Filters for Spatial Point Processes

Sumeetpal S. Singh, Ba-Ngu Vo, Adrian Baddeley
and Sergei Zuyev

CUED/F-INFENG/TR-591 (2007)

first version December 2007, revised February 2009

FILTERS FOR SPATIAL POINT PROCESSES

SUMEETPAL S. SINGH*, BA-NGU VO†, ADRIAN BADDELEY‡, AND SERGEI ZUYEV§

Abstract. Let $\mathbf{X}$ and $\mathbf{Y}$ be two jointly distributed spatial Point Processes on $\mathcal{X}$ and $\mathcal{Y}$ respectively (both complete separable metric spaces). We address the problem of estimating $\mathbf{X}$, which is the *hidden* Point Process (PP), given the realisation $\mathbf{y}$ of the *observed* PP $\mathbf{Y}$. We characterise the posterior distribution of $\mathbf{X}$ when it is marginally distributed according to a Poisson and Gauss-Poisson prior and when the transformation from $\mathbf{X}$ to $\mathbf{Y}$ includes thinning, displacement and augmentation with extra points. These results are then applied in a filtering context when the hidden process evolves in discrete time in a Markovian fashion. The dynamics of $\mathbf{X}$ considered are general enough for many target tracking applications.

Key words. PHD filter, target tracking, hidden point process inference, online filtering, Poisson point process prior, Gauss-Poisson point process

AMS subject classifications. 62M30, 62L12, 60G55, 60G35

1. Introduction.

Many engineering problems involve the on-line estimation of the state vector of a system that changes over time using a sequence of noisy observation vectors. Often a recursive filtering approach [1], [8] is employed as it allows an estimate to be made each time a new observation is received without having to reprocess past observations. Arguably, the most intuitively appealing example, which is the application motivating this work, is target tracking. Here the state vector contains the kinematic characteristics of the target under surveillance, and the observation vector is a (noise corrupted) sensor measurement such as a radar return. Recursive state estimation is also an important problem in other scientific disciplines, see for example [23], [10], [3].

An interesting and important generalisation of the above filtering problem arises when the state of the system and observations are no longer vectors but *random finite sets*. In the context of target tracking, this corresponds to a multiple target scenario. Instead of a single target there are now many targets whose states are to be estimated. The number of targets changes with time due to new targets constantly entering the surveillance region and old ones leaving it. Like the single target case, a sensor collects measurements from the multiple targets. However, some of the targets may not be detected by the sensor. Additionally, the sensor also receives a random number of false measurements. (In the case of a radar, this may is due to non-target generated reflections.) As a result, the observation at each time step is a set of measurements of unknown origin, only some of which are generated by the targets while the remainder are false (or clutter). The initial number of targets is not known and only a prior distribution on their number is available. Since not all the targets are observed at each time and new targets are constantly being introduced to the surveillance region, the number of targets are not known and also has to be estimated. More generally, the aim is to estimate, at each time step, the time-varying state set from the entire history of observation sets.

*Dept. of Engineering, University of Cambridge, UK (sss40@cam.ac.uk).

†Dept. of Electrical and Electronic Engineering, University of Melbourne, Australia (bv@ee.unimelb.edu.au).

‡School of Mathematics and Statistics, University of Western Australia, Australia (adrian@maths.uwa.edu.au).

§Dept. of Statistics and Modeling Science, University of Strathclyde, Glasgow, UK (sergei@stams.strath.ac.uk).

We commence with the study of the following static inference problem. We regard the unknown number of targets and their states as a hidden Point Process (PP) $\mathbf{X}$ on a complete separable metric space (CSMS) $\mathcal{X}$, and the collection of sensor measurements as the observed point process $\mathbf{Y}$ on a CSMS $\mathcal{Y}$. The problem of interest can be simplified to that estimating the hidden PP $\mathbf{X}$ given a realisation of the observed PP $\mathbf{Y}$. A prior for the hidden PP $\mathbf{X}$ together with the likelihood for $\mathbf{Y}$ gives the posterior distribution for $\mathbf{X}$ via the application of Bayes rule. Although such a PP formulation for multi-target tracking is not new, see [14], [22], [17], [7], there are very few works in the open literature that aim to characterise the posterior of $\mathbf{X}$ given $\mathbf{Y}$. This problem was studied by [11], [12] for an unrelated application in Forestry. For the transformation from $\mathbf{X}$ to $\mathbf{Y}$, the authors considered several disturbance mechanisms which included random thinning, displacement and augmentation of extra points (more details in Section 3.1). The prior distribution was assumed to be a regular PP such as a Strauss process. In this case the posterior does not admit a computationally tractable analytic characterisation and it was approximated numerically using Markov Chain Monte Carlo [12]. In the tracking literature the problem was studied in [13]. For this same observation model and a Poisson prior for $\mathbf{X}$, an expression relating the intensity (or the first moment) of the posterior and the prior was derived. (The Poisson PP is completely characterised by its first moment and it can be shown that all one needs to do to find the best Poisson approximation to the posterior is to characterise the posterior intensity [13]; see also Lemma A.3.) In addition, under the assumption of a time varying hidden process with Markovian dynamics, the author also derived the intensity for the posterior predicted one step ahead in time. These results were combined to yield a filter that propagates the intensity of the hidden PP and is known in the tracking literature as the Probability Hypothesis Density (PHD) filter [13].

To the best of our knowledge, [13] is the only work geared towards a practical on-line filtering approach for a Markov-in-time hidden spatial PP observed in noise. Detailed numerical studies using Sequential Monte Carlo approximations [20, 21] (and references therein), as well as Gaussian approximations [19] to the PHD filter have since demonstrated its potential as a powerful new approach to multi-target tracking. Partially motivated by this, in this paper we extend the results in [13] in several directions.

- In the case of the Poisson prior, the posterior was characterised only *indirectly* in [13] by providing the formula for its Probability Generating Functional (p.g.fl.) The author arrived at this formula by differentiating the joint p.g.fl. of the observed and hidden process. While this is a general proof technique, it is a rather technical approach (see the penultimate paragraph of this section) that does not exploit the structure of the problem - a Poisson prior and an observed process constructed via thinning, displacement and augmentation allows for a considerably stronger result with a simpler proof by calling upon several well known results concerning the Poisson PP [9]. In doing so, we are able to provide a closed-form expression for the posterior which is quite revealing on the structure of the conditional process $\mathbf{X}$ given the observed process $\mathbf{Y}$. Corollaries of this result include the expression relating the intensity of the posterior and prior (Cor. 4.2) as well as the law of the association of the points of the observed process (Cor. 4.3).
- While the result in [13] is only for a Poisson prior for $\mathbf{X}$, we extend the result to a Gauss-Poisson prior which covers the Poisson prior as a special case. (This extension is interesting for several reasons as detailed in the next

paragraph.) Using the characterisation of the posterior for a Gauss-Poisson hidden process, we apply the result in an online filtering context. We derive a new filter that propagates the first and second moment of a Markov-in-time hidden PP observed in noise which generalises the first moment filter proposed in [13].

A Gauss-Poisson PP generalises the Poisson PP by allowing for two-point clusters in its realization. This has important modeling implications. We consider a multiple-target tracking model where in addition to birth and deaths of targets, existing targets may spawn new targets. Such a model has already been proposed [13] but the difference is that our model can explicitly account for correlated movement between the parent and spawned target. Under this model, it is shown that the marginal distribution of the hidden process which represents the targets is Gauss-Poisson for all time; see Prop. 5.1. It is proper to approximate the posterior in this case by retaining the first two moments to more faithfully capture the correlation between the targets, which is of course not possible by just retaining the first moment. A PP which is completely characterised by its first two moments is the Gauss-Poisson process. Therefor this motivates the characterisation of the posterior for a Gauss-Poisson prior and the subsequent approximation of it by the closest fitting Gauss-Poisson process. (See Section 5.1 for more details.)

In the absence of spawning, our generalisation to a Gauss-Poisson hidden process yields a more flexible observation model. The observed process is generated assuming that either no measurement or only a single measurement is recorded from a target during the measurement acquisition phase by the sensor. This has been noted to be restrictive [6]. For example, when a target is closer in proximity to the measurement acquisition sensor, the resolution is improved and the sensor may be able to resolve individual features on the target [6]. Pre-processing of data from sensors prior to statistical analysis may also result in more than one measurement per target as, for example, during the process of converting an image based observation to point process observation by clustering intensity peaks in the image and extracting the center of these peaks. Our Gauss-Poisson generalisation results in a model where a target can generate at most two measurements per time.

We also clarify some important technicalities concerning the use of the derivatives of the joint p.g.fl. to characterise the posterior. These issues, which are fundamental to the proof technique, were not considered in [13].

This paper is organised as follows. Section 2 contains necessary definitions and notation as well as some basic facts about Point Processes. Section 3 describes the observation and underlying process models we will be dealing with. In the main Section 4 we characterise the posterior distributions for Poisson and Gauss-Poisson priors. Finally, Section 5 addresses the practical problem of online filtering and develops on the theoretical results obtained in the previous section. The Appendix contains some necessary auxiliary statements and technical proofs.

2. Definitions and Notation.

2.1. Preliminaries. Our primary motivation are applications that involve an unseen simple finite point process (PP) $\mathbf{X}$ which is observed indirectly through another finite PP $\mathbf{Y}$. It can be defined with the help of *disjoint union* as the canonical probability space (see, e.g., [15, 4]).

Given a complete separable metric space (CSMS) $\mathcal{X}$, denote by $\mathcal{B}(\mathcal{X})$ the Borel

σ-algebra on $\mathcal{X}$. Define the *state space* for the hidden PP $\mathbf{X}$ to be the disjoint union

$$\mathcal{X}^{\cup} = \bigcup_{n\in\mathbb{N}} \mathcal{X}^n,$$

where $\mathcal{X}^0 = \{\emptyset\}$ corresponds to empty configuration, $\mathcal{X}^n$ denotes a n-fold Cartesian product of $\mathcal{X}$ and $\mathbb{N}$ denotes the set of non-negative integers. Hence, any realisation $\mathbf{x}$ of $\mathbf{X}$ belongs to $\mathcal{X}^n$ for some $n \in \mathbb{N}$.

The collection of measurable sets of $\mathcal{X}^{\cup}$, denoted by $\sigma(\mathcal{X}^{\cup})$, is the class of all sets $A = \bigcup_{n\in\mathbb{N}} A^{(n)}$ in $\mathcal{X}^{\cup}$ such that $A^{(n)} \in \mathcal{B}(\mathcal{X}^n)$. A probability measure $P_{\mathbf{X}}$ on $(\mathcal{X}^{\cup}, \sigma(\mathcal{X}^{\cup}))$ can be defined by specifying its restrictions $P_{\mathbf{X}}^{(n)}$ on $(\mathcal{X}^n, \mathcal{B}(\mathcal{X}^n)$ for each $n \in \mathbb{N}$. The probability that there are n points is $p_n = P_{\mathbf{X}}^{(n)}(\mathcal{X}^n)$ and p_n should sum to one, i.e. $\sum_{n\in\mathbb{N}} p_n = 1$.

Let $A^{(n)}$ be a subset of $\mathcal{X}^n$ and σ a permutation $i_1, \ldots, i_n$ of $1, \ldots, n$. Let $A_\sigma^{(n)}$ be the set obtained from $A^{(n)}$ by the permutation $(x_{i_1}, \ldots, x_{i_n})$ of the coordinates of each vector in $A^{(n)}$. The measure $P_{\mathbf{X}}^{(n)}$ is symmetric if $P_{\mathbf{X}}^{(n)}(A_\sigma^{(n)}) = P_{\mathbf{X}}^{(n)}(A^{(n)})$ for all permutations σ. There is no preference in the ordering of the points in a realisation of a PP which may be thought of as random finite set. To be consistent with this interpretation we require $P_{\mathbf{X}}^{(n)}$ (for all n) to be a symmetric measure on $\mathcal{B}(\mathcal{X}^n)$, which implies that equal weight is given to all $n!$ permutations of a realisation $\mathbf{x} \in \mathcal{X}^n$.

A real valued function f on $\mathcal{X}^{\cup}$ is measurable if and only if each f_n, which is the restriction of f to $\mathcal{X}^n$, is a measurable function on $\mathcal{X}^n$, $n \in \mathbb{N}$. The integral of f with respect to $P_{\mathbf{X}}$ is defined to be

$$\int_{\mathcal{X}^{\cup}} P_{\mathbf{X}}(d\mathbf{x}) f(\mathbf{x}) = \sum_{n=0}^{\infty} \int_{\mathcal{X}^n} P_{\mathbf{X}}^{(n)}(dx_1, ..., dx_n) f_n(x_1, ..., x_n).$$

When for all n $f_n(x_1, ..., x_n) = \sum_{i=1}^{n} \mathbb{I}_A(x_i)$, $A \in \mathcal{B}(\mathcal{X})$, the above integral returns the expected number of points of $\mathbf{X}$ that fall in the set A [4, 18]. We denote this measure by $V_{\mathbf{X}}(dx)$, which is known as the *intensity measure* or *first moment* of the PP $\mathbf{X}$ [4].

Hereon, the notation $G[h_1, \ldots, h_n]$ will be used to denote a functional G evaluated at $h_1, \ldots, h_n$. A measure $\mu^{(n)}$ on the nth-product space $(\mathcal{X}^n, \mathcal{B}(\mathcal{X}^n))$, can be treated as a multi-linear functional, in which case the following functional notation applies

$$\mu^{(n)}[h_1, \ldots, h_n] = \int_{\mathcal{X}^n} \mu^{(n)}(dx_1, ..., dx_n) h_1(x_1) \cdots h_n(x_n).$$

A kernel $K: \mathcal{B}(\mathcal{Y}) \times \mathcal{X} \to \mathbb{R}_+$ is a measurable function $x \mapsto K(B|x)$ for every set $B \in \mathcal{B}(\mathcal{Y})$ and a measure $K(\cdot|x)$ on $\mathcal{B}(\mathcal{Y})$ for every $x \in \mathcal{X}$. For a kernel K and a function h on $\mathcal{Y}$ the following linear operator notation applies:

$$(Kh)(x) = \int_{\mathcal{Y}} K(dy|x) h(y).$$

2.2. Probability Generating Functional. The probability generating functional (p.g.fl.) is a fundamental descriptor of a PP and will play an important role in the development of some of the results. Let $\mathcal{U}(\mathcal{X})$ be class of all bounded measurable complex valued functions g on $\mathcal{X}$ which differ from 1 only on a compact set. The p.g.fl. $G_{\mathbf{X}}$ of a point process $\mathbf{X}$ on $\mathcal{X}$ is a functional defined by [15], [4],

$$G_{\mathbf{X}}[g] \equiv \sum_{n=0}^{\infty} P_{\mathbf{X}}^{(n)}[g, \ldots, g], \tag{2.1}$$

on a domain $D \subset \mathcal{U}(\mathcal{X})$ that includes the unit ball $\{g : \|g\| \leq 1\}$, where $\|\cdot\|$ denotes the supremum norm. Let g and ζ be fixed elements of $\mathcal{U}(\mathcal{X})$. The functional derivative at g in the direction ζ can be defined, if the limit exists, as follows

$$G_{\mathbf{X}}^{(1)}[g;\zeta] = \lim_{\varepsilon \downarrow 0} \frac{G_{\mathbf{X}}[g+\varepsilon\zeta] - G_{\mathbf{X}}[g]}{\varepsilon}$$

The domain of $G_{\mathbf{X}}^{(1)}[\cdot;\zeta]$ contains $\{g : \|g\| < 1\}$, since $\|g + \varepsilon\zeta\| < 1$ for all sufficiently small ε. Similarly, we can define the mth $(m > 1)$ iterated derivative as:

$$G_{\mathbf{X}}^{(m)}[g;\zeta_1,...,\zeta_m] = \lim_{\varepsilon \to 0} \frac{G_{\mathbf{X}}^{(m-1)}[g+\varepsilon\zeta_k;\zeta_1,...,\zeta_{m-1}] - G_{\mathbf{X}}^{(m-1)}[g;\zeta_1,...,\zeta_{m-1}]}{\varepsilon}.$$

The law of $\mathbf{X}$ can be recovered from $G_{\mathbf{X}}$ by differentiation [15]

$$G_{\mathbf{X}}^{(m)}[0;\zeta_1,...,\zeta_m] = m! P_{\mathbf{X}}^{(m)}[\zeta_1,\ldots,\zeta_m]. \tag{2.2}$$

The intensity measure $V_{\mathbf{X}}$ can also be obtained by differentiating $G_{\mathbf{X}}$ (see [15, 4])

$$V_{\mathbf{X}}(A) = \lim_{g \uparrow 1} G_{\mathbf{X}}^{(1)}[g;\mathbb{I}_A]. \tag{2.3}$$

Some of the results we present are specialised to Poisson PPs. A Poisson PP $\mathbf{X}$ is completely characterised by its intensity measure $V_{\mathbf{X}}$. To sample $\mathbf{X}$, first the number of points are drawn from a (discrete) Poisson distribution with mean $V_{\mathbf{X}}(\mathcal{X})$, then the location of the points are drawn independently according to $V_{\mathbf{X}}(\cdot)/V_{\mathbf{X}}(\mathcal{X})$. Hence for a Poisson PP,

$$P_{\mathbf{X}}^{(n)} = e^{-V_{\mathbf{X}}(\mathcal{X})} \frac{1}{n!} V_{\mathbf{X}}^{(n)}, \tag{2.4}$$

where $V_{\mathbf{X}}^{(n)}$ is the nth product measure, and its p.g.fl. is

$$G_{\mathbf{X}}[g] = e^{V_{\mathbf{X}}[g-1]}. \tag{2.5}$$

2.3. Bivariate and Marked Point Process. A similar setting applies to the *observed* PP with state space $\mathcal{Y}^{\cup} = \bigcup_{n \in \mathbb{N}} \mathcal{Y}^n$ and measurable sets $\sigma(\mathcal{Y}^{\cup})$ which is defined through realisations of $\mathbf{X}$. The process $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$, denoted $\mathbf{Y}(\mathbf{x})$, has distribution $P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x})$. We assume that the family of distributions $\{P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) : \mathbf{x} \in \mathcal{X}^{\cup}\}$ satisfies the usual regularity assumptions: for each $A \in \sigma(\mathcal{Y}^{\cup})$, $P_{\mathbf{Y}|\mathbf{X}}(A|\cdot)$ is a measurable function on $\mathcal{X}^{\cup}$. The joint probability distribution $P_{\mathbf{Y},\mathbf{X}}$ of the bivariate point process $(\mathbf{X},\mathbf{Y})$ on the product space $(\mathcal{X}^{\cup} \times \mathcal{Y}^{\cup}, \sigma(\mathcal{X}^{\cup}) \otimes \sigma(\mathcal{Y}^{\cup}))$ is then given by

$$P_{\mathbf{X},\mathbf{Y}}(d\mathbf{x}, d\mathbf{y}) = P_{\mathbf{Y}|\mathbf{X}}(d\mathbf{y}|\mathbf{x}) P_{\mathbf{X}}(d\mathbf{x}). \tag{2.6}$$

Applications involve specific realisations of $\mathbf{Y}$, hence we require the likelihood function of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$. Let $\lambda_{\mathcal{Y}}$ be a diffuse reference measure on $\mathcal{Y}$ (the standard case is when $\mathcal{Y}$ is a bounded subset of $\mathbb{R}^d$ and $\lambda_{\mathcal{Y}}$ is the Lebesgue measure). Let $P_{\mathbf{Y}|\mathbf{X}}^{(n)}(\cdot|\mathbf{x})$ denote the restriction of $P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x})$ to $\mathcal{Y}^n$. We assume that for each $n \in \mathbb{N}$, $P_{\mathbf{Y}|\mathbf{X}}^{(n)}(\cdot|\mathbf{x})$ is a symmetric measure on $\mathcal{Y}^n$ which admits a symmetric density $p_{\mathbf{Y}|\mathbf{X}}^{(n)}(\cdot|\mathbf{x})$ relative to $\lambda_{\mathcal{Y}}^n$ and for each $(y_1,\ldots,y_n) \in \mathcal{Y}^n$, $p_{\mathbf{Y}|\mathbf{X}}^{(n)}(y_1,\ldots,y_n|\cdot)$ is a symmetric measurable function on $\mathcal{X}^{\cup}$.

The p.g.fl. $G_{\mathbf{X},\mathbf{Y}}$ of the bivariate point process $(\mathbf{X},\mathbf{Y})$ [15, Section 5] is given by

$$G_{\mathbf{X},\mathbf{Y}}[g,h] = \sum_{n=0}^{\infty} \int_{\mathcal{X}^n} P_{\mathbf{X}}^{(n)}(dx_1 \dots dx_n) G_{\mathbf{Y}|\mathbf{X}}[h|\,(x_1,\dots,x_n)]\, g(x_1)\cdots g(x_n) \tag{2.7}$$

provided the integral exists. $G_{\mathbf{X},\mathbf{Y}}[g,h]$ is well defined for all complex valued measurable g,h satisfying $\|g\| \le 1, ||h|| \le 1$ since the latter implies $\left|G_{\mathbf{Y}|\mathbf{X}}[h|\mathbf{x}]\right| \le 1$. In the interest of brevity henceforth, we denote

$$G_{\mathbf{X},\mathbf{Y}}[g,h] = \int P_{\mathbf{X}}(d\mathbf{x}) G_{\mathbf{Y}|\mathbf{X}}[h|\mathbf{x}]\Pi_{\mathbf{x}}[g] \tag{2.8}$$

where $\Pi_{(x_1,\dots,x_n)}[g] = \prod_{i=1}^n g(x_i)$. We use the notation $G^{(0,m)}_{\mathbf{X},\mathbf{Y}}[g,h;\zeta_1,\dots,\zeta_m]$ to indicate that we differentiate w.r.t. the first argument of $G_{\mathbf{X},\mathbf{Y}}$ zero times and differentiate w.r.to the second argument m times.

An important case of the bivariate PP is when there is a bijection m between the points of the process $\mathbf{X}$ and $\mathbf{Y}$ for almost all their realisations. In this case each realisation $(\mathbf{x},\mathbf{y})$ can be represented as a collection of points $(x_i, m(x_i))$, $x_i \in \mathbf{x}, m(x_i) \in \mathbf{y}$ and all $m(x_i)$ are distinct. It is common to refer to this representation as a *marked PP* with *position space* $\mathcal{X}$ and *mark space* $\mathcal{Y}$. In the sequel we will be dealing with a particular case of bivariate processes where the marks $m(x_i)$ for each point x_i in a realisation $\mathbf{x}$ of $\mathbf{X}$ are drawn independently from distributions $M(\cdot|x_i)$ for a given kernel $M(dy|x)$. This defines the conditional distribution of $\mathbf{Y}$ as

$$P_{\mathbf{Y}|\mathbf{X}}(d\mathbf{y}|\mathbf{x}) = \left[\prod_{x_i\in\mathbf{x}} M(dy_i|x_i)\right]^{\mathrm{sym}} \tag{2.9}$$

if the cardinalities of $\mathbf{x}$ and $\mathbf{y}$ are equal and 0 otherwise. By construction, $P_{\mathbf{Y}|\mathbf{X}}(\emptyset|\emptyset) = 1$. We used above the *symmetrisation* notation meaning averaging over all permutations of indices: if $\mathbf{x}$ has n points and σ is a permutation of $(1,\dots,n)$ then

$$\left[\prod_{x_i\in\mathbf{x}} M(dy_i|x_i)\right]^{\mathrm{sym}} = \frac{1}{n!}\sum_{\sigma}\prod_{i=1}^n M(dy_i|x_{\sigma(i)})\,.$$

If $\mathbf{X}$ is Poisson, then substituting (2.4) and (2.9) into (2.8), the terms combine to produce $G_{\mathbf{X},\mathbf{Y}}[g,h] = e^{V_{\mathbf{X},\mathbf{Y}}[gh-1]}$, corresponding to a Poisson PP on the product space with intensity measure $V_{\mathbf{X},\mathbf{Y}}(dx\,dy) = V_{\mathbf{X}}(dx)M(dy|x)$. This fact is known as the *Marking theorem* for a Poisson process and reflects independence of its points (see, e.g., [9, p.55]).

3. Problem Statement. We aim to characterise the probability distribution of $\mathbf{X}$ conditional on a realisation $\mathbf{y} = (y_1,\dots,y_m)$ of $\mathbf{Y}$ which is defined to be

$$P_{\mathbf{X}|\mathbf{y}}(d\mathbf{x}) = \frac{P_{\mathbf{X}}(d\mathbf{x})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}{\int P_{\mathbf{X}}(d\mathbf{x})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})} \tag{3.1}$$

provided that $\int P_{\mathbf{X}}(d\mathbf{x})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) > 0$, where $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ is the likelihood of the realisation $\mathbf{y}$ of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$. (We use the subscript $\mathbf{X}|\mathbf{y}$ as an abbreviation of $\mathbf{X}|\mathbf{Y} = \mathbf{y}$.) We will consider both a Poisson and Gauss-Poisson prior for $\mathbf{X}$. We will also characterise the p.g.fl. of the posterior:

$$G_{\mathbf{X}|\mathbf{y}}[g] = \frac{\int P_{\mathbf{X}}(d\mathbf{x})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\Pi_{\mathbf{x}}[g]}{\int P_{\mathbf{X}}(d\mathbf{x})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}. \tag{3.2}$$

Our interest in characterising $G_{\mathbf{X}|\mathbf{y}}$ is because one then obtains the intensity measure $V_{\mathbf{X}|\mathbf{y}}(\cdot)$ of $P_{\mathbf{X}|\mathbf{y}}$ straightforwardly by differentiation. The intensity measure $V_{\mathbf{X}|\mathbf{y}}$ plays a central role in online filtering of a PP observed in noise (see Section 5.) It is possible to compute $V_{\mathbf{X}|\mathbf{y}}$ using numerical methods such as Sequential Monte Carlo [5, 21]. The intensity is regarded as a "sufficient statistic" for $P_{\mathbf{X}|\mathbf{y}}$. In fact it can be shown that a Poisson PP with intensity $V_{\mathbf{X}|\mathbf{y}}$ is the best Poisson approximation to $P_{\mathbf{X}|\mathbf{y}}$ in the Kullback-Leibler (KL) sense [13] (see also Lemma A.3).

3.1. Observation Process Model. Our goal is to characterise $P_{\mathbf{X}|\mathbf{y}}$ for a specific (but quite general) observation model and different priors on $\mathbf{X}$. In this section we present the model for the generation of the observed process $\mathbf{Y}$.

Our model for the observed process covers thinning, Markov shifts and superposition of false observations. Applications of this model include multi-target tracking [13] and spatial statistics [11, 12].

(Detections.) Consider a realisation $\mathbf{x} = (x_1, \dots, x_n)$ of $\mathbf{X}$. Each point x_i in the realisation generates a $\mathcal{Y}$-valued observation in the set $A \in \mathcal{B}(\mathcal{Y})$ with probability

$$L(A|x_i) \tag{3.3}$$

where $L: \mathcal{B}(\mathcal{Y}) \times \mathcal{X} \to [0,1]$ is a regular transition kernel with density $l(\cdot|x)$ relative to the reference measure $\lambda_{\mathcal{Y}}$. This happens independently for each x_i. The set of such points x_i is denoted by $\widehat{\mathbf{X}}$. Note that $1 - L(\mathcal{Y}|x) = \overline{L}(\mathcal{Y}|x)$, is the probability that the point x generates no observation. Subset of the points of $\mathbf{X}$ generating no observation is denoted by $\widetilde{\mathbf{X}}$. Finally, $\mathbf{\Theta}$ stands for the finite PP on $\mathcal{Y}$ of observations induced by $\widehat{\mathbf{X}}$.

(Clutter.) In addition to the detection PP $\mathbf{\Theta}$ generated from $\widehat{\mathbf{X}}$ we also observe points unrelated to $\mathbf{X}$, which are termed *clutter*. We model clutter by a finite Poisson PP $\mathbf{K}$ on $\mathcal{Y}$ with intensity measure $V_{\mathbf{K}}$. We assume that $\mathbf{K}$ is independent of the hidden PP $\mathbf{X}$ and the detection PP $\mathbf{\Theta}$. We assume that $V_{\mathbf{K}}$ admits a density with respect to the reference measure $\lambda_{\mathcal{Y}}$ on $\mathcal{Y}$,

$$V_{\mathbf{K}}(A) = \int_A \lambda_{\mathcal{Y}}(dy) v_{\mathbf{K}}(y). \tag{3.4}$$

The observed process $\mathbf{Y}$ is then a superposition of the detection PP $\mathbf{\Theta}$ and the clutter PP $\mathbf{K}$.

4. Posterior Characterisation. This section presents the main results of the paper where we characterise $P_{\mathbf{X}|\mathbf{y}}$ and $G_{\mathbf{X}|\mathbf{y}}[g]$ for the observation model described in Section 3.1. We commence first with a Poisson hidden process and then present the results for a Gauss-Poisson hidden process.

4.1. Characterisation of the Posterior for a Poisson Prior. The characterisation of $G_{\mathbf{X}|\mathbf{y}}$ for a Poisson $\mathbf{X}$ was first established in [13]. The author characterised conditional p.g.fl. in closed form by differentiating the joint p.g.fl. This is a general proof technique that we follow up on in Section 4.2 when dealing with a Gauss-Poisson hidden process. However, it is a rather technical approach that does not exploit the structure of the problem nor does it shed light on the structure of the conditional process $\mathbf{X}$ given the observation $\mathbf{y}$. In this section we derive $P_{\mathbf{X}|\mathbf{y}}$ for a Poisson prior *explicitly* (and not implicitly through its p.g.fl.) and express it in closed-form. The result is new and the proof is novel as well as intuitive. We commence by outlining the proof technique.

The basic idea is to use the above mentioned decomposition of the hidden PP $\mathbf{X}$ into superposition of $\widetilde{\mathbf{X}}$ and $\widehat{\mathbf{X}}$, where $\widetilde{\mathbf{X}}$ is *unobserved* while $\widehat{\mathbf{X}}$ is observed in noise (through $\mathbf{Y}$). Since $\widehat{\mathbf{X}}$ is obtained from $\mathbf{X}$ by independent marking then both $\widetilde{\mathbf{X}}$ and $\widehat{\mathbf{X}}$ are independent Poisson with respective intensity measures $V_{\widetilde{\mathbf{X}}}(dx) = V_{\mathbf{X}}(dx)\overline{L}(\mathcal{Y}|x)$ and $V_{\widehat{\mathbf{X}}}(dx) = V_{\mathbf{X}}(dx)L(\mathcal{Y}|x)$, see, e.g., [9, p.55]. This decomposition sheds light on the structure of the posterior: since $\widetilde{\mathbf{X}}$ is unobserved, its law is unchanged after observing $\mathbf{Y}$. As for $\widehat{\mathbf{X}}$, let its posterior be $P_{\widehat{\mathbf{X}}|\mathbf{y}}$. Thus, the desired posterior $P_{\mathbf{X}|\mathbf{y}}$ is

$$P_{\mathbf{X}|\mathbf{y}} = P_{\widetilde{\mathbf{X}}} * P_{\widehat{\mathbf{X}}|\mathbf{y}} \tag{4.1}$$

where $*$ denotes convolution, which follows since $\mathbf{X}$ is the superposition of $\widetilde{\mathbf{X}}$ and $\widehat{\mathbf{X}}$. In the special case when there is no clutter, $P_{\widehat{\mathbf{X}}|\mathbf{y}}$ is a Binomial process comprising of the same number of points as in $\mathbf{y}$ as every point in $\mathbf{y}$ is an observation of one and only one point of $\widehat{\mathbf{X}}$. When there is clutter, assuming there were m observations in total, $P_{\widehat{\mathbf{X}}|\mathbf{y}}$ is a distribution on $\mathcal{X}^{\cup}$ with support entirely on $\bigcup_{n\leq m} \mathcal{X}^n$. In Proposition 4.1 below, $P_{\widehat{\mathbf{X}}|\mathbf{y}}$ is characterised explicitly by (4.3) and the result of the convolution is expanded in (4.4).

To formulate the result, we need a further notation. Introduce $M = \{1, \dots, m\}$, $N = \{1, \dots, n\}$ and let $\mathcal{S}(K, K')$ be the set of all bijections between two finite sets K and K' of the same size. It is also convenient to extend the phase space $\mathcal{X}$ to $\mathcal{X}' = \mathcal{X} \cup \{\Delta_1\}$ by adding an isolated point Δ_1. All the points of the clutter process $\mathbf{K}$ are artificially put at Δ_1.

PROPOSITION 4.1. *Suppose that both processes* $\mathbf{X}$ *and* $\mathbf{K}$ *are Poisson with intensity measures* $V_{\mathbf{X}}$ *and* $V_{\mathbf{K}}$ *admitting densities* $v_{\mathbf{X}}(x)$ *and* $v_{\mathbf{K}}(y)$ *w.r.t. reference measures* $\lambda_{\mathcal{X}}$ *on* $\mathcal{X}$ *and* $\lambda_{\mathcal{Y}}$ *on* $\mathcal{Y}$*, respectively. Then the observation process* $\mathbf{Y}$ *is also Poisson with the density* $v_{\mathbf{Y}}$ *given by*

$$v_{\mathbf{Y}}(y) = v_{\mathbf{K}}(y) + V_{\mathbf{X}}[l(y|\cdot)]\,. \tag{4.2}$$

The conditional distribution of $\mathbf{X}$ *given observation* $\mathbf{y} = (y_1, \dots, y_m)$ *of* $\mathbf{Y}$ *coincides with the distribution of the superposition of two independent processes: Poisson process* $\widetilde{\mathbf{X}}$ *of unobserved points with intensity* $\overline{L}(\mathcal{Y}|x)v_{\mathbf{X}}(x)$, $x \in \mathcal{X}$*, and the restriction onto* $\mathcal{X}$ *of a binomial process* $\widehat{\mathbf{X}}'$ *with* m *points in* $\mathcal{X}' = \mathcal{X} \cup \{\Delta_1\}$ *with conditional distribution*

$$P_{\widehat{\mathbf{X}}'|\mathbf{y}}(dx'_1, \dots, dx'_m) = \left[\prod_{i=1}^{m} \mu(dx'_i|y_i)\right]^{sym}, \quad x'_i \in \mathcal{X} \cup \{\Delta_1\},, \tag{4.3}$$

where $\mu(dx'|y)$ *is given by (4.7). Equivalently,*

$$P_{\mathbf{X}|\mathbf{y}}(dx_1, \dots, dx_n) = \sum_{\substack{M_1 \subseteq M \\ N_1 \subseteq N \\ |M_1| = |N_1|}} L_1 L_2 L_3\,, \tag{4.4}$$

where

$$L_1 = \prod_{j \in M \setminus M_1} \frac{v_{\mathbf{K}}(y_j)}{v_{\mathbf{Y}}(y_j)}$$

$$L_2 = \sum_{\sigma \in \mathcal{S}(N_1, M_1)} \prod_{i \in N_1} \frac{l(y_{\sigma(i)}|x_i) v_{\mathbf{X}}(x_i)\, \lambda_{\mathcal{X}}(dx_i)}{v_{\mathbf{Y}}(y_{\sigma(i)})}$$

$$L_3 = \frac{e^{-V_{\mathbf{X}}[\overline{L}(\mathcal{Y}|\cdot)]}}{n!} \prod_{i \in N \setminus N_1} \overline{L}(\mathcal{Y}|x_i) v_{\mathbf{X}}(x_i) \lambda_{\mathcal{X}}(dx_i)\,.$$

The intensity (or first moment) of $P_{\mathbf{X}|\mathbf{y}}$ *has density*

$$v_{\mathbf{X}|\mathbf{y}}(x) = \overline{L}(\mathcal{Y}|x)v_{\mathbf{X}}(x) + \sum_{y_i \in \mathbf{y}} \frac{l(y_i|x)v_{\mathbf{X}}(x)}{v_{\mathbf{Y}}(y_i)}, \quad x \in \mathcal{X} \tag{4.5}$$

Proof. Define a new point process which we call Ξ. Formally, Ξ is a marked point process with position space $\mathcal{X}' = \mathcal{X} \cup \{\Delta_1\}$ and mark space $\mathcal{Y}' = \mathcal{Y} \cup \{\Delta_2\}$, where $\{\Delta_1\}, \{\Delta_2\}$ are two isolated one-point sets not elements of neither $\mathcal{X}$ nor $\mathcal{Y}$. Specifically, every point of the clutter $y \in \mathbf{K}$ gives rise to a point (Δ_1, y) and thus carries its position y as its mark. Then each point $x \in \mathbf{X}$ receives the mark corresponding to its observation y and the mark Δ_2 if it generates no observation, see Figure 4.1.

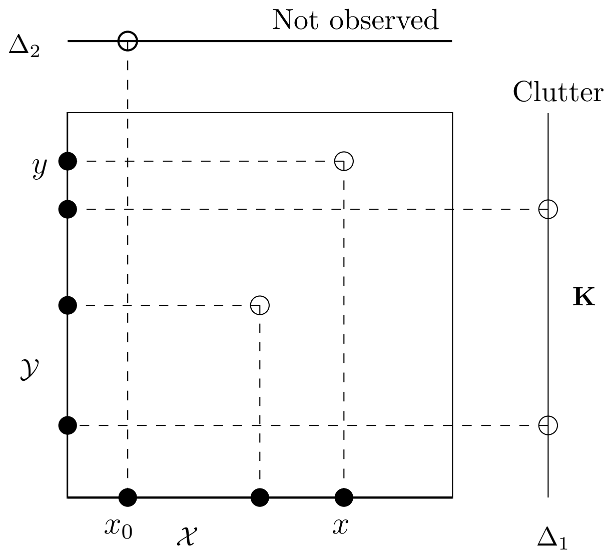

FIG. 4.1. *Process* Ξ *(void dots) and the corresponding processes:* $\mathbf{X}$ *(on horizontal axis),* $\mathbf{K}$ *(vertical line on the right) and* $\mathbf{Y}$ *(vertical axis). The upmost point corresponds to* $x_0 \in \mathbf{X}$ *which generated no observation.*

Both processes $\mathbf{X}$ and $\mathbf{K}$ are Poisson and observation positions are determined independently for each point. Thus, appealing to the Superposition and Marking Theorem for a Poisson process (see, e.g., [9, p.16,p.55]) it can be seen that Ξ is also a Poisson process in the product space $\mathcal{X}' \times \mathcal{Y}'$ driven by an intensity measure $V_\Xi(dx', dy')$. Its restriction on $\mathcal{X} \times \mathcal{Y}$ is the skew product of measures $V_{\mathbf{X}}(dx)L(dy|x)$, while $V_\Xi(\Delta_1, dy) = V_{\mathbf{K}}(dy)$ and $V_\Xi(dx, \Delta_2) = \overline{L}(\mathcal{Y}|x)V_{\mathbf{X}}(dx)$. Finally, $V_\Xi(\Delta_1, \Delta_2) = 0$ as the clutter is observed by definition. The projection of points of Ξ that fall in $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{X} \times \{\Delta_2\}$ onto $\mathcal{X}$ decomposes the process Ξ into the Poisson processes of observed points $\widehat{\mathbf{X}}$ and unobserved points $\widetilde{\mathbf{X}}$ mentioned above.

Now, conversely, consider the second coordinate $\mathcal{Y}'$ as position space and the corresponding first coordinates as marks. Then, using $V_\Xi(dx', dy')$, we can express the mark distributions $\mu(dx'|y')$, $y' \in \mathcal{Y}'$ which, as before, has the meaning of the conditional distribution of the mark x' of a point of $\mathbf{Y}'$ in location y'. We are interested in the observed locations, i.e. when $y' = y \in \mathcal{Y}$.

Although it is possible to deal with non-diffuse measures, for the sake of simplicity and keeping in mind applications considered in this paper we assumed above that $V_{\mathbf{X}}$,

$V_{\mathbf{K}}$ and almost all $L(\cdot|x)$ have the densities $v_{\mathbf{X}}$, $v_{\mathbf{K}}$ and $l(\cdot|x)$ with respect to their corresponding reference measures $\lambda_{\mathcal{X}}$ and $\lambda_{\mathcal{Y}}$. Then the intensity function of $\mathbf{\Xi}$ (the density of $V_{\mathbf{\Xi}}$) in $\mathcal{X} \times \mathcal{Y}$ is

$$v_{\mathbf{\Xi}}(x, y) = l(y|x)v_{\mathbf{X}}(x), \quad x \in \mathcal{X}, y \in \mathcal{Y} \tag{4.6}$$

and $v_{\mathbf{\Xi}}(\Delta_1, y) = v_{\mathbf{K}}(y)$. The marginal density at y is then given by (4.2). Hence, the mark distribution $\mu(dx'|y)$ has two components: diffuse supported by $\mathcal{X}$ and atomic concentrated on $\{\Delta_1\}$:

$$\mu(dx'|y) = \frac{v_{\mathbf{\Xi}}(x', y)\, \lambda_{\mathcal{X}}(dx')}{v_{\mathbf{Y}}(y)}\, \mathbb{I}_{\mathcal{X}}(x') + \frac{v_{\mathbf{K}}(y)}{v_{\mathbf{Y}}(y)}\, \mathbb{I}_{\{\Delta_1\}}(x')\,, \quad x' \in \mathcal{X} \cup \{\Delta_1\} \tag{4.7}$$

which, in view of (2.9), gives rise to (4.3). Expression (4.7) can be interpreted as follows: an observed point $y \in \mathbf{Y}$ is a clutter point with probability $1 - q(y) = v_{\mathbf{K}}(y)/v_{\mathbf{Y}}(y)$ and with complimentary probability $q(y)$ is an observation of some x distributed with probability density $q^{-1}(y)\, v_{\mathbf{\Xi}}(x, y)/v_{\mathbf{Y}}(y)$ in $\mathcal{X}$. Again, marks received by points of $\mathbf{Y}$ and, generally, multiple points at Δ_2 are independent since they are defined through the bivariate Poisson process $\mathbf{\Xi}$. As a consequence, the conditional distribution of $\mathbf{X}'$ which is $\mathbf{X}$ and multiple points at Δ_1 given realisation $\mathbf{y} = (y_1, \ldots, y_m)$ of $\mathbf{Y}$ corresponds to the superposition of two independent processes on $\mathcal{X}'$. One is the Poisson process $\widetilde{\mathbf{X}}$ of non-observed points with density $\overline{L}(\mathcal{Y}|x)v_{\mathbf{X}}(x), x \in \mathcal{X}$ which projects onto $\mathcal{X}$ from $\mathcal{X} \times \{\Delta_2\}$. The other is the Binomial process of m independent points in $\mathcal{X}'$ each following the mark distribution (4.7) of one of the points of $\mathbf{y}$. These two processes give rise to the corresponding two terms in the expression for the conditional intensity (4.5). Taking into account all possible associations of points of $\mathbf{x} = (x_1, \ldots, x_n)$ with points of $\mathbf{y}$ and Δ_2, one arrives at (4.4) with L_1 corresponding to the clutter points, L_2 – to the set M_1 of $\mathbf{X}$-points generating observations in different combinations and L_3 – to the set $N \setminus N_1$ of unobserved points. □

Corollary 4.2 (Mahler [13]). *Conditional p.g.fl for a Poisson* $\mathbf{X}$ *given observations* $\mathbf{y}$ *is given by*

$$G_{\mathbf{X}|\mathbf{y}}[g] = e^{V_{\mathbf{X}}\left[\overline{L}(\mathcal{Y}|\cdot)(g-1)\right]} \prod_{y_i \in \mathbf{y}} \left[\frac{V_{\mathbf{X}}[l(y_i|\cdot)g]}{v_{\mathbf{Y}}(y_i)} + \frac{v_{\mathbf{K}}(y_i)}{v_{\mathbf{Y}}(y_i)}\right] \tag{4.8}$$

for any non-negative real valued bounded function g *and* $v_{\mathbf{Y}}$ *given by (4.2).*

Proof. Extend the function g onto $\mathcal{X} \cup \{\Delta_1\}$ by setting $g'(x') = g(x)\, \mathbb{I}_{\mathcal{X}}(x') + \mathbb{I}_{\Delta_1}(x')$. According to the statement of Proposition 4.1, $P_{\mathbf{X}|\mathbf{y}}$ corresponds to the superposition of two independent processes: a Poisson process $\widetilde{\mathbf{X}}$ with intensity measure $V_{\mathbf{X}}(dx)\overline{L}(\mathcal{Y}|x)$ and the restriction to $\mathcal{X}$ of a binomial process $\widehat{\mathbf{X}}'$. Hence

$$G_{\mathbf{X}|\mathbf{y}}[g] = G_{\widetilde{\mathbf{X}}}[g] G_{\widehat{\mathbf{X}}'|\mathbf{y}}[g'].$$

Now using (4.7) we come to (4.8). Note that the conditional intensity expression (4.5) can also be obtained by differentiating (4.8) at $g = 1$. □

The proof of Proposition 4.1 also characterizes the distribution of the origin of the measurements, i.e., whether they are target generated or clutter. (The notation below was defined prior to Proposition 4.1.)

Corollary 4.3. *Let* $M_1 \subseteq M$ *index the observations that are target generated while* $M \setminus M_1$ *indexes clutter.* $P(M_1|\mathbf{y}) = \prod_{i \in M_1} \frac{V_{\mathbf{X}}[l(y_i|\cdot)]}{v_{\mathbf{Y}}(y_i)} \times \prod_{j \in M \setminus M_1} \frac{v_{\mathbf{K}}(y_j)}{v_{\mathbf{Y}}(y_j)}$.

Proof. Consider the restriction of the process Ξ defined in the proof of Proposition 4.1 to $\mathcal{X}' \times \mathcal{Y}$. Consider the second coordinate $\mathcal{Y}$ as position space and the corresponding first coordinates as marks. Then conditioned on the observed positions, the marks are independently distributed according to (4.7) from which the result follows. ∎

4.2. Characterisation of the Posterior for a Gauss-Poisson Prior. Consider two independent Poisson PPs, one of which is on $\mathcal{X}$ and the other on $\mathcal{X} \times \mathcal{X}$. The Gauss-Poisson PP $\mathbf{X}$, introduced in [16], can be represented as the superposition of the realisations of the two independent PPs to form a single unordered set in $\mathcal{X}$ [2]. Let $\mathbf{W}$ be the Poisson PP on $\mathcal{X}$ with intensity $V_{\mathbf{W}}(dx)$ while $\mathbf{Z}$ the Poisson PP on $\mathcal{X} \times \mathcal{X}$ with intensity $V_{\mathbf{Z}}(dx_1, dx_2)$. Associating a pair of points x_1, x_2 in $\mathcal{X}$ to every point (x_1, x_2) in a realisation of $\mathbf{Z}$, the superposition of these pairs and the points of $\mathbf{W}$ gives rise to a GP $\mathbf{X}$ with p.g.fl.

$$G_{\mathbf{X}}[g] = G_{\mathbf{W}}[g]\; G_{\mathbf{Z}}[f]\big|_{f=g\times g} = \exp V_{\mathbf{W}}[g-1]\, \exp V_{\mathbf{Z}}[f-1]\big|_{f=g\times g}. \tag{4.9}$$

The Brillinger's representation [2] of a GP process implies that, instead of $P_{\mathbf{X}|\mathbf{y}}$, it is equivalent to characterise the posterior $P_{\mathbf{W},\mathbf{Z}|\mathbf{y}}(d\mathbf{w} \times d\mathbf{z})$ which is a joint probability distribution on the product space $\mathcal{X}^{\cup} \times (\mathcal{X} \times \mathcal{X})^{\cup}$ while the prior was the product of independent Poisson distributions $P_{\mathbf{W}}(d\mathbf{w})$ and $P_{\mathbf{Z}}(d\mathbf{z})$. As in the Poisson case of Section 4.1, the prior is not conjugate for the observation model and the posterior will no longer be GP. Thus, we aim to find the best GP approximation to it in Kullback-Leibler (KL) sense.

Let the reference measure on $\mathcal{X}^{\cup} \times (\mathcal{X} \times \mathcal{X})^{\cup}$ be the direct product of the reference measures P_{ref} and Q_{ref} defined on $\mathcal{X}^{\cup}$ and $(\mathcal{X} \times \mathcal{X})^{\cup}$ respectively, where the priors satisfy $P_{\mathbf{W}} \ll P_{\text{ref}}$ and $P_{\mathbf{Z}} \ll Q_{\text{ref}}$. Using standard results concerning densities on product spaces, minimising the KL criterion,

$$KL(P_{\mathbf{W},\mathbf{Z}|\mathbf{y}} \| P_1 \otimes P_2) = \int dP_{\mathbf{W},\mathbf{Z}|\mathbf{y}} \log \frac{dP_{\mathbf{W},\mathbf{Z}|\mathbf{y}}}{d(P_1 \otimes P_2)} \tag{4.10}$$

with respect to $P_1 \otimes P_2$ is equivalent to maximising (see the proof of Lemma A.3)

$$\begin{aligned}&\int dP_{\mathbf{W},\mathbf{Z}|\mathbf{y}} \log \frac{d(P_1 \otimes P_2)}{d(P_{\text{ref}} \otimes Q_{\text{ref}})} \\ &= \int P_{\mathbf{W},\mathbf{Z}|\mathbf{y}}(d\mathbf{w}, (\mathcal{X} \times \mathcal{X})^{\cup}) \log \frac{dP_1}{dP_{\text{ref}}}(\mathbf{w}) + \int P_{\mathbf{W},\mathbf{Z}|\mathbf{y}}(\mathcal{X}^{\cup}, d\mathbf{z}) \log \frac{dP_2}{dQ_{\text{ref}}}(\mathbf{z}),\end{aligned}$$

where the second line follows from the fact that the density of $P_1 \otimes P_2$ w.r.to $P_{\text{ref}} \otimes Q_{\text{ref}}$ is the product $\frac{dP_1}{dP_{\text{ref}}}(\mathbf{w}) \frac{dP_2}{dQ_{\text{ref}}}(\mathbf{z})$. We see immediately that the marginals $P_1 = P_{\mathbf{W}|\mathbf{y}}(d\mathbf{w})$ and $P_2 = P_{\mathbf{Z}|\mathbf{y}}(d\mathbf{z})$ is the solution. We may now combine this observation and Lemma A.3 in the Appendix to obtain the following result. (Lemma A.3 characterises the best Poisson fit to the marginals.)

PROPOSITION 4.4. *Minimising $KL(P_{\mathbf{W},\mathbf{Z}|\mathbf{y}} \| P_1 \otimes P_2)$ with respect to $P_1 \otimes P_2$ with the additional restriction that P_1 and P_2 are Poisson is solved by the Poisson distributions with intensity measures of $P_{\mathbf{W}|\mathbf{y}}(d\mathbf{w})$ and of $P_{\mathbf{Z}|\mathbf{y}}(d\mathbf{z})$.*

The proof technique employed in Section 4.1 to characterise the posterior equally applies for $P_{\mathbf{W},\mathbf{Z}|\mathbf{y}}$. However the convolution of the law of the unobserved and the posterior of the observed process would be cumbersome in this case. It is algebraically more convenient for us to complete the result by first characterising $G_{\mathbf{W}|\mathbf{y}}[g]$ and

$G_{\mathbf{Z}|\mathbf{y}}[f]$ in closed-form. (The intensity measure of the marginals may then be obtained by differentiating these p.g.fl.s.) To do so, we will require the following definitions.

Let

$$\Lambda(h)(x) = 1 - L(\mathcal{Y}|x) + (Lh)(x) = \overline{L}(\mathcal{Y}|x) + (Lh)(x). \tag{4.11}$$

It follows from the defined observation model in Section 3.1, the conditional p.g.fl. $G_{\mathbf{Y}|\mathbf{W},\mathbf{Z}}[h|\mathbf{w},\mathbf{z}]$ is

$$\begin{aligned} &G_{\mathbf{Y}|\mathbf{W},\mathbf{Z}}[h|\mathbf{w} = (w_1,\ldots,w_n), \mathbf{z} = ((z_{1,1},z_{1,2}),\ldots,(z_{m,1},z_{m,2}))] \\ &= G_{\mathbf{K}}[h]\prod_{i=1}^{n}\Lambda(h)(w_i)\prod_{j=1}^{m}\Lambda(h)(z_{j,1})\Lambda(h)(z_{j,2}). \end{aligned} \tag{4.12}$$

Specifically, the expression follows from the fact that each point w_i in $\mathbf{W}$ (and $z_{j,i}$ in $\mathbf{Z}$) generates an observation independently of the remaining points and the clutter process is independent. We now define the joint p.g.fl. $G_{\mathbf{W},\mathbf{Z},\mathbf{Y}}[g,f,h]$ to be

$$G_{\mathbf{W},\mathbf{Z},\mathbf{Y}}[g,f,h] = \int P_{\mathbf{W}}(d\mathbf{w})P_{\mathbf{Z}}(d\mathbf{z})G_{\mathbf{Y}|\mathbf{W},\mathbf{Z}}[h|\mathbf{w},\mathbf{z}]\Pi_{\mathbf{w}}[g]\Pi_{\mathbf{z}}[f] \tag{4.13}$$

where $\Pi_{\mathbf{w}}$ was defined immediately after (2.8). Define the following p.g.fl.s:[1]

$$\begin{aligned} \omega_{i,j}[f] = &\int V_{\mathbf{Z}}(dx_1,dx_2)l(y_i|x_1)l(y_j|x_2)f(x_1,x_2) \\ &+ \int V_{\mathbf{Z}}(dx_1,dx_2)l(y_j|x_1)l(y_i|x_2)f(x_1,x_2), \end{aligned} \tag{4.14}$$

$$\begin{aligned} \Omega_i[g,f] = &\int V_{\mathbf{W}}(dx)g(x)l(y_i|x) + v_{\mathbf{K}}(y_i) \\ &+ \int V_{\mathbf{Z}}(dx_1,dx_2)\overline{L}(\mathcal{Y}|x_1)l(y_i|x_2)f(x_1,x_2) \\ &+ \int V_{\mathbf{Z}}(dx_1,dx_2)l(y_i|x_1)\overline{L}(\mathcal{Y}|x_2)f(x_1,x_2). \end{aligned} \tag{4.15}$$

Let $\sigma\{j;m\}$ denote the set of all distinct partitions of $\{1,2,\ldots,m\}$ into j sets of cardinality 2 and a single set comprised of the remaining $m-2j$ elements. We denote an element of $\sigma\{j;m\}$ by
$\sigma = (\{\sigma_1,\sigma_2\},\ldots,\{\sigma_{2j-1},\sigma_{2j}\},\{\sigma_{2j+1},\ldots,\sigma_m\})$. The proof of the following result appears in the Appendix.

Proposition 4.5. *Let $G_{\mathbf{W},\mathbf{Z},\mathbf{Y}}$ be defined as in (4.13). Then, conditioned on the*

[1] The elements of $\mathbf{w}$, w_i, are not to be confused with the functional $\omega_{i,j}[f]$ defined in (4.14), which is lower case omega.

realisation $\mathbf{y} = (y_1, \ldots, y_m)$ *of* $\mathbf{Y}$,

$$G_{\mathbf{W}|\mathbf{y}}[g] = \frac{\sum_{j=0}^{\lfloor m/2 \rfloor} \sum_{\sigma\{j;m\}} \prod_{i=1}^{j} \omega_{\sigma_{2i-1},\sigma_{2i}}[1] \prod_{i'=2j+1}^{m} \Omega_{\sigma_{i'}}[g,1]}{\sum_{j=0}^{\lfloor m/2 \rfloor} \sum_{\sigma\{j;m\}} \prod_{i=1}^{j} \omega_{\sigma_{2i-1},\sigma_{2i}}[1] \prod_{i'=2j+1}^{m} \Omega_{\sigma_{i'}}[1,1]} \exp V_{\mathbf{W}}[(g-1)\overline{L}(\mathcal{Y}|\cdot)], \tag{4.16}$$

$$G_{\mathbf{Z}|\mathbf{y}}[f] = \frac{\sum_{j=0}^{\lfloor m/2 \rfloor} \sum_{\sigma\{j;m\}} \prod_{i=1}^{j} \omega_{\sigma_{2i-1},\sigma_{2i}}[f] \prod_{i'=2j+1}^{m} \Omega_{\sigma_{i'}}[1,f]}{\sum_{j=0}^{\lfloor m/2 \rfloor} \sum_{\sigma\{j;m\}} \prod_{i=1}^{j} \omega_{\sigma_{2i-1},\sigma_{2i}}[1] \prod_{i'=2j+1}^{m} \Omega_{\sigma_{i'}}[1,1]} \times \exp V_{\mathbf{Z}}[(f-1)\left(\overline{L}(\mathcal{Y}|\cdot) \times \overline{L}(\mathcal{Y}|\cdot)\right)], \tag{4.17}$$

where g, f *are real valued non-negative functions satisfying* $\|g\| \le 1$, $\|f\| \le 1$.

We may differentiate $G_{\mathbf{W}|\mathbf{y}}[g]$ and $G_{\mathbf{Z}|\mathbf{y}}[f]$ to obtain the intensity measures of Proposition 4.4:

$$V_{\mathbf{W}|\mathbf{y}}(A) = V_{\mathbf{W}}[\overline{L}(\mathcal{Y}|\cdot)I_A(\cdot)] + C^{-1}\sum_{i=1}^{m} a_i V_{\mathbf{W}}[l(y_i|\cdot)I_A(\cdot)], \tag{4.18}$$

$$\begin{aligned} V_{\mathbf{Z}|\mathbf{y}}(A) = &\int V_{\mathbf{Z}}(dx_1, dx_2)\overline{L}(\mathcal{Y}|x_1)\overline{L}(\mathcal{Y}|x_2)I_A(x_1,x_2) \\ &+ C^{-1}\sum_{i=1}^{m} a_i \int V_{\mathbf{Z}}(dx_1, dx_2)\left(\overline{L}(\mathcal{Y}|x_1)l(y_i|x_2) + \overline{L}(\mathcal{Y}|x_2)l(y_i|x_1)\right) I_A(x_1,x_2) \\ &+ \sum_{i=1}^{m}\sum_{j=i+1}^{m} \frac{b_{i,j}}{C} \int V_{\mathbf{Z}}(dx_1, dx_2)\left(l(y_i|x_1)l(y_j|x_2) + l(y_j|x_1)l(y_i|x_2)\right) I_A(x_1,x_2) \end{aligned} \tag{4.19}$$

where a_i, $b_{i,j}$ and C are constants and the number of terms in the double sum is $m(m-1)/2$. (The exact expression for these constants can be easily deduced from the expressions for $G_{\mathbf{W}|\mathbf{y}}[g]$ and $G_{\mathbf{Z}|\mathbf{y}}[f]$.) However, computing these constants is computationally prohibitive for large number of observations m and below we detail a simple approximation scheme. The basic idea is to retain only those terms that contribute most to the evaluation of a_i, $b_{i,j}$ and C in the sense to be made precise now.

A convenient modeling assumption is that the observation kernel admits the following representation:

$$L(A|x) = p_D(x)\int_A \lambda_{\mathcal{Y}}(dy)\widetilde{l}(y|x), \tag{4.20}$$

where $p_D(x)$ is the detection probability and $\widetilde{l}(y|x)$, which integrates to one, determines the distribution of the observations. Let also the detection probability be uniformly bounded by a constant $p_D(x) \le \beta$. The parameter β will subsequently be varied from 0 to 1, but it will have no effect on $\widetilde{l}(y|x)$ in (4.20); it serves only to diminish the detection probability.

In the definition of Ω_i and $\omega_{i,j}$, replace all instances of $\overline{L}(\mathcal{Y}|x)$ with $1 - p_D(x)$ and $l(y|x)$ with $p_D(x)\widetilde{l}(y|x)$. $G_{\mathbf{W}|\mathbf{y}}[g]$, omitting the $\exp V_{\mathbf{W}}[(g-1)\overline{L}(\mathcal{Y}|\cdot)]$ term, is a weighted sum of p.g.fl.s of the following form, $\prod_{i'=2j+1}^{m} \Omega_{\sigma_{i'}}[g,1]$ where the weighting term is proportional to $\prod_{i=1}^{j} \omega_{\sigma_{2i-1},\sigma_{2i}}[1]$. Thus $V_{\mathbf{W}|\mathbf{y}}$ is also a mixture of intensity measures (which are uniformly bounded in β), one from each p.g.fl., with the same weight as its corresponding p.g.fl. All intensities in the numerator and denominator of $V_{\mathbf{W}|\mathbf{y}}$ having one or more ω term in its weight is $o(\beta)$. Ignoring all terms in the mixture with an $o(\beta)$ weight would give rise to an $o(\beta)$ approximation to $V_{\mathbf{W}|\mathbf{y}}$.[2](A similar argument applies for an $o(\beta)$ approximation to $V_{\mathbf{Z}|\mathbf{y}}$.) Thus an $o(\beta)$ approximation to the first moments in (4.18)-(4.19) is $C = 1$, $b_{i,j} = 0$ for all i, j and

$$a_i = \frac{1}{\Omega_i[1,1]}. \tag{4.21}$$

Proceeding similarly, to obtain an $o(\beta^3)$ approximation, note that any weight with two or more ω terms is $o(\beta^3)$. The corresponding $o(\beta^3)$ approximation to $V_{\mathbf{W}|\mathbf{y}}$ and $V_{\mathbf{Z}|\mathbf{y}}$ is obtained by omitting mixture terms with $o(\beta^3)$ weights, which yields (4.18)-(4.19) with

$$a_k = \prod_{\substack{i=1\\ i\neq k}}^{m} \Omega_i[1,1] + \sum_{\substack{i=1\\ i\neq k}}^{m} \sum_{\substack{j=i+1\\ j\neq k}}^{m} \omega_{i,j}[1] \prod_{\substack{l=1\\ l\neq i,j,k}}^{m} \Omega_l[1,1], \tag{4.22}$$

$$b_{i,j} = \prod_{\substack{k=1\\ k\neq i,j}}^{m} \Omega_k[1,1], \tag{4.23}$$

$$C = \prod_{i=1}^{m} \Omega_i[1,1] + \sum_{i=1}^{m} \sum_{j=i+1}^{m} \omega_{i,j}[1] b_{i,j}, \tag{4.24}$$

5. Application to Filtering. In this section we apply the formulae for the posterior derived in Section 4.2 to an online filtering problem defined as follows. We do so for the case when the hidden process is GP as it yields a new algorithm that generalises the Poisson hidden process case originally derived in [13]; the latter is recovered as a special case.

Let $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k\geq 0}$ be a discrete-time bivariate process where

- the hidden process $\{\mathbf{X}_k\}_{k\geq 0}$ is a Markov chain with state space $\mathcal{X}^{\cup}$.
- The observed process $\{\mathbf{Y}_k\}_{k\geq 0}$ takes values in $\mathcal{Y}^{\cup}$ and conditioned on $\{\mathbf{X}_k\}_{k\geq 0}$, $\{\mathbf{Y}_k\}_{k\geq 0}$ is an independent sequence. Furthermore, for any k the distribution of $\mathbf{Y}_k$ depends on $\mathbf{X}_k$ only.

This is the standard hidden Markov model (HMM) framework but defined on a general state space; see [8, 3] for background theory. The distribution of $\mathbf{Y}_k$ given $\mathbf{X}_k$ has been described in Section 3.1. Below, we described the hidden process dynamics.

5.1. Hidden Process Dynamics. In Section 4.2, when considering the GP hidden process, it was equivalent to represent $\mathbf{X}_k$ as $(\mathbf{W}_k, \mathbf{Z}_k)$ where $\mathbf{W}_k$ and $\mathbf{Z}_k$ are independent Poisson processes on $\mathcal{X}$ and $\mathcal{X} \times \mathcal{X}$ respectively. The dynamics of $\{(\mathbf{W}_k, \mathbf{Z}_k)\}_{k\geq 0}$ now described, which are general enough for many target tracking applications, has the important property that the distribution of $(\mathbf{W}_k, \mathbf{Z}_k)$, for all k, is the product of independent Poisson distributions.

---

[2]More precisely the remainder $V_{\mathbf{W}|\mathbf{y}}(dx) - \widehat{V}_{\mathbf{W}|\mathbf{y}}(dx)$ is $o(\beta)$.

Given a realisation $\mathbf{w} = (w_1, \ldots, w_n)$ and $\mathbf{z} = (z_1, \ldots, z_m)$ of $\mathbf{Z}_k$ and $\mathbf{W}_k$ respectively, we describe the transition to $(\mathbf{W}_{k+1}, \mathbf{Z}_{k+1})$. The hidden process dynamics is comprised of *Markov shifts*, *spawning* and independent *birth* . All points of $\mathbf{W}_k$ and $\mathbf{Z}_k$ undergo a Markov shift while only points of $\mathbf{W}_k$ may spawn. The independent birth process is defined on $\mathcal{X}$ only although one may trivially do so for $\mathcal{X} \times \mathcal{X}$ as well. We assume $\mathbf{W}_k$ and $\mathbf{Z}_k$ are independent Poisson processes with intensities $V_{\mathbf{W}_k}$ and $V_{\mathbf{Z}_k}$.

Consider a realisation $\mathbf{w} = (w_1, \ldots, w_n)$ and $\mathbf{z} = (z_1, \ldots, z_m)$ of $\mathbf{W}_k$ and $\mathbf{Z}_k$ respectively. The transition to $(\mathbf{W}_{k+1}, \mathbf{Z}_{k+1})$ is defined as follows:

(Markov shift.) Let $F_1 : \mathcal{B}(\mathcal{X} \cup \mathcal{X}^2) \times \mathcal{X} \to [0, 1]$ be a regular Markov kernel. Let $F_{1,1}(\cdot|x)$ denote the restriction of $F_1(\cdot|x)$ to $\mathcal{B}(\mathcal{X})$ while $F_{1,2}(\cdot|x)$ its restriction to $\mathcal{B}(\mathcal{X}^2)$. Each point w_i of $\mathbf{w}$ undergoes a Markov shift through $F_1(\cdot|w_i)$ independently of the remaining points in $\mathbf{w}$ and that of $\mathbf{z}$. The probability w_i is deleted is $1 - F_{1,1}(\mathcal{X}|w_i) - F_{1,2}(\mathcal{X}^2|w_i)$, the probability it survives and does not spawn is $F_{1,1}(\mathcal{X}|w_i)$, while the probability it survives and spawns is $F_{1,2}(\mathcal{X}^2|w_i)$. If w_i survives and does not spawn the kernel $F_{1,1}(\cdot|w_i)$, upon normalisation, determines the new location of w_i. For example, w_i is shifted into $A \in \mathcal{B}(\mathcal{X})$ with probability $F_{1,1}(A|w_i)$. If w_i survives and spawns then the new location of w_i and its child are jointly determined by the kernel $F_{1,2}(\cdot|w_i)$ (upon normalisation).

Similarly, let $F_2 : \mathcal{B}(\mathcal{X} \cup \mathcal{X}^2) \times \mathcal{X}^2 \to [0, 1]$ be a regular Markov kernel, let $F_{2,1}(\cdot|z)$ be its restriction to $\mathcal{B}(\mathcal{X})$ and $F_{2,2}(\cdot|z)$ its restriction to $\mathcal{B}(\mathcal{X}^2)$. Each point z_i of $\mathbf{z}$ is shifted into $A \in \mathcal{B}(\mathcal{X})$ with probability $F_{2,1}(A|z_i)$ or $B \in \mathcal{B}(\mathcal{X}^2)$ with probability $F_{2,2}(B|z_i)$. This occurs independently of the remaining points in $\mathbf{z}$ and that of $\mathbf{w}$. Since for each z, $F_2(\mathcal{X} \cup \mathcal{X}^2|z) = F_{2,1}(\mathcal{X}|z) + F_{2,2}(\mathcal{X}^2|z) \leq 1$, the point z itself may be deleted.

(Independent Birth.) New points unrelated to $\mathbf{w}$ and $\mathbf{z}$ are generated. Let $\mathbf{\Gamma}$ be a Poisson PP on $\mathcal{X}$ with intensity $V_{\mathbf{\Gamma}}$ that describes the new points generated, i.e., the realisation of $\mathbf{\Gamma}$ is the new points.

The specification of the kernel $F_{2,2}$ which governs the motion of the parent and spawned target is flexible. If targets that spawn have a more constrained motion compared to childless targets, e.g. they move slower or move in unison with their child, then this may be effected with a suitable choice of kernel $F_{2,2}$. If in the application targets do not spawn but can generate more than one measurement and the probability of doing so is location dependent, then the kernel $F_{1,2}$ may be selected so that the "spawned" target coincides in location with its parent. It is natural to assume that the probability of generating more than one measurement will be location dependent. For example, when a target is closer in proximity to the measurement acquisition sensor, the resolution is improved and the sensor may be able to resolve individual features on the target [6]. The limitation in our model is the fact that no more than two measurements are allowed.

Given that each point of $\mathbf{w}$ and $\mathbf{z}$ undergoes a Markov shift independently of the remaining points, and the birth process is independent, we may invoke the Superposition and Marking theorems for a Poisson process [9] to conclude that:

PROPOSITION 5.1. *$\mathbf{W}_{k+1}$ and $\mathbf{Z}_{k+1}$ are independent Poisson processes with intensities*

$$V_{\mathbf{W}_{k+1}}(A) = V_{\mathbf{\Gamma}}(A) + \int_A V_{\mathbf{W}_k}(dw) F_{1,1}(A|w) + \int_A V_{\mathbf{Z}_k}(dz) F_{2,1}(A|z), \tag{5.1}$$

$$V_{\mathbf{Z}_{k+1}}(B) = \int_B V_{\mathbf{Z}_k}(dz) F_{2,2}(B|z) + \int_B V_{\mathbf{W}_k}(dw) F_{1,2}(B|w). \tag{5.2}$$

5.2. Online Filtering. The prediction step in (5.1)-(5.2) may be combined with update formulae in (4.18)-(4.19) to give the following algorithm for the online filtering of the HMM $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k\geq 0}$ where the observation process was defined in Section 3.1. The basic idea behind the filter below is as follows. We assume that the law of $(\mathbf{W}_{k-1}, \mathbf{Z}_{k-1})$ given $\mathbf{Y}_1, \dots \mathbf{Y}_{k-1}$ is the product of independent Poisson distributions with intensities $(V_{\mathbf{W},k-1}, V_{\mathbf{Z},k-1})$. The hidden process dynamics is such that the predicted distribution, i.e. the law of $(\mathbf{W}_k, \mathbf{Z}_k)$ given $\mathbf{Y}_{1:k-1}$, is also the product of independent Poisson distributions with intensities given by (5.1)-(5.2). As the prior is not conjugate for the observation model, upon observing $\mathbf{Y}_k$, we have to invoke Proposition 4.4 to maintain this simple product of Poisson representation of the filter.

ALGORITHM 5.2.

Initialisation: *Let* $(\mathbf{W}_0, \mathbf{Z}_0)$ *be independent Poisson processes with intensities* $(V_{\mathbf{W},0}, V_{\mathbf{Z},0})$. *At iteration* $k \geq 1$, *perform the following two steps:*

Prediction:*From* $V_{\mathbf{W},k-1}$ *and* $V_{\mathbf{Z},k-1}$, *define the predicted intensities* $V_{\mathbf{W},k|k-1}$ *and* $V_{\mathbf{Z},k|k-1}$ *to be*

$$V_{\mathbf{W},k|k-1}(A) = V_{\mathbf{\Gamma}}(A) + \int_A V_{\mathbf{W},k-1}(dw)F_{1,1}(A|w) + \int_A V_{\mathbf{Z},k-1}(dz)F_{2,1}(A|z),$$

$$V_{\mathbf{Z},k|k-1}(B) = \int_B V_{\mathbf{Z},k-1}(dz)F_{2,2}(B|z) + \int_B V_{\mathbf{W},k-1}(dw)F_{1,2}(B|w).$$

Update: *Upon observing* $\mathbf{Y}_k = (y_1, \dots, y_m)$ *(m is not fixed!) update the intensities* $V_{\mathbf{W},k|k-1}$ *and* $V_{\mathbf{Z},k|k-1}$ *using (4.18) and (4.19) to obtain* $V_{\mathbf{W},k}$ *and* $V_{\mathbf{Z},k}$. *For the* $o(\beta)$ *approximation,* $C = 1$, $b_{i,j} = 0$ *for all* i,j *and* $a_i = 1/\Omega_i[1,1]$ *where*

$$\Omega_i[1,1] = \int V_{\mathbf{W},k|k-1}(dx)l(y_i|x) + \int \widetilde{V}_{\mathbf{K}}(dx)l(y_i|x) + \int V_{\mathbf{Z},k|k-1}(dx_1, dx_2)\left(\overline{L}(\mathcal{Y}|x_1)l(y_i|x_2) + l(y_i|x_1)\overline{L}(\mathcal{Y}|x_2)\right).$$

For the order $o(\beta^3)$ *approximation,* a_k, $b_{i,j}$ *and* C *are given by (4.22)-(4.24) with*

$$\omega_{i,j}[1] = \int V_{\mathbf{Z},k|k-1}(dx_1, dx_2)\left(l(y_i|x_1)l(y_j|x_2) + l(y_j|x_1)l(y_i|x_2)\right).$$

6. Conclusion. This paper was geared towards a practical on-line filtering approach for a Markov-in-time hidden spatial PP observed in noise with applications to target tracking. We considered both a Poisson and Gauss-Poisson hidden process, and an observed process constructed via thinning, displacement and augmentation with clutter. The intensity measure played a central role in online filtering. While it is not practical to compute the posterior numerically, the intensity can be computed using Sequential Monte Carlo based numerical schemes [5, 21]. In the case of the Poisson prior, we were able to provide a closed-form expression for the posterior of the corresponding static problem, from which several corollaries followed. For a Gauss-Poisson hidden process, we characterised the best product of Poissons approximation to the posterior. This result was then applied to a Markov-in-time Gauss-Poisson hidden process which undergoes thinning, displacement and augmentation to yield a new filter.

Appendix A. Appendix.

Our proof technique below is inspired by [13] where for a Poisson prior, $G_{\mathbf{X}|\mathbf{y}}[g]$ was given in closed-form by differentiating the joint p.g.fl. (Note that $P_{\mathbf{X}|\mathbf{y}}$ for this case was characterised in closed-form in Section 4.1.) While we will also characterize the posterior by differentiating the joint p.g.fl., the technical arguments used in the proof of Proposition 4.5 below are entirely different.

We will need the following technical Lemma concerning the validity of interchanging the order of integration and differentiation. (This issue was not addressed in [13].) The proof of this lemma is omitted as it is a standard (but tedious) application of the Dominated Convergence Theorem.

Recall the definition of the observed process in Section 3.1. The observed process $\mathbf{Y}$ is the superposition of the detection PP $\boldsymbol{\Theta}$ and the clutter PP $\mathbf{K}$. The various independence assumptions in the definition of these processes give rise to the following expression for the conditional p.g.fl. of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$:

$$G_{\mathbf{Y}|\mathbf{X}}[h|\mathbf{x}] = G_{\mathbf{K}}[h]G_{\boldsymbol{\Theta}|\mathbf{X}}[h|\mathbf{x}] = e^{V_{\mathbf{K}}[h-1]}\Pi_{\mathbf{x}}[\Lambda(h)] \tag{A.1}$$

where $G_{\boldsymbol{\Theta}|\mathbf{X}}[h|\varnothing] = 1$ by convention. For any integer m, $\zeta_i \in \mathcal{U}(\mathcal{X})$, $i = 1, \dots, m$, and h such that $\|h\| < \infty$, $G^{(m)}_{\mathbf{Y}|\mathbf{X}}[h; \zeta_1, \dots, \zeta_m|\mathbf{x}]$ exists. Let

$$n^{[r]} = \begin{cases} n(n-1)\cdots(n-r+1) & r = 0, \dots, n, \\ 0 & r > n. \end{cases} \tag{A.2}$$

Lemma A.1. *Consider the p.g.fl.* $G_{\mathbf{Y}|\mathbf{X}}[h|\mathbf{x}]$ *defined in (A.1). If* $\int P_{\mathbf{X}}(d\mathbf{x})|\mathbf{x}|^{[m]} < \infty$ *, where* $|\mathbf{x}|$ *denotes the dimension of* $\mathbf{x}$*, then for any* $\zeta_i \in \mathcal{U}(\mathcal{Y})$, $i = 1, \dots, m$, $h \in \mathcal{U}(\mathcal{Y})$ *with* $\|h\| < 1$ *and* $g \in \mathcal{U}(\mathcal{X})$ *with* $\|g\| \le 1$,

$$G^{(0,m)}_{\mathbf{X},\mathbf{Y}}[g, h; \zeta_1, \dots, \zeta_m] = \int P_{\mathbf{X}}(d\mathbf{x}) G^{(m)}_{\mathbf{Y}|\mathbf{X}}[h; \zeta_1, \dots, \zeta_m|\mathbf{x}]\Pi_{\mathbf{x}}[g].$$

Remark A.2. *In the expression (A.1) for* $G_{\mathbf{Y}|\mathbf{X}}[h|\mathbf{x}]$*, a* Poisson *clutter process* $\mathbf{K}$ *was included. We remark that the lemma still holds true also for non-Poisson clutter as long as* $G_{\mathbf{K}}$ *is k-times differentiable. This is indeed satisfied for a Poisson* $\mathbf{K}$ *and in fact, for a Gauss-Poisson clutter process too. (The latter being important for Proposition 4.5 below).*

Proof. [Proof *of Proposition 4.5*] From the definition of $G_{\mathbf{W},\mathbf{Z},\mathbf{Y}}[g, f, h]$ in (4.13), it follows that

$$G_{\mathbf{Z},\mathbf{Y}}[f, h] = G_{\mathbf{W},\mathbf{Z},\mathbf{Y}}[g, f, h]|_{g=1} = G_{\mathbf{K}}[h]\, G_{\mathbf{W}}[\Lambda(h)]\, G_{\mathbf{Z}}[f(x_1, x_2)\Lambda(h)(x_1)\Lambda(h)(x_2)].$$

The term $G_{\mathbf{K}}[h]G_{\mathbf{W}}[\Lambda(h)]$ corresponds to the p.g.fl. of a Poisson PP on $\mathcal{Y}$ with intensity measure

$$V_{\mathbf{K}}(A) + V_{\mathbf{W}}[L(A|.)].$$

We outline the three main steps to derive the conditional functional $G_{\mathbf{Z}|\mathbf{y}}[f]$.

Step 1: The first step in the proof is to derive the explicit formula for $G^{(0,m)}_{\mathbf{Z},\mathbf{Y}}[f, h; \zeta_1, \dots, \zeta_m]$. Fixing f and viewing $G_{\mathbf{Z},\mathbf{Y}}[f, h]$ as a function of h, we note that $G_{\mathbf{Z},\mathbf{Y}}[f, h]$ is the p.g.fl. of a GP PP on $\mathcal{Y}$. As such, the formula of Newman [16] for the derivative of the p.g.fl. of a GP PP may used. Consider (4.14) and (4.15) for a fixed f and $g = 1$. We redefine the left hand side of (4.14) and (4.15) as

$$\omega(y_i, y_j) = \int V_{\mathbf{Z}}(dx_1, dx_2) l(y_i|x_1) l(y_j|x_2) f(x_1, x_2) + \int V_{\mathbf{Z}}(dx_1, dx_2) l(y_j|x_1) l(y_i|x_2) f(x_1, x_2), \tag{A.3}$$

$$\Omega(y_i) = \int V_{\mathbf{W}}(dx) g(x) l(y_i|x) + \int \widetilde{V}_{\mathbf{K}}(dx) l(y_i|x) + \int V_{\mathbf{Z}}(dx_1, dx_2) \overline{L}(\mathcal{Y}|x_1) l(y_i|x_2) f(x_1, x_2) + \int V_{\mathbf{Z}}(dx_1, dx_2) l(y_i|x_1) \overline{L}(\mathcal{Y}|x_2) f(x_1, x_2). \tag{A.4}$$

to make explicit the dependence on y_i, y_j. (Note that f and g are suppressed from the notation since they are fixed. Also, the dependence of (4.14) and (4.15) on y_i, y_j was not made explicit therein because all formulae for the updated moments are for a fixed realisation of $\mathbf{Y}$.) We now have [16]

$$\begin{aligned} &G^{(0,m)}_{\mathbf{Z},\mathbf{Y}}[f, h; \zeta_1, \ldots, \zeta_m] \\ &= G_{\mathbf{Z},\mathbf{Y}}[f, h] \sum_{j=0}^{\lfloor m/2 \rfloor} \sum_{\sigma\{j:m\}} \prod_{i=1}^{j} \int \lambda_{\mathcal{Y}}(dy_1) \lambda_{\mathcal{Y}}(dy_2) \omega(y_1, y_2) \zeta_{\sigma_{2i-1}}(y_1) \zeta_{\sigma_{2i}}(y_2) \\ &\quad \times \prod_{i'=2j+1}^{m} \int \lambda_{\mathcal{Y}}(dy) \Omega(y) \zeta_{\sigma_{i'}}(y) \\ &= G_{\mathbf{Z},\mathbf{Y}}[f, h] \int \lambda^{(m)}_{\mathcal{Y}}(dy_1, \ldots, dy_m) \zeta_1(y_1) \zeta_2(y_2) \cdots \zeta_m(y_m) \\ &\quad \times \left(\sum_{j=0}^{\lfloor m/2 \rfloor} \sum_{\sigma\{j:m\}} \prod_{i=1}^{j} \omega(y_{\sigma_{2i-1}}, y_{\sigma_{2i}}) \prod_{i'=2j+1}^{m} \Omega(y_{\sigma_{i'}}) \right) \end{aligned} \tag{A.5}$$

where the last equality follows from the penultimate by inspection.

Step 2: The second step is to prove the validity of the interchange of integration and differentiation, which Lemma A.1 asserted for the case when the hidden process is Poisson on $\mathcal{X}$. In the present setting, $\mathbf{Z}$ is a Poisson PP on $\mathcal{X} \times \mathcal{X}$. We assert that the interchange is still valid and the proof of its validity proceeds along the same lines as that of Lemma A.1. The key condition allowing this interchange is that for any k,

$$\int P_{\mathbf{Z}}(d\mathbf{z}) \, (2|\mathbf{z}|)^{[k]} < \infty.$$

This integral is indeed bounded since for any λ,

$$\begin{aligned} &\sum_{n \geq k/2} \frac{\lambda^n}{n!} \, (2n)^{[k]} \\ &= \sum_{k > n \geq k/2} \frac{\lambda^n}{n!} \, (2n)^{[k]} + \sum_{n \geq k} \frac{\lambda^n}{n!} \frac{(2n)!}{(2n-k)!} \\ &= \sum_{k > n \geq k/2} \frac{\lambda^n}{n!} \, (2n)^{[k]} + \lambda^k \sum_{n \geq k} \lambda^{n-k} \frac{2n}{n} \frac{2n-1}{n-1} \cdots \frac{2n-k+1}{n-k+1} \frac{1}{(n-k)!}. \end{aligned}$$

The second sum of the last line is bounded since

$$\sup_n \left| \frac{2n}{n} \frac{2n-1}{n-1} \cdots \frac{2n-k+1}{n-k+1} \right| < \infty.$$

Step 3: Since the interchange of differentiation and integration is valid we have

$$G^{(0,m)}_{\mathbf{Z},\mathbf{Y}}[f, 0; \zeta_1, \ldots, \zeta_m] = \int P_{\mathbf{Z}}(d\mathbf{z}) J^{(m)}_{\mathbf{Y}|\mathbf{Z}}[\zeta_1, \ldots, \zeta_m|\mathbf{z}] f(\mathbf{z}).$$

Since $J^{(m)}_{\mathbf{Y}|\mathbf{Z}}$ admits a density by assumption we may write

$$\begin{aligned} &G^{(0,m)}_{\mathbf{Z},\mathbf{Y}}[f, 0; \zeta_1, \ldots, \zeta_m] \\ &= \int P_{\mathbf{Z}}(d\mathbf{z}) \int_{\mathcal{Y}^m} \lambda^{(m)}_{\mathcal{Y}}(dy_1, \ldots, dy_m)\, p^{(m)}_{\mathbf{Y}|\mathbf{Z}}(y_1, \ldots, y_m|\mathbf{z}) \zeta_1(y_1) \cdots \zeta_m(y_m) f(\mathbf{z}) \\ &= \int_{\mathcal{Y}^m} \lambda^{(m)}_{\mathcal{Y}}(dy_1, \ldots, dy_m)\, \zeta_1(y_1) \cdots \zeta_m(y_m) \int P_{\mathbf{Z}}(d\mathbf{z}) p^{(m)}_{\mathbf{Y}|\mathbf{X}}(y_1, \ldots, y_m|\mathbf{z}) f(\mathbf{z}) \quad (A.6) \end{aligned}$$

where the interchange of the order of integration in the last line follows from Fubini's theorem. Comparing (A.5) and (A.6), the proof is complete for the numerator of $G_{\mathbf{Z}|\mathbf{y}}[f]$ (denominator in the case $f = 0$) if we confirm that

$$\left(\sum_{j=0}^{\lfloor m/2 \rfloor} \sum_{\sigma\{j:m\}} \prod_{i=1}^{j} \omega(y_{\sigma_{2i-1}}, y_{\sigma_{2i}}) \prod_{i'=2j+1}^{m} \Omega(y_{\sigma_{i'}}) \right) G_{\mathbf{Z},\mathbf{Y}}[f, 0] \tag{A.7}$$

is indeed a valid version of

$$\int P_{\mathbf{Z}}(d\mathbf{z}) p^{(m)}_{\mathbf{Y}|\mathbf{X}}(y_1, \ldots, y_m|\mathbf{z}) f(\mathbf{z}) \tag{A.8}$$

in the sense that they agree $\lambda^{(m)}_{\mathcal{Y}}$ almost everywhere. Note that (A.7) and (A.8) are both integrable w.r.to $\lambda^{(m)}_{\mathcal{Y}}$ and therefore define measures on $\mathcal{B}(\mathcal{Y}^m)$. (Hence the need for real-valued non-negative f.) Let $A_i \in \mathcal{B}(\mathcal{Y})$, $i = 1, \ldots, m$, and set $\zeta_i = \mathbb{I}_{A_i}$. We see that these measures agree on the measurable rectangles of $\mathcal{Y}^m$. Since the measurable rectangles form a π-system, the measures agree on all the measurable sets of $\mathcal{Y}^m$ and the desired result follows.

The proof for $G_{\mathbf{W}|\mathbf{y}}[g]$ commences with the joint functional

$$\begin{aligned} G_{\mathbf{W},\mathbf{Y}}[g, h] &= G_{\mathbf{W},\mathbf{Z},\mathbf{Y}}[g, f, h]\big|_{f=1} \\ &= G_{\mathbf{K}}[h] G_{\mathbf{Z}}[\Lambda(h)(x_1)\Lambda(h)(x_2)] G_{\mathbf{W}}[g\Lambda(h)] \end{aligned}$$

and proceeds in three steps exactly as that for $G_{\mathbf{Z}|\mathbf{y}}[f]$ just outlined. The term $G_{\mathbf{K}}[h] G_{\mathbf{Z}}[\Lambda(h)(x_1)\Lambda(h)(x_2)]$ can be viewed as the p.g.fl. of a Gauss Poisson clutter process. Thus the hidden process is Poisson on $\mathcal{X}$ while the clutter is GP. For the first step, Lemma A.1 applies straightaway; see comments immediately following Lemma A.1. The remaining steps proceed as above. □

Let $V_{\mathbf{X}}$ be the intensity of a PP $\mathbf{X}$. The following result, proved in [13], states that the Poisson PP with intensity $V_{\mathbf{X}}$ is the best Poisson approximation to $P_{\mathbf{X}}$ in the Kullback-Leibler (KL) sense. The result is restated below and proof included for completeness as it was required for Proposition 4.4.

LEMMA A.3. *The solution to*

$$\arg_P \min_{Poisson} \int dP_{\mathbf{X}} \log \frac{dP_{\mathbf{X}}}{dP}$$

satisfies $V^* = V_{\mathbf{X}}$ *where* V^* *is the intensity of the minimising* P^* *and* $V_{\mathbf{X}}$ *is the intensity of* $P_{\mathbf{X}}$.

Proof. Let λ be a finite diffuse measure on $\mathcal{X}$ and the reference measure on $(\mathcal{X}^{\cup}, \sigma(\mathcal{X}^{\cup}))$, P_{ref}, be the distribution of the unit rate Poisson PP with intensity measure $V_{\text{ref}} = \lambda(\cdot)/\lambda(\mathcal{X})$. Since $P_{\mathbf{X}} \ll P \ll P_{\text{ref}}$, using the standard result concerning densities we have,
$\frac{dP_{\mathbf{X}}}{dP} = \frac{dP_{\mathbf{X}}}{dP_{\text{ref}}} \left(\frac{dP}{dP_{\text{ref}}}\right)^{-1}$, and it is equivalent to solve

$$\arg_P \max_{\text{Poisson}} \int dP_{\mathbf{X}} \log \frac{dP}{dP_{\text{ref}}}$$

Expanding the integral gives (while ignoring constant terms)

$$\begin{aligned}
&\sum_{n=0}^{\infty} \int_{\mathcal{X}^{(n)}} P_{\mathbf{X}}^{(n)}(dx_1, \ldots, dx_n) \log \left[e^{-V(\mathcal{X})} \frac{V(\mathcal{X})^n}{n!} \frac{v(x_1)}{V(\mathcal{X})} \cdots \frac{v(x_n)}{V(\mathcal{X})} \right] \\
&= \sum_{n=0}^{\infty} \int_{\mathcal{X}^{(n)}} P_{\mathbf{X}}^{(n)}(dx_1, \ldots, dx_n) \sum_{i=1}^{n} \log \frac{v(x_i)}{V(\mathcal{X})} + \sum_{n=0}^{\infty} P_{\mathbf{X}}^{(n)}(\mathcal{X}^n) \log \left[\frac{e^{-V(\mathcal{X})}}{n!} V(\mathcal{X})^n \right] \\
&= \int_{\mathcal{X}} V_{\mathbf{X}}(dx) \log \frac{v(x)}{V(\mathcal{X})} + \sum_{n=0}^{\infty} P_{\mathbf{X}}^{(n)}(\mathcal{X}^n) \log \left[\frac{e^{-V(\mathcal{X})}}{n!} V(\mathcal{X})^n \right] \\
&\leq \int_{\mathcal{X}} V_{\mathbf{X}}(dx) \log \frac{v_{\mathbf{X}}(x)}{V_{\mathbf{X}}(\mathcal{X})} + \sum_{n=0}^{\infty} P_{\mathbf{X}}^{(n)}(\mathcal{X}^n) \log \left[\frac{e^{-V_{\mathbf{X}}(\mathcal{X})}}{n!} V_{\mathbf{X}}(\mathcal{X})^n \right]
\end{aligned}$$

where $v_{\mathbf{X}}$ is the density of $V_{\mathbf{X}}$ w.r.to λ. The first integral of the third line follows from Campbell's theorem [18]. The first term of the upper bound in the last line follows from the non-negativity of the KL criterion while the second term follows from the fact that $\arg\max_\rho \sum_n p_n \log(e^{-\rho}\frac{\rho^n}{n!}) = \sum_n n p_n$ where $\{p_n\}$ is a probability on the non-negative integers. □

Acknowledgments. S.S. Singh would like to thank Daniel Clark for carrying out the differentiation of the p.g.fl.s $G_{\mathbf{W}|\mathbf{y}}[g]$ and $G_{\mathbf{Z}|\mathbf{y}}[f]$ to get the $o(\beta)$ update formula. The works of the 2nd and 3rd authors are support by Discovery Grant DP0878158 of the Australian Research Council. S. Zuyev acknowledges hospitality of the University of Western Australia and the University of Melbourne.

REFERENCES

[1] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, New Jersey, 1979.
[2] D. R. Brillinger, *A note on a representation for the Gauss-Poisson process*, Stoch. Proc. Applications, 6 (1978), pp. 135–137.
[3] O. Cappe, T. Ryden, and E. Moulines, *Inference in Hidden Markov Models*, Springer, New York, 2005.
[4] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, Springer, New York, 1988.
[5] A. Doucet, N. de Freitas, and N. Gordon, eds., *Sequential Monte Carlo Methods in Practise*, Springer, New York, 2001.

[6] K. Gilholm and D. Salmond, *Spatial distribution model for tracking extended objects*, IEE Proc. Radar, Sonar and Navigation, 152 (2005), pp. 364– 371.
[7] R. M. I. Goodman and H. Nguyen, *Mathematics of Data Fusion*, Kluwer Academic Publishers, 1997.
[8] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic, New York, 1970.
[9] J. F. C. Kingman, *Poisson Processes*, Clarendon Press, Oxford, 1993.
[10] G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series*, Springer-Verlag, New York, 1996.
[11] J. Lund and M. Rudemo, *Models for point processes observed with noise*, Biometrika, 87 (2000), pp. 235–249.
[12] J. Lund and E. Thonnes, *Perfect simulation and inference for point processes given noisy observations*, Comput. Stat., 19 (2004), pp. 317–336.
[13] R. Mahler, *Multi-target Bayes filtering via first-order multi-target moments*, IEEE Transactions of Aerospace and Electronic Systems, 39 (2003), pp. 1152–1178.
[14] S. Mori, C. Y. Chong, E. Tse, and R. Wishner, *Tracking and identifying multiple targets without apriori identifications*, IEEE Trans. Automatic Control, AC-21 (1986), pp. 401–409.
[15] J. E. Moyal, *The general theory of stochastic population processes*, Acta Mathematica, 108 (1962), pp. 1–31.
[16] D. S. Newman, *A new family of point processes which are characterized by their second moment properties*, J. Appl. Probab., 7 (1970), pp. 338–358.
[17] N. Portenko, H. Salehi, and A. Skorokhod, *On optimal filtering of multitarget systems based on point process observations*, Random Operators and Stochastic Equations, 5 (1997), pp. 1–34.
[18] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and its applications*, Wiley, 2nd ed. ed., 1995.
[19] B.-N. Vo and W.-K. Ma, *The gaussian mixture probability hypothesis density filter*, IEEE Trans. Signal Processing, 54 (2005), pp. 4091–4104.
[20] B.-N. Vo, S. Singh, and A. Doucet, *Sequential monte carlo implementation of the phd filter for multi-target tracking*, in Proc. Int'l Conf. on Information Fusion, Cairns, Australia, 2003, pp. 792–799.
[21] ———, *Sequential monte carlo methods for multi-target filtering with random finite sets*, IEEE Trans. Aerospace and Electronic Systems, 41 (2005), pp. 1224–1245.
[22] R. Washburn, *A random point process approach to multi-object tracking*, in Proc. American Control Conf., vol. 3, 1987, pp. 1846–1852.
[23] M. West and J. Harrison, *Bayesian forecasting and dynamic models*, Springer Series in Statistics, Springer-Verlag, 2nd ed., 1997.