arXiv:2503.15821v2 [stat.AP] 27 Aug 2025

Temporal Point Process Modeling of Aggressive Behavior Onset in Psychiatric Inpatient Youths with Autism

Michael Potter[1,*], Michael Everett[1], Ashutosh Singh[1], Georgios Stratis[1], Yuna Watanabe[2], Ahmet Demirkaya[1], Deniz Erdogmus[1], Tales Imbiriba[3,+], and Matthew S. Goodwin[2,+]

[1]Northeastern University, Electrical and Computer Engineering, Boston, 02115, USA
[2]Northeastern University, Bouvé College of Health Sciences and Khoury College of Computer Sciences, Boston, 02115, USA
[3]University of Massachusetts - Boston, Computer Science, Boston, 02125, USA
[*]potter.mi@northeastern.edu
[+]these authors contributed equally to this work

ABSTRACT

Aggressive behavior, including aggression towards others and self-injury, occurs in up to 80% of children and adolescents with autism, making it a leading cause of behavioral health referrals and a major driver of healthcare costs. Predicting when autistic youth will exhibit aggression can be challenging due to their communication difficulties. Many are minimally verbal or have poor emotional insight. Recent advances in Machine Learning and wearable biosensing demonstrate the ability to predict aggression within a limited future window (typically one to three minutes) in autistic individuals. However, existing works don't estimate aggression onset probability or the expected number of aggression onsets over longer periods, nor do they provide interpretable insights into onset dynamics. To address these limitations, we apply Temporal Point Processes (TPPs) – particularly self-exciting Hawkes processes – to model the timing of aggressive behavior onsets in psychiatric inpatient autistic youth. We benchmark several TPP models by evaluating their goodness-of-fit and predictive metrics. Our results demonstrate that self-exciting TPPs more accurately captures the irregular and clustered nature of aggression onsets, especially compared to traditional Poisson models. These incipient findings suggest that TPPs can provide interpretable, probabilistic forecasts of aggression onset along a time continuum, supporting future clinical decision-making and preemptive intervention.

Introduction

Autism is one of the most prevalent childhood disorders, affecting approximately 1 in 36 children[1]. A significant proportion (up to 80%) of autistic children and adolescents exhibit aggressive behaviors, including Self-Injurious Behavior (SIB), tantrums, meltdowns, property destruction, and Aggression Towards Others (ATO).[2–4]. These behaviors are among the leading causes of referral to behavioral health services[5] and contribute substantially to healthcare costs[6]. Many autistic youth struggle with emotional regulation and self-reporting of their internal states[7]. Thirty to forty percent are minimally verbal, and others experience difficulties with emotional insight and self-awareness[8]. As a result, aggressive behaviors appear unpredictable, occur at irregular, seemingly random times, and thus create barriers to accessing community resources, therapy, education, and clinical services. Families caring for children with autism often face increased stress, social isolation, and financial strain due to concerns about unexpected aggressive behaviors in different environments[9,10]. In addition, these behaviors can negatively affect support professionals, leading to increased compensation for work-related injuries, higher absenteeism, and staff turnover[11]. The cumulative effect of these challenges can demoralize caregivers and clinicians, disrupt patient care trajectories, and, in severe cases, require home-bound or residential placement, reducing quality of life while increasing costs.

Temporal Point Process (TPP) models capture irregular stochastic event times in various domains. TPP methods are widely applied in seismology[12], finance[13], and epidemiology[14]. In seismology, TPPs model earthquake occurrences by capturing the stochastic nature of event times and locations while accounting for aftershock sequences and spatial clustering to estimate the probability of future seismic activity. In finance, they capture the irregular timing of discrete price changes, trade executions, and market orders while accounting for serial dependence of events, offering insight into market stability and volatility. In epidemiology, TPPs model the spread of disease by treating infections as stochastic events that trigger secondary infections, helping to estimate transmission rate, predict outbreaks, and evaluate interventions.

Despite their broad applicability, TPPs have not yet been used to model aggressive behavior onsets in inpatient youths with autism. These onsets exhibit irregular timing and temporal clustering, where bursts of aggressive behavior occur within short intervals. Compared to point-estimate classifier methods, TPPs offer several key advantages. They generate probabilistic forecasts by sampling entire trajectories, account for irregular inter-onset intervals, and support causal inference and knowledge

discovery[15]. Applying TPP to aggressive behavior onsets in inpatient youths with autism allows us to address questions such as "How many aggressive behavior onsets will occur between 5–10 minutes of an observation session?" and "Do inpatient youths with autism exhibit self-excitation in aggressive behavior onsets?"[16].

Several studies have explored automated aggressive behavior detection in youths with autism using time-series data. Convolutional Neural Network (CNN)-based models analyzing video streams[17] have linked the likelihood of aggressive behavior to hand movement speed, while hybrid CNN-Long Short Term Memory (LSTM) models leveraging physiological signals from wearable biosensors[18] have been used to detect meltdowns and tantrums. While these methods improve aggressive behavior detection in autistic youth, they do not improve safety for clinicians, caregivers, or patients. Predicting when a subsequent aggressive behavior will occur before it happens could enable just-in-time preventive interventions to mitigate harm.

Recent advances in Machine Learning (ML) and Deep Learning (DL) demonstrate promising capabilities predicting imminent aggressive behavior in inpatient youths with autism. Studies have employed Logistic Regression (LR), Support Vector Machine (SVM), and Neural Networks (NNs) to forecast aggressive behavior one to three minutes into the future using three minutes of prior real-time peripheral physiological data from wearable biosensors[19–21]. While the results are promising, these methods do not estimate the probability of aggressive behavior onsets over longer future time windows and only utilize partial histories of physiological data and behavior onsets during prediction. Forecasting larger future time windows requires modeling the physiological data autoregressively. Additionally, using only three minutes of prior physiological data and onset information disregards potentially valuable historical data. Enabling long-term and adaptable forecasting of aggressive behaviors could improve resource allocation and intervention planning for clinicians and caregivers who support autistic youths.

This paper uses TPPs to forecast the number of proximal aggressive behavior events over longer periods than currently explored in the literature. Unlike discrete-time representations commonly used in time series applications, TPP modeling handles time between aggressive behaviors as random variables. This approach enables precise temporal modeling without requiring a fixed time window to aggregate events, which can introduce discretization errors[22]. Prior work has applied a discretized Non-Homogeneous Poisson Process (NHPP) model to analyze skin conductance responses alongside observed regulatory behaviors, identifying statistical differences in the physiological states of autistic youth[23]. The study examined self, co-, and combined regulatory behaviors to assess dyadic physiological regulation patterns. However, it focuses on statistical hypothesis testing relating to observed interpersonal regulatory behaviors rather than modeling or predicting future behaviors.

The present work applies TPPs to model and predict aggressive behavior onsets in inpatient youths with autism. We evaluate modeling assumptions using Goodness-of-Fit (GOF) statistics such as Quantile-Quantile (QQ) plots, raw residual plots, and Wassertein Distance (WD). Model performance is assessed through predictive performance metrics such as Pareto Smoothed Importance Sampling - Leave One Out (PSIS-LOO) Expected Log Density (ELPD) and Mean Absolute Percent Error (MAPE). Additionally, we interpret TPP model parameters for knowledge discovery, focusing on the branching factor of the Hawkes Point Process (HawkesPP) to understand the degree of self-excitation. This foundational study analyses different TPP strategies, including homogeneous, non-homogeneous, and self-exciting TPPs, demonstrating the potential of TPPs in addressing complex questions about aggressive behavior onset in inpatient youths with autism.

Methods

Data and Participants

Sites and study protocol: This study is a secondary analysis of data reported in[19] acquired from psychiatric inpatients serially enrolled at clinical inpatient sites (Bradley Hospital, Providence, Rhode Island; Cincinnati Children's Hospital, Cincinnati, Ohio; Western Psychiatric Hospital, Pittsburgh, Pennsylvania; and Spring Harbor Hospital, Portland, Maine) participating in the Autism Inpatient Collection (AIC). This prognostic study was designed to estimate and validate a predictive model and is reported following the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines. The Institutional Review Board (IRB) at each participating site reviewed and approved the study, including the AIC protocols and aggressive behavior prediction protocols. The sites and their IRBs are: Bradley Hospital (Providence, RI), Cincinnati Children's Hospital (Cincinnati, OH), Spring Harbor Hospital (Portland, ME), and Western Psychiatric Hospital (Pittsburgh, PA). IRB approval of the AIC extended to this study with an amendment. Guardians of all study participants provided informed consent and were remunerated. The IRB approved this retrospective study in compliance with the Health Information Portability and Accountability Act. All methods were performed in accordance with relevant guidelines and regulations following the Declaration of Helsinki.

Inclusion/Exclusion criteria: In total, 86 inpatients were enrolled. Inclusion criteria included confirmation of autism via research-reliable administration of the Autism Diagnostic Observation Schedule-2 (ADOS-2) (see[19]) and parent-reported, staff-reported, or staff-observed physical aggression or self-injurious behavior. Exclusion criteria included not having a parent proficient in English or the individual with autism having prisoner status. Of the 86 participants enrolled in the study, 16 (18%) were excluded from the analysis: 8 were unable to wear the physiological biosensor, and 8 were discharged before any

behavioral observation could occur. The final analytic sample consisted of 70 participants.

Participant statistics: Participants ranged in age from 5 to 19 years (M = 11.85, SD = 3.5), and were predominantly male (88%), white (90%), and non-Hispanic (92%)[19]. Nearly half of the participants (46%) were minimally verbal, as determined by ADOS-2 Module 1 or 2, and 57% met criteria for intellectual disability based on Leiter-3 global IQ scores (M = 72.96, SD = 26.12). The length of inpatient stay varied from 8 to 201 days (M = 37.28, SD = 33.95).

Observational procedures and behavioral coding: Research staff performed observational coding in the inpatient units. At the same time, inpatient study participants with autism wore the commercially available and regulatory-compliant E4 biosensor by Empatica, Inc. on their non-dominant wrist. In the present study, we focus solely on annotations of operationally defined (see[19]) aggressive behavior (i.e., SIB, Emotion Dysregulation (ED), ATO) episode's start (onset) and stop (offset) times within observation periods using a custom mobile application.

A total of 429 behavioral observation sessions were conducted between March 2019 and March 2020, with a median of 5 sessions per participant (IQR = 6). These sessions totaled approximately 497 hours of observation time (Median = 4.4 hours, IQR = 4.9 hours per session).

Across these sessions, 6,665 aggressive behavior episodes were annotated:

- **Self-injurious behavior (SIB)**: 3,983 episodes (60%); Median = 2 per participant (IQR = 23); Median duration = 1.97 sec (IQR = 4.25 sec)
- **Emotional Dysregulation (ED)**: 2,063 episodes (31%); Median = 8 per participant (IQR = 27); Median duration = 10.09 sec (IQR = 20.08 sec)
- **Aggression towards others (ATO)**: 619 episodes (9%); Median = 1 per participant (IQR = 8); Median duration = 2.31 sec (IQR = 3.75 sec)

To assess inter-rater reliability, 20% of the total dataset was randomly selected and independently double-coded by two trained research staff members at each inpatient site. Agreement was high across all categories: SIB ($\kappa = .93$), ED ($\kappa = .95$), and ATO ($\kappa = .86$).

Additional details on the dataset are available in[19] and summarized in Supplementary Table S7.

Temporal Point Processes

TPPs provide a principled statistical framework for modeling the generative process of events over continuous time. In this study, we leverage TPPs to model the timing of aggressive behavior onsets in autistic youth. A TPP is a stochastic process that represents the occurrence of discrete events (in our case, aggressive behavior onsets) over a fixed time window $[0,T]$[24]. A session, or realization, of a TPP, is an ordered sequence of aggressive behavior onset times, denoted as $S_i = \{t_1^{(i)}, t_2^{(i)}, \ldots, t_{J_i}^{(i)}\}$, where J_i is the number of onsets in session i, and $t_j^{(i)}$ represents the time elapsed since session i's start when the j-th aggressive behavior onset occurred. Equivalently, a TPP can be formulated as a counting process $N(t)$, representing the cumulative number of aggressive behavior onsets up to time t in a given observation session. Formally, we define this as:

$$N(t) = \max\{j : t_j \leq t, t_j \in S\}, \tag{1}$$

where $N(t)$ tracks the number of onsets within $[0,t]$. For a realization $S = \{t_1, t_2, \ldots, t_J\}$ of a TPP, the history up to (but not including) time t is $H_t = \{t_j \in S \mid t_j < t\}$. When an event at time t_j is included in the history we write $H_{t_j^+} = \{t_1, \ldots, t_j\}$, and when it is excluded we write $H_{t_j^-} = \{t_1, \ldots, t_{j-1}\}$.

A TPP is completely characterized by the conditional intensity function $\lambda(t \mid H_t)$: it defines the cumulative intensity, the Probability Density Function (PDF) and Cumulative Density Function (CDF) of the next event time, and the likelihood function of observed events. The conditional intensity function gives the instantaneous expected rate of events at time t given the past history. Formally, for $t > t_j$ in an interval following the last observed event,

$$\lambda(t \mid H_{t_j^+}) = \lim_{\Delta t \to 0} \frac{\mathbb{E}\left[N(t+\Delta t) - N(t) \mid H_{t_j^+}\right]}{\Delta t}. \tag{2}$$

Equivalently, in terms of the conditional density $f(\cdot \mid H_{t_j^+})$ and distribution $F(\cdot \mid H_{t_j^+})$ of the next event time,

$$\lambda(t \mid H_{t_j^+}) = \frac{f(t \mid H_{t_j^+})}{1 - F(t \mid H_{t_j^+})} \qquad \text{for } t > t_j. \tag{3}$$

Intuitively, for an infinitesimal interval Δt, $\lambda(t \mid H_t)\Delta t$ is the approximate probability (or expected count) of an event occurring in $[t, t+\Delta t)$ conditional on the observed history up to t.

The conditional cumulative intensity function (Equation (4)) is the integral of the conditional intensity function and represents the expected number of aggressive behavior onsets within the time window $(t_j, t]$, conditioned on the history of onsets up to (and including) t_j, where the last onset occurred at t_j. The cumulative intensity function (Equation (5)) is the average number of aggressive behavior onsets that occur within the observation session $[0, t]$.

$$\Lambda(t|H_{t_j^+}) = \int_{t_j}^{t} \lambda(s|H_s)ds \quad (4) \qquad \Lambda(t) = \int_{0}^{t} \lambda(s|H_s)ds \quad (5)$$

The conditional PDF (Equation (6)) and conditional CDF (Equation (7)) of an aggressive behavior onset occurring at and before time $t > t_j$, respectively, given the past history of aggressive behavior onsets within an observation session, is described by the conditional intensity function:

$$f(t|H_{t_j^+}) = \lambda(t|H_{t_j^+}) \exp\left(-\int_{t_j}^{t} \lambda(s|H_s)ds\right) \quad (6) \qquad F(t|H_{t_j^+}) = 1 - \exp\left(-\int_{t_j}^{t} \lambda(s|H_s)ds\right), \quad (7)$$

The conditional PDF and CDF map the conditional intensity—often hard to interpret for decision making—into familiar probability quantities (a density and a cumulative probability), which are the standard inputs for decision-making frameworks[25]. To simplify notation, we use $*$ to denote the conditional intensity function, conditional cumulative function, conditional CDF, and conditional PDF conditioned on the history of aggressive behavior onsets H_t, as $\lambda^*(t) = \lambda(t|H_t)$, $\Lambda^*(t) = \int_0^t \lambda^*(s)ds$, $f^*(t) = f(t|H_t)$, and $F^*(t) = F(t|H_t)$ respectively.

Throughout this study we benchmark the following parametric TPPs: the homogeneous Poisson process (Homogeneous Poisson Process (HPP)), a non-homogeneous Poisson process with a power-law intensity (NHPP), and several self-exciting Hawkes variants (exponential, power-law, and multi-exponential kernels). Specifying a parametric conditional intensity yields closed-form expressions for the conditional pdf, cdf, and cumulative intensity, and concentrates estimation on a compact parameter vector θ. Moreover, parametric TPPs are data efficient unlike non-parametric TPPs. We emphasize self-exciting models because they naturally capture temporal clustering of onsets, and are consistent with clinical theories that rising distress and physiological arousal can trigger cascades of aggressive behavior in autistic youth[26–28].

Data Preprocessing

We treat SIB, ED, and ATO as three recorded TPPs but collapse them into a single event class for modeling (see Fig. 1a). This is mathematically equivalent to the superposition of the behavior TPPs. At the 250 ms annotation resolution, we define an *isolated event* as an episode consisting of a single positive annotation sample (i.e., episode length = 1 sample), whereas a *continuous episode* is a run of consecutive positive samples. For analysis we reduce each continuous episode to its onset (the left boundary) and model only onset times and isolated events (that is, episodes of length one sample) (Fig. 1b).

Onset timestamps were recorded in calendar time and are normalized per session by subtracting the session start time so each realization begins at 0 (Fig. 1c). Observation sessions are treated as independent realizations; we do not concatenate sessions from the same participant because the unobserved gaps between sessions would bias parameter estimates and exacerbate edge effects[13].

Likelihood Function

The dataset comprises I observed sessions $\mathbb{S} = \{S_i\}_{i=1}^{I}$ with corresponding observation lengths $\mathbb{T} = \{T_i\}_{i=1}^{I}$. For a given session S_i, the likelihood function is a joint density function of all aggressive behavior onsets observed within the observation window $[0, T_i)$, factorized into conditional densities of each aggressive behavior onset given all the aggressive behavior onsets prior to it[29]:

$$L(\theta \mid S_i, T_i) = \left(1 - F(T_i|H_{T_i})\right) \prod_{j=1}^{J_i} f(t_j \mid H_{t_j}) \quad (8)$$

where the last term $(1 - F(T_i|H_{t_{T_i}}))$ corresponds to the probability that no aggressive behavior onsets are observed between $[t_{J_i}, T_i)$. Using Equation (7) and Equation (6), Equation (8) becomes

$$L(\theta|S_i, T_i) = \left(\prod_{j=1}^{J_i} \lambda_\theta^*(t_j)\right) \exp\left(-\int_0^{T_i} \lambda_\theta^*(s)ds\right) \qquad t_0 = 0. \quad (9)$$

Then, the joint likelihood function of all observed sessions (and therefore all aggressive behavior onsets), assuming independence between the observation sessions, is

$$L(\theta \mid \mathbb{S}, \mathbb{T}) = \prod_{i=1}^{I} L(\theta | S_i, T_i) \tag{10}$$

In the following sections, we explore different choices for the conditional intensity function and their implications for aggressive behavior onset analysis.

Intensity Function Parameterizations

Unlike typical ML and DL models, we employ several parametric TPPs that are capable of modeling complex random sequences of aggressive behavior onsets while remaining interpretable to clinicians. We aim to assess the feasibility of modeling and forecasting aggressive behavior onsets using only the history of previous onsets within an observation session. Since our focus is on event timing rather than additional attributes, we consider unmarked TPPs, which rely solely on onset timestamps. For this study, we exclude physiological signals from our analysis, as work in progress using Shapley values[30] and stepwise regression[31] suggests that the timing of aggression onset and offset is a key predictor of future aggression. This aligns with findings by Imbiriba et al[19] (see Fig. 1A–C in[19]), where Area Under the Receiver Operating Characteristic Curve (ROC-AUC) scores with physiological features alone are as low as 55, but exceed 80 when the most recent previous aggression onset is included as a covariate. Future work will incorporate physiological signals and contextual variables.

Homogeneous Poisson Process

The intensity function remains constant over time, meaning it does not change throughout the observation session. Additionally, the HPP assumes independence from past aggressive behavior onsets, implying that the timestamps of future aggressive behavior onsets are independent of previous onsets. Therefore, for a HPP , the aggressive behavior onsets within an observation session of length T are distributed uniformly at random over the interval $[0, T]$, where the number of aggressive behavior onsets in the session follows a Poisson distribution with rate μ[32]. The HPP is described by the single parameter, μ, giving the conditional intensity and cumulative conditional intensity functions

$$\lambda^*(t) = \mu \tag{11}$$

$$\Lambda^*(t) = \mu t \tag{12}$$

where $\theta = [\mu]$, $\theta \geq 0$, and μ is the average number of aggressive behavior onsets per unit time.

Non-Homogeneous Poisson Process

The NHPP has many parameterizations, but we choose the Power Law (PL) characterization to describe conditional intensity and conditional cumulative intensity functions. The conditional intensity function is independent of past aggressive behavior onsets, but it varies over time throughout the observation session. We choose the PL parameterization because it is flexible enough to model a conditional intensity function which is either increasing, decreasing, or remaining constant over time.

$$\lambda^*(t) = \alpha k t^{k-1} \tag{13}$$

$$\Lambda^*(t) = \alpha t^{-k} \tag{14}$$

where $\theta = [\alpha, k]$, $\theta \geq 0$, α is the scale parameter and k is the shape parameter. The parameter k controls the shape of the conditional intensity function: when $k > 1$, the intensity increases over time; when $k < 1$, the intensity decreases; and when $k = 1$, the intensity is constant, which reduces the model to the HPP. The parameter α scales the overall conditional intensity function, affecting the steepness of the intensity decay or growth over time.

Self-Exciting Point Processes: Hawkes Process

The HawkesPP is a self-exciting point process that removes the independence assumptions of traditional models mentioned previously by allowing future aggressive behavior onsets to depend on past occurrences within an observation session. This dependence results in a stochastic process, where both event times and intensity function are evolving random variables conditioned on the history of previous onsets[33].

The conditional intensity function of a HawkesPP is defined as:

$$\lambda^*(t) = \underbrace{\mu}_{\text{baseline intensity}} + \underbrace{\sum_{t_j < t} \phi(t - t_j)}_{\text{excitation trigger}}, \tag{15}$$

where μ represents baseline intensity, ensuring events occur even in the absence of prior onsets, and the summation term captures the self-exciting nature of the process. Another way to interpret this is that μ models the exogenous onset occurrences,

while the triggering kernel $\phi(t)$ captures endogenous onsets. Each past event t_j contributes to the intensity at time t through the triggering kernel $\phi(t-t_j)$, which is nonnegative and causal, meaning it only influences future events ($\phi(t) \geq 0$ for $t \geq 0$, and $\phi(t) = 0$ for $t < 0$). Typically, $\phi(\cdot)$ is chosen to be a monotonically decreasing function, ensuring the influence of past events diminishes over time. This formulation models the cascading effect of aggressive behavior, where each onset increases the likelihood of subsequent onsets, with this effect gradually decaying. As a result, the HawkesPP captures the temporal clustering of aggressive behavior onsets, making it a suitable framework for forecasting onset occurrences based on historical patterns.

We evaluate two kernel functions of a HawkesPP: the Omori-Law (or PL) kernel $\phi(t) = k(c+t)^{-p}$ and the exponential kernel $\phi(t) = \alpha\beta\exp(-\beta t)$. The Hawkes Point Process with Exponential Kernel (HawkesExpPP) is Markovian, enabling the conditional intensity at a given onset to be computed from the conditional intensity evaluated at the previous onset and the current onset.It is suitable for processes with rapid self-excitation decay, where the process quickly "forgets" past events. The parameter β governs the delay density $\beta\exp(-\beta t)$, determining the time between excitations. A larger β leads to faster decay of influence from previous onsets and shorter inter-event times between onsets, while a smaller β results in slower decay and longer influence of past onsets[34]. The parameter α is known as the infectivity factor, which is the average number of new aggressive behavior onsets triggered by any given onset occurrence[34].

$$\lambda^*(t) = \mu + \alpha\beta \sum_{t_j<t} \exp\left(-\beta(t-t_j)\right) \tag{16}$$

$$\Lambda^*(t) = \mu t + \alpha \sum_{t_j<t} \left(1 - \exp\left(-\beta(t-t_j)\right)\right) \tag{17}$$

where $\theta = [\mu, \alpha, \beta]$ and $\theta \geq 0$.

In contrast, the Hawkes Point Process with Power Law Kernel (HawkesPLPP) is used when self-excitation follows a heavy-tailed distribution, making it suitable for processes with long-term triggering effects, such as epidemics or earthquakes. Being non-Markovian, it requires more computation but exhibits scale-free behavior, effectively modeling long-range temporal dependencies and varying event frequencies. The parameter p controls the decay rate of self-excitation, with slower decay allowing past events to influence future occurrences over extended periods. The time constant c regularizes the behavior of the power law kernel at very short times[13].

$$\lambda^*(t) = \mu + k \sum_{t_j<t} (c+t-t_j)^{-p} \tag{18}$$

$$\Lambda^*(t) = \mu t + \sum_{t_j<t} \frac{k}{p-1} \left(c^{1-p} - (c+t-t_j)^{1-p}\right) \tag{19}$$

where $\theta = [\mu, k, c, p]$ and $[\mu, k, c] \geq 0, p > 1$.

The Omori-Law kernel can be approximated by a weighted sum of exponential kernels, combining the Markovian computation of the conditional intensity from the exponential kernel with the scale-free behavior of the Omori-Law kernel. We use a Hawkes Point Process with Two Exponential Kernels (Hawkes2ExpPP) model to capture multiple temporal patterns, such as high- and low-frequency aggressive behavior onsets, which a single exponential kernel HawkesPP struggles to represent. By incorporating multiple kernels with different temporal scales, the multi-kernel HawkesPP models distinct behavior patterns, such as rapid bursts or sporadic aggression, improving onset prediction across timescales[35]. Adding more than two exponential kernels yields only marginal gains while increasing model complexity.

$$\lambda^*(t) = \mu + \sum_{k=1}^{2} \alpha_k\beta_k \sum_{t_j<t} \exp\left(-\beta_k(t-t_j)\right) \tag{20}$$

$$\Lambda^*(t) = \mu t + \sum_{k=1}^{2} \alpha_k \sum_{t_j<t} \left(1 - \exp\left(-\beta_k(t-t_j)\right)\right) \tag{21}$$

where $\theta = [\mu, \alpha_1, \beta_1, \alpha_2, \beta_2]$ and $\theta \geq 0$.

In our dataset, we observe different intensity functions, as shown in Fig. 2.

We acknowledge that biases inherent to the data collection process, such as "edge effects" from finite observation session durations, are unavoidable and may impact model parameter inference. Edge effects in TPPs arise when events before the observation window remain unobserved, leading to incomplete event histories. In HawkesPPs, this can result in underestimating self-excitation parameters and overestimating baseline intensity, as future events depend on past occurrences[13]. However, if past events have only short-term influence – in keeping with our hypothesis that aggressive behavior onsets influence proximal events – edge effects have minimal impact on parameter inference because the triggering kernel decays rapidly, making transient periods short. In preliminary analyses, we performed parameter inference using a modified exponentially decaying HawkesPP[33]:

$$\lambda^*(t) = \mu + (\mu_0 - \mu)\beta\exp(-\beta t) + \alpha\beta \sum_{t_j<t} \exp(-\beta(t-t_j)) \tag{22}$$

where the conditional intensity function starts at μ_0. Unlike traditional formulations assuming a clear system start at $t = 0$, this approach acknowledges that many real-world processes have no well-defined beginning and are only observed from a certain point onward. Here, $t = 0$ marks the start of the recording of the observation session rather than the process itself. Our preliminary analyses supported the assumption that edge effects do not alter the infectivity factor or time scale; however, the

baseline intensity μ decreases from 0.022 to 0.018. This outcome is expected, as edge effects, such as left truncation, are known to upwardly bias baseline intensity. Introducing μ_0 may help compensate for this bias. However, we find that while the additional parameter provides only marginal benefits to the model, it significantly alters the interpretation of the model parameters. For this reason, we opted to not perform further experiments with this formulation or present further results on this TPP parameterization.

Model Inference

After preprocessing the data (Section "Data Preprocessing"), we perform Bayesian inference using Monte-Carlo Markov-Chain (MCMC) to estimate the posterior distribution of the TPP model parameters. By Bayes' theorem, the posterior distribution of the model parameters θ given the observed sessions $\mathbb{S}$ and session durations $\mathbb{T}$ is $\pi(\theta \mid \mathbb{S}, \mathbb{T}) \propto L(\theta \mid \mathbb{S}, \mathbb{T}) p(\theta)$, where $L(\theta \mid \mathbb{S}, \mathbb{T})$ is the likelihood function (Equation (10)), and $p(\theta)$ is the prior PDF. We assign Gamma priors to all TPP model parameters, except for the HawkesPLPP, as the Gamma distribution serves as a conjugate prior for the Poisson Process and offers flexibility over the positive real line. For the HawkesPLPP, we use wide uniform priors. To underscore how we chose weakly informative and wide prior PDFs, the prior PDF versus the posterior PDF for the model parameters is show in Supplementary Figures S7 to S10. Since direct computation of the posterior is often intractable due to high-dimensional integration or non-analytical integrals required for the partition function, we approximate it using MCMC. MCMC constructs an infinite dimension Markov chain whose stationary distribution converges to the target posterior: $p(\theta'|\mathbb{S}) = \int_\theta p(\theta'|\theta) p(\theta|\mathbb{S}) d\theta$. The transition function, $p(\theta'|\theta)$ of the Markov Chain is denoted as the transition kernel. Thus, each sampled state of the Markov chain represents a specific realization of the TPP model parameters sampled from the posterior PDF.

To generate samples from this Markov chain, we employ the No U-Turn Sampler (NUTS)[36], an adaptive variant of Hamiltonian Monte Carlo (HMC), as the transition kernel. In MCMC, the transition kernel defines the probability of moving from one state to another while ensuring ergodicity and convergence to the posterior. Unlike random-walk-based methods, NUTS leverages Hamiltonian dynamics and gradients of the log-posterior to explore the posterior efficiently, proposing new states that are both distant from the current state (reducing autocorrelation in the MCMC samples) and have high posterior probability (exploring the typical set of the posterior distribution[37]). For more details on the MCMC sampling process, we refer readers to[36,38]. NUTS's adaptive approach improves convergence and sampling efficiency in MCMC, especially for posteriors with strong correlations, poor geometry, and nonlinear dependencies[39]. However, efficient mixing of the sampling chains and accurate posterior distribution sampling must still be validated using MCMC diagnostics.

The Effective Sample Size (ESS) measures the number of effectively independent samples. The bulk ESS captures exploration of the high-density region of the posterior, while the tail ESS assesses sampling efficiency at the 5% and 95% quantiles[40] of the posterior. The ESS directly impacts the Monte-Carlo Standard Error (MCSE), which quantifies the uncertainty in the expectation estimates of the TPP model parameters concerning the posterior distribution due to finite sampling: MCSE $\approx \frac{\sigma}{\sqrt{\text{ESS}}}$, where σ represents the standard deviation of the posterior samples. Higher ESS reduces MCSE, leading to more accurate estimates of the posterior mean of the TPP model parameters. Conversely, low ESS increases MCSE, indicating strong autocorrelation in the samples. The Gelman-Rubin statistic ($\hat{R}$) compares within-chain and between-chain variances of the MCMC sample chains to assess whether the Markov Chain explored the entire posterior distribution and has reached a stationary distribution. Values close to 1 suggest well-mixed chains. We also monitor divergences, which indicate numerical instabilities when the sampler struggles to explore the posterior, typically due to regions of high curvature or poor adaptation.

We use 6000 warm up samples and collect 4000 post-warmup samples per MCMC chain, and set a target acceptance probability of 0.99 for the Metropolis-Hastings acceptance criterio and NUTS kernel tuning[36,38]. In all experiments, we observe $\hat{R} \approx 1$ for all TPP model parameters, zero divergences, high bulk and tail ESS for all TPP model parameters, and low MCSE for all TPP. Full MCMC sampling diagnostic results are in Supplementary "Monte-Carlo Markov Chain Diagnostics", Supplementary Tables S1 to S5, and Supplementary Figures S1 to S5. These MCMC diagnostics ensure efficient sampling of the posterior distribution; however, separate model validation metrics are required to assess the GOF and predictive performance of the selected parametric models.

Model Validation

A TPP model for aggressive behavior onsets should effectively handle irregular inter-onset intervals, flexibly capture complex temporal dependencies among onsets within an observation session, maintain interpretability, and provide probabilistic forecasting of future onsets. Evaluation of these qualities relies on both GOF and forecasting performance metrics. GOF tools, such as QQ plots, raw residual plots, and count distributions, assess how well the TPP captures the data-generating process of aggressive behavior onsets[41]. Forecasting performance is evaluated using metrics like PSIS-LOO ELPD, and MAPE, which quantify the model's ability to provide accurate probabilistic predictions of future onsets[42,43].

QQ Plots

A QQ plot is a residual analysis tool for comparing the quantiles of two distributions, often used to assess whether a dataset follows a specific theoretical distribution, such as an exponential distribution. The plot visualizes the relationship between observed data and the expected distribution by plotting the quantiles of the observed data against those of the theoretical distribution. If the scatter plot of the observed and theoretical quantiles closely follows a 45-degree line, it indicates strong alignment, suggesting that the observed data closely matches the theoretical distribution. To apply the QQ plot to TPP models that are not HPP and may have complex theoretical inter-onset distributions, we utilize the Random Time Change (RTC) theorem[44]. The RTC theorem asserts that, for a TPP characterized by a conditional intensity function $\lambda^*(t)$ defined over the history of aggressive event onset times $t_1,\ldots,t_J$, the transformed times $\tau(t_1),\ldots,\tau(t_J)$, obtained using the mapping function

$$\tau(t) = \int_0^t \lambda^*(s)ds = \Lambda^*(t), \tag{23}$$

will follow a standard Poisson process. In this case, the RTC theorem inter-onsets $\tau(t_{j+1}) - \tau(t_j)$ are independent and identically distributed (IID) and exponentially distributed with mean = 1 (representing one aggressive behavior onset per transformed time unit). Thus, the QQ plot compares the empirical distribution of inter-onsets in the observed data after applying the RTC theorem mapping (Equation (23)), while the theoretical distribution is an exponential distribution with mean = 1.

Raw Residual Plots

The raw residual plot is based on the Doob–Meyer decomposition theorem[45]. For a counting process $N(t)$ with conditional intensity $\lambda^*(t)$, the raw residual process is defined as

$$M(t) = N(t) - \Lambda(t) \tag{24}$$

where $\Lambda(t) = \int_0^t \lambda^*(s)\,ds$ is the cumulative intensity function. By the theorem, $M(t)$ is a martingale with mean zero at all times[41]. Therefore, the raw residuals should be centered at zero. To assess model fit, for each session we evaluated the raw residual process at the end of the observation period, $M(T)$, where T denotes the session's end time, and plotted $M(T)$ against the total number of aggressive behavior onsets observed in that session, $N(T)$. If the TPP model fits well, the raw residuals should be symmetrically distributed around zero, indicating that the model accurately captures the expected number of aggressive behavior onsets. Systematic deviations from zero would suggest that the model is either overestimating or underestimating the number of onsets, indicating a poor fit to the data[41].

Count Distribution

A well-fitting TPP should reproduce the empirical distribution of onset counts per session. For session windows $[0,T_i)$ let $N(T_i)$ denote the observed count. The empirical (true) count distribution is the distribution of $\{N(T_i)\}_{i=1}^{I}$. The model count distribution is obtained by simulating onsets in each $[0,T_i)$ using Ogata's thinning algorithm[33].

$$\begin{aligned}
&T_i \sim \mathscr{U}(\{T_1,T_2,\ldots T_I\}) \\
&\theta \sim p(\theta|\mathbb{S},\mathbb{T}) \\
&t_1 \sim f_\theta(t_1|H_{t_0^+}), \quad t_2 \sim f_\theta(t_2|H_{t_1^+}), \quad \ldots, \quad t_j \sim f_\theta(t_j|H_{t_{j-1}^+}) \quad \text{for } t_j < T_i \\
&n_i = N(T_i)
\end{aligned} \tag{25}$$

where $\mathscr{U}$ is a uniform distribution , θ are the model parameter samples from MCMC, t_j are sampled onset times conditioned on past history $H_{t_{j-1}^+}$, and n_i is the total number of events within the window T_i. We then compare the empirical and simulated count distributions (e.g., via Kernel Density Estimators (KDE)) and quantify their discrepancy with the WD[42,46].

Unlike benchmark metrics such as log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), that provide only relative model performance comparisons, QQ plots, raw residual plots, and count distribution plots assess how well the chosen model approximates the true generative process (in the present case, aggressive behavior onsets).

Pareto Smoothed Importance Sampling - Leave One Out Expected Log Predictive Density

Informally, the leave-one-out (LOO) ELPD evaluates the ability of TPP model to predict aggressive behavior onset times in a new session based on other sessions:

$$\text{ELPD}_{\text{loo}} = \sum_{i=1}^{N} \log p(S_i,T_i \mid \mathbb{S}_{-i},\mathbb{T}_{-i}) \tag{26}$$

$$p(S_i,T_i,|\,\mathbb{S}_{-i},\mathbb{T}_{-i}) = \int p(S_i,T_i \mid \theta)p(\theta \mid \mathbb{S}_{-i},\mathbb{T}_{-i})d\theta \tag{27}$$

where $\mathbb{S}_{-i} = \mathbb{S}\backslash S_i$ represents all observation sessions except observation session i, and $\mathbb{T}_{-i} = \mathbb{T}\backslash T_i$ represents all session durations except duration i. A higher LOO ELPD value indicates a better predictive model. The PSIS-LOO ELPD approximates

the LOO expected log predictive density using Pareto-smoothed Importance Sampling (IS), avoiding refitting the model for each held-out session. Using IS, the ELPD may be defined as

$$p(S_i, T_i|\mathbb{S}_{-i}, \mathbb{T}_{-i}) \approx \frac{1}{\sum_{m=1}^{M} p(S_i, T_i \mid \theta^{(m)})} \tag{28}$$

where M is the number of post-warmup MCMC samples from the posterior distribution $p(\theta \mid \mathbb{S})$. However, raw IS can be unstable because the importance ratios may have high or infinite variance. The Pareto Smoothed Importance Sampling (PSIS) procedure stabilizes the estimate by fitting a generalized Pareto distribution to the upper tail of the IS replacing extreme weights with Pareto-smoothed quantiles. This reduces IS weight variance and produces more reliable IS ELPD estimates, especially when the IS weights are heavy-tailed. We visualize each session PSIS-LOO ELPD estimate normalized by the number of onsets in the session, $\log(p(S_i|\mathbb{S}_{-i}))/|S_i|$, in Supplementary Figure S6. However, log-likelihood values may be challenging to interpret; thus, we also include the MAPE calculation for better interpretability.

Mean Absolute Percentage Error

To evaluate forecasting performance, we assess the TPP model's ability to predict the number of aggressive behavior onsets within a time window $[t, t+\Delta t]$. A starting point $t_{\text{start}} \in [\Delta t, T-\Delta t]$ is uniformly at random selected, with aggressive behavior onsets in $[0, t_{\text{start}}]$ forming the history H, and those in $[t_{\text{start}}, t_{\text{start}}+\Delta t]$ constituting the forecast goal. The MAPE is calculated as

$$\text{MAPE}(N(t,t+\Delta t), \tilde{N}(t,t+\Delta t)) = \frac{|N(t,t+\Delta t) - \tilde{N}(t,t+\Delta t)|}{\max(1, N(t,t+\Delta t))} \times 100 \tag{29}$$

where $N(t,t+\Delta t)$ is the observed number of aggressive behavior onsets in $[t,t+\Delta t)$, and $\tilde{N}(t,t+\Delta t)$ is the median predicted count in $[t,t+\Delta t)$, conditioned on the history H_t. Note that the denominator of MAPE has $\max(1, N(t,t+\Delta t))$ to ensure that 0 is not a divisor. The median prediction is

$$\tilde{N}(t,t+\Delta t) = \text{med}\left(\tilde{N}(t,t+\Delta t)^{(m)} \mid m = 1:M\right) \tag{30}$$

where each $\tilde{N}^{(m)}$ is sampled using Ogata's thinning algorithm. For each observation session S_i, we sample 25 uniformly at random t_{start} locations and their corresponding forecasting windows to compute the MAPE. A lower MAPE indicates a better predictive model for the number of aggressive behavior onsets within a future time window.

Results

Count Distribution

Figure 3a compares count distributions between HPP, NHPP, HawkesPLPP, HawkesExpPP, and Hawkes2ExpPP and the true count distribution. Fig. 3a(a,b) compares the count distributions of aggressive behavior onsets per session for the true data and the HPP and NHPP models. The empirical distribution (orange) is highly right-skewed, resembling an exponential distribution with a large probability mass at zero and a long, heavy tail. In contrast, the HPP and NHPP model-generated distributions (blue) exhibit a truncated Gaussian-like shape, centered around 15-16 counts per session, failing to capture the zero-inflation and heavy tail of the true distribution. This discrepancy suggests that both models struggle to account for the bursty and sparse nature of aggressive behavior onsets observed in the data. For instance, in a 100-minute observation session with no aggressive behavior onsets, the HPP model would still predict an average of 16 occurrences, highlighting its inability to capture zero-inflation in the data. However, as shown in Fig. 3a (c,d,e), the distributions generated by the self-exciting TPP models—HawkesPLPP, HawkesPP, and Hawkes2ExpPP (blue)—closely match the empirical distribution (orange). Notably, the KDE of the observed and generated session counts exhibit complete overlap for the HawkesPP and Hawkes2ExpPP, demonstrating the models' ability to accurately capture the true count distribution. Quantitatively, using the WD metric, Table 1 shows that the self-exciting processes (HawkesPLPP, HawkesExpPP and Hawkes2ExpPP) clearly lead to better estimations of the underlying count distribution compared to HPP and NHPP, with Hawkes2ExpPP leading to the smallest WD. Specifically, the Hawkes2ExpPP model achieves a Wasserstein distance of 3.3, more than three times smaller than the Poisson model's distance of 9.8. This metric was computed from 40 Monte Carlo (MC) trials.

QQ Plots

Figure 3b displays the RTC theorem transformed inter-onset quantiles (Equation (23)) against the $\exp(1)$ theoretical quantiles for each model of interest: the HPP, NHPP, HawkesPLPP, HawkesExpPP, and Hawkes2ExpPP models, Figure 3b(a,b) illustrates that the HPP and NHPP fail to capture the bursty nature of aggressive behavior onsets, as their inter-onset distributions allocate excessive probability density to larger onset counts compared to the true data. This leads to overestimating the number

of aggressive behavior onsets during low-intensity periods (when few or no onsets occur over extended durations) and underestimating them during self-excitation phases, highlighting their limitations in modeling self-excitatory dynamics. Fig. 3b(c,d) shows that the HawkesPLPP and HawkesExpPP tend to generate slightly shorter inter-onset times than observed aggressive behavior onsets. This deviation is particularly pronounced in the tail of the distribution, where the model underestimates longer inter-onset times, leading to a bias toward shorter predicted intervals. However, the HawkesPLPP captures the tail of the distribution more effectively than the HawkesExpPP due to the scale-free nature and long tail of the PL kernel. This suggests that in sessions with extended gaps between aggressive events, the HawkesExpPP tends to assume that subsequent events will occur sooner than they do. Fig. 3b(e) shows that the Hawkes2ExpPP effectively corrects the tail deviation observed in Fig. 3b(c,d), reducing the underestimation of longer inter-onset times. This improvement stems from the Hawkes2ExpPP's use of two scaled exponential kernels—one capturing short-term and the other long-term dependencies—enabling rapid and gradual variations in the intensity function. Additionally, the Markovian property of the exponential kernel facilitates better parameter inference. Consequently, the Hawkes2ExpPP provides a more flexible representation of temporal dynamics, better aligning with the observed inter-onset structure.

Raw Residual Plots

Figure 3c plots the raw residuals (Equation (24)) against the number of aggressive behavior onsets per session for each TPP model. By the Doob-Meyer decomposition theorem, residuals, Equation (24), should be centered around 0 (orange line in Fig. 3c). Figure 3c(a,b) shows that the HPP and NHPP models overestimate onsets in sessions with few occurrences and underestimate them in bursty sequences, consistent with the QQ plots. In contrast, Fig. 3c(c,d,e) indicates that the HawkesPLPP, HawkesExpPP, and Hawkes2ExpPP residuals are closer to 0. However, they exhibit a slight positive drift in raw residual error as the number of onsets in an observation session increases.

Pareto Smoothed Importance Sampling - Leave One Out Expected Log Predictive Density

For the PSIS-LOO ELPD estimates, the Pareto k diagnostic values were overwhelmingly favorable. All Pareto k values fell in the "good" range for the self-exciting models (HawkesPLPP, HawkesExpPP, Hawkes2ExpPP) and for the fitted TPP models overall. For the HPP and NHPP models, about 99% of the Pareto k values were "good" and the remainder were "ok," with a single NHPP sample classified as "bad." These categories follow the standard thresholds for the fitted Pareto k parameter used in PSIS-LOO diagnostics[43,47].

This pattern indicates that PSIS-LOO provides a stable and accurate approximation to full LOO ELPD cross-validation for our models. In terms of predictive performance, the Hawkes2ExpPP model attained the highest PSIS-LOO ELPD, followed by the HawkesExpPP and then the HawkesPLPP. The HPP and NHPP models showed substantially lower predictive performance – roughly a two-fold reduction relative to the best self-exciting models.

Mean Absolute Percentage Error

We evaluate the MAPE for the forecasted number of aggressive behavior onsets across time windows of $\Delta t = 1, 5, 10, 15, 20, 25$ minutes (Table 1). As the forecasting window increases, the performance gap between the NHPP and HPP models versus the self-exciting TPPs widens. Specifically, for the Hawkes2ExpPP the MAPE is 11% at 1 minute, rising to about 68% at 25 minutes, while the HPP model quickly becomes much less accurate, with MAPE exceeding 260% at 25 minutes. This is because the HPP and NHPP models do not incorporate the history of aggressive behavior onsets or account for self-excitation. Consequently, when many observation sessions have no aggressive behavior onsets, the HPP and NHPP models tend to overestimate the number of future onsets significantly. In contrast, the HawkesPLPP, HawkesExpPP, and Hawkes2ExpPP models, with a low background intensity μ, forecast fewer aggressive behavior onsets when there is little or no history of such onsets.

Although MAPE may be high for long-horizon probabilistic forecasts, Fig. 4 demonstrates the potential of estimating future aggressive behavior onset counts using only the history of past onsets. Notably, the median of the probabilistic forecast closely follows the true aggressive behavior counting process. Additionally, the upper bound of the 90% credible interval for the Hawkes2ExpPP model is significantly higher than its lower bound, in contrast to the NHPP. This discrepancy may be attributed to the branching factor, or infectivity, of the Hawkes2ExpPP being in the critical (≈ 1) or above-critical (> 1) regime[48], which leads to a nonzero probability of an infinite number of triggered onsets, otherwise stated as an explosion. In the next section, we further discuss the branching factor of the Hawkes2ExpPP and HawkesPLPP.

A corresponding analysis using the ROC-AUC[49], instead of the MAPE, is provided in the Supplementary "Classification Performance of TPPs" and Supplementary Table S6. Rather than estimating the number of onsets occurring within a randomly sampled window of size Δt, we predict the probability of at least one onset occurring in the window using Equation (7). The ground truth label is 1 if any onsets occur within the window and 0 otherwise (Supplementary Fig. S12). The Hawkes2ExpPP achieves ROC-AUC values of 0.85, 0.82, 0.80, 0.79, 0.78, and 0.77 at 1, 5, 10, 15, 20, or 25 minute time windows. In contrast, the HPP hovers around 0.50 for all windows, equivalent to randomly guessing if an aggressive episode will occur.

Branching Factor

The branching factor of a HawkesPP is the average number of "child" (endogenous) events triggered by each "parent" (exogenous) event. It is a key parameter that determines the self-exciting behavior of the process. Mathematically, the branching factor is defined as $BF = \int_0^\infty \phi(t)dt$, where $\phi(t)$ is the triggering kernel in Equation (15). The branching factor equation for the HawkesExpPP, Hawkes2ExpPP, and HawkesPLPP is shown in the x-axis label of Fig. 5 (a),(b),(c), respectively. The branching factor BF determines the stability of a HawkesPP. If $BF < 1$, the process is subcritical, meaning each onset triggers fewer than one new child onset on average, leading to a stable process. If $BF = 1$, the process is critical, where each onset triggers exactly one child onset on average, maintaining a delicate balance between exploding versus stability of onset generation. If $BF > 1$, the process is supercritical, where each onset triggers more than one child onset on average, potentially resulting in an explosive cascade of events. Understanding BF is essential for assessing the long-term behavior and stability of a HawkesPP.

Estimating the branching factor requires careful consideration of potential biases inherent in data collection and model parameter estimation processes. As shown in[13], branching factor estimates, particularly for HawkesPLPP and Hawkes2ExpPP, can be inflated by outliers. To verify that our branching factor estimation is not biased by outliers, we follow the approach in[13] by comparing quantile ratios (95%/90%, 99%/95%, and max/90%) between synthetic and empirical inter-onset sequences, using inverse CDF interpolation for observation sessions with at least five events. Additionally, we define outliers as inter-onset times exceeding 2× the 99% quantile[13]. As shown in Table 2, no substantial evidence suggests that outliers significantly affect our branching factor estimation. However, even if there were outliers, the branching factor estimate for the HawkesExpPP model has been shown to remain robust to a significant number of outlier inter-onsets in the data[13]. Another potential source of bias in branching factor estimation is edge effects[13]. However, we assume their impact on model parameter estimation is minimal (refer to Section "Intensity Function Parameterizations" for further details).

When the branching factor is in the subcritical regime, the BF is interpreted as the percentage of onsets in an observation session triggered by an exogenous onset. Since the HawkesExpPP has an average branching factor of $0.85 - 0.95$, this suggests that $85\% - 95\%$ of all aggressive behavior onsets are triggered by a past aggressive behavior onset rather than external factors. For both HawkesPLPP and Hawkes2ExpPP, there is a non-zero probability of the BF exceeding 1, though only slightly.

Discussion

Our results demonstrate that self-exciting point process models, such as the HawkesPP, fit the observed data well and capture the irregular and bursty nature of aggressive behavior onsets in psychiatric inpatient autistic youth. Compared to the standard HPP, the Hawkes model substantially outperforms in both predictive accuracy and its ability to represent the underlying generative process.

When we frame the prediction task as a classification problem – asking, "Will any aggressive episode occur in the next 1, 5, 10, 15, 20, or 25 minutes?" – the Hawkes2ExpPP reliably distinguishes between windows where aggression is likely versus unlikely, while the standard HPP cannot. We also note that our ROC-AUC values for predicting aggression within 1 to 5 minutes are very similar to those reported by Imbiriba et al.[19] who included physiological features, further validating our modeling approach and results. Beyond classification, the HawkesPP model also provides more accurate forecasts of the actual number of aggressive episodes in each window. In practical terms, our model gives much more realistic estimates of how many aggressive episodes will occur in a given period, especially as the forecast window grows. Previous studies[19–21] cannot estimate the expected number of aggressive behavior episodes in a given observation window. We demonstrated using WD – a measure of how well a model captures the generative process of the real-world distribution of aggressive onset counts – that the HawkesPP model generates data that more closely matches the true pattern of aggressive behavior counts in sessions observed in the hospital compared to traditional HPP. These findings are further supported by the PSIS-LOO ELPD, which directly measures out-of-sample predictive performance. The Hawkes model's PSIS-LOO ELPD is nearly twice as high (less negative) as the Poisson model's, demonstrating that it generalizes substantially better to new observation sessions and is more reliable for forecasting aggressive behavior onsets in real-world clinical settings.

Additionally, a key advantage of using TPP models is their interpretability and explainability. For example, the inferred parameters for the Hawkes process show that after an initial aggressive behavior onset, there is a high probability of subsequent onsets occurring in quick succession—a cascade effect. Thus, interventions that prevent the first instance or mitigate early occurrences of aggressive behavior onsets, or "mother" onsets, should be prioritized. This suggests that careful interactions and interventions should be provided to inpatient youths with autism to prevent a cascade of subsequent aggressive behavior onsets. These insights also introduce new research questions, such as whether we can generate the branching structure of the cluster representation of a HawkesPP to identify all the "mother" onsets and, through post-analysis, understand what triggers them.

Our collective results demonstrate that our modeling approach improves predictive performance and provides potentially meaningful clinical insights and actionable forecasts. Accurately forecasting the number of aggressive behavior onsets within a future window (Fig. 4), predicting whether any event will occur in that window (Fig. 6), and estimating the precise timing of future onsets all have translational clinical potential. For example, the model's predicted probability that at least one aggressive

episode will occur in the next 5 or 10 minutes could be used to trigger real-time alerts for clinicians and caregivers, prompting increased supervision or preventive action. Forecasts of the expected number of events in a window could inform staff allocation and resource planning, ensuring that support is available when risk is highest. Precise timing predictions could allow for just-in-time adaptive interventions, such as de-escalation strategies or environmental modifications, to be delivered before an episode occurs. Additionally, the model's ability to generate synthetic data that closely matches real-world patterns could enable safe simulation and testing of intervention protocols, as well as staff training, without risk to patients. Together, these outputs support proactive, data-driven decision-making to reduce harm and improve care in clinical settings.

Furthermore, our GOF visualizations – including the QQ plot (Fig. 3b), raw residual plot (Fig. 3c), and count distribution (Fig. 3a) – demonstrate that self-exciting point process models, such as the HawkesPP, can closely capture the complex, clustered patterns of aggressive behavior onset observed in psychiatric inpatient youth with autism. This strong fit is consistent with both expert clinical observation and prior analyses using the Ripley K estimator[50], which highlight the temporal clustering of aggression in this population (Supplementary "Ripley K Estimator" and Supplementary Fig. S11). Predictive metrics, including PSIS-LOO ELPD and MAPE, further confirm promise of TPPs for forecasting aggression risk over clinically relevant time windows.

Limitations and Future Work

Following this initial analysis, we intend to explore three future directions. First, we plan to divide our sample into sub-samples based on key variables such as the number of observation sessions, total observation time, and total aggressive behavior onsets, duration, and intensity. By analyzing differences among these sub-samples, we aim to gain deeper insights into the variability of aggressive behavior and its recurrence. Second, we will investigate statistical differences in inferred model parameters and examine variations in modeling choices based on the type of condition label, including SIB, ED, and ATO. Thirdly, we will utilize a hierarchical HawkesPP to account for between-person heterogeneity and disparities in data sample sizes among participants.

The distribution of aggressive behavior onsets per participant appears to follow a power-law pattern. Specifically, one participant accounts for 26.5% of the aggressive behavior onsets (1,287 onsets), two participants contribute 25% (475 to 910 onsets), five participants contribute 20.8% (106 to 476 onsets), and 62 participants account for 27.7% (0 to 105 onsets). This distribution suggests the existence of subpopulations within the data, conceptualized as chronically aggressive, hyper-aggressive, aggressive, and mildly aggressive. Given the substantial disparity in onsets exhibited by individual participants, hierarchical HawkesPP could address intra-individual and inter-individual variability. A single-level hierarchical HawkesPP could refine the analysis of the branching factor by modeling inter-individual heterogeneity and homogeneity. In contrast, a two-level hierarchical model would capture intra-individual and inter-individual session variability, further enhancing the analysis of edge effects, as discussed with the modified HawkesPP in Section "Intensity Function Parameterizations". This two-level model would also enable the inference of a μ_0 per observation session. Finally, we aggregated three distinct behaviors into a single aggression onset label. A more nuanced analysis would involve modeling ATO, SIB, and ED separately as each type of aggression may follow a distinct generative process, with potentially different baseline intensities and interacting infectivity factors. This modeling of interacting point processes could be achieved via a Multivariate Hawkes process[51].

While our approach demonstrates strong promise in learning a TPP that closely captures the generative process of aggressive behavior onsets, it still has limitations in predictive performance, mainly when answering questions like "how many events will occur within the next 5 minutes?". This modeling gap may be due to limited data sample size, the need for more expressive TPP models (such as those with advanced kernel functions or deep learning-based architectures), or the exclusion of important covariates like physiological signals.

Furthermore, a limitation of generative modeling is that although we may learn the generative process of aggressive behavior onsets with reasonable accuracy, this does not necessarily translate into improved predictive capability. To illustrate with a simple example, consider inferring the probability of heads in a fair coin toss. Estimating this probability as 0.5 from multiple flips is straightforward. However, even with an accurate estimate, it does not enable precise prediction of the outcome of the next flip or a sequence of flips. Our model may effectively capture the underlying generative process of aggressive behavior onsets but may still struggle to predict proximal occurrence given the history of onsets within an observation session. Thus, while our models provide probabilistic forecasts that could support real-time clinical decision-making, there is still a lack of standardized, interpretable metrics for evaluating temporal point process performance in decision-based frameworks – particularly metrics that are meaningful and actionable for clinicians[16,25]. Addressing these challenges will be essential for translating TPP modeling into practical, impactful tools for clinical care.

Lastly, a limitation towards real-time model parameter estimation is that MCMC inference is notoriously slow, with online parameter update mechanisms still under research. However, we may explore online inference methods, such as Cubature Kalman Filters, for estimating the parameters of the intensity function in a HawkesPP. This would allow us to perform online updates of the model parameters with streaming data, bringing us a step closer to enabling just-in-time preventative interventions

to reduce aggressive behavior in autistic youth.

Software Packages

All analyses were conducted in Python, (version 3.9) following the *Cookiecutter Data Science* project structure and workflow[52]. Bayesian inference was performed using `NumPyro=0.16.1`[39], a probabilistic programming language built on `JAX=0.4.30`[53] for efficient NUTS sampling. Model diagnostics (including PSIS-LOO cross-validation) and posterior exploratory analysis utilized `ArviZ=0.17.1`[54], a toolkit for exploratory analysis of Bayesian models. Evaluation metrics such as ROC-AUC and WD were calculated using `scikit-learn=1.5.2` and `scipy=1.12.0`, respectively. Data preprocessing employed `pandas=2.2.3`, `numpy=1.26.4`, and `xarray=2024.7.0`. Visualizations were created using `matplotlib=3.9.3`, and `seaborn=0.13.2`.

Acknowledgements
This study was supported by grant R01LM014191 from the National Institutes of Health, the Simons Foundation for Autism Research Initiative, and the Nancy Lurie Marks Family Foundation. We are grateful to participating families. Families were offered $75 for their child's participation in the study. All project personnel listed were paid as research assistants for the duration of their involvement.

Author Contributions Statement
M.P.,A.S.,D.E.,T.I.,M.S.G conceived the experiments. M.S.G.,Y.W. conducted the data collection. M.P.,M.E.,T.I. conducted the experiments. M.P.,M.E.,T.I.,M.S.G. analyzed the results. All authors reviewed the manuscript. D.E., M.S.G obtained funding.

Competing Interests
The author(s) declare no competing interests.

Data Availability
The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References
1. Maenner, M. J. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2020. *MMWR. Surveillance Summ.* **72** (2023).
2. Hattier, M. A., Matson, J. L., Belva, B. C. & Horovitz, M. The occurrence of challenging behaviours in children with autism spectrum disorders and atypical development. *Dev. Neurorehabilitation* **14**, 221–229 (2011).
3. Kanne, S. M. & Mazurek, M. O. Aggression in children and adolescents with asd: Prevalence and risk factors. *J. autism developmental disorders* **41**, 926–937 (2011).
4. Matson, J. L. & Cervantes, P. E. Assessing aggression in persons with autism spectrum disorders: An overview. *Res. developmental disabilities* **35**, 3269–3275 (2014).
5. Arnold, L. E. *et al.* Parent-defined target symptoms respond to risperidone in rupp autism study: customer approach to clinical trials. *J. Am. Acad. Child & Adolesc. Psychiatry* **42**, 1443–1450 (2003).
6. Croen, L. A., Najjar, D. V., Ray, G. T., Lotspeich, L. & Bernal, P. A comparison of health care utilization and costs of children with and without autism spectrum disorders in a large group-model health plan. *Pediatrics* **118**, e1203–e1211 (2006).
7. Mazefsky, C. A. & White, S. W. Emotion regulation: Concepts & practice in autism spectrum disorder. *Child adolescent psychiatric clinics North Am.* **23**, 10–1016 (2013).
8. Tager-Flusberg, H. *et al.* Conducting research with minimally verbal participants with autism spectrum disorder. *Autism* **21**, 852–861 (2017).
9. Davis, N. O. & Carter, A. S. Parenting stress in mothers and fathers of toddlers with autism spectrum disorders: Associations with child characteristics. *J. autism developmental disorders* **38**, 1278–1291 (2008).
10. Hodgetts, S., Nicholas, D. & Zwaigenbaum, L. Home sweet home? families' experiences with aggression in children with autism spectrum disorders. *Focus. on autism other developmental disabilities* **28**, 166–174 (2013).
11. Kiely, J. & Pankhurst, H. Violence faced by staff in a learning disability service. *Disabil. Rehabil.* **20**, 81–89 (1998).
12. Ogata, Y. Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. association* **83**, 9–27 (1988).
13. Filimonov, V. & Sornette, D. Apparent criticality and calibration issues in the hawkes self-excited point process model: application to high-frequency financial data. *Quant. Finance* **15**, 1293–1314 (2015).
14. Gatrell, A. C., Bailey, T. C., Diggle, P. J. & Rowlingson, B. S. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions Inst. Br. geographers* 256–274 (1996).
15. Yan, J. Recent advance in temporal point process: from machine learning perspective. *SJTU Tech. Rep.* (2019).
16. Shchur, O., Türkmen, A. C., Januschowski, T. & Günnemann, S. Neural temporal point processes: A review. *arXiv preprint arXiv:2104.03528* (2021).

17. Fadhel, Z. M., Hussein, Q. M. & Askari, A. O. A. R. A. Detecting aggressive behavior in children with autism spectrum disorder using machine learning. In *AIP Conference Proceedings*, vol. 2845 (AIP Publishing, 2023).

18. Khullar, V., Singh, H. P. & Bala, M. Meltdown/tantrum detection system for individuals with autism spectrum disorder. *Appl. Artif. Intell.* **35**, 1708–1732 (2021).

19. Imbiriba, T., Demirkaya, A., Singh, A., Erdogmus, D. & Goodwin, M. S. Wearable biosensing to predict imminent aggressive behavior in psychiatric inpatient youths with autism. *JAMA network open* **6**, e2348898–e2348898 (2023).

20. Goodwin, M. S. *et al.* Predicting imminent aggression onset in minimally-verbal youth with autism spectrum disorder using preceding physiological signals. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '18, 201–207, DOI: 10.1145/3240925.3240980 (Association for Computing Machinery, New York, NY, USA, 2018).

21. Imbiriba, T. *et al.* Biosensor prediction of aggression in youth with autism using kernel-based methods. In *Proceedings of the 13th ACM international conference on PErvasive technologies related to assistive environments*, 1–6 (2020).

22. Xiao, S. *et al.* Wasserstein learning of deep generative point process models. *Adv. neural information processing systems* **30** (2017).

23. Chaspari, T. *et al.* A non-homogeneous poisson process model of skin conductance responses integrated with observed regulatory behaviors for autism intervention. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 1611–1615 (IEEE, 2014).

24. Bae, W., Ahmed, M. O., Tung, F. & Oliveira, G. L. Meta temporal point processes. *arXiv preprint arXiv:2301.12023* (2023).

25. Shachter, R. D. & Peot, M. A. Decision making using probabilistic inference methods. In *Uncertainty in Artificial Intelligence*, 276–283 (Elsevier, 1992).

26. Lindsay, J. J. & Anderson, C. A. From antecedent conditions to violent actions: A general affective aggression model. *Pers. Soc. Psychol. Bull.* **26**, 533–547 (2000).

27. Scarpa, A. Physiological arousal and its dysregulation in child maladjustment. *Curr. Dir. Psychol. Sci.* **24**, 345–351 (2015).

28. Cohen, I. L., Yoo, J. H., Goodwin, M. S. & Moskowitz, L. Assessing challenging behaviors in autism spectrum disorders: Prevalence, rating scales, and autonomic indicators. *Int. handbook autism pervasive developmental disorders* 247–270 (2011).

29. Rasmussen, J. G. Lecture notes: Temporal point processes and the conditional intensity function. *arXiv preprint arXiv:1806.00221* (2018).

30. Nohara, Y., Matsumoto, K., Soejima, H. & Nakashima, N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput. Methods Programs Biomed.* **214**, 106584 (2022).

31. Zhang, Z. Variable selection with stepwise and best subset approaches. *Annals translational medicine* **4**, 136 (2016).

32. Gallager, R. G. Discrete stochastic processes. *J. Oper. Res. Soc.* **48**, 103–103 (1997).

33. Laub, P. J., Lee, Y., Pollett, P. K. & Taimre, T. Hawkes models and their applications. *Annu. Rev. Stat. Its Appl.* **12** (2024).

34. Hawkes, A. G. Point spectra of some mutually exciting point processes. *J. Royal Stat. Soc. Ser. B: Stat. Methodol.* **33**, 438–443 (1971).

35. Lee, K. Multi-kernel property in high-frequency price dynamics under hawkes model. *Stud. Nonlinear Dyn. & Econom.* **28**, 605–624 (2024).

36. Hoffman, M. D., Gelman, A. *et al.* The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).

37. Betancourt, M. A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434* (2017).

38. Barber, D. *Bayesian reasoning and machine learning* (Cambridge University Press, 2012).

39. Phan, D., Pradhan, N. & Jankowiak, M. Composable effects for flexible and accelerated probabilistic programming in numpyro. *arXiv preprint arXiv:1912.11554* (2019).

40. Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. statistical software* **76**, 1–32 (2017).

41. Wu, J., Smith, A. L. & Zheng, T. Diagnostics and visualization of point process models for event times on a social network. *Appl. modeling techniques data analysis 1: computational data analysis methods tools* **7**, 129–145 (2021).

42. Lüdke, D., Biloš, M., Shchur, O., Lienen, M. & Günnemann, S. Add and thin: Diffusion for temporal point processes. *Adv. Neural Inf. Process. Syst.* **36**, 56784–56801 (2023).

43. Vehtari, A., Gelman, A. & Gabry, J. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Stat. computing* **27**, 1413–1432 (2017).

44. Brown, T. C. & Nair, M. G. A simple proof of the multivariate random time change theorem for point processes. *J. Appl. Probab.* **25**, 210–214 (1988).

45. Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N. *Statistical models based on counting processes* (Springer Science & Business Media, 2012).

46. Kolouri, S. *et al.* Optimal transport and wasserstein distance.

47. Vehtari, A., Simpson, D., Gelman, A., Yao, Y. & Gabry, J. Pareto smoothed importance sampling. *J. Mach. Learn. Res.* **25**, 1–58 (2024).

48. Hardiman, S. J., Bercot, N. & Bouchaud, J.-P. Critical reflexivity in financial markets: a hawkes process analysis. *The Eur. Phys. J. B* **86**, 1–9 (2013).

49. Narkhede, S. Understanding auc-roc curve. *Towards data science* **26**, 220–227 (2018).

50. Karlis, D. Proposer of the vote of thanks to narayanan, kosmidis and dellaportas and contribution to the discussion of 'flexible marked spatio-temporal point processes with applications to event sequences from association football'. *J. Royal Stat. Soc. Ser. C: Appl. Stat.* **72**, 1127–1128 (2023).

51. Liniger, T. *Multivariate hawkes processes*. Ph.D. thesis, ETH Zurich (2009).

52. DrivenData. Drivendata: Data science & ai competitions to build a better world. https://www.drivendata.org/ (2025). Accessed: 2025-07-11.

53. Bradbury, J. *et al.* JAX: composable transformations of Python+NumPy programs (2018).

54. Kumar, R., Carroll, C., Hartikainen, A. & Martin, O. Arviz a unified library for exploratory analysis of bayesian models in python. *J. Open Source Softw.* **4**, 1143, DOI: 10.21105/joss.01143 (2019).

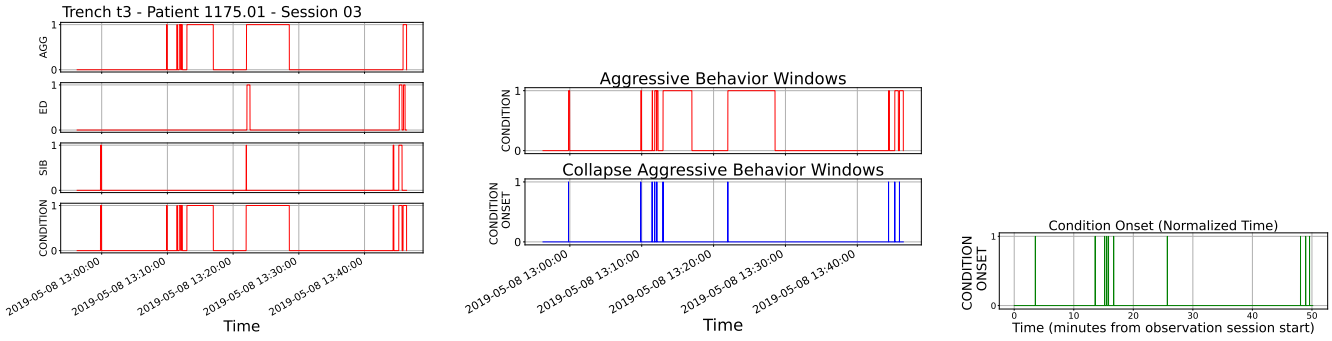


Figure 1. Three stages of data preprocessing. (a) Aggregating SIB,ATO,ED behavior labels into a condition label. (b) Preproccessing only the aggressive behavior onset timestamps. (c) Normalizing the aggressive behavior onset timestamps to be at the begining of the observation session.

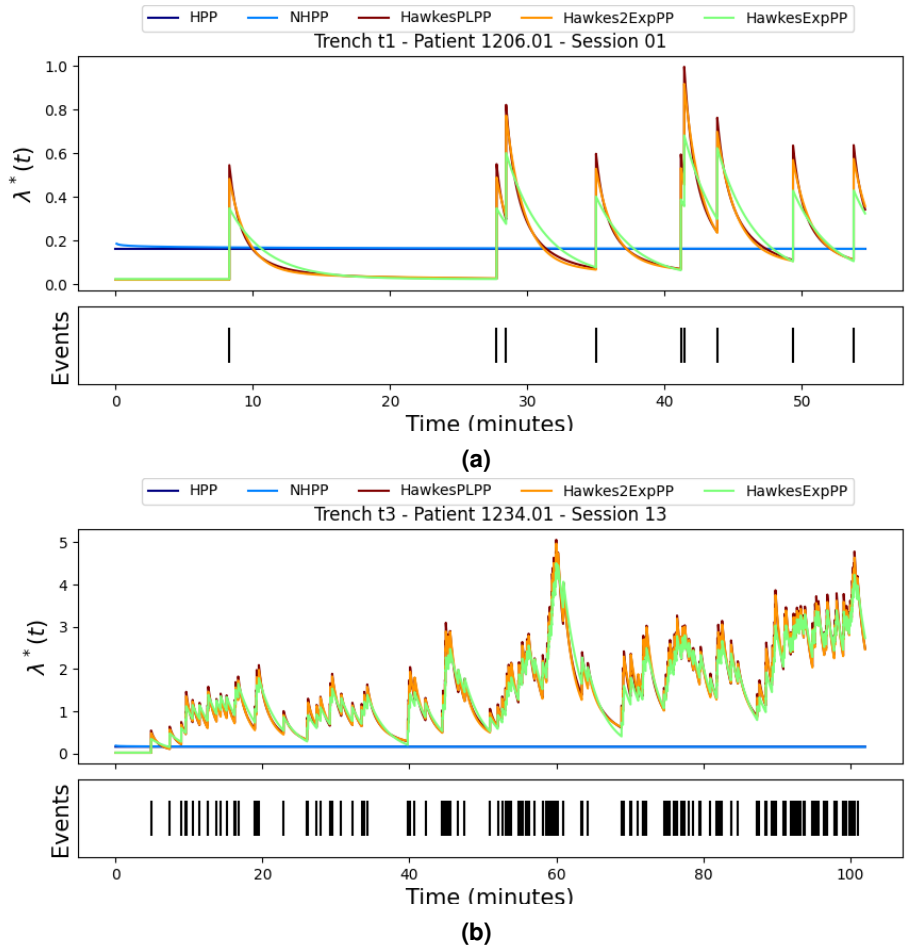


Figure 2. TPPs conditional intensity curves for two observation sessions. (a) The conditional intensity curve for each TPP model in participant 1206.01, observation session 1. (b) The conditional intensity curve for each TPP model in participant 1234.01 observation session 13.

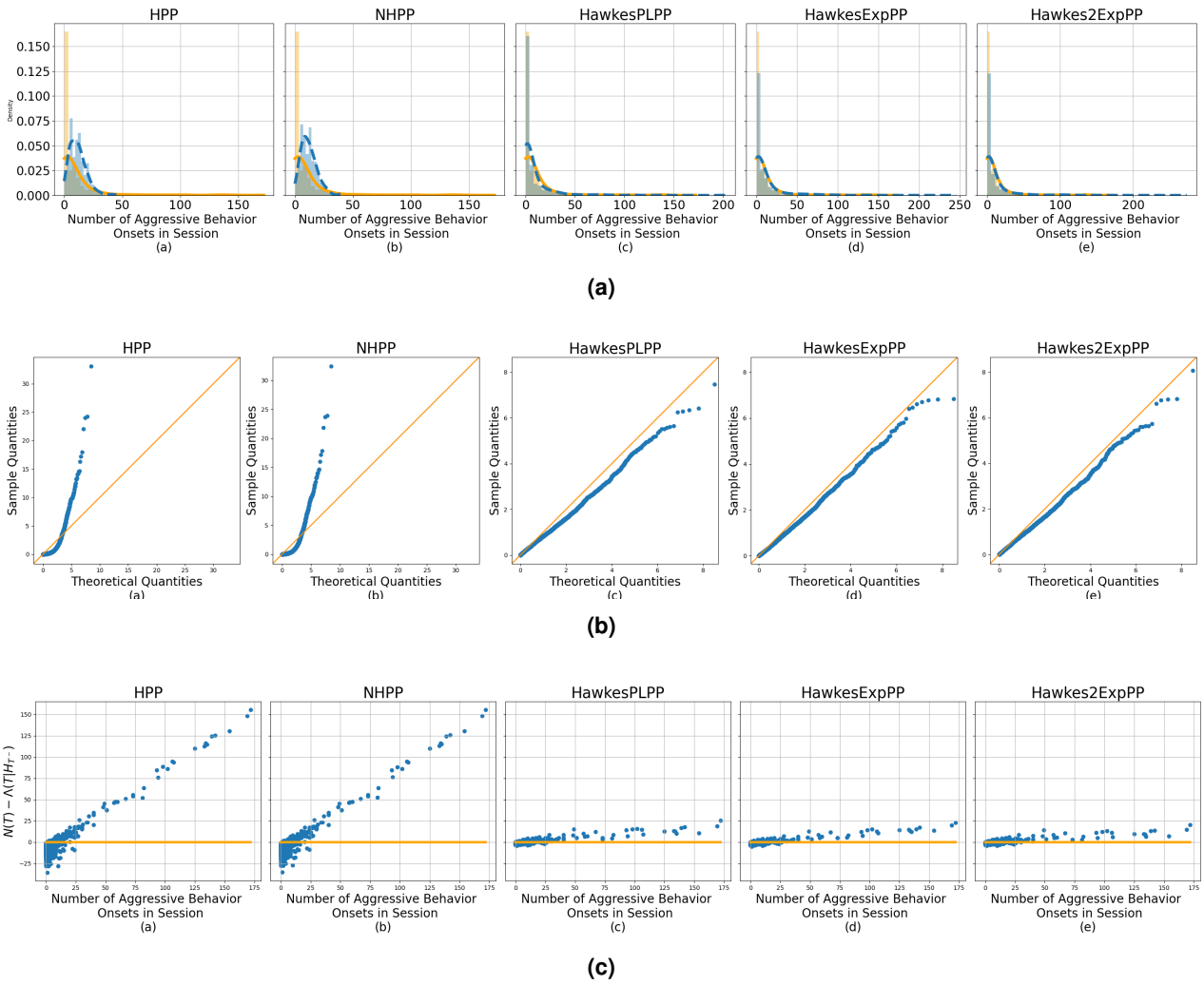
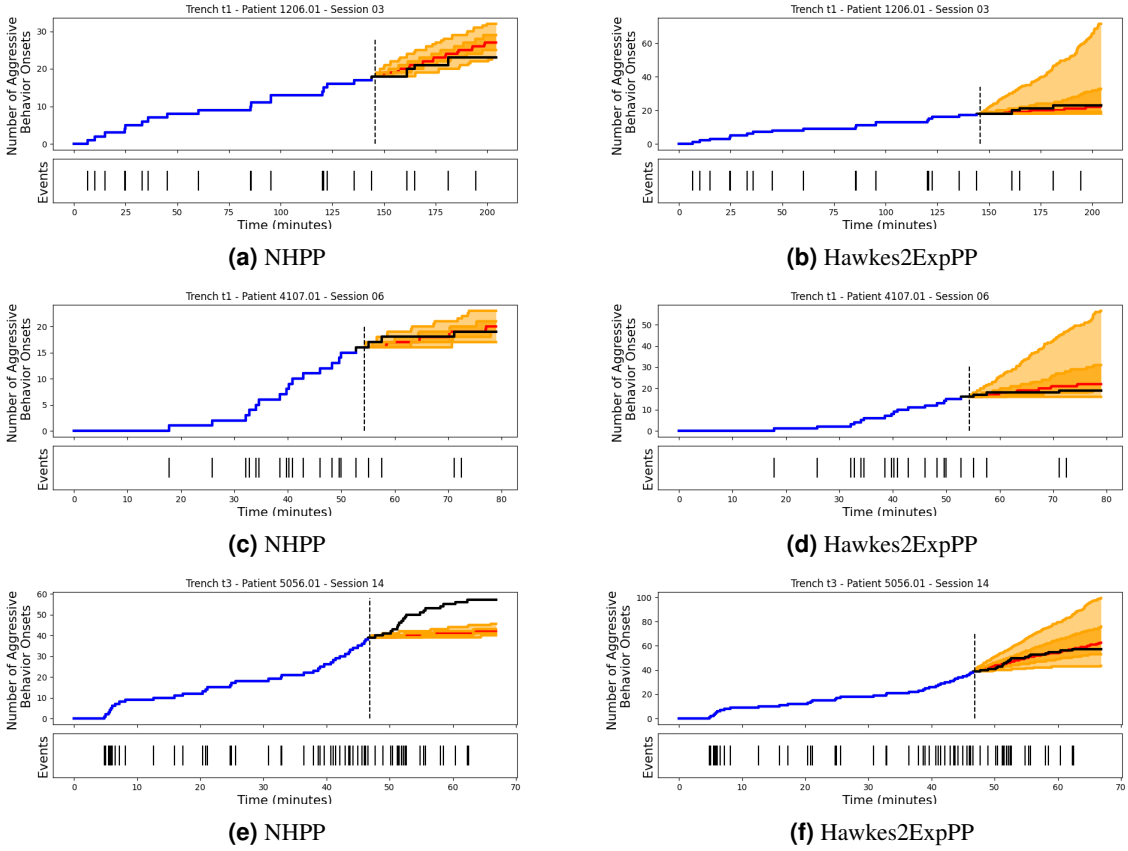


Figure 3. Goodness-of-fit evaluation metrics for the TPPs. Subfigure (a): The empirical data count distribution (orange) versus TPP generated count distribution (blue) for the number of aggressive behavior onsets in an observation session. Subfigure (b): the QQ plot of RTC theorem inter-arrivals for different TPP fits. The x-axis is the theoretical quantiles of an exponential distribution with mean 1, and the y-axis is the sample quantiles from all the RTC theorem inter-arrivals in the dataset. Subfigure (c): the scatter plot of the raw residuals (Equation (24)) versus the number of aggressive behavior onsets in an observation session (blue). In each subfigure, (a)-(e) refer to the HPP, NHPP, HawkesPLPP, HawkesExpPP, and Hawkes2ExpPP models, respectively.

Metric	|S|	HPP	NHPP	HawkesPLPP	HawkesExpPP	Hawkes2ExpPP
WD $\pm$ SE	429	9.75 ± 0.29	9.83 ± 0.31	3.53 ± 1.15	3.66 ± 1.5	3.27 ± 1.18
PSIS-LOO $\text{ELPD}_{\text{LOO}} \pm$ SE	429	-13727 ± 1003	-13731 ± 1005	-7832 ± 466	-7807 ± 481	-7710 ± 471
MAPE@$\Delta t = 1$	429	11.43	11.43	11.16	11.69	11.22
MAPE@$\Delta t = 5$	424	78.20	77.34	27.48	29.06	27.92
MAPE@$\Delta t = 10$	417	116.44	107.04	36.46	37.88	37.97
MAPE@$\Delta t = 15$	409	156.67	148.35	46.22	45.42	49.20
MAPE@$\Delta t = 20$	399	210.44	205.66	53.42	50.95	57.60
MAPE@$\Delta t = 25$	385	262.15	258.25	61.85	61.89	67.88

Table 1. GOF and predictive metrics for TPPs. $|S|$ is the number of observation sessions used to calculate each metric, and SE is the standard error. The first row is WD (mean±std) between the empirical data count distribution versus the TPP generated count distribution for number of aggressive behavior onsets within an observation session over 40 MC trials. The second row is PSIS-LOO ELPD over $|S| = 429$ observation sessions. The third row onward, i.e. any row with MAPE@$\Delta t = X$, is the MAPE for forecasting the number of aggressive behavior onsets in a time interval of size Δt (min). The forecast is the median count from 75 samples aggressive behavior onset sequences (for each observation session).



(a) NHPP **(b)** Hawkes2ExpPP

(c) NHPP **(d)** Hawkes2ExpPP

(e) NHPP **(f)** Hawkes2ExpPP

Figure 4. Visualization of the forecasted number of aggressive behavior onsets within a future time window, conditioned on the history of prior onsets. The blue line represents the observed history of onset counts, while the black line denotes the true future onset count. The black vertical dotted line marks the start of forecasting. The red line corresponds to the median prediction over 250 sampled forecasts, with the darker orange shaded region indicating the interquartile range (25th to 75th quantile) and the lighter orange shaded region representing the 90% credible interval (5th to 95th quantile). (a,b) show the forecasted number of onsets in participant 1206.01, session 3, (c,d) correspond to participant 4107.01, session 6, while (e,f) represent participant 5056.01, session 14, with each pair displaying forecasts for the NHPP and Hawkes2ExpPP, respectively.

Statistic Model	Q90	Q95	Q99	Max	N	Q95/Q90	Q99/Q95	Max/Q99	outlier
Hawkes2ExpPP	12.69 ± 12.09	19.37 ± 16.32	30.71 ± 21.21	35.99 ± 23.47	31.28 ± 44.40	1.89 ± 1.63	2.35 ± 2.57	1.40 ± 1.55	0.06 ± 0.24
HawkesPLPP	13.75 ± 12.53	21.14 ± 16.67	32.55 ± 21.14	37.10 ± 23.19	25.56 ± 30.87	1.86 ± 1.29	2.21 ± 2.45	1.24 ± 0.89	0.04 ± 0.21
HawkesExpPP	12.50 ± 12.52	19.66 ± 17.21	31.26 ± 21.79	36.48 ± 23.91	28.28 ± 36.39	1.91 ± 1.30	2.63 ± 3.20	1.40 ± 1.85	0.06 ± 0.26
True Data	9.86 ± 10.07	14.25 ± 13.01	20.49 ± 17.34	22.75 ± 18.92	29.13 ± 36.15	1.57 ± 0.51	1.70 ± 1.41	1.14 ± 0.26	0.01 ± 0.09

Table 2. Comparison of summary statistics for data generated from the fitted TPP versus the true data summary statistics. Q90, Q95, Q99 are the 90th, 95th, and 99th quantiles, whereas N is the average number of onsets per observation session.

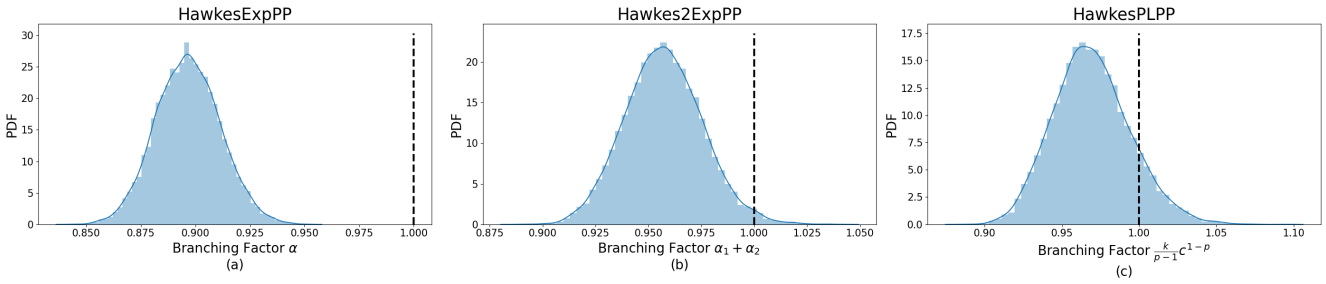


Figure 5. Posterior PDF over the branching factor for the HawkesPPs. (a), (b), (c) are the branching factors for the HawkesExpPP, Hawkes2ExpPP and HawkesPLPP, respectively.

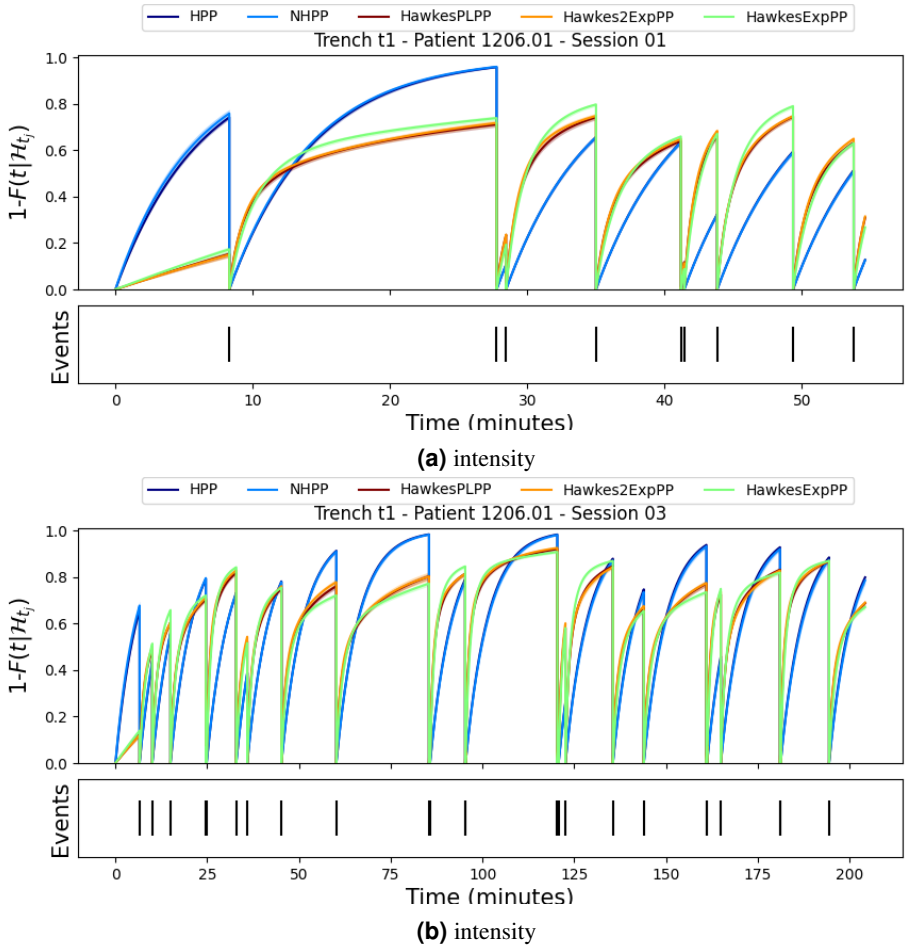


Figure 6. TPPs conditional CDF curves for two observation sessions. (a) The conditional CDF curve for each TPP model in participant 1206.01, observation session 1. (b) The conditional CDF curve for each TPP model in participant 1234.01 observation session 13.

Legends

- **Figure 1**: Three stages of data preprocessing. (a) Aggregating SIB,ATO,ED behavior labels into a condition label. (b) Preproccessing only the aggressive behavior onset timestamps. (c) Normalizing the aggressive behavior onset timestamps to be at the begining of the observation session.
- **Figure 2**: TPPs conditional intensity curves for two observation sessions. (a) The conditional intensity curve for each TPP model in participant 1206.01, observation session 1. (b) The conditional intensity curve for each TPP model in participant 1234.01 observation session 13.
- **Figure 3**: Goodness-of-fit evaluation metrics for the TPPs. Subfigure (a): The empirical data count distribution (orange) versus TPP generated count distribution (blue) for the number of aggressive behavior onsets in an observation session. Subfigure (b): the QQ plot of RTC theorem inter-arrivals for different TPP fits. The x-axis is the theoretical quantiles of an exponential distribution with mean 1, and the y-axis is the sample quantiles from all the RTC theorem inter-arrivals in the dataset. Subfigure (c): the scatter plot of the raw residuals (Equation (24)) versus the number of aggressive behavior onsets in an observation session (blue). In each subfigure, (a)-(e) refer to the HPP, NHPP, HawkesPLPP, HawkesExpPP, and Hawkes2ExpPP models, respectively.
- **Figure 4**: Visualization of the forecasted number of aggressive behavior onsets within a future time window, conditioned on the history of prior onsets. The blue line represents the observed history of onset counts, while the black line denotes the true future onset count. The black vertical dotted line marks the start of forecasting. The red line corresponds to the median prediction over 250 sampled forecasts, with the darker orange shaded region indicating the interquartile range (25th to 75th quantile) and the lighter orange shaded region representing the 90% credible interval (5th to 95th quantile). (a,b) show the forecasted number of onsets in participant 1206.01, session 3, (c,d) correspond to participant 4107.01, session 6, while (e,f) represent participant 5056.01, session 14, with each pair displaying
- **Figure 5**: Posterior PDF over the branching factor for the HawkesPPs. (a), (b), (c) are the branching factors for the HawkesExpPP, Hawkes2ExpPP and HawkesPLPP, respectively.
- **Figure 6**: TPPs conditional CDF curves for two observation sessions. (a) The conditional CDF curve for each TPP model in participant 1206.01, observation session 1. (b) The conditional CDF curve for each TPP model in participant 1234.01 observation session 13.
- **Table 1**: GOF and predictive metrics for TPPs. $|S|$ is the number of observation sessions used to calculate each metric, and SE is the standard error. The first row is WD (mean±std) between the empirical data count distribution versus the TPP generated count distribution for number of aggressive behavior onsets within an observation session over 40 MC trials. The second row is PSIS-LOO ELPD over $|S| = 429$ observation sessions. The third row onward, i.e. any row with MAPE@$\Delta t = X$, is the MAPE for forecasting the number of aggressive behavior onsets in a time interval of size Δt (min).
- **Table 2**: Comparison of summary statistics for data generated from the fitted TPP versus the true data summary statistics. Q90, Q95, Q99 are the 90th, 95th, and 99th quantiles, whereas N is the average number of onsets per observation session.

Supplementary Information

Supplementary Methods

Monte-Carlo Markov Chain Diagnostics

We provide the Monte-Carlo Markov-Chain (MCMC) inference diagnostics such as $\hat{R}$, Monte-Carlo Standard Error (MCSE) on the posterior mean and posterior variance of the Temporal Point Process (TPP) parameters, the Effective Sample Size (ESS) for the bulk and tail as well as the parameter expectations, standard deviations, and quantiles. Furthermore, we provide the autocorrelation plots of each MCMC chain along with the posterior Probability Density Function (PDF) of the TPP parameters. The MCMC inference diagnostics and MCMC sample summary statistics for the Homogeneous Poisson Process (HPP), Non-Homogeneous Poisson Process (NHPP), Hawkes Point Process with Exponential Kernel (HawkesExpPP), Hawkes Point Process with Two Exponential Kernels (Hawkes2ExpPP), and Hawkes Point Process with Power Law Kernel (HawkesPLPP) are found in Supplementary Tables S1 to S5 respectively. The MCMC sampling chain autocorrelation plots and MCMC sample histograms for each parameter for HPP, NHPP, HawkesExpPP, Hawkes2ExpPP, and HawkesPLPP are shown in Supplementary Figs. S1 to S5 respectively.

Pareto Smoothed Importance Sampling - Leave One Out (PSIS-LOO) Expected Log Density (ELPD)

We visualize the ELPD normalized by the number of aggressive behaviors in the observation session for each observation session against the number of observed aggressive behavior onsets to examine whether poor ELPD estimates correlate with the onset count (Supplementary Fig. S6). Notably, the TPP model performs worst when the number of onsets per session is between 1 and 10, indicating greater difficulty in capturing event dynamics in sparsely populated sessions.

Posterior versus Prior PDF of TPP parameters

We visualize the posterior PDF from MCMC inference samples and the prior PDF for each model parameter to compare how much the data shifts the posterior PDF from the prior PDF. This gives some indication on how much the data drives the posterior PDF, and if the prior distribution is biasing the results. The red lines and blue lines in the figures correspond to the prior PDF and posterior PDF respectively on the TPP parameters. As shown in Supplementary Figs. S7 to S10, with the exception of the HawkesPLPP, the prior PDFs are widely uninformative priors and the data drives the posterior PDF. In future work, we will try wider prior PDFs for the HawkesPLPP.

Ripley K Estimator

For self-exciting point processes, each onset triggers other onsets with high probability, but it is not necessarily true that the onset times cluster temporally. A method to investigate the aggregation / clustering of onsets is to use the non-parametric Ripley's K-function summary[50], which is a reduced second-moment measure. It may be thought of as a normalized (with respect to the rate of events) version of the expected number of onsets within a window of interval $2t$ of any random onset time selected (Supplementary Equation S1). For a HPP, theoretically $K(t) = 2t$. If $K(t) > 2t$, the process is said to be over-dispersed relative to the HPP and thus exhibits some degree of clustering. If $K(t) < 2t$, the process is said to be under-dispersed related to the HPP and thus onsets occur towards regular intervals. As shown in Supplementary Fig. S11, many of the observation sessions display some form of temporal clustering, indicating that self-exciting processes are an appropriate model choice.

Classification Performance of TPPs

To evaluate classification performance, we assess the TPP model's capability to predict if any onset will occur within a time window $[t, t+\Delta t]$. A starting point $t_{\text{start}} \in [\Delta t, T-\Delta t]$ is randomly selected, with aggressive behavior onsets in $[0, t_{\text{start}}]$ forming the history H, and those in $[t_{\text{start}}, t_{\text{start}}+\Delta t]$ constituting the forecast goal. We calculate the probability of atleast one onset occurring within an arbitrary window as $P[t_{\text{start}} < t_{\text{onset}} <= t_{\text{start}}+\Delta t] = 1-\exp(\Lambda(t_{\text{start}}+\Delta t)-\Lambda(t_{\text{start}}))$ with the history subsetted as $H_{t_s start}$. The ground truth label for the corresponding sampled time window is 1 if any onsets occur within the window and 0 otherwise. Again, for each observation session S_i, we sample 100 random t_{start} locations and their corresponding probability of atleast one onset to compute the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). An example of two sampled time windows within an observation session is shown in Supplementary Fig. S12. A higher ROC-AUC indicates a better predictive model for the number of aggressive behavior onsets within a future time window. We see in Supplementary Table S6 that the self-exciting processes have the best ROC-AUC, with a large gap between HPP and NHPP. This is because HPP and NHPP do not use the history when making a prediction if any event happens in a given future window. Furthermore, the HPP and NHPP cannot handle the temporal clustering of events that was shown in Supplementary "Ripley K Estimator".

Supplementary Tables

θ	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat	mean	std	median_std	5%	median	95%
μ	0.000037	0.000026	3905.527691	3690.888832	1.001562	0.163399	0.002321	0.002321	0.159582	0.163388	0.167197

Supplementary Table S1. HPP TPP MCMC diagnostics. Table includes the MCMC diagnostic statistics and the summary statistics over MCMC samples of the model parameters.

θ	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat	mean	std	median_std	5%	median	95%
α	0.000107	0.000076	8845.538507	7476.553304	1.000653	0.179480	0.010042	0.010048	0.163490	0.179137	0.196577
k	0.000132	0.000093	8806.309742	7355.212776	1.000527	0.978940	0.012380	0.012380	0.958507	0.978955	0.999088
μ	0.000066	0.000047	8900.544522	7482.490484	1.000722	0.172877	0.006248	0.006252	0.162838	0.172649	0.183549

Supplementary Table S2. NHPP TPP MCMC diagnostics. Table includes the MCMC inference diagnostic statistics and the summary statistics over MCMC samples of the model parameters.

θ	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat	mean	std	median_std	5%	median	95%
α	0.000134	0.000095	12090.076945	9771.538738	0.999996	0.896642	0.014732	0.014733	0.872701	0.896454	0.921383
β	0.000125	0.000089	12491.371402	10076.651507	1.000372	0.362435	0.013974	0.013977	0.339953	0.362132	0.385803
μ	0.000010	0.000007	12836.346675	9788.062905	1.000588	0.022929	0.001080	0.001080	0.021188	0.022921	0.024746

Supplementary Table S3. HawkesExpPP TPP MCMC diagnostics. Table includes the MCMC inference diagnostic statistics and the summary statistics over MCMC samples of the model parameters.

θ	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat	mean	std	median_std	5%	median	95%
α_0	0.000460	0.000325	9360.267378	7190.472066	1.001703	0.555669	0.044355	0.044365	0.481628	0.556617	0.627850
α_1	0.000403	0.000286	9467.247474	7473.615216	1.001967	0.401525	0.039339	0.039349	0.338327	0.400637	0.468482
β_0	0.000787	0.000566	9771.149405	7521.710565	1.000970	0.769212	0.076171	0.076372	0.654237	0.763688	0.902572
β_1	0.000170	0.000120	9193.668140	6901.310115	1.000258	0.089091	0.016494	0.016509	0.063173	0.088405	0.116994
c	0.000148	0.000105	10319.143976	6733.623358	0.999982	0.115363	0.015064	0.015065	0.090902	0.115163	0.140543
μ	0.000009	0.000006	13996.544463	8639.769974	1.000117	0.019584	0.001033	0.001034	0.017925	0.019566	0.021304

Supplementary Table S4. Hawkes2ExpPP TPP MCMC diagnostics. Table includes the MCMC inference diagnostic statistics and the summary statistics over MCMC samples of the model parameters.

θ	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat	mean	std	median_std	5%	median	95%
c	0.004642	0.003451	2989.559861	1857.791296	1.001330	1.608953	0.243644	0.244322	1.240103	1.590764	2.042069
k	0.008509	0.006804	2919.654330	1770.599360	1.001483	1.259888	0.394105	0.401316	0.773996	1.184151	2.005617
μ	0.000012	0.000009	7164.001917	6350.680372	1.002622	0.019955	0.001051	0.001051	0.018258	0.019942	0.021690
p	0.002076	0.001497	2940.716305	2072.339469	1.002068	1.837775	0.109316	0.109461	1.666977	1.832133	2.029415

Supplementary Table S5. HawkesPLPP TPP MCMC diagnostics. Table includes the MCMC inference diagnostic statistics and the summary statistics over MCMC samples of the model parameters.

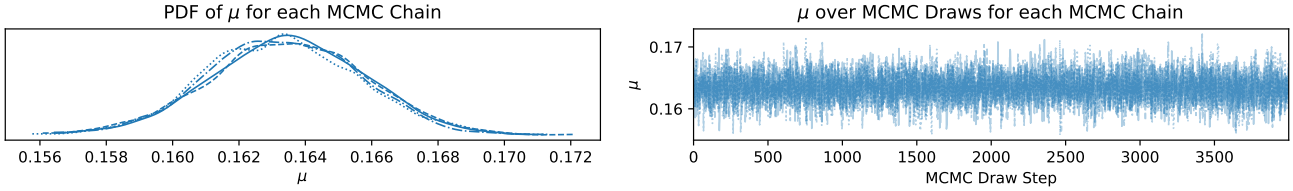
Δt (min)	# of Samples	# of + Samples	# of - Samples	HPP	NHPP	HawkesExpPP	HawkesPLPP	Hawkes2ExpPP
1	42900	4764	38136	0.499	0.545	0.858	0.852	0.846
5	42400	11088	31312	0.500	0.558	0.822	0.817	0.818
10	41700	14381	27319	0.494	0.549	0.800	0.803	0.800
15	40900	16040	24860	0.506	0.540	0.782	0.787	0.786
20	39900	17300	22600	0.498	0.526	0.769	0.777	0.776
25	38500	17577	20923	0.500	0.501	0.763	0.771	0.769

Supplementary Table S6. ROC-AUC for predicting if any onset will occur in a time interval of size Δt (min). Number of samples used to calculate the ROC-AUC is $|S| \times 100$, where $|S|$ may change depending on if it is valid or not because the size of the sampled window may exceed the observation session duration. Number of positive and negative samples is the number of samples with class 1 and 0, respectively.

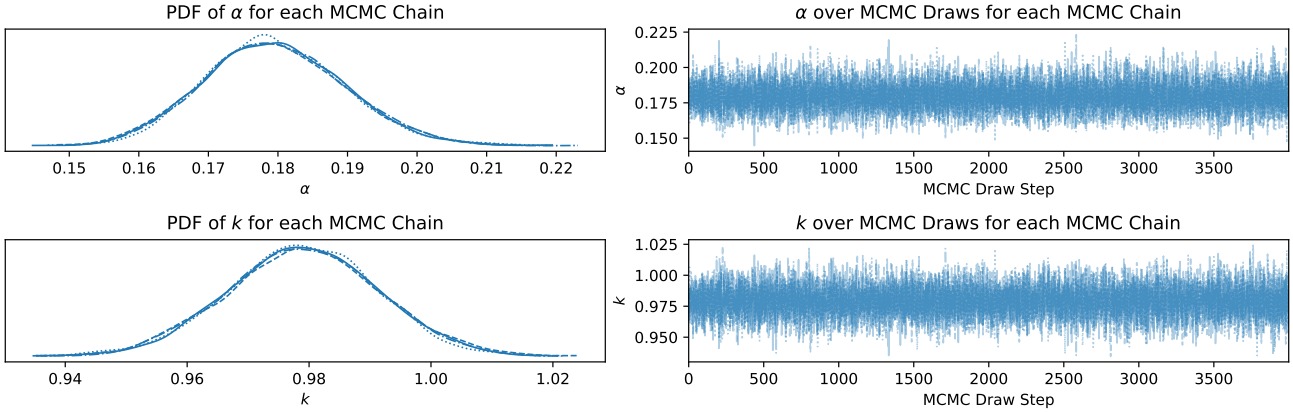
Patient ID	# Of Onsets	# of Observation Sessions	Total Observation Time (min)	Time (min)/ Session
1234.01	1287	24	2000.08	83.34
3207.01	741	7	528.88	75.55
3007.01	481	7	629.65	89.95
5056.01	407	17	940.49	55.32
3216.01	195	7	434.00	62.00
4107.01	150	16	1052.95	65.81
3204.01	140	9	931.94	103.55
3069.01	118	3	262.33	87.44
3174.01	98	8	376.91	47.11
4020.04	94	16	719.15	44.95
4371.01	79	5	339.70	67.94
4262.01	79	7	211.61	30.23
1094.01	77	25	2263.22	90.53
1234.02	65	3	253.31	84.44
4358.01	64	20	746.63	37.33
1206.01	57	4	386.73	96.68
1224.01	55	6	740.04	123.34
4356.01	55	4	177.53	44.38
4280.01	54	5	203.55	40.71
1247.01	50	20	1597.28	79.86
3141.01	44	5	446.34	89.27
3172.01	42	7	377.38	53.91
1229.01	38	11	949.90	86.35
1236.01	37	13	987.37	75.95
4348.01	33	7	550.03	78.58
1235.01	33	9	1088.57	120.95
4343.01	28	4	164.43	41.11
4293.04	26	13	792.47	60.96
1253.01	22	5	558.22	111.64
4365.01	22	9	330.09	36.68
3171.01	21	6	267.56	44.59
1241.01	21	8	657.87	82.23
3238.01	19	3	204.75	68.25
4407.01	19	4	296.57	74.14
1205.01	14	7	772.40	110.34
4354.01	13	7	516.60	73.80
1175.01	12	6	384.54	64.09
4401.01	10	6	256.91	42.82
4409.01	9	1	77.82	77.82
1233.01	9	18	908.44	50.47
4384.01	9	6	278.54	46.42
1223.01	8	5	686.19	137.24
1181.01	6	10	644.56	64.46
4183.01	6	6	252.44	42.07
4353.01	5	5	241.64	48.33
4090.01	5	1	75.77	75.77
4398.01	4	1	71.95	71.95
2036.01	3	1	22.32	22.32
4342.01	2	2	129.08	64.54
4381.01	2	3	196.33	65.44
4380.01	1	2	103.88	51.94
4395.01	1	2	80.94	40.47
1220.01	1	6	762.71	127.12
1244.01	0	1	4.98	4.98
1238.01	0	1	112.40	112.40
1240.01	0	1	64.95	64.95
1242.01	0	1	49.75	49.75
4392.01	0	1	66.48	66.48
4391.01	0	1	40.00	40.00
4389.01	0	1	26.78	26.78
4169.01	0	1	46.63	46.63
1245.01	0	1	73.85	73.85
1249.01	0	1	27.18	27.18
1255.01	0	1	71.05	71.05
4357.01	0	1	45.85	45.85
1232.01	0	1	79.50	79.50
1230.01	0	1	39.83	39.83
3177.01	0	1	67.88	67.88
3237.01	0	1	21.15	21.15
3205.01	0	1	57.98	57.98
Mean	69.59	6.13	426.10	65.84
Median	13.5	5	264.94	65.20
Std	187.71	5.84	452.16	27.11

Supplementary Table S7. Data Statistics. The number of aggressive behavior onsets, the total observation time across all observation sessions, and the total number of observation sessions, per patient.

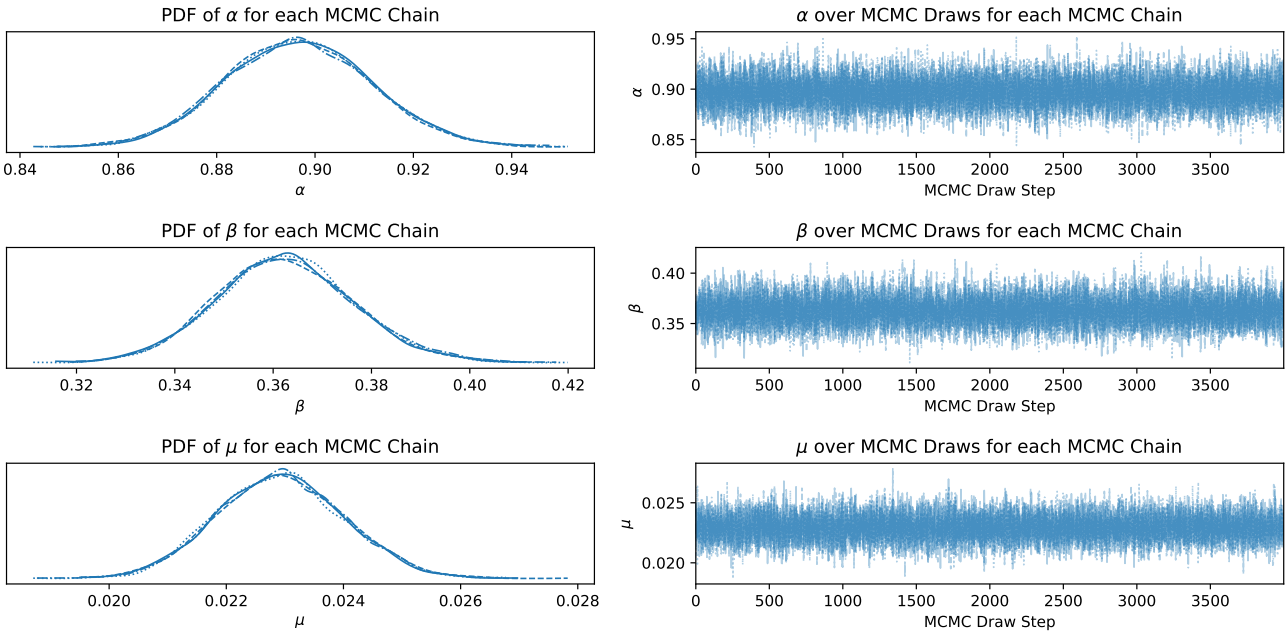
Supplementary Figures



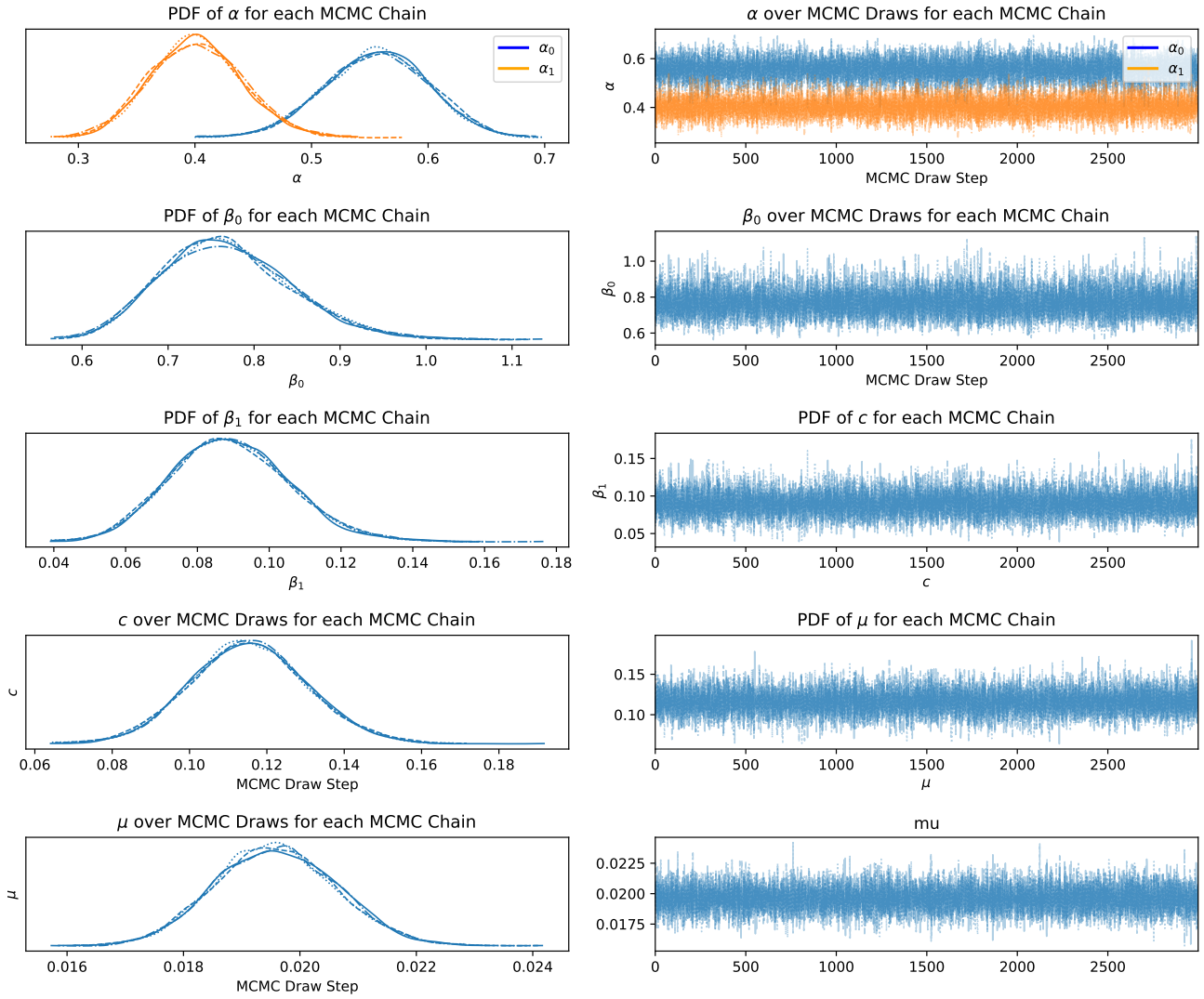
Supplementary Figure S1. Trace plots for HPP. Left column: Posterior probability density functions (PDFs) for each TPP parameter, estimated from MCMC samples. Right column: Autocorrelation plots for each parameter's MCMC chain. Each subplot shows results from four independent MCMC chains, represented by four different lines.



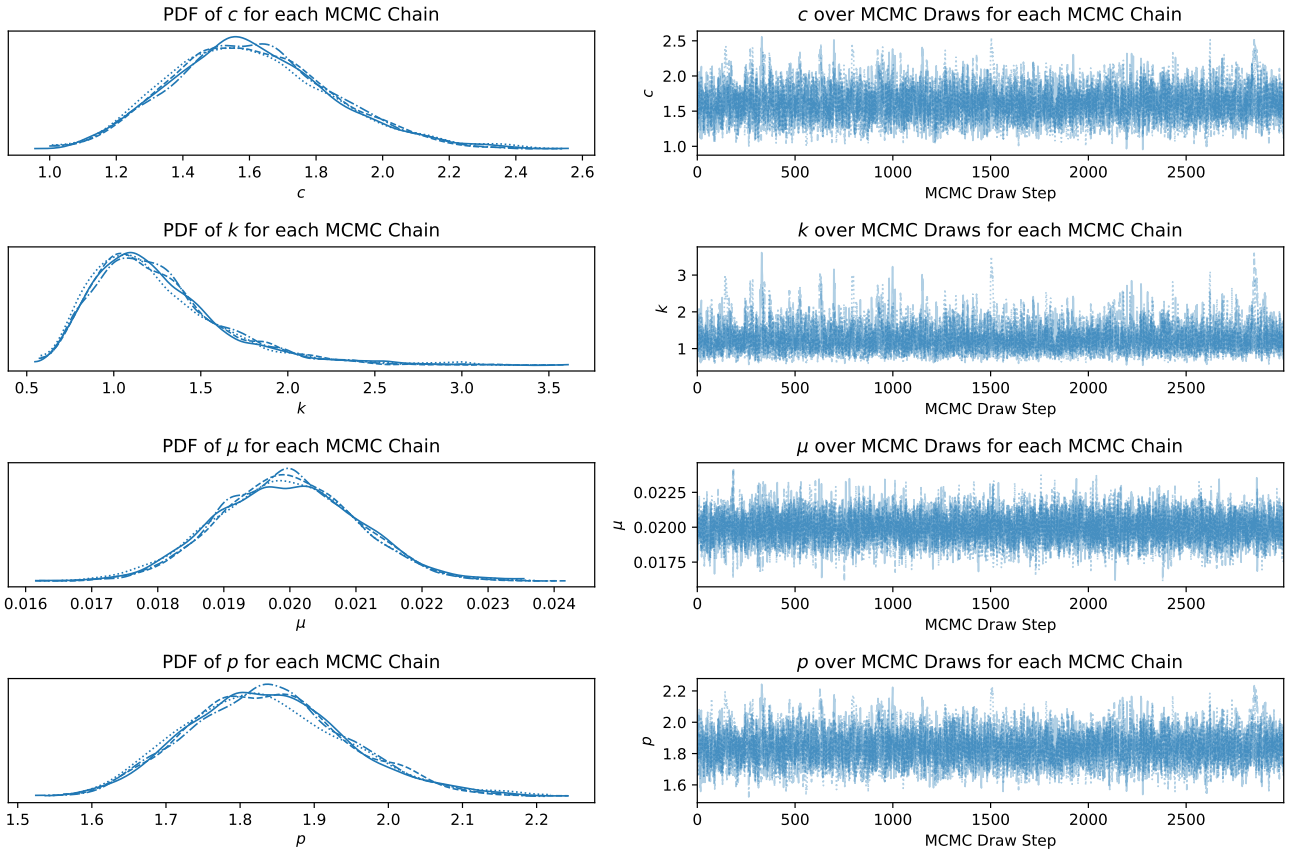
Supplementary Figure S2. Trace plots for NHPP. Left column: Posterior probability density functions (PDFs) for each TPP parameter, estimated from MCMC samples. Right column: Autocorrelation plots for each parameter's MCMC chain. Each subplot shows results from four independent MCMC chains, represented by four different lines.



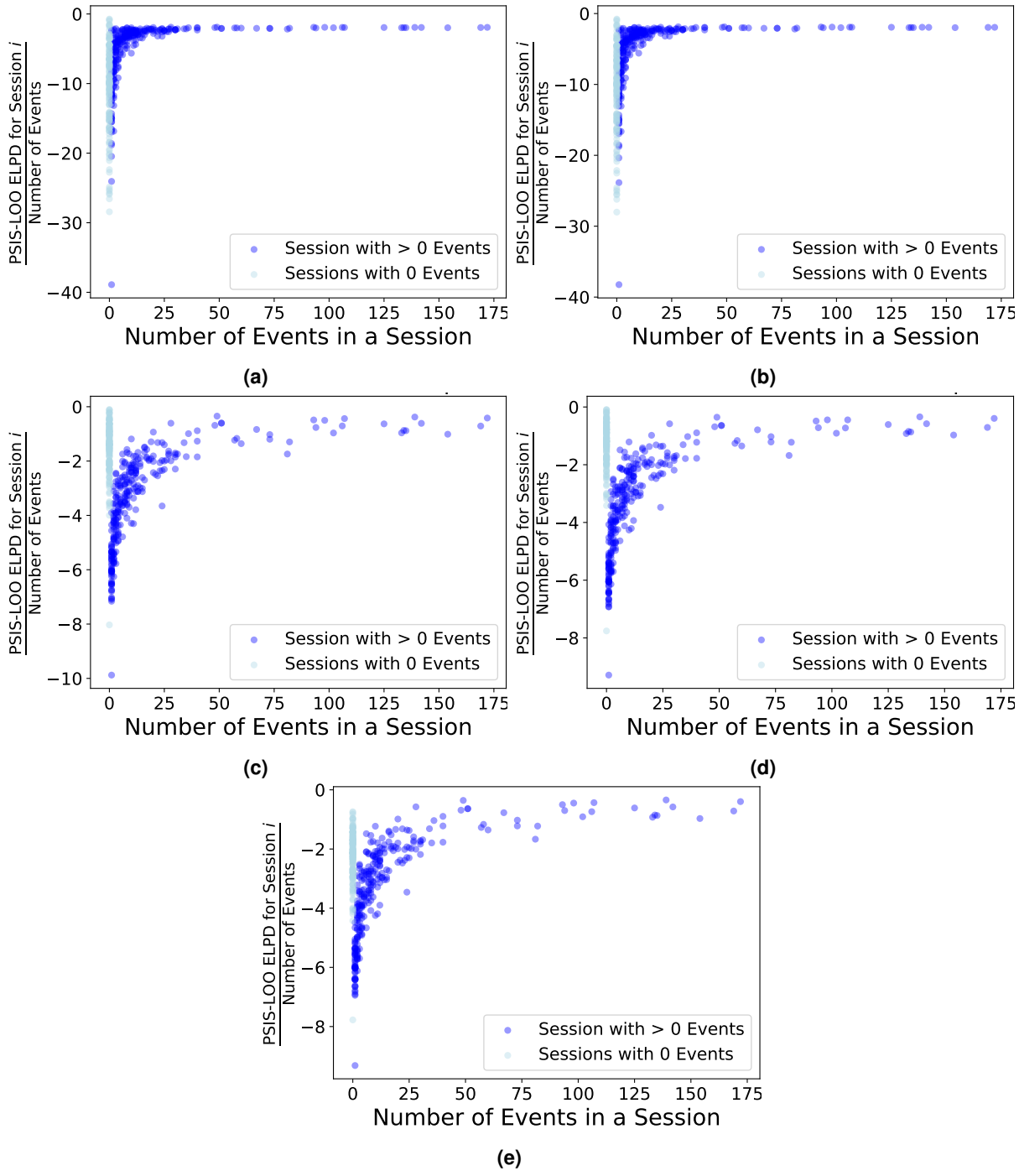
Supplementary Figure S3. Trace plots for HawkesExpPP. Left column: Posterior probability density functions (PDFs) for each TPP parameter, estimated from MCMC samples. Right column: Autocorrelation plots for each parameter's MCMC chain. Each subplot shows results from four independent MCMC chains, represented by four different lines.



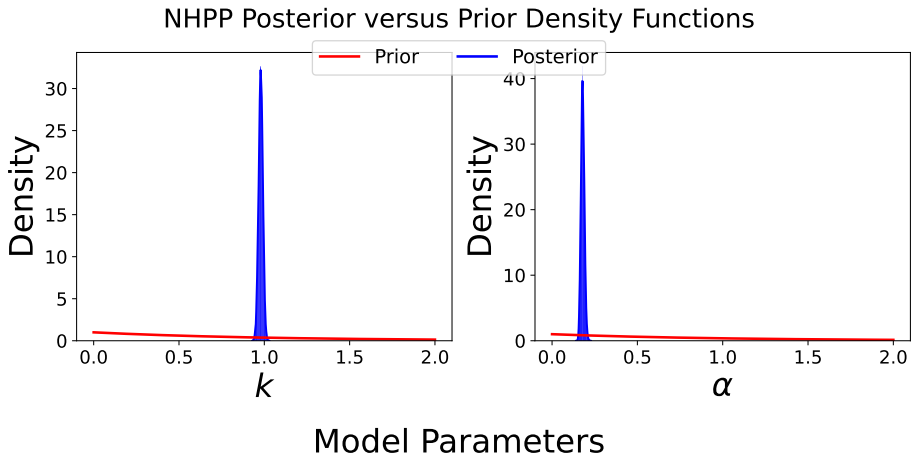
Supplementary Figure S4. Trace plots for Hawkes2ExpPP. Left column: Posterior probability density functions (PDFs) for each TPP parameter, estimated from MCMC samples. Right column: Autocorrelation plots for each parameter's MCMC chain. Each subplot shows results from four independent MCMC chains, represented by four different lines.



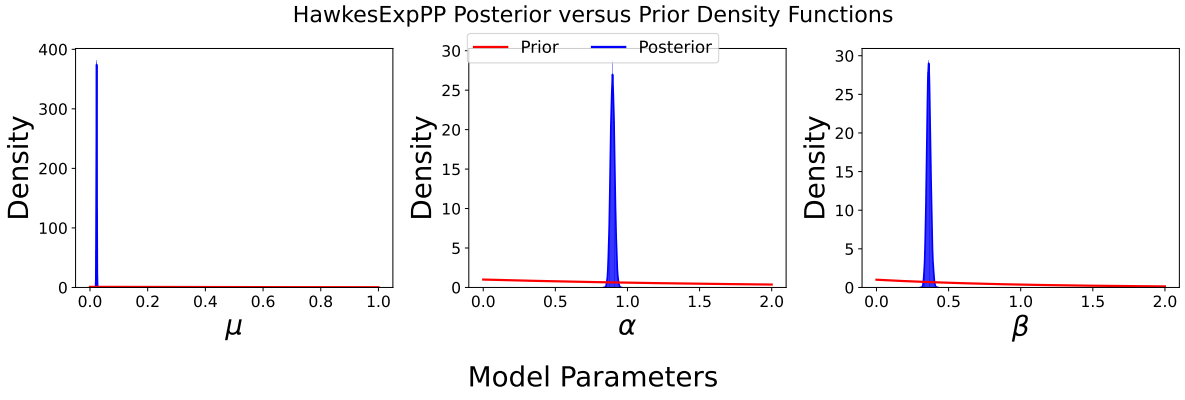
Supplementary Figure S5. Trace plots for HawkesPLPP. Left column: Posterior probability density functions (PDFs) for each TPP parameter, estimated from MCMC samples. Right column: Autocorrelation plots for each parameter's MCMC chain. Each subplot shows results from four independent MCMC chains, represented by four different lines.



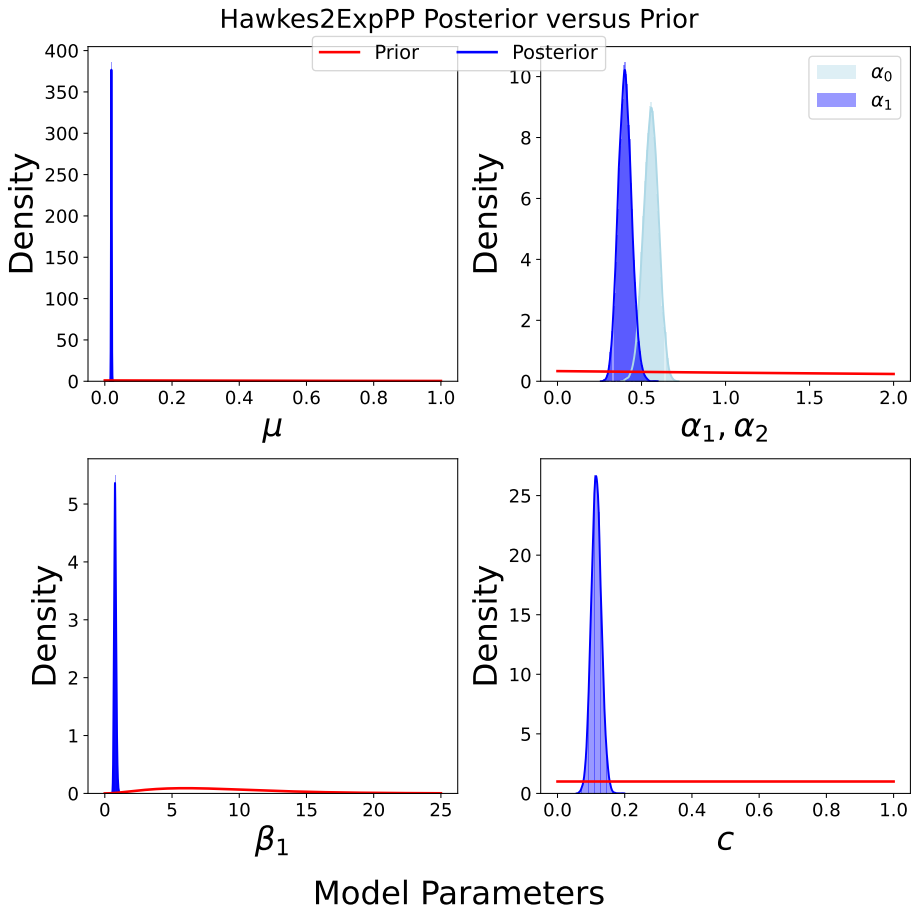
Supplementary Figure S6. Visualization of the PSIS-LOO ELPD for each observation session normalized by number of aggressive behavior onsets versus the number of aggressive behavior onsets observed within an observation session. The light blue scatter points correspond to the PSIS-LOO ELPD for observation sessions with no onsets, whereas the darker blue scatter points correspond to PSIS-LOO ELPD for observation sessions with atleast one onset. For sessions with no events, we simply divide the PSIS-LOO ELPD by 1. (a),(b),(c),(d),(e) correspond to the HPP, NHPP, HawkesExpPP, Hawkes2ExpPP and HawkesPLPP TPPs respectively.



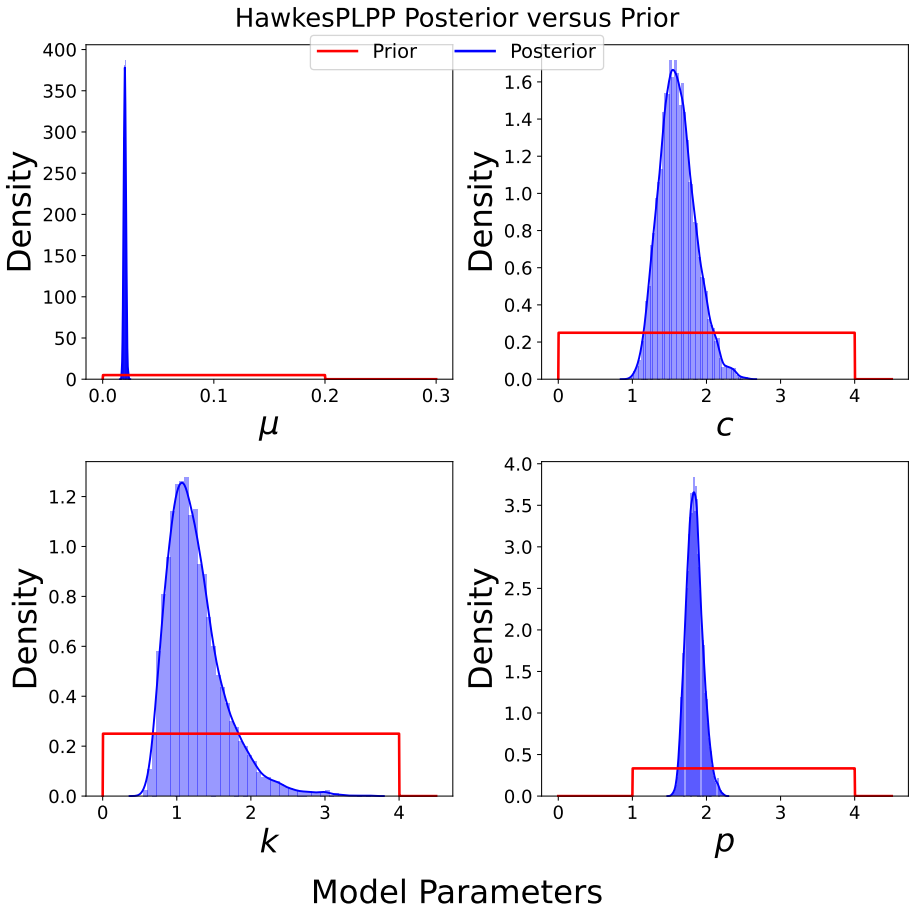
Supplementary Figure S7. Posterior versus prior PDF for NHPP parameters. The prior distributions for the NHPP parameters are: $k \sim \text{Gamma}(1,1)$, $\alpha \sim \text{Gamma}(1,1)$, and $\mu \sim \text{Gamma}(1,1)$.



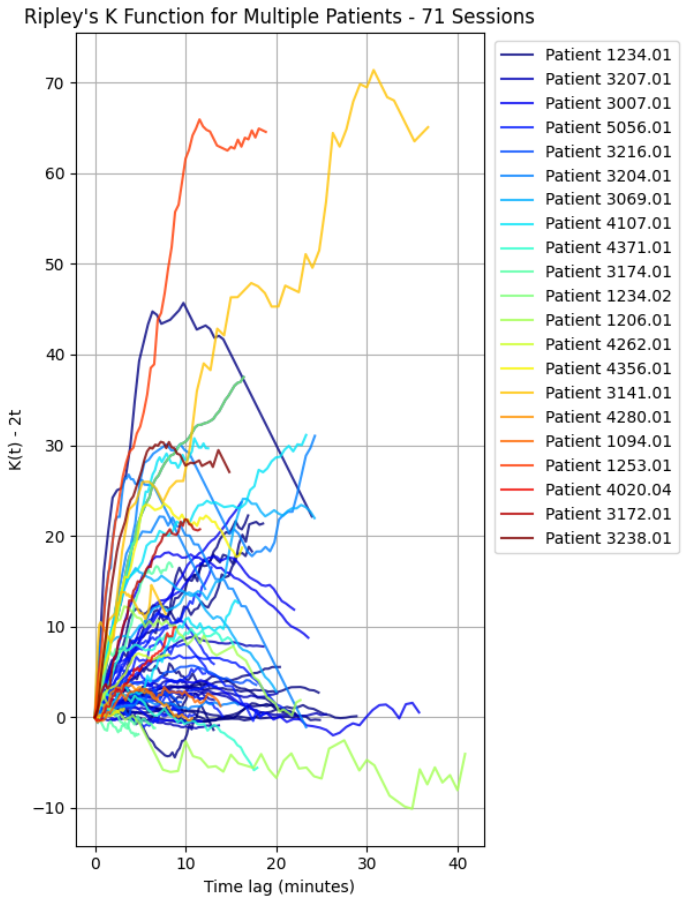
Supplementary Figure S8. Posterior versus prior PDF for HawkesExpPP parameters. The prior distributions for the HawkesExpPP parameters are: $\mu \sim \text{Gamma}(1,1)$, $\alpha \sim \text{Gamma}(1,1)$, and $\beta \sim \text{Gamma}(1,1)$.



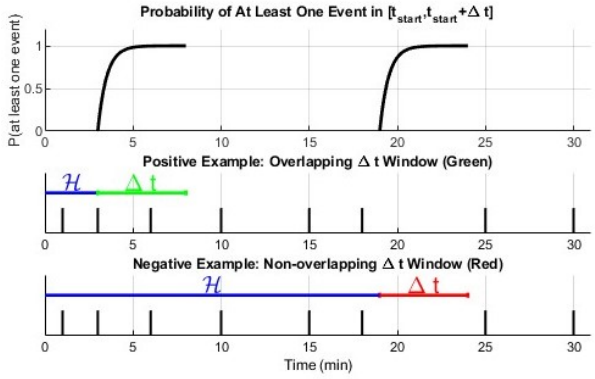
Supplementary Figure S9. Posterior versus prior PDF for Hawkes2ExpPP parameters. The prior distributions for the Hawkes2ExpPP parameters are: $\mu \sim \text{Gamma}(1,1)$, $\alpha_0 \sim \text{Gamma}(1,3)$, $\alpha_1 \sim \text{Gamma}(1,3)$, $\beta_0 \sim \text{Gamma}(3,3)$, $\beta_1 \sim \text{Gamma}(3,3)$, and $c \sim \text{Uniform}(0,1)$.



Supplementary Figure S10. Posterior versus prior PDF for HawkesPLPP parameters. The prior distributions for the HawkesPLPP parameters are: $\mu \sim \text{Uniform}(0,2)$, $c \sim \text{Uniform}(0,4)$, $k \sim \text{Uniform}(0,4)$, and $p \sim \text{Uniform}(1,4)$.



Supplementary Figure S11. For each observation session with atleast 15 onsets, we visualize the Ripley K estimator for varying values of t with reference to the HPP Ripley K estimator $2t$. The time lag is t.



Supplementary Figure S12. Two examples of sampling random windows and corresponding history. The black vertical lines in the bottom two axes are the timestamps where onsets occur. The history of the window is the blue horizontal line, while the positive and negative window is the green and red respectively horizontal line. A postive window means the label for the window is 1 because at least 1 event occurred in the window, while a negative window means the label for a window is 0 because no events happen in the window.

Supplementary Equations

$$\hat{K}(t) = \frac{T}{J^2} \sum_{j=1}^{J} \sum_{i \neq j} w_{ij} \mathbb{1}(|t_i - t_j| \leq t) \tag{S1}$$

where $\frac{T}{J}$ is number of onsets per unit time, J is the number of onsets observed in an observation session, w_{ij} is an edge correction taking values $w_{ij} = 1$ if $|t_i - t_j| \leq \min(t_i, T - t_i)$ and $w_{ij} = 2$, otherwise.