ORIGINAL ARTICLE

State estimation for aoristic models

Maria N. M. van Lieshout[1,2] | **Robin L. Markwitz[1,2]**

[1]Stochastics Research Group, Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands

[2]Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

Correspondence
Robin L. Markwitz, Centrum Wiskunde & Informatica (CWI), P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.
Email: Robin.Markwitz@cwi.nl

Abstract

Aoristic data can be described by a marked point process in time in which the points cannot be observed directly but are known to lie in observable intervals, the marks. We consider Bayesian state estimation for the latent points when the marks are modeled in terms of an alternating renewal process in equilibrium and the prior is a Markov point process. We derive the posterior distribution, estimate its parameters and present some examples that illustrate the influence of the prior distribution. The model is then used to estimate times of occurrence of interval censored crimes.

KEYWORDS

alternating renewal process, aoristic data, criminological data, marked temporal point process, Markov chain Monte Carlo methods, Markov point process state estimation

1 | INTRODUCTION

Inference for point processes on the real line has been dominated by a dynamic approach based on the stochastic intensity or hazard function (Brémaud, 1972; Karr, 1991; Last & Brandt, 1995). Such an approach is quite natural, is amenable to likelihood-based inference and allows the utilization of powerful tools from martingale theory. However, it breaks down completely when censoring breaks the orderly progression of time. In such cases, state estimation techniques are needed that are able to fill in the gaps (Brix & Diggle, 2001; Lieshout, 2016).

In this article, we concentrate on aoristic data (Ratcliffe & McCullagh, 1998) in which the points may not be observed directly but upper and lower bounds exist. Such data are commonplace in criminology. Suppose, for example, that a working person leaves his place of residence

early in the morning and returns late in the afternoon to discover that the residence has been burgled. Then the exact time of the break-in cannot be determined, but it must have happened during the absence of the resident. In certain cases, a burglar may also be caught in the act, in which case the time of break-in coincides with the bounds. The actual break-in times tend to be estimated by ad hoc, naive approaches, for example, the mid-point of the reported interval (Helms, 2008) or the length-weighted empirical probability mass function of the interval lengths. An obvious disadvantage of such methods is that dependencies between offence times, such as the near-repeat effect (Bernasco, 2009), are ignored.

The focus of this research is to develop a Bayesian inference framework for aoristic data that is able to infer missing information and takes into account expert knowledge and interaction. Specifically, in Section 2 we formalise aoristic censoring as a marked point process in time in which the points cannot be observed directly but are known to lie in observable intervals, the marks. Upon employing a Markov point process prior (Lieshout, 2000), the posterior distribution of the point locations is derived in Section 3. In Section 4 we turn to Monte Carlo-based inference. The article is concluded by simulated examples that demonstrate the influence of the prior (Section 5), an application of the model to a criminological dataset (Section 6), and some reflections on future research.

2 | MARKED POINT PROCESS FORMULATION

We are interested in developing statistical inference methods for aoristic crime data. To do this, we require models that describe the underlying behavior of the victims as well as the times at which crimes occur. We assume that residents leave and return to their properties independently of each other. To model this coming and going behavior, we will be using a stochastic process to generate time intervals within which residents alternate between being home and away from their residencies. Since burglaries can occur while the resident is home as well as when they are away, we have chosen an alternating renewal process to model these two distinct states. The sojourn times follow typical length distributions, such as the Gamma and Weibull. It is assumed that every resident has their own alternating renewal process, and burglars will very occasionally turn up at the victim's place of residence to commit property crimes. Time intervals in which burglaries occur are then noted and form part of the dataset, if the resident is not home or the exact time of burglary cannot be determined. Otherwise, the exact burglary time is recorded.

Note that the underlying behavior of the burglar is not modeled directly. Instead, we draw relationships between possible burglary times in a given neighborhood or city. Given n reported intervals, we can be sure that n burglaries occurred. However, we do not know when exactly within a given interval the crime occurred. Using Bayesian methods, we can impose a certain structure on the burglary times that may correspond to expert knowledge or wisdom. For example, one might have a good reason to believe that in a given region, criminals often burgle at very similar times, when they know that victims will not be home. Additionally, the near-repeat effect, which also would assume a dependence between burglary times, can be accounted for (Bernasco (2009)). By modeling these potential burglary times using a point process prior, this information can be embedded in the complete model. This complete model is a mixture of two classes: the proposed or tentative burglary times along with their corresponding intervals, as well as fully observed atoms when there is no uncertainty relating to the time of occurrence. In the following subsections, we introduce the mathematical concepts required to build this model.
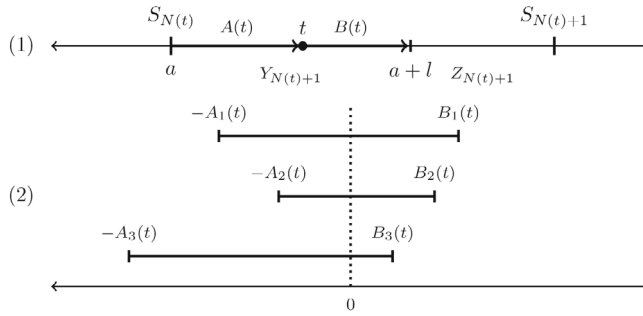
FIGURE 1 Diagram (1) shows an alternating renewal process where t falls in a Y-phase. The age $A(t)$ and excess $B(t)$ distributions describe the elapsed and remaining time, a and $a + l$ the two ends of the interval. Diagram (2) shows the parameterization of a number of intervals after shifting to the origin.

2.1 | Alternating renewal processes for censoring

In this article, we consider a censoring mechanism based on an alternating renewal process. Let $C_1, C_2, \ldots$ be a sequence of random 2-vectors such that $C_i = (Y_i, Z_i)$, $i \in \mathbb{N}$, are independent and identically distributed (Asmussen, 2003; Ross, 1996). Furthermore, assume that C_i has joint probability density function f on $(\mathbb{R}^+)^2$. Introduce $T_i = Y_i + Z_i$, set $S_0 = 0$ and let, for $n \in \mathbb{N}$, $S_n = \sum_{i=1}^{n} T_i$ be the time of the nth renewal. Note that no renewal occurs at the end of a Y-phase, and that Y_i occurs before the corresponding Z_i. Furthermore, assume that $0 < \mathbb{E}T_1 < \infty$. Then, by the strong law of large numbers,

$$N(t) = \sup \{ n \in \mathbb{N}_0 : S_n \le t \}, t \ge 0, \tag{1}$$

is well-defined and the supremum is attained with probability one, where $\mathbb{N}_0$ denotes $\mathbb{N} \cup \{0\}$. Furthermore, the renewal function

$$M(t) = \mathbb{E}N(t) = \sum_{n=1}^{\infty} \mathbb{P}(S_n \le t), t \ge 0,$$

is finite and absolutely continuous with respect to Lebesgue measure (Ross, 1996, chapter 3).

An alternating renewal process can be used for censoring in the following way. Consider a point $t \in \mathbb{R}^+$. If t happens to fall in some Z-phase, t is observed perfectly, whereas t is observed aoristically if it falls in a Y-phase, that is, only the beginning and end of the phase are recorded. The censoring mechanism is illustrated in Figure 1. In terms of the criminology context that motivates this work, Y_i and Z_i are the away and home phases of a resident, respectively. The point t is the time of a break-in.

2.2 | Age and excess distribution

Aoristic data generated by the censoring mechanism described in Section 2.1 can be expressed formally in terms of the age and excess (also referred to as residual lifetime) with respect to the Y-process. Recall that for an alternating renewal process, the age with respect to the Y-process is

defined as

$$A(t) = (t - S_{N(t)})\mathbf{1}\{S_{N(t)} + Y_{N(t)+1} > t\}, t \geq 0,$$

the excess with respect to the Y-process as

$$B(t) = (S_{N(t)+1} - Z_{N(t)+1} - t)\mathbf{1}\{S_{N(t)} + Y_{N(t)+1} > t\}, t \geq 0.$$

Let the pair $(a, l) \in \mathbb{R} \times \mathbb{R}^+$ correspond to the closed interval $[a, a + l]$. In this way, an interval on the real line is parameterized by its left-most point and length. A recorded interval $[a, a + l]$ in which a given latent point $t \geq 0$ falls is then written as $t + [-A(t), B(t)]$, seeing that $a = t - A(t)$ and $a + l = t + B(t)$. The closed interval $[-A(t), B(t)]$ is called the mark attached to t and corresponds to the two-dimensional parameter vector $I(t) = (-A(t), A(t) + B(t)) \in \mathbb{R} \times \mathbb{R}^+$. Note that the mark always contains the origin 0. For a visualization see Figure 1.

We would like to derive the form and nature of the mark distribution. To achieve this, the joint distribution of age and excess with respect to the Y-process is first considered.

Proposition 1. *Let N be an alternating renewal process as in (1). Assume that T_1 is absolutely continuous with respect to Lebesgue measure and that $0 < \mathbb{E}T_1 < \infty$. Then, for $t \geq 0$, the joint distribution of $(A(t), B(t))$ has an atom at $(0, 0)$ of size*

$$c(t) = F_Y(t) - \int_0^t [1 - F_Y(t - s)] \, dM(s),$$

and, for $0 \leq u \leq t, v \geq 0$,

$$\mathbb{P}(A(t) \leq u; B(t) \leq v) = c(t) + [F_Y(t + v) - F_Y(t)] \mathbf{1}\{u = t\}$$
$$+ \int_{t-u}^t [F_Y(t + v - s) - F_Y(t - s)] \, dM(s).$$

Here F_Y denotes the cumulative distribution function of Y_1 and M is the renewal function.

Proof. Write F_n for the cumulative distribution function of S_n, $n \in \mathbb{N}$. By partitioning over the number of renewals up to time t and upon noting that $N(t) = n$ if and only if $S_n \leq t$ and $S_n + Y_{n+1} + Z_{n+1} > t$, one obtains that

$$\mathbb{P}(A(t) \leq u; B(t) \leq v) = c(t) + \mathbb{P}(t - S_{N(t)} \leq u; t < S_{N(t)} + Y_{N(t)+1} \leq t + v)$$

$$= c(t) + \sum_{n=0}^\infty \mathbb{P}(t - u \leq S_{N(t)}; t < S_{N(t)} + Y_{N(t)+1} \leq t + v; N(t) = n)$$

$$= c(t) + \mathbb{P}(t - u \leq S_{N(t)}; t < S_{N(t)} + Y_{N(t)+1} \leq t + v; N(t) = 0)$$

$$+ \sum_{n=1}^\infty \int_{t-u}^t \mathbb{P}(t - s < Y_{n+1} \leq t + v - s) dF_n(s).$$

The claim follows by an application of Fubini's theorem for the last term, the observation that

$$\mathbb{P}(t - u \leq S_{N(t)}; t < S_{N(t)} + Y_{N(t)+1} \leq t + v; N(t) = 0) = \mathbb{P}(t < Y_1 \leq t + v)$$

if $u = t$ and zero otherwise, and because

$$c(t) = 1 - \mathbb{P}(S_{N(t)} + Y_{N(t)+1} > t)$$

$$= 1 - \mathbb{P}(Y_1 > t) - \sum_{n=1}^{\infty} \int_0^t \mathbb{P}(Y_{n+1} > t - s) dF_n(s) = F_Y(t) - \int_0^t [1 - F_Y(t - s)] \, dM(s).$$

∎

The long-run behavior as time goes to infinity can be obtained by appealing to the key renewal theorem. Recall that this theorem applies to functions $h : \mathbb{R}^+ \to \mathbb{R}$ such that $h(t) \geq 0$ for all $t \geq 0$, h is monotonically nonincreasing and integrable. For such functions, it states that

$$\lim_{t \to \infty} \int_0^t h(t - s) dM(s) = \frac{1}{\mathbb{E}[T_1]} \int_0^\infty h(s) ds$$

(Ross, 1996, theorem 3.4.2).

Specialising to the parameter vector $I(t)$, the following theorem holds.

Theorem 1. *Let N be an alternating renewal process as in (1). Assume that Y_1 and T_1 are absolutely continuous with respect to Lebesgue measure with probability density functions f_Y and f and that $0 < \mathbb{E}T_1 < \infty$. As $t \to \infty$, the joint probability distribution of $(-A(t), A(t) + B(t))$ tends in distribution to ν, the mixture of an atom at $(0, 0)$ and an absolutely continuous component that has probability density function $f_Y(l)/\mathbb{E}Y_1$ on $\{(a, l) \in \mathbb{R} \times \mathbb{R}^+ : a \leq 0 \leq a + l\}$. The mixture weights are, respectively, $\mathbb{E}Z_1/\mathbb{E}T_1$ and $\mathbb{E}Y_1/\mathbb{E}T_1$.*

Proof. First, let us consider the limit behavior of the joint cumulative distribution function of $A(t)$ and $B(t)$ as $t \to \infty$. With the notation of Proposition 1, by theorem 3.4.4 in Ross (1996) or the key renewal theorem applied to $1 - F_Y$, $c(t)$ converges to $\mathbb{E}Z_1/\mathbb{E}T_1$. Also, for $t > u$, the second term in the joint cumulative distribution function of $A(t)$ and $B(t)$ is zero. For the last term, note that for $v \geq 0$ the function $h_v : \mathbb{R}^+ \to \mathbb{R}$ defined by $h_v(s) = 1 - F_Y(v + s)$ is nonnegative, monotonically nonincreasing and integrable. Hence the key renewal theorem implies that

$$\lim_{t \to \infty} \int_0^t [1 - F_Y(t + v - s)] \, dM(s) = \frac{1}{\mathbb{E}T_1} \int_0^\infty [1 - F_Y(v + s)] \, ds.$$

Analogously, for fixed $u \geq 0$, $t - u \to \infty$ if and only if $t \to \infty$. Writing $s = t - u$,

$$\lim_{s \to \infty} \int_0^s [1 - F_Y(s + u + v - r)] \, dM(r) = \frac{1}{\mathbb{E}T_1} \int_0^\infty [1 - F_Y(u + v + r)] \, dr$$

$$= \lim_{t \to \infty} \int_0^{t-u} [1 - F_Y(t + v - r)] \, dM(r).$$

We conclude that $Q(u, v) = \lim_{t \to \infty} \mathbb{P}(A(t) \leq u; B(t) \leq v)$ exists and equals

$$Q(u, v) = \frac{\mathbb{E}Z_1}{\mathbb{E}T_1} + \frac{1}{\mathbb{E}T_1} \int_0^u [F_Y(v + s) - F_Y(s)] \, ds.$$

Note that Q is a cumulative distribution function, corresponding to the mixture of an atom at $(0, 0)$ and an absolutely continuous component with probability density function $f_Y(u + v)/\mathbb{E}Y_1$ on $(\mathbb{R}^+)^2$. By Helly's continuity theorem, $(A(t), B(t))$ converges in distribution.

Turning to the parameterization $I(t) = (-A(t), A(t) + B(t))$, its limit distribution inherits an atom at $(0, 0)$ from Q of size $\mathbb{E}Z_1/\mathbb{E}T_1$. By the change of variables bijection $h : (\mathbb{R}^+)^2 \to \mathbb{R}^- \times \mathbb{R}^+$ defined by $h(u, v) = (-u, u + v)$, since h is differentiable, the absolutely continuous part has probability density function $f_Y(h^{-1}(a, l))|\det J_{h^{-1}}(a, l)|/\mathbb{E}Y_1 = f_Y(-a + a + l)/\mathbb{E}Y_1$, where $J_{h^{-1}}$ is the Jacobian of h^{-1}. ∎

2.3 | Complete model formulation

We are now ready to formulate a model. Let $\mathcal{X}$ be an open subset of the positive half-line $\mathbb{R}^+$. The state space of X, denoted by $\mathcal{N}_{\mathcal{X}}$, consists of finite sets $\{t_1, \ldots, t_n\} \subset \mathcal{X}$, $n \in \mathbb{N}_0$, which we equip with the Borel σ-algebra of the weak topology (Daley & Vere-Jones, 2003). Note that the $\{t_1, \ldots, t_n\}$ represent the burglary times. We will assume that the distribution of X is specified in terms of a probability density function p_X with respect to the distribution of a unit rate Poisson process on $\mathcal{X}$ (Lieshout, 2000).

Recall that we want to model two occurrences—partial and total observation. Returning to the criminological context, these correspond to knowing a time span within which a burglary occurred (t falls in a Y-phase) and knowing the exact time (t falls in a Z-phase), respectively. Using the alternating renewal process in equilibrium to simulate the behavior of the victim, the points of X are independently marked according to the mixture distribution of Theorem 1. In this way, the complete model W is obtained. Its realisations are sets $\{(t_1, I_1), \ldots, (t_n, I_n)\} \subset \mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+)$, where $I_j = (a_j, l_j)$ is the jth mark parameterization. The pair (t_j, I_j) thus defines an interval $[t_j + a_j, t_j + a_j + l_j]$. The ensemble of all realisations is denoted by $\mathcal{N}_{\mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+)}$ and equipped with the Borel σ-algebra of the weak topology. Note that W has probability density function p_X with respect to the distribution of a Poisson process on $\mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+)$ with intensity measure $\ell \times \nu$ where ℓ is Lebesgue measure.

Due to the censoring, one does not observe the complete model W but rather the set

$$U = \bigcup_{(t, I) \in W} ((t, 0) + I) \tag{2}$$

of interval parameterisations. When $I = (0, 0)$, we have a full observation of the burglary time. For any other value of I, only the corresponding interval is observed. Our aim is to reconstruct X or W from U. In order to do so, the posterior distribution of X or W given U is needed. This will be the topic of the next section.

3 | THE BAYESIAN FRAMEWORK

In a Bayesian framework, the posterior distribution updates prior forms in the light of data gathered (Gamerman & Lopes, 2006). Heuristically,

$$p_{X|U}(\mathbf{x}|\mathbf{u}) \propto p_{U|X}(\mathbf{u}|\mathbf{x}) p_X(\mathbf{x}), \tag{3}$$

through the use of Bayes' theorem. The term $p_{U|X}(\mathbf{u} \mid \mathbf{x})$ describes the likelihood that the points of $\mathbf{x}$ generate the intervals in $\mathbf{u}$. In the literature this term is referred to as a forward term, forward density or forward model (Lieshout, 1995; Lieshout & Baddeley, 2002). The term $p_X(\mathbf{x})$ captures

prior beliefs about the geometry of **x**; in our case, the burglary times. In our context, since the forward model is a mixture of discrete and absolutely continuous components, some care is required in handling (3).

Theorem 2. *Let W be a point process on the open set $\mathcal{X} \subset \mathbb{R}$ with probability density function p_X with respect to the distribution of a unit rate Poisson process on $\mathcal{X}$ marked independently with mark distribution ν defined in Theorem 1. Write X for the ground process of locations in $\mathcal{X}$ and consider the forward model (2). Let $\mathbf{u}$ be a realization of U that consists of an atomic part $\{(a_1, 0), \ldots, (a_m, 0)\}$, $m \in \mathbb{N}_0$, and a nonatomic part $\{(a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$, $n \geq m$. Then the posterior distribution of X given $U = \mathbf{u}$ satisfies, for A in the Borel σ-algebra of the weak topology on $\mathcal{N}_{\mathcal{X}}$,*

$$\mathbb{P}(X \in A | U = \mathbf{u}) = c(\mathbf{u}) \int_{\mathcal{X}^{n-m}} p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) \mathbf{1}_A(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) \times$$

$$\times \left(\sum_{\substack{D_1, \ldots, D_{n-m} \\ \cup_i \{D_i\} = \{1, \ldots, n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\} \right) \prod_{i=1}^{n-m} dx_i,$$

provided that $c(\mathbf{u})^{-1}$ defined by

$$\int_{\mathcal{X}^{n-m}} p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) \left(\sum_{\substack{D_1, \ldots, D_{n-m} \\ \cup_i \{D_i\} = \{1, \ldots, n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\} \right) \prod_{i=1}^{n-m} dx_i$$

exists in $(0, \infty)$.

Proof. We must show that for each A in the Borel σ-algebra of $\mathcal{N}_{\mathcal{X}}$ with respect to the weak topology and each F in the Borel σ-algebra of the weak topology on $\mathcal{N}_{\mathbb{R} \times \mathbb{R}^+}$ the following identity holds:

$$\mathbb{E}\left[\mathbf{1}_F(U)\mathbb{P}(X \in A | U)\right] = \mathbb{E}\left[\mathbf{1}_F(U)\mathbf{1}_A(X)\right]. \tag{4}$$

Let

$$q_x(a, l) = \frac{f_Y(l)}{\mathbb{E}Y_1} \mathbf{1}\{a \leq x \leq a + l\}$$

describe a probability density function for parameterizations of intervals generated by $x \in \mathcal{X}$, noting that it is jointly measurable as a function on $\mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+)$. Then, denoting the cardinality of a set by $|\cdot|$ and Lebesgue measure by ℓ, $\mathbb{E}[\mathbf{1}_F(U)\mathbf{1}_A(X)]$ can be expanded as

$$\sum_{n=0}^{\infty} \frac{e^{-\ell(\mathcal{X})}}{n!} \sum_{C_0 \subset \{1, \ldots, n\}} \left(\frac{\mathbb{E}Z_1}{\mathbb{E}T_1}\right)^{|C_0|} \left(\frac{\mathbb{E}Y_1}{\mathbb{E}T_1}\right)^{n-|C_0|} \int_{\mathcal{X}^n} \mathbf{1}_A(\{x_1, \ldots, x_n\}) p_X(\{x_1, \ldots, x_n\}) \sum_{\substack{C_1, \ldots, C_{n-|C_0|} \\ \cup_j \{C_j\} = \{1, \ldots, n\} \setminus C_0}}$$

$$\frac{1}{(n-|C_0|)!} \left(\int_{(\mathbb{R} \times \mathbb{R}^+)^{n-|C_0|}} \prod_{j=1}^{n-|C_0|} q_{x_{C_j}}(u_j) \mathbf{1}_F(\{u_1, \ldots, u_{n-|C_0|}\} \cup \{(x_k, 0) : k \in C_0\}) \prod_{j=1}^{n-|C_0|} du_j \right) \prod_{i=1}^{n} dx_i. \tag{5}$$

For the left-hand side of (4), expanding as before, we get

$$\sum_{n=0}^{\infty} \frac{e^{-\ell(\mathcal{X})}}{n!} \sum_{C_0 \subset \{1,\ldots,n\}} \left(\frac{\mathbb{E}Z_1}{\mathbb{E}T_1} \right)^{|C_0|} \left(\frac{\mathbb{E}Y_1}{\mathbb{E}T_1} \right)^{n-|C_0|} \int_{\mathcal{X}^n} p_X(\{x_1,\ldots,x_n\}) \sum_{\substack{C_1,\ldots,C_{n-|C_0|} \\ \cup_j \{C_j\}=\{1,\ldots,n\} \setminus C_0}} \frac{1}{(n-|C_0|)!}$$

$$\left(\int_{(\mathbb{R} \times \mathbb{R}^+)^{n-|C_0|}} \prod_{j=1}^{n-|C_0|} q_{x_{C_j}}(u_j) \mathbb{P}\left(X \in A | U = \{u_1,\ldots,u_{n-|C_0|}\} \cup \{(x_k,0) : k \in C_0\} \right) \times \right.$$

$$\left. \times \mathbf{1}_F(\{u_1,\ldots,u_{n-|C_0|}\} \cup \{(x_k,0) : k \in C_0\}) \prod_{j=1}^{n-|C_0|} du_j \right) \prod_{i=1}^{n} dx_i.$$

Plugging in the claimed expression for the conditional expectation, one obtains

$$\sum_{n=0}^{\infty} \frac{e^{-\ell(\mathcal{X})}}{n!} \sum_{C_0 \subset \{1,\ldots,n\}} \left(\frac{\mathbb{E}Z_1}{\mathbb{E}T_1} \right)^{|C_0|} \left(\frac{\mathbb{E}Y_1}{\mathbb{E}T_1} \right)^{n-|C_0|} \int_{\mathcal{X}^n} p_X(\{x_1,\ldots,x_n\}) \sum_{\substack{C_1,\ldots,C_{n-|C_0|} \\ \cup_j \{C_j\}=\{1,\ldots,n\} \setminus C_0}} \frac{1}{(n-|C_0|)!}$$

$$\int_{(\mathbb{R} \times \mathbb{R}^+)^{n-|C_0|}} \prod_{j=1}^{n-|C_0|} q_{x_{C_j}}(u_j) c(\{u_1,\ldots,u_{n-|C_0|}\} \cup \{(x_k,0) : k \in C_0\})$$

$$\left(\int_{\mathcal{X}^{n-|C_0|}} p_X(\{y_1,\ldots,y_{n-|C_0|}\} \cup \{x_k : k \in C_0\}) \mathbf{1}_A(\{y_1,\ldots,y_{n-|C_0|}\} \cup \{x_k : k \in C_0\}) \sum_{\substack{D_1,\ldots,D_{n-|C_0|} \\ \cup_j \{D_j\}=\{1,\ldots,n\} \setminus C_0}} \right.$$

$$\left. \prod_{k=1}^{n-|C_0|} \mathbf{1}\{y_{D_k} \in [u_{k,1}, u_{k,1}+u_{k,2}]\} dy_k \right) \mathbf{1}_F(\{u_1,\ldots,u_{n-|C_0|}\} \cup \{(x_k,0) : k \in C_0\}) \prod_{j=1}^{n-|C_0|} du_j \prod_{i=1}^{n} dx_i.$$

Note that in order to cancel terms, the order of integration must be changed. As evidently the first term in $q_x(a,l)$, that is $f_Y(l)/\mathbb{E}Y_1$, does not depend on x. By Fubini's theorem,

$$\mathbb{E}[\mathbf{1}_F(U)\mathbb{P}(X \in A|U)] = \sum_{n=0}^{\infty} \frac{e^{-\ell(\mathcal{X})}}{n!} \sum_{C_0 \subset \{1,\ldots,n\}} \left(\frac{\mathbb{E}Z_1}{\mathbb{E}T_1} \right)^{|C_0|} \left(\frac{\mathbb{E}Y_1}{\mathbb{E}T_1} \right)^{n-|C_0|}$$

$$\int_{\mathcal{X}^n} p_X(\{y_1,\ldots,y_{n-|C_0|}\} \cup \{x_k : k \in C_0\}) \mathbf{1}_A(\{y_1,\ldots,y_{n-|C_0|}\} \cup \{x_k : k \in C_0\})$$

$$\sum_{\substack{D_1,\ldots,D_{n-|C_0|} \\ \cup_j \{D_j\}=\{1,\ldots,n\} \setminus C_0}} \frac{1}{(n-|C_0|)!} \int_{(\mathbb{R} \times \mathbb{R}^+)^{n-|C_0|}} \mathbf{1}_F(\{u_1,\ldots,u_{n-|C_0|}\} \cup \{(x_k,0) : k \in C_0\}) \prod_{j=1}^{n-|C_0|} q_{y_{D_j}}(u_j)$$

$$\left(\int_{\mathcal{X}^{n-|C_0|}} p_X(\{x_1,\ldots,x_n\}) \sum_{\substack{C_1,\ldots,C_{n-|C_0|} \\ \cup_j \{C_j\}=\{1,\ldots,n\} \setminus C_0}} \prod_{i=1}^{n-|C_0|} \mathbf{1}\{x_{C_i} \in [u_{i,1}, u_{i,1}+u_{i,2}]\} dx_{C_i} \right) \times$$

$$\times c(\{u_1,\ldots,u_{n-|C_0|}\} \cup \{(x_k,0) : k \in C_0\}) \prod_{j=1}^{n-|C_0|} du_j \prod_{k=1}^{n-|C_0|} dy_k \prod_{i \in C_0} dx_i = \mathbb{E}[\mathbf{1}_F(U)\mathbf{1}_A(X)],$$

after canceling and rearranging terms and noting that the term in between brackets cancels out against the normalization constant $c(\{u_1, \ldots, u_{n-|C_0|}\} \cup \{(x_k, 0) : k \in C_0\})$. ∎

Theorem 2 states that the posterior distribution of X given $U = \mathbf{u}$ is the union of m atoms combined with $n - m$ points that are distributed on $\mathcal{X}^{n-m}$ according to the symmetric probability density function

$$c(\mathbf{u})p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) \sum_{\substack{D_1, \ldots, D_{n-m} \\ \cup_i \{D_i\} = \{1, \ldots, n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\} \tag{6}$$

with respect to Lebesgue measure.

Corollary 1. *In the framework of Theorem 2, the conditional distribution of the mark assignments $D_1, \ldots, D_{n-m}$ for nonatomic marks is as follows. For $d_1, \ldots, d_{n-m} \in \{1, \ldots, n-m\}$ such that $\{d_1, \ldots, d_{n-m}\} = \{1, \ldots, n-m\}$,*

$$\mathbb{P}(D_1 = d_1, \ldots, D_{n-m} = d_{n-m} | X = \{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}, U = \mathbf{u})$$

$$= \frac{\prod_{i=1}^{n-m} \mathbf{1}\{x_{d_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\}}{\sum_{\substack{C_1, \ldots, C_{n-m} \\ \cup_i \{C_i\} = \{1, \ldots, n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{C_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\}},$$

provided that $x_i \in [a_{m+i}, a_{m+i} + l_{m+i}]$ for $i = 1, \ldots, n-m$ and zero otherwise.

As a special case, let us consider an inhomogeneous Poisson process with integrable intensity function $\lambda : \mathcal{X} \to \mathbb{R}^+$. Then, under the posterior distribution, X consists of n independent points, one in each interval of $\mathbf{u}$, with probability density function

$$\frac{\lambda(x)}{\int_{[a_i, a_i + l_i] \cap \mathcal{X}} \lambda(s)ds}$$

on $[a_i, a_i + l_i] \cap \mathcal{X}$ for intervals with $l_i > 0$. To see this, recall that for a Poisson process (Lieshout, 2000)

$$p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) = \exp\left[\int_{\mathcal{X}} (1 - \lambda(s))ds\right] \prod_{j=1}^{m} \lambda(a_j) \prod_{i=1}^{n-m} \lambda(x_i),$$

which factorises over terms associated with each interval. Hence (6) is proportional to

$$\sum_{\substack{D_1, \ldots, D_{n-m} \\ \cup_i \{D_i\} = \{1, \ldots, n-m\}}} \prod_{i=1}^{n-m} \lambda(x_{D_i}) \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\}.$$

4 | STATISTICAL INFERENCE

In this section we will consider statistical inference for aoristically censored data. Our main aim is to reconstruct the latent point process X from observed parameterized intervals U, that may or

may not be censored. In tandem, the censoring probability as well as the parameters η of the distribution of the nondegenerate intervals must be estimated. Parameters of the prior distribution may either be treated as fixed or subject to estimation.

4.1 | Forward model parameters

Suppose that we observe a realization $\mathbf{u} = \{(a_1, 0), \ldots, (a_m, 0), (a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$ of U, where $a_i \in \mathbb{R}$, $l_i > 0$ and $n \neq 0$. Our first aim is to estimate the parameters η of the mark distribution ν (cf. Theorem 1). The parameter vector η comprises the parameters ζ of the probability density function f_Y as well as any other parameters χ involved in the joint distribution of the random vector $C_1 = (Y_1, Z_1)$ that defines the alternating renewal process (cf. Section 2.1).

The likelihood function can be obtained from the proof of Theorem 2 by taking A equal to $\mathcal{N}_\mathcal{X}$ in Equation (5). On a logarithmic scale,

$$L(\eta; \mathbf{u}) = m \log\left(\frac{\mathbb{E}[Z_1; \zeta, \chi]}{\mathbb{E}[T_1; \zeta, \chi]}\right) + (n - m) \log\left(\frac{\mathbb{E}[Y_1; \zeta]}{\mathbb{E}[T_1; \zeta, \chi]}\right) + \sum_{i=1}^{n-m} \log\left(\frac{f_Y(l_i; \zeta)}{\mathbb{E}[Y_1; \zeta]}\right), \quad (7)$$

upon ignoring terms that do not depend on η.

Equation (7) simplifies greatly if we assume that the mixture weight $p = \mathbb{E}[Z_1; \zeta, \chi]/\mathbb{E}[T_1; \zeta, \chi]$ does not depend on ζ. This is the case, for example, when Y_1 and Z_1 are independent and Gamma distributed with the same shape parameter k and rate parameters λ for Y_1 and $\chi\lambda$ for Z_1. Then $p = p(\chi) = 1/(1 + \chi)$ does not depend on $\zeta = (k, \lambda)$ and

$$L(p, \zeta; \mathbf{u}) = m \log p + (n - m) \log(1 - p) + \sum_{i=1}^{n-m} \log\left(\frac{f_Y(l_i; \zeta)}{\mathbb{E}[Y_1; \zeta]}\right).$$

The atom probability p may be estimated by the sample estimate m/n, the fraction of atoms in the sample $\mathbf{u}$. For ζ, we need the following result.

Proposition 2. *Let ν be as in Theorem 1. Then the distribution of the lengths of nondegenerate intervals is given by the length-weighted marginal distribution $f(l) = l f_Y(l)/\mathbb{E}Y_1$ and the left-most points are, conditionally on $L = l$, uniformly distributed on $[-l, 0]$.*

Proof. Let $f(l)$ be the marginal probability density function of the length l. Evaluating,

$$f(l) = \int \frac{f_Y(l)}{\mathbb{E}Y_1} \mathbf{1}\{a \leq 0 \leq a + l\} da = \int_{-l}^{0} \frac{f_Y(l)}{\mathbb{E}Y_1} da = \frac{l f_Y(l)}{\mathbb{E}Y_1}.$$

Let $f_{A|L=l}(a)$ be the conditional probability density function for the left-most point A of an interval given its length L. Using the definition of conditional intensity for $l > 0$,

$$f_{A|L=l}(a) = \frac{f(a, l)}{f(l)} = \frac{\frac{f_Y(l)}{\mathbb{E}Y_1} \mathbf{1}\{a \leq 0 \leq a + l\}}{\frac{l f_Y(l)}{\mathbb{E}Y_1}} = \frac{1}{l} \mathbf{1}\{a \in [-l, 0]\}.$$

Thus $A \sim \text{Unif}[-l, 0]$. ∎

Assume that the mixture weights do not depend on ζ. Then, the parameters $\zeta = (k, \lambda)$ may be estimated by treating the nondegenerate intervals as an independent sample from $f(l)$ and applying the maximum likelihood method. For example, if f_Y is the probability density function of a Gamma(k, λ) distribution with shape parameter $k > 0$ and rate parameter $\lambda > 0$, $f(l)$ is the probability density of a Gamma distribution with parameters $k + 1$ and λ.

4.2 | State estimation

Since the posterior distribution of X or W given U (cf. Theorem 2) is intractable because of the normalisation constant $c(\mathbf{u})$, we will use Markov chain Monte Carlo methods (Brooks et al., 2011; Møller & Waagepetersen, 2003) for simulation. These methods construct a Markov chain in such a way that the stationary distribution of the chain is exactly the posterior distribution. Of these methods, a Metropolis–Hastings algorithm with a fixed number of points will be used. Since the transition probabilities depend on likelihood ratios, the benefit is that one can sample from unnormalized densities.

Let us return to the framework of Theorem 2. Note that sampling from the posterior distribution of X given U is cumbersome due to the presence of the permutation sum term in (6). Therefore our approach is to sample from the posterior distribution of the complete model W and project on its ground process of locations. Doing so avoids attributing points to intervals and therefore avoids the intractable sum. Moreover, as we saw in Section 2.3, W has probability density function p_X with respect to a unit rate Poisson process on $\mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+)$ with intensity measure $\ell \times \nu$. Upon observing $U = \mathbf{u}$ for $\mathbf{u} = \{(a_1, 0), \ldots, (a_m, 0), (a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$, by (6) this means that we must sample from a probability density function π on $\mathcal{X}^{n-m}$ that is proportional to $p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})$. The ordering of the points inherent in working on $\mathcal{X}^{n-m}$ represents the unique correspondence between points in X and intervals in U in the complete model. We will use the notation $\overline{\mathbf{x}}$ to indicate that we look at vectors rather than sets $\mathbf{x}$. In the special case that $n = m$, all points are observed perfectly and there is no need for any simulation.

We will use the Metropolis–Hastings algorithm (Brooks et al., 2011) when $n > m$, that is when there are density-admitting points. The state space is given by

$$\overline{E}(\mathbf{u}) = \{(x_1, \ldots, x_{n-m}) \in \mathcal{X}^{n-m} : x_i \in \mathcal{X} \cap [a_{m+i}, a_{m+i} + l_{m+i}], p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) > 0\}.$$

From now on, we shall assume that the state space is nondegenerate in the sense that

$$\int_{\overline{E}(\mathbf{u})} p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) dx_1 \ldots dx_{n-m} > 0. \tag{8}$$

Now, the Metropolis–Hastings algorithm is defined as follows. Let $q : \overline{E}(\mathbf{u}) \times \overline{E}(\mathbf{u}) \to \mathbb{R}^+$ be a Markov kernel. Iteratively, if the current state is $\overline{\mathbf{x}} \in \overline{E}(\mathbf{u})$, propose a new state $\overline{\mathbf{y}} \in \overline{E}(\mathbf{u})$ according to the probability density function $q(\overline{\mathbf{x}}, \cdot)$ and accept the proposal to move to $\overline{\mathbf{y}}$ with probability

$$\alpha(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = \begin{cases} 1 & \text{if } p_X(\{a_1, \ldots, a_m, y_1, \ldots, y_{n-m}\}) q(\overline{\mathbf{y}}, \overline{\mathbf{x}}) \\ & \geq p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) q(\overline{\mathbf{x}}, \overline{\mathbf{y}}); \\ \frac{p_X(\{a_1, \ldots, a_m, y_1, \ldots, y_{n-m}\}) q(\overline{\mathbf{y}}, \overline{\mathbf{x}})}{p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) q(\overline{\mathbf{x}}, \overline{\mathbf{y}})\}} & \text{otherwise.} \end{cases} \tag{9}$$

When the proposal is rejected, stay in the current state $\overline{\mathbf{x}}$. The choice of q depends on p_X. In our simulations in Section 5, we will use the following algorithm which is valid when the prior density function p_X is strictly positive.

Algorithm 1. Suppose that $p_X > 0$ and $n > m$. Iteratively, if the current state is $\overline{\mathbf{x}} \in \overline{E}(\mathbf{u})$,

- pick an interval $[a_{m+i}, a_{m+i} + l_{m+i}]$, $i = 1, \ldots, n - m$, uniformly at random from the nondegenerate ones;
- generate a uniformly randomly distributed point y_i on $\mathcal{X} \cap [a_{m+i}, a_{m+i} + l_{m+i}]$ and propose to update x_i to y_i;
- accept the proposal with probability

$$\alpha_i((x_1, \ldots, x_{n-m}), y_i) = \min\left(1, \frac{p_X((\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\} \setminus \{x_i\}) \cup \{y_i\})}{p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})}\right) \quad (10)$$

and otherwise stay in the current state.

A few remarks are in order. First, note that since $\mathcal{X}$ is open, its intersection with closed intervals that contain a point in $\mathcal{X}$ is also nondegenerate when $l_i > 0$. Secondly, when p_X may take the value zero, the proposal mechanism in Algorithm 1 might result in a new state that does not belong to $\overline{E}(\mathbf{u})$, even when $\overline{\mathbf{x}}$ does. Moreover, only changing one component at a time might lead to nonirreducible Markov chains. For example, if $\mathbf{u}$ contains the parameterizations of the intervals $[0, 1]$ and $[0.1, 1]$ and $p_X(\mathbf{x}) = 0$ for realisations $\mathbf{x}$ that contain components separated by a distance less than 0.55, then states such as $\overline{\mathbf{x}} = (0.3, 0.9)$ and $\overline{\mathbf{y}} = (0.9, 0.3)$ cannot be reached from one another.

In the next propositions, basic properties of the algorithm are considered. The proofs are modifications to our context of classic Metropolis–Hastings proofs found in, for example, Mengersen and Tweedie (1996), Roberts and Smith (1994) or Møller and Waagepetersen (2003, chapter 7).

We will write Y_i for subsequent states and denote by $P(\overline{\mathbf{x}}, F) = P(Y_{i+1} \in F | Y_i = \overline{\mathbf{x}})$ the transition probability from state $\overline{\mathbf{x}} \in \overline{E}(\mathbf{u})$ into Borel set $F \subset \overline{E}(\mathbf{u})$. Similarly, for $\tau = 2, 3, \ldots, P^\tau(\overline{\mathbf{x}}, F)$ denotes the τ-step transition probability.

Proposition 3. *Consider the setup of Theorem 2 with $n > m$ and assume that condition (8) is met. Then, the Metropolis–Hastings algorithm defined by Markov kernel q on $\overline{E}(\mathbf{u})$ and acceptance probabilities (9), is reversible with respect to π.*

Proof. Take $\overline{\mathbf{x}}, \overline{\mathbf{y}}$ in $\overline{E}(\mathbf{u})$ and assume that $\pi(\overline{\mathbf{y}})q(\overline{\mathbf{y}}, \overline{\mathbf{x}}) > \pi(\overline{\mathbf{x}})q(\overline{\mathbf{x}}, \overline{\mathbf{y}}) \geq 0$. Then

$$\pi(\overline{\mathbf{x}})q(\overline{\mathbf{x}}, \overline{\mathbf{y}})\alpha(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = c(\mathbf{u})p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})q(\overline{\mathbf{x}}, \overline{\mathbf{y}}) =$$

$$c(\mathbf{u})p_X(\{a_1, \ldots, a_m, y_1, \ldots, y_{n-m}\})q(\overline{\mathbf{y}}, \overline{\mathbf{x}})\frac{p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})q(\overline{\mathbf{x}}, \overline{\mathbf{y}})}{p_X(\{a_1, \ldots, a_m, y_1, \ldots, y_{n-m}\})q(\overline{\mathbf{y}}, \overline{\mathbf{x}})} = \pi(\overline{\mathbf{y}})q(\overline{\mathbf{y}}, \overline{\mathbf{x}})\alpha(\overline{\mathbf{y}}, \overline{\mathbf{x}}),$$

writing $c(\mathbf{u})$ for the normalization constant. We conclude that the chain is in detailed balance and therefore reversible with respect to π. ∎

Recall that the Markov chain is called π-irreducible (Meyn & Tweedie, 2009) if for every $\overline{\mathbf{x}} \in \overline{E}(\mathbf{u})$ and every Borel set $F \subset \overline{E}(\mathbf{u})$ with $\pi(F) > 0$ there exists some natural number τ such that $P^\tau(\overline{\mathbf{x}}, F) > 0$.

Proposition 4. *Consider the setup of Theorem 2 with $n > m$ and assume that condition (8) is met. Let Q be the one-step transition kernel of the Markov chain on $\overline{E}(\mathbf{u})$ generated by Markov kernel $q : \overline{E}(\mathbf{u}) \times \overline{E}(\mathbf{u}) \to \mathbb{R}^+$ in which every proposal is accepted. If the chain defined by Q is π-irreducible and $q(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = 0$ if and only if $q(\overline{\mathbf{y}}, \overline{\mathbf{x}}) = 0$, then the Metropolis–Hastings algorithm defined by q and (9) is π-irreducible. In particular, the chain of Algorithm 1 is π- irreducible when $p_X > 0$.*

Proof. The first part follows from Roberts & Smith, 1994, theorem 3.ii.

For Algorithm 1, $q(\overline{\mathbf{x}}, \overline{\mathbf{y}}) > 0$ only if $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$ in $\overline{E}(\mathbf{u})$ differ in at most a single component. Thus, assume that $x_j = y_j$ for all $j \neq i \in \{1, \ldots, n - m\}$ and $x_i \neq y_i$. Then $q(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = q(\overline{\mathbf{y}}, \overline{\mathbf{x}})$ so they are strictly positive or zero together. Write Q^τ for the τ-step transition kernel of the always-accept chain. Then, for $\overline{\mathbf{x}}, \overline{\mathbf{y}} \in \overline{E}(\mathbf{u})$,

$$q^{n-m}(\overline{\mathbf{x}}, \overline{\mathbf{y}}) \geq \left(\frac{1}{n-m} \right)^{n-m} \prod_{i=1}^{n-m} \frac{1}{\ell(\mathcal{X} \cap [a_{m+i}, a_{m+i} + l_{m+i}])} > 0,$$

by changing each component in turn. We conclude that the Markov chain of Algorithm 1 is π-irreducible. ∎

Recall that a π-irreducible Markov chain is called aperiodic (Meyn & Tweedie, 2009) if the state space ($\overline{E}(\mathbf{u})$ in our case) cannot be partitioned into measurable sets $B_0, B_1, \ldots, B_{d-1}$ such that $\pi(\overline{E}(\mathbf{u}) \setminus \cup_{j=0}^{d-1} B_j) = 0$ and $P(\overline{\mathbf{x}}, B_{j+1 \bmod d}) = 1$ for all $\overline{\mathbf{x}} \in B_j$ (for some $d > 1$, the period). By Møller and Waagepetersen (2003, proposition 7.6), a π-irreducible Markov chain with invariant probability distribution π is aperiodic if and only if for some small set D with $\pi(D) > 0$ and some $\tau \in \mathbb{N}$, the following holds: $P^i(\overline{\mathbf{x}}, D) > 0$ for all $\overline{\mathbf{x}} \in D$ and $i \geq \tau$.

Proposition 5. *Consider the setup of Theorem 2 with $n > m$ and assume that condition (8) is met. If $0 < p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) \leq \delta$ for some $\delta > 0$ and all $\overline{\mathbf{x}} \in \overline{E}(\mathbf{u})$, then the Markov chain of Algorithm 1 is aperiodic.*

Proof. Let ξ be the point on $\mathcal{X} \cap [a_{m+1}, a_{m+1} + l_{m+1}]$ that replaces x_1. By (8), there exist $x_2, \ldots, x_{n-m}$ such that

$$\int_{\mathcal{X} \cap [a_{m+1}, a_{m+1}+l_{m+1}]} c(\mathbf{u}) p_X(\{a_1, \ldots, a_m, \xi, x_2, \ldots, x_{n-m}\}) d\xi = \int_{\mathcal{X} \cap [a_{m+1}, a_{m+1}+l_{m+1}]} \pi(\xi, x_2, \ldots, x_{n-m}) d\xi$$

is strictly positive, where $c(\mathbf{u})$ is the normalization constant. Define a measure μ on the Borel σ-algebra on $\overline{E}(\mathbf{u})$ by

$$\mu(F) = \int_{\mathcal{X} \cap [a_{m+1}, a_{m+1}+l_{m+1}]} \mathbf{1}\{(\xi, x_2, \ldots, x_{n-m}) \in F\} \pi(\xi, x_2, \ldots, x_{n-m}) d\xi$$

and note that $\mu(\overline{E}(\mathbf{u})) > 0$. Set $C = \{(\xi, x_2, \ldots, x_{n-m}) : \xi \in \mathcal{X} \cap [a_{m+1}, a_{m+1} + l_{m+1}]\}$. We claim that C is small with respect to μ. To see this, take $\overline{\mathbf{y}} = (y, x_2, \ldots, x_{n-m}) \in C$ and note that for $F \subset \overline{E}(\mathbf{u})$, the transition probability $P(\overline{\mathbf{y}}, F)$ is at least

$$\frac{1}{n-m} \frac{1}{\ell(\mathcal{X} \cap [a_{m+1}, a_{m+1} + l_{m+1}])} \int_{\mathcal{X} \cap [a_{m+1}, a_{m+1}+l_{m+1}]} \mathbf{1}_F(\xi, x_2, \ldots, x_{n-m}) \alpha_1((y, x_2, \ldots, x_{n-m}), \xi) d\xi.$$

If $\pi(\xi, x_2, \ldots, x_{n-m}) \geq \pi(y, x_2, \ldots, x_{n-m})$, then $\alpha_1((y, x_2, \ldots, x_{n-m}), \xi) = 1 \geq \pi(\xi, x_2, \ldots, x_{n-m})/(c(\mathbf{u})\delta)$. Otherwise, $\alpha_1((y, \ldots, x_{n-m}), \xi) = \pi(\xi, x_2, \ldots, x_{n-m})/ \pi(y, x_2, \ldots, x_{n-m}) \geq \pi(\xi, x_2, \ldots, x_{n-m})/(c(\mathbf{u})\delta)$. In summary,

$$P(\overline{\mathbf{y}}, F) \geq \frac{1}{n-m} \frac{1}{\ell(\mathcal{X} \cap [a_{m+1}, a_{m+1} + l_{m+1}])} \frac{\mu(F)}{c(\mathbf{u})\delta},$$

so C is small with respect to μ. Moreover, $\pi(C) > 0$ because of the choice of $x_2, \ldots, x_{n-m}$.

Iterating the above argument one notices that for $\overline{\mathbf{y}} \in C$, $P^\tau(\overline{\mathbf{y}}, C)$ is at least as large as the τth power of the bound above with $F = \overline{E}(\mathbf{u})$, an observation that completes the proof. ∎

In conclusion, under mild conditions, from almost all initial states, Algorithm 1 converges in total variation to the invariant probability distribution. Conditions for general proposal kernels q can be found in Møller and Waagepetersen (2003, chapter 7) or Roberts and Smith (1994, theorem 3).

4.3 | Prior model parameters

In Section 4.2, we discussed Monte Carlo methods to sample from the posterior distribution of W or X given $U = \mathbf{u}$. This distribution is defined in terms of the prior probability density function p_X. Typically, p_X is given in unnormalised form and depends on a parameter vector θ, that is, $p_X(\mathbf{x}; \theta) = c(\theta)h_X(\mathbf{x}; \theta)$ for an explicit function $h_X : \mathcal{N}_\mathcal{X} \to \mathbb{R}^+$. Since the likelihood function for θ contains the latent marked point process W, we call on techniques from missing data analysis.

The likelihood function $l(\theta)$ is obtained from the proof of Theorem 2 by taking A equal to $\mathcal{N}_\mathcal{X}$ in Equation (5). Disregarding terms that do not depend on θ, one obtains $l(\theta; \mathbf{u}) = c(\theta)c(\theta|\mathbf{u})^{-1}$ where

$$c(\theta)^{-1} = \sum_{k=0}^{\infty} \frac{e^{-\ell(\mathcal{X})}}{k!} \int_{\mathcal{X}^k} h_X(\{x_1, \ldots, x_k\}; \theta) dx_1 \ldots dx_k,$$

and $c(\theta|\mathbf{u})^{-1}$ is given by

$$\int_{\mathcal{X}^{n-m}} h_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}; \theta) \left(\sum_{\substack{D_1, \ldots, D_{n-m} \\ \cup_i\{D_i\}=\{1, \ldots, n-m\}}} \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\} \right) \prod_{i=1}^{n-m} dx_i.$$

One observes that $c(\theta|\mathbf{u})$ is equal to the normalization constant of the nonatomic part of the posterior distribution of X given $U = \mathbf{u}$.

To handle the two normalization constants, it is necessary to look at the log relative likelihood $L(\theta)$ of U with respect to some fixed and user-selected reference parameter θ_0, Then, as in Gelfand and Carlin (1993) and Geyer (1999),

$$L(\theta) = \log\left[\frac{c(\theta)c(\theta_0|\mathbf{u})}{c(\theta_0)c(\theta|\mathbf{u})}\right] = \log \mathbb{E}_{\theta_0}\left[\frac{h_X(X; \theta)}{h_X(X; \theta_0)}|U = \mathbf{u}\right] - \log \mathbb{E}_{\theta_0}\left[\frac{h_X(X; \theta)}{h_X(X; \theta_0)}\right]. \tag{11}$$

Being expressible in terms of expectations under the reference parameter, the log-likelihood ratio can be approximated by Markov chain Monte Carlo methods. Note that two samples are

required: one from the posterior distribution of X and one from the prior. For the latter, provided p_X is locally stable, classic Metropolis–Hastings methods based on births and deaths apply (Geyer, 1999; Møller & Waagepetersen, 2003). If the conditional intensity is monotone, exact simulation can be carried out (Kendall & Møller, 2000; Lieshout & Baddeley, 2002). In Section 6, a Monte Carlo likelihood algorithm is used to perform this parameter estimation method on a burglary data set.

5 | SIMULATIONS WITH DIFFERENT PRIORS

In this section, we present a few simulations to illustrate how the choice of prior affects state estimation. Calculations were carried out using the C++ marked point process library MPPLIB, developed by Steenbeek et al. For p_X we choose the area-interaction point process (Baddeley & Lieshout, 1995; Widom & Rowlinson, 1970), a model that favors clustered, regular and random realizations depending on parameter values. Specifically, this model has probability density function

$$p(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \exp\left[-\log \gamma \ell(\mathcal{X} \cap U_r(\mathbf{x}))\right] \tag{12}$$

with respect to a unit rate Poisson process on $\mathcal{X}$. Here $U_r(\mathbf{x}) = \bigcup_{i=1}^{n} B(x_i, r)$ where $B(x_i, r)$ is the closed interval $[x_i - r, x_i + r]$. When $\gamma < 1$, realizations tend to be regular, for $\gamma > 1$ clustered. When $\gamma = 1$, one has a Poisson process with intensity β. The scalar $\alpha = c(\beta, \gamma)$ is the normalization constant. Realizations can be obtained by Kendall's dominated coupling from the past (CFTP) algorithm (Kendall, 1998) developed initially from the perfect simulation methods of Propp and Wilson for coupled Markov chains (Propp & Wilson, 1996).

5.1 | Toy example

Consider data $\mathbf{u} = \{(0.45, 0.4), (0.51, 0), (0.58, 0)\}$ that consist of two atoms and a single nondegenerate interval. By the discussion at the end of Section 3, for a Poisson prior ($\gamma = 1$), the posterior distribution of the location X_3 in $\mathcal{X} = (0, 1)$ that generated the nondegenerate interval is uniformly distributed. To see the effect of informative priors, Figure 2 plots the posterior distribution of X_3 when the prior is an area-interaction model with $\eta = 2r \log \gamma = 1.2$ and $r = 0.1$. Note that mass is shifted to the left side of the interval due to the presence of atoms. For $\eta = -1.2$ and $r = 0.1$, the atoms repel X_3, resulting in mass being shifted to the right side of the interval (cf. Figure 3). To carry out the state estimation, we ran Algorithm 1 with a burn-in of 10,000 steps and calculated the histograms based on the subsequent 100,000 steps.

5.2 | Area-interaction gamma model

The left-most panel in Figure 4 shows a simulation in $\mathcal{X} = (0, 1)$ from U in a model where X is an area-interaction process with parameters $\beta = 12$, $\eta = 2r \log \gamma = 0$ and $r = 0.05$ marked by a mixture distribution ν in which the atom probability is $p = 0.2$ and f_Y is the probability density function of a Gamma distribution with shape parameter $k = 2.5$ and rate parameter $\lambda = 0.07$. The points shown as black dots are the points of X in the simulated pattern, the red
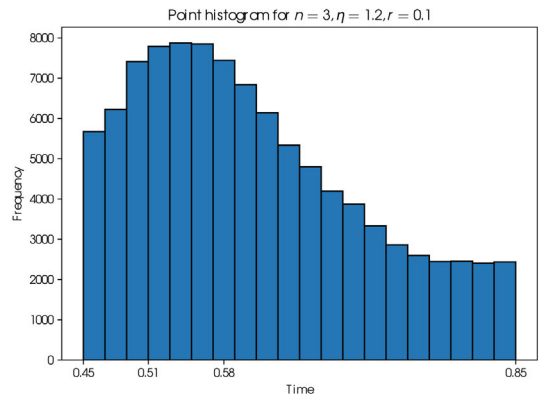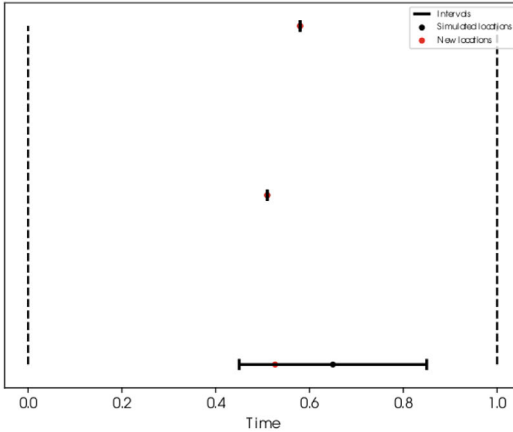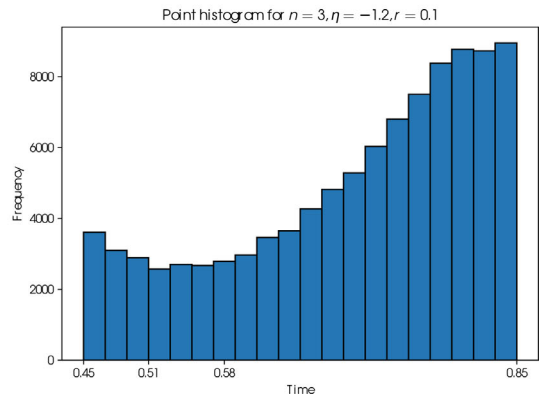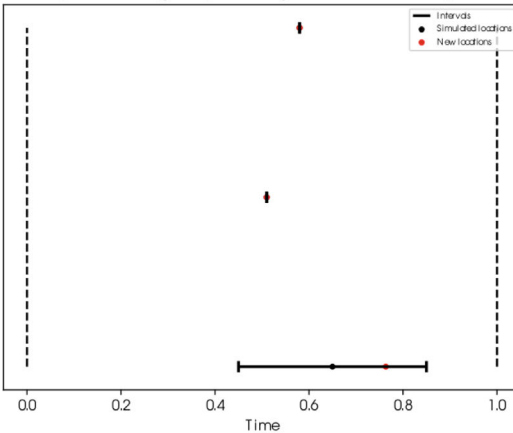
FIGURE 2 Locations of two atoms and a spanning interval together with a histogram of point locations. The interval start and end points as well as the atom locations are marked on the histogram x-axis.

FIGURE 3 Locations of two atoms and a spanning interval together with a histogram of point locations. The interval start and end points as well as the atom locations are marked on the histogram x-axis.

points constitute a realisation from the posterior distribution of X given U obtained by running Algorithm 1 for 10,000 steps. The points seem to settle in a random manner within the intervals.

A simulation using a prior favoring clustering can be found in black in Figure 5. The parameter settings were as before except that $\eta = 1.2$. In red, a realization from the posterior distribution is shown, obtained after 10,000 steps from the Metropolis–Hastings algorithm. Figure 5 shows the effect of the complex underlying geometry when choosing proposal points within parameterized intervals. The algorithm tends to move proposed times to areas where multiple intervals intersect, leading to clustering within these regions.
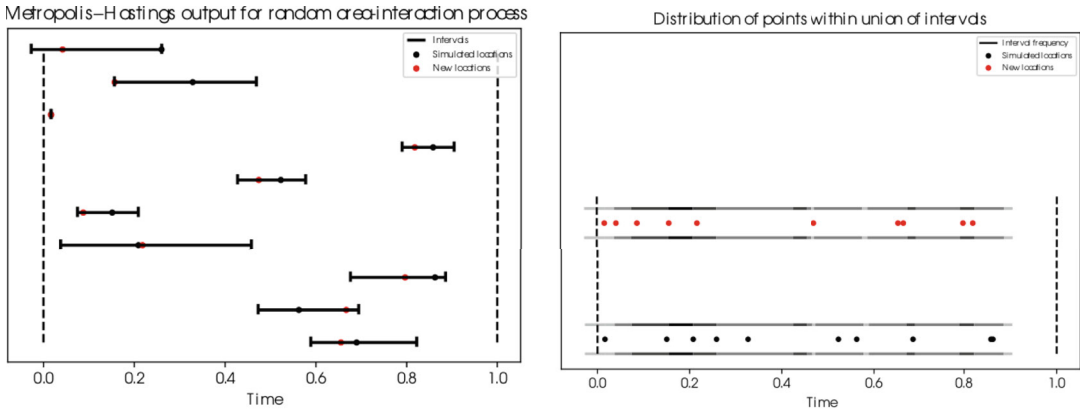
FIGURE 4 Plots of the simulated and new locations of a random area-interaction point process. Parameter values: $(\beta, \eta, r, \lambda, k, p) = (12, 0, 0.05, 0.07, 2.5, 0.2)$
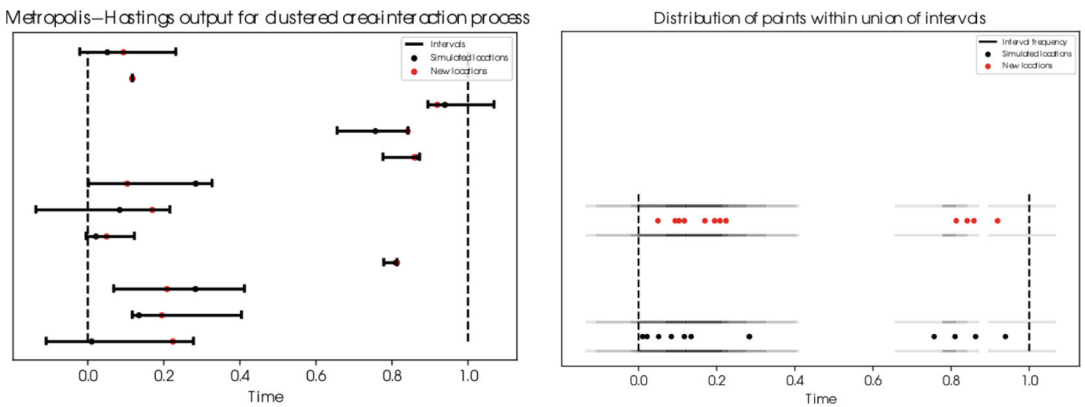

FIGURE 5 Parameter values: $(\beta, \eta, r, \lambda, k, p) = (12, 1.2, 0.05, 0.07, 2.5, 0.2)$

Figure 6 shows, in black, a simulation from U for a regular area-interaction prior with $\eta = 2r \log \gamma = -1.2$. A realisation from the posterior distribution of X is shown in red. The structure of the prior point process is maintained in the posterior, with points being spread out from each other.

6 | APPLICATION TO CRIMINOLOGICAL DATA

Having verified the effect of different priors on the model, a natural next step is to run the model on an actual dataset. Returning to the criminological context that inspired the model, times of property crimes are often not known due to victims not being home. Deducing the posterior distribution of offense times would aid improvements in predictive policing, as well as assist local governments in providing recommendations to residents in high-crime areas.
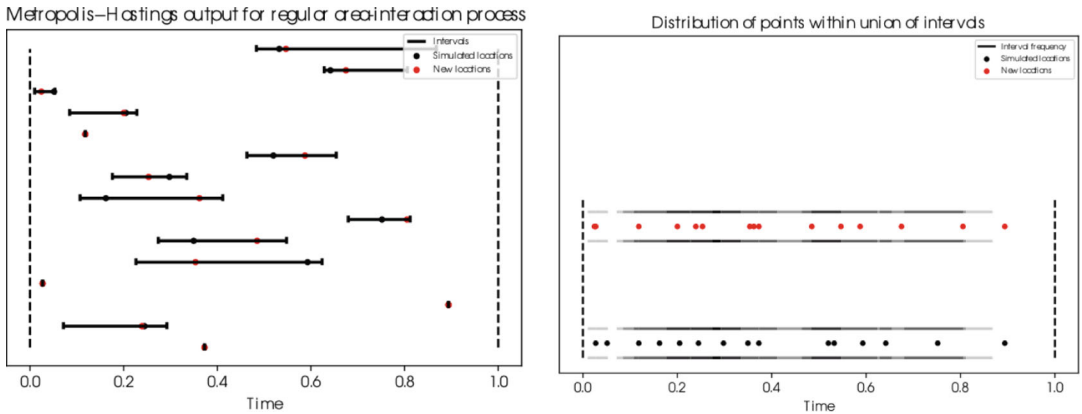
FIGURE 6   Parameter values: $(\beta, \eta, r, \lambda, k, p) = (12, -1.2, 0.05, 0.07, 2.5, 0.2)$

The data set we have chosen collects burglaries in Washington D.C. in February 2016 (see https://opendata.dc.gov/datasets/DCGIS::crime-incidents-in-2016/) after applying filters for month and type of crime. We scale the time so that $\mathcal{X} = (0, 1)$. Intervals smaller than 30 min in length are treated as atoms. The data also contain many fields, including exact date and time of a resident leaving and re-appearing at the scene of the crime.

6.1 | Parameter estimation

In Section 5 we considered simulated examples with pre-set parameter values to generate interval realizations. In the case of a real-data example, the aforementioned parameter values need to be estimated.

6.1.1 | Forward model parameters

In the case of the Washington D.C. data, a Gamma model proved to be unsatisfactory. Specifically, the salient feature of the collection of observed nondegenerate intervals is that it combines a large number of small intervals with a few very long ones, which cannot be captured by a Gamma distribution with shape parameter larger than one. Due to the length-weighted nature of the observed non-degenerate intervals, this criterion must be satisfied (see Proposition 2). Therefore, we instead use the heavy-tailed Weibull distribution.

Since we have independence of the mixture weight $\mathbb{E}[Y_1]/\mathbb{E}[T_1]$ on the parameter vector $\zeta = (k, \theta)$, we may use Proposition 2. The length-weighted marginal distribution generated by a Weibull random variable Y_1 has probability density function

$$f(l) = \frac{l f_Y(l)}{\mathbb{E} Y_1} = \frac{k}{\lambda \Gamma(1 + \frac{1}{k})} \left(\frac{l}{\lambda} \right)^k e^{-\left(\frac{l}{\lambda} \right)^k}, l > 0,$$

where k is the shape parameter and λ the scale parameter.

To infer the values of k and λ from the data, maximum likelihood estimation has been used. Using the log-likelihood, after some calculation we find that

$$\hat{\lambda} = \left(\frac{\hat{k}}{n(\hat{k}+1)} \sum_{i=1}^{n} l_i^{\hat{k}} \right)^{1/\hat{k}},$$

and we can hence solve the score equation

$$\frac{n}{k} + \frac{n}{k^2} \log \left(\sum_{i=1}^{n} l_i^k \right) + \sum_{i=1}^{n} \log (l_i) + \frac{n}{k^2} \log \left(\frac{k}{n(k+1)} \right)$$

$$+ \frac{n}{k^2} \psi \left(1 + \frac{1}{k} \right) - \frac{n(k+1)}{k} \frac{\sum_{i=1}^{n} l_i^k \log (l_i)}{\sum_{i=1}^{n} l_i^k} = 0,$$

numerically for k and subsequently λ, where $\psi(x) = \frac{d}{dx} \log(\Gamma(x))$ is the digamma function.

For the Washington D.C. data, $\hat{k} = 0.27$ and $\hat{\lambda} = 0.000046$.

6.1.2 | Prior parameters

We assume that the prior is an area-interaction model as in Section 5 with parameters β and η. To simplify calculations and for scaling reasons, we use $\kappa = \log \beta$ instead. Thus, $\theta = (\kappa, \eta)$ becomes the parameter vector. In this experiment, we have set $r = 0.008$, which corresponds to 5.5 h.

To estimate the parameter vector θ, we proceed as in Section 4.3 and consider (11). For the area-interaction model,

$$h_X(\mathbf{x}, \theta) = e^{\kappa n(\mathbf{x}) - \eta \frac{A(\mathbf{x})}{2r}},$$

where $A(\mathbf{x}) = \ell(\mathcal{X} \cap U_r(\mathbf{x}))$. The expectations in (11) must be approximated by Monte Carlo samples. Let N denote the number of samples. Then the Monte Carlo log relative likelihood $l_N(\theta)$ reads

$$l_N(\theta) = \log \left(\frac{1}{N} \sum_{i=1}^{N} \frac{h_X(X_{\mathbf{u},i}; \theta)}{h_X(X_{\mathbf{u},i}; \theta_0)} \right) - \log \left(\frac{1}{N} \sum_{i=1}^{N} \frac{h_X(X_i; \theta)}{h_X(X_i; \theta_0)} \right), \tag{13}$$

where $X_{\mathbf{u},i}$ are samples from the posterior and X_i from the prior. An approximate maximum likelihood estimator can be obtained by taking the gradient with respect to the parameter θ and equating to zero.

The parameters $\theta_0 = (\kappa_0, \eta_0)$ are known as reference parameters, and safeguard the algorithm from instability issues. A common method to choose θ_0 is through the Monte Carlo EM procedure (Dempster et al., 1977; Gelfand & Carlin, 1993; Geyer, 1999). In this process, one iteratively optimises the Monte Carlo approximation of

$$Q(\theta, \theta_k) = \mathbb{E}_{\theta_k}[\log p_X(X \cup \mathbf{a}; \theta)|U = \mathbf{u}],$$

writing **a** as the atoms of **u**. When, as in our case, the gradient of h is well-defined, this can be carried out by solving

$$\nabla Q(\theta, \theta_k) = \frac{1}{N} \sum_{i=1}^{N} \frac{\nabla h_X(X_{\mathbf{u},i}; \theta)}{h_X(X_{\mathbf{u},i}; \theta)} - \frac{\sum_{i=1}^{N} \frac{\nabla h_X(X_i; \theta)}{h_X(X_i; \theta_k)}}{\sum_{i=1}^{N} \frac{h_X(X_i; \theta)}{h_X(X_i; \theta_k)}} = 0$$

over θ (Geyer, 1999). When the reference parameters θ_0 have been found, the most likely value of θ is calculated by maximizing Equation (13). This parameter vector is then used to perform the state estimation procedure outlined in Section 4, which generates samples from the posterior distributions. In this way, realizations of burglary times are attained from real-life aoristic crime data sets.

6.2 | Results for application

Using the methods outlined in Section 6.1.2, the model was run on the Washington D.C. burglary dataset for February 2016. With model parameters tuned, prior parameters were estimated and realisations were generated with this set of parameter values after the Monte Carlo EM procedure was run until convergence. Ten such realizations have been plotted in Figure 7, with each realization being denoted in black. Each realiation corresponds to one configuration of burglary times. Note that for demonstration purposes, only 3 days of February are shown.
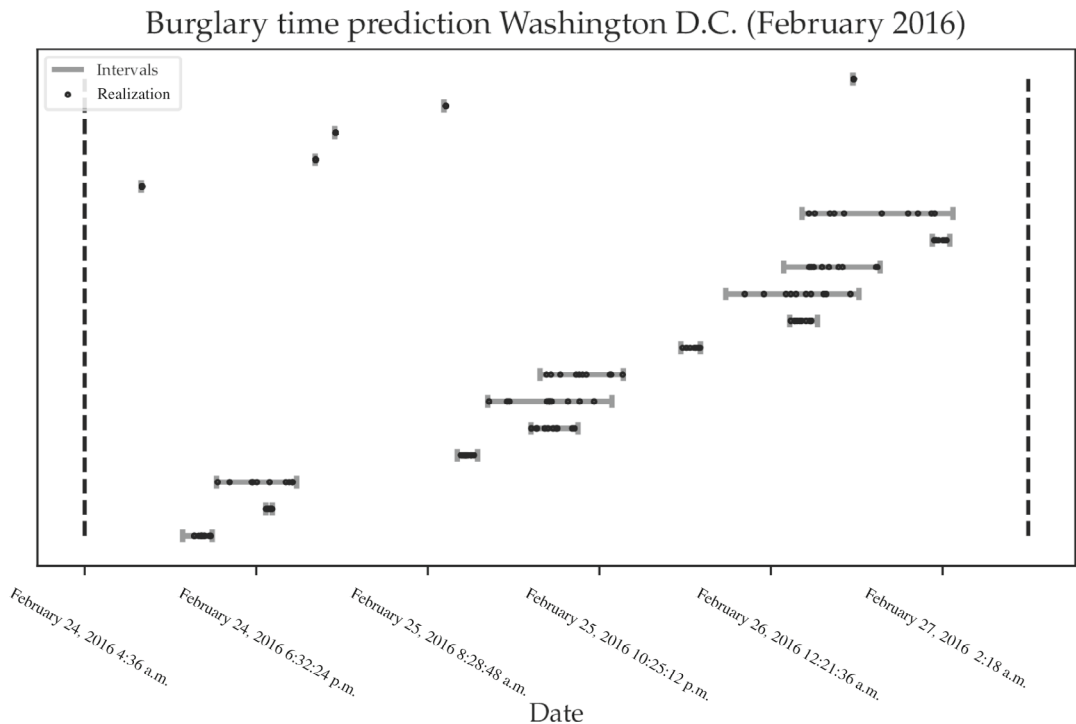

FIGURE 7 Model output for 3 days of February 2016 with 10 realizations in black. Prior parameters are estimated to be $\beta = 115.469$, $\eta = -0.256$ with model parameters $r = 0.008$, $k = 0.27$ and $\lambda = 0.000046$.

As in the simulation example, an area-interaction prior was used. The prior parameter estimation led to results of $\beta = 115.469$ and $\eta = -0.256$, meaning that the model assumed a regular underlying structure. While one may assume that criminology datasets would lend themselves better to clustered priors, this may not always be the case. For example, it may be the case that burglars wish not to interfere with each other, leading to burglary times being more spread out. Future work will look into these phenomena as well as relaxing certain, perhaps restrictive model assumptions.

7 | CONCLUSION

In this work, a Bayesian inference framework for aoristic data was introduced in which an alternating renewal process is used to model the interval censoring of temporal data, converting it into a marked point process model. A prospective point, which cannot be observed directly, was paired with an interval within which the point lies. State estimation was then applied to best estimate the location of this point. Theory was developed regarding the distribution of these marks based on this renewal framework and the posterior distribution deduced. The fact that the forward model allows for a mixture of discrete and absolutely continuous components makes this process nontrivial. A state estimation procedure was outlined in the form of a Metropolis–Hastings algorithm for a fixed number of points, after which ergodicity properties were verified. Using an area-interaction prior, this procedure was applied to sample from the posterior distribution. Effects of the prior were clearly present when sampling from the complete model. Finally, we demonstrated that the model can be applied to a criminological data set.

Throughout, we assumed that all intervals corresponding to a point in $\mathcal{X}$ were observed. Returning to a criminology context, sampling bias may arise since the data may contain only intervals whose right-most point is in a given interval. Additionally, a random labeling regime was assumed. It might be more realistic to have location-dependent independent marking, for example based on a semi-Markov process rather than an alternating renewal process. Furthermore, spatial aspects were completely ignored. These generalisations will form the topic for our future research.

ORCID
Maria N. M. van Lieshout https://orcid.org/0000-0002-7941-2176
Robin L. Markwitz https://orcid.org/0000-0001-5459-1692

REFERENCES
Asmussen, S. (2003). *Applied probability and queues.* Springer.
Baddeley, A., & Lieshout, M. N. M. van. (1995). Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, *47*, 601–619.
Bernasco, W. (2009). *Burglary.* In M. Tonry (Ed.), *Oxford handbook of crime and public policy* (pp. 165–190). Oxford University Press.
Brémaud, P. (1972). *A martingale approach to point processes* (PhD thesis). University of Berkeley.

Brix, A., & Diggle, P. J. (2001). Spatio-temporal prediction for Log-Gaussian Cox processes. *Journal of the Royal Statistical Society: Series B*, *63*, 823–841.

Brooks, S., Gelman, A., Jones, G., & Meng, X. L. (2011). *Handbook of Markov chain Monte Carlo*. CRC Press.

Daley, D. J., & Vere-Jones, D. (2003). *An introduction to the theory of point processes* (2nd ed.). Springer.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*, 1–38.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. CRC Press.

Gelfand, A. E., & Carlin, B. P. (1993). Maximum-likelihood estimation for constrained- or missing-data models. *Canadian Journal of Statistics*, *21*, 303–311.

Geyer, C. J. (1999). *Likelihood inference for spatial point processes*. In O. Barndorff-Nielsen, W. S. Kendall, & M. N. M. van Lieshout (Eds.), *Stochastic geometry, likelihood, and computation* (pp. 141–172). CRC Press.

Helms, D. (2008). *Temporal analysis*. In S. L. Gwinn, C. Bruce, J. P. Cooper, & S. Hick (Eds.), *Exploring crime analysis: Readings on essential skills* (pp. 214–257). International Association of Crime Analysts.

Karr, A. F. (1991). *Point processes and their statistical inference* (2nd ed.). Marcel Dekker.

Kendall, W. (1998). *Perfect simulation for the area-interaction point process*. In L. Accardi & C. Heyde (Eds.), *Probability towards the year 2000* (pp. 218–234). Springer.

Kendall, W. S., & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, *32*, 844–865.

Last, G., & Brandt, A. (1995). *Marked point processes on the real line: The dynamical approach*. Springer.

Mengersen, K. L., & Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics*, *24*, 101–121.

Meyn, S., & Tweedie, R. L. (2009). *Markov chains and stochastic stability* (2nd ed.). Cambridge University Press.

Møller, J., & Waagepetersen, R. (2003). *Statistical inference and simulation for spatial point processes*. CRC Press.

Propp, J. G., & Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, *9*, 223–252.

Ratcliffe, J. H., & McCullagh, M. J. (1998). Aoristic crime analysis. *International Journal of Geographical Information Science*, *12*, 751–764.

Roberts, G. O., & Smith, A. F. M. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications*, *49*, 207–216.

Ross, S. M. (1996). *Stochastic processes* (2nd ed.). Wiley.

Lieshout, M. N. M. van. (1995). *Stochastic geometry models in image analysis and spatial statistics*. CWI Tracts.

Lieshout, M. N. M. van. (2016). Likelihood based inference for partially observed renewal processes. *Statistics and Probability Letters*, *118*, 190–196.

Lieshout, M. N. M.van, & Baddeley, A. J. (2002). *Extrapolating and interpolating spatial patterns*. In A. B. Lawson & D. G. T. Denison (Eds.), *Spatial cluster modelling* (pp. 61–86). CRC Press.

Lieshout, M. N. M. van. (2000). *Markov point processes and their applications*. Imperial College Press.

Widom, B., & Rowlinson, J. S. (1970). A new model for the study of liquid-vapor phase transitions. *Journal of Chemical Physics*, *52*, 1670–1684.