Hawkes Processes with Variable Length Memory: Existence, Inference and Application to Neuronal Activity

Sacha Quayle*, Anna Bonnet*, Maxime Sangnier*

November 12, 2026

Abstract

Multivariate Hawkes processes are past-dependant point processes originally introduced to model excitation effects, later extended to a nonlinear framework to account for the opposite effect, known as inhibition. Motivated by applications in neuroscience, where the memory of a neuron may reset upon firing, we introduce a new class of nonlinear Hawkes processes with variable length memory. Our model generalises classical Hawkes processes, with or without inhibition, describing the situation where the probability of an event occurring within a given subprocess may depend differently on the history before and after its last event. In particular, if the subprocess does not depend on the history before its last event, it is said to have a variable length memory. Our main contributions are to prove existence of such processes, and to derive a workable likelihood maximisation method, capable of identifying both classical and variable memory dynamics. We demonstrate the effectiveness of our approach both on synthetic data, and on a neuronal activity dataset.

1 Introduction

In neuroscience, the activity of each neuron is characterised by its membrane potential. When this potential reaches a certain threshold, the neuron fires and emits a spike, which is extremely brief in time. It is therefore natural to model a spike as a random event and to represent the entire neuronal activity as a point process consisting of the sequence of spike times (Galves et al., 2024), often using a Hawkes process framework (Lambert et al., 2018). In Classical Hawkes process models, the spiking intensity, which caracterises the probability of emitting a spike at a given time, depends on a function of the history, the length of which is deterministically defined to be either ∞ or the size of the function support. However, after emitting a spike, a neuron's membrane potential often resets to an equilibrium level, causing the spiking intensity to depend on a variable length memory, specifically, on the activity since its last spike. Classical Hawkes processes do not capture this feature, which motivates the class of models studied in this paper.

Originally, Hawkes processes were univariate self-exciting processes (Hawkes, 1971), but they have since been generalised to a nonlinear multivariate framework (Brémaud & Massoulié, 1996) to take into account inhibition, where the appearance of events can decrease the probability of occurrence of subsequent events. This flexibility allows Hawkes processes to model a large variety of phenomena, from the time occurrences of earthquakes (Ogata, 1988) to social media (Rizoiu et al., 2017), criminology (Olinde & Short, 2020), ecology (Nicvert et al., 2024), or physiological data (Loison et al., 2025), to cite a few. Hawkes processes are particularly relevant in applications such as neuroscience (Lambert et al., 2018), where both excitatory and inhibitory effects co-exist. However, traditional Hawkes process models assume a deterministic memory length: the probability of an event occurring at a given time depends either on the full past history or, when kernel functions have bounded support, on a fixed finite window of past events (Sulem et al., 2024). To better capture the memory reset behaviour observed in neurons, a variant of Hawkes processes with variable length

*Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, Paris, France

memory was proposed. The variable memory structure, in which a neuron's spiking probability depends only on the activity since its last spike, was first introduced in a discrete-time setting in Galves & Löcherbach (2013). However, a formal generalisation of this idea to continuous time, closer in nature to Hawkes processes, was proposed in Galves & Löcherbach (2016); Hodara & Löcherbach (2017), in a nonlinear and interacting particle system framework on an infinite graph. However, it is possible some neurons exhibit memory reset, while others do not, which motivates the development of a model that accommodates both fixed and variable memory dynamics, taking into account a possible diversity in memory behaviors. To the best of our knowledge, only a few works address statistical inference for these models. In particular, Hodara & Löcherbach (2017) focus on graph estimation, i.e., inferring the set of presynaptic neurons for each neuron based on its spiking history, while accounting for nonlinearity and memory reset effects. Other aspects such as parameter estimation, or distinguishing neurons with variable versus fixed memory, remain largely unexplored.

Classical multivariate nonlinear Hawkes process have been widely studied since their introduction in Brémaud & Massoulié (1996) which also prove their existence and stability. Limit theorems have been established by using renewal theory when the kernel functions are of bounded support (Costa et al., 2020), and there are mean-field results on the behavior of two neuronal populations (Duval et al., 2022). Regarding statistical inference, in the frequentist setting estimation methods consist of maximum likelihood estimation for exponential kernels (Bonnet et al., 2023) and feed-forward neural networks (Joseph & Jain, 2024a,b), a least-squares approach (Bacry et al., 2020), nonparametric approaches based on Bernstein-type polynomials (Lemonnier & Vayatis, 2014) and reproducing kernel Hilbert spaces (Bonnet & Sangnier, 2025). In the neural network community, inference of nonlinear Hawkes processes has also been addressed thanks to specific architectures, such as long short-term memory recurrent neural networks (Mei & Eisner, 2017), self-attention (Zhang et al., 2020) and transformers (Zuo et al., 2020; Meng et al., 2024).

In the Bayesian framework, Sulem et al. (2024) proposed a nonparametric estimation procedure for kernel functions with bounded support. The authors then developed a variational inference procedure (Sulem et al., 2023) in order to reduce the computational cost of their method. Neural networks have also been used based on Gaussian process priors (Pan et al., 2021). Finally, Deutsch & Ross (2025) investigated a parametric inference method based on a new reparametrisation of the process. Despite these advancements, asymptotic results for the estimators in the inhibitory case are still lacking, highlighting an area for future research.

In this paper, we introduce a generalised model that includes both the classic Hawkes process and the variable length memory Hawkes process, in a finite multidimensional setting. We study existence conditions to ensure the non-explosion of these processes, generalising spectral radius conditions outlined in Brémaud & Massoulié (1996). For the variable length memory model, in which all neurons exhibit memory reset, we establish existence under weaker conditions, when the kernel functions are simply bounded. We then present a maximum likelihood estimation method for the generalised model using exponential kernel functions, inspired by the methodology presented in Bonnet et al. (2023), focusing on intervals where the intensity is positive to provide an explicit expression for the log-likelihood, which can be computed recursively. This enables inference on both the model parameters and the type of memory dynamics involved, implemented in a five-step procedure that includes identifying significant interactions within the entire process, and performing tests to detect classic or variable memory dynamics. The Markovian properties for the intensities are also useful for computation of the compensator, which enables us to assess the goodness-of-fit of our model using the time change theorem (Daley & Vere-Jones, 2008, Proposition 7.4.IV). Our numerical procedure is implemented in Python and is freely available on GitHub[1]. We conduct a numerical study on simulated data, consisting of 2 then 10 subprocesses, demonstrating that our method provides accurate estimations and can correctly identify interaction types. We also apply our method to neuronal activity data to demonstrate the practical relevance of our framework.

Section 2 presents the generalised model, outlines conditions for non-explosion, and derives Markovian properties for the intensities. Section 3 details the estimation procedure, including maximum likelihood estimation, interaction tests, goodness-of-fit, and the resampling procedure.

[1] https://github.com/sachaquayle/hawkesGVM

Section 4 illustrates the procedure on synthetic datasets, and Section 5 applies it to neuronal data. Finally, Section 6 discusses the limitations of this paper and future perspectives.

2 Hawkes Processes with Variable Length Memory

2.1 Definitions

In what follows, we denote $\mathbb{N}^* = \mathbb{N}\backslash\{0\}$ the set of integers without 0, and we consider a multivariate point process $N = (N^1, \dots, N^d)$ of dimension $d \geq 1$, where each component N^i, for $i \in \{1, \dots, d\}$, is a point process on $\mathbb{R}_+$, defined on the Borel σ-algebra $\mathcal{B}(\mathbb{R}_+)$. Each subprocess N^i is characterised by its sequence of event times $(T_k^i)_{k\in\mathbb{N}^*}$, and its law is uniquely determined by its conditional intensity function λ^i, defined by:

$$\lambda^i(t) = \lim_{h\to 0} \frac{\mathrm{E}\left[N^i[t, t+h)|\mathcal{F}_t\right]}{h},$$

where $(\mathcal{F}_t)_{t\geq 0}$ is the canonical filtration of N. For any $t \geq 0$, we also denote $N^i(t) = \sum_{k\in\mathbb{N}^*} 1_{\{T_k^i \leq t\}}$. The process N can be seen as a point process on $\mathbb{R}_+$, obtained by superposition of $N^1, \dots, N^d$, with associated event times $(T_{(l)})_{l\in\mathbb{N}^*}$, and $(d_l)_{l\in\mathbb{N}^*}$ the sequence of indexes indicating the dimension to which each event time belongs.

We consider multivariate Hawkes processes that allow for inhibition, meaning that the occurrence of events can decrease the probability of appearance of future events. The conditional intensities are thus defined as:

$$\lambda^i(t) = (\lambda^{i*}(t))^+,$$

where $x^+ = \max(0, x)$ and $\lambda^{i*}(t)$ is the underlying (possibly negative) intensity of subprocess N^i.

A generalised Hawkes process with variable length memory (abbreviated to GVM) is a specific class of multivariate point processes $N = (N^1, \dots, N^d)$ on $\mathbb{R}_+$, where each component N^i has an underlying intensity λ^{i*} of the form:

$$\begin{aligned}\lambda^{i*}(t) &= \mu_i + \sum_{j=1}^{d} \left[\int_{[T^i_{N^i(t)}, t[} h_{ij}(t-s)\,\mathrm{d}N^j(s) + \int_{[0, T^i_{N^i(t)}[} \widetilde{h}_{ij}(t-s)\,\mathrm{d}N^j(s) \right] \\ &= \mu_i + \sum_{j=1}^{d} \left[\sum_{k\in\mathbb{N}^*,\, T^i_{N^i(t)} \leq T^j_k < t} h_{ij}(t - T^j_k) + \sum_{k\in\mathbb{N}^*,\, T^j_k < T^i_{N^i(t)}} \widetilde{h}_{ij}(t - T^j_k) \right]. \end{aligned} \tag{1}$$

where $\mu_i > 0$ is the baseline intensity, which quantifies the rate of event times in the absence of cross-interactions, and $h_{ij}, \widetilde{h}_{ij} : \mathbb{R}_+ \to \mathbb{R}$ are measurable interaction functions describing the influence of N^j on N^i.

In the case where $h_{ij} = \widetilde{h}_{ij}$ for any $i, j \in \{1, \dots, d\}$, Model (GVM) resorts to the classical nonlinear Hawkes process model, referred to from now on as Model (HP), in which the entire past contributes to the intensity.

When $\widetilde{h}_{ij} = 0$ for any $i, j \in \{1, \dots, d\}$, Model (GVM) resorts to a Hawkes process model with variable length memory, similar in spirit to the models introduced in Galves & Löcherbach (2013, 2016); Hodara & Löcherbach (2017), our formulation being in continuous time, involving a finite number of neurons, and defining the conditional intensity through a slightly different form. In what follows, we refer to this model as Model (VM). In this case, the underlying intensity λ^{i*} of N^i is:

$$\lambda^{i*}(t) = \mu_i + \sum_{j=1}^{d} \sum_{k\in\mathbb{N}^*,\, T^i_{N^i(t)} \leq T^j_k < t} h_{ij}(t - T^j_k).$$

In contrast to Model (HP), Model (VM) restricts the memory of subprocess N^i to events occurring after $T^i_{N^i(t)}$, its most recent event, which allows us to model memory reset dynamics, where the memory of each subprocess resets once an event from that subprocess occurs.

Model (GVM) allows for statistical inference without committing to a specific choice between Models (HP) and (VM), and also allows to account for other memory dynamics than the two previous models, taking into account the possibility of partial memory reset, where some neurons may exhibit memory reset while others do not.

2.2 Existence results

In this section, we investigate conditions ensuring the existence of the process, i.e., preventing explosion. Given a multivariate point process $N = (N^1, \dots, N^d)$, we say that N is non-explosive, or finite, if, a.s.: $\forall t > 0$, $N[0,t] < \infty$. Equivalently, the process is non-explosive if (Asmussen, 2003, p.40): $\Pr\left(\sup_{l\in\mathbb{N}^*} T_{(l)} < \infty\right) = 0$. In the context of multivariate Hawkes processes under Model (HP), (Brémaud & Massoulié, 1996, Theorem 7) showed that non-explosion is guaranteed under the condition:

$$\rho\left(\int_0^\infty |h_{ij}|(t)\,\mathrm{d}t\right)_{1\le i,j\le d} < 1\,,$$

where $\rho(A)$ denotes the spectral radius of the matrix A. The authors further established uniqueness, stationarity and stability of the process, which are detailed in their work. A similar existence result can be obtained for Model (GVM):

Proposition 1 (Existence for Model (GVM))**.** *Let $h_{ij}, \widetilde{h}_{ij} : \mathbb{R}_+ \to \mathbb{R}$ be measurable functions for $i, j \in \{1, \dots, d\}$. Assume that:*

$$\rho\left(\int_0^\infty \max(h_{ij}, \widetilde{h}_{ij})^+(t)\,\mathrm{d}t\right)_{1\le i,j\le d} < 1\,.$$

Then there exists a non-explosive multivariate point process N under Model (GVM).

The result follows by dominating the kernels h_{ij} and $\widetilde{h}_{ij}$ pointwise by $\max(h_{ij}, \widetilde{h}_{ij})^+$, and applying a thinning argument based on (Ogata, 1981, Proposition 1). The detailed proof is provided in Appendix A.

By taking $\widetilde{h}_{ij} = 0$, Proposition 1 also gives a spectral radius condition for existence under Model (VM). However, a key feature of Model (VM) is that the memory of each subprocess resets at every event time of that subprocess, which naturally prevents explosion. As a consequence, non-explosion can be ensured under the sole assumption that the interaction functions are bounded.

Theorem 1 (Existence for Model (VM))**.** *Let $h_{ij} : \mathbb{R}_+ \to \mathbb{R}$ be measurable functions for $i, j \in \{1, \dots, d\}$. Assume that the functions h_{ij} are bounded. Then there exists a non-explosive multivariate point process N under Model (VM).*

The proof relies on dominating the conditional intensity process $\lambda = (\lambda^1, \dots, \lambda^d)$ by a time-continuous Markov chain whose non-explosive behaviour follows from standard Markov chain theory, and again by applying a thinning argument based on (Ogata, 1981, Proposition 1). The detailed proof is provided in Appendix B.

Let us note that this result exclusively concerns Model (VM), in which the functions $\widetilde{h}_{ij}$ are null. In this setting, the assumption required for non-explosion is simply the boundedness of the functions h_{ij}. This condition is mild and is satisfied by most commonly used interaction kernels in practice, such as exponential or compactly supported kernels. For Model (GVM), we expect that a similar but weaker result could hold, potentially with conditions primarily involving the functions $\widetilde{h}_{ij}$, but a formal justification is still under development.

2.3 Markovian property

The goal of this section is to establish Markovian properties for the conditional intensities under Model (GVM), which will be used later for simulation algorithms and likelihood computation. In the following, we suppose that the interaction functions are exponential: for any $i, j \in \{1, \dots, d\}$,

$$h_{ij}(t) = \alpha_{ij} e^{-\beta_i t} \text{ and } \widetilde{h}_{ij}(t) = \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i t},$$

with $\alpha_{ij}, \widetilde{\alpha}_{ij} \in \mathbb{R}$ and $\beta_i, \widetilde{\beta}_i > 0$.

Several variants of the multivariate exponential Hawkes model exist. One common approach is to assume that all interaction functions share a common decay rate $\beta > 0$ (Bacry et al., 2020). In this work, we adopt a less restrictive assumption, where the decay rates in each interaction function h_{ij} and $\widetilde{h}_{ij}$ depend only on the receiving process N^i. This assumption, also used in the self-exciting case of a 2-dimensional Hawkes process by Ogata (1981), allows a balance between model flexibility and computational efficiency, particularly for evaluating the conditional intensities. Biologically, this assumption is also reasonable, as it implies that the rate at which a given process "forgets" an event depends only on the receiving process and not on the source of the event, or whether the triggering event occurred in the recent or distant past.

In this case, under Model (GVM), the underlying intensities are, for any $i \in \{1, \dots, d\}$:

$$\lambda^{i*}(t) = \mu_i + \sum_{j=1}^{d} \left[\sum_{k \in \mathbb{N}^*,\, T^i_{N^i(t)} \leq T^j_k < t} \alpha_{ij} e^{-\beta_i (t - T^j_k)} + \sum_{k \in \mathbb{N}^*,\, T^j_k < T^i_{N^i(t)}} \widetilde{\alpha}_{ij} e^{\widetilde{\beta}_i (t - T^j_k)} \right]. \tag{2}$$

We introduce, for any $i \in \{1, ..., d\}$:

$$\begin{cases} \eta^i(t) = \sum\limits_{j=1}^{d} \sum\limits_{k \in \mathbb{N}^*,\, T^i_{N^i(t)} \leq T^j_k < t} \alpha_{ij} e^{-\beta_i (t - T^j_k)} \\ \widetilde{\eta}^i(t) = \sum\limits_{j=1}^{d} \sum\limits_{k \in \mathbb{N}^*,\, T^j_k < T^i_{N^i(t)}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i (t - T^j_k)} \\ \widetilde{\eta}^i_{\text{aux}}(t) = \sum\limits_{j=1}^{d} \sum\limits_{k \in \mathbb{N}^*,\, T^i_{N^i(t)} \leq T^j_k < t} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i (t - T^j_k)} \end{cases} \tag{3}$$

where $\eta^i, \widetilde{\eta}^i$ are such that:

$$\lambda^{i*}(t) = \mu_i + \eta^i(t) + \widetilde{\eta}^i(t)\,,$$

and $\widetilde{\eta}^i_{\text{aux}}$ is introduced for computation purposes.

Proposition 2. *Let $N = (N^1, ..., N^d)$ be a non-linear multivariate Hawkes process under Model (GVM) with underlying intensities defined by Equation (2). Let $\eta^i, \widetilde{\eta}^i$ and $\widetilde{\eta}^i_{aux}$ be defined by (3) for any $i \in \{1, \dots, d\}$. We suppose that two different processes cannot jump at the same time, i.e., for any $i \neq j \in \{1, \dots, d\}$ and $k_1, k_2 \in \mathbb{N}^*$:*

$$T^i_{k_1} \neq T^j_{k_2} \ \text{ a.s.}$$

Then, for any $i \in \{1, \dots, d\}$:

$$\forall t \in (T_{(l)}, T_{(l+1)}), \qquad \lambda^{i*}(t) = \mu_i + \left[\eta^i(T_{(l)}) + \alpha_{i,d_l}\right] e^{-\beta_i (t - T_{(l)})} + \widetilde{\eta}^i(T_{(l)}) e^{-\widetilde{\beta}_i (t - T_{(l)})}\,,$$

and

$$\lambda^{i*}(T_{(l+1)}) = \begin{cases} \mu_i + \left[\widetilde{\eta}^i(T_{(l)}) + \widetilde{\eta}^i_{aux}(T_{(l)}) + \widetilde{\alpha}_{i,d_l}\right] e^{-\widetilde{\beta}_i (T_{(l+1)} - T_{(l)})} & \text{if } i = d_{l+1} \\ \lambda^{i*}(T^-_{(l+1)}) & \text{if } i \neq d_{l+1}\,. \end{cases}$$

Proof of Proposition 2 can be found in Appendix C.

Let us remark that if $\widetilde{\alpha}_{ij} = 0$ for any $i, j \in \{1, \dots, d\}$, which corresponds to Model (VM), then, for any $t \in (T_{(l)}, T_{(l+1)})$:

$$\lambda^{i*}(t) = \mu_i + (\lambda^{i*}(T_{(l)}) + \alpha_{i,d_l} - \mu_i) e^{-\beta_i (t - T_{(l)})}\,,$$

thus, strictly between two jump times, the recursive formulas remain the same as in Model (HP). The only difference lies in the values $\lambda^{i*}(T_{(l+1)})$ of the intensities precisely at the event times, which reset to μ_i if $i = d_{l+1}$, reflecting the memory reset.

3 Estimation

3.1 Maximum likelihood estimation

As in the previous section, we suppose the kernel functions are exponential:

$$h_{ij}(t) = \alpha_{ij} e^{-\beta_i t} \text{ and } \widetilde{h}_{ij}(t) = \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i t} \, .$$

Our objective is to estimate the parameters $(\mu_i, \alpha_{ij}, \beta_i, \widetilde{\alpha}_{ij}, \widetilde{\beta}_i)_{1 \leq i,j \leq d}$ using Maximum Likelihood Estimation. To do so efficiently, it is crucial to compute the log-likelihood recursively. For any multivariate point process $N = (N^1, \ldots, N^d)$, the log-likelihood of the subprocess N^i up to time $t \geq 0$ is given by (Daley & Vere-Jones, 2003, Proposition 7.3.III.):

$$\ell_{\lambda^i}(t) = \sum_{k \geq 0, \, T_k^i < t} \ln \lambda^i(T_k^{i-}) - \Lambda^i(t) \, ,$$

where $\Lambda^i(t) = \int_0^t \lambda^i(s) \, \mathrm{d}s$ is the compensator of subprocess N^i. Additionally, the total log-likelihood of the process N is: $\ell_\lambda(t) = \sum_{i=1}^d \ell_{\lambda^i}(t)$.

In the case of Hawkes processes under Model (GVM), the terms $\ln \lambda_\theta^i(T_k^{i-})$ can be computed recursively using Proposition 2. The main difficulty therefore lies in the computation of the compensator, which requires identifying the intervals over which the underlying intensities remain positive. As detailed in Bonnet et al. (2023), this computation relies on the monotony of the underlying intensities between consecutive event times, which can be ensured by further assuming, for any $i \in \{1, \ldots, d\}$, that:

$$\widetilde{\beta}_i = \beta_i \, .$$

This condition implies that the decay rates governing the influence of both recent and distant past events are identical, which is a plausible assumption, and ensures that, between two consecutive event times of N, the underlying intensities λ^{i*} are increasing or decreasing exponential functions.

In this case, the parameter vector to be estimated is:

$$\theta = (\mu_i, \alpha_{ij}, \beta_i, \widetilde{\alpha}_{ij})_{1 \leq i,j \leq d} = (\mu, \alpha, \beta, \widetilde{\alpha}) \in (\mathbb{R}_+^*)^d \times \mathbb{R}^{d^2} \times (\mathbb{R}_+^*)^d \times \mathbb{R}^{d^2} \, .$$

In the following, we denote by λ_θ^{i*} and Λ_θ^i respectively the underlying intensity and the compensator of subprocess N^i associated with parameter θ. We also denote by λ_θ, Λ_θ and ℓ_θ respectively the intensity, the compensator, and the log-likelihood of N.

Proposition 3. *Let $N = (N^1, ..., N^d)$ be a non-linear multivariate Hawkes process under Model (GVM) with underlying intensities defined by Equation (2). For any $l \in \mathbb{N}^*$ and for any $i \in \{1, ..., d\}$, define:*

$$T_{(l),\theta}^{i*} = \min \left(T_{(l+1)}, \, \inf \left\{ t \geq T_{(l)}, \, \lambda_\theta^{i*}(t) \geq 0 \right\} \right) .$$

Then, for any $k \in \mathbb{N}^*$ and for any $i \in \{1, ..., d\}$:

$$T_{(l),\theta}^{i*} = \min \left(T_{(l+1)}, \, t_{l,\theta}^{i*} \right) ,$$

where:

$$t_{l,\theta}^{i*} = \left(T_{(l)} + \beta_i^{-1} \ln \left(\frac{\mu_i - \lambda_\theta^{i*}(T_{(l)}) - \alpha_{i,d_l}}{\mu_i} \right) 1_{\lambda_\theta^{i*}(T_{(l)}^+) < 0} \right) ,$$

and the compensator of the subprocess N^i is:

$$\Lambda_\theta^i(t) = \begin{cases} \mu_i t & \text{if } t < T_{(1)} \\ \mu_i T_{(1)} + \sum\limits_{l=1}^{N(t)} J_{l,\theta}^i(t) & \text{if } t \geq T_{(1)} \end{cases} ,$$

where, for any $l \in \{1, \ldots, N(t)\}$:

$$J_{l,\theta}^i(t) = \mu_i \left[\min(T_{(l+1)}, t) - T_{(l)}^{i*} \right] + \frac{\lambda_\theta^{i*}(T_{(l)}) - \mu_i + \alpha_{i,d_l}}{\beta_i} \left[e^{-\beta_i (T_{(l)}^{i*} - T_{(l)})} - e^{-\beta_i (\min(T_{(l+1)}, t) - T_{(l)})} \right] .$$

This result, obtained by using Proposition 2, therefore enables a recursive computation of the restart times and makes it possible to evaluate the compensator with the same computational complexity as in the purely self-exciting case.

3.2 Post hoc estimation of interactions

In the following estimation procedure, we suppose that we observe n independent realisations $N_1, \dots, N_n$ of a Hawkes process on $[0, T]$. Our goal is to identify significant interactions between subprocesses and to test whether these interactions follow classic or variable memory dynamics. To this end, we test various hypotheses on the coefficients of the interaction matrices α and $\widetilde{\alpha}$, while accounting for the multiple testing setting through the Benjamini-Hochberg (BH) procedure (Benjamini & Hochberg, 1995) to control the false discovery rate. We prefer to control the false discovery rate (FDR) instead of the probability of making at least one false discovery (FWER) for two reasons. First, our method is intended to be applied to situations were the number of hypotheses to test is relatively large, in which FWER is known to be too conservative. Second, the FDR controlling procedure is used for interaction detection before re-estimation, so we can allow without damage a few false discoveries (in which case, re-estimation would lead to low intensity interactions). We are exactly in the motivating situation of FDR, where we want to discover promising avenues of investigation, i.e, estimation, while controlling the proportion of wrong leads.

In each testing setting, the approach consists of three steps: construct confidence intervals for the parameters of interest, compute the corresponding p-values for the relevant null hypotheses, then apply the BH procedure to select significant effects.

Two types of confidence intervals can be considered. The first type relies on the asymptotic normality of the estimators, using the Student distribution or the Hotelling distribution. This assumption is true in the case of purely exciting interactions but is not guaranteed theoretically under Model (GVM). For Model (GVM), asymptotic normality is only conjectured and can be checked empirically (see Section S1 of the Supplementary Material). This is discussed in the following section. The second type of confidence intervals that can be considered is empirical confidence intervals, constructed directly from the empirical distribution of the estimators across all realisations. More precisely, for a significance level $\eta \in (0, 1)$, we use the $(\lfloor \frac{\eta}{2} n \rfloor)$ and $(\lceil (1-\frac{\eta}{2}) n \rceil)$-th smallest values. For example, a confidence interval for α_{ij} is

$$\left[\alpha_{ij}^{(\lfloor \frac{\eta}{2} n \rfloor)}, \alpha_{ij}^{(\lceil (1-\frac{\eta}{2}) n \rceil)} \right],$$

where $(\alpha_{ij}^{(k)})_{1 \leq k \leq n}$ is the ordered sequence of estimations of α_{ij}, with the convention $\alpha_{ij}^{(0)} = -\infty$.

Our experience is that existence of interactions is often well detected even in misspecified models (such as a simple nonlinear exponential model). Armed with this a priori, we propose to test first existence of interactions, then its nature (with or without memory reset). The full testing procedure consists of:

Estimation procedure under Model (GVM):

1. **Step 1** (Initial estimation): Compute maximum likelihood estimators $\hat{\theta}_1, \dots, \hat{\theta}_n$ under Model (GVM) for each realisation.
2. **Step 2** (Estimation of interactions): Estimate interacting subprocesses by testing, for any $i, j \in \{1, \dots, d\}$, the null hypotheses $(\alpha_{ij}, \widetilde{\alpha}_{ij}) = (0, 0)$ (see Test 1) with levels corrected by the BH procedure.
3. **Step 3** (Re-estimation): Set insignificant coefficients of the matrix α to zero and re-compute $\hat{\theta}_1, \dots, \hat{\theta}_n$.
4. **Step 4** (Nature of interactions): Estimate the nature of interactions by testing, for any $i, j \in \{1, \dots, d\}$, the null hypotheses $\widetilde{\alpha}_{ij} = 0$ (see Test 2) and $\alpha_{ij} = \widetilde{\alpha}_{ij}$ (see Test 3) with levels corrected by the BH procedure.
5. **Step 5** (Final estimation): Compute final estimators $\hat{\theta}_1, \dots, \hat{\theta}_n$ for each realisation, or final estimator $\hat{\theta}$ over all realisations, considering the test results from Step 4.

In Step 2, Test 1 is performed for the joint hypotheses $(\alpha_{ij}, \widetilde{\alpha}_{ij}) = (0, 0)$ in order to directly detect the absence of interaction between subprocesses. This approach also reduces the number of multiple tests, as it requires d^2 tests instead of $2d^2$ if we were to test each parameter separately.

In Step 5, the final estimation can be obtained using two different methods. The first approach is by averaging the estimators $(\hat{\theta}_1, \dots, \hat{\theta}_n)$ obtained for each realisation. The second consists in maximising the sum of the log-likelihoods over all realisations, $\sum_{k=1}^{n} \ell_\theta^k$. The advantages of both methods are discussed in the following section.

Test 1: Null hypothesis $(\alpha_{ij}, \widetilde{\alpha}_{ij}) = (0, 0)$.

1. For estimations $\widehat{\alpha}_{ij}^k$ and $\widehat{\widetilde{\alpha}}_{ij}^k$ (of α_{ij} and $\widetilde{\alpha}_{ij}$) computed on the k^{th} repetition, define $\gamma_{ij}^k = (\widehat{\alpha}_{ij}^k \widehat{\widetilde{\alpha}}_{ij}^k)$.
2. Set:
   - $\overline{\gamma_{ij}} := \frac{1}{n} \sum_{k=1}^{n} \gamma_{ij}^k$;
   - $S_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} \left(\gamma_{ij}^k - \overline{\gamma_{ij}}\right) \left(\gamma_{ij}^k - \overline{\gamma_{i,j}}\right)^\top$.
3. Defining the Hotelling t-squared statistic $t_{ij}^2 = n \overline{\gamma_{ij}}^\top S_{ij}^{-1} \overline{\gamma_{ij}}$, compute:
   - the p-value for an asymptotic confidence interval: $\eta_0 = 1 - F(t_{ij}^2)$, where F is the cumulative distribution function of the Hotelling T-squared distribution with parameters 2 and $n-1$;
   - the p-values for an empirical confidence interval: $\eta_0 = 2\min(k_0^+, k_0^-)/n$, where k_0^+ (resp. k_0^-) is the number of strictly positive (resp. negative) estimations α_{ij}^k, $k \in \{1, \dots, n\}$, and $\widetilde{\eta}_0 = 2\min(\widetilde{k}_0^+, \widetilde{k}_0^-)/n$, where $\widetilde{k}_0^+$ (resp. $\widetilde{k}_0^-$) is the number of strictly positive (resp. negative) estimations $\widetilde{\alpha}_{ij}^k$, $k \in \{1, \dots, n\}$.

Test 2: Null hypothesis $\widetilde{\alpha}_{ij} = 0$.

1. For any $i, j \in \{1, \dots, d\}$ where an interaction was detected in Test 1, set:
   - $\overline{\widetilde{\alpha}}_{ij} = \frac{1}{n} \sum_{k=1}^{n} \widehat{\widetilde{\alpha}}_{ij}^k$;
   - $\widetilde{S}_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} \left(\widehat{\widetilde{\alpha}}_{ij}^k - \overline{\widetilde{\alpha}}_{ij}\right)^2$.
2. Defining the statistic $t_{ij} = \overline{\widetilde{\alpha}}_{ij} \sqrt{n/\widetilde{S}_{ij}}$, compute the p-value:
   - for an asymptotic confidence interval: $\eta_0 = 2(1 - F(t_{ij}))$, where F is the cumulative distribution function of the Student distribution with parameter $n-1$;
   - for an empirical confidence interval: $\eta_0 = 2\min(k_0^+, k_0^-)/n$, where k_0^+ (resp. k_0^-) is the number of strictly positive (resp. negative) estimations α_{ij}^k, $k \in \{1, \dots, n\}$.

Test 3: Null hypothesis $\alpha_{ij} = \widetilde{\alpha}_{ij}$.

1. For any $i, j \in \{1, \dots, d\}$ where an interaction was detected in Test 1, set:
   - $\overline{\alpha}_{ij} = \frac{1}{n} \sum_{k=1}^{n} \widehat{\alpha}_{ij}^{k}$ and $\overline{\widetilde{\alpha}}_{ij}^{(n)} = \frac{1}{n} \sum_{k=1}^{n} \widehat{\widetilde{\alpha}}_{ij}^{k}$;
   - $S_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} \left(\widehat{\alpha}_{ij}^{k} - \widehat{\widetilde{\alpha}}_{ij}^{k} - \overline{\alpha}_{ij} + \overline{\widetilde{\alpha}}_{ij} \right)^2$.
2. Defining the statistic $t_{ij} = \left(\overline{\alpha}_{ij} - \overline{\widetilde{\alpha}}_{ij} \right) \sqrt{n / S_{ij}}$, compute the p-value:
   - for an asymptotic confidence interval: $\eta_0 = 2(1 - F(t_{ij}))$, where F is the cumulative distribution function of the Student distribution with parameter $n - 1$;
   - for an empirical confidence interval: $\eta_0 = 2 \min(k_0^+, k_0^-)/n$, where k_0^+ (resp. k_0^-) is the number of strictly positive (resp. negative) estimations $\alpha_{ij}^k - \widetilde{\alpha}_{ij}^k$, $k \in \{1, \dots, n\}$.

3.3 Goodness-of-fit

When trying to model certain phenomena through a Hawkes process model, it is important to be able to verify the goodness-of-fit of an estimated model obtained. The computation of the compensator in Proposition 3 enables the use of the Time Change Theorem for inhomogeneous Poisson processes (Daley & Vere-Jones, 2008, Proposition 7.4.IV), which states that a sequence of times $(T_{(l)})_{l \in \mathbb{N}^*}$ is a realisation of a point process with compensator Λ if and only if the sequence of transformed times $(\Lambda(T_{(l)}))_{l \in \mathbb{N}^*}$ is a realisation of a unit-rate Poisson process.

Given n realisations $N_1, \dots, N_n$ with event times denoted $\left(T_{(l),k}\right)_{l \in \mathbb{N}^*}$ for any $k \in \{1, \dots, n\}$, a first approach would therefore be to compute, for each realisation N_k, an estimator $\hat{\theta}_k$, and to test whether $(T_{(l),k})_{l \in \mathbb{N}^*}$ is a realisation of a point process with conditional intensity $\lambda_{\hat{\theta}_k}$. This can be done using an equality test between the empirical distribution $(\Lambda_{\hat{\theta}_k,k}(T_{(l+1),k}) - \Lambda_{\hat{\theta}_k,k}(T_{(l),k}))_{l \in \mathbb{N}^*}$ and the unit exponential distribution, where $\Lambda_{\hat{\theta}_k,k}$ is the compensator associated to the conditional intensity $\lambda_{\hat{\theta}_k,k}$ of N_k with parameter $\hat{\theta}_k$. As deeply discussed in Reynaud-Bouret et al. (2014), it is explained that doing the goodness-of-fit test with the same sequence of event times as the one used for building the MLE will tend to over-estimate the p-value, whilst doing the test with an independent sample of event times will tend to under-estimate the p-value. Therefore, to correct the calculated p-values, a method based on resampling is proposed by the authors (Reynaud-Bouret et al., 2014, Test 4). In this procedure, the event times of each realisation are transformed using the compensator, a random subsample of the realisations is drawn then concatenated using the appropriate shifts, and the test is performed between the resulting times and the unit-rate exponential distribution. The full procedure is decribed below.

Resampling procedure for goodness-of-fit:

1. Observe n realisations $N_1, \dots, N_n$ of multivariate Hawkes processes on $[0, T]$. The event times for each realisation are denoted $\left(T_{(l),k}\right)_{1 \leq l \leq n_k}$ for any $k \in \{1, \dots, n\}$, where $n_k := N_k[0, T]$.

2. Estimate the final parameter $\hat{\theta}$ using the n realisations.

3. For any $1 \leq k \leq n$, apply the time-transformation to obtain the points $\widehat{\mathcal{N}}_k = \left(\Lambda_{\hat{\theta},k}(T_{(l),k})\right)_{1 \leq l \leq n_k}$ on $[0, \Lambda_{\hat{\theta},k}(T)]$.

4. Randomly select a subsample $S = \{i_1, \dots, i_{p_n}\}$ from $\{1, \dots, n\}$ with cardinality p_n such that $p_n/n \underset{n \to \infty}{\longrightarrow} 0$ (for example, $p_n = \lfloor \sqrt{n} \rfloor$ or $p_n = \lfloor n^{2/3}) \rfloor$.

5. Concatenate the p_n processes $\widehat{\mathcal{N}}_k$ for $k \in S$ to form $\widehat{\mathcal{N}}^{c,p_n}$.

6. Denote $M = p_n^{-1} \sum_{k \in S} \Lambda_{\hat{\theta},k}(T)$, and fix $0 < \kappa < M$ (for example, $\kappa = 0.99M$).

7. Use a goodness-of-fit test to check whether the times $\{t \in \widehat{\mathcal{N}}^{c,p_n},\, t \leq p_n \kappa\}$ are drawn from a homogeneous Poisson process with intensity 1.

4 Numerical results on synthetic datasets

4.1 Simulation procedure

We consider non-linear Hawkes processes under Model (GVM) with conditional intensities defined by Equation (1), with $h_{ij}(t) = \alpha_{ij} e^{-\beta_i t}$ and $\widetilde{h}_{ij}(t) = \widetilde{\alpha}_{ij} e^{-\beta_i t}$, such that the parameters $\alpha := (\alpha_{ij})_{1 \leq i,j \leq d}$, $\beta := (\beta_i)_{1 \leq i \leq d}$, and $\widetilde{\alpha} := (\widetilde{\alpha}_{ij})_{1 \leq i,j \leq d}$ satisfy the existence conditions presented in Section 2.2. We denote $\theta = (\mu, \alpha, \beta, \widetilde{\alpha})$.

The subsequent sections analyse parameter estimation under each model using simulated data, with an emphasis on estimation under Model (GVM). For each scenario considered, 25 independent realisations are generated via the thinning method, ensuring each dataset contains 5000 event times. Parameter estimation under Model (GVM) follows the 5-Step procedure described in Section 3.2, with parameters estimated by minimising the negative log-likelihood computed via Proposition 2 and Proposition 3, using the L-BFGS method from the scipy.optimize.minimize library (Virtanen et al., 2020). Parameter estimation under Models (HP) and (VM) simply follows Steps 1 to 3, by replacing the null hypothesis in Test 1 by $\alpha_{ij} = 0$.

For all of the considered tests, we use asymptotic normality-based confidence intervals since the number of realisations in our dataset is limited. As explained in Section 3.2, it is also possible to use empirical confidence intervals, however they require a large number of realisations to be reliable. Additional comparisons between empirical and asymptotic confidence intervals, as well as a study of the empirical distributions to support the asymptotic normality conjecture, are provided in Section S1 of the Supplementary Material.

In Step 5, as discussed in Section 3.2, the final estimations can be computed in two different ways. Estimations for each realisation are done in order to compare and produce the boxplots of the relative squared errors. However, to assess goodness-of-fit, the final estimator $\hat{\theta}$ is computed by maximising the sum of the log-likelihoods, instead of averaging the estimations for each realisation. This second approach is preferred to obtain a final estimator, as it ensures that the resulting parameters define a conditional intensity that remains non-negative at the observed event times. Additional details on the comparison of the different estimators are provided in Section S2 of the Supplementary Material.

Finally, the p-values are obtained using the resampling procedure described in Section 3.3. The equality test considered here is the Cramer-von-Mises (CvM) test, using the CvM test implemented in the scipy.cramervonmises library. This test is based on the integrated squared difference between the empirical and theoretical cumulative distribution functions, and is generally more powerful than the usual Kolmogorov-Smirnov (KS) test. A detailed comparison between the KS and CvM tests is provided in Section S3 of the Supplementary Material to further support this choice.

4.2 Estimation on bivariate processes

We first consider an exponential bivariate Hawkes process with parameters

$$\mu = \begin{pmatrix} 0.7 \\ 1 \end{pmatrix}, \quad \alpha = \begin{pmatrix} 0.2 & 0 \\ -0.6 & 1.2 \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

From it, two different scenarios are distinguished: the first one, named Scenario (HP), corresponds to simulating under Model (HP), that is taking

$$\widetilde{\alpha} = \begin{pmatrix} 0.2 & 0 \\ -0.6 & 1.2 \end{pmatrix}.$$

The second one, named Scenario (VM), corresponds to simulating under Model (VM), that is taking $\widetilde{\alpha} = 0$.

For each scenario, we display the outcomes from Tests 1 to 3, and the relative squared errors for each group of parameters (μ, α, β and $\widetilde{\alpha}$) after Steps 1, 3 and 5, by considering vector norms. The results show that Model (GVM) achieves accurate parameter estimation in both scenarios. As illustrated in Fig. 1, Tests 1 to 3 successfully recover the support and the correct interaction type. Fig. 2 illustrates that the relative errors progressively decrease after each test, highlighting the effectiveness of our five-step procedure. Most of the error reduction occurs between Steps 3 and 5, primarily affecting the estimates of β and $\widetilde{\alpha}$. The error on α remains relatively stable throughout the estimation procedure, which validates our procedure consisting of testing absence of interactions first. Although the errors on $\widetilde{\alpha}$ are higher than those on α, it is guessed that, since $\widetilde{\alpha}$ represents distant past influence, its contribution is substantially smaller than recent past when the kernel is exponential, and is thus more difficult to estimate.

In Table 1, we display the average p-value from the outcomes of the goodness-of-fit test for Model (GVM), by applying the resampling procedure 50 times. These tests are performed with the final estimation obtained after Step 5, computed over all realisations. In comparison, p-values are also provided for the true parameters as well as with estimations under Models (HP) and (VM), where the final estimates are taken after Step 3. These results indicate that Model (GVM) provides consistently high p-values, similar to those obtained with the true parameters in both scenarios, demonstrating its flexibility and ability to fit the data well. Model (HP) performs well when data is from Scenario (HP) but fails when the data comes from Scenario (VM), as reflected by lower p-values. Similarly, Model (VM) fits (VM) data accurately but performs poorly for (HP) data. This highlights the limitations of each specific model and the robustness of the generalised Model (GVM).

Table 1: Average p-values obtained after resampling realisations from Scenarios (HP) and (VM) in dimension 2.

Scenario	(HP)			(VM)		
Model	(HP)	(VM)	(GVM)	(HP)	(VM)	(GVM)
True	0.46	6e-8	0.48	1e-8	0.43	0.4
Estimations	0.5	0.17	0.5	0.2	0.4	0.41

4.3 Estimation on 10-dimensional processes

We consider 10-dimensional parameters μ, α, β and $\widetilde{\alpha}$ whose heatmaps are displayed in Fig. 3 and whose exact values can be found on GitHub. The value for α has been chosen to include sparsity and both excitatory and inhibitory interactions. The chosen value for $\widetilde{\alpha}$ reflects that all subprocesses exhibit classic Hawkes behaviour, except for Subprocesses 1, 3 and 8 which have a variable length memory. We consider four scenarios listed below: corresponding to scenarios under Models (HP), (VM) and (GVM), and an additional scenario where the values of β are multiplied by a factor 10 (thus indicating the rate at which the process forgets past events).

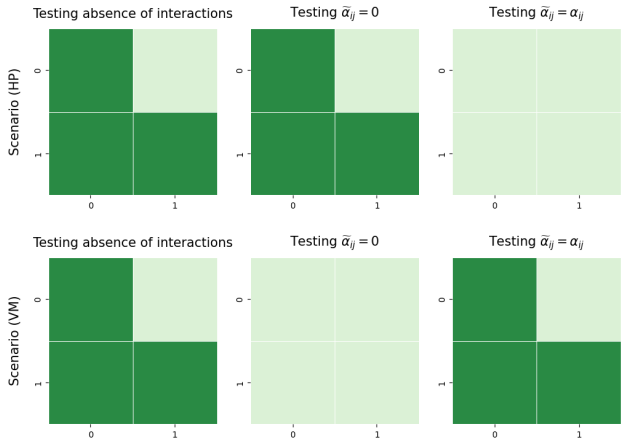


Fig. 1: Test results from estimations under (GVM) for Scenarios (HP) and (VM) in dimension 2. Colours mean true positive (detected 0 or equal value), true negative (detected non-null value or non-equal value), false positive (non-null value set to 0 or non-equal value set equal), false negative (undetected 0 or equal value).

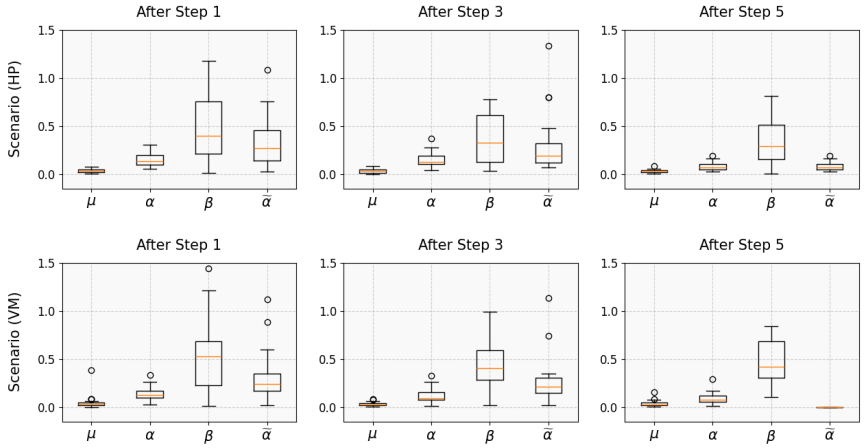


Fig. 2: Boxplots of the relative squared error for each group of parameters (μ, α, β and $\widetilde{\alpha}$) under (GVM) for Scenarios (HP) and (VM) in dimension 2, after Steps 1, 3 and 5 of the estimation procedure.

- Scenario (HP): Parameters μ, α, β are given by Fig. 3 and $\widetilde{\alpha} = \alpha$.
- Scenario (VM): Parameters μ, α, β are given by Fig. 3 and $\widetilde{\alpha} = 0$.
- Scenario (GVM): Parameters μ, α, β and $\widetilde{\alpha}$ are given by Fig. 3.
- Scenario (HP-short-mem): Parameters μ, α, β are given by Fig. 3, β is multiplied by a factor 10, and $\widetilde{\alpha} = \alpha$.

Table 2: Average p-values over 25 estimations of realisations from Scenarios (HP), (VM), (GVM) and (HP-short-mem) in dimension 10.

Scenario	(HP)			(VM)			(GVM)			(HP-short-mem)		
Model	(HP)	(VM)	(GVM)	(HP)	(VM)	(GVM)	(HP)	(VM)	(GVM)	(HP)	(VM)	(GVM)
MLE	0.53	0.39	0.59	0.36	0.55	0.47	0.44	0.49	0.54	0.12	0.24	0.18

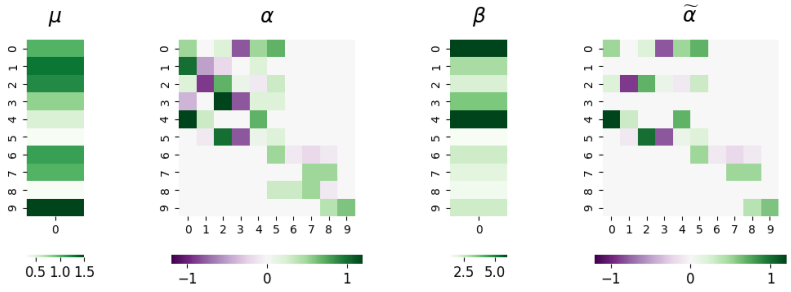


Fig. 3: Heatmaps of true parameters μ, α, β and $\widetilde{\alpha}$ in dimension 10.

For each scenario, we display the outcomes from Tests 1 to 3, performed using Student-based confidence intervals, and the relative squared errors for each group of parameters (μ, α, β and $\widetilde{\alpha}$) after Step 5, by considering vector norms (unlike the bivariate case, errors after Steps 1 and 3 are omitted). As shown in Fig. 4, Model (GVM) accurately recovers parameters and interaction structure in Scenarios (HP), (VM) and (GVM), with Tests 1 to 3 producing reliable results, and Fig. 5 confirming low relative errors after Step 5. In Scenario (HP-short-mem), however, relative errors for $\widetilde{\alpha}$ are high, and Fig. 4 shows that Test 2 tends to set a lot of values of $\widetilde{\alpha}$ to zero. This is expected, as large values of β cause the exponential kernel to decay rapidly, causing the processes to quickly lose memory of past events, even in Model (HP), leading Model (GVM) to overestimate values of α and assign zero values to $\widetilde{\alpha}$.

In Table 2, we also display the average p-value from the outcomes of the goodness-of-fit test for Model (GVM), by applying the resampling procedure 50 times. These tests are performed with the final estimation obtained after Step 5 over all realisations. In comparison, p-values are also provided for estimations under Models (HP) and (VM), in which case the final estimates are taken after Step 3. As before, Model (GVM) yields consistently high p-values across all scenarios, while Models (HP) and (VM) only perform reliably when the data match their model. In Scenario (HP-short-mem), p-values are comparable across models, indicating no clear distinction, even when the model is misspecified.

4.4 Summary of the numerical study

Overall, estimation under Model (GVM) performs consistently well across all models and scenarios, producing satisfactory test results, low relative errors after Step 5, and high p-values. In contrast, Models (HP) and (VM) yield reliable results only when the data is simulated using the same model: when applied to data generated by a different model, these two models exhibit poor performance, particularly in terms of p-values. However, these discrepancies are most pronounced for small values of β, where the differences between the three models are more distinct. For larger values of β, the rapid decay of the exponential distribution causes all three models to quickly lose memory of past events, reducing the impact of model differences.

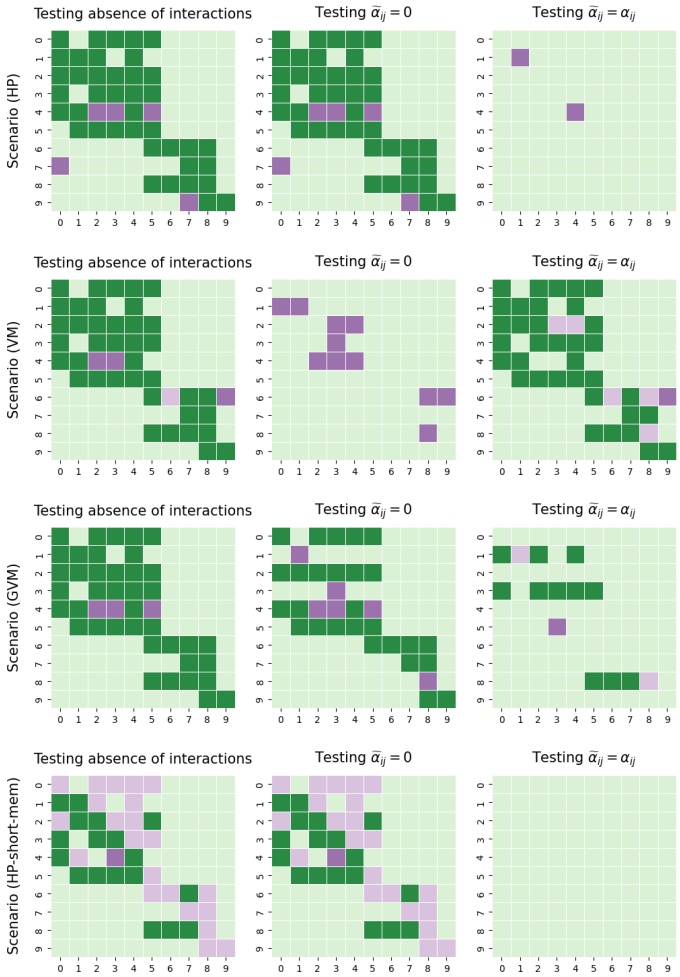


Fig. 4: Test results from estimations under (GVM) for Scenarios (HP), (VM), (GVM) and (HP-short-mem) in dimension 10. Colours mean true positive (detected 0 or equal value), true negative (detected non-null value or non-equal value), false positive (non-null value set to 0 or non-equal value set equal), false negative (undetected 0 or equal value).

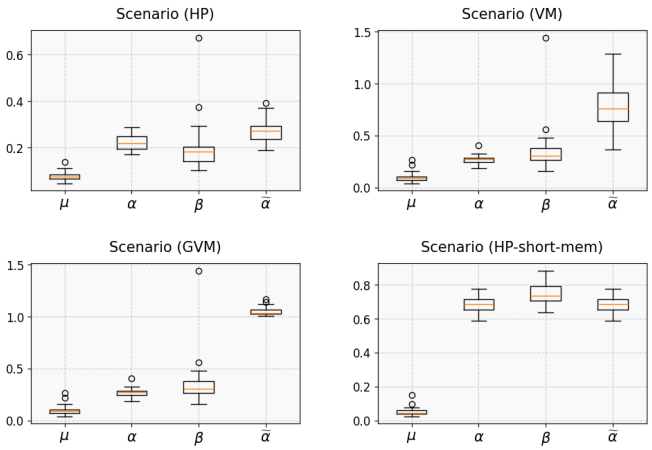


Fig. 5: Boxplots of the relative squared error for each group of parameters (μ, α, β and $\widetilde{\alpha}$) under (GVM) for 25 realisations of Scenarios (HP), (VM), (GVM) and (HP-short-mem) in dimension 10, after Step 5 of the estimation procedure.

5 Application on neuronal data

5.1 Data preprocessing

In this section, we apply our estimation procedure under Model (GVM) to a collection of 10 trials consisting in the measurement of spike trains of 250 neurons from the lumbar spinal circuits of a red-eared turtle (Petersen & Berg, 2016; Radosevic et al., 2019). Each trial events were recorded for 40 seconds. We begin with a pre-processing approach inspired by Bonnet et al. (2022), detailed in Steps 1 to 3 below, which involves trimming the trials to ensure stationarity, removing trials with too many non-active neurons, and retaining only sufficiently active neurons. After Step 3, the dataset is reduced to observations within $[0, 10]$, a total of 7 trials, and 61 neurons. Due to the limited number of trials, and in order to perform tests in our estimation procedure, we then follow a resampling approach described in Bonnet et al. (2023) and Reynaud-Bouret et al. (2014), corresponding to Steps 4 to 6, in order to generate a larger number of trials. The full procedure is described below. The choices made at each step are further motivated by the visualisations provided in Appendix D, which illustrate the normalised cumulative spike counts $N^i(t)/t$, for the raw, filtered, and resampled data, corresponding respectively to the outcomes of Steps 1, 3, and 6.

Data preprocessing steps:

1. **Step 1** (Trimming): Data is processed in order to take into account stationarity by considering only events that took place on $[11, 21]$. Consequently, each trial is considered within the normalised time window $[0, 10]$.
2. **Step 2** (Non-relevant trials removal): Only trials with fewer than 10 non-active neurons are retained.
3. **Step 3** (Non-active neurons removal): Only neurons with at least 50 jumps across all trials are kept.
4. **Step 4** (Subsampling): Three realisations, denoted (N_1, N_2, N_3), are sampled at random without replacement, preserving their order.
5. **Step 5** (Concatenation): The three realisations are concatenated by considering that N_1 takes place in $[0, 10]$, N_2 takes place in $[10, 20]$ and N_3 takes place in $[20, 30]$.
6. **Step 6** (Resampling): Steps 4 and 5 are repeated 25 times to obtain a sample of 25 realisations.

5.2 Estimation results

We follow Steps 1 to 5 of the estimation procedure under Model (GVM) as described in Section 3.2. Yet, since only 7 trials are available but Tests 1, 2 and 3 require a larger number of trials, interaction detection is performed on the 25 resampled trials obtained after Step 6 of the preprocessing described in Section 5.1, whilst the final estimation is done on the 7 original trials, obtained by maximising the sum of the log-likelihoods. The results of Tests 1 (on the left), 2, and 3 (on the right) are summarised in Fig. 6. The results of Test 1 indicate that many significant interactions persist (3264 out of 3721), which can be attributed to our preprocessing procedure that retains only sufficiently active neurons. On the diagonal values, our test detects 56 interactions out of 61, highlighting the major effect of self-interactions. The results from Tests 2 and 3 demonstrate that both classic memory dynamics and variable memory dynamics are detected. In particular, we detect 1045 interactions where $\widetilde{\alpha}_{ij} = 0$ and $\alpha_{ij} \neq \widetilde{\alpha}_{ij}$, and 630 interactions where $\alpha_{ij} = \widetilde{\alpha}_{ij}$ and $\widetilde{\alpha}_{ij} \neq 0$, supporting the relevance of our model and its ability to capture different types of dynamics. Furthermore, this finding aligns with our model's framework, where some interactions have a memory reset while others do not. Estimation results for final estimations μ, α, β and $\widetilde{\alpha}$ after Step 5 of the estimation procedure are presented in Fig. 7. The heatmap matrices of the signs of α and $\widetilde{\alpha}$ show numerous significant interactions, both excitatory and inhibitory.

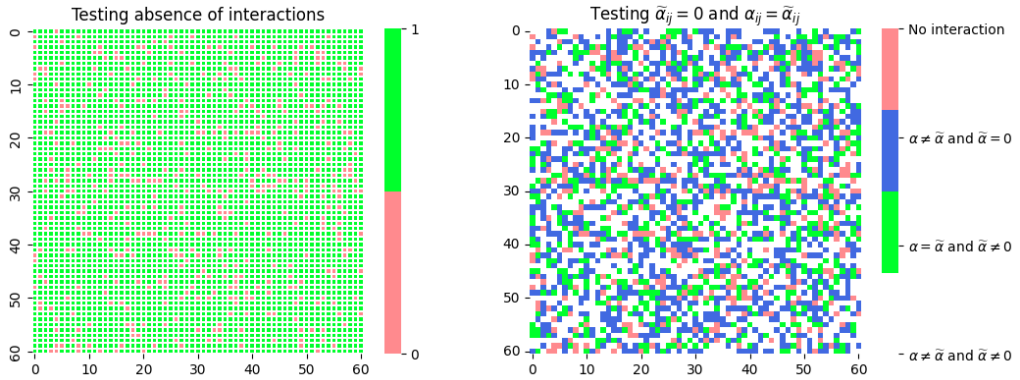


Fig. 6: Test results from estimations under (GVM) for the 25 resampled neuronal trials. On the left: red (0) indicates no interaction, while green (1) indicates a detected interaction. On the right: summary of the results of Tests 2 and 3.

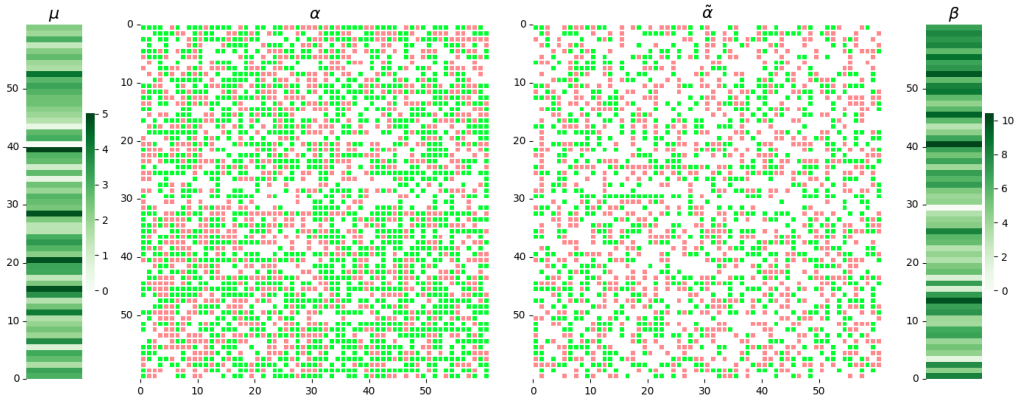


Fig. 7: Heatmaps of estimated parameters after Step 5 of the estimation procedure under (GVM) on the neuronal data. For α and $\widetilde{\alpha}$: white represents no interaction, green represents an excitatory interaction, and red represents an inhibitory interaction.

We also display in Table 3 the average p-value and the rejection rates at the 5% significance level from the outcomes of the goodness-of-fit test, by applying the resampling procedure from Section 3.3 a total of 50 times for the final estimator, under Model (GVM), Model (HP) and Model (VM). We observe that all average p-values are far from 0.5, indicating that there is room for improvement on all models. Yet, the rejection rates indicate that the null hypothesis is less frequently rejected under Model (GVM) than Models (HP) and (VM), with rejection rates of 0.56, 1, and 0.95, respectively.

Table 3: Average p-values and rejection rates at 5% for final
estimations on neuronal data over 50 resampled trials.

Model	(HP)	(VM)	(GVM)
p-value	1e-5	0.01	0.051
Rejection rate	1	0.95	0.56

6 Conclusion

In this paper, we introduced a generalised model for nonlinear Hawkes processes in a finite multidimensional setting, taking into account possible variable memory dynamics. We proved existence of processes in this framework and provided in particular a weak and nontrivial condition for the variable length memory model. The question of stationarity remains open and would be a valuable direction for future work, especially in order to derive long-term properties and to justify our statistical procedures.

On the inference side, we derived explicit formulas for the conditional intensities and developed a methodology for parameter estimation, interaction detection, and distinguishing between classic and variable memory dynamics. Our numerical results on synthetic data suggest the relevance and the flexibility of this generalised model, and its ability to adapt to both variable and fixed memory scenarios.

We also applied our inference procedure to neuronal spike data, which revealed several cases of variable memory dynamics, and less rejections under Model (GVM) than for the two pre-existing ones. However, the rejection rates remains substantial, indicating that Model (GVM) can still be refined in order to better suit neuronal data. In particular, this work focuses on two main simplifying assumptions: simple exponential kernels and decay rates depending only on the receiving process. Future research could aim at extending the model to other types of kernels and more general interaction structures. Besides, neurons exhibit a refractory period following a spike, during which they cannot fire again (Galves et al., 2024). It would therefore be relevant to account for this biological feature in the model to better distinguish between true memory resets and the effects of refractoriness, for instance by considering that the self-interaction functions are sums of two exponential functions, or are non-parametric functions, as it is done in Bonnet & Sangnier (2025) for classical Hawkes processes without variable memory.

Additionally, according to Dale's principle (Strata & Harvey, 1999; Galves et al., 2024), neurons are generally assumed to exert either an excitatory or inhibitory effect on other neurons, so it would be of interest to integrate statistical tests to determine these interaction types.

Finally, our testing procedures rely on the asymptotic normality of the estimators. Establishing consistency results, as has been done for classic self-exciting Hawkes processes (Ogata, 1978; Lotz, 2024), would be an important next step. A natural starting point would be to derive theoretical guarantees for Model (HP) with inhibition, or for Model (GVM) restricted to self-exciting interactions with positive kernels. In both cases, computing the expected number of events in a finite interval is already nontrivial, making the extension of existing consistency results particularly challenging.

Acknowledgements

The three authors would like to thank the Editor, Associate Editor and anonymous reviewers for their constructive comments, which helped improve the estimation procedure and the manuscript. They would also like to warmly thank Eva Löcherbach for insightful suggestions regarding variable length memory models for neuroscience, and Arnaud Guyader for fruitful discussions on continuous Markov chains. This work is part of the project HAPPY ANR-23-CE40-0007 and 2022 DAE 103 EMERGENCE(S) - PROCECO.

References

Asmussen, S. (2003). *Applied Probability and Queues*, volume 51 of *Stochastic Modelling and Applied Probability*. New York: Springer.

Bacry, E., Bompaire, M., Gaïffas, S., & Muzy, J.-F. (2020). Sparse and low-rank multivariate Hawkes processes. *Journal of Machine Learning Research*, 21(50), 1–32.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.

Bonnet, A., Dion, C., Gindraud, F., & Lemler, S. (2022). Neuronal Network Inference and Membrane Potential Model using Multivariate Hawkes Processes. *Journal of Neuroscience Methods*, 372, 109550.

Bonnet, A., Herrera, M. M., & Sangnier, M. (2023). Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity. *Statistics and Computing*, 33(4), 91.

Bonnet, A. & Sangnier, M. (2025). Nonparametric estimation of Hawkes processes with RKHSs. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics* (pp. 3574–3582).: PMLR.

Brémaud, P. & Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3), 1563–1588.

Costa, M., Graham, C., Marsalle, L., & Tran, V. C. (2020). Renewal in Hawkes processes with self-excitation and inhibition. *Advances in Applied Probability*, 52(3), 879–915.

Daley, D. J. & Vere-Jones (2003). *An Introduction to the Theory of Point Processes*, volume 1 of *Probability and its Applications*. New York: Springer-Verlag, 2 edition.

Daley, D. J. & Vere-Jones, D. (2008). *An introduction to the Theory of Point Processes*, volume 2 of *Probability and its Applications*. Springer-Verlag, 2 edition.

Deutsch, I. & Ross, G. J. (2025). Estimating Product Cannibalisation in Wholesale using Multivariate Hawkes Processes with Inhibition. *Annals of Applied Statistics*, 19(1), 235–260.

Duval, C., Luçon, E., & Pouzat, C. (2022). Interacting Hawkes processes with multiplicative inhibition. *Stochastic Processes and their Applications*, 148, 180–226.

Galves, A. & Löcherbach, E. (2013). Infinite Systems of Interacting Chains with Memory of Variable Length—A Stochastic Model for Biological Neural Nets. *Journal of Statistical Physics*, 151(5), 896–921.

Galves, A. & Löcherbach, E. (2016). Modeling networks of spiking neurons as interacting processes with memory of variable length. *Journal de la Société Française de Statistique*, 157(1), 17–32.

Galves, A., Löcherbach, E., & Pouzat, C. (2024). *Probabilistic Spiking Neuronal Nets: Neuromathematics for the Computer Era*. Lecture Notes on Mathematical Modelling in the Life Sciences. Cham: Springer International Publishing.

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

Hodara, P. & Löcherbach, E. (2017). Hawkes processes with variable length memory and an infinite number of components. *Advances in Applied Probability*, 49, 84–107.

Joseph, S. & Jain, S. (2024a). A neural network based model for multi-dimensional non-linear Hawkes processes. *Journal of Computational and Applied Mathematics*, 447, 115889.

Joseph, S. & Jain, S. (2024b). Non-Parametric Estimation of Multi-dimensional Marked Hawkes Processes.

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

Lemonnier, R. & Vayatis, N. (2014). Nonparametric Markovian Learning of Triggering Kernels for Mutually Exciting and Mutually Inhibiting Multivariate Hawkes Processes. In T. Calders, F. Esposito, E. Hüllermeier, & R. Meo (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 161–176). Berlin, Heidelberg: Springer.

Loison, V., Staerman, G., & Moreau, T. (2025). UNHaP: Unmixing Noise from Hawkes Processes. In *International Conference on Artificial Intelligence and Statistics* Phuket, Thailand.

Lotz, A. (2024). A sparsity test for multivariate Hawkes processes.

Mei, H. & Eisner, J. (2017). The neural hawkes process: a neurally self-modulating multivariate point process. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17 (pp. 6757–6767). Red Hook, NY, USA: Curran Associates Inc.

Meng, Z., Wan, K., Huang, Y., Li, Z., Wang, Y., & Zhou, F. (2024). Interpretable Transformer Hawkes Processes: Unveiling Complex Interactions in Social Networks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24 (pp. 2200–2211). New York, NY, USA: Association for Computing Machinery.

Nicvert, L., Donnet, S., Keith, M., Peel, M., Somers, M. J., Swanepoel, L. H., Venter, J., Fritz, H., & Dray, S. (2024). Using the multivariate Hawkes process to study interactions between multiple species from camera trap data. *Ecology*, 105(4), e4237.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(2), 243–261.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1), 23–31.

Ogata, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Association*, 83(401), 9–27.

Olinde, J. & Short, M. (2020). A Self-limiting Hawkes Process: Interpretation, Estimation, and Use in Crime Modeling. *2020 IEEE International Conference on Big Data*, (pp. 3219).

Pan, Z., Wang, Z., Phillips, J. M., & Zhe, S. (2021). Self-adaptable point processes with nonparametric time decays. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21 (pp. 4594–4606). Red Hook, NY, USA: Curran Associates Inc.

Petersen, P. C. & Berg, R. W. (2016). Lognormal firing rate distribution reveals prominent fluctuation–driven regime in spinal motor networks. *eLife*, 5, e18805.

Radosevic, M., Willumsen, A., Petersen, P. C., Lindén, H., Vestergaard, M., & Berg, R. W. (2019). Decoupling of timescales reveals sparse convergent CPG network in the adult spinal cord. *Nature Communications*, 10(1).

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

Rizoiu, M.-A., Lee, Y., Mishra, S., & Xie, L. (2017). A Tutorial on Hawkes Processes for Events in Social Media.

Strata, P. & Harvey, R. (1999). Dale's principle. *Brain Res. Bull.*, 50, 349–50.

Sulem, D., Rivoirard, V., & Rousseau, J. (2023). Scalable and adaptive variational Bayes methods for Hawkes processes.

Sulem, D., Rivoirard, V., & Rousseau, J. (2024). Bayesian estimation of nonlinear Hawkes processes. *Bernoulli*, 30(2), 1257–1286.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., & van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272.

Zhang, Q., Lipani, A., Lipani, A., & Yilmaz, E. (2020). Self-attentive Hawkes process. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *ICML'20* (pp. 11183–11193).: JMLR.org.

Zuo, S., Jiang, H., Li, Z., Zhao, T., & Zha, H. (2020). Transformer Hawkes process. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *ICML'20* (pp. 11692–11702).: JMLR.org.

Appendix

A Proof of Proposition 1

Let us keep notations from Section 2.1. Denote $(T_k^j)_{k\in\mathbb{N}^*}$ the arrival times for each process N^j. Then, by monotony and sublinearity of $x \mapsto x^+$ on $\mathbb{R}$, for any $i \in \{1,\dots,d\}$, for any $t \geq 0$, a.s.:

$$\begin{aligned}
\lambda^i(t) &= \left(\mu_i + \sum_{j=1}^d \left[\sum_{k\in\mathbb{N}^*,\, T^i_{N^i(t)} \leq T_k^j < t} h_{ij}(t - T_k^j) + \sum_{k\in\mathbb{N}^*,\, T_k^j < T^i_{N^i(t)}} \widetilde{h}_{ij}(t - T_k^j)\right]\right)^+ \\
&\leq \left(\mu_i + \sum_{j=1}^d \left[\sum_{k\in\mathbb{N}^*,\, T^i_{N^i(t)} \leq T_k^j < t} \max(h_{ij}, \widetilde{h}_{ij})(t - T_k^j) + \sum_{k\in\mathbb{N}^*,\, T_k^j < T^i_{N^i(t)}} \max(h_{ij}, \widetilde{h}_{ij})(t - T_k^j)\right]\right)^+ \\
&= \left(\mu_i + \sum_{j=1}^d \sum_{T_k^j < t} \max(h_{ij}, \widetilde{h}_{ij})(t - T_k^j)\right)^+ \leq \mu_i + \sum_{j=1}^d \sum_{T_k^j < t} \max(h_{ij}, \widetilde{h}_{ij})^+(t - T_k^j)
\end{aligned}$$

thus, a.s.:, $\sum_{i=1}^d \lambda^i(t) \leq \sum_{i=1}^d \overline{\lambda}^i(t)$, where

$$\overline{\lambda}^i(t) = \mu_i + \sum_{j=1}^d \sum_{T_k^j < t} \max(h_{ij}, \widetilde{h}_{ij})^+(t - T_k^j)\,.$$

By the spectral radius assumption and by (Brémaud & Massoulié, 1996, Theorem 7), there exists a family $(N^1, ..., N^d)$ of point processes with respective conditional intensities $\overline{\lambda}^1, ..., \overline{\lambda}^d$. By locally upper bounding the sum $\sum_{i=1}^d \overline{\lambda}^d$ between two consecutive jump times, we can construct a one-dimensional pathwise constant $(\mathcal{F}_t^-)_{t\geq 0}$-predictable process $\overline{\lambda}(t)$ (where $(\mathcal{F}_t)_{t\geq 0}$ is the canonical filtration of N), and a one-dimensional stationary point process $\overline{N}$ with conditional intensity $\overline{\lambda}$, such that $\sum_{i=1}^d \lambda^i(t) \leq \overline{\lambda}(t)$ a.s. Thus, by thinning according to (Ogata, 1981, Proposition 1), there exists a multivariate point process $N = (N^1, ..., N^d)$ with conditional intensities $\lambda^1, ..., \lambda^d$.

B Proof of Theorem 1

For any $i \in \{1,\dots,d\}$, denote $(T_k^i)_{k\in\mathbb{N}^*}$ the arrival times for N^i. Then, for any $t \geq 0$:

$$\begin{aligned}
\lambda^i(t) &= \left(\mu_i + h_{ii}(t - T^i_{N^i(t)}) + \sum_{j\neq i} \sum_{T^i_{N^i(t)} \leq T_k^j < t} h_{ij}(t - T_k^j)\right)^+ \\
&\leq \mu_i + \left|h_{ii}(t - T^i_{N^i(t)})\right| + \sum_{j\neq i} \sum_{T^i_{N^i(t)} \leq T_k^j < t} \left|h_{ij}(t - T_k^j)\right| \\
&\leq C\left(1 + \sum_{j\neq i} N^j\left[T^i_{N^i(t)}, t\right)\right) =: \overline{\lambda}^i(t),
\end{aligned}$$

where $C := \sup_{1\leq i,j\leq d}(\mu_i + \|h_{ij}\|_\infty)$. Therefore, we can dominate the original intensity $\lambda = (\lambda^1, \dots, \lambda^d)$ by a piecewise constant stochastic process $\overline{\lambda} = (\overline{\lambda}^1, \dots, \overline{\lambda}^d)$, constant between two consecutive event times. By (Ogata, 1981, Proposition 1), the Hawkes process N can then be constructed via thinning from a dominating process $\overline{N}$ with intensity $\overline{\lambda}$. In particular:

$$\forall A \in \mathcal{B}(\mathbb{R}^d),\ \overline{N}(A) < \infty \implies N(A) < \infty\,.$$

We now focus on the process $\overline{N} = (\overline{N}^1, \dots, \overline{N}^d)$, defined by its intensity $\overline{\lambda}$. Our goal is to prove that $\overline{N}$ does not explode (which implies that the process N does not explode). To do so, we will first show that the intensity process $\overline{\lambda}$ forms a time-continuous Markov chain. This structure then allows us to apply results from Markov process theory to establish the non-explosion of $\overline{N}$. The intensity $\overline{\lambda}$ takes values in:

$$E = \bigcup_{i=1}^{d} \{(C(1+k_1), \dots, C, \dots, C(1+k_d)),\ (k_1, \dots, k_d) \in \mathbb{N}^d\}\,.$$

Here, the entry C at position i indicates that process i has just experienced a jump. The values $C(1+k_j)$ represent the self-excitation effect (the "+1") combined with the cumulative contribution from the $d-1$ other processes since the last event in dimension j.

For any $l \in \mathbb{N}$, denote $Z^{(l)} \in E$ the value of the intensity $\overline{\lambda}$ on the interval $(\overline{T}_{(l)}, \overline{T}_{(l+1)})$:

$$Z^{(l)} = \left(\overline{\lambda}^1(T^+_{(l)}), \dots, \overline{\lambda}^d(T^+_{(l)})\right)\,.$$

Initially, we have $Z^{(0)} = (C, \dots, C)$. Now suppose that $Z^{(l)} = (Z_1^{(l)}, \dots, Z_d^{(l)})$. If the next event time is from dimension $j \in \{1, \dots, d\}$, which occurs with probability:

$$Z_j^{(l)} / \sum_{k=1}^{d} Z_k^{(l)}\,,$$

the next state is given by:

$$Z^{(l+1)} = (Z_1^{(l)} + C, \dots, C, \dots, Z_d^{(l)} + C)\,,$$

where the value C appears at position j (corresponding to the dimension that just jumped), while the others are incremented by C. This defines a Markov chain $(Z^{(l)})_{l\in\mathbb{N}^*}$ taking values in E, with transition probabilities $(p_{zz'})_{z,z'\in E}$ as described above. Thus, the process $(\overline{\lambda}(t))_{t\geq 0}$ is a time-continuous Markov chain with discrete state space. The chain starts at $z_0 := (C, \dots, C)$, which is visited only once at time 0. After the first jump, the chain evolves within an irreducible class $E' = E\backslash\{z_0\}$. Recurrence of the chain can thus be studied by restricting to E'. According to (Asmussen, 2003, Proposition 2.4 (p.44)), a sufficient condition to ensure non-explosion of the process is that this restricted chain is recurrent.

The remainder of the proof focuses on establishing the recurrence of the chain. We treat the cases $d = 2$ and $d > 2$ separately: for $d = 2$, the transition graph is simple, making the proof more intuitive. For $d > 2$, the graph becomes significantly more complex, and the proof relies on a recurrence criterion.

Proof of recursion for $d = 2$

If $d = 2$, the state space is:

$$\begin{aligned} E :=& \{(C, C(1+k)),\ k \in \mathbb{N}\} \cup \{(C(1+k), C),\ k \in \mathbb{N}\} \\ =:& \{z_k^{(1)},\ k \in \mathbb{N}\} \cup \{z_k^{(2)},\ k \in \mathbb{N}\}\,, \end{aligned}$$

and the transition graph of the chain $(Z^{(l)})_{l\in\mathbb{N}}$ is illustrated below, with transition probabilities $(p_{z,z'})_{z,z'\in E}$ as described above. Intuitively, the chain's dynamics are governed by the occurrence

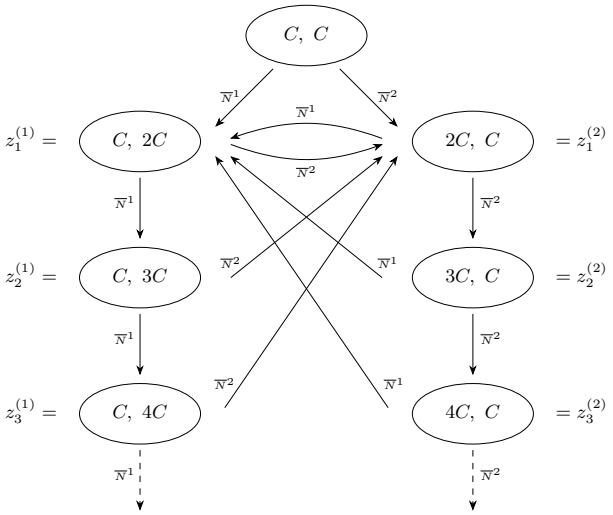


Fig. 8: Transition graph of the Markov chain $(Z^{(l)})_{l\in\mathbb{N}}$ for $d = 2$.

of events in the two underlying processes: when events are generated by the same process, the chain follows one of the two possible branches and continues to evolve in that direction. When an event is triggered by the other process, the chain returns to the state corresponding to the other branch.

Let us show that the chain restricted to $E\backslash\{z_0\}$ is recurrent. Since this chain is irreducible, it suffices to show that the state $z_1^{(1)}$ is recurrent. We define the return time to $z_1^{(1)}$:

$$T_{z_1^{(1)}} = \inf\{n \in \mathbb{N}^*,\, Z^{(n)} = z_1^{(1)} \,|Z^{(0)} = z_1^{(1)}\}\,.$$

Denote $\Pr_1 := \Pr_{z_1^{(1)}}$. Our goal is to show that $\Pr_1(T_{z_1^{(1)}} = \infty) = 0$, meaning that the chain returns to $z_1^{(1)}$ with probability 1. Examining the graph of the Markov chain, for $T_{z_1^{(1)}}$ to be infinite, one of two scenarios must occur: either only event times from $\overline{N}^1$ are observed indefinitely, or at some point, an event time from $\overline{N}^2$ occurs, after which only event times from $\overline{N}^2$ are observed. Therefore, we can decompose the probability as follows:

$$\begin{aligned}\Pr_1(T_{z_1^{(1)}} = \infty) &= \Pr_1\left(\forall l \geq 1,\, \overline{d}_l = 1\right) + \Pr_1\left(\exists k \geq 1,\, \forall l \leq k,\, \overline{d}_l = 1,\text{ and } \forall l > k,\, \overline{d}_l = 2\right)\\ &\leq \Pr_1\left(\forall l \geq 1,\, \overline{d}_l = 1\right) + \sum_{k\geq 1} \Pr_1\left(\forall l > k,\, \overline{d}_l = 2\right).\end{aligned}$$

However, $\Pr_1\left(\forall l \geq 1,\, \overline{d}_l = 1\right) = \lim\limits_{p\to\infty} \Pr_1\left(\forall 1 \leq l \leq p,\, \overline{d}_l = 1\right)$, with:

$$\begin{aligned}\Pr_1\left(\forall 1 \leq l \leq p,\, \overline{d}_l = 1\right) &= p_{z_1^{(1)},z_2^{(1)}} \dots p_{z_{p-1}^{(1)},z_p^{(1)}} = \prod_{l=1}^{p-1} \frac{C}{C + lC}\\ &= \frac{1}{p!} \underset{p\to\infty}{\longrightarrow} 0.\end{aligned}$$

Hence, we have $\Pr_1\left(\forall l \geq 1,\, \overline{d}_l = 1\right) = 0$. Similarly, for any $k \geq 1$, $\Pr_1\left(\forall l > k,\, \overline{d}_l = 2\right) = 0$. It follows that $\Pr_1(T_{z_1^{(1)}} = \infty) = 0$, i.e., $z_1^{(1)}$ is recurrent.

Proof of recursion for $d > 2$

According to (Asmussen, 2003, Proposition 5.3), the chain is recurrent if there exists a finite subset E_0 of E and a function $g : E \to \mathbb{R}$ such that $g(z) \underset{\|z\|\to\infty}{\longrightarrow} +\infty$ and:

$$\forall z \notin E_0,\, \sum_{z'\in E} p_{zz'} g(z') \leq g(z)\,.$$

Let us consider the following set E_0 and the following function $g : E \to \mathbb{R}$:

$$E_0 = \{z \in E,\, \|z\|_1 \leq Cd^2\}\,, \quad g(z) = \sum_{i=1}^{d} z_i\,.$$

Then E_0 is finite and $g(z) \underset{\|z\|\to\infty}{\longrightarrow} +\infty$. Let $z = (z_1, \ldots, z_d) \notin E_0$. Then the future possible states for the Markov chain are $z^i := (z_1 + C, \ldots, C, \ldots, z_d + C)$ for any $1 \leq i \leq d$. Therefore:

$$\begin{aligned}\sum_{z'\in E} p_{zz'} g(z') &= \sum_{i=1}^{d} \frac{z_i}{g(z)} g(z^i) = \sum_{i=1}^{d} \frac{z_i}{g(z)} \left(\sum_{j\neq i} z_j + dC\right) \\ &= \sum_{i=1}^{d} \frac{z_i}{g(z)} \left(g(z) - z_i + dC\right) = g(z) + \sum_{i=1}^{d} \frac{z_i}{g(z)} \left(-z_i + dC\right) \\ &= g(z) - \frac{\|z\|_2^2}{g(z)} + dC \sum_{i=1}^{d} \frac{z_i}{g(z)} \leq g(z) - \frac{\|z\|_2^2}{g(z)} + dC\,.\end{aligned}$$

However:

$$\frac{\|z\|_2^2}{g(z)} = \frac{\|z\|_2^2}{\|z\|_1} \geq \frac{\|z\|_1}{d} > dC,$$

which implies:

$$\sum_{z'\in E} p_{zz'} g(z') \leq g(z)\,,$$

thus the chain is recurrent.

C Proof of Proposition 2

We keep the same notations: we denote $(T_k^j)_{k\in\mathbb{N}^*}$ the arrival times for each process N^j, and $(T_{(l)})_{l\in\mathbb{N}^*}$ the ordered sequence of arrival times of the process $N = (N^1, ..., N^d)$ with d_l being the dimension which the arrival time $T_{(l)}$ belongs to. Let $l \in \mathbb{N}^*$. We separate cases $t \in]T_{(l)}, T_{(l+1)}[$ and $t = T_{(l+1)}$.

If $t \in]T_{(l)}, T_{(l+1)}[$: we start by noting that, for any $i \in \{1, \ldots, d\}$: $T^i_{N^i(t)} = T^i_{N^i(T_{(l)})}$, therefore:

$$\eta^i(t) = \sum_{j=1}^{d} \sum_{T^i_{N^i(t)} \leq T_k^j < t} \alpha_{ij} e^{-\beta_i(t - T_k^j)} = e^{-\beta_i(t-T_{(l)})} \sum_{j=1}^{d} \sum_{T^i_{N^i(T_{(l)})} \leq T_k^j < t} \alpha_{ij} e^{-\beta_i(T_{(l)} - T_k^j)}\,,$$

and, by the assumption that two different processes cannot jump at the same time:

$$\sum_{T^i_{N^i(T_{(l)})} \leq T_k^j < t} \alpha_{ij} e^{-\beta_i(T_{(l)} - T_k^j)} = \begin{cases} \displaystyle\sum_{T^i_{N^i(T_{(l)})} \leq T_k^j < T_{(l)}} \alpha_{ij} e^{-\beta_i(T_{(l)} - T_k^j)} & \text{if } j \neq d_l \\ \displaystyle\sum_{T^i_{N^i(T_{(l)})} \leq T_k^j < T_{(l)}} \alpha_{ij} e^{-\beta_i(T_{(l)} - T_k^j)} + \alpha_{i,d_l} & \text{if } j = d_l \end{cases},$$

therefore:

$$\eta^i(t) = (\eta^i(T_{(l)}) + \alpha_{i,d_l}) e^{-\beta_i(t-T_{(l)})}\,.$$

Similarly, we also have: $\widetilde{\eta}^i_{aux}(t) = (\widetilde{\eta}^i_{aux}(T_{(l)}) + \widetilde{\alpha}_{i,d_l}) e^{-\widetilde{\beta}_i(t-T_{(l)})}$. For $\widetilde{\eta}^i$, we have:

$$\widetilde{\eta}^i(t) = \sum_{j=1}^{d} \sum_{T_k^j < T^i_{N^i(t)}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i(t - T_k^j)} = e^{-\widetilde{\beta}_i(t-T_{(l)})} \sum_{j=1}^{d} \sum_{T_k^j < T^i_{N^i(T_{(l)})}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i(T_{(l)} - T_k^j)} = e^{-\widetilde{\beta}_i(t-T_{(l)})} \widetilde{\eta}^i(T_{(l)})\,.$$

If $t = T_{(l+1)}$, we have:

$$T^i_{N^i(T_{(l+1)})} = \begin{cases} T^i_{N^i(T_{(l)})} & \text{if } i \neq d_{l+1} \\ T_{(l+1)} & \text{if } i = d_{l+1} \end{cases},$$

therefore if $i \neq d_{l+1}$, then the results for $t \in]T_{(l)}, T_{(l+1)}[$ hold, i.e.,

$$\begin{aligned} \eta^i(T_{(l+1)}) &= (\eta^i(T_{(l)}) + \alpha_{i,d_l}) e^{-\beta_i (T_{(l+1)} - T_{(l)})}, \\ \widetilde{\eta}^i_{aux}(T_{(l+1)}) &= (\widetilde{\eta}^i_{aux}(T_{(l)}) + \widetilde{\alpha}_{i,d_l}) e^{-\widetilde{\beta}_i (T_{(l+1)} - T_{(l)})}, \text{ and} \\ \widetilde{\eta}^i(T_{(l+1)}) &= \widetilde{\eta}^i(T_{(l)}) e^{-\widetilde{\beta}_i (T_{(l+1)} - T_{(l)})}. \end{aligned}$$

If $i = d_{l+1}$, then $T^i_{N^i(T_{(l)})} = T_{(l+1)}$, so there are no jumps from any process in $[T^i_{N^i(T_{(l+1)})}, T_{(l+1)}[$, therefore:

$$\eta^i(T_{(l+1)}) = 0 \text{ and } \widetilde{\eta}^i_{aux}(T_{(l+1)}) = 0\,,$$

and, for $\widetilde{\eta}$:

$$\begin{aligned} \widetilde{\eta}^i(T_{(l+1)}) &= \sum_{j=1}^{d} \sum_{T^j_k < T^i_{N^i(T_{(l+1)})}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i (T_{(l+1)} - T^j_k)} \\ &= e^{-\widetilde{\beta}_i (T_{(l+1)} - T_{(l)})} \sum_{j=1}^{d} \sum_{T^j_k < T_{(l+1)}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i (T_{(l)} - T^j_k)} \\ &= e^{-\widetilde{\beta}_i (T_{(l+1)} - T_{(l)})} \left(\sum_{j=1}^{d} \sum_{T^j_k < T^i_{N^i(T_{(l)})}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i (T_{(l)} - T^j_k)} + \sum_{j=1}^{d} \sum_{T^i_{N^i(T_{(l)})} \leq T^j_k < T_{(l+1)}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i (T_{(l)} - T^j_k)} \right) \\ &= e^{-\widetilde{\beta}_i (T_{(l+1)} - T_{(l)})} \left(\widetilde{\eta}^i(T_{(l)}) + \sum_{j=1}^{d} \sum_{T^i_{N^i(T_{(l)})} \leq T^j_k < T_{(l+1)}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i (T_{(l)} - T^j_k)} \right), \end{aligned}$$

and:

$$\begin{aligned} \sum_{j=1}^{d} \sum_{T^i_{N^i(T_{(l)})} \leq T^j_k < T_{(l+1)}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i (T_{(l)} - T^j_k)} &= \sum_{j=1}^{d} \sum_{T^i_{N^i(T_{(l)})} \leq T^j_k < T_{(l)}} \widetilde{\alpha}_{ij} e^{-\widetilde{\beta}_i (T_{(l)} - T^j_k)} + \widetilde{\alpha}_{i,d_l} \\ &= \widetilde{\eta}^i_{aux}(T_{(l)}) + \widetilde{\alpha}_{i,d_l}\,, \end{aligned}$$

therefore:

$$\widetilde{\eta}^i(T_{(l+1)}) = e^{-\widetilde{\beta}_i (T_{(l+1)} - T_{(l)})} \left(\widetilde{\eta}^i(T_{(l)}) + \widetilde{\eta}^i_{aux}(T_{(l)}) + \widetilde{\alpha}_{i,d_l} \right).$$

D Illustration of Preprocessing Steps

Figures 9, 10 and 11 respectively display the normalised cumulative spike counts $N^i(t)/t$ for the original trials on the normalised time window $[0, 10]$, for the trials after removing non-relevant ones and non-active neurons, and for the 25 resampled trials on the time window $[0, 30]$, as explained in Section 5.

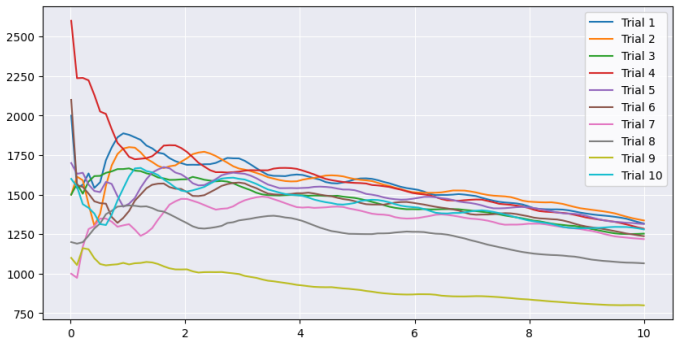


Fig. 9: Normalised cumulative spike counts $N^i(t)/t$ for the 10 original neuronal trials, shown on the normalised time interval $[0, 10]$, obtained after Step 1.

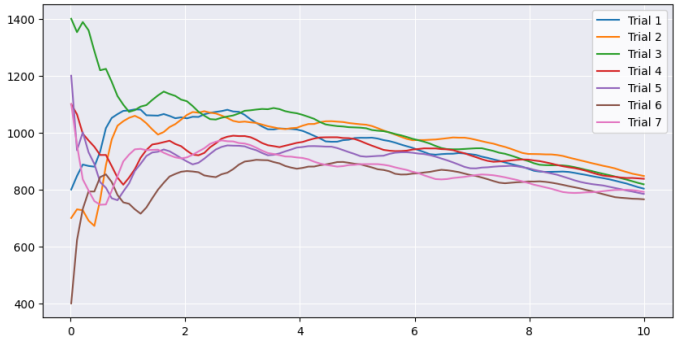


Fig. 10: Normalised cumulative spike counts $N^i(t)/t$ for the 7 neuronal trials retained after Step 3.

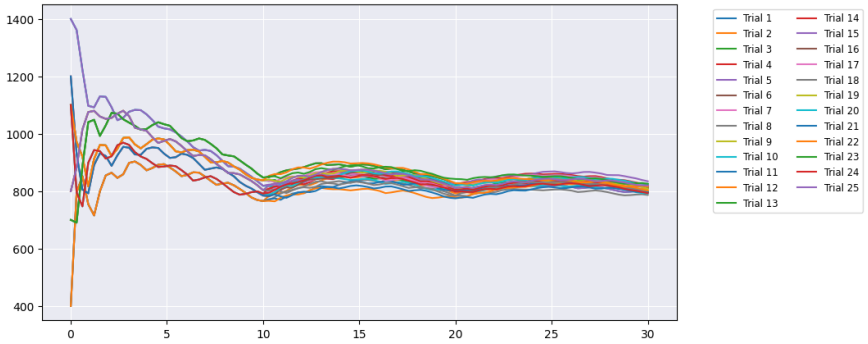


Fig. 11: Normalised cumulative spike counts $N^i(t)/t$ for the 25 trials obtained after the resampling procedure in Step 6.