Multi-object filtering with second-order moment statistics

Isabel Christiane Schlangen

A thesis submitted for the degree of

Doctor of Philosophy

Heriot-Watt University, Edinburgh
School of Engineering and Physical Sciences (EPS)

· September 15, 2017 ·

Abstract

The focus of this work lies on multi-object estimation techniques, in particular the Probability Hypothesis Density (PHD) filter and its variations. The PHD filter is a recursive, closed-form estimation technique which tracks a population of objects as a group, hence avoiding the combinatorics of data association and therefore yielding a powerful alternative to methods like Multi-Hypothesis Tracking (MHT). Its relatively low computational complexity stems from strong modelling assumptions which have been relaxed in the Cardinalized PHD (CPHD) filter to gain more flexibility, but at a much higher computational cost. We are concerned with the development of two suitable alternatives which give a compromise between the simplicity and elegance of the PHD filter and the versatility of the CPHD filter. The first alternative generalises the clutter model of the PHD filter, leading to more accurate estimation results in the presence of highly variable numbers of false alarms; the second alternative provides a closed-form recursion of a second-order PHD filter that propagates variance information along with the target intensity, thus providing more information than the PHD filter while keeping a much lower computational complexity than the CPHD filter. The discussed filters are applied on simulated data, furthermore their practicality is demonstrated on live-cell super-resolution microscopy images to provide powerful techniques for molecule and cell tracking, stage drift estimation and estimation of background noise.

Acknowledgements

First of all, I would like to thank my supervisor Daniel Clark for letting me be part of his amazing research team, and my second supervisor Colin Rickman for always being there for me in times of need. This project would certainly not have flourished so well without the wonderful Manu Delande and Jérémie Houssineau, your unconditional support and your friendship are very dear to me. I am in deep gratitude to the Institute of Biological Chemistry, Biophysics and Bioengineering (IB3) at Heriot-Watt University and to my sponsors, the Edinburgh Super-Resolution Imaging Consortium (ESRIC), for giving me the best opportunities to thrive as a young researcher and scientist, and to the postgraduate society PoBBBs for all the lovely Friday cakes, flat parties and company. A special thank you to Moira Ewing who helped me regain my balance, thanks to you my work felt comfortable like a deep inhalation and the end of my project like a long, slow, smooth exhalation. Mariia Dmitrieva, Miguel Hermida, Dirk-Jan Cornelissen and Corina Barbalata, thank you all for sharing so many wonderful moments with me and for making the hard times seem easier, and thanks to Flávio Eler, Vibhav Bharti, David Cormack, Alexey Narykov, Mark Campbell, Elizabeth Vargas, Salvatore Caporale and so many others in and around EM 3.31 for making me look forward to come to the lab every day. Certainly, there is not enough space on this page to mention all the other lovely people that I met over the last three years on campus, in the Chaplaincy and the Edinburgh Festival Chorus, but be assured that I am lucky to have you all in my life.

Contents

List of Tables

List of Figures

List of Acronyms

PHD Probability Hypothesis Density

SO-PHD Second-Order PHD

CPHD Cardinalized PHD

MHT Multi-Hypothesis Tracking

JPDA Joint Probabilistic Data Association

HISP Hypothesised filter for Independent Stochastic Populations

SLAM Simultaneous Localisation and Mapping

EKF Extended Kalman Filter

UKF Unscented Kalman Filter

GM Gaussian Mixture

SMC Sequential Monte Carlo

n.c.v. near constant velocity

i.i.d. independent and identically distributed

pmf probability mass function

PGFL Probability Generating Functional

FISST Finite Set Statistics

RFS Random Finite Set

PALM Photo-Activated Localisation Microscopy

sptPALM single particle tracking PALM

STORM Stochastic Optical Reconstruction Microscopy

SIM Structured Illumination Microscopy

STED Stimulated Emission Depletion

TIRF Total Internal Reflection Fluorescence

psf point spread function

IUWT Isotropic Undecimated Wavelet Transform

SNR signal-to-noise ratio

RMSE Root Mean Square Error

MC Monte Carlo

MAP Maximum A Posteriori

List of Publications

1. Schlangen, I., Houssineau, J., and Clark, D. (2014). A novel approach to image calibration in super-resolution microscopy. In *Control, Automation and Information Sciences (ICCAIS), 2014 International Conference on*, pages 111–116. IEEE

2. McKenna, I., Tonolini, F., Tobin, R., Houssineau, J., Bridle, H., McDougall, C., Schlangen, I., McGrath, J. S., Jimenez, M., and Clark, D. E. (2015). Observing the dynamics of waterborne pathogens for assessing the level of contamination. In *Sensor Signal Processing for Defence (SSPD), 2015*, pages 1–5. IEEE

3. Schlangen, I., Delande, E., Houssineau, J., and Clark, D. E. (2016a). A PHD filter with negative binomial clutter. In *Information Fusion (FUSION), 2016 19th International Conference on*, pages 658–665. IEEE

4. Schlangen, I., Franco, J., Houssineau, J., Pitkeathly, W. T. E., Clark, D. E., Smal, I., and Rickman, C. (2016b). Marker-less stage drift correction in super-resolution microscopy using the single-cluster PHD filter. *Selected Topics in Signal Processing, IEEE Journal of*, 10(1):193–202

5. Hagen, O., Houssineau, J., Schlangen, I., Delande, E. D., Franco, J., and Clark, D. E. (2016). Joint estimation of telescope drift and space object tracking. In *Aerospace Conference, 2016 IEEE*, pages 1–10. IEEE

6. Schlangen, I., Delande, E. D., Houssineau, J., and Clark, D. E. (2017c). A second-order PHD filter with mean and variance in target number. *Accepted to: IEEE Transactions on Signal Processing.* arXiv:1704.02084

7. Schlangen, I., Bharti, V., Delande, E., and Clark, D. E. (2017a). Joint multi-object and clutter rate estimation with the single-cluster PHD filter. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 1087–1091. IEEE

8. Schlangen, I., Clark, D. E., and Delande, E. D. (2017b). Single-cluster PHD filter methods for joint multi-object filtering and parameter estimation. *Submitted to: IEEE Transactions on Signal Processing.* arXiv:1705.05312

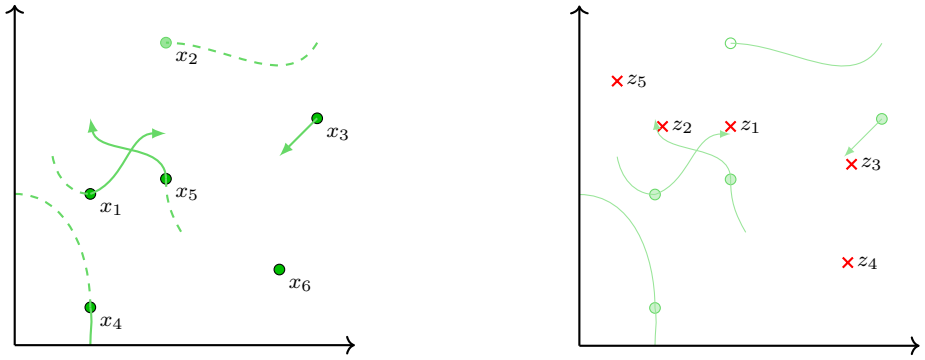
Introduction

Multi-object estimation on challenging scenarios

Multi-object estimation and tracking is an active, thriving field of research in the context of signal processing which has manifold applications in different areas of science and engineering. It was first inspired by surveillance tasks for both military and civil applications, mainly to track missiles or vehicles by land, sea and air using radar or sonar systems. Recent emerging fields such as robotics and computer vision benefit greatly from the findings of the target tracking community: a few examples are self-localisation of robotic vehicles with respect to their environment (also referred to as Simultaneous Localisation and Mapping (SLAM), see e.g. Thrun and Leonard [2008]), surveillance of public spaces for security or commercial applications [Andriluka et al., 2008; Urtasun et al., 2006], path planning and navigation of autonomous underwater vehicles [Clark et al., 2005] or even tracking of dolphin chirps [Gruden and White, 2016]. Also in biomedical applications, automated methods for identifying and tracking specimens of interest have become of increasing importance since they avoid tedious hand-labelling and linking in the presence of typically hundreds of objects in a dataset [Wood et al., 2012; McKenna et al., 2015; Wilson et al., 2016; Meijering et al., 2012].

Unfortunately, any type of sensor system produces imperfect data which raises several issues for all multi-object estimation techniques alike. First of all, information is usually binned into resolution cells which is typical for radar but also for digital images whose resolution cells are pixels; therefore, the precision in the localisation of an object is limited by the size of the resolution cells. Apart from techniques like track-before-detect [Boers et al., 2006; Rutten et al., 2005] or active contours [Yokoyama and Poggio, 2005; Sundaramoorthi et al., 2006] which use the full sensor input for the estimation of the target states, most tracking techniques are based on the assumption that the objects are reduced to points with no volume in the state space. In the latter case, detection algorithm is usually included before the actual tracker which extracts reference points (or bounding boxes) representing the objects as single elements in some measurement coordinate system. These detection methods of course induce an additional level of uncertainty on the true state of the objects since the extracted point-like measurements are rarely perfectly aligned with the actual targets. Apart from uncertainty in the extracted object locations, the data acquired by the sensors and detection methods is usually corrupted with varying amounts of noise, either induced by the environment which might have similar properties to the objects of interest and therefore produces spurious measurements, or induced by the sensor itself due to imperfections

of its physics (dead pixels, readout noise, high sensitivity etc.). On the other hand, both sensors and detectors are prone to missed detections which might be due to a low signal-to-noise ratio, occlusions in the monitored scene, signal overlap, etc., which makes it even harder for tracking algorithms to connect information along the time axis. Apart from these challenges, multi-target tracking techniques also have to cope with appearance and disappearance of objects which are a priori not distinguishable from false positives and false negatives for the sensor/detector, and target crossings or targets with close proximity to each other raise the question of data association. Figure 1 illustrates all these challenges: one can see that the measurement space (1b) rarely shows the same number of objects that are present in the real world (1a).



(a) Object state space.

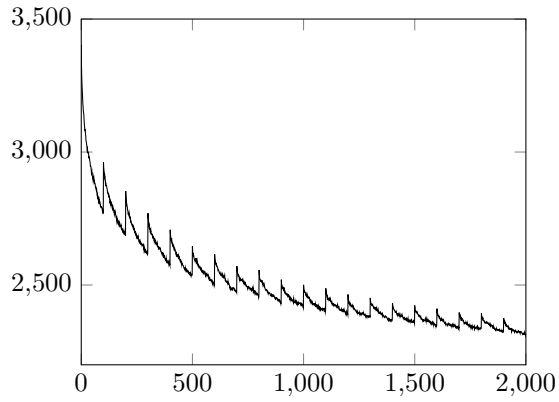
(b) Measurement space.

Figure 1: Illustration of the data association problem in multi-object filtering. True objects and their trajectories are marked in green, measurements in red. Dashed lines refer to movement at previous times, whereas the solid lines show the current movement of the targets.

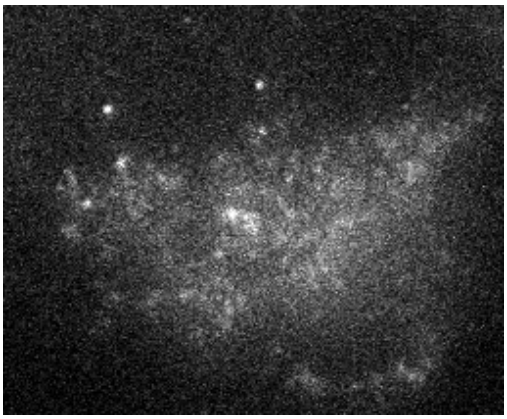
As demonstrated in Meijering et al. [2012], the interest in target tracking for biomedical research, in particular cell and particle tracking in microscopy image sequences, has been increasing exponentially since the 1970s. However, it was only very recently that advanced tracking techniques have been utilised for biomedical applications due to increased collaborations between scientists, mathematicians and engineers. In fact, many biomedical research labs still rely on rather ad-hoc heuristic methods in the manner of nearest-neighbour search [Husain et al., 2012; Casuso et al., 2012; Wilson et al., 2016] which cannot cope well with ambiguities in the data as described above. Other methods [Coraluppi and Carthel, 2004; Liang et al., 2010] use Multi-Hypothesis Tracking (MHT) or Joint Probabilistic Data Association (JPDA) which track each object individually using a group of single-object Bayesian filters. The problem of measurement-target association is solved by keeping all hypotheses of possible combinations (in case of the MHT filter, see Reid [1979]) or by selecting the most likely of those combinations (in case of the JPDA filter, see Fortmann et al. [1983]). These techniques can become computationally complex when the number of objects in the frame is high or there are many ambiguities in the data.

So far, all of the considered techniques focus on the data linking (i.e. associating measurements with specific objects) without paying much attention to false positives or fluctuating amounts of new objects in the scene. However, one issue that is particularly prominent in fluorescence microscopy is the high variability in the number of objects and in the amount of background noise: as explained in detail in Chapter 3, fluorescence in the samples gets excited repeatedly, i.e. every couple of

frames, the number of detections suddenly increases considerably and then slowly decreases again as more and more fluorescent molecules move to the dark state (cf. the intensity profile of a typical dataset in Figure 2a). Furthermore, the obtained images usually have a very low signal-to-noise ratio and hence naturally produce a lot of false positives and false negatives; an example frame is given in Figure 2b. In order to aid the analysis of such challenging scenarios, it would be therefore desirable to find a mathematically sound description of true objects and false positives/negatives alike which allows great flexibility in modelling the observed behaviours.



(a) Mean intensity in each image over time.



(b) A sample frame.

Figure 2: A typcial Photo-Activated Localisation Microscopy dataset (16 bit), showing the activity of N-type calcium channels on the membrane of PC12 cells.

Tracking groups of objects with the PHD filter

One of the objectives of this thesis is to provide principled methods for multi-object estimation and tracking, in particular for biomedical applications, which focuses on the problem of high variability in the number of false alarms and true objects, which is usually not given much attention in the existing techniques. Mahler [2003] introduced a novel type of multi-target filters in form of the PHD filter which shall be the focus of this work. This class of filters (including the more general Cardinalized PHD (CPHD) filter [Mahler, 2007a]) does not regard the set of objects as a collection of individuals, like in the MHT or JPDA filters, but rather tracks the target population *as a whole*. This interpretation makes it possible to avoid direct data association on the one hand and it gives mathematical tools to model the size of the target and clutter populations on the other hand. Although the original formulation of the PHD filter was embedded in the Finite Set Statistics (FISST) framework [Mahler, 2007b] which uses Random Finite Set (RFS)s to describe the target population, we prefer the more general concept of point processes in this work since it gives more flexibility to describe interesting phenomena such as regional correlation and it can recover all results of FISST in a natural way. Note that point processes were also used in Streit and Stone [2008] to formulate a point process variation of the PHD filter which was coined as the *intensity filter* there.

While the chosen class of filters already provides the ideal basis for the task of modelling changes in the number of objects properly, the PHD filter can be very restrictive in its modelling

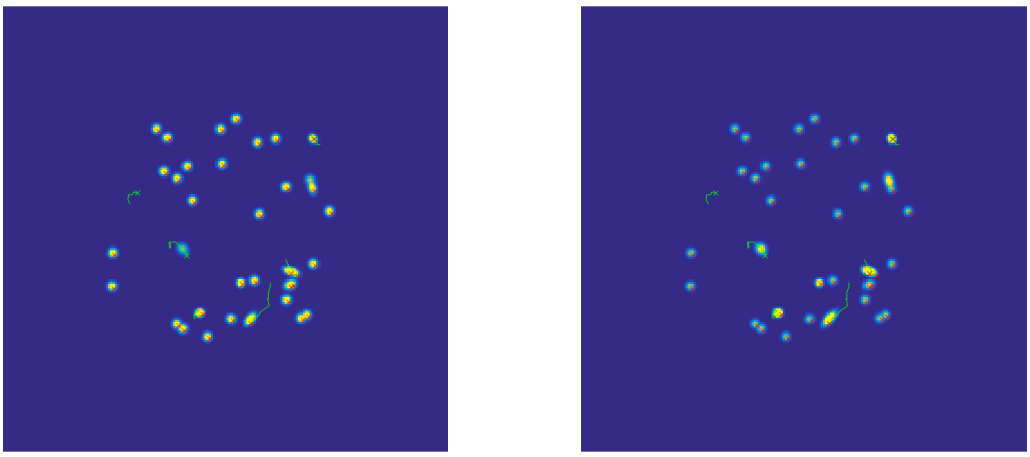
assumptions since it assumes that the number of true objects and false alarms is Poisson distributed. The Poisson assumption is problematic since it has equal mean and variance, which means that if, for example, the estimated number of targets is low, the filter also assumes a low variance in the number of targets; consequently, a spontaneous burst of new objects is not covered by the model and is therefore miss-interpreted as clutter. If, on the other hand, the estimated number of targets is reasonably high and there is an unexpected burst of false positives, the filter tends to assign high weight to these false detections very easily since the variance in the number of targets is assumed to be high as well. The existing alternative, the CPHD filter, does not make any assumption on the probability mass function (pmf) of the number of targets and false alarms, hence it offers maximum flexibility to circumvent the described issues of the PHD filter. Unfortunately, it has a significantly higher computational complexity since the whole cardinality distribution of the target group is propagated over time which involves a lot of additional calculations in every set of incoming data. Furthermore, it has been noticed by Fränken et al. [2009] that the CPHD filter produces the so-called "spooky effect at a distance" which causes a shift of probability mass from miss-detected objects in the scene to the detected ones, disregarding their mutual distance.

For all those reasons, we felt the need of an alternative filter whose modelling assumptions are more general than that of the PHD filter while still being fully described using only a small set of parameters to avoid the long runtimes that occur with the CPHD filter. Taking the Poisson distribution as a starting point, a suitable generalisation is found in the well-known negative binomial distribution whose variance is always greater than its mean [Daley and Vere-Jones, 2003]. Further research on the Poisson and the negative binomial distributions brought our attention to Harry H. Panjer's discovery of a unified distribution which encompasses the Poisson, binomial and negative binomial distributions in one single mathematical formulation [Panjer, 1981]. By taking the binomial distribution on board, the Panjer distribution admits also the case where the variance is less than the mean, hence giving full flexibility in the description of groups of objects in terms of the mean and variance in their cardinality (cf. Figure 2.1 in Chapter 2). Moreover, the binomial, negative binomial and Panjer distributions depend on only two parameters which stand in one-to-one correspondence with their mean and variance, a property which will be exploited in the techniques introduced in this work.

Key contributions

Our first attempt to generalise the PHD filter, published in Schlangen et al. [2016a] and presented below in Section 2.2.1, replaces the Poisson assumption on the population of false alarms by a negative binomial model. This PHD filter with negative binomial clutter proved to show a more robust performance in the presence of bursts of false alarms than the original PHD filter as shown in Figure 3; while the PHD filter assigns high weight to all received measurements alike, the new technique only gives significance to the already confirmed targets while downweighting the false measurements.

After these encouraging results, the more general Panjer distribution appeared to be the perfect candidate for a modelling assumption that yields a generalisation of the PHD filter while keeping



(a) PHD filter assuming Poisson clutter.

(b) PHD filter assuming NB clutter.

Figure 3: The PHD filter with Poisson and negative binomial clutter model in comparison. The ground truth is plotted in green, the measurements in red. The underlying map shows the probability distribution of the target population, where yellow hues represent a high probability of target presence, whereas blue hues represent a low probability.

the complexity reasonably low in contrast to the CPHD filter. Using the Panjer assumption for both the group of targets and false alarms resulted in the formulation of the Second-Order PHD (SO-PHD) filter, which is published in Schlangen et al. [2017c] and presented in Section 2.2.4. It is a second-order filter since the nature of the Panjer distribution makes it possible to propagate the variance in the number of objects along with the mean; since the Poisson, binomial and negative binomial cases are covered, the filter is able to choose the appropriate distributions on the fly which represents the observed behaviour most accurately.

Since the SO-PHD filter exploits variance information on the number of objects, it felt natural to study covariance and correlation as well. Even though none of the considered filters propagates those statistics directly, it is still possible to find the respective filter-specific formulae to extract this information after every iteration. Using point processes instead of RFSs, any statistics on specific regions in the surveillance scene can be explicitly formulated, thus we could study the spooky effect mentioned above [Fränken et al., 2009] by computing the regional correlation for the PHD, CPHD and SO-PHD filters (see Schlangen et al. [2017c] and Section 2.4 below). Thanks to regional correlation, it is possible to quantify the spooky effect, showing that the PHD filter is not affected when the objects are far away from each other, in contrast to the CPHD which shows a strong (mostly negative) correlation. The proposed SO-PHD filter, on the other hand, tends to have a much weaker correlation.

In addition to basic filtering, this work also explores the use of the PHD, CPHD and SO-PHD filters in a single-cluster framework which makes it possible to estimate filter parameters simultaneously with the target population. Based on the idea of the single-cluster PHD filter [Swain, 2013], the CPHD filter and the proposed SO-PHD filter were also embedded in a hierarchical algorithm in the same manner, see Schlangen et al. [2017b] and Section 2.3. First results on simulated data show that the obtained algorithms are capable alternatives to the single-cluster PHD filter which outperform other state-of-the-art techniques like Leung et al. [2016].

Regarding fluorescence microscopy applications, the discussed techniques also showed promising

results when run on real data. In Section 3.2.1, first studies on real data show that the proposed SO-PHD filter provides a capable low-complexity alternative to the CPHD filter which is more flexible in its modelling assumptions than the PHD filter. Moreover, the single-cluster PHD filter has been successfully utilised to eliminate stage drift in super-resolution microscopy videos (see Schlangen et al. [2016b] and Section 3.2.2) and to estimate unknown clutter rates (see Schlangen et al. [2017a] and Section 3.2.3).

This work is organised as follows: Chapter 1 gives a comprehensive introduction to point processes and it describes the PHD, CPHD and single-cluster PHD filters in this framework. Chapter 2 introduces the Panjer distribution and point process with their interesting properties, and provides the formulation and derivation of the PHD filter with Panjer clutter and the full SO-PHD filter. After that, the CPHD and SO-PHD filters are embedded in a single-cluster framework and compared with the single-cluster PHD filter in simulations. Furthermore, correlation and covariance of the PHD, CPHD and SO-PHD filters are presented in the same chapter, alongside the study on the spooky effect. Finally, Chapter 3 describes the technique of PALM and provides studies of the filters of interest on real microscopy data. Details on implementation are given in the appendix.

Chapter 1

Multi-object filtering using point processes

This chapter gives a summary about the state of the art concerning multi-object estimation using point processes. Its findings will lay the foundation for Chapter 2 which presents generalisations of the concepts introduced here. The notations closely follow Schlangen et al. [2016a], Schlangen et al. [2017c] and Schlangen et al. [2017b].

1.1 Probabilistic methods for representing groups of objects

In many applications where it is required to monitor multiple objects simultaneously, not only are their states unknown and time-varying, but also their number. For this reason, this section focuses on the concept of *point processes*, assuming that the targets can be represented by point states in a suitable state space. It will be shown in Chapter 3 how to extract point-like detections from single-molecule fluorescence microscopy images which represent the molecules of interest, but the following theory equally applies to other multi-object scenarios that can be described in this way.

For the remainder of this work, let us assume that a scene of interest is monitored with one or multiple sensors that produce measurements in some d_z-dimensional measurement space $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$ at discrete points in time, henceforth indexed by the natural numbers $k \in \mathbb{N}$. Targets are described in a d_x-dimensional topological state space $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ which is equipped with the Borel σ-algebra $\mathcal{B}(\mathcal{X})$. Recall that the Borel σ-algebra is the algebra generated by all open sets in $\mathcal{X}$ which respects the whole space $\mathcal{X}$, complements, and infinite unions of sets. The elements of $\mathcal{B}(\mathcal{X})$ are called *Borel sets* in $\mathcal{X}$.

1.1.1 Point processes

One convenient way to describe a group of objects probabilistically in $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ is by using a random variable Φ whose states and cardinality are both random; if the realisations of Φ are considered to be finite *sets*, Φ is called a *Random Finite Set (RFS)*, and if they are considered to be *sequences*, Φ is called a *point process* on $\mathcal{X}$. More formally, we can write:

Definition 1.1.1 (Random Finite Sets)**.** *A* Random Finite Set (RFS) $\mathbf{X}$ *taking values in* $\mathcal{X}$ *is a random variable on the collection of all finite subsets in* $\mathcal{X}$*. A realisation of* $\mathbf{X}$ *is a set* $\mathbf{x} = \{x_1, \ldots, x_n\} \subset \mathcal{X}$*, describing a group of* n *objects with states* $x_i \in \mathcal{X}$ *where both* n *and all* x_i *are random.*

Definition 1.1.2 (Point processes)**.** *A* point process Φ *on* $\mathcal{X}$ *is a random variable on the space* $\mathfrak{X} = \bigcup_{n\geq 0} \mathcal{X}^n$ *of finite sequences in* $\mathcal{X}$*. A realisation of* Φ *is a sequence* $\varphi = (x_1, \ldots, x_n) \in \mathcal{X}^n$*, describing a group of* n *objects with states* $x_i \in \mathcal{X}$ *where both* n *and all* x_i *are random.*[1]

The necessary theory to handle common statistics about RFSs is given in Mahler's Finite Set Statistics (FISST) framework [Mahler, 2007b]. This framework uses exclusively *probability densities* which are more commonly adopted in the engineering community. A major drawback of densities, however, is their inability to describe statistics on whole regions instead of single points. Since the main findings of this work involve regional statistics, it is reasonable to move from *densities* to *measures* and, consequently, to abandon the set-based formulation in favour of (unordered) sequences. Every probability density admits a probability measure by definition such that all findings within FISST can be easily recovered using measures. Note, however, that the reverse is not necessarily true; only probability measures that are absolutely continuous with respect to some reference measure λ admit a density according to the theorem of Radon and Nikodym [Nikodym, 1930]; this density is then called *Radon-Nikodym derivative*.

A probability measure P_Φ can be defined on the measurable space $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ through the projection measures $P_\Phi^{(n)}$ which define probability measures on the product spaces $\mathcal{X}^n$ for all $n \geq 0$, i.e. on the space of sequences of length n [Stoyan et al., 1997]. The order of the sequences is implicitly removed by assuming that the probability of permutations of the same sequence φ is equal, i.e.

$$P_\Phi^{(n)}\Big((x_1, \ldots, x_n)\Big) = P_\Phi^{(n)}\Big((x_{\pi(1)}, \ldots, x_{\pi(n)})\Big) \tag{1.1}$$

for any permutation π of the index set $\{1, \ldots, n\}$.

For handling discrete values, in particular the set of acquired measurements at a given time, it will be natural to exploit the notion of densities when necessary. Therefore, let us assume that the probability distribution P_Φ admits a density p_Φ with respect to a suitable reference measure λ. The Radon-Nikodym derivatives of the projection measures $P_\Phi^{(n)}$ will be denoted by $p_\Phi^{(n)}$.

A point process is called *simple* if all elements x_i of a realisation φ are distinct. In this sense, a RFS is a simple point process since sets do not allow repetition by definition. In classical tracking scenarios with a continuous state space, two objects almost surely do not share the same state, therefore it will be assumed in the following that the considered point processes are simple.[2]

[1] Note that as an abuse of notation, the formulation "$x \in \varphi$" will be used in this work to indicate that the element $x \in \mathcal{X}$ is a part of the sequence φ.

[2] It was argued in [Mahler, 2007b] that repetition does not have a physical meaning and is therefore "unneccessary" [Mahler, 2007b, p.710], but it is of utmost relevance for experiments with discrete outcomes and should therefore not be discarded in the general case.

1.1.2 Statistical moments

Just like for real-valued random variables, it is possible to formulate *statistical moments* for a point process Φ in the following way:

Definition 1.1.3 (Factorial and non-factorial moment measures [Stoyan et al., 1997])**.**

(a) The n-th order (non-factorial) moment measure $\mu_\Phi^{(n)}$ *of a point process* Φ *is a measure on* $\mathcal{X}^n$ *such that*

$$\int f_n(x_{1:n})\mu_\Phi^{(n)}(\mathrm{d}(x_{1:n})) = \mathbb{E}\Bigg[\sum_{(x_1,\ldots,x_n)\in\Phi} f_n(x_{1:n})\Bigg] \tag{1.2a}$$

$$= \int \Bigg[\sum_{(x_1,\ldots,x_n)\in\Phi} f_n(x_{1:n})\Bigg] P_\Phi(\mathrm{d}\varphi) \tag{1.2b}$$

for any bounded measurable n*-variate function* f_n *on* $\mathcal{X}^n$*, where* $x_{1:n} := (x_1, \ldots, x_n)$*.*

(b) The n-th order factorial moment measure $\nu_\Phi^{(n)}$ *of a point process* Φ *is a measure on* $\mathcal{X}^n$ *such that*

$$\int f_n(x_{1:n})\nu_\Phi^{(n)}(\mathrm{d}(x_{1:n})) = \mathbb{E}\Bigg[\sum_{(x_1,\ldots,x_n)\in\Phi}^{\neq} f_n(x_{1:n})\Bigg] \tag{1.3a}$$

$$= \int \Bigg[\sum_{(x_1,\ldots,x_n)\in\Phi}^{\neq} f_n(x_{1:n})\Bigg] P_\Phi(\mathrm{d}\varphi) \tag{1.3b}$$

for any bounded measurable n*-variate function* f_n *on* $\mathcal{X}^n$*, where* $\Sigma^{\neq}$ *indicates that the points* $x_1, \ldots, x_n$ *are all pairwise distinct.*

This definition is also known as *Campbell's theorem* in the literature [Stoyan et al., 1997].

Based on Definition 1.1.3, it is possible to describe the expectation of an event where the points $x_1, \ldots, x_n$ fall in the regions $B_1, \ldots, B_n$ simultaneously: As a consequence of the definition of the indicator function

$$\mathbb{1}_B(x) = \begin{cases} 1 & \text{for } x \in B, \\ 0 & \text{otherwise,} \end{cases} \tag{1.4}$$

it follows that

$$\mu_\Phi^{(n)}(B_1 \times \cdots \times B_n) = \mathbb{E}\Bigg[\sum_{x_1,\ldots,x_n\in\Phi} \mathbb{1}_{B_1}(x_1)\ldots\mathbb{1}_{B_n}(x_n)\Bigg], \tag{1.5}$$

$$\nu_\Phi^{(n)}(B_1 \times \cdots \times B_n) = \mathbb{E}\Bigg[\sum_{x_1,\ldots,x_n\in\Phi}^{\neq} \mathbb{1}_{B_1}(x_1)\ldots\mathbb{1}_{B_n}(x_n)\Bigg], \tag{1.6}$$

for any regions $B_i \in \mathcal{B}(\mathcal{X})$, $1 \leq i \leq n$. Note that Eq. (1.5) allows for repetitions in the sequence $x_1, \ldots, x_n$, whereas Eq. (1.6) does not.

It is easy to see from the definition that the first-order factorial and non-factorial measures coincide, i.e. $\mu_\Phi^{(1)}(\cdot) = \nu_\Phi^{(1)}(\cdot)$; this quantity is also known as the *intensity measure* of Φ. Both the intensity measure and its associated density will be denoted by μ_Φ in this work, dropping the superscript "(1)" where it is unambiguous. In the context of RFS, μ_Φ is also known as the

Probability Hypothesis Density [Mahler, 2003], which inspired the name of the class of Bayesian filters which is studied below. Note that for any regions $B, B' \in \mathcal{B}(\mathcal{X})$ [Stoyan et al., 1997] the second-order moment measure can be decomposed into

$$\mu_\Phi^{(2)}(B \times B') = \mu_\Phi(B \cap B') + \nu_\Phi^{(2)}(B \times B'). \tag{1.7}$$

Since they will play a fundamental role in this work, let us define the following second-order statistics.

Definition 1.1.4 (Covariance, variance and correlation of Φ [Stoyan et al., 1997; Illian et al., 2008])**.** *The* covariance*,* variance*, and* correlation *of a point process Φ are defined with*

$$\mathrm{cov}_\Phi(B, B') = \mu_\Phi^{(2)}(B \times B') - \mu_\Phi(B)\mu_\Phi(B'), \tag{1.8}$$

$$\mathrm{var}_\Phi(B) = \mu_\Phi^{(2)}(B \times B) - [\mu_\Phi(B)]^2, \tag{1.9}$$

$$\mathrm{corr}_\Phi(B, B') = \frac{\mathrm{cov}_\Phi(B, B')}{\sqrt{\mathrm{var}_\Phi(B)}\sqrt{\mathrm{var}_\Phi(B')}}, \tag{1.10}$$

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

All three of these quantities yield scalar values in $\mathbb{R}$. While the intensity $\mu_\Phi(B)$ describes the expected number of objects in B, the variance $\mathrm{var}_\Phi(B)$ stands for the variability of the target number in B around the expected value [Delande et al., 2014].[3] Formally, the variance is a special case of the more abstract covariance $\mathrm{cov}_\Phi(B, B')$ for $B = B'$. Furthermore, the correlation $\mathrm{corr}_\Phi(B, B')$, yielding a scalar in the interval $[-1, 1]$, describes the relation between the estimated number of targets in B and B', where positive/negative values stand for positive/negative proportionality, respectively. The correlation will be exploited in the simulations presented in Section 2.4 below to quantify a filtering effect known as the "spooky effect" of multi-object filters [Fränken et al., 2009].

1.1.3 Probability generating functionals

It is possible to study point processes in an alternative representation, i.e. via the *Probability Generating Functional (PGFL)* and the *Laplace functional*. They play a similar role to the Fourier transform in signal processing and the probability generating function for discrete, real-valued random variables. The notation of PGFLs was already previously exploited in Streit [2013] to write down the Probability Hypothesis Density (PHD) and intensity filter equations, and in Streit et al. [2015] to formulate various other multi-object filters in the same framework.

Definition 1.1.5 (Probability generating functional and Laplace functional)**.** *Let $f : \mathcal{X} \to \mathbb{R}^+$ and $h : \mathcal{X} \to [0, 1]$ be two test functions. The* Laplace functional $\mathcal{L}_\Phi$ *and the* Probability Generating

[3] In general, var_Φ is non-additive and does not admit a density, which is one of the reasons why a measure-theoretic approach was chosen for this work; see also Section 2.4.

Functional (PGFL) $\mathcal{G}_\Phi$ *of a point process* Φ *are defined by*

$$\mathcal{L}_\Phi(f) = \sum_{n\geq 0} \int \exp\left(-\sum_{i=1}^{n} f(x_i)\right) P_\Phi^{(n)}(\mathrm{d}x_{1:n}), \tag{1.11}$$

$$\mathcal{G}_\Phi(h) = \sum_{n\geq 0} \int \left[\prod_{i=1}^{n} h(x_i)\right] P_\Phi^{(n)}(\mathrm{d}x_{1:n}). \tag{1.12}$$

From the definitions (1.11) and (1.12), one easily verifies the relation

$$\mathcal{G}_\Phi(h) = \mathcal{L}_\Phi(-\ln h). \tag{1.13}$$

Functional differentiation is an important tool to work with PGFLs and Laplace functionals as we will see below. There are different operators available, e.g. the Gâteaux, the Fréchet and the chain differential [Bernhard, 2005]. In this work, the chain differential is exploited which admits a general higher-order product and chain rule [Clark and Houssineau, 2013; Clark et al., 2015].

Definition 1.1.6 (Chain differential [Bernhard, 2005])**.** *Let* $(\eta_n : \mathcal{X} \to \mathbb{R}^+)_{n\in\mathbb{N}}$ *be a sequence of positive, bounded functions converging pointwise to a function* $\eta : \mathcal{X} \to \mathbb{R}^+$*, and let* $(\varepsilon_n)_{n\in\mathbb{N}}$ *be a sequence of positive real values converging to* 0*. The chain differential of a functional* G *with respect to its functional argument* $h : \mathcal{X} \to \mathbb{R}^+$ *in the direction of* η *is defined as*

$$\delta G(h;\eta) = \lim_{n\to\infty} \frac{G(h+\varepsilon_n\eta_n) - G(h)}{\varepsilon_n}. \tag{1.14}$$

If the limit exists, it is unique for any sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ *and* $(\eta_n : \mathcal{X} \to \mathbb{R}^+)_{n\in\mathbb{N}}$ *with the above properties.*

It can be shown that the projection measures $P_\Phi^{(n)}$ and the nth order non-factorial and factorial moments $\mu_\Phi^{(n)}$ and $\nu_\Phi^{(n)}$ can be recovered from the PGFL or the Laplace functional of Φ through differentiation with

$$P_\Phi^{(n)}(B_1\times\cdots\times B_n) = \frac{1}{n!}\delta^n\mathcal{G}_\Phi(h;\mathbb{1}_{B_1},\ldots,\mathbb{1}_{B_n})|_{h=0}, \tag{1.15}$$

$$\mu_\Phi^{(n)}(B_1\times\cdots\times B_n) = (-1)^n\delta^n\mathcal{L}_\Phi(f;\mathbb{1}_{B_1},\ldots,\mathbb{1}_{B_n})|_{f=0}, \tag{1.16}$$

$$\nu_\Phi^{(n)}(B_1\times\cdots\times B_n) = \delta^n\mathcal{G}_\Phi(h;\mathbb{1}_{B_1},\ldots,\mathbb{1}_{B_n})|_{h=1}, \tag{1.17}$$

for any set of regions $B_i \in \mathcal{B}(\mathcal{X})$, $1 \leq i \leq n$ [Stoyan et al., 1997]. As mentioned before, the chain differential admits a *product rule!first order* [Bernhard, 2005]

$$\delta(F\cdot G)(h;\eta) = \delta F(h;\eta)G(h) + F(h)\delta G(h;\eta), \tag{1.18}$$

and a *chain rule* [Bernhard, 2005]

$$\delta(F\circ G)(h;\eta) = \delta F(G(h);\delta G(h;\eta)). \tag{1.19}$$

Similarly, one can show the *n-fold product rule* [Clark et al., 2015]

$$\delta^n(F \cdot G)(h;\eta_1,\ldots,\eta_n) = \sum_{\omega\subseteq\{1,\ldots,n\}} \delta^{|\omega|}F\Big(h;(\eta_i)_{i\in\omega}\Big)\delta^{|\bar{\omega}|}G\Big(h;(\eta_j)_{j\in\bar{\omega}}\Big), \tag{1.20}$$

where $\bar{\omega} = \{1,\ldots,n\} \setminus \omega$, and the *n-fold chain rule* [Clark and Houssineau, 2013; Clark et al., 2015]

$$\delta^n(F \circ G)(h;\eta_1,\ldots,\eta_n) = \sum_{\pi\in\Pi_n} \delta^{|\pi|}F\left(G(h);\left(\delta^{|\omega|}G(h;(\eta_i)_{i\in\omega})\right)_{\omega\in\pi}\right), \tag{1.21}$$

where Π_n is the set of partitions of the index set $\{1,\ldots,n\}$. Eq. (1.21) is also known as *Faà di Bruno's formula for chain differentials*, and it has recently been exploited in different contexts, see e.g. [Clark and Houssineau, 2013, 2014; Bryant et al., 2016; Schlangen et al., 2016a]. The generalised product rule for set derivatives [Mahler, 2007b, p. 389] can be seen as the RFS equivalent of Eq. (1.20).

Remark 1.1.7. *Linear and continuous differential operators are commonly referred to as* derivatives*; therefore, the chain differential will be called* chain derivative *or simply* derivative *for the remainder of this work since in the cases of interest, it behaves as such as we will see below.*

Remark 1.1.8. *As mentioned earlier, only measures which are absolutely continuous with respect to some reference measure admit a density, namely their Radon-Nikodym derivative. In that sense, the notation $\delta G(f,\delta_x)$ is defined as the Radon-Nikodym derivative of the measure $\mu' : B \mapsto \delta G(f,\mathbb{1}_B)$ evaluated at point x, i.e.*

$$\delta G(f,\delta_x) := \frac{\mathrm{d}\mu'}{\mathrm{d}\lambda}(x), \tag{1.22}$$

for any suitable function f on $\mathcal{X}$ and any point $x \in \mathcal{X}$. The PGFL and the Laplace functional of any point process Φ hold this property since the probability distribution P_Φ is assumed to admit a density with respect to the reference measure λ. The following notation will be extensively used in the derivations shown later on:

$$p_\Phi^{(n)}(x_1,\ldots,x_n) = \frac{1}{n!}\delta^n\mathcal{G}_\Phi(h;\delta_{x_1},\ldots,\delta_{x_n})|_{h=0}, \tag{1.23}$$

for any $x_i \in \mathcal{X}$, $1 \leq i \leq n$. Again, there is an equivalent formulation for RFSs, found in Eq. (52) in [Mahler, 2003].

1.1.4 Examples of point processes

Three different point processes are discussed below to illustrate the formulation of point processes on specific examples. All of them will be used later on to model different phenomena, resulting in different multi-object filters.

Definition 1.1.9 (The i.i.d. cluster process)**.** *An independent and identically distributed (i.i.d.) cluster process with cardinality distribution ρ on $\mathbb{N}$ and spatial distribution s on $\mathcal{X}$ describes a*

group of objects whose size is described by ρ, and whose states are i.i.d. according to s. Its PGFL is given by

$$\mathcal{G}_{\text{i.i.d.}}(h) = \sum_{n\geq 0} \rho(n) \left[\int h(x)s(\mathrm{d}x)\right]^n . \tag{1.24}$$

The formulation i.i.d. cluster process is quite general since it does not make any assumptions on the cardinality distribution ρ. The Cardinalized PHD (CPHD) filter (see Section 1.3.3 and [Mahler, 2007a]) is constructed with the assumption that the predicted target process $\Phi_{k|k-1}$ is an i.i.d. cluster process, i.e. this filter allows a great flexibility to describe the cardinality distribution of the target process.

Definition 1.1.10 (The Bernoulli process)**.** *A Bernoulli point process with parameter $0 \leq p \leq 1$ 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 $(1-p)$. Its PGFL is given by*

$$\mathcal{G}_{\text{Bernoulli}}(h) = (1-p) + p\int h(x)s(\mathrm{d}x). \tag{1.25}$$

Binary events like the detection or the survival process of a target are commonly modelled using a Bernoulli point process.

Definition 1.1.11 (The 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*

$$\mathcal{G}_{\text{Poisson}}(h) = \exp\left(\int [h(x)-1]\mu(\mathrm{d}x)\right), \tag{1.26}$$

where the intensity measure μ of the process is such that $\mu(\mathrm{d}x) = \lambda s(\mathrm{d}x)$.

The Poisson point process has convenient properties that made it an attractive modelling choice for the PHD filter (see Section 1.3.2 and [Mahler, 2003]): First of all, the Poisson model is prevalent in many applications and therefore highly relevant; furthermore, it is completely described by only one parameter λ. Its mean and variance are equal and depend directly on λ, i.e. $\mu_{\text{Poisson}}(B) = \text{var}_{\text{Poisson}}(B) = \lambda s(B)$ for any $B \in \mathcal{B}(\mathcal{X})$. Moreover, convenient differentiation rules exist to handle exponential functionals (see Proposition 1.1.13 below in Section 1.1.5), i.e. the Poisson point process (1.26) is easy to manipulate.

1.1.5 Derivation rules for exponential functionals

Let us find the first- and nth order differentiation rules for the exponential functional $\exp(\cdot)$ which will be helpful for the formulation of the results in the subsequent chapters.

Lemma 1.1.12 ([Clark et al., 2015])**.** *The first-order derivative of the exponential functional* exp *in composition with a general functional $G(h)$ is the functional chain derivative*[4]

$$\delta(\exp\circ G)(h;\eta) = \exp(G(h))\delta G(h;\eta). \tag{1.27}$$

[4]Note that this differential can indeed be called derivative since it is linear and continuous, cf. Remark 1.1.7.

Proof. Choose a series of functions $(\eta_n)_{n\in\mathbb{N}}$ which converges pointwise to $\delta G(h;\eta)$ for $n\to\infty$, and a series $(\varepsilon_n)_{n\in\mathbb{N}}$ which converges to 0. Then

$$\begin{aligned}
\delta(\exp\circ G)(h;\eta) &\overset{(1.19)}{=} \delta\exp(G(h);\delta G(h;\eta)) \\
&= \lim_{n\to\infty}\frac{1}{\varepsilon_n}[\exp(G(h)+\varepsilon_n\eta_n)-\exp(G(h))] \\
&= \exp(G(h))\lim_{n\to\infty}\frac{1}{\varepsilon_n}[\exp(\varepsilon_n\eta_n)-1] \\
&= \exp(G(h))\lim_{n\to\infty}\frac{1}{\varepsilon_n}\left[\sum_{j\geq 0}\frac{(\varepsilon_n\eta_n)^j}{j!}-1\right] \\
&= \exp(G(h))\lim_{n\to\infty}\frac{1}{\varepsilon_n}\left[\varepsilon_n\eta_n+\left(\sum_{j\geq 2}\frac{(\varepsilon_n\eta_n)^j}{j!}\right)\right] \\
&= \exp(G(h))\lim_{n\to\infty}\left[\eta_n+\underbrace{\sum_{j\geq 2}\frac{\varepsilon_n^{j-1}\eta_n^j}{j!}}_{=0(n\to\infty)}\right] \\
&= \exp(G(h))\delta G(h;\eta).
\end{aligned}$$

□

Proposition 1.1.13. *Let G be a linear functional. The nth-order chain derivative of the composition $\exp(G(h))$ is found to be*

$$\delta^n(\exp\circ G)(h;\eta_1,\ldots,\eta_n) = \exp(G(h))\prod_{i=1}^{n}\delta G(h;\eta_i). \tag{1.28}$$

Proof. The nth-order derivative of $\exp(G(h))$ can be easily seen from Faà di Bruno's formula (1.21): 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 $\eta_1,\ldots,\eta_n$. □

1.2 The Bayes filter

In biomedical applications, it is still very common to use heuristic nearest-neighbour techniques for target tracking[5] [Chenouard et al., 2014], despite the fact that biomedical images often suffer from a low signal-to-noise ratio (SNR) and therefore, data association ambiguities, missed detections and false positives are extremely common (see the discussion in Chapter 3). Allowing for *uncertainty* in the analysis generally leads to more robust tracking results by incorporating statistics on the spatial distribution of the objects, the detection and survival rates, birth and clutter models, expected motion behaviour, etc. For this reason, the Bayes recursion provides a useful framework since it accommodates a probabilistic solution to the state estimation problem. It can be formulated for a single object or multiple objects jointly.

[5] In biomedical literature, target tracking is usually referred to as *data linking*.

1.2.1 Single-object Bayes filter

In the single-object scenario, let us assume that the monitored object exists with probability 1 at all times and produces at most one measurement per time scan. Furthermore, it is assumed that the sensor does not produce false alarms. For the multi-object case below, the existence assumption will be released to account for a time-varying number of objects.

Let $x_{k-1} \in \mathcal{X}$ be the hidden target state at time $k-1$, which is described through the probability measure

$$p_{k-1}(\mathrm{d}x_{k-1}|z_0, \ldots, z_{k-1}) \tag{1.29}$$

depending on all measurements $z_0, \ldots, z_{k-1} \in \mathcal{Z}$. The Bayes recursion first *predicts* the next target state x_k at time k using an appropriate dynamical model, e.g. Brownian motion or a near constant velocity (n.c.v.) model and then *corrects* this belief based on the measurement model describing the sensor. Usually, the dynamical model $t_{k|k-1}$ is assumed to be Markovian, i.e.

$$t_{k|k-1}(\mathrm{d}x_k|x_0 \ldots x_{k-1}) = t_{k|k-1}(\mathrm{d}x_k|x_{k-1}). \tag{1.30}$$

Using the prior p_{k-1} and the Markov transition $t_{k|k-1}$, the prediction step is performed with the *Chapman-Kolmogorov equation*

$$p_{k|k-1}(\mathrm{d}x_k|z_0, \ldots, z_{k-1}) = \int t_{k|k-1}(\mathrm{d}x_k|x) p_{k-1}(\mathrm{d}x|z_0, \ldots, z_{k-1}). \tag{1.31}$$

The Chapman-Kolmogorov equation propagates the object state from time $k-1$ to time k by marginalising over all possible previous states $x_{k-1} \in \mathcal{X}$.

Assume that the sensor detects the object at time k, producing a measurement $z_k \in \mathcal{Z}$. Uncertainty about the sensor accuracy is incorporated using a likelihood function $l_k(x_k|z_k)$ that determines how likely measurement z_k was produced by the target state x_k. Using l_k and the predicted distribution $p_{k|k-1}$, the Bayes update is given by

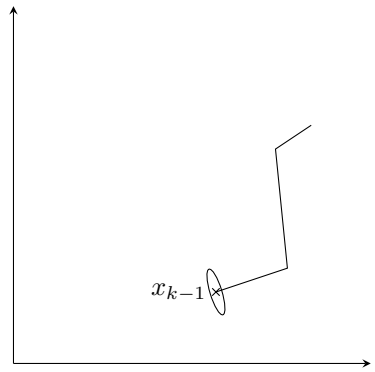
$$p_k(\mathrm{d}x_k|z_0, \ldots, z_k) = \frac{l_k(x_k|z_k) p_{k|k-1}(\mathrm{d}x_k|z_0, \ldots, z_{k-1})}{\int l_k(z_k|x) p_{k|k-1}(x|z_0, \ldots, z_{k-1}) dx}, \tag{1.32}$$

which is also known as *Bayes' rule*.

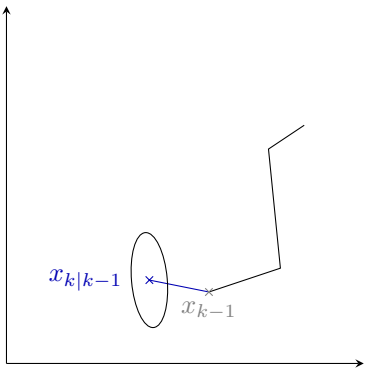
Fig. 1.1 gives a visual description of the Bayesian recursion, representing uncertainty by ellipses. While the prediction usually increases the uncertainty of the object state, the update decreases the uncertainty again by involving new information. This behaviour is very intuitive and makes it possible to recover the object position even after some time of no detections since the increasing uncertainty gives credit to a wider area where the object might have travelled.

In general, the Bayes filter is computationally intractable without further assumptions. However, if the state transition $t_{k|k-1}$ is a linear operator and furthermore, the likelihood function l_k and the prior and predicted probability densities are Gaussian, the single-object Bayes filter becomes the well-known Kalman filter [Kalman, 1960] (given in Appendix A.1) which is widely used in single- and multi-object tracking applications since it provides an optimal solution to

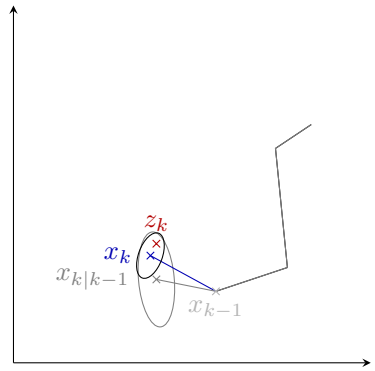
the single-target estimation problem. If the transition is non-linear but can be locally described by a linear transformation, the Extended Kalman Filter (EKF) provides a useful approximation [Anderson and Moore, 1979, Ch. 8], and for highly non-linear problems the Unscented Kalman Filter (UKF) is a non-optimal alternative [Julier and Uhlmann, 1997]. Increasingly cheap processing power has facilitated the development of Sequential Monte Carlo (SMC) techniques which draw and propagate random samples from the probability density [Del Moral, 1996]. This approach has the advantage that it does not rely on any modelling assumptions, however the samples have to be drawn carefully to avoid divergence, and increasing the accuracy by using more particles comes with an increase of computational time.



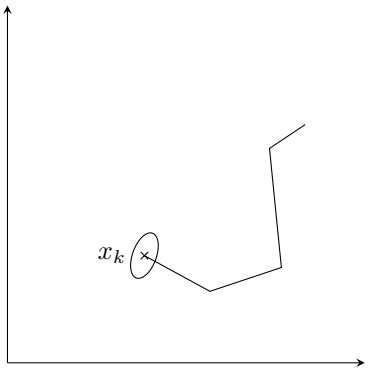
(a) Prior belief x_{k-1} of the target position.



(b) *Prediction* of state x_{k-1} to state $x_{k|k-1}$ using a suitable dynamical model.



(c) *Update* of predicted state $x_{k|k-1}$ to x_k with measurement z_k.



(d) New belief x_k of the current target position.

Figure 1.1: Single-target Bayesian filtering.

1.2.2 Multi-object Bayes filter

A measure-theoretic formulation of the multi-target Bayesian recursion can be found via

$$P_{k|k-1}(\mathrm{d}\xi) = \int T_{k|k-1}(\mathrm{d}\xi|\varphi)P_{k-1}(\mathrm{d}\varphi), \tag{1.33}$$

$$P_k(\mathrm{d}\xi|Z_k) = \frac{L_k(Z_k|\xi)P_{k|k-1}(\mathrm{d}\xi)}{\int L_k(Z_k|\varphi)P_{k|k-1}(\mathrm{d}\varphi)}, \tag{1.34}$$

where $P_\bullet$ is the probability distribution describing the multi-target state, $T_{k|k-1}$ is the multi-target Markov transition kernel from time $k-1$ to time k, and L_k is the multi-target/multi-measurement

likelihood at time k.[6] Note that Eqns (1.33) and (1.34) are based on the RFS-based formulation in [Mahler, 2007b], where *set* integrals are used in favour of using densities rather than measures.

Again, the general Bayes filter formulation is computationally intractable without any additional assumptions, even more so because the combinatorial problem of data association has to be solved. The oldest approaches such as the well-established Multi-Hypothesis Tracking (MHT) [Reid, 1979] and Joint Probabilistic Data Association (JPDA) [Fortmann et al., 1983] perform single-target tracking on each individual and create multiple hypotheses for possible associations between measurements and targets. A more recent development is the Multi-Bernoulli filter [Mahler, 2007b] and its derivatives [Vo et al., 2009; Williams, 2012, 2015] which incorporate a probability of existence in the target state and describe every object with a Bernoulli experiment on its existence. A completely different philosophy lies behind the PHD filter [Mahler, 2003] and its derivatives since the association problem is eliminated by tracking the target population as a whole without caring about individual trajectories.[7] While the RFS framework was used traditionally to formulate the PHD filter, it is possible to find an equivalent formulation with point processes as demonstrated in Section 1.3 below.

1.3 Multi-object estimation using point processes

This section presents the PHD and CPHD filters in the mathematical framework of point processes. These methods do not consider the full probability distribution of the process they describe, but they propagate first-order moment information instead which makes it possible to find closed-form solutions.

The oldest recursion propagating a point process through its first-order moment, recalled in Section 1.3.2, is the well-established PHD filter [Mahler, 2003], which will be referred to with the superscript $\flat$. Based on the assumption that the cardinalities of both the predicted target process and the clutter process follow a Poisson model, this filter provides an elegant closed-form solution to the multi-object estimation problem which propagates the first-order moment of the underlying target process, which gave the filter its name. After the need for higher-order information was expressed in [Erdinc et al., 2005], Mahler derived the CPHD filter [Mahler, 2007a] which propagates the full cardinality distribution instead of choosing a particular one like in the PHD filter (see Section 1.3.3). This method is very flexible in the description of the population cardinality, but it also comes with a much greater computational cost since the full cardinality distribution has to be computed in every iteration of the algorithm. The notation $\sharp$ will be used below to mark terms specific to the CPHD filter.

Before stating the filtering equations, let us define some useful notation which will be used for the remainder of this work.

[6] For two measures μ, μ' on some space X, the notation $\mu(\mathrm{d}x) = \mu'(\mathrm{d}x)$ for $x \in X$ will be used 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.

[7] Note that although it is not a natural output of the algorithm, capable track extraction methods exist nonetheless, see Appendix A.2.

1.3.1 Useful notations

In the single-target case, the monitored object was assumed to exist with probability 1 at all times. In most multi-object scenarios, however, the number of objects is generally unknown and time-varying, i.e. objects might randomly leave or enter the scene. Therefore, let us introduce a function $p_{\mathrm{s}}(\cdot)$ describing the state-dependent *probability of survival* and a *birth process* $\Phi_k^{\flat}$ with intensity $\mu_{\mathrm{b},k}$ that models new incoming targets. The single-target Markov transition will be denoted by $t_{k|k-1}$.

Regarding the incoming sensor data, let us denote by $Z_k \subset \mathcal{Z}$ the set of measurements obtained at time k, and let $z \in Z_k$ denote a single measurement. Since the information obtained by the sensor is never flawless, let us incorporate a *probability of detection* for the targets, denoted by $p_{\mathrm{d}}(\cdot)$, and a *clutter process* Φ_k^{c} with intensity $\mu_{\mathrm{c},k}$ that describes the occurrence of false positives over time. The function ℓ_k will denote the likelihood of associating a single measurement with a specific target at time k.

Finally, let us define some specific terms which will be common to all filters described below, up to the specific prior and predicted intensities $\mu_{k-1}^{\bullet}$ and $\mu_{k|k-1}^{\bullet}$ with $\bullet \in \{\flat, \sharp\}$.[8] The *survival term* is defined with

$$\mu_{\mathrm{s},k}(B) = \int_{\mathcal{X}} p_{\mathrm{s},k}(x) t_{k|k-1}(B|x) \mu_{k-1}^{\bullet}(\mathrm{d}x). \tag{1.35}$$

and the *missed detection* and *association terms* with

$$\mu_k^{\phi}(B) = \int_B (1 - p_{\mathrm{d},k}(x)) \mu_{k|k-1}^{\bullet}(\mathrm{d}x) \tag{1.36}$$

$$\mu_k^{z}(B) = \int_B p_{\mathrm{d},k}(x) \ell_k(x|z) \mu_{k|k-1}^{\bullet}(\mathrm{d}x), \tag{1.37}$$

for any measurement $z \in Z_k$ on any region $B \in \mathcal{B}(\mathcal{X})$.

1.3.2 The PHD filter [Mahler, 2003]

Proposition 1.3.1 (PHD recursion [Mahler, 2003])**.** *(a) The predicted first-order moment measure is given by*

$$\mu_{k|k-1}^{\flat}(B) = \mu_{\mathrm{b},k}(B) + \mu_{\mathrm{s},k}(B) \tag{1.38}$$

with survival intensity (1.35)*, where* $\bullet = \flat$*.*

(b) The updated first-order moment measure with Poisson distributed prediction and clutter model is derived as

$$\mu_k^{\flat}(B) = \mu_k^{\phi}(B) + \sum_{z \in Z_k} \frac{\mu_k^{z}(B)}{\mu_{\mathrm{c},k}(z) + \mu_k^{z}(\mathcal{X})} \tag{1.39}$$

with missed detection term (1.36) *and association term* (1.37)*, where* $\bullet = \flat$*.*

[8] In Chapter 2, this notation will also be used in the formulation of two new PHD-like methods, where $\bullet \in \{\natural, \natural\}$.

1.3.3 The CPHD filter [Mahler, 2007a]

As mentioned before, the CPHD filter recursion propagates both the intensity and the cardinality distribution of the target process. Therefore, define ρ_k to be the cardinality distribution of the target population at time k and let ρ_{b} and ρ_{c} be the birth and clutter cardinality distributions, respectively.

Let us denote the discrete and continuous inner product by

$$\langle f, g \rangle = \int f(x)g(x)\mathrm{d}(x) \quad \text{(continuous case)}, \tag{1.40}$$

$$\langle f, g \rangle = \sum_{n \geq 0} f(n)g(n) \qquad \text{(discrete case)}. \tag{1.41}$$

Furthermore, define the terms $\Upsilon^d[\mu, Z]$ with [Vo et al., 2007]

$$\Upsilon^d[\mu, Z](n) = \sum_{j=0}^{\min(|Z|, n-d)} \frac{n!(|Z|-j)!}{(n-(j+d))!} \rho_{\mathrm{c}}(|Z|-j) \frac{\mu_k^{\phi}(\mathcal{X})^{n-(j+d)}}{\mu_{k|k-1}^{\sharp}(\mathcal{X})^n} e_j(Z), \tag{1.42}$$

where $e_j(Z)$ denote the elementary symmetric functions defined by

$$e_j(Z) := \sum_{\substack{Z' \subseteq Z \\ |Z'|=j}} \prod_{z \in Z'} \frac{\mu_k^z(\mathcal{X})}{s_{\mathrm{c},k}(z)}. \tag{1.43}$$

These terms are used to define the factors $l_1^{\sharp}(\phi)$ and $l_1^{\sharp}(z)$ as

$$l_1^{\sharp}(\phi) = \frac{\langle \Upsilon^1[\mu, Z], \rho_{k|k-1} \rangle}{\langle \Upsilon^0[\mu, Z], \rho_{k|k-1} \rangle}, \tag{1.44}$$

$$l_1^{\sharp}(z) = \frac{\langle \Upsilon^1[\mu, Z \setminus \{z\}], \rho_{k|k-1} \rangle}{\langle \Upsilon^0[\mu, Z], \rho_{k|k-1} \rangle}. \tag{1.45}$$

Proposition 1.3.2 (CPHD recursion [Mahler, 2007a]). *(a) In the manner of* (1.38)*, the predicted first-order moment measure is found to be*

$$\mu_{k|k-1}^{\sharp}(B) = \mu_{\mathrm{b},k}(B) + \mu_{\mathrm{s},k}(B), \tag{1.46}$$

with $\bullet = \sharp$ *in* (1.35)*. The predicted target cardinality distribution is given by*

$$\rho_{k|k-1}(n) = \sum_{j=0}^{n} \rho_{\mathrm{b}}(n-j) S[\mu_{k-1}^{\sharp}, \rho_{k-1}](j) \tag{1.47}$$

for any $n \in \mathbb{N}$ *with*

$$S[\mu, \rho](j) = \sum_{l=j}^{\infty} \binom{l}{j} \frac{\langle p_{\mathrm{s},k}, \mu \rangle^j \langle (1-p_{\mathrm{s},k}), \mu \rangle^{l-j}}{\langle 1, \mu \rangle^l} \rho(l). \tag{1.48}$$

(b) The updated first-order moment measure is found to be

$$\mu_k^\sharp(B) = \mu_k^\phi(B) l_1^\sharp(\phi) + \sum_{z \in Z_k} \frac{\mu_k^z(B)}{s_{\mathrm{c},k}(z)} l_1^\sharp(z) \tag{1.49}$$

with missed detection term (1.36) *and association term* (1.37) *setting* $\bullet = \sharp$*. For any* $n \in \mathbb{N}$*, the updated target cardinality distribution is given by*

$$\rho_k(n) = \frac{\Upsilon^0[\mu_{k|k-1}^\sharp, Z](n)\rho_{k|k-1}(n)}{\langle \Upsilon^0[\mu_{k|k-1}^\sharp, Z], \rho_{k|k-1} \rangle}. \tag{1.50}$$

1.4 Joint parameter and multi-object estimation with hierarchical processes

In many situations, not all filtering parameters are known a priori, especially if the sensor specifications are unknown, e.g. if the sensor is physically moving or if it has to be calibrated in terms of noise level or of its relative position to other sensors. Applications where sensor-specific parameters are unknown include sensor calibration [Houssineau et al., 2016; Ristic and Clark, 2012; Ristic et al., 2012, 2013], drift estimation [Hagen et al., 2016; Schlangen et al., 2016b], Simultaneous Localisation and Mapping (SLAM) [Lee et al., 2013], or clutter rate estimation [Schlangen et al., 2017a]. Instead of estimating the unknown parameters separately, it is possible to formulate a framework that performs the sensor state estimation and the state estimation of the monitored targets in a *joint* manner by regarding the sensor parameter as common behaviour amongst the target group [Swain, 2013].

More formally, let us denote by Ψ the point process describing the hidden parameter y in a suitable space $\mathcal{Y}$ with probability distribution P_Ψ; similarly let Φ be the point process on the space $\mathcal{X}$ describing the multi-target configuration via the probability distributions $P_\Phi^{(n)}$ for every cardinality n of the target population. Their individual PGFLs can be written as

$$G_\Psi(h) = \int h(y) P_\Psi(\mathrm{d}y), \tag{1.51}$$

$$\mathcal{G}_\Phi(g) = \sum_{n \geq 0} \int \left[\prod_{i=1}^n g(x_i) \right] P_\Phi^{(n)}(\mathrm{d}x_{1:n}), \tag{1.52}$$

and hence, the joint PGFL of the processes Ψ and Φ is found to be the composition

$$\mathcal{G}_{\Phi,\Psi}(g,h) = G_\Psi(hG_\Phi(g|\cdot)) = \int h(y) \left[\sum_{n \geq 0} \int \left[\prod_{i=1}^n g(x_i|y) \right] P_\Phi^{(n)}(\mathrm{d}x_{1:n}|y) \right] P_\Psi(\mathrm{d}y). \tag{1.53}$$

This formulation suggests to estimate the distribution P_Ψ of the sensor process and the *conditional* distribution $P_{\Phi|\Psi}$, i.e. the distribution of the multi-target configuration conditioned on the sensor state. Hence, the formulation (1.53) imposes a hierarchy on the joint multi-target and single-parameter process, where the process Ψ is called the *high-level* or *parent process*, and the process Φ is referred to as the *low-level* or *daughter process*. A graphical description is given in Figure 1.2.

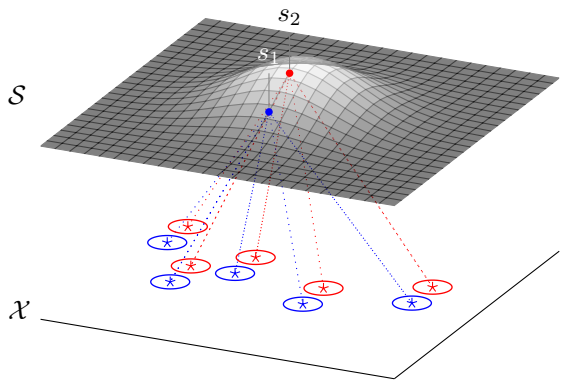


Figure 1.2: Visual description of the structure of the hierarchical filter. Based on the state of the sensor in $\mathcal{S}$, the multi-object configuration is affected accordingly.

1.4.1 The parent process

In the following, the notation $\hat{\cdot}$ will be used to refer to the parent process unless stated otherwise. The functions $\hat{t}_{k|k-1}$ and $\hat{\ell}_k$ denote the Markov transition of the sensor state $y \in \mathcal{Y}$ and the multi-object likelihood based on the underlying filter-specific multi-object estimation, respectively. The probability density $\hat{P}_k$ at time k is propagated in time with the Bayes recursion

$$\hat{P}_{k|k-1}(\mathrm{d}y) = \int \hat{t}_{k|k-1}(\mathrm{d}y|y')\hat{P}_{k-1}(\mathrm{d}y'), \tag{1.54}$$

$$\hat{P}_k(\mathrm{d}y) = \frac{\hat{\ell}_k(y|Z_k)\hat{P}_{k|k-1}(\mathrm{d}y)}{\int \hat{\ell}_k(y'|Z_k)\hat{P}_{k|k-1}(\mathrm{d}y')}. \tag{1.55}$$

1.4.2 The daughter process

For the daughter process, the Markov transition and single-object likelihood functions are denoted by $t_{k|k-1}$ and ℓ_k, respectively. With those, the density P_k of the multi-object state $\varphi \in \mathcal{X}^{n_k}$ at time k is propagated over time with Bayes recursion

$$P_{k|k-1}(\mathrm{d}\varphi|y) = \int t_{k|k-1}(\mathrm{d}\varphi|\varphi', y)P_{k-1}(\mathrm{d}\varphi'|y), \tag{1.56}$$

$$P_k(\mathrm{d}\varphi|y) = \frac{\ell_k(\varphi|Z_k, y)P_{k|k-1}(\mathrm{d}\varphi|y)}{\int \ell_k(\varphi'|Z_k, y)P_{k|k-1}(\mathrm{d}\varphi'|y)}. \tag{1.57}$$

It is possible to use the recursion of the PHD filter for the daughter process, conditioned on the sensor state s like in (1.56) and (1.57) [Swain, 2013]; the only missing link between the parent and the daughter processes is the multi-object likelihood function $\hat{\ell}_k$ that describes the likelihood profile of the sensor state. This function is specific to the daughter process, i.e. it stems from the same assumptions that are used to derive the filter.

1.4.3 The single-cluster PHD filter [Swain, 2013]

In context of parameter estimation, the PHD filter can be easily utilised for multi-object estimation using the following multi-object likelihood function [Swain, 2013].

Proposition 1.4.1 (Multi-object likelihood of the PHD filter [Swain, 2013])**.** *The likelihood func-*

tion of the PHD filter for a given sensor state y is found to be

$$\hat{\ell}_k^{\flat}(y|Z) = \frac{\prod_{z\in Z}\left[\mu_{\mathrm{c},k}(z|y) + \mu_{k|k-1}^{z}(\mathcal{X}|y)\right]}{\exp\left[\int_{\mathcal{Z}} \mu_{\mathrm{c},k}(z|y)\mathrm{d}z + \int_{\mathcal{X}} p_{\mathrm{d},k}(x|y)\mu_{k|k-1}^{\flat}(\mathrm{d}x|y)\right]}. \tag{1.58}$$

Chapter 2

Beyond the PHD filter: generalising the modelling assumptions

The findings of Chapter 1 have described two powerful state-of-the-art multi-object filtering methods, namely the PHD and the CPHD filter, which work well in many real-life scenarios. In terms of computational complexity, the PHD filter is a low-cost method whose complexity depends linearly on the number of measurements; the CPHD filter, on the other hand, has a computational complexity of $\mathcal{O}(n_{\text{card}} \cdot (|Z_k|^2 \log^2 |Z_k|))$ [Delande et al., 2014], i.e. it is dependent on the maximum cardinality n_{card} which is always chosen considerably greater than the average number of measurements. Out of runtime considerations, the PHD filter is therefore the preferred choice; however, the Poisson assumption can be quite restrictive as it will be shown below; here, the CPHD filter gives much greater flexibility in the modelling since it does not pose any assumptions on the form of the cardinality distributions of the target and clutter populations.

In this dissertation, we seek to find an alternative PHD filter recursion that is computationally less demanding than the CPHD filter but whose modelling assumptions are more general than those of the PHD filter. The following section presents the so-called Panjer distribution which encompasses the Poisson, binomial and negative binomial distributions in one single mathematical description [Fackler, 2009]. This distribution accounts for all ratios of mean and variance, which makes it a much more flexible choice than the Poisson distribution whose variance always equals its mean. Moreover, the Panjer distribution depends on two parameters which stand in one-to-one correspondence with the mean and variance. This fact makes it natural to formulate a PHD recursion which not only propagates the first-order moment, but also the variance of the target process as requested by Erdinc et al. [2005]. The considerations about the variance further inspire the exploitation of other second-order information for the analysis of filter properties; we will see in Section 2.4 that statistics like regional correlation and covariance can help to understand a certain weight shift effect called the *spooky effect* which was first noticed on the CPHD filter by Fränken et al. [2009]. Finally, the new filter, as well as the CPHD filter, will be embedded in the single-cluster filtering framework described in Section 1.4.

2.1 Generalising the Poisson process

2.1.1 The Panjer distribution

Before working with the Panjer distribution, let us recall the definitions of the rising and falling factorials, which can be seen as a generalised factorial, and the generalised binomial coefficient for any real value $x \in \mathbb{R}$ and any non-negative integer $n \in \mathbb{N}$.

Definition 2.1.1 (Generalised factorial and binomial coefficient)**.** *Consider a real number $x \in \mathbb{R}$ and a non-negative integer $n \in \mathbb{N}$.*

(i) The Pochhammer symbol *or* rising factorial *$x_{n\uparrow}$ is given by*

$$x_{n\uparrow} := \prod_{i=0}^{n-1}(x+i), \quad x_{0\uparrow} := 1. \tag{2.1}$$

(ii) In the same manner, the falling factorial *$x_{n\downarrow}$ is given by*

$$x_{n\downarrow} := \prod_{i=0}^{n-1}(x-i), \quad x_{0\downarrow} := 1. \tag{2.2}$$

(iii) Using (2.2)*, the* generalised binomial coefficient *$\binom{x}{n}$ is defined as*

$$\binom{x}{n} = \frac{x_{n\downarrow}}{n!}. \tag{2.3}$$

The Panjer process was recently formulated in Schlangen et al. [2017c], inspired by the close relation between the Poisson, binomial and negative binomial distributions found by Panjer [1981]. In the following, let us consider the following definition of the Panjer probability mass function (pmf).

Definition 2.1.2 (Panjer distribution)**.** *For any $n \in \mathbb{N}$, the Panjer distribution is defined via its pmf*

$$\rho_{\text{Panjer}}(n) = \binom{-\alpha}{n}\left(1+\frac{1}{\beta}\right)^{-\alpha}\left(\frac{-1}{\beta+1}\right)^{n}, \tag{2.4}$$

where either $\alpha, \beta \in \mathbb{R}_{>0}$ or $\alpha \in \mathbb{Z}_{<0}$ and $\beta \in \mathbb{R}_{<0}$. Note that negative, non-integer values of α could result in complex values, therefore they are excluded here.

Remark 2.1.3. *A different definition is given in form of the probability mass function (1) in Fackler [2009] which takes the form*

$$\rho(n) = \left(1+\frac{\lambda}{\alpha}\right)^{-\alpha} \frac{\lambda^n}{n!} \prod_{i=0}^{n-1}\frac{\alpha+i}{\alpha+\lambda}. \tag{2.5}$$

To show that the two formulations are in fact equivalent, substitute λ *with* $\frac{\alpha}{\beta}$*:*

$$\rho(n) = \left(1+\frac{1}{\beta}\right)^{-\alpha} \frac{\alpha^n}{\beta^n(\alpha+\frac{\alpha}{\beta})^n} \frac{\prod_{i=0}^{n-1}(\alpha+i)}{n!}(-1)^{2n} \tag{2.6a}$$

$$= \left(1+\frac{1}{\beta}\right)^{-\alpha} \left(\frac{-\alpha}{\alpha\beta+\alpha}\right)^n \frac{\prod_{i=0}^{n-1}(-\alpha-i)}{n!} \tag{2.6b}$$

$$= \left(1+\frac{1}{\beta}\right)^{-\alpha} \left(\frac{-1}{\beta+1}\right)^n \binom{-\alpha}{n}. \tag{2.6c}$$

In (2.6a)*, a factor* $(-1)^{2n} = 1$ *was added to produce the negative values in the terms of Equation* (2.4)*; the equality* (2.6c) *follows from the definition of the generalised binomial coefficient* (2.3)*.*

As mentioned above, this definition encompasses three special cases:

Proposition 2.1.4.

1. *If* $\alpha, \beta \in \mathbb{R}_{>0}$*, the pmf* ρ_{Panjer} *represents the negative binomial distribution.*
2. *If* $\alpha \in \mathbb{Z}_{<0}$ *and* $\beta \in \mathbb{R}_{<0}$*, the pmf* ρ_{Panjer} *represents the binomial distribution.*
3. *If* $\alpha \to \infty$ *with constant* $\lambda = \frac{\alpha}{\beta}$*, expression* (2.4) *converges to the Poisson pmf*

$$\rho_{\text{Poisson}}(n) = e^{-\lambda}\frac{\lambda^n}{n!}. \tag{2.7}$$

Proof. The first two special cases follow directly from the definition of the Panjer pmf, see Fackler [2009]. For the limit case, assume that $\lambda = \frac{\alpha}{\beta}$ is constant and let α grow to ∞:

$$\begin{aligned}
\lim_{\alpha\to\infty} \rho_{\text{Panjer}}(n) &= \lim_{\alpha\to\infty} \binom{-\alpha}{n}\left(1+\frac{1}{\beta}\right)^{-\alpha}\left(\frac{-1}{\beta+1}\right)^n \\
&= \lim_{\alpha\to\infty} \frac{-\alpha_{n\downarrow}}{n!} \exp\left(-\alpha\ln\left(1+\frac{\lambda}{\alpha}\right)\right)\left(-\frac{\lambda}{\alpha+\lambda}\right)^n \\
&= \lim_{\alpha\to\infty} \frac{1}{n!}\frac{\lambda^n\alpha_{n\uparrow}}{(\alpha+\lambda)^n} \exp\left(-\alpha\ln\left(1+\frac{\lambda}{\alpha}\right)\right) \\
&= \lim_{\alpha\to\infty} \frac{\lambda^n}{n!} \underbrace{\frac{\alpha}{\alpha+\lambda}\frac{\alpha+1}{\alpha+\lambda}\cdots\frac{\alpha+n-1}{\alpha+\lambda}}_{=1 \text{ for } \alpha\to\infty} \exp\Big(-\alpha \underbrace{\ln\left(1+\frac{\lambda}{\alpha}\right)}_{\equiv\frac{\lambda}{\alpha} \text{ for } \frac{\lambda}{\alpha}\to 0}\Big) \\
&= \frac{\lambda^n}{n!}\exp(-\lambda).
\end{aligned}$$

□

As shown in Fackler [2009], mean and variance of the Panjer distribution are given by

$$\mu_{\text{Panjer}} = \frac{\alpha}{\beta}, \tag{2.8}$$

$$\text{var}_{\text{Panjer}} = \mu_{\text{Panjer}}\left(1+\frac{1}{\beta}\right). \tag{2.9}$$

It is easy to see from Equation (2.9) that the variance is always greater than the mean if $0 < \beta < \infty$ (which corresponds to the negative binomial case) and less than the mean if $-\infty < \beta < 0$ (which is the binomial case). Together with the Poisson limit where the mean equals the variance, the Panjer distribution makes it therefore possible to describe any ratio of mean and variance which

makes it a convenient modelling choice. The following table summarises the special cases of the Panjer distribution at a glance.

negative binomial	$\alpha, \beta \in \mathbb{R}_{>0}$	$\mu_{\text{Panjer}} < \text{var}_{\text{Panjer}}$
binomial	$\alpha \in \mathbb{Z}_{<0}$ and $\beta \in \mathbb{R}_{<0}$	$\mu_{\text{Panjer}} > \text{var}_{\text{Panjer}}$
Poisson	$\lambda := \frac{\alpha}{\beta} = \text{const.}$ and $\alpha \to \infty$	$\mu_{\text{Panjer}} = \text{var}_{\text{Panjer}}$

Table 2.1: Recovering the Poisson, binomial and negative binomial distributions from the Panjer distribution [Panjer, 1981; Fackler, 2009; Klugman et al., 2012].

In order to demonstrate the different parametrisations visually, Figure 2.1 shows plots of all three cases for the same mean value $\mu_{\text{Panjer}} = 5$. It can be seen that the binomial distribution, having the smallest variance, has the lightest tail, i.e. probabilities of values away from the mean become 0 very quickly and the probability mass highly concentrates around the mean. The negative binomial distribution, on the other hand, has the largest tail, i.e. values far away from the mean still have considerable probability mass, and it has no peak when the variance is very different from the mean. The Poisson distribution builds the intermediate case since it is the limit of both the binomial and negative binomial distributions when $\text{var}_{\text{Panjer}} \to \mu_{\text{Panjer}}$.

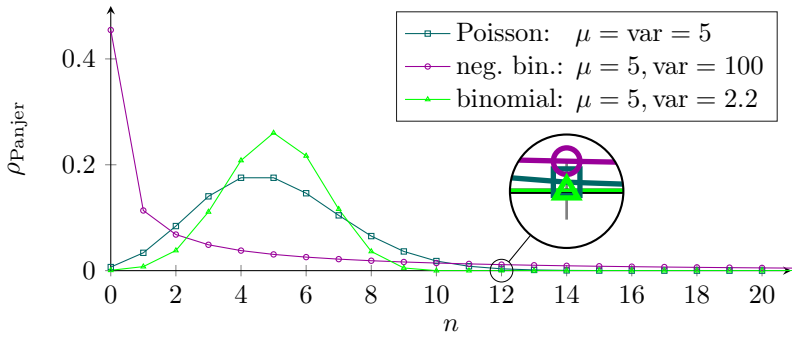


Figure 2.1: The Poisson, binomial and negative binomial distributions with identical mean in comparison. Here, we use the short-hand notation $\mu := \mu_{\text{Panjer}}$ and $\text{var} := \text{var}_{\text{Panjer}}$.

2.1.2 Derivation rules for the potentiation of functionals

In order to define and work with a point process whose cardinality is Panjer distributed, it will be convenient to define some basic differentiation rules in the manner of Equations (1.27) and (1.28), but for the exponentiation of a functional G with a parameter α.

Lemma 2.1.5. *The first-order derivative of the power functional* $(\,\cdot\,)^\alpha$ *in composition with a general functional* $G(h)$ *is the functional*

$$\delta(G^\alpha)(h;\eta) = \alpha G(h)^{\alpha-1}\delta G(h;\eta). \tag{2.10}$$

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

$$(x+y)^\alpha = \sum_{k\geq 0} \binom{\alpha}{k} x^{\alpha-k} y^k. \tag{2.11}$$

with the generalised binomial coefficient $\binom{\alpha}{k}$ as defined in (2.3).

In analogy to Lemma 1.1.12, 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,

$$\begin{aligned}
\delta(G^\alpha)(h;\eta) &\overset{(1.19)}{=} \delta\left((\,\cdot\,)^\alpha\right)(G(h);\delta G(h;\eta)) \\
&= \lim_{n\to\infty}\frac{1}{\varepsilon_n}[(G(h)+\varepsilon_n\eta_n)^\alpha-(G(h))^\alpha] \\
&\overset{(2.11)}{=} \lim_{n\to\infty}\frac{1}{\varepsilon_n}\left[\sum_{j\geq 0}\binom{\alpha}{j}G(h)^{\alpha-j}(\varepsilon_n\eta_n)^j - G(h)^\alpha\right] \\
&= \lim_{n\to\infty}\frac{1}{\varepsilon_n}\left[\sum_{j\geq 1}\binom{\alpha}{j}G(h)^{\alpha-j}(\varepsilon_n\eta_n)^j\right] \\
&= \lim_{n\to\infty}\left[\alpha G(h)^{\alpha-1}\eta_n + \underbrace{\varepsilon_n\sum_{j\geq 2}\binom{\alpha}{j}G(h)^{\alpha-j}\varepsilon_n^{j-2}\eta_n^j}_{=0(n\to\infty)}\right] \\
&= \alpha G(h)^{\alpha-1}\delta G(h;\eta). \qquad \square
\end{aligned}$$

Proposition 2.1.6. *For any linear functional G, the nth-order derivative of the composition $G(h)^\alpha$ can be expressed as*

$$\delta^n(G^\alpha)(h;\eta_1,\ldots,\eta_n) = \alpha_{n\downarrow}\; G(h)^{\alpha-n}\prod_{i=1}^n \delta G(h;\eta_i). \tag{2.12}$$

Proof. Thanks to the structure of (2.10), the nth-order differential of G^α can again be obtained inductively in analogy to Prop. 1.1.13, based on the assumption that G is linear. The derivative of the outer function creates a factor $(\alpha-i+1)$ for each differentiation step. $\square$

2.1.3 The Panjer point process

With the findings of the previous sections, it is now possible to formulate the PGFL of a point process whose cardinality is Panjer distributed.

Proposition 2.1.7 (The Panjer process [Daley and Vere-Jones, 2003; Panjer, 1981; Fackler, 2009; Klugman et al., 2012; Schlangen et al., 2017c])**.** *A Panjer point process with parameters α and β and spatial distribution s is an i.i.d. cluster process with spatial distribution s and cardinality distribution ρ_{Panjer} (2.4) with parameters α and β. Its PGFL is of the form*

$$\mathcal{G}_{\text{Panjer}}(h) = \left(1+\frac{1}{\beta}\int[1-h(x)]s(\mathrm{d}x)\right)^{-\alpha}. \tag{2.13}$$

Proof. Substitute (2.4) for ρ in the definition of an i.i.d. point process (1.24):

$$\mathcal{G}_{\text{Panjer}}(h) \overset{(1.24)}{=} \sum_{n\geq 0} \binom{-\alpha}{n} \left(1+\frac{1}{\beta}\right)^{-\alpha} \left(\frac{-1}{\beta+1}\right)^n \left[\int h(x)s(\mathrm{d}x)\right]^n \tag{2.14a}$$

$$= \left(1+\frac{1}{\beta}\right)^{-\alpha} \sum_{n\geq 0} \binom{-\alpha}{n} \left[\frac{-1}{\beta+1}\int h(x)s(\mathrm{d}x)\right]^n \tag{2.14b}$$

$$= \left(1+\frac{1}{\beta}\right)^{-\alpha} \left[1-\frac{1}{\beta+1}\int h(x)s(\mathrm{d}x)\right]^{-\alpha} \tag{2.14c}$$

$$= \left[1+\frac{1}{\beta}-\frac{1}{\beta}\int h(x)s(\mathrm{d}x)\right]^{-\alpha} \tag{2.14d}$$

$$= \left[1+\frac{1}{\beta}\int [1-h(x)]s(\mathrm{d}x)\right]^{-\alpha}. \tag{2.14e}$$

Equality (2.14c) follows from the binomial theorem (2.11). □

Let us now derive the regional mean and variance of the Panjer process and find an interesting relation between them and the parameters of the process. For this purpose, let $B, B' \in \mathcal{B}(\mathcal{X})$ be two arbitrary regions. The mean $\mu(B)$ is found using Equation (1.16):

$$\mu(B) \overset{(1.16)}{=} \delta\mathcal{G}_{\text{Panjer}}(e^{-f}; \mathbb{1}_B)\Big|_{h=1} \tag{2.15a}$$

$$\overset{(2.13)}{=} \delta\left(\left[1+\frac{1}{\beta}\int [1-e^{-f(x)}]s(\mathrm{d}x)\right]^{-\alpha}; \mathbb{1}_B\right)\Big|_{h=1} \tag{2.15b}$$

$$\overset{(2.12)}{=} \alpha\left[1+\frac{1}{\beta}\int [1-e^0]s(\mathrm{d}x)\right]^{-\alpha-1}\left[\frac{1}{\beta}\int_B e^0 s(\mathrm{d}x)\right] \tag{2.15c}$$

$$= \frac{\alpha}{\beta}\int_B s(\mathrm{d}x). \tag{2.15d}$$

Similarly, the second-order moment $\mu^{(2)}(B\times B')$ is found to be

$$\mu^{(2)}(B\times B') = \delta^2\mathcal{G}_{\text{Panjer}}(e^{-f}; \mathbb{1}_B, \mathbb{1}_{B'})\Big|_{f=0} \tag{2.16a}$$

$$= \frac{\alpha_{2\uparrow}}{\beta^2}\left[1+\frac{1}{\beta}\int [1-e^0]s(\mathrm{d}x)\right]^{-\alpha-2}\int_B e^0 s(\mathrm{d}x)\int_{B'} e^0 s(\mathrm{d}x) \tag{2.16b}$$

$$+ \frac{\alpha}{\beta}\left[1+\frac{1}{\beta}\int [1-e^0]s(\mathrm{d}x)\right]^{-\alpha-1}\int_{B\cap B'} e^0 s(\mathrm{d}x) \tag{2.16c}$$

$$= \frac{\alpha_{2\uparrow}}{\beta^2}\int_B s(\mathrm{d}x)\int_{B'} s(\mathrm{d}x) + \frac{\alpha}{\beta}\int_{B\cap B'} s(\mathrm{d}x). \tag{2.16d}$$

Therefore, the variance $\mathrm{var}(B)$ is found as follows:

$$\mathrm{var}(B) \overset{(1.9)}{=} \mu^{(2)}(B\times B) - [\mu(B)]^2 \tag{2.17a}$$

$$= \mu(B)\left(1+\frac{1}{\beta}\int_B s(\mathrm{d}x)\right). \tag{2.17b}$$

From (2.15) and (2.17) we obtain for $B=\mathcal{X}$ that

$$\mu(\mathcal{X}) = \frac{\alpha}{\beta}, \qquad \mathrm{var}(\mathcal{X}) = \mu(\mathcal{X})\left(1+\frac{1}{\beta}\right). \tag{2.18}$$

These considerations lead to the following relation between mean and variance of a Panjer process and its parameters α and β.

Proposition 2.1.8. *For a Panjer process with intensity $\mu(\cdot)$ and variance $\mathrm{var}(\cdot)$, its parameters α and β are found with the equalities*

$$\alpha = \frac{\mu(\mathcal{X})^2}{\mathrm{var}(\mathcal{X}) - \mu(\mathcal{X})}, \tag{2.19}$$

$$\beta = \frac{\mu(\mathcal{X})}{\mathrm{var}(\mathcal{X}) - \mu(\mathcal{X})}. \tag{2.20}$$

Equations (2.19) and (2.20) show again that the negative binomial case requires the mean to be less than the variance, whereas the binomial case requires the variance to be less than the mean (and the ratio (2.19) to be a negative integer). Together with the Poisson limit case, the Panjer point process can therefore have (almost) all possible values of mean and variance which makes it much more flexible than the Poisson point process. It will be exploited for two generalisations of the PHD filter which will be introduced in the following section.

2.2 Filtering with the Panjer point process

Two novel PHD filter alternatives are proposed below[1] to find a compromise between the relative computational simplicity of the PHD filter and the modelling flexibility of the CPHD filter. They use the *Panjer point process* (see Sec. 2.1.3) to allow for cases where the variance in the number of objects is smaller, greater, or equal to the mean. The first filter that uses this generalised model is the PHD filter with Panjer clutter [Schlangen et al., 2016a], described in Sec. 2.2.1 and denoted by the superscript *; here, the Panjer-specific parameters are understood as the user-defined clutter parameters of the algorithm. The other PHD variation is the Second-Order PHD (SO-PHD) filter [Schlangen et al., 2017c], presented in Sec. 2.2.4 below, which exploits the Panjer distribution for the cardinality of both the predicted target and the false alarm processes; here, the Panjer-specific parameters of the target process are *propagated*, and therefore also its mean and variance due to Proposition 2.1.8. The SO-PHD filter is henceforth marked with the superscript $^\natural$.

2.2.1 The PHD filter with Panjer clutter [Schlangen et al., 2016a]

The PHD filter with Panjer clutter was originally introduced as the *PHD filter with negative binomial clutter* in Schlangen et al. [2016a]. In fact, due to the form of the Panjer distribution, the PGFLs of the Panjer and the negative binomial point process are the same (only differing by the choice of α and β), hence the formulation of Schlangen et al. [2016a] is generalised here without need to change the equations. The proof of the recursion equations is detailed below.

Theorem 2.2.1 (PHD recursion with Panjer clutter [Schlangen et al., 2016a])**.** *Let $\alpha_{\mathrm{c},k}$ and $\beta_{\mathrm{c},k}$ be the Panjer clutter parameters at time k, and let $s_{\mathrm{c},k}$ denote the corresponding spatial distribution of the false alarms.*

[1] These works are published in Schlangen et al. [2016a] and Schlangen et al. [2017c].

(a) Like in the case of the PHD filter prediction (1.38)*, the predicted first-order moment measure is given by*

$$\mu^{\bullet}_{k|k-1}(B) = \mu_{\mathrm{b},k}(B) + \mu_{\mathrm{s},k}(B), \tag{2.21}$$

setting $\bullet = \times$ *in the survival term* (1.35)*.*

(b) The updated first-order moment measure with Poisson prediction and Panjer false alarm model is found to be

$$\mu^{\bullet}_{k}(B) = \mu^{\phi}_{k}(B) + \sum_{z \in Z_k} \frac{\mu^{z}_{k}(B)}{s_{\mathrm{c},k}(z)} \frac{Y(Z_k \backslash \{z\})}{Y(Z_k)} \tag{2.22}$$

with missed detection term (1.36) *and association term* (1.37) *where* $\bullet = \times$*. For any* $Z \subset \mathcal{Z}$*, the terms* $Y(Z)$ *are defined as*

$$Y(Z) = \sum_{j=0}^{|Z|} \frac{(\alpha_{\mathrm{c},k})_{j\uparrow}}{(\beta_{\mathrm{c},k}+1)^{j}} \; e_{|Z|-j}(Z) \tag{2.23}$$

with the so-called elementary symmetric functions *(see Vo et al. [2007])*

$$e_j(Z) = \sum_{\substack{\bar{Z} \subseteq Z \\ |\bar{Z}|=j}} \prod_{z \in \bar{Z}} \frac{\mu^{z}_{k}(\mathcal{X})}{s_{\mathrm{c},k}(z)}. \tag{2.24}$$

Note that the CPHD filter equations contain similar terms to (2.23), but the latter has a much simpler form. In fact, the missed detection term $\mu^{\phi}_{k}(B)$ is not rescaled at all, similarly to the respective term in the PHD filter update (1.39); the association terms, however, are now rescaled with the terms $\dfrac{Y(Z_k \backslash \{z\})}{s_{\mathrm{c},k}(z) Y(Z_k)}$.

The recursion of the PHD filter with Panjer clutter is derived analogously to that of the original PHD filter [Mahler, 2003], hence the proof of the prediction is completely identical to that of the PHD filter since the modified assumption on the false alarm process only affects the update. The proof of the update step, on the other hand, incorporates the following three main steps:

1. Find the joint PGFL $\mathcal{G}_{J,k}$ of the target and measurement processes at time k using a Panjer point process to describe the false positives. It takes the general form

$$\mathcal{G}_{J,k}(g,h) = \mathcal{G}_{k|k-1}\left(\mathcal{G}_{\mathrm{d}}(g,h)\right) \mathcal{G}_{\mathrm{c}}(g) \tag{2.25}$$

   and its detailed formulation is found by applying the filter-specific assumptions. This step is provided in Lemma 2.2.2.

2. Find the conditional PGFL $\mathcal{G}_k$ which describes the observation process using

$$\mathcal{G}_k(h|Z_k) = \frac{\delta^m \mathcal{G}_{J,k}(g,h;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}{\delta^m \mathcal{G}_{J,k}(g,1;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}, \tag{2.26}$$

   which follows from Bayes rule [Mahler, 2003]. This step is performed in Lemma 2.2.3.

3. Find the first-order moment of $\mathcal{G}_k$ which is the numerator of the updated intensity

$$\mu_k(B) = \frac{\delta^{m+1}\mathcal{G}_{J,k}(g,h;\delta_{z_1},\ldots,\delta_{z_m},\mathbb{1}_B)|_{g=0,h=1}}{\delta^m\mathcal{G}_{J,k}(g,1;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}. \tag{2.27}$$

This step is found in Lemma 2.2.4. Note that in general, (1.16) requires the Laplace functional, not the PGFL, however the first factorial and non-factorial moments of a process are equal as stated earlier, so Equation (2.27) is valid.

A few detailed remarks are given in the end of the section to explain the final form of Equation (2.22). Note that exactly the same approach will be used later on in Section 2.2.5.

2.2.2 Derivation of the PHD filter with Panjer clutter

Lemma 2.2.2. *Assume that*

- *the predicted process with PGFL $\mathcal{G}^{\ast}_{k|k-1}$ is a Poisson process;*
- *the detection process with PGFL $\mathcal{G}_{\mathrm{d}}$ is a Bernoulli process;*
- *the clutter process with PGFL $\mathcal{G}_{\mathrm{c}}$ is a Panjer process which is independent of the predicted process.*

Then, the joint target and measurement process for the update step of the PHD filter with Panjer clutter leads to a joint PGFL of the form

$$\mathcal{G}^{\ast}_{J,k}(g,h) = \exp\Big(F_{\mathrm{d}}(g,h)\Big)\Big(F_{\mathrm{c}}(g)\Big)^{-\alpha_{\mathrm{c},k}}, \tag{2.28}$$

where the functionals F_{d} and F_{c} are defined as

$$F_{\mathrm{d}}(g,h) = \int_{\mathcal{X}}\left[h(x)\left(1-p_{\mathrm{d}}(x)+p_{\mathrm{d}}(x)\int_{\mathcal{Z}} g(z)l(x|z)\mathrm{d}z\right)-1\right]\mu^{\ast}_{k|k-1}(x)\mathrm{d}x \tag{2.29}$$

and

$$F_{\mathrm{c}}(g) = 1+\frac{1}{\beta_{\mathrm{c},k}}\int_{\mathcal{Z}}(1-g(z))s_{\mathrm{c},k}(z)\mathrm{d}z. \tag{2.30}$$

Proof. The predicted and the clutter processes are independent, thus their PGFLs are superimposed using multiplication; the detection process, on the other hand, branches from the predicted process, thus the detection and prediction processes are concatenated. The specific form of the functions F_{d} and F_{c} follow from the definitions (1.25), (1.26), and (2.13). □

Lemma 2.2.3. *The mth order derivative of* (2.28) *with respect to g in the directions $\delta_{z_1},\ldots,\delta_{z_m}$ can be written as*

$$\begin{aligned}&\delta^m\mathcal{G}^{\ast}_{J,k}(g,h;\delta_{z_1},\ldots,\delta_{z_m})\\ &\quad=\exp(F_{\mathrm{d}}(g,h))\sum_{j=0}^{|Z_m|}\frac{(\alpha_{\mathrm{c},k})_{j\uparrow}}{\beta^j_{\mathrm{c},k}}F_{\mathrm{c}}(g)^{-\alpha_{\mathrm{c},k}-j}\sum_{\substack{Z\subseteq Z_k\\|Z|=m-j}}\left(\prod_{z\in Z}F^z_{\mathrm{d}}(h)\prod_{z'\in Z_k\setminus Z}s_{\mathrm{c},k}(z')\right),\end{aligned} \tag{2.31}$$

where

$$F_{\mathrm{d}}^{z}(h) \;:=\; \delta F_{\mathrm{d}}(g,h;\delta_z) \;=\; \int_{\mathcal{X}} h(x)p_{\mathrm{d}}(x)l(x|z)\mu_{k|k-1}^{\text{※}}(x)\mathrm{d}x. \tag{2.32}$$

Proof. With the help of the general chain rule (1.21), let us first state the j-fold derivatives of the factors $\mathcal{G}_{\mathrm{d}}(g,h) := \exp\Big(F_{\mathrm{d}}(g,h)\Big)$ and $\mathcal{G}_{\mathrm{c}}(g) := \Big(F_{\mathrm{c}}(g)\Big)^{-\alpha_{\mathrm{c},k}}$ before incorporating the results in the general product rule (1.20). Note that since F_{d} and F_{c} are linear with respect to the argument g, higher-order derivatives disappear. In the following, let $Z \in Z_k$, where $|Z| = j$.

$$\begin{aligned}
\delta^j\mathcal{G}_{\mathrm{d}}(g,h;(\delta_z)_{z\in Z}) &\overset{(1.21)}{=} \sum_{\pi\in\Pi(Z)} \delta^{(|\pi|)} \exp\left(F_{\mathrm{d}}(g,h);\Big(\underbrace{\delta^{(|Z'|)}F_{\mathrm{d}}\big(g,h;(\delta_{z'})_{z'\in Z'}\big)}_{=0 \text{ for } |Z'|>1}\Big)_{Z'\in\pi}\right)\\
&= \delta^j \exp\left(F_{\mathrm{d}}(g,h);\Big(\delta F_{\mathrm{d}}(g,h;\delta_z)\Big)_{z\in Z}\right)\\
&\overset{(1.28)}{=} \exp(F_{\mathrm{d}}(g,h)) \prod_{z\in Z} \delta F_{\mathrm{d}}(g,h;\delta_z)\\
&= \exp(F_{\mathrm{d}}(g,h)) \prod_{z\in Z} F_{\mathrm{d}}^{z}(h).
\end{aligned} \tag{2.33}$$

Here, $\Pi(Z)$ denotes the partition set of Z and π is one possible partition. For the $m-j$th derivative of $\mathcal{G}_{\mathrm{c}}$, define the compement $\bar{Z}$ of $Z \in Z_k$ via $\bar{Z} = Z_k \setminus Z$.

$$\begin{aligned}
\delta^{m-j}\mathcal{G}_{\mathrm{c}}(g;(\delta_z)_{z\in\bar{Z}}) &\overset{(1.21)}{=} \sum_{\pi\in\Pi(\bar{Z})} \delta^{(|\pi|)}\left(F_{\mathrm{c}}(g)^{-\alpha_{\mathrm{c},k}};\Big(\underbrace{\delta^{(|Z'|)}F_{\mathrm{c}}(g;(\delta_{z'})_{z'\in Z'})}_{=0 \text{ for } |Z'|>1}\Big)_{Z'\in\pi}\right)\\
&= \delta^{m-j}\left(F_{\mathrm{c}}(g)^{-\alpha_{\mathrm{c},k}};\Big(\delta F_{\mathrm{c}}(g;\delta_z)\Big)_{z\in\bar{Z}}\right)\\
&\overset{(2.12)}{=} \frac{(\alpha_{\mathrm{c},k})_{j\uparrow}}{\beta_{\mathrm{c},k}^{j}} F_{\mathrm{c}}(g)^{-\alpha_{\mathrm{c},k}-j} \prod_{z\in Z} s_{\mathrm{c},k}(z).
\end{aligned} \tag{2.34}$$

The desired result is obtained by inserting (2.33) and (2.34) into the product rule (1.20). □

Lemma 2.2.4. *Let us write $\alpha := \alpha_{\mathrm{c},k}$ and $\beta := \beta_{\mathrm{c},k}$ for the sake of brevity. The first-order derivative of (2.31) with respect to h in the direction $\mathbb{1}_B$ for some region $B \in \mathcal{B}(\mathcal{X})$ is found to be*

$$\begin{aligned}
&\delta^{m+1}\mathcal{G}_{J,k}^{\text{※}}(g,h;\delta_{z_1},\dots,\delta_{z_m},\mathbb{1}_B)\\
&= \exp(F_{\mathrm{d}}(g,h))\Bigg[\delta F_{\mathrm{d}}(g,h;\mathbb{1}_B)\sum_{j=0}^{m}\frac{\alpha_{j\uparrow}}{\beta^j}F_{\mathrm{c}}(g)^{-\alpha-j}\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}s_{\mathrm{c},k}(z)\prod_{z'\in Z_m\setminus Z}F_{\mathrm{d}}^{z'}(h)\\
&\quad+\sum_{j=0}^{m}\frac{\alpha_{j\uparrow}}{\beta^j}F_{\mathrm{c}}(g)^{-\alpha-j}\prod_{z\in Z}s_{\mathrm{c},k}(z)\Bigg(\sum_{z'\in Z_k\setminus Z}\mu_k^{z'}(x)\prod_{\substack{z''\in Z_k\setminus Z\\z''\neq z'}}F_{\mathrm{d}}^{z''}(h)\Bigg)\Bigg]
\end{aligned} \tag{2.35}$$

with

$$\delta F_{\mathrm{d}}(g,h;\mathbb{1}_B) = \int_B \left(1-p_{\mathrm{d},k}(x)+p_{\mathrm{d},k}(x)\int_{\mathbb{Z}} g(z)l_k(x|z)\mathrm{d}z\right)\mu_{k|k-1}^{\text{※}}(x)dx. \tag{2.36}$$

Proof. Use the product rule (1.18) on Equation (2.31). □

Proof of Theorem 2.2.1. Lemmata 2.2.3 and 2.2.4 provide the denominator and the numerator of 2.27, respectively, when setting $h = 1$ and $g = 0$. The terms $\exp(F_{\mathrm{d}}(0,1))$ and $F_{\mathrm{c}}(0)^{-\alpha}$ cancel

out in the fraction, and expanding with $\left(\prod_{z\in Z_m} s_{\mathrm{c},k}(z)\right)^{-1}$ eliminates the second product in the elementary symmetric functions. The first sum in (2.71) forms the missed detection term, whereas the second sum describes the associations for each measurement (after rearranging the sums). Furthermore, the term $F_{\mathrm{c}}(0) = (1 + \frac{1}{\beta})$ simplifies with the term $\frac{\alpha_{j\uparrow}}{\beta^j}$. □

2.2.3 Simulations with the PHD filter with Panjer clutter [Schlangen et al., 2016a]

To compare the performance of the new filter with the original PHD filter in presence of spontaneous bursts of false alarms, two artificial datasets are considered below which provide ground truth. The data was simulated on a two-dimensional, square state space of size 50 m × 50 m with a time lapse of 1 s between each pair of consecutive frames.[2] Targets were generated independently and identically distributed according to a Poisson birth model with mean 0.5, where between 0 and 5 targets were introduced in the first frame. The initial target velocity was set to 0 with a small white noise with standard deviation $0.5\,\mathrm{m\,s^{-1}}$ in both image dimensions. The targets were propagated over time according to a near-constant velocity model, assuming a small white noise with standard deviation 0.5 m and $0.01\,\mathrm{m\,s^{-1}}$ on each dimension of the position and velocity, respectively. The survival rate was set to $p_{\mathrm{s}} = 0.99$ for the whole experiment.

For the simulation of the detection process, the probability of detection was set to $p_{\mathrm{d}} = 0.9$ and a white measurement noise with standard deviation 0.4 m was imposed. The two considered scenarios differ from each other as follows:

S1.1 The first scenario was simulated over 15 time steps, where exactly 9 false alarms were randomly placed in the scene in the first 14 time steps. At time 15, the number of false alarms was set artificially to values between 0 and 130 to simulate bursts of clutter with varying intensity.

S1.2 In the second scenario, the false alarms were generated according to a negative binomial point process with mean 9.5 and variance 100. Data was simulated across 100 time instances to analyse the long-term effect of false alarms with high variance in cardinality.

Both the original PHD filter and the proposed method were initialised with the same parameters that were used to create the simulated data, assuming a slightly higher birth intensity of 10 for the first frame. Due to its nature, the PHD filter with Panjer clutter is able to incorporate the information about the variance of the clutter process. The clutter variance of the original PHD filter, on the other hand, is bound to the value of the assumed clutter intensity, therefore it was initialised both with a mean of 9.5 and 50 to evaluate if different intensities improve the estimation result. The corresponding distributions used by both filters are shown in Fig. 2.2. For all experiments, a Gaussian Mixture (GM) implementation[3] was chosen for both methods in the manner of Vo and Ma [2006], where the standard deviation of the process and measurement noise of the underlying Kalman filters was set to 0.1 m and 0.4 m, respectively. The fact that both filters are implemented with a GM approach and initialised with the same parameters suggests that the

[2] Note that the choice of units is arbitrary.
[3] A description of Gaussian Mixture implementations and the pseudocode are found in the appendix.

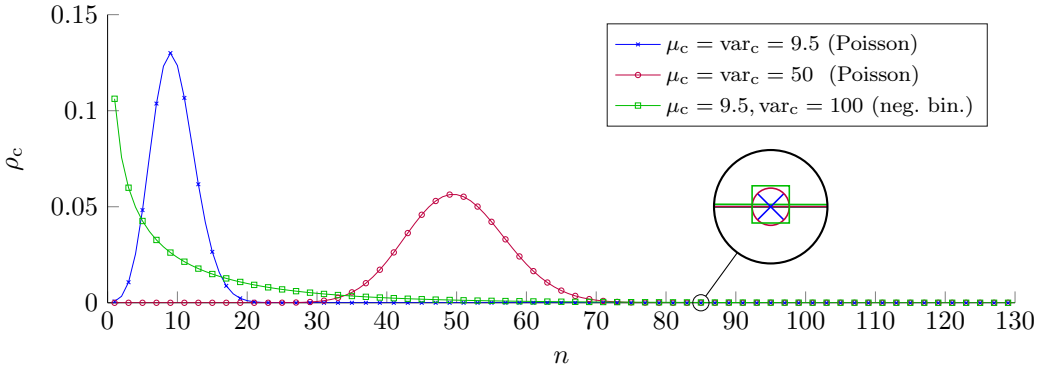


Figure 2.2: The different clutter distributions ρ_c used in the experiment. As shown in the magnified area, the negative binomial distribution still carries a significantly positive probability mass even beyond $n = 80$, whereas both Poisson distributions already reached mass 0.

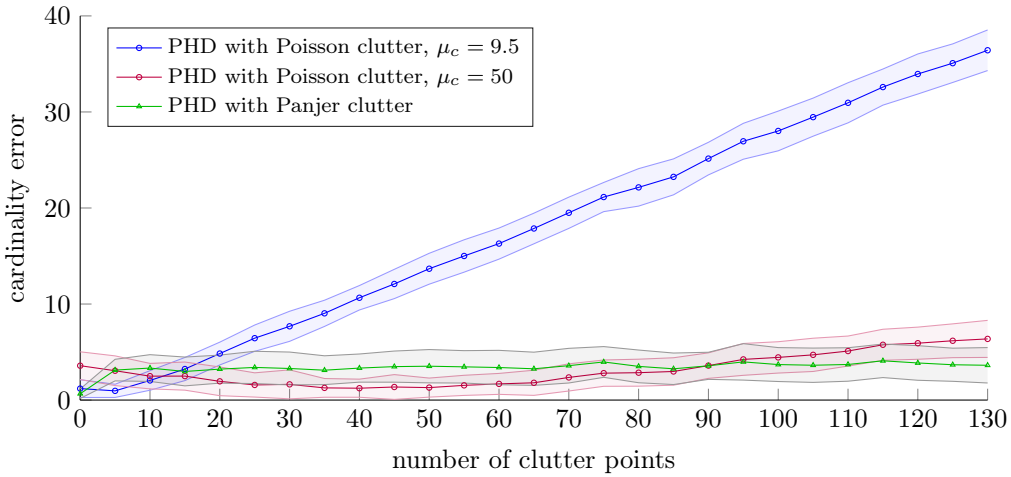


Figure 2.3: Simulation S1.1: Estimation error of both algorithms at time 15 dependent on the present number of clutter points. The results are averaged over 100 Monte Carlo runs for each false alarm number and plotted along with the variance over those runs.

localisation error of the objects are very similar, therefore Figures 2.3 and 2.4 show only the error in the cardinality estimate, averaged over 100 and 500 Monte Carlo (MC) runs, respectively.

Fig. 2.3 shows the errors in the estimated target number for Scenario 1 over the simulated number of clutter points at time 15, together with the variance over the 100 MC runs. It can be noted that the PHD filter with smaller mean performs best for small amounts of clutter close to the assumed mean 9.5 since they are in accordance with the model, but cannot cope with high amounts of false alarms which are miss-interpreted as new-born targets. The PHD filter that assumed the high clutter intensity also performs best when the number of false alarms falls within a confidence interval around the mode but shows increasing errors in the estimated number of targets when the amount of false alarms falls out of the model. The proposed filter, in contrast, shows a consistent performance for any amount of occurring false alarms.

Fig. 2.4 displays the mean and variance of the errors in the estimated target number for Scenario 2 over all 100 time steps, averaged over 500 MC runs. It can be seen that the PHD filter with the low clutter intensity shows by far the highest variance in performance since it is unable to cope with sudden bursts of false alarms as seen above. Still, the PHD filter with the high clutter intensity performs worse on average since in most time scans, the actual number of false alarms per frame

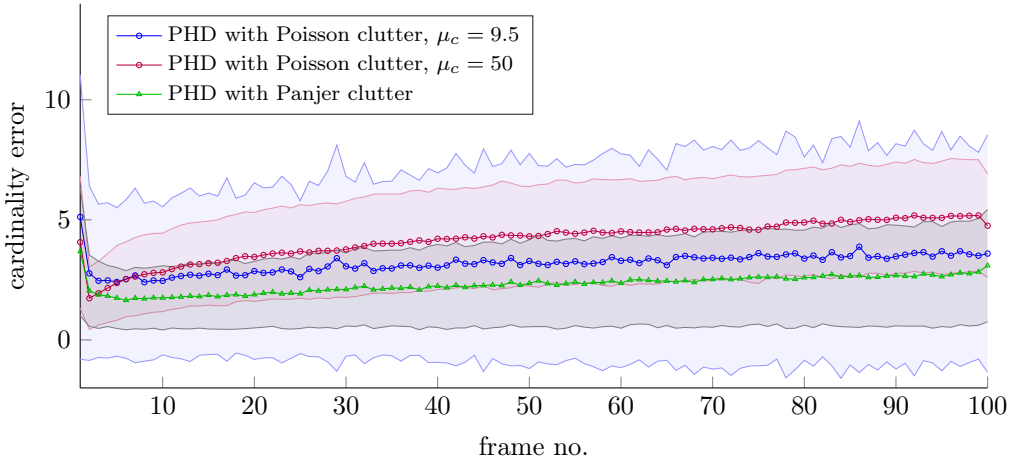


Figure 2.4: Simulation S1.2: Mean and standard deviation over 500 Monte Carlo runs of the cardinality errors obtained with both algorithms.

is much lower than the assumed value of 50. The proposed filter, in contrast, consistently shows the lowest estimation error in cardinality since the negative binomial occurrence of false alarms is well modelled with the Panjer assumption.

In terms of runtime, the proposed filter took on average 8.0 s for one iteration of Scenario S1.2, whereas the original PHD filter took 5.7 s on a dual-core Dell Precision M4800 workstation with Intel(R) Core(TM) i7-4710MQ CPU @ 2.50GHz using a Matlab implementation.

2.2.4 The Second-Order PHD filter [Schlangen et al., 2017c]

The promising results of Sections 2.2.1 - 2.2.3 inspired the idea to change the predicted target process of the PHD filter to a Panjer process as well. We have seen in Equations (2.19) and (2.20) that the mean and variance of a Panjer point process stand in one-to-one correspondence with its two parameters α and β. As a pleasant side effect of this, it is possible to derive a PHD filter recursion which propagates both mean and variance of the target process through the Panjer assumption. The data flow of the proposed recursion is depicted in Figure 3.1; it can be seen that mean and variance are directly propagated in the prediction, whereas the predicted Panjer parameters $\alpha_{k|k-1}$ and $\beta_{k|k-1}$ are calculated from mean and variance for the update equations. Note that this filter only propagates the variance *on the full state space* $\mathcal{X}$, however regional statistics can still be extracted at all times. The proposed filter will be called the *Second-Order PHD (SO-PHD) filter* (or alternatively the *Panjer PHD filter* or simply the *Panjer filter*) since it propagates second-order information using a Panjer process to describe the target and clutter populations. It was introduced by Schlangen et al. [2017c] and is formulated below using the same notation. Note that a similar idea was attempted in Mahler [2006] in form of the so-called *binomial filter* (without proof) which uses the binomial assumption for the target process twice (i.e. once before both the prediction and the update steps) and models the false alarms with a Poisson process. However, it was acknowledged that the practicality is restricted due to the constraint on the admissible parameters (recall that α is restricted to the negative integers in that case); furthermore, the binomial assumption requires a lot of certainty in the target process due to the restricted variance.

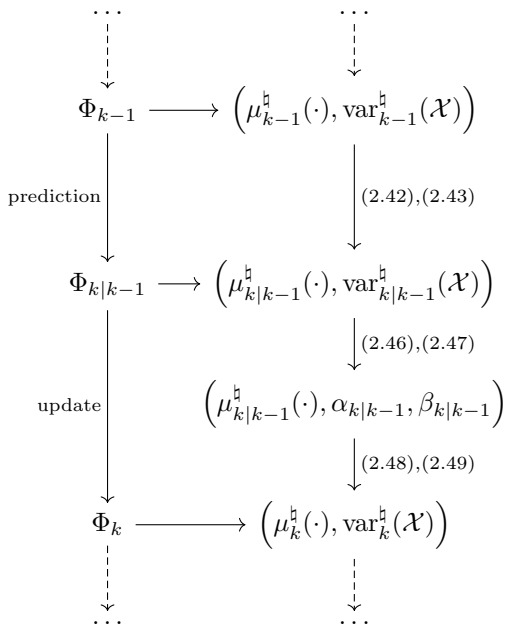


Figure 2.5: Data flow of the second-order PHD filter at time k. In addition to the intensity function μ it propagates the scalar $\mathrm{var}(\mathcal{X})$, describing the variance in the estimated number of targets on the whole state space.

Since the recursion of the SO-PHD filter are more involved through the additional propagation of the process variance, let us state the prediction and update equations separately. As mentioned earlier, the notation $\natural$ will be used below to refer to the SO-PHD filter. Let $\alpha_{k|k-1}, \beta_{k|k-1}$ and $\alpha_{\mathrm{c},k}, \beta_{\mathrm{c},k}$ be the parameters of the predicted target and clutter processes at time k, respectively. Define the terms

$$Y_u(Z) := \sum_{j=0}^{|Z|} \frac{(\alpha_{k|k-1})_{(j+u)\uparrow}}{(\beta_{k|k-1})^{j+u}} \frac{(\alpha_{\mathrm{c},k})_{(|Z|-j)\uparrow}}{(\beta_{\mathrm{c},k}+1)^{|Z|-j}} F_{\mathrm{d}}^{-j-u} e_j(Z) \tag{2.37}$$

for any $Z \subseteq Z_k$, where F_{d} is the scalar given by

$$F_{\mathrm{d}} := \int \left[1 + \frac{p_{\mathrm{d},k}(x)}{\beta_{k|k-1}}\right] \mu^{\natural}_{k|k-1}(\mathrm{d}x), \tag{2.38}$$

and e_j denotes the elementary symmetric functions (cf. Eq. (2.24))

$$e_j(Z) := \sum_{\substack{Z' \subseteq Z \\ |Z'|=j}} \prod_{z \in Z'} \frac{\mu^z_k(\mathcal{X})}{s_{\mathrm{c},k}(z)}, \tag{2.39}$$

where $s_{\mathrm{c},k}$ is the spatial false alarm distribution at time k and the association term is defined via (1.37) with $\bullet = \natural$.[4] Furthermore, define the expression $l^{\natural}_d$ for $d = 1, 2$ via

$$l^{\natural}_d(\phi) := \frac{Y_d(Z_k)}{Y_0(Z_k)} \quad \text{and} \quad l^{\natural}_d(z) := \frac{Y_d(Z_k \backslash \{z\})}{Y_0(Z_k)}. \tag{2.40}$$

[4]Note that the elementary symmetric functions of the PHD filter with Panjer clutter are identical (up to $\bullet = \natural$), cf. Equation (2.24).

In a similar manner, define

$$l_2^{\natural,\neq}(z,z') := \begin{cases} \dfrac{Y_2(Z_k \backslash \{z,z'\})}{Y_0(Z_k)} & \text{if} \quad z \neq z', \\ 0 & \text{otherwise.} \end{cases} \tag{2.41}$$

The prediction of the variance involves the second-order factorial moment $\nu_k^{(2)}$ which, in general, cannot be retrieved from the predicted information $\mu_{k|k-1}^{\natural}, \mathrm{var}_{k|k-1}^{\natural}$ only. The assumption that $p_{\mathrm{s},k}(x) = p_{\mathrm{s},k}$ shall be uniform for all x in the state space $\mathcal{X}$, however, facilitates the formulation of a closed-form recursion. Theorem 2.2.5 states both the general and the simplified prediction equations.

Theorem 2.2.5 (Second-order PHD recursion, prediction [Schlangen et al., 2017c])**.**

(a) In the manner of (1.38) *and* (1.46)*, the predicted first-order moment measure of the Panjer filter is given by*

$$\mu_{k|k-1}^{\natural}(B) = \mu_{\mathrm{b},k}(B) + \mu_{\mathrm{s},k}(B), \tag{2.42}$$

using $\bullet = \natural$ *in* (1.35)*. The predicted variance in the whole state space* $\mathcal{X}$ *is given by*

$$\mathrm{var}_{k|k-1}^{\natural}(\mathcal{X}) = \mathrm{var}_{\mathrm{b},k}(\mathcal{X}) + \mathrm{var}_{\mathrm{s},k}(\mathcal{X}), \tag{2.43}$$

where $\mathrm{var}_{\mathrm{b},k}$ *denotes the variance of the birth process and* $\mathrm{var}_{\mathrm{s},k}$ *is the variance of the predicted process concerning the surviving targets which is found to be*

$$\mathrm{var}_{\mathrm{s},k}(B) = \mu_{\mathrm{s},k}(B)\Big[1 - \mu_{\mathrm{s},k}(B)\Big] + \int p_{\mathrm{s},k}(x) p_{\mathrm{s},k}(x') t_{k|k-1}(B|x) t_{k|k-1}(B|x') \nu_{k-1}^{(2)}(\mathrm{d}(x,x')), \tag{2.44}$$

again setting $\bullet = \natural$ *in* (1.35)*.*

(b) Assume additionally that $p_{\mathrm{s},k}(x) = p_{\mathrm{s},k}$ *is constant for all* $x \in \mathcal{X}$ *at time* k*. Then, the variance of the predicted process in* $\mathcal{X}$ *reduces to the expression*

$$\mathrm{var}_{\mathrm{s},k}(\mathcal{X}) = p_{\mathrm{s},k}^2 \mathrm{var}_{k-1}(\mathcal{X}) + p_{\mathrm{s},k}[1 - p_{\mathrm{s},k}]\mu_{k-1}(\mathcal{X}). \tag{2.45}$$

Theorem 2.2.6 (Second-order PHD recursion, update [Schlangen et al., 2017c])**.** *With the help of Equations* (2.19) *and* (2.20)*, find* $\alpha_{k|k-1}$ *and* $\beta_{k|k-1}$ *using the expressions*

$$\alpha_{k|k-1} = \frac{\mu_{k|k-1}^{\natural}(\mathcal{X})^2}{\mathrm{var}_{k|k-1}^{\natural}(\mathcal{X}) - \mu_{k|k-1}^{\natural}(\mathcal{X})}, \tag{2.46}$$

$$\beta_{k|k-1} = \frac{\mu_{k|k-1}^{\natural}(\mathcal{X})}{\mathrm{var}_{k|k-1}^{\natural}(\mathcal{X}) - \mu_{k|k-1}^{\natural}(\mathcal{X})}. \tag{2.47}$$

Then, the updated first-order moment measure becomes

$$\mu_k^{\natural}(B) = \mu_k^{\phi}(B) l_1^{\natural}(\phi) + \sum_{z \in Z_k} \frac{\mu_k^z(B)}{s_{\mathrm{c},k}(z)} l_1^{\natural}(z), \tag{2.48}$$

and the updated variance is obtained with

$$\begin{aligned}\mathrm{var}_k^{\natural}(B) = \mu_k^{\natural}(B) + \mu_k^{\phi}(B)^2 \left[l_2^{\natural}(\phi) - l_1^{\natural}(\phi)^2\right] + 2\mu_k^{\phi}(B) \sum_{z \in Z_k} \frac{\mu_k^z(B)}{s_{\mathrm{c},k}(z)} \left[l_2^{\natural}(z) - l_1^{\natural}(\phi) l_1^{\natural}(z)\right] \\ + \sum_{z,z' \in Z_k} \frac{\mu_k^z(B)}{s_{\mathrm{c},k}(z)} \frac{\mu_k^{z'}(B)}{s_{\mathrm{c},k}(z')} \left[l_2^{\natural,\neq}(z,z') - \hat{l}_1^{\natural}(z)\hat{l}_1(z')\right].\end{aligned} \tag{2.49}$$

In Section 2.1, it was already mentioned that the Poisson distribution is a limit case of the Panjer distribution for constant $\lambda = \frac{\alpha}{\beta}$ and $\alpha \to \infty$; hence, it was already shown that the PHD filter recursion can be recovered from the PHD filter with Panjer clutter. The following corollary shows how the intensity update of the SO-PHD filter is connected to those of the original PHD filter and the PHD filter with Panjer clutter.

Corollary 2.2.7 (Limit cases [Schlangen et al., 2016a, 2017c])**.** *If 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* (2.48) *converges to the intensity update of the PHD filter with Panjer clutter* (2.22)*. Furthermore, if the clutter process is assumed Poisson as well, i.e., $\alpha_{\mathrm{c},k}, \beta_{\mathrm{c},k} \to \infty$ with constant ratio $\lambda_{\mathrm{c},k} = \frac{\alpha_{\mathrm{c},k}}{\beta_{\mathrm{c},k}}$, then the intensity update* (2.48) *of the SO-PHD filter converges to the intensity update* (1.39) *of the original PHD filter.*

Remark 2.2.8. *Note, however, that since the PHD filter with Panjer clutter only propagates the mean of the target process, it can be seen as a generalisation of the PHD filter, whereas the SO-PHD filter (just like the CPHD filter) propagates additional information and is hence conceptually different.*

2.2.5 Derivation of the SO-PHD filter

Proof of Theorem 2.2.5. (a) In the following, we denote by $\mathcal{G}_{\mathrm{s},k}$ the PGFL of the point process describing the evolution of a target, given that it survives from the previous to the present time step, and let us denote the birth point process by $\mathcal{G}_{\mathrm{b},k}$. For the sake of simplicity, the time subscripts on the birth and surviving target processes will be omitted in this proof.

First, let us write down the PGFL $\mathcal{G}_{k|k-1}^{\natural}$ of the prediction process which takes the form

$$\mathcal{G}_{k|k-1}^{\natural}(h) = \mathcal{G}_{\mathrm{b}}(h)\mathcal{G}_{k-1}^{\natural}(\mathcal{G}_{\mathrm{s}}(h|\cdot)). \tag{2.50}$$

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 branches from the prior target process Φ_{k-1} [Daley and Vere-Jones, 2003, Eq. 5.5.18]. Apart from the assumption that $\mathcal{G}_{\mathrm{s}}$ is a Bernoulli process and that the birth process is independent of the target process, *no additional assumptions have to be made up to this point* (in contrast to the approach of Mahler [2006] which requires $\mathcal{G}_{k-1}$ to be a binomial process). Therefore, the predicted intensity $\mu_{k|k-1}^{\natural}$ is found to be identical to the predicted intensity of the PHD filter in Equation (1.38).

Following Equation (1.9), the second-order moment of $\mathcal{L}_{k|k-1}$ is computed and the square of

the predicted intensity (1.38) be subtracted from the result to obtain the predicted variance $\mathrm{var}_{k|k-1}$. Therefore, the first step is to construct the second-order moment $\mu^{(2)}_{k|k-1}(B \times B')$ in arbitrary regions $B, B' \in \mathcal{B}(\mathcal{X})$ using Equation (1.16):

$$\mu^{(2)}_{k|k-1}(B \times B') = \delta^2 \mathcal{L}^{\natural}_{k|k-1}(f; \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0} \tag{2.51a}$$

$$= \delta^2 \mathcal{G}^{\natural}_{k|k-1}(e^{-f}; \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0}. \tag{2.51b}$$

The product rule (1.18) leads to the expression

$$\begin{aligned}
\mu^{(2)}_{k|k-1}(B \times B') \overset{(1.18)}{=}\ & \delta^2 \mathcal{G}_{\mathrm{b}}(e^{-f}; \mathbb{1}_B; \mathbb{1}_{B'})\big|_{f=0} \mathcal{G}^{\natural}_{k-1}(\mathcal{G}_{\mathrm{s}}(1|\cdot)) \\
& + \delta \mathcal{G}_{\mathrm{b}}(e^{-f}; \mathbb{1}_B)\big|_{f=0} \delta(\mathcal{G}^{\natural}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_{B'})\big|_{f=0} \\
& + \delta \mathcal{G}_{\mathrm{b}}(e^{-f}; \mathbb{1}_{B'})\big|_{f=0} \delta(\mathcal{G}^{\natural}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B)\big|_{f=0} \\
& + \mathcal{G}_{\mathrm{b}}(1) \delta^2(\mathcal{G}^{\natural}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B, \mathbb{1}_{B'})\big|_{f=0} \qquad (2.51c) \\
\overset{(1.16)}{=}\ & \mu^{(2)}_{\mathrm{b}}(B \times B') \\
& - \mu_{\mathrm{b}}(B) \delta(\mathcal{G}^{\natural}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_{B'})\big|_{f=0} \\
& - \mu_{\mathrm{b}}(B') \delta(\mathcal{G}^{\natural}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B)\big|_{f=0} \\
& + \delta^2(\mathcal{G}^{\natural}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B, \mathbb{1}_{B'})|_{f=0}, \qquad (2.51d)
\end{aligned}$$

where μ_{b} and $\mu^{(2)}_{\mathrm{b}}$ are the first- and second-order moment measures of the birth process, respectively. Let us first compute $\delta(\mathcal{G}^{\natural}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B)\big|_{f=0}$ in (2.51d). The general definition of PGFLs (1.12) leads to the expression

$$\delta(\mathcal{G}^{\natural}_{k-1}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot)); \mathbb{1}_B)\big|_{f=0} \overset{(1.12)}{=} \sum_{n \geq 0} \int_{\mathcal{X}^n} \delta\left(\left[\prod_{i=1}^{n} \mathcal{G}_{\mathrm{s}}(e^{-f}|x_i) \right]; \mathbb{1}_B \right) \Bigg|_{f=0} P^{(n)}_{k-1}(\mathrm{d}x_{1:n}) \tag{2.52a}$$

$$\overset{(1.18)}{=} \sum_{n \geq 0} \int_{\mathcal{X}^n} \sum_{i=1}^{n} \delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x_i); \mathbb{1}_B)\big|_{f=0} P^{(n)}_{k-1}(\mathrm{d}x_{1:n}) \tag{2.52b}$$

$$\overset{(1.2a)}{=} \int \delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x); \mathbb{1}_B)\big|_{f=0} \mu_{k-1}(\mathrm{d}x). \tag{2.52c}$$

As mentioned previously, the survival process for a target with prior state x is modelled with a Bernoulli point process with (state-dependent) parameter $p_{\mathrm{s}}(x)$ and spatial distribution $t(\cdot|x)$, therefore (1.25) yields

$$\mathcal{G}_{\mathrm{s}}(e^{-f}|x) = 1 - p_{\mathrm{s}}(x) + p_{\mathrm{s}}(x) \int e^{-f(y)} t(\mathrm{d}y|x) \tag{2.53}$$

whose derivative is found to be

$$\delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x); \mathbb{1}_B) = p_{\mathrm{s}}(x) \int \delta(e^{-f(y)}; \mathbb{1}_B) t(\mathrm{d}y|x) \tag{2.54a}$$

$$= -p_{\mathrm{s}}(x) \int \mathbb{1}_B(y) e^{-f(y)} t(\mathrm{d}y|x) \tag{2.54b}$$

which leads to

$$\begin{aligned}\delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x);\mathbb{1}_B)\big|_{f=0} &= \delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x);\mathbb{1}_B)\big|_{f=0} \\ &= -p_{\mathrm{s}}(x)t(B|x).\end{aligned} \tag{2.54c}$$

Substituting (2.54c) in (2.52c) yields

$$\delta(\mathcal{G}_{k-1}^{\natural}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot));\mathbb{1}_B)\big|_{f=0} = -\int p_{\mathrm{s}}(x)t(B|x)\mu_{k-1}(\mathrm{d}x). \tag{2.55}$$

In a similar manner, let us write out the expression $\delta^2(\mathcal{G}_{k-1}^{\natural}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot));\mathbb{1}_B,\mathbb{1}_{B'})|_{f=0}$ in (2.51d), again using the general definition (1.12) of PGFLs:

$$\delta^2(\mathcal{G}_{k-1}^{\natural}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot));\mathbb{1}_B,\mathbb{1}_{B'})|_{f=0} \tag{2.56a}$$

$$= \sum_{n\geq 0}\int_{\mathcal{X}^n}\delta^2\left(\left[\prod_{i=1}^{n}\mathcal{G}_{\mathrm{s}}(e^{-f}|x_i)\right];\mathbb{1}_B,\mathbb{1}_{B'}\right)\Bigg|_{f=0} P_{k-1}^{(n)}(\mathrm{d}x_{1:n}) \tag{2.56b}$$

$$\begin{aligned}&\overset{(1.18)}{=} \sum_{n\geq 0}\int_{\mathcal{X}^n}\sum_{i=1}^{n}\delta^2(\mathcal{G}_{\mathrm{s}}(e^{-f}|x_i);\mathbb{1}_B,\mathbb{1}_{B'})\big|_{f=0} P_{k-1}^{(n)}(\mathrm{d}x_{1:n}) \\ &\quad + \sum_{n\geq 0}\int_{\mathcal{X}^n}\sum_{\substack{1\leq i,j\leq n\\ i\neq j}}\delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x_i);\mathbb{1}_B)\big|_{f=0}\delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x_j);\mathbb{1}_{B'})\big|_{f=0}P_{k-1}^{(n)}(\mathrm{d}x_{1:n})\end{aligned} \tag{2.56c}$$

$$\begin{aligned}&= \int\delta^2(\mathcal{G}_{\mathrm{s}}(e^{-f}|x);\mathbb{1}_B,\mathbb{1}_{B'})\big|_{f=0}\mu_{k-1}(\mathrm{d}x) \\ &\quad + \int\delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x);\mathbb{1}_B)\big|_{f=0}\delta(\mathcal{G}_{\mathrm{s}}(e^{-f}|x');\mathbb{1}_{B'})\big|_{f=0}\nu_{k-1}^{(2)}(\mathrm{d}(x,x')).\end{aligned} \tag{2.56d}$$

It can be deducted with the help of Equation (2.54) that

$$\delta^2(\mathcal{G}_{\mathrm{s}}(e^{-f}|x);\mathbb{1}_B,\mathbb{1}_{B'})\big|_{f=0} = p_{\mathrm{s}}(x)t(B\cap B'|x), \tag{2.57}$$

so that (2.56d) simplifies to the expression

$$\delta^2(\mathcal{G}_{k-1}^{\natural}(\mathcal{G}_{\mathrm{s}}(e^{-f}|\cdot));\mathbb{1}_B,\mathbb{1}_{B'})|_{f=0} = \mu_{\mathrm{s}}(B\cap B') + \int p_{\mathrm{s}}(x)t(B|x)p_{\mathrm{s}}(x')t(B'|x')\nu_{k-1}^{(2)}(\mathrm{d}(x,x')). \tag{2.58}$$

The second-order moment for $B = B'$ is found by inserting (2.55) and (2.58) in (2.51d):

$$\begin{aligned}\mu_{k|k-1}^{(2)}(B\times B) = \mu_{\mathrm{b}}^{(2)}(B\times B) &+ 2\mu_{\mathrm{b}}(B)\mu_{\mathrm{s}}(B) + \mu_{\mathrm{s}}(B) \\ &+ \int p_{\mathrm{s}}(x)t(B|x)p_{\mathrm{s}}(x')t(B|x')\nu_{k-1}^{(2)}(\mathrm{d}(x,x')).\end{aligned} \tag{2.59}$$

Finally, Equation (1.9) leads to

$$\begin{aligned}\mathrm{var}_{k|k-1}^{\natural}(B) = \mathrm{var}_{\mathrm{b}}(B) &+ [\mu_{\mathrm{b}}(B)]^2 - [\mu_{k|k-1}^{\natural}(B)]^2 \\ &+ 2\mu_{\mathrm{b}}(B)\mu_{\mathrm{s}}(B) + \mu_{\mathrm{s}}(B) + \int p_{\mathrm{s}}(x)t(B|x)p_{\mathrm{s}}(x')t(B|x')\nu_{k-1}^{(2)}(\mathrm{d}(x,x')),\end{aligned} \tag{2.60}$$

and the predicted variance is finally obtained by substituting the predicted intensity (1.38) in

Equation (2.60).

(b) Let us simplify the predicted variance assuming that $p_{\mathrm{s},k}$ is uniform over the state space. Under this assumption, Equation (1.35) with $B = \mathcal{X}$ and $\bullet = \natural$ simplifies to

$$\mu_{\mathrm{s},k}(\mathcal{X}) = p_{\mathrm{s},k} \int \underbrace{t_{k|k-1}(\mathcal{X}|x)}_{=1} \mu_{k-1}^{\natural}(\mathrm{d}x) \tag{2.61a}$$

$$= p_{\mathrm{s},k} \mu_{k-1}^{\natural}(\mathcal{X}). \tag{2.61b}$$

With the help of Equation (2.61), the predicted variance (2.43) reduces to

$$\mathrm{var}_{\mathrm{s},k}(\mathcal{X}) = \mu_{\mathrm{s},k}(\mathcal{X})[1 - \mu_{\mathrm{s},k}(\mathcal{X})] + p_{\mathrm{s},k}^2 \int \underbrace{t_{k|k-1}(\mathcal{X}|x)}_{=1} \underbrace{t_{k|k-1}(\mathcal{X}|x')}_{=1} \nu_{k-1}^{(2)}(\mathrm{d}(x, x')) \tag{2.62a}$$

$$\overset{(2.61)}{=} p_{\mathrm{s},k} \mu_{k-1}^{\natural}(\mathcal{X}) \Big[1 - p_{\mathrm{s},k} \mu_{k-1}^{\natural}(\mathcal{X})\Big] + p_{\mathrm{s},k}^2 \nu_{k-1}^{(2)}(\mathcal{X} \times \mathcal{X}) \tag{2.62b}$$

$$\overset{(1.7)}{=} p_{\mathrm{s},k} \mu_{k-1}^{\natural}(\mathcal{X}) \Big[1 - p_{\mathrm{s},k} \mu_{k-1}^{\natural}(\mathcal{X})\Big] + p_{\mathrm{s},k}^2 \Big[\mu_{k-1}^{(2)}(\mathcal{X} \times \mathcal{X}) - \mu_{k-1}^{\natural}(\mathcal{X})\Big] \tag{2.62c}$$

$$\overset{(1.9)}{=} p_{\mathrm{s},k} \mu_{k-1}^{\natural}(\mathcal{X}) \Big[1 - p_{\mathrm{s},k} \mu_{k-1}^{\natural}(\mathcal{X})\Big] + p_{\mathrm{s},k}^2 \Big[\mathrm{var}_{k-1}^{\natural}(\mathcal{X}) + [\mu_{k-1}^{\natural}(\mathcal{X})]^2 - \mu_{k-1}^{\natural}(\mathcal{X})\Big] \tag{2.62d}$$

$$= p_{\mathrm{s},k}^2 \mathrm{var}_{k-1}^{\natural}(\mathcal{X}) + p_{\mathrm{s},k}[1 - p_{\mathrm{s},k}] \mu_{k-1}^{\natural}(\mathcal{X}). \tag{2.62e}$$

□

Proof of Theorem 2.2.6. For the proof of the SO-PHD update, let $\mathcal{G}_{\mathrm{c},k}$ and $\mathcal{G}_{\mathrm{d},k}$ denote the PGFL of the false alarm and detection processes, respectively. The SO-PHD filter is constructed with the following assumptions:

1. The predicted target process is a Panjer process with parameters $\alpha := \alpha_{k|k-1}$, $\beta := \beta_{k|k-1}$ and spatial distribution $s := s_{k|k-1}$.
2. Each target produces measurements independently.
3. The detection process is modelled as a Bernoulli point process with state-dependent detection rate $p_{\mathrm{d},k}(x)$ and association likelihood $l_k(x|\cdot)$ for $x \in \mathcal{X}$.
4. The false alarm process is Panjer with parameters $\alpha_{\mathrm{c}} := \alpha_{\mathrm{c},k}$, $\beta_{\mathrm{c}} := \beta_{\mathrm{c},k}$ and spatial distribution $\mathrm{s_c} := \mathrm{s}_{\mathrm{c},k}$.

Just like for the PHD filter with Panjer clutter, the general formulation for the joint PGFL of the predicted and false alarm processes is found with Equation (2.25). The Panjer assumption on the predicted target and the false alarm processes in (2.25) leads to a PGFL of the form

$$\mathcal{G}_{\mathrm{J},k}^{\natural}(g, h) = \mu(\mathcal{X})^{\alpha} \Big(F_{\mathrm{d}}(g, h)\Big)^{-\alpha} \Big(F_{\mathrm{c}}(g)\Big)^{-\alpha_{\mathrm{c}}}, \tag{2.63}$$

where

$$F_{\mathrm{d}}(g,h) := \mu(\mathcal{X})\left(1+\frac{1}{\beta}\int(1-h(x)\mathcal{G}_{\mathrm{d}}(g|x))s(\mathrm{d}x)\right) \tag{2.64a}$$

$$= \int\left[1+\frac{1-h(x)\mathcal{G}_{\mathrm{d}}(g|x)}{\beta}\right]\mu(\mathrm{d}x), \tag{2.64b}$$

and

$$F_{\mathrm{c}}(g) := 1+\frac{1}{\beta_{\mathrm{c}}}\int(1-g(z))s_{\mathrm{c}}(z)\mathrm{d}z \tag{2.65}$$

Here, the PGFL of the detection process is found via (1.25), taking the form

$$\mathcal{G}_{\mathrm{d}}(g|x) = q_{\mathrm{d}}(x)+p_{\mathrm{d}}(x)\int_{\mathcal{Z}} g(z)l(z|x)\mathrm{d}z. \tag{2.66}$$

Note that an additional factor $\frac{\mu(\mathcal{X})^{\alpha}}{\mu(\mathcal{X})^{\alpha}}$ was added to Equation (2.63) so that the notations of the SO-PHD filter match those of the PHD and CPHD filters [Vo et al., 2007; Delande et al., 2014]. Furthermore, F_{d} and F_{c} are linear in g, therefore their higher-order derivatives vanish, and for $z \in Z_k$ their first-order derivatives are given by

$$\delta F_{\mathrm{d}}(g,h;\delta_z) = -\int\frac{h(x)p_{\mathrm{d}}(x)l(x|z)}{\beta}\mu(\mathrm{d}x), \tag{2.67}$$

$$\delta F_{\mathrm{c}}(g;\delta_z) = -\frac{1}{\beta_{\mathrm{c}}}s_{\mathrm{c}}(z). \tag{2.68}$$

Again, the conditional PGFL is found using Equation (2.25) in Bayes' rule (2.26). Its denominator is found with the help of the higher-order product and chain rules (1.20) and (1.21) which leads to the expression

$$\delta^{|Z_k|}\mathcal{G}^{\natural}_{\mathrm{J},k}(g,h;(\delta_z)_{z\in Z_k})$$

$$= \mu(\mathcal{X})^{\alpha}\sum_{j=0}^{|Z_k|}\frac{(\alpha)_j}{\beta^j}\frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{\beta_{\mathrm{c}}^{|Z_k|-j}}\frac{F_{\mathrm{d}}(g,h)^{-\alpha-j}}{F_{\mathrm{c}}(g)^{\alpha_{\mathrm{c}}+|Z_k|-j}}\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\left[\prod_{z\in Z}F_{\mathrm{d}}^{z}(h)\prod_{z'\in Z_k\setminus Z}s_{\mathrm{c}}(z')\right] \tag{2.69a}$$

$$\propto \sum_{j=0}^{|Z_k|}\frac{(\alpha)_j}{\beta^j}\frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{(\beta_{\mathrm{c}}F_{\mathrm{c}}(g))^{|Z_k|-j}}F_{\mathrm{d}}(g,h)^{-j}\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F_{\mathrm{d}}^{z}(h)}{s_{\mathrm{c}}(z)}, \tag{2.69b}$$

where

$$F_{\mathrm{d}}^{z}(h) := \int h(x)p_{\mathrm{d}}(x)l(z|x)\mu(\mathrm{d}x). \tag{2.70}$$

The proportionality in (2.69b) is with respect to the constant $\mu(\mathcal{X})^{\alpha}F_{\mathrm{d}}(g,h)^{-\alpha}F_{\mathrm{c}}(g)^{-\alpha_{\mathrm{c}}}\prod_{z\in Z_k}s_{\mathrm{c}}(z)$ which will cancel out with the numerator of (2.26).

Just like in the proof of the PHD filter with Panjer clutter, the updated intensity is found with Equation (2.27), therefore Equation (2.69) has to be differentiated another time with respect to h

in the direction of $\mathbb{1}_B$. With the product rule (1.18), one obtains

$$\begin{aligned}
&\delta^{|Z_k|+1}\mathcal{G}_{\mathrm{J},k}(g,h;(\delta_z)_{z\in Z_k},\mathbb{1}_B)\\
&\propto(-\beta\delta F_{\mathrm{d}}(g,h;\mathbb{1}_B))\sum_{j=0}^{|Z_k|}\frac{(\alpha)_{j+1}}{\beta^{j+1}}\frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{(\beta_{\mathrm{c}}F_{\mathrm{c}}(g))^{|Z_k|-j}}F_{\mathrm{d}}(g,h)^{-j-1}\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}F_{\mathrm{d}}^{z}(h)\\
&\quad+\sum_{z\in Z_k}F_{\mathrm{d}}^{z}(\mathbb{1}_B)\sum_{j=0}^{|Z_k|-1}\frac{(\alpha)_{j+1}}{\beta^{j+1}}\frac{(\alpha_{\mathrm{c}})_{(|Z_k|-1)-j}}{(\beta_{\mathrm{c}}F_{\mathrm{c}}(g))^{(|Z_k|-1)-j}}F_{\mathrm{d}}(g,h)^{-j-1}\sum_{\substack{Z\subseteq Z_k\setminus\{z\}\\|Z|=j}}\prod_{z'\in Z_k}F_{\mathrm{d}}^{z'}(h),
\end{aligned}\tag{2.71}$$

where

$$\delta F_{\mathrm{d}}(g,h;\mathbb{1}_B)=-\frac{1}{\beta}\int_B\left[q_{\mathrm{d}}(x)+p_{\mathrm{d}}(x)\int_{\mathcal{Z}}g(z)\ell(x|z)\mathrm{d}z\right]\mu(\mathrm{d}x).\tag{2.72}$$

By substituting (2.69) and (2.71) into (2.27) and setting $g=0$ and $h=1$, the updated intensity of the SO-PHD filter is found.

Apart from the first-order moment $\mu_k^{\natural}$, also the second-order moment $\mu_k^{(2)}$ is necessary to compute the updated variance of the SO-PHD filter in an arbitrary region $B\in\mathcal{B}(\mathcal{X})$ which is found using (cf. (1.16)) [Delande et al., 2014]

$$\mu_k^{(2)}(B\times B')=\frac{\delta^{|Z_k|+2}\mathcal{G}_{\mathrm{J},k}(0,e^{-f};(\delta_z)_{z\in Z_k},\mathbb{1}_B,\mathbb{1}_{B'})|_{g=0,f=0}}{\delta^{|Z_k|}\mathcal{G}_{\mathrm{J},k}(g,1;(\delta_z)_{z\in Z_k}))|_{g=0}}.\tag{2.73}$$

The denominator of the expression (2.73) has already been computed in (2.69); we shall thus focus here on the numerator

$$\delta^{|Z_k|}\mathcal{G}_{\mathrm{J},k}(0,e^{-f};(\delta_z)_{z\in Z_k})\propto\sum_{j=0}^{|Z_k|}\frac{(\alpha)_j}{\beta^j}\frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{(1+\beta_{\mathrm{c}})^{|Z_k|-j}}F_{\mathrm{d}}(0,e^{-f})^{-j}\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F_{\mathrm{d}}^{z}(e^{-f})}{s_{\mathrm{c}}(z)},\tag{2.74}$$

which also stems from Equation (2.69). The first-order derivative of (2.74) in direction $\mathbb{1}_B$ is found to be

$$\begin{aligned}
&\delta^{|Z_k|+1}\mathcal{G}_{\mathrm{J},k}(0,e^{-f};(\delta_z)_{z\in Z_k},\mathbb{1}_B)\\
&\propto-\sum_{j=0}^{|Z_k|}\frac{(\alpha)_{j+1}}{\beta^{j+1}}\frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{(\beta_{\mathrm{c}}+1)^{|Z_k|-j}}F_{\mathrm{d}}(0,e^{-f})^{-j-1}F_{\mathrm{md}}(e^{-f}\mathbb{1}_B)\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}F_{\mathrm{d}}^{z}(e^{-f})\\
&\quad-\sum_{j=1}^{|Z_k|}\frac{(\alpha)_j}{\beta^j}\frac{(\alpha_{\mathrm{c}})_{|Z_k|-j}}{(\beta_{\mathrm{c}}+1)^{|Z_k|-j}}F_{\mathrm{d}}(0,e^{-f})^{-j}\sum_{z\in Z_k}\frac{F_{\mathrm{d}}^{z}(e^{-f}\mathbb{1}_B)}{s_{\mathrm{c}}(z)}\sum_{Z\subseteq Z_k\setminus\{z\}}\prod_{z'\in Z}\frac{F_{\mathrm{d}}^{z'}(e^{-f})}{s_{\mathrm{c}}(z)},
\end{aligned}\tag{2.75}$$

where

$$F_{\mathrm{md}}(h):=\int h(x)q_{\mathrm{d}}(x)\mu(\mathrm{d}x).\tag{2.76}$$

Equation (2.75) contains three functionals that are dependent on f in each of the two terms,

therefore the second-order derivative will be a sum of six terms, i.e. it takes the form

$$
\begin{aligned}
&\delta^{|Z_k|+2}\mathcal{G}_{\mathrm{J},k}(0,e^{-f};(\delta_z)_{z\in Z_k},\mathbb{1}_B,\mathbb{1}_{B'}) \\
&\propto \sum_{j=0}^{|Z_k|}\frac{(\alpha)_{j+2}}{\beta^{j+2}}\frac{(\alpha_\mathrm{c})_{|Z_k|-j}}{(\beta_\mathrm{c}+1)^{|Z_k|-j}}F_\mathrm{d}(0,e^{-f})^{-j-2}F_\mathrm{md}(e^{-f}\mathbb{1}_B)F_\mathrm{md}(e^{-f}\mathbb{1}_{B'})\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F_\mathrm{d}^z(e^{-f})}{s_\mathrm{c}(z)} \\
&\quad+\sum_{j=1}^{|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}F_\mathrm{md}(e^{-f}\mathbb{1}_B)\sum_{z\in Z_k}\frac{F_\mathrm{d}^z(e^{-f}\mathbb{1}_{B'})}{s_\mathrm{c}(z)}\sum_{\substack{Z\subseteq Z_k\setminus\{z\}\\|Z|=j-1}}\prod_{z'\in Z}\frac{F_\mathrm{d}^{z'}(e^{-f})}{s_\mathrm{c}(z)} \\
&\quad+\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}F_\mathrm{md}(e^{-f}\mathbb{1}_{B\cap B'})\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F_\mathrm{d}^z(e^{-f})}{s_\mathrm{c}(z)} \\
&\quad+\sum_{j=1}^{|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}F_\mathrm{md}(e^{-f}\mathbb{1}_{B'})\sum_{z\in Z_m}\frac{F_\mathrm{d}^z(e^{-f}\mathbb{1}_B)}{s_\mathrm{c}(z)}\sum_{\substack{Z\subseteq Z_k\setminus\{z\}\\|Z|=j-1}}\prod_{z'\in Z}\frac{F_\mathrm{d}^{z'}(e^{-f})}{s_\mathrm{c}(z)} \\
&\quad+\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}\sum_{z\in Z_k}\frac{F_\mathrm{d}^z(e^{-f}\mathbb{1}_{B\cap B'})}{s_\mathrm{c}(z)}\sum_{\substack{Z\subseteq Z_k\setminus\{z\}\\|Z|=j-1}}\prod_{z'\in Z}\frac{F_\mathrm{d}^{z'}(e^{-f})}{s_\mathrm{c}(z)} \\
&\quad+\sum_{j=2}^{|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}\sum_{\substack{z,z'\in Z_k\\z\neq z'}}\frac{F_\mathrm{d}^z(e^{-f}\mathbb{1}_B)}{s_\mathrm{c}(z)}\frac{F_\mathrm{d}^{z'}(e^{-f}\mathbb{1}_{B'})}{s_\mathrm{c}(z)}\sum_{\substack{Z\subseteq Z_k\setminus\{z,z'\}\\|Z|=j-2}}\prod_{z''\in Z}\frac{F_\mathrm{d}^{z''}(e^{-f})}{s_\mathrm{c}(z)}.
\end{aligned}
\tag{2.77}
$$

Note that the third and fifth terms in (2.77) correspond exactly to the updated first-order moment $\mu_k^\natural$. Subsituting and (2.69) and (2.77) into (2.73) yields the much shorter notation

$$
\begin{aligned}
\mu_k^{(2)}(B\times B') &= \mu_k^\natural(B\cap B')+\mu_k^\phi(B)\mu_k^\phi(B')\ell_2(\phi)+\mu_k^\phi(B)\sum_{z\in Z}\frac{\mu_k^z(B')}{s_\mathrm{c}(z)}\ell_2(z) \\
&\quad+\mu_k^\phi(B')\sum_{z\in Z}\frac{\mu_k^z(B)}{s_\mathrm{c}(z)}\ell_2(z)+\sum_{z,z'\in Z_k}\frac{\mu_k^z(B)}{s_\mathrm{c}(z)}\frac{\mu_k^{z'}(B')}{s_\mathrm{c}(z')}\ell_2^{\neq}(z,z').
\end{aligned}
\tag{2.78}
$$

Equation (1.9) finally yields the variance $\mathrm{var}_k^\natural$ using the result (2.78) for $B=B'$ and the definition of $\mu_k^\natural$. □

Proof of Corollary 2.2.7. Again, the time subscripts are omitted where there are no ambiguities. Assume that the predicted point process is Poisson with constant parameter $\lambda=\frac{\alpha}{\beta}$, where $\alpha\to\infty$. First of all, with the identity $\mu(\mathrm{d}x)=\lambda s(\mathrm{d}x)$ one obtains the limit

$$
\lim_{\alpha,\beta\to\infty}F_\mathrm{d}=\lim_{\alpha,\beta\to\infty}\int\Big[1+\underbrace{\frac{p_{\mathrm{d},k}(x)}{\beta}}_{\to 0}\Big]\lambda s(\mathrm{d}x) \tag{2.79a}
$$

$$
=\lambda. \tag{2.79b}
$$

In order to check the convergence of the intensity update equation (1.39) with $\bullet=\natural$, the only

relevant term is $Y_u(Z)$ (2.23) which is the only term that contains α or β. Therefore,

$$\lim_{\alpha,\beta\to\infty} Y_u(Z) = \lim_{\alpha,\beta\to\infty} \sum_{j=0}^{|Z|} \frac{\alpha_{(j+u)\uparrow}}{(\beta)^{j+u}} \frac{(\alpha_c)_{(|Z|-j)\uparrow}}{(\beta_c+1)^{|Z|-j}} F_d^{-j-u}\ e_j(Z) \tag{2.80a}$$

$$\overset{(2.1)}{=} \lim_{\alpha,\beta\to\infty} \sum_{j=0}^{|Z_k|} \lambda \underbrace{\left(\lambda+\frac{1}{\beta}\right)}_{\to\lambda} \cdots \underbrace{\left(\lambda+\frac{j+u-1}{\beta}\right)}_{\to\lambda} \frac{(\alpha_c)_{(|Z|-j)\uparrow}}{(\beta_c+1)^{|Z|-j}} \underbrace{F_d^{-j-u}}_{\to\lambda^{-j-u}} e_j(Z) \tag{2.80b}$$

$$= \sum_{j=0}^{|Z|} \frac{(\alpha_c)_{(|Z|-j)\uparrow}}{(\beta_c+1)^{|Z|-j}} e_j(Z). \tag{2.80c}$$

In particular, note that the limit of $Y_u(Z)$ is independent of the value of u; the corrective terms (2.40) hence converge to

$$\begin{cases} \lim\limits_{\alpha,\beta\to\infty} \ell_1(\phi) = 1 \\ \lim\limits_{\alpha,\beta\to\infty} \ell_1(z) = \dfrac{\sum_{j=0}^{|Z_k|-1} \frac{(\alpha_c)_{(|Z_k|-j-1)\uparrow}}{(\beta_c+1)^{|Z_k|-j-1}} e_j(Z_k\setminus\{z\})}{\sum_{j=0}^{|Z_k|} \frac{(\alpha_c)_{(|Z_k|-j)\uparrow}}{(\beta_c+1)^{|Z|-j}} e_j(Z_k)}, \end{cases} \tag{2.81}$$

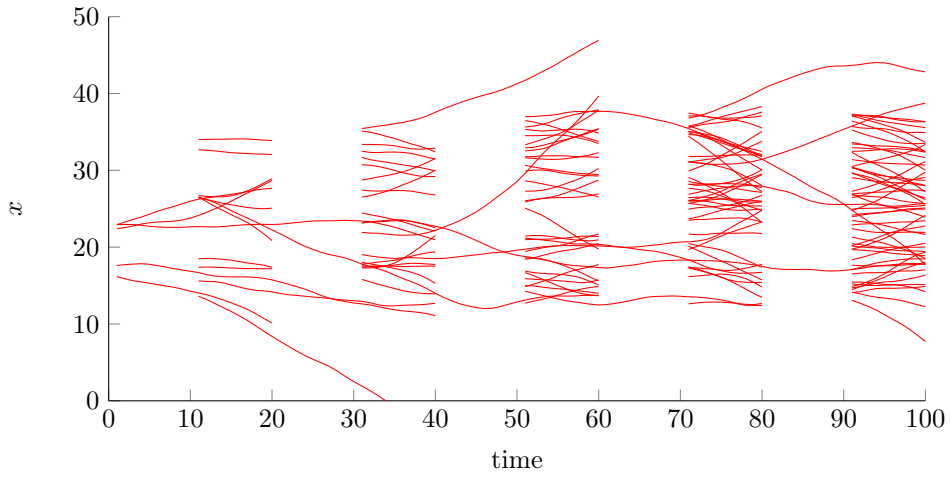
which coincides with the terms for the intensity update (2.22) of the PHD filter with Panjer clutter.

If we further assume that the clutter process is Poisson, i.e. $\lambda_c = \frac{\alpha_c}{\beta_c}$ is constant and $\alpha\to\infty$, the terms (2.81) further reduce to

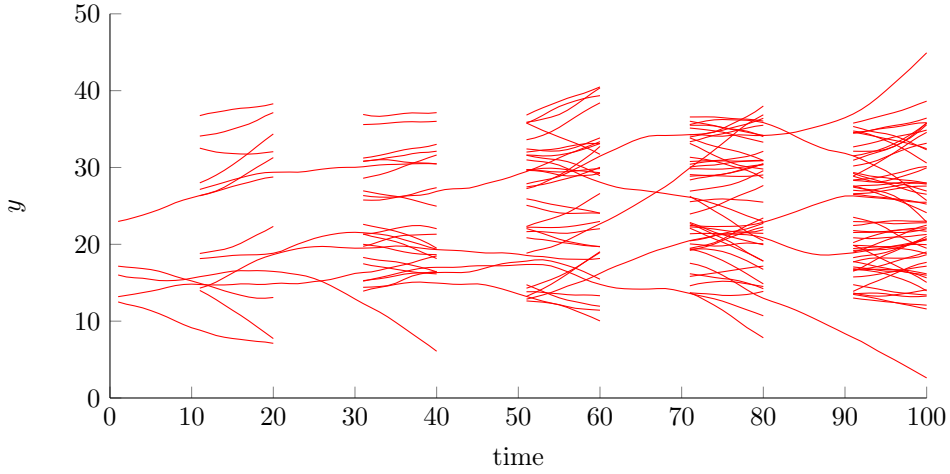
$$\begin{aligned} \lim_{\alpha\to\infty} Y(Z) &= \lim_{\alpha\to\infty} \sum_{k=0}^{|Z|} \frac{(\alpha_c)_{j\uparrow}}{(\frac{\alpha_c+\lambda_c}{\lambda_c})^j}\ e_{m-j}(Z) \\ &= \lim_{\alpha\to\infty} \sum_{j=0}^{|Z|} \lambda_c^j \underbrace{\left[\frac{\alpha}{\alpha+\lambda_c}\frac{\alpha+1}{\alpha+\lambda_c}\cdots\frac{\alpha+j-1}{\alpha+\lambda_c}\right]}_{=1 \text{ for } \alpha\to\infty}\ e_{m-j}(Z) \\ &= \sum_{k=0}^{|Z|} \lambda_c^j\ e_{m-j}(Z). \end{aligned} \tag{2.82}$$

In the limit $\alpha\to\infty$, the fraction $\frac{Y(Z_k\setminus\{z\})}{Y(Z_k)}$ becomes

$$\begin{aligned} \lim_{\alpha\to\infty} \frac{Y(Z_k\setminus\{z\})}{Y(Z_k)} &= \frac{\sum_{j=0}^{|Z|-1} \lambda_c^j\ e_{m-1-j}(Z\setminus\{z\})}{\sum_{j=0}^{|Z_k|} \lambda_c^j\ e_{m-j}(Z_k)} \cdot \frac{\frac{1}{\lambda_c^{|Z_k|}}}{\frac{1}{\lambda_c^{|Z_k|}}} \\ &= \frac{1}{\lambda_c} \frac{\sum_{Z\subseteq Z_k\setminus\{z\}} \prod_{z'\in Z} \frac{\mu_k^{z'}(\mathcal{X})}{\lambda_c s_c(z')}}{\sum_{Z\subseteq Z_m} \prod_{z'\in Z} \frac{\mu_k^{z'}(\mathcal{X})}{\lambda_c s_c(z')}} \\ &\overset{(*)}{=} \frac{1}{\lambda_c} \frac{\sum_{Z\subseteq Z_m\setminus\{z\}} \prod_{z'\in Z} \frac{\mu_k^{z'}(\mathcal{X})}{\lambda_c s_c(z')}}{\sum_{Z\subseteq Z_k\setminus\{z\}} \prod_{z'\in Z} \frac{\mu_k^{z'}(\mathcal{X})}{\lambda_c s_c(z')} \left(1+\frac{\mu_k^z(\mathcal{X})}{\lambda_c s_c(z)}\right)} \\ &= \frac{s_c(z)}{\mu_k^z(\mathcal{X}) + \lambda_c s_c(z)}. \end{aligned} \tag{2.83}$$



(a) x coordinates (in m) over time.



(b) y coordinates (in m) over time.

Figure 2.6: Ground truth of the step experiment S2.3.

Equation $(*)$ was obtained by rearranging the sum in the denominator to separate the terms with $Z \ni z$. By inserting the result of (2.83) into (2.22), we obtain

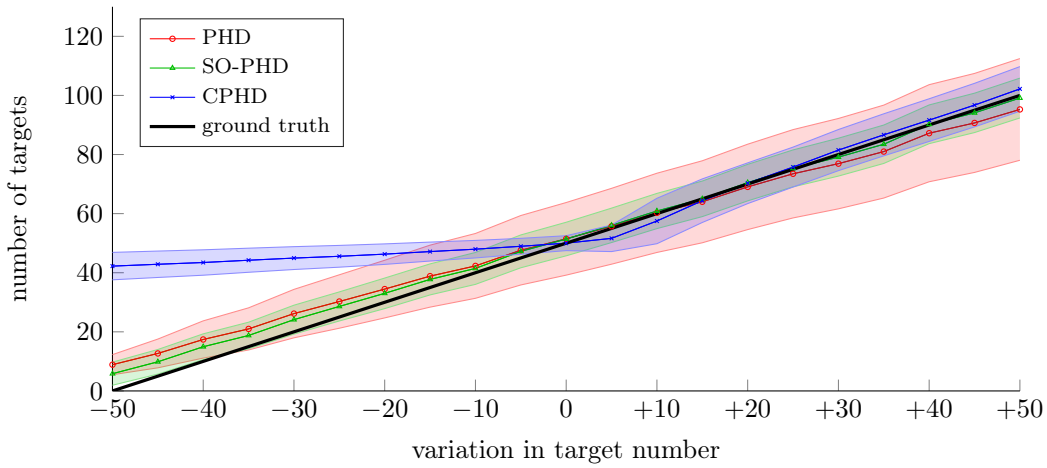
$$\begin{aligned} \lim_{\alpha \to \infty} \mu_k^{\boldsymbol{*}}(B) &= \mu_k^{\phi}(B) + \sum_{z \in Z_k} \frac{\mu_k^z(x)}{s_{\mathrm{c}}(z)} \left(\frac{s_{\mathrm{c}}(z)}{\mu_k^z(\mathcal{X}) + \lambda_{\mathrm{c}} s_{\mathrm{c}}(z)} \right) \\ &= \mu_k^{\phi}(B) + \sum_{z \in Z_k} \frac{\mu_k^z(B)}{\mu_k^z(\mathcal{X}) + \lambda_{\mathrm{c}} s_{\mathrm{c}}(z)} \end{aligned} \tag{2.84}$$

which is the update intensity (1.39) of the PHD filter [Mahler, 2003]. □

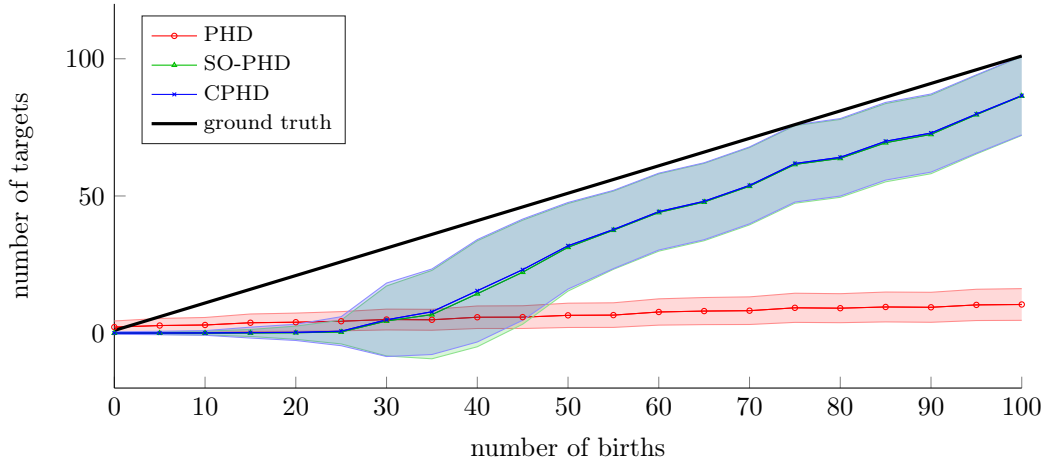
2.2.6 Simulations with the SO-PHD filter [Schlangen et al., 2017c]

In the manner of Section 2.2.3, the SO-PHD filter is analysed on simulated data and compared with the PHD and CPHD filters in order to assess how each of the filters adapts to spontaneous large changes in the number of objects.

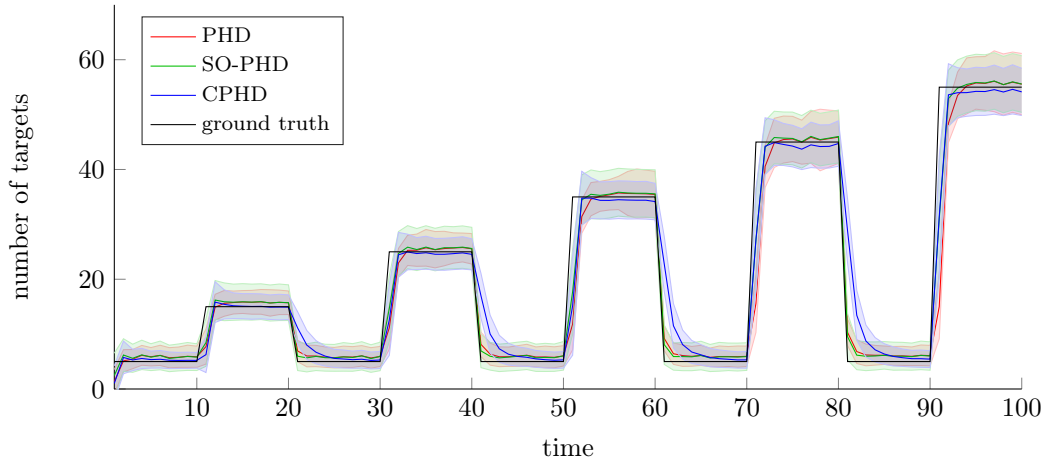
Again, the state space is assumed two-dimensional and of size $50\,\mathrm{m} \times 50\,\mathrm{m}$. Objects are only generated in the centre of the image in order to prevent them from leaving the surveillance region during the experiment. They move according to an n.c.v. model, having a small acceleration noise



(a) Experiment 1.1.



(b) Experiment 1.2.



(c) Experiment 1.3.

Figure 2.7: 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).

with a standard deviation of $0.3\,\mathrm{m\,s^{-2}}$ and Gaussian distributed initial velocity with mean 0 and standard deviation $0.5\,\mathrm{m\,s^{-1}}$ in both dimensions. Measurements are generated from the targets with a detection rate of $p_\mathrm{d} = 0.9$ and measurement noise with standard deviation 0.2 m in each dimension. The filters assume a survival rate of $p_\mathrm{s} = 0.99$ throughout all experiments. Let us consider three different scenarios:

S2.1 Similarly to Simulation S1.1, the experiment runs for 15 time steps. Exactly 50 targets are propagated for the first 14 time scans, until the number of targets suddenly changes at time 15, either by adding or removing up to 50 targets randomly in the state space. The birth model of the SO-PHD and CPHD filters is negative binomial with mean $\mu_{\mathrm{b}} = 25$ and variance $\mathrm{var}_{\mathrm{b}} = 100$, the PHD filter assumes Poisson birth with mean $\mu_{\mathrm{b}} = 25$. This scenario assumes a Poisson clutter process with mean $\mu_{\mathrm{c}} = 5$.

S2.2 This scenario is created in the same manner as Scenario S2.1, but propagating only one target through the first 14 time steps and then adding up to 100 targets at time 15 across the state space. The birth process is assumed negative binomial with mean $\mu_{\mathrm{b}} = 1$ and $\mathrm{var}_{\mathrm{b}} = 100$ for the three filters, though the PHD filter cannot utilise variance information. This scenario assumes a Poisson clutter process with mean $\mu_{\mathrm{c}} = 20$.

S2.3 In contrast to the first two scenarios, this scenario is designed to analyse how the filters adapt over time to spontaneous bursts of target birth (cf. Simulation S1.2). There are at least 5 targets in the scene at all times, but n additional targets are introduced every 20th time step for 10 time steps each, where $n \in \{10, 20, 30, 40, 50\}$. A plot of the true x and y positions of the targets over time is given in Figure 2.6. This scenario assumes a Poisson clutter process with mean $\mu_{\mathrm{c}} = 5$.

Fig. 2.7 depicts the estimated number of targets for the three experiments along with the computed filter variance. In experiment S2.1, the PHD filter benefits from the fact that the number of targets was high in the previous time steps and consequently its mean and variance, such that it is equally reactive to extreme target birth and death. In experiment S2.2, on the other hand, the filter cannot adapt fast to big increases in the number of objects since its mean and variance has stabilised around 1 previously. The CPHD filter, on the other hand, copes well with target birth in both scenarios since a negative binomial birth model could be incorporated in the cardinality information; in case of extreme target death, however, it fails to adapt fast. The SO-PHD filter, on the other hand, copes equally well with spontaneous target birth and death in both scenarios. It has to be noted that both the SO-PHD and the CPHD filter show a displacement in reaction to spontaneous target birth in scenario S2.2 since the false alarm rate is set to 20 and the filters tend to interpret unexplained measurements rather as false alarms than as newborn objects.

Experiment S2.3 analyses how long it takes the three filters to adapt to sudden changes in the number of objects of variable intensity. After each change in the number of objects, the filters have 10 scans to adapt to the new number of targets. This experiment confirms the findings of the first two results: the PHD filter needs up to three time steps to adapt to sudden birth but adapts to target death very quickly; the CPHD filter, on the other hand, behaves in the opposite manner, adapting quickly to target birth and taking up to five time steps to adapt to target death. The proposed method is reactive to both target birth and death.

The average runtimes of one iteration for Scenario S2.3 in Matlab on a dual-core Dell Precision M4800 workstation with Intel(R) Core(TM) i7-4710MQ CPU @ 2.50GHz were 0.9 s for the PHD filter, 1.2 s for the SO-PHD filter and 6.5 s for the CPHD filter.

2.3 Two alternative single-cluster PHD filter methods

The formulation of the SO-PHD filter motivates its exploitation for applications that involve, but go beyond, multi-object estimation. In Section 1.4, it was shown on the example of the PHD filter that multi-object filters can be embedded in a hierarchical process to estimate unknown parameters on top of the multi-target state. In general, this hierarchical architecture can take on any daughter process, so this section attempts to utilise the CPHD and SO-PHD filters to formulate their respective single-cluster processes.[5]

2.3.1 The single-cluster CPHD and SO-PHD filters

Theorem 2.3.1 (Multi-object likelihood of the CPHD filter)**.** *The multi-object likelihood function of the CPHD filter for a given sensor state s is found to be*

$$\ell_k^\sharp(s|Z) = \langle \tilde{\Upsilon}^0[\mu_{k|k-1}^\sharp, Z], \rho_{k,k-1} \rangle \tag{2.85}$$

with

$$\begin{aligned}&\tilde{\Upsilon}^0[\mu_{k|k-1}^\sharp, Z](n) = \\ &\sum_{j=0}^{\min(|Z|,n-u)} \frac{n!(|Z|-j)!}{(n-(j+d))!} \rho_{\mathrm{c}}(|Z|-j) \mu_k^\phi(\mathcal{X}|s)^{n-(j+d)} \sum_{\substack{Z' \subseteq Z \\ |Z'|=j}} \prod_{z \in Z'} \mu_k^z(\mathcal{X}|s) \prod_{z' \in (Z')^c} s_{\mathrm{c},k}(z),\end{aligned} \tag{2.86}$$

where $(Z')^c = Z \setminus Z'$.

Theorem 2.3.2 (Multi-object likelihood of the second-order PHD filter)**.** *Write $\alpha = \alpha_{k|k-1}$ and $\beta = \beta_{k|k-1}$ for the sake of brevity, and let*

$$\tilde{F}_{\mathrm{d},s} = 1 - \frac{1}{\beta} \int_{\mathcal{X}} p_{\mathrm{d},k}(x|s) s_{k|k-1}(\mathrm{d}x, s), \tag{2.87}$$

$$F_{\mathrm{c}} = 1 + \frac{1}{\beta_{\mathrm{c}}} \tag{2.88}$$

for a given sensor state s. The multi-object likelihood function of the Panjer PHD filter for s is found to be

$$\ell_k^\natural(s|Z) = \sum_{j=0}^{|Z|} \frac{\alpha_{j\uparrow}}{\beta^j} \frac{(\alpha_{\mathrm{c},k})_{(|Z|-j)\uparrow}}{(\beta_{\mathrm{c},k}+1)^{|Z|-j}} \tilde{F}_{\mathrm{d},s}^{-\alpha-j} F_{\mathrm{c}}^{-\alpha_{\mathrm{c}}-|Z|-j} \sum_{\substack{Z' \subseteq Z \\ |Z'|=j}} \prod_{z \in Z'} \mu_k^z(\mathcal{X}) \prod_{z' \in (Z')^c} s_{\mathrm{c},k}(z|s). \tag{2.89}$$

Again, let us assess the performance of the filters on simulated data first. For the sake of comparison, an alternative likelihood function is consulted which is introduced by Leung et al. [2016]. This multi-object likelihood takes all possible associations between the measurement set and the target population into account. First of all, write the short-hand notation

$$p_{\mathrm{c}}(Z) = \frac{\prod_{z \in Z} s_{\mathrm{c}}(z|s_k)}{\exp\left(\int s_{\mathrm{c}}(z'|s_k) dz'\right)}. \tag{2.90}$$

[5] In Schlangen et al. [2014], the so-called Hypothesised filter for Independent Stochastic Populations (HISP) [Houssineau, 2015] was already successfully embedded in a single-cluster process approach.

Furthermore, define the association function $\theta : \mathcal{B}(\mathcal{Z}) \to \mathcal{B}(\mathcal{X})$ which maps a selection of measurements $Z = \{z_1, \ldots, z_{m_k}\}$ to a selection of targets $X = \{x_1, \ldots, x_{n_k}\}$ via

$$\theta(z_j) = \begin{cases} x_i & \text{if } z_j \text{ is associated with } x_i, \\ 0 & \text{otherwise,} \end{cases} \tag{2.91}$$

where the x_i are extracted from the predicted intensity $\mu^{\bullet}_{k|k-1}$ with $\bullet \in \{\flat, \sharp, \natural\}$. Then,

$$\begin{aligned} \hat{\ell}_k(s|Z) = p_{\mathrm{c}}(Z) &\prod_{i=1}^{n_k}(1 - p_{\mathrm{d}}(x_i|s)) \\ &\cdot \sum_{\theta} \prod_{\substack{j=1 \\ \theta(j)\neq 0}}^{m_k} \frac{p_{\mathrm{d}}(\theta(z_j)|s)\ell(\theta(z_j)|z_j, s)}{(1 - p_{\mathrm{d}}(\theta(z_j|s))p_{\mathrm{c}}(z_j)}. \end{aligned} \tag{2.92}$$

Note that in contrast to (1.58), the associations in the last term of (2.92) are *not* marginalised over all possible states, but this algorithm depends on the extraction of specific object locations x_i. In other words, this approach does not use the full information available through the predicted intensity.

2.3.2 Derivation of the multi-object likelihood for the PHD, CPHD and SO-PHD filters

Each of the following proofs follows the same structure. Just like in the proofs of the PHD filter with Panjer clutter and the SO-PHD filter, the PGFL of the observation process (2.26) is found using the specific assumptions of the respective filter, and the multi-object likelihood is found via the $|Z|$th order derivative with respect to each measurement.

Alternative proof of PHD likelihood (1.58). The PHD likelihood was demonstrated in [Swain, 2013] in the context of random finite sets, for the sake of completeness this section provides an alternative proof using point processes. In case of the PHD filter, both $G_{k|k-1}$ and G_{c} are Poisson PGFLs such that (2.25) takes the form

$$\begin{aligned} G^{\flat}_{J,k}(g|s) = \exp\Big(&\int (g(z) - 1)\mu_{\mathrm{c}}(z|s) \\ &+ \int \left(\left[1 - p_{\mathrm{d}}(x|s) + p_{\mathrm{d}}(x|s) \int g(z)\ell(x|z, s)\mathrm{d}z \right] - 1 \right) \mu_{k|k-1}(x|s)\mathrm{d}x \Big). \end{aligned} \tag{2.93}$$

Differentiation by g requires repeated applications of the chain rule (1.19) which pulls out one multiplicative term $\mu_{\mathrm{c}}(z) + \int p_{\mathrm{d}}(x|s) l(x|z, s)\mu_{k|k-1}(x|s)\mathrm{d}x$ for each measurement $z \in Z_k$. The final result is obtained by setting $g = 0$. □

Proof of CPHD likelihood (2.85). The CPHD filter assumes the target and clutter processes to be

i.i.d. cluster processes. This results in the following variation of (2.25):

$$\begin{aligned} G^{\sharp}_{\text{obs}}(g|s) &= G^{\sharp}_{\text{c}}(g|s)G^{\sharp}_{k|k-1}(G_{\text{d}}(g|\cdot,s)) := \left(\sum_{n\geq 0}\rho_{\text{c}}(n)\left[\int g(z)s_{\text{c}}(z|s)\mathrm{d}z\right]^n\right) \\ &\cdot\left(\sum_{n\geq 0}\rho_{k|k-1}(n)\left[\int\left(1-p_{\text{d}}(x|s)+p_{\text{d}}(x|s)\int g(z)\ell(x|z,s)\mathrm{d}z\right)s^{\sharp}_{k|k-1}(x|s)\mathrm{d}x\right]^n\right) \end{aligned} \tag{2.94}$$

Since the product in (2.94) cannot be simplified like in (2.93) above, it requires the use of the general product rule (1.20) leading to

$$\begin{aligned} &\delta^m G^{\sharp}_{\text{obs}}(g;\delta_{z_1},\ldots,\delta_{z_m}|s)|_{g=0} \\ &= \sum_{Z\subseteq Z_m}\left(\delta^{|Z^c|}G^{\sharp}_{\text{c}}(g;(\delta_z)_{z\in Z^c}|s)|_{g=0}\delta^{|Z|}G^{\sharp}_{k|k-1}(G_{\text{d}}(g;(\delta_z)_{z\in Z}|\cdot,s))|_{g=0}\right), \end{aligned} \tag{2.95}$$

where $Z^c = Z_m \setminus Z$. The first term in (2.95) creates the product $\prod_{z\in Z_m\setminus Z} s_{\text{c}}(z|s)\rho_{\text{c}}(Z^c)$ for all $Z \subseteq Z_k$ since setting $g = 0$ eliminates almost every term of the i.i.d. cluster process. The second term is evaluated using the chain rule (1.21) which leads to the sum

$$\begin{aligned} &\delta^{|Z|}G^{\sharp}_{k|k-1}(G_{\text{d}}(g;(\delta_z)_{z\in Z}|\cdot,s))|_{g=0} \\ &= \sum_{n\geq|Z|}\frac{n!\rho_{k|k-1}(n)}{(n-|Z|)!}\prod_{z\in Z}\int p_{\text{d}}(x)\ell(x|z,s)s^{\sharp}_{k|k-1}(x|s)\mathrm{d}x\left(\int(1-p_{\text{d}}(x))s^{\sharp}_{k|k-1}(x)\mathrm{d}x\right)^{n-|Z|}. \end{aligned} \tag{2.96}$$

The result is obtained by switching summations and rearranging the terms. □

Proof of SO-PHD likelihood (2.89). The Panjer assumption of the second-order PHD filter leads to the PGFL

$$\begin{aligned} G^{\natural}_{\text{obs}}(g|s) &= G^{\natural}_{\text{c}}(g|s)G^{\natural}_{k|k-1}(G_{\text{d}}(g|\cdot,s)) = \left(1+\frac{1}{\beta_{\text{c}}}\int(1-g(z))s_{\text{c}}(z|s)\mathrm{d}z\right)^{-\alpha_{\text{c}}} \\ &\cdot\left(1+\frac{1}{\beta_{k|k-1}}\int\left[1-\left(1-p_{\text{d}}(x|s)+p_{\text{d}}(x|s)\int\ell(x|z,s)g(z)\mathrm{d}z\right)\right]s^{\natural}_{k|k-1}(x|s)\mathrm{d}x\right)^{-\alpha_{k|k-1}}. \end{aligned} \tag{2.97}$$

The general product rule (1.20) is needed for the product in (2.97) in the same manner as (2.95), replacing $\sharp$ by $\natural$. Using the notations $G^{\natural}_{k|k-1}(g|s) := F_{\text{d}}(g|s)^{-\alpha_{k|k-1}}$ and $G^{\natural}_{\text{c}}(g|s) := F_{\text{c}}(g|s)^{-\alpha_{\text{c}}}$, the corresponding derivatives are

$$\delta^{|Z|}G^{\natural}_{k|k-1}(g;(\delta_z)_{z\in Z}|s) = F_{\text{d}}(g|s)^{-\alpha_{k|k-1}-|Z|}\frac{(\alpha_{k|k-1})_{|Z|}}{\beta^{|Z|}_{k|k-1}}\prod_{z\in Z}\int p_{\text{d}}(x|s)\ell(x|z,s)s_{k|k-1}(x|s)\mathrm{d}x \tag{2.98}$$

and

$$\delta^{|Z^c|}G^{\natural}_{\text{c}}(g;(\delta_z)_{z\in Z^c}|s) = F_{\text{d}}(g|s)^{-\alpha_{\text{c}}-|Z^c|}\frac{(\alpha_{\text{c}})_{|Z|}}{\beta^{|Z|}_{\text{c}}}\prod_{z\in Z^c}s_{\text{c}}(z). \tag{2.99}$$

Including (2.98) and (2.99) into (2.97), switching the summations and rearranging the terms leads to the desired result. □

2.3.3 Simulations using the single-cluster CPHD and SO-PHD filters

All experiments presented in this section are performed using a Gaussian Mixture implementation of the algorithms described earlier, see [Vo and Ma, 2006; Vo et al., 2007; Schlangen et al., 2017c] and the Appendix for more detail. The calibration is conducted for all filters with their respective full multi-object likelihoods (1.58), (2.89) and (2.85), henceforth globally denoted by L1, as well as with the likelihood (2.92), labeled by L2 in the following. The parent process is implemented in all cases with a SMC filter approach in the manner of [Lee et al., 2013; Schlangen et al., 2016b] and others (see Appendix A.3.2), using 300 MC particles for each run and performing basic roulette resampling if the effective sample size falls below 150. All results presented in the following are averaged over 50 MC runs, and the sensor estimate as well as the estimated number of targets is computed as the weighted mean over all particles.

As [Leung et al., 2016] suggests, the likelihood (2.92) needs additional assumptions to make it computationally feasible. Firstly, only associations are taken into account that lead to a single object - single association likelihood above a threshold $\tau_0 = 10^{-7}$ in the first time step and $\tau = 10^{-3}$ otherwise. Furthermore, a connected component analysis is performed to find groups of object-measurement clusters that are worth being associated, and these groups are restricted contain at most 3 measurements and 3 targets. Note that the three multi-object likelihood functions (1.58), (2.89) and (2.85) do not require any restrictions.

2.3.3.1 Experiment 1

For the first experiment, a global ground truth is simulated over 100 time steps (of unit 1 s) for both the multi-target configuration and the sensor trajectory, and the 50 MC runs are performed on different measurement sets extracted from this ground truth. Both the target and the sensor state space are assumed four-dimensional, accounting for position and velocity in a two-dimensional environment measured in m. The targets follow a Poisson birth model with mean 4, and the objects move according to a n.c.v. model with acceleration noise of $0.3\,\mathrm{m\,s^{-2}}$ and with an initial velocity of 0 with Gaussian noise of $0.1\,\mathrm{m\,s^{-1}}$ in both x and y. Each target survives with a probability of $\mathrm{p_s} = 0.95$. The sensor follows an n.c.v. model with acceleration noise $0.2\,\mathrm{m\,s^{-2}}$ and initial velocity $0\,\mathrm{m\,s^{-1}}$ in both x and y. The simulated sensor trajectory is depicted in Fig. 2.8a.

From this scenario, measurements are generated with a uniform detection probability of $p_\mathrm{d} = 0.99$ over the whole state space. The observation space is assumed two-dimensional, accounting for the two dimensions of the environment that contains the objects. Measurements are superimposed with a measurement noise of 0.1 m in both dimensions and with the created sensor drift, and false alarms are generated uniformly over the state space according to a Poisson noise model with mean 10.

In this scenario, the filter parameters are set to the same parameters that were used to generate the simulation. Fig. 2.9a shows the root mean square error in the estimation of the sensor trajectory over all 100 time steps for all filters. It can be seen that the three filters do not differ greatly among each other since the generated measurements fit the filter parameters. The full multi-object likelihoods associated to each filter, however, bring a much better estimation of the sensor state in

filter	prediction	update	likelihood
PHD L1	0.0022	0.2805	0.0182
PHD L2	0.0022	0.7205	0.5007
SO-PHD L1	0.0024	0.5914	0.0191
SO-PHD L2	0.0019	0.8681	0.4285
CPHD L1	0.2256	4.5857	0.0218
CPHD L2	0.2141	4.8107	0.4630

Table 2.2: Averaged runtimes in seconds, Experiment 1.

	filter	prediction	update	likelihood
14 deaths	PHD L1	0.0014	0.0812	0.0033
	PHD L2	0.0014	0.0897	0.0173
	SO-PHD L1	0.0014	0.2003	0.0036
	SO-PHD L2	0.0013	0.1957	0.0153
	CPHD L1	0.3281	1.6374	0.0067
	CPHD L2	0.3079	1.5674	0.0166
15 births	PHD L1	0.0016	0.2323	0.0109
	PHD L2	0.0016	0.2664	0.0484
	SO-PHD L1	0.0017	0.7661	0.0117
	SO-PHD L2	0.0017	0.7389	0.0431
	CPHD L1	0.3185	5.1335	0.0143
	CPHD L2	0.3066	4.8654	0.0427

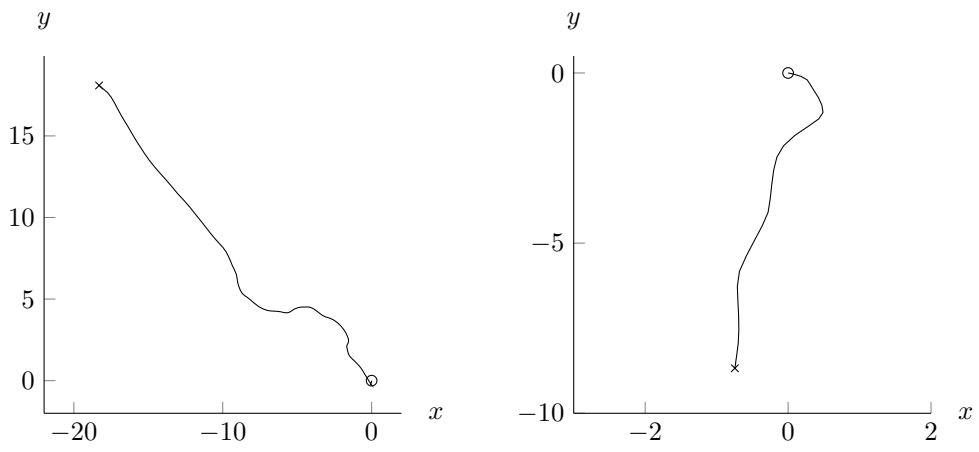
Table 2.3: Averaged runtimes in seconds, Experiment 2.

comparison to the likelihood suggested by [Leung et al., 2016] which is consistently diverging after time step 10 for all three filters. In terms of the estimated number of targets (Fig. 2.9b and 2.9c), all filters seem to monitor the ground truth consistently, apart from the second-order PHD filter with the alternative likelihood that slightly underestimates the number of targets in comparison to the other filters.

Tab. 2.2 shows the averaged runtimes for the prediction, update and likelihood functions of all filters, averaged over 50 MC runs. The CPHD filter update is up to 16 times slower than the updates of the PHD filters of first and second order. Moreover, the alternative likelihood function L2 also performs considerably slower than the full multi-object likelihood function.

2.3.3.2 Experiment 2

The second experiment aims at analysing the effect of model mismatches on the robustness of the filters. While the filters follow the same target birth/death model as in Experiment 1, the ground truth does not: 15 targets are created at the initial step, and stay alive until time step $t = 15$; then, either 14 objects are artificially removed (Experiment 2.1) or 15 objects are added to the existing population (Experiment 2.2), such that the number of targets changes to 1 or 30, respectively, all of which stay alive until the end of the scenario. Little is known for the filters about the birth/death model, thus the second-order PHD and the CPHD filters are fed with a negative binomial birth with mean 2 and variance 20, accounting for a large uncertainty. The PHD filter, on the other hand, can only describe the number of newborn targets through its mean value, which is set to 2 as well. The probability of survival in the three filters is set to $p_{\mathrm{s}} = 0.99$. There are no model mismatches for the remaining parameters, whose values are set as in Experiment 1, except for a



(a) Experiment 1.

(b) Experiment 2.

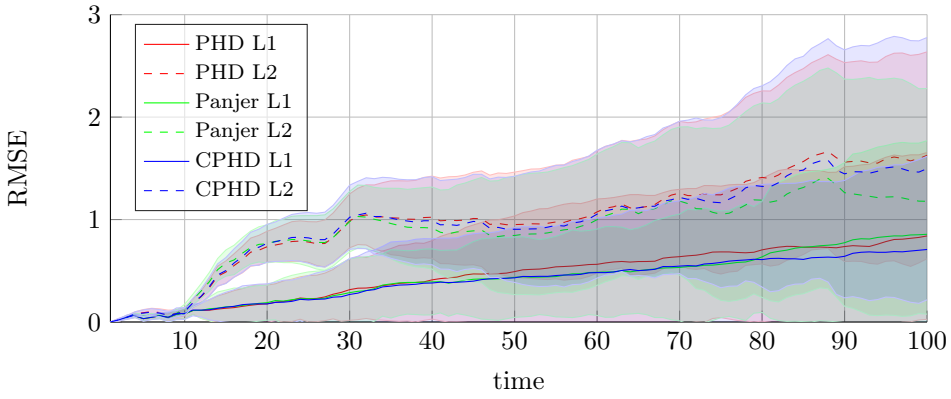
Figure 2.8: Ground truth for the sensor state for Experiments 1 and 2, both originating at $(0, 0)$.

slightly smaller acceleration noise ($0.1\,\mathrm{m\,s^{-2}}$).

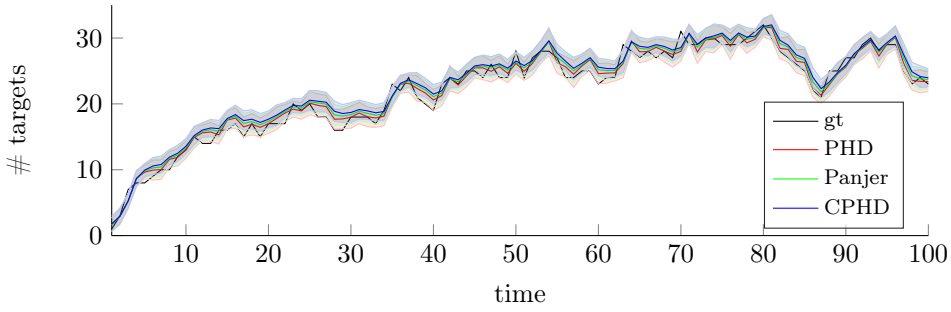
Fig. 2.10 shows the estimation results for Experiment 2.1, with unexpected target death at time $t = 15$. As for the first experiment, the proposed likelihood L1 leads to a significantly more accurate estimation of the sensor state. Regardless of the chosen filter or likelihood, the estimation error increases sharply after time $t = 15$; it might be explained by the sudden target death which drastically decreases the amount of information for the estimation of the sensor state. Fig. 2.10b and 2.10c suggest that the CPHD filter is less reactive to unexpected target disappearances; this might be explained by its lower flexibility to out-of-model target deaths since it maintains a full cardinality distribution on the number of targets. On the other hand, it remains unclear why the second-order PHD filter slightly underestimates the number of targets if combined with the alternative likelihood L2, but shows accurate results with the proposed likelihood L1.

The results for Experiment 2.2 are displayed in Fig. 2.11. Again, the proposed method using likelihood L1 shows better performances in the estimation of the sensor state. Unlike Experiment 2.1, the unexpected increase in the number of targets provides more information and seems to facilitate the estimation of the sensor state, resulting in an improvement for all the filters and likelihoods immediately after time step $t = 15$. Again, the second-order PHD filter underestimates the number of targets when combined with the alternative likelihood L2.

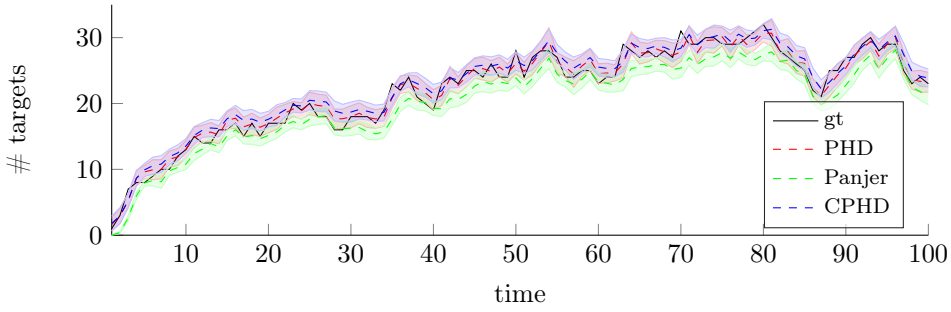
Tab. 2.3 confirms the findings on the runtime analysis of Experiment 1. The decrease in target number in Experiment 2.1 leads to a generally faster performance for all three filters, whereas the increase in target number in Experiment 2.2 results in longer runtimes. In the update, the CPHD filter again runs up to 20 times slower than the PHD filter, whereas the second-order PHD filter needs only twice as much time to run as the PHD filter. Furthermore, the proposed likelihood L1 is consistently faster than the alternative likelihood L2.



(a) Error in the estimated sensor state.



(b) Estimated number of targets, proposed likelihood (L1).



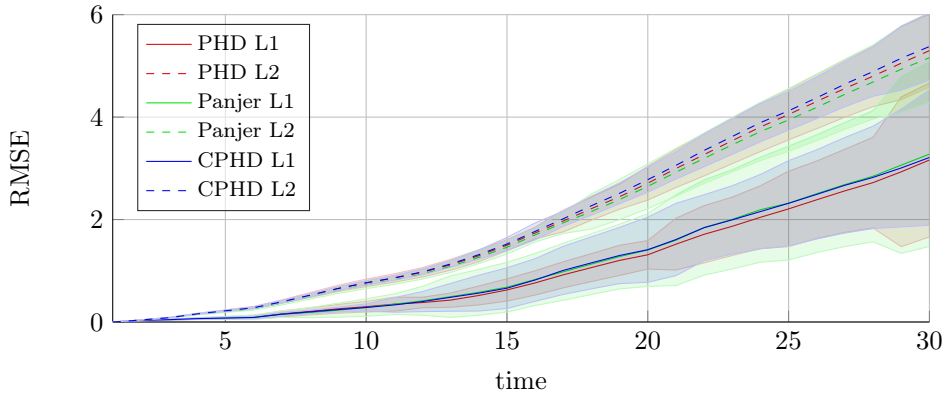
(c) Estimated number of targets, alternative likelihood (L2).

Figure 2.9: Results for Experiment 1 (no model mismatches).

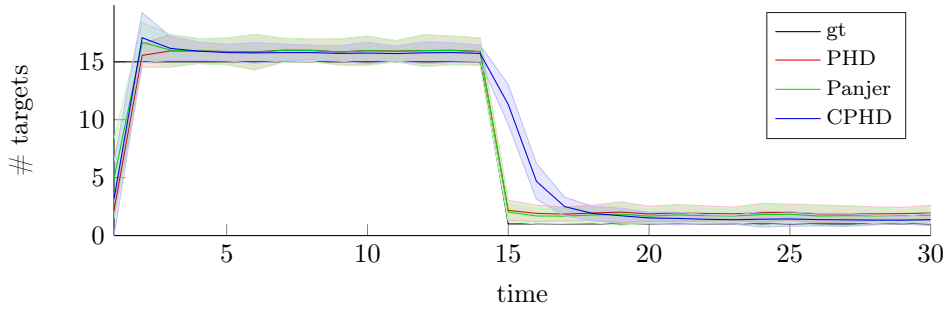
2.4 Exploiting second-order information

2.4.1 Variance, covariance and correlation of point processes

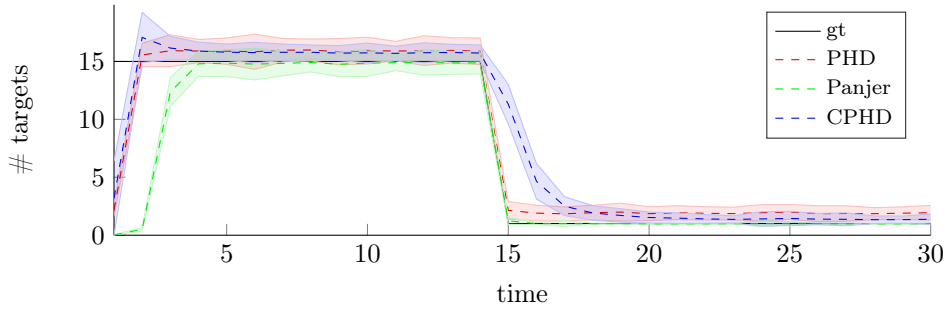
Within the SO-PHD filter framework, the variance was incorporated in the propagation of the target process additionally to the propagation of the intensity. In fact, higher-order information can also be extracted as an additional statistic at any time from any filter as a means of analysing its behaviour. In [Delande et al., 2014], the regional variance for the PHD and CPHD filters was derived and it was demonstrated in simulations that the regional variance is not proportional to the size of the region; in fact, it is not even additive and consequently not a measure. This can be



(a) Error in the estimated sensor state.



(b) Estimated number of targets, proposed likelihood (L1).



(c) Estimated number of targets, alternative likelihood (L2).

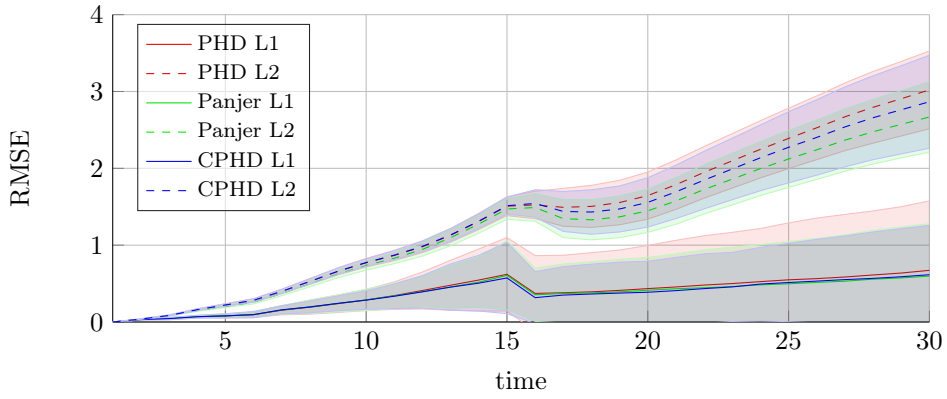
Figure 2.10: Experiment 2.1 (out-of-model target deaths).

demonstrated by looking at the definition of the variance (1.9):

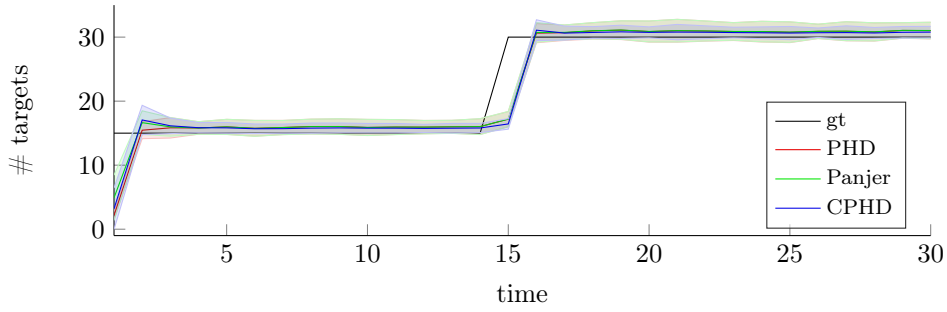
$$\begin{aligned}
\operatorname{var}_\Phi(A \cup B) &= \mu_\Phi^{(2)}((A \cup B) \times (A \cup B)) - [\mu_\Phi(A \cup B)]^2 && \text{(2.100a)} \\
&= \mu_\Phi^{(2)}((A \times A) \cup (A \times B) \cup (B \times A) \cup (B \times B)) - [\mu_\Phi(A \cup B)]^2 && \text{(2.100b)} \\
&= \mu_\Phi^{(2)}(A \times A) + \mu_\Phi^{(2)}(A \times B) + \mu_\Phi^{(2)}(B \times A) + \mu_\Phi^{(2)}(B \times B) - [\mu_\Phi(A) + \mu_\Phi(B)]^2 && \text{(2.100c)} \\
&= \operatorname{var}_\Phi(A) + \operatorname{var}_\Phi(B) + \mu_\Phi^{(2)}(A \times B) + \mu_\Phi^{(2)}(B \times A) - 2\mu_\Phi(A)\mu_\Phi(B) && \text{(2.100d)}
\end{aligned}$$

Here, Equation (2.100c) follows from the additivity of the two moment measures μ_Φ and $\mu_\Phi^{(2)}$. From the result (2.100) it is easy to see that the variance of the union of two sets $A, B \in \mathcal{B}(\mathcal{X})$ can be less, more or equal to the sum of the individual variances, depending on the values of the moments.

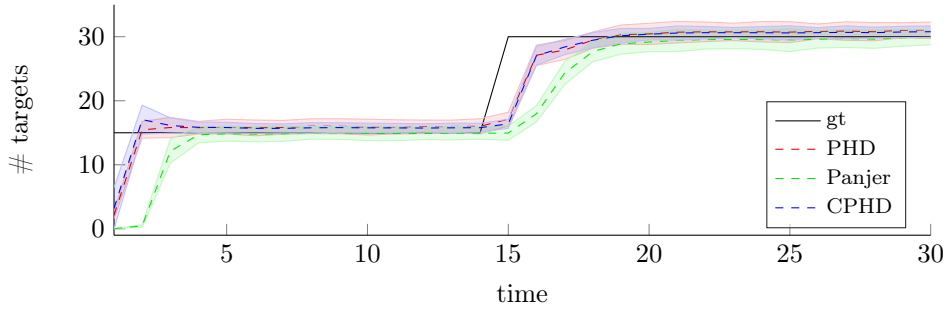
Example 2.4.1. *Let Φ be a point process describing exactly one object, furthermore consider two disjoint regions $A, B \in \mathcal{B}(\mathcal{X})$. Assume that the considered object is either in A with probability p or*



(a) Error in the estimated sensor state.



(b) Estimated number of targets, proposed likelihood (L1).



(c) Estimated number of targets, alternative likelihood (L2).

Figure 2.11: Experiment 2.2 (out-of-model target births).

in B with probability $(1-p)$. The variance in cardinality in each individual region is greater than 0, but the variance in the union $A \cup B$ equals 0 since the object is in the union with probability 1.

Statistical variance is not the only second-order information that is widely studied throughout the literature; also the covariance (1.8) and correlation (1.10) are useful statistical quantities. In the following, we will introduce these concepts in the context of point processes to utilise them to yield a better understanding of the multi-object filters discussed in Sections 1.3 and 2.2.4. Since variance and correlation can be computed via the covariance, let us state the regional covariance for the PHD, CPHD and SO-PHD filters.

Proposition 2.4.2 (Regional covariance [Delande et al., 2014] and [Schlangen et al., 2017c])**.** *Let $B, B' \in \mathcal{B}(\mathcal{X})$ be two Borel sets.*

(a) The PHD filter covariance in B, B' after the update is found to be

$$\mathrm{cov}_k^{\flat}(B \times B') = \mu_k^{\flat}(B \cap B') - \sum_{z \in Z_k} \frac{\mu_k^z(B)\mu_k^z(B')}{\left[\mu_k^z(\mathcal{X}) + \mu_{\mathrm{c},k}(z)\right]^2}, \tag{2.101}$$

where $\bullet = \flat$ *in the definition of the association terms* μ_k^z.

(b) The CPHD filter covariance in B, B' *after the update is given by*

$$\begin{aligned}\mathrm{cov}_k^{\sharp}(B \times B') &= \mu_k^{\sharp}(B \cap B') + \mu_k^{\phi}(B)\mu_k^{\phi}(B')[\ell_2^{\sharp}(\phi) - \ell_1^{\sharp}(\phi)^2] \\ &\quad + \sum_{z \in Z_k} \left[\mu_k^{\phi}(B)\frac{\mu_k^z(B')}{s_{\mathrm{c},k}(z)} + \mu_k^{\phi}(B')\frac{\mu_k^z(B)}{s_{\mathrm{c},k}(z)}\right]\left[\ell_2^{\sharp}(z) - \ell_1^{\sharp}(z)\ell_1(\phi)\right] \\ &\quad + \sum_{z,z' \in Z_k} \left[\frac{\mu_k^z(B)}{s_{\mathrm{c},k}(z)}\frac{\mu_k^{z'}(B')}{s_{\mathrm{c},k}(z')}\right]\left[\ell_2^{\sharp,\neq}(z,z') - \ell_1^{\sharp}(z)\ell_1^{\sharp}(z')\right],\end{aligned} \tag{2.102}$$

where $\bullet = \sharp$ *in the definition of the association and missed detection terms* μ_k^z *and* mu_k^{ϕ} *and*

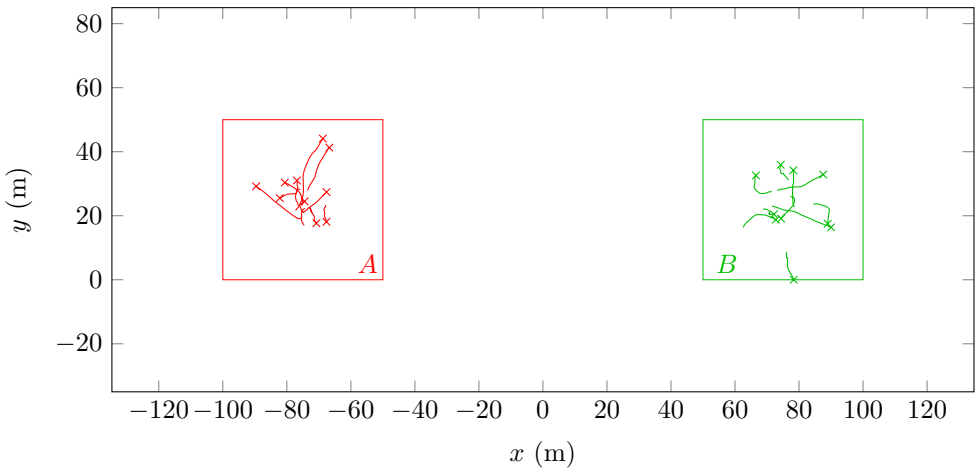
$$\ell_2^{\sharp,\neq}(z,z') := \begin{cases} \dfrac{\langle \Upsilon^2[\mu, Z \setminus \{z,z'\}], \rho_{k|k-1}\rangle}{\langle \Upsilon^0[\mu, Z], \rho_{k|k-1}\rangle}. & \text{if} \quad z \neq z', \\ 0 & \text{otherwise.} \end{cases} \tag{2.103}$$

(c) The covariance $\mathrm{cov}_k^{\natural}(B \times B')$ *of the SO-PHD filter in* B, B' *after the update is equal to* (2.102)*, where* $\sharp$ *is globally replaced by* $\natural$.

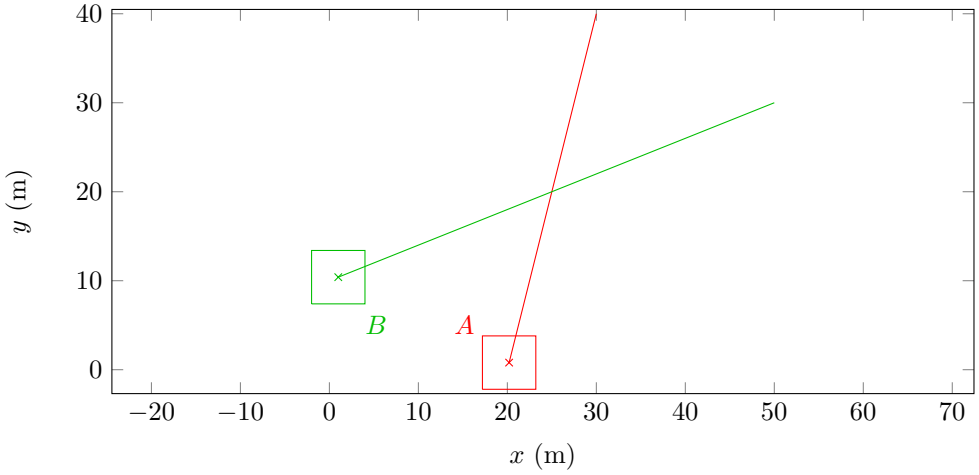
Equations (2.101) and (2.102) stating the covariances $\mathrm{cov}_k^{\bullet}(B, B')$, $\bullet \in \{\flat, \sharp, \natural\}$, are a direct consequence of Equation (1.8), using the first- and second-order moments of the target process. For the PHD and CPHD filters, the second-order moment is given in Equations (29)/(31) and (19)/(29) in [Delande et al., 2014], respectively; the full derivations of the second-order moments and covariances for the SO-PHD filter was already provided in (2.78). From the covariance, the variance is found immediately by setting $B = B'$, see Equation (1.9), and the correlation is found via (1.10) using the definition of variance and covariance.

2.4.2 Application example: analysing the "spooky effect at a distance" [Fränken et al., 2009]

In [Ulmke et al., 2008; Fränken et al., 2009], it was discovered that the CPHD filter shows a counter-intuitive behaviour in the presence of missed detections: for example, if there are two objects that move independently at a great distance from each other and only one of them gets detected at some point in time, then the CPHD filter shifts some of the weight of the miss-detected object to the detected one. An analytical explanation in [Fränken et al., 2009] revealed that the cardinality distribution is updated correctly in such a case but the intensity update draws probability weight towards the detections. This behaviour was coined ibid. as the "spooky effect at a distance" or simply "spookiness" (referring to a similar effect in quantum physics), however a mathematical definition was not established that made this behaviour measurable. Furthermore, the authors stated as a side remark that the PHD filter shows a similar behaviour; this claim was adopted later in [Vo and Vo, 2012] without verification, and [Mahler, 2014] concluded based on a short theoretical analysis on a minimal example that while being superpositional, the PHD filter shows even more "spookiness" than the CPHD filter since "the weight of the undetected track is even smaller than it should be, and this missing weight has been shifted to the detected track" [Mahler, 2014, p. 221].



(a) Experiment S3.1: Static regions A (red) and B (green).



(b) Experiment S3.2: Moving regions A (red) and B (green).

Figure 2.12: Ground truth for the two experiments that analyse the spooky effect of the three filters.

So far, while this interesting effect has been identified and studied in the aforementioned works, the literature still lacks an objective means of quantifying it.

In this section, two experiments will be conducted which exploit the regional correlation (1.10) as a quantitative measure of spookiness in the PHD, CPHD and SO-PHD filters. For this purpose, the filter-dependent regional covariance equations (2.101) and (2.102) are utilised. The first experiment focuses on the presence of missed detections on the intensities in two well-separated regions, whereas the second experiment analyses the effect of the distance and overlap of two regions on their filter correlation.

S3.1 Spooky effect at a distance:

Consider two two-dimensional regions $A, B \in \mathcal{B}(\mathcal{X})$ of size $50\,\mathrm{m} \times 50\,\mathrm{m}$ which are $100\,\mathrm{m}$ apart along the x axis to avoid interferences. In each of these regions, 10 objects are created in the first time step and propagated with survival rate $p_\mathrm{s} = 1$ for 100 time steps using a near-constant velocity model assuming an acceleration noise of $0.1\,\mathrm{m\,s^{-2}}$ in each dimension. Their initial velocities are Gaussian normal distributed with mean 0 and standard deviation $0.3\,\mathrm{m\,s^{-1}}$ along each dimension of the state space. The setup is depicted in Figure 2.12a. From

these objects, measurements are generated using a detection rate of $p_{\mathrm{d}} = 0.9$ and measurement noise with standard deviation 0.2 m in each dimension. Furthermore, the clutter process is assumed Poisson with a clutter rate of 20 in each region. In order to provoke the spooky effect, all objects in region B are intentionally miss-detected every tenth time scan.

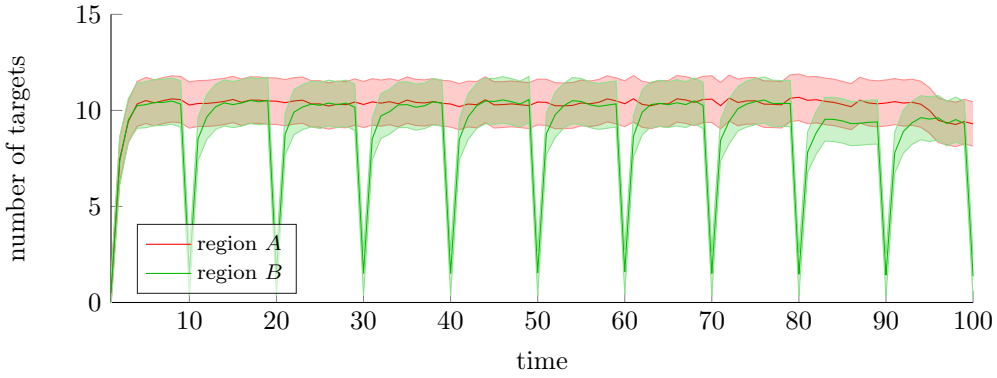
The PHD, CPHD and SO-PHD filters are initialised with the parameters used for the simulation, assuming a survival rate of 0.98 and a negative binomial birth process with intensity $\mu_{\mathrm{b}}(\mathcal{X}) = 1$ and variance $\mathrm{var}_b(\mathcal{X}) = 100$. The regional means and variances in each region estimated by the three filters are shown separately in Figures 2.13a, 2.13b and 2.13c. The PHD filter shows no effect on region A when the objects in region B are miss-detected (see Figure 2.13a); the CPHD filter, in contrast, experiences considerable peaks in intensity in region A when the intensity in region B drops (see Figure 2.13b). For the SO-PHD filter, on the other hand, the estimated number of objects in A drops slightly when the objects in B are not detected, and it shows a slight overshoot when the intensity in B goes back up in the subsequent time step (see Figure 2.13c). In other words, the change in intensity in both regions is strongly negatively proportional for the CPHD filter and slightly positively proportional for the SO-PHD filter. These findings can be visualised neatly in terms of correlation as seen in Figure 2.13d: Here, the correlation of the PHD filter between A and B stays at 0 over the whole experiment, whereas it shows negative values for the CPHD filter, especially at the time instances where region B is undetected, and slightly positive values for the SO-PHD filter.

S3.2 Influence of distance on the spooky effect:

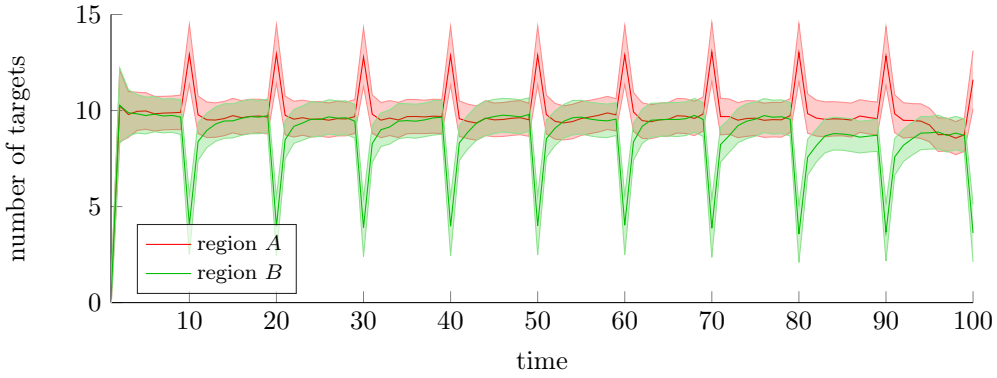
The results of Scenario S3.1 let us wonder what would happen in terms of correlation if the regions draw closer to each other, and which conclusions one could draw for the spooky effect of the three filters, and how the birth variance affects the results. Therefore, let us consider again a square, two-dimensional state space of size $50\,\mathrm{m} \times 50\,\mathrm{m}$ that is monitored for 50 time steps with time lapse 1 s. In the beginning, two objects are generated at the initial positions $(30, 40)$ and $(50, 30)$, surviving throughout the whole experiment with probability $p_{\mathrm{s}} = 1$ and moving according to a constant velocity model with velocities $(-0.2, -0.8)$ and $(-1, -0.4)$, respectively. Their true trajectories are shown in Figure 2.12b. This time, measurements are generated with detection rate $p_{\mathrm{d}} = 0.9$ without artificial missed detections, and the measurement noise was assumed to be 1 m in each dimension of the state space. The false alarm rate was set to 2 false alarms per frame.

The three filters were initialised with the same parameters used for the simulation, setting the initial velocities to be Gaussian normal distributed with mean $0\,\mathrm{m\,s^{-1}}$ and standard deviation $1\,\mathrm{m\,s^{-1}}$. Furthermore, the birth model was set negative binomial with mean 0.4 and three different variances 1, 2 and 10, and the survival rate was assumed to be 0.98. The acceleration noise was set to $0.01\,\mathrm{m\,s^{-2}}$.

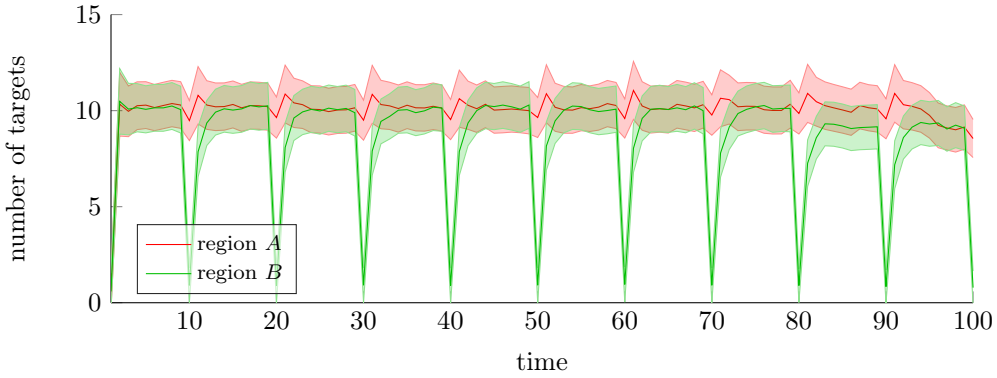
Instead of assuming fixed regions like in Scenario S3.1, two square regions A and B of size $6\,\mathrm{m} \times 6\,\mathrm{m}$ were centered on the true position of the targets at every time scan, hence moving them along the target trajectories. Furthermore, the trajectories cross in $(25, 20)$ at time 26



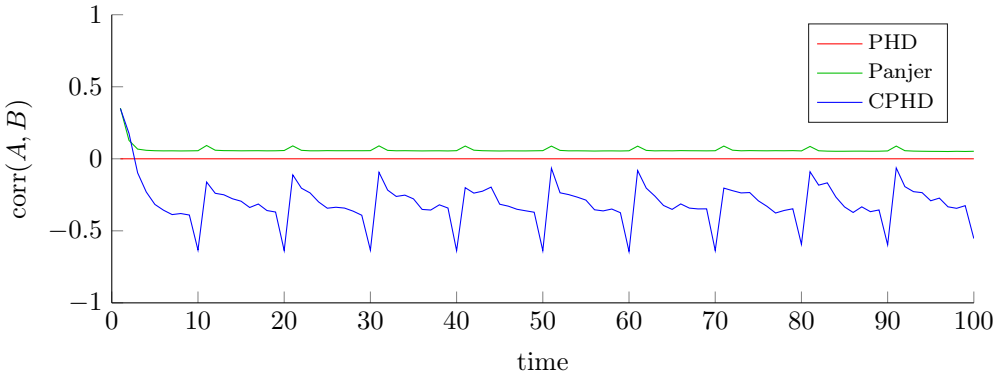
(a) Mean and standard deviation of the estimated target number, PHD filter.



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



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

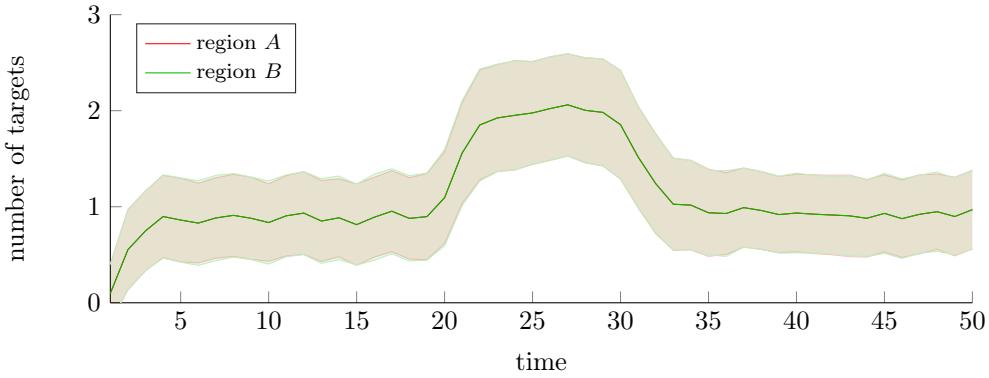


(d) Correlation between the estimated number of targets in regions A and B.

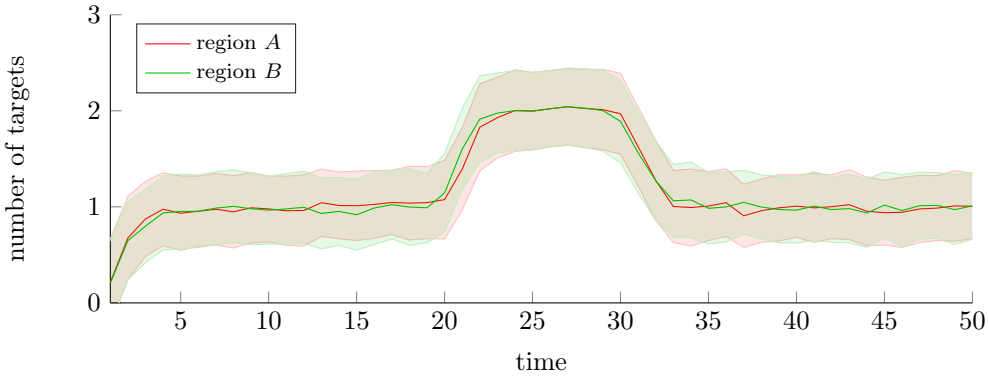
Figure 2.13: Results for Scenario S3.1, averaged over 100 MC runs. Figures 2.13a, 2.13b and 2.13c show the mean and standard deviation of the estimated number of targets per region for the three filters. Figure 2.13d shows the respective correlations of A and B according to each filter.

such that the two regions perfectly align there. Figures 2.14, 2.15 and 2.16 show the simulation results in terms of the estimated number of targets in both regions and correlation over time. The correlation plots additionally show the percentage-wise overlap of the two regions and one minus the Hellinger distance between the filtered intensities of the two objects.

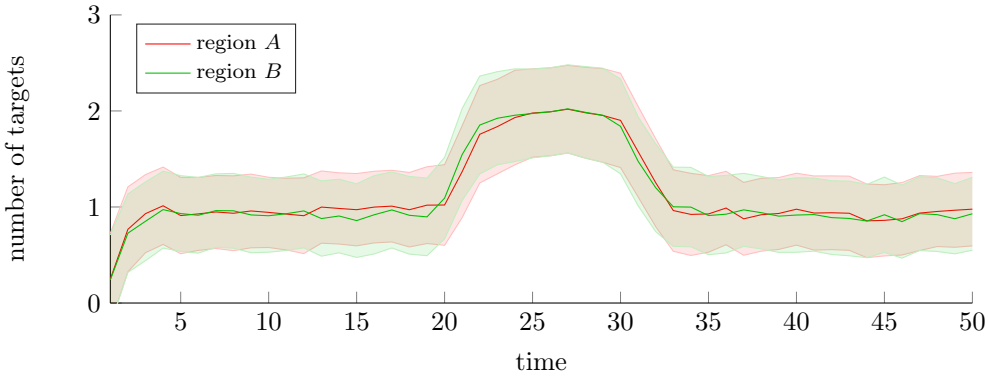
It can be seen in all experiments that the estimated intensities change from 1 to 2 as soon as both targets are present in each region. For increasing birth variance, the estimated intensities drop slightly in case of the CPHD filter and grow slightly for the SO-PHD filter; this behaviour might be tied to the regional correlation of the filters: for the CPHD filter, the correlation decreases considerably with increasing birth variance, hence suppressing the estimated cardinalities in both regions. The regional correlation of the SO-PHD filter, in contrast, increases for increasing birth variance which seems to have an enhancing effect on the estimated cardinality. The PHD filter cannot incorporate variance information, furthermore its correlation equals 0 while there is no overlap in the regions. Regarding the changing distance of the two regions, it can be noted that for all three filters, the correlation already grows before the regions actually start to overlap, so the vicinity around the regions already influences the estimation inside the regions.



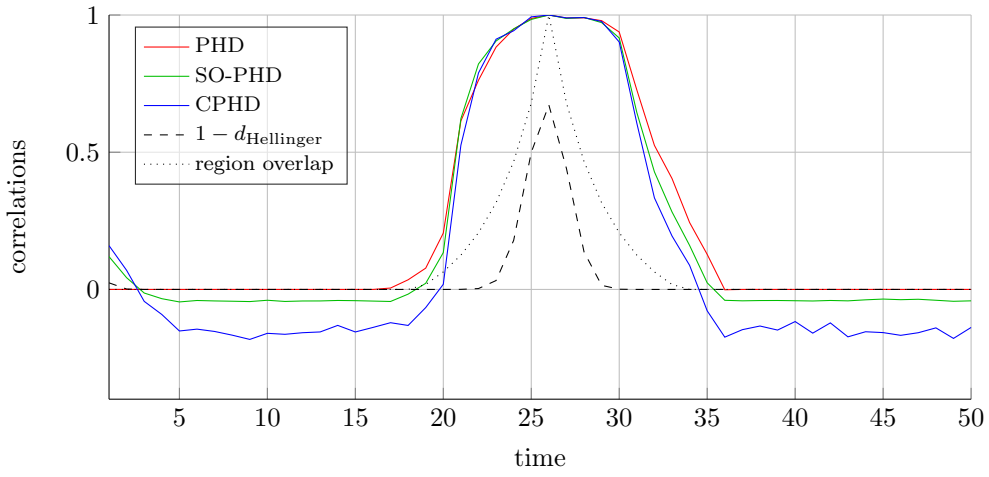
(a) Mean and standard deviation of the estimated target number, PHD filter.



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

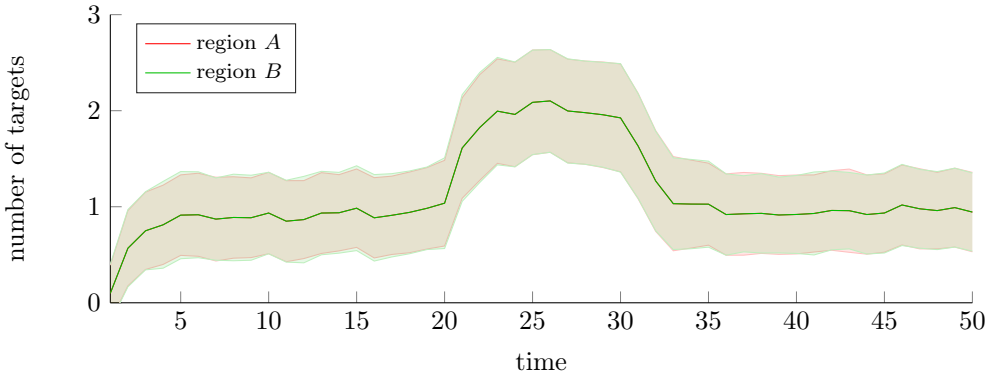


(c) Mean and standard deviation of the estimated target number, SO-PHD filter.

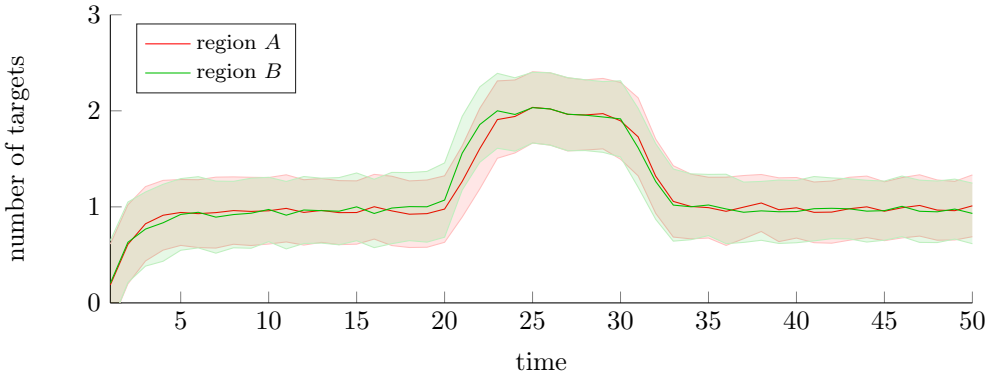


(d) Correlation results between the estimated number of targets in regions A and B.

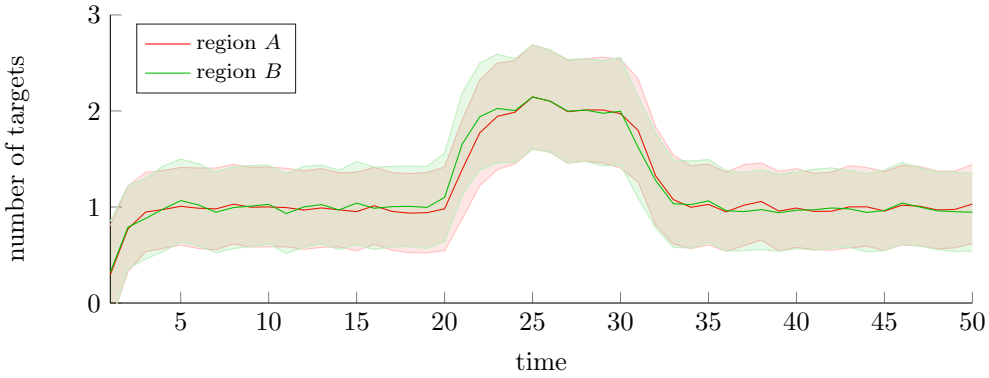
Figure 2.14: Results for Scenario S3.2 with birth variance $\text{var}_\text{b} = 1$, averaged over 100 MC runs. Figures 2.14a, 2.14b and 2.14c show the mean and standard deviation of the estimated number of targets per region for the three filters. Figure 2.14d shows the respective correlations of A and B according to each filter.



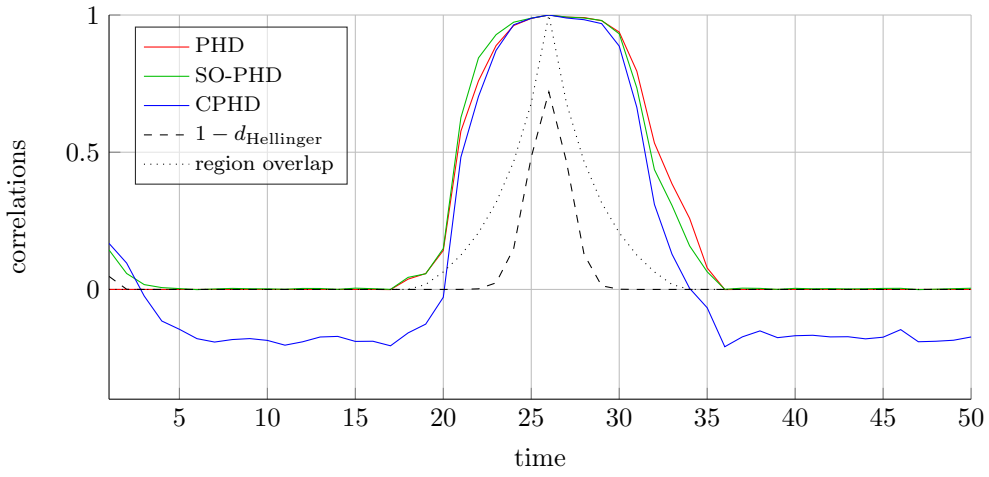
(a) Mean and standard deviation of the estimated target number, PHD filter.



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



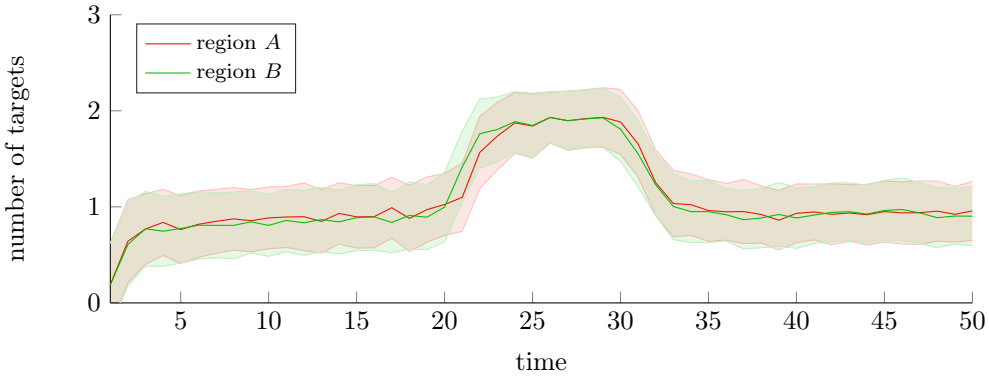
(c) Mean and standard deviation of the estimated target number, SO-PHD filter.



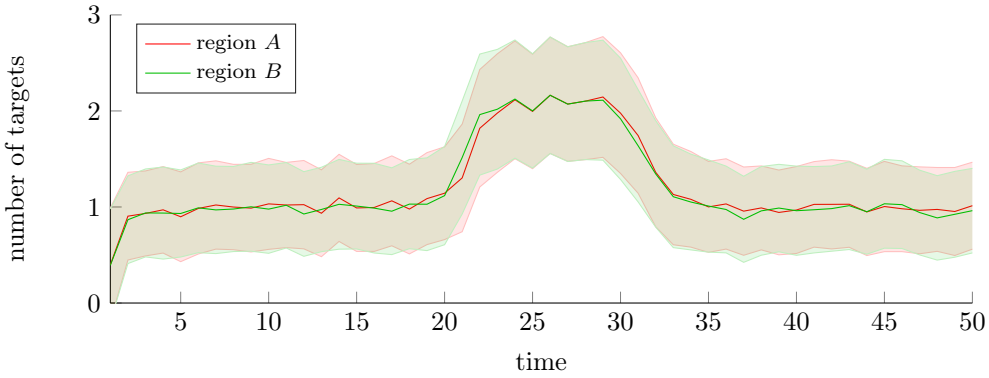
(d) Correlation results between the estimated number of targets in regions A and B.

Figure 2.15: Results for Scenario S3.2 with birth variance $\text{var}_\text{b} = 2$, averaged over 100 MC runs. Figures 2.15a, 2.15b and 2.15c show the mean and standard deviation of the estimated number of targets per region for the three filters. Figure 2.15d shows the respective correlations of A and B according to each filter.

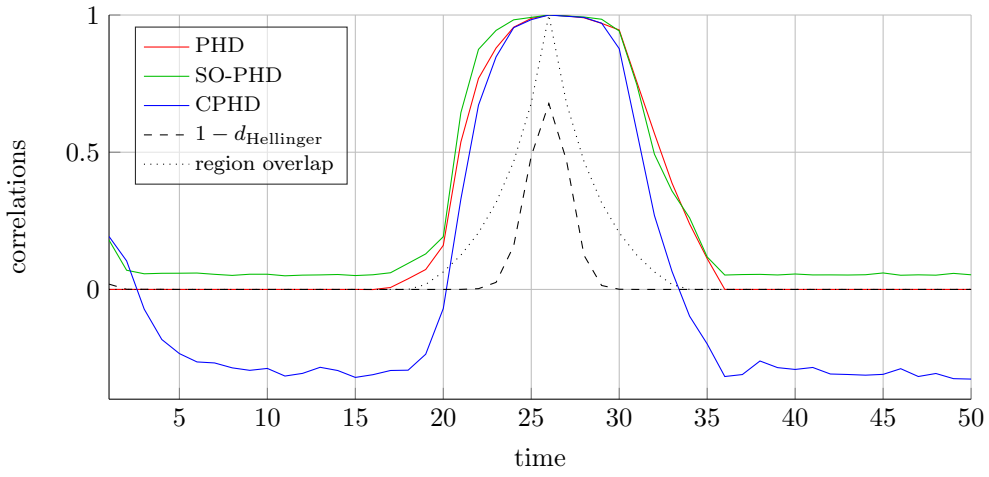
(a) Mean and standard deviation of the estimated target number, PHD filter.



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



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



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

Figure 2.16: Results for Scenario S3.2 with birth variance $\text{var}_\text{b} = 10$, averaged over 100 MC runs. Figures 2.16a, 2.16b and 2.16c show the mean and standard deviation of the estimated number of targets per region for the three filters. Figure 2.16d shows the respective correlations of A and B according to each filter.

Chapter 3

Challenges in super-resolution microscopy

So far, the findings of Chapter 2 have assessed the performance of the family of PHD filters only in simulations, however it will be important to study their potential on real data as well. As mentioned in the introductory remarks, one type of applications that is particularly prone to non-Poisson behaviour in the number of objects and false alarms is single-molecule fluorescence microscopy, which we will focus on in this chapter.

Recent developments in optical microscopy have made it possible to monitor intracellular processes at the nano-scale, either by using superlenses [Luo and Ishihara, 2004; Fang et al., 2005], or by crossing the diffraction limit of light using active illumination techniques. The basic idea of active illumination is to label some protein of interest with a photo-activatable fluorescent marker and resolve occurring signal overlaps using, for example, excitation and depletion (Stimulated Emission Depletion (STED) microscopy [Hell and Wichmann, 1994]), structured illumination (Structured Illumination Microscopy (SIM) [Gustafsson, 2000]), or stochastic downsampling of the signal (Photo-Activated Localisation Microscopy (PALM) [Betzig et al., 2006; Hess et al., 2006] and Stochastic Optical Reconstruction Microscopy (STORM) [Rust et al., 2006]). We will focus on PALM below since this technique makes it possible to study molecule dynamics over time in live cells, as opposed to many other techniques which only work on fixed (dead) cells. Target tracking (which is usually called *linking* in the biomedical literature) makes it possible to extract desired statistics such as object number, dynamics, life time, speed, etc. from PALM image sequences in an automatic and principled manner. If PALM is combined with tracking techniques, it is referred to as single particle tracking PALM (sptPALM) [Manley et al., 2008]. Note that in this context, the term *particle tracking* has nothing to do with SMC techniques but simply refers to tracking movements of point-like structures in a medium. In the following, however, the term *particle* will always denote a random sample of a probability distribution, and the tracked objects will be called *molecules*, *objects* or *targets*.

As mentioned earlier, tracking/linking is usually still performed using heuristic nearest-neighbour techniques [Meijering et al., 2006], however the manifold ambiguities in single-molecule fluorescence

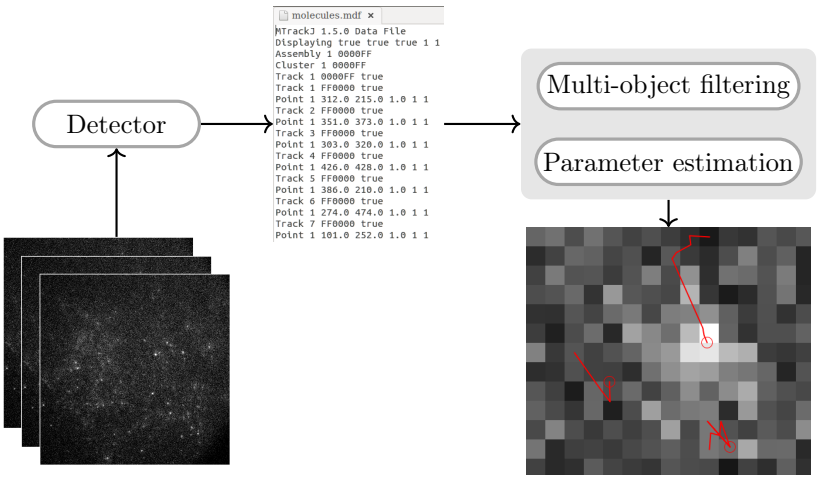


Figure 3.1: Data flow of the data processing of PALM image sequences.

microscopy images makes those techniques highly error-prone since they cannot cope well with target crossings, false alarms and missed detections. In this chapter, we propose an algorithm to process PALM data using the PHD filter methods discussed in Chapters 1 and 2. The basic data flow is shown in Figure 3.1; first, the image sequence is pre-processed using a suitable detection algorithm [Smal et al., 2010] and point-like measurements are extracted from the image. These measurements usually describe 2-dimensional position coordinates in the image, but other statistics such as intensity, size, shape, etc. could be extracted and estimated as well. These detections are forwarded a the multi-object estimation algorithm of choice (e.g. the PHD, CPHD, or SO-PHD filter or their respective single-cluster versions for additional parameter estimation) which gives the desired statistics as its final output.

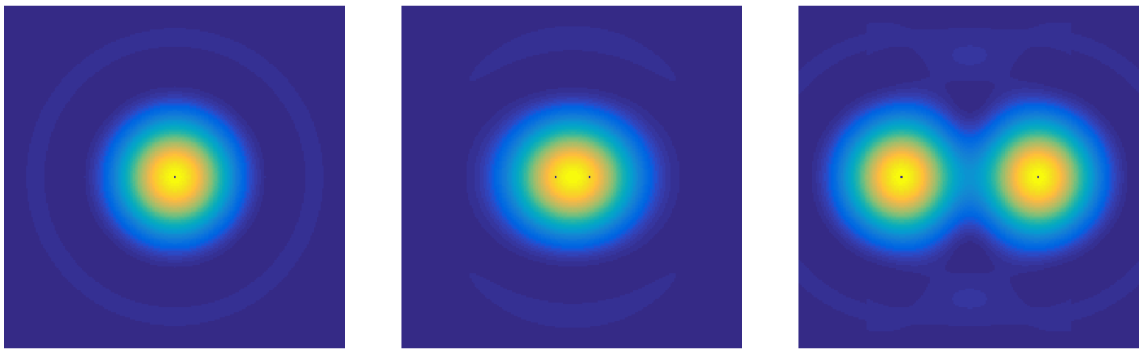
In Section 3.1.1, the PALM technique will be very briefly explained to highlight the difficulties that arise in the data processing. Section 3.1.2 proposes a simple technique to extract detections from PALM images, and the subsequent sections describe the conducted experiments on real data.

3.1 Acquisition and processing of PALM images

3.1.1 On the nature of PALM data

In PALM [Betzig et al., 2006], the molecules of interest are transfected with a fluorescent marker which is *photo-activatable*, i.e. fluorescence is activated through illumination with laser light of a certain wavelength which is specific to the used fluorophore, then switched on by laser light of a second specific wavelength and switched off permanently with laser light of a third specific wave length. Since the fluorophores are tiny point-shaped light sources beyond the diffraction limit of light, their emitted signal is blurred with the point spread function of the microscope which is dependent on the numerical aperture and the optics of the lenses used in the microscope. This convolution leads to an optical signal in the shape of an Airy disk (Fig. 3.2a) [Betzig et al., 2006; Rust et al., 2006]. Now, if two such signals are too close to each other, i.e. closer than half the emitted wavelength, their Airy disks overlap significantly so that they cannot be resolved (Figure 3.2b). To achieve super-resolution, a method has to be found to detect each signal separately,

either by spatial or temporal distance between the detections.



(a) Airy disk originating from one point source.

(b) Airy disks of two point sources close to each other.

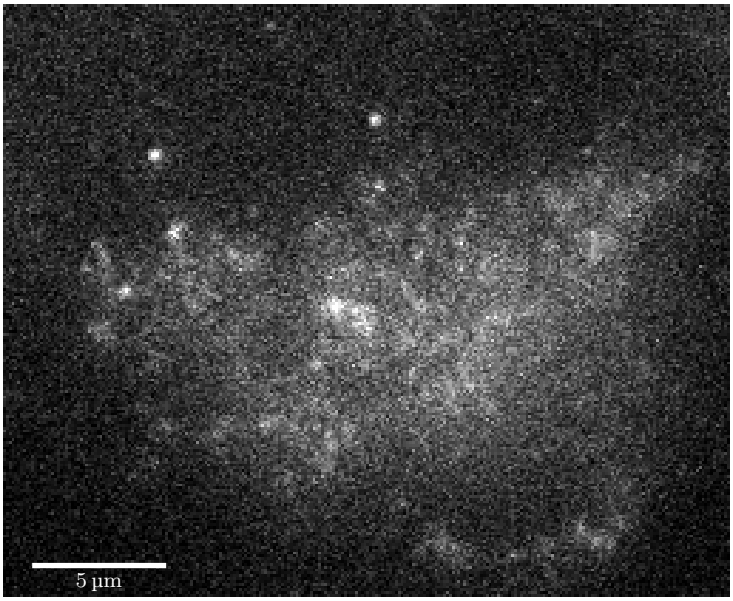
(c) Airy disks of two point sources far from each other.

Figure 3.2: Airy disks of objects beyond the diffraction limit. If the objects of interest are so close together that their Airy disks overlap substantially, it is almost impossible to distinguish them (Fig. 3.2b), however if they are reasonably far away from each other, their locations can still be resolved accurately (Fig. 3.2c).

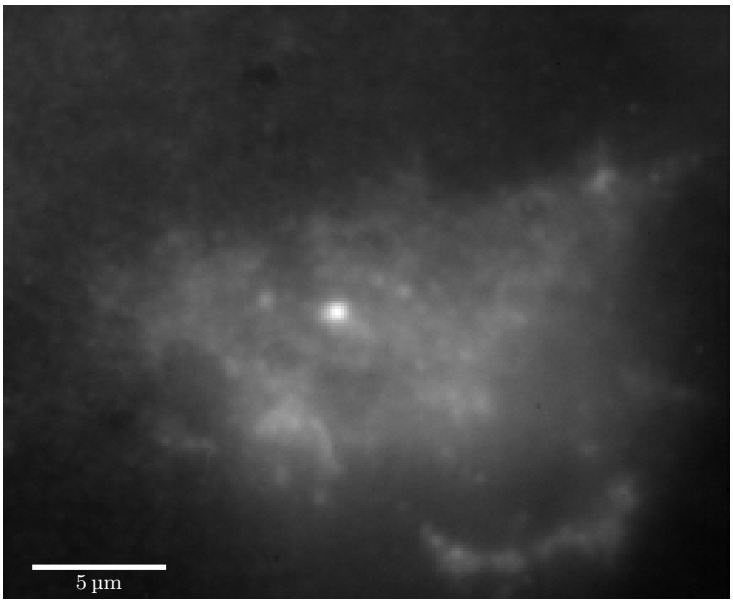
The photo-activation in PALM is in a way "stochastic", i.e. only a small, random subset of all fluorophores is brought to the active state, whereas the de-activation laser switches all activity off at once. Like this, sub-diffraction structures can be resolved using temporal separation since clustered molecules are imaged successively, hoping to keep simultaneous excitation of objects in close proximity to each other to a minimum. Furthermore, once a fluorophore changes to the dark state, it is bleached for the rest of the experiment, therefore each detection belongs to a distinct object. STORM is quite similar to PALM, however fluorophores are attached through immuno-labelling and they can switch between the light and dark state multiple times [Rust et al., 2006] which makes it impossible to count or label objects; it is rather used for imaging subcellular structures like microtubules. Both techniques and their various derivatives result in images similar to Fig. 3.3, showing the molecules of interest as bright spots on noisy background.

While the idea behind PALM works perfectly fine in theory, there are a number of challenges that makes the resulting images hard to interpret. First of all, there is no guarantee that objects with close proximity to each other are not activated at once, and it is hard to tell from an image like 3.3 how much signal overlap is still recorded. Furthermore, since the fluorescent signal is quite weak and high frame rates are needed to monitor quick intracellular processes, the recorded PALM images usually suffer from a low SNR and autofluorescence of the background is a common problem leading to a non-homogeneous background profile. Another challenge arising from the small size of the monitored objects is the problem of stage drift which can be caused by small movements in the system such as thermal expansion or motor vibrations.

One way of reducing the background signal is by using Total Internal Reflection Fluorescence (TIRF) microscopy [Ambrose, 1956]: in this imaging technique, a high-power laser beam is totally internally reflected on the cover slip and the evanescent field around the beam only excites a very thin slice (usually less than 200 nm) of the cell membrane that touches the cover slip. Therefore, TIRF microscopy can remove most of the background fluorescence which is particularly useful when cell signalling at the membrane is studied. All examples studied in this chapter are acquired using the TIRF technique on an Olympus Cell Excellence wide-field microscope and an EMCCD



(a) One frame.



(b) Sum across the stack. Note that most of the detail is blurred out.

Figure 3.3: PALM dataset showing fluorescent N-type calcium channels, image dimensions 27.14 µm× 22.15 µm. Figure 3.3a shows one frame of the sequence, and Figure 3.3b shows the sum of all 4000 frames. The latter is equivalent to the output if all molecules were fluorescing at once, demonstrating how PALM makes it possible to increase resolution by looking only at a few molecules at a time.

camera with a pixel width corresponding to 106 nm.

Regarding stage drift, the findings of Sections 1.4 and 2.3 about hierarchical processes can be exploited to track the movement of the sensor simultaneously with the molecular movement. The same technique can also be applied to estimate other unknown parameters such as the false alarm rate [Schlangen et al., 2017a].

3.1.2 Image processing

Given a sequence of PALM images, the first step is to extract sets of point-like measurements representing the molecules of interest per frame. Even though in general, signal overlaps cannot be ruled out completely, let us assume in the following that each blob corresponds to one distinct fluorophore. Following Smal et al. [2010], spot detection is performed in three steps:

1. *Noise reduction*: Readout and photon noise have to be removed from the input image, using techniques like Gaussian blurring or wavelet denoising.

2. *Signal enhancement*: An image enhancement step can be used to emphasise the structures of interest in the de-noised image and to remove the background.

3. *Signal thresholding*: Thresholding will give a binary output that identifies regions of interest in the image.

The method of choice for the experiments below is the so-called Isotropic Undecimated Wavelet Transform (IUWT) which uses a wavelet decomposition to extract blob-like structures of a certain size/frequency from the images through band-pass filtering. Since low-frequency information (i.e. background activity) can be selectively left out in the reconstruction of the enhanced image, this technique combines all three enhancement steps described above in the following manner: Let $\mathcal{I}$ denote the pixel matrix of the input image which is given by

$$\mathcal{I} = \{I(i,j)\}_{i\in\{1\dots m\},j\in\{1\dots n\}} \in \mathbb{R}^{m\times n}, \tag{3.1}$$

where $I(i,j)$ is the intensity of the pixel at position (i,j), and define the one-dimensional kernel

$$H = \left[\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16}\right]. \tag{3.2}$$

By convolving $\mathcal{I}$ row- and columnwise with H for K times, we obtain a sequence $\{\mathcal{I}_k\}_{k\in\{1\dots K\}}$ which leads in turn to the sequence of the form

$$\mathcal{W}_k = \mathcal{I}_{k-1} - \mathcal{I}_k, \tag{3.3}$$

obtained with the pixel-wise differences $W_k(i,j) = I_{k-1}(i,j) - I_k(i,j)$ for $i \in \{1\dots m\}$ and $j \in \{1\dots n\}$. The sequence $\{W_k\}_{k\in\{1\dots K\}}$ is called the *à trous wavelet decomposition* with the property

$$\mathcal{I} = \mathcal{I}_K + \sum_{k=1}^{K} \mathcal{W}_k. \tag{3.4}$$

The matrices $\mathcal{W}_k$ encode information of different frequencies in the image, i.e. structures of different sizes. The background can be suppressed by removing the low-frequency bottom layer $\mathcal{I}_k$ from the sequence, furthermore the high-frequency noise contained in the top layer $\mathcal{W}_1$ can be removed as well. A sequence $\{\hat{\mathcal{W}}_k\}_{k\in\{2\dots K\}}$ is obtained from the remaining layers by using a user-defined threshold, hence producing the final output

$$\mathcal{I}_{\mathrm{IUWT}} = \sum_{k=2}^{K} \hat{\mathcal{W}}_k. \tag{3.5}$$

Figure 3.4a shows the IUWT of Figure 3.3. There are different ways of extracting point-like measurements from the enhanced image $\mathcal{I}_{\mathrm{IUWT}}$. The most naive approach is to extract local maxima or the centre of mass of each blob in the image, another way is to fit a sum of Gaussians on the image, using the fact that the Airy disk can be approximated well by a Gaussian bell curve [Zhang et al., 2007]. Both the centroid and the Gaussian fit approaches are demonstrated in Figures 3.4b and 3.4c; in the following, the centroid approach is preferred since, apart from being easier to compute, it detects all shapes in the image whereas the Gaussian fit approach seems to miss some of the objects.

3.2 Applications

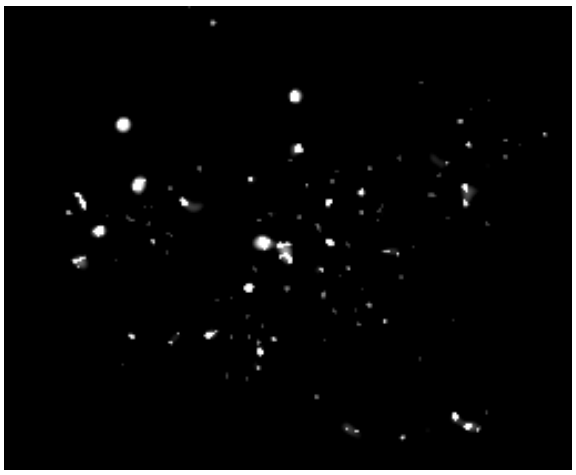
This chapter gives three examples in which the PHD filter methods are shown to be useful alternatives to the current state of the art used in cell and molecule tracking. The first application (Section 3.2.1) shows a comparison of the PHD, CPHD and SO-PHD filters for pure molecule tracking; the other two experiments use variations of the single-cluster PHD filter to estimate stage drift (Section 3.2.2) and the unknown clutter rate in the image sequence (Section 3.2.3).

3.2.1 Application 1: Tracking molecules

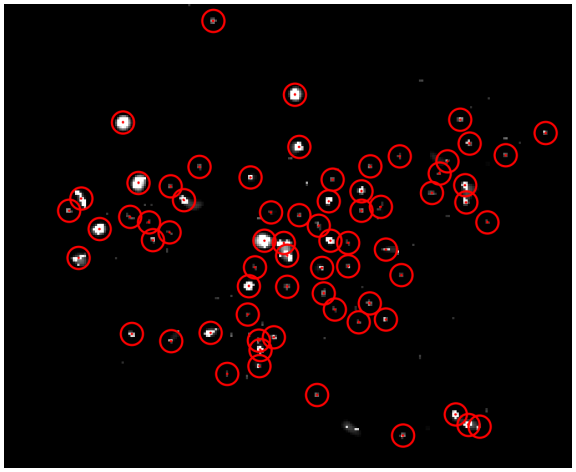
The following experiment is conducted on live PALM data showing N-type calcium channels[1] on the membrane of a tumour cell of the PC12 cell line.[2] In order to make the channels visible in the data, the cell culture is transfected with photoactivatable mCherry (PAmCherry), a red-fluorescent protein which is brought to the active state using laser light with a wavelength of around $350-400\,\mathrm{nm}$ and whose fluorescence is then triggered using laser light with a wavelength of maximally $595\,\mathrm{nm}$. In this particular experiment, $405\,\mathrm{nm}$ laser-activated PAmCherry was visualized using $561\,\mathrm{nm}$ wavelength laser excitation. Each activation-excitation cycle lasts for 100 frames; the total number of frames is 4000, but for this experiment, only the first 500 frames are considered. The extracted images are of size $27.14\,\mu\mathrm{m}\times 22.15\,\mu\mathrm{m}$ which corresponds to image dimensions of $256\,\mathrm{px}\times 209\,\mathrm{px}$. Figure 3.3a shows a sample frame of the considered sequence.

From this dataset, detections are extracted across all 500 frames in the manner of Section 3.1.2. The PHD, CPHD and SO-PHD filters are run on the extracted measurements based on the following parametrisation: Targets are assumed to move according to an n.c.v. model with an

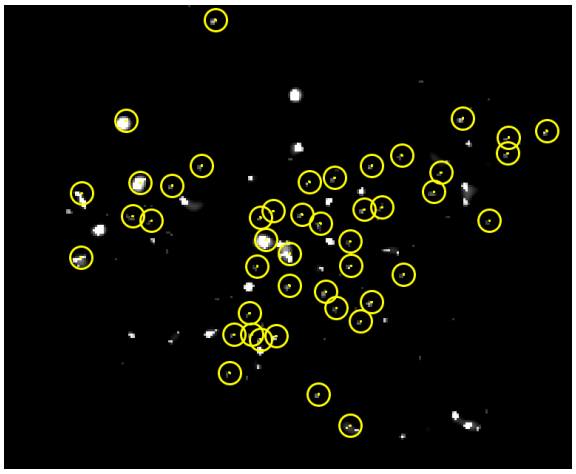
[1] N-type calcium channels are important for signalling of the nervous system.
[2] Thanks to Katarzyna Ciałowicz for providing the images.



(a) À trous wavelet denoising.



(b) Centroids of the detected blobs, excluding those with less than 3 pixels.



(c) Gaussian fits on the denoised image.

Figure 3.4: The detection process on the example of Fig. 3.3.

acceleration noise of 0.1 px per frame squared, and the initial target velocity is assumed Gaussian normal distributed with mean 0 and a small standard derivation of 0.05 px per frame for both image dimensions. Molecules are supposed to stay fluorescent with survival probability $p_s = 0.9$, and new fluorophores are assumed to start emitting light according to a negative binomial birth model of mean 10 and variance 80, accounting for unexpected bursts in the number of objects due to repeated activation-excitation cycles. Furthermore, the measurement noise is set to 1 px in each

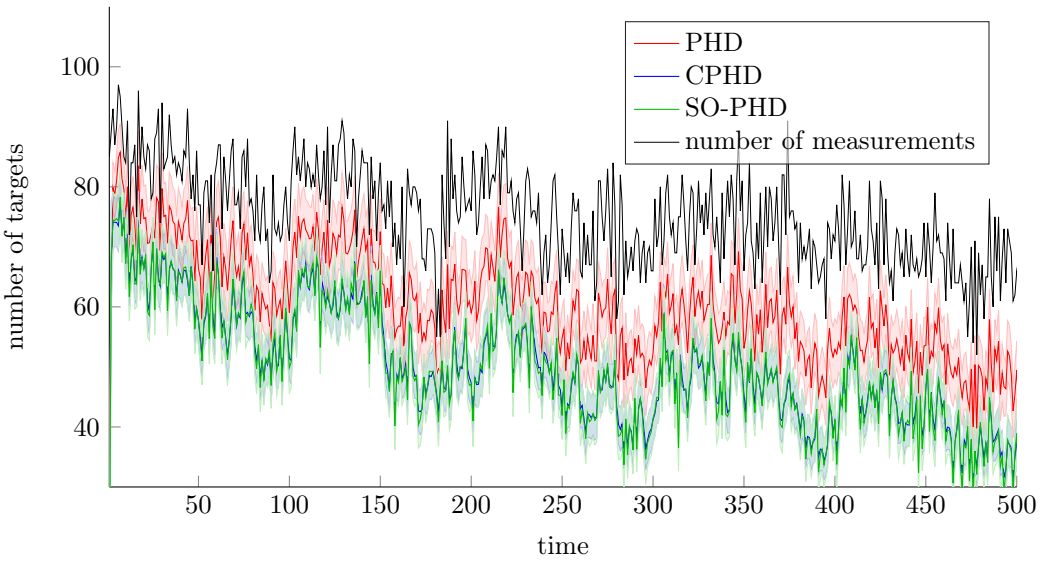


Figure 3.5: Estimated mean and variance in the number of targets over time for the live PALM dataset.

image dimension, and false alarms are modelled using a negative binomial distribution with mean 10 and variance 60. The estimation results for all three filters are shown in Figure 3.5 in terms of the estimated mean and variance in the number of targets. The number of extracted measurements is plotted in the same coordinate system to give an idea of the data flow.

The results reflect the experimental setup of PALM as the plot shows: in every excitation cycle, the number of active objects is growing quickly and then slowly reducing over time; the overall number of objects also decreases since more and more objects are transferred permanently to the dark state. It can be noticed that the PHD filter consistently produces a higher estimate in the number of objects than the CPHD and SO-PHD filters; it seems that the missing information on the variance in the number of births and false alarms brings the PHD filter to trust the observations more, whereas the other two filters are more reluctant to interpret incoming data as true targets. This confirms the findings in Section 2.2.6 and Vo et al. [2007].

3.2.2 Application 2: Stage drift estimation

The following experiment was conducted in [Schlangen et al., 2016b] to demonstrate the value of the single-cluster PHD filter with multiple motion models for the calibration of PALM data which is subject to stage drift. The idea is to utilise only the static objects in the image for the calibration since they give more reliable feedback on the sensor movement. Therefore, multiple motion models are incorporated in the PHD filter recursion as follows: for all N_{mm} considered motion models with index $i \in 1, \ldots, N_{\mathrm{mm}}$, the prediction is performed according to (1.38) using a model-specific birth intensity $\mu_{\mathrm{b},k}^{(i)}$ and exchanging the Markov transition $t_{k|k-1}^{(i)}$ in the survival term, such that the prediction becomes

$$\mu_{k|k-1}^{\flat,(i)}(B) = \mu_{\mathrm{b},k}^{(i)}(B) + \mu_{\mathrm{s},k}^{(i)}(B) \tag{3.6}$$

with modified survival term

$$\mu_{\mathrm{s},k}^{(i)}(B) = \int_{\mathcal{X}} p_{\mathrm{s},k}(x) t_{k|k-1}^{(i)}(B|x) \mu_{k-1}^{\flat}(\mathrm{d}x). \tag{3.7}$$

In the update, the normalisation is performed over *all motion models*, so that (1.39) is changed into

$$\mu_k^{\flat,(i)}(B) = \mu_k^{\phi,(i)}(B) + \sum_{z \in Z_k} \frac{\mu_k^{z,(i)}(B)}{\mu_{\mathrm{c},k}(z) + \sum_{j=1}^{N_{\mathrm{mm}}} \mu_k^{z,(j)}(\mathcal{X})}, \tag{3.8}$$

where $\mu_k^{\phi,(i)}$ and $\mu_k^{z,(i)}$ are obtained by performing Equations (1.36) and (1.37) on the prediction intensities $\mu_{k|k-1}^{\flat,(i)}$ according the respective motion model i.

For the experimental setup, a PALM dataset is used which shows SNAP25[3] activity in human embryonic kidney cells from the 293 cell line.[4] The cells are transfected with the red-fluorescent protein PAmCherry again, and the cover slip contains randomly placed gold beads which serve as fiducial markers to aid with the detection of possible stage drift. The images represent a square area of size $(54.272\,\mu\mathrm{m})^2$ which corresponds to an image size of $(512\,\mathrm{px})^2$, and they are recorded at a sampling rate of 16.6 Hz over 100 frames. An example frame is shown in Figure 3.6a. Detections are again obtained using the technique described in Section 3.1.2.

Since the fiducial markers are placed randomly over the image, it is possible find two sub-images one of which contains fiducial markers and one of which does not; like this, it can be assessed to which extent the single-cluster PHD filter relies on fixed points in comparison to image calibration methods which are conventionally used for the drift estimation. The two chosen sub-images are presented in Figure 3.6c, having the following specifications:

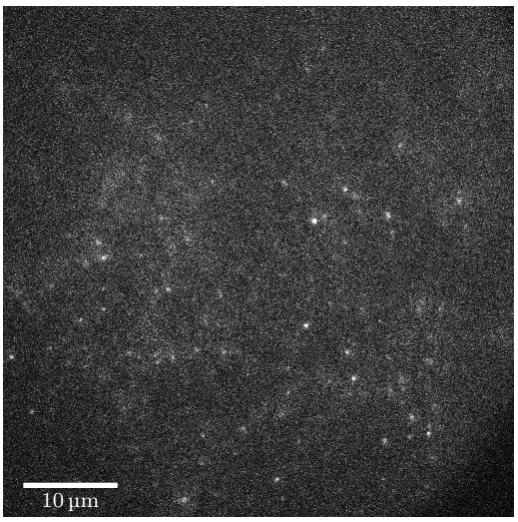
(i) This sub-image contains 4 beads, and it is of size $90\,\mathrm{px} \times 75\,\mathrm{px}$.

(ii) This sub-image does not contain beads, and it is of size $100\,\mathrm{px} \times 90\,\mathrm{px}$.

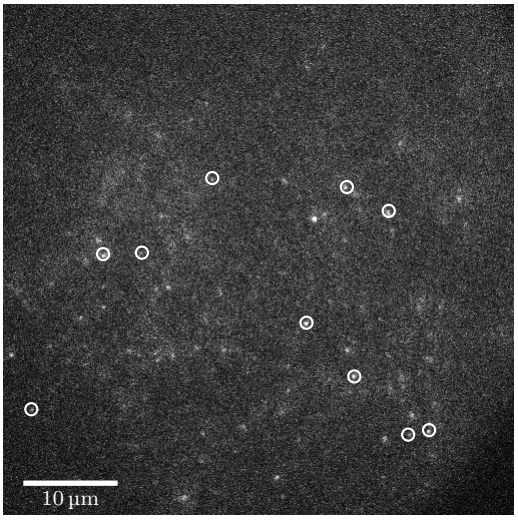
In order to have reliable ground truth available for the assessment of the drift estimation, the measurements obtained from both subframes are corrupted with simulated stage drift. To ensure that there is no detectable image drift in the sequence a priori, 10 fiducial markers (shown in Figure 3.6b) are selected in the full image frames and tracked across the whole image sequence.[5] Figure 3.7 shows the displacement of each marker with respect to its location in the first frame, averaged over all 10 objects, and it can be seen that the average displacement consistently falls below 0.1 px which is to be expected due to the imperfection of the detection method.

After showing that the dataset is drift-free, 20 artificial sensor trajectories are simulated in a 2-dimensional state space using a n.c.v. and a Brownian motion model, respectively, to be added to the measurements extracted from the two subframes. The Brownian motion is created using a standard deviation of 1 px per frame on each dimension of the sensor position; the n.c.v. model assumes a Gaussian normal distributed initial velocity with mean 0 standard deviation 0.25 px per frame and state space dimension and acceleration noise of 0.2 px per frame squared and dimension.

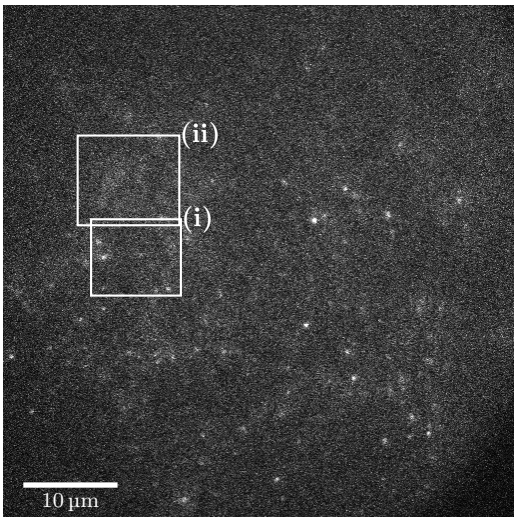
[3] SNAP25 is a protein which is involved in the process of exocytosis.
[4] Thanks to Colin Rickman for providing the images.
[5] For this purpose, an implementation of the HISP filter was used, see Houssineau et al. [2013].



(a) Original image.



(b) A selection of beads used for the analysis in Fig. 3.7.



(c) Chosen sub-images: (i) contains 4 beads, (ii) marker-free.

Figure 3.6: Example image of PALM data used for the estimation of the sensor drift.

Finally, the single-cluster multi-model PHD filter is run on the two sub-images (corrupted with the different random instances of Brownian and linear drift) using two different motion models for the objects, i.e. constant position and Brownian motion. For each simulated sensor trajectory, 20

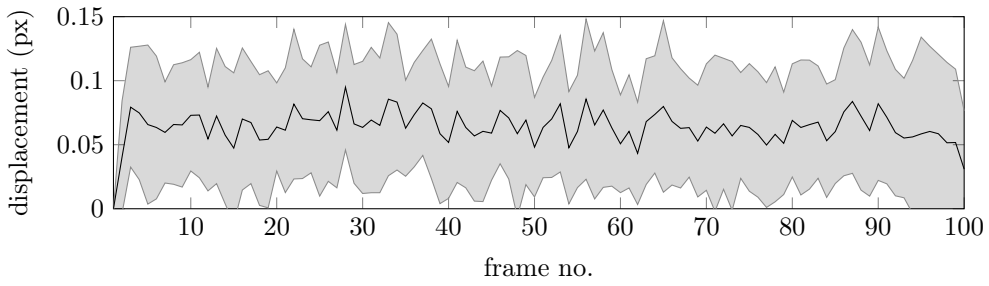
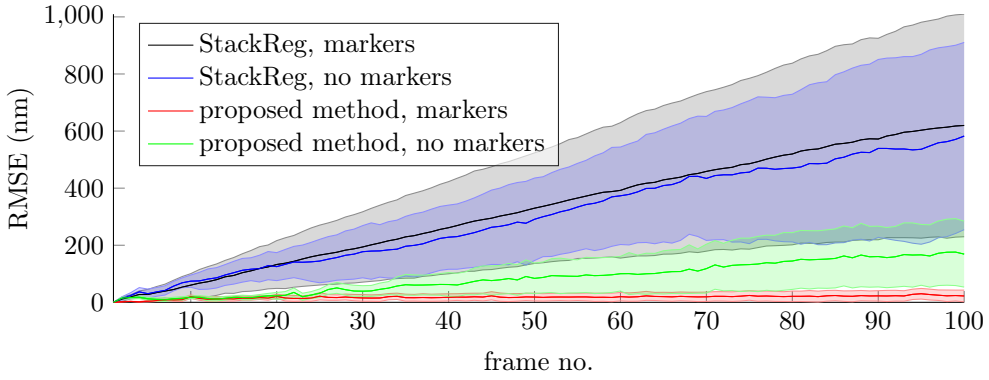
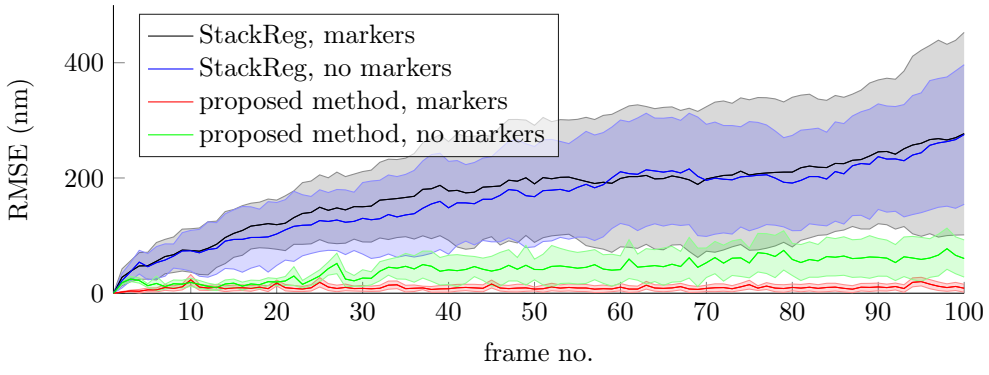


Figure 3.7: Average drift in the PALM dataset shown in Figure 3.6a over time, based on the displacement of 10 gold beads (Figure 3.6b) with respect to their positions in the first frame.

MC runs are performed since an SMC implementation is used for the parent process; see Appendix A.3.2 for details on the implementation. The outcome is compared with the calibration results of the ImageJ plugin *StackReg* which calibrates the frames of the image sequence with an intensity-based least-squares optimisation method [Thevenaz et al., 1998]. The results for Brownian and linear drift, averaged over all trajectories and MC runs, are demonstrated in Figure 3.8 in terms of the Root Mean Square Error (RMSE) of the estimated sensor position and the true position. The proposed method shows very little error when it is aided by fiducial markers, and a slightly increased error when it solely relies on the static molecules appearing spontaneously; the image-based registration, on the other hand, leads to estimation errors which are around four times higher than the marker-free drift estimation with the proposed algorithm.



(a) Estimation of linear drift.



(b) Estimation of Brownian drift.

Figure 3.8: Results of the sensor drift estimation the two subimages of 3.6c for Brownian and linear motion, all plotted with their standard deviation over all MC runs.

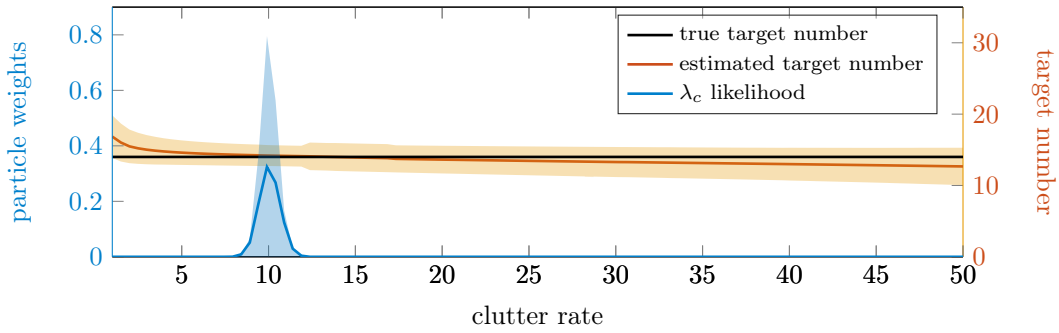
3.2.3 Application 3: Clutter rate estimation

In practice, parameters are often chosen by the user based on intuition and careful manual tweaking, however it would be useful to have a technique at hand which chooses the optimal parameter automatically based on the incoming data instead. Inspired by this idea, we used the single-cluster PHD filter in Schlangen et al. [2017a] for the estimation of the background noise in PALM data, assuming that it follows a Poisson model. In contrast to the experiment in Section 3.2.2, the classic PHD filter with only one motion model is used, i.e. $N_{\mathrm{mm}} = 1$; furthermore, the particles of the SMC implementation of the parent process are neither moved in the prediction nor resampled in the update since we are looking for a parameter which we assume to be static, i.e. the (one-dimensional) false alarm rate λ. Instead, a fixed grid of integer values between 0 and $\lambda_{\max}$ is assumed whose weights are simply updated using the multi-object likelihood of the PHD filter using the respective false alarm rates.

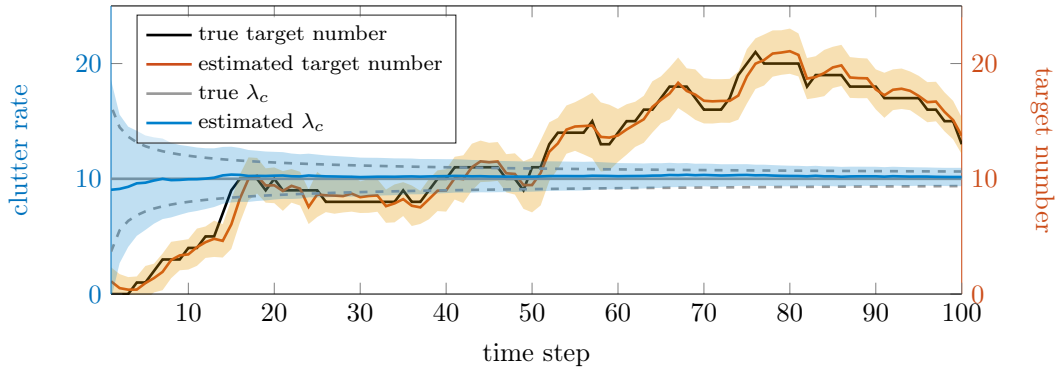
To validate this idea, the filter is first tested on 100 MC instances of a simulated dataset of 100 frames, created on a square two-dimensional state space with a width of 50 µm. Objects are generated according to a Poisson birth process with a birth rate of 0.5 targets per frame, having a Gaussian distributed initial velocity with mean 0 and standard deviation 500 nm per frame and moving according to an n.c.v. model with acceleration noise of 100 nm per frame in both image dimensions, respectively. Furthermore, the objects survive with probability $p_{\mathrm{s}} = 0.98$ and are detected with probability $p_{\mathrm{d}} = 0.99$. False alarms are generated according to a Poisson model with an average of 10 false positives per frame, which is assumed to be unknown for the filtering process.

The maximum admissible clutter parameter for this scenario is set to $\lambda_{\max} = 50$; a typical output of the filter at time 60 is shown in Figure 3.9a, where 14 targets are currently present. The plot shows the estimated number of objects for different clutter rates in red, and the corresponding multi-object likelihood in blue. As it can be seen, the multi-object likelihood is well-concentrated around the true false alarm rate 10, and the number of objects is also most accurately estimated around this value. When run over time, the Maximum A Posteriori (MAP) value is extracted as the filter output, i.e. the false alarm rate which leads to the highest multi-object likelihood; similarly, the corresponding estimated number of objects is extracted as well. Both MAP outputs over time are shown in Figure 3.9b, along with the ground truth (plotted in black): as it can be noticed, the estimated clutter rate converges quickly to the true value, and the true number of objects falls comfortably into the 2σ gate around the PHD estimate (using the PHD filter variance as formulated in Delande et al. [2014]).

After this sanity check, the method is tested on the same PALM dataset used in Section 3.2.2 for the drift estimation. Again, no ground truth is available, so we choose to run the algorithm on the full frame and a square subframe of width 200 px (i.e. 21.2 nm), shown in Figure 3.10, expecting that the ratio of the image sizes will approximately match with the ratio of the estimated clutter values and number of targets. The filter parameters are set as follows: The objects are assumed to appear according to a Poisson model with a birth rate of 5 objects per frame, and to survive with probability $p_{\mathrm{s}} = 0.95$. Their motion model is assumed to be n.c.v. with an acceleration noise

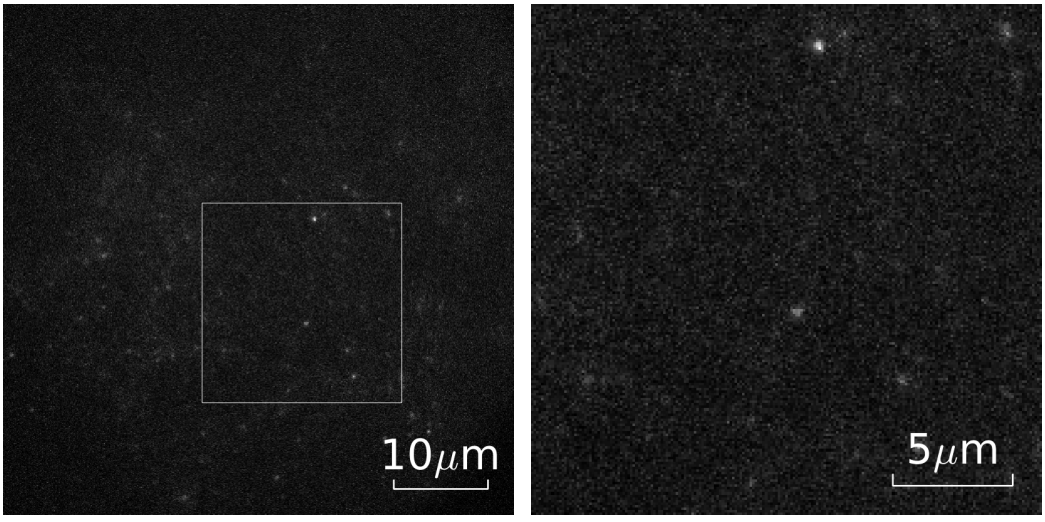


(a) Simulation result in frame 60 over the considered clutter rates for one MC run, where 14 targets are currently present. The estimated mean and standard deviation in the number of targets for different clutter rates is shown in red, and the multi-object likelihood is plotted in blue. The light blue area shows a Gaussian confidence area with standard deviation $\pm\sqrt{\frac{\lambda_c}{k}}$ around the true mean 10.



(b) Simulation results over time (MAP value), averaged over 100 MC runs. Again, the estimated mean and standard deviation in the number of targets are plotted in red, and the average clutter rate with corresponding standard deviation over all MC runs is plotted in blue. The dashed line stands for a $\pm\sqrt{\frac{\lambda_c}{k}}$ confidence interval around the true value.

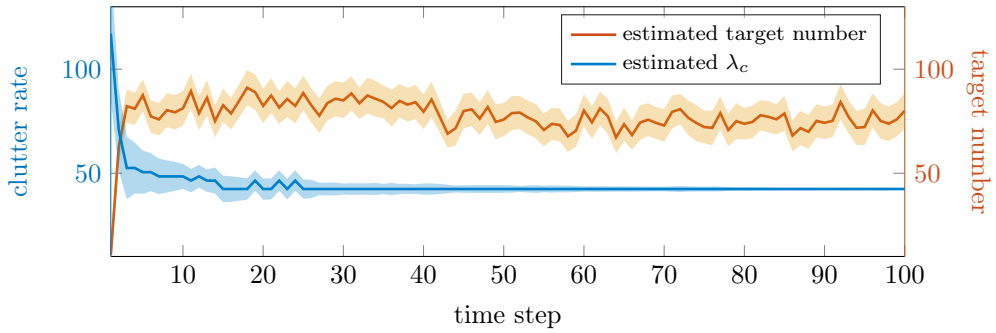
Figure 3.9: Simulation results for the clutter rate estimation.



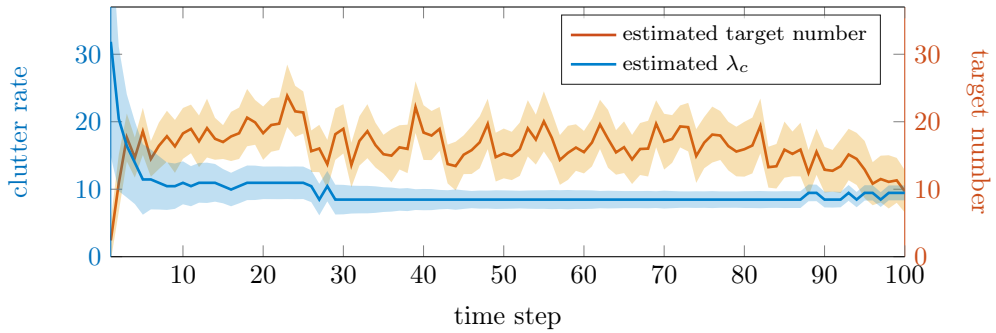
(a) Full image frame with subframe marked.

(b) The chosen subframe.

Figure 3.10: Sample frame of the PALM dataset of Section 3.2.2 with selected subframe.



(a) Results for the full images.



(b) Results for the subimages.

Figure 3.11: Estimation results for the PALM dataset with the same colour coding as in Figure 3.9b.

of 106 nm per frame squared and Gaussian distributed initial velocity with mean 0 and standard deviation 5.3 nm per frame and image dimension. The detection rate was assumed to be $p_{\mathrm{d}} = 0.95$, and the measurement noise was set to 212 nm. Since there are around 80 measurements in each frame, the maximum admissible false alarm rate was set to 100 in this case.

As Figure 3.11 shows, the estimated clutter rate for the full frame converges to 42.4 false alarms per frame, whereas the sub-images give a value of 8.5; the estimated number of targets oscillates around 75 and 15 for the full and cropped sequences, respectively. Since the image area of the cropped frame is approximately $\frac{1}{5}$ of the full frame, the estimated values are consistent with the expectation.

Conclusion

The focus of this work was on multi-object estimation techniques that are based on the representation of targets through point processes. It was demonstrated that the PHD filter, while being a powerful multi-object estimator in many applications, can be restrictive in its modelling assumptions and therefore might not cope well with unexpected changes in the number of detections. The CPHD filter, which was designed to overcome the limitations of the PHD filter by propagating the whole cardinality distribution of the target process in addition to the intensity, was shown to perform considerably slower and sometimes being over-confident in the estimation of the number of targets. Therefore, alternative estimation methods were designed to relax the modelling assumptions of the PHD filter while avoiding the computational complexity of the CPHD filter. For this purpose, the so-called Panjer point process was introduced which generalises the Poisson point process by incorporating the Poisson, binomial and negative binomial distributions in one single formulation and therefore allowing for a more flexible description of the number of objects by parametrising the mean and variance separately. As a first attempt, the PHD filter with Panjer clutter was introduced and it has been demonstrated in simulations that the generalised clutter model makes it possible to estimate the target population considerably better than using the Poisson model in the presence of greatly varying numbers of false alarms. Based on these promising results, the SO-PHD filter was formulated which uses a Panjer model for both the predicted target process and the false alarm process. Utilising the properties of the Panjer point process, the SO-PHD filter propagates second-order information in form of the variance in the number of targets in addition to the process intensity at a much lower computational cost than the CPHD filter. Furthermore, simulation studies have shown that the spooky effect at a distance which was diagnosed for the CPHD filter previously is present in the SO-PHD filter as well, but to a much smaller extent. The CPHD and SO-PHD filters were also successfully embedded in a single-cluster estimation algorithm based on the idea of the single-cluster PHD filter. The low computational time of the SO-PHD filter makes it a suitable alternative to the PHD filter in challenging parameter estimation scenarios. The presented experiments on simulated and real data showed that all three methods are powerful techniques to extract meaningful statistics from microscopy images, hence providing helpful analysis tools to aid biomedical research and other fields of science and engineering.

Appendices

Appendix A

Implementation

A.1 The Kalman filter and Gaussian mixtures

The Kalman filter was first introduced by Kalman [1960] as an exact closed-form solution of the Bayesian recursion for one object. It is based on the assumption that the described dynamical system is linear and represented fully by Gaussian distributions. In the following, let $\mathcal{N}(\cdot; m_k, C_k)$ denote a Gaussian distribution with d_x-dimensional mean $m_k \in \mathbb{R}^{d_x}$ and a covariance matrix $C_k \in \mathbb{R}^{d_x \times d_x}$. The prediction step in the Kalman filter involves a state transition matrix $F_k \in \mathbb{R}^{d_x \times d_x}$, a process noise matrix $Q_k \in \mathbb{R}^{d_x \times d_x}$ an optional control input matrix $B_k \in \mathbb{R}^{d_x \times d_x}$ if a control vector $u_k \mathbb{R}^{d_x}$ is involved. With these notations, the prediction step of the Kalman filter is found to be (cf. Figure 1.1b for illustration)

$$m_{k|k-1} = F_k m_{k-1} + B_k u_k, \tag{A.1}$$

$$C_{k|k-1} = F_k C_{k-1} F_k^T + Q_k. \tag{A.2}$$

Suppose that at time k, a d_z-dimensional measurement $z_k \in \mathbb{R}^{d_z}$ is obtained. First, the so-called innovation with mean $y_k \in \mathbb{R}^{d_z}$ and covariance $S_k \in \mathbb{R}^{d_z \times d_z}$ is computed which represents the information gain through the measurement. Its value is dependent on the process noise matrix $R_k \in \mathbb{R}^{d_z \times d_z}$ and on the projection of the predicted Gaussian distribution onto the measurement space using the projection matrix $H_k \in \mathbb{R}^{d_z \times d_x}$. As an intermediate step, the Kalman gain $K_k \in \mathbb{R}^{d_x \times d_z}$ is computed from the covariances of the predicted distribution and the innovation, which is used in turn to find the updated distribution with mean m_k and covariance C_k. The equations of the Kalman filter update are given as (cf. Figure 1.1c for illustration)

$$y_k = z_k - H_k m_{k|k-1}, \tag{A.3}$$

$$S_k = H_k C_{k|k-1} H_k^T + R_k, \tag{A.4}$$

$$K_k = C_{k|k-1} H_k^T S_k^{-1}, \tag{A.5}$$

$$m_k = m_{k|k-1} + K_k y_k, \tag{A.6}$$

$$C_k = (I - K_k H_k) C_{k|k-1}. \tag{A.7}$$

The formulation of the Kalman filter can be exploited when working with Gaussian mixture models. A Gaussian Mixture (GM) is given by the sum

$$p(\,\cdot\,) = \sum_{i=1}^{N_{k-1}} w_{k-1}^{(i)} \mathcal{N}(\,\cdot\,; m_{k-1}^{(i)}, C_{k-1}^{(i)}). \tag{A.8}$$

for a collection $\left\{w_{k-1}^{(i)}, m_{k-1}^{(i)}, C_{k-1}^{(i)}\right\}_{i=1}^{N_{k-1}}$ of Gaussian distributions with weights $0 \leq w_{k-1}^{(i)} \leq 1$. In the prediction, each of the Gaussian components is predicted using Equations (A.1) and (A.2) and updated with Equations (A.6) and (A.7). The prediction and update of the component weights $w_{k-1}^{(i)}$, however, depends on the specific filter in which the Gaussian mixture model is used. In the Gaussian sum filter [Alspach and Sorenson, 1972], for example, the weights are updated using the innovation given by (A.3) and (A.4) and then normalised. In case of the PHD filter methods discussed in this work, the reweighting of the Gaussians is more intricate, involving probabilities of detection and survival and other normalisation constants. An example is shown in Appendix A.3.1, showing pseudocode for a GM implementation of the SO-PHD filter.

A.2 Track extraction and the PHD filter class

Since the class of PHD filter techniques is concerned with tracking target populations rather than individuals and avoids data association, track extraction is not a natural part of the estimation. However, it can be useful for biomedical applications to extract information for specific objects, e.g. in order to calculate statistics on individual molecule speeds. Some methods were suggested in the literature, e.g. by Lin et al. [2006] and Li et al. [2014] for SMC implementations of the PHD filter. If a GM implementation is used like in the experiments of this work, the involved Kalman filters can be utilised for the track extraction in the following manner. First of all, additionally to its weight $w \in [0, 1]$, mean $m \in \mathcal{X} \subseteq \mathbb{R}^4$ and covariance matrix $C \in \mathbb{R}^{4\times4}$, accounting for positions and velocities in x and y, each Gaussian in the mixture can be equipped with a unique identification number $i \in \mathbb{N}$ and a sequence $h = (x_{k_0}, \ldots, x_{k-1})$ of previous state estimates, where k_0 is the time step where the object was created. Now, if an object is updated at time k, its history h is augmented by the new estimate x_k. In the merging step, two objects are merged if the Hellinger distance between their Gaussian distributions falls below a user-defined threshold and if their identification numbers are identical. Like this, it is ensured that objects with significantly different trajectories are maintained even if their distributions become similar. In the end, only the trajectories of those objects are extracted whose weight w_k raises above a user-defined threshold.

A.3 Pseudocode

In the following, let N_{k-1}, $N_{\mathrm{b},k-1}$ denote the respective numbers of Gaussians in the prior and birth mixtures.

A.3.1 Second-order GM-PHD filter

Algorithm 1: SO-PHD prediction (time k).

Input

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

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

Prior intensity

$\mu_{k-1}(\mathcal{X}) = \sum_{i=1}^{N_{k-1}} w_{k-1}^{(i)}$

Survival process

for $1 \leq i \leq N_{k-1}$ **do**

$w_{k|k-1}^{(i)} = p_{\mathrm{s},k} w_{k-1}^{(i)}$

$m_{k|k-1}^{(i)} = F_{k-1} m_{k-1}^{(i)}$

$P_{k|k-1}^{(i)} = F_{k-1} P_{k-1}^{(i)} F_{k-1}^{T} + Q_{k-1}$

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 $1 \leq j \leq N_{\mathrm{b},k-1}$ **do**

$w_{k|k-1}^{(n_{k-1}+j)} = w_{\mathrm{b},k-1}^{(j)}$

$m_{k|k-1}^{(n_{k-1}+j)} = m_{\mathrm{b},k-1}^{(j)}$

$P_{k|k-1}^{(n_{k-1}+j)} = P_{\mathrm{b},k-1}^{(j)}$

end for

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

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

Output

Predicted process: $\left\{w_{k|k-1}^{(i)}, m_{k|k-1}^{(i)}, P_{k|k-1}^{(i)}\right\}_{i=1}^{N_{k|k-1}}, \mathrm{var}_{k|k-1}(\mathcal{X})$

Algorithm 2: Computation of the elementary symmetric functions using Vieta's theorem.

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 $0 \leq j \leq m_k$

Output

Elementary symmetric functions $\{e_j(Z_k)\}_{0 \leq j \leq m_k}$

Algorithm 3: SO-PHD data update (time k).

Input

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

Current measurements: $Z_k = \{z_j\}_{j=1}^{M_k}$

Panjer parameters

$\mu_{k|k-1}(\mathcal{X}) = \sum_{i=1}^{N_{k|k-1}} w_{k|k-1}^{(i)}$

$\alpha_{k|k-1} = \mu_{k|k-1}(\mathcal{X})^2/(\mathrm{var}_{k|k-1}(\mathcal{X}) - \mu_{k|k-1}(\mathcal{X}))$

$\beta_{k|k-1} = \mu_{k|k-1}(\mathcal{X})/(\mathrm{var}_{k|k-1}(\mathcal{X}) - \mu_{k|k-1}(\mathcal{X}))$

Missed detection and measurement terms

for $1 \leq i \leq N_{k|k-1}$ **do**

$w_{\phi,k}^{(i)} = (1 - p_{\mathrm{d},k}) w_{k|k-1}^{(i)}$

$m_{\phi,k}^{(i)} = m_{k|k-1}^{(i)}$

$P_{\phi,k}^{(i)} = P_{k|k-1}^{(i)}$

end for

$\mu_k^{\phi}(\mathcal{X}) = (1 - p_{\mathrm{d},k}) \mu_{k|k-1}(\mathcal{X})$

for $1 \leq j \leq M_k$ **do**

for $1 \leq i \leq N_{k|k-1}$ **do**

$y_k^{(i,j)} = z_j - H_k m_{k|k-1}^{(i)}$

$S_k^{(i)} = H_k P_{k|k-1}^{(i)} H_k^T + R_k$

$K_k^{(i)} = P_{k|k-1}^{(i)} H_k^T [S_k^{(i)}]^{-1}$

$w_{\mathrm{d},k}^{(i,j)} = p_{\mathrm{d},k} w_{\mathrm{d},k|k-1}^{(i,j)} \mathcal{N}(z; y_k^{(i,j)}, S_k^{(i)}) / s_{\mathrm{c},k}$

$m_{\mathrm{d},k}^{(i,j)} = m_{k|k-1}^{(i)} + K_k^{(i)} y_k^{(i,j)}$

$P_{\mathrm{d},k}^{(i,j)} = (I - K_k^{(i)} H_k) P_{k|k-1}^{(i)}$

end for

$\mu_k^{z_j}(\mathcal{X}) = \sum_{i=1}^{N_{k|k-1}} w_{\mathrm{d},k}^{(i,j)}$

end for

Corrective terms

$F_{\mathrm{d}} = (1 + \frac{p_{\mathrm{d},k}}{\beta_{k|k-1}}) \sum_{z \in Z_k} \mu_k^{z_j}(\mathcal{X})$

Compute $\{e_d(Z_k)\}_{0 \leq d \leq M_k}$ using Algorithm 2

for $0 \leq u \leq 2$ **do**

$\Upsilon_u(Z_k) = \sum_{j=0}^{M_k} \frac{(\alpha_{k|k-1})_{j+u}}{(\beta_{k|k-1})^{j+u}} \frac{(\alpha_{\mathrm{c},k})_{m_k-j}}{(\beta_{\mathrm{c},k}+1)^{m_k-j}} F_{\mathrm{d}}^{-j-u} e_j(Z_k)$

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 $1 \leq i \leq M_k$ **do**

Compute $\{e_d(Z_k \setminus z_i)\}_{0 \leq d \leq M_k - 1}$ using Algorithm 2

for $1 \leq u \leq 2$ **do**

$\Upsilon_u(Z_k \setminus z_i) = \sum_{d=0}^{M_k-1} \frac{(\alpha_{k|k-1})_{d+u}}{(\beta_{k|k-1})^{d+u}} \frac{(\alpha_{\mathrm{c},k})_{m_k-1-d}}{(\beta_{\mathrm{c},k}+1)^{m_k-1-d}} F_{\mathrm{d}}^{-d-u} e_d(Z_k \setminus z_i)$

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 $1 \leq i < j \leq M_k$ **do**

Compute $\{e_d(Z_k \setminus \{z_i, z_j\})\}_{0 \leq d \leq M_k - 2}$ using Algorithm 2

$\Upsilon_2(Z_k \setminus \{z_i, z_j\}) = \sum_{d=0}^{M_k-2} \frac{(\alpha_{k|k-1})_{d+2}}{(\beta_{k|k-1})^{d+2}} \frac{(\alpha_{\mathrm{c},k})_{m_k-2-d}}{(\beta_{\mathrm{c},k}+1)^{m_k-2-d}} F_{\mathrm{d}}^{-d-2} e_d(Z_k \setminus \{z_i, z_j\})$

$\ell_2^{\neq}(z_i, z_j) = \Upsilon_2(Z_k \setminus \{z_i, z_j\})/\Upsilon_0(Z_k)$

end for

end for

Missed detection terms

for $1 \leq i \leq N_{k|k-1}$ **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 $1 \leq j \leq M_k$ **do**

$w_k^{(i \cdot n_{k|k-1}+j)} = \ell_1(z_j) w_{\mathrm{d},k}^{(i,j)}$

$m_k^{(i \cdot n_{k|k-1}+j)} = m_{\mathrm{d},k}^{(i,j)}$

$P_k^{(i \cdot n_{k|k-1}+j)} = P_{\mathrm{d},k}^{(i,j)}$

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

$$\mathrm{var}_k(\mathcal{X}) = \mu_k(\mathcal{X}) + \mu_k^{\phi}(\mathcal{X})^2 \left[\ell_2(\phi) - \ell_1(\phi)^2\right] + 2\mu_k^{\phi}(\mathcal{X}) \sum_{z \in Z_k} \frac{\mu_k^{z}(\mathcal{X})}{s_{\mathrm{c},k}(z)} \left[\ell_2(z) - \ell_1(\phi)\ell_1(z)\right] + \sum_{z \neq z' \in Z_k} \frac{\mu_k^{z}(\mathcal{X})}{s_{\mathrm{c},k}(z)} \frac{\mu_k^{z'}(\mathcal{X})}{s_{\mathrm{c},k}(z')} \left[\ell_2^{\neq}(z, z') - \ell_1(z)\ell_1(z')\right],$$

Output

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

A.3.2 The single-cluster multi-object filter

The parent process of single-cluster multi-object filters is usually implemented with a SMC approach to avoid the burden of estimating correlations of the objects with the estimated parameter (i.e. the sensor state, clutter rate, etc.). Therefore, the sensor state is represented by a fixed number N of random samples $(\{(y_k^i, w_k^i, Д_k^i)\}_{i=1}^N$, called *particles*, where y_k^i denotes the ith instance of the estimated parameter at time k, w_k^i is its corresponding weight, and $Д_k^i$ is the set of statistics maintained by the multi-object filter of choice, dependent on the particle state y_k^i. If the PHD filter is chosen, $Д_k^i$ is the intensity $\mu_k^\flat$ of the target population only, for the CPHD filter it denotes the intensity $\mu_k^\sharp$ with the cardinality distribution ρ_k, and for the SO-PHD filter it stands for the intensity $\mu_k^\natural$ plus the variance $\mathrm{var}_k^\natural$. The multi-object likelihood function can either be the filter-specific function (1.58), (2.85) or (2.89), respectively, or a suitable alternative as given in Leung et al. [2016]. To avoid particle degeneracy [Doucet et al., 2001], a resampling step is performed if the number of effective particles N_{eff} falls below a certain value $r \cdot N$ with $0 \leq r \leq 1$; the number of effective particles is computed using the formula

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^N (w_k^i)^2}. \tag{A.9}$$

One of the easiest methods of resampling is called *roulette resampling* [Doucet et al., 2001], which is chosen for all experiments conducted in this work; to improve the results, one could consider alternative approaches like *importance sampling* and others. These considerations lead to the following algorithm.

Algorithm 3: Algorithm for parameter estimation.

Input

Set of particles $\{(y_k^i, w_k^i, Д_k^i)\}_{i=1}^N$

Set of measurements Z_{k+1}

Prediction

for $1 \leq i \leq N$ **do**

Sample $y_{k+1}^i \sim \mathcal{N}(\mathbf{y}_k^i, \Sigma_s)$

$Д_{k+1|k}^i \leftarrow$ `Prediction`$(Д_k^i)$

end for

Update

for $1 \leq i \leq N$ **do**

$Д_{k+1}^i \leftarrow$ `Update`$(Д_{k+1|k}^i, Z_{k+1})$

$w_{k+1}^i \leftarrow$ `MultiObjectLikelihood`$(Д_{k+1|k}^i, Z_{k+1})$

end for

Resampling

$$N_{\text{eff}} \leftarrow \frac{1}{\sum_{i=1}^{N}(w_{k+1}^{i})^2}$$

if $N_{\text{eff}} \leq r \cdot N$ **then**

  $\{y_{k+1}^{i}\}_{i=1}^{N} \leftarrow \texttt{Resampling}(\{(y_{k+1}^{i}, w_{k+1}^{i})\}_{i=1}^{N})$

  for $1 \leq i \leq N$ **do**

    $w_{k+1}^{i} \leftarrow \frac{1}{N}$

  end for

end if

Output

Set of particles $\{(y_{k+1}^{i}, w_{k+1}^{i}, Д_{k+1}^{i})\}_{i=1}^{N}$

References

Alspach, D. and Sorenson, H. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE transactions on automatic control*, 17(4):439–448.

Ambrose, E. J. (1956). A surface contact microscope for the study of cell movements. *Nature*, 178(4543):1194–1194.

Anderson, B. D. O. and Moore, J. B. (1979). Optimal filtering. *Englewood Cliffs*, 21:22–95.

Andriluka, M., Roth, S., and Schiele, B. (2008). People-tracking-by-detection and people-detection-by-tracking. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE.

Bernhard, P. (2005). Chain differentials with an application to the mathematical fear operator. *Nonlinear Analysis: Theory, Methods & Applications*, 62(7):1225–1233.

Betzig, E., Patterson, G., Sougrat, R., Lindwasser, O., Olenych, S., Bonifacino, J., Davidson, M., Lippincott-Schwartz, J., and Hess, H. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, 313, No. 5793:1642–1645.

Boers, Y., Driessen, H., Torstensson, J., Trieb, M., Karlsson, R., and Gustafsson, F. (2006). Track-before-detect algorithm for tracking extended targets. *IEE Proceedings-Radar, Sonar and Navigation*, 153(4):345–351.

Bryant, D. S., Delande, E. D., Gehly, S., Houssineau, J., Clark, D. E., and Jones, B. A. (2016). The CPHD filter with target spawning. *IEEE Transactions on Signal Processing*, 65(5):13124–13138.

Casuso, I., Khao, J., Chami, M., Paul-Gilloteaux, P., Husain, M., Duneau, J.-P., Stahlberg, H., Sturgis, J. N., and Scheuring, S. (2012). Characterization of the motion of membrane proteins using high-speed atomic force microscopy. *Nature nanotechnology*, 7(8):525–529.

Chenouard, N., Smal, I., De Chaumont, F., Maška, M., Sbalzarini, I. F., Gong, Y., Cardinale, J., Carthel, C., Coraluppi, S., Winter, M., et al. (2014). Objective comparison of particle tracking methods. *Nature methods*, 11(3):281–289.

Clark, D. E., Bell, J., de Saint-Pern, Y., and Petillot, Y. (2005). PHD filter multi-target tracking in 3d sonar. In *Oceans 2005-Europe*, volume 1, pages 265–270. IEEE.

Clark, D. E. and Houssineau, J. (2013). Faa di BrunoâĂŹs formula and spatial cluster modelling. *Spatial Statistics*, 6:109–117.

Clark, D. E. and Houssineau, J. (2013). Faa di Bruno's formula for chain differentials. *ArXiv e-prints*. arXiv:1202.0264v4.

Clark, D. E. and Houssineau, J. (2014). Faà Di Bruno's formula and Volterra series. In *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, pages 217–219. IEEE.

Clark, D. E., Houssineau, J., and Delande, E. D. (2015). A few calculus rules for chain differentials. *ArXiv e-prints*. arXiv:1506.08626v1.

Coraluppi, S. and Carthel, C. (2004). Recursive track fusion for multi-sensor surveillance. *Information Fusion*, 5(1):23–33.

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

Del Moral, P. (1996). Non-linear filtering: interacting particle resolution. *Markov processes and related fields*, 2(4):555–581.

Delande, E. D., Üney, M., Houssineau, J., and Clark, D. E. (2014). Regional variance for multi-object filtering. *Signal Processing, IEEE Transactions on*, 62(13):3415–3428.

Doucet, A., Smith, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Information Science and Statistics. Springer New York.

Erdinc, O., Willett, P., and Bar-Shalom, Y. (2005). Probability hypothesis density filter for multitarget multisensor tracking. In *Information Fusion, Proceedings of the 7th International Conference on*, volume 1, page 8.

Fackler, M. (2009). Panjer class united – one formula for the Poisson, Binomial, and Negative Binomial distribution. *Proceedings of the 39th ASTIN Colloquium*.

Fang, N., Lee, H., Sun, C., and Zhang, X. (2005). Sub–diffraction-limited optical imaging with a silver superlens. *Science*, 308(5721):534–537.

Fortmann, T., Bar-Shalom, Y., and Scheffe, M. (1983). Sonar tracking of multiple targets using joint probabilistic data association. *IEEE journal of Oceanic Engineering*, 8(3):173–184.

Fränken, D., Schmidt, M., and Ulmke, M. (2009). Spooky Action at a Distance in the Cardinalized Probability Hypothesis Density Filter. *Aerospace and Electronic Systems, IEEE Transactions on*, 45(4):1657–1664.

Gruden, P. and White, P. R. (2016). Automated tracking of dolphin whistles using Gaussian mixture probability hypothesis density filters. *The Journal of the Acoustical Society of America*, 140(3):1981–1991.

Gustafsson, M. G. L. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy*, 198(2):82–87.

Hagen, O., Houssineau, J., Schlangen, I., Delande, E. D., Franco, J., and Clark, D. E. (2016). Joint estimation of telescope drift and space object tracking. In *Aerospace Conference, 2016 IEEE*, pages 1–10. IEEE.

Hell, S. W. and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11):780–782.

Hess, S. T., Girirajan, T. P. K., and Mason, M. D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–4272.

Houssineau, J. (2015). *Representation and estimation of stochastic populations*. PhD thesis, Heriot-Watt University.

Houssineau, J., Clark, D. E., Ivekovic, S., Lee, C. S., and Franco, J. (2016). A unified approach for multi-object triangulation, tracking and camera calibration. *IEEE Transactions on Signal Processing*, 64(11):2934–2948.

Houssineau, J., Del Moral, P., and Clark, D. E. (2013). General multi-object filtering and association measure. In *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2013 IEEE 5th International Workshop on*, pages 33–36. IEEE.

Husain, M., Boudier, T., Paul-Gilloteaux, P., Casuso, I., and Scheuring, S. (2012). Software for drift compensation, particle tracking and particle analysis of high-speed atomic force microscopy image series. *Journal of Molecular Recognition*, 25(5):292–298.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons.

Julier, S. J. and Uhlmann, J. K. (1997). New extension of the Kalmans filter to nonlinear systems. In *AeroSense'97*, pages 182–193. International Society for Optics and Photonics.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

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

Lee, C.-S., Clark, D. E., and Salvi, J. (2013). SLAM with dynamic targets via single-cluster PHD filtering. *Selected Topics in Signal Processing, IEEE Journal of*, PP(99):1–1.

Leung, K. Y. K., Inostroza, F., and Adams, M. (2016). Multifeature-based importance weighting for the PHD SLAM filter. *IEEE Transactions on Aerospace and Electronic Systems*, 52(6):2697–2714.

Li, T., Sun, S., Corchado, J. M., and Siyau, M. F. (2014). A particle dyeing approach for track continuity for the SMC-PHD filter. In *Information Fusion (FUSION), 2014 17th International Conference on*, pages 1–8. IEEE.

Liang, L., Shen, H., De Camilli, P., and Duncan, J. S. (2010). Tracking clathrin coated pits with a multiple hypothesis based method. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 315–322. Springer.

Lin, L., Bar-Shalom, Y., and Kirubarajan, T. (2006). Track labeling and PHD filter for multitarget tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 42(3):778–795.

Luo, X. and Ishihara, T. (2004). Surface plasmon resonant interference nanolithography technique. *Applied Physics Letters*, 84(23):4780–4782.

Mahler, R. (2006). PHD filters of second order in target number. In *Proc. SPIE Defense and Security Symposium*, volume 6236, page 1. International Society for Optics and Photonics.

Mahler, R. P. S. (2003). Multitarget Bayes filtering via first-order multitarget moments. *Aerospace and Electronic Systems, IEEE Transactions on*, 39(4):1152–1178.

Mahler, R. P. S. (2007a). PHD filters of higher order in target number. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(4):1523–1543.

Mahler, R. P. S. (2007b). *Statistical multisource-multitarget information fusion*. Artech House, Boston.

Mahler, R. P. S. (2014). *Advances in statistical multisource-multitarget information fusion*. Artech House.

Manley, S., Gillette, J. M., Patterson, G. H., Shroff, H., Hess, H. F., Betzig, E., and Lippincott-Schwartz, J. (2008). High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature methods*, 5(2):155–157.

McKenna, I., Tonolini, F., Tobin, R., Houssineau, J., Bridle, H., McDougall, C., Schlangen, I., McGrath, J. S., Jimenez, M., and Clark, D. E. (2015). Observing the dynamics of waterborne pathogens for assessing the level of contamination. In *Sensor Signal Processing for Defence (SSPD), 2015*, pages 1–5. IEEE.

Meijering, E., Dzyubachyk, O., and Smal, I. (2012). 9 methods for cell and particle tracking. *Methods in enzymology*, 504(9):183–200.

Meijering, E., Smal, I., and Danuser, G. (2006). Tracking in molecular bioimaging. *IEEE Signal Process. Mag.*, 23(3):46–53.

Nikodym, O. (1930). Sur une généralisation des intégrales de MJ Radon. *Fundamenta Mathematicae*, 15(1):131–179.

Panjer, H. H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin: The Journal of the IAA*, 12(1):22–26.

Reid, D. (1979). An algorithm for tracking multiple targets. *IEEE transactions on Automatic Control*, 24(6):843–854.

Ristic, B. and Clark, D. (2012). Particle filter for joint estimation of multi-object dynamic state and multi-sensor bias. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3877–3880. IEEE.

Ristic, B., Clark, D. E., and Gordon, N. (2012). Calibration of tracking systems using detections from non-cooperative targets. In *Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2012 Workshop on*, pages 25–30. IEEE.

Ristic, B., Clark, D. E., and Gordon, N. (2013). Calibration of multi-target tracking algorithms using non-cooperative targets. *IEEE Journal of Selected Topics in Signal Processing*, 7(3):390–398.

Rust, M., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3:793–796.

Rutten, M. G., Ristic, B., and Gordon, N. J. (2005). A comparison of particle filters for recursive track-before-detect. In *Information Fusion, 2005 8th International Conference on*, volume 1, pages 169–175. IEEE.

Schlangen, I., Bharti, V., Delande, E., and Clark, D. E. (2017a). Joint multi-object and clutter rate estimation with the single-cluster PHD filter. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 1087–1091. IEEE.

Schlangen, I., Clark, D. E., and Delande, E. D. (2017b). Single-cluster PHD filter methods for joint multi-object filtering and parameter estimation. *Submitted to: IEEE Transactions on Signal Processing*. arXiv:1705.05312.

Schlangen, I., Delande, E., Houssineau, J., and Clark, D. E. (2016a). A PHD filter with negative binomial clutter. In *Information Fusion (FUSION), 2016 19th International Conference on*, pages 658–665. IEEE.

Schlangen, I., Delande, E. D., Houssineau, J., and Clark, D. E. (2017c). A second-order PHD filter with mean and variance in target number. *Accepted to: IEEE Transactions on Signal Processing*. arXiv:1704.02084.

Schlangen, I., Franco, J., Houssineau, J., Pitkeathly, W. T. E., Clark, D. E., Smal, I., and Rickman, C. (2016b). Marker-less stage drift correction in super-resolution microscopy using the single-cluster PHD filter. *Selected Topics in Signal Processing, IEEE Journal of*, 10(1):193–202.

Schlangen, I., Houssineau, J., and Clark, D. (2014). A novel approach to image calibration in super-resolution microscopy. In *Control, Automation and Information Sciences (ICCAIS), 2014 International Conference on*, pages 111–116. IEEE.

Smal, I., Loog, M., Niessen, W., and Meijering, E. (2010). Quantitative comparison of spot detection methods in fluorescence microscopy. *IEEE Transactions on Medical Imaging*, 29(2):282–301.

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

Streit, R., Degen, C., and Koch, W. (2015). The pointillist family of multitarget tracking filters. *arXiv preprint arXiv:1505.08000*.

Streit, R. L. (2013). The probability generating functional for finite point processes, and its application to the comparison of phd and intensity filters. *J. Adv. Inf. Fusion*, 8(2):119–132.

Streit, R. L. and Stone, L. D. (2008). Bayes derivation of multitarget intensity filters. In *Information Fusion, 2008 11th International Conference on*, pages 1–8. IEEE.

Sundaramoorthi, G., Jackson, J. D., Yezzi, A., and Mennucci, A. C. (2006). Tracking with Sobolev active contours. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 674–680. IEEE.

Swain, A. J. (2013). *Group and Extended Target Tracking with the Probability Hypothesis Density Filter*. PhD thesis, Heriot-Watt University.

Thevenaz, P., Ruttimann, U. E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE transactions on image processing*, 7(1):27–41.

Thrun, S. and Leonard, J. J. (2008). Simultaneous localization and mapping. In *Springer handbook of robotics*, pages 871–889. Springer.

Ulmke, M., Fränken, D., and Schmidt, M. (2008). Missed detection problems in the cardinalized probability hypothesis density filter. In *Information Fusion, 2008 11th International Conference on*, pages 1–7. IEEE.

Urtasun, R., Fleet, D. J., and Fua, P. (2006). 3d people tracking with gaussian process dynamical models. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 238–245. IEEE.

Vo, B.-N. and Ma, W.-K. (2006). The Gaussian Mixture Probability Hypothesis Density Filter. *Signal Processing, IEEE Transactions on*, 54(11):4091–4104.

Vo, B. T. and Vo, B. N. (2012). The para-normal Bayes multi-target filter and the spooky effect. In *Information Fusion (FUSION), 2012 15th International Conference on*, pages 173–180. IEEE.

Vo, B.-T., Vo, B.-N., and Cantoni, A. (2007). Analytic Implementations of the Cardinalized Probability Hypothesis Density Filter. *Signal Processing, IEEE Transactions on*, 55(7):3553–3567.

Vo, B.-T., Vo, B.-N., and Cantoni, A. (2009). The cardinality balanced multi-target multi-Bernoulli filter and its implementations. *IEEE Transactions on Signal Processing*, 57(2):409–423.

Williams, J. L. (2012). Hybrid poisson and multi-Bernoulli filters. In *Information Fusion (FUSION), 2012 15th International Conference on*, pages 1103–1110. IEEE.

Williams, J. L. (2015). An efficient, variational approximation of the best fitting multi-Bernoulli filter. *IEEE Transactions on Signal Processing*, 63(1):258–273.

Wilson, R. S., Yang, L., Dun, A., Smyth, A. M., Duncan, R. R., Rickman, C., and Lu, W. (2016). Automated single particle detection and tracking for large microscopy datasets. *Royal Society open science*, 3(5):160225.

Wood, T. M., Yates, C. A., Wilkinson, D. A., and Rosser, G. (2012). Simplified multitarget tracking using the PHD filter for microscopic video data. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(5):702–713.

Yokoyama, M. and Poggio, T. (2005). A contour-based moving object detection and tracking. In *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*, pages 271–276. IEEE.

Zhang, B., Zerubia, J., and Olivo-Marin, J.-C. (2007). Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics*, 46(10):1819–1829.

Index