dot{X}_k|Z_{0:k-1}) = \int f_{k|k-1}(\dot{X}_k|\dot{X}_{k-1})p(\dot{X}_{k-1}|Z_{0:k-1})\delta\dot{X}_{k-1}, \quad (18)$$

$$p(\dot{X}_k|Z_{0:k}) \propto g_k(Z_k|\dot{X}_k)p(\dot{X}_k|Z_{0:k-1}), \quad (19)$$

where $\delta\dot{X}$ denotes the set integral, $f_{k|k-1}(\dot{X}_k|\dot{X}_{k-1})$ and $g_k(Z_k|\dot{X}_k)$ are the multi-object Markov and likelihood densities which can be constructed as in [8, eq. 13.43] and [8, eq. 12.140] respectively (see also [14]), except that the transition of an element of $\dot{X}_k$ is given by $p_{k|k-1}(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})$ and the likelihood of a measurement $\mathbf{z} \in Z_k$ with $\boldsymbol{\xi}_k$ by $\delta_{\mathbf{z}}(\mathbf{y}_k)$.

We are now ready to derive a PHD filter for PMC models. We first need to adapt assumptions **A.1-A.3** (see section III-A2) to the PMC context. So let **A.1-A.3** become **A'.1-A'.3**. **A'.1** assumes that each given pair (target, observation) is a PMC and evolves independently of any other pair; **A'.2** coincides with **A.2**; and **A'.3** states that the predicted-couple process which governs the birth and surviving targets is Poisson.

From the previous assumptions and the multi-object Bayes Filter (18)-(19), it follows that PHD $v_k(\mathbf{x}, \mathbf{y})$ can be propagated via a prediction and an update equations:

$$v_{k|k-1}(\mathbf{x}, \mathbf{y}) = \int p_{s,k}(\mathbf{x}_{k-1})p_{k|k-1}(\mathbf{x}, \mathbf{y}|\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) \times v_{k-1}(\mathbf{x}_{k-1}, \mathbf{y}_{k-1})\mathrm{d}\mathbf{x}_{k-1}\mathrm{d}\mathbf{y}_{k-1} + \gamma_k(\mathbf{x}, \mathbf{y}), \quad (20)$$

$$v_k(\mathbf{x}, \mathbf{y}) = [1 - p_{d,k}(\mathbf{x})]\, v_{k|k-1}(\mathbf{x}, \mathbf{y}) + \sum_{\mathbf{z}\in Z_k} \frac{p_{d,k}(\mathbf{x})v_{k|k-1}(\mathbf{x}, \mathbf{z})\delta_{\mathbf{z}}(\mathbf{y})}{\kappa_k(\mathbf{z}) + \int p_{d,k}(\mathbf{x})v_{k|k-1}(\mathbf{x}, \mathbf{z})\,\mathrm{d}\mathbf{x}}. \quad (21)$$

Finally remember that $v_k(\mathbf{x}) = \int v_k(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{y}$. Note that this integral w.r.t. variable $\mathbf{y}$ can reduce to a sum, depending on the nature of v_k: the first term in (21) is associated to non-detected couples $(\mathbf{x}, \mathbf{y})$, while the second term is associated to detected couples. Remember that if a target is detected, its associated measurement is necessarily in Z_k and it is thus not surprising that the expectation of of the number of detected couples in region where $\mathbf{y} \neq \mathbf{z}$, $\mathbf{z} \in Z_k$, is null.

Remark 1 Let us analyze how $v_k(\mathbf{x}) = \int v_k(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{y}$ propagates in time when the parameters of the PMC PHD filter reduce to those of the HMC one, i.e. when for given functions $f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1})$, $g_k(\mathbf{y}|\mathbf{x})$ and $\tilde{\gamma}_k(\mathbf{x})$, $p_{k|k-1}(\mathbf{x}, \mathbf{y}|\mathbf{x}_{k-1}, \mathbf{y}_{k-1})$ and $\gamma_k(\mathbf{x}, \mathbf{y})$ are built as

$$p_{k|k-1}(\mathbf{x}, \mathbf{y}|\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) = f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1})g_k(\mathbf{y}|\mathbf{x}), \quad (22)$$

$$\gamma_k(\mathbf{x}, \mathbf{y}) = g_k(\mathbf{y}|\mathbf{x})\tilde{\gamma}_k(\mathbf{x}) \quad (23)$$

(this choice is justified by the fact that in an HMC model, the likelihood of a measurement $\mathbf{y}$ with state $\mathbf{x}$ is given by $g_k(\mathbf{y}|\mathbf{x})$). Writing

$$v_{k-1}(\mathbf{x}, \mathbf{y}) = \frac{v_{k-1}(\mathbf{x}, \mathbf{y})}{\int v_{k-1}(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{y}} v_{k-1}(\mathbf{x}), \quad (24)$$

and applying (20)-(21), we show that $v_{k-1}(\mathbf{x})$ is propagated via (14)-(15), which means that the marginalized PMC PHD filter reduces to the classical HMC PHD one.

C. A GM implementation of the PMC-PHD filter for linear and Gaussian PMC

Linear and Gaussian models are widely used in multi-object filtering [11] [10]. So in this section, we focus on general linear and Gaussian PMC models (9)-(11). We assume that the birth intensity $\gamma_k(\mathbf{x}, \mathbf{y})$ is a GM:

$$\gamma_k(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{J_{\gamma_k}} w_{\gamma_k}^{(i)} \mathcal{N}(\mathbf{x}, \mathbf{y}; \mathbf{m}_{\gamma_k}^{(i)}, \mathbf{P}_{\gamma_k}^{(i)}), \quad (25)$$

and that the probabilities of survival $p_{s,k}$ and of detection $p_{d,k}$ are independent of state $\mathbf{x}$. Let us assume that at time $k-1$, PHD v_{k-1} writes as

$$v_{k-1}(\mathbf{x}, \mathbf{y}) = v_{k-1}^1(\mathbf{x}, \mathbf{y}) + v_{k-1}^2(\mathbf{x}, \mathbf{y}), \quad (26)$$

where

$$v_{k-1}^1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{J_{k-1}^1} w_{k-1}^{1,(i)} \mathcal{N}(\mathbf{x}, \mathbf{y}; \mathbf{m}_{k-1}^{1,(i)}, \mathbf{P}_{k-1}^{1,(i)}), \quad (27)$$

$$v_{k-1}^2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{J_{k-1}^2} w_{k-1}^{2,(i)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{k-1}^{2,(i)}, \mathbf{P}_{k-1}^{2,(i)})\delta_{\mathbf{z}^{(i)}}(\mathbf{y}), \quad (28)$$

where $\mathbf{z}^{(i)}$ belongs to a given set $Z = \{\mathbf{z}^{(1)}, \cdots, \mathbf{z}^{(J_{k-1}^2)}\}$ with possible repetitions.

1) Prediction Step: Plugging (26), (27) and (28) in (20), PHD $v_{k|k-1}(\mathbf{x}, \mathbf{y})$ is a GM:

$$v_{k|k-1}(\mathbf{x}, \mathbf{y}) = v_{k|k-1}^1(\mathbf{x}, \mathbf{y}) + v_{k|k-1}^2(\mathbf{x}, \mathbf{y}) + \gamma_k(\mathbf{x}, \mathbf{y}), \quad (29)$$

$$v_{k|k-1}^1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{J_{k-1}^1} w_{k|k-1}^{1,(i)} \mathcal{N}(\mathbf{x}, \mathbf{y}; \mathbf{m}_{k|k-1}^{1,(i)}, \mathbf{P}_{k|k-1}^{1,(i)}), \quad (30)$$

$$v_{k|k-1}^2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{J_{k-1}^2} w_{k|k-1}^{2,(i)} \mathcal{N}(\mathbf{x}, \mathbf{y}; \mathbf{m}_{k|k-1}^{2,(i)}, \mathbf{P}_{k|k-1}^{2,(i)}), \quad (31)$$

where

$$w_{k|k-1}^{j,(i)} = p_{s,k} w_{k-1}^{j,(i)} \text{ for } j = \{1, 2\}, \quad (32)$$

$$\mathbf{m}_{k|k-1}^{1,(i)} = \mathbf{B}_k \mathbf{m}_{k-1}^{1,(i)}, \;\; \mathbf{m}_{k|k-1}^{2,(i)} = \mathbf{B}_k \begin{bmatrix} \mathbf{m}_{k-1}^{2,(i)} \\ \mathbf{z}^{(i)} \end{bmatrix}, \quad (33)$$

$$\mathbf{P}_{k|k-1}^{1,(i)} = \mathbf{\Sigma}_k + \mathbf{B}_k \mathbf{P}_{k-1}^{1,(i)} \mathbf{B}_k^T, \quad (34)$$

$$\mathbf{P}_{k|k-1}^{2,(i)} = \mathbf{\Sigma}_k + \begin{bmatrix} \mathbf{F}_k^1 \\ \mathbf{H}_k^1 \end{bmatrix} \mathbf{P}_{k-1}^{2,(i)} [\mathbf{F}_k^{1\,T} \mathbf{H}_k^{1\,T}]. \quad (35)$$

This result is based on Lemma 1 (see Appendix A) applied to $\boldsymbol{\zeta} = (\mathbf{x}, \mathbf{y})$, $\boldsymbol{\eta} = (\mathbf{x}_{k-1}, \mathbf{y}_{k-1})$ for the derivation of $v_{k|k-1}^1(\mathbf{x}, \mathbf{y})$ and to $\boldsymbol{\zeta} = (\mathbf{x}, \mathbf{y})$, $\boldsymbol{\eta} = \mathbf{x}_{k-1}$ for that of $v_{k|k-1}^2(\mathbf{x}, \mathbf{y})$. Finally, by reordering the terms of GM (29) and setting $J_{k|k-1} = J_{k-1}^1 + J_{k-1}^2 + J_{\gamma_k}$, $v_{k|k-1}$ can be rewritten as

$$v_{k|k-1}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^{(i)} \mathcal{N}(\mathbf{x}, \mathbf{y}; \mathbf{m}_{k|k-1}^{(i)}; \mathbf{P}_{k|k-1}^{(i)}), \quad (36)$$

where we introduce the following notations:

$$\mathbf{m}_{k|k-1}^{(i)} = \begin{bmatrix} \mathbf{m}_{k|k-1}^{\mathbf{x},(i)}, \\ \mathbf{m}_{k|k-1}^{\mathbf{y},(i)} \end{bmatrix}, \; \mathbf{P}_{k|k-1}^{(i)} = \begin{bmatrix} \mathbf{P}_{k|k-1}^{\mathbf{x},(i)} & \mathbf{P}_{k|k-1}^{\mathbf{xy},(i)} \\ \mathbf{P}_{k|k-1}^{\mathbf{xy},(i)\,T} & \mathbf{P}_{k|k-1}^{\mathbf{y},(i)} \end{bmatrix}. \quad (37)$$

2) Update Step: Plugging (36) in (21) and using Lemma 2 (see Appendix A), we get the following expression for $v_k(\mathbf{x},\mathbf{y})$:

$$v_k(\mathbf{x},\mathbf{y}) = v_k^1(\mathbf{x},\mathbf{y}) + v_k^2(\mathbf{x},\mathbf{y}), \quad (38)$$

where

$$v_k^1(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{J_{k|k-1}} w_k^{1,(i)} \mathcal{N}(\mathbf{x},\mathbf{y};\mathbf{m}_{k|k-1}^{(i)};\mathbf{P}_{k|k-1}^{(i)}), \quad (39)$$

$$v_k^2(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{z}\in Z_k} \sum_{i=1}^{J_{k|k-1}} w_k^{2,(i)}(\mathbf{z}) \mathcal{N}(\mathbf{x};\mathbf{m}_k^{2,(i)}(\mathbf{z});\mathbf{P}_k^{2,(i)})\delta_{\mathbf{z}}(\mathbf{y}), \quad (40)$$

where

$$w_k^{1,(i)} = (1-p_{d,k}) w_{k|k-1}^{(i)}, \quad (41)$$

$$w_k^{2,(i)}(\mathbf{z}) = \frac{p_{d,k} w_{k|k-1}^{(i)} q_k^{(i)}(\mathbf{z})}{\kappa_k(\mathbf{z}) + p_{d,k}\sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^{(i)} q_k^{(i)}(\mathbf{z})}, \quad (42)$$

$$q_k^{(i)}(\mathbf{z}) = \mathcal{N}(\mathbf{z};\mathbf{m}_{k|k-1}^{\mathbf{y},(i)};\mathbf{P}_{k|k-1}^{\mathbf{y},(i)}), \quad (43)$$

$$\mathbf{m}_k^{2,(i)}(\mathbf{z}) = \mathbf{m}_{k|k-1}^{\mathbf{x},(i)} + \mathbf{K}_k^{(i)}(\mathbf{z} - \mathbf{m}_{k|k-1}^{\mathbf{y},(i)}), \quad (44)$$

$$\mathbf{P}_k^{2,(i)} = \mathbf{P}_{k|k-1}^{\mathbf{x},(i)} - \mathbf{K}_k^{(i)} {\mathbf{P}_{k|k-1}^{\mathbf{xy},(i)}}^T, \quad (45)$$

$$\mathbf{K}_k^{(i)} = \mathbf{P}_{k|k-1}^{\mathbf{xy},(i)} (\mathbf{P}_{k|k-1}^{\mathbf{y},(i)})^{-1}. \quad (46)$$

By reordering the terms of mixture (40) and setting $J_k^2 = J_{k|k-1} \times \mathrm{Card}(Z_k)$, $v_k^2(\mathbf{x},\mathbf{y})$ can be rewritten as

$$v_k^2(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{J_k^2} w_k^{2,(i)} \mathcal{N}(\mathbf{x};\mathbf{m}_k^{2,(i)};\mathbf{P}_k^{2,(i)})\delta_{\mathbf{z}^{(i)}}(\mathbf{y}); \quad (47)$$

so the form of PHD v_k is the same as that of v_{k-1}. Contrary to the classical GM implementation [11], PHD v_k is no longer a GM but a sum of a GM for components which correspond to mis-detections and of a normal-discrete (ND) mixture (i.e., a GM w.r.t. $\mathbf{x}$ and a discrete mixture w.r.t. $\mathbf{y}$) for detection terms. However, $v_k(\mathbf{x}) = \int v_k(\mathbf{x},\mathbf{y})\mathrm{d}\mathbf{y}$ remains a GM.

Remark 2 Remember that if the PMC reduces to an HMC and assumption (23) on $\gamma_k(\mathbf{x},\mathbf{y})$ holds, then the PMC PHD filter reduces to the classical one. So here, if

$$\mathbf{B}_k = \begin{bmatrix} \mathbf{F}_k & \mathbf{0} \\ \mathbf{H}_k\mathbf{F}_k & \mathbf{0} \end{bmatrix}, \mathbf{\Sigma}_k = \begin{bmatrix} \mathbf{Q}_k & (\mathbf{H}_k\mathbf{Q}_k)^T \\ (\mathbf{H}_k\mathbf{Q}_k) & \mathbf{R}_k + \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^T \end{bmatrix}, \quad (48)$$

and for all i, $1 \leq i \leq J_{\gamma_k}$ $m_{\gamma_k}^{(i)} = \begin{bmatrix} \mathbf{m}_{\gamma_k}^{\mathbf{x},(i)} & \mathbf{H}_k\mathbf{m}_{\gamma_k}^{\mathbf{x},(i)} \end{bmatrix}^T$, $\mathbf{P}_{\gamma_k}^{(i)} = \begin{bmatrix} \mathbf{P}_{\gamma_k}^{\mathbf{x},(i)} & (\mathbf{H}_k\mathbf{P}_{\gamma_k}^{\mathbf{x},(i)})^T \\ \mathbf{H}_k\mathbf{P}_{\gamma_k}^{\mathbf{x},(i)} & \mathbf{R}_k + \mathbf{H}_k\mathbf{P}_{\gamma_k}^{\mathbf{x},(i)}\mathbf{H}_k^T, \end{bmatrix}$ then the GM-PMC PHD filter reduces to the classical GM PHD one [11] with $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k;\mathbf{F}_k\mathbf{x}_{k-1};\mathbf{Q}_k)$, $g_k(\mathbf{y}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k;\mathbf{H}_k\mathbf{x}_k;\mathbf{R}_k)$ and $\gamma_k(\mathbf{x}) = \sum_{i=1}^{J_{\gamma_k}} w_{\gamma_k}^{(i)}\mathcal{N}(\mathbf{x};\mathbf{m}_{\gamma_k}^{(i)};\mathbf{P}_{\gamma_k}^{(i)})$.

Comparing to the GM HMC PHD filter [11], the extra computational cost induced by our GM PMC PHD implementation is negligible since it is only due to matrices $\mathbf{F}_k^2$ and $\mathbf{H}_k^2$ which, for example, add some extra terms in (33). On the other hand, this GM implementation for linear and Gaussian PMC models suffers from the same implementation issues as the GM one for linear and Gaussian HMC ones: due to the update step, the number of components of the mixture grows exponentially. So one should adapt existing solutions to reduce the number of components. The pruning and cutting techniques are well known [11]: we delete components of the mixture, the weights $w_k^{1,(i)}$ and $w_k^{2,(i)}$ of which are under a given threshold T_p, and one can limit the number of components of $v_k^1(\mathbf{x},\mathbf{y})$ and of $v_k^2(\mathbf{x},\mathbf{y})$ by keeping respectively the $J_{\max}^1$ and $J_{\max}^2$ components with highest weights. However, due to the nature of $v_k(\mathbf{x},\mathbf{y})$, the merging step is slightly modified, as explained in the next section.

3) Merging: A well known technique for reducing the cardinality of the mixture consists in merging components which are close. However here, contrary to the merging step of the GM implementation in the HMC case [11], one should not merge the components of $v_k^1(\mathbf{x},\mathbf{y})$ which are due to mis-detections with those of $v_k^2(\mathbf{x},\mathbf{y})$ which are due to detections because their forms are different: $v_k^1(\mathbf{x},\mathbf{y})$ is a GM while $v_k^2(\mathbf{x},\mathbf{y})$ is an ND mixture. This argument can be illustrated with the following scenario. Assume that $v_{k|k-1}(\mathbf{x},\mathbf{y}) = \mathcal{N}(\mathbf{x},\mathbf{y};\mathbf{m}_{k|k-1};\mathbf{P}_{k|k-1})$, $\kappa_k = 0$ and $Z_k = \{\mathbf{z}\}$. Then $v_k(\mathbf{x},\mathbf{y}) = (1-p_{d,k})\mathcal{N}(\mathbf{x},\mathbf{y};\mathbf{m}_{k|k-1};\mathbf{P}_{k|k-1}) + \mathcal{N}(\mathbf{x},\mathbf{y};\mathbf{m}_k(\mathbf{z});\mathbf{P}_k)\delta_{\mathbf{z}}(\mathbf{y})$. So if we assume that $\mathcal{N}(\mathbf{x};\mathbf{m}_{k|k-1}^{\mathbf{x}};\mathbf{P}_{k|k-1}^{\mathbf{x}})$ and $\mathcal{N}(\mathbf{x};\mathbf{m}_k^{\mathbf{x}};\mathbf{P}_k^{\mathbf{x}})$ are close enough, they would be merged in the classical GM implementation [11]. The state is estimated by $((1-p_{d,k})\mathbf{m}_{k|k-1}^{\mathbf{x}} + \mathbf{m}_k^{\mathbf{x}}(\mathbf{z}))/(2-p_{d,k})$; however, we are sure that $\mathbf{z}$ is due to the target described by Gaussian $\mathcal{N}(\mathbf{x},\mathbf{y};\mathbf{m}_{k|k-1};\mathbf{P}_{k|k-1})$ so we should estimate the state with $\mathbf{m}_k^{\mathbf{x}}(\mathbf{z})$.

Remark 3 In the scenario described above, an estimate of the number of targets is $2-p_{d,k}$, whereas it should be 1, since measurement $\mathbf{z}$ is inevitably due to the predicted state $\mathbf{m}_{k|k-1}$ (there is no false alarm measurement) and the expectation of the predicted number of targets is 1. As is well-known [31], this problem is due to the Poisson assumption upon which the derivation of the update step of the PHD filter [7] relies, and is corrected by the CPHD filter [17].

So we actually consider two merging steps: one for $v_{k|k}^1(\mathbf{x},\mathbf{y})$ and one for $v_{k|k}^2(\mathbf{x},\mathbf{y})$. For $v_{k|k}^1(\mathbf{x},\mathbf{y})$, we will merge Gaussians (i) and (j) (which are of dimension $m+q$) if for a given threshold T_m we have:

$$(\mathbf{m}_k^{1,(i)} - \mathbf{m}_k^{1,(j)})^T(\mathbf{P}_k^{1,(i)})^{-1}(\mathbf{m}_k^{1,(i)} - \mathbf{m}_k^{1,(j)}) \leq T_m. \quad (49)$$

Finally, two components (i) and (j) (now of dimension m) of v_k^2 will be merged if $\mathbf{m}_k^{2,(i)}$, $\mathbf{m}_k^{2,(j)}$ and $\mathbf{P}_k^{2,(i)}$ satisfy (49) and $\mathbf{z}^{(i)} = \mathbf{z}^{(j)}$. In other words, we only merge Gaussians which share the same observation $\mathbf{z}$.

Once the pruning and merging steps have been performed, one can extract states: a Gaussian with weights above a given threshold, typically 0.5, will be considered as a target with state given by the mean of the Gaussian.

IV. Applications and simulations

In this section, we describe some statistical models which are of practical interest and which do not fit the HMC framework. The practical interest of some particular PMC models has been pointed out in a recent contribution [22]. In addition, we present another practical interest of PMC models: we show that linear and Gaussian PMC models enable us to to keep the physical constraints of a given HMC model while relaxing its independence assumptions. Finally, we compare both the HMC and PMC based PHD filters on simulations.

A. Example 1 - Correlated process and measurement noises

In the linear and Gaussian model (2)-(3), we have assumed that process and observation noises were independent. Let us now assume that $\mathbf{u}_k$ and $\mathbf{v}_k$ are dependent with $\mathrm{E}(\mathbf{u}_k\mathbf{v}_k^T) = \mathbf{S}_k \neq \mathbf{0}$. Such a correlation may appear during the discretization step of a continuous model [22]. This model is a linear and Gaussian PMC model in which matrix $\mathbf{B}_k$ is unchanged when compared to the linear and Gaussian HMC model (4), but the covariance matrix $\mathbf{\Sigma}_k = \mathrm{E}(\mathbf{w}_k\mathbf{w}_k^T)$ in (5) becomes

$$\mathbf{\Sigma}_k = \begin{bmatrix} \mathbf{Q}_k & (\mathbf{H}_k\mathbf{Q}_k + \mathbf{S}_k^T)^T \\ \mathbf{H}_k\mathbf{Q}_k + \mathbf{S}_k^T & \mathbf{R}_k + \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^T + \mathbf{H}_k\mathbf{S}_k^T + \mathbf{S}_k\mathbf{H}_k^T \end{bmatrix}.$$

Indeed, using Lemma 2 in Appendix A, we see that $p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_{k-1})$ depends on $\mathbf{x}_k$ *and* $\mathbf{x}_{k-1}$ via its mean $\mathbf{m}^{\mathbf{y}_k}(\mathbf{x}_{k-1}, \mathbf{x}_k)$, which reads $\mathbf{m}^{\mathbf{y}_k}(\mathbf{x}_{k-1}, \mathbf{x}_k) = \mathbf{H}_k\mathbf{F}_k\mathbf{x}_{k-1} + (\mathbf{H}_k\mathbf{Q}_k + \mathbf{S}_k^T)\mathbf{Q}_k^{-1}(\mathbf{x}_k - \mathbf{F}_k\mathbf{x}_{k-1})$ (of course, $\mathbf{m}^{\mathbf{y}_k}(\mathbf{x}_{k-1}, \mathbf{x}_k)$ reduces to $\mathbf{H}_k\mathbf{x}_k$ if $\mathbf{S}_k = \mathbf{0}$.) So property **P.2** is not satisfied, which ensures that this simple model is not an HMC.

B. Example 2 - Colored measurement noise

Another example [32, section 8.4] [33] is the case where we relax the mutual independence hypothesis of noises $\mathbf{v}_k$ in (3) and assume that $\{\mathbf{v}_k\}_{k\geq 0}$ is an MC (observation noise $\mathbf{v}_k$ is colored), $\mathbf{v}_k = \mathbf{H}_k^n\mathbf{v}_{k-1} + \tilde{\mathbf{v}}_k$, where $\tilde{\mathbf{v}}_k \sim \mathcal{N}(.; \mathbf{0}; \tilde{\mathbf{R}}_k)$ is a sequence of independent noises, independent of $\mathbf{u}_1, \cdots, \mathbf{u}_k$ and of $\mathbf{x}_0$. Then, using (2) and $\mathbf{v}_{k-1} = \mathbf{y}_{k-1} - \mathbf{H}_{k-1}\mathbf{x}_{k-1}$, (3) becomes

$$\begin{aligned} \mathbf{y}_k &= \mathbf{H}_k\mathbf{x}_k + \mathbf{H}_k^n\mathbf{v}_{k-1} + \tilde{\mathbf{v}}_k \\ &= (\mathbf{H}_k\mathbf{F}_k - \mathbf{H}_k^n\mathbf{H}_{k-1})\,\mathbf{x}_{k-1} + \mathbf{H}_k^n\mathbf{y}_{k-1} + \mathbf{H}_k\mathbf{u}_k + \tilde{\mathbf{v}}_k. \end{aligned}$$

This model is still a linear and Gaussian PMC (9) where

$$\mathbf{B}_k = \begin{bmatrix} \mathbf{F}_k & \mathbf{0} \\ \mathbf{H}_k\mathbf{F}_k - \mathbf{H}_k^n\mathbf{H}_{k-1} & \mathbf{H}_k^n \end{bmatrix}, \quad (50)$$

$$\mathbf{\Sigma}_k = \begin{bmatrix} \mathbf{Q}_k & (\mathbf{H}_k\mathbf{Q}_k)^T \\ \mathbf{H}_k\mathbf{Q}_k & \tilde{\mathbf{R}}_k + \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^T \end{bmatrix}. \quad (51)$$

Again, $p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_{k-1})$ depends on $\mathbf{x}_k$, but also on $\mathbf{x}_{k-1}$ and $\mathbf{y}_{k-1}$, via its mean $\mathbf{m}^{\mathbf{y}_k}(\mathbf{x}_{k-1}, \mathbf{x}_k, \mathbf{y}_{k-1})$ which here reads $\mathbf{m}^{\mathbf{y}_k}(\mathbf{x}_{k-1}, \mathbf{x}_k, \mathbf{y}_{k-1}) = \mathbf{H}_k\mathbf{x}_k - \mathbf{H}_k^n\mathbf{H}_{k-1}\mathbf{x}_{k-1} + \mathbf{H}_k^n\mathbf{y}_{k-1}$.

C. A particular class of linear and Gaussian PMC models

HMC are popular partly because $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $g_k(\mathbf{y}_k|\mathbf{x}_k)$ describe a given physical problem of interest for practitioners. The derivation of linear and Gaussian PMC models which share the same local physical properties as given HMC ones, but in which some independence assumptions would be relaxed, and in particular properties **P.1** and **P.2**, is thus a major stake.

So let us assume that $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1}; \mathbf{Q}_k)$ and $g_k(\mathbf{y}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{H}_k\mathbf{x}_k; \mathbf{R}_k)$ are given. Our goal in this section is to construct PMC models, such that the transition pdf $p_{k|k-1}(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})$ is independent of the initial conditions at time $k = 0$, and such that for all k, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ coincides with $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$, and $p(\mathbf{y}_k|\mathbf{x}_k)$ with $g_k(\mathbf{y}_k|\mathbf{x}_k)$; but remember that in general $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ will not be equal to $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_{k-1})$, and $p(\mathbf{y}_k|\mathbf{x}_k)$ will not be equal to $p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_{k-1})$. We have the following result (the proof is given in Appendix B).

Proposition 1 *Let us consider an HMC model with*

$$p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; \mathbf{m}_0; \mathbf{P}_0), \quad (52)$$

$$f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1}; \mathbf{Q}_k) \text{ and} \quad (53)$$

$$g_{k|k-1}(\mathbf{y}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{H}_k\mathbf{x}_k; \mathbf{R}_k), \quad (54)$$

for all k. Then the linear and Gaussian PMC models which satisfy $p(\mathbf{x}_k|\mathbf{x}_{k-1}) = f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $p(\mathbf{y}_k|\mathbf{x}_k) = g_{k|k-1}(\mathbf{y}_k|\mathbf{x}_k)$, and such that $p_{k|k-1}(\mathbf{x}_k, \mathbf{y}_k|\mathbf{x}_{k-1}, \mathbf{y}_{k-1})$ does not depend on parameters $(\mathbf{m}_0, \mathbf{P}_0)$, are described by the following equations:

$$p(\boldsymbol{\xi}_0) = \mathcal{N}\left(\boldsymbol{\xi}_0; \begin{bmatrix} \mathbf{m}_0 \\ \mathbf{H}_0\mathbf{m}_0 \end{bmatrix}; \begin{bmatrix} \mathbf{P}_0 & (\mathbf{H}_0\mathbf{P}_0)^T \\ \mathbf{H}_0\mathbf{P}_0 & \mathbf{R}_0 + \mathbf{H}_0\mathbf{P}_0\mathbf{H}_0^T \end{bmatrix}\right), \quad (55)$$

$$p_{k|k-1}(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) = \mathcal{N}(\boldsymbol{\xi}_k; \mathbf{B}_k\boldsymbol{\xi}_{k-1}; \mathbf{\Sigma}_\mathbf{k}), \quad (56)$$

where matrices $\mathbf{B}_k$ and $\mathbf{\Sigma}_k$ are defined by

$$\mathbf{B}_k = \begin{bmatrix} \mathbf{F}_k - \mathbf{F}_k^2\mathbf{H}_{k-1} & \mathbf{F}_k^2 \\ \mathbf{H}_k\mathbf{F}_k - \mathbf{H}_k^2\mathbf{H}_{k-1} & \mathbf{H}_k^2 \end{bmatrix}, \quad (57)$$

$$\mathbf{\Sigma}_k = \begin{bmatrix} \mathbf{\Sigma}_k^{11} & {\mathbf{\Sigma}_k^{21}}^T \\ \mathbf{\Sigma}_k^{21} & \mathbf{\Sigma}_k^{22} \end{bmatrix}, \quad (58)$$

$$\mathbf{\Sigma}_k^{11} = \mathbf{Q}_k - \mathbf{F}_k^2\mathbf{R}_{k-1}(\mathbf{F}_k^2)^T, \quad (59)$$

$$\mathbf{\Sigma}_k^{21} = \mathbf{H}_k\mathbf{Q}_k - \mathbf{H}_k^2\mathbf{R}_{k-1}(\mathbf{F}_k^2)^T, \quad (60)$$

$$\mathbf{\Sigma}_k^{22} = \mathbf{R}_k - \mathbf{H}_k^2\mathbf{R}_{k-1}(\mathbf{H}_k^2)^T + \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k, \quad (61)$$

and where $\mathbf{F}_k^2$ and $\mathbf{H}_k^2$ can be arbitrarily chosen provided $\mathbf{\Sigma}_\mathbf{k}$ is positive definite.

Let us comment this result. Setting $\mathbf{F}_k^2 = \mathbf{0}$ and $\mathbf{H}_k^2 = \mathbf{0}$ we get model (4)-(5) again, which is not surprising, since the classical linear and Gaussian HMC model described by $f_{k|k-1}$ and $g_{k|k}$ is of course one particular element out of this general class of PMC models. Next, if we set $\mathbf{H}_k^2 = \mathbf{H}_k\mathbf{F}_k^2$ then $p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_{k-1}) = g_k(\mathbf{y}_k|\mathbf{x}_k)$ and we are in a situation where **P.2** and **P.3** are satisfied, but not **P.1**, i.e. $\mathbf{y}_k$ is independent of $\mathbf{x}_{k-1}$ and $\mathbf{y}_{k-1}$ given $\mathbf{x}_k$; if furthermore we set $\mathbf{F}_k^2 = \mathbf{0}$ then $\mathbf{x}_k$ becomes independent of $\mathbf{y}_{k-1}$ given $\mathbf{x}_{k-1}$, so **P.1** is satisfied.

D. Performance Analysis

Let us now study the benefit of a PMC-PHD approach when indeed data no longer follow a classical HMC model. In order to compare the algorithms, we compute at each time step the Optimal Subpattern Assignment (OSPA) metric [34] which takes into account the number of estimated targets and the estimated states, and compare them with the true set of targets. Let $X = \{\mathbf{x}_1, ..., \mathbf{x}_m\}$ and $\widehat{X} = \{\widehat{\mathbf{x}}_1, ..., \widehat{\mathbf{x}}_n\}$ be two finite sets. Here, $\widehat{X}$ represents the estimated finite set of targets and X represents the true finite set of targets. For $1 \leq p < +\infty$ and $c > 0$, let $d^{(c)}(\mathbf{x}, \widehat{\mathbf{x}}) = \min(c, ||\mathbf{x} - \widehat{\mathbf{x}}||)$ ($||.||$ is the Euclidean norm), Π_n be the set of permutations on $\{1, 2, ..., n\}$, and let $\pi(i)$ be the i-th component of a given permutation π. The OSPA metric is defined by:

$$\overline{d}_p^c(X, \widehat{X}) \triangleq \left(\frac{1}{n} \left(\min_{\pi \in \Pi_n} \sum_{i=1}^{m} d^{(c)}(\mathbf{x}_i, \widehat{\mathbf{x}}_{\pi(i)})^p + c^p(n-m) \right) \right)^{\frac{1}{p}} \quad (62)$$

if $m \leq n$ and $\overline{d}_p^c(X, \widehat{X}) \triangleq \overline{d}_p^c(\widehat{X}, X)$ if $m > n$. In our simulations, we set $p = 20$ and $c = 1$. Let also

$$\mathbf{F}_k = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{H}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad (63)$$

$$\mathbf{Q}_k = \begin{bmatrix} 100 & 1 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 0 & 0 & 100 & 1 \\ 0 & 0 & 1 & 10 \end{bmatrix}, \mathbf{R}_k = \begin{bmatrix} 10^2 & 0 \\ 0 & 10^2 \end{bmatrix}, \quad (64)$$

and let us generate trajectories from a model of Proposition 1 where

$$\mathbf{F}_k^2 = \begin{bmatrix} a & 0 \\ 0 & 0 \\ 0 & b \\ 0 & 0 \end{bmatrix}, \mathbf{H}_k^2 = \begin{bmatrix} c & 0 \\ 0 & d \end{bmatrix}. \quad (65)$$

Remember that in these models, the pdfs $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ of $\mathbf{x}_k$ given $\mathbf{x}_{k-1}$, and $p(\mathbf{y}_k|\mathbf{x}_k)$ of $\mathbf{y}_k$ given $\mathbf{x}_k$, respectively coincide with $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1}; \mathbf{Q}_k)$ and $g_k(\mathbf{y}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{H}_k\mathbf{x}_k; \mathbf{R}_k)$. So we actually compare a PMC-PHD filter based on model (56)-(61) to an HMC-PHD one based on $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $g_k(\mathbf{y}_k|\mathbf{x}_k)$, and in this case both algorithms rely on models which locally share the same physical properties.

We consider a scenario of 6 targets: targets 1 and 2 born at $k = 1$, targets 3 and 4 born at $k = 20$, and targets 5 and 6 born at $k = 50$. We track the positions and velocities of targets in Cartesian coordinates, $\mathbf{x}_k = [p_{x,k}, \dot{p}_{x,k}, p_{y,k}, \dot{p}_{y,k}]^T$.

We generate uniformly a mean of 20 clutter measurements on region $V = [-2000, 2000] \times [-2000, 2000]$, so the clutter intensity is taken as $\kappa(\mathbf{z}) = 2.5 \times 10^{-6} \times \mathbf{1}_V(\mathbf{z})$, where $\mathbf{1}_V(\mathbf{z})$ is the indicator function of V. We also set $p_{s,k} = 0.98$ and $p_{d,k} = 0.9$. A realization of such a scenario is displayed in Fig. 1.

For the GM-HMC implementation, the birth intensity is chosen as $\gamma_k(\mathbf{x}) = 0.01 \sum_{i=1}^{6} \mathcal{N}(\mathbf{x}; \mathbf{m}_{\gamma_k}^{(i)}; \mathbf{P}_{\gamma_k})$, where $\mathbf{m}_{\gamma_k}^{(i)}$ correspond to areas where target may appear (e.g. $\mathbf{m}_{\gamma_k}^{(1)} = [200, 0, 500, 0]^T$) and $\mathbf{P}_{\gamma_k} = \text{diag}(1000, 400, 1000, 400)$; for the GM-PMC implementation it is chosen as

$$\gamma_k(\boldsymbol{\xi}) = 0.01 \sum_{i=1}^{6} \mathcal{N}\left(\boldsymbol{\xi}; \begin{bmatrix} \mathbf{m}_{\gamma_k}^{(i)} \\ \mathbf{H}_k\mathbf{m}_{\gamma_k}^{(i)} \end{bmatrix}; \begin{bmatrix} \mathbf{P}_{\gamma_k} & (\mathbf{H}_k\mathbf{P}_{\gamma_k})^T \\ \mathbf{H}_k\mathbf{P}_{\gamma_k} & \mathbf{R}_k\mathbf{H}_k\mathbf{P}_{\gamma_k}\mathbf{H}_k^T \end{bmatrix} \right).$$

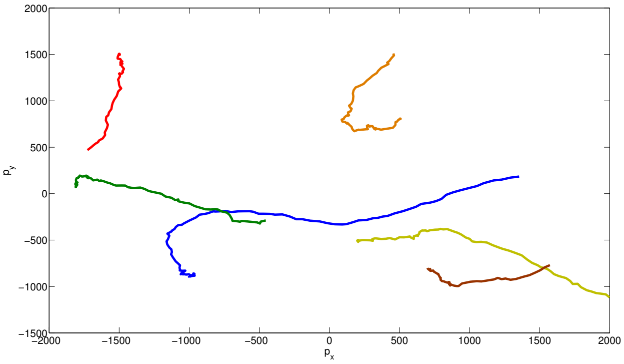

Fig. 1. Typical Scenario - 6 targets evolve according to a linear and Gaussian PMC model

The pruning and merging thresholds are respectively $T_p = 10^{-5}$ and $T_m = 5$ meters, we keep at most $J_{\max} = 100$ Gaussians for the GM-HMC implementation, and we set $J_{\max}^1 = J_{\max}^2 = 50$ for the GM-PMC one. A Gaussian is considered as a target as soon as its weight is above 0.5. Other values of these parameters do not affect the results.

We set $a = b = 0.7$ and $c = d = 0.1$ in (65) and we compute the OSPA distance averaged over 200 runs. Results are displayed in Fig. 2. We observe that the PMC PHD filter outperforms the HMC one although both algorithms share the same pdfs $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $p(\mathbf{y}_k|\mathbf{x}_k)$. However, the usual independence assumptions are no longer satisfied; the statistical properties of PMC models are not taken into account by the GM-HMC implementation. Because a b, c and d are different from 0, properties **P.1**-**P.2** are not satisfied. In particular, pdf

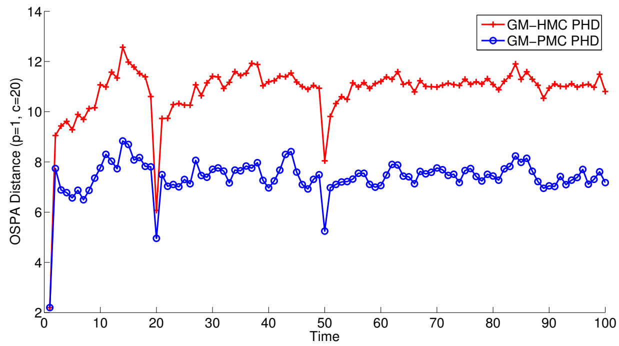

Fig. 2. OSPA distance with $p = 1$ and $c = 20$. Averaged over time, the OSPA distance is 7.3837 for the PMC-PHD filter and 10.9143 for the HMC-PHD one.

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}_k^1\mathbf{x}_{k-1} + \mathbf{F}_k^2\mathbf{y}_{k-1}; \boldsymbol{\Sigma}_k^{11}),$$

$$\boldsymbol{\Sigma}_k^{11} = \begin{bmatrix} 49 & 1 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 0 & 0 & 49 & 1 \\ 0 & 0 & 1 & 10 \end{bmatrix} \quad (66)$$

is approximated in the HMC formulation by $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1}; \mathbf{Q}_k)$, where $\mathbf{Q}_k$ is given in (64). So the information brought by observation

$\mathbf{y}_{k-1}$ is not taken into account in the HMC implementation, and the incertitude on state $\mathbf{x}_k$, given $\mathbf{x}_{k-1}$ and $\mathbf{y}_{k-1}$, increases. Such an approximation has consequences on the means and covariances of the GM, which will affect the estimates of the states, but also on the weights of the GM, which will affect the extracted number of targets. Indeed, the localization error (first term of r.h.s of (62)) averaged over time, is approximately 3.14 for the GM-PMC implementation, and 5.68 for the GM-HMC one; the cardinality error (second term of r.h.s of (62)) which, averaged over time, is 4.23 for the GM-PMC implementation and 5.12 for the GM-HMC one.

The estimated total mass of the PHD is well estimated for both implementations but that of the HMC one becomes less reliable when time increases, as is shown in Fig. 3 where we have displayed the standard deviation of the estimated total mass of the PHD for the GM-PMC implementation, normalized w.r.t. that of the GM-HMC one.

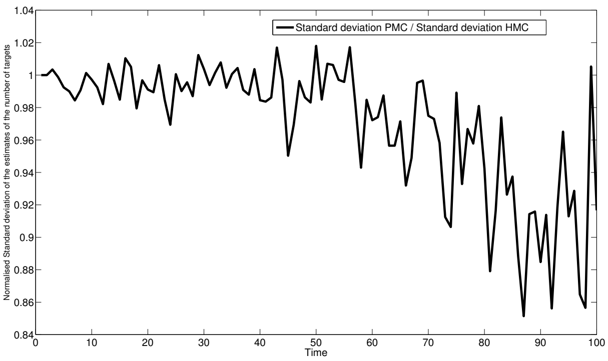

Fig. 3. Normalized Standard deviation of the estimate of the number of targets in the GM-PMC implementation of the PHD filter, w.r.t. that of the GM-HMC one.

The computational time is almost the same for both methods (approximately 0.65 second per iteration), which is not surprising since both filters only differ via terms $\mathbf{F}_k^2$ and $\mathbf{H}_k^2$ which are null in the HMC case. So the performances improvement requires a negligible extra computational cost.

We have shown that the HMC-PHD filter is not adapted for such a scenario although both implementations share the same pdfs $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $p(\mathbf{y}_k|\mathbf{x}_k)$. A common trick to get around modeling mismatches in the HMC context is to tune the process noise of the motion model. So for the same scenario, we compute the HMC-PHD filter by replacing $\mathbf{Q}$ in (64) by $\mathbf{\Sigma}_k^{11}$ in (66) and by $1.5\mathbf{Q}$. The time averaged OSPA distance are now respectively 10.9785 and 10.9092; so the trick does not improve the performances of the HMC-PHD filter. This is because the physical properties of the new HMC model, described by $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $g_k(\mathbf{y}_k|\mathbf{x}_k)$ no longer coincide with the true ones, described by $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $p(\mathbf{y}_k|\mathbf{x}_k)$.

V. Conclusion

In this paper we showed that the strong structural properties involved by the classical HMC model assumption in the multi-target filtering problem could indeed be relaxed. We computed a PHD filter adapted to PMC models, which is a class of stochastic models which embeds and generalizes HMC. We proposed a GM implementation of our PMC PHD filter for general linear and Gaussian PMC models, and we validated our results via simulations, in a scenario where we kept the local physical properties of linear and Gaussian HMC models, while relaxing their statistical constraints.

Appendix A
Conditioning among Gaussian variables

We recall in this section two classical results on Gaussian pdfs (see e.g. [35]).

Lemma 1 *Let* $\boldsymbol{\zeta} \in \mathbb{R}^m$, $\boldsymbol{\eta} \in \mathbb{R}^q$, $\mathbf{Q}$ *(resp.* $\mathbf{P}$*) be a* $m \times m$ *(resp.* $q \times q$*) positive definite matrix (other vectors and matrices are of appropriate dimensions), then*

$$\int \mathcal{N}(\boldsymbol{\zeta};\mathbf{F}\boldsymbol{\eta}+\mathbf{d};\mathbf{Q})\mathcal{N}(\boldsymbol{\eta};\mathbf{m};\mathbf{P})\mathrm{d}\boldsymbol{\eta} = \mathcal{N}(\boldsymbol{\zeta};\mathbf{F}\mathbf{m}+\mathbf{d};\mathbf{Q}+\mathbf{F}\mathbf{P}\mathbf{F}^T).$$

Lemma 2 *Let* $\boldsymbol{\zeta} \in \mathbb{R}^m$, $\boldsymbol{\eta} \in \mathbb{R}^q$, $\mathbf{P}^{\boldsymbol{\zeta}}$ *(resp.* $\mathbf{P}^{\boldsymbol{\eta}}$*) be a* $m \times m$ *(resp.* $q \times q$*) matrix and* $\mathbf{P}^{\boldsymbol{\zeta},\boldsymbol{\eta}}$ *a* $m \times q$ *matrix. Let us assume that pdf of* $(\boldsymbol{\zeta},\boldsymbol{\eta})$ *is a Gaussian,*

$$p(\boldsymbol{\zeta},\boldsymbol{\eta}) = \mathcal{N}(\boldsymbol{\zeta},\boldsymbol{\eta};\begin{bmatrix}\mathbf{m}^{\boldsymbol{\zeta}}\\ \mathbf{m}^{\boldsymbol{\eta}}\end{bmatrix};\begin{bmatrix}\mathbf{P}^{\boldsymbol{\zeta}} & \mathbf{P}^{\boldsymbol{\zeta},\boldsymbol{\eta}}\\ \mathbf{P}^{\boldsymbol{\zeta},\boldsymbol{\eta}^T} & \mathbf{P}^{\boldsymbol{\eta}}\end{bmatrix}).$$

Then

$$\begin{aligned} p(\boldsymbol{\zeta},\boldsymbol{\eta}) &= \mathcal{N}(\boldsymbol{\eta};\mathbf{m}^{\boldsymbol{\eta}};\mathbf{P}^{\boldsymbol{\eta}})\mathcal{N}(\boldsymbol{\zeta};\tilde{\mathbf{m}}^{\boldsymbol{\zeta}}(\boldsymbol{\eta});\tilde{\mathbf{P}}^{\boldsymbol{\zeta}}),\\ \tilde{\mathbf{m}}^{\boldsymbol{\zeta}}(\boldsymbol{\eta}) &= \mathbf{m}^{\boldsymbol{\zeta}} + \mathbf{P}^{\boldsymbol{\zeta},\boldsymbol{\eta}}(\mathbf{P}^{\boldsymbol{\eta}})^{-1}(\boldsymbol{\eta}-\mathbf{m}^{\boldsymbol{\eta}}),\\ \tilde{\mathbf{P}}^{\boldsymbol{\zeta}} &= \mathbf{P}^{\boldsymbol{\zeta}} - \mathbf{P}^{\boldsymbol{\zeta},\boldsymbol{\eta}}(\mathbf{P}^{\boldsymbol{\eta}})^{-1}\mathbf{P}^{\boldsymbol{\zeta},\boldsymbol{\eta}^T}. \end{aligned}$$

Appendix B
Proof of Proposition 1

We use an induction argument. So let us assume that at time k, $p(\mathbf{y}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k;\mathbf{H}_k\mathbf{x}_k;\mathbf{R}_k)$. Then the pdf of couple $\boldsymbol{\xi}_k = (\mathbf{x}_k,\mathbf{y}_k)$ can be written as

$$p(\boldsymbol{\xi}_k) = \mathcal{N}(\boldsymbol{\xi}_k;\begin{bmatrix}\mathbf{m}_k^{\mathbf{x}}\\ \mathbf{H}_k\mathbf{m}_k^{\mathbf{x}}\end{bmatrix};\begin{bmatrix}\mathbf{P}_k^{\mathbf{x}} & (\mathbf{H}_k\mathbf{P}_k^{\mathbf{x}})^T\\ \mathbf{H}_k\mathbf{P}_k^{\mathbf{x}} & \mathbf{R}_k+\mathbf{H}_k\mathbf{P}_k^{\mathbf{x}}\mathbf{H}_k^T\end{bmatrix}).$$

Next, according to (9) and (10) $p(\mathbf{x}_{k+1}|\mathbf{x}_k,\mathbf{y}_k) = \mathcal{N}(\mathbf{x}_{k+1};\mathbf{F}_{k+1}^1\mathbf{x}_k + \mathbf{F}_{k+1}^2\mathbf{y}_k;\mathbf{\Sigma}_{k+1}^{11})$. Using Lemma 1 (see Appendix A), $p(\mathbf{x}_{k+1}|\mathbf{x}_k) = \int p(\mathbf{x}_{k+1}|\mathbf{x}_k,\mathbf{y}_k)p(\mathbf{y}_k|\mathbf{x}_k)\mathrm{d}\mathbf{y}_k$ is computable in function of $\mathbf{F}_{k+1}^1$, $\mathbf{F}_{k+1}^2$, $\mathbf{\Sigma}_{k+1}^{11}$, $\mathbf{H}_k$ and $\mathbf{R}_k$, and condition $p(\mathbf{x}_k|\mathbf{x}_{k-1}) = f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ implies that

$$\begin{aligned} \mathbf{F}_{k+1}^1 + \mathbf{F}_{k+1}^2\mathbf{H}_k &= \mathbf{F}_{k+1},\\ \mathbf{\Sigma}_{k+1}^{11} + \mathbf{F}_{k+1}^2\mathbf{R}_k(\mathbf{F}_{k+1}^2)^T &= \mathbf{Q}_{k+1}. \end{aligned}$$

Next, using again Lemma 1 we compute the parameters of Gaussian $p(\mathbf{x}_{k+1},\mathbf{y}_{k+1}) = \int p(\boldsymbol{\xi}_{k+1}|\boldsymbol{\xi}_k)p(\boldsymbol{\xi}_k)\mathrm{d}\boldsymbol{\xi}_k$, and from Lemma 2 we compute $p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1})$ in function of the parameters of the PMC model. Using the fact that assumption $p(\mathbf{y}_k|\mathbf{x}_k) = g_{k|k-1}(\mathbf{y}_k|\mathbf{x}_k)$ is valid whatever $\mathbf{m}_0$, we get the following conditions on $\mathbf{H}_{k+1}^1$, $\mathbf{\Sigma}_{k+1}^{12}$ and $\mathbf{\Sigma}_{k+1}^{22}$:

$$\begin{aligned} \mathbf{H}_{k+1}^1 &= \mathbf{H}_{k+1}\mathbf{F}_{k+1} - \mathbf{H}_{k+1}^2\mathbf{H}_k,\\ \mathbf{\Sigma}_{k+1}^{12} &= \mathbf{H}_{k+1}\mathbf{Q}_{k+1} - \mathbf{H}_{k+1}^2\mathbf{R}_k(\mathbf{F}_{k+1}^2)^T,\\ \mathbf{\Sigma}_{k+1}^{22} &= \mathbf{R}_{k+1} - \mathbf{H}_{k+1}^2\mathbf{R}_k(\mathbf{H}_{k+1}^2)^T + \mathbf{H}_{k+1}\mathbf{Q}_{k+1}\mathbf{H}_{k+1}^T. \end{aligned}$$

References

[1] Y. Bar-Shalom, *Tracking and data association*. San Diego, CA, USA: Academic Press Professional, Inc., 1987.

[2] L. Svensson, D. Svensson, M. Guerriero, and P. Willett, "Set JPDA filter for Multitarget Tracking." *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4677–4691, 2011.

[3] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 2009.

[4] T. Long, L. Zheng, X. Chen, Y. Li, and T. Zeng, "Improved probabilistic multi-hypothesis tracker for multiple target tracking with switching attribute states." *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5721–5733, 2011.

[5] J. Goutsias, R. P. Mahler, and H. T. Nguyen, *Random Sets Theory and Applications*. Springer- Verlag New York, 1997.

[6] I. R. Goodman, R. P. Mahler, and H. T. Nguyen, *Mathematics of Data Fusion*. Norwell, MA, USA: Kluwer Academic Publishers, 1997.

[7] R. Mahler, "Multitarget Bayes Filtering via First-Order Multitarget Moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, October 2003.

[8] R.Mahler, *Statistical Multisource Multitarget Information Fusion*. Artech House, 2007.

[9] D. Clark, I. T. Ruiz, Y. Petillot, and J. Bell, "Particle PHD filter multiple target tracking in sonar image," *IEEE Trans. Aerospace and Electronic Systems*, vol. 43, no. 1, pp. 409–416, 2007.

[10] C. Lundquist, L. Hammarstrand, and F. Gustafsson, "Road intensity based mapping using radar measurements with a probability hypothesis density filter," *IEEE Transactions on Signal Processing*, no. 4, pp. 1397–1408, 2011.

[11] B.-N. Vo and W. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4091–4104, November 2006.

[12] D. Clark and B.-N. Vo, "Convergence Analysis of the Gaussian Mixture PHD Filter," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1204–1212, 2007.

[13] S. A. Pasha, B.-N. Vo, H. D. Tuan, and W.-K. Ma, "A Gaussian Mixture PHD Filter for Jump Markov System Models," *IEEE Trans. Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 919–936, 2009.

[14] 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, pp. 1224–1245, 2005.

[15] D. Clark and J. Bell, "Convergence results for the Particle-PHD Filter," *IEEE Transactions on Signal Processing*, vol. 54, 2006.

[16] N. Whiteley, S. Singh, and S. Godsill, "Auxiliary Particle Implementation of the Probability Hypothesis Density Filter," *IEEE Trans. Aerospace and Electronic Systems*, vol. 46, pp. 1437–1454, July 2010.

[17] R.Mahler, "A theory of PHD filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic Systems* , vol. 43, no. 4, pp. 1523–1543, October 2007.

[18] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic Implementations of the Cardinalized Probability Hypothesis Density Filter," *IEEE Transactions on Signal Processing* , vol. 55, no. 7, pp. 3553–3367, July 2007.

[19] R. P. S. Mahler, B.-T. Vo, and B.-N. Vo, "CPHD filtering with unknown clutter rate and detection profile." *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3497–3513, 2011.

[20] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. New York: Springer Verlag, 2001.

[21] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear / Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, February 2002.

[22] S. Saha and F. Gustafsson, "Particle filtering with dependent noise processes," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4497–4508, September 2012.

[23] W. Pieczynski, "Pairwise Markov chains," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 634–39, May 2003.

[24] S. Derrode and W. Pieczynski, "Signal and image segmentation using pairwise Markov chains," *IEEE Transactions on Signal Processing*, vol. 52, no. 9, pp. 2477–89, 2004.

[25] S. Julier and J. Uhlmann, "Unscented Filtering and Nonlinear Estimation," in *Proceedings of the IEEE*, vol. 92, March 2004, pp. 401–422.

[26] 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.

[27] F. Desbouvries and W. Pieczynski, "Particle filtering in pairwise and triplet Markov chains," in *Proc. IEEE - EURASIP Workshop on Nonlinear Signal and Image Processing*, Grado-Gorizia, Italy, June 8-11 2003.

[28] P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski, "Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise," *Signal Processing*, vol. 91, no. 2, pp. 163–175, February 2011.

[29] R. S. Lipster and A. N. Shiryaev, *Statistics of Random Processes, Vol. 2: Applications*. Berlin: Springer Verlag, 2001, ch. 13 : "Conditionally Gaussian Sequences : Filtering and Related Problems".

[30] W. Pieczynski and F. Desbouvries, "Kalman filtering using pairwise Gaussian models," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 03)*, Hong-Kong, 2003.

[31] O. Erdinc, P. Willett, and Y. Bar-Shalom, "The bin-occupancy filter and its connection to the PHD filters," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4232–4246, Nov. 2009.

[32] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New-York: John Wiley and sons, 2001.

[33] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Science and Engineering. San Diego: Academic Press, 1970, vol. 64.

[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.

[35] Rao, *Linear statistical inference and its applications*, 2nd ed., ser. Wiley series in probability and mathematical statistics. New York: Wiley, 1973.