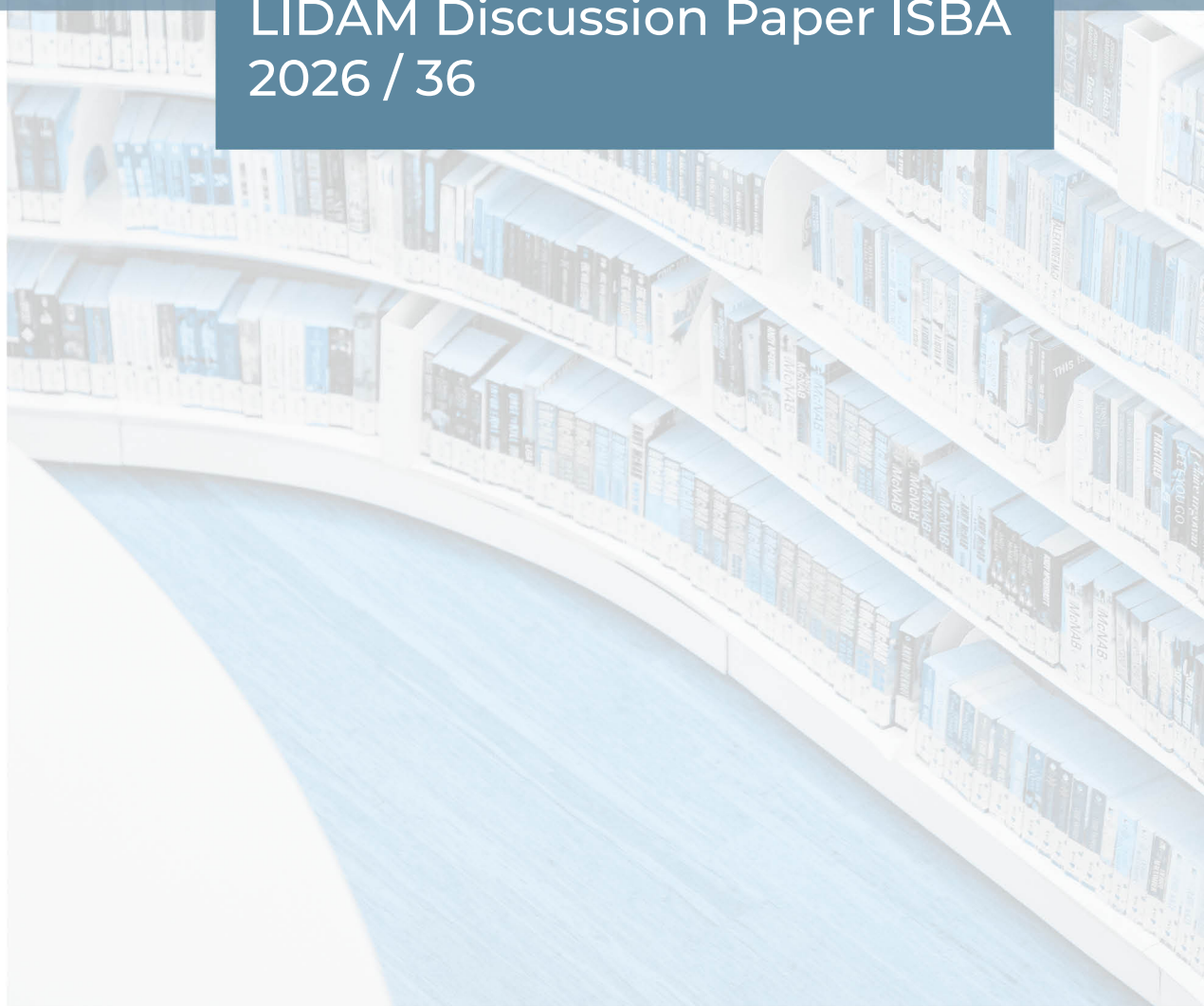
A SPATIAL MARKOV-MODULATED MARKED HAWKES PROCESS FOR PARAMETRIC INSURANCE

Donatien Hainaut

LIDAM Discussion Paper ISBA
2026 / 36



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

A spatial Markov-modulated marked Hawkes process for parametric insurance

Donatien Hainaut*
UCLouvain - LIDAM

October 20, 2026

Abstract This paper introduces a tractable Markov-modulated marked Hawkes process for modelling spatial clustered events and pricing parametric insurance. The model combines spatial self-excitation, event-specific marks and a latent continuous-time Markov chain whose state, updated at event times, governs the intensity, triggering mechanism and mark distribution. This specification preserves regime-switching dynamics while simplifying likelihood evaluation. We derive the likelihood of the observed process and develop forward-backward filtering and an expectation maximization (EM) algorithm for estimating the model when the regimes are unobserved. We also establish a sufficient stability condition and provide a simulation procedure for generating event times, locations, marks and latent regimes. The framework is applied to earthquake data from Greece using a switching Epidemic Type Aftershock Sequence (ETAS) model with Pareto-distributed magnitudes. The empirical analysis shows that the model captures temporal clustering and alternating periods of seismic activity, while reproducing the central tendency of earthquake magnitudes. Finally, we apply the model to price a regional parametric earthquake insurance contract covering an elliptical area around Athens. These results illustrate how latent regime switching, spatial dependence and event severity can be jointly incorporated into catastrophe risk modelling and parametric insurance pricing.

Keywords: Hawkes process, Markov-modulated process, spatial process, parametric insurance

1 Introduction

Hawkes processes are self-exciting point processes widely used to model events that occur in clusters, whereby the occurrence of an event temporarily increases the probability of subsequent events. Originally introduced by Hawkes (1971 a,b), these processes have become important tools for describing phenomena in which events are not independent over time. Their applications include financial modelling (Hainaut 2022 chapters 4 and 5), crime (Mohler et al. 2011, Zhu and Xie 2022), insurance (Jang and Dassios 2013, Hainaut 2016, 2017, Swishchuk et al. 2021), terrorism (Jun and Cook 2024) or social media (Yuan et al. 2019). In earthquake modelling, the Epidemic-Type Aftershock Sequence (ETAS) model developed by Ogata (1988) represents a particularly influential application, as it captures the triggering relationship between earthquakes and their aftershocks. However, earthquake occurrence depends not only on temporal

*Corresponding author. Postal address: Voie du Roman Pays 20, 1348 Louvain-la-Neuve Belgium. E-mail to: donatien.hainaut(at)uclouvain.be

clustering but also on the spatial distribution of seismic events and their magnitudes. Spatial and marked extensions of Hawkes processes (Ogata 1998, Zhuang et al. 2002) therefore provide a natural framework for representing the joint evolution of event times, locations and severities. By incorporating spatial triggering mechanisms, such models can describe how seismic activity propagates geographically and how the resulting risk varies across a region, making them particularly relevant for catastrophe risk assessment.

Markov-modulated models provide another framework for modelling insurance risk by allowing claim frequencies, claim severities or an insurer's surplus dynamics to depend on an underlying stochastic environment represented by a finite-state continuous-time Markov chain. This approach captures changes between regimes with different risk characteristics, such as periods of low and high claim activity, and can therefore provide a more realistic representation of insurance losses than models with constant parameters. In actuarial research, Markov-modulated risk processes have been studied in connection with ruin probabilities and solvency (Asmussen 1989, Ng and Yang 2006), while related work has considered hidden environmental regimes in insurance claims (Elliott et al. 2009, Shi et al. 2025), life insurance (Hainaut 2012) and regime-dependent reinsurance and dividend strategies (Wei et al. 2010). These studies motivate the use of regime-switching structures in insurance applications, particularly when the underlying risk environment is not directly observable and must be inferred from event data. Markov-modulated diffusions can also approximate a Hawkes process (Hainaut and Deelstra 2019). Introducing regime switching into a Hawkes process can complicate likelihood-based inference and the derivation of analytical properties, particularly when the latent regime affects the process dynamics. Nevertheless, Bonnet and Robin (2025) propose a discrete-time, non-spatial Markov-switching Hawkes process with an exponential kernel and develop an expectation-maximization (EM) algorithm for parameter estimation. Hainaut and Moraux (2019), in turn, investigate the properties of a Hawkes process with a Markov-modulated baseline intensity.

Beyond the classical ETAS framework, recent research has explored more flexible spatio-temporal Hawkes models to account for spatial heterogeneity, anisotropy and interactions between event times and locations (Kwon et al. 2023) or non-parametric frameworks (Fox et al. 2016). Spatial modelling has also been considered directly in the context of earthquake parametric insurance, for instance through Bayesian spatial quantile regression approaches (Pai et al. 2022). These developments highlight the importance of combining spatial dependence with a realistic representation of event severity when assessing catastrophe risk. Nevertheless, representing changes in the underlying seismic regime while maintaining a tractable estimation procedure remains a relevant modelling challenge. The approach developed in this paper addresses this challenge by introducing a latent regime-switching mechanism into a marked spatial Hawkes process and linking its estimation to the pricing of a geographically defined insurance contract.

A related challenge is to translate catastrophe risk models into reliable pricing tools for parametric insurance, where payouts are determined by predefined indices or event characteristics rather than by the policyholder's actual losses (see, e.g., Biffis et al. 2022; Louass and Picard 2026). In earthquake insurance, for example, a contract may pay a predetermined amount when an earthquake's magnitude (Gu et al. 2023, Bayliss et al. 2020), location or another measurable seismic characteristic (Radu and Alexandru 2022) satisfies specified conditions. Such contracts can facilitate rapid claims settlement and reduce loss-adjustment costs, although their effectiveness depends on the relationship between the trigger and the underlying economic losses, commonly referred to as basis risk (Clement et al. 2018). Pricing these products requires models that accurately represent both the frequency and severity of triggering events, as well as their spatial distribution relative to the insured area. This spatial dimension is particularly important when coverage is restricted to a specific geographical region, since earthquakes occurring at dif-

ferent distances from the insured territory can have substantially different implications for the probability of a payout and the aggregate cost of claims. Spatial seismicity models, including ETAS-type approaches, consequently offer a useful foundation for quantifying the expected cost and tail risk of parametric earthquake insurance (see e.g. Cummins and Mahul 2009).

This paper introduces a tractable Markov-modulated marked spatial Hawkes process designed to capture changes in seismic activity and support the pricing of spatially defined parametric insurance contracts. The model combines spatial self-excitation and event-specific marks with a latent continuous-time Markov chain whose regime governs the background intensity, triggering mechanism and mark distribution. The latent regime evolves at event times, allowing the process to alternate between distinct levels and patterns of seismic activity while retaining a structure that facilitates statistical inference. We derive the likelihood of the observed process and develop a forward-backward filtering procedure and an expectation-maximization (EM) algorithm to estimate the model when the regimes are unobserved. We also establish a sufficient stability condition and provide an algorithm for simulating event times, locations, marks and latent regimes. The framework is applied to Greek earthquake data through a switching ETAS specification with Pareto-distributed magnitudes. Finally, we use the fitted models to price a parametric earthquake insurance contract covering an elliptical region around Athens, comparing the resulting pure and risk-loaded premiums across models with different numbers of regimes. The main contribution is to combine latent regime switching, spatial self-excitation and marked event severity within a tractable framework that connects statistical modelling of seismicity with the assessment of insurance risk.

The remainder of the paper is organized as follows. Section 2 introduces the Markov-modulated marked spatial Hawkes process and presents its main properties. Sections 3 to 6 derive the likelihood, describe the filtering and the EM procedure used for parameter estimation. Section 7 presents simulation algorithm. Section 8 discusses the framework for parametric insurance with the Markov-modulated marked spatial Hawkes process. The next section focuses on the switching ETAS model. Section 10 presents the application to Greek earthquake data and the pricing of the parametric insurance contract covering the Athens region.

2 A modulated spatial marked point process

We consider a spatial self-exciting counting process $(N(t))_{t\geq 0}$ with event times $(T_k)_{k\in\mathbb{N}} \in \mathbb{R}^+$, locations $(\boldsymbol{X}_k)_{k\in\mathbb{N}} \in \mathcal{S}$ and random marks $(M_k)_{k\in\mathbb{N}} \in \mathcal{M}$. In applications, the spatial domain $\mathcal{S}$ is a subset of $\mathbb{R}^2$ and $\boldsymbol{X}_k$ contains the geographical coordinates of the k^{th} events. The probability space and the measure of probability are $(\Omega, \mathbb{P})$ whereas the filtration of events is denoted by

$$\mathcal{G}_t := \sigma\left\{(\boldsymbol{X}_k, T_k, M_k) \,:\, T_k \leq t\right\} .$$

Let $(\boldsymbol{Z}_t)_{t\geq 0}$ be a continuous-time Markov chain with d states. $\boldsymbol{Z}_t$ takes its value in $\{\boldsymbol{e}_1, ..., \boldsymbol{e}_d\}$, the set of canonical basis vectors of $\mathbb{R}^d$. The filtration of $\boldsymbol{Z}_t$ is denoted by $(\mathcal{H}_t)_{t\geq 0}$. The initial distribution of $\boldsymbol{Z}_0$ is $\boldsymbol{\pi}_0 = (\pi_0^{(1)}, ..., \pi_0^{(d)})^\top$, such that $\sum_{i=1}^d \pi_0^{(i)} = 1$ and $\pi_0^{(i)} \geq 0$. The $d \times d$ matrix of instantaneous transition probabilities of $\boldsymbol{Z}_t$ is denoted by $A = (a_{i,j})_{i,j=1,...,d}$. The a_{ij}'s fulfill the following constraints

$$\begin{cases} a_{ij} & \geq 0 \quad i \neq j\,, \\ a_{ii} & = -\sum_{i\neq j} a_{ij}\,. \end{cases}$$

The matrix A defines the expectation of the chain and the probabilities of transition between states:

$$\begin{aligned} \mathbb{E}\left(\boldsymbol{Z}_s \,|\, \boldsymbol{Z}_t = \boldsymbol{e}_i\right) &= \boldsymbol{e}_i^\top \exp\left(A(s-t)\right)\,, \\ \mathbb{P}\left(\boldsymbol{Z}_s = \boldsymbol{e}_j \,|\, \boldsymbol{Z}_t = \boldsymbol{e}_i\right) &= \boldsymbol{e}_i^\top \exp\left(A(s-t)\right)\boldsymbol{e}_j\,; \end{aligned}$$

We assume that the chain admits a stationary state with the following vector of probabilities:

$$\boldsymbol{\pi} := \lim_{u \to \infty} \boldsymbol{\pi}_0^\top \exp(Au) \ .$$

In order to define the dynamics of $N(t)$, we introduce the following notation:

$$j(t) := \max\{k \ : \ T_k < t\} \, .$$

This is the index of the most recent event occurring before t, with the convention $j(t) = 0$ if no event has occurred before t. The state governing the intensity between two successive arrivals is denoted by $\boldsymbol{Z}_{j(t)} := \boldsymbol{Z}_{T_{j(t)}}$. To simplify the notation, we also write $\boldsymbol{Z}_i := \boldsymbol{Z}_{T_i}$. The Markov chain $\boldsymbol{Z}_t$ modulates the parameters of the spatial self-exciting process. The modulation occurs exclusively at event times.

The intensity of the counting process $N(t)$, involves a d-vector of background or baseline intensities, $\boldsymbol{\mu}(\boldsymbol{x}) := \left(\mu^{(1)}(\boldsymbol{x}), ..., \mu^{(d)}(\boldsymbol{x})\right)^\top$, and

$$\boldsymbol{\phi}(\boldsymbol{u}, s, m) := \begin{pmatrix} \phi^{(1)}(\boldsymbol{u}, s, m) \\ \vdots \\ \phi^{(d)}(\boldsymbol{u}, s, m) \end{pmatrix} ,$$

a $d-$vector of positive kernel functions. The functions $\left(\phi^{(i)}(\boldsymbol{u}, s, m)\right)_{i=1,...,d}$ decreases as $||\boldsymbol{u}||_2$ or s increases. The variable $m \in \mathcal{M}$ allows the triggering effect to depend on the characteristics of the previous event. A sufficient condition that ensures the stability of the process is provided at the end of this section. The intensity of the counting process is:

$$\Lambda(t) = \int_{\mathcal{S}} \lambda(\boldsymbol{x}, t) \, d\boldsymbol{x} \, ,$$

where $\lambda(\boldsymbol{x}, t)$ is modulated by the Markov chain as follows:

$$\lambda(\boldsymbol{x}, t) = \left(\boldsymbol{\mu}(\boldsymbol{x}) + \sum_{T_k < t} \boldsymbol{\phi}\left(\boldsymbol{x} - \boldsymbol{X}_k, t - T_k, M_k\right)^\top \right) \boldsymbol{Z}_{j(t)}.$$

The random marks $(M_k)_{k \geq 0}$ are also modulated by $\boldsymbol{Z}_{j(t)}$. Conditionally to $\boldsymbol{Z}_{k-1}$, the pdf of M_k is $f_M\left(m \,|\, \boldsymbol{Z}_{k-1}\right)$. The intensity is adapted to the augmented filtration $\mathcal{F}_t = \mathcal{H}_t \vee \mathcal{G}_{j(t)}$.

This specification freezes the regime between arrivals. It differs from existing models, in which the Markov chain continuously modulate the background intensity and the kernel function. In our framework, a jump of the hidden Markov chain between arrivals has no impact until the next event. Our model has a greater analytical tractability for this reason. Since $\boldsymbol{Z}_{j(t)} \in \{\boldsymbol{e}_1, ..., \boldsymbol{e}_d\}$, this simply selects one regime and we can rewrite the intensity as:

$$\lambda(\boldsymbol{x}, t) = \mu^{(s(t))}(\boldsymbol{x}) + \sum_{T_k < t} \phi^{(s(t))}\left(\boldsymbol{x} - \boldsymbol{X}_k, t - T_k, M_k\right) \, ,$$

where $s(t) = (1, ..., d) \times \boldsymbol{Z}_{j(t)}$, is the active Markov state at time $T_{j(t)}$. Conditionally to the sample path of $\boldsymbol{Z}_t$, the counting process $N(t)$ may be seen as a sum of non-homogeneous Poisson processes:

- $N_b(t)$ with rate $\int_{\mathcal{S}} \boldsymbol{\mu}(\boldsymbol{x})^\top \boldsymbol{Z}_{j(t)} \, d\boldsymbol{x}$, the background process.
- $N_1(t)$, starting at time T_1, with intensity $\int_{\mathcal{S}} \boldsymbol{\phi}\left(\boldsymbol{x} - \boldsymbol{X}_1, t - T_1, M_1\right)^\top \boldsymbol{Z}_{j(t)} d\boldsymbol{x}$, the offspring of the first event.

- $N_k(t)$, starting at time T_k, with intensity $\int_{\mathcal{S}} \boldsymbol{\phi}\left(\boldsymbol{x}-\boldsymbol{X}_k, t-T_k, M_k\right)^\top \boldsymbol{Z}_{j(t)} d\boldsymbol{x}$, the offspring of the k^{th} event, for $T_k < t$.

The expected number of offspring of the k^{th} event, conditionally to $\{\boldsymbol{X}_k, T_k\}$ and to the sample path of $\boldsymbol{Z}_t$ is equal to

$$\lim_{t\to\infty} \mathbb{E}\left(N_k(t) \,|\, \{\boldsymbol{X}_k, T_k\} \vee \mathcal{H}_\infty\right) = \int_{\mathcal{S}} \int_{T_k}^{\infty} \int_{\mathcal{M}} \boldsymbol{\phi}\left(\boldsymbol{x}-\boldsymbol{X}_k, s-T_k, m\right)^\top \boldsymbol{Z}_{j(s)}\, f_M\left(m|\boldsymbol{Z}_k\right)\, dm\, ds\, d\boldsymbol{x}\,. \tag{1}$$

Let us denote by $||\boldsymbol{\phi}||_1 = \left(||\phi^{(1)}||_1, \ldots, ||\phi^{(d)}||_1\right)^\top$ the $d-$vector of L_1-norms of kernels,

$$||\phi^{(j)}||_1 := \int_{\mathcal{S}} \int_0^{\infty} \int_{\mathcal{M}} \phi^{(j)}(\boldsymbol{x}, u, m) f_M\left(m|\boldsymbol{e}_j\right)\, dm\, du\, d\boldsymbol{x}\,.$$

As $\boldsymbol{Z}_t$ is a unit basis vector, we infer from Equation (1) that the expected number of offspring of an event is then bounded by the largest L_1-norm of kernels:

$$\lim_{t\to\infty} \mathbb{E}\left(N_k(t)\right) \leq \max_{i\in(1,\ldots d)} \left(||\phi^{(i)}||_1\right)\,.$$

A sufficient condition of stability is therefore

$$||\phi_{\max}||_1 := \max\left(||\boldsymbol{\phi}||_1\right) < 1\,. \tag{2}$$

The next proposition is the direct analogue of the classical Hawkes bound on the stationary event rate.

Proposition 1. *If condition (2) is fulfilled, the stationary event rate, $\bar{\boldsymbol{\lambda}}$, satisfies the following inequality:*

$$\bar{\boldsymbol{\lambda}} \leq \frac{\mu_{max}}{1 - ||\phi_{\max}||_1}\,. \tag{3}$$

Proof. Each event in the (background) 0^{th} generation acts as a parent of a 1^{st}generation of offspring events, N_1, that generate on average at most $||\phi_{\max}||_1$ offspring. Each event in the 1^{st} generation acts as a parent of a 2^{nd} generation of offspring events, N_2 :

$$\begin{aligned}\mathbb{E}\left(N_2\right) &= \mathbb{E}\left(\mathbb{E}\left(N_2|N_1\right)\right)\\ &\leq \mathbb{E}\left(||\phi_{\max}||_1 N_1\right) \leq \left(||\phi_{\max}||_1\right)^2\,.\end{aligned}$$

Each event in the $(k-1)^{th}$ generation acts as a parent of a k^{th} generation of offspring events, N_k. The expectation of N_k is then bounded by $\mathbb{E}\left(N_k\right) \leq \left(||\phi_{\max}||_1\right)^k$. The total expected number of offspring events for a single background event is equal to:

$$\mathbb{E}\left(\sum_{k=1}^{\infty} N_k\right) \leq \sum_{k=1}^{\infty} \left(||\phi_{\max}||_1\right)^k = \frac{||\phi_{\max}||_1}{1 - ||\phi_{\max}||_1}\,.$$

Let us define $\mu_{\max} = \max_{i\in(1,\ldots,d)} \int_{\mathcal{S}} \mu^{(i)}(\boldsymbol{x})\, d\boldsymbol{x}$. Since $\boldsymbol{Z}_{j(t)}$ is a basis vector, the background intensity is bounded by $\mu_{\max}$. Each immigrant generates a cluster whose expected size is at most

$$1 + \frac{||\phi_{\max}||_1}{1 - ||\phi_{\max}||_1} = \frac{1}{1 - ||\phi_{\max}||_1}\,.$$

Consequently, the stationary event rate satisfies Equation (3). □

3 Likelihood when Z_t is observable

In this section, we present the likelihood of the modulated spatial self-exciting process when the Markov chain is visible at event times. The results are needed to develop the expectation-maximization algorithm and the filtering of $\boldsymbol{Z}_t$ when the Markov chain is hidden. Let us assume that we observe n events on the time interval $[0, T]$:

$$0 \leq T_1 < \ldots < T_n \leq T \, .$$

The corresponding positions and marks are denoted by $\boldsymbol{X}_k$ and M_k. Given T_{k-1} and $\boldsymbol{Z}_{k-1}$, the state remains fixed over $[T_{k-1}, T_k)$. The k^{th} events is a triplet denoted by

$$\boldsymbol{E}_k = (T_k, \boldsymbol{X}_k, M_k) \, .$$

The sample set of the first k events is:

$$\mathcal{E}_k := (\boldsymbol{E}_1, ..., \boldsymbol{E}_k) \, .$$

We denote by $\lambda_k(\boldsymbol{x}, t)$, the intensity function over the interval $[T_{k-1}, T_k)$, that is equal to

$$\begin{aligned} \lambda_k(\boldsymbol{x}, t) &:= \lambda_k \left(\boldsymbol{x}, t \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right) \\ &= \left(\boldsymbol{\mu}(\boldsymbol{x})^\top + \sum_{T_j < t} \boldsymbol{\phi} \left(\boldsymbol{x} - \boldsymbol{X}_j, t - T_j, M_j \right)^\top \right) \boldsymbol{Z}_{k-1} \, , \end{aligned}$$

for $T_{k-1} \leq t \leq T_k$. Conditionally on $\mathcal{E}_{k-1}$ and $\boldsymbol{Z}_{k-1}$, The probability that the k^{th} event occurs after time t is given by

$$\mathbb{P} \left(\boldsymbol{X}_k \in \mathcal{S}, T_k \geq t \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right) = \exp \left(- \int_{\mathcal{S}} \int_{T_{k-1}}^{t} \lambda_k(\boldsymbol{x}, s) \, ds \, d\boldsymbol{x} \right) \, . \tag{4}$$

We denote by $f_k \left(\boldsymbol{x}, t \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right)$ and $F_k \left(t \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right)$, respectively the conditional probability density function (pdf) and cumulative distribution function (cdf) of the arrival time of the k^{th} event. These pdf and cdf are related to Equation (4) as follows:

$$\begin{aligned} F_k \left(t \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right) &= 1 - \mathbb{P} \left(X_k \in \mathcal{S}, T_k \geq t \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right) \\ &= \int_{\mathcal{S}} \int_{T_{k-1}}^{t} f_k \left(\boldsymbol{x}, s \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right) \, ds \, d\boldsymbol{x}. \end{aligned}$$

Deriving this last expression w.r.t. t leads to the analytical expression of the pdf:

$$f_k \left(\boldsymbol{x}, t \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right) = \lambda_k \left(\boldsymbol{x}, t \right) \exp \left(- \int_{\mathcal{S}} \int_{T_{k-1}}^{t} \lambda_k \left(\boldsymbol{x}, s \right) \, ds \, d\boldsymbol{x} \right) \, . \tag{5}$$

Since conditionally to $\boldsymbol{Z}_{k-1}$, marks are assumed independent from event times, the joint likelihood of $E_k = (T_k, \boldsymbol{X}_k, M_k)$ is equal to the product of pdfs :

$$f_k \left(\boldsymbol{x}, t, m \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right) = f_M \left(m \,|\, \boldsymbol{Z}_{k-1} \right) f_k \left(\boldsymbol{x}, t \,|\, \mathcal{E}_{k-1} \,,\, \boldsymbol{Z}_{k-1} \right) \, .$$

Furthermore, the probability of no-event between T_n and T admits an analytical expression:

$$\begin{aligned} S_n^{(i)} &:= \mathbb{P} \left(T_{n+1} \geq T \,|\, \mathcal{E}_n, \, \boldsymbol{Z}_n = \boldsymbol{e}_i \right) \\ &= \exp \left(- \int_{\mathcal{S}} \int_{T_n}^{T} \lambda_{n+1} \left(\boldsymbol{x}, s \,|\, \mathcal{E}_n \,,\, \boldsymbol{e}_i \right) ds \, d\boldsymbol{x} \right) \, . \end{aligned} \tag{6}$$

The vector of these probabilities is $\boldsymbol{S}_n := \left(S_n^{(1)}, \ldots, S_n^{(d)} \right)$. The next proposition develops the likelihood of the model.

Proposition 2. *Suppose the sequence $(\boldsymbol{Z}_i)_{i=0,...,n}$ is observed. Let Θ_N be the parameters of the counting process $N(t)$. The likelihood of the sample $\mathcal{E}_n$ is the following product*

$$\mathcal{L}_N(\Theta_N) = \boldsymbol{S}_n^\top \boldsymbol{Z}_n \prod_{k=0}^{n-1} f_{k+1}\left(\boldsymbol{X}_{k+1}, T_{k+1}, M_{k+1} \,|\, \mathcal{E}_k\,,\, \boldsymbol{Z}_k\right)\,, \tag{7}$$

and the log-likelihood is given by

$$l_N\left(\Theta_N\right) \;=\; \sum_{k=1}^{n} \left(\ln \lambda(\boldsymbol{X}_k, T_{k-}) + \ln f_M\left(M_k|\boldsymbol{Z}_{k-1}\right)\right) - \int_{\mathcal{S}} \int_0^T \lambda(\boldsymbol{x}, t)\, dt\, d\boldsymbol{x}. \tag{8}$$

Proof. Using the Bayes rules, the likelihood $\mathcal{L}_N(\Theta_N) = f\left(\mathcal{E}_n\right)$ of the sample is proportional to the product of densities $f_{k+1}(.)$:

$$\mathcal{L}_N(\Theta_N) \;=\; \mathbb{P}\left(T_{n+1} \geq T \,|\, \mathcal{E}_n, \boldsymbol{Z}_n\right) \prod_{k=0}^{n-1} f_{k+1}\left(\boldsymbol{X}_{k+1}, T_{k+1}, M_{k+1} \,|\, \mathcal{E}_k\,,\, \boldsymbol{Z}_k\right)\,.$$

We replace the pdf $f_{k+1}\left(\boldsymbol{X}_{k+1}, T_{k+1}, M_{k+1} \,|\, \mathcal{T}_k\,,\, \boldsymbol{Z}_k\right)$ by its expression (5) and directly infer that

$$\begin{aligned}\mathcal{L}_N(\Theta_N) \;=\; & \exp\left(-\int_{\mathcal{S}}\int_{T_n}^{T} \lambda_{n+1}(\boldsymbol{x}, s)\, ds\, d\boldsymbol{x}\right) \prod_{k=0}^{n-1} \lambda_{k+1}\left(\boldsymbol{X}_{k+1}, T_{k+1}\right) \\ & \exp\left(-\int_{\mathcal{S}}\int_{T_k}^{T_{k+1}-} \lambda_{k+1}(\boldsymbol{x}, s)\, ds\, d\boldsymbol{x}\right) f_M\left(M_{k+1}|\boldsymbol{Z}_k\right)\,.\end{aligned}$$

The log-likelihood $l_N\left(\Theta_N\right) = \ln \mathcal{L}_N(\Theta_N)$ is therefore equal to the following sum

$$\begin{aligned} l_N\left(\Theta_N\right) \;=\; & \sum_{k=1}^{n} \left(\ln \lambda_k(\boldsymbol{X}_k, T_k) + \ln f_M\left(M_k|\boldsymbol{Z}_{k-1}\right)\right) \\ & - \sum_{k=0}^{n-1} \left(\int_{\mathcal{S}}\int_{T_k}^{T_{k+1}-} \lambda_{k+1}(\boldsymbol{x}, s)\, ds\, d\boldsymbol{x}\right) - \left(\int_{\mathcal{S}}\int_{T_n}^{T} \lambda_{n+1}(\boldsymbol{x}, s)\, ds\, d\boldsymbol{x}\right)\,.\end{aligned}$$

As $\lambda_k(\boldsymbol{X}_k, T_k) = \lambda(\boldsymbol{X}_k, T_k-)$, we conclude that the log-likelihood is well given by Equation (8) . □

We denote by $\Delta_{k+1} := T_{k+1} - T_k$, the time interval between the k^{th} and $(k+1)^{th}$ events. We define $p_{i,j}(\Delta_{k+1})$, as the probability of transition of the Markov chain over this time interval:

$$\begin{aligned} p_{i,j}(\Delta_{k+1}) &:= \mathbb{P}\left(Z_{k+1} = \boldsymbol{e}_j \,|\, Z_k = \boldsymbol{e}_i\right) \\ &= \left[e^{A\Delta_{k+1}}\right]_{i,j}\,.\end{aligned} \tag{9}$$

The matrix of probabilities of transition is $P^{(k+1)} := e^{A\Delta_{k+1}}$. The Markov chain is visible at times $\mathcal{T}_n = \{0, T_1, ..., T_n\}$ and the states at these times are denoted by $s_k := s(T_k)$ for $k = 0, ..., n$. If $\Theta_{mc} = (A, \boldsymbol{\pi}_0)$ is the set of parameters defining $\boldsymbol{Z}_t$, the likelihood and log-likelihood of the Markov chain are respectively given by:

$$\begin{aligned}\mathcal{L}_{mc}(\Theta_{mc}) &= \boldsymbol{\pi}_0^\top Z_0 \prod_{k=0}^{n-1} p_{s_k, s_{k+1}}(\Delta_{k+1})\,, \\ l_{mc}(\Theta_{mc}) &= \ln\left(\boldsymbol{\pi}_0^\top Z_0\right) + \sum_{k=0}^{n-1} \ln\left(p_{s_k, s_{k+1}}(\Delta_{k+1})\right)\,.\end{aligned}$$

The full likelihood is therefore the product of log-likelihoods of the counting process $\mathcal{L}_N(\Theta_N)$ and of the Markov chain $\mathcal{L}_{mc}(\Theta_{mc})$:

$$\begin{aligned}&\mathcal{L}_v\left(\Theta_N, \Theta_{mc}\right) = \\ &\boldsymbol{\pi}_0^\top Z_0 \left(\prod_{k=0}^{n-1} p_{s_k,s_{k+1}}(\Delta_{k+1}) f_{k+1}\left(\boldsymbol{X}_{k+1}, T_{k+1}, M_{k+1} \,|\, \mathcal{E}_k\,,\, \boldsymbol{Z}_k\right)\right) \boldsymbol{S}_n^\top Z_n\,.\end{aligned} \tag{10}$$

If the Markov chain is visible, We can estimate the parameters by maximizing the log-likelihood numerically. Nevertheless, in real applications, the Markov chain is usually a hidden process. Therefore, we need to filter $\boldsymbol{Z}_t$ for calculating the likelihood, as detailed in the next section.

4 Likelihood with a hidden Markov chain

Computing the likelihood when the Markov chain is unobserved requires to guess the state of $\boldsymbol{Z}_t$ using the sample of past events. We introduce the joint density of all observed arrivals up to T_k and the hidden state after the $(k-1)^{th}$ event:

$$\begin{aligned}\alpha_k(j) &:= f\left(\mathcal{E}_k\,,\, \boldsymbol{Z}_{k-1} = \boldsymbol{e}_j\right) \\ &= f_k\left(\boldsymbol{E}_k \,|\, \mathcal{E}_{k-1},\, \boldsymbol{Z}_{k-1} = \boldsymbol{e}_j\right) f\left(\mathcal{E}_{k-1},\, Z_{k-1} = \boldsymbol{e}_j\right)\,.\end{aligned} \tag{11}$$

The joint density of $\mathcal{E}_{k-1}$, $\boldsymbol{Z}_{k-1}$ can be rewritten as sum of joint densities of the triplet $\mathcal{E}_{k-1}$, $\boldsymbol{Z}_{k-1}$, $\boldsymbol{Z}_{k-2}$ and the Bayes' rule allows us to relate this density to $\alpha_{k-1}(.)$:

$$\begin{aligned}f\left(\mathcal{E}_{k-1},\, \boldsymbol{Z}_{k-1} = \boldsymbol{e}_j\right) &= \sum_{i=1}^{d} f\left(\mathcal{E}_{k-1},\, \boldsymbol{Z}_{k-1} = \boldsymbol{e}_j,\, \boldsymbol{Z}_{k-2} = \boldsymbol{e}_i\right) \\ &= \sum_{i=1}^{d} \mathbb{P}\left(\boldsymbol{Z}_{k-1} = \boldsymbol{e}_j \,|\mathcal{E}_{k-1},\, \boldsymbol{Z}_{k-2} = \boldsymbol{e}_i\right) \alpha_{k-1}(i)\,.\end{aligned}$$

Substituting this expression into Equation 11, leads to the following recursion for $k = 1, ..., n$ and $j = 1, ..., d$:

$$\alpha_k(j) = f_k\left(\boldsymbol{E}_k|\mathcal{E}_{k-1}\,,\, \boldsymbol{Z}_{k-1} = \boldsymbol{e}_j\right) \sum_{i=1}^{d} p_{i,j}(\Delta_{k-1}) \alpha_{k-1}(i)\,. \tag{12}$$

The pdf $f_k\left(\boldsymbol{E}_k|\mathcal{E}_{k-1}\,,\, \boldsymbol{Z}_{k-1}\right)$ and transition probabilities $p_{i,j}(\Delta_{k-1})$ are defined by Equations (5) and (9) (with the convention that $\Delta_0 = 0$). $\alpha_{k-1}(i)$ contains all information up to event $k-1$, $p_{i,j}(\Delta_{k-1})$ propagates the hidden state and f_k provides the emission density. We retrieve a similar structure to a classical Hidden Markov Model (HMM). Interestingly, this would not be the case for a standard Markov-modulated Hawkes process with intensity depending on the full path $(\boldsymbol{Z}_t)_{t\geq 0}$, because the emission distribution would depend on the unobserved evolution of $\boldsymbol{Z}_t$ between events.

In later developments, we denote by $\boldsymbol{\alpha}_k := (\alpha_k(1), \ldots, \alpha_k(d))^\top$ the vector of joint densities of $\mathcal{E}_{k-1}$, $\boldsymbol{Z}_{k-1}$. Since the first arrival depends on Z_0, $\boldsymbol{\alpha}_0 := \boldsymbol{\pi}_0$. In order to rewrite the recursion (12) in matrix form, we define the following diagonal matrix:

$$D_k := \operatorname{diag}\left(f_k(\boldsymbol{E}_k|\mathcal{E}_{k-1}\,,\, \boldsymbol{e}_1), \ldots, f_k(\boldsymbol{E}_k|\mathcal{E}_{k-1}\,,\, \boldsymbol{e}_d)\right)\,. \tag{13}$$

The recursion (12) in matrix form is then equal to:

$$\boldsymbol{\alpha}_k = D_k\, P^{(k-1)\top} \boldsymbol{\alpha}_{k-1}\,. \tag{14}$$

This is nearly a standard Hidden Markov Model forward recursion. After processing all arrivals we must account for the absence of further events over $[T_n, T]$.

Proposition 3. *Suppose the sequence* $(\boldsymbol{Z}_i)_{i=0,...,n}$ *is hidden. Let* Θ_N *and* Θ_{mc} *be respectively the parameters of the counting process and of the Markov chain.* $N(t)$*. The likelihood of the sample* $\mathcal{E}_n$ *is the following product*

$$\begin{aligned}\mathcal{L}_h\left(\Theta_N, \Theta_{mc}\right) &= \boldsymbol{S}_n^\top P^{(n)\top} \boldsymbol{\alpha}_n \\ &= \boldsymbol{S}_n^\top P^{(n)\top} D_n P^{(n-1)\top} \dots D_1 P^{(0)\top} \boldsymbol{\pi}_0 \,.\end{aligned} \tag{15}$$

where $\boldsymbol{S}_n$ *is the vector of* $S_n^{(i)}$ *such as defined in Equation (6).*

Proof. Let us define by $\mathcal{L}_h^{(j)}\left(\Theta_N, \Theta_{mc}\right)$ the likelihood if the Markov chain is in the j^{th} state at time T_n. Using the Bayes rule, we develop it as follows:

$$\begin{aligned}\mathcal{L}_h^{(j)}\left(\Theta_N, \Theta_{mc}\right) &:= f\left(\mathbf{1}_{\{T_{n+1}>T\}}, \mathcal{E}_n\,,\, \boldsymbol{Z}_n = \boldsymbol{e}_j\right) \\ &= \mathbb{P}\left(T_{n+1} > T \,|\, \mathcal{E}_n\,,\, \boldsymbol{Z}_n = \boldsymbol{e}_j\right) f\left(\mathcal{E}_n,\, \boldsymbol{Z}_n = \boldsymbol{e}_j\right)\,.\end{aligned}$$

As $f\left(\mathcal{E}_n,\, \boldsymbol{Z}_n\right)$ is equal a sum over marginal pdfs of $(\mathcal{E}_n, \boldsymbol{Z}_n, \boldsymbol{Z}_{n-1})$, we rewrite this likelihood as:

$$\begin{aligned}\mathcal{L}_h^{(j)}\left(\Theta_N, \Theta_{mc}\right) &= S_n^{(j)} \sum_{i=1}^{d} f\left(\mathcal{E}_n,\, \boldsymbol{Z}_n = \boldsymbol{e}_j,\, \boldsymbol{Z}_{n-1} = \boldsymbol{e}_i\right) \\ &= S_n^{(j)} \sum_{i=1}^{d} \mathbb{P}\left(\boldsymbol{Z}_n = \boldsymbol{e}_j \,|\, \mathcal{E}_n,\, \boldsymbol{Z}_{n-1} = \boldsymbol{e}_i\right) \alpha_n(i)\,,\end{aligned}$$

which is equivalent to Equation (15). □

In the estimation procedure, we never evaluate the matrix product (15) directly because of underflow. Instead, we normalize the $\boldsymbol{\alpha}_k$. For $k = 1, ..., n$, we define the normalized forward vectors as follows

$$\tilde{\boldsymbol{\alpha}}_k \quad = \quad \frac{D_k P^{(k-1)\top} \tilde{\boldsymbol{\alpha}}_{k-1}}{c_k}\,,$$

where $\tilde{\boldsymbol{\alpha}}_0 = \boldsymbol{\pi}_0$ and c_k is a normalizing constant:

$$c_k = \mathbf{1}^\top D_k P^{(k-1)\top} \tilde{\boldsymbol{\alpha}}_{k-1}\,.$$

The likelihood is then rewritten as the product of scaling constants and of $\boldsymbol{S}_n^\top P^{(n)\top} \tilde{\boldsymbol{\alpha}}_n$:

$$\mathcal{L}_h\left(\Theta_N, \Theta_{mc}\right) = \left(\prod_{k=1}^{n} c_k\right) \boldsymbol{S}_n^\top P^{(n)^\top} \tilde{\boldsymbol{\alpha}}_n\,.$$

Consequently, the log-likelihood is a function of scaled quantities:

$$l_h\left(\Theta_N, \Theta_{mc}\right) = \sum_{k=1}^{n} \ln c_k + \ln\left(\boldsymbol{S}_n^\top P^{(n)^\top} \tilde{\boldsymbol{\alpha}}_n\right)\,. \tag{16}$$

We can estimate parameters by direct maximization of the log-likelihood (16) using for instance the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.

5 Forward-Backward filtering

Using (14), we can filter the state probabilities forward. Summing over all possible states gives the marginal density of the observations:

$$f\left(\mathcal{E}_k\right) = \sum_{j=1}^{d} \alpha_k(j) = \boldsymbol{\alpha}_k^\top \mathbf{1}\,.$$

Therefore, by Bayes' formula, we infer the posterior filtered state probabilities:

$$\mathbb{P}\left(Z_{k-1} = \boldsymbol{e}_j \,|\, \mathcal{E}_k\right) = \frac{f\left(\mathcal{E}_k\,,\, Z_{k-1} = \boldsymbol{e}_j\right)}{f\left(\mathcal{E}_k\right)} = \frac{\boldsymbol{\alpha}_k^\top \boldsymbol{e}_j}{\boldsymbol{\alpha}_k^\top \mathbf{1}}\,.$$

These probabilities are calculated with the information up to time T_k. We can then improve the estimation of state probabilities by using the whole sample path of the process, using a forward-backward recursion. Let us define the a-posteriori state probabilities $\gamma_k(i)$ and transition probabilities $\xi_k(i,j)$ using the full sample path:

$$\begin{aligned}\gamma_k(i) &:= \mathbb{P}\left(Z_k = \boldsymbol{e}_i \,|\, \mathcal{E}_n\right)\,,\\ \xi_k(i,j) &:= \mathbb{P}\left(Z_k = \boldsymbol{e}_i\,,\, Z_{k+1} = \boldsymbol{e}_j | \,\mathcal{E}_n\right)\,.\end{aligned}$$

We note $\mathcal{E}_k^c = (E_{k+1}, ..., E_n)$, the sample of events occurring later than T_k and we define the backward state probabilities, $\beta_k(i)$, as the density of $\mathcal{E}_k^c$, conditionally to $\mathcal{E}_k$ and to $Z_{k-1} = \boldsymbol{e}_i$:

$$\beta_k(i) \quad = \quad f\left(\mathcal{E}_k^c \,|\, \mathcal{E}_k\,,\, Z_{k-1} = \boldsymbol{e}_i\right)\,.$$

The corresponding vector is $\boldsymbol{\beta}_k := (\beta_k(1), \ldots, \beta_k(d))^\top$. Notice that we condition with respect to $\mathcal{E}_k$ because the intensity depends on the event history. Conditioning on $(\mathcal{E}_k, Z_{k-1})$ restores the conditional independence structure needed for a forward-backward recursion. By definition, we have $\boldsymbol{\beta}_n = \frac{P^{(n)}\boldsymbol{S}_n}{\mathbf{1}^\top P^{(n)}\boldsymbol{S}_n}$. Using the Bayes rule, we develop $\beta_k(i)$ as follows:

$$\begin{aligned}\beta_k(i) &= \sum_{j=1}^{d} \mathbb{P}\left(\boldsymbol{Z}_k = \boldsymbol{e}_j \,|\, \mathcal{E}_k\,,\, Z_{k-1} = \boldsymbol{e}_i\right)\\ &\quad \times f\left(T_{k+1}\,,\, \mathcal{E}_{k+1}^c\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_j \,|\, \mathcal{E}_k\,,\, \boldsymbol{Z}_{k-1} = \boldsymbol{e}_i\right)\\ &= \sum_{j=1}^{d} p_{ij}(\Delta_k) f\left(\boldsymbol{E}_{k+1} \,|\, \mathcal{E}_k\,,\, \boldsymbol{Z}_{k-1} = \boldsymbol{e}_i,\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_j\right)\\ &\quad \times f\left(\mathcal{E}_{k+1}^c \,|\, \mathcal{E}_{k+1}\,,\, \boldsymbol{Z}_{k-1} = \boldsymbol{e}_i,\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_j\right)\,.\end{aligned}$$

We infer from this last relation that $\beta_k(i)$ fulfills the following recursion:

$$\beta_k(i) = \sum_{j=1}^{d} p_{ij}(\Delta_k) f_{k+1}\left(\boldsymbol{E}_{k+1} \,|\, \mathcal{E}_k\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_j\right) \beta_{k+1}(j)\,, \tag{17}$$

for $k = 0, ..., n-1$ and $i = 1, ..., d$. In matrix form, this recursive equation becomes

$$\boldsymbol{\beta}_k = P^{(k)} D_{k+1} \boldsymbol{\beta}_{k+1}\,,$$

where D_{k+1} is the diagonal matrix defined in equation (13) and $P^{(k)} = e^{A\Delta_k}$ is the matrix of transition probabilities. Using again the Bayes' rule, we infer that the joint pdf of the sample of events and Markov chain at time T_k is equal to

$$\begin{aligned}f\left(\mathcal{E}_n\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_i\right) &= f\left(\mathcal{E}_{k+1}\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_i\right) f\left(\mathcal{E}_{k+1}^c | \mathcal{E}_k,\, \boldsymbol{Z}_k = \boldsymbol{e}_i\right)\\ &= \alpha_{k+1}(i)\beta_{k+1}(i)\,.\end{aligned}$$

The a-posteriori state probabilities $\gamma_k(i) = \mathbb{P}\left(Z_k = \boldsymbol{e}_i \,|\, \mathcal{E}_n\right)$ are therefore estimated by the following ratio:

$$\gamma_k(i) \quad = \quad \frac{f\left(\mathcal{E}_n\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_j\right)}{f\left(\mathcal{E}_n\right)} = \frac{\alpha_{k+1}(i)\beta_{k+1}(i)}{\sum_{j=1}^{d} \alpha_{k+1}(j)\beta_{k+1}(j)}\,. \tag{18}$$

It remains to estimate the probabilities of transition $\xi_k(i,j)$. By applying the Bayes' rule, the joint probability of $\mathcal{E}_n$, $\boldsymbol{Z}_k$ and $\boldsymbol{Z}_{k+1}$ is the product:

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_n\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_i\,, \boldsymbol{Z}_{k+1} = \boldsymbol{e}_j\right) &= f\left(\mathcal{E}_{k+1}\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_i\right) p_{i,j}(\Delta_{k+1}) \\ &\times f_{k+2}\left(\boldsymbol{E}_{k+2} | \mathcal{E}_{k+1}\,, \boldsymbol{Z}_{k+1} = \boldsymbol{e}_j\right) f\left(\mathcal{E}_{k+2}^c \,|\, \mathcal{E}_{k+2}\,,\, \boldsymbol{Z}_{k+1} = \boldsymbol{e}_j\right)\,.\end{aligned}$$

The a-posteriori transition probabilities are therefore equal to the ratio:

$$\begin{aligned}\xi_k(i,j) &= \frac{\mathbb{P}\left(\mathcal{E}_n\,,\, \boldsymbol{Z}_k = \boldsymbol{e}_i\,, \boldsymbol{Z}_{k+1} = \boldsymbol{e}_j\right)}{f\left(\mathcal{E}_n\right)} \\ &= \frac{\alpha_{k+1}(i) p_{i,j}(\Delta_{k+1})\, f_{k+2}\left(\boldsymbol{E}_{k+2} | \mathcal{E}_{k+1}\,, \boldsymbol{Z}_{k+1} = \boldsymbol{e}_j\right) \beta_{k+2}(j)}{\sum_{l=1}^{d} \alpha_{k+1}(l)\beta_{k+1}(l)}\,.\end{aligned} \tag{19}$$

The probabilities $\gamma_k(i)$ and $\xi_k(i,j)$ plays a central role in the EM algorithm, presented in the next section.

6 EM Algorithm

A nice property of the model is that the likelihood admits an analytical expression. Hence, we can directly estimate parameters by numerical maximization of this quantity. Nevertheless, there exists an alternative based on an expectation-maximization approach. We denote by $\Theta^{(m)} = \left(\Theta_N^{(m)}, \Theta_{mc}^{(m)}\right)$ the set of all parameters, including instantaneous transition probabilities, after m iterations of the estimation procedure. If the Markov chain is visible, the complete log-likelihood is provided by Equation (10). Taking logs leads to

$$\begin{aligned}l_v\left(\Theta^{(m)}\right) = \ln\left(\boldsymbol{\pi}_0^\top Z_0\right) + \sum_{k=0}^{n-1} \ln\left(p_{s_k, s_{k+1}}(\Delta_{k+1})\right) + \\ \sum_{k=0}^{n-1} \ln\left(f_{k+1}\left(\boldsymbol{X}_{k+1}, T_{k+1}, M_{k+1} \,|\, \mathcal{E}_k\,,\, \boldsymbol{Z}_k\right)\right) + \ln\left(\boldsymbol{S}_n^\top Z_n\right)\,.\end{aligned} \tag{20}$$

This is the quantity that we average in the expectation (E) step. In the maximization (M) step the parameters are are updated to increase the expected log-likelihood. In the E-step, we calculate

$$Q\left(\Theta\,,\, \Theta^{(m)}\right) \quad := \quad \mathbb{E}\left(l_v\left(\Theta\right) \,|\, \mathcal{E}_n\,,\, \Theta^{(m)}\right).$$

From Equation (20), we rewrite $Q\left(\Theta\,,\, \Theta^{(m)}\right)$ as a function of a-posteriori state probabilities $\gamma_k^{(m)}(i)$ and transition probabilities $\xi_k^{(m)}(i,j)$ with parameters $\Theta^{(m)}$:

$$\begin{aligned}Q\left(\Theta\,,\, \Theta^{(m)}\right) = \sum_{i=1}^{d} \gamma_0^{(m)}(i) \ln\left(\boldsymbol{\pi}_0^\top \boldsymbol{e}_i\right) + \sum_{k=0}^{n-1} \sum_{i,j=1}^{d} \xi_k^{(m)}(i,j) \ln\left(p_{i,j}(\Delta_{k+1})\right) + \\ \sum_{k=0}^{n-1} \sum_{i=1}^{d} \gamma_k^{(m)}(i) \ln\left(f_{k+1}\left(\boldsymbol{X}_{k+1}, T_{k+1}, M_{k+1} \,|\, \mathcal{E}_k\,,\, \boldsymbol{e}_i\right)\right) + \sum_{i=1}^{d} \gamma_n^{(m)}(i) \ln\left(S_n^{(i)}\right)\,.\end{aligned} \tag{21}$$

In the M-step, we find the set of parameters which maximize $Q\left(\Theta\,,\,\Theta^{(m)}\right)$:

$$\Theta^{(m+1)} \;=\; \arg\max_{\Theta} Q\left(\Theta\,,\,\Theta^{(m)}\right)\,.$$

Given the form of $Q\left(\Theta\,,\,\Theta^{(m)}\right)$, we can split this optimization step in three sub-steps. First, we optimize parameters $\Theta_{mc}^{(m+1)} = \left(A^{(m+1)}, \boldsymbol{\pi}_0^{(m+1)}\right)$ of the hidden Markov chain. The estimate of the initial distribution of the Markov chain is

$$\boldsymbol{\pi}_0^{(m+1)} = \arg\max_{\boldsymbol{\pi}_0} \sum_{i=1}^{d} \gamma_0^{(m)}(i) \ln\left(\boldsymbol{\pi}_0^{\top}\boldsymbol{e}_i\right)\,,$$

which admits as solution $\boldsymbol{\pi}_0^{(m+1)} = \boldsymbol{\gamma}_0^{(m)}$ where $\boldsymbol{\gamma}_0^{(m)}$ is the vector $\left(\gamma_0^{(m)}(1)\,,...,\,\gamma_0^{(m)}(d)\right)$. The estimate of the matrix of transition probabilities solves the following optimization problem:

$$A^{(m+1)} = \arg\max_{A} \sum_{k=0}^{n-1}\sum_{i,j=1}^{d} \xi_k^{(m)}(i,j) \ln\left(\left(e^{A\Delta_{k+1}}\right)_{i,j}\right)\,. \tag{22}$$

Unlike the standard discrete-time HMM, the generator matrix A does not admit a closed-form EM update because the transition probabilities are obtained through the matrix exponential $e^{A\Delta_k}$. Nevertheless, if Δ_k are small, we can use a first order approximation and find an analytical expression of $A^{(m+1)}$.

Proposition 4. *Let us denote $N_{ij} = \sum_{k=0}^{n-1} \xi_k^{(m)}(i,j)$ for $i,j \in \{1,..,d\}$ and $B_i = \sum_{k=0}^{n-1} \xi_k^{(m)}(i,i)\Delta_{k+1}$. Using a first order approximation of the matrix exponential, $A^{(m+1)}$, solution of (22), is such that*

$$\begin{cases} a_{ij}^{(m+1)} = \frac{N_{ij}}{B_i} & i \neq j\,, \\ a_{ii}^{(m+1)} = -\sum_{j\neq i}^{d} \frac{N_{ij}}{B_i} & . \end{cases} \tag{23}$$

Proof. If Δ_k are small, we use the following linear approximation of the matrix exponential:

$$e^{A\Delta_{k+1}} = I_d + A\Delta_{k+1} + \mathcal{O}\left(\Delta_k^2\right)\,.$$

Since the elements of this matrix are equal to

$$\left(e^{A\Delta_{k+1}}\right)_{i,j} = p_{i,j}(\Delta_{k+1}) \approx \begin{cases} a_{ij}\Delta_{k+1}\,, & i \neq j \\ 1 + a_{ii}\Delta_{k+1}\,, & i = j \end{cases}\,,$$

we can rewrite the objective (22) to maximize as follows

$$\begin{aligned} &\sum_{k=0}^{n-1}\sum_{i,j=1}^{d} \xi_k^{(m)}(i,j) \ln\left(\left(e^{A\Delta_{k+1}}\right)_{i,j}\right) \\ &\quad\approx \sum_{k=0}^{n-1}\left[\sum_{i\neq j}^{d} \xi_k^{(m)}(i,j)\ln\left(a_{ij}\Delta_{k+1}\right) + \sum_{i}^{d} \xi_k^{(m)}(i,i)\ln\left(1 + a_{ii}\Delta_{k+1}\right)\right]\,. \end{aligned}$$

By construction, the diagonal of A is such that $a_{ii} = -\sum_{i\neq j} a_{ij}$, for $i = 1,...,d$ and if $N_{ij} = \sum_{k=0}^{n-1} \xi_k^{(m)}(i,j)$, the objective (22) is proportional to:

$$\begin{aligned} &\sum_{k=0}^{n-1}\sum_{i,j=1}^{d} \xi_k^{(m)}(i,j) \ln\left(\left(e^{A\Delta_{k+1}}\right)_{i,j}\right) \\ &\quad\propto \left[\sum_{i\neq j}^{d} N_{i,j}\ln\left(a_{ij}\right) + \sum_{i}^{d}\sum_{k=0}^{n-1} \xi_k^{(m)}(i,i)\ln\left(1 - \sum_{i\neq j} a_{ij}\Delta_{k+1}\right)\right]\,. \end{aligned}$$

Using again a first order approximation for the logarithm yields:

$$\sum_{k=0}^{n-1}\sum_{i,j=1}^{d}\xi_k^{(m)}(i,j)\ln\left(\left(e^{A\Delta_{k+1}}\right)_{i,j}\right) \propto \left[\sum_{i\neq j}^{d} N_{i,j}\ln\left(a_{ij}\right)+\sum_{i}^{d} a_{ii}B_i\right]$$
$$= \sum_{i\neq j}^{d}\left(N_{i,j}\ln\left(a_{ij}\right)-a_{ij}B_i\right) ,$$

where $B_i = \sum_{k=0}^{n-1}\xi_k^{(m)}(i,i)\Delta_{k+1}$. We retrieve Equation (23) by cancelling the first order derivative of this last expression. □

The final step consists of maximizing the likelihood with respect to the parameters $\Theta_N^{(m+1)}$ driving the Hawkes process in each regime:

$$\Theta_N^{(i+1)} = \arg\max_{(\boldsymbol{\mu},\boldsymbol{\phi})}\left[\sum_{i=1}^{d}\sum_{k=0}^{n-1}\gamma_k^{(m)}(i)\ln\left(f_{k+1}\left(\boldsymbol{X}_{k+1},T_{k+1},M_{k+1}\,|\,\mathcal{E}_k\,,\,\boldsymbol{e}_i\right)\right)\right. \tag{24}$$
$$\left.+\sum_{k=0}^{n-1}\sum_{i=1}^{d}\gamma_n^{(m)}(i)\ln\left(S_n^{(i)}\right)\right] .$$

In practice, this optimization is performed numerically by running a limited number of iterations of e.g. the BFGS algorithm. The full EM algorithm is summarized below.

EM Algorithm

1) Random initialization of $\Theta^{(0)} = \left(\Theta_N^{(0)},\Theta_{mc}^{(0)}\right)$.

2) Loop on m :

2.1) Calculation of $\gamma_k^{(m)}(i)$ and transition probabilities $\xi_k^{(m)}(i,j)$ with parameters $\Theta^{(m)}$.

2.2) We update $\boldsymbol{\pi}_0^{(m+1)} = \boldsymbol{\gamma}_0^{(m)}$ and $A^{(m+1)} = \left(a_{i,j}^{(m+1)}\right)_{i,j=1,\dots,d}$:

$$\begin{cases} a_{ij}^{(m+1)} = \frac{N_{ij}}{B_i} & i\neq j\,, \\ a_{ii}^{(m+1)} = -\sum_{j\neq i}^{d}\frac{N_{ij}}{B_i} & . \end{cases}$$

2.3) We run a few iterations (e.g. a dozen) of an optimization algorithm to find an approximated solution $\Theta_N^{(i+1)}$ of Equation (24).

7 Simulation algorithm

We adapt Ogata's algorithm to simulate the counting process. It is based on the thinning property of a non-homogeneous (NH) Poisson process. A NH Poisson process, $N(t)$, with intensity $\Lambda(t)$ can be split into m independent NH Poisson processes $N_j(t)$, with intensities $\lambda_j(t)$ for $j = 1,...,m$ such that $\Lambda(t) = \sum_{j=1}^{m}\lambda_j(t)$. An event is caused by the j^{th} process with a probability $\frac{\lambda_j(t)}{\Lambda(t)}$.

We describe how to sample the $(k+1)^{th}$ event $\boldsymbol{E}_{k+1} = (\boldsymbol{X}_{k+1},T_{k+1},\boldsymbol{M}_{k+1})$, after having already sampled $\boldsymbol{E}_1,...,\boldsymbol{E}_k$ and $\boldsymbol{Z}_1,...,\boldsymbol{Z}_k$. Let us assume that $\boldsymbol{Z}_k = \boldsymbol{e}_i$.

Simulation algorithm:

1) We start our time counter $S = T_{k+}$. We sample an inter-arrival time τ, according to an exponential distribution with parameter

$$\Lambda^* := \Lambda(T_{k+}) = \int_{\mathcal{S}} \lambda(\boldsymbol{x}, T_{k+})\, d\boldsymbol{x}\,.$$

2) We update the time counter $S = S + \tau$. We accept or reject this inter-arrival time according to the ratio of the true event rate $\Lambda(S_-)$ to the thinning rate λ^* .

3) If accepted, we record the $(k+1)$ event time as $T_{k+1} = S$. Otherwise, we repeat the sampling of an inter-arrival time until one is accepted.

4) We sample the mark M_{k+1} from $f_M\left(m \,|\, \boldsymbol{Z}_k = \boldsymbol{e}_i\right)$.

5) To sample the location $\boldsymbol{X}_{k+1}$, we rewrite $\Lambda(T_{k+1-}) = W_0^{(i)} + \sum_{j=1}^{k} W_j^{(i)}$ where

$$\begin{cases} W_0^{(i)} = \int_{\mathcal{S}} \mu^{(i)}(\boldsymbol{x})\, d\boldsymbol{x}\,, \\ W_j^{(i)} = \int_{\mathcal{S}} \phi^{(i)}\left(\boldsymbol{x} - \boldsymbol{X}_j, T_{k+1} - T_j, M_j\right)\, d\boldsymbol{x} \quad j = 1, ..., k\,. \end{cases} \tag{25}$$

We choose the parent by sampling a discrete variable $J \in \{0, 1, ..., k\}$ with probabilities

$$\mathbb{P}(J = j) = \frac{W_j^{(i)}}{\Lambda(T_{k+1-})}\,,$$

where $J = 0$ is a new background event and $J = j \in \{1, ..., k\}$ is the offspring of the j^{th} event. Next we sample the location conditionally to J with the pdf:

$$\begin{cases} f_{\boldsymbol{X}}(\boldsymbol{x} \,|\, J = 0) = \frac{\mu^{(i)}(\boldsymbol{x})}{W_0^{(i)}}\,, \\ f_{\boldsymbol{X}}(\boldsymbol{x} \,|\, J = j) = \frac{\phi^{(i)}(\boldsymbol{x} - \boldsymbol{X}_j, T_{k+1} - T_j, M_j)}{W_j^{(i)}} \quad j = 1, ..., k\,. \end{cases}$$

6) Finally, we sample $\boldsymbol{Z}_{k+1}$ from a multinomial distribution with probabilities

$$p^* = \boldsymbol{Z}_k^\top \exp\left(A(T_{k+1} - T_k)\right)$$

We repeat this loop until $T_{k+1} \geq T$.

If the kernel is separable into time, space and mark independent kernels:

$$\phi^{(i)}(\boldsymbol{u}, s, m) = \phi_{\mathcal{S}}^{(i)}(\boldsymbol{u})\phi_{\mathcal{T}}^{(i)}(s)\phi_{\mathcal{M}}^{(i)}(m)$$

and if the space kernel is a pdf, i.e. $\int_{\mathcal{S}} \phi_{\mathcal{S}}^{(i)}(\boldsymbol{u})\, d\boldsymbol{u} = 1$, the $W_j^{(i)}$ such as defined in Equation (25) becomes :

$$W_j^{(i)} = \phi_{\mathcal{T}}^{(i)}\left(T_{k+1} - T_j\right) \phi_{\mathcal{M}}^{(i)}\left(M_j\right)\ ,\ j = 1, ..., k\,.$$

The spatial pdf of the new event location, conditionally to J simplifies as follows:

$$f_{\boldsymbol{X}}(\boldsymbol{x} \,|\, J = j) = \phi_{\mathcal{S}}^{(i)}\left(\boldsymbol{x} - \boldsymbol{X}_j\right) \quad , \quad j = 1, ..., k\,.$$

8 Parametric insurance based on regional mark

The spatial marked structure of the switching Hawkes process provides a natural framework for the design of parametric insurance contracts. In contrast with indemnity-based insurance, a parametric contract pays a predetermined amount when an observable physical index exceeds a specified threshold. A particularly attractive index is the mark of an event occurring within a predefined geographical region. The main advantage of this approach is that the payment can be determined immediately from publicly available observations, without requiring an assessment of the actual economic damage. In the numerical illustration, we consider a parametric insurance contract covering earthquake damage, whose the index is the magnitude on the Richter scale.

Let $D \subset \mathbb{R}^2$ denote the geographical region covered by the contract. To account for geographical heterogeneity, we consider an elliptical region centered at $x_0 = (x_{0,1}, x_{0,2})$, with semi-major and semi-minor axes a and b, and orientation angle θ. A location x belongs to the insured region if

$$(x - x_0)^\top R_\theta \begin{pmatrix} a^{-2} & 0 \\ 0 & b^{-2} \end{pmatrix} R_\theta^\top (x - x_0) \leq 1, \tag{26}$$

where

$$R_\theta \;=\; \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},$$

is the rotation matrix associated with angle θ. The use of an ellipse is useful when the insured exposure is geographically elongated, for example along a coastline, a fault line, or a group of municipalities. More generally, the methodology applies to any measurable spatial domain D. Let T denote the duration of the insurance contract and define

$$N_D(T) = \sum_{T_k \leq T} \mathbf{1}_{\{X_k \in D\}},$$

the number of events occurring inside the insured region during the contract period. We define the regional maximum mark during the contract period as

$$M_D(T) = \max_{\{k: T_k \leq T,\, \boldsymbol{X}_k \in D\}} M_k,$$

with $M_D(T) = m_0$ when no events occur in D. The use of the maximum mark is particularly natural for parametric insurance because the mark is usually not additive across events in the same way as an ordinary monetary claim. Consider a contract with trigger mark m_L, exhaustion mark $m_U > m_L$, and maximum payment L. A simple linear payout function is

$$Y_T = L \times g\left(M_D(T)\right), \tag{27}$$

where

$$g(m) = \begin{cases} 0, & m < m_L, \\ \dfrac{m - m_L}{m_U - m_L}, & m_L \leq m < m_U, \\ 1, & m \geq m_U. \end{cases} \tag{28}$$

Thus, the contract pays nothing below m_L, increases linearly between m_L and m_U, and reaches its maximum payment L at m_U. For example, a contract insuring damages caused by an earthquake could specify a payment of 0 below magnitude 5.5, a linearly increasing payment between

magnitudes 5.5 and 6.5, and a full payment above magnitude 6.5. Importantly, the payment depends exclusively on the observed magnitude and location of the earthquake and does not require an estimate of the actual damage suffered by the policyholder. The actuarial pure premium is the expected value of the contract payment,

$$\Pi^{\text{pure}} = \mathbb{E}\left(Y_T|\mathcal{G}_0\right) = \sum_{i=1}^{d} \pi_0^{(i)} \mathbb{E}\left(Y_T \mid \boldsymbol{Z}_0 = e_i\right).$$

More generally, if information about the current regime is available at the valuation date, the insurer can replace the unconditional probabilities $\pi_0^{(i)}$, by filtered probabilities $\mathbb{P}\left(\boldsymbol{Z}_t = e_i \mid \mathcal{G}_t\right)$, and obtain a dynamically updated estimate of the expected claim cost. Consequently, pricing the contract requires the probability that at least one event with mark exceeding m_L occurs inside D during the contract period. This probability is directly related to the spatial switching Hawkes process. Conditionally on the event history and the active regime $Z_{j(t)} = e_i$, the spatial intensity is

$$\lambda(\boldsymbol{x}, t) = \mu^{(i)}(\boldsymbol{x}) + \sum_{T_k < t} \phi^{(i)}\left(\boldsymbol{x} - \boldsymbol{X}_k, t - T_k, M_k\right).$$

Therefore, the instantaneous intensity of events occurring inside the insured region is $\Lambda_D(t) = \int_D \lambda(\boldsymbol{x}, t)\, d\boldsymbol{x}$. The corresponding intensity of events satisfying both the geographical and mark trigger can be written as

$$\Lambda_{D,m_L}(t) = \int_D \int_{m_L}^{\infty} f_M(m \mid \boldsymbol{Z}_{j(t)})\, \lambda(\boldsymbol{x}, t)\, dm\, d\boldsymbol{x},$$

with the precise expression depending on the specifications of the model. Conditional on the complete history and regime path, the probability of observing no trigger over $[0, T]$ is therefore determined by the integrated trigger intensity,

$$\mathbb{P}\left(M_D(T) < m_L \mid \mathcal{F}_T\right) = \exp\left(-\int_0^T \Lambda_{D,m_L}(t)\, dt\right).$$

This expression highlights an important actuarial feature of the proposed framework: the probability of an insurance trigger depends not only on the marginal frequency of earthquakes but also on their spatial concentration, mark distribution and temporal clustering.

Although some conditional quantities can be calculated analytically, the self-exciting structure makes the unconditional distribution of the regional maximum mark difficult to obtain in closed form. A practical actuarial pricing procedure is therefore to use the simulation algorithm developed in the previous section. For each simulated trajectory $r = 1, \ldots, R$, we generate the sequence

$$\{(T_k^{(r)}, \boldsymbol{X}_k^{(r)}, M_k^{(r)}, \boldsymbol{Z}_k^{(r)})\},$$

over the contract period. The regional maximum magnitude is then

$$M_D^{(r)} = \max_{\{T_k^{(r)} \le T,\, X_k^{(r)} \in D\}} M_k^{(r)},$$

and the corresponding simulated payment is $Y_T^{(r)} = L\, g\left(M_D^{(r)}\right)$. The Monte Carlo estimator of the pure premium is

$$\widehat{\Pi}^{\text{pure}} = \frac{1}{R} \sum_{r=1}^{R} Y_T^{(r)}.$$

The simulation algorithm already developed in this paper is particularly convenient for this purpose because it jointly generates the event time, location and mark and subsequently updates the hidden regime. Thus, no additional stochastic model is required to generate the variables entering the insurance trigger. The pure premium does not represent the final premium charged to the policyholder. An insurer must also account for expenses, capital requirements, model uncertainty and the cost of bearing catastrophe risk. For instance, a common approach is to use a risk measure, such as the Value-at-Risk (VaR) of level α, on the simulated annual payment distribution:

$$\Pi^{\text{gross}} = \mathbb{E}\left(Y_T\right) + \theta\, \text{VaR}_\alpha\left(Y_T\right), \tag{29}$$

The principal actuarial limitation of a parametric contract is basis risk. The insurance payment is determined by the physical trigger rather than by the actual economic loss. Two events of identical mark can produce substantially different economic consequences depending on their precise location, depth, local construction characteristics and exposure concentration. Nevertheless, the spatial component of the proposed model provides a mechanism for reducing this basis risk. Instead of defining the trigger solely through mark, the contract conditions the mark trigger on the event occurring inside the insured region D. The choice of D therefore becomes an important actuarial parameter. In the numerical illustration, we compare prices of the parametric insurance detailed in this section, for covering earthquake damages in the area of Athens.

9 Switching ETAS model with Pareto marks

In spatial applications, it is convenient to split the kernel into spatial, time and mark components:

$$\phi^{(i)}(\boldsymbol{x}-\boldsymbol{u}, t-s, m) = \phi^{(i)}_{\mathcal{S}}(\boldsymbol{x}-\boldsymbol{u})\phi^{(i)}_{\mathcal{T}}(t-s)\phi^{(i)}_{\mathcal{M}}(m) \quad , \quad i = 1, ..., d\,.$$

In the numerical illustration, we focus on the Epidemic Type Aftershock Sequence (ETAS) kernel, that is modulated by the Markov chain $\boldsymbol{Z}_t$. We denote by $\|\boldsymbol{x}-\boldsymbol{u}\|_2 = \sqrt{\sum_{j=1}^{2}(x_j-u_j)^2}$,the Euclidean distance between locations $\boldsymbol{x} = (x_1, x_2)$ and $\boldsymbol{u} = (u_1, u_2)$. In the numerical illustration, the mark is the magnitude of earthquakes that is modeled by a Pareto distribution. We consider a slightly different version of the ETAS model, in which the kernel of marks is proportional to the power of the spread between the mark and a lower bound. More precisely, the spatial, time and mark kernels are respectively defined by

$$\begin{cases} \phi^{(i)}_{\mathcal{M}}(m) = \kappa_i\left(1 + (m-m_0)^{b_i}\right) & , \kappa_i,\, b_i > 0\,,\, m_0 > 0\,, \\ \phi^{(i)}_{\mathcal{T}}(t-s) = \frac{1}{((t-s)+c_i)^{p_i}} & , c_i > 0\,,\, p_i > 1\,, \\ \phi^{(i)}_{\mathcal{S}}(\boldsymbol{x}-\boldsymbol{u}) = \frac{1}{\left(\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + l_i^2\right)^{q_i}} & , l_i > 0\,,\, q_i > 1\,. \end{cases}$$

for $i = 1, ..., d$. The quantity m_0 denotes the mark of completeness of the catalog, so that all marks are greater than m_0. In the classical version of the ETAS model, the kernel of marks is exponential $\phi_{\mathcal{M}}(m) = \kappa e^{b(m-m_0)}$ with $b > 0$ and marks are usually exponential random variables. Using a power kernel is motivated by the choice of a Pareto distribution for marks. We can in this case find the conditions that ensure the stability of the process. Furthermore, empirically, we obtain an excellent goodness of fit of Greeks earthquakes with a power kernel and Pareto distributed marks.

Proposition 5. *The switching ETAS model is stable if the following inequality holds for all* $i = 1, ..., d$*:*

$$||\phi^{(i)}||_1 = \frac{\pi\, l_i^{2-2q_i}}{(q_i-1)\,(p_i-1)\,(c_i)^{p_i-1}}\,\mathbb{E}\left(\phi^{(i)}_{\mathcal{M}}(M)\right) < 1\,. \tag{30}$$

Proof. For all $j \in \{1, ..., d\}$, the process is stable if the norm of $\phi^{(i)}$ is upper bounded by one:

$$||\phi^{(i)}||_1 = \int_{\mathcal{S}} \phi_{\mathcal{S}}^{(i)}(\boldsymbol{x})\, d\boldsymbol{x} \int_0^{\infty} \phi_{\mathcal{T}}^{(i)}(s)\, ds\, \mathbb{E}\left(\phi_{\mathcal{M}}^{(i)}(M)\right) < 1\,.$$

The integral of the time kernel is equal to

$$\int_0^{\infty} \phi_{\mathcal{T}}^{(i)}(s)\, ds \;\; = \;\; \frac{1}{(p_i - 1)\,(c_i)^{p_i - 1}}\,. \tag{31}$$

Under the assumption that the spatial domain is large, we approximate $\mathcal{S}$ by $\mathbb{R}^2$. The integral of the spatial kernel is therefore approximated by it upper bound:

$$\begin{aligned}\int_{\mathcal{S}} \phi_{\mathcal{S}}^{(i)}(\boldsymbol{x})\, d\boldsymbol{x} &\approx \int_{\mathbb{R}^2} \phi_{\mathcal{S}}^{(i)}(\boldsymbol{x})\, d\boldsymbol{x} \\ &= \int_{\mathbb{R}^2} \left(\left(x_1^2 + x_2^2\right) + l_i^2\right)^{-q_i}\, d\boldsymbol{x}\,.\end{aligned}$$

Switching to polar coordinates , $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$, we have $d\boldsymbol{x} = r\, dr\, d\theta$ and

$$\begin{aligned}\int_{\mathbb{R}^2} \phi_{\mathcal{S}}^{(i)}(\boldsymbol{x})\, d\boldsymbol{x} &= \int_0^{2\pi} \int_0^{\infty} r\left(r^2 + l_i^2\right)^{-q_i}\, dr\, d\theta \\ &= 2\pi \int_0^{\infty} r\left(r^2 + l_i^2\right)^{-q_i}\, dr\,.\end{aligned}$$

We perform a second change of variable: $u = r^2 + l_i^2$ and $du = 2r\, dr$. We finally obtain that:

$$\int_{\mathbb{R}^2} \phi_{\mathcal{S}}^{(i)}(\boldsymbol{x})\, d\boldsymbol{x} = \pi \int_{d_i^2}^{\infty} u^{-q_i}\, du = \frac{\pi}{q_i - 1} l_i^{2-2q_i}\,. \tag{32}$$

Combining (31) and (32) leads to the condition (30). □

As mentioned earlier, we assume that marks are Pareto distributed, with parameters $\boldsymbol{\eta} = (\eta_1, ..., \eta_d)^{\top}$ in each regime. The pdf of marks is in this case equal to

$$f_M^{(i)}(m) = \frac{\eta_i m_0^{\eta_i}}{m^{\eta_i + 1}} \quad , m \geq m_0\,,\, i = 1, ..., d\,.$$

The stability of the process depends upon the expectation of the kernel of marks that is provided in the next result.

Proposition 6. *If marks are Pareto distributed and $\phi_{\mathcal{M}}^{(i)}(m) = \kappa_i\left(1 + (m - m_0)^{b_i}\right)$, for $i \in \{1, ..., d\}$:*

$$\mathbb{E}\left(\phi_{\mathcal{M}}^{(i)}(M)\right) = \kappa_i + \kappa_i\, \eta_i\, m_0^{b_i}\, B\left(\eta_i - b_i\,,\, b_i + 1\right)\,, \tag{33}$$

where $B(u, v) = \int_0^1 x^{u-1}\,(1 - x)^{v-1}\, dx$ is the Beta function.

Proof. By construction, this expectation is:

$$\mathbb{E}\left(\phi_{\mathcal{M}}^{(i)}(M)\right) \;\; = \;\; \kappa_i + \kappa_i \eta_i m_0^{\eta_i} \int_{m_0}^{\infty} (u - m_0)^{b_i} \frac{1}{u^{\eta_i + 1}}\, du\,,\, i = 1, ..., d\,. \tag{34}$$

We perform a change of variables: $u = \frac{m_0}{x}$ and $du = -\frac{m_0}{x^2} dx$. The integral in the expectation becomes then

$$\begin{aligned}\int_{m_0}^{\infty} (u - m_0)^{b_i} \frac{1}{u^{\eta_i + 1}}\, du &= \int_0^1 \left(\frac{m_0}{x} - m_0\right)^{b_i} \left(\frac{m_0}{x}\right)^{-(\eta_i + 1)} \frac{m_0}{x^2}\, dx \\ &= (m_0)^{b_i - \eta_i} \int_0^1 (1 - x)^{b_i}\, x^{\eta_i - b_i - 1}\, dx \\ &= (m_0)^{b_i - \eta_i}\, B\left(\eta_i - b_i\,,\, b_i + 1\right)\,.\end{aligned}$$

This result combined with Equation (34) leads to the result. □

From (33), we infer the following additional condition on the Pareto and kernel parameters:

$$\eta_i - b_i > 0\,,\; i \in \{1, ..., d\}\,. \tag{35}$$

The calculation of the log-likelihood entails to integrate the density over S and a time interval $[T_{k-1}\,,\,t)$. The next corollary provides the analytical expression of this integral.

Corollary 7. *If $\boldsymbol{Z}_{k-1} = \boldsymbol{e}_i$, the integral of the intensity over S and a time interval $[T_{k-1}\,,\,t)$ is:*

$$\int_{\mathcal{S}}\int_{T_{k-1}}^{t} \lambda_k\left(\boldsymbol{x}, s\right)\, ds\, d\boldsymbol{x} = \left(t - T_{k-1}\right) \int_{\mathcal{S}} \mu^{(i)}(\boldsymbol{x})\, d\boldsymbol{x} + \frac{\pi\, l_i^{2-2q_i}}{\left(q_i - 1\right)\left(p_i - 1\right)} \tag{36}$$
$$\times \left(\sum_{T_j < t} \left(\left(T_{k-1} - T_j + c_i\right)^{1-p_i} - \left(t - T_j + c_i\right)^{1-p_i} \right) \kappa_i \left(1 + \left(M_j - m_0\right)^{b_i} \right) \right)\,.$$

Proof. Let us assume that $\boldsymbol{Z}_{k-1} = \boldsymbol{e}_i$. The integral of the intensity on the spatial domain from T_{k-1} to t, is in this case equal to the following sum

$$\int_{\mathcal{S}}\int_{T_{k-1}}^{t} \lambda_k\left(\boldsymbol{x}, s\right)\, ds\, d\boldsymbol{x} = \int_{\mathcal{S}}\int_{T_{k-1}}^{t} \mu^{(i)}(\boldsymbol{x})\, ds\, d\boldsymbol{x}$$
$$+ \sum_{T_j < t} \int_{\mathcal{S}} \phi_{\mathcal{S}}^{(i)}\left(\boldsymbol{x} - \boldsymbol{X}_j\right) d\boldsymbol{x} \int_{T_{k-1}}^{t} \phi_{\mathcal{T}}^{(i)}\left(s - T_j\right)\, ds\, \phi_{\mathcal{M}}^{(i)}\left(M_j\right)\,,$$

where the integral of the spatial kernel on $\mathcal{S}$ is approximated by an integral on $\mathbb{R}^2$:

$$\int_{\mathcal{S}} \phi_{\mathcal{S}}^{(i)}\left(\boldsymbol{x} - \boldsymbol{X}_j\right) d\boldsymbol{x} \approx \frac{\pi}{q_i - 1} l_i^{2-2q_i}\,.$$

Since the integral of the temporal kernel is equal to

$$\begin{aligned} \int_{T_{k-1}}^{t} \phi_{\mathcal{T}}^{(i)}\left(s - T_j\right)\, ds &= \int_{T_{k-1}}^{t} \frac{1}{\left(s - T_j + c_i\right)^{p_i}}\, ds \\ &= \frac{1}{\left(1 - p_i\right)} \left(\left(t - T_j + c_i\right)^{1-p_i} - \left(T_{k-1} - T_j + c_i\right)^{1-p_i} \right)\,, \end{aligned}$$

we obtain the expression (36). □

10 Numerical illustration

10.1 Estimation

We fit the model to a dataset of earthquakes recorded in Greece and the surrounding region over the period from 3 October 2016 to 2 October 2026. The data are obtained from the U.S. Geological Survey (USGS)[1]. We restrict the sample to earthquakes occurring within the geographical window defined by latitudes 34.727°–40.800° N and longitudes 19.094°–26.521° E, with magnitudes of at least 4.5 and focal depths below 60 km. These selection criteria yield a sample of 636 earthquakes. Figure 1 provides two complementary views of the data: the left panel displays the temporal evolution of earthquake magnitudes, while the right panel shows the geographical distribution of the recorded events.

[1] Source : https://earthquake.usgs.gov/

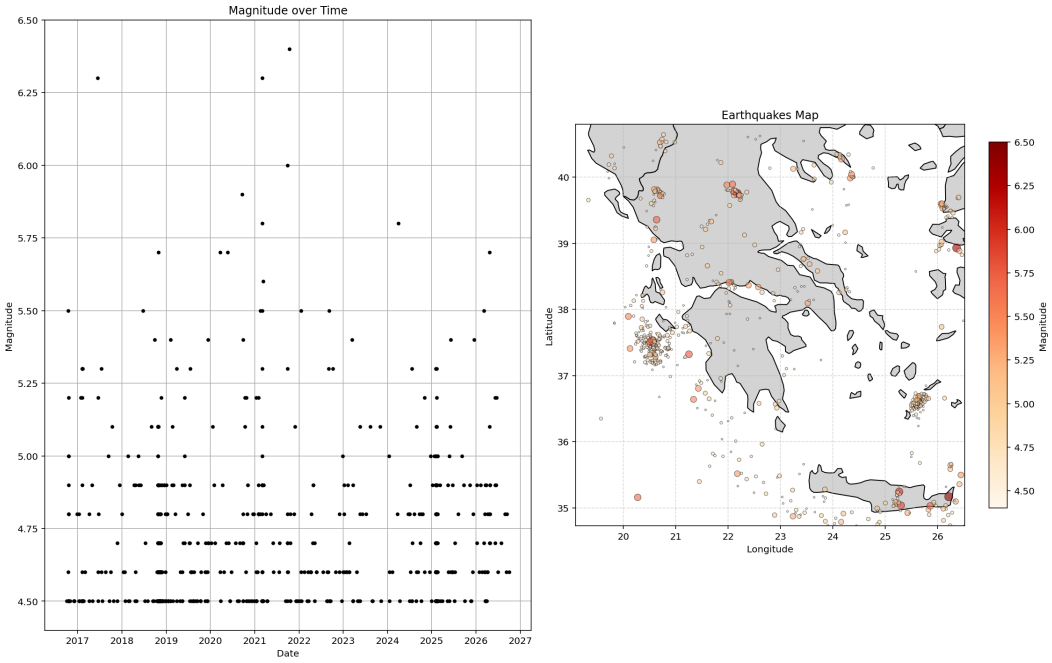


Figure 1: Left: time series of earthquakes observed in Greece from 3/10/16 to 2/10/26. Right: locations of earthquakes . The magnitude is measured on the Richter scale.

We model the baseline earthquake intensity as the product of a positive constant, μ_i, and a spatial density, $h(\boldsymbol{x})$, so that

$$\mu^{(i)}(\boldsymbol{x}) \quad = \quad \mu_i \times h(\boldsymbol{x}) \quad , i = 1, ..., d\,.$$

The spatial density is estimated using a Gaussian kernel estimator fitted to earthquakes recorded in the study area between 25 September 2016 and 2 October 2026. The bandwidth is selected by maximizing the log-likelihood over a grid of candidate values, yielding an estimated bandwidth of 22.122 km. Figure 2 displays the resulting spatial density estimate, highlighting two apparent hotspots south of the islands of Laganas and Naxos.

To avoid parameter estimates that violate the stability condition, we add a quadratic penalty to the log-likelihood whenever theL^1-norm of an excitation kernel exceeds the threshold of 0.95. The penalized log-likelihood is defined as

$$l_h\left(\Theta_N, \Theta_{mc}\right) - \omega \left(\max_{i \in (1,...d)} \left(||\phi^{(i)}||_1 - 0.95\,;\, 0 \right) \right)^2 .$$

where $\omega > 0$ controls the strength of the penalty. The penalty is zero when all excitation-kernel norms are at most 0.95, and increases quadratically when the largest norm exceeds this threshold.

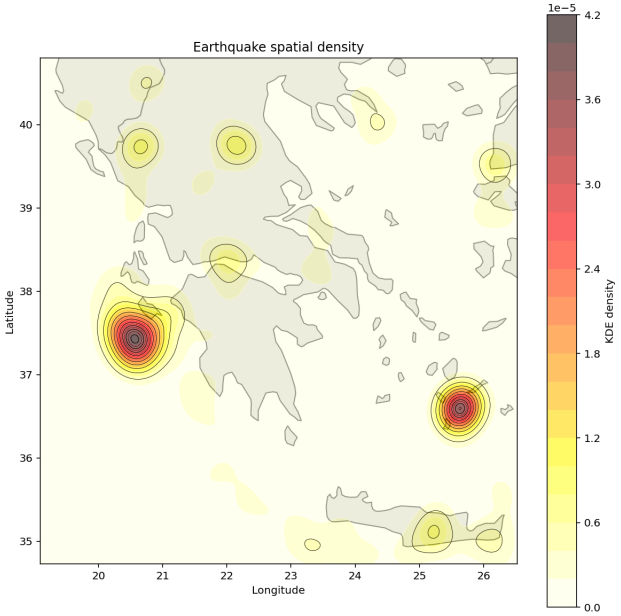


Figure 2: Logarithms of the estimated spatial kernel density.

We also evaluate the fitted models on a validation dataset comprising earthquakes of magnitude at least 4.5 and focal depths below 60 kilometers, recorded between 1 October 2006 and 1 October 2016. This dataset contains 396 events. We compute the log-likelihood on these observations to assess model performance on data outside the training period.

		Poisson	Log-likelihood			EM algorithm	
Train	d	n.a.	1	2	3	2	3
	dofs	3	9	19	31	19	31
	log-L	-9866.14	-3495.71	-3425.79	-3423.23	-3434.54	-3429.98
	AIC	19738.27	7009.42	6889.58	6908.46	6907.09	6921.95
	BIC	19751.64	7049.52	6974.23	7046.58	6991.74	7060.06
Test	log-L.	-6994.51	-3315.34	-3254.01	-3265.08	-3267.94	-3272.48

Table 1: log-likelihoods (log-L), AIC, BIC of switching models with 1, 2 and 3 regimes. The estimation is performed by direct maximization of the log-likelihood and by using the EM-algorithm. "dofs" is the number of parameters.

Table 1 compares the log-likelihood, Akaike information criterion (AIC), and Bayesian information criterion (BIC) of models with one, two and three regimes, and of the Poisson model. We also compare our results with those of a spatial Poisson model with a baseline spatial intensity $\lambda \times h(\boldsymbol{x})$ and Pareto distributed magnitudes estimated by log-likelihood maximization. For the two- and three-regime specifications, we consider both direct maximization of the log-likelihood and estimation using the EM algorithm. The spatial Poisson model achieves the lowest goodness of fit because it fails to explain the spatial dependence of earthquakes. The three-regime model estimated by direct maximization achieves the highest log-likelihood, indicating the best in-sample fit among the models considered. However, both the AIC and BIC attain their lowest values for the two-regime model. This suggests that the improvement in fit obtained by introducing a third regime does not sufficiently compensate for the additional model complexity. According to these information criteria, the two-regime specification therefore provides the preferred balance between goodness of fit and parsimony.

	1D	2D		3D		
		State 1	State 2	State 1	State 2	State 3
μ_i	23.7231	22.9633	24.4141	24.5290	24.2427	22.3955
c_i	3.4e-05	0.0147	5.5e-05	4.7e-05	0.0087	6.7e-05
p_i	1.1156	1.0587	1.2057	1.1989	1.0432	2.4482
l_i	10.5315	9.0621	9.3697	10.1486	9.6454	71.4331
q_i	2.5315	1.9828	2.3736	2.4844	2.2857	6.6332
κ_i	16.8512	0.6786	4.9503	11.1143	4.1766	0.0559
b_i	4.0323	2.9140	3.8963	3.9216	2.1488	0.0599
m_0	4.5	4.5	4.5	4.5	4.5	4.5
η_i	20.5557	19.6066	21.5685	21.5239	19.9739	18.7948

Table 2: Estimated parameters by log-likelihood maximization, one- (1D), two- (2D) and three- (3D) regime models.

The results also indicate that the estimation method affects the quality of the fit. For both the two- and three-regime models, direct maximization yields a higher log-likelihood than the EM algorithm. This difference is consistent with the approximation used in the EM implementation, which relies on a linear approximation of the matrix exponential. Overall, the results favour the two-regime model as a more parsimonious specification, while the three-regime model provides a slightly better fit to the training data.

Tables 2 and 3 report the estimated parameters and the associated one-day transition probabilities. In the two-regime model, although the baseline intensities are similar across regimes (22.96 versus 24.41), the excitation parameter differs by more than one order of magnitude. Consequently, the distinction between regimes is driven primarily by their contagion behaviour rather than by differences in spontaneous earthquake occurrence. Regime 2 corresponds to periods during which earthquakes generate substantially larger aftershock cascades, whereas regime 1 is closer to a weakly self-exciting process.

The three-regime specification provides a more nuanced characterization of seismic activity. The third regime exhibits almost no self-excitation (κ_3=0.0559), resulting in dynamics that are close to those of a Poisson process, where events occur largely independently of previous shocks. The second regime is characterized by a moderate level of self-excitation, corresponding to periods in which earthquakes generate limited aftershock activity. In contrast, the first regime displays the strongest self-exciting behaviour, indicating episodes of pronounced clustering during which earthquakes are more likely to trigger subsequent events. The three-regime model therefore distinguishes between background seismicity, moderately clustered activity, and highly contagious seismic episodes.

	2D	
A	State 1	State 2
State 1	-14.1676	14.1676
State 2	46.1159	-46.1159

	3D		
A	State 1	State 2	State 3
State 1	-50.0207	49.969	0.0517
State 2	17.7013	-18.4352	0.7339
State 3	2.3828	0.1467	-2.5295

	2D	
P(1 day)	State 1	State 2
State 1	0.9642	0.0358
State 2	0.1165	0.8835

	3D		
P(1 day)	State 1	State 2	State 3
State 1	0.8749	0.1248	0.0003
State 2	0.0442	0.9538	0.0020
State 3	0.0061	0.0008	0.9931

Table 3: Estimated matrix of instantaneous and 1-day transition probabilities, (2D) and three (3D) regime models.

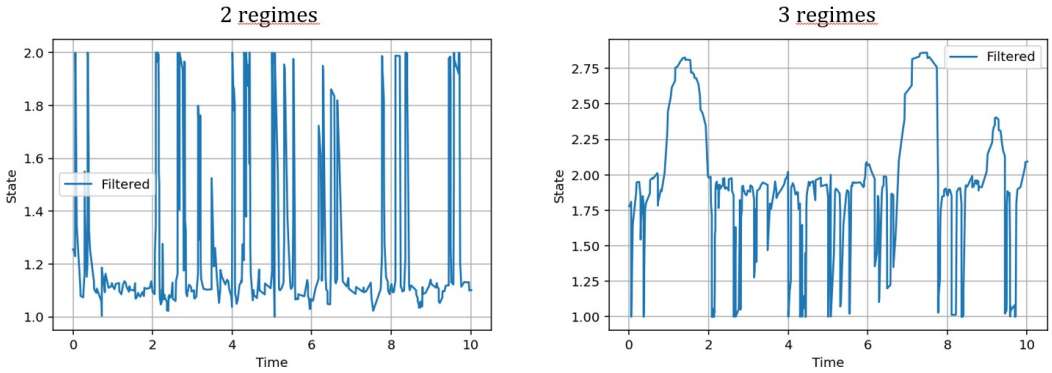


Figure 3: Filtered probabilities of regimes in the two- and three-regime models.

Figure 3 displays the smoothed probabilities associated with each regime, obtained using the forward-backward algorithm. For the two-regime model, the process spends most of the time in the first regime, while transitions to the second regime occur during episodes of heightened seismic activity. This result is consistent with the larger self-excitation parameter estimated for the second regime, suggesting that periods of intense earthquake clustering are captured by this state.

The three-regime model provides a more detailed description of the dynamics. The periods identified as regime 1 largely coincide with those assigned to the second regime in the two-state specification, indicating that both models detect the same episodes of elevated seismic activity. Regime 2 corresponds to an intermediate state, characterized by a lower level of self-excitation and representing moderately clustered seismic activity. In contrast, regime 3 can be interpreted as a background seismicity state, where earthquake occurrences are largely independent of previous events. Overall, the three-state model reveals that highly active periods can be separated from moderately clustered and nearly independent seismic episodes.

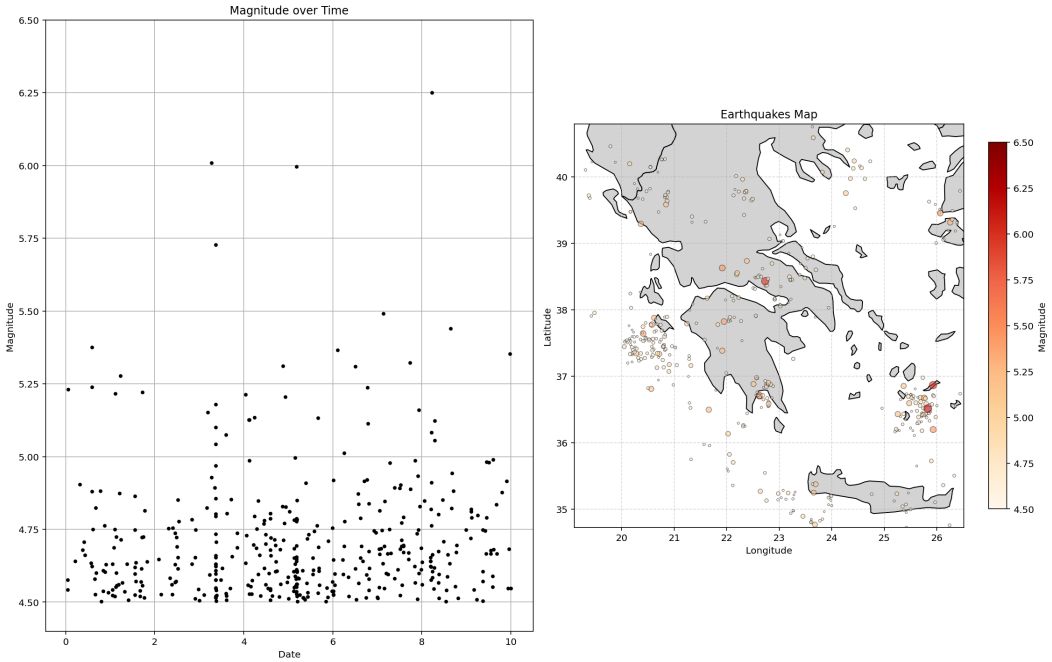


Figure 4: Left: time series of simulated earthquakes over ten years. Right: locations of simulated earthquakes. The magnitude is measured on the Richter scale.

10.2 Simulation

Figure 4 presents a sample path generated from the fitted two-regime model over a ten-year period. The simulated trajectory exhibits several characteristics observed in the real earthquake catalogue shown in Figure 1. In particular, the model reproduces alternating periods of relatively low activity and episodes of intense clustering, reflecting the switching behaviour of the latent Markov chain and the self-exciting nature of the process. The occurrence of bursts of events followed by quieter periods is consistent with the temporal patterns observed in the historical data, suggesting that the model captures the main mechanisms governing the evolution of seismic activity.

To evaluate the reliability of the Forward-Backward filter, Figure 5 displays both the simulated hidden Markov chain, updated at event times, and the associated filtered regime probabilities. The filtered probabilities closely track the true state sequence and correctly identify most regime switches. Discrepancies are primarily observed near transition periods, where uncertainty regarding the underlying state is naturally higher. The overall agreement between the simulated and inferred regimes demonstrates the ability of the filtering procedure to recover the latent state governing earthquake dynamics.

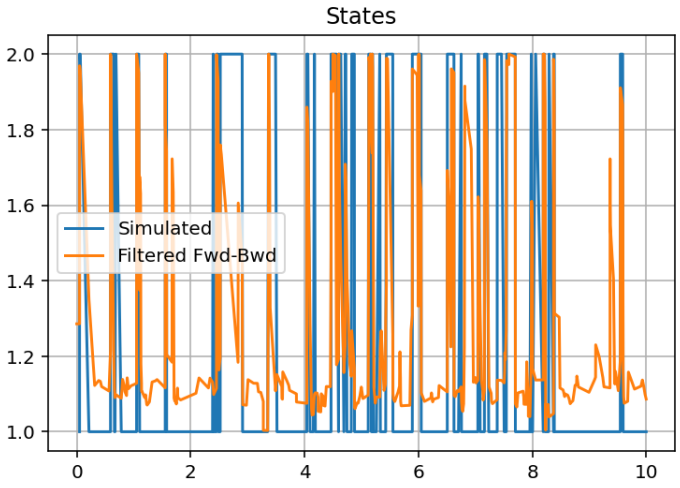


Figure 5: Simulated versus filtered states

Table 4 reports summary statistics computed from 1000 sample paths simulated over a one-year horizon under the one- (1D), two- (2D), and three-regime (3D) specifications. The average annual numbers of earthquakes generated by the 2D and 3D models are very similar, both being close to 39 events per year, whereas the 1D model produces a higher average of 48.26 events. These figures can be compared with the empirical averages observed over the validation period 2006-2016 (39.53 events per year) and the training period 2016-2026 (63.49 events per year). The simulated averages obtained from the 2D and 3D models closely match the validation sample, while remaining below the average observed during the calibration period. This discrepancy is not surprising, as the period 2016-2026 includes several episodes of unusually intense seismic activity, notably the heightened seismicity observed in 2020-2021 and the exceptional earthquake swarm associated with volcanic activity in 2025. Nevertheless, the upper tail of the simulated distributions remains compatible with such active periods, with the 95th and 99th percentiles of the annual event count lying around 70 and 100 earthquakes, respectively. These results indicate that all three models are capable of generating highly active seismic years comparable to those observed in the most recent decade, while preserving realistic long-run frequencies.

	1D	2D	3D
Average number of events/year	48.26	39.0	38.49
Standard deviation of events/year	19.47	17.76	18.96
95% percentile events/year	73.0	69.0	72.0
95% percentile events/year	95.01	100.03	106.03
Average of magnitudes	4.73	4.73	4.73
Standard deviation of magnitudes	0.24	0.24	0.24
95% percentile of magnitudes	5.20	5.21	5.21
99% percentile of magnitudes	5.61	5.65	5.64

Average number of events/year, training set	63.49
Average number of events/year, validation set	39.53
Average number of events/year, full dataset	51.51
Average of magnitudes, full dataset	4.76
Standard deviation of magnitudes, full dataset	0.33
95% percentile of magnitudes, full dataset	5.4
99% percentile of magnitudes, full dataset	6.17

Table 4: Statistics for 1000 simulated events, 1 year period, 1D, 2D, 3D regime models.

The simulated magnitude distributions are also in good agreement with the observed data. Across all specifications, the mean simulated magnitude is approximately 4.73, which is very close to the empirical average of 4.76 computed from the full 2006-2026 earthquake catalogue. The models therefore reproduce the central tendency of earthquake magnitudes accurately. However, the upper tail is somewhat underestimated, as the simulated 99th percentiles remain below the empirical value of 6.17 obtained from the 1032 observed earthquakes. This suggests that the magnitude component of the model captures typical events well but reproduces extreme earthquakes less frequently than observed in the data.

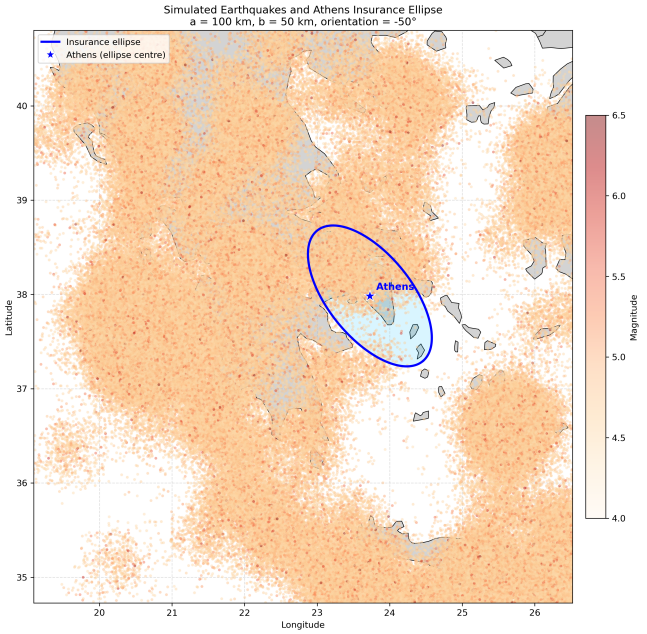


Figure 6: 10000 one-year scenarios.

10.3 Pricing

In this section, we use our model to evaluate the annual premium of the parametric insurance contract introduced in Section 8, indexed the magnitude of earthquakes occurring in an ellipse centered on Athens with parameters a =100 km, b= 50 km and $\theta = -50°$. The maximum payment is L =100 M€ and the lower and upper bounds of the function $g(m)$ defined by Equation (28) are $m_L = 5.0$ and $m_U = 6.0$. We price the insurance contract with 10000 one-year simulations with the one-, two- and three-regime models. We also compare with the spatial Pareto-Poisson model. Figure 6 shows the 389663 earthquakes simulated with the model with 2 states and the elliptic area that is considered in the insurance contract.

Full area	**Poisson**	**1D**	**2D**	**3D**
Number of eartquakes	635 081	481 320	389 663	384 513
Number of events in state 1	n.a.	481 320	215 641	164 149
Number of events in state 2	n.a.	n.a.	174 022	180 376
Number of events in state 3	n.a.	n.a.	n.a.	39988
Athens area	**Poisson**	**1D**	**2D**	**3D**
Number of events	13077	10281	7837	8076
Average magnitude	4.73	4.73	4.73	4.73
Maximum magnitude	7.75	7.51	7.32	7.82
Standard deviation of magnitudes	0.24	0.24	0.24	0.25
99.5% percentile of magnitudes	5.84	5.85	5.78	5.82
Pure premium	3.59	2.5	2.03	2.14
99.5% benefits	88.69	84.63	72.37	77.52
Gross premium ($\theta = 0.10$ and $\alpha = 99.5\%$)	12.10	10.72	9.06	9.68

Table 5: Statistics of pricing simulations and premiums.

Tables 5 presents the number of earthquakes simulated at the global level and in the elliptical area centered on Athens. As mentioned in the previous section, the two- and three- regime models yield on average less events that the one state process. This one regime model itself simulates on average less earthquakes that a pure Poisson model. We observe a good balance of earthquakes generated in each regime. We simulate 10281, 7837 and 8076 earthquakes in the ellipse with the one-, two- and three-regime models. The magnitude of simulated earthquakes in the Athens area reaches a maximum between 7.32 to 7.51. Nevertheless, events of such large amplitude are rare since the 99.5% percentile ranges from 5.748 to 5.85. The Poisson model leads to the highest pure and gross premium due to an overestimation of earthquake frequency in the elliptic area. Indeed, the Poisson model does not replicate the spatial contagion of earthquakes and then generates more randomly dispersed events than the spatial Hawkes process.

The highest pure premium, apart from the Poisson model, is obtained with the one regime model, 2.5 M€ per year, while the cheapest price of 2.03 M€ is yield by the 2D model. The gross premium, calculated with formula (29), a loading $\theta = 0.10$ and a 99.5% VaR ranges from 9.06M€ to 10.72M€, depending upon the model. These figures can be compared to the benefits that would have been paid if such a contract was subscribed over the period 2006-2026. During this time interval, 19 earthquakes were recorded in the area of Athens as shown in the left plot of Figure 7. The claim costs would have been of 40 M€ in 2014, 30M€ in 2015 and 30M€ in 2019, as illustrated in the right plot of the same Figure. The annual pure premium based on the 20 year history is equal to 4.76 M€ which is twice higher than the pure premium computed with the model but twice lower than the premium with a risk loading.

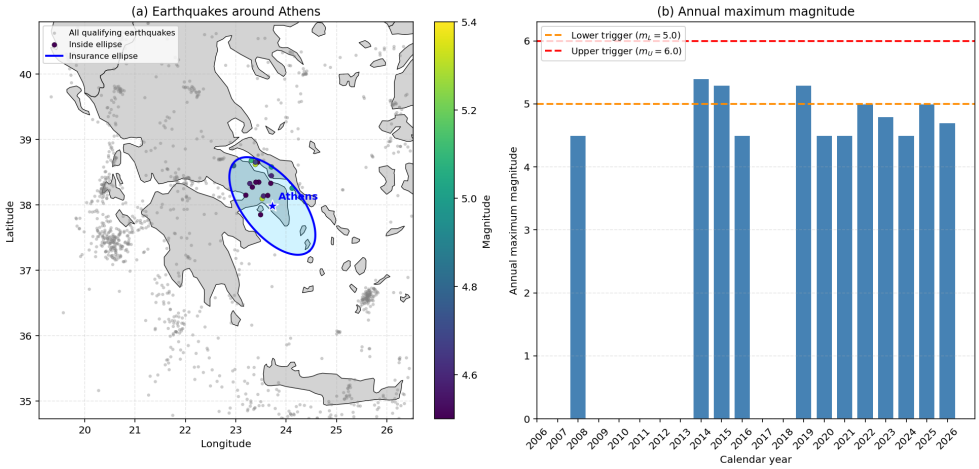


Figure 7: Left: location of the 19 earthquakes recorded in the Athens area between 2006 and 2026. Right: earthquakes that would trigger the parametric insurance coverage.

11 Conclusions

This paper introduces a Markov-modulated marked spatial Hawkes process that combines temporal self-excitation, spatial dependence, event-specific marks and latent regime switching within a tractable statistical framework. By allowing the hidden regime to govern the background intensity, triggering mechanism and mark distribution, the model accommodates changes in seismic activity while preserving a likelihood structure suitable for inference. The derivation of the likelihood, together with the forward–backward filtering procedure and the expectation–maximization algorithm, provides a practical approach to estimating the model when the regimes are unobserved. Furthermore, the stability condition and simulation procedure establish the basis for analysing the process and applying it to catastrophe risk assessment.

The application to Greek earthquake data illustrates the relevance of this framework for representing heterogeneous seismic dynamics. The estimated regimes distinguish periods of intense earthquake clustering from moderately active and relatively quiet periods, while the simulated trajectories reproduce alternating episodes of high and low seismic activity. The two-regime specification provides a parsimonious representation of these dynamics, whereas the three-regime model offers a more detailed characterization of seismic activity. The simulations also reproduce the central tendency of earthquake magnitudes, with an average close to that observed in the data. Nevertheless, the upper tail of the magnitude distribution remains slightly underestimated, highlighting a limitation in the representation of extreme events with a Pareto distributed magnitude.

The insurance application further demonstrates how the proposed model can connect seismic risk assessment with the pricing of geographically defined parametric contracts. For the elliptical region centered on Athens, the simulated annual pure premiums range from €2.03 million to €2.50 million across the three specifications, while the corresponding gross premiums range from €9.06 million to €10.72 million under the considered risk-loading rule. These differences show that the choice of model specification can materially affect the assessment of insurance costs. By jointly modelling event frequency, spatial concentration, magnitude and temporal clustering, the framework provides a coherent basis for evaluating regional trigger probabilities and the distribution of contractual payments. However, the intrinsic basis risk of parametric insurance call for caution when interpreting the resulting premiums.

Several extensions could improve the framework and broaden its actuarial applications. In particular, more flexible magnitude distributions or tail-modelling techniques could improve the representation of rare, severe earthquakes. Further developments could also incorporate parameter and model uncertainty more explicitly, and investigate the sensitivity of premiums to the insured region. Overall, this study provides a tractable link between regime-switching spatial point-process modelling and parametric catastrophe insurance, with potential applications to other risks characterised by clustered events, changing risk conditions and geographically dependent losses.

Data availability statements

The data and code that support the findings of this study are available from the corresponding author, upon reasonable request.

Statements and declarations

Authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

References

[1] Asmussen, S. 1989. Risk theory in a Markovian environment. Scandinavian Actuarial Journal, 1989(2), 69–100.

[2] Bayliss C., Guidotti R., Estrada-Moreno A., Franco G., Juan A.A. 2020. A biased-randomized algorithm for optimizing efficiency in parametric earthquake (Re) insurance solutions. Computers & Operations Research, 123, 105033

[3] Biffis E., Chavez E., Louaas A., Picard P., 2022. Parametric insurance and technology adoption in developing countries. Geneva Risk and Insurance Review, 47, 7-44

[4] Cummins J. D., Mahul O., 2009. Catastrophe risk financing in developing countries: Principles for Public Intervention. World Bank. ISBN 978082137739

[5] Clement K.Y., Botzen W.J., Brouwer R., Aerts J.C. 2018. A global review of the impact of basis risk on the functioning of and demand for index insurance. International Journal for Disaster Management & Risk Reduction, 28, 845-853

[6] Elliott R. J., Chen Z., Duan Q., 2009. Insurance claims modulated by a hidden Brownian marked point process. Insurance: Mathematics and Economics, 45(2), 163–172

[7] Jang J., Dassios A. 2013. A bivariate shot noise self-exciting process for insurance. Insurance: Mathematics & Economics, 53 (3), 524-532

[8] Fox E.W., Schoenberg F.P., Gordon J.S. 2016. Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric hawkes point process models of earthquake occurrences. The Annals of Applied Statistics, 10 (3), 1725-1756

[9] Gu Z., Li Y., Zhang M., Liu Y. 2023. Modelling economic losses from earthquakes using regression forests: Application to parametric insurance. Economic Modelling, 125, 106350

[10] Hainaut D. 2012. Multidimensional Lee–Carter model with switching mortality processes. Insurance: Mathematics & Economics, 50, 236-246

[11] Hainaut D. 2016. Impact of volatility clustering on equity indexed annuities. Insurance: Mathematics & Economics, 71, 367-381

[12] Hainaut D. 2017. Contagion modeling between the financial and insurance markets with time changed processes. Insurance: Mathematics & Economics, 74, 63-77

[13] Hainaut D., Deelstra G. 2019. A self-exciting switching jump diffusion: properties, calibration and hitting time. Quantitative Finance, 19, 407-426

[14] Hainaut D., Moraux F., 2019. A switching self-exciting jump diffusion process for stock prices. Annals of Finance, 15, 267-306

[15] Hainaut D. 2022. Continuous time processes for finance. Switching, Self-exciting, fractional and other recent dynamics. Springer & Bocconi Series in mathematics, statistics, finance and Economics. Springer Nature Switzerland AG

[16] Hawkes A., 1971(a). Point sprectra of some mutually exciting point processes. Journal of the Royal Statistical Society Series B, 33, 438-443

[17] Hawkes A., 1971(b). Spectra of some self-exciting and mutually exciting point processes. Biometrika 58, 8390

[18] Jun M., Cook S., 2024. Flexible multivariate spatio-temporal hawkes process models of terrorism. The Annals of Applied Statistics 18 (2), 1378-1403Kwon J., Zheng Y., and Jun M., 2023. Flexible spatio-temporal Hawkes process models for earthquake occurrences. Spatial Statistics, 54, 100728

[19] Louaas A., Picard P. 2026. On the design of optimal parametric insurance. The Geneva Risk and Insurance Review. 51, 173-200

[20] Mohler G.O., Short M.B., Brantingham P.J., Schoenberg F.P., Tita G.E., 2011. Self-exciting point process modeling of crime Journal of the American Statistical Association, 106 (493), 100-108

[21] Ng, A. C. Y., & Yang, H. 2006. On the joint distribution of surplus before and after ruin under a Markovian regime switching model. Stochastic Processes and their Applications, 116 (2), 244–266

[22] Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes. Journal of the American Statistical Association, 83(401), 9-27

[23] Ogata, Y., 1998. Space-time point-process models for earthquake occurrences. Annals of the Institute of Statistical Mathematics 50 (2), 379-402

[24] Pai J., Li Y., Yang A., Li C., 2022. Earthquake parametric insurance with Bayesian spatial quantile regression. Insurance: Mathematics and Economics, 106, 1-12

[25] Radu N., Alexandru F. 2022. Parametric insurance - A possible and necessary solution to insure the earthquake risk of Romania. Risk, 10 (3), 59

[26] Shi Y., Punzo A., Otneim H., Maruotti A. 2025. Hidden semi-Markov models for rainfall-related insurance claims. Insurance: Mathematics and Economics, 120, 91-106

[27] Swishchuk A., Zagst R., Zeller G. 2021. Hawkes processes in insurance: Risk model, application to empirical data and optimal investment. Insurance: Mathematics and Economics, 101, 107-124

[28] Yuan B., Li H., Bertozzi A.L., Brantingham P.J., Porter M.A., 2019. Multivariate spatiotemporal hawkes processes and network reconstruction. SIAM Journal on Mathematics of Data Science, 1 (2), 356-382

[29] Zhuang J., Ogata Y., Vere-Jones, D. 2002. Stochastic declustering of space-time earthquake occurrences, Journal of the American Statistical Association, 97 (458), 369-3

[30] Zhu S., Xie Y., 2022. Spatiotemporal-textual point processes for crime linkage detection. The Annals of Applied Statistics, 16 (2), 1151-1170.

[31] Wei J., Yang H., Wang R., 2010. Optimal reinsurance and dividend strategies under the Markov-modulated insurance risk model. Stochastic Analysis and Applications, 28(6), 1078–1105