Sequential Monte Carlo methods for Multi-target Filtering with Random Finite Sets

Ba-Ngu Vo, Sumeetpal Singh, and Arnaud Doucet

Abstract— Random finite sets are natural representations of multi-target states and observations that allow multi-sensor multi-target filtering to fit in the unifying random set framework for Data Fusion. Although the foundation has been established in the form of Finite Set Statistics (FISST), its relationship to conventional probability is not clear. Furthermore, optimal Bayesian multi-target filtering is not yet practical due to the inherent computational hurdle. Even the Probability Hypothesis Density (PHD) filter, which propagates only the first moment (or PHD) instead of the full multi-target posterior, still involves multiple integrals with no closed forms in general. This article establishes the relationship between FISST and conventional probability that leads to the development of a sequential Monte Carlo (SMC) multi-target filter. In addition, a SMC implementation of the PHD filter is proposed and demonstrated on a number of simulated scenarios. Both of the proposed filters are suitable for problems involving non-linear non-Gaussian dynamics. Convergence results for these filters are also established.

Index Terms— Multi-target Tracking, Optimal Filtering, Particle Filter, Point Processes, Random Sets, Sequential Monte Carlo.

I. INTRODUCTION

Multi-target filtering is a class of dynamic state estimation problems in which the entity of interest is a finite set that is random in the number of elements as well as the values of individual elements [4]–[6]. Random finite sets are therefore natural representations of multi-target states and multi-target measurements. The modelling of multi-target dynamics using random sets naturally leads to algorithms which incorporate track initiation and termination, a procedure that has mostly been performed separately in traditional tracking algorithms. More importantly, random sets provide a rigorous unified framework for the seemingly unconnected sub-disciplines of data fusion [15], [17], [25].

Although stochastic geometrical models, including deformable templates and random finite sets (or simple finite point processes) have long been used by statisticians to develop techniques for object recognition in static images [2], their use has been largely overlooked in the data fusion and tracking literature until recently [24]. The earliest published work using a point process formalism for multi-target filtering

B. Vo is with the Electrical Engineering Department the University of Melbourne, Melbourne, Vic. 3010, Australia. (email: bnvo@unimelb.edu.au).
S. Singh is with the Department of Engineering, Cambridge University, CB2 1PZ Cambride, U.K. (email: sss40@eng.cam.ac.uk).
A. Doucet is with the Department of Engineering, Cambridge University, CB2 1PZ Cambride, U.K. (email: ad2@eng.cam.ac.uk).

appears to be [35]. A point process based filter was also proposed in [47] to estimate an unknown but fixed number of targets. In [32], [33], [41], a jump process was combined with stochastic diffusion equations on a non-Euclidean manifold to track a time varying number of targets. The same problem with continuous state evolution and marked-point process observations was considered in [37]. However, the first systematic treatment of multi-sensor multi-target filtering, as part of a unified framework for data fusion using random set theory was finite set statistics (FISST) [15], [17], [24], [25], [27], [28]. The key to a rigorous formulation of multi-target estimation as a Bayesian filtering problem is to conceptually view the target set as a single meta-target and the set of observations collected by the sensor as a single meta-observation. FISST provides a set of mathematical tools that allows direct application of Bayesian inferencing to multi-target problems.

From a theoretical standpoint, central FISST concepts such as set integral and set derivative are not conventional probabilistic concepts. Since measure theoretic probability is the foundation for Bayesian filtering of any kind, it is important to understand its connection to FISST. This article establishes the relationship between FISST and conventional probability. In particular, it is shown that a unitless set derivative of a belief mass function is a probability density, and that a set integral is closely related to the conventional (measure theoretic) integral.

From an implementation view point, analogous to single-target filtering, Bayes multi-target filtering propagates the multi-target posterior density recursively in time [15], [25], [27], [28]. This involves the evaluation of multiple set-integrals (see Section II-C) and the computational intractability is far more severe than its single-target counterpart. A more tractable alternative to optimal multi-target filtering is the Probability Hypothesis Density (PHD) filter [26], [27], [29]. It is a recursion propagating the 1st moment, called the intensity function or PHD, associated with the multi-target posterior. Since the domain of the intensity function is the space where individual targets live, its propagation requires much less computational power than the multi-target posterior. Unfortunately, this still involves multiple integrals that have no closed form expressions in general. This article proposes Sequential Monte Carlo (SMC) implementations for both the Bayes multi-target filter and the PHD filter together with convergence results.

SMC methods are powerful tools in Bayesian filtering [1], [10], [11], [16] and have been applied to multi-target problems [3], [12], [21], [22], [30]. However, no principled SMC implementations have been proposed in the context of FISST. The relationship between FISST and conventional probability established in this paper leads to a principled SMC

implementation of the Bayes multi-target filter, for which the approximation error is inversely proportional to the number of samples. The multi-target posterior is represented by a large set of weighted random samples (particles), possibly of different dimensions, and are propagated over time using the importance sampling and resampling strategy of the standard particle filter. This algorithm was first described in [46]. Around the same time, it was proposed in [39] to truncate the set integral and apply standard SMC to each of the ordinary integrals in the truncated expression [39]. However, convergence analysis for this approach was not available. Although SMC implementations of the Bayes multi-target filter are computationally tractable, they are still expensive, especially when the number of targets is large. Thus, it is important to search for computationally cheaper alternatives.

The PHD filter is a cheaper alternative. However, direct application of standard SMC methods to propagate the intensity function would fail because firstly, the intensity function is not a probability density function; and secondly, the PHD recursion is not a standard Bayes recursion. In this paper, a particle interpretation of the PHD recursion is given, which allows a SMC implementation of the PHD filter. The intensity function or PHD is represented by a large set of weighted random samples (of the same dimension) which are propagated over time using a generalised importance sampling and resampling strategy. It is also shown that the approximation error vanishes as the number of particles increases. The proposed algorithm is general enough to cover non-linear non-Gaussian dynamics. Moreover, the number of particles can be adapted to maintain a constant ratio of particles to expected number of targets. This approach first appeared in [46] around the same time as two other independent works [40] and [48]. In [40], only the special case without clutter for ground target filtering was considered. On the other hand, [48] describes an implementation for the special case with neither birth nor spawning.

The rest of the paper is organised as follows. Section II describes the connection between the FISST formulation and the conventional probabilistic formulation of the multi-target filtering problem. In addition, the particle multi-target filter and convergence analysis are also presented. Section III reviews the PHD filter and describes a particle implementation of the PHD recursion with convergence analysis. Simulation results are presented in Section IV. Finally, some conclusions and potential extensions are discussed in Section V. The necessary probability background for the convergence analysis of the proposed filters is given in Appendix A. For completeness, the basics of random finite sets are given in Appendix B and mathematical proofs are given in Appendix C.

II. RANDOM FINITE SET AND BAYES MULTI-TARGET FILTERING

This section provides a discussion on the relationships between FISST and conventional (measure theoretic) probability which leads to a principled SMC implementation of the Bayes multi-target filter. The Random Finite Set (RFS) model for multi-target filtering is first described in subsection II-A. A measure theoretic Bayes formulation (using conventional probability theory) is then outlined in subsection II-B. Subsection II-C summarizes central ideas in the FISST formulation, followed by a discussion of its relationship with the measure theoretic formulation in subsection II-D. Finally, subsection II-E presents a SMC implementation of the Bayes multi-target filter. Readers who are only interested in SMC implementations can skip subsections II-B, II-C, II-D without loss of continuity.

A. Random Finite Set Model

In a single-target system, the state and measurement at time k are two vectors of possibly different dimensions. These vectors evolve in time, but their dimensions are fixed. However, this is not the case in a multi-target system, where the multi-target state and multi-target measurement are two collections of individual targets and measurements. As the multi-target state and multi-target measurement evolve in time, the number of individual targets and measurements may change, i.e. the dimensions of the multi-target state and multi-target measurement also evolve in time. Moreover, there is no ordering for the elements of the multi-target state and measurement.

The multi-target state and multi-target measurement at time k are naturally represented as finite subsets X_k and Z_k respectively. For example, if at time k there are $M(k)$ targets located at $x_{k,1}, \ldots, x_{k,M(k)}$ in the single-target state space E_s (e.g. $\mathbf{R}^{n_x}$) then,

$$X_k = \{x_{k,1}, \ldots, x_{k,M(k)}\} \in \mathcal{F}(E_s)$$

is the multi-target state, where $\mathcal{F}(E)$ denotes the collection of all finite subsets of the space E. Similarly, if $N(k)$ observations $z_{k,1}, \ldots, z_{k,N(k)}$ in the single-target observation space E_o (e.g. $\mathbf{R}^{n_z}$) are received at time k, then

$$Z_k = \{z_{k,1}, \ldots, z_{k,N(k)}\} \in \mathcal{F}(E_o)$$

is the multi-target measurement, in which some of the $N(k)$ observations may be due to clutter.

Analogous to single target systems, where uncertainty is characterised by modelling the states and measurements by random vectors, uncertainty in a multi-target system is characterised by modelling multi-target states and multi-target measurements as random finite sets (RFS) Ξ_k and Σ_k on the (single-target) state and observation spaces E_s and E_o respectively. (A formal definition of a RFS is given in Appendix B).

The multi-target dynamics and observation can be described as follows. Given a realisation X_{k-1} of the RFS Ξ_{k-1} at time $k-1$, the multi-target state at time k is modelled by the RFS

$$\Xi_k = S_k(X_{k-1}) \cup N_k(X_{k-1}) \tag{1}$$

where $S_k(X_{k-1})$ denotes the RFS of targets that have survived at time k, $N_k(X_{k-1})$ is the RFS of new targets comprising of the RFS $B_k(X_{k-1})$ of targets spawned from X_{k-1} and the RFS Γ_k of targets that appear spontaneously at time k. i.e.

$$N_k(X_{k-1}) = B_k(X_{k-1}) \cup \Gamma_k.$$

Details on how $S_k(X_{k-1})$, $B_k(X_{k-1})$ and Γ_k can be determined from models of individual target dynamics, target births and deaths, for various assumptions, are available in [15], [25], [28]. The RFS Ξ_k encapsulates all aspects of multi-target motion such as the time-varying number of targets, individual target motion, target birth, spawning and target interactions. Similarly, given a realisation X_k of Ξ_k at time k, the multi-target observation is modelled by the RFS

$$\Sigma_k = \Theta_k(X_k) \cup C_k(X_k) \qquad (2)$$

where $\Theta_k(X_k)$ denotes the RFS of measurements generated by X_k, and $C_k(X_k)$ denotes the RFS of clutter or false alarms. The reader is referred to [15], [25], [28] for details on determining $\Theta_k(X_k)$ and $C_k(X_k)$ from the underlying physical models of the sensors. The RFS Σ_k encapsulates all sensor characteristics such as measurement noise, sensor field of view (i.e. state-dependent probability of detection) and false alarms.

The multi-target filtering problem concerns the estimation of the multi-target state X_k at time step k given the collection $Z_{1:k} \equiv (Z_1, ..., Z_k)$ of all multi-target observations up to time k.

B. Measure theoretic formulation

This subsection outlines a measure theoretic Bayesian formulation of the multi-target filtering problem. The object of interest in Bayesian estimation is the posterior probability density. Hence, the application of Bayesian reasoning to multi-target estimation hinges on a suitable notion of probability density for RFS.

The probability density p_Ξ of a RFS Ξ is given by the Radon-Nikodým derivative of the probability distribution P_Ξ with respect to an appropriate dominating measure μ, i.e. $P_\Xi(\mathcal{T}) = \int_\mathcal{T} p_\Xi(X)\mu(dX)$[1], for any Borel subset $\mathcal{T} \subseteq \mathcal{F}(E)$. Suppose that volume in the space E is measured in units of K. Then, one such μ is the unnormalised distribution of a Poisson point process with a uniform rate of K^{-1}(see also Appendix B)

$$\mu(\mathcal{T}) = \sum_{i=0}^{\infty} \frac{\lambda^i(\chi^{-1}(\mathcal{T}) \cap E^i)}{i!}, \qquad (3)$$

where λ^i is the ith product (unitless) Lebesque measure, and $\chi : \uplus_{i=0}^{\infty} E^i \to \mathcal{F}(E)$ is a mapping of vectors to sets defined by $\chi([x_1, ..., x_i]^T) = \{x_1, ..., x_i\}$. The measure (3) is commonly used in point process theory as a dominating measure [13], [34]. The integral of a non-negative function $f : \mathcal{F}(E) \to [0, \infty)$ with respect to μ is given by (see also Appendix B)

$$\int_\mathcal{T} f(X)\mu(dX)$$
$$= \sum_{i=0}^{\infty} \frac{1}{i!} \int_{\chi^{-1}(\mathcal{T}) \cap E^i} f(\{x_1, ..., x_i\})\lambda^i(dx_1...dx_i). \quad (4)$$

For any Borel subsets $\mathcal{U} \subseteq \mathcal{F}(E_s)$, $\mathcal{V} \subseteq \mathcal{F}(E_o)$ let

$$P_{k|k}(\mathcal{U}|Z_{1:k}) \equiv P(\Xi_k \in \mathcal{U}|Z_{1:k})$$

---

[1] It is implicitly assumed that P_Ξ is absolutely continuous with respect to μ.

denote the (posterior) probability measure of the RFS Ξ_k given all the observations $Z_{1:k} = (Z_1, ..., Z_k)$ up to time k,

$$P_{k|k-1}(\mathcal{U}|X_{k-1}) \equiv P(\Xi_k \in \mathcal{U}|X_{k-1})$$

denote the probability measure of the RFS Ξ_k modelled by (1), and

$$P_k(\mathcal{V}|X_k) \equiv P(\Sigma_k \in \mathcal{V}|X_k)$$

denote the probability measure of the RFS Σ_k modelled by (2). Let μ_s and μ_o be dominating measures of the form (3) on the Borel subsets of $\mathcal{F}(E_s)$ and $\mathcal{F}(E_o)$ respectively. Then, the *multi-target posterior density* $p_{k|k}(\cdot|Z_{1:k})$, *multi-target transition density* $f_{k|k-1}(\cdot|X_{k-1})$ and *multi-target likelihood* $g_k(\cdot|X_k)$ are the Radon-Nikodým derivatives of $P_{k|k}(\cdot|Z_{1:k})$ w.r.t. μ_s, $P_{k|k-1}(\cdot|X_{k-1})$ w.r.t. μ_s, and $P_k(\cdot|X_k)$ w.r.t. μ_o respectively i.e.

$$
\begin{aligned}
P_{k|k}(\mathcal{U}|Z_{1:k}) &= \int_\mathcal{U} p_{k|k}(X_k|Z_{1:k})\mu_s(dX_k), \\
P_{k|k-1}(\mathcal{U}|X_{k-1}) &= \int_\mathcal{U} f_{k|k-1}(X_k|X_{k-1})\mu_s(dX_k), \\
P_k(\mathcal{V}|X_k) &= \int_\mathcal{V} g_k(Z_k|X_k)\mu_o(dZ_k).
\end{aligned}
$$

The statistical behaviour of the RFS Ξ_k, modelled by (1), is now characterised by the multi-target transition density $f_{k|k-1}(\cdot|X_{k-1})$ in an analogous fashion to the single-target transition density. Likewise, the statistical behaviour of the RFS Σ_k, modelled by (2), can now be described by the multi-target likelihood $g_k(\cdot|X_k)$ in an analogous fashion to the single-target likelihood function. The multi-target transition density $f_{k|k-1}(\cdot|\cdot)$ incorporates all aspects of motion of multiple targets such as the time-varying number of targets, individual target motion, target birth, spawning and target interactions. The multi-target likelihood $g_k(\cdot|\cdot)$ incorporates all sensor behaviour such as measurement noise, sensor field of view (i.e. state-dependent probability of detection) and clutter models. Here an existence-type definition for the densities $f_{k|k-1}(\cdot|X_{k-1})$ and $g_k(\cdot|X_k)$ has been used. It will be shown in subsections II-C and II-D how such densities can be computed from the dynamical model (1)-(2).

The optimal multi-target Bayes filter is given by the recursion

$$p_{k|k-1}(X_k|Z_{1:k-1})$$
$$= \int f_{k|k-1}(X_k|X)p_{k-1|k-1}(X|Z_{1:k-1})\mu_s(dX) \quad (5)$$
$$p_{k|k}(X_k|Z_{1:k})$$
$$= \frac{g_k(Z_k|X_k)p_{k|k-1}(X_k|Z_{1:k-1})}{\int g_k(Z_k|X)p_{k|k-1}(X|Z_{1:k-1})\mu_s(dX)}. \quad (6)$$

The main difference between the recursion (5)-(6) and standard clutter-free single-target filtering is that X_k and Z_k can change dimension as k changes. In most cases, the recursion (5)-(6) cannot be done analytically. In subsection II-E a Sequential Monte Carlo (SMC) implementation of this recursion is proposed.

C. Finite Set Statistics

This section summarizes key concepts such as set integrals and set derivatives in the finite set statistics (FISST) formulation of the multi-target filtering problem. Individual target motion in a multi-target problem is often modelled by a transition density on the single-target state space E_s while the measurement process is modelled as a likelihood on the single-target observation space E_o. Consequently, it is difficult to construct multi-target transition density and likelihood as Radon-Nikodým derivatives of probability measures on the Borel subsets of $\mathcal{F}(E_s)$ and $\mathcal{F}(E_o)$. FISST, on the other hand, is based on belief mass functions (see Appendix B) defined directly on the closed subsets of E_s and E_o [15], [25]. This allows descriptions of multi-target motion and measurement to be systematically constructed from (1) and (2) respectively (see [25]). However, belief mass functions are non-additive, hence their Radon-Nikodým derivatives (or densities) are not defined. FISST introduces a non-measure theoretic notion of 'density' through set integrals and set derivatives [15].

Let $\mathcal{C}(E)$ denote the collection of closed subsets of E. The set derivative of a function $F : \mathcal{C}(E) \to [0, \infty)$ at a point $x \in E$ is a mapping $(dF)_x : \mathcal{C}(E) \to [0, \infty)$ defined as

$$(dF)_x(S) \equiv \lim_{\lambda_K(\Delta_x) \to 0} \frac{F(S \cup \Delta_x) - F(S)}{\lambda_K(\Delta_x)},$$

where $\lambda_K(\Delta_x)$ is the volume (Lebesgue measure) of a neighbourhood Δ_x of x in units of K (note $\lambda_K = K\lambda$). This is a simplified version of the complete definition given in [15]. Furthermore, the set derivative at a finite set $X = \{x_1, ..., x_n\}$ is defined by the recursion

$$(dF)_{\{x_1,...,x_n\}}(T) \equiv (d(dF)_{\{x_1,...,x_{n-1}\}})_{x_n}(T),$$

where $(dF)_\emptyset \equiv F$ by convention[2]. Note that $(dF)_X(S)$ has unit of $K^{-|X|}$, where $|X|$ denote the cardinality or number of elements of X. Hence, for a fixed $S \subseteq E$ the set derivatives $(dF)_X(S)$ and $(dF)_Y(S)$ have different units if $|X| \neq |Y|$. The set derivative can also be defined as iterated Frechét derivatives of the probability generating functional [27], [29].

Let f be a function defined by $f(X) = (dF)_X(\emptyset)$. Then the set integral of f over a closed subset $S \subseteq E$ is defined as follows [15], [25], [29][3]

$$\int_S f(X)\delta X \equiv \sum_{i=0}^{\infty} \frac{1}{i!} \int_{S^i} f(\{x_1, ..., x_i\})\lambda_K^i(dx_1...dx_i).$$

The set integral and set derivative are related by the following generalised fundamental theorem of calculus,

$$f(X) = (dF)_X(\emptyset) \text{ if and only if } F(S) = \int_S f(X)\delta X,$$

which allows the 'density' of a non-additive set function to be determined constructively.

For any closed subsets $S \subseteq E_s$ and $T \subseteq E_o$, let

$$\beta_{k|k}(S|Z_{1:k}) \equiv P(\Xi_k \subseteq S|Z_{1:k})$$

---

[2]In [15] the notation $\frac{\delta F}{\delta X}(S)$ was used for the set derivative $(dF)_X(S)$.

[3]In [29], [15] pp. 141-142, the set integral is defined for any real or vector valued function f. This is probably a typographical error as it implies that the terms in the sum have different unit of measurement.

denote the (posterior) belief mass function of the RFS Ξ_k given all the observation sets $Z_{1:k} = (Z_1, ..., Z_k)$ up to time k,

$$\beta_{k|k-1}(S|X_{k-1}) \equiv P(\Xi_k \subseteq S|X_{k-1})$$

denote the belief mass function of the RFS Ξ_k modelled by (1), and

$$\beta_k(T|X_k) \equiv P(\Sigma_k \subseteq T|X_k)$$

denote the belief mass function of the RFS Σ_k modelled by (2). Then, the *FISST multi-target posterior density* $\pi_{k|k}(\cdot|Z_{1:k})$, *FISST multi-target transition density* $\varphi_{k|k-1}(\cdot|X_{k-1})$ and *FISST multi-target likelihood* $\rho_k(\cdot|X_k)$ are the set derivatives of $\beta_{k|k}(\cdot|Z_{1:k})$, $\beta_{k|k-1}(\cdot|X_{k-1})$ and $\beta_k(\cdot|X_k)$ respectively. The FISST multi-target Bayes filter proposed in [15], [25], [29] is given by the recursion

$$\pi_{k|k-1}(X_k|Z_{1:k-1})$$
$$= \int \varphi_{k|k-1}(X_k|X)\pi_{k-1|k-1}(X|Z_{1:k-1})\delta X \quad (7)$$
$$\pi_{k|k}(X_k|Z_{1:k})$$
$$= \frac{\rho_k(Z_k|X_k)\pi_{k|k-1}(X_k|Z_{1:k-1})}{\int \rho_k(Z_k|X)\pi_{k|k-1}(X|Z_{1:k-1})\delta X}. \quad (8)$$

Observe the resemblance between the recursions (7)-(8) and (5)-(6). The difference is that the integrals used in (7)-(8) are set integrals and the functions involved have units. In particular, $\pi_{k|k}(X|Z_{1:k})$, $\pi_{k-1|k-1}(X|Z_{1:k-1})$, $\pi_{k|k-1}(X|Z_{1:k-1})$, $\varphi_{k|k-1}(X|X_{k-1})$ have units of $K_s^{-|X|}$, and $\rho_k(Z_k|X_k)$ has units of $K_o^{-|Z_k|}$, whereas the corresponding functions in (5)-(6) namely $p_{k|k}(X|Z_{1:k})$, $p_{k-1|k-1}(X|Z_{1:k-1})$, $p_{k|k-1}(X|Z_{1:k-1})$, $f_{k|k-1}(X|X_{k-1})$ and $g_k(Z|X_k)$ are all unitless.

At this stage the mathematically minded reader might question the validity of the recursion (7)-(8). Even though the standard Bayes recursion generalises to more general spaces with sufficiently 'nice' structures and consistent notions of integration, it is not obvious that Bayes rule for probability densities also applies to set derivatives of belief mass functions. A rigorous treatment of Bayes rule for probability densities requires deep results in conditional probability (see [38] pp. 230-231). To the best of our knowledge, no such rigorous treatment of Bayes rule for set derivatives of belief mass functions is available. Nevertherless, the result established in the next subsection allows rigorous justification of the FISST Bayes recursion for multi-target filtering.

D. Relationship between FISST and conventional probability

This subsection establishes the relationship between finite set statistics (FISST) and conventional probability theory. In particular, it is shown that the set derivative of a belief mass function of a RFS is closely related to its probability density. This relationship allows the conditional densities $f_{k|k-1}(\cdot|\cdot)$ and $g_k(\cdot|\cdot)$ used in the recursion (5)-(6) to be constructed from the underlying physical model of the sensors, individual target dynamics, target births and deaths using the tools of

FISST. It also consolidates the validity of the FISST multi-target recursion (7)-(8).

Proposition 1. *Suppose that Ξ is a RFS on E with probability distribution P_Ξ and belief mass function β_Ξ. If P_Ξ is absolutely continuous with respect to μ, the unnormalised distribution of a Poisson point process with rate K^{-1} defined by (3), then*

$$\frac{dP_\Xi}{d\mu}(X) = K^{|X|}(d\beta_\Xi)_{(X)}(\emptyset). \qquad (9)$$

Proof: First, note that using the integral defined in (4), for any closed $S \subseteq E$, and any measurable $f : \mathcal{F}(E) \to [0, \infty)$

$$
\begin{aligned}
&\int_{\chi(\uplus_{i=0}^\infty S^i)} f(X)\mu(dX) \\
&= \sum_{i=0}^\infty \frac{1}{i!} \int_{S^i} f(\{x_1, ..., x_i\})\lambda^i(dx_1...dx_i) \\
&= \sum_{i=0}^\infty \frac{1}{i!K^i} \int_{S^i} f(\{x_1, ..., x_i\})\lambda_K^i(dx_1...dx_i) \\
&= \int_S K^{-|X|} f(X)\delta X, \qquad (10)
\end{aligned}
$$

where $\uplus$ denotes disjoint union. Let $p_\Xi = dP_\Xi/d\mu$. Then, it follows from Eq. (10) that, for any closed $S \subseteq E$,

$$
\begin{aligned}
\int_S K^{-|X|} p_\Xi(X)\delta X &= \int_{\chi(\uplus_{i=0}^\infty S^i)} p_\Xi(X)\mu(dX) \\
&= P_\Xi(\chi(\uplus_{i=0}^\infty S^i)) = \beta_\Xi(S).
\end{aligned}
$$

Since S is arbitrary, from the FISST fundamental theorem of calculus $K^{-|X|}p_\Xi(X) = (d\beta_\Xi)_{(X)}(\emptyset)$. ∎

Eq. (10) relates the set integral on the closed sets of E to a conventional integral on Borel sets of $\mathcal{F}(E)$. Recall that $(d\beta_\Xi)_{(X)}(\emptyset)$ has unit of $K^{-|X|}$, hence $K^{|X|}(d\beta_\Xi)_{(X)}(\emptyset)$ is unitless. Proposition 1 implies that the set derivative of the belief mass function β_Ξ without its unit is the probability density p_Ξ with respect to the dominating measure μ given in (3). In other words, the unitless set derivative of the belief mass function of a RFS is its probability density. It is important to note that the probability density p_Ξ is unit dependent, since the dominating measure μ depends on the choice of units.

Through Proposition 1, FISST converts the construction of multi-target densities from multi-target models (1)-(2) into computing set derivatives of belief mass functions. In particular, $f_{k|k-1}(X_k|X_{k-1})$ and $g_k(Z_k|X_k)$ can be determined explicitly by

$$
\begin{aligned}
f_{k|k-1}(X_k|X_{k-1}) &= K_s^{|X_k|}(d\beta_{k|k-1}(\cdot|X_{k-1}))_{X_k}(\emptyset), \\
g_k(Z_k|X_k) &= K_o^{|Z_k|}(d\beta_{k|k}(\cdot|X_k))_{Z_k}(\emptyset).
\end{aligned}
$$

where K_s and K_o denote the units of volume in the spaces E_s and E_o respectively. Procedures for analytically differentiating belief mass functions have also been developed in [15], [25] to facilitate the task for tracking engineers. In general, the multi-target dynamic model (1)-(2) yields the following multi-target

Markov transition and likelihood

$$
\begin{aligned}
&f_{k|k-1}(X_k|X_{k-1}) \\
&= \sum_{W \subseteq X_k} s_{k|k-1}(W|X_{k-1})n_{k|k-1}(X_k - W|X_{k-1}) \quad (11) \\
&g_k(Z_k|X_k) \\
&= \sum_{W \subseteq Z_k} \theta_k(W|X_k)c_k(Z_k - W|X_k), \qquad (12)
\end{aligned}
$$

where $s_{k|k-1}(\cdot|X_{k-1})$ is the density of the RFS $S_k(X_{k-1})$ of surviving targets, $n_{k|k-1}(\cdot|X_{k-1})$ is the density of the RFS $N_k(X_{k-1})$ of new-born targets, $\theta_k(\cdot|X_k)$ is the density of the RFS $\Theta_k(X_k)$ of target generated observations and $c_k(\cdot|X_k)$ is the density of the RFS $C_k(X_k)$ of false alarms. Note that the difference operation used in (11)-(12) is the set difference. Details on how $s_{k|k-1}(\cdot|X_{k-1})$, $n_{k|k-1}(\cdot|X_{k-1})$, $\theta_k(\cdot|X_k)$ and $c_k(\cdot|X_k)$ are derived from the underlying physical model of the sensors, individual target dynamics, target births and deaths, for various assumptions, are available in [15], [25], [28].

Proposition 1 can also be used to justify the validity of the FISST Bayes propagation equations (7)-(8). Using (9) to substitute,

$$
\begin{aligned}
p_{k-1|k-1}(X|Z_{1:k-1}) &= K_s^{|X|}\pi_{k-1|k-1}(X|Z_{1:k-1}), \\
p_{k|k-1}(X|Z_{1:k-1}) &= K_s^{|X|}\pi_{k|k-1}(X|Z_{1:k-1}), \\
f_{k|k-1}(X|X_{k-1}) &= K_s^{|X|}\varphi_{k|k-1}(X|X_{k-1}), \\
p_{k|k}(X|Z_{1:k}) &= K_s^{|X|}\pi_{k|k}(X|Z_{1:k}), \\
g_k(Z|X_k) &= K_o^{|Z|}\rho_k(Z|X_k),
\end{aligned}
$$

into the Bayes recursion (5)-(6) and using (10) to convert a conventional integral on $\mathcal{F}(E_s)$ into a set integral on E_s yield the FISST Bayes recursion.

E. Particle Multi-target Filter

This subsection presents a SMC implementation of the Bayes multi-target filter. The propagation of the multi-target posterior density recursively in time involves the evaluation of multiple set integrals and hence the computational requirement is much more intensive than single-target filtering. Sequential Monte Carlo (SMC) filtering techniques permits recursive propagation of the set of weighted particles that approximate the posterior density.

Central in Monte Carlo methods is the notion of approximating the integrals of interest using random samples. In the context of FISST, what does a sample from a belief mass function mean? Can these samples be used to approximate the set integrals of interest? For the purpose of numerical integration, the FISST multi-target density and its unitless version are equivalent. Hence, it suffices to implement the unitless version of the FISST Bayes multi-target filter, i.e. (5)-(6). Since the unitless FISST multi-target density is indeed a probability density (see subsection II-D), Monte Carlo approximations of the integrals of interest can be constructed using random samples. The single-target particle filter can thus be directly generalised to the multi-target case. In the multi-target

context however, each particle is a finite set and the particles themselves can thus be of varying dimensions.

Assume at time $k-1$, a set of weighted particles $\{w_{k-1}^{(i)}, X_{k-1}^{(i)}\}_{i=1}^{N}$ representing the multi-target posterior $p_{k-1|k-1}$ is available, i.e.

$$p_{k-1|k-1}(X_{k-1}|Z_{1:k-1}) \approx \sum_{i=1}^{N} w_{k-1}^{(i)} \delta_{X_{k-1}^{(i)}}(X_{k-1}).$$

The particle filter proceeds to approximate the multi-target posterior $p_{k|k}$ at time k by a new set of weighted particles $\{w_k^{(i)}, X_k^{(i)}\}_{i=1}^{N}$ as follows

---

Particle Multi-Target Filter

At time $k \geq 1$,

Step 1: _Sampling Step_

- For $i = 1, ..., N$, sample $\widetilde{X}_k^{(i)} \sim q_k\left(\cdot \,|\, X_{k-1}^{(i)}, Z_k\right)$ and set

$$\widetilde{w}_k^{(i)} = \frac{g_k\left(Z_k \,|\, \widetilde{X}_k^{(i)}\right) f_{k|k-1}\left(\widetilde{X}_k^{(i)} \,\middle|\, X_{k-1}^{(i)}\right)}{q_k\left(\widetilde{X}_k^{(i)} \,\middle|\, X_{k-1}^{(i)}, Z_k\right)} w_{k-1}^{(i)}. \tag{13}$$

- Normalise weights: $\sum_{i=1}^{N} \widetilde{w}_k^{(i)} = 1$.

Step 2: _Resampling Step_

- Resample $\left\{\widetilde{w}_k^{(i)}, \widetilde{X}_k^{(i)}\right\}_{i=1}^{N}$ to get $\left\{w_k^{(i)}, X_k^{(i)}\right\}_{i=1}^{N}$.

The importance sampling density $q_k\left(\cdot | X_{k-1}, Z_k\right)$ is a multi-target density and $\widetilde{X}_k^{(i)}$ is a sample from a RFS or point process. Details on sampling from a point process can be found in the spatial statistics literature, see for example [8], [43], [44] and references therein. It is implicit in the above algorithm description that

$$\sup_{X,X'} \left| \frac{f_{k|k-1}(X|X')}{q_k(X|X',Z_k)} \right|$$

is finite and so the weights are well-defined.

There are various ways to perform the resampling step. Most methods consist of making $\zeta_k^{(i)}$ copies of each particle $\widetilde{X}_k^{(i)}$, under the constraint $\sum_{i=1}^{N} \zeta_k^{(i)} = N$, to obtain $\{X_k^{(i)}\}_{i=1}^{N}$. The (random) resampling mechanism is chosen such that $E[\zeta_k^{(i)}] = N a_k^{(i)}$ where $a_k^{(i)} > 0$, $\sum_{i=1}^{N} a_k^{(i)} = 1$ is a sequence of weights set by the user. This resampling step could be achieved using multinomial resampling but the efficient stratified resampling algorithm described in [23] has better statistical properties. The new weights are set to $w_k^{(i)} \propto \widetilde{w}_k^{(i)}/a_k^{(i)}$, $\sum_{i=1}^{N} w_k^{(i)} = 1$. Typically, $a_k^{(i)} = \widetilde{w}_k^{(i)}$ but alternatively we can select $a_k^{(i)} \propto (\widetilde{w}_k^{(i)})^\nu$ where $\nu \in (0,1)$.

After the resampling step, an optional Markov Chain Monte Carlo (MCMC) step can also be applied to increase particle diversity [14]. Since the particles belong in spaces of different dimensions, a reversible jump MCMC step [18] is required. Under standard assumptions, the mean squared error of the SMC approximation is inversely proportional to the number of particles. This is stated more concisely as follows.

Proposition 2. _Consider the particle multi-target filter. Assuming that for all $k \geq 1$ the unnnormalized importance weights (13) are bounded, then there exists a constant c_k independent of N such that for any bounded Borel measurable function f on $\mathcal{F}(E_s)$_

$$\mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} f(X_k^{(i)}) - \int f(X_k) P_{k|k}(dX_k|Z_{1:k})\right)^2\right] \leq c_k \frac{\|f\|_\infty^2}{N},$$

where $\|f\|_\infty = \sup_X |f(X)|$. Under additional mixing assumptions on the dynamic model, it can be shown that $c_k \leq c$ for any k [9].[4]

Having obtained the posterior density, consider the problem of obtaining an estimate of the multi-target state. There is no multi-target analogue of the expected a posteriori (EAP) estimator, since there is no notion of addition for sets. However, by treating a RFS as a random counting measure, an alternative estimator can be defined. The EAP estimator $V_{k|k}$ of the corresponding random counting measure is a measure defined by

$$V_{k|k}(S) = \int \left(\sum_{x \in X} \mathbf{1}_S(x)\right) P_{k|k}(dX|Z_{1:k}),$$

for any Borel $S \subseteq E_s$, where $\mathbf{1}_S$ denotes the indicator function $\mathbf{1}_S(x) = 1$ if $x \in S$ and $\mathbf{1}_S(x) = 0$ otherwise. It is important to emphasize that $V_{k|k}$ is not an EAP estimator for RFS (see Section III-A for more details). However, assuming that the measure $V_{k|k}$ admits a density $D_{k|k}$, then the peaks of $D_{k|k}$ can be used to provide estimates of individual targets. Note that given a particle approximation $\left\{w_k^{(i)}, X_k^{(i)}\right\}_{i=1}^{N}$ of $p_{k|k}$, the particle approximation of $D_{k|k}$ can be obtained by

$$\begin{aligned} D_{k|k}(x) &\approx \sum_{i=1}^{N} w_k^{(i)} \left(\sum_{y \in X_k^{(i)}} \delta_y(x)\right) \\ &= \sum_{i=1}^{N} \sum_{y \in X_k^{(i)}} w_k^{(i)} \delta_y(x). \end{aligned}$$

The main practical problem with the multi-target particle filter is the need to perform importance sampling in very high dimensional spaces if many targets are present. Moreover, it can be difficult to find an efficient importance density. A naive choice of importance density such as $q_k(\cdot | X_{k-1}^{(i)}, Z_k) = f_{k|k-1}(\cdot | X_{k-1}^{(i)})$ will typically lead to an algorithm whose efficiency decreases exponentially with the number of targets for a fixed number of particles. The approached used in [36] can be adopted to mitigate this problem, but this requires intensive computations and is still inefficient with large number of targets. Also, the weight update can be expensive due to the combinatorial nature of the multi-target Markov transition and likelihood (11)-(12).

---

[4] The above result is stated in [9] and [11] for $\mathbf{R}^n$, i.e. f is a Borel measureable function on $\mathbf{R}^n$, but the proof for $\mathcal{F}(E_s)$ follows in a straightforward manner.

III. Sequential Monte Carlo Implementation of the PHD Filter

This section describes a novel Sequential Monte Carlo (SMC) implementation of the Probability Hypothesis Density (PHD) filter [26], [29]. The idea is to propagate the intensity function (or PHD) of the multi-target posterior. Sections III-A and III-B provide a brief review of the PHD filter. A particle interpretation of the propagation equation is then presented in Sections III-C, III-D, III-E followed by description of the algorithm and a convergence result in Section III-F.

A. The 1st Moment of a RFS

The 1st moment of a RFS is the analogue of the expectation of a random vector. However, the expectation of a RFS has no meaning since there is no notion of addition for sets. Nevertheless, the 1st moment can be indirectly constructed by representing the RFS as a random counting measure or random density function.

A finite subset $X \in \mathcal{F}(E)$ can also be equivalently represented by the counting measure N_X (on the Borel subsets of E) defined by $N_X(S) = \sum_{x \in X} \mathbf{1}_S(x) = |X \cap S|$, where the notation $|A|$ is used to denote the number of elements in A. Consequently, the random finite set Ξ can also be represented by a random counting measure N_Ξ defined by $N_\Xi(S) = |\Xi \cap S|$. This representation is commonly used in the point process literature [8], [43], [44].

Using the random counting measure representation, the 1st moment or *intensity measure* V_Ξ of a RFS Ξ is defined by

$$V_\Xi(S) \equiv \mathbf{E}[N_\Xi(S)] = \int \left(\sum_{x \in X} \mathbf{1}_S(x) \right) P_\Xi(dX), \quad (14)$$

for each Borel measurable $S \subseteq E$. The intensity measure over a region S, i.e $V_\Xi(S)$ gives the expected number of elements of Ξ that are in S. Although the intensity measure V_Ξ is an average of counting measures, V_Ξ itself is not a counting measure and hence does not necessarily have a finite set representation. Consequently, V_Ξ cannot be used as an expected a posteriori (EAP) estimator for RFS although it can be defined as the EAP estimator of the corresponding random counting measure.

The density D_Ξ of the intensity measure V_Ξ w.r.t. the Lebesgue measure (if it exists) i.e.

$$D_\Xi = \frac{dV_\Xi}{d\lambda}, \quad (15)$$

is called the *intensity function*, and is also known in the tracking literature as the Probability Hypothesis Density (PHD), a term first introduced in [42] (hence the name PHD filter). It can be shown that [26], [29]

$$D_\Xi(x) = (d\beta_\Xi)_x(E).$$

The intensity function D_Ξ is a unique function on E, except on a set of measure zero. Moreover, $\int_S D_\Xi(x)\lambda(dx) = \mathbf{E}[|\Xi \cap S|]$ is the expected number of elements of Ξ that are in a given measurable region $S \subseteq E$. Consequently, the peaks of D_Ξ are points in E with the highest local concentration of expected number of targets, and hence, can be used to generate estimates

for the elements of Ξ. Since the total mass of the intensity $V_\Xi(E)$ gives the expected number of targets, the simplest approach is to round $V_\Xi(E)$ and choose the resulting number of highest peaks from the intensity function.

B. The PHD Filter

Let $D_{k|k}$ denote the intensity function associated with the multi-target posterior $p_{k|k}$ at time k for each $k \geq 0$. The PHD filter involves a prediction step and an update step that propagates the intensity function $D_{k|k}$ recursively in time. This recursion can be succinctly described by introducing the following prediction and update operators (on the space of integrable functions on E_s).

The PHD prediction operator $\Phi_{k|k-1}$ is defined by

$$(\Phi_{k|k-1}\alpha)(x) = \gamma_k(x) + \int \phi_{k|k-1}(x,\xi)\alpha(\xi)\lambda(d\xi), \quad (16)$$

for any integrable function α on E_s, where γ_k denotes the intensity function of the spontaneous birth RFS Γ_k,

$$\phi_{k|k-1}(x,\xi) = b_{k|k-1}(x\,|\,\xi) + e_{k|k-1}(\xi)f_{k|k-1}(x\,|\,\xi),$$

with $b_{k|k-1}(\cdot\,|\,\xi)$ denoting the intensity function of the RFS $B_{k|k-1}(\{\xi\})$ of targets spawned from the previous state ξ, $e_{k|k-1}(\xi)$ denoting the probability that the target still exists at time k given that it has previous state ξ, and $f_{k|k-1}(\cdot\,|\,\cdot)$ denoting the transition probability density of individual targets[5].

The PHD update operator Ψ_k is defined by

$$(\Psi_k\alpha)(x) = \left[\upsilon(x) + \sum_{z \in Z_k} \frac{\psi_{k,z}(x)}{\kappa_k(z) + \langle \psi_{k,z}, \alpha \rangle} \right] \alpha(x), \quad (17)$$

for any integrable function α on E_s, where $\kappa_k(\cdot)$ is the intensity function of the clutter RFS,

$$\begin{aligned} \upsilon(x) &= 1 - P_D(x), \\ \psi_{k,z}(x) &= P_D(x)g_k(z\,|\,x), \\ \langle f, h \rangle &= \int f(x)h(x)\lambda(dx), \end{aligned}$$

with $P_D(x)$ denoting the (state-dependent) probability of detection, $g_k(\cdot\,|\,\cdot)$ denoting the likelihood of individual targets[6]. Note that $\kappa_k(\cdot)$ can also be written as $r_k c_k(\cdot)$, where r_k is the average number of clutter points per scan and c_k is the probability distribution of each clutter point.

Assuming that the targets evolve independently of each other, the intensity function $D_{k|k-1}$ associated with the predicted multi-target density $p_{k|k-1}$ in the prediction equation (5) is given by [26], [29]

$$D_{k|k-1} = \Phi_{k|k-1}D_{k-1|k-1}. \quad (18)$$

Moreover, assuming that the predicted multi-target density $p_{k|k-1}$ is Poisson, and that the detection and measurement of a target is independent from that of other targets, it was

---

[5]For notational convenience, the same notation as the multi-target transition density is used. There is no danger of confusion as multi-target densities do not appear in this section.

[6]The same notation as the multi-target likelihood is used.

shown in [26], [29] that the intensity function $D_{k|k}$ associated with the posterior density $p_{k|k}$ in the update equation (6) is

$$D_{k|k} = (\Psi_k \circ \Phi_{k|k-1})D_{k-1|k-1}, \qquad (19)$$

where "$\circ$" denotes composition of functions. It is implicitly assumed that there are no merged nor split measurements. The Poisson assumption is a mathematical simplification that allows a closed-form expression for the update step of the PHD filter. In using the intensity function $D_{k|k}$ to characterize the multi-target posterior density $p_{k|k}$ it is implicitly assumed that the higher order moments are negligible. These assumptions are justifiable when the false alarm rate and measurement noise are small.

Since the intensity function is a function defined on the space where individual targets live, its propagation requires much less computational power than the multi-target posterior. Unfortunately, this still involves multiple integrals that have no closed form expressions in general. A SMC implementation of the PHD recursion (16)-(17) is described next.

C. The prediction operator

At time step $k-1$, consider the prediction of a non-negative integrable function α_{k-1} to time k, i.e.

$$(\Phi_{k|k-1}\alpha_{k-1})(x_k)$$
$$= \int \phi_{k|k-1}(x_k, x_{k-1})\alpha_{k-1}(x_{k-1})\lambda(dx_{k-1}) + \gamma_k(x_k). \quad (20)$$

Given a particle representation of α_{k-1} i.e.

$$\hat{\alpha}_{k-1}(x_{k-1}) = \sum_{i=1}^{L_{k-1}} w_{k-1}^{(i)} \delta_{x_{k-1}^{(i)}}(x_{k-1})$$

then

$$(\Phi_{k|k-1}\hat{\alpha}_{k-1})(x_k) = \sum_{i=1}^{L_{k-1}} w_{k-1}^{(i)} \phi_{k|k-1}(x_k, x_{k-1}^{(i)}) + \gamma_k(x_k). \quad (21)$$

A particle approximation of $\Phi_{k|k-1}\hat{\alpha}_{k-1}$ in (21) can then be derived by applying importance sampling to each of its terms. Given the importance (or proposal) densities $p_k(\cdot | Z_k)$, $q_k(\cdot | x_{k-1}, Z_k)$ such that $\gamma_k(x_k) > 0$ implies $p_k(x_k | Z_k) > 0$ and $\phi_{k|k-1}(x_k, x_{k-1}) > 0$ implies $q_k(x_k | x_{k-1}, Z_k) > 0$, Eq. (21) can be rewritten as

$$(\Phi_{k|k-1}\hat{\alpha}_{k-1})(x_k)$$
$$= \sum_{i=1}^{L_{k-1}} w_{k-1}^{(i)} \frac{\phi_{k|k-1}(x_k, x_{k-1}^{(i)})}{q_k(x_k | x_{k-1}^{(i)}, Z_k)} q_k(x_k | x_{k-1}^{(i)}, Z_k)$$
$$+ \frac{\gamma_k(x_k)}{p_k(x_k | Z_k)} p_k(x_k | Z_k).$$

Thus, the following Monte Carlo (particle) approximation can be obtained

$$(\widehat{\Phi}_{k|k-1}\hat{\alpha}_{k-1})(x_k) \equiv \sum_{i=1}^{L_{k-1}+J_k} w_{k|k-1}^{(i)} \delta_{x_k^{(i)}}(x_k)$$

where

$$x_k^{(i)} \sim \begin{cases} q_k\left(\cdot | x_{k-1}^{(i)}, Z_k\right), & i = 1, ..., L_{k-1} \\ p_k\left(\cdot | Z_k\right), & i = L_{k-1}+1, ..., L_{k-1}+J_k \end{cases}$$

$$w_{k|k-1}^{(i)} = \begin{cases} \dfrac{\phi_{k|k-1}\left(x_k^{(i)}, x_{k-1}^{(i)}\right) w_{k-1}^{(i)}}{q_k\left(x_k^{(i)} | x_{k-1}^{(i)}, Z_k\right)}, & i = 1, ..., L_{k-1} \\ \dfrac{\gamma_k\left(x_k^{(i)}\right)}{J_k p_k\left(x_k^{(i)} | Z_k\right)}, & i = L_{k-1}+1, ..., L_{k-1}+J_k \end{cases}$$

It is easy to verify that for any integrable test function h,

$$\mathbf{E}\left[\left\langle \widehat{\Phi}_{k|k-1}\alpha_{k-1}, h \right\rangle\right] = \left\langle \Phi_{k|k-1}\alpha_{k-1}, h \right\rangle.$$

Note that we started with α_{k-1} having L_{k-1} particles, which are then predicted forward by the kernel $\phi_{k|k-1}$ to another set of L_{k-1} particles. Additionally, J_k new particles arise from the birth process. The number of new particles J_k can be a function of k to accommodate the varying number of new targets at each time step. Assuming that the total mass of γ_k has a closed form, then typically J_k is chosen to be proportional to this mass, i.e. $J_k = \rho \int \gamma_k(x)dx$, so that on average we have ρ particles per new born target.

D. The update operator

For the update step of the recursion, assume that prediction step yields a function $\alpha_{k|k-1}$ characterised by $\{w_{k|k-1}^{(i)}, x_k^{(i)}\}_{i=1}^{L_{k-1}+J_k}$. Applying the update operator gives

$$(\Psi_k \alpha_{k|k-1})(x) = \sum_{i=1}^{L_{k-1}+J_k} w_k^{(i)} \delta_{x_k^{(i)}}(x),$$

where

$$w_k^{(i)} = \left[\upsilon(x^{(i)}) + \sum_{z \in Z_k} \frac{\psi_{k,z}(x_k^{(i)})}{\kappa_k(z) + C_k(z)}\right] w_{k|k-1}^{(i)}, \quad (22)$$

$$C_k(z) = \sum_{j=1}^{L_{k-1}+J_k} \psi_{k,z}(x_k^{(j)}) w_{k|k-1}^{(j)}. \quad (23)$$

The update operator maps the function with particle representation $\{w_{k|k-1}^{(i)}, x_k^{(i)}\}_{i=1}^{L_{k-1}+J_k}$ into one with particle representation $\{w_k^{(i)}, x_k^{(i)}\}_{i=1}^{L_{k-1}+J_k}$ by modifying the weights of these particles according to Eq. (22).

E. Particle propagation

For any $k \geq 0$, let $\hat{\alpha}_k = \{w_k^{(i)}, x_k^{(i)}\}_{i=1}^{L_k}$ denote a particle approximation of $D_{k|k}$. The algorithm is designed such that the concentration of particles in a given region of the single-target state space, say S, represents the expected number of targets in S, i.e. $E[|\Xi_k \cap S|| Z_{1:k}] \approx \sum_{j=1}^{L_k} \mathbf{1}_S(x_k^{(i)}) w_k^{(j)}$.

Using the PHD recursion, a particle approximation of the intensity function at time step $k > 0$ can be obtained from a particle approximation at the previous time step by

$$\hat{\alpha}_k = (\Psi_k \circ \widehat{\Phi}_{k|k-1})\hat{\alpha}_{k-1}. \quad (24)$$

Note that since $\hat{\alpha}_k$ has $L_k = L_{k-1} + J_k$ particles, the number L_k of particles may increase over time even if the number of targets does not. This is very inefficient, since computational

resource is wasted in exploring regions of the (single-target) state space where there are no targets. On the other hand if L_k is fixed then the ratio of particles to targets would fluctuate as the number of targets changes. Consequently, at times there may be an insufficient number of particles to resolve the targets (up to the PHD filter limitations) while at other times there may be an excess of particles for a small number of targets or no target at all. It would be computationally more efficient to adaptively allocate say ρ particles per target at each time epoch.

Since the expected number of targets $N_{k|k}$ (given by the total mass $\int D_{k|k}(\xi)\lambda(d\xi)$) can be estimated by $\widehat{N}_{k|k} = \sum_{j=1}^{L_{k-1}+J_k} w_k^{(j)}$, it is intuitive to have the number of particles $L_k \cong \rho\widehat{N}_{k|k}$. Furthermore, we also want to eliminate particles with low weights and multiply particles with high weights to focus on the important zones of the space. This can be achieved by resampling $L_k \approx \rho\widehat{N}_{k|k}$ particles from $\{w_k^{(i)}, x_k^{(i)}\}_{i=1}^{L_{k-1}+J_k}$ and redistributing the total mass $\widehat{N}_{k|k}$ among the L_k resampled particles.

F. Algorithm

Based on the elements presented above, it is possible to propose the following generic particle filtering algorithm for the PHD recursion.

Particle PHD filter

At time $k \geq 1$,

Step 1: *Prediction*

- For $i = 1, ..., L_{k-1}$, sample $\widetilde{x}_k^{(i)} \sim q_k(\cdot | x_{k-1}^{(i)}, Z_k)$ and compute the predicted weights

$$\widetilde{w}_{k|k-1}^{(i)} = \frac{\phi_{k|k-1}(\widetilde{x}_k^{(i)}, x_{k-1}^{(i)})}{q_k(\widetilde{x}_k^{(i)} | x_{k-1}^{(i)}, Z_k)} w_{k-1}^{(i)}. \qquad (25)$$

- For $i = L_{k-1}+1, ..., L_{k-1}+J_k$, sample $\widetilde{x}_k^{(i)} \sim p_k(\cdot | Z_k)$ and compute the weights of new born particles

$$\widetilde{w}_{k|k-1}^{(i)} = \frac{1}{J_k} \frac{\gamma_k(\widetilde{x}_k^{(i)})}{p_k(\widetilde{x}_k^{(i)} | Z_k)}. \qquad (26)$$

Step 2: *Update*

- For each $z \in Z_k$, compute

$$C_k(z) = \sum_{j=1}^{L_{k-1}+J_k} \psi_{k,z}(\widetilde{x}_k^{(j)})\widetilde{w}_{k|k-1}^{(j)}. \qquad (27)$$

- For $i = 1, ..., L_{k-1}+J_k$, update weights

$$\widetilde{w}_k^{(i)} = \left[v(\widetilde{x}_k^{(i)}) + \sum_{z \in Z_k} \frac{\psi_{k,z}(\widetilde{x}_k^{(i)})}{\kappa_k(z) + C_k(z)} \right] \widetilde{w}_{k|k-1}^{(i)}. \qquad (28)$$

Step 3: *Resampling*

- Compute the total mass $\widehat{N}_{k|k} = \sum_{j=1}^{L_{k-1}+J_k} \widetilde{w}_k^{(j)}$,
- Resample $\left\{ \frac{\widetilde{w}_k^{(i)}}{\widehat{N}_{k|k}}, \widetilde{x}_k^{(i)} \right\}_{i=1}^{L_{k-1}+J_k}$ to get $\left\{ \frac{w_k^{(i)}}{\widehat{N}_{k|k}}, x_k^{(i)} \right\}_{i=1}^{L_k}$,
- Rescale (multiply) the weights by $\widehat{N}_{k|k}$ to get $\left\{ w_k^{(i)}, x_k^{(i)} \right\}_{i=1}^{L_k}$.

It is implicit in the prediction step of the above algorithm that for each k,

$$\sup_{\xi, x} \left| \frac{\phi_{k|k-1}(x, \xi)}{q_k(x | \xi, Z_k)} \right| \leq Q_k, \qquad (29)$$

$$\sup_x \left| \frac{\gamma_k(x)}{p_k(x | Z_k)} \right| \leq P_k, \qquad (30)$$

where Q_k and P_k are finite and so the weights (25)-(26) are well-defined.

Care must be taken when implementing the resampling step for the particle PHD filter. In this case, the new weights $\{w_k^{(i)}\}_{i=1}^{L_k}$ are not normalised to 1 but sum to $\widehat{N}_{k|k} = \sum_{i=1}^{L_{k-1}+J_k} \widetilde{w}_k^{(i)}$. Similar to the standard particle filter, each particle $\widetilde{x}_k^{(i)}$ is copied $\zeta_k^{(i)}$ times under the constraint $\sum_{i=1}^{L_{k-1}+J_k} \zeta_k^{(i)} = L_k$ to obtain $\{x_k^{(i)}\}_{i=1}^{L_k}$. The (random) resampling mechanism is chosen such that $\mathbf{E}[\zeta_k^{(i)}] = L_k a_k^{(i)}$ where $a_k^{(i)} > 0$, $\sum_{i=1}^{L_{k-1}+J_k} a_k^{(i)} = 1$ is a sequence of weights set by the user. However, the new weights are set to $w_k^{(i)} \propto \widetilde{w}_k^{(i)}/a_k^{(i)}$ with $\sum_{i=1}^{L_k} w_k^{(i)} = \widehat{N}_{k|k}$ instead of $\sum_{i=1}^{L_k} w_k^{(i)} = 1$. Typically, $a_k^{(i)} = \widetilde{w}_k^{(i)}/\widehat{N}_{k|k}$ but alternatively we can select $a_k^{(i)} \propto (\widetilde{w}_k^{(i)})^\nu$ where $\nu \in (0, 1)$.

It is standard to assume (see [7]) that, for all $|q_i| \leq 1$ and for some constant c_k

$$\sum_{i=1}^{L_{k-1}+J_k} \sum_{j=1}^{L_{k-1}+J_k} q_i [A_k]_{i,j} q_j \leq L_k c_k, \qquad (31)$$

where

$$[A_k]_{i,j} = \mathbf{E}\left[\left(\zeta_k^{(i)} - \frac{L_k \widetilde{w}_k^{(i)}}{\widehat{N}_{k|k}} \right) \left(\zeta_k^{(j)} - \frac{L_k \widetilde{w}_k^{(j)}}{\widehat{N}_{k|k}} \right) \right]. \qquad (32)$$

For example, in the popular multinomial resampling scheme of [7],

$$\mathbf{E}\left[\zeta_k^{(i)} - \frac{L_k \widetilde{w}_k^{(i)}}{\widehat{N}_{k|k}} \right] = 0$$

and (31) is satisfied for $c_k = 1$. Additionally, the branching scheme proposed in [7] also satisfies this assumption.

Under standard assumptions a similar result to that of [7] also holds for the particle PHD filter.

Proposition 3. *Suppose that for all $k \geq 1$, $e_{k|k-1}$ is continuous, the single-target transition $f_{k|k-1}(\cdot | \cdot)$, the spawning intensity $b_{k|k-1}(\cdot | \cdot)$ are Feller[7], and that the single-target likelihood g_k is bounded, continuous and strictly positive. Consider the particle PHD filter. If the weights (25)-(26) are bounded, the resampling satisfies (31) and the number of particles L_k is fixed to L, such that L/J_k is finite for all $k \geq 1$, then for any bounded and continuous function f on E_s*

$$\lim_{L \to \infty} \mathbf{E}\left[\left| \sum_{i=1}^{L} f(x_k^{(i)}) w_k^{(i)} - \int f(x) D_{k|k}(x)\lambda(dx) \right| \right] = 0.$$

The proof of this result is detailed in Appendix C. A stronger result for the particle PHD filter showing the mean

---

[7]See Appendix C for the Feller property.

squared error diminishing at a rate inversely proportional to the number of samples has also been established and will be published elsewhere.

For initialisation, importance sampling can be applied to obtain a particle approximation of the initial intensity function. If no prior information is available, the initial intensity function can be set to zero and hence no particles are needed. In this case the algorithm starts to sample from the birth process at the next iteration. A better strategy is to guess the number of targets $\widehat{N}_0$ (from the observations) and set the initial intensity function to a uniform intensity with total mass $\widehat{N}_0$.

Remark: The particle PHD filter reduces to the standard particle filter in the case where there is only one target with no birth, no death, no clutter and unity probability of detection.

In the standard particle filtering context, choosing the importance distribution so as to minimise the (conditional) variance of the weights is well known. In the context of the PHD filter, this becomes much more difficult and is the subject of further study.

IV. SIMULATIONS

A. Linear-Gaussian examples

For illustration purposes, consider a two-dimensional scenario with an unknown and time varying number of targets observed in clutter over the region $[-100, 100] \times [-100, 100]$. The state of each target consists of position and velocity, while only position measurements are available. Each target moves according to the following linear Gaussian dynamics i.e.

$$x_k = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} x_{k-1} + \begin{bmatrix} \frac{T^2}{2} & 0 \\ T & 0 \\ 0 & \frac{T^2}{2} \\ 0 & T \end{bmatrix} \begin{bmatrix} v_{1,k} \\ v_{2,k} \end{bmatrix}$$

where $x_k = [x_{1,k}, x_{2,k}, x_{3,k}, x_{4,k}]^T$; $[x_{1,k}, x_{3,k}]^T$ is the position, while $[x_{2,k}, x_{4,k}]^T$ is the velocity at time k, and $T = 1$ is the sampling period. The process noise $\{v_{1,k}\}$, $\{v_{2,k}\}$ are mutually independent zero-mean Gaussian white noise with respective standard deviations $\sigma_{v_1} = 1$ and $\sigma_{v_2} = 0.1$.

Targets can appear or disappear in the scene at any time. Each existing target has a (state independent) probability of survival $e_{k|k-1} = 0.95$. For simplicity no spawning is considered in these examples. New targets can appear spontaneously according to a Poisson point process with intensity function $\gamma_k = 0.2\mathcal{N}(\cdot; \bar{x}, Q)$, where

$$\bar{x} = \begin{bmatrix} 0 \\ 3 \\ 0 \\ -3 \end{bmatrix}, Q = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\mathcal{N}(\cdot; \bar{x}, Q)$ denotes a normal density with mean $\bar{x}$ and covariance Q.

Position ground truth of 4 tracks over 40 scans are displayed in Figure 1. These 4 tracks start in the vicinity of the origin and move radially outwards. The start and finish times of the tracks can be seen from Figure 2, which plots the individual x and y components of each track against time. Note that the
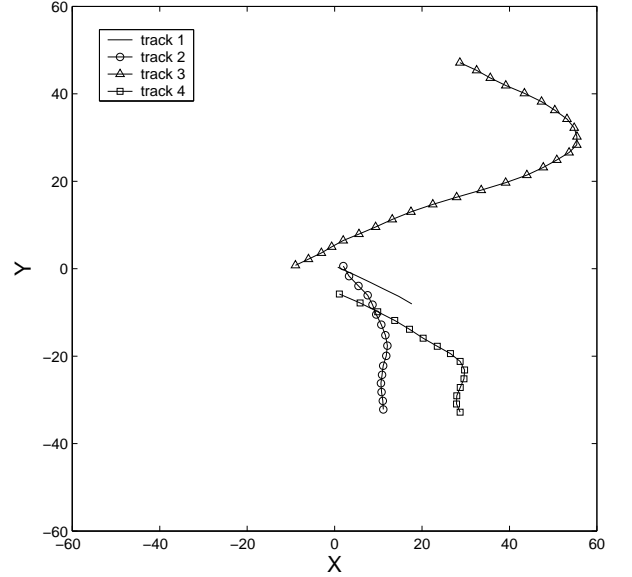

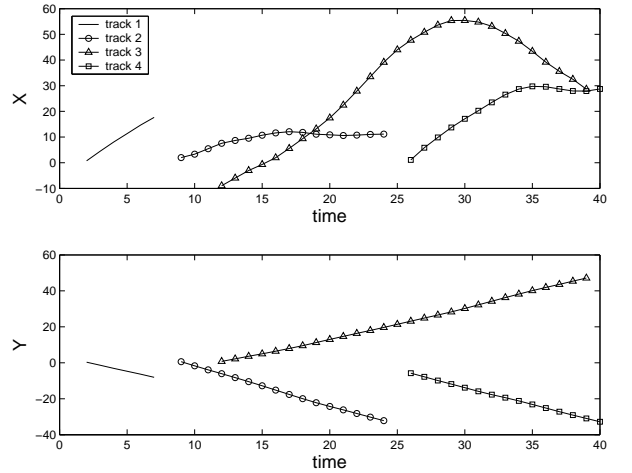Fig. 1.   Ground truth: position plots of 4 tracks superimposed over 40 time steps.

Fig. 2.   Ground truth: plots of x and y components of the 4 true tracks against time, showing the different start and finish times of the tracks.

velocity is relatively constant in the y-direction while the x-direction shows more fluctuations.

The target-originated measurements are given by

$$y_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x_k + \begin{bmatrix} w_{1,k} \\ w_{2,k} \end{bmatrix}$$

with $\{w_{1,k}\}$ and $\{w_{2,k}\}$ mutually independent zero-mean Gaussian white noise with standard deviations $\sigma_{w_1} = \sigma_{w_1} = 2.5$. The measurement noise is assumed independent of the process noise. To demonstrate the mechanics of the particle PHD filter, we can consider a unity probability of detection without loss of generality from an algorithmic viewpoint. Clutter is uniformly distributed over the region $[-100, 100] \times [-100, 100]$ with an average rate of r points per scan, i.e. a Poisson point process with a uniform intensity function $\kappa = r/200^2$. Hence, we have an average of $0.0005r$ clutter points per σ-gate.

1000 particles per expected target are used in the proposed particle PHD filter, thus, the number of particles varies throughout the simulation. Also, the number of particles is
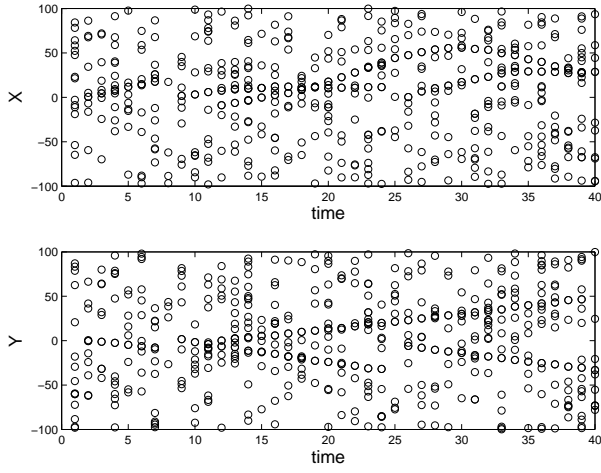
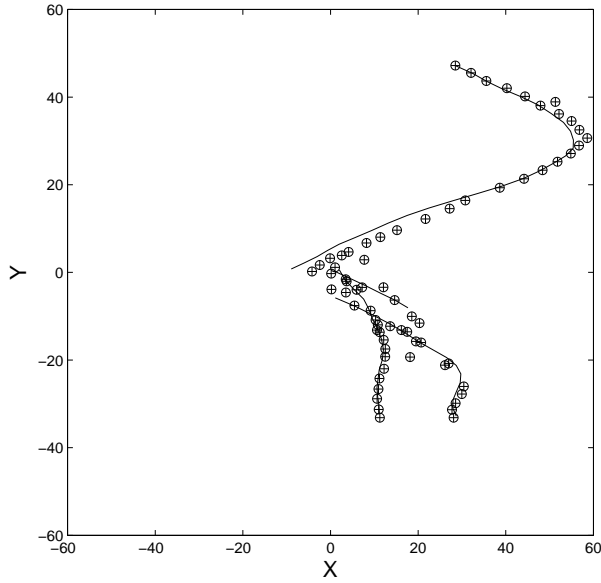Fig. 3.   x and y components of position observations immersed in clutter of rate $r = 10$.

Fig. 4.   Filter output ($r = 10$), position estimates ($\oplus$) superimposed on ground truth (solid-line) over 40 time steps.

Fig. 5.   Filter output ($r = 10$), plots of x and y components of position estimates (circle) against time, superimposed on ground truth (solid-line).

Fig. 6.   Target number estimate and multi-target miss-distance at each time step ($r = 10$).

hard-limited so that it does not fall below 500 when the expected number of target is less than 0.5. The importance sampling densities used are $q_k = f_{k|k-1}$ and $p_k = \mathcal{N}(\cdot; \bar{x}, Q)$. Target state estimates are extracted from the particle approximation of the intensity function by applying standard clustering techniques [19] to the set of particles at each time step.

The x and y coordinates of observations in clutter with an average rate of $r = 10$ points per scan are shown for each time step in Figure 3. Figure 4 shows the position estimates superimposed on the true tracks over 40 time steps. Figure 5 shows the individual x and y coordinates of the true tracks and the estimated targets at each time step. Observe the close proximity of the estimated positions to the true tracks. Quantitatively, the standard root-mean-squared error used in single target problems is not appropriate as a measure of performance since this requires correct association. It was proposed in [20] to use the Wasserstein distance as a multi-target miss-distance. The Wasserstein distance is defined for
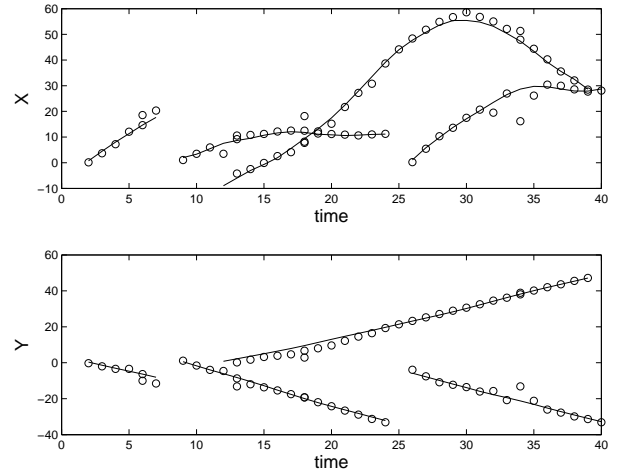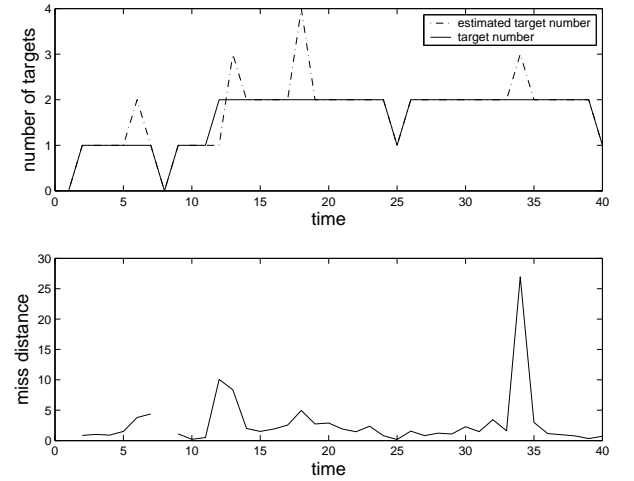
any two non-empty subsets $\hat{X}$, X as

$$d_p(\hat{X}, X) = \min_{C} \sqrt[p]{\sum_{i=1}^{|\hat{X}|} \sum_{j=1}^{|X|} C_{i,j} \|\hat{x}_i - x_j\|^p},$$

where the minimum is taken over the set of all transportation matrices C, (a transportation matrix is one whose entries $C_{i,j}$ satisfy $C_{i,j} \geq 0, \sum_{j=1}^{|X|} C_{i,j} = 1/|\hat{X}|, \sum_{i=1}^{|\hat{X}|} C_{i,j} = 1/|X|$). Note that the Wasserstein miss-distance is not defined if either the estimate $\hat{X}$ or the ground truth X is empty. When $\hat{X}$ and X have the same number of elements, the Wasserstein distance gives the distance for the best association. Figure 6 plots the estimated target number (obtained by rounding the total intensity measure) and true target number along with the Wasserstein distance between the positions estimates of the multi-target state and ground truth at each time step. When the estimated number is incorrect, the clustering output becomes unreliable. Moreover, the Wasserstein distance tends to penalise sets of different cardinalities. This results in high peaks in the multi-target miss-distance at the instances where the estimated number is incorrect as indicated in Figure 6. When the estimated number of targets is correct, the Wasserstein miss-distance stays below 3.5, which turns out to be

approximately equal to the root-mean-squared measurement noise. The PHD filter is effective in this scenario because the false alarm rate and measurement noise are sufficiently small (0.005 clutter points per σ-gate) so that it is unlikely for a false alarm to fall within a one σ-gate surrounding any target track.
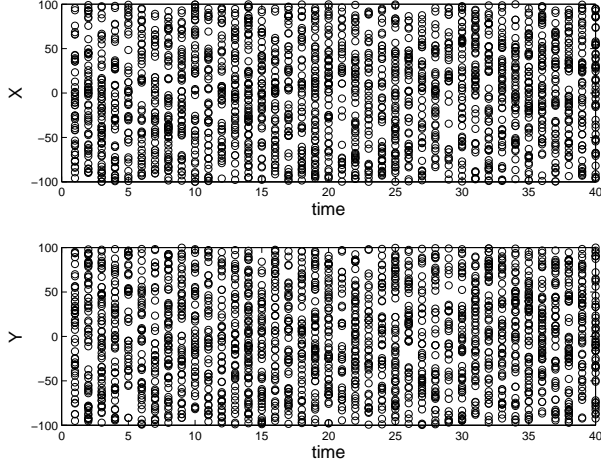

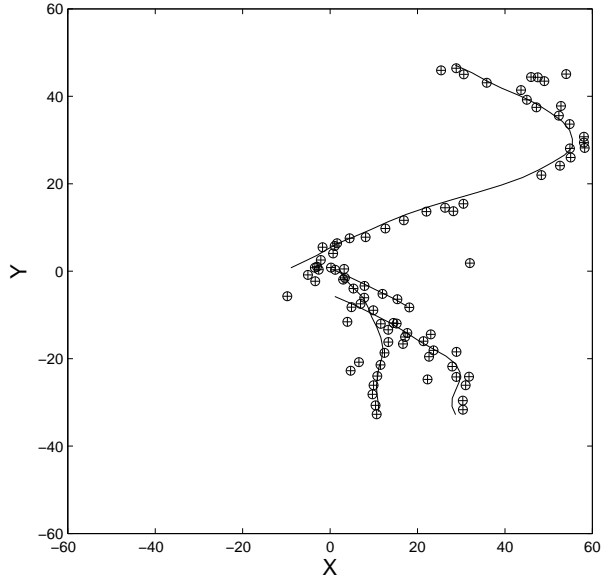Fig. 7. x and y components of position observations immersed in clutter of rate $r = 50$.

Fig. 8. Filter output ($r = 50$), position estimates ($\oplus$) superimposed on ground truth (solid-line) over 40 time steps.

Figure 7 shows the observations for the tracks of Figure 2 observed in denser clutter with rate $r = 50$. The degradation in performance of the proposed particle PHD filter is evident from Figures 8 and 9. This is confirmed in Figure 10, which shows an increased error in the estimated target number and the Wasserstein multi-target miss-distance. The degradation in performance can be attributed to the increased average number of 0.025 clutter points per σ-gate.

B. Bearing and range tracking example

To demonstrate the versatility of the proposed particle PHD filter on nonlinear problems, consider a bearing and range
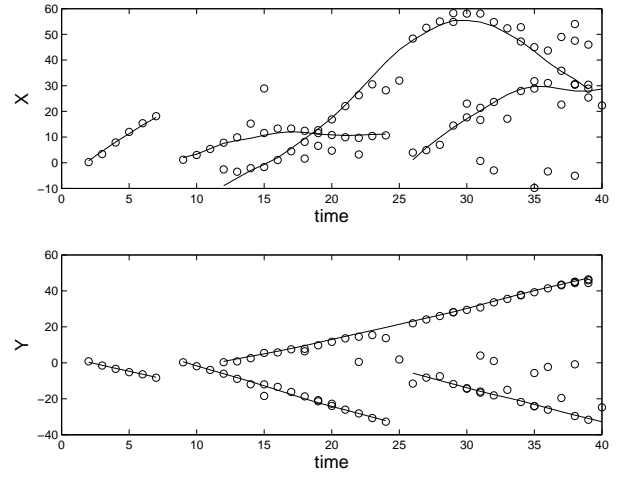

Fig. 9. Filter output ($r = 50$), plots of x and y components of position estimates (circle) against time, superimposed on ground truth (solid-line).

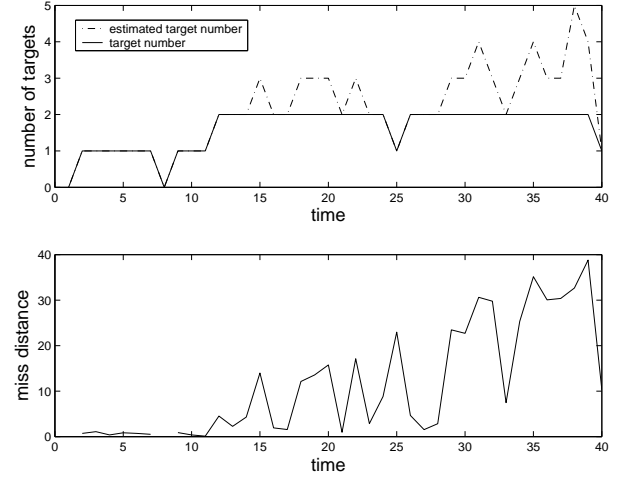

Fig. 10. Target number estimate and multi-target miss-distance at each time step ($r = 50$).
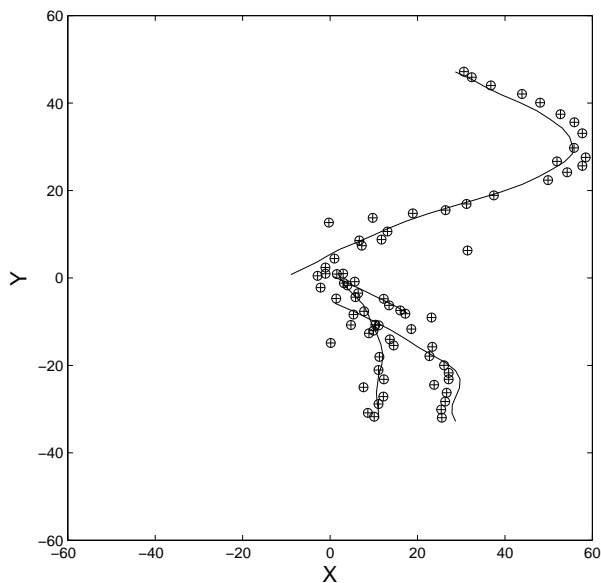
tracking application. The same target tracks as in the linear example over the region $[-100, 100] \times [-100, 100]$ are used (see Figure 2). The sensor is located at $[0, -100]^T$ and the measurement equations are

$$\theta_k = \arctan\left(\frac{[\ 1\ \ 0\ \ 0\ \ 0\ ]x_k}{[\ 0\ \ 0\ \ 1\ \ 0\ ]x_k + 100}\right) + w_{1,k},$$

$$r_k = \left\|\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x_k - \begin{bmatrix} 0 \\ -100 \end{bmatrix}\right\| + w_{2,k}.$$
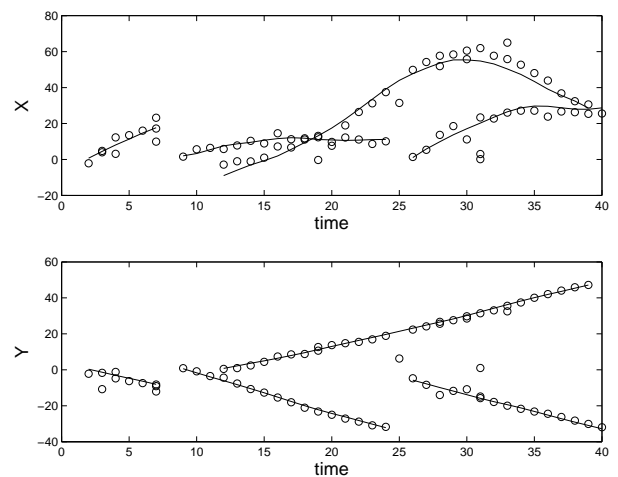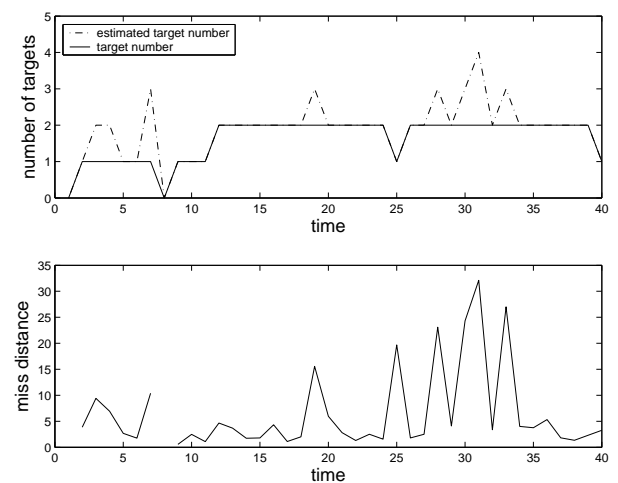
The measurement noise $\{w_{1,k}\}$ and $\{w_{2,k}\}$ are zero-mean Gaussian white noise with respective standard deviations $\sigma_{w_1} = 0.05$ (i.e. approximately 3 degrees) and $\sigma_{w_2} = 2$, are independent of each other and the process noise. Without loss of generality we use a unity probability of detection. Clutter is uniformly distributed over the observation space $[-\pi/2, \pi/2] \times [0, 200]$ with an average rate of r points per scan, i.e. a Poisson point process on $[-\pi/2, \pi/2] \times [0, 200]$ with a uniform intensity function $\kappa = r/200\pi$. Again this gives an average of $0.0005r$ clutter points per σ-gate as in the linear example.

Figure 11 shows, at each time step, the bearing and range observations in clutter with an average rate of $r = 10$ points per scan. Figure 12 shows the positions of the estimated
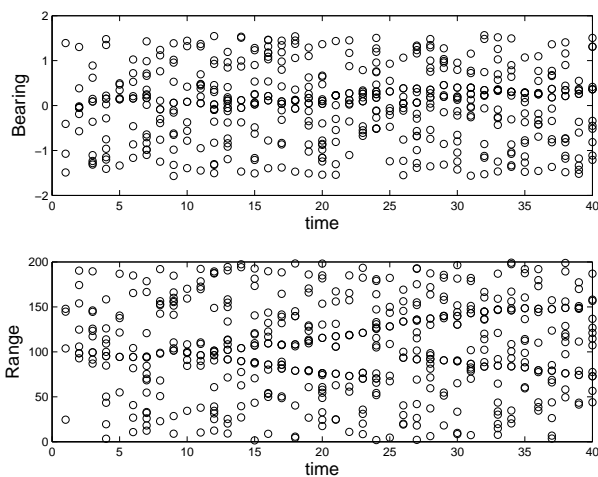
Fig. 11.   Bearing and range observations immersed in clutter of rate $r = 10$.

Fig. 13.   Filter output for bearing and range tracking ($r = 10$), plots of x and y components of position estimates (circle) against time, superimposed on ground truth (solid-line).

Fig. 14.   Target number estimate and multi-target miss-distance for bearing and range tracking ($r = 10$).

Fig. 12.   Filter output for bearing and range tracking ($r = 10$), position estimates ($\oplus$) superimposed on ground truth (solid-line) over 40 time steps.
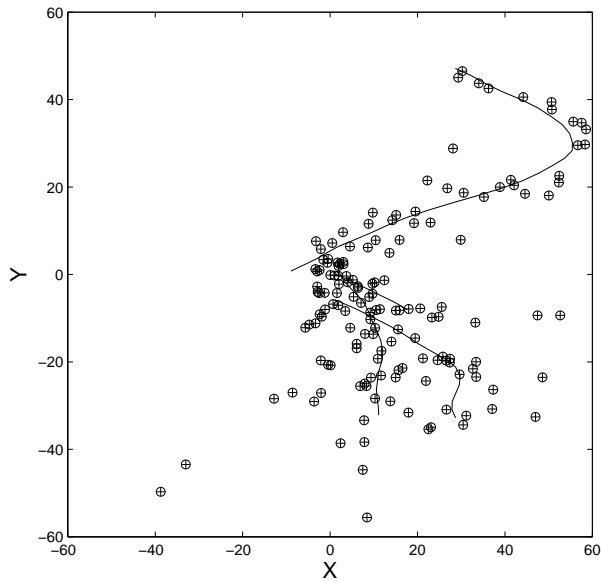
targets superimposed on the tracks over 40 time steps. The individual x and y coordinates of the tracks and estimated targets for each time step are shown in Figure 13. Similar to the linear example for $r = 10$, the estimated positions are close to the true tracks. This is quantified in Figure 14, which compares the estimated targets against ground truth in terms of target number and Wasserstein multi-target miss distance at each time step. The multi-target miss distance exhibits peaks at the instances where the estimated number is incorrect. The peaking in Figure 14 from the 25th to 33rd time steps coincides with the approximate alignment of targets and clutter along a single bearing. When the estimated number of targets is correct, the Wasserstein miss-distance is approximately 5 (higher than that for the linear example). Again the PHD filter is effective in this low clutter scenario because the false alarm rate and measurement noise are sufficiently small (0.005 clutter points per σ-gate) so that it is unlikely for a false alarm to fall within a one σ-gate surrounding any target track.

Figure 15 shows the observations for the same tracks observed in denser clutter of rate $r = 50$. In this case the

average number of clutter points per σ-gate is 0.025. The degradation in performance can be seen from Figures 16 and 17. Figure 18 compares the estimated targets against ground truth in terms of target number and Wasserstein multi-target miss-distance. Note the much larger error in the estimated target number and the Wasserstein miss-distance.
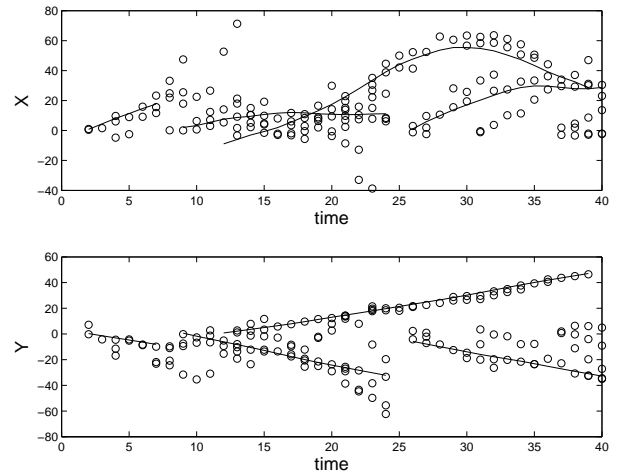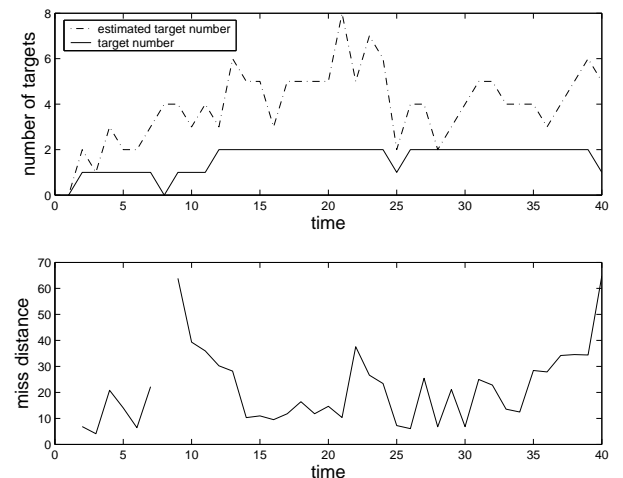
Remark: The Wasserstein miss-distance provides a measure of the overall performance of the PHD filter and the point estimate extraction process. Since the PHD filter only produces an intensity function estimate, the measure of performance (using the Wasserstein miss-distance) depends on the extraction of point estimates from the filter output. The extraction of point estimates from the particle approximation using standard clustering [19] or peak finding techniques can be unreliable, especially when the estimated number of targets is incorrect. A fundamental measure of performance for the PHD filter should not depend on the quality of the point estimate extractions. To directly measure the performance of the PHD filter, it is possible to treat the PHD filter output and ground truth as multi-dimensional images and use the Wasserstein distance as a similarity measure between these two images. In this approach, the miss-distance can be defined for filter
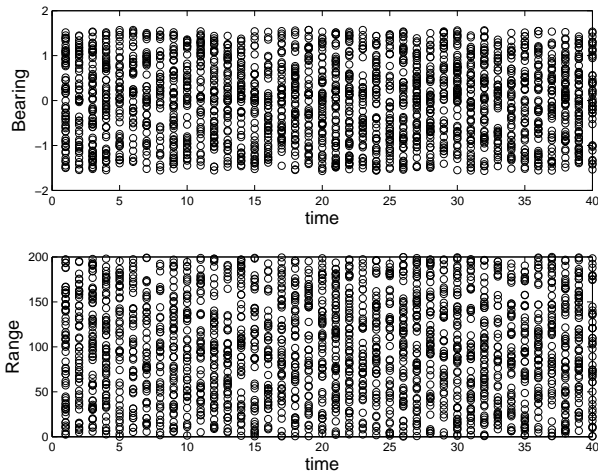
Fig. 15.　Bearing and range observations immersed in clutter of rate $r = 50$.

Fig. 17.　Filter output for bearing and range tracking ($r = 50$), plots of x and y components of position estimates (circle) against time, superimposed on ground truth (solid-line).

Fig. 16.　Filter output for bearing and range tracking ($r = 50$), position estimates ($\oplus$) superimposed on ground truth (solid-line) over 40 time steps.

Fig. 18.　Target number estimate and multi-target miss-distance for bearing and range tracking ($r = 50$).

estimates and ground truth that involve empty sets. However, the similarity measure between two images can expensive to compute.

V. CONCLUSION

This paper has established that FISST concepts such as set integral and set derivative are closely related to the measure theoretic integral and density. This provides an important connection between Finite Set Statistics (FISST) and standard probability theory. In particular, it has been shown that the difficult task of computing probability densities of Random Finite Sets can be achieved via FISST. This result also allows us to develop a principled and computationally tractable SMC implementation of the Bayes multi-target filter. In addition, we developed a generalised importance sampling and resampling strategy to implement the Probability Hypothesis Density (intensity function) filter, a much cheaper alternative than the Bayes multi-target filter. Convergence results for these SMC implementations have also been established. Both of the proposed approaches are general enough to handle non-linear non-Gaussian target dynamics. Bearing in mind that no

multi-target filtering algorithm is capable of performing well in an arbitrarily adverse environment, the PHD filter shows good promise under a reasonable level of measurement noise and false alarm rates. However, the viability of the proposed approach needs to be tested in real applications.

There are various potential extensions to this work. First, choosing the importance distributions so as to minimise the (conditional) variance of the weights is a challenging problem. Second, the PHD filter output is an intensity function estimate, not point estimates of individual target states. As end users may only be interested in the number of targets, their locations and the associated confidence, efficient and reliable algorithms for computing point estimates from intensity function estimate become important. Thirdly, as the FISST framework does not yield track-valued estimates, it would be useful in many applications to incorporate association functionality to random set based filters to yield track-valued estimates rather than point estimates of the states.

VI. APPENDIX A (MEASURE AND PROBABILITY)

The pair $(\mathcal{X}, \sigma(\mathcal{X}))$ in which $\sigma(\mathcal{X})$ denotes a σ-algebra of subsets of $\mathcal{X}$ is called a *measurable space*. If $\mathcal{X}$ is equipped

with a topology $\tau(\mathcal{X})$, then the σ-algebra of interest is the smallest one that contains $\tau(\mathcal{X})$, and is called the *Borel σ-algebra* or *Borel sets* of $\mathcal{X}$. A set $\mathcal{T} \subseteq \mathcal{X}$ is said to be *measurable* if $\mathcal{T} \in \sigma(\mathcal{X})$. A function $f : \mathcal{X} \to \mathbf{R}$ is said to be *measurable* if the inverse images of the Borel sets of $\mathbf{R}$ under f are measurable. The triple $(\mathcal{X}, \sigma(\mathcal{X}), \mu)$ in which μ is a measure on $\sigma(\mathcal{X})$ is called a *measure space*. If $\mathcal{X} = \cup_{i=0}^{\infty} \mathcal{T}_i$ for some countable sequence of $\mathcal{T}_i \in \sigma(\mathcal{X})$ with $\mu(\mathcal{T}_i) < \infty$, then μ is said to be *σ-finite*.

The *integral* of a measurable function $f : \mathcal{X} \to \mathbf{R}$,

$$\int f(X)\mu(dX),$$

is defined as a limit of integrals of simple (or step) functions. The integral of f over any measurable $\mathcal{T} \subseteq \mathcal{X}$ is defined as

$$\int_{\mathcal{T}} f(X)\mu(dX) = \int \mathbf{1}_{\mathcal{T}}(X)f(X)\mu(dX),$$

where $\mathbf{1}_{\mathcal{T}}$ denotes the indicator function $\mathbf{1}_{\mathcal{T}}(X) = 1$ if $X \in \mathcal{T}$ and $\mathbf{1}_{\mathcal{T}}(X) = 0$ otherwise.

Two σ-finite measures μ_1 and μ_2 on the same measurable space $(\mathcal{X}, \sigma(\mathcal{X}))$ may be given in terms of the other by

$$\mu_2(\mathcal{T}) = \int_{\mathcal{T}} g(X)\mu_1(dX), \quad \forall \mathcal{T} \in \sigma(\mathcal{X}),$$

if $\mu_1(\mathcal{T}) = 0$ implies $\mu_2(\mathcal{T}) = 0$. In this case μ_2 is said to be *absolutely continuous* with respect to (w.r.t.) μ_1 (or equivalently $\mu_2 << \mu_1$) and $g : \mathcal{X} \to [0, \infty)$ is called the *Radon-Nikodým derivative* or *density* of μ_2 with respect to μ_1, which is also denoted as $g = d\mu_2/d\mu_1$.

Let $(\Omega, \sigma(\Omega), P)$ be a *probability space*, i.e. a measure space with $P(\Omega) = 1$, and $(\mathcal{X}, \sigma(\mathcal{X}), \mu)$ be a measure space. An *$\mathcal{X}$-valued random variable* (or *random variable* for simplicity) Ξ is a measurable mapping

$$\Xi : \Omega \to \mathcal{X}.$$

The *probability distribution* of the random variable Ξ is a measure P_Ξ on $\sigma(\mathcal{X})$ defined for each $\mathcal{T} \in \sigma(\mathcal{X})$ by $P_\Xi(\mathcal{T}) = P(\Xi^{-1}(\mathcal{T}))$. If P_Ξ is absolutely continuous w.r.t. the measure μ, then $p_\Xi = dP_\Xi/d\mu$ is called the *probability density* of Ξ. w.r.t. μ.

Let the measurable spaces $(\mathcal{X}, \sigma(\mathcal{X}))$ and $(\mathcal{Z}, \sigma(\mathcal{Z}))$ denote the state space and observation space respectively. The *joint probability distribution* $P_{\Xi,\Sigma}(\cdot, \cdot)$ of the random variables Ξ on $\mathcal{X}$ and Σ on $\mathcal{Z}$ is a probability measure on $\sigma(\mathcal{X}) \otimes \sigma(\mathcal{Z})$, the σ-algebra generated by measurable rectangles $\mathcal{T} \times \mathcal{U}$ with $\mathcal{T} \in \sigma(\mathcal{X})$ and $\mathcal{U} \in \sigma(\mathcal{Z})$, that satisfies

$$P_{\Xi,\Sigma}(\mathcal{T},\mathcal{U}) = P(\Xi^{-1}(\mathcal{T}) \cap \Sigma^{-1}(\mathcal{U})).$$

The joint probability distribution of a finite number of random variables can be defined in a similar way.

VII. APPENDIX B (RANDOM FINITE SETS)

For completeness, this Appendix outlines the basics of random finite sets (RFS) or simple finite point processes[8].

---

[8]A simple finite point process set does not allow repeated elements and only contains a finite number of elements.

Background material on RFS are abundant in the point processes literature; see for example [8], [43]. However, works with an inclination to multi-target filtering are quite new; the major body of work appears to be that of Mahler [15], [24]. The monograph [25] and the thesis [45] are excellent introductions accessible to a wide range of readers.

Given a locally compact Hausdorff separable space E (e.g. $\mathbf{R}^n$), let $\mathcal{F}(E)$ denote the collection of finite subsets of E. The topology on $\mathcal{F}(E)$ is taken to be the myopic or Mathéron topology [31]. A *random finite set* Ξ on E is defined as a measurable mapping

$$\Xi : \Omega \to \mathcal{F}(E),$$

where Ω is a sample space with a probability measure P defined on $\sigma(\Omega)$. The probability measure P induces a *probability law* for Ξ, which can be specified in terms of a probability distribution, a void probability or a belief function. The most natural description of the probability law for Ξ is the probability distribution P_Ξ defined for any Borel subset $\mathcal{T}$ of $\mathcal{F}(E)$ by

$$P_\Xi(\mathcal{T}) = P(\Xi^{-1}(\mathcal{T})) = P(\{\omega : \Xi(\omega) \in \mathcal{T}\}).$$

However, from random set theory [15], [31], the probability law for Ξ can also be given in terms of the *belief mass function* β_Ξ defined for any closed subset S of E by

$$\beta_\Xi(S) = P(\{\omega : \Xi(\omega) \subseteq S\}).$$

A third equivalent description, closely related to the belief mass function, is the *void probability* ς_Ξ [8], [43], [44], which is defined for any closed subset S of E by

$$\varsigma_\Xi(S) = P(\{\omega : |\Xi(\omega) \cap S| = 0\}) = \beta_\Xi(S^c),$$

where $|X|$ denotes the number of elements in X.

The simplest class of RFSs are the *Poisson point processes*. A Poisson point process Υ is a RFS characterised by the property that for any k disjoint Borel subsets $S_1, ..., S_k$ of E, the random variables $|\Upsilon \cap S_1|, ..., |\Upsilon \cap S_k|$ are independent and Poisson. Let $v_\Upsilon(S)$ denote the mean of the Poisson random variable $|\Upsilon \cap S|$. Then v_Υ defines a (unitless) measure on the Borel subsets of E, and is called the *intensity measure* of Υ [8], [43], [44]. The probability distribution of Υ is given by [13], [34]

$$P_\Upsilon(\mathcal{T}) = e^{-v_\Upsilon(E)} \sum_{i=0}^{\infty} \frac{v_\Upsilon^i(\chi^{-1}(\mathcal{T}) \cap E^i)}{i!}, \qquad (33)$$

where v_Υ^i denotes the ith product measure of v_Υ and $\chi : \uplus_{i=0}^{\infty} E^i \to \mathcal{F}(E)$ is the mapping of vectors to finite sets defined for each i by $\chi([x_1, ..., x_i]^T) = \{x_1, ..., x_i\}$. The mapping χ is measurable [44] and hence P_Υ is well defined. A word of caution, it is common practice in the stochastic geometry literature to write the measure (33) with the following abuse of notation [13], [34]

$$P_\Upsilon(\mathcal{T}) = e^{-v_\Upsilon(E)} \sum_{i=0}^{\infty} \frac{v_\Upsilon^i(\mathcal{T} \cap E^i)}{i!}, \qquad (34)$$

where it is implicit that $\mathcal{T} := \chi^{-1}(\mathcal{T})$, i.e. vectors are considered as finite sets and vice-versa depending on the context of the expression.

The integral of a measurable function $f : \mathcal{F}(E) \to \mathbf{R}$ with respect to the measure

$$\mu(\mathcal{T}) = \sum_{i=0}^{\infty} \frac{v_{\Upsilon}^i(\chi^{-1}(\mathcal{T}) \cap E^i)}{i!}, \qquad (35)$$

is given by [13], [34]

$$\int_{\mathcal{T}} f(X)\mu(dX)$$
$$= \sum_{i=0}^{\infty} \frac{1}{i!} \int_{\chi^{-1}(\mathcal{T}) \cap E^i} f(\{x_1,...,x_i\})v_{\Upsilon}^i(dx_1...dx_i). \quad (36)$$

This is straight forward to verify using the countable additivity of measure. Decompose $\mathcal{T} = \biguplus_{i=0}^{\infty} \mathcal{T}_i$, where $\mathcal{T}_i$ is the subset of $\mathcal{T}$ which contains all (finite) subsets with i elements, and note that $\chi^{-1}(\mathcal{T}_i) \cap E^i = \chi^{-1}(\mathcal{T}) \cap E^i$, then

$$\int_{\mathcal{T}} f(X)\mu(dX)$$
$$= \sum_{i=0}^{\infty} \int_{\mathcal{T}_i} f(X)\mu(dX)$$
$$= \sum_{i=0}^{\infty} \frac{1}{i!} \int_{\chi^{-1}(\mathcal{T}_i) \cap E^i} f(\{x_1,...,x_i\})v_{\Upsilon}^i(dx_1...dx_i)$$
$$= \sum_{i=0}^{\infty} \frac{1}{i!} \int_{\chi^{-1}(\mathcal{T}) \cap E^i} f(\{x_1,...,x_i\})v_{\Upsilon}^i(dx_1...dx_i).$$

For any Borel subset S of E, let $\lambda_K(S)$ denote the Lebesgue measure (or volume) of S in units of K. The density of v_{Υ} w.r.t. λ_K (if one exists) is called an *intensity function* or *rate* of Υ and has units of K^{-1}. A Poisson point process is completely characterised by its intensity measure (or equivalently its rate). A Poisson point process with a uniform rate of K^{-1} has intensity measure $\lambda = \lambda_K/K$ and its probability distribution is given by [13], [34] i.e.

$$P_{\Upsilon}(\mathcal{T}) = e^{-\lambda(E)} \sum_{i=0}^{\infty} \frac{\lambda^i(\chi_i^{-1}(\mathcal{T}) \cap E^i)}{i!}. \qquad (37)$$

VIII. APPENDIX C (PROOF OF PROPOSITION 3)

To simplify notations, let $\alpha_k = D_{k|k}$, $\pi_k = D_{k|k-1}$, $\phi_k = \phi_{k|k-1}$, $\alpha_k^L = \sum_{i=1}^{L} w_k^{(i)} \delta_{x_k^{(i)}}$, $\tilde{\alpha}_k^{L+J} = \sum_{i=1}^{L+J} \tilde{w}_k^{(i)} \delta_{\tilde{x}_k^{(i)}}$, $\pi_k^{L+J} = \sum_{i=1}^{L+J} \tilde{w}_{k|k-1}^{(i)} \delta_{\tilde{x}_k^{(i)}}$ and $d_k^L = \widehat{N}_{k|k}$. Note that $d_k^L = \sum_{i=1}^{L} w_k^{(i)} = \sum_{i=1}^{L+J} \tilde{w}_k^{(i)}$.

Given $\phi : E_s \times E_s \to \mathbf{R}$, for a function $f : E_s \to \mathbf{R}$ and a density[9] α on E_s, define $f\alpha \in \mathbf{R}$, $\phi\alpha : E_s \to \mathbf{R}$ and $f\phi : E_s \to \mathbf{R}$ by

$$f\alpha = \int f(x)\alpha(x)\lambda(dx),$$
$$(\phi\alpha)(x) = \int \phi(x,\xi)\alpha(\xi)\lambda(d\xi),$$
$$(f\phi)(x) = \int f(\xi)\phi(\xi,x)\lambda(d\xi).$$

[9] All densities on $\mathbf{R}^{n_x}$ are with respect to the Lebesgue measure.

Note that $f\phi\alpha = f(\phi\alpha) = (f\phi)\alpha$, and $\Phi_{k|k-1}\alpha = \phi_k\alpha$. Since $f_{k|k-1}$ and $b_{k|k-1}$ are Feller (a standard assumption [11, Chapter 2]) $\phi_k = \phi_{k|k-1}$ is also Feller, i.e. $f\phi_k \in C_b(E_s)$ for all $f \in C_b(E_s)$, where $C_b(E_s)$ denotes the space of bounded and continuous functions on E_s.

Define the σ-algebras

$$\bar{\mathcal{F}}_k = \sigma\left\{x_r^{(i)}, \tilde{x}_s^{(i)}; r < k, \ s \le k, i = 1,...,L+J\right\}, (38)$$
$$\mathcal{F}_k = \sigma\left\{x_r^{(i)}, \tilde{x}_s^{(i)}; r \le k, \ s \le k, i = 1,...,L+J\right\}. (39)$$

Observe that $\left(\tilde{w}_k^{(i)}\right)_{i=1}^{L+J}$ are $\bar{\mathcal{F}}_k$-measurable. All expectations defined in this Appendix assume a fixed sequence of observations $Z_{1:k}$. Conditioning on $Z_{1:k}$ is omitted to simplify the notation.

Proposition 3 is established by proving Lemma C.1 below, and then showing that the premises for Lemma C.1 are satisfied.

Lemma C.1: *If the following three conditions hold for all* $f \in C_b(E_s)$:

$$\lim_{L\to\infty} \mathbf{E}\left[\left|f\alpha_0^L - f\alpha_0\right|\right] = 0, \qquad (40)$$
$$\lim_{L\to\infty} \mathbf{E}\left[\left|f\pi_k^{L+J} - f(\phi_k\alpha_{k-1}^L)\right|\right] = 0, \qquad (41)$$
$$\lim_{L\to\infty} \mathbf{E}\left[\left|f\alpha_k^L - f\tilde{\alpha}_k^{L+J}\right|\right] = 0. \qquad (42)$$

Then, for all $f \in C_b(E_s)$,

$$\lim_{L\to\infty} \mathbf{E}\left[\left|f\pi_k^{L+J} - f\pi_k\right|\right] = 0, \qquad (43)$$
$$\lim_{L\to\infty} \mathbf{E}\left[\left|f\alpha_k^L - f\alpha_k\right|\right] = 0. \qquad (44)$$

Proof: To establish (43), recall that $f\pi_k = f\phi_k\alpha_{k-1}$, and write

$$\left|f\pi_k^{L+J} - f\pi_k\right| \le \left|f\pi_k^{L+J} - f\phi_k\alpha_{k-1}^L\right|$$
$$+ \left|f\phi_k\alpha_{k-1}^L - f\phi_k\alpha_{k-1}\right|. \quad (45)$$

The expectation of the first term on the right of (45) converges to zero as $L \to \infty$ by virtue of (41). Since $f\phi_k \in C_b(E_s)$ for all $f \in C_b(E_s)$, the expected value of the second term converges to zero by induction and (40).

Similarly, to establish (44) write

$$\left|f\alpha_k^L - f\alpha_k\right| \le \left|f\alpha_k^L - f\tilde{\alpha}_k^{L+J}\right| + \left|f\tilde{\alpha}_k^{L+J} - f\alpha_k\right|. \quad (46)$$

Using (42), the expectation of the first term on the right of (46) converges to zero. For the second term, write

$$\left|f\tilde{\alpha}_k^{L+J} - f\alpha_k\right|$$
$$= \left|f(\Psi_k\pi_k^{L+J}) - f(\Psi_k\pi_k)\right|$$
$$\le \left|(f\times\upsilon_k)\pi_k^{L+J} - (f\times\upsilon_k)\pi_k\right|$$
$$+ \sum_{z\in Z_k} \left|\frac{(f\times\psi_{k,z})\pi_k^{L+J}}{\kappa_k(z)+C_k(z)} - \frac{(f\times\psi_{k,z})\pi_k}{\kappa_k(z)+\psi_{k,z}\pi_k}\right|. (47)$$

Using (43) just established, the expected value of the first term on the right of (47) converges to zero. Moreover, each term

in the sum over Z_k is bounded by

$$
\left| \frac{(f \times \psi_{k,z}) \pi_k^{L+J}}{\kappa_k(z) + C_k(z)} - \frac{(f \times \psi_{k,z}) \pi_k}{\kappa_k(z) + \psi_{k,z} \pi_k} \right|
$$
$$
\leq \left| \frac{(f \times \psi_{k,z}) \pi_k^{L+J}}{\kappa_k(z) + C_k(z)} - \frac{(f \times \psi_{k,z}) \pi_k^{L+J}}{\kappa_k(z) + \psi_{k,z} \pi_k} \right|
$$
$$
+ \left| \frac{(f \times \psi_{k,z}) \pi_k^{L+J}}{\kappa_k(z) + \psi_{k,z} \pi_k} - \frac{(f \times \psi_{k,z}) \pi_k}{\kappa_k(z) + \psi_{k,z} \pi_k} \right| . \quad (48)
$$

Again using (43), the expectation of the second term on the right of (48) converges to zero. The first term on the right of (48) can be rearranged to give (49). Noting that $\left| (f \times \psi_{k,z}) \pi_k^{L+J} \right| \leq \|f\|_\infty \left| \psi_{k,z} \pi_k^{L+J} \right|$, and $C_k(z) = \psi_{k,z} \pi_k^{L+J}$ yields (50)

$$
\left| \frac{(f \times \psi_{k,z}) \pi_k^{L+J}}{\kappa_k(z) + C_k(z)} - \frac{(f \times \psi_{k,z}) \pi_k^{L+J}}{\kappa_k(z) + \psi_{k,z} \pi_k} \right|
$$
$$
\leq \left| \frac{(f \times \psi_{k,z}) \pi_k^{L+J}}{\kappa_k(z) + C_k(z)} \right| \left| \frac{\psi_{k,z} \pi_k - C_k(z)}{\kappa_k(z) + \psi_{k,z} \pi_k} \right| \quad (49)
$$
$$
\leq \|f\|_\infty \left| \frac{\psi_{k,z} \pi_k - \psi_{k,z} \pi_k^{L+J}}{\kappa_k(z) + \psi_{k,z} \pi_k} \right| . \quad (50)
$$

The expectation of the term on right of (50) converges to zero by virtue of (43). Consequently, the expectation of the sum over Z_k in (47) converges to zero and hence, (44) follows. ∎

The premises of Lemma C.1 are now established by the following lemmas.

Lemma C.2: *If* $\lim_{L \to \infty} L^{-1} \mathbf{E} \left[\left(d_{k-1}^L \right)^2 \right] = 0$ *then,* $\lim_{L \to \infty} L^{-1} \mathbf{E} \left[\left(d_k^L \right)^2 \right] = 0$.

Proof: Substitute for $\tilde{w}_k^{(i)}$ from (28) into $d_k^L = \sum_{i=1}^{L+J} \tilde{w}_k^{(i)}$ and using (27) for $C_k(z)$ gives.

$$
d_k^L = \sum_{i=1}^{L+J} \tilde{w}_{k|k-1}^{(i)} \upsilon_k(\tilde{x}_k^{(i)}) + \sum_{z \in \mathcal{Z}_k} \frac{C_k(z)}{\kappa_k(z) + C_k(z)}
$$
$$
\leq \sum_{i=1}^{L+J} \tilde{w}_{k|k-1}^{(i)} \upsilon_k(\tilde{x}_k^{(i)}) + |Z_k| \quad (51)
$$

Substituting for $\tilde{w}_{k|k-1}^{(i)}$ from (25-26) into the first term on the right of (51) and using $w_{k-1}^{(i)} = d_{k-1}^L / L$ for $i = 1, ..., L$ gives

$$
d_k^L \leq \frac{d_{k-1}^L}{L} \sum_{i=1}^{L} \left[\frac{\phi_k(\tilde{x}_k^{(i)}, x_{k-1}^{(i)})}{q_k(\tilde{x}_k^{(i)} | x_{k-1}^{(i)})} \right] \upsilon_k(\tilde{x}_k^{(i)})
$$
$$
+ \frac{1}{J} \sum_{i=L+1}^{L+J} \left[\frac{\gamma_k(\tilde{x}_k^{(i)})}{p_k(\tilde{x}_k^{(i)})} \right] \upsilon_k(\tilde{x}_k^{(i)}) + |Z_k|
$$
$$
\leq d_{k-1}^L \sup_{\xi, x} \left| \frac{\phi_k(x, \xi)}{q_k(x|\xi)} \upsilon_k(x) \right|
$$
$$
+ \sup_x \left| \frac{\gamma_k(x)}{p_k(x)} \upsilon_k(x) \right| + |Z_k| .
$$

Using the bounds (29)-(30), the result then follows. ∎

Lemma C.3: *For all* $f \in C_b(E_s)$, $\lim_{L \to \infty} \mathbf{E} \left[\left| f \alpha_k^L - f \tilde{\alpha}_k^{L+J} \right| \right]$.

Proof:

$$
\mathbf{E}_{\bar{\mathcal{F}}_k} \left[\left| f \tilde{\alpha}_k^{L+J} - f \alpha_k^L \right|^2 \right]
$$
$$
= \mathbf{E}_{\bar{\mathcal{F}}_k} \left[\left(\frac{d_k^L}{L} \right)^2 \left(\sum_{i=1}^{L+J} f(\tilde{x}_k^{(i)}) \left(\frac{L \tilde{w}_k^{(i)}}{d_k^L} - \zeta_k^{(i)} \right) \right)^2 \right]
$$
$$
= \left(\frac{d_k^L}{L} \right)^2 \sum_{i=1}^{L+J} \sum_{j=1}^{L+J} f(\tilde{x}_k^{(i)}) \mathbf{E}_{\bar{\mathcal{F}}_k} \left[\left(\frac{L \tilde{w}_k^{(i)}}{d_k^L} - \zeta_k^{(i)} \right) \right.
$$
$$
\left. \times \left(\frac{L \tilde{w}_k^{(j)}}{d_k^L} - \zeta_k^{(j)} \right) \right] f(\tilde{x}_k^{(j)})
$$
$$
= \left(\frac{d_k^L}{L} \right)^2 \|f\|_\infty^2 \sum_{i=1}^{L+J} \sum_{j=1}^{L+J} \frac{f(\tilde{x}_k^{(i)})}{\|f\|_\infty} [A_k]_{i,j} \frac{f(\tilde{x}_k^{(j)})}{\|f\|_\infty}
$$
$$
\leq \left(\frac{d_k^L}{L} \right)^2 \|f\|_\infty^2 L c_k,
$$

where the last 2 steps follow from (31) and (32). Thus, using Lemma C.2, $\mathbf{E} \left[\left| f \tilde{\alpha}_k^{L+J} - f \alpha_k^L \right|^2 \right] \to 0$ as $L \to \infty$. ∎

Lemma C.4: *For all* $f \in C_b(E_s)$, $\lim_{L \to \infty} \mathbf{E} \left[\left| f \pi_k^{L+J} - f(\phi_k \alpha_{k-1}^L) \right| \right] = 0$.

Proof:

$$
f \pi_k^{L+J} - f \phi_k \alpha_{k-1}^L
$$
$$
= \sum_{i=1}^{L+J} f(\tilde{x}_k^{(i)}) \tilde{w}_{k|k-1}^{(i)} - \sum_{i=1}^{L} (f \phi_k)(x_{k-1}^{(i)}) w_{k-1}^{(i)} - f \gamma_k
$$
$$
= \sum_{i=1}^{L} \left[f(\tilde{x}_k^{(i)}) \tilde{w}_{k|k-1}^{(i)} - (f \phi_k)(x_{k-1}^{(i)}) w_{k-1}^{(i)} \right]
$$
$$
+ \sum_{i=L+1}^{L+J} f(\tilde{x}_k^{(i)}) \tilde{w}_{k|k-1}^{(i)} - f \gamma_k. \quad (52)
$$

For the first term on the right of (52), substitute for $\tilde{w}_{k|k-1}^{(i)}$ from (25) and $w_{k-1}^{(i)} = d_{k-1}^L / L$. Take the expectation of its square and using the independence of the cross terms conditional on $\mathcal{F}_{k-1}$ gives (53). Expanding the square inside the expectation on the right of (53) and taking the expectation inside this expansion gives (54)

$$
\mathbf{E}_{\mathcal{F}_{k-1}} \left\{ \left[\sum_{i=1}^{L} \left(f(\tilde{x}_k^{(i)}) \tilde{w}_{k|k-1}^{(i)} - (f \phi_k)(x_{k-1}^{(i)}) w_{k-1}^{(i)} \right) \right]^2 \right\}
$$
$$
= \left(\frac{d_{k-1}^L}{L} \right)^2 \sum_{i=1}^{L} \mathbf{E}_{\mathcal{F}_{k-1}} \left[\left(f(\tilde{x}_k^{(i)}) \frac{\phi_k(\tilde{x}_k^{(i)}, x_{k-1}^{(i)})}{q_k(\tilde{x}_k^{(i)} | x_{k-1}^{(i)})} \right. \right.
$$
$$
\left. \left. - (f \phi_k)(x_{k-1}^{(i)}) \right)^2 \right] \quad (53)
$$
$$
= \left(\frac{d_{k-1}^L}{L} \right)^2 \sum_{i=1}^{L} \left[\int \frac{f^2(x) \phi_k^2(x, x_{k-1}^{(i)})}{q_k(x | x_{k-1}^{(i)})} dx \right.
$$
$$
\left. - \left((f \phi_k)(x_{k-1}^{(i)}) \right)^2 \right] \quad (54)
$$
$$
\leq \frac{(d_{k-1}^L)^2}{L} \|f\|_\infty^2 \sup_{\xi, x} \left| \frac{\phi_k(\xi, x)}{q_k(\xi | x)} \right| + \frac{(d_{k-1}^L)^2}{L} \sup_x (f \phi_k)^2 (x).
$$

Using (29), it follows that the expectation of the first term on the right of (52) converges to zero as $L \to \infty$.

Similarly, for the second term on the right of (52), substitute for $\tilde{w}_{k|k-1}^{(i)}$ from (26) and take the expectation of its square, keeping in mind the independence of the cross terms conditional on $\mathcal{F}_{k-1}$, gives (55). Expanding the square inside the expectation on the right of (55) and taking the expectation inside this expansion gives (56)

$$\mathbf{E}_{\mathcal{F}_{k-1}}\left[\left(\frac{1}{J}\sum_{i=L+1}^{L+J} f(\tilde{x}_k^{(i)})\frac{\gamma_k(\tilde{x}_k^{(i)})}{p_k(\tilde{x}_k^{(i)})} - f\gamma_k\right)^2\right]$$

$$= \frac{1}{J^2}\sum_{i=1}^{L+J}\mathbf{E}_{\mathcal{F}_{k-1}}\left[\left(f(\tilde{x}_k^{(i)})\frac{\gamma_k(\tilde{x}_k^{(i)})}{p_k(\tilde{x}_k^{(i)})} - f\gamma_k\right)^2\right] \quad (55)$$

$$= \frac{1}{J^2}\sum_{i=1}^{L+J}\left[\int \frac{f^2(x)\gamma_k^2(x)}{p_k(x)}dx - (f\gamma_k)^2\right] \quad (56)$$

$$\leq \frac{\|f\|_\infty^2}{J}\left[\int \frac{\gamma_k^2(x)}{p_k(x)}dx\right] - \frac{(f\gamma_k)^2}{J}.$$

Using (30), $\int \frac{\gamma_k^2(x)}{p_k(x)}dx \leq P_k \int \gamma_k(x)dx$. Moreover, $J \to \infty$ as $L \to \infty$, hence, the above expectation converges to zero. Consequently, the second term on the right of (52) also converges to zero. ∎

REFERENCES

[1] Arulampalam S., Maskell S., Gordon N. J., and Clapp T., "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking", *IEEE Trans. Signal Processing*, Vol. 50, No. 2, pp. 174-188, 2002.

[2] Baddeley A.J., and van Lieshout M.N., "ICM for object recognition". In Y. Dodge and J. Whittaker (eds.), *Computational Statistics*, Vol. 2. Heidelberg: Physica/Springer, pp. 271-286, 1992.

[3] Ballantyne D. J., Chan H. Y., and Kouritzin M. A., "A branching particle-based nonlinear filter for multi-target tracking", *Proc. Int'l Conf. on Information Fusion*, Montreal 2001.

[4] Bar-Shalom Y., and Fortmann T. E., *Tracking and Data Association*, Academic Press, San Diego, 1988.

[5] Bar-Shalom Y., and Li X-R., *Multitarget-Multisensor Tracking: Principles and Techniques*, Storrs, CT: YBS Publishing, 1995.

[6] Blackman S., Popoli R., *Design and Analysis of Modern Tracking Systems*, Artech House, Boston, 1999.

[7] Crisan D., "Particle filters-A theoretical perspective," in *Sequential Monte Carlo Methods in Practice*, Doucet A., de Freitas N. and Gordon N. J., (eds.), pp. 17-41, Springer-Verlag, May 2001.

[8] Daley D., and Vere-Jones D., *An Introduction to the Theory of Point Processes*, Springer-Verlag, 1988.

[9] Del Moral P., *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer-Verlag: New York, 2004.

[10] Doucet A., Godsill S. J., and Andrieu C., "On sequential Monte Carlo sampling methods for Bayesian filtering," *Stat. Comp.*, **10**, pp. 197-208, 2000.

[11] Doucet A., de Freitas N., and Gordon N. J., *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, May 2001.

[12] Doucet A., Vo B., Andrieu C., and Davy M., "Particle filtering for multitarget tracking and sensor management", *Proc. Int'l Conf. on Information Fusion*, pp. 474-481, Annapolis Maryland, 2002.

[13] Geyer C., "Likelihood inference for spatial point processes", in *Stochastic Geometry Likelihood and Computation*, Barndorff-Nielsen et. al (eds.), pp. 79-140, 1999.

[14] Gilks W., and Berzuini C., "Following a moving target-MonteCarlo inference for dynamic Bayesian models", Journal of the Royal Statistical Society, Series B, Vol. 63, pp. 127-146, 2001.

[15] Goodman I., Mahler R., and Nguyen H., *Mathematics of Data Fusion*, Kluwer Academic Publishers, 1997.

[16] Gordon N. J., Salmond D., and Smith A., "Novel approach to nonlinear/non-Gaussian Bayesian state estimation", *IEE Proceedings-F* Vol. 140, No. 2, pp. 107-113, 1993.

[17] Goutsias J., Mahler R., and Nguyen H. (eds.), *Random Sets Theory and Applications*, Springer-Verlag New York, 1997.

[18] Green P. J., "Reversible jump MCMC computation and Bayesian model determination", *Biometrika*, Vol. 82, pp. 711-732, 1995.

[19] Gupta M. M., Jin L., and Homma N., *Static and Dynamic neural networks*, John Wiley and Sons, Inc., 2003.

[20] Hoffman J., and Mahler R., "Multitarget miss distance via optimal assignment", *IEEE Trans. Sys., Man, and Cybernetics-Part A*, Vol. 34, No. 3, pp. 327-336, 2004.

[21] Hue C., Le Cadre J.-P., and Perez P., "Tracking multiple objects with particle filtering", *IEEE Trans. AES*, Vol. 38, No. 3, pp. 791-812, 2002.

[22] Isard M., and MacCormick J., "BraMBLe: A Bayesian multi-blob tracker", *Proc. Eighth Int'l Conf. on Computer Vision*, Vancouver, Vol. 3, pp. 43-41, July 2001.

[23] Kitagawa G., "Monte Carlo filter and smoother for non-Gaussian non-linear state space models", *J. Comp. Graph. Stat.*, Vol. 5, No. 1, pp. 1-25, 1996.

[24] Mahler R., "Global integrated data fusion" , *Proc. 7th Nat. Symp. on Sensor Fusion*, Vol. 1, (Unclassified) Sandia National Laboratories, Albuquerque, ERIM Ann Arbor MI, pp. 187-199, 1994.

[25] Mahler R., *An Introduction to Multisource-Multitarget Statistics and Applications*, Lockheed Martin Technical Monograph. March 15, 2000.

[26] Mahler R., "A theoretical foundation for the Stein-Winter Probability Hypothesis Density (PHD) multi-target tracking approach," *Proc.2002 MSS Nat'l Symp. on Sensor and Data Fusion*, Vol. I (Unclassified), San Antonio TX, June 2000.

[27] Mahler R., "Multi-target moments and their application to multi-target tracking", *Proc. Workshop on Estimation, Tracking and Fusion*: A tribute to Yaakov Bar-Shalom, Monterey, pp. 134-166, 2001.

[28] Mahler R., "Random set theory for target tracking and identification", *Data Fusion Hand Book*, D. Hall and J. Llinas (eds.), CRC press Boca Raton, pp. 14/1-14/33, 2001.

[29] Mahler R., "Multi-target Bayes filtering via first-order multi-target moments", *IEEE Trans. AES*, Vol. 39, No. 4, pp. 1152-1178, 2003.

[30] Maskell S., Gordon N. J., Rollason M., and Salmond D., "Efficient particle filtering for multi-target tracking with application to tracking in structured images," *Signal and Data Processing of Small Targets, Proc. SPIE,* Vol. 4728, 2002.

[31] Mathéron G., *Random Sets and Integral Geometry*, J. Wiley, 1975.

[32] Miller M. I., Teichman R. S., Srivastava A., O'Sullivan J. A., and Snyder D. L., "Jump-Diffusion processes for automated tracking-target recognition," *Proc. Conf. on Information Science and Systems,* Baltimore, Maryland, March 1993.

[33] Miller M. I., Srivastava A., and Grenander U., "Conditional-mean estimation via jump-diffusion processes in multiple target tracking/recognition," *IEEE Trans. Signal Processing*, Vol. 43, No. 11, pp. 2678-2690, 1995.

[34] Moller J., "Markov chain Monte Carlo and spatial point processes", in *Stochastic Geometry Likelihood and Computation*, Barndorff-Nielsen et. al (eds.), pp. 141-172, 1999.

[35] Mori S., Chong C., Tse E., and Wishner R., "Tracking and identifying multiple targets without apriori identifications", *IEEE Trans. Automatic Control*, Vol. AC-21, pp. 401-409, 1986.

[36] Orton M., and Fitzgerald W., "A Bayesian approach to tracking multiple targets using sensor arrays and particle filters", *IEEE Trans. Signal Processing*, Vol. 50, No. 2, pp. 216-223, 2002.

[37] Portenko N., Salehi H., and Skorokhod A., "On optimal filtering of multitarget systems based on point process observations", *Random Operators and Stochastic Equations,* Vol. 5, No. 1, pp. 1-34, 1997.

[38] Shiryaev A., *Probability*, 2nd Ed., Springer-Verlag, 1996.

[39] Sidenbladh H., and Wirkander S. L., "Tracking random sets of vehicles in terrain", *Proc. 2003 IEEE Workshop on Multi-Object Tracking*, Madison Wisconsin, 2003.

[40] Sidenbladh H., "Multi-target particle filtering for the Probability Hypothesis Density", *Proc. Int'l Conf. on Information Fusion*, Cairns, Australia, pp. 800-806, 2003.

[41] Srivastava A., Miller M.I., and Grenander U., "Jump-Diffusion processes for tracking and direction finding", *Proc. 29th Allerton Conf. on Communication, Control, and Computing,* University of Illinois Urbana, pp. 563-570, 1991.

[42] Stein M. C., and Winter C. L., "An adaptive theory of probabilistic evidence accrual," *Los Alamos National Laboratotries Report,* LA-UR-93-3336, 1993.

[43] Stoyan D., Kendall D., and Mecke J., *Stochastic Geometry and its Applications*, John Wiley & Sons, 1995.

[44] van Lieshout M. N., *Markov Point Processes and their Applications*, Imperial College Press, 2000.

[45] Vihola M., *Random Sets for Multitarget Tracking and Data Fusion*, Licentiate Thesis, Tampere University of Technology, 2004.

[46] Vo B., Singh S., and Doucet A., "Sequential Monte Carlo implementation of the PHD filter for multi-target tracking," *Proc. Int'l Conf. on Information Fusion*, Cairns, Australia, pp. 792- 799, 2003.

[47] Washburn R., "A random point process approach to multi-object tracking", *Proc. American Control Conf.*, vol. 3, pp. 1846-1852, 1987.

[48] Zajic T., and Mahler R., "A Particle-Systems implementation of the PHD multi-target tracking filter," in I. Kadar (ed.), *Signal Processing, Sensor Fusion and Target Recognition XII, Proc. SPIE,* Vol. 5096, pp. 291-299, 2003.