# Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity

Anna Bonnet, Miguel Martinez Herrera, Maxime Sangnier

# Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity

Anna Bonnet
Univ. de Paris and Sorbonne Université, CNRS, LPSM, F-75013 Paris, France
and
Miguel Martinez Herrera
Univ. de Paris and Sorbonne Université, CNRS, LPSM, F-75013 Paris, France
and
Maxime Sangnier
Univ. de Paris and Sorbonne Université, CNRS, LPSM, F-75013 Paris, France

April 26, 2022

**Abstract**

The Hawkes process is a multivariate past-dependent point process used to model the relationship of event occurrences between different phenomena. Although the Hawkes process was originally introduced to describe excitation interactions, which means that one event increases the chances of another occurring, there has been a growing interest in modeling the opposite effect, known as inhibition. In this paper, we propose a maximum likelihood approach to estimate the interaction functions of a multivariate Hawkes process that can account for both exciting and inhibiting effects. To the best of our knowledge, this is the first exact inference procedure designed for such a general setting in the frequentist framework. Our method includes a thresholding step in order to recover the support of interactions and therefore to infer the connectivity graph. A benefit of our method is to provide an explicit computation of the log-likelihood, which enables in addition to perform a goodness-of-fit test for assessing the quality of estimations. We compare our method to classical approaches, which were developed in the linear framework and are not specifically designed for handling inhibiting effects. We show that the proposed estimator performs better on synthetic data than alternative approaches. We also illustrate the application of our procedure to a neuronal activity dataset, which highlights the presence of both exciting and inhibiting effects between neurons.

*Keywords: Non-linear Hawkes process, point process, maximum likelihood estimation, goodness-of-fit.*

# 1 Introduction

A Hawkes process is a point process in which each point is commonly associated with event occurrences in time. In this past-dependent model, every event time impacts the probability that other events take place subsequently. These processes are characterised by the conditional intensity function, seen as an instantaneous measure of the probability of event occurrences. Since their introduction in Hawkes [1971], Hawkes processes have been applied in a wide variety of fields, for instance in seismology [Ogata, 1988], social media [Rizoiu et al., 2017], criminology [Olinde and Short, 2020] and neuroscience [Reynaud-Bouret et al., 2018].

The multidimensional version of this model, referred to as the multivariate Hawkes process, models the interaction between different kinds of events through kernel functions. Originally this model takes only into account mutually exciting interactions (an event increases the chances of others occurring) by assuming that all kernel functions are non-negative. A specificity of self-exciting Hawkes processes is their branching structure, also known as cluster structure. Introduced in Hawkes and Oakes [1974], it has been used in order to build the theoretical background of Hawkes processes (the most known results being existence and the value of the expected number of points) by leveraging the branching theory. Estimation methods in the literature are vast including maximum likelihood estimators [Ozaki, 1979, Guo et al., 2018] and method of moments [Da Fonseca and Zaatour, 2013]. Non-parametric approaches include an EM procedure introduced in Lewis and Mohler [2011], estimations obtained via the solution of Wiener-Hopf equations [Bacry and Muzy, 2016] or by approximating the process through autoregressive models [Kirchner, 2017] or through functions in reproducing kernel Hilbert spaces [Yang et al., 2017].

Although the self-exciting Hawkes process remains widely studied, there has been a growing interest in modeling the opposite effect, known as inhibition, in which the probability of observing an event is lowered by the apparition of certain events. In practice, this amounts to considering negative kernel functions. In order to maintain the positivity of the intensity function, a non-linear operator is added to the expression which in turns entails the loss of the cluster representation. This model known as the non-linear Hawkes process was first presented in Brémaud and Massoulié [1996], where existence of such processes was proved via construction using bi-dimensional marked Poisson processes. Such approach of analysis has been used in the literature as in Chen et al. [2017], where a coupling process is established and leveraged to obtain theoretical guarantees on cross-analysis covariance. Another approach is presented in Costa et al. [2020], where renewal theory allows to obtain limit theorems for processes with bounded support kernel functions. Estimation methods focus mainly on non-parametric methods for general interactions and non-linear functions, as found in Bacry and Muzy [2016], Sulem et al. [2021].

In the last years, alternative models have been designed in order to take into account inhibiting effects in Hawkes processes. An example is the neural Hawkes process, presented in Mei and Eisner [2017], Zuo et al. [2020], which combines a multivariate Hawkes process and a recurrent neural network architecture. In Duval et al. [2021], a multiplicative model considers two sets of neuronal populations, one exciting and another inhibiting, and each intensity function is the product of two non-linear functions (one for each group). Another model is presented in Olinde and Short [2020] and called self-limiting Hawkes process. It includes the inhibition as a multiplicative term in front of a the traditional self-exciting intensity function.

In this paper, we present a maximum likelihood estimation method for multivariate Hawkes processes with exponential kernel functions, that works for both exciting and inhibiting interactions, as modelled by Brémaud and Massoulié [1996], Chen et al. [2017]. This work builds upon the methodology for the univariate case, presented in Bonnet et al. [2021], by focusing in the intervals where the intensity function is positive. We show that, under a weak assumption on the kernel functions, these intervals can be exactly determined. We can then write the compensator for each dimension which in turn provides an explicit expression of the log-likelihood. This enables to build the corresponding maximum likelihood estimator, the numerical procedure of which is implemented in Python and freely available on GitHub.[1] As a by-product, the closed-form expression of the compensator also allows to assess goodness-of-fit via the Time Change Theorem and multiple testing. We

[1] https://github.com/migmtz/multivariate-hawkes-inhibition

carry out a numerical study on simulated data and on a neuronal activity dataset [Petersen and Berg, 2016, Radosevic et al., 2019]. The performance of our approach is compared to estimations obtained via approximations from Bacry et al. [2020] and Lemonnier and Vayatis [2014], and we show that our method not only achieves better estimations but is capable of identifying correctly the interaction network of the process.

To outline this paper, Section 2 presents the multivariate Hawkes process framework and reviews the literature regarding inference of non-linear Hawkes processes. In Section 3, we give a simple condition of existence of a multivariate Hawkes process with inhibition and we present the estimation procedure with an exact computation of the log-likelihood when the kernel functions are exponential. We also explain in the same section, how our approach enables to perform a goodness-of-fit test in order to assess the quality of estimations. The whole procedure is illustrated on simulated data in Section 4 and applied to a neuronal activity dataset in Section 5.

# 2 The multivariate Hawkes process

## 2.1 Definition

A multivariate Hawkes process $N = (N^1, N^2, \dots, N^d)$ of dimension $d$ is defined by $d$ point processes on $\mathbb{R}_+^*$, denoted $N^i \colon \mathcal{B}(\mathbb{R}_+^*) \to \mathbb{N}$, where $\mathcal{B}(\mathbb{R}_+^*)$ is the Borel algebra on the set of positive numbers. The process $N$ is supposed to be orderly, in other words, two events cannot occur at the same time.

Each process $N^i$ can be characterised by its associated event times $(T_k^i)_k$ and its conditional intensity function, defined for all $t \geq 0$ by

$$\lambda^i(t) = \left(\mu^i + \sum_{j=1}^{d} \int_0^t h_{ij}(t-s)\, \mathrm{d}N^j(s)\right)^+ = \left(\mu^i + \sum_{j=1}^{d} \sum_{T_k^j \leq t} h_{ij}(t - T_k^j)\right)^+ , \qquad (1)$$

where $x^+ = \max(0, x)$. Here, the quantity $\mu^i \in \mathbb{R}_+^*$ is called the baseline intensity and each interaction function $h_{ij} \colon \mathbb{R}_+^* \to \mathbb{R}$ represents the influence of the process $N_j$ on the process $N_i$.

**Remark.** *The positive-part function in Equation (1) is needed to insure the non-negativity of $\lambda^i$ in the presence of strong inhibiting effects (that is when some interaction functions $h_{ij}$ are sufficiently negative compared to positive contributions). Concretely, the positive part does not affect the intensity function if inhibiting effects are in minority compared to the positive contributions (exciting effects or the baseline intensities).*

For each process $N_i$ and for all $t \geq 0$, let us note $N^i(t) = \sum_{k\geq 1} \mathbb{1}_{T_k^i \leq t}$ the measure of $(0, t]$ and the compensator

$$\Lambda^i(t) = \int_0^t \lambda^i(u) \, \mathrm{d}u \, .$$

The process $N$ can be seen as a point process on $\mathbb{R}_+^*$, where for any $B \in \mathcal{B}(\mathbb{R}_+^*)$, $N(B) = \sum_{i=1}^d N^i(B)$ (and we may define, for every $t \geq 0$, $N(t) = \sum_{k\geq 1} \mathbb{1}_{T_{(k)} \leq t} = \sum_{i=1}^d N^i(t)$). Similarly to a univariate process, $N$ can be characterised by its conditional intensity $\lambda$ (also called total intensity) [Daley and Vere-Jones, 2003]:

$$\lambda(t) = \sum_{i=1}^d \lambda^i(t) \, , \tag{2}$$

and by its compensator

$$\Lambda(t) = \int_0^t \lambda(u) \, \mathrm{d}u = \sum_{i=1}^d \Lambda^i(t) \, .$$

From this point of view, the process $N$ is associated to event times $\left(T_{(k)}\right)_k = \left(T_{u_k}^{m_k}\right)_k$, corresponding to the ordered sequence composed of $\bigcup_{i=1}^d \{T_k^i \mid k > 0\}$. Here, $(u_k)_k$ is the random ordering sequence and $(m_k)_k$ the sequence of *marks* that make it possible to identify to which dimension each time corresponds to. These marks can be written as

$$m_k = \sum_{j=1}^d j \mathbb{1}_{N^j\left(\{T_{(k)}\}\right) = 1} \, .$$

As the aim of this paper is to describe a practical methodology for estimating the conditional intensities $\lambda_1, \ldots, \lambda_d$ via maximizing the log-likelihood, the latter quantity has to be made explicit. Let $t \geq 0$ and assume that event times $\{T_k^i : 1 \leq k \leq N_i(t), 1 \leq i \leq d\}$ are observed in the interval $(0, t]$. Then, given a parametric model $\mathcal{P} = \{\lambda_\theta \colon \theta \in \Theta\}$ (and associated compensators $\Lambda_\theta$) for each conditional intensity function $\lambda^i$, for every $\theta \in \Theta$, the log-likelihood $\ell_t(\theta)$ reads [Daley and Vere-Jones, 2003, Proposition 7.3.III.]

$$\ell_t(\theta) = \sum_{i=1}^{d} \ell_t^i(\theta)\,, \qquad \text{with} \qquad \ell_t^i(\theta) = \sum_{k=1}^{N^i(t)} \log \lambda_\theta^i(T_k^{i-}) - \Lambda_\theta^i(t)\,, \tag{3}$$

where $\lambda_\theta^i(T_k^{i-}) = \lim_{t \to T_k^{i-}} \lambda_\theta^i(t)$ and with convention $\log(x) = -\infty$ for $x \leq 0$.

The heart of the problem in deriving a maximum likelihood estimator for the conditional intensities $\lambda^i$ is being able to evaluate exactly the compensator values $\Lambda_\theta^i(t)$ for every possible $\theta \in \Theta$, which requires to determine when the conditional intensities $\lambda^i$ are non-zero. The forthcoming sections clear this point up.

## 2.2 Related work

Estimation methods for Hawkes processes have focused mainly on self-exciting interactions (by assuming $h_{ij} \geq 0$). In Ozaki [1979], the author presents the maximum likelihood estimation method for univariate processes with exponential kernel, the same method is established in Mishra et al. [2016] for the power law kernel function. In Chen et al. [2018] the maximum likelihood method is presented for the multivariate version with exponential kernel. In Bacry et al. [2020], estimations for the exponential multivariate case are obtained by optimising a least-squares criterion. Other methods in the parametric setting include estimations obtained via spectral analysis in Adamopoulos [1976], by using an EM algorithm in Veen and Schoenberg [2006] or via the method of moments in Da Fonseca and Zaatour [2013].

Estimators of the interaction functions are also presented in non-parametric settings. For instance, Yang et al. [2017] proposes a non-parametric online algorithm for multivariate Hawkes processes via functions in a reproducing kernel Hilbert space and an alternate version of the likelihood. In Reynaud-Bouret et al. [2014], an approximation of $h_{ij}$ is obtained by considering piece-wise constant functions on a bounded interval $[0, A]$, $A > 0$. This method consists in optimising a least-squares loss by considering the histograms formed by the observed event times. Hawkes processes with excitation have also been studied in Bayesian contexts, as in Rasmussen [2013] for the univariate case and in Donnet et al. [2020] for multivariate processes using the loglikelihood.

Although inhibiting effects in Hawkes processes where first mentioned in Brémaud and Massoulié [1996], they have only met a growing interest in the last decade. Concerning inference, most of the known methods are not designed for handling the inhibiting case, but some are capable in practice to estimate negative interactions. For instance Reynaud-Bouret et al. [2014] in the non-parametric setting and Bacry et al. [2020] in the parametric one, as mentioned above. Another similar method is proposed in Lemonnier and Vayatis [2014] where an approximation of the compensator $\Lambda$ is obtained by integrating the intensity function $\lambda$ without the positive part function and then considering a family of exponential functions. However, it is unclear how these approaches will perform when the function $\lambda$ is frequently equal to zero due to inhibiting terms. This remark is mentioned in Bacry and Muzy [2016] where by assuming that there is a negligible chance of the intensity function being negative, their proposed method provides negative-valued estimations. Obviously, this is not a problem while these intensities remain non-negative. However, as soon as the latter condition is violated, the numerical procedures fail to produce an admissible estimation.

Inference procedures that are adapted specifically to Hawkes processes with inhibition are scarcer in the literature. Sulem et al. [2021] presents various results for non-linear Hawkes processes including inhibition effects regarding existence, stability and Bayesian estimation for kernel functions with bounded support. Deutsch and Ross [2022] presents choices of priors for Bayesian estimation based on a new reparametrisation of the process.

Lastly, Bonnet et al. [2021] presents a maximum likelihood estimation adapted to the univariate Hawkes process with inhibition and monotone kernel functions. The decisive contribution of this work is to give, for an exponential kernel $h(t) = \alpha \mathrm{e}^{-\beta t}$ ($\alpha \in \mathbb{R}$, $\beta > 0$), a closed-form expression of *restart times*, which are basically the instants at which the single conditional intensity becomes non-zero. This makes possible to compute explicitly the compensator and then the log-likelihood. Yet, this study is limited to the univariate case. It has to be noted that a formalism similar to Bonnet et al. [2021] but for multivariate Hawkes processes is mentioned in Deutsch and Ross [2022]. However, the authors chose to put this framework aside and prefer to focus on non-parametric Bayesian estimation.

This paper goes a step forward in estimation of multivariate Hawkes processes with inhibition, by providing the first exact maximum likelihood method for exponential interactions $h_{ij}(t) = \alpha_{ij} \mathrm{e}^{-\beta_{ij} t}$. As it will be explained in the next section, this also enables to perform standard goodness-of-fit tests.

# 3 Existence and estimation

## 3.1 Existence

Before motivating and explaining the estimation procedure proposed in this paper, existence, uniqueness and stationarity of a counting process characterised by conditional intensities $\lambda^i$ as defined in Equation (1) need to be confirmed.

A sufficient condition for the existence of a unique stationary Hawkes process with such intensities is for the matrix $S = (\|h_{ij}\|_1)_{ij}$ to have a spectral radius $\rho(S)$ strictly smaller than 1 [Brémaud and Massoulié, 1996, Theorem 7]. However, this appears to be a strong assumption in the framework considered here, as it takes into account negative interactions, which in theory do not compromise the existence of the process. Thus, we point out a weaker assumption, which considers instead the exciting contribution of each kernel function $h_{ij}$.

**Proposition 3.1.** *If the matrix $S^+ = (\|h_{ij}^+\|_1)_{ij}$ satisfies $\|S^+\|_\infty < 1$, then there exists a unique distribution of the process $N$ with finite average intensity.*

This result is a straightforward adaptation of [Sulem et al., 2021, Lemma 2.1], which originally states the existence of Hawkes processes with inhibition tailored by interaction functions $h_{ij}$ with bounded supports. Proposition 3.1 is the extension to interaction functions with unbounded supports and is immediately obtained by changing $\int_{s-A}^{s}$ to $\int_{0}^{s}$ in [Sulem et al., 2021, Appendix D].

While quite simple, Proposition 3.1 is of importance to theoretically support the simulation and the estimation of multivariate Hawkes processes with exponential interaction functions. In particular, it allows for studying numerically processes with strong inhibiting effects, for which the interaction matrix $S$ has a spectral radius $\rho(S) \geq 1$, but which verifies $\|S^+\|_\infty < 1$ (see Section 4).

### 3.2 Introductive example

Figure 1 depicts (in red) conditional intensities $\lambda^1$ and $\lambda^2$ for a realisation of 2-dimensional Hawkes process. The simulation has been carried out with baselines $\mu^1 = 0.5$ and $\mu^2 = 1.0$, and with exponential kernels $h_{ij}(t) = \alpha_{ij} \mathrm{e}^{-\beta_{ij} t}$ parameterised by:

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \begin{pmatrix} -1.9 & 3.0 \\ 0.9 & -0.7 \end{pmatrix}, \qquad \text{and} \qquad \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = \begin{pmatrix} 2.0 & 20.0 \\ 3.0 & 2.0 \end{pmatrix}.$$

These kernels have been chosen such that both processes are self-inhibiting ($\alpha_{11}, \alpha_{22} < 0$) but inter-exciting ($\alpha_{12}, \alpha_{21} > 0$).

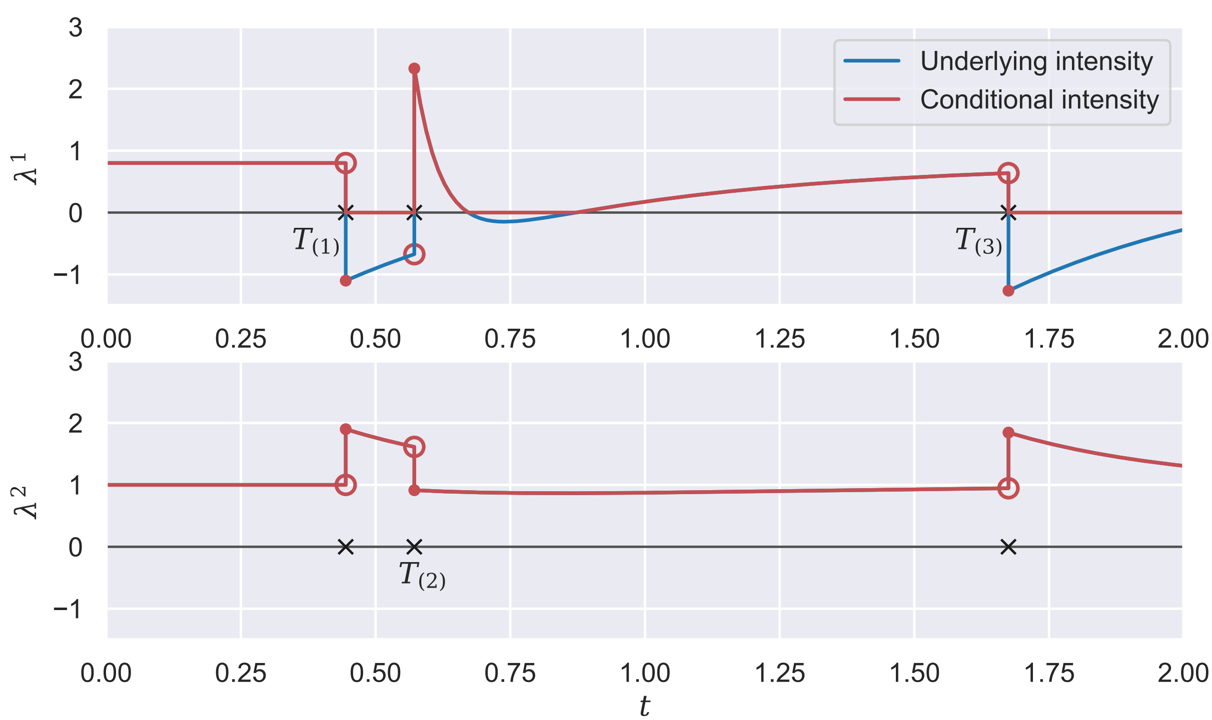



Figure 1: Simulation of a 2-dimensional Hawkes process. Each cross corresponds to an event time, and each $T_{(k)}$ is shown in its corresponding process.

The goal of this paper is to establish a parametric estimation method, via maximum likelihood estimation, that is able to handle both excitation and inhibition frameworks in the multivariate case. For this purpose, it is necessary to compute explicitly the log-likelihood $\ell_t(\theta)$ (see Equation (3)) and in particular to evaluate the compensator $\Lambda^i_\theta$, expressed as an integral of $\lambda^i_\theta$. For the latter, the main challenge is to determine when conditional intensities $\lambda^i$ are non-zero, that is on which intervals they are tailored by the exponential interaction functions and not by the positive-part operator.

In Bonnet et al. [2021], the authors solved this challenge for univariate processes by remarking that the conditional intensity is monotone between two event times. Figure 1 illustrates that this is not necessarily true for multivariate processes (here, between $T_{(2)}$ and $T_{(3)}$). This constitutes the major difficulty we have to cope with.

### 3.3 Underlying intensity and restart times in the multivariate setting

From now on, let us focus the study to the exponential model [Hawkes, 1971], where each interaction function $h_{ij}$ is then defined as

$$h_{ij}(t) = \alpha_{ij} \mathrm{e}^{-\beta_{ij} t}\,,$$

with $\alpha_{ij} \in \mathbb{R}$ and $\beta_{ij} \in \mathbb{R}^*_+$ for $i, j \in \{1, \dots, d\}$. For each $i \in \{1, \dots, d\}$, we define the underlying intensity function $\lambda^{i\star}$ to be

$$\lambda^{i\star}(t) = \mu^i + \sum_{j=1}^{d} \int_0^t h_{ij}(t-s)\, \mathrm{d}N^j(s)\,.$$

This quantity coincides with the conditional intensity $\lambda^i$ when it is non-zero, and is non-positive otherwise. In particular, we can observe that $\lambda^i(t) = (\lambda^{i\star})^+$ (see Figure 1).

As explained in the previous section, the main difficulty of the multivariate exponential setting is the non-monotony of conditional intensities $\lambda^i$ between two event times. Determining intervals where $\lambda^i$ is non-zero (that is when $\lambda^{i\star}$ is positive) would require to numerically find the roots of a high-degree polynomial, which is expensive and inexact. To alleviate this problem, we introduce Assumption 1.

**Assumption 1.** *For each $i \in \{1, \dots, d\}$, there exists $\beta_i \in \mathbb{R}^*_+$ such that $\beta_{ij} = \beta_i$ for all $j \in \{1, \dots, d\}$.*

As we will see in Lemma 3.1, this assumption enables to recover the monotony of the conditional intensities between two times. It remains now to determine when the underlying intensity $\lambda^{i\star}$ is negative. To do so, we leverage the work done in Bonnet et al. [2021] and define the restart times in the multivariate framework, to be, for any $k$ and $i$:

$$T_{(k)}^{i\star} = \min\left(\inf\{t \geq T_{(k)} \colon \lambda^{i\star}(t) \geq 0\}, T_{(k+1)}\right).$$

Lemma 3.1 formalises when $\lambda^i(t) > 0$ with respect to restart times $T_k^i$.

**Lemma 3.1.** *If Assumption 1 is granted, then for each $i \in \{1, \dots, d\}$ and any $k > 1$:*

$$T_{(k)}^{i\star} = \min\left(T_{(k)} + \beta_i^{-1} \log\left(\frac{\mu^i - \lambda^{i\star}(T_{(k)})}{\mu^i}\right)\mathbb{1}_{\{\lambda^{i\star}(T_{(k)})<0\}}, T_{(k+1)}\right). \tag{4}$$

*Furthermore, if $T_{(k)}^{i\star} < T_{(k+1)}$, then*

$$\lambda^i(t) = \lambda^{i\star}(t) > 0 \quad \text{for any } t \in (T_{(k)}^{i\star}, T_{(k+1)}).$$

*Proof.* Let $i \in \{1, \dots, d\}$. For any $k \geq 1$, the underlying intensity function $\lambda^{i\star}$ in the interval $[T_{(k)}, T_{(k+1)})$ can be written:

$$\lambda^{i\star}(t) = \mu^i + \sum_{j=1}^{d} \sum_{\ell=1}^{N^j(t)} \alpha_{ij} \mathrm{e}^{-\beta_{ij}(t-T_\ell^j)},$$

this function is differentiable and we obtain:

$$(\lambda^{i\star})'(t) = -\sum_{j=1}^{d} \beta_{ij} \sum_{\ell=1}^{N^j(t)} \alpha_{ij} \mathrm{e}^{-\beta_{ij}(t-T_\ell^j)}.$$

By using Assumption 1 that for all $j \in \{1, \dots, d\}$, $\beta_{ij} = \beta_i \in \mathbb{R}_+^*$, we obtain the following differential equation:

$$(\lambda^{i\star})'(t) = -\beta_i \left(\lambda^{i\star}(t) - \mu^i\right),$$

which by solving on the interval gives:

$$\lambda^{i\star}(t) = \mu^i + \left(\lambda^{i\star}(T_{(k)}) - \mu^i\right)\mathrm{e}^{-\beta_i(t-T_{(k)})}. \tag{5}$$

In particular, the derivative of the underlying intensity function is of opposite sign as $(\lambda^{i\star}(T_{(k)}) - \mu^i)$. Let us distinguish two cases:

- If $\lambda^{i\star}(T_{(k)}) \geq 0$, then,

$$T^{i\star}_{(k)} = T_{(k)} = \min\left(T_{(k)}, T_{(k+1)}\right)$$
$$= \min\left(T_{(k)} + \beta_i^{-1}\log\left(\frac{\mu^i - \lambda^{i\star}(T_{(k)})}{\mu^i}\right)\mathbb{1}_{\{\lambda^{i\star}(T_{(k)})<0\}}, T_{(k+1)}\right).$$

  If $(\lambda^{i\star}(T_{(k)}) - \mu^i) \geq 0$, then $\lambda^{i\star}$ is decreasing and lower-bounded by $\mu^i$. If $(\lambda^{i\star}(T_{(k)}) - \mu^i) < 0$ then $\lambda^{i\star}$ is increasing and lower-bounded by zero. In both cases, for any $t \in (T^{i\star}_{(k)}, T_{(k+1)})$, $\lambda^{i\star}(t) > 0$ and then $\lambda^{i\star} = \lambda^i(t)$.

- If $\lambda^{i\star}(T_{(k)}) < 0$, then $\left(\lambda^{i\star}\left(T_{(k)}\right) - \mu^i\right) < 0$ so $\lambda^{i\star}$ is strictly increasing. By denoting $\lambda^i(T_k^{i-}) := \lim_{t\to T_k^{i-}} \lambda^i(t)$:

  - If $\lambda^i(T^{i-}_{(k+1)}) \leq 0$, then for any $t \in [T_{(k)}, T_{(k+1)})$, $\lambda^{i\star}(t) < 0$ and so $T^{i\star}_{(k)} = T_{(k+1)}$.
  - If $\lambda^i(T^{i-}_{(k+1)}) > 0$, then by the intermediate value theorem, there exists $t^\star \in (T_{(k)}, T_{(k+1)})$ such that $\lambda^{i\star}(t^\star) = 0$. By using Equation (5), we obtain:

$$t^\star = T_{(k)} + \beta_i^{-1}\log\left(\frac{\mu^i - \lambda^{i\star}(T_{(k)})}{\mu^i}\right),$$

    and by definition $T^{i\star}_{(k)} = t^\star < T_{(k+1)}$. Lastly, for any $t \in (T^{i\star}_{(k)}, T_{(k+1)})$, $\lambda^{i\star}(t) > 0$ and then $\lambda^{i\star}(t) = \lambda^i(t)$.

Combining all scenarios achieves the proof.

□

Figure 2 illustrates all three different scenarios. In particular, the term

$$T_{(k)} + \beta_i^{-1}\log\left(\frac{\mu^i - \lambda^{i\star}(T_{(k)})}{\mu^i}\right)$$

can be seen as the solution to the equation $\mu^i + (\lambda^{i\star}(T_{(k)}) - \mu^i)\mathrm{e}^{-\beta_i(t-T_{(k)})} = 0$ on the interval $[T_{(k)}, +\infty)$ when $\lambda^{i\star}(T_{(k)}) < 0$.

**Remark.** *This model with constant recovery rates $\beta_i$ has been studied before in the works of Ogata [1981] in the self-exciting version of the process. Intuitively, this assumption considers the situation where the rate of "dissipation" of any internal or external effect is*

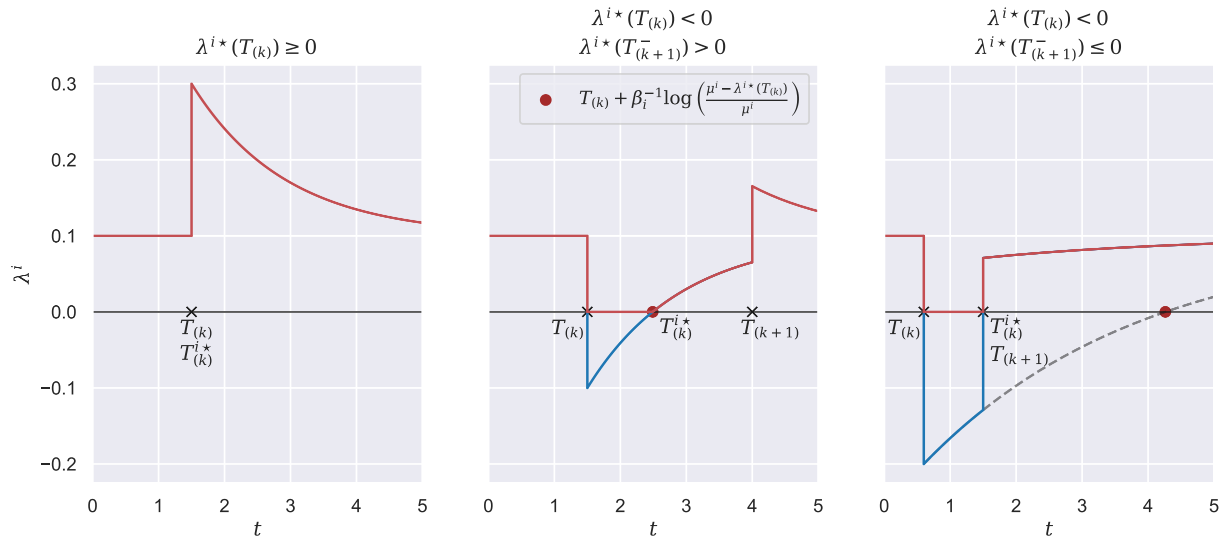



Figure 2: Illustration of three possible scenarios for restart times $T^{i\star}_{(k)}$ depending on the sign of $\lambda^{i\star}(T_{(k)})$ and $\lambda^{i\star}(T^-_{(k+1)})$. The pointed line in the last scenario shows the equation $\mu^i + (\lambda^{i\star}(T_{(k)}) - \mu^i)\mathrm{e}^{-\beta_i(t-T_{(k)})}$ and the term $T_{(k)} + \beta_i^{-1}\log\left(\frac{\mu^i-\lambda^{i\star}(T_{(k)})}{\mu^i}\right)$ as its only root.

*dependent only on the receiving phenomenon. For instance, for neuronal interactions, each activation from neuron $j$ will have an impact on a connected neuron $i$ dependent on both neurons $(\alpha_{ij})_{ij}$ but the "recovery" time can be assumed to depend only on the connected neuron $i$ $(\beta_i)_i$.*

**Proposition 3.2.** *[Compensator for multivariate exponential kernels] Let $N$ be a multivariate Hawkes process with exponential kernels. We suppose that Assumption 1 is granted. For any $i \in \{1, \dots, d\}$ the compensator $\Lambda^i$ of process $N^i$ reads:*

$$\Lambda^i(t) = \begin{cases} \mu^i t & \text{if } t < T_{(1)} \\ \mu^i T_{(1)} + \sum_{k=1}^{N(t)-1} \int_{T^{i\star}_{(k)}}^{T_{(k+1)}} \lambda^{i\star}(u)\,\mathrm{d}u + \int_{T^{i\star}_{(N(t))}}^{t} \lambda^{i\star}(u)\,\mathrm{d}u & \text{if } t \geq T_{(1)}\,, \end{cases} \tag{6}$$

*where:*

$$\int_{T^{i\star}_{(k)}}^{\tau} \lambda^{i\star}(u)\,\mathrm{d}u = \mu^i(\tau - T^{i\star}_{(k)}) + \beta_i^{-1}(\lambda^{i\star}(T_{(k)}) - \mu^i)(\mathrm{e}^{-\beta_i(T^{i\star}_{(k)} - T_{(k)})} - \mathrm{e}^{-\beta_i(\tau - T_{(k)})})\,, \tag{7}$$

*for* $\tau \in [T^{i\star}_{(k)}, T_{(k+1)}]$ *and with the conventions that the sum is equal to* $0$ *if* $N(t) = 1$ *and the last integral is equal to 0 if* $t < T^{i\star}_{(N(t))}$.

*Proof.* Equation (6) is obtained by splitting the expression of $\Lambda^i$ on each interval $[T_{(k)}, T_{(k+1)})$ for any $k \in \{0, N(t) - 1\}$. By using Lemma 3.1 we replace $\lambda$ by $\lambda^{i\star}$. For Equation (7), the integral can be directly obtained from the expression of $\lambda^{i\star}$ (5) from the proof in Lemma 3.1. □

### 3.4 Maximum likelihood estimation

As expected, Proposition 3.2 makes it possible to compute explicitly the log-likelihood expressed in Equation (3) for multivariate exponential Hawkes processes. This is formalised in Corollary 3.2.1.

**Corollary 3.2.1.** *Let* $\mathcal{P}$ *be the parametric exponential model for a multivariate Hawkes process of dimension* $d$*, defined through a parametric model* $\mathcal{P}_i$ *for each process* $N^i$*:*

$$\mathcal{P}^i = \left\{ \lambda^i_{\theta_i} = \left( \mu^i + \sum_{j=1}^{d} \int_{-\infty}^{t} \alpha_{ij} \mathrm{e}^{-\beta_i (t-s)} \, \mathrm{d}N^j(s) \right)^+ : \theta_i = (\mu^i, \alpha_{i1}, \dots, \alpha_{id}, \beta_i) \in \Theta \right\},$$

*with* $\Theta = \mathbb{R}^\star_+ \times \mathbb{R}^d \times \mathbb{R}^\star_+$*. We can then write* $\mathcal{P}$ *as:*

$$\mathcal{P} = \left\{ \lambda_\theta = \sum_{i=1}^{d} \lambda^i_{\theta_i} : \theta = (\theta_i) \in \Theta^d \right\},$$

*with underlying intensity functions* $\lambda^{i\star}_\theta$*, restart times* $(T^{i\star}_{\theta,(k)})_{k \in \{1,\dots,N(t)\}}$ *and compensator functions* $\Lambda^i_\theta$*.*

*For any* $\theta \in \Theta$*, the log-likelihood of the* $i$*-th process is equal to:*

$$\ell^i_t(\theta) = \log \mu^i + \sum_{k=2}^{N^i(t)} \log \left( \mu^i + (\lambda^{i\star}_\theta (T_{(N(T^i_k)-1)}) - \mu^i) \mathrm{e}^{-\beta_i (T^i_k - T_{(N(T^i_k)-1)})} \right) - \Lambda^i_\theta(t) \,, \qquad (8)$$

*with* $\Lambda^i_\theta$ *given by Equation (6) and with convention* $\log(x) = -\infty$ *for* $x \leq 0$.

*Proof.* In order to obtain $\lambda^{i\star}_\theta(T^{i-}_k)$ let us start by noting that if $k = 1$ then $\lambda^{i\star}_\theta(T^{i-}_k) = \mu^i$. For $k \geq 2$, we are working in the interval $[T_{(N(T^i_k)-1)}, T_{(N(T^i_k))})$ as $T_{(N(T^i_k))} = T^i_k$ by properties of a point process. Using then Equation (5) in the proof of Lemma 3.1 we obtain for any $i$:

$$\lambda_\theta^{i\star}(T_k^{i-}) = \begin{cases} \mu^i & \text{if } k = 1\,, \\ \mu^i + (\lambda_\theta^{i\star}(T_{(N(T_k^i)-1)}) - \mu^i)\mathrm{e}^{-\beta_i(T_k^i - T_{(N(T_k^i)-1)})} & \text{if } k \geq 2\,. \end{cases}$$

This result along with Proposition 3.2 gives the result. □

Algorithm 1 in Appendix A presents the iterative computation of the likelihood using (8). In particular, the complexity of the computation is $O(N(t) \times d)$.

### 3.5 Goodness-of-fit

As a benefit of our approach, it is possible to perform a goodness-of-fit test for assessing the quality of estimations. Indeed, the closed-form expression of the compensator given in Proposition 3.2 enables to use the Time Change Theorem for inhomogeneous Poisson processes [Daley and Vere-Jones, 2003, Proposition 7.4.IV]. For any $i$, the sequence of transformed times $(\Lambda^i(T_k^i))_k$ is a realisation of a homogeneous Poisson process with unit-intensity if and only if $(T_k^i)_k$ is a realisation of a point process with intensity $\lambda^i$.

We can then define for any $\theta \in \Theta$ the null hypothesis

$$\mathcal{H}_i : \text{“}(T_k^i)_k \text{ is a realisation of a point process with intensity } \lambda_\theta^i\text{”}.$$

The hypothesis is then tested via a Kolmogorov-Smirnov test between the empirical distribution $(\Lambda_\theta^i(T_{k+1}^i) - \Lambda_\theta^i(T_k^i))_{k\geq 1}$ and an exponential distribution with parameter 1. We obtain then $d$ different tests with $p$-values $(p_i)_{i\geq 1}$. Using multiple testing approaches can help in determining correctly estimated processes.

Lastly, we can obtain an additional test by considering the entire sequence of times $(T_{(k)})_{k\geq 1}$ and the total intensity $\lambda$. We obtain then the null hypothesis

$$\mathcal{H}_{tot} : \text{“}(T_{(k)})_k \text{ is a realisation of a point process with intensity } \lambda_\theta\text{”},$$

with corresponding $p$-value $p_{tot}$.

In the forthcoming sections, this testing procedure is applied to several realisations of event times, that are independent of the considered estimator. This enables to assess properly the accuracy of estimations, without knowing the true conditional intensities. This is particularly interesting for real-world data.

# 4 Illustration on synthetic datasets

## 4.1 Simulation procedure

In order to assess the performance of the maximum likelihood estimation method, we simulate different data by using Ogata's thinning method Ogata [1981]. This method consists in defining a piecewise constant function $\lambda^+$ such that for any $k > 1$ and any $t \in [T_{(k)}, T_{(k+1)})$, $\lambda^+(t) \geq \lambda(t)$. For this, we define $\lambda^+$ for any $t \in [T_{(k)}, T_{(k+1)})$ as

$$\lambda^+(t) = \sum_{i=1}^{d} \left( \mu^i + \sum_{j=1}^{d} \int_0^t \alpha_{ij}^+ \mathrm{e}^{-\beta_i (T_{(k)} - s)} \, \mathrm{d}N^j(s) \right),$$

which corresponds to considering only the positive interactions.

Four different parameter sets are considered: three sets for 2-dimensional Hawkes processes and a last one for a 10-dimensional process. Table 1 presents the parameters used in Dimension 2. All scenarios contain at least one negative interaction ($\alpha_{ij} < 0$). Scenario (1) is a Hawkes process where all parameters are non-null whereas Scenarios (2) and (3) are chosen to study the performance of our method when estimating null interactions ($\alpha_{12}$ for Scenario (2) and $\alpha_{21}$ for Scenario (3)). All simulations have exactly 5000 event times in total.

| Scenario | (1) | (2) | (3) |
|---|---|---|---|
| $\begin{pmatrix} \mu^1 \\ \mu^2 \end{pmatrix}$ | $\begin{pmatrix} 0.5 \\ 1.0 \end{pmatrix}$ | $\begin{pmatrix} 0.7 \\ 1.0 \end{pmatrix}$ | $\begin{pmatrix} 1.2 \\ 1.0 \end{pmatrix}$ |
| $\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}$ | $\begin{pmatrix} -1.9 & 3.0 \\ 1.2 & 1.5 \end{pmatrix}$ | $\begin{pmatrix} 0.2 & 0.0 \\ -0.6 & 1.2 \end{pmatrix}$ | $\begin{pmatrix} -1.0 & 0.1 \\ 0.0 & -0.8 \end{pmatrix}$ |
| $\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ | $\begin{pmatrix} 5.0 \\ 8.0 \end{pmatrix}$ | $\begin{pmatrix} 3.0 \\ 2.0 \end{pmatrix}$ | $\begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ |

Table 1: Parameters for simulations of two-dimensional Hawkes processes.

In order to carry out the hypothesis testing procedure, we simulate a sample of Hawkes processes independent from the one used for the estimation. Each testing sample contains as many realisations as the estimation sample. All $p$-values presented in the paper correspond to the average obtained over all realisations.

## 4.2 Proposed methods and comparison to existing procedures

The main focus of this paper is to assess the performance of the maximum likelihood estimator to correctly detect the interacting functions of our processes (without ambiguity, estimators are denoted with a tilde: $(\tilde{\mu}^i)_i$, $(\tilde{\alpha}_{ij})_{ij}$, $(\tilde{\beta}_i)_i$ and $(\tilde{h}_{ij})_{ij}$). To do so, we consider two versions of the method presented here:

- (MLE) The estimator obtained by minimising the opposite of the log-likelihood $-\sum_{i=1}^{d} \ell_t^i(\theta)$ (see Equation (8)). The log-likelihood is computed via Algorithm 1 and the minimisation is done with the L-BFGS-B method [Byrd et al., 1995].
- (MLE-$\varepsilon$) The estimator obtained by adding a thresholding step to the previous method, similar to the cumulative percentage of total variation approach used in Principal Component Analysis [Joliffe, 2002, Section 6.1.1]. All estimated values $|\tilde{\alpha}_{ij}|$ are arranged in increasing order, the cumulative sum $(s_k)_k$ is computed and all estimations $\tilde{\alpha}_{ij}$ such that $s_k < \varepsilon \sum_{p,q} |\tilde{\alpha}_{pq}|$ are set to zero, for a threshold $\varepsilon \in (0,1)$. The value of $\varepsilon$ is chosen over a grid of values as the one achieving the highest mean over all $p$-values associated to the goodness-of-fit tests described in Section 3.5. All non-null estimations $\tilde{\alpha}_{ij}$ are then re-estimated by maximising the log-likelihood.

The method MLE-$\varepsilon$ is proposed as a way of estimating the null interactions, as in the cases of Scenarios (2) and (3) for the 2-dimensional processes, and also for the 10-dimensional setting.

**Remark.** *Another option considered for MLE-$\varepsilon$ is to use instead the values $|\tilde{\alpha}_{ij}/\tilde{\beta}_i|$ for the thresholding. Numerical results slightly differ between the two methods, with the retained method showing better overall $p$-values.*

We compare our estimations to three popular approaches from the literature.

1. (Approx) The first one [Lemonnier and Vayatis, 2014] is obtained by approaching the compensator $\Lambda^i(t)$ (in each log-likelihood $\ell_t^i(\theta)$) by

$$\int_0^t \lambda^{i\star}(u)\,\mathrm{d}u\,.$$

In the case where all interactions are positive, this integral is equal to the compensator. The difference is when interactions are negative as this integral takes into account the negative values of the underlying intensity function.

2. The other two methods minimise the least-squares loss approximation defined as Reynaud-Bouret et al. [2014], Bacry et al. [2020]:

$$R_t(\theta) = \int_0^t (\lambda_\theta(u))^2\,\mathrm{d}u - \frac{2}{t}\sum_{k=1}^{N(t)} \lambda_\theta^{m_k}(T_{(k)}^-)\,,$$

which is an observable approximation of $\|\lambda_\theta - \lambda\|_t^2 = \int_0^t (\lambda_\theta(u) - \lambda(u))^2\,\mathrm{d}u$ up to a constant term. In Bacry et al. [2020], all interactions are assumed to be positive, however the implemented version of this method in the package `tick` Bacry et al. [2018] allows to retrieve negative values. For this, we consider two different kernel functions from this implementation:

   - (Lst-sq) $h_{ij}(t) = \alpha_{ij}\beta_{ij}e^{-\beta_{ij}t}$, where $\beta_{ij}$ is fixed beforehand by the practitioner. In practice, we fix $\beta_{ij} = \beta_i$ to be consistent with our model (see Assumption 1). The only solver in the implementation that provides negative values is BFGS, which is limited to work with an $\ell_2$-penalty. The grid of values $\{1, 10, \ldots, 10^6\}$ is considered for the regularisation constant. To obtain the best estimation for this method, we choose the constant that minimises the relative squared error over all estimated parameters.

   - (Grid-lst-sq) $h_{ij}(t) = \sum_{u=1}^{U} \alpha_{ij}^u \beta^u e^{-\beta^u t}$, with $(\beta^u)_u$ a fixed grid of parameters. In our case, we choose $U = d$ and the grid contains each parameter $\beta_i$. Intuitively, by applying an $\ell_1$ penalty, this method would be able to retrieve the corresponding parameter $\beta_i$ for each process. However, in practice, the implementation uses BFGS as optimiser and is limited to work with an $\ell_2$-penalty.

As for Lst-sq, the regularisation parameter is chosen over the grid of values by minimising the relative squared error.

### 4.2.1 Results on bivariate Hawkes processes

For the two-dimensional Hawkes processes, we simulate 25 realisations for each parameter set and an estimation is obtained for each individual simulation. All estimations are then averaged. We begin by comparing methods MLE and MLE-$\varepsilon$ along with both Approx and Lst-sq which are the two methods with the same kernel functions considered in this paper. Figure 3 represents the relative squared errors for each group of parameters by considering vector norms.

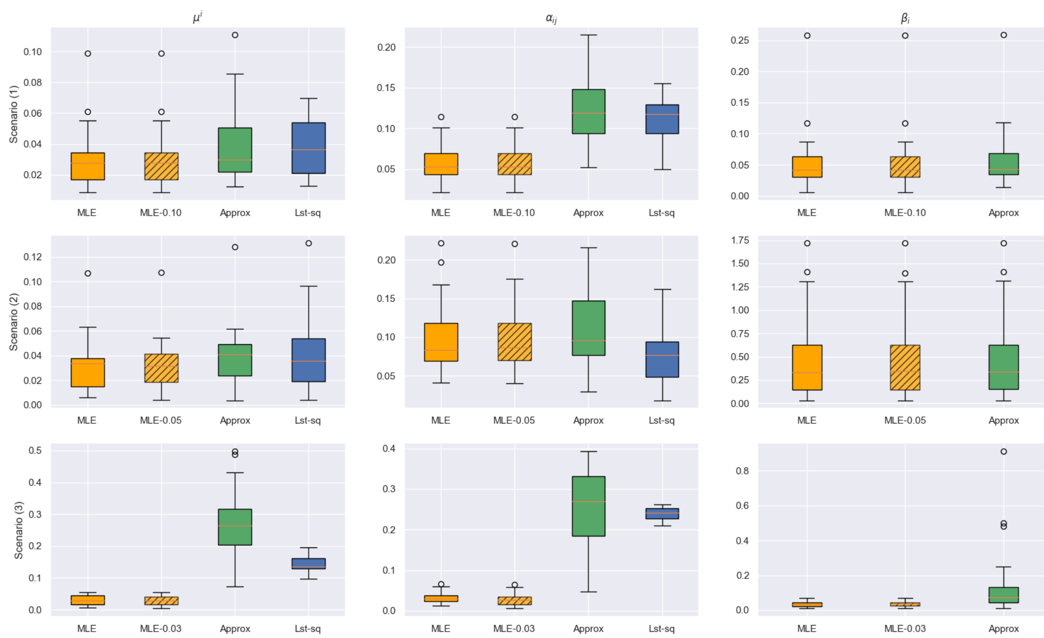



Figure 3: Boxplots representing the relative squared error for each group of parameters for two-dimensional Hawkes processes. Lst-sq does not appear in the last column because it is parameterised with the true values of $(\beta_i)_i$. The proposed methods are MLE and MLE-$\varepsilon$.

First, we observe that delay factors $(\beta_i)_i$ (last column of Figure 3) are similarly estimated by all approaches. Let us recall that Lst-sq is not included in the comparison of delay factors: since it requires to provide a value for these parameters (they are not estimated), it was given the true values of $(\beta_i)_i$ as input. An alternative offered by `tick` is to provide a grid of values, but this approach, denoted Grid-lst-sq, is included in the comparison at the end of the section because of its difference with the exponential model considered here.

Then, regarding the baseline intensities $(\mu_i)_i$ and the interaction factors $(\alpha_{ij})_{ij}$, both MLE and MLE-$\varepsilon$ outperform the two other approaches. In Scenario (2), all estimation methods perform reasonably well. This can be explained by the weak inhibiting effect of the interaction $1 \to 2$, leaving the intensity almost always positive.

In Scenario (1), estimations of Approx and Lst-sq are altered because the inhibiting effect is stronger than in Scenario (2). The major changes appear in Scenario (3), where both Approx and Lst-sq obtain very high relative errors. More precisely, they fail to explain the interactions between the two processes (see the estimations $(\tilde{\alpha}_{ij})_{ij}$ in the middle column of Figure 3), which is compensated by a wrong estimation $(\tilde{\mu}_i)_i$ of baseline intensities. This is not surprising since Scenario (1), and even more Scenario (3), were designed so that the intensity functions are frequently equal to zero, which induces major differences between true and underlying intensities. Since Approx and Lst-sq are both based on assuming that these two functions are almost equal, the violation of this assumption causes large estimation errors. As expected, MLE and MLE-$\varepsilon$, which are developed to handle such inhibiting scenarios, provide accurate estimations.

These results are confirmed by the outcomes of the goodness-of-fit test displayed in Table 2. It shows indeed the averaged $p$-values for each scenario using both the true parameters and all four estimations from Figure 3 with 25 simulations different from the ones used for estimation. In particular, we can see that MLE and MLE-$\varepsilon$ obtain high $p$-values, being very close to those obtained using the true parameters. Table 2 also highlights when parameters are incorrectly estimated. For instance, in Scenario (1), Approx correctly estimate Process 2 but fails to estimate Process 1 (the $p$-value is almost half the one obtained with the true parameters), which is the one characterised by a self-inhibiting behaviour. In addition, at least one of the two proposed methods obtains the best value for $p_{tot}$ in each scenario, which illustrates the ability of these procedures to reconstruct the complete process $N$. Let us note that the very low $p$-values obtained by Approx and Lst-sq for Scenario (3) confirm the ability of the goodness-of-fit procedure to detect when the parameter estimations strongly differ from the true parameters.

| | Scenario (1) | | | Scenario (2) | | | Scenario (3) | | |
|---|---|---|---|---|---|---|---|---|---|
| $p$-value | $p_1$ | $p_2$ | $p_{tot}$ | $p_1$ | $p_2$ | $p_{tot}$ | $p_1$ | $p_2$ | $p_{tot}$ |
| True | 0.492 | 0.438 | 0.430 | 0.535 | 0.468 | 0.479 | 0.510 | 0.623 | 0.338 |
| MLE | 0.440 | 0.442 | **0.398** | 0.483 | 0.461 | 0.485 | 0.549 | 0.638 | **0.357** |
| MLE-$\varepsilon$ | 0.440 | 0.442 | **0.398** | 0.488 | 0.461 | **0.491** | 0.549 | 0.574 | 0.327 |
| Approx | 0.257 | 0.442 | 0.358 | 0.483 | 0.452 | 0.459 | 0.0 | 0.007 | 0.0 |
| Lst-sq | 0.154 | 0.438 | 0.392 | 0.534 | 0.463 | 0.478 | 0.0 | 0.0 | 0.0 |

Table 2: Average $p$-values for estimations of two-dimensional Hawkes processes for all scenarios. The values are averaged over 25 simulations. In bold the highest $p_{tot}$ value obtained among all estimation methods.

Lastly, let us investigate the estimations obtained via Grid-lst-sq, which can be used in practice as a way to estimate the parameters $\beta_i$ by providing a grid of possible parameters. Let us mention that both of the previous comparisons (boxplots and $p$-values) cannot be done here due to the difference in the number of parameters, but we can compare the methods in terms of reconstructions $\tilde{h}_{ij}$ of the interaction functions $h_{ij}$. For this purpose, we analyse Figure 4, which represents the estimated interaction functions $\tilde{h}_{ij}$ for all five methods in Scenario (3). Interestingly, we see that Grid-lst-sq performs similarly to Lst-sq, while being fed with all true values $(\beta_i)_i$ for each interaction. However, we see that Grid-lst-sq suffers from the same difficulties than Approx and Lst-sq, which was expected since it relies on the same unvalid assumption. Let us note that we chose to display the results for Scenario (3) since it highlights the main differences between the compared approaches but the reconstructions for Scenarios (1) and (2) can be found in Appendix B (Figures 10 and 11).

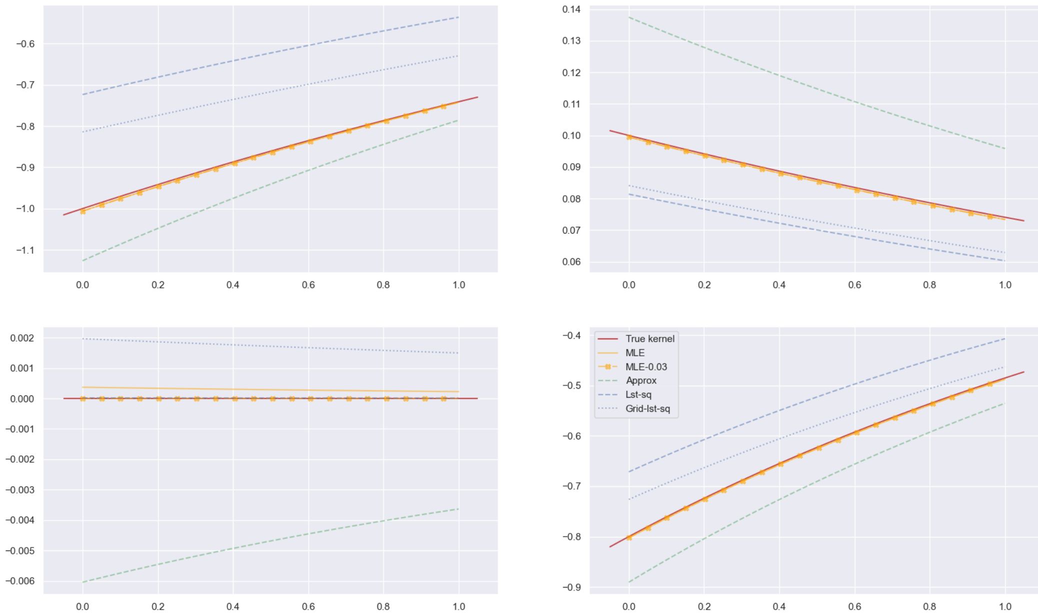


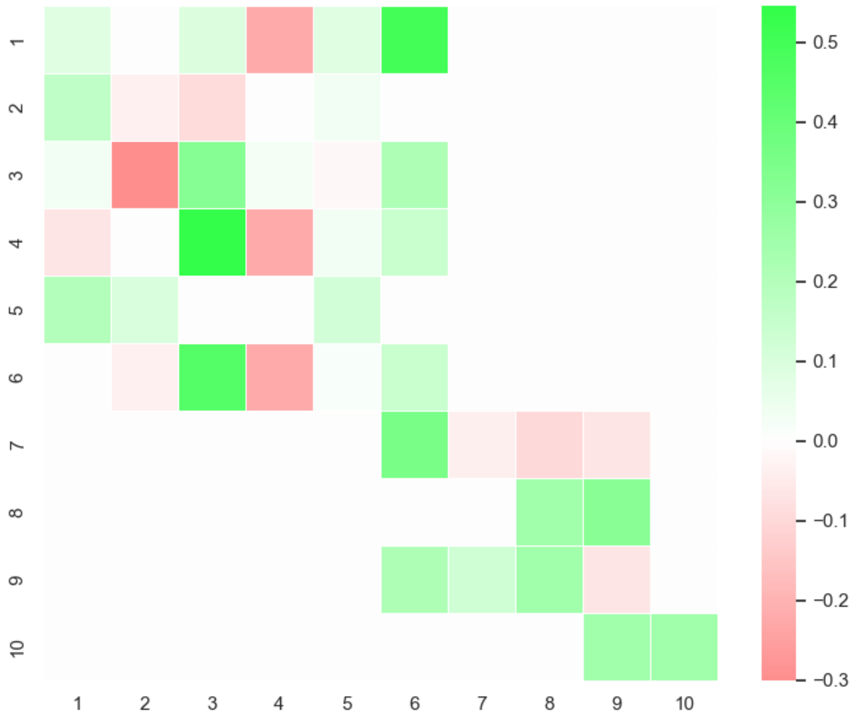

Figure 4: Reconstruction of interaction functions $h_{ij}$ for Scenario (3) of two-dimensional Hawkes processes along with all estimated functions $\tilde{h}_{ij}$. The real function is plotted in red and 25 estimations are averaged for each method.

#### 4.2.2 A 10-dimensional Hawkes process

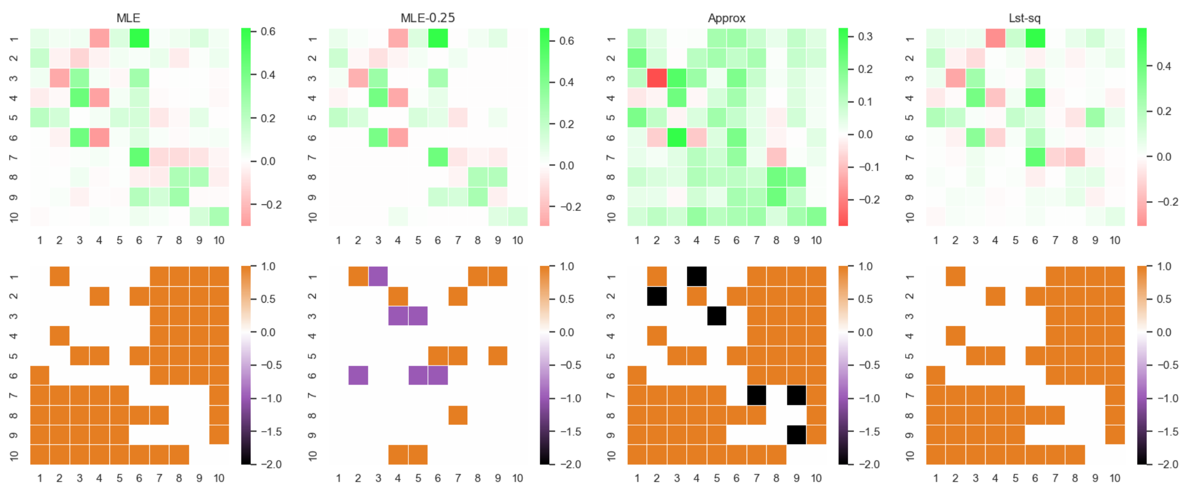
A 10-dimensional Hawkes process is simulated based on a set of parameters corresponding to the quantities $(sign(\alpha_{ij})\|h_{ij}\|_1)_{ij} = (\alpha_{ij}/\beta_i)_{ij}$ displayed in Figure 5. The chosen parameters fulfill the existence condition $\|S^+\|_\infty < 1$.

For the purpose of the numerical illustration and in order to keep the computation time of both MLE and MLE-$\varepsilon$ (which increases with the dimension $d$) affordable, only 5 different realisations are simulated. The corresponding estimations $\tilde{\alpha}_{ij}$ and $\tilde{\beta}_{ij}$ are then averaged and displayed in Figure 5. The heatmap representation is convenient for high-dimensional processes and it allows us to see whether the signs of each interaction are well-estimated and whether the null-interactions are correctly detected.

In this example we decided to keep only Approx and Lst-sq as comparison methods as these are the ones with the same parametrisation for the kernel functions. Among the four methods considered, Approx is the only one that wrongly estimates the sign of some interactions, represented by the black boxes in the second row matrix. MLE and



(a) Heatmap of real parameters.




(b) Top row corresponds to heatmap for each estimation method. Bottom row corresponds to errors made with respect to real parameters. A value of 1 (orange) shows an undetected 0 (non-null estimation for $\alpha_{ij} = 0$), a value of -1 (purple) shows a non-null value set to 0 and a value of -2 (black) shows a non-null value whose sign is wrongly estimated.

Figure 5: Heatmaps of $(sign(\alpha_{ij})\|h_{ij}\|_1)_{ij} = (\alpha_{ij}/\beta_i)_{ij}$ (for true and estimated parameters) for the 10-dimensional simulation.

Lst-sq correctly retrieve the sign of each interactions but are unable to detect the null interactions. MLE-$\varepsilon$ is the best concerning both the estimations of the interactions by identifying correctly excitation and inhibition, along with obtaining a matrix with a similar support as the original. Table 3 summarises the $p$-values for each hypothesis as described in Section 3.5. Both MLE and MLE-$\varepsilon$ obtain overall better $p$-values with no particularly low values, which is not the case for Approx (see $p_4$ and $p_9$) and for Lst-sq (see $p_8$ and $p_{10}$). In particular, MLE-$\varepsilon$ achieves the best $p_{tot}$ value, suggesting it is the best at reconstructing not only each individual sub-process but the entire process seen as a whole.

| $p$-value | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ | $p_7$ | $p_8$ | $p_9$ | $p_{10}$ | $p_{tot}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| True | 0.403 | 0.708 | 0.273 | 0.728 | 0.545 | 0.386 | 0.701 | 0.234 | 0.668 | 0.632 | 0.462 |
| MLE | 0.407 | 0.743 | 0.287 | 0.640 | 0.458 | 0.384 | 0.638 | 0.240 | 0.617 | 0.123 | 0.336 |
| MLE-0.25 | 0.439 | 0.721 | 0.289 | 0.665 | 0.497 | 0.311 | 0.631 | 0.269 | 0.621 | 0.119 | **0.403** |
| Approx | 0.354 | 0.407 | 0.450 | 0.035 | 0.394 | 0.219 | 0.390 | 0.339 | 0.080 | 0.135 | 0.282 |
| Lst-sq | 0.427 | 0.712 | 0.204 | 0.604 | 0.447 | 0.313 | 0.607 | 0.043 | 0.572 | 0.036 | 0.282 |

Table 3: $p$-values for estimations of a ten-dimensional Hawkes process. The values are averaged over 5 simulations. In bold the highest $p_{tot}$ value obtained among all estimation methods.

# 5 Application on neuronal data

## 5.1 Preprocessing and data description

In this section we present the results obtained by our estimation method applied to a collection of 10 trials consisting in the measurement of spike trains of 223 neurons from the lumbar spinal of a red-eared turtle. This data are first presented in Petersen and Berg [2016] and then also analysed in Bonnet et al. [2022] to study how the activity of a group of neurons impacts the membrane potential's dynamic of another neuron. Events were registered for 40 seconds and in order to take into account eventual stationarity we only consider the events that took place on the interval $[11, 24]$ (see Bonnet et al. [2022] for further details). Among all trials, each neuron recording contains between 54 and 4621 time events. Furthermore, we divide our samples in a training set consisting on all events in half the interval $[11, 17.5]$ and a test set consisting on the remaining window $[17.5, 24]$, in particular each neuron has at least 15 time events in each set. The training sets are used for obtaining the estimations and the test sets for performing the goodness-of-fit tests.

### 5.2 Estimation results

In this section, we present the results obtained via MLE and MLE-$\varepsilon$. The estimation method is the same as in the previous section: an estimation is obtained for each individual trial and then a single estimation is obtained by averaging over each neuron. Figure 6 presents the heatmap matrix $(\tilde{\alpha}_{ij}/\tilde{\beta}_i)_{ij}$ of the estimation obtained through MLE. The estimation presents both excitation and inhibition effects between the neurons and in particular a negative diagonal equivalent to self-inhibition. An interpretation of this diagonal may be the refractory period of a neuron (minimal time before being able to be activated again).

However, in practice not all neurons interact with each other and an important part of the study of such interactions is to estimate the interaction graph between neurons. For this we apply then the MLE-$\varepsilon$ method. In the absence of any information about the existing neuronal connections, to determine the best threshold level $\varepsilon$ we compare for each estimation the $p$-values obtained via the method described in Section 3.5. Figure 7 shows the ordered $p$-values for each hypothesis $\mathcal{H}_i$ along with hypothesis $\mathcal{H}_{tot}$ for different values of $\varepsilon$. Being in a high-dimensional context, it is important to take into account a multiple testing procedure and for this purpose we choose the Benjamini-Hochberg method consisting in adapting the threshold of each $p$-value, represented in the figure by a blue line. Let us recall here how this method works : the Benjamini-Hochberg (B-H) method controls the false discovery rate (FDR). If we denote $V$ the number of rejected true null hypothesis and $S$ the number of rejected true alternative hypothesis, the FDR is defined

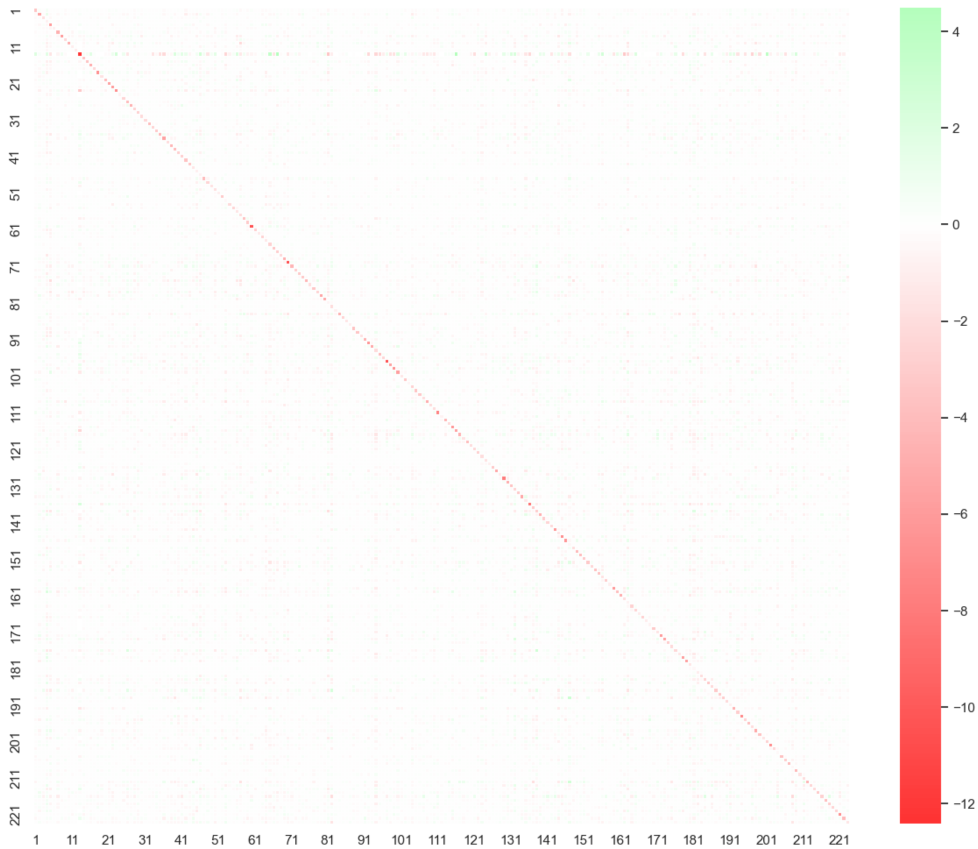



Figure 6: Heatmap matrix $(\tilde{\alpha}_{ij}/\tilde{\beta}_i)_{ij}$ for the MLE estimation on 223 neurons.

as

$$FDR = \mathbb{E}\left[\frac{V}{S+V}\right].$$

In other words we control the expected number of true null hypothesis (i.e. the process is indeed a Hawkes process) rejected by our testing method. The B-H procedure considers the ordered $p$-values $(p_{(k)})_{k\in\{1,\ldots,d+1\}}$ and compares each one to the adapted rejection threshold $\frac{\alpha k}{d+1}$. Then we determine the largest $K \in \{1,\ldots,d+1\}$ such that $p_{(K)} < \frac{\alpha K}{d+1}$ and we reject all hypothesis such that $p_{(k)} < p_{(K)}$.

We begin by noticing that as we increase the threshold $\varepsilon$ the $p$-values appear to increase progressively and that for values like $\varepsilon = 0.9$ and $\varepsilon = 0.95$ most hypothesis are not rejected after the multiple testing adjustment.

This suggests that the simpler the model the better $p$-values we obtain so we decided to include another estimation, named "Diag" on both graphics, consisting on fixing all $\alpha_{ij} = 0$

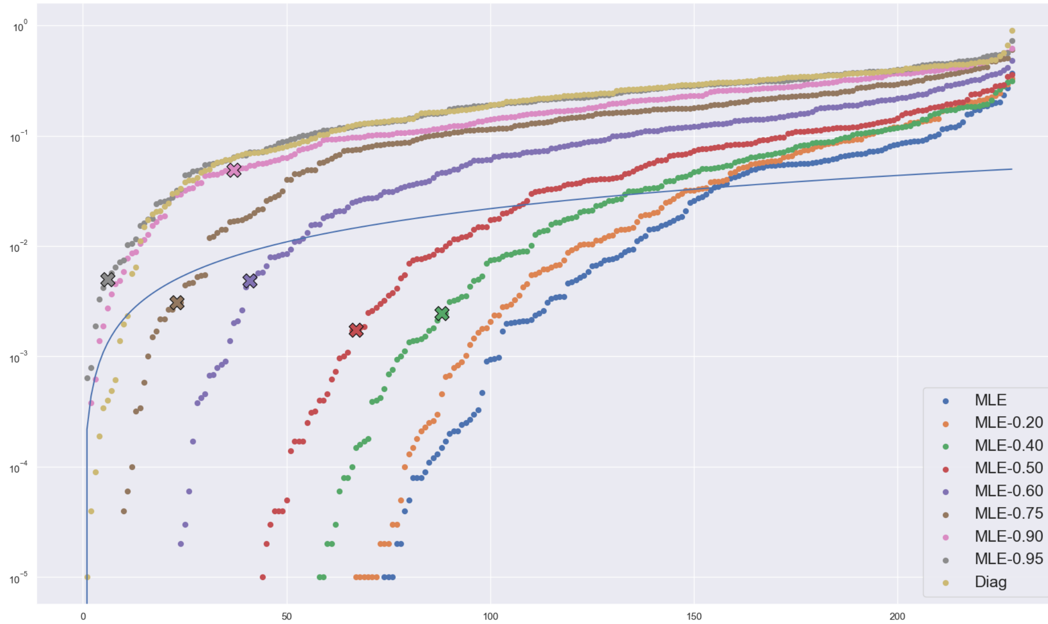



Figure 7: Ordered $p$-values for all hypothesis tests $\mathcal{H}_i$ and $\mathcal{H}_{tot}$. $p_{tot}$ appears as a cross for each model. The blue curve corresponds to the adapted rejection threshold from the B-H procedure, so all tests whose $p$-value are over the line are not rejected.

for $i \neq j$. This corresponds to a model where no interaction between neurons exists and we keep only self-interactions: in other words, each neuron is seen as a univariate Hawkes process with three parameters $(\mu, \alpha, \beta)$. Although most hypothesis $\mathcal{H}_i$ are not rejected by the method, the total $p$-value $p_{tot}$ is zero which suggests that although such a model could explain each dimension individually, it is unable to explain the neurons' interactions as a whole interconnected process. The values of $p_{tot}$ for all models are presented in Figure 7 by a cross and are summarised in Table 4. The model that best represents the complete process $N$ corresponds to $\varepsilon = 0.90$ with the highest value for $p_{tot}$ and with almost all hypothesis, including $\mathcal{H}_{tot}$, not rejected. This suggests that the estimations provided by MLE-0.90 are the best fit for explaining the entire process as well as each individual subprocess.

Figure 8 shows the heatmap matrix for $\varepsilon = 0.90$. This matrix contains around 10.2% of non-null entries in matrix $(\tilde{\alpha}_{ij})_{ij}$. First, we notice that among the non-zero interactions,

| | MLE | MLE-0.20 | MLE-0.40 | MLE-0.50 | MLE-0.60 | MLE-0.75 | MLE-0.90 | MLE-0.95 | Diag |
|---|---|---|---|---|---|---|---|---|---|
| $p_{tot}$ | 0.0 | 0.0 | 0.002 | 0.002 | 0.005 | 0.003 | **0.049** | **0.005** | 0.0 |

Table 4: Values of $p_{tot}$ for each estimation method for the neuronal dataset. In bold appear the $p$-values above the rejection threshold after Benjamini-Hochberg procedure.

we detect all types of interactions: mutual excitation, mutual inhibition, self-excitation, self-inhibition. This supports the relevance of carefully accounting for inhibition when developing inference procedures. Interestingly, we observe that, similarly to the results obtained with MLE (Figure 6), the diagonal still contains the larger estimated coefficients, though we notice that some signs changed after reestimation. This would suggest that some of the self-inhibiting effects that were detected without thresholding could be spurious and due to compensations of other interactions. Although the interpretation of this phenomenon is not straightforward, it highlights the necessity of correctly recover the support of interactions in order to provide accurate estimations. This point will be further discussed in Section 6.

Figure 9 displays the graph of interaction between neurons. We observe that a large number of neurons are densely connected with only some of them having a limited number of entering or exiting connections.

# 6 Discussion

In this paper, a workable methodology for estimating a multivariate exponential Hawkes process with potential inhibiting interactions is introduced. Our approach relies on the mild assumption that the delay factors $\beta_{ij}$ only depend on the affected process $N_i$, which enables to derive a closed-form expression for the compensator, appearing both in the log-likelihood and in the Time Change Theorem. Up to our knowledge, this constitutes the first exact frequentist inference and testing methods which are able to handle negative interaction functions.

While the independence of the delay factors with respect to the interacting processes is

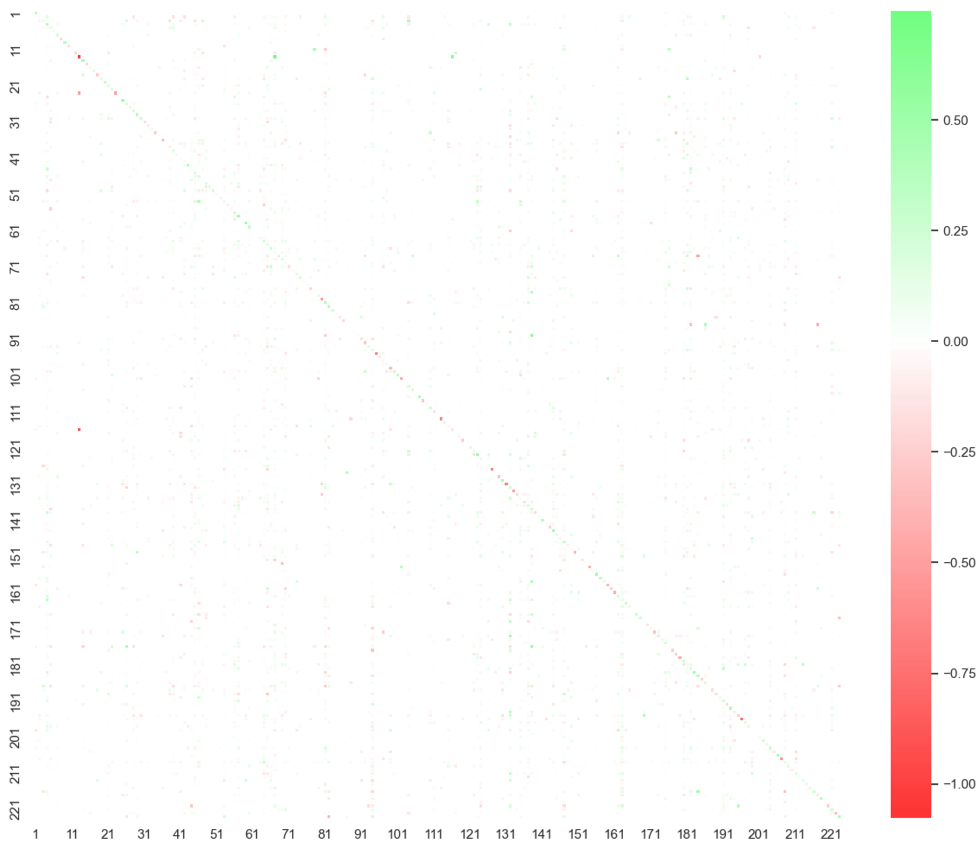
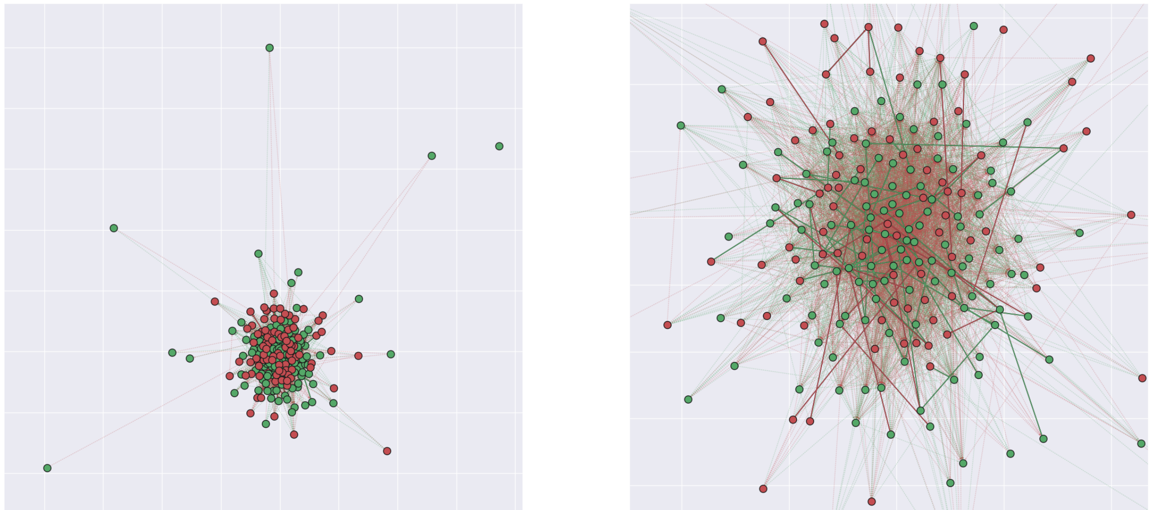


Figure 8: Heatmap of MLE-0.90 estimation on 223 neurons.

quite consistent with practical applications such as neuronal activity, it could be a limitation of our approach when considering heterogeneous phenomena. Going over this assumption would force us to use numerical integration methods and would considerably increase the computational time of the estimation procedure. This is obviously detrimental since, in practice, time sequences are increasingly abundant and large. On the other hand, improving the computational effectiveness of estimation procedures for Hawkes processes is a current direction of research [Bompaire et al., 2018].

This work focuses on a computational perspective of maximum likelihood estimation. It is naturally of interest to provide a theoretical study of the asymptotic behaviour of our estimator, as done for exciting Hawkes processes [Guo et al., 2018]. This work is currently under investigation.

From the applicative point of view, estimating accurately the support of the interactions



(a) Total graph.

(b) Zoomed version of graph of interactions.

Figure 9: Graph of interactions. A green edge corresponds to an exciting interaction and a red edge to an inhibiting one. Edges such that $|\tilde{\alpha}_{ij}|/\tilde{\beta}_i > 0.25$ are opaque. Self-interaction is represented by color of the nodes.

is a central question, besides knowing their nature. Indeed, the support estimation is crucial both for recovering the connectivity graph and for estimating the interaction functions since the non-zero coefficients are eventually re-estimated on the support. This matter deserves further investigations regarding for instance post hoc selection or regularised estimation [Bacry et al., 2020].

Let us also highlight that, because of the physical constraints of the experiment, only a fraction of the neuronal network is observed, which raises the question of interpretability of the estimated interactions. Indeed, the latter do not take into account the interactions with neurons that are outside the observed network. Very recent results tackle the consistency of estimated interactions in a partially observed network [Reynaud-Bouret et al., 2021]. A necessary condition to recover interactions in the subnetwork requires in particular to have a large number of interactions within the full network. Regarding the neuronal application, it could be of great interest to further investigate the interpretability of the inferred interactions and connectivity graph in light of the aforementioned work.

## Conflict of interest

The authors report there are no competing interests to declare.

### SUPPLEMENTARY MATERIAL

**Python package for multivariate Hawkes processes:** a Python package containing code for manipulation of multivariate Hawkes process, including:

1. simulation;
2. estimation via the proposed methods;
3. goodness-of-fit as described in this article;

is freely available online at the address

`https://github.com/migmtz/multivariate-hawkes-inhibition`.

The package also contains synthetic datasets for reproducibility of the results analysed in this article.

## References


A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83:9–27, 1988.

M.A. Rizoiu, Y. Lee, S. Mishra, and L. Xie. A tutorial on Hawkes processes for events in social media. *arXiv:1708.06401*, 2017.

J. Olinde and M.B. Short. A self-limiting Hawkes process: Interpretation, estimation, and use in crime modeling. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 3212–3219, 2020.

P. Reynaud-Bouret, R.C. Lambert, C. Tuleau-Malot, T. Bessaih, V. Rivoirard, Y. Bouret, and N. Leresche. Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *Journal of Neuroscience Methods*, 297:9–21, 2018.

A.G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11:493–503, 1974.

T. Ozaki. Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31:145–155, 1979.

X. Guo, A. Hu, R. Xu, and J. Zhang. Consistency and computation of regularized mles for multivariate Hawkes processes. *arXiv:1810.02955*, 2018.

J. Da Fonseca and R. Zaatour. Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 34:548–579, 2013.

E. Lewis and G. Mohler. A nonparametric em algorithm for multiscale Hawkes processes. *Journal of Nonparametric Statistics*, 1:1–20, 2011.

E. Bacry and J.F. Muzy. First- and Second-Order Statistics Characterization of Hawkes Processes and Non-Parametric Estimation. *IEEE Transactions on Information Theory*, 62:2184–2202, 2016.

M. Kirchner. An estimation procedure for the Hawkes process. *Quantitative Finance*, 17: 571–595, 2017.

Y. Yang, J. Etesami, N. He, and N. Kiyavash. Online learning for multivariate Hawkes processes. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24:1563–1588, 1996.

S. Chen, A. Shojaie, E. Shea-Brown, and D. Witten. The multivariate Hawkes process in high dimensions: Beyond mutual excitation. *arXiv:1707.04928v2*, 2017.

M. Costa, C. Graham, L. Marsalle, and V. Tran. Renewal in Hawkes processes with self-excitation and inhibition. *Advances in Applied Probability*, 52:879–915, 2020.

D. Sulem, V. Rivoirard, and J. Rousseau. Bayesian estimation of nonlinear Hawkes process. *arXiv:2103.17164v2*, 2021.

H. Mei and J. Eisner. The neural Hawkes process: A neurally self-modulating multivariate point process. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 6757–6767. Curran Associates Inc., 2017.

S. Zuo, H. Jiang, Z. Li, T. Zhao, and H. Zha. Transformer Hawkes process. In *Proceedings of the 37th International Conference on Machine Learning*, pages 11692–11702. PMLR, 2020.

C. Duval, E. Luçon, and C. Pouzat. Interacting Hawkes processes with multiplicative inhibition. *arXiv:2105.10597*, 2021.

A. Bonnet, M. Martinez Herrera, and M. Sangnier. Maximum likelihood estimation for Hawkes processes with self-excitation or inhibition. *Statistics & Probability Letters*, 179:109214, 2021.

P.C. Petersen and R.W. Berg. Lognormal firing rate distribution reveals prominent fluctuation-driven regime in spinal motor networks. *Elife*, 5:e18805, 2016.

M. Radosevic, A. Willumsen, P.C. Petersen, H. Lindén, M. Vestergaard, and R.W. Berg. Decoupling of timescales reveals sparse convergent cpg network in the adult spinal cord. *Nature Communications*, 10:1–14, 2019.

E. Bacry, M. Bompaire, S. Gaïffas, and J.F. Muzy. Sparse and low-rank multivariate Hawkes processes. *Journal of Machine Learning Research*, 21:1–32, 2020.

R. Lemonnier and N. Vayatis. Nonparametric markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate Hawkes processes. In *Machine Learning and Knowledge Discovery in Databases*, page 161–176. Springer Berlin Heidelberg, 2014.

D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications (New York). Springer-Verlag, second edition, 2003.

S. Mishra, M.A. Rizoiu, and L. Xie. Feature driven and point process approaches for popularity prediction. *arXiv:1608.04862v2*, 2016.

J.M. Chen, A.G. Hawkes, E. Scalas, and M. Trinh. Performance of information criteria for selection of Hawkes process models of financial data. *Quantitative Finance*, 18:225–235, 2018.

L. Adamopoulos. Cluster models for earthquakes: Regional comparisons. *Journal of the International Association for Mathematical Geology*, 8(4):463–475, 1976.

A. Veen and F.P. Schoenberg. Estimation of space–time branching process models in seismology using an em–type algorithm. *Journal of the American Statistical Association*, 103(482):614–624, 2006.

P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience*, 4:3, 2014.

J.G. Rasmussen. Bayesian inference for Hawkes processes. *Methodology and Computing in Applied Probability*, 15:623–642, 2013.

S. Donnet, V. Rivoirard, and J. Rousseau. Nonparametric bayesian estimation of multivariate Hawkes processes. *Annals of Statistics*, 48(5):2698 – 2727, 2020.

I. Deutsch and G.J. Ross. Bayesian estimation of multivariate Hawkes processes with inhibition and sparsity. *arXiv:2201.05009*, 2022.

Y. Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27:23–30, 1981.

R.H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16:1190–1208, 1995.

I.T. Joliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, second edition, 2002.

E. Bacry, M. Bompaire, P. Deegan, S. Gaïffas, and S.V. Poulsen. tick: a python library for statistical learning, with an emphasis on Hawkes processes and time-dependent models. *Journal of Machine Learning Research*, 18(214):1–5, 2018.

A. Bonnet, C. Dion-Blanc, F. Gindraud, and S. Lemler. Neuronal network inference and membrane potential model using multivariate hawkes processes. *Journal of Neuroscience Methods*, 372:109550, 2022. ISSN 0165-0270.

M. Bompaire, E. Bacry, and S. Gaïffas. Dual optimization for convex constrained objectives without the gradient-lipschitz assumption. *arXiv:1807.03545*, 2018.

P. Reynaud-Bouret, A. Muzy, and I. Bethus. Towards a mathematical definition of functional connectivity. *Comptes rendus. Mathématique*, 359:481–492, 2021.

# A Algorithm for computing the log-likelihood

**Algorithm 1:** Computation of the log-likelihood $\ell_t(\theta)$ of a multivariate exponential Hawkes process.

**Input** Parameters $\mu^i$, $\alpha_{ij}$, $\beta_i$ for $i, j \in \{1, \ldots, d\}$, list of event times and marks $(T_{(k)}, m_k)_{k=1:N(t)}$;

**Initialisation** Initialise for all $i$, $\Lambda_k^i = \mu^i T_{(1)}$, $\lambda^{i\star}(T_{(k)}^-) = \mu^i$, $\lambda_k^{i\star} = \mu^i + \alpha_{im_1}$ and $\ell_t(\theta) = \log(\lambda^{m_1\star}(T_{(k)}^-)) - \sum_{i=1}^d \Lambda_k^i$;

**for** $k = 2$ *to* $N(t)$ **do**

  Compute for all $i$, $T_{(k-1)}^{i\star} = \min\left(T_{(k-1)} + \beta_i^{-1} \log\left(\frac{\mu^i - \lambda_k^{i\star}}{\mu^i}\right) \mathbb{1}_{\{\lambda_k^{i\star} < 0\}}, T_{(k)}\right)$;

  Compute for all $i$,

  $\Lambda_k^i = \mu^i (T_{(k)} - T_{(k-1)}^{i\star}) + \beta_i^{-1} (\lambda_k^{i\star} - \mu^i)(\mathrm{e}^{-\beta_i (T_{(k-1)}^{i\star} - T_{(k-1)})} - \mathrm{e}^{-\beta_i (T_{(k)} - T_{(k-1)})})$;

  Compute for all $i$, $\lambda^{i\star}(T_{(k)}^-) = \mu^i + (\lambda_k^{i\star} - \mu^i) \mathrm{e}^{-\beta_i (T_{(k)} - T_{(k-1)})}$;

  Update $\ell_t(\theta) = \ell_t(\theta) + \log(\lambda^{m_k\star}(T_{(k)}^-)) - \sum_{i=1}^d \Lambda_k^i$;

  Compute for all $i$, $\lambda_k^{i\star} = \lambda^{i\star}(T_{(k)}^-) + \alpha_{im_k}$;

**end**

Compute for all $i$, $T_{(N(t))}^{i\star} = \min\left(T_{(N(t))} + \beta_i^{-1} \log\left(\frac{\mu^i - \lambda_k^{i\star}}{\mu^i}\right) \mathbb{1}_{\{\lambda_k^{i\star} < 0\}}, t\right)$;

Compute for all $i$,

$\Lambda_k^i = \left[\mu^i (t - T_{(N(t))}^{i\star}) + \beta_i^{-1} (\lambda_k^{i\star} - \mu^i)(\mathrm{e}^{-\beta_i (T_{N(t)}^{i\star} - T_{(N(t))})} - \mathrm{e}^{-\beta_i (t - T_{(N(t))})})\right] \mathbb{1}_{\{t > T_{(N(t))}^{i\star}\}}$;

Update $\ell_t(\theta) = \ell_t(\theta) - \sum_{i=1}^d \Lambda_k^i$;

**return** Log-likelihood $\ell_t(\theta)$.

# B Reconstructed interaction functions for synthetic data

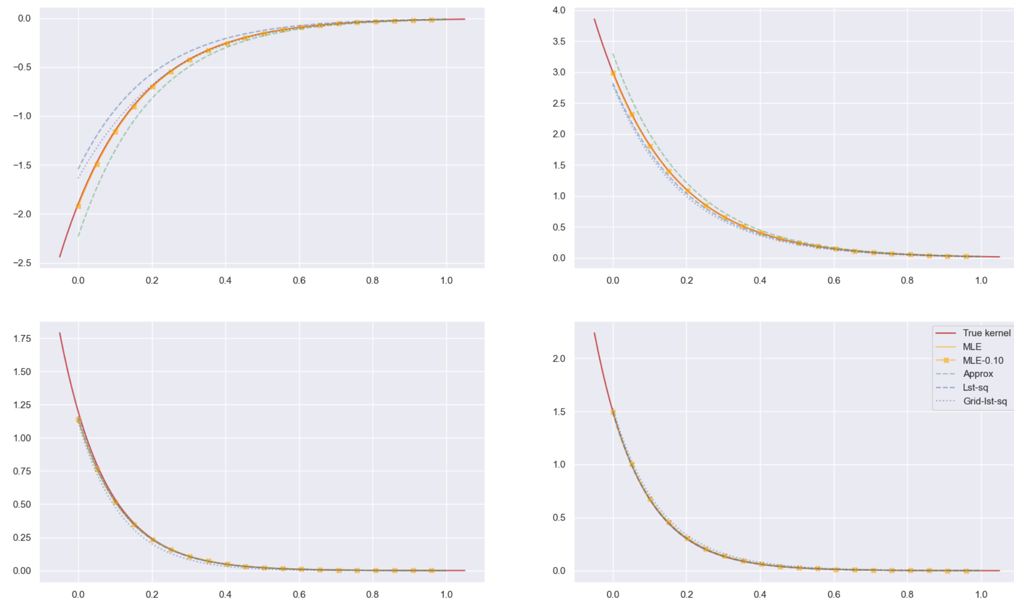



Figure 10: Reconstruction of interaction functions $h_{ij}$ for Scenario (1) of two-dimensional Hawkes processes along with all estimated functions $\tilde{h}_{ij}$. The real function is plotted in red and 25 estimations are averaged for each method.

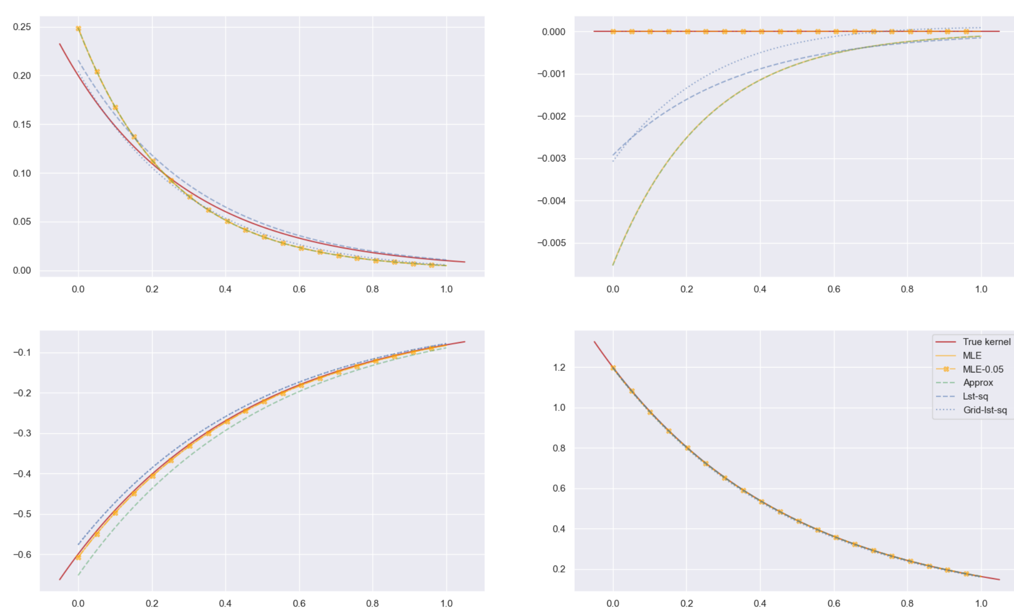



Figure 11: Reconstruction of interaction functions $h_{ij}$ for Scenario (2) of two-dimensional Hawkes processes along with all estimated functions $\tilde{h}_{ij}$. The real function is plotted in red and 25 estimations are averaged for each method.