19th International Conference on Information Fusion
Heidelberg, Germany - July 5-8, 2016

A PHD Filter With Negative Binomial Clutter

Isabel Schlangen*, Emmanuel Delande*, Jérémie Houssineau* and Daniel E. Clark*
* School of Electrical and Physical Sciences
Heriot-Watt University, Edinburgh EH14 4AS, UK
Email: {is117, E.D.Delande, J.Houssineau, D.E.Clark}@hw.ac.uk

***Abstract*—The Probability Hypothesis Density (PHD) filter has brought significant advances in multi-object estimation since it not only estimates the spatial distribution of a population of objects but also provides estimates about the actual object number which is unknown in many scenarios. However, strict assumptions have to be made for the framework to be practical, in particular through the choice of the distribution for target and false alarm numbers. The exponential form of the Poisson distribution, for example, offers great simplicity in the derivation of the filter, but in many applications, this assumption is very restrictive, e.g. when there is a lot of variability in the number of measurements. This paper introduces a variation of the original PHD filter which assumes a negative binomial distribution for the false alarm number. It will be demonstrated that the altered formulation of the filter simply leads to an additional factor in the update equation and that the original PHD filter is a special case of the proposed method. A Gaussian Mixture (GM) implementation of both filters is used to test and compare their performance on simulated data.**

I. Introduction

Multi-object estimation has become of increasing importance in a large variety of applications, ranging from maritime or aerial surveillance to space situational awareness and even biomedicine. The quality and quantity of the data that is recorded over time varies greatly depending on the nature of the sensor design and the monitored scene, and it is almost impossible to avoid the occurrence of clutter. Therefore, clutter compensation techniques have to be found in order to accommodate the different sensor modalities and to improve the estimation results.

Some clutter can be highly dynamic with a large variation in the number of clutter points. In the case of radar surveillance on the open sea for instance, various types of false alarms are recorded which are known as *sea clutter*. Some reflections on the sea water might give more or less static interferences, but especially the crests of waves in rough sea conditions can lead to spontaneous outbursts of false alarms, called *burst scattering* [1]. A similar phenomenon can be observed in super-resolution fluorescence microscopy which is subject to severe intensity changes due to repeated excitation illumination with a laser beam [2], [3].

In Probability Hypothesis Density (PHD) filtering which was introduced by Mahler in 2003 [4], the number of false alarms is usually assumed to be Poisson distributed which is a reasonable constraint for scenarios with a moderately varying number of clutter points in the acquired data. However, one important property of the Poisson distribution is that its variance and mean are equal, and in some scenarios like described above, the restriction on the variance of the Poisson distribution might lead to a poor description of the actual clutter behaviour that might show a very high variance in the false alarm occurrence. Previous works suggested methods to estimate the clutter intensity and its cardinality statistics [5]–[7], where the number of clutter points is either Poisson or binomial distributed, however both of those distributions assume that the variance is less than or equal to the mean.

In this paper, a variation of the original PHD filter is presented which assumes a Negative Binomial (NB) distribution of the clutter number, allowing a large variance in the occurrence of clutter points. In Sec. II, basic notions on finite set statistics are given to facilitate the formulation of the different models for possible target populations and the new filtering method, given in Sec. III. A Gaussian-Mixture implementation of both filters is used to give a performance comparison in Sec. IV before the article concludes. An appendix delivers additional formulae that are used in the proof of the filter.

II. Point processes and probability generating functionals

The PHD filter stems from Finite Set Statistics (FISST) which is extensively treated in [8]. Some basic concepts will be stated in the following since they are important for the derivation of the new filter.

Point processes are random objects describing a random number of points in a given state space $\mathcal{X}$, usually $\mathbb{R}^d$. They are useful for scenarios where both the object number and their spatio-temporal behaviour are unknown since both aspects are modelled together. Just like for other random objects, the probability density p_Φ of a point process Φ is represented by its Probability Generating Functional (PGFL) as an infinite sum of the form

$$G_\Phi(h) = \sum_{n\geq 0} \int_{\mathcal{X}^n} \left[\prod_{i=1}^{n} h(x_i)\right] p_\Phi(x_{1:n}) \mathrm{d}(x_{1:n}) \qquad (1)$$

where $x_{1:n} = (x_1, \ldots, x_n) \in \mathcal{X}^n$. This notation is important for the modelling of such processes since it expresses both a variation over all possible numbers of objects in a target space $\mathcal{X}$ and the actual target states themselves. Furthermore, it gives back the first-order moment density (or intensity, or Probability Hypothesis Density) $\mu_\Phi(x)$ and the probability density $p_\Phi(x_1, \ldots, x_n)$ for each cardinality n immediately via

$$\mu_\Phi(x) = \delta G_\Phi(h; \delta_x)|_{h=1}, \qquad (2)$$

$$p_\Phi(x_1, \ldots, x_n) = \frac{1}{n!} \delta^n G_\Phi(h; \delta_{x_1}, \ldots, \delta_{x_n})\Big|_{h=0} \qquad (3)$$

978-0-9964527-4-8©2016 ISIF

where $\delta_{x_i}, i \in \{1, \ldots, n\}$ are Dirac delta functions and δG denotes the differentiation operator applied on a functional G.

Different ways have been described in the past to differentiate a functional $G(.)$, for example the Fréchet and Gâteaux differentials. However, the choice of this article is the so called *chain differential* since it allows the description of a chain rule [9]–[11]. This has been adopted in FISST [12], [13] since the original four chain rules were insufficiently general [8, p. 390-1].

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

$$\delta G(h; \eta) = \lim_{n \to \infty} \frac{G(h + \varepsilon_n \eta_n) - G(h)}{\varepsilon_n}, \tag{4}$$

where $(\varepsilon_n)_{n \in \mathbb{N}}$ is a series of real numbers converging to 0 and $(\eta_n : \mathcal{X} \to \mathbb{R}^+)_{n \in \mathbb{N}}$ is a series of functions which converge pointwise to η.

Just like the differential for ordinary functions, the chain differential is linear in its functional argument. Furthermore, it yields similar differentiation rules

$$\delta(F \cdot G)(h; \eta) = \delta F(h; \eta) G(h) + F(h) \delta G(h; \eta) \tag{5}$$

and

$$\delta(F \circ G)(h; \eta) = \delta F(G(h); \delta G(h; \eta)) \tag{6}$$

which are also called the *product rule* and *chain rule*, respectively. In case of the chain rule, however, note the nested structure in contrast to the multiplicative structure in the ordinary chain rule for functions. Both rules can be generalised to the nth-order derivatives for functional products and compositions as follows:

$$\begin{aligned} &\delta^n (F \cdot G)(h; \eta_1, \ldots, \eta_n) \\ &\quad = \sum_{\omega \subseteq \{1, \ldots, n\}} \delta^{|\omega|} F\Big(h; (\eta_i)_{i \in \omega}\Big) \delta^{n - |\omega|} G\Big(h; (\eta_j)_{j \in \bar{\omega}}\Big), \end{aligned} \tag{7}$$

where $\bar{\omega} = \{1, \ldots, n\} \setminus \omega$ is the complement of ω, and

$$\begin{aligned} &\delta^n (F \circ G)(h; \eta_1, \ldots, \eta_n) \\ &\quad = \sum_{\pi \in \Pi_n} \delta^{|\pi|} F\left(G(h); \Big(\delta^{|\omega|} G(h'; (\eta_i)_{i \in \omega})\Big)_{\omega \in \pi}\right), \end{aligned} \tag{8}$$

where Π_n is the set of partitions of the index set $\{1, \ldots, n\}$. Eq. (8) is also known as *Faà di Bruno's formula for chain differentials* [9], [10].

III. The PHD filter with Negative Binomial clutter model

In this section, the proposed filter will be presented. It branches from the classical PHD formulation in the modelling phase already, so it is crucial to state first the different models that will be used.

A. Three fundamental point processes

A suitable model is essential for the performance of a filter since prior knowledge about the behaviour of the targets can help to enhance the interpretation of the observations. For that purpose, let us first define three basic distributions and their PGFLs, namely the Bernoulli, Poisson and Negative Binomial distributions. These will be crucial in the formulation of the new filter introduced in Sec. III-B.

1) **Bernoulli point processes**: A Bernoulli point process has only two possible outcomes: either there is an object with spatial distribution s or there is no object, where both events occur with the probabilities p and $(1-p)$, respectively. In this case, the general PGFL in (1) simplifies to the following expression:

$$G_{\text{Bernoulli}}(h) = (1 - p) + p \int_{\mathcal{X}} h(x) s(x) \mathrm{d}x. \tag{9}$$

In multi-target tracking, two different examples of such processes are found, namely the target survival and the target detection processes.

2) **Poisson point processes**: Poisson point processes describe populations of random objects whose cardinality is Poisson distributed. The objects are assumed to be independent and identically distributed (i.i.d.) according to some spatial distribution s and their number is Poisson distributed with parameter λ. The PGFL of the Poisson process is easily derived to be of the form

$$G_{\text{Poisson}}(h) = \exp\left(\int_{\mathcal{X}} [h(x) - 1] \mu(x) \mathrm{d}x \right) \tag{10}$$

where $\mu(x) = \lambda s(x)$ is the intensity of the process.
In the FISST framework, it is common to model the number of objects appearing in the Field of View (FoV) during one time step with a Poisson point process. The functional (10) has very convenient properties w.r.t. differentiation due to its exponential form. Furthermore, it has the important property that its mean and variance are equal.

3) **Negative Binomial point processes**: NB point processes describe populations of random objects whose cardinality is NB distributed. The objects are again i.i.d. according to some spatial distribution s. The PGFL of a NB process is derived from a Poisson process with Gamma distributed mean, resulting in the formulation

$$G_{\text{NB}}(h) = \left(1 + \frac{1}{\beta} \int_{\mathcal{X}} [1 - h(x)] s(x) \mathrm{d}x \right)^{-\alpha} \tag{11}$$

with strictly positive real values $\alpha, \beta \in \mathbb{R}_+$ [14]. In contrast to Poisson point processes, the variance of this process is always greater than the mean, however different ratios of the latter can be obtained using appropriate choices of α and β.

Note that the parameters of the NB distribution stand in correspondence with the mean and the variance via the following equations which can be derived straightforwardly from the PGFL (11):

$$\begin{aligned} \mu(B) &= \frac{\alpha}{\beta} \int_B s(x) dx, \\ \operatorname{var}(B) &= \mu(B) \left(1 + \frac{1}{\beta} \int_B s(x) dx \right). \end{aligned} \tag{12}$$

Furthermore, there is a close relationship between the Poisson and NB distributions; in fact, one can show that the Poisson distribution is the limiting case of a NB distribution [15]. This suggests experimentation with negative binomial assumptions in the modelling in order to lift the restrictiveness of the Poisson case. The fact that the variance of a NB point process is always greater than (or in the Poisson limit case, equal to) its mean might be beneficial for cases with highly fluctuating object number because different ratios of mean to variance can be studied.

B. Derivation of the PHD filter with NB clutter model

The PHD filter is a Bayesian recursion for multi-target state estimation which involves two stages of operation, a *prediction* and an *update* step. The prediction gives an estimate of the target state at a given time based on prior knowledge, whereas the update corrects this belief by incorporating additional information that might be given through sensor input. The PHD filter got its name from the fact that instead of the whole distribution, the first order moment is propagated which is also called the *intensity* or *probability hypothesis density* of the process.

Since false alarms only affect the update phase of the filtering framework, the prediction phase is the same as for the classical PHD filter [4]. In this paper, however, a measurement-driven birth model is considered for the practical implementation, resulting in a shortened prediction equation without spontaneous appearance of new targets [16].

The PHD update, on the other hand, is modelled based on the following assumptions:

Assumptions III.1.

(a) New observations are made independently.

(b) The predicted point process Φ is Poisson with intensity $\mu_\Phi(x) = \lambda_\Phi s_\Phi(x)$ where λ_Φ is the defining parameter and s_Φ the spatial distribution of the process.

(c) The detection process is Bernoulli distributed, i.e. an object x is detected with probability $p_{\mathrm{d}}(x)$ producing a measurement z with likelihood $l(z|x)$, whereas x remains undetected with probability $(1 - p_{\mathrm{d}}(x))$.

(d) The birth process is assumed to be Poisson of the form (10) *with intensity $\mu_{\mathrm{b}}(x) = \lambda_{\mathrm{b}} s_{\mathrm{b}}(x)$, where λ_{b} is the defining parameter and s_{b} the spatial distribution of the process which is dependent on the measurements.*

(e) The clutter process is a Negative Binomial distributed point process with spatial distribution s_{c} and positive parameters $\alpha, \beta \in \mathbb{R}_+$.

These assumptions lead to the main result of this article, the update equation of the PHD filter with NB clutter model. For that, let us define the *Pochhammer symbol* or *falling factorial* [17] of x by n with

$$(x)_n = x(x-1)\cdots(x-n+1), \quad (x)_0 = 1. \tag{13}$$

Theorem III.2. *Let Z be a set of measurements provided by a sensor at a given time, and let Ξ denote the measurement process. Under the assumptions III.1 listed above, the updated first-order moment density with Poisson prediction and negative binomial false alarm model is found to be*

$$\mu_{\Phi|\Xi}(x|Z) = \mu_\Phi^\phi(x) + \sum_{z\in Z} \frac{\mu_\Phi^z(x)}{s_{\mathrm{c}}(z)} \frac{Y(Z\backslash\{z\})}{Y(Z)} \tag{14}$$

with missed detection term

$$\mu_\Phi^\phi(x) = (1 - p_{\mathrm{d}}(x))\mu_\Phi(x) \tag{15}$$

and association term

$$\mu_\Phi^z(x) = p_{\mathrm{d}}(x) l(z|x) \mu_\Phi(x). \tag{16}$$

For a given measurement set Z of cardinality $|Z|$, the terms $Y(Z)$ are defined as

$$Y(Z) = \sum_{k=0}^{|Z|} \frac{(\alpha+k-1)_k}{(\beta+1)^k} \; e_{|Z|-k}(Z) \tag{17}$$

with

$$e_k(Z) = \sum_{\substack{\tilde{Z}\subseteq Z \\ |\tilde{Z}|=k}} \prod_{z\in\tilde{Z}} \frac{\int_{\mathcal{X}} \mu_\Phi^z(x)\mathrm{d}x + \mu_{\mathrm{b}}(z)}{s_{\mathrm{c}}(z)}. \tag{18}$$

Proof sketch. The proof of Thm. III.2 follows the methodology of [4] for the derivation of the PHD filter. It will can divided in three fundamental steps that are obtained using in the Lemmas VI.5, VI.6 and VI.7 in the appendix.

1) First of all, the joint PGFL $G_{\Xi,\Phi}$ of the target and measurement processes Φ and Ξ has to be stated explicitly using the modelling assumptions III.1.
2) With the help of this, the conditional PGFL $G_{\Phi|\Xi}$ has to be derived using Bayes rule which leads to the formulation

$$G_{\Phi|\Xi}(h|Z_m) = \frac{\delta^m G_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}{\delta^m G_{\Xi,\Phi}(g,1;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}} \tag{19}$$

where $Z_m = \{z_1, \ldots, z_m\}$ is a measurement set of cardinality m consisting of the observations z_i [4].

3) Lastly, the first-order moment of (19) will be calculated which gives the final result. □

As mentioned earlier, the Poisson distribution can be expressed as a special case of the NB distribution [15]. Although the variance of NB processes is always strictly greater than the mean, the Poisson limit case leads to a ratio of 1:1. In other words, the parametrisation of NB processes can simulate all possible ratios between mean and variance given that the variance is not smaller than the mean. Let us state the following corollary of Thm. III.2.

Corollary III.3. *Under the assumptions III.1, let the ratio $\mu_{\mathrm{c}} = \frac{\alpha}{\beta}$ be an arbitrary but fixed number. Then, by taking the*

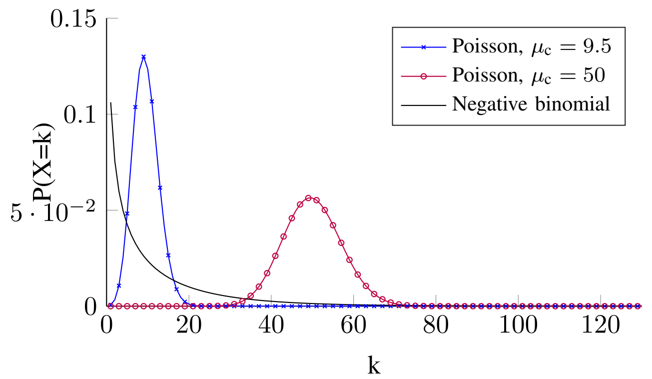

Fig. 1: The NB and the Poisson distributions in comparison. Note that high variance of the NB distribution allows for relatively high probabilities over a large range whereas the Poisson distribution is concentrated around its respective mean.

limit $\alpha \to \infty$, Eq. (14) *reduces to the update equation of the original PHD filter.*

Proof. Cf. VI-C. □

IV. Experiments

Two different experiments were conducted to analyse the behaviour of the new filter in the presence of a highly fluctuating number of false alarms. For all experiments with the NB based filter, the mean false alarm number is assumed to be 9.5, and the chosen values for the NB distribution are set accordingly to be $\alpha = 0.5$ and $\beta = 0.0526$, leading to a variance of 190.0. The original PHD filter was first run with the same mean number of false alarm, i.e. 9.5, and then with a larger value of 50 accounting for the high variability in the number of false alarms.

The corresponding distributions are shown in Fig. 1.

A. Simulation of data

The FoV was assumed to be a square image frame of size 50 m $\times$ 50 m. Between 0 and 5 targets were initialised independently and identically distributed over the FoV. The survival probability was set to 0.99 and the initial velocity of the targets was normally distributed with mean 0 m and standard deviation 0.5 m in both image dimensions. Across 100 time steps, the existing targets were propagated using a nearly constant velocity model, assuming small white acceleration noise with standard deviation 0.01 m/s^2 on each dimension. Additional white noise on the target states with a standard deviation of 0.5 m simulated measurement noise. In each frame, a Poisson distributed number of new-born targets (with mean 0.5) was initiated randomly in the FoV. The detection process was simulated by occasionally discarding measurements, using a detection probability of 0.9. Furthermore, the following two alternatives for the clutter model were simulated:

- **S.1** The first scenario is created using exactly 9 false alarms in each time step until time step 15 where the number of false alarms suddenly changes to different values ranging from 0 to 130. This scenario examines the change in the estimated number of targets of both filters under the influence of various amounts of clutter.
- **S.2** The second scenario was created using a NB distribution for the false alarm number with the parameters as listed above, simulating 100 time steps for each run. This scenario helps to compare the overall behaviour of the new filter in comparison to the original PHD filter in presence of a highly fluctuating number of false alarms.

B. Estimation results

For all experiments, a Gaussian Mixture (GM) implementation was chosen for both methods in the manner of [18]. Furthermore, the parametrisation was chosen equally for both PHD versions to make them comparable.

For the underlying Kalman filtering, the dynamics noise for the prediction was set to 0.1 m and the standard deviation of the measurement noise was set to 0.4 m. The survival and detection probabilities were set to the true values, and the mean number μ_b of newborn targets per time step was chosen to be 10 in the initial step and 3 afterwards. Both PHD iterations contained a merging and a pruning step, where two components were merged with a Hellinger distance below 0.8 and a component was discarded with a weight below 10^{-7}.

Furthermore, the parametrisation of the clutter model was used as stated above, where the proposed method assumed a clutter intensity of 9.5 clutter points at all times whereas the PHD with Poisson clutter model was tested with the two different intensities 9.5 and 50 false alarms, respectively.

The two algorithms differ only in the modelling of the number of false alarms and not in their spatial distribution, thus it can be expected that they yield similar performances w.r.t. the estimation of the target states whereas they differ greatly in the estimation of the expected target number. Therefore, the conducted experiments are analysed based on the cardinality error obtained through the comparison of the estimation results with the ground truth.

Fig. 2 shows the cardinality errors of scenario **S.1** in the time step where the clutter number changes, plotted against the true number of clutter points at that time. The presented results are averaged over 100 Monte Carlo runs per final false alarm number; the graphs indicate both the mean cardinality error and the corresponding variance.

Regarding the reactivity on the amount of clutter, it can be noticed that the original PHD filter performs slightly better than the proposed method for clutter values in a sigma range around the respective assumed Poisson clutter intensity where the NB distribution is much lower (cf. Fig. 1). For values outside that range, however, the NB based algorithm outperforms the original method thanks to its large variance, especially for numbers of clutter points that are far away from the intensity expected by the Poisson distribution.

In order to analyse the performance the proposed method on a more realistic scenario, the algorithms were further applied to 500 randomly generated instances of scenario **S.2**. Fig. 3

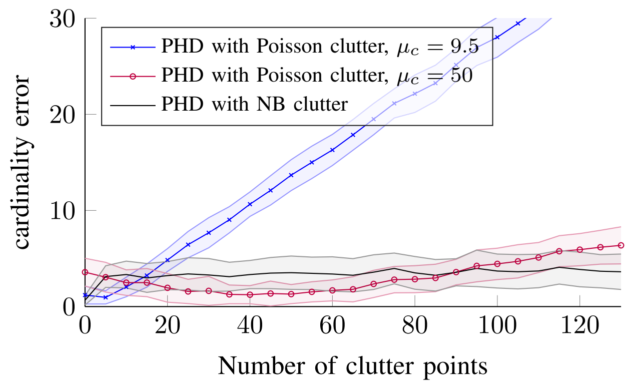

Fig. 2: Simulation 2: Estimation error of both algorithms at time 15 dependent on the present number of clutter points. The results are averaged over 100 Monte Carlo runs for each false alarm number and plotted along with the variance over those runs.

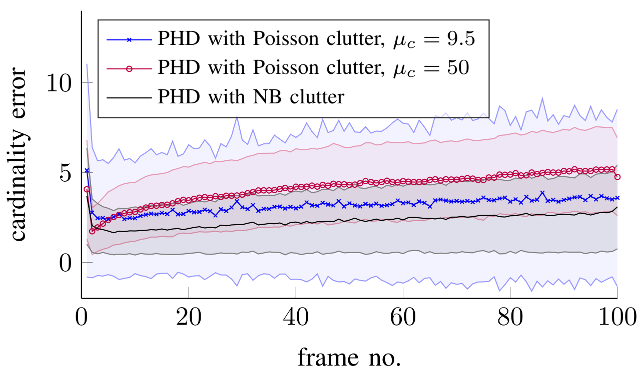

Fig. 3: Mean and standard deviation over 500 Monte Carlo runs of the cardinality errors obtained with both algorithms.

shows the mean cardinality error over time together with the variance over all runs.

For both assumed intensities, the original PHD filter produces higher average cardinality errors than the NB based filter, and an increased intensity actually penalises the overall performance since low numbers of clutter points are not modelled well in that case. Furthermore, its performance varies greatly across the runs for a modelled mean false alarm number of 9.5, and it has a more stable behaviour for a value of 50, even though the false alarms are generated with a NB distribution with mean 9.5. In contrast, the proposed filter demonstrates its robustness to high variations in the false alarm number since the Poisson assumption is insufficient to model the variability in the data.

Finally, the increase of complexity induced by the proposed method was analysed by comparing the runtime of both algorithms. All experiments were conducted on a dual-core Dell Precision M4800 workstation with Intel(R) Core(TM) i7-4710MQ CPU @ 2.50GHz using a Matlab implementation. For one run of **S.2**, the proposed filter takes 8.02959 s averaged over 100 runs, whereas the original PHD filter needs 6.7237 s. This slight increase appears reasonable for scenarios that benefit greatly from an improved clutter model. Note that the combinatorial summation in (18) can be easily computed using Vieta's theorem, avoiding heavy computational cost [19].

V. Conclusion

A new version of the PHD filter is introduced and derived in this article which assumes a negative binomial distribution of the false alarm number. The algorithm leads to an additional factor in the original formulation of the filter which makes it easy to realise on top of an existing implementation. Furthermore, the algorithm is tested on two different sets of simulated data that examine its reactivity to one particular measurement spike of various heights, as well as its general performance on a scenario with Negative Binomial distributed clutter. In both cases, the results of the proposed method were contrasted against the performance of the original PHD filter with two different clutter intensities. The experiments demonstrate that the alternative PHD version provides a much more stable estimation of the true target number than the original filter in the presence of spontaneous bursts of clutter. It is further proved that the proposed formulation converges to the original Poisson case by adjusting the parameters, so it is possible to customise the filter through its parametrisation in order to choose how much credibility will be given to incoming measurements.

VI. Appendix

The appendix collects all mathematical details that are needed for the result shown in Thm. III.2. The first part demonstrates some general differentiation rules for exponential and negatively exponentiated functionals. In the second part, three lemmas will be stated that supplement the proof sketch provided in Sec. III-B. The last part demonstrates that the original PHD filter is a limit case of the introduced method.

A. Preliminary results

Since the two PGFLs (10) and (11) are of special interest in this article, let us determine the nth order derivatives of the compositions $\exp(G(h))$ and $(G(h))^{-\alpha}$ for linear functionals G which are stated in Prop. VI.2 and VI.4.

Lemma VI.1. *The first-order derivative of the exponential functional* $\exp$ *in composition with a general functional* $G(h)$ *is the functional chain differential*

$$\delta(\exp \circ G)(h;\eta) = \exp(G(h))\delta G(h;\eta). \tag{20}$$

Proof. Cf. [10]. $\square$

Proposition VI.2. *Let* G *be a linear functional. Then the* n*th-order derivative of the composition* $\exp(G(h))$ *is found to be*

$$\delta^n(\exp \circ G)(h;\eta_1,\ldots,\eta_n) = \exp(G(h))\prod_{i=1}^{n}\delta G(h;\eta_i). \tag{21}$$

Proof. The nth-order derivative of $\exp(G(h))$ can be easily seen from Faà di Bruno's formula (8): since $\delta^{(2)}G = 0$ due to the linearity of G, the only partition that leads to a non-zero term is the set of singletons such that one factor $\delta G(h;\eta_i)$ is drawn from the exponential term for all directions $\eta_1,\ldots,\eta_n$. $\square$

The PGFL of the negative binomial distribution involves a power $-\alpha$ where $\alpha\in\mathbb{R}_+$ is a positive real number. It will be shown in the following that the functional composition with a power yields a similarly multiplicative result as seen above.

Lemma VI.3. *The first-order derivative of the power functional* $(.)^{-\alpha}$ *in composition with a general functional* $G(h)$ *is the functional*

$$\delta(G^{-\alpha})(h;\eta)=(-\alpha)G(h)^{-\alpha-1}\delta G(h;\eta). \tag{22}$$

Proof. For the proof of the lemma, the binomial theorem for general exponents $\alpha\in\mathbb{R}$ is used:

$$(x+y)^{\alpha}=\sum_{k\geq 0}\frac{(\alpha)_k}{k!}\,x^{\alpha-k}y^{k}. \tag{23}$$

with the truncated factorial $(\alpha)_k$ defined in (13).

In analogy to Lem. VI.1, consider a series of functions $(\eta_n)_{n\in\mathbb{N}}$ converging pointwise to $\delta G(h;\eta)$ for $n\to\infty$ and a series $(\varepsilon_n)_{n\in\mathbb{N}}$ converging to 0. Therefore,

$$\begin{aligned}
&\delta(G^{-\alpha})(h;\eta)\\
&\overset{(6)}{=}\delta\left((.)^{-\alpha}\right)(G(h);\delta G(h;\eta))\\
&=\lim_{n\to\infty}\frac{1}{\varepsilon_n}[(G(h)+\varepsilon_n\eta_n)^{-\alpha}-(G(h))^{-\alpha}]\\
&\overset{(23)}{=}\lim_{n\to\infty}\frac{1}{\varepsilon_n}\left[\sum_{k\geq 0}\frac{(-\alpha)_k}{k!}G(h)^{-\alpha-k}(\varepsilon_n\eta_n)^k-G(h)^{-\alpha}\right]\\
&=\lim_{n\to\infty}\frac{1}{\varepsilon_n}\left[\sum_{k\geq 1}\frac{(-\alpha)_k}{k!}G(h)^{-\alpha-k}(\varepsilon_n\eta_n)^k\right]\\
&=\lim_{n\to\infty}\Bigg[(-\alpha)G(h)^{-\alpha-1}\eta_n\\
&\qquad\qquad+\underbrace{\varepsilon_n\sum_{k\geq 2}\frac{(-\alpha)_k}{k!}G(h)^{-\alpha-k}\varepsilon_n^{k-2}\eta_n^k}_{=0(n\to\infty)}\Bigg]\\
&=(-\alpha)G(h)^{-\alpha-1}\delta G(h;\eta).
\end{aligned}$$

$\square$

Proposition VI.4. *For any linear functional* G*, the* n*th-order derivative of the composition* $G(h)^{-\alpha}$ *can be expressed as*

$$\begin{aligned}
&\delta^n(G^{-\alpha})(h;\eta_1,\ldots,\eta_n)\\
&\quad=(-1)^n(\alpha+n-1)_n\;G(h)^{-\alpha-n}\prod_{i=1}^{n}\delta G(h;\eta_i).
\end{aligned} \tag{24}$$

Proof. Thanks to the structure of (22), the nth-order differential of $G^{-\alpha}$ can again be obtained inductively in analogy to Prop. VI.2 if G is linear. The derivative of the outer function creates a factor $(-\alpha-i+1)$ for each differentiation step. $\square$

B. Derivation of the filter

The following proof follows the sketch given in Sec. III. It is divided into three main lemmas, VI.5, VI.6 and VI.7.

Lemma VI.5. *With assumptions III.1, the joint target and measurement process leads to a joint PGFL of the form*

$$\begin{aligned}
G_{\Xi,\Phi}(g,h)&=G_{\Phi}\left(G_{\mathrm{d}}(g,h)\right)G_{\mathrm{b}}(g)G_{\mathrm{c}}(g)\\
&=\exp\Big(F_1(g,h)\Big)\Big(F_2(g)\Big)^{-\alpha},
\end{aligned} \tag{25}$$

where the functionals F_1 *and* F_2 *are defined as*

$$\begin{aligned}
F_1(g,h)&=\int_{\mathcal{X}}\big[h(x)f(x)-1\big]\mu_{\Phi}(x)\mathrm{d}x\\
&\quad+\int_{\mathcal{Z}}(g(z)-1)\mu_{\mathrm{b}}(z)\mathrm{d}z,
\end{aligned} \tag{26}$$

$$f(x)=1-p_{\mathrm{d}}(x)+p_{\mathrm{d}}(x)\int_{\mathcal{Z}}g(z)l(z|x)\mathrm{d}z$$

and

$$F_2(g)=1+\frac{1}{\beta}\int_{\mathcal{Z}}(1-g(z))s_{\mathrm{c}}(z)\mathrm{d}z. \tag{27}$$

Proof. The structure in (25) is obtained as follows. Since the predicted, the birth and the clutter processes are assumed to be independent, their PGFLs are simply multiplied to describe their superposition. Furthermore, the detection process branches from the predicted process which leads to the nested structure of their two PGFLs. By taking the assumptions III.1 into account, one easily verifies the functionals F_1 and F_2 by combining the two Poisson processes to a sum in the exponential term. $\square$

Note that both F_1 and F_2 are linear in g and F_1 is linear in h, such that higher-order derivatives become zero. For the sake of readability throughout following steps, let us define

$$\begin{aligned}
F_3(h|z)&:=\delta F_1(g,h;\delta_z)\\
&=\int_{\mathcal{X}}h(x)p_{\mathrm{d}}(x)l(z|x)\mu_{\Phi}(x)\mathrm{d}x+\mu_{\mathrm{b}}(z)
\end{aligned} \tag{28}$$

and note that $\delta F_2(g;\delta_z)=-\frac{1}{\beta}s_{\mathrm{c}}(z)$.

In order to write down expression (19), let us first find the mth order derivative of (25) w.r.t. g.

Lemma VI.6. *The* m*th order derivative of* (25) *w.r.t.* g *can be written as*

$$\begin{aligned}
&\delta^m G_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m})\\
&\quad=\exp(F_1(g,h))\sum_{k=0}^{m}\frac{(\alpha-k+1)_k}{\beta^k}F_2(g)^{-\alpha-k}\\
&\qquad\cdot\sum_{\substack{Z\subseteq Z_m\\|Z|=k}}\left(\prod_{z\in Z}s_{\mathrm{c}}(z)\prod_{z'\in Z_m\setminus Z}F_3(h|z')\right).
\end{aligned} \tag{29}$$

Proof. Since $G_{\Xi,\Phi}(g,h)=\exp\Big(F_1(g,h)\Big)\Big(F_2(g)\Big)^{-\alpha}$ is a product of two functionals dependent on g, the general higher-order product rule (7) can be applied. First of all, let us define $G_1(g,h)=\exp\Big(F_1(g,h)\Big)$ and $G_2(g)=\Big(F_2(g)\Big)^{-\alpha}$ for the sake of compactness. For the first term G_1, it follows for $\omega \subseteq \{1, \ldots, m\}$ with $|\omega| = m - k$ (and the corresponding set $Z \subseteq Z_m$ of observations indexed by ω) using (8) and (21) that

$$\begin{aligned}
&\delta^{m-k} G_1(g, h; (\eta_i)_{i\in\omega}) \\
&\overset{(8)}{=} \sum_{\pi\in\Pi(\omega)} \delta^{(|\pi|)} \exp\Bigg(F_1(g,h); \\
&\qquad \Big(\underbrace{\delta^{(|\omega'|)} F_1\big(g,h;(\eta_j)_{j\in\omega'}\big)}_{=0 \text{ for } |\omega'|>1} \Big)_{\omega'\in\pi} \Bigg) \\
&= \delta^{m-k} \exp\Bigg(F_1(g,h); \Big(\delta F_1(g,h;\eta_i)\Big)_{i\in\omega} \Bigg) \\
&\overset{(21)}{=} \exp(F_1(g,h)) \prod_{z\in Z} \delta F_1(g,h;\delta_z) \\
&= \exp(F_1(g,h)) \prod_{z\in Z} F_3(h|z).
\end{aligned} \tag{30}$$

where $\Pi(\omega)$ denotes the partition set of ω. On the other hand, the second term G_2 has the higher-order derivative

$$\begin{aligned}
&\delta^{k} G_2(g; (\eta_j)_{j\in\bar\omega}) \\
&\overset{(8)}{=} \sum_{\pi\in\Pi(\bar\omega)} \delta^{(|\pi|)} \Bigg(F_2(g)^{-\alpha}; \\
&\qquad \Big(\underbrace{\delta^{(|\omega'|)} F_2\big(g;(\eta_k)_{k\in\omega'}\big)}_{=0 \text{ for } |\omega'|>1} \Big)_{\omega'\in\pi} \Bigg) \\
&= \delta^{k} \Bigg(F_2(g)^{-\alpha}; \Big(\delta F_2(g;\eta_j)\Big)_{j\in\bar\omega} \Bigg) \\
&\overset{(24)}{=} \frac{(\alpha+k-1)_k}{\beta^k} F_2(g)^{-\alpha-k} \prod_{z\in Z} s_{\mathrm{c}}(z).
\end{aligned} \tag{31}$$

with the complement $\bar\omega = \{1, \ldots, m\} \backslash \omega$ and corresponding measurement set $\bar Z \subseteq Z_m \setminus Z$ indexed by $\bar\omega$. These two results (30) and (31) used in the product rule (7) give the desired result. ☐

Since the mean of a PGFL is its first-order moment at $h = 1$ as stated in (2), the PHD of (19) is found via

$$\mu_{\Phi|\Xi}(x|Z_m) = \frac{\delta^{m+1} G_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m},\delta_x)|_{g=0,h=1}}{\delta^m G_{\Xi,\Phi}(g,1;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}. \tag{32}$$

The derivative in the denominator of Eq. (32) was already determined in Lem. VI.6, so one is left with another differentiation of (29) w.r.t. h which leads to the derivative in the numerator of expression (32) as follows:

Lemma VI.7. *The first-order derivative of* (29) *w.r.t. h is found to be*

$$\begin{aligned}
&\delta^{m+1} G_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m},\delta_x) \\
&\quad= \exp(F_1(g,h)) \sum_{j=0}^{m} \frac{(\alpha+k-1)_k}{\beta^k} F_2(g)^{-\alpha-k} \\
&\quad\cdot \Bigg[\sum_{\substack{Z\subseteq Z_m\\|Z|=k}} F_4(g) \prod_{z\in Z} s_{\mathrm{c}}(z) \prod_{z'\in Z_m\setminus Z} F_3(h|z') \\
&\quad+ \prod_{z\in Z} s_{\mathrm{c}}(z) \Bigg(\sum_{z'\in Z_m\setminus Z} \mu_\Phi^{z'}(x) \prod_{\substack{z''\in Z_m\setminus Z\\ z''\neq z'}} F_3(h|z'') \Bigg) \Bigg].
\end{aligned} \tag{33}$$

The functions $F_4(g)$ and $\mu_\Phi^z(B)$ are defined as

$$\begin{aligned}
F_4(g) &= \delta F_1(g,h;\delta_x) \\
&= \left(1 - p_{\mathrm{d}}(x) + p_{\mathrm{d}}(x)\int_{\mathcal{Z}} g(z) l(z|x)\mathrm{d}z\right) \mu_\Phi(x)
\end{aligned} \tag{34}$$

and

$$\mu_\Phi^z(x) = p_{\mathrm{d}}(x) l(z|x) \mu_\Phi(x). \tag{35}$$

Proof. Use the product rule (5). ☐

Note that the first term of the r.h.s. of (33) corresponds to the missed detection term, whereas the remaining expression deals with the associations of x with the measurements z_j.

In the end, the result Thm. III.2 is obtained by setting $h = 1$ and $g = 0$ and writing the last two results (29) and (33) in the form (32).

Proof of Thm. III.2. To simplify the explicit form of fraction (32), note that the term $\exp(F_1(g,h))$ cancels out immediately such that one is left with a sum of two big fractions coming from (33). Setting $g = 0$ eradicates the term $p_{\mathrm{d}}(x)\int_{\mathcal{Z}} g(z)l(z|x)\mathrm{d}z$, furthermore a factor $F_2(0)^{-\alpha}$ can be removed in all terms. Since $F_2(0) = (1 + \frac{1}{\beta})$, it simplifies with the fraction in (17). To eliminate the product $\prod s_{\mathrm{c}}(z)$ in each term, it is helpful to expand the fraction with $\frac{1}{\prod_{z\in Z_m} s_{\mathrm{c}}(z)}$ which leads to the functions $e_k(Z_m)$ as defined above. ☐

C. The original PHD filter as a limit of the PHD filter with NB clutter

It is possible to show that one can recover the classic formulation of *the PHD filter as a limit case of the PHD filter with negative binomial clutter* by fixing the ratio $\mu_{\mathrm{c}} = \frac{\alpha}{\beta}$ and taking the limit $\alpha \to \infty$. The only term in (14) which contains

α or β is $Y(Z)$, thus let us have a look at this term in detail first:

$$\begin{aligned}
&\lim_{\alpha\to\infty} Y(Z_m) \\
&= \lim_{\alpha\to\infty} \sum_{k=0}^{m} \tfrac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{(\beta+1)^k}\ e_{m-k}(Z) \\
&= \lim_{\alpha\to\infty} \sum_{k=0}^{m} \frac{(\alpha)_k}{(\frac{\alpha+\mu_c}{\mu_c})^k}\ e_{m-k}(Z) \\
&= \lim_{\alpha\to\infty} \sum_{k=0}^{m} \mu_c^k \left[\tfrac{\alpha}{\alpha+\mu_c}\tfrac{\alpha+1}{\alpha+\mu_c}\cdots\tfrac{\alpha+k-1}{\alpha+\mu_c}\right]\ e_{m-k}(Z) \\
&\overset{(\sharp)}{=} \sum_{k=0}^{m} \mu_c^k\ e_{m-k}(Z)
\end{aligned} \tag{36}$$

where l'Hôspital's rule [20] was used in ($\sharp$). For the fraction $\frac{Y(Z_m\setminus\{z\})}{Y(Z_m)}$, the limit becomes

$$\begin{aligned}
&\lim_{\alpha\to\infty} \frac{Y(Z_m\setminus\{z\})}{Y(Z_m)} \\
&= \frac{\sum_{k=0}^{m-1}\mu_c^k\ e_{m-1-k}(Z_m\setminus\{z\})}{\sum_{k=0}^{m}\mu_c^k\ e_{m-k}(Z_m)} \\
&= \frac{\sum_{k=0}^{m-1}\mu_c^k\ e_{m-1-k}(Z_m\setminus\{z\})}{\sum_{k=0}^{m}\mu_c^k\ e_{m-k}(Z_m)} \cdot \frac{\frac{1}{\mu_c^m}}{\frac{1}{\mu_c^m}} \\
&= \frac{1}{\mu_c}\frac{\sum_{Z\subseteq Z_m\setminus\{z\}}\prod_{z'\in Z}\dfrac{F_3(1|z')}{\mu_c s_c(z')}}{\sum_{Z\subseteq Z_m}\prod_{z'\in Z}\dfrac{F_3(1|z')}{\mu_c s_c(z')}} \\
&\overset{(*)}{=} \frac{1}{\mu_c}\frac{\sum_{Z\subseteq Z_m\setminus\{z\}}\prod_{z'\in Z}\dfrac{F_3(1|z')}{\mu_c s_c(z')}}{\sum_{Z\subseteq Z_m\setminus\{z\}}\prod_{z'\in Z}\dfrac{F_3(1|z')}{\mu_c s_c(z')}\left(1+\dfrac{F_3(1|z)}{\mu_c s_c(z)}\right)} \\
&= \frac{s_c(z)}{F_3(1|z)+\mu_c s_c(z)}.
\end{aligned} \tag{37}$$

Equation ($*$) was obtained by rearranging the sum in the denominator to separate the terms with $Z \ni z$. By inserting the result of (37) into (14), we obtain

$$\begin{aligned}
&\lim_{\alpha\to\infty}\mu_{\Phi|\Xi}(x|Z_m) \\
&\quad= \mu_\Phi^\phi(x) + \sum_{z\in Z_m}\frac{\mu_\Phi^z(x)}{s_c(z)}\left(\frac{s_c(z)}{F_3(1|z)+\mu_c s_c(z)}\right) \\
&\quad= \mu_\Phi^\phi(x) + \sum_{z\in Z_m}\frac{\mu_\Phi^z(x)}{F_3(1|z)+\mu_c s_c(z)}
\end{aligned} \tag{38}$$

which is the update of the PHD filter [4].

ACKNOWLEDGEMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), Grant number EP/K014277/1 and the MOD University Defence Research Collaboration in Signal Processing. Isabel Schlangen has a PhD studentship funded by the Edinburgh Super-Resolution Imaging Consortium (ESRIC), Grant number MR/K01563X/1.

REFERENCES

[1] S. Watts, "Radar sea clutter: Recent progress and future challenges," in *Radar, 2008 International Conference on*, Sept 2008, pp. 10–16.

[2] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[3] S. T. Hess, T. P. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophysical Journal*, vol. 91, no. 11, pp. 4258 – 4272, 2006.

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

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

[6] X. Chen, R. Tharmarasa, M. Pelletier, and T. Kirubarajan, "Integrated clutter estimation and target tracking using poisson point processes," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1210–1235, APRIL 2012.

[7] R. Mahler and B.-T. Vo, "An improved CPHD filter for unknown clutter backgrounds," pp. 90 910B–90 910B–12, 2014.

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

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

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

[11] P. Bernhard, "Chain differentials with an application to the mathematical fear operator," *Nonlinear Analysis*, vol. 62, p. 1225?1233, 2005.

[12] D. Clark and R. Mahler, "Generalized phd filters via a general chain rule," in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 157–164.

[13] E. Delande, M. Uney, J. Houssineau, and D. Clark, "Regional variance for multi-object filtering," *Signal Processing, IEEE Transactions on*, vol. 62, no. 13, pp. 3415–3428, 2014.

[14] 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.

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

[16] B. Ristic, D. E. Clark, B.-N. Vo, and B.-T. Vo, "Adaptive Target Birth Intensity for PHD and CPHD Filters," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 2, p. 1656 1668, Apr. 2012.

[17] P. J. Olver, *Classical Invariant Theory*. Cambridge University Press, 1999, cambridge Books Online.

[18] 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.

[19] B. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3553 – 3567, 2007.

[20] F. Lettenmeyer, "Über die sogenannte Hospitalsche Regel." *Journal für die reine und angewandte Mathematik*, vol. 174, pp. 246–247, 1936.