



KALIKOW DECOMPOSITION FOR COUNTING PROCESSES WITH STOCHASTIC INTENSITY AND APPLICATION TO SIMULATION ALGORITHMS

TIEN CUONG PHI,* *Université Côte d'Azur, LJAD, France*

EVA LÖCHERBACH,** *Université Paris 1 Panthéon-Sorbonne, France*

PATRICIA REYNAUD-BOURET,*** *Université Côte d'Azur, CNRS, France*

Abstract

We propose a new Kalikow decomposition for continuous-time multivariate counting processes, on potentially infinite networks. We prove the existence of such a decomposition in various cases. This decomposition allows us to derive simulation algorithms that hold either for stationary processes with potentially infinite network but bounded intensities, or for processes with unbounded intensities in a finite network and with empty past before zero. The Kalikow decomposition is not unique, and we discuss the choice of the decomposition in terms of algorithmic efficiency in certain cases. We apply these methods to several examples: the linear Hawkes process, the age-dependent Hawkes process, the exponential Hawkes process, and the Galves–Löcherbach process.

Keywords: Kalikow decomposition; counting process; Hawkes process; perfect simulation; simulation algorithms

2020 Mathematics Subject Classification: Primary 60G55

Secondary 60K35

1. Introduction

Multivariate point (or counting) processes on networks have been used to model a large variety of situations, e.g. social networks [16], financial prices [2], and genomics [31]. One of the most complex network models comes from neuroscience, where the number of nodes can be as large as billions [24, 28, 32]. Several counting process models have been used to model such large networks, e.g. Hawkes processes [17, 18] and Galves–Löcherbach models [13]. The simulation of such large and potentially infinite networks is of fundamental importance in computational neuroscience [24, 28]. From a more mathematical point of view, the existence of such processes in a stationary regime and within a potentially infinite network has also attracted considerable interest (see [13], [19], and [27] in discrete or continuous time).

Kalikow decompositions [22] have been introduced and used mainly in discrete time. Such a decomposition provides a decomposition of the transition probabilities into a mixture of more elementary transitions. The whole idea is that even if the process is complex (infinite memory, infinite network), the elementary transitions look only at what happens in a finite

Received 31 March 2021; revision received 20 February 2023; accepted 24 February 2023.

* Postal address: Université Côte d'Azur, LJAD, France. Email: cuong.tienphi@gmail.com

** Postal address: Université Paris 1 Panthéon-Sorbonne, Statistique, Analyse et Modélisation Multidisciplinaire EA 4543 et FR FP2M 2036 CNRS, France. Email: eva.locherbach@univ-paris1.fr

*** Postal address: Université Côte d'Azur, CNRS, LJAD, France. Email: Patricia.Reynaud-Bouret@univ-cotedazur.fr

© The Author(s), 2023. Published by Cambridge University Press on behalf of Applied Probability Trust.

neighborhood in time and space. Once the decomposition is proved for a given process, this can be used to write algorithms to simulate the process. Indeed, by the Kalikow decomposition, the process can be decomposed into random elementary transitions that do not need access to the whole information to move forward in time. This is useful in two ways. First one can simulate the points appearing for a given node in the network, without needing the whole past history or even the whole network. This leads to *perfect simulation* algorithms [13, 19, 27]. Here the word ‘perfect’ refers to the fact that it is possible in finite time to simulate what happens on one or a finite number of the nodes of the potentially infinite network in a stationary regime. Second, this decomposition can drastically reduce the time complexity of the simulation algorithms, because we do not need to store or compute at each step what happens in the whole network to proceed. To our knowledge, all existing papers referring to Kalikow decomposition are theoretical and focus on the first aim: indeed, if one can prove that such a perfect simulation algorithm ends after a finite number of steps, it means at the same time that the process exists in a stationary regime. This is of tremendous theoretical importance when dealing with infinite networks [13, 19].

In the present paper we propose going from discrete to continuous time. Therefore we decompose conditional intensities rather than transition probabilities. This leads to serious difficulties that usually prevent a more practical application of the simulation algorithms. Indeed, to our knowledge, only a few papers deal with continuous-time counting processes. Dassios and Zhao [10] deal with Hawkes processes having a finite number of interacting components in the case of exponential memory kernels, and rely heavily on the underlying Markovian structure of the intensity. Their approach does not work for an infinite number of interacting components or for more general memory kernels. Hodara and Löcherbach [19] work with an infinite number of interacting components, for general non-linear Hawkes processes. But their decomposition is constructed under the assumption that there is a dominating Poisson process on each of the nodes, from which the points of the processes of interest can be thinned by rejection sampling (see also [26] for another use of thinning in the simulation of counting processes). To prove the existence of a Kalikow decomposition and go back to a more classical discrete-time setting, the authors need to freeze the dominating Poisson process, leading to a mixture, in the Kalikow decomposition, that depends on the realization of the dominating Poisson process. Such a mixture is not accessible in practice, and this prevents the use of their perfect simulation algorithm for more concrete purposes than mere existence.

More recently, in a previous computational article [28], we used another type of Kalikow decomposition, which does not depend on the dominating Poisson process. This leads to a perfect simulation algorithm, which can be used as a concrete way for computational neuroscience to simulate neuronal networks as an open physical system, where we do not need to simulate the whole network to simulate what happens in a small part of it [28]. However, this approach was mainly done in the direction of computer science, and the definition of the Kalikow decomposition was not sufficiently broad to encompass the classical counting process examples such as classical linear Hawkes processes.

In the present work we want to go further, by proposing a Kalikow decomposition for general continuous-time counting processes, under appropriate conditions on its intensity, without assuming the existence of a dominating Poisson process at all. We also prove (and this is not done in [28]) that such a decomposition exists for various interesting examples, even if it is not unique. More precisely, we show that processes having an intensity which is either continuous or can be approximated by a series of compactly supported intensities allows for such a decomposition. Finally we propose two algorithms. The first is essentially the one proposed in [28], except that it is now proved to work for the more general definition of the Kalikow decomposition that we have introduced here. This is a perfect simulation algorithm in the sense that it

can simulate the point process of a given node in a potentially infinite network in a stationary regime (it moves backwards to create the past that is needed to obtain a stationary process). To do so, we need to assume the existence of a dominating Poisson measure. The second algorithm only moves forwards from an empty past (meaning no points before zero) in a finite network and therefore simulates outside the stationary regime. In this case the intensity does not need to be bounded, as in the more classical Ogata's algorithm [26].

The paper is organized as follows. In Section 2 we introduce the basic notation and give the precise definition of a Kalikow decomposition. In Section 3 we present a very general method to obtain a Kalikow decomposition for a counting process having stochastic intensity. We exhibit this method on various examples including the linear Hawkes process [17, 18], the age-dependent Hawkes process [29], more general non-linear Hawkes processes with analytic rate function, and Galves–Löcherbach models [13]. Finally, in Section 4 we present the two algorithms based on the Kalikow decomposition given in Section 3, and we discuss the efficiency of the perfect simulation algorithm with respect to the Kalikow decomposition.

2. Notation and Kalikow decomposition

2.1. Notation and definition

We start this section by recalling the definition of simple locally finite counting processes and stochastic intensities. We refer the reader to [3] and [9] for more complete statements.

Let \mathbf{I} be a countable index set. We start by introducing a canonical path space for the sequence of points of a counting process having interacting components indexed by $i \in \mathbf{I}$. This space is given by

$$\mathcal{X}_\infty = \left\{ (\{t_n^i\}_{n \in \mathbb{Z}})_{i \in \mathbf{I}} : \text{for all } i \neq j \in \mathbf{I}, n, m \in \mathbb{Z}, \right. \\ \left. t_n^i \in [-\infty, \infty], \lim_{n \rightarrow \pm\infty} t_n^i = \pm\infty, t_n^i < t_{n+1}^i \text{ and } t_n^i \neq t_m^j \text{ if they are not infinite} \right\}.$$

Notice that we allow for the choices $t_n^i = \pm\infty$ such that elements of \mathcal{X}_∞ may have only a finite number of (finite) points before time 0 or only a finite number of (finite) points after time 0, or both. Henceforth, whenever we speak of the points of a counting process, we implicitly mean finite points.

We then introduce, for any $t \in \mathbb{R}$, $i \in \mathbf{I}$ and $x = (\{t_n^i\}_{n \in \mathbb{Z}})_{i \in \mathbf{I}}$,

$$\begin{cases} Z_t^i(x) = \sum_{n \geq 1} \mathbf{1}_{t_n^i \leq t}, & \text{if } t \geq 0, \\ Z_t^i(x) = -\sum_{n \leq 0} \mathbf{1}_{t < t_n^i}, & \text{if } t \leq 0, \end{cases}$$

and for brevity we write $Z = (Z^i)_{i \in \mathbf{I}}$ for the associated collection of counting processes, indexed by $i \in \mathbf{I}$. Note that $Z \in D(\mathbb{R}, \mathbb{Z})^{\mathbf{I}}$, where $D(\mathbb{R}, \mathbb{Z})$ is the space of non-decreasing càdlàg piecewise constant functions; see [21]. We let \mathcal{X}_t denote the canonical path space of points of Z before time t , given by

$$\mathcal{X}_t = \mathcal{X}_\infty \cap (-\infty, t)^{\mathbf{I}},$$

and we identify $(Z_s)_{s < t}$ with the past configuration $X_t \in \mathcal{X}_t$ defined by

$$X_t(x) = (\{t_n^i\}_{t_n^i < t})_{i \in \mathbf{I}}.$$

We consider $(\mathcal{F}_t)_{t \in \mathbb{R}}$, the past filtration of the process $Z = (Z^i)_{i \in \mathbf{I}}$, defined by

$$\mathcal{F}_t = \sigma(Z^i_s, i \in \mathbf{I}, s \leq t).$$

Moreover, for any $x = (\{t^i_n\}_{n \in \mathbb{Z}_-})_{i \in \mathbf{I}} \in \mathcal{X}$ and any $i \in \mathbf{I}$, we denote the point measure associated to index i by

$$dx^i_s = \sum_{m \in \mathbb{Z}^-} \delta_{t^i_m}(ds),$$

which is an element of the space \mathcal{N} of locally finite point measures on \mathbb{R}_- , endowed with the topology of vague convergence, and we endow \mathcal{X} with the metric induced by the product metric on $\mathcal{N}^{\mathbf{I}}$. Finally, throughout this article and without further mention, the integral \int_a^b stands for $\int_{[a,b)}$ with $a, b \in \mathbb{R}$, and $Z^i([a, b))$ (resp. $Z^i((a, b])$) stands for the number of points in the configuration with index i in $[a, b)$ (resp. in $(a, b]$).

In the present article we are only interested in time-homogeneous counting processes, that is, informally, processes that at each time t depend in the same way on the past configuration x_t . To be more rigorous, let us define, for each $x_t = (\{t^i_n\}_{t^i_n < t})_{i \in \mathbf{I}} \in \mathcal{X}_t$,

$$x_t^{\leftarrow t} := (\{t^i_n - t\}_n)_i,$$

which is the shifted configuration at time 0. The generic space for such a past configuration $x_t^{\leftarrow t}$ that is rooted at time 0 is denoted by $\mathcal{X} := \mathcal{X}_0$.

Under suitable assumptions, the evolution of the counting process $Z = (Z^i)_{i \in \mathbf{I}}$ with respect to $(\mathcal{F}_t)_{t \in \mathbb{R}}$ is fully characterized by its stochastic intensity, which depends on the past configuration; see Proposition 7.2.IV of [9]. Hence, in this paper, for any $x_t \in \mathcal{X}_t$, given that the past before time t is x_t , we let $\phi^i_t(x_t)$ denote the corresponding stochastic intensity of the process Z^i at time t for any $i \in \mathbf{I}$. More precisely, for any $x_t \in \mathcal{X}_t$, we have

$$\mathbb{P}(Z^i \text{ has a jump in } [t, t + dt) \mid \text{past before time } t = x_t) = \phi^i_t(x_t) dt.$$

Definition 2.1. For a given $i \in \mathbf{I}$, a counting process Z^i with stochastic intensity $(\phi^i_t(x_t))_{t \in \mathbb{R}}$ is said to be time-homogeneous if there exists a measurable function $\phi^i: \mathcal{X} \rightarrow \mathbb{R}_+$, called the *generic intensity*, such that

$$\phi^i_t(x_t) = \phi^i(x_t^{\leftarrow t})$$

for all $t \in \mathbb{R}$ and $x_t \in \mathcal{X}_t$.

Note that a counting process that is time-homogeneous is not necessarily stationary. For instance, exponential Hawkes processes [5] (see Section 3.3 below for more details, in particular (3.9) for the precise form of the intensity) starting with empty past before time 0 (i.e. empty past history before time 0, or in other words, no points before 0) may explode in finite time, but they are still time-homogeneous in the sense of the above definition. One can also think of simple linear Hawkes processes (see (3.4) below for the precise form of the intensity) with empty past before time 0 in the supercritical regime, that is, for which the interaction function has L^1 -norm larger than 1 and thus produces an exponentially growing number of points as time increases [2]. Therefore, if the process of interest is stationary, one can think of \mathcal{X} as \mathcal{X}_0 , the set of configurations at time 0. However, if this is not the case, \mathcal{X} has to be thought of as

just a generic space whose configurations are cut at time 0 (i.e. no points exist after 0) and which is the set on which ϕ^i is defined.

To define the Kalikow decomposition, we need to define neighborhoods and cylindrical functions on neighborhoods. A neighborhood v is a Borel subset of $\mathbf{I} \times (-\infty, 0)$. This neighborhood is said to be finite if there exists a finite subset $J \subset \mathbf{I}$ and a finite interval $[a, b]$ such that

$$v \subset J \times [a, b].$$

Definition 2.2. For any neighborhood v and $x, y \in \mathcal{X}$, we say that $x \stackrel{v}{=} y$ whenever $x = y$ in v . This means that, for all $i \in \mathbf{I}, n \in \mathbb{Z}$, such that $t_n^i \in x$ and $(i, t_n^i) \in v$, we have $t_n^i \in y$ and vice versa.

A real-valued function f is said to be cylindrical in v if $f(x) = f(y)$ for any $x \stackrel{v}{=} y$, and we usually stress the dependence on v by writing $f_v(x)$.

A family of neighborhoods \mathbf{V} is a countable collection of finite neighborhoods v . It usually includes the empty set \emptyset . One might have a different family of neighborhoods for each i , even if in several examples the same family works for all i .

Definition 2.3. A time-homogeneous counting process $(Z^i)_{i \in \mathbf{I}}$ of generic intensity ϕ^i admits the Kalikow decomposition with respect to the collection of neighborhood families $(\mathbf{V}^i)_{i \in \mathbf{I}}$ and a given subspace \mathcal{Y} of \mathcal{X} if, for any $i \in \mathbf{I}$ and any $v \in \mathbf{V}^i$, there exists a cylindrical function $\phi_v^i(\cdot)$ on v taking values in \mathbb{R}_+ and a probability $\lambda^i(\cdot)$ on \mathbf{V}^i such that

$$\phi^i(x) = \sum_{v \in \mathbf{V}^i} \lambda^i(v) \phi_v^i(x) \quad \text{for all } x \in \mathcal{X} \cap \mathcal{Y}.$$

Remark 2.1. Note that the probability $\lambda^i(\cdot)$ in Definition 2.3 is a deterministic function, which is why this decomposition is unconditional, whereas in [19], $\lambda^i(\cdot)$ depends on the dominating Poisson processes (see the discussion in the Introduction). Second, we do not restrict ourselves to a bounded intensity, and we do not force all the ϕ_v^i to be bounded with the same bound, which is a notable improvement compared to [28].

2.2. On the subspace \mathcal{Y}

A Kalikow decomposition does not exist for all intensities and all subspaces \mathcal{Y} , and we stress the fact that it depends on the choice of \mathcal{Y} . The role of \mathcal{Y} is to make the Kalikow decomposition achievable. There are many possible choices for such a subspace, depending on the model under consideration and the precise form of the intensity. In this paper we will discuss two main examples. The first example is the choice $\mathcal{Y} = \mathcal{X}^{>\delta}$, which is the subspace of \mathcal{X} where the distance between any two consecutive possible points is greater than δ , that is,

$$\mathcal{X}^{>\delta} = \{x = (\{t_n^i\}_{n \in \mathbb{Z}^-})_{i \in \mathbf{I}} \in \mathcal{X} : \text{for all } n, i, t_{n+1}^i - t_n^i > \delta\}. \quad (2.1)$$

Such a choice is convenient for counting processes with a hard exclusion role where, by definition of the intensity, any two consecutive points need to be at a distance at least equal to δ ; see Section 3.2 below, where we discuss the example of age-dependent Hawkes processes with hard refractory period.

In the case when ϕ^i is continuous for each i and we want to simulate the process starting from the empty past before time 0, during some finite time interval $[0, T]$ and up to some

activity level $K > 0$, another possible choice of a subspace \mathcal{Y} that we consider is

$$\mathcal{Y} = \mathcal{X}^{T,K} = \{x \in \mathcal{X} : \text{for all } i \in \mathbf{I}, Z_T^i(x) \leq K\}.$$

By continuity of ϕ^i , it is clear that the intensities are bounded on $\mathcal{X}^{T,K}$. In this case we will be able to simulate the process, using the Kalikow decomposition, up to the first exit time of $\mathcal{X}^{T,K}$; see Section 4.1 below.

2.3. Representation, thinning, and simulation

Locally finite simple counting processes with intensity $\phi_t^i(x_t)$ can always be represented as thinning of a bivariate Poisson measure. More precisely, if $(\pi^i)_{i \in \mathbf{I}}$ are independent Poisson random measures on $\mathbb{R} \times \mathbb{R}_+$ with intensity 1, the point measures defined by

$$dX_t^i = \int_{u \geq 0} \mathbf{1}_{u \leq \phi_t^i(x_t)} \pi^i(dt, du)$$

define points of counting processes $(Z^i)_{i \in \mathbf{I}}$ having intensity $\phi_t^i(x_t)$ at time t , given that the past before time t is x_t ; see e.g. Lemma 3 and 4 of [4] and Chapter 14 of [20].

This representation has been used for a long time to simulate processes forwards in time; see e.g. [23] and [26]. More precisely, consider for the moment the easy case where we have empty past before time 0 and intensities that are bounded for any i by a fixed constant $\Gamma^i > 0$. The simulation of the Poisson random measure π^i then consists in a homogeneous Poisson process, N^i , in time, having intensity Γ^i (which can also be built by a succession of independent exponential jumps of parameter Γ^i), and then to attach to each point T of this process, independently of anything else, independent marks U_T which are uniformly distributed on $[0, \Gamma^i]$. The pair (T, U_T) for $T \in N^i$ forms a bivariate Poisson random measure π^i in the band of height Γ^i .

The classical thinning algorithm – for intensities bounded by Γ^i – then consists in saying that the points of N^i are accepted if $U_T \leq \phi_T^i(X_T)$ and that these accepted points correspond to a counting process of intensity $\phi_T^i(X_T)$. Equivalently, one can attach independent uniform marks \mathcal{U}_T on $[0,1]$ and say that we accept T if $\mathcal{U}_T \leq \phi_T^i(X_T)/\Gamma^i$, or one can even just say that we accept a point T in N^i with probability $\phi_T^i(X_T)/\Gamma^i$.

The Kalikow decomposition allows us to go one step further thanks to the following result.

Proposition 2.1. *Let $(\mathbf{V}^i)_{i \in \mathbf{I}}$ be a collection of families of finite neighborhoods, and for any $i \in \mathbf{I}$, let $(\lambda^i(v))_{v \in \mathbf{V}^i}$ be probabilities on \mathbf{V}^i and let ϕ_v^i be cylindrical functions on v for each $v \in \mathbf{V}^i$. Moreover, let $(\Pi^i)_{i \in \mathbf{I}}$ be independent Poisson measures on $\mathbb{R} \times \mathbb{R}_+ \times \mathbf{V}^i$ with intensity measure $dt du \lambda^i(dv)$.*

Then, for every $i \in \mathbf{I}$,

$$dZ_t^i = \int_{u \geq 0, v \in \mathbf{V}^i} \mathbf{1}_{u \leq \phi_v^i(X_t^{\leftarrow t})} \Pi^i(dt, du, dv)$$

defines a time-homogeneous counting process Z^i having generic intensity given by

$$\phi^i = \sum_{v \in \mathbf{V}^i} \lambda^i(v) \phi_v^i. \tag{2.2}$$

From a more algorithmic point of view, the construction of Π^i is equivalent to attaching to each atom (T, U_T) of π^i (the bivariate Poisson random measure) a mark V_T , independently of

everything else and distributed according to λ^i . The above theoretical result can be interpreted in the following way. It is sufficient to draw the neighborhood V_T at random for each point T according to λ^i , and to do so as if the intensity were just $\phi_{V_T}^i(X_T^{\leftarrow T})$ instead of having to compute the full sum in (2.2).

From a computational point of view, the main interest of this is to diminish drastically the number of computations to be done, replacing the summation with a random selection. From a theoretical point of view, the above representation enables us to define reset events at which the process forgets its past, at least in a local way. These resets take place whenever, for instance, the empty set is picked as a neighborhood, which happens with probability $\lambda^i(\emptyset)$, for a given $i \in \mathbf{I}$. Indeed, this means that at this time t , the history of the i th component is reset and becomes independent of the past. In a nutshell, the whole theoretical interest of using the Kalikow decomposition for perfect simulation is therefore to prove that such resets happen often enough and for sufficiently many coordinates i , and that one can simulate the distribution between resets without needing to simulate outside this zone. The interest of this strategy is twofold. First of all, if this procedure works, it shows that the stationary distribution exists and is unique (see Theorem 4.1 below). Moreover, it also allows us to perfectly simulate from this stationary distribution.

The above representation needs a three-dimensional Poisson random measure. As we have seen above, if it is easy to simulate such a measure within a band, we cannot simulate it without having an upper bound on the intensity. That is why we distinguish two cases for the algorithms. Here is a brief informal overview.

- If we are looking for a stationary distribution, we need to ‘propose’ a first point before thinning it in a ‘Kalikow’ way. To do so, we need a fixed upper bound, say Γ^i , that holds for all times, for a given coordinate i . Then we will be able to go back in time recursively as follows: (i) pick the random neighborhood, (ii) simulate the points in the neighborhood according to a Poisson process of intensity Γ^j if the neighborhood is on node j , and (iii) search again for the neighborhoods of the points that we just simulated and go on. In this backward step, we create a clan of ancestors for the first point. Under some conditions, this recursion ends in finite time, because either the empty neighborhood is picked or because the simulation of the Poisson process inside the neighborhood is empty. Hence the status of the first point, even inside the stationary distribution, only depends on a finite set of points that we have been able to create. It remains to accept or reject, in a forward movement, all these points according to the rule $U_T \leq \phi_{V_T}^i(x_T^{\leftarrow T})$. See Section 4.2 below.
- If there is no bound, which is typically the case for the linear or exponential Hawkes process, one cannot ‘propose’ a first point in a stationary manner. However, if we start with an empty past before 0, the intensity at time 0 is usually bounded and we can ‘propose’ a first point. The main strategy is now to update this upper bound as time goes by and to change the size of the steps we are making, one point after another. This algorithm can only move in forward time and we need to know all the network to make it move forwards. Hence this approach only works for finite networks with known past (typically, empty past before time 0). See Section 4.1 below.

Note that in particular for linear (multivariate) Hawkes processes, there is another way to do perfect simulation, relying on the cluster representation of these processes; see [6], [7], and [25]. However, the approach we propose here is much more general, because we do not need

this cluster representation, but rely on the Kalikow decomposition which exists for a broader class of processes (e.g. non-linear Hawkes processes with hard refractory period and Lipschitz rate function; see Section 3.2 below).

The above proposition of course holds as long as $\sum_{v \in \mathbf{V}^i} \lambda^i(v) \phi_v^i$ exists, and it furnishes a way to simulate a process with intensity ϕ^i . Below, we will see how we combine it with a proper definition of \mathcal{Y} to make the simulation algorithms work in practice.

Proof of Proposition 2.1. To avoid confusion, we consider two filtrations. We write $(\mathcal{F}_t)_{t \in \mathbb{R}}$ for the canonical filtration of $(\Pi^i)_{i \in \mathbf{I}}$, containing the filtration $(\mathcal{F}_t^Z)_{t \in \mathbb{R}}$ which corresponds to the one given by the counting processes $(Z^i)_{i \in \mathbf{I}}$.

To prove that $(Z^i)_{i \in \mathbf{I}}$ has the correct intensity, let us look at $\mathbb{E}(Z^i((a, b]) | \mathcal{F}_a^Z)$ for any $a < b$. We have

$$\mathbb{E}(Z^i((a, b]) | \mathcal{F}_a^Z) = \mathbb{E}\left(\int_{s \in (a, b], u \geq 0, v \in \mathbf{V}^i} \mathbf{1}_{u \leq \phi_v^i(X_s^{\leftarrow s})} \Pi^i(ds, du, dv) | \mathcal{F}_a^Z\right).$$

The integral in v is independent of the rest, so we can integrate it and replace it with its corresponding intensity measure, which leads to

$$\mathbb{E}(Z^i((a, b]) | \mathcal{F}_a^Z) = \mathbb{E}\left(\sum_{v \in \mathbf{V}^i} \lambda^i(v) \int_{s \in (a, b], u \geq 0} \mathbf{1}_{u \leq \phi_v^i(X_s^{\leftarrow s})} \pi^i(ds, du) | \mathcal{F}_a^Z\right),$$

where π^i is the bivariate Poisson random measure of rate 1. Therefore we get

$$\mathbb{E}(Z^i((a, b]) | \mathcal{F}_a^Z) = \sum_{v \in \mathbf{V}^i} \lambda^i(v) \mathbb{E}\left(\int_{s \in (a, b]} \phi_v^i(X_s^{\leftarrow s}) ds | \mathcal{F}_a^Z\right),$$

that is,

$$\mathbb{E}(Z^i((a, b]) | \mathcal{F}_a^Z) = \mathbb{E}\left(\int_{s \in (a, b]} \phi_s^i(X_s) ds | \mathcal{F}_a^Z\right),$$

which concludes the proof. □

3. How to compute the Kalikow decomposition in various cases

To obtain the Kalikow decomposition, we try to rewrite the intensity as a convergent sum of cylindrical functions over a suitable family of neighborhoods. To define such a family of neighborhoods properly, let us start by introducing the minimal information that is needed to compute the intensity.

Let us consider a time-homogeneous counting process Z^i having intensity $\phi_t^i(x_t) = \phi^i(x_t^{\leftarrow t})$. Let $\mathcal{V}^i \subset \mathbf{I} \times (-\infty, 0)$ be minimal such that ϕ^i is cylindrical on \mathcal{V}^i . We interpret \mathcal{V}^i as the support of dependence for ϕ^i .

Definition 3.1. A coherent family of finite neighborhoods \mathbf{V}^i for Z^i is such that

$$\mathcal{V}^i \subset \bigcup_{v \in \mathbf{V}^i} v.$$

Proposition 3.1. Let $Z = (Z^i)_{i \in \mathbf{I}}$ be a time-homogeneous counting process having intensity $\phi_t^i(x_t) = \phi^i(x_t^{\leftarrow t})$ at time t , let \mathbf{V}^i be an associated coherent family of neighborhoods, and let

$\mathcal{Y} \subset \mathcal{X}$ be a subspace of \mathcal{X} . If there exists a family of non-negative functions $\Delta_v^i(x)$ that are cylindrical on v for each $v \in \mathbf{V}^i$ and $i \in \mathbf{I}$ such that

$$\sum_{v \in \mathbf{V}^i} \Delta_v^i(x) \text{ is convergent and } \phi^i(x) = \sum_{v \in \mathbf{V}^i} \Delta_v^i(x) \text{ for all } x \in \mathcal{X} \cap \mathcal{Y}, \tag{3.1}$$

then we have the following.

(i) For any choice of probabilities λ^i on \mathbf{V}^i such that

$$\lambda^i(v) = 0 \text{ only if } \sup_{x \in \mathcal{X} \cap \mathcal{Y}} \Delta_v^i(x) = 0,$$

then $(Z^i)_{i \in \mathbf{I}}$ admits the Kalikow decomposition with respect to the collection of neighborhood families $(\mathbf{V}^i)_{i \in \mathbf{I}}$ and the subspace \mathcal{Y} , with weights given by $\lambda^i(v)$ and cylindrical functions ϕ_v^i , where

$$\phi_v^i(x) = \Delta_v^i(x) / \lambda^i(v),$$

with the convention $0/0 = 0$.

(ii) If, for all $i \in \mathbf{I}$ and $v \in \mathbf{V}^i$, there exist non-negative deterministic constants Γ_v^i such that

$$\sup_{x \in \mathcal{X} \cap \mathcal{Y}} \Delta_v^i(x) \leq \Gamma_v^i < \infty$$

and $\Gamma^i = \sum_{v \in \mathbf{V}^i} \Gamma_v^i \neq 0$ is finite, then one can in particular choose

$$\lambda^i(v) = \frac{\Gamma_v^i}{\Gamma^i} \text{ and } \phi_v^i(x) = \frac{\Gamma_v^i}{\Gamma^i} \Delta_v^i(x) \text{ for all } i \in \mathbf{I}, v \in \mathbf{V}^i. \tag{3.2}$$

In this case, all functions ϕ_v^i and ϕ^i are upper-bounded by Γ^i .

Proof. The first point is obvious. Note that for the second one, since $\sum_{v \in \mathbf{V}^i} \Gamma_v^i = \Gamma^i$, the choice $\lambda^i(v) = \Gamma_v^i / \Gamma^i$ defines a probability. \square

The weights $\lambda^i(v)$ in the Kalikow decomposition are not unique, and this is so even in the bounded case. Indeed, Γ_v^i can always be chosen much bigger than $\sup_{x \in \mathcal{X} \cap \mathcal{Y}} \Delta_v^i(x)$. If we order the neighborhoods from the most simple (the empty set) to the most complex (by size, or by distance to $(i, 0)$), this means that it is always possible to choose large weights $\lambda^i(v)$ on complex neighborhoods, but not that easy to choose small weights on them. However, we would like the backward steps of the perfect simulation algorithm to end after a small finite number of iterations. Hence we typically need larger weights on the empty set and on the less complex neighborhoods such that fewer computations are done or less memory is stored in our algorithms. This means that one can try to optimize the weights to minimize the complexity of the algorithms that we develop (see Section 4.2.6).

Note also that the above result leads directly to a very general statement for continuous generic intensities.

Corollary 3.1. Let $Z = (Z^i)_{i \in \mathbf{I}}$ be a time-homogeneous counting process with generic intensity ϕ^i , let $\mathbf{V}^i = \{v_k^i, k \in \mathbb{N}\}$ be an associated coherent family of neighborhoods which is increasing, i.e. $v_k^i \subset v_{k+1}^i$ for all k , and let \mathcal{Y} be a subspace of \mathcal{X} , such that for any $i \in \mathbf{I}$, ϕ^i is strongly

continuous on \mathcal{Y} , that is,

$$\sup_{i \in \mathbf{I}} \sup_{x, y \in \mathcal{Y}: x \stackrel{v_i}{=} y} |\phi^i(x) - \phi^i(y)| \rightarrow 0 \tag{3.3}$$

as $k \rightarrow \infty$.

Then item (i) of Proposition 3.1 holds with the choice

$$\begin{cases} \Delta_{v_0}^i(x) = \inf \{ \phi^i(y) : y \in \mathcal{Y}, y \stackrel{v_0}{=} x \}, \\ \Delta_{v_k}^i(x) = \inf \{ \phi^i(y) : y \in \mathcal{Y}, y \stackrel{v_k}{=} x \} - \inf \{ \phi^i(y) : y \in \mathcal{Y}, y \stackrel{v_{k-1}}{=} x \} \text{ for } k > 0. \end{cases}$$

Let us briefly resume the above discussion. All processes having a generic intensity which is strongly continuous in the sense of (3.3) admit a Kalikow decomposition. If, moreover, we impose a sufficiently rapid decay of the modulus of continuity of the rate function, as the neighborhoods approach the whole space, it is possible in this case to perform the perfect simulation algorithm of Section 4.2. The processes discussed below in Sections 3.2 and 3.4 belong to this class of processes.

Moreover, all processes having a generic intensity that can be approximated by a series of local intensities having compact interaction support (in both time and space) admit a Kalikow decomposition. In general, in this case we will only be able to simulate the process outside the stationary regime. The processes discussed below in Sections 3.1 and 3.3 belong to this class of processes.

In what follows we discuss how this method can be implemented for various specific examples of processes and for various examples of neighborhood families. The choice of the neighborhood family depends heavily on the dynamic considered, and has to be ‘guessed’ for each concrete example.

3.1. Linear Hawkes process

In the following, we consider a linear Hawkes process [17, 18] for a finite number of interacting components, that is, \mathbf{I} is finite. In this framework, for any $x \in \mathcal{X}$, we have

$$\phi^i(x) = \mu^i + \sum_{j \in \mathbf{I}} \int_{-\infty}^0 h_j^i(-s) dx_s^j, \tag{3.4}$$

where the non-negative interaction functions $h_j^i(\cdot)$ measure the local dependence of process Z^i on Z^j and the non-negative parameters μ^i refer to the spontaneous rate of process Z^i . The classical assumption to ensure a stationary Hawkes process is to assume that the spectral radius of $(\int_0^\infty h_j^i(s) ds)_{i,j \in \mathbf{I}}$ is strictly smaller than 1. We do not need to make this assumption here: we just assume that for any i, j , $h_j^i \in L_{loc}^1$. In this case we have

$$\mathcal{V}^i = \bigcup_{j \in \mathbf{I}: h_j^i \neq 0} \{j\} \times \text{Supp}(h_j^i). \tag{3.5}$$

Cutting the support of h_j^i into small pieces of length ϵ , for some fixed $\epsilon > 0$, we are led to consider atomic neighborhoods of the type

$$w_{j,n} = \{j\} \times [-n\epsilon, -(n-1)\epsilon), \tag{3.6}$$

which are supported by one single neuron within a small interval of length ϵ . The most generic family of neighborhoods that we can consider in this sense is

$$\mathbf{V}_{\text{atom}} = \{\emptyset\} \cup \{w_{j,n} : j \in \mathbf{I}, n \in \mathbb{N}^*\},$$

which we choose for all i . For this family one can prove the following result as a straightforward corollary of Proposition 3.1.

Corollary 3.2. *The multivariate linear Hawkes process defined by (3.4) admits the Kalikow decomposition given by Proposition 3.1(i), with respect to the collection of neighborhood families $(\mathbf{V}^i)_{i \in \mathbf{I}}$, where for all i , $\mathbf{V}^i = \mathbf{V}_{\text{atom}}$, and with respect to the subspace*

$$\mathcal{Y} = \left\{ x \in \mathcal{X} : \sum_{i \in \mathbf{I}} \phi^i(x) < \infty \right\},$$

with

$$\begin{cases} \Delta_{w_{j,n}}^i(x) := \int_{-n\epsilon}^{-n\epsilon+\epsilon} h_j^i(-s) dx_s^j, & \text{if } v = w_{j,n}, \\ \Delta_{\emptyset}^i = \mu^i, & \text{if } v = \emptyset. \end{cases}$$

Note in particular that the linear Hawkes process has an intensity which is not bounded by a fixed constant, so we cannot use Proposition 3.1(ii). In particular, we stress that we will not be able to use the above decomposition for the purpose of perfect simulation of the stationary version of the linear Hawkes process (which does not necessarily exist under our minimal assumptions). However, we will be able to use the decomposition for an efficient sampling algorithm of the process in its non-stationary regime, starting from the empty past before time 0; see Section 4.1 below.

Remark 3.1. The choice of $\mathcal{Y} = \{x \in \mathcal{X} : \sum_{i \in \mathbf{I}} \phi^i(x) < \infty\}$ is the minimal choice to ensure that (3.1) holds, by monotone convergence. By classical results on linear Hawkes processes, it is well known that, starting from the empty past at time 0, $Z_t = (Z_t^i)_{i \in \mathbf{I}}$ stays within \mathcal{Y} almost surely, for any $t \geq 0$; see e.g. [11].

3.2. Age-dependent Hawkes process with hard refractory period

In this section we are interested in writing a Kalikow decomposition for age-dependent Hawkes processes with a hard refractory period, for the purpose of perfect simulation of its stationary version. This process was first introduced in [8], and no Kalikow decomposition has been established for this process, even in a conditional framework. In our setting, the stochastic intensity of an age-dependent Hawkes process with hard refractory of length $\delta > 0$ can be written as follows. For any $i \in \mathbf{I}$ and $x \in \mathcal{X}$,

$$\phi^i(x) = \psi^i \left(\sum_{j \in \mathbf{I}} \int_{-\infty}^0 h_j^i(-s) dx_s^j \right) \mathbb{1}_{a^i(x) > \delta}. \quad (3.7)$$

In the above formula, the age of the process is defined by

$$a^i(x) = - \sup \{t_k^i \in x : t_k^i < 0\},$$

that is, the delay since the last point. If there is no such last point, we simply put $a^i(x) = +\infty$. The function ψ^i is called the rate function; it is assumed to be positive and at least

locally Lipschitz-continuous. The locally integrable functions h_j^i model the interactions as in the framework of the classical Hawkes process. In this setting we allow for an infinite number of components, that is, \mathbf{I} might be infinite. The set \mathcal{V}^i remains defined by (3.5) but the set \mathcal{Y} is not the same. Indeed, by definition of the stochastic intensity (3.7), we observe that the distance between any two consecutive jumps has to be larger than δ . This observation leads us to consider the subspace

$$\mathcal{Y} = \mathcal{X}^{>\delta},$$

which was introduced in (2.1).

We also change the family of neighborhoods. Now we envision a family of nested and increasing neighborhoods. More specifically, from a nested family of subsets of \mathbf{I} given by $(\omega_k^i)_{k \in \mathbb{N}^*}$, with $\omega_1^i = \{i\}$, $\omega_k^i \subset \omega_{k+1}^i$ and $\cup_{k \geq 1} \omega_k^i = \mathbf{I}$, we can design a nested family of neighborhoods $\mathbf{V}_{\text{nested}}^i$ by

$$\mathbf{V}_{\text{nested}}^i = \{v_k^i = \omega_k^i \times [-k\delta, 0) \text{ for } k \in \mathbb{N}^*\}.$$

Note that this family does not include the empty set. This is due to the presence of the hard refractory period. As we will see in Section 4, the backward steps in the perfect simulation algorithm do not end only because of the probability of picking the empty set but also because some neighborhoods might just be empty (in the sense that no points appear in them). This is the main difference compared to similar perfect simulation algorithms that exist in discrete time; see e.g. [19] or [27].

For this family, we can prove the following result as a corollary of Proposition 3.1.

Corollary 3.3. *Assume that for all $i, j \in \mathbf{I}$, $h_j^i(\cdot)$ is a non-negative, non-increasing L^1 -function and that for every i*

$$\sum_{j \in \mathbf{I}} \|h_j^i\|_1 < \infty \quad \text{and} \quad \sum_{j \in \mathbf{I}} h_j^i(0) < \infty.$$

Then the multivariate age-dependent Hawkes process with hard refractory period defined by (3.7) admits a Kalikow decomposition with respect to $\mathbf{V}^i = \mathbf{V}_{\text{nested}}^i$ and $\mathcal{Y} = \mathcal{X}^{>\delta}$, under the following conditions.

- (i) *If, for every i , $\psi^i(\cdot)$ is an increasing, non-negative continuous function, then the Kalikow decomposition of Proposition 3.1(i) applies with*

$$\Delta_1^i(x) = \Delta_{v_1^i}^i(x) = \psi^i(0) \mathbf{1}_{a^i(x) > \delta},$$

and for all $k \geq 2$,

$$\begin{aligned} \Delta_k^i(x) &= \Delta_{v_k^i}^i(x) \\ &= \psi^i \left(\sum_{j \in \omega_k^i} \int_{-k\delta}^0 h_j^i(-s) dx_s^j \right) \mathbf{1}_{a^i(x) > \delta} - \psi^i \left(\sum_{j \in \omega_{k-1}^i} \int_{-(k-1)\delta}^0 h_j^i(-s) dx_s^j \right) \mathbf{1}_{a^i(x) > \delta}. \end{aligned}$$

- (ii) *If, in addition, $\psi^i(\cdot)$ is L -Lipschitz, the Kalikow decomposition of Proposition 3.1(ii) applies for the choices*

$$\Gamma_1^i = \Gamma_{v_1^i}^i \geq \bar{\Gamma}_1^i := \psi^i(0),$$

and for all $k \geq 2$,

$$\Gamma_k^i = \Gamma_{v_k^i}^i \geq \bar{\Gamma}_k^i := L \left[\sum_{j \in \omega_k^i \setminus \omega_{k-1}^i} (h_j^i(0) + \delta^{-1} \|h_j^i\|_1) + \sum_{j \in \omega_{k-1}^i} h_j^i((k-1)\delta) \right],$$

as long as $\sum_{k=1}^\infty \Gamma_k^i < \infty$.

In particular, since

$$\sum_{k=1}^\infty \bar{\Gamma}_k^i \leq \psi^i(0) + 2L \left[\sum_{j \in \mathbf{I}} h_j^i(0) + \delta^{-1} \sum_{j \in \mathbf{I}} \|h_j^i\|_1 \right],$$

we find that

$$\Gamma_k^i = \bar{\Gamma}_k^i$$

is a valid choice.

Proof. For part (i) of the result, by assumption, $(\Delta_k^i(x))_{k \geq 1}$ is well-defined, non-negative and cylindrical on $v_k^i = \omega_k^i \times [-k\delta, 0)$, since the family of neighborhoods is nested. Let

$$r_n^i(x) := \sum_{k=1}^n \Delta_k^i(x) = \psi^i \left(\sum_{j \in \omega_n^i} \int_{-n\delta}^0 h_j^i(-s) dx_s^j \right) \mathbb{1}_{\alpha^i(x) > \delta}$$

and let us show that $r_n^i(x) \rightarrow \phi^i(x)$ when $n \rightarrow \infty$. Consider the inner term of the parenthesis,

$$\sum_{j \in \omega_n^i} \int_{-n\delta}^0 h_j^i(-s) dx_s^j = \int_{\mathbf{I} \times \mathbb{R}^-} h_j^i(-s) \mathbb{1}_{(j,s) \in v_n^i} dx_s^j d\kappa_j,$$

where we let $d\kappa$ denote the counting measure on the discrete set \mathbf{I} .

We note that $(h_j^i(-s) \mathbb{1}_{(j,s) \in v_n^i})_{n \in \mathbb{Z}}$ is a non-negative and non-decreasing sequence in n . In addition, it converges to $h_j^i(-s) \mathbb{1}_{(j,s) \in \mathbf{I} \times (-\infty, 0)}$ as $n \rightarrow \infty$. Moreover, since $\psi^i(\cdot)$ is a continuous and increasing function, the monotone convergence theorem for Lebesgue Stieltjes measures implies that $r_n^i(x) \rightarrow \phi^i(x)$ as $n \rightarrow \infty$. As a consequence, Proposition 3.1(i) applies.

For part (ii), ψ^i is L -Lipschitz. Hence we have

$$\Delta_i^k(x) \leq L \times \left[\sum_{j \in \omega_k^i \setminus \omega_{k-1}^i} \int_{-k\delta}^0 h_j^i(-s) dx_s^j + \sum_{j \in \omega_{k-1}^i} \int_{-k\delta}^{-(k-1)\delta} h_j^i(-s) dx_s^j \right]. \tag{3.8}$$

So let us fix $0 \leq k < l, j \in \mathbf{I}, x \in \mathcal{X}^{>\delta}$ and $t \in \mathbb{R}$ and let us concentrate first on upper-bounding $\int_{t-l\delta}^{t-k\delta} h_j^i(t-s) dx_s^j$ by adapting Lemma 2.4 of [29]. For any $\epsilon > 0$, we have

$$\begin{aligned} \int_{[t-l\delta, t-k\delta-\epsilon]} h_j^i(t-s) dx_s^j &= \sum_{k \leq m < l} \int_{[t-(m+1)\delta-\epsilon, t-m\delta-\epsilon]} h_j^i(t-s) dx_s^j \\ &\leq \sum_{k \leq m < l} h_j^i(m\delta + \epsilon), \end{aligned}$$

because h_j^i is non-increasing and because there is at most one jump in the interval of length δ . Therefore

$$\begin{aligned} \int_{[t-l\delta, t-k\delta)} h_j^i(t-s) dx_s^j &= \lim_{\epsilon \downarrow 0} \int_{[t-l\delta, t-k\delta-\epsilon]} h_j^i(t-s) dx_s^j \\ &\leq \lim_{\epsilon \downarrow 0} \sum_{k \leq m < l} h_j^i(m\delta + \epsilon) \\ &\leq \sum_{k \leq m < l} h_j^i(m\delta), \end{aligned}$$

by monotone convergence and the fact that $h_j^i(\cdot)$ is a decreasing function.

Going back to (3.8), we therefore have

$$\int_{-k\delta}^{-(k-1)\delta} h_j^i(-s) dx_s^j \leq h_j^i((k-1)\delta))$$

and

$$\int_{-k\delta}^0 h_j^i(-s) dx_s^j \leq \sum_{0 \leq m \leq k-1} h_j^i(m\delta) \leq h_j^i(0) + \delta^{-1} \int_0^{k-1} h_j^i(t) dt \leq h_j^i(0) + \delta^{-1} \|h_j^i\|_1.$$

This shows that if $\Gamma_k^i \geq \bar{\Gamma}_k^i$, we are indeed upper-bounding $\sup_{x \in \mathcal{X}^{>\delta}} \Delta_k^i(x)$. The last upper bound on $\sum_k \bar{\Gamma}_k^i$ is done in a similar way. \square

Remark 3.2. In particular, this proves that $\Delta_k^i(x) \rightarrow 0$ when $k \rightarrow \infty$ for fixed $x \in \mathcal{X}^{>\delta}$, and that this convergence is uniform on $\mathcal{X}^{>\delta}$ if ψ^i is Lipschitz.

Remark 3.3. It is straightforward to see that

$$\inf_{\substack{v_k^i \\ z \stackrel{\delta}{=} x}} \psi^i \left(\sum_{j \in \mathbf{I}} \int_{-\infty}^0 h_j^i(-s) dz_s^j \right) \mathbb{1}_{a^i(x) > \delta} = \psi^i \left(\sum_{j \in \omega_k^i} \int_{-k\delta}^0 h_j^i(-s) dx_s^j \right) \mathbb{1}_{a^i(x) > \delta},$$

because of the monotonicity property of ψ^i and since $h_j^i \geq 0$. Indeed, the minimizing configuration is obtained by having no points outside v_k^i . Hence, for $k \geq 2$, we observe that

$$\begin{aligned} &\Delta_k^i(x) \\ &= \inf_{\substack{v_k^i \\ z \stackrel{\delta}{=} x}} \left[\psi^i \left(\sum_{j \in \mathbf{I}} \int_{-\infty}^0 h_j^i(-s) dz_s^j \right) \mathbb{1}_{a^i(x) > \delta} \right] - \inf_{\substack{v_{k-1}^i \\ z \stackrel{\delta}{=} x}} \left[\psi^i \left(\sum_{j \in \mathbf{I}} \int_{-\infty}^0 h_j^i(-s) dz_s^j \right) \mathbb{1}_{a^i(x) > \delta} \right]. \end{aligned}$$

The above prescription corresponds to the classical method of obtaining a Kalikow decomposition in discrete time, as discussed in [13] and [19] (see also Corollary 2.1).

Remark 3.4. Notice that it is not possible to let δ tend to 0 and to recover Corollary 3.3 for linear or even general non-linear Hawkes processes having Lipschitz-continuous rate function, since typically the bounds in Corollary 3.3 are exploding.

3.3. Non-linear Hawkes process with an analytic rate function

In the above examples we have shown how one can easily derive the Kalikow decomposition if the intensity is already in the shape of a sum, as is the case for the linear Hawkes process, with an atomic family of neighborhoods. In this case the complexity of all neighborhoods is very small. In the case of the age-dependent Hawkes process, thanks to the refractory period and the monotonicity and continuity properties of the underlying functions, we have been able to derive a Kalikow decomposition as well, but with respect to a more complex family of neighborhoods which is nested. Moreover, we have shown that under Lipschitz properties, the intensities are bounded. In both cases the processes were non-exploding, that is, they only have a finite number of jumps within finite time intervals, almost surely.

Now we want to reach much more erratic processes, which can even explode in finite time. To do so, we cannot keep the refractory period, and we consider non-linear Hawkes processes $Z = (Z^i)_{i \in \mathbf{I}}$ with intensity given by

$$\phi^i(x) = \psi^i \left(\sum_{j \in \mathbf{I}} \int_{-\infty}^0 h_j^i(-s) dx_s^j \right) \tag{3.9}$$

for any $x \in \mathcal{X}$, where $\psi^i: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ are measurable functions and where $h_j^i: \mathbb{R} \rightarrow \mathbb{R}_+$ belong to L^1 .

Brémaud and Massoulié [4] proved that if ψ^i is Lipschitz, then – under suitable additional assumptions on $\|h_j^i\|_1$ – a stationary version exists. However, in the case where ψ^i is only locally Lipschitz, the existence of a non-exploding or even a stationary solution is not guaranteed. For example, when choosing $\psi^i(x) = e^x$, then the process may explode in finite time with strictly positive probability; see [5].

We will use analytical properties of ψ^i to perform a Taylor expansion of ψ^i . To do so, we consider a family of neighborhoods which are iterated tensor products of \mathbf{V}_{atom} . For fixed ϵ , recall that an atomic neighborhood (see (3.6)) is such that $w_{j,n} = \{j\} \times [-n\epsilon, -(n-1)\epsilon]$, for $j \in \mathbf{I}$ and $n \in \mathbb{N}^*$. A neighborhood of order k is then a union of k atomic neighborhoods. For $\alpha_1 = (j_1, n_1), \dots, \alpha_k = (j_k, n_k)$, we put

$$v_{\alpha_1; k} = \bigcup_{k=1}^n w_{j_k, n_k}.$$

The family of all possible such unions is denoted by $\mathbf{V}_{\otimes k}$, with the convention that $\mathbf{V}_{\otimes 0} = \{\emptyset\}$.

The Taylor family of neighborhoods is then defined by

$$\mathbf{V}_{\text{Taylor}} = \bigcup_{k=0}^{\infty} \mathbf{V}_{\otimes k}.$$

Note that if, for instance, some of the α_i are equal, then the neighborhood $v_{\alpha_1; k}$ collapses on a smaller union of less than k intervals and it is also possible that two neighborhoods with different indices are in fact equal. We do not simplify the family and remove redundancies. It is important that neighborhoods with different indices are considered different in order to define properly the probability λ^i .

We are limited by the radius of convergence of the function ψ^i , say K . This is why we are working within the space

$$\mathcal{Y} = \mathcal{X}^{<K} = \left\{ x \in \mathcal{X} : \sup_{i \in \mathbf{I}} \sum_{j \in \mathbf{I}} \int_{-\infty}^0 h_j^i(-s) dx_s^j < K \right\}.$$

Corollary 3.4. *Let us assume that for any $i, j \in \mathbf{I}$, the function $h_j^i(\cdot)$ is non-negative and that for every i ,*

$$\sup_i \sum_{j \in \mathbf{I}} h_j^i(t) < \infty.$$

Let us also assume that for every $i \in \mathbf{I}$, $\psi^i(\cdot)$ is an analytic function on \mathbb{R} with radius of convergence $K > 0$ around 0, such that its derivative of order k , $[\psi^i]^{(k)}(0)$, is non-negative for all $k \geq 0$. Then the multivariate non-linear Hawkes process defined by (3.9) admits the Kalikow decomposition of Proposition 3.1(i) with respect to $\mathbf{V}^i = \mathbf{V}_{\text{Taylor}}$ for all i and with respect to the subspace $\mathcal{Y} = \mathcal{X}^{<K}$. Specifically, we have the following.

- If $v = \emptyset$, $\Delta_v^i(x) = \psi^i(0)$.
- If $v = v_{\alpha_1:k}$, for $\alpha_1 = (j_1, n_1), \dots, \alpha_k = (j_k, n_k)$, then

$$\Delta_v^i(x) = \frac{[\psi^i]^{(k)}(0)}{k!} a_{\alpha_1}(x) \cdot \dots \cdot a_{\alpha_k}(x),$$

where for $\alpha = (j, n)$,

$$a_\alpha(x) := \int_{-(n+1)\epsilon}^{-n\epsilon} h_j^i(-s) dx_s^j.$$

Proof. The condition ensures that the intensities are at least locally defined if, starting from an empty past, one first point occurs. The proof of the convergence is then straightforward and consists in using the analytic properties of ψ^i . □

Note that the linear Hawkes process is a particular case of this corollary with $\psi^i(u) = \mu^i + u$ and $K = \infty$. But the above strategy can also be applied to the exponential Hawkes process, $\psi^i(u) = \exp(u)$, $K = \infty$ and $[\psi^i]^{(k)}(0) = 1$ for all k , which is only locally Lipschitz. One can also use it for other functions, e.g. $\psi^i(u) = \text{ch}(u) = (e^u + e^{-u})/2$ with $K = \infty$, for which even derivatives $[\psi^i]^{(2k)}(0) = 1$ and for which the other derivatives are null.

3.4. Galves–Löcherbach process with saturation thresholds

We close this section with an example which is not directly related to Hawkes processes and for which, using the same arguments as above, we may establish a Kalikow decomposition as well. This is the model with saturation threshold, which has already been considered in [19]. More precisely, we put

$$\phi^i(x) = \psi^i \left(\sum_{j \in \mathbf{I}} (\beta_j^i Z^j((-a^i(x), 0))) \wedge K_j^i \right) \tag{3.10}$$

for $\psi^i: \mathbb{R} \rightarrow \mathbb{R}_+$ a Lipschitz-continuous non-decreasing rate function, $\beta_j^i \geq 0$ the weight of j on i , where we choose $\beta_i^i = 0$, and $K_j^i \geq 0$ the saturation threshold. In this framework, analogously to Corollary 3.3, we can prove the following

Corollary 3.5. *Consider the Galves–Löcherbach process having generic intensity ϕ^i given by (3.10), and suppose that*

$$\sup_{i \in \mathbf{I}} \sum_{j \in \mathbf{I}} K_j^i < \infty.$$

Then ϕ^i admits the Kalikow decomposition given by Proposition 3.1(i), with respect to the collection of neighborhood families $(\mathbf{V}^i)_{i \in \mathbf{I}}$, where for all i , $\mathbf{V}^i = \mathbf{V}_{\text{nested}}^i \cup \{v_0^i = \emptyset\}$, and with respect to the subspace $\mathcal{Y} = \mathcal{X}$, with $\Delta_0^i(x) = \psi^i(0)$ and for any $k \geq 1$,

$$\begin{aligned} \Delta_k^i(x) &= \Delta_{v_k^i}^i(x) \\ &= \psi^i \left(\sum_{j \in \omega_k^i} \left[\beta_j^i \int_{-(k\delta) \wedge a^i(x)}^0 dx_s^j \right] \wedge K_j^i \right) - \psi^i \left(\sum_{j \in \omega_{k-1}^i} \left[\beta_j^i \int_{-((k-1)\delta) \wedge a^i(x)}^0 dx_s^j \right] \wedge K_j^i \right), \end{aligned}$$

where $w_0^i = \emptyset$.

4. Simulation algorithms

In this section we present two types of algorithm corresponding to the two items of Proposition 3.1. First we present a simulation algorithm that simulates the time-homogeneous counting process Z for finite networks, starting from the empty past before time 0, up to some finite time, under the conditions of item (i) of Proposition 3.1. Second, we present a perfect simulation algorithm that simulates the process Z in its stationary regime, within a finite space–time window, under the conditions of item (ii) of Proposition 3.1.

4.1. Simulating forwards in time, starting from a fixed past

Suppose that we are in the situation of item (i) of Proposition 3.1, with \mathbf{I} finite, and that we wish to simulate the process until a certain $t_{\max} > 0$ starting from the empty past before time 0. We might never reach t_{\max} because the process might typically explode, so we introduce

$$\tau := t_{\max} \wedge \inf\{t \geq 0 : Z_t \notin \mathcal{Y}\} \wedge \inf\left\{t \geq 0 : \sum_i Z_t^i \geq N_{\max}\right\}$$

for some fixed deterministic number N_{\max} . We make the following assumption.

Assumption 4.1. For all $i \in \mathbf{I}$ and for all jump times T of the process Z , there exists $\tilde{\Gamma}_T^i := \tilde{\Gamma}^i(x_T)$ which depends only on the configuration up to time T (including T), such that

$$\sup_{t \geq T} \sup_v \phi_v^i(x_t^{\leftarrow t}) \mathbf{1}_{x_t \cap (T, t) = \emptyset} \leq \tilde{\Gamma}_T^i < \infty.$$

Intuitively, as soon as no point is produced, for any chosen neighborhood v , the intensity function restricted to this neighborhood is bounded by a predictable function. That allows us, at the same time, to construct the dominating intensity and do a thinning step.

Example 4.1. The above assumption is satisfied for any non-linear Hawkes process with non-decreasing rate function and with decreasing interaction functions h_j^i , in the framework of Sections 3.1 or 3.3 above. For instance, in the framework of Section 3.1, recalling that $\lambda^i(v)$ denotes the probability of choosing the neighborhood $v = \{j\} \times [-n\epsilon, -n\epsilon + \epsilon)$, we put

$$\tilde{\Gamma}_T^i = \sup_{j, n} \sup_J \frac{h_j^i(0)Z^j(J)}{\lambda^i(v)}, \tag{4.1}$$

where $J \subset [0, T]$ is any interval of length ϵ .

Let us now describe how we simulate forwards in time, starting from the empty past before time 0, up to some maximal number of points $N_{\max} > 0$. In the following we will add subscript T to $\tilde{\Gamma}^i$ to distinguish it from the old version.

Step 0. We initialize the set of points with $X = \mathbf{0} \in \mathcal{X}$, where $\mathbf{0} \in \mathcal{X}$ designs the configuration having no points, and $T = N = 0$.

While $T \leq \tau$ and $N \leq N_{\max}$, do the following.

Step 1. For each $i \in \mathbf{I}$, compute

$$\tilde{\Gamma}_T^i \geq \sup_{t \geq T} \sup_v \phi_v^i(X^{\leftarrow t}) \mathbf{1}_{X \cap (T, t) = \emptyset}.$$

Step 2. Simulate the next jump

$$T \leftarrow T + \text{Exp} \left(\sum_{i \in \mathbf{I}} \tilde{\Gamma}_T^i \right).$$

If $T > \tau$ stop, else choose the associated index i with probability

$$\mathbb{P}(I = i) = \frac{\tilde{\Gamma}_T^i}{\sum_{j \in \mathbf{I}} \tilde{\Gamma}_T^j}.$$

Step 3. Choose the associated interaction neighborhood $V_T = v$ with probability $\lambda^i(v)$.

Step 4. Accept T as a jump of index i with probability $(\phi_v^i(X^{\leftarrow T})) / \tilde{\Gamma}_T^i$ and add (i, T) to X in this case. Put $N \leftarrow N + 1$. If the point is rejected, do nothing. Go back to Step 1.

Remark 4.1. The bound $\tilde{\Gamma}_T^i$, computed in Step 1, is random since it depends on the configuration X that has been simulated so far.

Corollary 4.1. (Corollary of Proposition 2.1.) *The previous algorithm simulates a time-homogeneous counting process Z of generic intensity given by item (i) of Proposition 3.1.*

Proof. Note that the random variable $\tilde{\Gamma}_T^i$ is a bound on the generic intensity ϕ^i in the absence of the appearance of a new point in the future after time T . Hence the ‘new’ T (see also Section 2.3) computed at Step 2 can be seen as the abscissa of a point of an hidden bivariate Poisson process in a band of height $\sum_{i \in \mathbf{I}} \tilde{\Gamma}_T^i$. The association of a particular index i is quite usual (see e.g. Ogata’s algorithm [26]), and it means that if the chosen index is i then the point T can also be seen as the next point of the bivariate Poisson process π^i discussed in Section 2.3 in the band of height $\tilde{\Gamma}_T^i$. Since this is an upper bound on ϕ^i , the thinning procedure will not consider points of π^i outside this band, at least until a new point is accepted. The association of a neighborhood in Step 3 is therefore consistent with the three-dimensional Poisson process of Proposition 2.1, and the points that are accepted at Step 4 are the ones accepted in Proposition 2.1. Therefore this algorithm indeed simulates points with the desired intensities as long as the process stays in \mathcal{Y} and the Kalikow decomposition holds. \square

Remark 4.2. Note that this algorithm will indeed save computational time if both the computation of $\tilde{\Gamma}_T^i$ and the verification that one stays in \mathcal{Y} is easy. In the particular example of linear Hawkes processes (Corollary 3.2), with decreasing h_j^i , and bounded support on $[0, \epsilon]$, say, one can take $\tilde{\Gamma}_T^i$ as defined in (4.1) and we know that the process will stay in \mathcal{Y} with probability 1.

Further algorithmic work, beyond the scope of the present article, might include a bespoke design of Γ_T^i to save computational time.

4.2. Perfect simulation of the process in its stationary regime

Throughout this section we suppose that we are in the situation of item (ii) of Proposition 3.1, that is, the generic intensities ϕ^i are upper-bounded by deterministic constants Γ^i and the subspace \mathcal{Y} is invariant under the dynamics, that is, $Z_t \in \mathcal{Y}$ implies $Z_{t+s} \in \mathcal{Y}$ for all $s \geq 0$. This, for example, is the case for the age-dependent Hawkes process with hard refractory period considered in Section 3.2 above. In particular, in this case it is possible to restrict the dynamic to \mathcal{Y} , and in what follows we will show how to simulate from the unique stationary version of the process within this restricted state space \mathcal{Y} .

In what follows, we propose simulating, in its stationary regime, the process Z^i for a fixed $i \in \mathbf{I}$, on an interval $[0, t_{\max}]$, for some fixed $t_{\max} > 0$. Our algorithm is a modification of the method described in [28], which works in the case where all the Γ^i are equal. The procedure consists of backward and forward steps. In the backward steps, thanks to the Kalikow decomposition, we create a set of ancestors, which is a list of all the points that might influence the point under consideration. On the other hand, in the forward steps, where we go forwards in time, by using the thinning method [26] in its refined version stated in Theorem 2.1, we give the decision (acceptance or rejection) for each visited point based on its neighborhood, until the state of all considered points is decided.

The idea of relying on such a two-step procedure is not new and has already been proposed in the literature, even in a continuous-time setting (see e.g. [12], [14], and [15]), where such an approach is used to simulate from infinite-range Gibbs measures and/or from the steady state of interacting particle systems, relying on a decomposition of the spin flip rates as a convex combination of local flip rates. The main difference of our present approach with respect to these results is twofold. Firstly, in these articles, only spatial interactions are considered, whereas we have to decompose both with respect to the spatial interactions and the history in the present article. And secondly and more importantly, in all these articles the authors manage to go back to transition probabilities and then establish a Kalikow decomposition for them. So somehow this means that we are back in the framework of discrete-time processes as in [22], where these ideas were introduced. Such a discrete-time approach is used in [19]. The idea of decomposing the intensities directly rather than going back to probabilities, and to decompose both with respect to time, i.e. history, and space, i.e. the interactions, is – at least to our knowledge – completely new.

4.2.1. Backward procedure. Recall that according to item (ii) of Proposition 3.1, all functions ϕ^i and ϕ_v^i are bounded by a fixed deterministic constant Γ^i .

For the sake of understandability from a theoretical point of view, we will assume that the N^i are independent homogeneous Poisson processes on the real line with intensity Γ^i and that these points are fixed before one starts the algorithm. Indeed, following what was said in Proposition 2.1, we will recover the points of Z^i as a thinning of a three-dimensional Poisson process. The fact that the intensities are bounded allows us to thin only in a band of height Γ^i , and that is why we are going to thin N^i into Z^i .

In practice, of course, infinite knowledge of the N^i is not possible, and Steps 1' and 2' are there to explain concretely how this is done in practice.

Step 0. Fix i , set the initial time to be $T_0 = 0$.

Step 1. Take T to be the first point of N^i after 0. This is the first possible jump of Z^i after T_0 .

Step 1'. Simulate this T as

$$T \leftarrow T_0 + \text{Exp}(\Gamma^i).$$

In particular, this implies that the underlying N^i is empty on $[T_0, T)$.

If $T > t_{\max}$, stop.

Step 2. Independently of anything else, pick a random neighborhood V_T^i of (i, T) according to the distribution $(\lambda^i(v))_{v \in \mathbf{V}^i}$ given in (3.2), that is,

$$\mathbb{P}(V_T^i = v) = \lambda_i(v).$$

On $\{V_T^i = v\}$, consider the shift of this neighborhood v by T defined by

$$v^{\rightarrow T} := \{(j, u + T) : (j, u) \in v\}$$

and, for any $j \in \mathbf{I}$, the projection to the second coordinate of $v^{\rightarrow T}$ by

$$p_j(v^{\rightarrow T}) := \{t \in \mathbb{R} : (j, t) \in v^{\rightarrow T}\}.$$

Note that if, for some j , $(j, t) \notin v^{\rightarrow T}$ for all t , then $p_j(v^{\rightarrow T}) = \emptyset$. Finally, we look for the set of points of N that might directly influence the decision of acceptance/rejection of T :

$$\mathcal{C}_1^{(i,T)} = \bigcup_{j \in \mathbf{I}} \{(j, t) : t \in p_j(v^{\rightarrow T}) \text{ is a jump of } N^j\}.$$

Step 2'. Simulate independent Poisson processes in $v^{\rightarrow T}$, that is, for each $j \in \mathbf{I}$, we simulate the points of N^j independently of anything else as a Poisson process of intensity Γ^j on $p_j(v^{\rightarrow T})$, to identify $\mathcal{C}_1^{(i,T)}$ as before.

It is possible that at this point in the algorithm (especially in the iteration below), $v^{\rightarrow T}$ intersects with neighborhoods that have already been picked. In this case, simulate only on the portion of the neighborhood that has never been visited before. Note in particular that we assume the set $\{i\} \times [T_0, T)$ to be an already visited region, and that within this region there are no points, by definition of T .

Step 3. Recursively, we define the n th set of ancestors of (i, T) by

$$\mathcal{C}_n^{(i,T)} = \bigcup_{(j,s) \in \mathcal{C}_{n-1}^{(i,T)}} \mathcal{C}_1^{(j,s)} \setminus (\mathcal{C}_1^{(i,T)} \cup \dots \cup \mathcal{C}_{n-1}^{(i,T)}),$$

by performing Step 2 or Step 2' for each $(j, s) \in \mathcal{C}_{n-1}^{(i,T)}$.

Note that $\mathcal{C}_n^{(i,T)}$ exactly corresponds to the points that are really simulated as new in the iterated Step 2'.

We denote

$$N^{(i,T)} = \inf\{n : \mathcal{C}_n^{(i,T)} = \emptyset\},$$

where $\inf \emptyset := +\infty$. The backward scheme stops when $N^{(i,T)} < \infty$, and below we give sufficient conditions guaranteeing this fact (see Proposition 4.1 below). In this case we say that the total clan of ancestors of (i, T) is given by

$$\mathcal{C}^{(i,T)} = \bigcup_{k=1}^{N^{(i,T)}} \mathcal{C}_k^{(i,T)}.$$

In what follows we also consider the associated interaction support given by

$$\mathcal{V}^{(i,T)} = \bigcup_{(j,t) \in \mathcal{C}^{(i,T)}} (V_t^j)^{\rightarrow t}$$

and we put

$$T^{(i,T)} := T - \inf\{s : \exists j \in \mathbf{I} : (j, s) \in \mathcal{V}^{(i,T)}\}, \quad (4.2)$$

which is the total time the backward steps need to look back in the past.

Remark 4.3. Let us emphasize that $\lambda_i(\emptyset)$ does not need to be strictly positive in order to guarantee that $N^{(i,T)} < \infty$. If $\lambda_i(\emptyset) = 0$, at every step of the backward scheme we always need to simulate a Poisson process in a non-empty neighborhood. However, if there is no point simulated in such a neighborhood, then we do not add any points to the clan of ancestors, and if this happens sufficiently often, then the backward step ends as well. This is one of the main advantages of our approach and a major difference with respect to [13] and [19]. In Proposition 4.1 below we give sufficient conditions implying that the algorithm stops after a finite number of steps almost surely.

Remark 4.4. Steps 1' and 2' are consistent with recreating the processes N^i in the neighborhoods of interest, instead of taking them for granted beforehand. In particular, this is the reason why the algorithm is very careful when dealing with overlapping neighborhoods not to simulate the same process twice on the same portion of the space.

4.2.2. *Forward procedure* Supposing that $N^{(i,T)} < \infty$, we now use a forward procedure to accept or reject recursively each point in $\mathcal{C}^{(i,T)}$, until the status of T is decided.

We start with the point $(j, s) \in \mathcal{C}^{(i,T)}$, which is the smallest in time, so that its associated neighborhood is either empty ($v = \emptyset$) or non-empty but without any point of the Poisson process in it.

Step 1. Accept (j, s) with probability $\phi_{V_s^j}^j(X_s^{\leftarrow s})/\Gamma^j$, where V_s^j is the neighborhood of (j, s) .

Step 2. Move to the next point of $\mathcal{C}^{(i,T)}$ in increasing time order. Repeat Steps 1 and 2 until the status of T is determined.

Update step. To simulate on $[0, t_{\max}]$, go back to Step 0 of the backward procedure and replace the starting time of the initial step T_0 with T . Repeat the backward and forward procedures until $T > t_{\max}$.

Remark 4.5. If one wants to adapt the algorithm to the simulation of a generic finite subset F of $\mathbf{I} \times \mathbb{R}$, it is sufficient to shift everything to the smallest time in F (which will be the initial T_0) and to repeat the process on all the i in F . Again in Step 2', we need to be careful to simulate only on new parts of $\mathbf{I} \times \mathbb{R}$ and not parts that have been already discovered and where one of the N^i has already been simulated.

4.2.3. *Why the backward procedure ends after a finite number of steps.* Our procedure only works if $N^{(i,T)} < \infty$ almost surely. In what follows we give sufficient conditions implying this. To do so, we compare our process to a spatiotemporal branching process. Let us define this more mathematically.

The initial directed graph of ancestors \mathcal{T} . We start with the ancestor which is the point (i, T) constituting generation 0. All points belonging to $\mathcal{C}_n^{(i,T)}$ are called points of the n th generation.

For each point (j, t) belonging to generation n , all elements of

$$\mathcal{C}_1^{(j,t)} = \bigcup_{k \in \mathbf{I}} \{(k, s) : s \in p_k((V_t^j)^{\rightarrow t}) \text{ is a jump of } N^k\}$$

define the children of the point (j, t) . The choices of the children of (j, t) and (j', t') for any two distinct elements of generation n are not necessarily independent, since the associated neighborhoods might overlap, that is, we may have $V_t^j \cap V_{t'}^{j'} \neq \emptyset$.

Note that \mathcal{T} is not a tree, precisely because of this potential overlapping: two different parents might have the same child.

The dominating branching process $\tilde{\mathcal{T}}$. To obtain sufficient conditions implying that $N^{(i,T)} < \infty$, we therefore construct a dominating spatiotemporal branching process starting from the same ancestor (i, T) , where the choices of children are independent. To go from generation n to the next generation $n + 1$, to any point (j, t) belonging to generation n we associate the same children $\mathcal{C}_1^{(j,t)}$, as before, whenever the chosen neighborhood V_t^j does not overlap with parts of $\mathbf{I} \times \mathbb{R}$ where we have already simulated in previous steps. However, if there are parts of the neighborhood V_t^j that intersect with parts already visited, we simulate – independently of anything else – on the whole neighborhood. By doing so, we make the number of children larger, but for each point, these choices are independent of anything else. If this dominating branching process goes extinct in finite time, then so does the original process $\mathcal{C}_n^{(i,T)}$, and the backward part of the algorithm terminates.

To formulate a sufficient criterion that implies the almost sure extinction of the dominating branching process $\tilde{\mathcal{T}}$, let us denote the product measure P on $\mathbf{I} \times \mathbb{R}$, defined on the Borel subsets of $\mathbf{I} \times \mathbb{R}$, as follows:

$$P(J \times A) := \sum_{j \in J} \Gamma^j \mu(A)$$

for any $J \subset \mathbf{I}$, where A is a Borel subset of \mathbb{R} and where μ is the Lebesgue measure. The following proposition is already proved in [28] in a particular case where all the Γ^j are equal.

Proposition 4.1. *If*

$$\sup_{i \in \mathbf{I}} \sum_{v \in \mathbf{V}^i} P(v) \lambda^i(v) =: \gamma < 1, \tag{4.3}$$

then the backward steps in the perfect simulation algorithm terminate almost surely in finite time.

Proof. For any neighborhood v , we have

$$\sum_{j \in v} \mathbb{E}(\text{card}\{t : t \in p_j(v) \text{ is a jump of } N^j\}) = \sum_{j \in v} \Gamma^j \mu(p_j(v)) = P(v).$$

This implies that $\sum_{v \in \mathbf{V}^i} P(v) \lambda^i(v)$ is the mean number of children issued from one point of type i . Then condition (4.3) implies that the mean number of children is less than one in each step, which is the classical sub-criticality condition for branching processes; see [1]. The result then follows from the fact that according to the above discussion, we can dominate $L_n^{(i,T)} := \text{card}(\mathcal{C}_n^{(i,T)})$ by a classical Galton–Watson branching process having offspring mean $\gamma < 1$ which goes extinct in finite time almost surely. \square

4.2.4. Why we sample from the stationary distribution.

Theorem 4.1. *Suppose we are in the situation of item (ii) of Proposition 3.1 and the subspace \mathcal{Y} is invariant under the dynamics, that is, the configuration associated to Z, X , satisfies that $X_t^{\leftarrow t} \in \mathcal{Y}$ implies $X_{t+s}^{\leftarrow t+s} \in \mathcal{Y}$, for all positive s . If (4.3) holds, then the process possesses a unique stationary distribution in restriction to \mathcal{Y} , and the accepted points of the forward procedure yield a perfect sample of this stationary distribution within the space–time window $\{i\} \times [0, t_{\max}]$.*

Proof. In this proof we adapt the ideas of the proof of Theorem 2 in [14] to the present framework. For any finite subset F of $\mathbf{I} \times \mathbb{R}$ the backward and forward procedures produce a sample of a point process within the space–time window F , and we write μ_F for the law of the output. By construction, the family of probability laws $\{\mu_F, F \subset \mathbf{I} \times \mathbb{R} \text{ finite}\}$ is a consistent family of finite-dimensional distributions. Hence there exists a unique probability measure μ on $(\mathcal{X}_\infty, \mathcal{B}(\mathcal{X}_\infty))$ such that μ_F is the projection onto F of μ , for any fixed finite set $F \subset \mathbf{I} \times \mathbb{R}$.

We show that μ is the unique stationary distribution of the process Z within \mathcal{Y} . In order to do so, we use a slight modification of our algorithm in order to construct Z starting from some fixed past, say $x \in \mathcal{Y}$, before time 0. The modification is defined as follows.

We fix $t, t_{\max} > 0$ and put $F = \{i\} \times [t, t + t_{\max}]$. Recall that the original backward procedure relies on the *a priori* realization of all the Poisson processes N^i on $(-\infty, t + t_{\max}]$. In our modified procedure we replace, for any $i \in \mathbf{I}$, the points of N^i within $(-\infty, 0)$ by those of x^i , where $x = (x^i)_{i \in \mathbf{I}}$ is our initial condition.

Step 0. Put $T_0 = t$.

Step 1. Perform Steps 1–3 of the backward procedure, replacing the Poisson processes N^i in restriction to $(-\infty, 0)$ with the corresponding points of x , and stop this procedure at time

$$\tilde{N}^{(i,T)} = \inf\{n: \mathcal{C}_n^{(i,T)} \subset \mathbf{I} \times (-\infty, 0)\} \wedge N^{(i,T)}.$$

Indeed, on the set $\{\tilde{N}^{(i,T)} < N^{(i,T)}\}$, at this time, only points with negative times have to be considered, and all these points are determined by the initial condition x . In this modified version, when we stop the algorithm, the output set $\mathcal{C}_{\tilde{N}^{(i,T)}}^{(i,T)}$ might not be empty. This set is exactly the set of points before time 0 that have an influence on the acceptance or rejection of the point (i, T) .

Notice that the time

$$\inf\{n: \exists(j, t) \in \mathcal{C}_n^{(i,T)}: (V_t^j)^{\rightarrow t} \cap \mathbf{I} \times (-\infty, 0) \neq \emptyset\},$$

if it is finite, is the first time where the modified algorithm starts to be different from the original one. In particular, if the backward steps stop before reaching the negative sites, that is, if we are on the event $T^{(i,T)} \leq T$ (recall (4.2)), then

$$\mathcal{C}_{\tilde{N}^{(i,T)}}^{(i,T)} = \emptyset, \quad \tilde{N}^{(i,T)} = N^{(i,T)},$$

and the two procedures produce the same sets of points.

The forward procedure is performed as before, replacing the unknown points before time 0 with the fixed past configuration x .

Then the law of the set $\{(\tilde{\tau}_n^i)\}$ stated at the end of the modified algorithm 2 is the law of $Z_{[t, t+t_{\max}]}^i$, starting from the fixed past x before time 0. The output of the modified algorithm equals the output of the unmodified perfect simulation algorithm 2 if $T^{(i,T)} \leq T$.

We now give a formal argument proving that μ is indeed a stationary distribution of the process. Let $f: \mathcal{X}_\infty \rightarrow \mathbb{R}_+$ be a bounded measurable function that is cylindrical on $\{i\} \times [t, t + t_{\max}]$. Then

$$\begin{aligned} \mathbb{E}[f(Z_{[t, t+t_{\max}}^i) \mid Z_{\mathbb{R}_-} = x] &= \mathbb{E}[f(\{(\tilde{\tau}_n^i)\}, T^{(i,T)} \leq T) + \mathbb{E}[f(\{(\tilde{\tau}_n^i)\}, T^{(i,T)} > T) \\ &= \mathbb{E}[f(\{(\tau_n^i)\}, T^{(i,T)} \leq T) + \mathbb{E}[f(\{(\tau_n^i)\}, T^{(i,T)} > T], \end{aligned}$$

where (τ_n^i) is the output of the original perfect simulation algorithm.

But

$$\mathbb{E}[f(\{(\tilde{\tau}_n^i)\}, T^{(i,T)} > T) \leq \|f\|_\infty \mathbb{P}(T^{(i,T)} \geq t) \rightarrow 0 \quad \text{as } t \rightarrow \infty,$$

since finiteness of the tree implies the finiteness of $T^{(i,T)}$, and since by shift invariance the law of $T^{(i,T)}$ does not depend on T . Hence we obtain that

$$\lim_{t \rightarrow \infty} \mathbb{E}[f(Z_{[t, t+t_{\max}}^i) \mid Z_{\mathbb{R}_-} = x] = \mathbb{E}[f(Z_{[t, t+t_{\max}}^i)],$$

since $\mathbf{1}_{T^{(i,T)} \leq T} \rightarrow 1$ almost surely.

This implies that μ is a stationary distribution of the process. Replacing the initial condition x with any stationary initial condition, chosen within \mathcal{Y} , we finally also get uniqueness of the stationary distribution. □

4.2.5. The complexity of the algorithm Throughout this section we suppose once more that we are in the situation of item (ii) of Proposition 3.1 and that the subspace \mathcal{Y} is invariant under the dynamics, that is, $Z_t \in \mathcal{Y}$ implies $Z_{t+s} \in \mathcal{Y}$ for all $s \geq 0$. Our goal is to study the effect of the choice of the $(\lambda_i(v))_{v \in \mathcal{V}_i}$ on the number of points simulated by our algorithm. Until the end of this section, we make the following assumption.

Assumption 4.2.

- (i) *The index set $\mathbf{I} = \{1, \dots, N\}$ is finite.*
- (ii) *The sub-criticality assumption (4.3) is satisfied.*

Let us fix several notations that will be useful below. We let e_i denote the i th unit vector of \mathbb{R}^N , $\mathbf{1}$ is the vector $(1, 1, \dots, 1)^\top$, and μ still stands for the Lebesgue measure. Finally, by a positive vector we mean that all its components are positive.

We will rely on the multitype branching process \tilde{T} introduced at the beginning of Section 4.2.3, which is a space–time-valued process starting from the ancestor (i, T) in generation 0. We will refer to these points as ‘particles’ and we say the type of a particle is its associated index value i . Recall that in the definition of this branching process, each particle (j, t) belonging to generation n , independently of anything else, gives rise to offspring particles which are chosen as independent copies of

$$C_1^{(j,t)} = \bigcup_{k \in \mathbf{I}} \{(k, s) : s \in p_k((V_t^j)^{\rightarrow t}) \text{ is a jump of a Poisson process of intensity } \Gamma^k\}.$$

For $n \geq 1$, let $K^i(n)$ be the N -dimensional vector containing the numbers of offspring particles of different types belonging to the n th generation of the process issued from (i, T) , that is, $K^i(n) = (K_1^i(n), \dots, K_N^i(n))^\top$, where $K_k^i(n)$ is the number of particles of type k within generation n . We use the convention that $K^i(0) = e_i$ for every $1 \leq i \leq N$. For every

$j \in \{1, \dots, N\}$, let

$$X_j^i := K_j^i(1)$$

be the number of offspring particles of type j of the initial particle (i, T) . We have already seen that X_j^i is the cardinal of the points that a Poisson process of intensity Γ^j produces on $p_j(V_T^i)$, where V_T^i is the random neighborhood chosen for particle (i, T) . In other words, if we let \mathcal{P} denote the Poisson distribution, given that $V_T^i = v$, we have

$$\mathcal{L}(X_j^i \mid V_T^i = v) = \mathcal{P}(\Gamma^j \mu(p_j(v))). \tag{4.4}$$

Letting $X^i = K^1(1) \in \mathbb{R}^N$ denote the associated vector, we consider for any $\theta \in \mathbb{R}^N$ the log-Laplace transform of X^i given by

$$\phi_i(\theta) := \log \mathbb{E}_i(e^{\theta^\top X^i}),$$

where \mathbb{P}_i denotes the law of the branching process starting from a single ancestor having type i and \mathbb{E}_i is the corresponding expectation.

Moreover, we consider

$$W^i(n) = \sum_{k=0}^n K^i(k)$$

to be the total number of offspring particles within the first n generations. The log Laplace transform associated to the random vector $W^i(n)$ is given by

$$\Phi_i^{(n)}(\theta) := \log \mathbb{E}_i(e^{\theta^\top W^i(n)}),$$

where

$$\Phi^{(n)}(\theta) = (\Phi_1^{(n)}(\theta), \dots, \Phi_N^{(n)}(\theta))^\top.$$

Since we are working under the sub-criticality condition (4.3),

$$W^i(\infty) := \lim_{n \rightarrow \infty} W^i(n) \quad \text{and} \quad W^i = \sum_{j=1}^N W_j^i(\infty)$$

are well-defined and almost surely finite. In the above formula, $W_j^i(\infty)$ denotes the j th coordinate of the N -dimensional vector $W^i(\infty)$, that is, the total number of offspring of type j issued from one ancestor particle of type i . In particular, W^i is the total number of offspring particles issued from one ancestor particle of type i . We introduce the associated log-Laplace transforms

$$\phi_i(\theta) := \log \mathbb{E}_i(e^{\theta^\top W^i(\infty)}), \quad \Phi(\theta) = (\Phi_1(\theta), \Phi_2(\theta), \dots, \Phi_N(\theta))^\top.$$

We are now going to state an exponential inequality, inspired by Lemma 1 of [30]. To do so, let $\|\cdot\|_\infty$ be the ∞ -norm on \mathbb{R}^N , that is, for $x = (x_1, \dots, x_N)$, $\|x\|_\infty = \max\{|x_i|, 1 \leq i \leq N\}$, and let $B(0, r)$ denote the open balls having center $0 \in \mathbb{R}^N$ and radius r with respect to this norm. We have to introduce an additional assumption stating that the log-Laplace transform ϕ_i is finite within an open ball around 0 .

Assumption 4.3. *There exists $R > 0$ such that for all positive vectors θ belonging to $B(0, R)$ we have*

$$\sup_i \phi_i(\theta) = \sup_i \sum_{j=1}^N \log \left(\sum_{v \in \mathbf{V}^i} \lambda_i(v) \exp[(e^{\theta_j} - 1)\Gamma^j \mu(p_j(v))] \right) < \infty.$$

Proposition 4.2. *Given Assumptions 4.2 and 4.3, the following holds.*

(i) *There exists $r \in (0, R)$ such that for all $\theta \in B(0, r)$, $\Phi_i(\theta) < \infty$, and moreover*

$$\Phi(\theta) = \theta + \phi(\Phi(\theta)).$$

(ii) *In particular, for $0 \leq \vartheta < r$, there exists a constant c_0 that depends on M and i such that the following deviation inequality holds for the total number of offspring particles:*

$$\mathbb{P}_i(W^i > \mathbb{E}_i(W^i) + x) \leq c_0 e^{-\vartheta x} \tag{4.5}$$

for all $x > 0$.

Remark 4.6. Let M be the Jacobian matrix of $\phi(\cdot)$ at 0. Since

$$\phi_i(\theta) = \log \mathbb{E}_i(e^{\theta^\top X^i}) = \log \mathbb{E}_i \left(\prod_{j=1}^N e^{\theta_j X_j^i} \right),$$

we deduce that, using (4.4),

$$M_{ij} = \mathbb{E}_i(X_j^i) = \sum_{v \in \mathbf{V}^i} \Gamma^j \mu(p_j(v)) \lambda_i(v), \tag{4.6}$$

which is the mean number of offspring particles of type j , issued by a particle of type i . It is a well-known fact in branching processes (see e.g. Chapter V of [1]) that

$$\mathbb{E}_i((K^i(n))^\top) = e_i^\top M^n,$$

and that

$$\mathbb{E}_i((W^i(n))^\top) = \mathbb{E}_i \left(\sum_{k=0}^n (K^i(k))^\top \right) = e_i^\top \left(\sum_{k=0}^n M^k \right).$$

Thus, by monotone convergence,

$$\mathbb{E}_i(W^i) = e_i^\top \left(\sum_{k=0}^{\infty} M^k \right) \mathbf{1}, \tag{4.7}$$

which is finite by (4.3). Hence (4.5) tells us that the number of points produced in the backward steps is well concentrated around its mean. If we think that the overall complexity of the algorithm is governed by the number of points produced in the backward steps, then (4.7) gives us a good way to determine which distribution λ^i can lead to the less costly algorithm. We are looking basically at the λ^i that are minimizing (4.7), with M defined by (4.6)

Proof of Proposition 4.2. Step 1. First we prove that $\phi_i(\theta)$ is well-defined for all $\theta \in B(0, r)$. Indeed, since the $(X_j^i)_{j=1, \dots, N}$ are independent, we have

$$\phi_i(\theta) = \log \mathbb{E}_i \left(\prod_{j=1}^N \exp(\theta_j X_j^i) \right) = \sum_{j=1}^N \log \mathbb{E}_i (\exp(\theta_j X_j^i)).$$

By (4.4),

$$\mathbb{E}_i (\exp(\theta_j X_j^i) \mid V_T^i = v) = \exp[(e^{\theta_j} - 1)\Gamma^j \mu(p_j(v))],$$

and integrating with respect to the choice of V_T^i , we obtain

$$\phi_i(\theta) = \sum_{j=1}^N \log \left(\sum_{v \in \mathbf{V}^i} \lambda_i(v) \exp[(e^{\theta_j} - 1)\Gamma^j \mu(p_j(v))] \right) < \infty.$$

Step 2. In the following, we prove that $\Phi_i^{(n)}(\theta)$ satisfies the recursion

$$\Phi_i^{(n)}(\theta) = \theta^\top K^i(0) + \phi_i(\Phi^{(n-1)}(\theta)).$$

Indeed, by definition of $W^i(n)$ we have

$$\mathbb{E}_i (e^{\theta^\top W^i(n)}) = e^{\theta^\top K^i(0)} \mathbb{E}_i (e^{\theta^\top \sum_{k=1}^n K^i(k)}).$$

Now, let us introduce, for any j and any $1 \leq p \leq X_j^i$, the vector $K_{(p)}^j(n-1)$ of offspring particles within the $(n-1)$ th generation, issued from the p th particle of type j in the first generation. Notice that by the branching property, for $p = 1, \dots, X_j^i$, we have that the $K_{(p)}^j(n-1)$ are independent copies of $K^j(n-1)$. Therefore, conditioning on the first generation of offspring particles,

$$\begin{aligned} \mathbb{E}_i (e^{\theta^\top \sum_{k=1}^n K^i(k)}) &= \mathbb{E}_i \left(e^{\theta^\top \sum_{k=1}^n \sum_{j=1}^N \sum_{p=1}^{X_j^i} K_{(p)}^j(k-1)} \right) \\ &= \mathbb{E}_i \left(\prod_{j=1}^N \prod_{p=1}^{X_j^i} e^{\theta^\top \sum_{k=1}^n K_{(p)}^j(k-1)} \right) \\ &= \mathbb{E}_i \left[\mathbb{E} \left(\prod_{j=1}^N \prod_{p=1}^{X_j^i} e^{\theta^\top \sum_{k=1}^n K_{(p)}^j(k-1)} \mid X^i \right) \right]. \end{aligned}$$

Since the $(K_{(p)}^j(k-1))_{1 \leq j \leq N}$ are independent and independent of X^i , we obtain

$$\begin{aligned} \mathbb{E}_i \left[\mathbb{E} \left(\prod_{j=1}^N \prod_{p=1}^{X_j^i} e^{\theta^\top \sum_{k=1}^n K_{(p)}^j(k-1)} \mid X^i \right) \right] &= \mathbb{E}_i \left[\prod_{j=1}^N \prod_{p=1}^{X_j^i} \mathbb{E} (e^{\theta^\top \sum_{k=1}^n K_{(p)}^j(k-1)} \mid X_j^i) \right] \\ &= \mathbb{E}_i \left[\prod_{j=1}^N \prod_{p=1}^{X_j^i} \mathbb{E} (e^{\theta^\top \sum_{k=1}^n K_{(p)}^j(k-1)}) \right] \end{aligned}$$

$$\begin{aligned}
 &= \mathbb{E}_i \left[\prod_{j=1}^N (\mathbb{E} (e^{\theta^\top W^j(n-1)}))^{X_j^i} \right] \\
 &= \mathbb{E}_i \left[\prod_{j=1}^N e^{\Phi_j^{(n-1)}(\theta) X_j^i} \right] \\
 &= \mathbb{E}_i [e^{\Phi^{(n-1)}(\theta)^\top X_i}] \\
 &= e^{\phi_i(\Phi^{(n-1)}(\theta))},
 \end{aligned}$$

implying

$$\Phi_i^{(n)}(\theta) = \theta^\top K^i(0) + \phi_i(\Phi^{(n-1)}(\theta)),$$

or, in vector form,

$$\Phi^{(n)}(\theta) = \theta + \phi(\Phi^{(n-1)}(\theta)). \tag{4.8}$$

Step 3. Let us consider the sums of the elements of the i th line of the matrix M ,

$$\sum_{j=1}^N M_{ij} = \sum_{j=1}^N \mathbb{E}_i(X_j^i).$$

By definition, this is the mean number of offspring particles (of any type), issued from a particle of type i , and by the arguments presented in the proof of Proposition 4.1, this mean number is given by

$$\sum_{j=1}^N M_{ij} = \sum_{v \in \mathbf{V}^i} \lambda^i(v) P(v) \leq \gamma < 1,$$

where the last upper bound holds by our assumptions. Hence

$$\|M\|_\infty = \sup_{\|x\|_\infty \leq 1} \{\|Mx\|_\infty\} = \sup_i \sum_{j=1}^N |M_{ij}| \leq \gamma < 1,$$

where $\|\cdot\|_\infty$ is the induced norm for matrices on $\mathbb{R}^{N \times N}$. Therefore the derivative of ϕ in 0 satisfies $\|D\phi(0)\|_\infty \leq \gamma < 1$. Moreover, since the norm is continuous and so is $D\phi(s)$, for any $\gamma < C < 1$, there is a $\tilde{r} \in (0, R)$, such that, for $\|s\|_\infty < \tilde{r}$,

$$\|D\phi(s)\|_\infty \leq C < 1.$$

Hence $\phi(s)$ is Lipschitz-continuous in the ball $B(0, \tilde{r})$ and moreover $\phi(0) = 0$, which implies that

$$\|\phi(s)\|_\infty \leq C \|s\|_\infty$$

for $\|s\|_\infty < \tilde{r}$.

Now, take θ such that $|\theta_i| \leq \tilde{r}(1 - C)$ for any $1 \leq i \leq N$. Using (4.8), we can show by induction that for all n ,

$$\|\Phi^{(n)}(\theta)\|_\infty \leq \|\theta\|_\infty (1 + C + \dots + C^n) \leq \tilde{r} < \infty.$$

Hence, by monotone convergence of $\Phi^n(\theta) \rightarrow \Phi(\theta)$ and the limit in (4.8), we have

$$\|\Phi(\theta)\|_\infty \leq r \quad \text{and} \quad \Phi(\theta) = \theta + \phi(\Phi(\theta)),$$

with $r = \tilde{r}(1 - C)$. In particular, choosing $\theta = \vartheta \mathbf{1} \in B(0, r)$ with $\vartheta \in \mathbb{R}$, we have

$$\mathbb{E}_i(e^{\vartheta W^i}) < \infty,$$

where we recall that W^i is the total number of offspring particles of i .

Step 4. We use Markov's inequality and obtain for any $\vartheta \in (0, r)$

$$\mathbb{P}_i(W^i > \mathbb{E}_i(W^i) + x) = \mathbb{P}_i(e^{\vartheta W^i} > e^{\vartheta \mathbb{E}_i(W^i) + \vartheta x}) \leq e^{\tilde{\Phi}_i(\vartheta) - \vartheta \mathbb{E}_i(W^i) - \vartheta x},$$

where $\tilde{\Phi}_i(\vartheta) = \log \mathbb{E}_i(e^{\vartheta W^i})$. Using Taylor's formula, we have

$$\tilde{\Phi}_i(\vartheta) = \tilde{\Phi}'_i(0)\vartheta + \frac{1}{2}\tilde{\Phi}''_i(\tilde{\vartheta})\vartheta^2 = \mathbb{E}_i(W^i)\vartheta + \frac{1}{2}\tilde{\Phi}''_i(\tilde{\vartheta})\vartheta^2$$

for some $\tilde{\vartheta} \in (0, \vartheta)$. Choosing $c_0 = \sup_{\vartheta \leq r} e^{\frac{1}{2}\tilde{\Phi}''_i(\tilde{\vartheta})\vartheta^2}$ implies the assertion. □

4.2.6. Choice of the weights on a particular example In this section we discuss a very particular example to show how to calibrate the choice of the λ^i in terms of minimizing the number of points produced in the backward steps. Our example is the age-dependent Hawkes process with hard refractory period $\delta > 0$ defined by (3.7) and with non-decreasing ψ^i which is L -Lipschitz for all i , so that Corollary 3.3(ii) applies. The following choices are merely directed to have the simplest possible computations on a non-trivial infinite case.

In what follows, to simplify the computations, we consider that $\mathbf{I} = \mathbb{Z}$ and $L = 1$. Moreover, for all i , we make the following assumptions.

- (1) $\psi^i(0) = 1$.
- (2) $h_j^i(t) = \beta_j^i \exp(-t/\delta)$, where $\beta_j^i = 1/(2|j - i|^\gamma)$ for $j \neq i$, for some fixed positive parameter $\gamma > 1$, and $\beta_j^i = 1$, where $\beta_j^i \geq 0$. Note that $\|h_j^i\|_1 = \beta_j^i \delta$ and that $h_j^i(0) = \beta_j^i$.
- (3) With the notation of Corollary 3.3, we use $\omega_1^i = \{i\}$ and

$$\omega_k^i = \{i - k + 1, \dots, i, i + 1, \dots, i + k - 1\} \quad \text{for all } k \geq 2.$$

With this choice, one can see that $\bar{\Gamma}_k^i$ defined in Corollary 3.3(ii) is given for $k \geq 2$ by

$$\bar{\Gamma}_k^i = \frac{2}{(k - 1)^\gamma} + e^{-(k-1)} \left(1 + \sum_{m=1}^{k-2} \frac{1}{m^\gamma} \right).$$

Hence, for some constant $C_\gamma > 0$ depending on γ , one can always choose, for all $k \geq 1$,

$$\Gamma_k^i \geq C_\gamma k^{-\gamma},$$

as long as $\sum_k \Gamma_k^i < \infty$. So let us take

$$\Gamma_k^i = C_\gamma k^{-p}$$

for some $1 < p \leq \gamma$. Note that in this case, Γ^i is independent of i , since

$$\Gamma^i = \sum_{k=1}^{\infty} \Gamma_k^i = C_\gamma c_p, \quad \text{where } c_p = \sum_{k=1}^{\infty} k^{-p}.$$

By applying Proposition 3.3(ii) combined with Corollary 3.3(ii), we obtain

$$\lambda^i(v_k^i) = (c_p)^{-1} k^{-p}.$$

Now let us turn to finding p such that the λ^i are minimizing (4.7), that is,

$$\mathbb{E}_i(W^i) = e_i^\top \left(\sum_{k=0}^{\infty} M^k \right) \mathbf{1},$$

with M defined by

$$M_{ij} = \mathbb{E}_i(X_j^i) = \sum_{k=1}^{\infty} \Gamma^j \mu(p_j(v_k^i)) \lambda^i(v_k^i).$$

Summing over all possible types j , we obtain the mean number of offspring particles of a particle of type i , which is given by

$$M_i = \sum_{j=1}^{\infty} M_{ij} = C_\gamma \delta \sum_{k=1}^{\infty} (2k - 1) k^{1-p} =: C_\gamma \delta f(p), \quad \text{where } f(p) = \sum_{k=1}^{\infty} (2k - 1) k^{1-p}.$$

Clearly, $f(p) < \infty$ if and only if $p > 3$, and thus, *a fortiori*, $\gamma > 3$. Since f is a decreasing function of p , the optimal choice for p is thus $p = \gamma$.

Acknowledgements

This paper is dedicated to the Institute of Mathematics Hanoi, where T. C. Phi was honored to work for more than three years. We also warmly thank the referees of an earlier version of this paper for the very detailed comments that helped us to improve the manuscript.

Funding information

This work was supported by the French government, through the UCAJedi and 3IA Côte d’Azur Investissements d’Avenir managed by the National Research Agency (ANR-15-IDEX-01 and ANR- 19-P3IA-0002), by the CNRS through the ‘Mission pour les Initiatives Transverses et Interdisciplinaires’ (Projet DYNAMO, ‘APP Modélisation du Vivant’), by the interdisciplinary Institute for Modeling in Neuroscience and Cognition (NeuroMod) of the Université Côte d’Azur, and directly by the National Research Agency (ANR-19-CE40-0024) with the ChaMaNe project. Moreover, this work is part of the FAPESP project Research, Innovation and Dissemination Center for Neuromathematics (grant 2013/07699-0).

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

- [1] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes* (Grundlehren der mathematischen Wissenschaften **196**). Springer, New York and Heidelberg.
- [2] BACRY, E., DELATTRE, S., HOFFMANN, M. AND MUZY, J.-F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stoch. Process. Appl.* **123**, 2475–2499.
- [3] BRÉMAUD, P. (1981). *Point Processes and Queues: Martingale Dynamics* (Springer Series in Statistics). Springer, New York and Berlin.
- [4] BRÉMAUD, P. AND MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Prob.* **24**, 1563–1588.
- [5] CARSTENSEN, L., SANDELIN, A., WINTHER, O. AND HANSEN, N. R. (2010). Multivariate Hawkes process models of the occurrence of regulatory elements. *BMC Bioinform.* **11**, 1–19.
- [6] CHEN, X. (2021). Perfect sampling of Hawkes processes and queues with Hawkes arrivals. *Stoch. Systems* **11**, 264–283.
- [7] CHEN, X. AND WANG, X. (2020). Perfect sampling of multivariate Hawkes processes. In *Proceedings of the Winter Simulation Conference (WSC '20)*, pp. 469–480. IEEE Press.
- [8] CHEVALLIER, J. (2017). Mean-field limit of generalized Hawkes processes. *Stoch. Process. Appl.* **127**, 3870–3912.
- [9] DALEY, D. J. AND VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*, Vol. I, *Elementary Theory and Methods*. Springer.
- [10] DASSIOS, A. AND ZHAO, H. (2013). Exact simulation of Hawkes process with exponentially decaying intensity. *Electron. Commun. Prob.* **18**, 1–13.
- [11] DELATTRE, S., FOURNIER, N. AND HOFFMANN, M. (2016). Hawkes processes on large networks. *Ann. Appl. Prob.* **26**, 216–261.
- [12] FERRARI, P. A. (1990). Ergodicity for spin systems with stirrings. *Ann. Prob.* **18**, 1523–1538.
- [13] GALVES, A. AND LÖCHERBACH, E. (2013). Infinite systems of interacting chains with memory of variable length: a stochastic model for biological neural nets. *J. Statist. Phys.* **151**, 896–921.
- [14] GALVES, A., GARCIA, N., LÖCHERBACH, E. AND ORLANDI, E. (2013). Kalikow-type decomposition for multicolor infinite range particle systems. *Ann. Appl. Prob.* **23**, 1629–1659.
- [15] GALVES, A., LÖCHERBACH, E. AND ORLANDI, E. (2010). Perfect simulation of infinite range Gibbs measures and coupling with their finite range approximations. *J. Statist. Phys.* **138**, 476–495.
- [16] HALL, E. C. AND WILLET, R. M. (2016). Tracking dynamic point processes on networks. *IEEE Trans. Inform. Theory* **62**, 4327–4346.
- [17] HAWKES, A. G. (1971). Point spectra of some mutually exciting point processes. *J. R. Statist. Soc. B* **33**, 438–443.
- [18] HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90.
- [19] HODARA, P. AND LÖCHERBACH, E. (2017). Hawkes processes with variable length memory and an infinite number of components. *Adv. Appl. Prob.* **49**, 84–107.
- [20] JACOD, J. (1979). *Calcul Stochastique et Problèmes de Martingales* (Lecture Notes Math. **714**). Springer.
- [21] JACOD, J. AND SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd edn. Springer, Berlin, Heidelberg and New York.
- [22] KALIKOW, S. (1990). Random Markov processes and uniform martingales. *Israel J. Math.* **71**, 33–54.
- [23] LEWIS, P. W. AND SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Res. Logistics Quart.* **26**, 403–413.
- [24] MASCART, C., MUZY, A. AND REYNAUD-BOURET, P. (2020). Discrete event simulation of point processes: a computational complexity analysis on sparse graphs. Available at [arXiv:2001.01702](https://arxiv.org/abs/2001.01702).
- [25] MØLLER, J. AND RASMUSSEN, J. G. (2005). Perfect simulation of Hawkes processes. *Adv. Appl. Prob.* **37**, 629–646.
- [26] OGATA, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans. Inform. Theory* **27**, 23–31.
- [27] OST, G. AND REYNAUD-BOURET, P. (2020). Sparse space–time models: concentration inequalities and Lasso. *Ann. Inst. H. Poincaré Prob. Statist.* **56**, 2377–2405.
- [28] PHI, T. C., MUZY, A. AND REYNAUD-BOURET, P. (2020). Event-scheduling algorithms with Kalikow decomposition for simulating potentially infinite neuronal networks. *SN Comput. Sci.* **1**, 35.
- [29] RAAD, M. B., DITLEVSEN, S. AND LÖCHERBACH, E. (2020). Stability and mean-field limits of age dependent Hawkes processes. *Ann. Inst. H. Poincaré Prob. Statist.* **56**, 1958–1990.
- [30] REYNAUD-BOURET, P. AND ROY, E. (2006). Some non asymptotic tail estimates for Hawkes processes. *Bull. Belg. Math. Soc. Simon Stevin* **13**, 883–896.

- [31] REYNAUD-BOURET, P. AND SCHBATH, S. (2010). Adaptive estimation for Hawkes processes: application to genome analysis. *Ann. Statist.* **38**, 2781–2822.
- [32] SCARELLA, G., MASCART, C., MUZY, A., PHI, T. C. AND REYNAUD-BOURET, P. (2021). Reconstruction de la connectivité fonctionnelle en neurosciences: une amélioration des algorithmes actuels. In *52èmes Journées de Statistique de la Société Française de Statistique (SFdS)*.