A SPECTRAL ANALYSIS APPROACH FOR ESTIMATION OF A NOISY HAWKES PROCESS

Anna Bonnet [1], Félix Cheysson [2], Miguel Martinez Herrera [3] & Maxime Sangnier[4]

[1] *Sorbonne Université, France, anna.bonnet@sorbonne-universite.fr*
[3] *Gustave Eiffel University, France, felix.cheysson@univ-eiffel.fr*
[3] *Sorbonne Université, France, miguel.martinez_herrera@sorbonne-universite.fr*
[4] *Sorbonne Université, France, maxime.sangnier@sorbonne-universite.fr*

Résumé. Les méthodes classiques d'estimation pour les processus de Hawkes auto-excitants sont fondées sur l'hypothèse que les événements observés correspondent à une réalisation d'un processus de Hawkes, sans perturbation aucune. Néanmoins, il est raisonnable d'envisager qu'en pratique les observations sont bruitées, en un sens dépendant du contexte. Il devient alors essentiel de modéliser la source de bruit et d'adapter les procédures d'inférence pour obtenir des estimations pertinentes d'un tel processus, dit bruité. Bien que plusieurs modélisations existent, nous considérons dans ce travail un mécanisme de bruit consistant en l'ajout d'observations qui proviennent d'un processus externe. Plus précisément, on suppose que les événements correspondent à la réunion de points provenant d'un processus de Hawkes et d'un processus de Poisson indépendants, sans que l'origine de ces points ne soit connue. Puisque dans ce contexte les méthodes canoniques d'estimation, comme le maximum de vraisemblance ou l'algorithme *Expectation-Maximisation*, sont soit impossibles à mettre en œuvre soit numériquement trop coûteuses, nous proposons une nouvelle procédure d'inférence fondée sur l'analyse spectrale des moments de second ordre du processus de Hawkes bruité. Nos contributions incluent une caractérisation du spectre de Bartlett grâce à sa densité spectrale, ainsi que des conditions d'identifiabilité du modèle statistique dans les cas de processus univariés et bivariés à interactions exponentielles. Un nouvel estimateur construit sur la maximisation de la log-vraisemblance spectrale est présenté et son comportement est analysé numériquement sur des données synthétiques. Il est ainsi montré qu'en plus d'être implémentable sans connaître l'origine des événements (processus de Hawkes ou de Poisson), l'estimateur proposé estime correctement les deux processus.

Mots-clés. Processus de Hawkes, processus ponctuel, analyse spectrale, estimation paramétrique, superposition, identifiabilité.

Abstract. Classical estimation methods for Hawkes processes rely on the assumption that observed event times are indeed a realisation of a Hawkes process, without considering any potential perturbation of the model. However, in practice, observations are often altered by some noise, the form of which depends on the context. It is then required to model the alteration mechanism in order to infer accurately such a noisy Hawkes process. While several models exist, we consider, in this work, the observations to be the indistinguishable union of event times coming from a Hawkes process and from an independent Poisson process. Since standard inference methods (such as maximum likelihood or Expectation-Maximisation) are either unworkable or numerically prohibitive in this context, we propose an estimation procedure based on the spectral analysis of second order properties of the noisy Hawkes process.

Novel results include a characterisation of the Bartlett spectrum thanks to its spectral density, and sufficient conditions for the identifiability of the ensuing statistical model with exponential interaction functions for both univariate and bivariate processes. A new estimator based on maximising the spectral log-likelihood is then described, and its behaviour is numerically illustrated on synthetic data. Besides being free from knowing the source of each observed time (Hawkes or Poisson process), the proposed estimator is shown to perform accurately in estimating both processes.

Keywords. Hawkes process, point process, spectral analysis, parametric estimation, superposition, identifiability.

1 Introduction

Hawkes processes, introduced in Hawkes (1971), are a class of point processes that have been originally used to model self-exciting phenomena and more recently other types of past-dependent behaviours. Their fields of applications are wide and include for instance seismology (Ogata, 1988, 1998), neuroscience (Chornoboy et al., 1988; Lambert et al., 2018), criminology (Olinde and Short, 2020), finance (Embrechts et al., 2011; Bacry et al., 2015) and biology (Gupta et al., 2018), to mention a few. Consequently, there has been a deep focus on estimation techniques for Hawkes processes. Among them, let us mention maximum likelihood approaches (Ogata, 1978; Ozaki, 1979; Guo et al., 2018), methods of moments (Da Fonseca and Zaatour, 2013), least-squares contrast minimisation (Reynaud-Bouret et al., 2014; Bacry et al., 2020), Expectation-Maximisation (EM) procedures (Lewis and Mohler, 2011), methods using approximations through autoregressive models (Kirchner, 2017).

As a consequence of this, there has been a deep focus on estimation techniques for Hawkes processes through maximum likelihood estimation (Ogata, 1978; Ozaki, 1979; Guo et al., 2018), method of moments (Da Fonseca and Zaatour, 2013), least-squares contrast minimisation (Bacry et al., 2020), EM procedure (Lewis and Mohler, 2011), using approximations through autoregressive models (Kirchner, 2017), by decomposition on histogram basis (Reynaud-Bouret et al., 2014).

All of these methods assume that the history of the point process has been accurately observed, which is untrue in many applications. For example, when reading spike train for the study of neuronal activation networks, it is likely to detect additional points not corresponding to real events.

Although the study of noised observations is a common issue in other contexts, it has been scarcely studied in the point processes setting. The works of Lund and Rudemo (2000) focus on a general model where a point process' event times are randomly thinned, displaced and an additional external "noise" is added in the form of an exogeneous point process.

In a recent contribution by Cheysson and Lang (2022), the authors employ the spectral analysis of point processes to study a noised version of a Hawkes process. The model consists on an aggregate count of the real event times within specific time intervals instead of precise occurrences in time. By leveraging the properties of the Bartlett spectrum, they propose an estimator obtained through means of maximisation of the spectral log-likehood. They illustrate its efficacy on synthetic and real-world data on a parametric context for different kernel functions.

In this communication, we consider the undistinguishable superposition of a Hawkes process and a homogeneous Poisson process. We derive a parametric estimation procedure by leveraging the general expression for the spectral density of this model. This approach has the advantage of not needing any knowledge on the source of the observed event times. By working with the exponential kernel function for the Hawkes process, we provide sufficient conditions for the identifiability of our model. Our numerical procedure is implemented in Python and we carry out a numerical study on synthetic data for all different identifiable scenarios illustrating the accuracy of our estimations.

2 Mathematical setting

Let $H = (H_1, \ldots, H_d)$ be a stationary multivariate Hawkes process on $\mathbb{R}$ defined by its conditional intensity functions λ_i^H ($i \in \{1, \ldots, d\}$): for all $t \in \mathbb{R}$,

$$\lambda_i^H(t) = \mu_i + \sum_{j=1}^{d} \int_{-\infty}^{t} h_{ij}(t-s)\, H_j(\mathrm{d}s) = \mu_i + \sum_{j=1}^{d} \sum_{T_k^{H_j} \leq t} h_{ij}(t - T_k^{H_j})\,, \tag{1}$$

where $\mu_i > 0$ is the baseline intensity of process H_i, $h_{ij} > 0$ is the interaction or kernel function describing the effect of process H_j on process H_i, and $(T_k^{H_j})_{k\geq 1}$ denotes the event times of H_j.

By defining the matrix $S^+ = (\|h_{ij}\|_1)_{ij}$ where

$$\|h_{ij}\|_1 = \int_{-\infty}^{+\infty} h_{ij}(t)\,\mathrm{d}t\,,$$

the stationarity condition of H reduces to controlling the spectral radius of S^+: $\rho(S^+) < 1$ (Brémaud and Massoulié, 1996). In this communication, we consider the exponential kernel function defined for any integers i, j as:

$$h_{ij}(t) = \alpha_{ij}\beta_i \mathrm{e}^{-\beta_i t}\,, \quad \text{for } t > 0, \tag{2}$$

with $\alpha_{ij} \geq 0$ and $\beta_i > 0$, and so $S^+ = (\alpha_{ij})_{ij}$.

Our goal is to study a noisy version of the Hawkes process where the sequence of event times $(T_k^H)_{k\geq 1}$ of H is contaminated by the event times from another process, which is chosen to be a multivariate Poisson process.

Formally, let $P = (P_1, \ldots, P_d)$ be a multivariate process made up with d independent homogeneous Poisson processes on $\mathbb{R}$ (denoted $P_1, \ldots, P_d$) with shared intensity $\lambda_0 > 0$, supposed to be independent from H. We note the event times $(T_k^{P_i})_{k\geq 1}$.

We consider then the point process $N = (N_1, \dots, N_d)$ defined as the superposition of H and P. The sequence of event times $(T_k^{N_i})_{k\geq 1}$ of N_i is the ordered union of $(T_k^{H_i})_{k\geq 1}$ and $(T_k^{P_i})_{k\geq 1}$. Throughout this paper we will refer to N as the **noisy Hawkes process**.

Our goal is to estimate both processes parameters (i.e. the baselines μ_i, the kernels h_{ij} and the shared Poisson intensity λ_0) from the sole observation of $(T_k^{N_i})_{k\geq 1}$. Our method consists in characterising a multivariate point process by its matrix-valued spectral density function, denoted $\mathbf{f} : \mathbb{R} \to \mathbb{C}^{d\times d}$ (with for all $\omega \in \mathbb{R}$, $\mathbf{f}(\omega) = (f_{ij}(\omega))_{1\leq i,j\leq d}$), which is related to second-order measures (Bartlett, 1963).

We define the second order moment measures M_{ij}^N for any i, j as:

$$M_{ij}^N(A, B) \coloneqq \mathbb{E}[N_i(A)N_j(B)] = \int_{A\times B} N_i(\mathrm{d}x)\, N_j(\mathrm{d}y)\,.$$

As the process is stationary, M_{ij}^N can be decomposed in a product of $\ell_{\mathbb{R}}$ and a so-called reduced measure $\breve{M}_{ij}^N$, such that for any bounded measurable function g of bounded support

$$\int_{\mathbb{R}^2} g(x, y)\, M_{ij}^N(\mathrm{d}x, \mathrm{d}y) = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x, x+u)\, \ell_{\mathbb{R}}(\mathrm{d}x)\breve{M}_{ij}^N(\mathrm{d}u)\,. \tag{3}$$

The spectral density function can then be expressed as:

$$f_{ij}^N(\omega) = \int_{\mathbb{R}} \mathrm{e}^{-2\pi i x\omega}\, \breve{M}_{ij}^N(\mathrm{d}x) - \mathbb{E}[\lambda_i^H(0)]\mathbb{E}[\lambda_j^H(0)]\mathbb{1}_{\omega=0}, \quad \forall\omega \in \mathbb{R}\,. \tag{4}$$

Given some observed multivariate times $(T_k^N)_{k\geq 1}$ (in a prescribed time window $[0, T]$), for any pair $(i, j) \in \{1, \dots, d\}^2$, spectral density functions can be approximated by the cross-periodograms, defined for all $\omega \in \mathbb{R}$ by:

$$I_{ij}^T(\omega) = \sum_{k=1}^{N_i(T)}\sum_{l=1}^{N_j(T)} \mathrm{e}^{-2\pi i\omega(T_k^{N_i} - T_l^{N_j})}\,, \tag{5}$$

where $N_i(t) = N_i([0, t))$. The matrix-valued function $\mathbf{I}^T : \omega \in \mathbb{R} \mapsto (I_{ij}^T(\omega))_{1\leq i,j\leq d}$ can be computed regardless of the knowledge of the source of the event times.

In the scope of statistical inference, a parametric model for spectral density functions is considered:

$$\mathcal{P} = \left\{\mathbf{f}_\theta^N : \mathbb{R} \to \mathbb{C}^{d\times d}, \theta = (\mu, h, \lambda_0) \in \Theta\right\}\,.$$

Then, the so-called Whittle estimator $\mathbf{f}_{\hat{\theta}}^N$ of $\mathbf{f}$ (Whittle, 1952), can be obtained by maximising the approximate spectral log-likelihood (Brillinger, 2012; Düker and Pipiras, 2019; Villani et al., 2022):

$$\ell_T(\theta) = -\frac{1}{T}\sum_{k=1}^{M}\left(\log(\det(\mathbf{f}_\theta^N(\omega_k))) + \mathrm{Tr}\left(\mathbf{f}_\theta^N(\omega_k)^{-1}\mathbf{I}^T(\omega_k)\right)\right), \tag{6}$$

where det and Tr are respectively the determinant and trace of matrices, $\omega_k = k/T$, $k \in \{1, 2, \dots\}$ and M is a hyperparameter.

3 Spectral density and estimation

3.1 Superposition of processes

As our model focuses on the superposition of two point processes, the following general result can be first established.

Proposition 3.1. *Let X, Y be two independent stationary multivariate point processes on $\mathbb{R}$, admitting respective spectral density matrices $\mathbf{f}^X$, $\mathbf{f}^Y$.*

Then $N = X + Y$ also admits a spectral density matrix and:

$$\mathbf{f}^N = \mathbf{f}^X + \mathbf{f}^Y . \tag{7}$$

We can then explicit the spectral matrix $\mathbf{f}^N$ by leveraging Proposition 3.1. Indeed, under stationarity conditions, the spectral matrix $\mathbf{f}^H$ of a Hawkes process is known to depend on the Fourier transform of the interaction functions and on the mean densities (Daley and Vere-Jones, 2003, Example 8.3(c)). More precisely, we define the matrix of Fourier transform interactions as:

$$\tilde{h}(\omega) := (\tilde{h}_{ij}(\omega))_{ij} ,$$

where $\tilde{h}_{ij}$ denotes the Fourier transform of h_{ij}.

The mean intensities m_i^H of each process H_i can be obtained by:

$$\begin{pmatrix} m_1^H \\ \vdots \\ m_d^H \end{pmatrix} = (I_d - \tilde{h}(0))^{-1} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} ,$$

where I_d is the identity matrix of dimension d, and the spectral matrix by (Daley and Vere-Jones, 2003, Equation 8.3.11):

$$\mathbf{f}^H(\omega) = (I_d - \tilde{h}(-\omega))^{-1} \operatorname{Diag}(m^H)(I_d - \tilde{h}(\omega)^T)^{-1} , \tag{8}$$

with $\operatorname{Diag}(m^H)$ defined as the diagonal matrix formed by the mean intensities m_i^H.

As this expression is still true for a homogeneous Poisson process (with $\tilde{h} \equiv 0$), the spectral matrix of a noisy Hawkes process N reads:

$$\mathbf{f}(\omega) = \mathbf{f}^H(\omega) + \lambda_0 I_d , \tag{9}$$

with $\mathbf{f}^H$ defined by Equation (8). With this result, we can establish an estimation method by optimising the spectral log-likelihood in Equation (6).

3.2 Identifiability in the univariate framework

When working in the univariate framework, we define the exponential model for noisy Hawkes process as:

$$\mathcal{Q} = \left\{ f_\theta^N : \theta := (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}_{>0} \times [0,1) \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \right\} ,$$

where f_θ^N is the single spectral density function of the candidate noisy Hawkes process, α and β are the parameters of the single kernel (see Equations (2) and (4)).

By leveraging Proposition 3.1, the spectral density reduces to:

$$f_\theta^N(\omega) = \frac{\mu}{1-\alpha}\alpha\beta^2(2-\alpha)\left(\frac{1}{(\beta(1-\alpha))^2 + (2\pi\omega)^2}\right) + \left(\frac{\mu}{1-\alpha} + \lambda_0\right) , \tag{10}$$

and we obtain the following result concerning identifiability.

Proposition 3.2. *The model $\mathcal{Q}$ is not identifiable. In particular, for any admissible parameter $\theta = (\mu, \alpha, \beta, \lambda_0)$ there exists an infinite number of parameters θ' such that $f_\theta = f_{\theta'}$.*

However, if we assume that one of the four quantities in θ is known, then the reduced model defined by the remaining triplet of parameters is identifiable.

3.3 Identifiability in the bivariate framework

In the bivariate framework, let now consider the model:

$$\mathcal{Q}_\Lambda = \left\{ \mathbf{f}_\theta : \theta := (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}^2_{>0} \times \Lambda \times \mathbb{R}^2_{>0} \times \mathbb{R}_{>0}, \rho(\alpha) < 1 \right\} ,$$

where $\Lambda \subset [0,1)^{2\times 2}$. Since identifiability when $\Lambda = [0,1)^{2\times 2}$ deserves an intricate analysis, only two particular situations are presented below.

Proposition 3.3 presents two non-identifiable scenarios and in Proposition 3.4 we establish sufficient conditions for identifiability of $\mathcal{Q}_\Lambda$.

Proposition 3.3. *The model $\mathcal{Q}_\Lambda$ is not identifiable in both situations:*

1. $\Lambda = \{\alpha \in [0,1)^{2\times 2} : \alpha_{12} = \alpha_{21} = 0\}$*, that is for diagonal matrices α (with possibly null entries).*

2. $\Lambda = \{\alpha \in [0,1)^{2\times 2} : \alpha_{11}\alpha_{12} = 0, \alpha_{21}\alpha_{22} > 0 \text{ or } \alpha_{11}\alpha_{12} > 0, \alpha_{21}\alpha_{22} = 0\}$*, that is for matrices α with a null row and a row with positive entries.*

Proposition 3.4. *The model $\mathcal{Q}_\Lambda$ is identifiable in both situations:*

1. $\Lambda = \{\alpha \in [0,1)^{2\times 2} : \alpha_{11} \geq 0, \alpha_{21} > 0, \alpha_{12} = \alpha_{22} = 0 \text{ or } \alpha_{11} = \alpha_{21} = 0, \alpha_{12} \geq 0, \alpha_{22} > 0\}$*, that is for matrices α with a null column and a positive entry on the antidiagonal.*

2. $\Lambda = \{\alpha \in [0,1)^{2\times 2} : \alpha_{11} > 0, \alpha_{22} > 0, \alpha_{12}\alpha_{21} = 0\}$*, that is for matrices α with a positive diagonal and at least a null entry on the antidiagonal.*

4 Numerical results

This section depicts a simple numerical illustration of the behaviour of the proposed estimator $\hat{\theta}$, obtained by maximising the spectral log-likelihood ℓ_T (see Equation (6)) thanks to the L-BFGS-B method implemented in the `scipy` Python package. A univariate process is considered, along with the exponential model $\mathcal{Q}$. It results that $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\lambda}_0)$, that is the base intensity $\hat{\mu}$ of the Hawkes process, the interaction magnitude $\hat{\alpha}$ and the decay rate $\hat{\beta}$ of the kernel; and the intensity $\hat{\lambda}_0$ of the noise.

Observations are simulated according to an exponential Hawkes process with $(\mu, \alpha, \beta) = (1.0, 0.5, 1.0)$ and a Poisson process with $\lambda_0 = 1.2$ (so the target parameter is $\theta^\star = (\mu, \alpha, \beta, \lambda_0) = (1.0, 0.5, 1.0, 1.2)$) and the behaviour of $\hat{\theta}$ is analysed via the relative error $\frac{\|\hat{\theta} - \theta^\star\|_2}{\|\theta^\star\|_2}$ averaged over 50 repetitions. This setting is without loss of generality and, in particular all conclusions are consistent with other settings such as $\lambda_0 = 0.4$ and $\lambda_0 = 2.0$ (either decreased or increased noises).

Since the considered model $\mathcal{Q}$ is not identifiable (see Proposition 3.2) unless fixing a parameter, four situations are presented (see Figure 1): estimating i) (α, β, λ) with μ fixed, ii) (μ, β, λ) with α fixed, iii) (μ, α, λ) with β fixed, iv) (μ, α, β) with λ_0 fixed. The fixed parameter is set to the target value each time.

Figure 1 presents the trend of the relative errors with respect to the number of time events (top panels) and to the computation time (bottom panels). As expected, they decrease to 0, suggesting that the estimator $\hat{\theta}$ converges (in quadratic mean) to $\theta^\star$ as the horizon T goes to infinity.

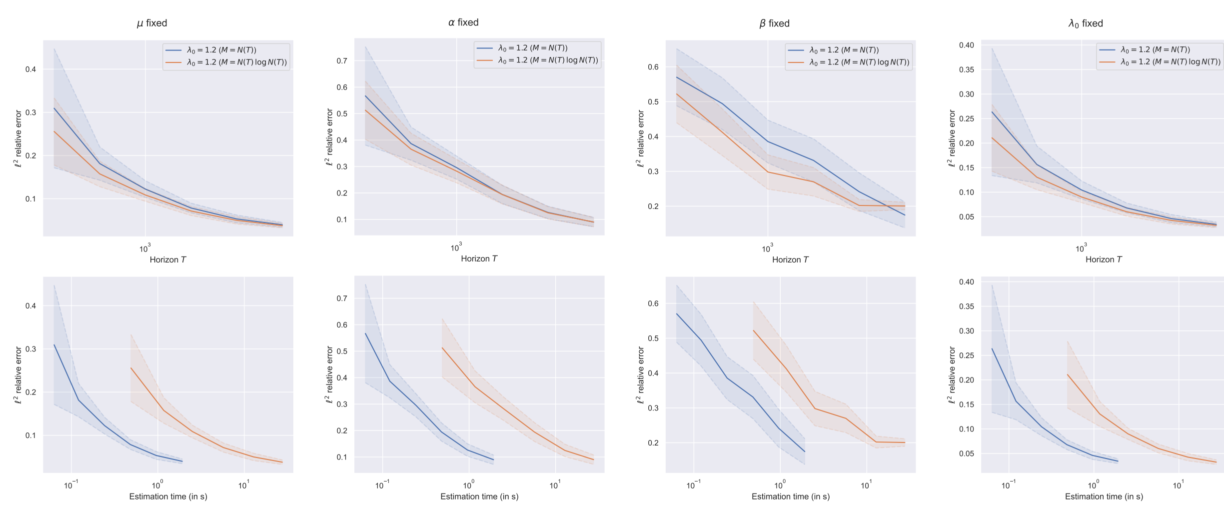

Figure 1: Relative estimation error respectively for μ, α, β and λ_0 fixed (columns from left to right) with respect to the time horizon T (top) and the computation time (bottom).

Moreover, as suggested in (Pham, 1996), the choice of the hyperparameter M is important. We follow (Cheysson and Lang, 2022), that proposes to consider M increasing with the number of observed times $N(T)$. Figure 1 illustrates the trend of the relative errors when $M = N(T)$ (blue curves) and when $M = N(T) \log N(T)$ (orange curves). It appears, first, that the relative errors are comparable for a given number of points $N(T)$ (top panels); second, that the case $M = N(T) \log N(T)$ requires a fitted time which is roughly 10 times greater than that for $M = N(T)$. Overall, it seams preferable to choose $M = N(T)$.

5 Conclusion

In this communication, we proposed an estimation procedure for the superposition of a Hawkes process and a Poisson process when missing information about the source of each process' event times. Our first contribution is providing explicit expressions for the spectral density of any superposition of two independent point processes. We established sufficient conditions for the identifiability of such a model. We then developed and implemented an estimator through the maximisation of the spectral log-likelihood in such scenarios and illustrated the efficiency of our method on synthetic data.

Future work include a detailed study of the multivariate setting, both from the theoretical and from the numerical points of view. On the first hand, it will be tried to extend the identifiability results to dimension $d \geq 3$. On the other hand, our estimation method will be numerically assessed on real datasets regarding neuronal activity with multiple subprocesses.

References

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

Bacry, E., Mastromatteo, I., and Muzy, J.-F. (2015), "Hawkes Processes in Finance," *Market Microstructure and Liquidity*, 01, 1550005.

Bartlett, M. S. (1963), "The Spectral Analysis of Point Processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, 25, 264–296.

Brillinger, D. R. (2012), *Statistical Inference for Stationary Point Processes*, New York, NY: Springer New York, 499–543.

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

Cheysson, F. and Lang, G. (2022), "Spectral estimation of Hawkes processes from count data," *The Annals of Statistics*, 50, 1722 – 1746.

Chornoboy, E. S., Schramm, L. P., and Karr, A. F. (1988), "Maximum likelihood identification of neural point process systems," *Biological Cybernetics*, 59, 265–275.

Da Fonseca, J. and Zaatour, R. (2013), "Hawkes Process: Fast Calibration, Application to Trade Clustering, and Diffusive Limit," *Journal of Futures Markets*, 34, 548–579.

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

Düker, M.-C. and Pipiras, V. (2019), "Asymptotic results for multivariate local Whittle estimation with applications," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*.

Embrechts, P., Liniger, T., and Lin, L. (2011), "Multivariate Hawkes processes: an application to financial data," *Journal of Applied Probability*, 48, 367–378.

Guo, X., Hu, A., Xu, R., and Zhang, J. (2018), "Consistency and Computation of Regularized MLEs for Multivariate Hawkes Processes," Preprint at https://arxiv.org/abs/1810.02955.

Gupta, A., Farajtabar, M., Dilkina, B., and Zha, H. (2018), "Discrete Interventions in Hawkes Processes with Applications in Invasive Species Management," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, International Joint Conferences on Artificial Intelligence Organization.

Hawkes, A. (1971), "Spectra of Some Self-Exciting and Mutually Exciting Point Processes," *Biometrika*, 58, 83–90.

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

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

Lewis, E. and Mohler, G. (2011), "A Nonparametric EM Algorithm for Multiscale Hawkes Processes," *Journal of Nonparametric Statistics*, 1, 1–20.

Lund, J. and Rudemo, M. (2000), "Models for Point Processes Observed with Noise," *Biometrika*, 87, 235–249.

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

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

— (1998), "Space-Time Point-Process Models for Earthquake Occurrences," *Annals of the Institute of Statistical Mathematics*, 50, 379–402.

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

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

Pham, D. (1996), "Blind separation of instantaneous mixture of sources via an independent component analysis," *IEEE Transactions on Signal Processing*, 44, 2768–2779.

Reynaud-Bouret, P., Rivoirard, V., Grammont, F., and Tuleau-Malot, C. (2014), "Goodness-of-Fit Tests and Nonparametric Adaptive Estimation for Spike Train Analysis," *The Journal of Mathematical Neuroscience*, 4, 3.

Villani, M., Quiroz, M., Kohn, R., and Salomone, R. (2022), "Spectral Subsampling MCMC for Stationary Multivariate Time Series with Applications to Vector ARTFIMA Processes," *Econometrics and Statistics*.

Whittle, P. (1952), "Some results in time series analysis," *Scandinavian Actuarial Journal*, 1952, 48–60.