A second-order PHD filter with mean and variance in target number

Isabel Schlangen*, Emmanuel Delande*, Jérémie Houssineau* and Daniel E. Clark*

Abstract—The Probability Hypothesis Density (PHD) and Cardinalized PHD (CPHD) filters are popular solutions to the multi-target tracking problem due to their low complexity and ability to estimate the number and states of targets in cluttered environments. The PHD filter propagates the first-order moment (i.e. mean) of the number of targets while the CPHD propagates the cardinality distribution in the number of targets, albeit for a greater computational cost. Introducing the Panjer point process, this paper proposes a Second-Order PHD (SO-PHD) filter, propagating the second-order moment (i.e. variance) of the number of targets alongside its mean. The resulting algorithm is more versatile in the modelling choices than the PHD filter, and its computational cost is significantly lower compared to the CPHD filter. The paper compares the three filters in statistical simulations which demonstrate that the proposed filter reacts more quickly to changes in the number of targets, i.e., target births and target deaths, than the CPHD filter. In addition, a new statistic for multi-object filters is introduced in order to study the correlation between the estimated number of targets in different regions of the state space, and propose a quantitative analysis of the spooky effect for the three filters.

I. INTRODUCTION

In the context of multi-target detection and tracking problems, methods based on the Random Finite Set (RFS) framework have recently attracted a lot of attention due to the development of low-complexity algorithms within this methodology [1]. The best-known algorithm is perhaps the Probability Hypothesis Density (PHD) filter that jointly estimates the number of targets and their states by propagating the first-order moment of a RFS [2]; a Gaussian Mixture (GM) and a Sequential Monte Carlo (SMC) implementation have been presented in [3] and [4].

Erdinc and Willett [5] suggested that only propagating the first-order moment did not provide sufficient information for applications where a high confidence in the target number was needed. Consequently, Mahler derived the Cardinalized PHD (CPHD) filter which propagates the cardinality distribution of the target point process alongside its first-order moment [6]. It thus provides higher-order information on the number of targets, but to the expense of a higher computational cost. Around the same time, he also proposed a filter restricted to the first two moments using a binomial approximation [7]. However, due to the binomial approximation it was suggested that restrictions were required on the relative number of false

alarms and targets. In 2007, Vo et al. showed that the CPHD filter can be overconfident in some cases [8], and in 2009, Fränken et al. identified a counter-intuitive property of the CPHD filter that occurs with the weights of the targets when they are miss-detected which they called the spooky effect [9]. An alternative approach for extending the PHD filter to a second-order filter was proposed by Singh et al. using a Gauss-Poisson prior [10], though this was designed specifically for tracking correlated pairs of targets.

Other developments in the Finite Set Statistics (FISST) framework have focussed on more advanced filtering solutions with higher complexity. The Multi-Bernoulli (MeMBer) filter [1] is based on a fully Bayesian approach where the system assumes that each target is modelled by a state estimate and a probability of existence. The bias in the number of targets in the original MeMBer filter was addressed in [11], and further developments around Bernoulli RFSs were introduced in [12], [13]. Various methods propagating information on individual targets within the FISST framework have been developed since [14]–[16].

This paper focuses on a filtering solution with low complexity. We introduce a second-order filter in which the predicted target process is assumed Panjer instead of Poisson. The Panjer distribution [17] is specified by two parameters and encompasses the binomial, Poisson and negative binomial distributions; unlike a Poisson distribution, it can describe a population of targets whose estimated size has a higher or lower variance than its mean. The proposed Second-Order PHD (SO-PHD) filter thus complements the original PHD filter with the variance in the estimated number of targets; it also propagates less information than the CPHD filter but has a lower computational cost. The Panjer distribution was studied for the analysis of the CPHD update in [9], though it was not used to develop a new filter. The proposed filter can also be seen as a generalisation of the PHD filter with a negative binomial-distributed false alarms [18], which was designed for scenarios with high variability in background noise.[1] We also introduce the correlation for point processes in disjoint regions of the state space, in order to analyse quantitatively the spooky effect [9] for the considered filters.

Sec. II introduces some background material on point processes. Sec. III presents four relevant point processes, then used in Sec. IV to construct the proposed SO-PHD filter. The regional correlation for the PHD, SO-PHD, and CPHD filters are introduced in Sec. V. A comparison of the GM

---

[1] A follow-on work based on the results in the current paper has been developed for joint parameter estimation and multi-target tracking [19].

implementations of the three filters is given in Sec. VI, and Sec. VII concludes. Pseudo-code for the proposed algorithms and detailed proofs are given in appendices.

II. BACKGROUND

Many recent works in multi-object filtering exploit Mahler's FISST framework [1], in which multi-target state configurations are described by RFSs. The FISST framework allows for the production of the densities of various statistical quantities describing a RFS (multi-object density, Probability Hypothesis Density, etc.) through the set derivative operator.

This paper considers higher-order statistical quantities whose expression arises naturally from probability *measures* rather than *densities*, such as the regional covariance, or does not admit a density altogether, such as the regional variance or correlation (see Sec. II-C). Hence we shall favour the measure-theoretical formulation originating from the point process theory, for which a specific methodology has been developed to construct higher-order statistical moment measures or densities through the chain derivative operator [20]. Let us furthermore assume that the theorem of Radon and Nikodým [21] holds for all measures studied in this article, i.e. that they admit densities.

In the section, we provide the necessary background material on point processes, and highlight the connections with the FISST framework when appropriate. For the rest of the paper, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space with sample space Ω, σ-algebra $\mathcal{F}$, and probability measure $\mathbb{P}$. Throughout the paper, all random variables are defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and we denote by $\mathbb{E}$ the expectation with respect to (w.r.t.) $\mathbb{P}$.

A. Point processes

We denote by $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ the d_x-dimensional state space describing the state of an individual object (position, velocity, etc.). A point process Φ on $\mathcal{X}$ is a random variable on the process space $\mathfrak{X} = \bigcup_{n \geq 0} \mathcal{X}^n$, i.e., the space of finite sequences of points in $\mathcal{X}$. A realisation of Φ is a sequence $\varphi = (x_1, \ldots, x_n) \in \mathcal{X}^n$, representing a population of n objects with states $x_i \in \mathcal{X}$. Point processes can be described using their probability distribution P_Φ on the measurable space $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$, where $\mathcal{B}(\mathfrak{X})$ denotes the Borel σ-algebra of the process space $\mathfrak{X}$ [22].

The projection measure $P_\Phi^{(n)}$ of the probability distribution P_Φ on $\mathcal{X}^n$, $n \geq 0$, describes the realisations of Φ with n elements; the projection measures of a point process are always defined as symmetrical functions, so that the permutations of a realisation φ are equally probable. Furthermore, a point process is called *simple* if φ does not contain repetitions, i.e. its elements are pairwise distinct almost surely. For the rest of the paper, all point processes are assumed simple. In that case, it is assumed that the probability distribution P_Φ of a point process admits a density p_Φ w.r.t. some reference measure λ. The densities of the projection measures $P_\Phi^{(n)}$ are denoted by $p_\Phi^{(n)}$, and both quantities will be exploited throughout the paper.

In the literature originating from Mahler's FISST framework [2], [3], an alternative construction of simple point processes

is a RFS, a random object whose realizations are *sets* of points $\{x_1, \ldots, x_n\}$, in which the elements are by construction *unordered*.

B. Multi-target Bayesian filtering

In the context of multi-target tracking, we make use of a *target point process* Φ_k to describe the information about the target population at time k. The scene is observed by a sensor system, providing sets of measurements at discrete times (indexed by $k \in \mathbb{N}$ in the following). The d_z-dimensional observation space describing the individual measurements produced by the sensor (range, azimuth, etc.) is denoted by $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$. The set of measurements collected at time k is denoted by Z_k.

Point processes can be cast into a Bayesian framework in order to propagate Φ_k over time [1]. Bayesian filtering consists of a *prediction* or *time update* step which is concerned with the motion model, birth and death of targets, and a *data update* step which models the observation process, missed detections and false alarms and exploits the current measurement set Z_k.

The full multi-target Bayesian recursion propagates the law P_k of the target process Φ_k. The time prediction and data update equations at time k are given by [1]

$$P_{k|k-1}(\mathrm{d}\xi) = \int T_{k|k-1}(\mathrm{d}\xi|\varphi) P_{k-1}(\mathrm{d}\varphi), \qquad (1)$$

$$P_k(\mathrm{d}\xi|Z_k) = \frac{L_k(Z_k|\xi) P_{k|k-1}(\mathrm{d}\xi)}{\int L_k(Z_k|\varphi) P_{k|k-1}(\mathrm{d}\varphi)}, \qquad (2)$$

where $T_{k|k-1}$ is the multi-target Markov transition kernel from time $k-1$ to time k, and L_k is the multi-measurement/multi-target likelihood at time step k.[2] Note that the formulation of the multi-target Bayesian recursion with *measure-theoretical* integrals (1), (2) is drawn from its original RFS-based formulation in [1] with *set* integrals.

C. Statistical moments

Similar to real-valued random variables, statistical moments can be defined for a point process Φ in order to provide an alternative description to its probability distribution P_Φ (or, equivalently, to its projection measures $P_\Phi^{(n)}$ for any $n \in \mathbb{N}$). Statistical moments will play an important role in this paper, for the construction of the SO-PHD filter in Sec. IV as well as for the study of the correlation in the estimated target number in distinct regions of the state space in Sec. V.

The *n-th order moment measure* $\mu_\Phi^{(n)}$ of a point process Φ is the measure on $\mathcal{X}^n$ such that, for any bounded measurable function f_n on $\mathcal{X}^n$, it holds that [22]

$$\int f_n(x_{1:n}) \mu_\Phi^{(n)}(\mathrm{d}(x_{1:n})) = \mathbb{E}\left[\sum_{x_1, \ldots, x_n \in \Phi} f_n(x_{1:n}) \right] \qquad (3)$$

where we use the shorter notation $x_{1:n}$ to denote the sequence $(x_1, \ldots, x_n)$.[3] In addition, the *n-th order factorial moment*

---

[2] When μ, μ' are two measures on some space X, we use the notation $\mu(\mathrm{d}x) = \mu'(\mathrm{d}x)$, where $x \in X$, to indicate that $\int f(x)\mu(\mathrm{d}x) = \int f(x)\mu'(\mathrm{d}x)$ for any bounded measurable function f on X.

[3] When $\varphi \in X^n$, $n \geq 0$, is a sequence of elements on some space X, the abuse of notation "$x \in \varphi$" is used to denote that the element $x \in X$ appears in the sequence φ.

measure $\nu_\Phi^{(n)}$ of a point process Φ is the measure on $\mathcal{X}^n$ such that, for any bounded measurable function f_n on $\mathcal{X}^n$, it holds that [22]

$$\int f_n(x_{1:n})\nu_\Phi^{(n)}(\mathrm{d}(x_{1:n})) = \mathbb{E}\left[\sum_{x_1,\ldots,x_n\in\Phi}^{\neq} f_n(x_{1:n})\right] \quad (4)$$

where $\Sigma^{\neq}$ indicates that the selected points $x_1,\ldots,x_n$ are all pairwise distinct. The last result is known as Campbell's theorem [22].

Setting $f_n(x_{1:n}) = \prod_{i=1}^n \mathbb{1}_{B_i}(x_i)$ in Eqs (3), (4), yields

$$\mu_\Phi^{(n)}(B_1\times\cdots\times B_n) = \mathbb{E}\left[\sum_{x_1,\ldots,x_n\in\Phi} \mathbb{1}_{B_1}(x_1)\ldots\mathbb{1}_{B_n}(x_n)\right], \quad (5)$$

$$\nu_\Phi^{(n)}(B_1\times\cdots\times B_n) = \mathbb{E}\left[\sum_{x_1,\ldots,x_n\in\Phi}^{\neq} \mathbb{1}_{B_1}(x_1)\ldots\mathbb{1}_{B_n}(x_n)\right], \quad (6)$$

for any regions $B_i \in \mathcal{B}(\mathcal{X})$, $1 \le i \le n$.[4] Eqs (5) and (6) provide some insight on the moment measures. The scalar $\mu_\Phi^{(n)}(B_1\times\cdots\times B_n)$ estimates the joint localisation of sequence points within the regions B_i, while $\nu_\Phi^{(n)}(B_1\times\cdots\times B_n)$ further imposes the sequence points to be pairwise distinct.

Note that the first-order moment measure $\mu_\Phi^{(1)}$ coincides with the first-order factorial moment measure $\nu_\Phi^{(1)}$; it is known as the *intensity measure* of the point process and simply denoted by μ_Φ. Its associated density, also denoted by μ_Φ, is called the *intensity* of the point process Φ, more usually called *Probability Hypothesis Density* in the context of RFSs [2]. In this paper we shall also exploit the second-order moment measures; similarly to real-valued random variables we can define the *covariance*, *variance*, and *correlation* of a point process Φ as [22], [23]

$$\mathrm{cov}_\Phi(B, B') := \mu_\Phi^{(2)}(B \times B') - \mu_\Phi(B)\mu_\Phi(B'), \quad (7)$$

$$\mathrm{var}_\Phi(B) := \mu_\Phi^{(2)}(B \times B) - [\mu_\Phi(B)]^2, \quad (8)$$

$$\mathrm{corr}_\Phi(B, B') := \frac{\mathrm{cov}_\Phi(B, B')}{\sqrt{\mathrm{var}_\Phi(B)}\sqrt{\mathrm{var}_\Phi(B')}}, \quad (9)$$

for any regions $B, B' \in \mathcal{B}(\mathcal{X})$. The scalar $\mu_\Phi(B)$ yields the expected (or *mean*) number of objects within B, while the scalar $\mathrm{var}_\Phi(B)$ quantifies the spread of the estimated number of objects within B around its mean value $\mu_\Phi(B)$ [24]. Finally, the scalar $\mathrm{corr}_\Phi(B, B')$ quantifies the correlation between the estimated number of targets within B and B'; it will be exploited in this paper to assess the so-called "spooky effect" of multi-object filters, coined in [9] for the CPHD filter.

Note that in the general case the variance var_Φ is a non-additive function, and does not admit a density. Note also that the second-order moment measure can be decomposed into the sum

$$\mu_\Phi^{(2)}(B \times B') = \mu_\Phi(B \cap B') + \nu_\Phi^{(2)}(B \times B'), \quad (10)$$

for any regions $B, B' \in \mathcal{B}(\mathcal{X})$.

---

[4]The notation $\mathbb{1}_B$ denotes the indicator function, i.e., $\mathbb{1}_B(x) = 1$ if $x \in B$, and zero otherwise.

D. Point processes and functionals

Similar to the Fourier transform for signals or the probability generating function for discrete real-valued random variables, convenient tools exist to handle operations on point processes. The *Laplace functional* $\mathcal{L}_\Phi$ and the *Probability Generating Functional (PGFL)* $\mathcal{G}_\Phi$ of a point process Φ are defined by

$$\mathcal{L}_\Phi(f) := \sum_{n\ge0} \int \exp\left(-\sum_{i=1}^n f(x_i)\right) P_\Phi^{(n)}(\mathrm{d}x_{1:n}), \quad (11)$$

$$\mathcal{G}_\Phi(h) := \sum_{n\ge0} \int \left[\prod_{i=1}^n h(x_i)\right] P_\Phi^{(n)}(\mathrm{d}x_{1:n}), \quad (12)$$

respectively for two test functions $f : \mathcal{X} \to \mathbb{R}^+$ and $h : \mathcal{X} \to [0, 1]$. Note that from (11) and (12) it holds that

$$\mathcal{G}_\Phi(h) = \mathcal{L}_\Phi(-\ln h). \quad (13)$$

Depending on the nature of the point process Φ, the expression of the functionals may reduce to simpler expressions that do not involve infinite sums (see examples in Sec. III).

E. Point processes and differentiation

In this paper we shall exploit the *chain differential* [20], a convenient operator that allows for the evaluation of both the statistical moments of a point process Φ and their corresponding densities through the differentiation of its Laplace functional $\mathcal{L}_\Phi$ or its PGFL $\mathcal{G}_\Phi$ [25]–[27].

Given a functional G and two functions $h, \eta : \mathcal{X} \to \mathbb{R}^+$, the (chain) differential of G w.r.t. h in the direction of η is defined as [20]

$$\delta G(h; \eta) := \lim_{n\to\infty} \frac{G(h + \varepsilon_n\eta_n) - G(h)}{\varepsilon_n}, \quad (14)$$

when the limit exists and is identical for any sequence of real numbers $(\varepsilon_n)_{n\in\mathbb{N}}$ converging to 0 and any sequence of functions $(\eta_n : \mathcal{X} \to \mathbb{R}^+)_{n\in\mathbb{N}}$ converging pointwise to η.

The statistical quantities described in Sec. II-A and Sec. II-C can then be extracted through the following differentiations:

$$P_\Phi^{(n)}(B_1\times\cdots\times B_n) = \frac{1}{n!}\delta^n\mathcal{G}_\Phi(h; \mathbb{1}_{B_1}, \ldots, \mathbb{1}_{B_n})|_{h=0}, \quad (15)$$

$$\mu_\Phi^{(n)}(B_1\times\cdots\times B_n) = (-1)^n\delta^n\mathcal{L}_\Phi(f; \mathbb{1}_{B_1}, \ldots, \mathbb{1}_{B_n})|_{f=0}, \quad (16)$$

$$\nu_\Phi^{(n)}(B_1\times\cdots\times B_n) = \delta^n\mathcal{G}_\Phi(h; \mathbb{1}_{B_1}, \ldots, \mathbb{1}_{B_n})|_{h=1}, \quad (17)$$

for any regions $B_i \in \mathcal{B}(\mathcal{X})$, $1 \le i \le n$ [22]. The chain differential has convenient properties and leads to rules similar to the classical derivative: namely, a *product rule* [20]

$$\delta(F \cdot G)(h; \eta) = \delta F(h; \eta)G(h) + F(h)\delta G(h; \eta), \quad (18)$$

and a *chain rule* [20]

$$\delta(F \circ G)(h; \eta) = \delta F(G(h); \delta G(h; \eta)). \quad (19)$$

They can be generalised to the *n-fold product rule* [27]

$$\delta^n(F \cdot G)(h; \eta_1, \ldots, \eta_n)$$
$$= \sum_{\omega\subseteq\{1,\ldots,n\}} \delta^{|\omega|}F\Big(h; (\eta_i)_{i\in\omega}\Big)\delta^{|\omega^c|}G\Big(h; (\eta_j)_{j\in\omega^c}\Big), \quad (20)$$

where $\omega^{\mathrm{c}} = \{1, \ldots, n\} \setminus \omega$ is the complement of ω, and the *n-fold chain rule* or *Faà di Bruno's formula for chain differentials* [26], [27]

$$\delta^n(F \circ G)(h; \eta_1, \ldots, \eta_n)$$
$$= \sum_{\pi \in \Pi_n} \delta^{|\pi|} F\left(G(h); \left(\delta^{|\omega|}G(h; (\eta_i)_{i \in \omega})\right)_{\omega \in \pi}\right), \quad (21)$$

where Π_n is the set of partitions of the index set $\{1, \ldots, n\}$. The equivalent of the *n*-fold product rule (20) in the FISST framework is called the generalised product rule for set derivatives [1, p. 389]. Faà di Bruno's formula (21) has recently been applied for spatial cluster modelling [28], Volterra series [29], multi-target spawning [30], and for negative binomial clutter modelling [18].

When the chain differential (14) is linear and continuous w.r.t. its argument, it is also called the *chain derivative* operator. For the rest of the paper, chain differentials will always assumed to be chain derivatives and called as such. Also, when a functional G is defined as an integral with respect to a measure μ on $\mathcal{X}$ which is absolutely continuous with respect to the reference measure λ, the term $\delta G(f, \delta_x)$ will be understood as the Radon-Nikodým derivative of the measure $\mu' : B \mapsto \delta G(f, \mathbb{1}_B)$ evaluated at point x, i.e. $\delta G(f, \delta_x) := \dfrac{\mathrm{d}\mu'}{\mathrm{d}\lambda}(x)$, for any appropriate function f on $\mathcal{X}$ and any point $x \in \mathcal{X}$. In the context of this paper, this property holds for the PGFL $\mathcal{G}_\Phi$ of any point process Φ since its probability distribution P_Φ admits a density w.r.t. the reference measure λ. In particular,

$$p_\Phi^{(n)}(x_1, \ldots, x_n) = \frac{1}{n!} \delta^n \mathcal{G}_\Phi(h; \delta_{x_1}, \ldots, \delta_{x_n})|_{h=0}, \quad (22)$$

for any points $x_i \in \mathcal{X}$, $1 \le i \le n$. This result is similar to the extraction rule (52) in [2], allowing for the evaluation of the multitarget density of a RFS in the *set* $\{x_1, \ldots, x_n\}$.

III. FOUR RELEVANT EXAMPLES OF POINT PROCESSES

This section presents three well-established point processes in the context of multi-object estimation, namely, the independent and identically distributed (i.i.d.), Bernoulli, and Poisson point processes. It then introduces the Panjer point process and its fundamental properties.

A. i.i.d. cluster process

An i.i.d. cluster process with cardinality distribution ρ on $\mathbb{N}$ and spatial distribution s on $\mathcal{X}$ describes a population whose size is described by ρ, and whose objects' states are i.i.d. according to s. Its PGFL is given by

$$\mathcal{G}_{\text{i.i.d.}}(h) = \sum_{n \ge 0} \rho(n) \left[\int h(x)s(\mathrm{d}x)\right]^n. \quad (23)$$

In the construction of the CPHD filter, the predicted target process $\Phi_{k|k-1}$ is assumed i.i.d. cluster [6].

B. Bernoulli process

A Bernoulli point process with parameter $0 \le p \le 1$ and spatial distribution s is an i.i.d. cluster process with spatial distribution s, whose size is 1 with probability p and 0 with probability $q = (1 - p)$. Its PGFL is given by

$$\mathcal{G}_{\text{Bernoulli}}(h) = q + p \int h(x)s(\mathrm{d}x). \quad (24)$$

In the context of target tracking, Bernoulli processes are commonly used to describe binary events such as the detection or survival of individual targets.

C. Poisson process

A Poisson process with parameter λ and spatial distribution s is an i.i.d. cluster process with spatial distribution s, whose size is Poisson distributed with rate λ. Its PGFL is given by

$$\mathcal{G}_{\text{Poisson}}(h) = \exp\left(\int [h(x) - 1]\mu(\mathrm{d}x)\right), \quad (25)$$

where the intensity measure μ of the process is such that $\mu(\mathrm{d}x) = \lambda s(\mathrm{d}x)$. Due to its simple form and its prevalence in many natural phenomena, the Poisson point process is a common and well-studied modelling choice. It can be shown that the intensity (5) and the variance (8) of a Poisson process are equal when evaluated in any region $B \in \mathcal{B}(\mathcal{X})$, i.e., $\mu_\Phi(B) = \text{var}_\Phi(B)$. In other words, the random variable describing the number of objects within B has equal mean and variance. This property holds in particular for $B = \mathcal{X}$. In the construction of the PHD filter, the predicted target process $\Phi_{k|k-1}$ is assumed Poisson [2].

D. Panjer process

A Panjer point process with parameters α and β and spatial distribution s is an i.i.d. cluster process with spatial distribution s, whose size is Panjer distributed with parameters α and β [17], i.e., whose cardinality distribution is given by

$$\rho(n) := \binom{-\alpha}{n}\left(1 + \frac{1}{\beta}\right)^{-\alpha}\left(\frac{-1}{\beta + 1}\right)^n, \quad (26)$$

for any $n \in \mathbb{N}$, where either $\alpha, \beta \in \mathbb{R}_{>0}$ or $\alpha \in \mathbb{Z}_{<0}$ and $\beta \in \mathbb{R}_{<0}$.[5] The particular nature of the Panjer process is determined by the values α and β:

- For finite and positive α and β, (26) describes a negative binomial distribution.
- For finite and negative α and β we obtain a binomial distribution.[6]
- The limit case $\alpha, \beta \to \infty$ with constant ratio $\lambda := \frac{\alpha}{\beta}$ yields a Poisson process with parameter λ [18], [31].

The PGFL of a negative binomial process is given in [32], and it can be extended to the Panjer point process as follows:

---

[5]Note that negative, non-integer values of α yield complex values, and are thus discarded.

[6]In [17], the binomial and negative binomial distributions are given in different forms but are equivalent to (26).

Proposition III.1. *The PGFL of a Panjer process with parameters α, β is given by*

$$\mathcal{G}_{\text{Panjer}}(h) = \left(1 + \frac{1}{\beta}\int [1 - h(x)]s(\mathrm{d}x)\right)^{-\alpha}. \qquad (27)$$

The proof is given in appendix. The parameters of a Panjer point process are linked to the first- and second-order moment of its cardinality distribution as follows:

Proposition III.2. *The parameters α_Φ, β_Φ of a Panjer process Φ are such that*

$$\alpha_\Phi = \frac{\mu_\Phi(\mathcal{X})^2}{\text{var}_\Phi(\mathcal{X}) - \mu_\Phi(\mathcal{X})}, \qquad (28)$$

$$\beta_\Phi = \frac{\mu_\Phi(\mathcal{X})}{\text{var}_\Phi(\mathcal{X}) - \mu_\Phi(\mathcal{X})}. \qquad (29)$$

The proof is given in appendix. It can be seen from Eqs. (28), (29) that binomial and negative binomial point processes have a size with *larger* and *smaller* variance than mean, respectively. In particular, a negative binomial point process can model a population whose size is highly uncertain, such as the clutter process in the PHD filter with negative binomial clutter [18].

IV. THE SO-PHD FILTER WITH VARIANCE IN TARGET NUMBER

The intensity measure of the target process μ_Φ (or its density) plays an important role in the construction of multi-object filters; it is propagated by both the PHD [2] and CPHD filters [6], and the quantity $\mu_\Phi(\mathcal{X})$ yields the expected number of targets in the scene. The CPHD also propagates the cardinality distribution of the target process ρ_Φ, describing the number of targets in the scene. As an intermediate solution, the SO-PHD filter propagates the variance in the number of targets $\text{var}_\Phi(\mathcal{X})$. In order to do so, the Poisson or i.i.d. cluster assumption on the predicted target process and clutter process is replaced by a Panjer assumption. The data flow of the SO-PHD filter is depicted in Fig. 1.

A. Time prediction step (time k)

In the time prediction step, the posterior target process Φ_{k-1} is predicted to $\Phi_{k|k-1}$ based on prior knowledge on the dynamical behaviour of the targets. The assumptions of the time prediction step can be stated as follows:

Assumptions IV.1.
- (a) *The targets evolve independently from each other;*
- (b) *A target with state $x \in \mathcal{X}$ at time $k-1$ survived to the current time k with probability $p_{\text{s},k}(x)$; if it did so, its state evolved according to a Markov transition kernel $t_{k|k-1}(\cdot|x)$;*
- (c) *New targets entered the scene between time $k-1$ and k, independently of the existing targets and described by a newborn point process $\Phi_{\text{b},k}$ with PGFL $\mathcal{G}_{\text{b},k}$.*

Assumptions IV.2.
- (a) *The probability of survival is uniform over the state space, i.e., $p_{\text{s},k}(x) := p_{\text{s},k}$ for any $x \in \mathcal{X}$.*

$$\cdots$$
$$\downarrow \qquad\qquad \downarrow$$
$$\Phi_{k-1} \longrightarrow \left(\mu_{k-1}(\cdot), \text{var}_{k-1}(\mathcal{X})\right)$$

prediction $\downarrow \qquad\qquad\qquad \downarrow (30,32)$

$$\Phi_{k|k-1} \longrightarrow \left(\mu_{k|k-1}(\cdot), \text{var}_{k|k-1}(\mathcal{X})\right)$$

$$\downarrow (28,29)$$

$$\left(\mu_{k|k-1}(\cdot), \alpha_{k|k-1}, \beta_{k|k-1}\right)$$

update $\downarrow \qquad\qquad\qquad \downarrow (41,42)$

$$\Phi_k \longrightarrow \left(\mu_k(\cdot), \text{var}_k(\mathcal{X})\right)$$

$$\downarrow \qquad\qquad\qquad \downarrow$$
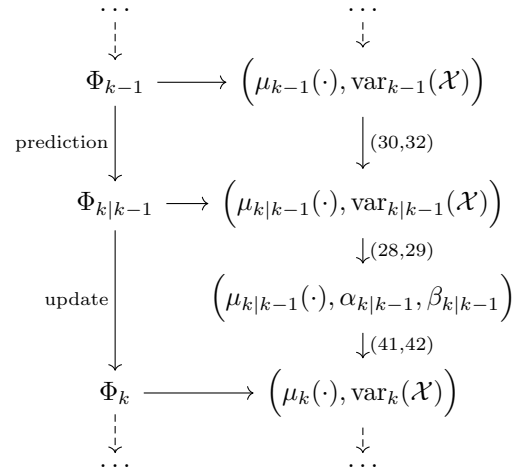$$\cdots \qquad\qquad \cdots$$

Fig. 1: Data flow of the SO-PHD filter at time k. In addition to the intensity function μ it propagates the scalar $\text{var}(\mathcal{X})$, describing the variance in the estimated number of targets in the whole state space.

Note that Assumptions IV.1 are those of the original PHD filter; in particular, the SO-PHD filter does not require a specific form for the posterior process Φ_{k-1} or the birth process $\Phi_{\text{b},k}$.

Theorem IV.3 (Intensity prediction [2]). *Under Assumptions IV.1, the intensity measure $\mu_{k|k-1}$ of the predicted target process $\Phi_{k|k-1}$ is given by*

$$\mu_{k|k-1}(B) = \mu_{\text{b},k}(B) + \mu_{\text{s},k}(B), \qquad (30)$$

in any $B \in \mathcal{B}(\mathcal{X})$, where $\mu_{\text{s},k}$ is the intensity measure of the process describing the surviving targets

$$\mu_{\text{s},k}(B) := \int p_{\text{s},k}(x) t_{k|k-1}(B|x) \mu_{k-1}(\mathrm{d}x), \qquad (31)$$

and $\mu_{\text{b},k}$ is the intensity measure of the newborn process $\Phi_{\text{b},k}$.

Theorem IV.4 (Variance prediction). *Under Assumptions IV.1, the variance $\text{var}_{k|k-1}$ of the predicted target process $\Phi_{k|k-1}$ is given by*

$$\text{var}_{k|k-1}(B) = \text{var}_{\text{b},k}(B) + \text{var}_{\text{s},k}(B), \qquad (32)$$

in any $B \in \mathcal{B}(\mathcal{X})$, where $\text{var}_{\text{s},k}$ is the variance of the process describing the surviving targets

$$\text{var}_{\text{s},k}(B) := \mu_{\text{s},k}(B)\left[1 - \mu_{\text{s},k}(B)\right]$$
$$+ \int p_{\text{s},k}(x)p_{\text{s},k}(x')t_{k|k-1}(B|x)t_{k|k-1}(B|x')\nu_{k-1}^{(2)}(\mathrm{d}(x,x')), \qquad (33)$$

and $\text{var}_{\text{b},k}$ is the variance of the newborn process $\Phi_{\text{b},k}$.

The proof is given in appendix. Note that the propagation of the regional variance (32) – i.e., the variance $\text{var}_{k|k-1}(B)$ in any $B \in \mathcal{B}(\mathcal{X})$ – requires the posterior second-order factorial moment $\nu_{k-1}^{(2)}$, which is not available from the posterior information $\mu_{k-1}(\cdot)$, $\text{var}_{k-1}(\mathcal{X})$ (see data flow in Fig. 1). However, considering the additional Assumption IV.2, the variance of the

predicted target process $\Phi_{k|k-1}$ evaluated in the whole state space becomes as follows.

Corollary IV.5 (Variance prediction, uniform $p_{s,k}$)**.** *Under Assumptions IV.1 and IV.2, the variance $\mathrm{var}_{k|k-1}$ of the predicted target process $\Phi_{k|k-1}$ evaluated in the whole state space $\mathcal{X}$ is given by*

$$\mathrm{var}_{k|k-1}(\mathcal{X}) = \mathrm{var}_{b,k}(\mathcal{X}) + \mathrm{var}_{s,k}(\mathcal{X}), \qquad (34)$$

where $\mathrm{var}_{s,k}$ is the variance of the process describing the surviving targets

$$\mathrm{var}_{s,k}(\mathcal{X}) = p_{s,k}^2 \mathrm{var}_{k-1}(\mathcal{X}) + p_{s,k}[1-p_{s,k}]\mu_{k-1}(\mathcal{X}), \quad (35)$$

and $\mathrm{var}_{b,k}$ is the variance of the newborn process $\Phi_{b,k}$.

The proof is given in appendix. The results in Thm. IV.3 and Cor. IV.5 produce the predicted quantities $\mu_{k|k-1}$, $\mathrm{var}_{k|k-1}(\mathcal{X})$ from their posterior values μ_{k-1}, $\mathrm{var}_{k-1}(\mathcal{X})$.

B. Data update step (time k)

In the data update step, the predicted process $\Phi_{k|k-1}$ is updated to Φ_k given the current measurement set Z_k, collected from the sensor. The date update step relies on the following assumptions:

Assumptions IV.6.

(a) *The predicted target process $\Phi_{k|k-1}$ is Panjer, with parameters $\alpha_{k|k-1}$, $\beta_{k|k-1}$ and spatial distribution $s_{k|k-1}$.*

(b) *The measurements originating from target detections are generated independently from each other.*

(c) *A target with state $x \in \mathcal{X}$ is detected with probability $p_{d,k}(x)$; if so, it produces a measurement whose state is distributed according to a likelihood $l_k(\cdot|x)$.*

(d) *The clutter process, describing the false alarms produced by the sensor, is Panjer with parameters $\alpha_{c,k}$, $\beta_{c,k}$ and spatial distribution $s_{c,k}$.*

Before stating the data update equations for the SO-PHD filter, recall the *Pochhammer symbol* or *rising factorial* $(\zeta)_n$ for any $\zeta \in \mathbb{R}$ and $n \in \mathbb{N}$:

$$(\zeta)_n := \zeta(\zeta+1)\cdots(\zeta+n-1), \quad (\zeta)_0 := 1. \qquad (36)$$

Following the notations used in [24] and introduced in [8], we define the corrective terms

$$\ell_u(z) := \frac{\Upsilon_u(Z_k\setminus\{z\})}{\Upsilon_0(Z_k)}, \quad \ell_u(\phi) := \frac{\Upsilon_u(Z_k)}{\Upsilon_0(Z_k)}, \qquad (37)$$

for any $u \in \mathbb{N}$ and any $z \in Z_k$, where

$$\Upsilon_u(Z) := \sum_{j=0}^{|Z|} \frac{(\alpha_{k|k-1})_{j+u}}{(\beta_{k|k-1})^{j+u}} \frac{(\alpha_{c,k})_{|Z|-j}}{(\beta_{c,k}+1)^{|Z|-j}} F_d^{-j-u} e_j(Z), \qquad (38)$$

for any $Z \subseteq Z_k$, where F_d is the scalar given by

$$F_d := \int \left[1 + \frac{p_{d,k}(x)}{\beta_{k|k-1}}\right]\mu_{k|k-1}(\mathrm{d}x), \qquad (39)$$

and e_j is the j-th elementary symmetric function

$$e_j(Z) := \sum_{\substack{Z'\subseteq Z \\ |Z'|=j}} \prod_{z\in Z'} \frac{\mu_k^z(\mathcal{X})}{s_{c,k}(z)}, \qquad (40)$$

with $\mu_k^z(B) := \int_B p_{d,k}(x)l_k(z|x)\mu_{k|k-1}(\mathrm{d}x)$, for any $B \in \mathcal{B}(\mathcal{X})$.[7]

Theorem IV.7 (Intensity update)**.** *Under Assumptions IV.6, the intensity measure μ_k of the updated target process Φ_k is given by*

$$\mu_k(B) = \mu_k^\phi(B)\ell_1(\phi) + \sum_{z\in Z_k} \frac{\mu_k^z(B)}{s_{c,k}(z)}\ell_1(z), \qquad (41)$$

in any $B \in \mathcal{B}(\mathcal{X})$, where the missed detection term μ_k^ϕ is given by $\mu_k^\phi(B) := \int_B (1-p_{d,k}(x))\mu_{k|k-1}(\mathrm{d}x)$.

Theorem IV.8 (Variance update)**.** *Under Assumptions IV.6, the variance var_k of the updated target process Φ_k is given by*

$$\begin{aligned}
\mathrm{var}_k(B) = {}& \mu_k(B) + \mu_k^\phi(B)^2 \left[\ell_2(\phi) - \ell_1(\phi)^2\right] \\
& + 2\mu_k^\phi(B) \sum_{z\in Z_k} \frac{\mu_k^z(B)}{s_{c,k}(z)} \left[\ell_2(z) - \ell_1(\phi)\ell_1(z)\right] \\
& + \sum_{z,z'\in Z_k} \frac{\mu_k^z(B)}{s_{c,k}(z)} \frac{\mu_k^{z'}(B)}{s_{c,k}(z')} \left[\ell_2^{\neq}(z,z') - \ell_1(z)\ell_1(z')\right],
\end{aligned} \qquad (42)$$

in any $B \in \mathcal{B}(\mathcal{X})$, with

$$\ell_2^{\neq}(z,z') := \begin{cases} \dfrac{\Upsilon_2(Z_k\setminus\{z,z'\})}{\Upsilon_0(Z_k)}, & z \neq z', \\ 0, & \text{otherwise.} \end{cases} \qquad (43)$$

The proofs of Thms IV.7 and IV.8 are given in appendix. Together with Eqs (28), (29), the results in Thms IV.7, IV.8 produce the updated quantities μ_k, $\mathrm{var}_k(\mathcal{X})$ from their predicted values $\mu_{k|k-1}$, $\mathrm{var}_{k|k-1}(\mathcal{X})$.

As mentioned earlier in Sec. III-D, a Panjer distribution converges to a Poisson distribution for suitable parameters α, β. An interesting consequence for the intensity update of the SO-PHD filter (41) follows.

Corollary IV.9 (Intensity update: limit cases)**.** *If, in addition to Assumptions IV.6, the predicted point process $\Phi_{k|k-1}$ is assumed Poisson, i.e., $\alpha_{k|k-1}, \beta_{k|k-1} \to \infty$ with constant ratio $\lambda_{k|k-1} := \frac{\alpha_{k|k-1}}{\beta_{k|k-1}}$, then the intensity update (41) converges to the intensity update of the PHD filter with Panjer clutter given in [18].*

Furthermore, if the clutter process is assumed Poisson as well, i.e., $\alpha_{c,k}, \beta_{c,k} \to \infty$ with constant ratio $\lambda_{c,k} := \frac{\alpha_{c,k}}{\beta_{c,k}}$, then the intensity update given in [18] converges to the intensity update of the original PHD filter [2].

With Cor. IV.9, the SO-PHD filter presented in this paper can be seen as a generalisation of the original PHD filter[8]. Note that the expression of the intensity (41) and variance (42) of the updated target process are remarkably similar to

---

[7] In these definitions, the time subscripts on the ℓ_u, Υ_u, F_d, and e_j terms are omitted for the sake of simplicity.

[8] Under the proviso that the additional assumption IV.2 is met, i.e., the probability of survival $p_{s,k}$ is uniform over the state space.

their counterpart in the CPHD filter, and only differ on the expressions of the corrective terms ℓ_u [24]. Both filters involve the computation of elementary symmetric functions $e_j(Z)$ on subsets Z of the measurement set Z_k. Each function has a computational cost of $\mathcal{O}(|Z|\log^2|Z|)$ [8]; the CPHD requires the computation for sets of the form Z_k, and $Z_k \setminus \{z\}$, for a total cost of $\mathcal{O}(|Z_k|^2\log^2|Z_k|)$, while the proposed solution requires the computation for sets of the form Z_k, $Z_k \setminus \{z\}$, $Z_k \setminus \{z,z'\}$, for a total cost of $\mathcal{O}(|Z_k|^3\log^2|Z_k|)$. However, while the CPHD filter requires the computation of the $\Upsilon^u(n)$ terms [8, Eq. (14)] for each possible target number n (to a maximum number $N_{\max}$ set as a parameter), the proposed filter requires the computation of the Υ_u terms (38) *only once*. The complexity of the proposed filter is thus significantly lower than for the CPHD filter, as it will be illustrated in the simulation results in Sec. VI, and the difference in complexity increases with the value $N_{\max}$.

V. REGIONAL CORRELATIONS FOR PHD FILTERS

In order to assess the mutual influence of the estimated number of targets in two regions $B, B' \in \mathcal{B}(\mathcal{X})$, we compute in this section the statistical correlation (9) of the updated target process Φ_k for the PHD, SO-PHD and CPHD filters.

Proposition V.1 (Covariance of the PHD filters)**.**
Let $B, B' \in \mathcal{B}(\mathcal{X})$ be two arbitrary regions in the state space.
(a) PHD filter:
Let $\lambda_{c,k}$ be the Poisson clutter rate at time k. The covariance of the updated target process Φ_k in B, B' is

$$\mathrm{cov}_k(B \times B') = \mu_k(B \cap B')$$
$$- \sum_{z \in Z_k} \frac{\mu_k^z(B)\mu_k^z(B')}{\left[\mu_k^z(\mathcal{X}) + \lambda_{c,k}s_{c,k}(z)\right]^2}. \quad (44)$$

(b) Second-order PHD filter:
The covariance of the updated target process Φ_k in B, B' is

$$\mathrm{cov}_k(B \times B')$$
$$= \mu_k(B \cap B') + \mu_k^\phi(B)\mu_k^\phi(B')[\ell_2(\phi) - \ell_1(\phi)^2]$$
$$+ \sum_{z \in Z_k} \left[\mu_k^\phi(B)\frac{\mu_k^z(B')}{s_{c,k}(z)} + \mu_k^\phi(B')\frac{\mu_k^z(B)}{s_{c,k}(z)} \right] [\ell_2(z) - \ell_1(z)\ell_1(\phi)]$$
$$+ \sum_{z,z' \in Z_k} \left[\frac{\mu_k^z(B)}{s_{c,k}(z)} \frac{\mu_k^{z'}(B')}{s_{c,k}(z')} \right] \left[\ell_2^{\neq}(z,z') - \ell_1(z)\ell_1(z') \right]. \quad (45)$$

(c) CPHD filter:
The covariance of the updated target process Φ_k in B, B' is given by (45)*, where the corrective terms ℓ_1, ℓ_2 and $\ell_2^{\neq}$ are replaced by the values in Eqns (20), (30) of [24].*

The proof is given in appendix. The correlations $\mathrm{corr}_\Phi(B, B')$ are a direct consequence of Eq. (9), using the regional variance stated in Eqns (35), (33) [24] for the PHD and CPHD filters and the regional variance (42) for the SO-PHD filter.

VI. EXPERIMENTS

A GM implementation [3], [8] was used for all algorithms to make them comparable. For the CPHD filter, the maximum number of targets $N_{\max}$ is set to 150 for all experiments. The Optimal Sub-Pattern Assignment (OSPA) metric per time step [33] is used with the Euclidean distance (i.e. $p = 2$) and the cutoff $c = 100$.

A. Scenario 1

This scenario examines the robustness of the PHD, CPHD, and SO-PHD filters to large variations in the number of targets and focuses on a single time step when the change in target number occurs.

The size of the surveillance scene is $50\,\mathrm{m} \times 50\,\mathrm{m}$. The generation of new objects is restricted to the centre of the image to prevent the objects from leaving the scene before the last time step. Their movement is generated using a nearly constant velocity model where the standard deviation of the acceleration noise is $0.3\,\mathrm{m\,s^{-2}}$ and the initial velocity is Gaussian normal distributed with mean 0 and standard deviation $0.5\,\mathrm{m\,s^{-1}}$ along each dimension of the state space. False alarms are generated according to a Poisson point process with uniform spatial distribution and clutter rate $\mu_c = 5$ for experiments 1.1, 1.2 and $\mu_c = 20$ for experiment 1.3. The probabilities of detection and survival are constant and set to 0.9 and 0.99, respectively.

1.1 50 targets are created in the first time step and propagated until time step 15 to give the algorithms time to settle. At time 15, the number of targets suddenly changes, either by removing some or all of the current targets without creating new objects or by creating up to 50 births while maintaining the old targets. The birth model is Poisson with uniform spatial distribution and birth rate $\mu_b = 25$, for the three filters.

1.2 The parameters are identical to experiment 1.1, except that the birth model is negative binomial with $\mu_b = 25$ and $\mathrm{var}_b = 100$ for the SO-PHD and CPHD filter.

1.3 Here, only one target is created in the beginning and maintained up to time 15. At this time, from 0 to 100 targets are spontaneously created in the scene. The birth model is a negative binomial point process with uniform spatial distribution, mean $\mu_b = 1$ and $\mathrm{var}_b = 100$ for the three filters, though the PHD filter cannot exploit the information on the variance.

Fig. 2 depicts the results of this scenario. In experiment 1.1 and 1.2, the three filters estimate target birth more accurately than target death since the high survival probability, together with a high birth rate, does not account for severe drops in the number of targets. In particular, the CPHD filter lacks flexibility and fails at recognising unexpected drops in the number of targets. Choosing negative binomial birth model allows for larger uncertainty in the number of targets and improves the quality of the estimate for the CPHD and SO-PHD filters. Furthermore, the variance of the SO-PHD filter is lower than that of the PHD filter. Experiment 1.3 highlights a limitation of the PHD filter, which reduces the prior information on the number of newborn targets to its mean value. The CPHD and SO-PHD filters, on the other hand, can exploit a birth process with high variability in target number – i.e., through a negative binomial process with large variance in target number – in
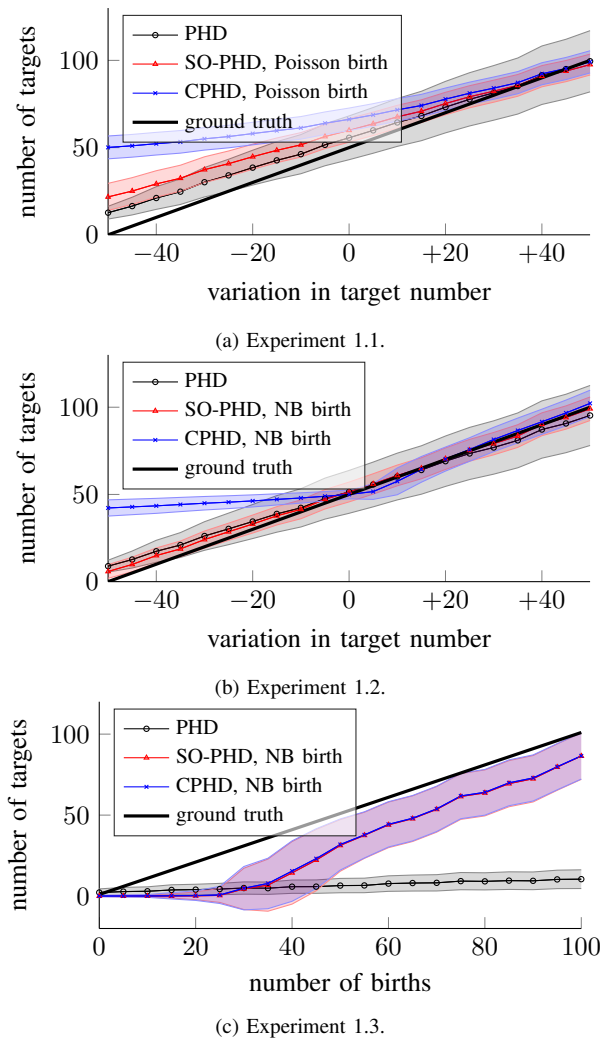
(a) Experiment 1.1.

(b) Experiment 1.2.

(c) Experiment 1.3.

Fig. 2: Results for Scenario 1, averaged over 20 MC runs. The lines depict the mean of the estimated number of targets, the coloured areas show the 2σ confidence region (estimated by the filter).

order to cope with a burst of target births. Fig. 2c suggests that the birth and false alarm processes are competing in the CPHD and SO-PHD filters when there is a significant influx in the number of newborn targets, resulting in an offset linked to the mean number of false alarms (recall that $\mu_c = 20$ in this case). The PHD filter, on the other hand, is unable to cope with an influx that is well beyond the Poisson model.

Average run times are omitted for this scenario as they change greatly with the different changes in target number and are therefore not very meaningful. The following scenarios will provide a more valuable insight in the computational performance.

B. Scenario 2

This scenario examines the behaviours of the PHD, CPHD and SO-PHD filters with increasing amounts of target birth and death.

The size of the surveillance scene is $50\,\text{m} \times 50\,\text{m}$. The number of targets is designed to follow a stair pattern starting

with 5 initial targets, and increasing the cardinality by 10, 15, 20 and 25 targets every ten time steps until time 40. From time 50 onwards up to time 90, the number of targets is decreased in reverse order, i.e. every ten time steps, the target population is reduced by 25, 20, 15, and 10 targets. The generation of new objects is restricted to the centre of the image to prevent the objects from leaving the scene before the last time step. Their movement is generated using a nearly constant velocity model where the standard deviation of the acceleration noise is $0.1\,\text{m}\,\text{s}^{-2}$ and the initial velocity is Gaussian normal distributed with mean 0 and standard deviation $0.3\,\text{m}\,\text{s}^{-1}$ along each dimension of the state space.

From the ground truth obtained as above, measurements are created with a constant probability of detection. For comparison, two different values are chosen, i.e. $p_\text{d} = 0.95$ in the first experiment and $p_\text{d} = 0.6$ in the second. Each detection is corrupted with white noise with standard deviation $0.2\,\text{m}$ in each dimension. Additionally, false alarms are generated according to a Poisson point process with uniform spatial distribution and clutter rate $\mu_\text{c} = 15$.

The three filters are parametrised with the simulation parameters above. In addition, the probability of survival is set to $p_\text{s} = 0.98$, and target birth is modelled using a negative binomial process with uniform spatial distribution, mean $\mu_\text{b}(\mathcal{X}) = 1$ and variance $\text{var}_b(\mathcal{X}) = 100$ to account for the big changes in the number of objects. Each experiment is averaged on 100 MC runs.

In Fig. 3, an example run of the first experiment is depicted. Fig. 4 shows the estimated means and variances for all filters and all experiments over time (left column), along with the mean and standard deviation of the respective OSPA distances over time (right column). The first experiment (Fig. 4a-4b) demonstrates that the three filters show a delay in the adjustment of the cardinality estimate when the population is growing, resulting in spikes of OSPA error. In general, the CPHD filter is closest to the true target number, however in case of target death, the PHD and SO-PHD filters prove to be more reactive despite setting the survival rate to 98%. In the second experiment (cf. Fig. 4c-4d), all three filters show a significant increase in the estimated variance in cardinality since target death and missed detections are hard to distinguish and therefore more missed detections lead to increased uncertainty in the number of targets. In terms of the estimated mean, on the other hand, the proposed method shows the highest reactivity to target birth and especially to target death, estimated poorly with the CPHD filter. Table I shows the averaged run time for both cases of this scenario. The prediction runs approximately 100 times slower for the CPHD than for the first- and second-order PHD filters; this is to be expected since the complexity of the former grows proportional to the range of cardinalities for which the cardinality distribution is estimated. The update performance, on the other hand, varies greatly for different probabilities of detection: if p_d is low, the weight for miss-detected objects does not plummet directly and therefore the information about dead tracks is kept and propagated for longer.

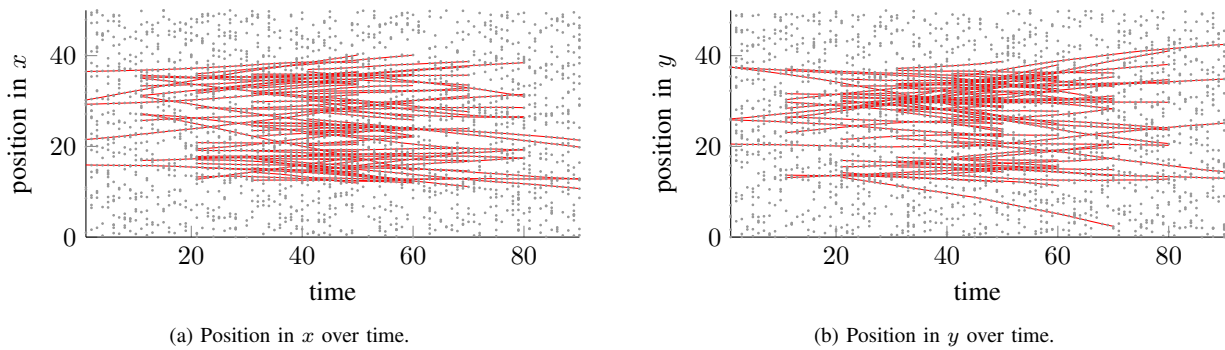(a) Position in x over time.

(b) Position in y over time.

Fig. 3: The setup of experiment 2.1, plotted separately for x and y over time (shown for one MC run). The ground truth is plotted in red, the measurements in grey.

(a) Estimated target number, experiment 2.1.

(b) OSPA results, experiment 2.1.

(c) Estimated target number, experiment 2.2.
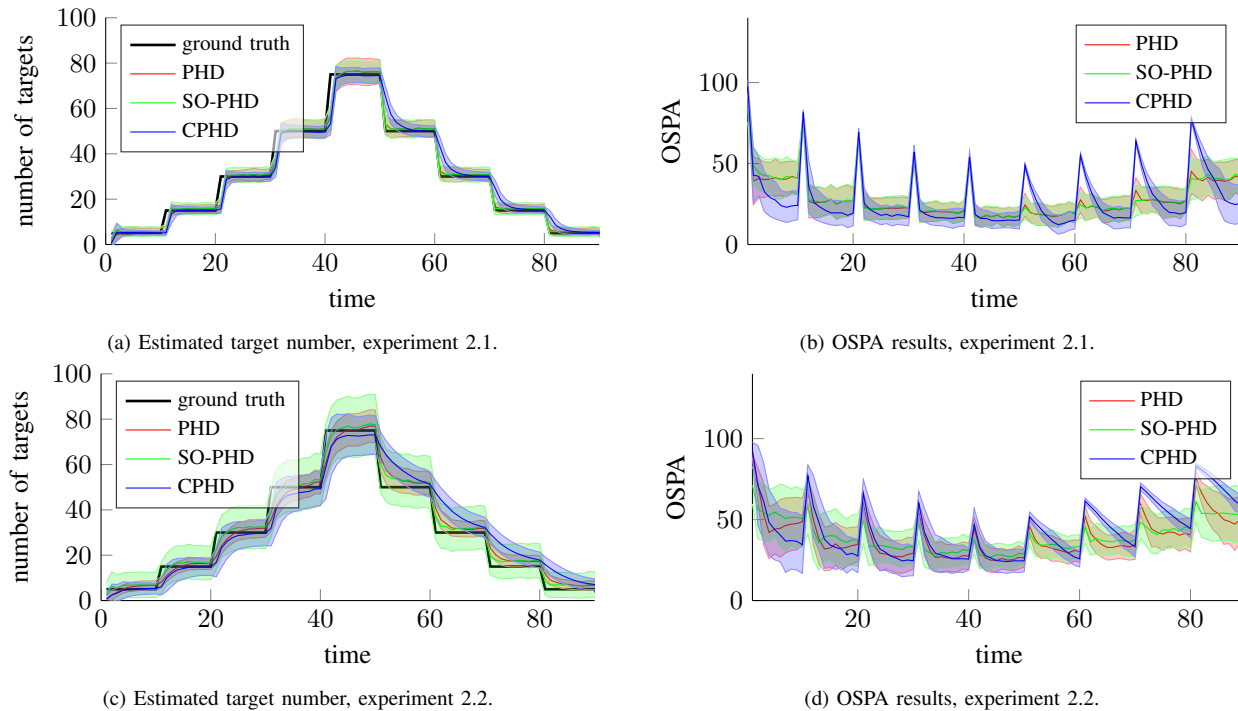
(d) OSPA results, experiment 2.2.

Fig. 4: Results for Scenario 2, averaged over 100 MC runs. Fig. 4a and 4c show the estimated means and variances of the number of targets, Fig. 4b and 4d displays the mean and standard deviation of the respective OSPA results. The rows depict the results of experiments 2.1 ($p_{\mathrm{d}} = 0.95$) and 2.2 ($p_{\mathrm{d}} = 0.6$), respectively.

	Exp.	PHD	SO-PHD	CPHD
Pred.	2.1	0.0143	0.0150 (×1.05)	0.9761 (×68.26)
	2.2	0.0266	0.0285 (×1.07)	1.0901 (×40.98)
	3	0.0121	0.0143 (×1.18)	0.6734 (×55.65)
Update	2.1	3.9233	6.2693 (×1.60)	23.0930 (×5.89)
	2.2	36.6506	40.9254 (×1.12)	46.9830 (×1.28)
	3	2.1956	2.3640 (×1.08)	10.3355 (×4.71)

TABLE I: Runtimes for experiments 2 and 3, averaged over all time steps and Monte Carlo runs, with relative time with respect to the PHD filter in brackets. The times are given in seconds.

C. Scenario 3

This scenario assesses the spooky effect of the PHD, CPHD, and SO-PHD filters through the regional covariance introduced in this paper. Two completely separate regions of interest, henceforth called A and B, are depicted in Fig. 5. Both regions are of size $50\,\mathrm{m} \times 50\,\mathrm{m}$, and they are $100\,\mathrm{m}$ apart horizontally. In each region, 10 targets are initialised in the first time step and they survive throughout 100 time steps. Again, the generation of new objects is restricted to the centre of each region to prevent the objects from leaving the scene before the last time step. Their movement is generated using a nearly constant velocity model where the standard deviation of the acceleration noise is $0.1\,\mathrm{m\,s^{-2}}$ and the initial velocity is Gaussian normal distributed with mean 0 and standard deviation $0.3\,\mathrm{m\,s^{-1}}$ along each dimension of the state space.

Measurements are created with the (constant) probability of detection $p_{\mathrm{d}} = 0.9$. Each detection is corrupted with white noise with standard deviation $0.2\,\mathrm{m}$ in each dimension. Additionally, false alarms are generated in each region according to a Poisson point process with uniform spatial distribution (in the region) and clutter rate $\mu_{\mathrm{c}}(A) = \mu_{\mathrm{c}}(B) = 20$.

The three filters are parametrised with the simulation parameters above. In addition, the probability of survival is set to $p_{\mathrm{s}} = 0.98$, and target birth is modelled using a negative binomial point process with uniform spatial distribution (in the region) with mean $\mu_{\mathrm{b}}(\mathcal{X}) = 1$ and variance $\mathrm{var}_b(\mathcal{X}) = 100$ to account for sudden changes in the number of objects.

In order to analyse the spooky effect on this scenario, *all* objects in region B are forced to be miss-detected every 10 time steps, additionally to the modelled natural missed detections in the scene. Fig. 6b-6d show the estimated regional means and regional variances for the three filters in both regions. In case of the PHD filter (cf. Fig. 6b), the intensity in region A is unaffected by the sudden drop in the intensity in region B. The proposed filter, in contrast, reacts with a slight drop in the intensity of region A when the targets in B are missed, and it compensates sightly in each subsequent time step (Fig. 6c). The biggest effect by far is noticed with the CPHD filter, as seen in Fig. 6d. Every time the objects in B stay undetected, the intensity in that region does not drop as low as for the other two filters, but the intensity in region A increases notably to approximately 12 targets.

The observed behaviour can be further illustrated by looking at the correlation of A and B under the PHD, SO-PHD and CPHD filters, exploiting the covariance of the three filters given in Sec. V. Eq. (44) shows that the covariance of the PHD filter is 0 if the two regions are disjoint and the region of origin of each measurement is unambiguous; this is clearly seen in the correlation depicted in Fig. 6a. The same figure shows a strongly negative correlation in the case of the CPHD filter, which highlights the spooky effect: the filter compensates for the lost intensity mass in region B by introducing it in region A. The SO-PHD filter shows a milder but positive correlation, as the sudden drop/increase in intensity mass in region B goes along with a smaller drop/increase in region A. These results suggest that, on these experiments, the SO-PHD filter exhibits a milder spooky effect than the CPHD filter.

Table I shows the averaged run time for this scenario, showing a coherent image with the findings above.
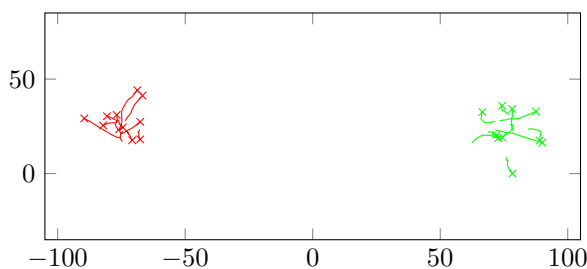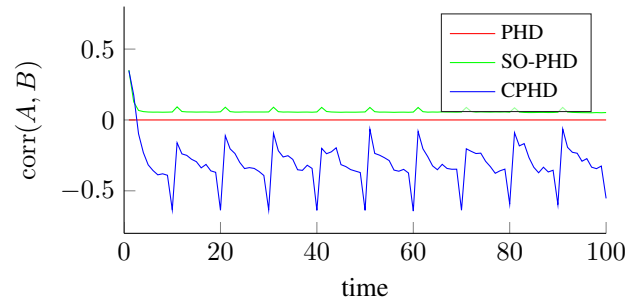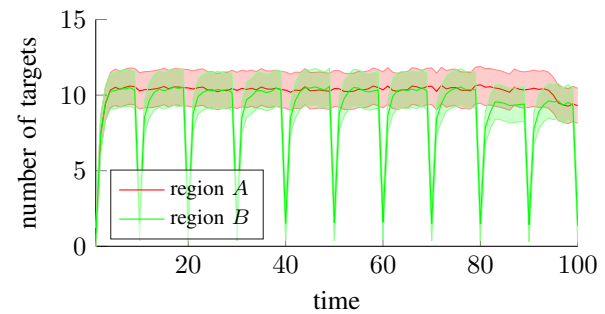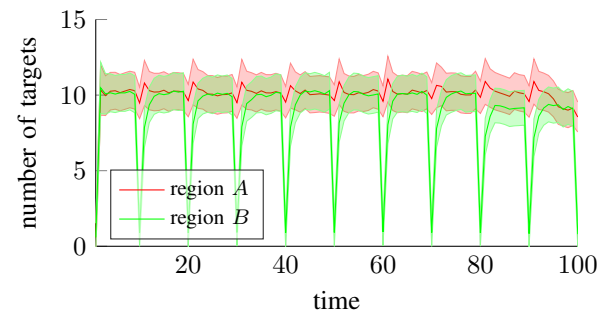

Fig. 5: Tracking scenario, with region A on the left and region B on the right.

(a) Correlation between the estimated number of targets in regions A and B.

(b) Mean and standard deviation of the estimated target number, PHD filter.

(c) Mean and standard deviation of the estimated target number, SO-PHD filter.
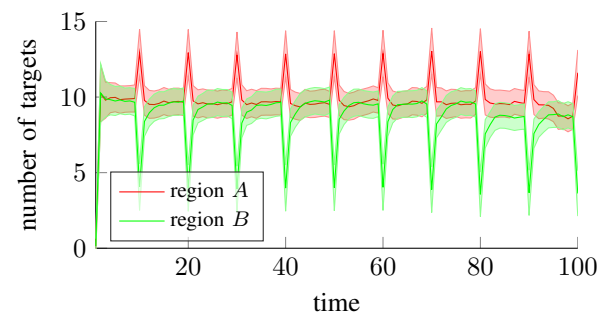
(d) Mean and standard deviation of the estimated target number, CPHD filter.

Fig. 6: Results for Scenario 3, averaged over 100 MC runs. Fig. 6a shows the correlation in A and B for all filters. Fig. 6b, 6c and 6d depict the mean and standard deviation of the estimated number of targets per region for the three filters.

VII. Conclusion

A new SO-PHD filter has been introduced, propagating the variance in the estimated number of targets alongside the first-order moment of the target process. The Panjer point process is introduced in order to approximate the multi-target predicted process and to model the false alarm process. Described with two parameters, a Panjer distribution encompasses the binomial, Poisson, and negative binomial distribution; the resulting SO-PHD filter provides more flexibility in the modelling phase than the PHD filter. The proposed filter is implemented with a Gaussian mixture algorithm, and compared to the PHD and CPHD filters on simulated data where it proved to be more robust to changes in the number of targets of unusually large extent. In a more usual scenario, the three filters showed similar performance; the proposed filter proved more reactive to the disappearance of targets than the CPHD filter, while having a significantly lower computational complexity.

The regional covariance of a point process is introduced in order to analyse the correlation between the estimated number of targets in disjoint regions of the state space, and to assess quantitatively the well-known spooky effect of the three filters on a simulated scenario. The results showed that the estimated targets in the two regions were uncorrelated with the PHD filter, strongly negatively correlated with the CPHD filter, and midlly positively correlated with the proposed SO-PHD filter.

APPENDIX A
PROOFS

The appendix provides the proofs for the results in Sec. III and IV. We first introduce the following differentiation rules, whose proofs are given in [18].

Lemma A.1 (Differentiation rules)**.** *Let G be a linear functional.*

(a) The nth order derivative of the composition $\exp(G(h))$ can be written as

$$\delta^n(\exp \circ G)(h; \eta_1, \dots, \eta_n) = \exp(G(h)) \prod_{i=1}^n \delta G(h; \eta_i). \tag{46}$$

(b) The nth order derivative of the composition $(G(h))^{-\alpha}$ is derived to be

$$\delta^n(G^{-\alpha})(h; \eta_1, \dots, \eta_n)$$
$$= (-1)^n (\alpha)_n \, G(h)^{-\alpha-n} \prod_{i=1}^n \delta G(h; \eta_i) \tag{47}$$

with $(\cdot)_n$ being the Pochhammer symbol (36).

Proof of Prop. III.1. Since a Panjer point process is an i.i.d. point process, let us start with equation (23), inserting (26) for ρ:

$\mathcal{G}_{\text{Panjer}}(h)$

$$\overset{(23)}{=} \sum_{n \geq 0} \binom{-\alpha}{n} \left(1 + \frac{1}{\beta}\right)^{-\alpha} \left(\frac{-1}{\beta+1}\right)^n \left[\int h(x) s(\mathrm{d}x)\right]^n \tag{48a}$$

$$= \left(1 + \frac{1}{\beta}\right)^{-\alpha} \sum_{n \geq 0} \binom{-\alpha}{n} \left[\frac{-1}{\beta+1} \int h(x) s(\mathrm{d}x)\right]^n \tag{48b}$$

$$= \left(1 + \frac{1}{\beta}\right)^{-\alpha} \left[1 - \frac{1}{\beta+1} \int h(x) s(\mathrm{d}x)\right]^{-\alpha} \tag{48c}$$

$$= \left[1 + \frac{1}{\beta} \int [1 - h(x)] s(\mathrm{d}x)\right]^{-\alpha} \tag{48d}$$

Equality (48c) follows from the binomial series. □

Proof of Prop. III.2. Let us derive the mean and variance of a Panjer process with parameters α, β and spatial distribution s, for arbitrary regions $B, B' \in \mathcal{B}(\mathcal{X})$:

$$\mu(B) \overset{(16)}{=} \delta \mathcal{G}_{\text{Panjer}}(h; \mathbb{1}_B) \Big|_{h=1} \tag{49a}$$

$$\overset{(27)}{=} \delta \left(\left[1 + \frac{1}{\beta} \int [1 - h(x)] s(\mathrm{d}x)\right]^{-\alpha}; \mathbb{1}_B\right) \Big|_{h=1} \tag{49b}$$

$$\overset{(47)}{=} -\alpha \left[1 + \frac{1}{\beta} \int [1 - 1] s(\mathrm{d}x)\right]^{-\alpha-1} \left[-\frac{1}{\beta} \int_B s(\mathrm{d}x)\right] \tag{49c}$$

$$= \frac{\alpha}{\beta} \int_B s(\mathrm{d}x). \tag{49d}$$

$$\mu^{(2)}(B \times B') = \delta^2 \mathcal{G}_{\text{Panjer}}(e^{-f}; \mathbb{1}_B, \mathbb{1}_{B'}) \Big|_{f=0} \tag{50a}$$

$$= \frac{(\alpha)_2}{\beta^2} \left[1 + \frac{1}{\beta} \int [1 - e^0] s(\mathrm{d}x)\right]^{-\alpha-2} \int_B e^0 s(\mathrm{d}x) \int_{B'} e^0 s(\mathrm{d}x)$$
$$+ \frac{\alpha}{\beta} \left[1 + \frac{1}{\beta} \int [1 - e^0] s(\mathrm{d}x)\right]^{-\alpha-1} \int_{B \cap B'} s(\mathrm{d}x) \tag{50b}$$

$$= \frac{(\alpha)_2}{\beta^2} \int_B s(\mathrm{d}x) \int_{B'} s(\mathrm{d}x) + \frac{\alpha}{\beta} \int_{B \cap B'} s(\mathrm{d}x). \tag{50c}$$

Therefore,

$$\text{var}(B) \overset{(8)}{=} \mu^{(2)}(B \times B) - [\mu(B)]^2 \tag{51a}$$

$$= \mu(B) \left(1 + \frac{1}{\beta} \int_B s(\mathrm{d}x)\right). \tag{51b}$$

From (49) and (51) we get

$$\begin{cases} \mu(\mathcal{X}) = \dfrac{\alpha}{\beta} \\ \text{var}(\mathcal{X}) = \mu(\mathcal{X}) \left(1 + \dfrac{1}{\beta}\right), \end{cases} \tag{52}$$

which yields the desired result when solved for α and β. □

Proof of Thm. IV.4. Let us denote by $\mathcal{G}_{s,k}$ the PGFL of the point process describing the evolution of a target from the previous time step. For the sake of simplicity, we shall omit the time subscripts on the quantities related to the survival and birth process.

The PGFL $\mathcal{G}_{k|k-1}$ of the predicted target process takes the form $\mathcal{G}_{k|k-1}(h) = \mathcal{G}_{\mathrm{b}}(h)\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(h|\cdot))$. Here, the multiplicative structure stems from the independence between the newborn targets and those surviving from the previous time step; the composition appears because the survival process applies to each preexisting target described by the updated target process Φ_{k-1} from the previous time step [32, Eq. 5.5.18].

In order to produce the variance $\mathrm{var}_{k|k-1}$ of the predicted process via (8) we first build the second-order moment $\mu_{k|k-1}^{(2)}(B \times B')$ in arbitrary regions $B, B' \in \mathcal{B}(\mathcal{X})$. From (16) we have

$$\mu_{k|k-1}^{(2)}(B \times B') = \delta^2 \mathcal{L}_{k|k-1}(f; \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0} \tag{53a}$$

$$= \delta^2 \mathcal{G}_{k|k-1}(e^{-f}; \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0}. \tag{53b}$$

The product rule (18) gives

$$\begin{aligned}
\mu_{k|k-1}^{(2)}(B \times B') &= \delta^2 \mathcal{G}_{\mathrm{b}}(e^{-f}; \mathbb{1}_B; \mathbb{1}_{B'})\big|_{f=0}\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(1|\cdot)) \\
&\quad + \delta\mathcal{G}_{\mathrm{b}}(e^{-f}; \mathbb{1}_B)\big|_{f=0}\delta(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_{B'})\big|_{f=0} \\
&\quad + \delta\mathcal{G}_{\mathrm{b}}(e^{-f}; \mathbb{1}_{B'})\big|_{f=0}\delta(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B)\big|_{f=0} \\
&\quad + \mathcal{G}_{\mathrm{b}}(1)\delta^2(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0},
\end{aligned} \tag{53c}$$

$$\begin{aligned}
&\overset{(16)}{=} \mu_{\mathrm{b}}^{(2)}(B \times B') - \mu_{\mathrm{b}}(B)\delta(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_{B'})\big|_{f=0} \\
&\quad - \mu_{\mathrm{b}}(B')\delta(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B)\big|_{f=0} \\
&\quad + \delta^2(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0},
\end{aligned} \tag{53d}$$

where μ_{b} and $\mu_{\mathrm{b}}^{(2)}$ are the first- and second-order moment measures of the birth process, respectively. Let us first focus on the term $\delta(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B)\big|_{f=0}$ in (53d). Using the definition of the PGFL (12) we can write

$$\delta(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B)\big|_{f=0}$$
$$= \sum_{n \geq 0}\int_{\mathcal{X}^n}\delta\left(\left[\prod_{i=1}^n \mathcal{G}_{\mathrm{s}}(e^{-f}|x_i)\right]; \mathbb{1}_B\right)\bigg|_{f=0} P_{k-1}^{(n)}(\mathrm{d}x_{1:n}) \tag{54a}$$

$$\overset{(18)}{=} \sum_{n \geq 0}\int_{\mathcal{X}^n}\sum_{i=1}^n \delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x_i); \mathbb{1}_B)\big|_{f=0} P_{k-1}^{(n)}(\mathrm{d}x_{1:n}) \tag{54b}$$

$$\overset{(4)}{=} \int \delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x); \mathbb{1}_B)\big|_{f=0}\mu_{k-1}(\mathrm{d}x). \tag{54c}$$

The survival process for a target with state x at the previous time step can be described with a Bernoulli point process with parameter $p_{\mathrm{s}}(x)$ and spatial distribution $t(\cdot|x)$, and thus (24) gives

$$\mathcal{G}_{\mathrm{s}}(e^{-f}|x) = 1 - p_{\mathrm{s}}(x) + p_{\mathrm{s}}(x)\int e^{-f(y)}t(\mathrm{d}y|x). \tag{55}$$

It follows that

$$\delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x); \mathbb{1}_B) = p_{\mathrm{s}}(x)\int \delta(e^{-f(y)}; \mathbb{1}_B)t(\mathrm{d}y|x) \tag{56a}$$

$$= -p_{\mathrm{s}}(x)\int \mathbb{1}_B(y)e^{-f(y)}t(\mathrm{d}y|x), \tag{56b}$$

which leads to

$$\delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x); \mathbb{1}_B)\big|_{f=0} = -p_{\mathrm{s}}(x)t(B|x). \tag{57}$$

Substituting (57) in (54c) yields

$$\delta(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B)\big|_{f=0} = -\int p_{\mathrm{s}}(x)t(B|x)\mu_{k-1}(\mathrm{d}x). \tag{58}$$

Let us write the last term $\delta^2(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0}$ in (53d) in a similar manner as above. From the definition of the PGFL (12) we can write

$$\delta^2(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0} \tag{59a}$$

$$\begin{aligned}
&\overset{(18)}{=} \sum_{n \geq 0}\int_{\mathcal{X}^n}\sum_{i=1}^n \delta^2(\mathcal{G}_{\mathrm{s}}(e^{-f}|x_i); \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0} P_{k-1}^{(n)}(\mathrm{d}x_{1:n}) \\
&\quad + \sum_{n \geq 0}\int_{\mathcal{X}^n}\sum_{\substack{1 \leq i,j \leq n \\ i \neq j}}\delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x_i); \mathbb{1}_B)\big|_{f=0} \\
&\qquad \cdot \delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x_j); \mathbb{1}_{B'})\big|_{f=0} P_{k-1}^{(n)}(\mathrm{d}x_{1:n})
\end{aligned} \tag{59b}$$

$$\begin{aligned}
&\overset{(4)}{=} \int \delta^2(\mathcal{G}_{\mathrm{s}}(e^{-f}|x); \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0}\mu_{k-1}(\mathrm{d}x) \\
&\quad + \int \delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x); \mathbb{1}_B)\big|_{f=0} \\
&\qquad \cdot \delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x'); \mathbb{1}_{B'})\big|_{f=0}\nu_{k-1}^{(2)}(\mathrm{d}(x, x')).
\end{aligned} \tag{59c}$$

From (56), the value of $\delta^2(\mathcal{G}_{\mathrm{s}}(e^{-f}|x); \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0}$ is found to be

$$\delta^2(\mathcal{G}_{\mathrm{s}}(e^{-f}|x); \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0} = p_{\mathrm{s}}(x)t(B \cap B'|x), \tag{60}$$

so that (59c) becomes

$$\begin{aligned}
&\delta^2(\mathcal{G}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B, \mathbb{1}_{B'})|_{f=0} = \mu_{\mathrm{s}}(B \cap B') \\
&\quad + \int p_{\mathrm{s}}(x)t(B|x)p_{\mathrm{s}}(x')t(B'|x')\nu_{k-1}^{(2)}(\mathrm{d}(x, x')).
\end{aligned} \tag{61}$$

Substituting (58) and (61) in (53d) and setting $B = B'$ yields

$$\begin{aligned}
\mu_{k|k-1}^{(2)}(B \times B) &= \mu_{\mathrm{b}}^{(2)}(B \times B) + 2\mu_{\mathrm{b}}(B)\mu_{\mathrm{s}}(B) + \mu_{\mathrm{s}}(B) \\
&\quad + \int p_{\mathrm{s}}(x)t(B|x)p_{\mathrm{s}}(x')t(B|x')\nu_{k-1}^{(2)}(\mathrm{d}(x, x')).
\end{aligned} \tag{62}$$

Using the definition of the variance (8) then yields

$$\begin{aligned}
\mathrm{var}_{k|k-1}(B) &= \mathrm{var}_{\mathrm{b}}(B) + [\mu_{\mathrm{b}}(B)]^2 - [\mu_{k|k-1}(B)]^2 \\
&\quad + 2\mu_{\mathrm{b}}(B)\mu_{\mathrm{s}}(B) + \mu_{\mathrm{s}}(B) \\
&\quad + \int p_{\mathrm{s}}(x)t(B|x)p_{\mathrm{s}}(x')t(B|x')\nu_{k-1}^{(2)}(\mathrm{d}(x, x')),
\end{aligned} \tag{63}$$

and substituting the expression of the predicted intensity (30) to $\mu_{k|k-1}(B)$ in (63) yields the desired result. $\qquad\square$

Proof of Cor. IV.5. Let us assume that the probability of survival $p_{\mathrm{s},k}$ is uniform over the state space. First of all, Eq. (31) with $B = \mathcal{X}$ simplifies to

$$\mu_{\mathrm{s},k}(\mathcal{X}) = p_{\mathrm{s},k}\int \underbrace{t_{k|k-1}(\mathcal{X}|x)}_{=1}\mu_{k-1}(\mathrm{d}x) \tag{64a}$$

$$= p_{\mathrm{s},k}\mu_{k-1}(\mathcal{X}). \tag{64b}$$

From Eq. (32) we can then compute the variance of the survival process $\text{var}_{s,k}$ in the whole state space, i.e.

$$\text{var}_{s,k}(\mathcal{X}) = \mu_{s,k}(\mathcal{X})[1 - \mu_{s,k}(\mathcal{X})] \tag{65a}$$

$$+ p_{s,k}^2 \int \underbrace{t_{k|k-1}(\mathcal{X}|x)}_{=1} \underbrace{t_{k|k-1}(\mathcal{X}|x')}_{=1} \nu_{k-1}^{(2)}(\text{d}(x,x')) \tag{65b}$$

$$\overset{(64)}{=} p_{s,k}\mu_{k-1}(\mathcal{X})\Big[1 - p_{s,k}\mu_{k-1}(\mathcal{X})\Big] + p_{s,k}^2 \nu_{k-1}^{(2)}(\mathcal{X} \times \mathcal{X}) \tag{65c}$$

$$\overset{(10)}{=} p_{s,k}\mu_{k-1}(\mathcal{X})\Big[1 - p_{s,k}\mu_{k-1}(\mathcal{X})\Big]$$
$$+ p_{s,k}^2\Big[\mu_{k-1}^{(2)}(\mathcal{X} \times \mathcal{X}) - \mu_{k-1}(\mathcal{X})\Big] \tag{65d}$$

$$\overset{(8)}{=} p_{s,k}\mu_{k-1}(\mathcal{X})\Big[1 - p_{s,k}\mu_{k-1}(\mathcal{X})\Big]$$
$$+ p_{s,k}^2\Big[\text{var}_{k-1}(\mathcal{X}) + [\mu_{k-1}(\mathcal{X})]^2 - \mu_{k-1}(\mathcal{X})\Big] \tag{65e}$$

so $\text{var}_{s,k}(\mathcal{X}) = p_{s,k}^2\text{var}_{k-1}(\mathcal{X}) + p_{s,k}[1 - p_{s,k}]\mu_{k-1}(\mathcal{X})$. $\square$

Proof of Thm. IV.7. For the sake of simplicity, time subscripts will be omitted when there is no ambiguity. In addition, $q_d(x) := 1 - p_d(x)$ will denote the probability of missed detection for a target with state $x \in \mathcal{X}$, and $\mathcal{G}_{c,k}$ and $\mathcal{G}_{d,k}$ will denote the PGFLs of the clutter and target detection process, respectively.

From Assumptions IV.6 we can write the joint PGFL describing the predicted targets and the observations [2]:

$$\mathcal{G}_{J,k}(g,h) = \mathcal{G}_{k|k-1}\left(h\mathcal{G}_d(g|\cdot)\right)\mathcal{G}_c(g), \tag{66}$$

where the multiplicative form stems from the independence between the target-generated and the clutter measurements; the composition appears because the detection process applies to each target described by $\Phi_{k|k-1}$. Since both the predicted target and the clutter process are assumed Panjer, (66) takes the more specific form

$$\mathcal{G}_{J,k}(g,h) = \mu(\mathcal{X})^\alpha\left(F_d(g,h)\right)^{-\alpha}\left(F_c(g)\right)^{-\alpha_c}, \tag{67}$$

where

$$F_d(g,h) := \mu(\mathcal{X})\left(1 + \frac{1}{\beta}\int(1 - h(x)\mathcal{G}_d(g|x))s(\text{d}x)\right) \tag{68}$$

$$= \int\left[1 + \frac{1 - h(x)\mathcal{G}_d(g|x)}{\beta}\right]\mu(\text{d}x), \tag{69}$$

and $F_c(g) := 1 + \frac{1}{\beta_c}\int(1 - g(z))s_c(z)\text{d}z$. Note that the clutter term F_c follows directly from the definition of a Panjer process (27); the detection term (67) stems from (27) as well but is then scaled by the predicted mean number of targets $\mu(\mathcal{X})$, so that the final result of the theorem exploits similar notations as the CPHD filter in [8], [24]. The detection process in state x can be described with a Bernoulli point process with parameter $p_d(x)$ and spatial distribution density $l(\cdot|x)$, and thus (24) gives $\mathcal{G}_d(g|x) = q_d(x) + p_d(x)\int_{\mathcal{Z}}g(z)l(z|x)\text{d}z$. Note that both F_d and F_c are linear w.r.t. to the argument g, and thus only their first-order derivatives are non-zero; given an arbitrary measurement $z \in Z_k$, we can write

$$\delta F_d(g,h;\delta_z) = -\int\frac{h(x)p_d(x)l(z|x)}{\beta}\mu(\text{d}x), \tag{70}$$

and $\delta F_c(g;\delta_z) = -s_c(z)/\beta_c$. Similar to the PHD filter update [2], the PGFL of the updated target process Φ_k is obtained from the differentiation of the joint PGFL (67) using Bayes' rule:

$$\mathcal{G}_k(h) = \frac{\delta^{|Z_k|}\mathcal{G}_{J,k}(g,h;(\delta_z)_{z\in Z_k})|_{g=0}}{\delta^{|Z_k|}\mathcal{G}_{J,k}(g,1;(\delta_z)_{z\in Z_k}))|_{g=0}}. \tag{71}$$

Using the higher-order product (20) and chain (21) rules, the $|Z_k|$-th derivative of the joint PGFL (67) in directions $(\delta_z)_{z\in Z_k}$ yields

$$\delta^{|Z_k|}\mathcal{G}_{J,k}(g,h;(\delta_z)_{z\in Z_k}) \propto \sum_{j=0}^{|Z_k|}\frac{(\alpha)_j}{\beta^j}\frac{(\alpha_c)_{|Z_k|-j}}{\beta_c^{|Z_k|-j}}F_d(g,h)^{-j}$$

$$\cdot F_c(g)^{-|Z_k|+j}\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\left(\prod_{z\in Z}F_d^z(h)\prod_{z'\in Z_k\setminus Z}s_c(z')\right) \tag{72a}$$

$$\propto \sum_{j=0}^{|Z_k|}\frac{(\alpha)_j}{\beta^j}\frac{(\alpha_c)_{|Z_k|-j}}{(\beta_c F_c(g))^{|Z_k|-j}}F_d(g,h)^{-j}\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F_d^z(h)}{s_c(z)}, \tag{72b}$$

where $F_d^z(h) := \int h(x)p_d(x)l(z|x)\mu(\text{d}x)$. The proportional constant in (72) is the quantity $\mu(\mathcal{X})^\alpha F_d(g,h)^{-\alpha}F_c(g)^{-\alpha_c}\prod_{z\in Z_k}s_c(z)$; since it is discarded in the ratio (71), it will be omitted from now on. Details of the developments leading to (72) can be found in Lem. VI.6 in [18], where a similar result is produced.

Similar to [2], we can finally compute the intensity of the updated target process Φ_k in any region $B \in \mathcal{B}(\mathcal{X})$ from the first-order derivative of its PGFL (71), i.e.

$$\mu_k(B) = \frac{\delta^{|Z_k|+1}\mathcal{G}_{J,k}(g,h;(\delta_z)_{z\in Z_k},\mathbb{1}_B)|_{g=0,h=1}}{\delta^{|Z_k|}\mathcal{G}_{J,k}(g,1;(\delta_z)_{z\in Z_k}))|_{g=0}}. \tag{73}$$

We first need to differentiate (72) in direction $\mathbb{1}_B$ through the product rule (18) and get

$$\delta^{|Z_k|+1}\mathcal{G}_{J,k}(g,h;(\delta_z)_{z\in Z_k},\mathbb{1}_B) \propto (-\beta\delta F_d(g,h;\mathbb{1}_B))$$

$$\cdot\sum_{j=0}^{|Z_k|}\frac{(\alpha)_{j+1}}{\beta^{j+1}}\frac{(\alpha_c)_{|Z_k|-j}}{(\beta_c F_c(g))^{|Z_k|-j}}F_d(g,h)^{-j-1}\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}F_d^z(h)$$

$$+\sum_{z\in Z_k}F_d^z(\mathbb{1}_B)\sum_{j=0}^{|Z_k|-1}\frac{(\alpha)_{j+1}}{\beta^{j+1}}\frac{(\alpha_c)_{(|Z_k|-1)-j}}{(\beta_c F_c(g))^{(|Z_k|-1)-j}}$$

$$\cdot F_d(g,h)^{-j-1}\sum_{\substack{Z\subseteq Z_k\setminus\{z\}\\|Z|=j}}\prod_{z'\in Z_k}F_d^{z'}(h), \tag{74}$$

where

$$\delta F_d(g,h;\mathbb{1}_B)$$
$$= -\frac{1}{\beta}\int_B\left[q_d(x) + p_d(x)\int_{\mathcal{Z}}g(z)\ell(z|x)\text{d}z\right]\mu(\text{d}x). \tag{75}$$

Substituting (72) and (74) into (73) yields the result. $\square$

Proof of Thm. IV.8. From (16), we can compute the second-order moment measure $\mu_k^{(2)}$ in any regions $B, B' \in \mathcal{B}(\mathcal{X})$ from the second-order derivative of the Laplace functional $\mathcal{L}_k$ of the updated target process Φ_k. Substituting $\exp(-f)$ to h in

the PGFL (71) yields the expression of the Laplace functional $\mathcal{L}_k$, and from (16) it follows that [24]

$$\mu_k^{(2)}(B \times B') = \frac{\delta^{|Z_k|+2}\mathcal{G}_{\mathrm{J},k}(g, e^{-f}; (\delta_z)_{z \in Z_k}, \mathbb{1}_B, \mathbb{1}_{B'})|_{g=0,f=0}}{\delta^{|Z_k|}\mathcal{G}_{\mathrm{J},k}(g, 1; (\delta_z)_{z \in Z_k}))|_{g=0}}. \tag{76}$$

The denominator in (76) has already been computed in (72); we shall thus focus here on the derivation in directions $\mathbb{1}_B, \mathbb{1}_{B'}$ of the numerator

$$\delta^{|Z_k|}\mathcal{G}_{\mathrm{J},k}(0, e^{-f}; (\delta_z)_{z \in Z_k}) \propto \sum_{j=0}^{|Z_k|} \frac{(\alpha)_j}{\beta^j} \frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{(1+\beta_{\mathrm{c}})^{|Z_k|-j}}$$
$$\cdot F_{\mathrm{d}}(0, e^{-f})^{-j} \sum_{\substack{Z \subseteq Z_k \\ |Z|=j}} \prod_{z \in Z} \frac{F_{\mathrm{d}}^z(e^{-f})}{s_{\mathrm{c}}(z)}. \tag{77}$$

The first-order derivative of (77) in direction $\mathbb{1}_B$ is

$$\delta^{|Z_k|+1}\mathcal{G}_{\mathrm{J},k}(0, e^{-f}; (\delta_z)_{z \in Z_k}, \mathbb{1}_B)$$
$$\propto -\sum_{j=0}^{|Z_k|} \frac{(\alpha)_{j+1}}{\beta^{j+1}} \frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{(\beta_{\mathrm{c}}+1)^{|Z_k|-j}} F_{\mathrm{d}}(0, e^{-f})^{-j-1}$$
$$\cdot F_{\mathrm{md}}(e^{-f}\mathbb{1}_B) \sum_{\substack{Z \subseteq Z_k \\ |Z|=j}} \prod_{z \in Z} F_{\mathrm{d}}^z(e^{-f})$$
$$-\sum_{j=1}^{|Z_k|} \frac{(\alpha)_j}{\beta^j} \frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{(\beta_{\mathrm{c}}+1)^{|Z_k|-j}} F_{\mathrm{d}}(0, e^{-f})^{-j}$$
$$\cdot \sum_{z \in Z_k} \frac{F_{\mathrm{d}}^z(e^{-f}\mathbb{1}_B)}{s_{\mathrm{c}}(z)} \sum_{Z \subseteq Z_k \setminus \{z\}} \prod_{z' \in Z} \frac{F_{\mathrm{d}}^{z'}(e^{-f})}{s_{\mathrm{c}}(z)}, \tag{78}$$

where $F_{\mathrm{md}}(h) := \int h(x) q_{\mathrm{d}}(x) \mu(\mathrm{d}x)$. In a similar manner, (78) can be differentiated another time in the direction of $\mathbb{1}_{B'}$ to yield the second-order derivative of (77) in directions $\mathbb{1}_B, \mathbb{1}_{B'}$. Substituting this second-order derivative and Eq. (72) into (76) yields

$$\mu_k^{(2)}(B \times B') = \mu_k(B \cap B') + \mu_k^\phi(B)\mu_k^\phi(B')\ell_2(\phi)$$
$$+ \mu_k^\phi(B) \sum_{z \in Z} \frac{\mu_k^z(B')}{s_{\mathrm{c}}(z)}\ell_2(z) + \mu_k^\phi(B') \sum_{z \in Z} \frac{\mu_k^z(B)}{s_{\mathrm{c}}(z)}\ell_2(z)$$
$$+ \sum_{z,z' \in Z_k} \frac{\mu_k^z(B)}{s_{\mathrm{c}}(z)} \frac{s_k^{z'}(B')}{s_{\mathrm{c}}(z')}\ell_2^{\neq}(z, z'). \tag{79}$$

Following (8), the intensity (41) is then squared and subtracted from the second-order moment (79) evaluated with $B' = B$ in order to yield the desired quantity $\mathrm{var}_k(B)$. $\quad\square$

Proof of Cor. IV.9. Let us assume that the predicted target process $\Phi_{k|k-1}$ is Poisson with rate $\lambda_{k|k-1}$, i.e., $\alpha_{k|k-1}, \beta_{k|k-1} \to \infty$, with constant ratio $\lambda_{k|k-1} = \frac{\alpha}{\beta}$. For the same of simplicity, the time subscripts on $\alpha_{k|k-1}, \beta_{k|k-1}, \lambda_{k|k-1}$ are omitted for the rest of the proof. Note first that, since $\mu(\mathrm{d}x) = \lambda s(\mathrm{d}x)$, we have

$$\lim_{\alpha,\beta \to \infty} F_{\mathrm{d}} = \lim_{\alpha,\beta \to \infty} \int \left[1 + \underbrace{\frac{p_{\mathrm{d},k}(x)}{\beta}}_{\to 0}\right]\lambda s(\mathrm{d}x) = \lambda. \tag{80}$$

In order to check the convergence of the intensity update equation (41), we only need to check the convergence of the

term $\Upsilon_u(Z)$ in (38) as it is the only term that contains α or β. We can write:

$$\lim_{\alpha,\beta \to \infty} \sum_{j=0}^{|Z|} \frac{(\alpha)_{j+u}}{(\beta)^{j+u}} \frac{(\alpha_{\mathrm{c}})_{|Z|-j}}{(\beta_{\mathrm{c}}+1)^{|Z|-j}} F_{\mathrm{d}}^{-j-u} e_j(Z) \tag{81}$$

$$\stackrel{(36)}{=} \lim_{\alpha,\beta \to \infty} \sum_{j=0}^{|Z_k|} \prod_{i=0}^{j+u-1} \underbrace{\left(\lambda + \frac{i}{\beta}\right)}_{\to \lambda} \frac{(\alpha_{\mathrm{c}})_{|Z|-j}}{(\beta_{\mathrm{c}}+1)^{|Z|-j}} \underbrace{F_{\mathrm{d}}^{-j-u}}_{\to \lambda^{-j-u}} e_j(Z) \tag{82}$$

$$= \sum_{j=0}^{|Z|} \frac{(\alpha_{\mathrm{c}})_{|Z|-j}}{(\beta_{\mathrm{c}}+1)^{|Z|-j}} e_j(Z). \tag{83}$$

Note in particular that the limit of $\Upsilon_u(Z)$ is independent of the value of u; the corrective terms (37) thus converge to

$$\begin{cases} \lim\limits_{\alpha,\beta \to \infty} \ell_1(\phi) = 1 \\ \lim\limits_{\alpha,\beta \to \infty} \ell_1(z) = \dfrac{\sum_{j=0}^{|Z_k|-1} \frac{(\alpha_{\mathrm{c}})_{|Z_k|-j-1}}{(\beta_{\mathrm{c}}+1)^{|Z_k|-j-1}} e_j(Z_k \setminus \{z\})}{\sum_{j=0}^{|Z_k|} \frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{(\beta_{\mathrm{c}}+1)^{|Z_k|-j}} e_j(Z_k)}, \end{cases} \tag{84}$$

which coincides with the results of Thm III.2 in [18].

If we further assume that the clutter process is Poisson, the intensity update equation (41) further converges to the intensity update equation of the original PHD filter [18]. $\quad\square$

Proof of Prop. V.1. The covariance is found with Eq. (7). For the SO-PHD filter, the first- and second-order moment measures are given by Eqns (41) and (79). For the PHD filter, they are given by (28) and (31) in [24], and for the CPHD filter by (19) and (29) ibid. $\quad\square$

APPENDIX B
SECOND-ORDER GM-PHD FILTER

Algorithm 1: Prediction (time k).

Input
Posterior: $\{w_{k-1}^{(i)}, m_{k-1}^{(i)}, P_{k-1}^{(i)}\}_{i=1}^{N_{k-1}}, \mathrm{var}_{k-1}(\mathcal{X})$
Birth: $\{w_{\mathrm{b},k-1}^{(i)}, m_{\mathrm{b},k-1}^{(i)}, P_{\mathrm{b},k-1}^{(i)}\}_{i=1}^{N_{\mathrm{b},k-1}}, \mathrm{var}_{\mathrm{b},k}(\mathcal{X})$

Survival process
$\mu_{k-1}(\mathcal{X}) := \sum_{i=1}^{N_{k-1}} w_{k-1}^{(i)}$
for $1 \le i \le n_{k-1}$ **do**
$\quad w_{k|k-1}^{(i)} := p_{\mathrm{s},k} w_{k-1}^{(i)}$
$\quad m_{k|k-1}^{(i)} := F_{k-1} m_{k-1}^{(i)}$
$\quad P_{k|k-1}^{(i)} := F_{k-1} P_{k-1}^{(i)} F_{k-1}^T + Q_{k-1}$
end for
$\mathrm{var}_{\mathrm{s},k}(\mathcal{X}) := p_{\mathrm{s},k}^2 \mathrm{var}_{k-1}(\mathcal{X}) + p_{\mathrm{s},k}[1 - p_{\mathrm{s},k}]\mu_{k-1}(\mathcal{X})$

Newborn process
for $1 \le j \le N_{\mathrm{b},k-1}$ **do**
$\quad \{w, m, P\}_{k|k-1}^{(N_{k-1}+j)} := \{w, m, P\}_{\mathrm{b},k-1}^{(j)}$
end for

$N_{k|k-1} := N_{k-1} + N_{b,k-1}$

$\text{var}_{k|k-1}(\mathcal{X}) := \text{var}_{b,k}(\mathcal{X}) + \text{var}_{s,k}(\mathcal{X})$

Output
Prediction: $\{w_{k|k-1}^{(i)}, m_{k|k-1}^{(i)}, P_{k|k-1}^{(i)}\}_{i=1}^{N_{k|k-1}}, \text{var}_{k|k-1}(\mathcal{X})$

Algorithm 2: Elementary symmetric functions.

Input
Collection of terms: $\{\mu_k^z(\mathcal{X})\}_{z \in Z_k}$

Vieta's theorem
Expand: $p(x) = \prod_{z \in Z_k}(x - \mu_k^z(\mathcal{X})) = \sum_{j=0}^{m_k} p_j x^j$
Set $e_j(Z_k) = p_j$ for all $0 \le j \le m_k$

Output
$\{e_j(Z_k)\}_{0 \le j \le m_k}$

Algorithm 3: Data update (time k).

Input
Prediction: $\{w_{k|k-1}^{(i)}, m_{k|k-1}^{(i)}, P_{k|k-1}^{(i)}\}_{i=1}^{N_{k|k-1}}, \text{var}_{k|k-1}(\mathcal{X})$
Current measurements: $Z_k = \{z_j\}_{j=1}^{M_k}$

Panjer parameters
$\mu_{k|k-1}(\mathcal{X}) := \sum_{i=1}^{N_{k|k-1}} w_{k|k-1}^{(i)}$
$\alpha_{k|k-1} := \mu_{k|k-1}(\mathcal{X})^2/(\text{var}_{k|k-1}(\mathcal{X}) - \mu_{k|k-1}(\mathcal{X}))$
$\beta_{k|k-1} := \mu_{k|k-1}(\mathcal{X})/(\text{var}_{k|k-1}(\mathcal{X}) - \mu_{k|k-1}(\mathcal{X}))$

Missed detection and measurement terms
for $1 \le i \le N_{k|k-1}$ **do**
    $w_{\phi,k}^{(i)} := (1 - p_{d,k})w_{k|k-1}^{(i)}$
    $m_{\phi,k}^{(i)} := m_{k|k-1}^{(i)}$
    $P_{\phi,k}^{(i)} := P_{k|k-1}^{(i)}$
end for
$\mu_k^\phi(\mathcal{X}) := (1 - p_{d,k})\mu_{k|k-1}(\mathcal{X})$
for $1 \le j \le M_k$ **do**
    for $1 \le i \le N_{k|k-1}$ **do**
        $y_k^{(i,j)} := z_j - H_k m_{k|k-1}^{(i)}$
        $S_k^{(i)} := H_k P_{k|k-1}^{(i)} H_k^T + R_k$
        $K_k^{(i)} := P_{k|k-1}^{(i)} H_k^T [S_k^{(i)}]^{-1}$
        $w_{d,k}^{(i,j)} := p_{d,k} w_{d,k|k-1}^{(i,j)} \mathcal{N}(z; y_k^{(i,j)}, S_k^{(i)})/s_{c,k}$
        $m_{d,k}^{(i,j)} := m_{k|k-1}^{(i)} + K_k^{(i)} y_k^{(i,j)}$
        $P_{d,k}^{(i,j)} := (I - K_k^{(i)} H_k)P_{k|k-1}^{(i)}$
    end for
    $\mu_k^{z_j}(\mathcal{X}) := \sum_{i=1}^{N_{k|k-1}} w_{d,k}^{(i,j)}$
end for

Corrective terms
$F_d := (1 + \frac{p_{d,k}}{\beta_{k|k-1}}) \sum_{z \in Z_k} \mu_k^{z_j}(\mathcal{X})$
Compute $\{e_d(Z_k)\}_{0 \le d \le M_k}$ using Alg. 1
for $0 \le u \le 2$ **do**
    $\Upsilon_u(Z_k) := \sum_{j=0}^{M_k} \frac{(\alpha_{k|k-1})_{j+u}}{(\beta_{k|k-1})^{j+u}} \frac{(\alpha_{c,k})_{m_k-j}}{(\beta_{c,k}+1)^{m_k-j}} F_d^{-j-u} e_j(Z_k)$
end for

$\ell_1(\phi) := \Upsilon_1(Z_k)/\Upsilon_0(Z_k), \quad \ell_2(\phi) := \Upsilon_2(Z_k)/\Upsilon_0(Z_k)$
for $1 \le i \le M_k$ **do**
    Compute $\{e_d(Z_k \setminus z_i)\}_{0 \le d \le M_k-1}$ using Alg. 1
    for $1 \le u \le 2$ **do**
        $\Upsilon_u(Z_k \setminus z_i) := \sum_{d=0}^{M_k-1} \frac{(\alpha_{k|k-1})_{d+u}}{(\beta_{k|k-1})^{d+u}}$
        $\cdot \frac{(\alpha_{c,k})_{m_k-1-d}}{(\beta_{c,k}+1)^{m_k-1-d}} F_d^{-d-u} e_d(Z_k \setminus z_i)$
    end for
    $\ell_1(z_i) := \Upsilon_1(Z_k \setminus z_i)/\Upsilon_0(Z_k)$
    $\ell_2(z_i) := \Upsilon_2(Z_k \setminus z_i)/\Upsilon_0(Z_k)$
    for $1 \le i < j \le M_k$ **do**
        Compute $\{e_d(Z_k \setminus \{z_i, z_j\})\}_{0 \le d \le M_k-2}$ using Alg. 1
        $\Upsilon_2(Z_k \setminus \{z_i, z_j\}) := \sum_{d=0}^{M_k-2} \frac{(\alpha_{k|k-1})_{d+2}}{(\beta_{k|k-1})^{d+2}}$
        $\cdot \frac{(\alpha_{c,k})_{m_k-2-d}}{(\beta_{c,k}+1)^{m_k-2-d}} F_d^{-d-2} e_d(Z_k \setminus \{z_i, z_j\})$
        $\ell_2^{\ne}(z_i, z_j) := \Upsilon_2(Z_k \setminus \{z_i, z_j\})/\Upsilon_0(Z_k)$
    end for
end for

Missed detection terms
for $1 \le i \le N_{k|k-1}$ **do**
    $w_k^{(i)} := \ell_1(\phi)w_{\phi,k}^{(i)}$
    $m_k^{(i)} := m_{\phi,k}^{(i)}$
    $P_k^{(i)} := P_{\phi,k}^{(i)}$
    Association terms
    for $1 \le j \le M_k$ **do**
        $w_k^{(i \cdot n_{k|k-1}+j)} := \ell_1(z_j)w_{d,k}^{(i,j)}$
        $m_k^{(i \cdot n_{k|k-1}+j)} := m_{d,k}^{(i,j)}$
        $P_k^{(i \cdot n_{k|k-1}+j)} := P_{d,k}^{(i,j)}$
    end for
end for
$N_k := N_{k|k-1} + N_{k|k-1}M_k$
$\mu_k(\mathcal{X}) := \sum_{i=1}^{N_k} w_k^{(i)}$

Variance update
$\text{var}_k(\mathcal{X}) := \mu_k(\mathcal{X}) + \mu_k^\phi(\mathcal{X})^2 [\ell_2(\phi) - \ell_1(\phi)^2]$
$+ 2\mu_k^\phi(\mathcal{X}) \sum_{z \in Z_k} \frac{\mu_k^z(\mathcal{X})}{s_{c,k}(z)} [\ell_2(z) - \ell_1(\phi)\ell_1(z)]$
$+ \sum_{z \ne z' \in Z_k} \frac{\mu_k^z(\mathcal{X})}{s_{c,k}(z)} \frac{\mu_k^{z'}(\mathcal{X})}{s_{c,k}(z')} [\ell_2^{\ne}(z,z') - \ell_1(z)\ell_1(z')]$

Output
Posterior: $\{w_k^{(i)}, m_k^{(i)}, P_k^{(i)}\}_{i=1}^{N_k}, \text{var}_k(\mathcal{X})$

REFERENCES

[1] R. P. S. Mahler, *Statistical multisource-multitarget information fusion*. Boston: Artech House, 2007.

[2] ——, "Multitarget Bayes filtering via first-order multitarget moments," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 39, no. 4, pp. 1152–1178, 2003.

[3] B.-N. Vo and W.-K. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4091–4104, 2006.

[4] B.-N. Vo, S. S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multitarget filtering with random finite sets," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 41, no. 4, pp. 1224–1245, 2005.

[5] O. Erdinc, P. Willett, and Y. Bar-Shalom, "Probability hypothesis density filter for multitarget multisensor tracking," in *Information Fusion, Proceedings of the 7th International Conference on*, vol. 1, Jul. 2005, p. 8.

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

[7] R. Mahler, "PHD filters of second order in target number," in *Proc. SPIE Defense and Security Symposium*, vol. 6236. International Society for Optics and Photonics, 2006.

[8] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic Implementations of the Cardinalized Probability Hypothesis Density Filter," *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3553–3567, Jul. 2007.

[9] D. Fränken, M. Schmidt, and M. Ulmke, "Spooky Action at a Distance in the Cardinalized Probability Hypothesis Density Filter," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 45, no. 4, pp. 1657–1664, Oct. 2009.

[10] S. S. Singh, B.-N. Vo, A. Baddeley, and S. Zuyev, "Filters for Spatial Point Processes," *SIAM Journal on Control and Optimization*, vol. 48, no. 4, pp. 2275–2295, 2009.

[11] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The cardinality balanced multi-target multi-Bernoulli filter and its implementations," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 409–423, 2009.

[12] J. L. Williams, "Hybrid Poisson and multi-Bernoulli filters," in *Information Fusion (FUSION), 2012 15th International Conference on*. IEEE, 2012, pp. 1103–1110.

[13] ——, "An efficient, variational approximation of the best fitting multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 258–273, 2015.

[14] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3460–3475, 2013.

[15] B.-N. Vo, B.-T. Vo, and D. Phung, "Labeled random finite sets and the Bayes multi-target tracking filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6554–6567, 2014.

[16] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, "The labeled multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3246–3260, 2014.

[17] M. Fackler, "Panjer class united – one formula for the Poisson, Binomial, and Negative Binomial distribution," *ASTIN colloquium*, 2009.

[18] I. Schlangen, E. D. Delande, J. Houssineau, and D. E. Clark, "A PHD filter with Negative Binomial Clutter," in *Information Fusion, Proceedings of the 16th International Conference on*, Jul. 2016, pp. 658–665.

[19] I. Schlangen, D. E. Clark, and E. D. Delande, "Single-cluster PHD filter methods for joint multi-object filtering and parameter estimation," 2017, arXiv:1705.05312.

[20] P. Bernhard, "Chain differentials with an application to the mathematical fear operator," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 62, no. 7, pp. 1225–1233, 2005.

[21] O. Nikodym, "Sur une généralisation des intégrales de MJ Radon," *Fundamenta Mathematicae*, vol. 15, no. 1, pp. 131–179, 1930.

[22] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 1997.

[23] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons, 2008, vol. 70.

[24] E. D. Delande, M. Üney, J. Houssineau, and D. E. Clark, "Regional Variance for Multi-Object Filtering," *Signal Processing, IEEE Transactions on*, vol. 62, no. 13, pp. 3415–3428, Jul. 2014.

[25] D. E. Clark and R. P. S. Mahler, "Generalized PHD filters via a general chain rule," in *Information Fusion, Proceedings of the 15th International Conference on*, Jul. 2012, pp. 157–164.

[26] D. E. Clark and J. Houssineau, "Faa di Bruno's formula for chain differentials," *ArXiv e-prints*, Oct. 2013, arXiv:1202.0264v4.

[27] D. E. Clark, J. Houssineau, and E. D. Delande, "A few calculus rules for chain differentials," *ArXiv e-prints*, Jun. 2015, arXiv:1506.08626v1.

[28] D. E. Clark and J. Houssineau, "Faa di Bruno's formula and spatial cluster modelling," *Spatial Statistics*, vol. 6, pp. 109–117, 2013.

[29] ——, "Faà Di Bruno's formula and Volterra series," in *2014 IEEE Workshop on Statistical Signal Processing (SSP)*. IEEE, 2014, pp. 217–219.

[30] D. S. Bryant, E. D. Delande, S. Gehly, J. Houssineau, D. E. Clark, and B. A. Jones, "The CPHD Filter with Target Spawning," *Signal Processing, IEEE Transactions on*, vol. 65, no. 5, pp. 13 124–13 138, 2017.

[31] S. A. Klugman, H. H. Panjer, and G. Willmot, *Loss Models: From Data to Decisions*, ser. Wiley Series in Probability and Statistics. Wiley, 2012.

[32] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes. vol. I. , Elementary theory and methods*, ser. Probability and its applications. New York, Berlin, Paris: Springer, 2003.

[33] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A Consistent Metric for Performance Evaluation of Multi-Object Filters," *Signal Processing, IEEE Transactions on*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.

Isabel Schlangen is a current Ph.D. studentship holder at the Edinburgh Super-Resolution Imaging Consortium, UK. She received a German diploma in Mathematics from the University of Bonn (Germany) in 2012 and a joint MSc degree in Vision and Robotics from the Universities of Burgundy (France), Girona (Spain), and Heriot-Watt (Edinburgh, UK) in 2014. Her current research interests are multi-target estimation, probability theory and image analysis in a mainly biomedical context.

Emmanuel D. Delande received an Eng. degree from the Ecole Centrale de Lille, Lille, and a M.Sc. degree in automatic control and signal processing from the University of Science & Technology, Lille, both in 2008. He was awarded his Ph.D. in 2012 from the Ecole Centrale de Lille. He is a research associate at Heriot-Watt University in Edinburgh. His research interests are in the design and the implementation of multi-object filtering solutions for multiple target tracking and sensor management problems.

Jérémie Houssineau received an Eng. degree in mathematical and mechanical modelling from MAT-MECA, Bordeaux, and a M.Sc. degree in mathematical modelling and statistics from the University of Bordeaux, both in 2009. From 2009 to 2011, he was a Research Engineer with DCNS, Toulon. He received his Ph.D. degree in statistical signal processing from Heriot-Watt University, Edinburgh, in 2015. He is currently a Research Fellow in the department of Statistics and Applied Probability of the National University of Singapore.

Daniel Clark is in the Departement CITI in Telecom SudParis in France. From 2007-2017 he was in the School of Engineering and Physical Sciences at Heriot-Watt University. His research interests are in the development of the theory and applications of multi-object estimation algorithms for sensor fusion problems. He has led a range of projects spanning theoretical algorithm development to practical deployment. He was awarded his Ph.D. in 2006 from Heriot-Watt University.