Functional marked Hawkes processes, with a view to model hurricanes data

Giada Lo Galbo[1], Giada Adelfio[1], Marcello Chiodi[1] and Mohammad Ghorbani[2]

[1] Department of Economics, Business and Statistics, University of Palermo, Italy, giada.logalbo@unipa.it, giada.adelfio@unipa.it, marcello.chiodi@unipa.it
[2] Department of Engineering Sciences and Mathematics, Luleå University of Technology, Sweden, mohammad.ghorbani@ltu.se

Abstract. While statistical methods and models for spatial point processes with scalar marks are rather well-established, their counterparts for functional marked spatio-temporal point processes remain in their infancy. As a result, this has become a growing research topic, driven by advances in data collection, storage, and availability. Representative instances include earthquake and hurricane data, which involve functional marks and exhibit self-exciting behaviour. We contribute to this emerging area by developing a marked functional Hawkes point process and applying the proposed model to analyse hurricane data in the Atlantic Basin.

Keywords: Conditional intensity function, Cubic B-spline, Functional marked Hawkes process, Hurricane data.

1 Introduction

Hawkes processes [6] model event occurrences in dynamic systems where past events influence future ones. Initially used to analyse seismic sequences [10], they have since been extended to finance, social dynamics, and weather [9]. In the context of natural hazards, self-exciting processes can describe extreme weather events, where the occurrence of an event alters the probability of occurrence for the future events in nearby spatio-temporal regions [9]. Marked spatio-temporal point processes further extend this framework by incorporating additional event attributes, such as magnitude or other relevant characteristics. Traditional approaches assume that the occurrence rate depends on exogenous covariates and scalar marks [1], yet the relationship between self-excitation and the functional marks which evolve over time like stochastic processes, remains largely unexplored. Namely, when the mark represents a dynamic quantity, its effect on the process introduces additional complexity, influencing both event intensity and clustering in ways that standard models fail to capture. Building on [4], which introduces a theoretical framework for spatio-temporal functional marked point processes (STFMPP) and proposes initial developments in inferential analysis and modelling for specific generative mechanisms, this work proposes a theoretical framework for spatio-temporal functional marked Hawkes point processes (STFMHPP), expanding their applications in weather science and other areas.

2 Methodology

2.1 Spatio-temporal functional marked Hawkes processes

Let $\mathbf{\Psi}_G = \{(\mathbf{u}_i, t_i)\}_{i=1}^N \subseteq \mathbb{R}^d \times \mathbb{R}$, $N \in \mathbb{N}_0 = \{1, \ldots, \infty\}$, be a ground process, defined as a random collection of points in some space. Suppose each point in $\mathbf{\Psi}_G$ is associated with a K-dimensional stochastic process, $F_i(\tau) = \{F_{i_k}(\tau)\}_{k=1}^K$, $\tau \in \mathcal{T} \subseteq [0, \infty)$, $K \geq 1$, referred to as a *functional mark*. We call the derived process $\mathbf{\Psi} = \{(\mathbf{u}_i, t_i), F_i\}_{i=1}^N$ a spatio-temporal functional marked point process (STFMPP) [4]. The underlying mark space, $\mathcal{F}$, is defined as the Polish function space, $\mathcal{U}^k$, generally constructed over the Euclidean ones [11]. An element $F \in \mathcal{F}$ consists of a collection of functions $\{F_i\}_{i=1}^k : \mathcal{T} \to \mathbb{R}$ [4].

A STFMPP [4] is uniquely characterized by its conditional intensity function (CIF), given by the product of the ground CIF, $\lambda_G(\mathbf{u}, t, \tau; \boldsymbol{\theta}|\mathcal{H}_{t_-})$, which corresponds to $\mathbf{\Psi}_G$, and the functional mark density, $f(F(\tau)|\mathcal{H}_{t_-})$, which depends on the functional mark F. The CIF depends on the past history of the process, $\mathcal{H}_{t_-}$, including all previous spatio-temporal locations and their associated functional marks, up to time t, but not including t.

Definition 1. *Let $\mathbf{\Psi} = \{((\mathbf{u}_i, t_i), F_i)\}_{i=1}^N$, $N \in \mathbb{N}_0$, be a* STFMPP *with a functional mark F_i assigned to each point $(t_i, \mathbf{u}_i)$. The process $\mathbf{\Psi}$ is called a* STFMHPP *if its CIF is given by*

$$\lambda(\mathbf{u}, t, F(\tau)|\mathcal{H}_{t_-}) = \left(\mu(\mathbf{u}, t) + \sum_{j: t_j < t} \phi\big(\mathbf{u} - \mathbf{u}_j, t - t_j, F_j(\tau)\big) \right) f\big(F(\tau)|\mathcal{H}_{t_-}\big) \tag{1}$$

where: $\mu(\mathbf{u}, t)$ is the baseline intensity; $\phi\big(\mathbf{u} - \mathbf{u}_j, t - t_j, F_j(\tau)\big)$ is the excitation function capturing self-excitation due to past events which depends on the functional mark; $f\big(F(\tau)|\mathcal{H}_{t_-}\big)$, a function modulating the intensity based on the functional mark $F(\tau)$ associated with the event at the spatio-temporal location $(\mathbf{u}, t)$.

The baseline intensity, also called *background intensity function* [10][1], $\mu(\mathbf{u}, t)$, can be specified as homogeneous or inhomogeneous in space and time, depending on the studied phenomenon [2]. In our application, we consider an inhomogeneous background intensity function. We assume that the excitation function can be factorized as $\phi\big(\mathbf{u} - \mathbf{u}_j, t - t_j, F_j(\tau)\big) = f(\mathbf{u} - \mathbf{u}_j) g(t - t_j) h\big(F_j(\tau)\big)$ [1]. Additionally, we define the spatial and temporal triggering kernels as $f\big(\mathbf{u} - \mathbf{u}_j\big) = (\|\mathbf{u} - \mathbf{u}_j\|^2 + d)^{-q}$ - with $\|\cdot\|$, the Euclidean distance - and $g(t - t_j) = \kappa_0(t - t_j + c)^{-p}$, respectively. In the latter, $\{d, c\}$ represent displacement parameters, $\{q, p\}$ are decaying parameters and κ_0 is the productivity parameter [10][1]. The functional mark, $F(\tau)$, is often projected onto a finite-dimensional space for practical application. For smoothing purposes [5], we consider projecting it onto a set of basis functions, expressed as $\tilde{F}(\tau) = \boldsymbol{\varphi}(\tau)\boldsymbol{\gamma} = \sum_{i=1}^I \gamma_i \varphi_i(\tau)$, where $\boldsymbol{\varphi}(\tau)$ represents a set of bases such as cubic spline, and $\boldsymbol{\gamma}$ denotes coefficients [11]. We choose an exponential function to ensure non-negativity and obtain $h\big(\tilde{F}(\tau)\big) = \exp\{\boldsymbol{\varphi}(\tau)\boldsymbol{\gamma}\}$.

In this work, model fitting focuses on the ground CIF, while the estimation of the functional mark density is left for future investigation.

2.2 Parameter estimation

Given a realization of the STFMPP $\boldsymbol{\Psi}$ within a bounded spatio-temporal window $W \times [0,T) \subset \mathbb{R}^d \times \mathbb{R}^+$ with marks $F \in \mathcal{F}$, we assume that $T > 0$ is fixed and no event has occurred before time 0. To estimate the parameters of the ground CIF, denoted by $\boldsymbol{\theta}$, assuming STFMHPP generating mechanism, we maximize the log-likelihood function:

$$\ell(\boldsymbol{\theta}; \boldsymbol{\Psi}_G) = \sum_{(\mathbf{u}_i, t_i, F_i) \in \boldsymbol{\Psi}} \left\{ \sum_{k=2}^{K} \log[\lambda_G(\mathbf{u}_i, t_i, \tau_k; \boldsymbol{\theta} | \mathcal{H}_{t_-})] - \int_{\tau_{k-1}}^{\tau_k} \int_0^T \int_W \lambda_G(\mathbf{u}, t, \zeta; \boldsymbol{\theta} | \mathcal{H}_{t_-}) d\mathbf{u} dt d\zeta \right\} \quad (2)$$

As in [3], the CIF is evaluated along the functional domain τ. Following [7], the estimation procedure is made up in two steps. First, the regression coefficients in the basis expansion of $F(\cdot)$, $\boldsymbol{\gamma}$, are estimated by fitting a linear model to the smoothed functional mark of each event. Then, following the idea in [3], the parameter values governing the Hawkes process with power-law decays, $\boldsymbol{\eta} = \{\mu, \kappa_0, d, q, c, p\}$, are obtained by maximizing the log-likelihood function. The second estimation step consists of maximizing the likelihood function by integrating the functional mark along the lifetime points, $\{\tau_k\}_{k=1}^K$, where the functional CIF incorporates the previously estimated cubic B-spline coefficients. The latter are substituted within the definition of the functional mark excitation function, $h\big(\tilde{F}(\tau)\big)$, with $h(\cdot) = \exp\{\cdot\}$.

3 Application to hurricanes data

We analyse hurricanes in the Atlantic Basin using the *HURDAT2* database [8], also known as *Best Track Data*. The dataset provides official records of tropical cyclones, including spatial and temporal locations, wind speed, and central pressure. Data are recorded every 6 hours during synoptic times, with updates including additional non-synoptic observations [8]. The dataset contains $1,972$ hurricane events from 1851 to 2023, with lifetimes ranging from 6 hours to 32.75 days. Wind speed values from 1949–1969 are approximated due to limited satellite data [8]. To ensure reliability, we select 391 hurricanes occurred between May 17, 1970, and October 10, 2023. The lifetime of selected hurricanes is limited to the first $K = 20$ records, and the analysis is conducted within a common lifetime interval of $\tau \in [0, 4.75]$ days.

Initially, we explore a larger number of events, consisting of hurricanes of at least 15 lifetime records. A starting analysis on this subset unveils the aggregated nature of the data throughout the spatio-temporal domain, which motivates the proposed STFMHPP model fitting to the data. Furthermore, we conduct simulation studies, showing that stable and reliable results are achieved when functional marks are observed in at least 20 lifetime records. This supports the decision to restrict the analysis to a common functional domain consisting of $K = 20$ records. Furthermore, spatial and temporal coordinates are respectively converted into kilometers and days from May 17, 1970. The functional mark, $F(\tau)$, is the maximum wind speed recorded every 6 hours, in miles per minute (mi/m) converted from knots, ranging from 1.111 to 17.779 mi/m.

Figure 1 shows the functional marked spatio-temporal distribution of the selected 391 hurricanes. Line color represents the occurrence time, t, whilst line shape and thickness reflect the maximum wind speed values, $F(\tau)$, over the lifetime. Namely, trajectories start at squares, marking the spatial locations of occurrence, and extend rightward, with positive and negative slopes showing increases and decreases, respectively, in the recorded maximum wind speed. To allow comparability, trajectories are displayed over the common functional domain, $\tau \in [0, 4.75]$ days.

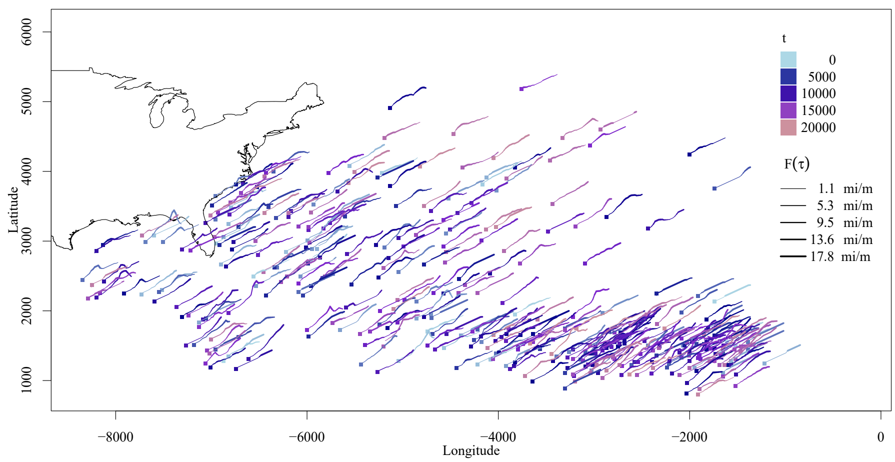

Fig. 1: Functional marked spatio-temporal distribution of the 391 selected hurricanes, occurred between May 17, 1970, and October 10, 2023, in the Atlantic Basin. The closer to pink the curve color, the more recent the hurricane; the more positive the slope of trajectory and the thicker the trajectory, $F(\tau)$, the stronger the wind. Trajectories start at squares, marking the spatial locations of occurrence, and extend rightward as the lifetime increases over the common functional domain $\tau \in [0, 4.75]$ days.

As shown in Figure 1, hurricanes tend to cluster in the Southern-East Atlantic Basin, most occurring after day 5000. Additionally, higher wind speeds are recorded later in their lifetime, highlighting a greater self-exciting effect in the later stages, namely after three days from their occurrence. A preliminary analysis reveals seasonal patterns in daily, weekly, and monthly occurrences, aligning with the *hurricane season*. For the spatial component, a slight departure from the assumption of independence and complete spatial randomness under a spatial Poisson process is observed. However, we deem negligible such spatial inhomogeneity, for the purpose of defining the background intensity function. To account for the temporal seasonality, a Fourier-based seasonal component, is incorporated into the inhomogeneous temporal background intensity function, $\mu(t)$, as follows:

$$\mu(t) = \exp\left\{\beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{M}\right) + \beta_3 \cos\left(\frac{2\pi t}{M}\right)\right\} \tag{3}$$

with t, the occurrence day; M, the seasonal period. Here, the parameter β_0 captures the spatial background effect, taking fixed and equal to $t = 0$ the temporal coordinate; the parameters β_1, β_2 and β_3 catch the temporal and seasonal variations. The baseline spatial background effect will therefore be given by $\exp\{\beta_0\}$. In our application, model fitting focuses on estimating the parameters set $\boldsymbol{\theta} = \{\boldsymbol{\gamma}, \boldsymbol{\eta}\}$, where: $\boldsymbol{\eta} = \{\boldsymbol{\beta}, \kappa_0, d, q, c, p\}$.

Table 1 shows the STFMHPP model fitting results in terms of parameters estimates during the second step, $\hat{\boldsymbol{\eta}}$, standard errors, $\hat{\sigma}_{\hat{\boldsymbol{\eta}}}$, and corresponding p-values, p, for data in Figure 1 using the two-step estimation method described in SubSection 2.2. The functional marks are fitted using cubic B-splines of order $d = 4$, with $I = 5$ knots.

Table 1: Results of the STFMHPP model fitting to the hurricanes data.

η	β_0	β_1	β_2	β_3	κ_0	d	q	c	p
$\hat{\boldsymbol{\eta}}$	-0.2263	-0.0009	-0.0016	-0.0007	0.0420	203.2276	1.1855	34.0905	1.1727
$\hat{\sigma}_{\hat{\boldsymbol{\eta}}}$	0.0001	$2 \cdot 10^{-5}$	0.0001	$2 \cdot 10^{-5}$	0.0002	0.1467	0.0009	0.1212	0.0016
p	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

The results indicate a low degree of self-excitation ($\hat{\kappa}_0 = 0.042$), paired with a strong presence of background events ($\exp\{\hat{\beta}_0\} = 0.7975$), if marginalizing over the temporal component. The background intensity varies by day of the year, and decreases over time ($\hat{\beta}_1 = -0.0009$). Seasonal effects indicate an average estimated temporal intensity of ≈ 0.0018 hurricanes, occurring on the ≈ 232.485 day of the year, that is, the 20 of August, aligning with the *hurricanes season* and the preliminary exploratory analysis. The estimated displacement from previous events is approximately $\hat{d} \approx 203$ kms in space and $\hat{c} \approx 34$ days in time. The short-range influence is stronger in time than in space, as indicated by the relationship $\hat{p} = 1.1727 \leq 1.1855 = \hat{q}$. As for the regression coefficients of the cubic B-spline bases, $\hat{\boldsymbol{\gamma}}$, ranging between -4.8345 and 22.7261, the majority are positive, while a small subset ($\approx 0.004\%$) takes negative values.

4 Discussion and conclusion

We defined STFMHPPs for modelling spatio-temporal events with stochastic process-driven marks. Model-based inference was performed using a two-step procedure and applied to hurricane data. The model was fitted to 391 hurricane events from the Atlantic Basin, using maximum wind speed as the functional mark.

The latter was smoothed using cubic B-spline bases, and temporal seasonality was captured through a Fourier-based representation. The results highlighted a low degree of self-excitation and a strong presence of background events, with spatial and temporal displacements of approximately 203 km and 34 days, respectively. Finally, short-range influence was seen stronger in time than in space.

Acknowledgements

Giada Lo Galbo is grateful for the financial support for her doctoral studies provided by the "Programma Operativo della Regione Siciliana, PR F.S.E. + Sicilia 2021/2027". CUP G71I22001190006. This project has been supported by the national grant of the Italian Ministry of University and Research MUR-PRIN 2022: Spatio-temporal Functional Marked Point Processes for probabilistic forecasting of earthquakes 2022BN7CJP. P. I. Giada Adelfio. CUP B53C24006340006. Mohammad Ghorbani was financially supported by the Kempe Foundations (grant number JCSMK22-0134) and the Olle Engkvists Foundation (grant number 224-0071). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

1. Adelfio, G., Chiodi, M.: Including covariates in a space-time point process with application to seismicity. Stat. Methods. Appl. 30, 947–971 (2021). doi:10.1007/s10260-020-00543-5
2. Bernabeu, A., Zhuang, J., Mateu, J.: Spatio-Temporal Hawkes Point Processes: A Review. J. Agric. Biol. Environ. Stat. 30(1), 89–119 (2024). doi:10.1007/s13253-024-00653-7
3. Fok, C.C.T., Ramsay, J.O., et Al.: A functional marked point process model for lupus data. Can. J. Stat. 40, 517–529 (2012). doi:10.1002/cjs.11136
4. Ghorbani, M., Cronie, O., et Al.: Functional marked point processes: a natural structure to unify spatio-temporal frameworks and to analyse dependent functional data. TEST 30, 529–568 (2021). doi:10.1007/s11749-020-00730-2
5. Giraldo, R., Mateu, J., Delicado, P.: `geofd`: An R Package for Function-Valued Geostatistical Prediction, Rev. Colomb. Estad. 35(**3**), 385-407 (2012).
6. Hawkes, A.G.: Spectra of some self-exciting and mutually exciting point processes. Biometrika 58, 83-90 (1971). doi:10.1016/0022-2836(81)90087-5
7. Kattamanchi, S., Tiwari, R.K., et Al.: Non-stationary ETAS to model earthquake occurrences affected by episodic aseismic transients, EPS 69 (2017).
8. Landsea, C.W., Franklin, J.L.: Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format, MWR 141: 3576–3592 (2013). doi:10.1175/MWR-D-12-00254.1
9. Lo Galbo, G., Chiodi, M.: Severe convective storms' reproduction: empirical analysis from the marked self-exciting point processes point of view. Environ. Ecol. Stat. 31, 409–431 (2024). doi:10.1007/s10651-023-00593-4
10. Ogata, Y.: Space-Time Point-Process Models for Earthquake Occurrences. AIMS 50, 379–402 (1998). doi:10.1023/A:1003403601725
11. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis. Springer, New York (2005). doi:10.1002/0470013192.bsa239