




Context-aware spatio-temporal event prediction via convolutional Hawkes processes

Maya Okawa^{1,3}  · Tomoharu Iwata² · Yusuke Tanaka² · Takeshi Kurashima¹ · Hiroyuki Toda¹ · Hisashi Kashima³

Received: 3 May 2021 / Revised: 14 December 2021 / Accepted: 6 February 2022 /

Published online: 21 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media LLC, part of Springer Nature 2022

Abstract

Massive spatio-temporal event data sets are now available that cover events such as disease outbreaks, armed conflicts and crimes. Predicting such events and revealing the underlying triggering patterns are a crucial task for many applications, ranging from disease control to global politics. Traditional event prediction models based on Hawkes processes capture the spatio-temporal relationships between events, but cannot incorporate complex and heterogeneous external features, including population distribution, weather and terrain. This paper proposes an event prediction method that effectively utilizes the rich external information present in sets of unstructured data (e.g., map images, satellite images and weather map). Specifically, we extend a convolutional neural network (CNN) by combining it with continuous kernel convolution; and design the conditional intensity of Hawkes process based on the extended neural network model that accepts images as its input. Our approach of using the continuous convolution kernel provides a flexible way to discover the complex effect of external factors on the triggering process, as well as yielding tractable optimization algorithms. We use real-world event data from different domains (i.e., disease outbreaks, armed conflicts and protests) to demonstrate that the proposed method has better prediction performance than existing methods.

Keywords Point process · Deep learning · Event prediction · Hawkes process

Editors: Annalisa Appice, Sergio Escalera, Jose A. Gamez, Heike Trautmann.

✉ Maya Okawa
maya.okawa.af@hco.ntt.co.jp

¹ NTT Human Informatics Laboratories, NTT Corporation, 1-1, Kanagawa, Yokosuka, Japan

² NTT Communication Science Laboratories, NTT Corporation, Kyoto, Japan

³ Department of Intelligence Science and Technology, Kyoto University, Kyoto, Japan

1 Introduction

Spatio-temporal event data are being accumulated in many important fields such as health care and public safety. Such data contains time and location, indicating when and where events have happened. For example, electronic health records are represented as a sequence of events with locations and times of disease outbreaks. Armed conflicts are recorded with locations and times at which the conflicts took place.

A wide range of event sequences are demonstrations of spatio-temporal processes that have “self-exciting” or triggering patterns. In all the aforementioned examples, event occurrence is triggered by preceding events. For instance, disease outbreak can ignite secondary outbreaks, often leading to epidemics. A conflict between rival ethnic groups may trigger a cycle of retaliation.

Modeling such triggering processes and predicting future events is crucial for realizing many applications such as disease control and harmonizing global politics. For instance, if local health authorities can predict when, where and which events will trigger disease outbreaks, they can make more effective intervention policies (Wagner et al., 2011). Better understanding and prediction of conflicts will help governments take more appropriate actions to reduce life and economic losses.

Hawkes process is a general mathematical framework for modeling triggering processes; it is characterized by a *conditional intensity* that describes the rate of events occurring at any location and at any time. Hawkes process has been adopted for modeling a wide spectrum of events, including infectious disease (Reinhart, 2018), terrorist attacks (Porter & White, 2012), crimes (Mohler et al., 2011) and earthquakes (Ogata et al., 2003). However, these models fail to adequately depict the real diffusion process, since its conditional intensity is modeled as a function of spatio-temporal distance, and the impact of external factors on triggering processes is ignored. Real-world triggering processes are determined not only by the spatio-temporal relationship between events but also by external factors such as population distribution, weather, road network and terrain. These external features can be spatially heterogeneous and change over time. For example, infectious diseases spread among high population areas (Morse, 2001). The transmission of diseases is also influenced by other external factors, including trading patterns (Nicolas et al., 2013), land use (Patz et al., 2004) and weather (Parham & Michael, 2010). Conflicts tend to be more accentuated in densely populated areas (Lee et al., 2019).

One promising approach to capturing the spatial heterogeneity of the triggering process is to incorporate external factors, e.g., population, weather and road network. Nowadays, rich external information sets are becoming accessible. For example, with the development of remote sensing techniques, high resolution satellite images are being collected and are available at various spectral, spatial and temporal resolutions. Also, open-source GIS platforms have become commonplace; they provide geographic features including road network and land use, in the form of a colored map. These images contain meaningful information that can rarely be found in traditional information sources, and offer detailed spatial patterns of various external factors, ranging from human demography to weather and land use, as well as their temporal variations.

Several studies (Kim et al. (2017; Meyer, 2018; Servadio et al., 2018) have extended Hawkes process to incorporate external factors, e.g., regional populations (Meyer, 2018), mobility flows between regions (Kim et al., 2017) and weather conditions (Servadio et al., 2018). But these methods are based on hand-crafted features engineered by domain experts, and make a simplified assumption on the conditional intensity as a function of

these features. Thus these methods cannot handle unstructured data like images, which contain rich, meaningful information.

In this paper, we propose an event prediction method that effectively utilizes the rich external features present in georeferenced images. Inspired by the recent success of deep learning models in computer vision (Vaswani et al., 2017; Zhang et al., 2018), we use them to enhance the Hawkes process model. The most straightforward way is to directly replace the Hawkes process intensity by a neural network that accepts these images as its input. Although this approach enables the automatic discovery of meaningful information from the images and thus improve event prediction performance, it suffers from the intractable optimization problem, as integral computations are required to determine the likelihood needed for estimation.

We solve this by introducing a novel architecture for Hawkes processes. In particular, we extend a convolutional neural network (CNN) by combining it with continuous kernel convolution; the conditional intensity of Hawkes process is designed on the extended model. Our approach of using the continuous convolution kernel provides a flexible way of learning the complex external features present in the images, allowing us to capture the spatial heterogeneity of the triggering process. Notably, our formulation permits the likelihood to be determined by tractable integration. In the proposed method, referred to as Convolutional Hawkes process (**ConvHawkes**), the parameters of the neural network and the convolutional kernel can be simultaneously optimized to maximize the likelihood by using gradient-based algorithms.

We conduct experiments on three real-world datasets from multiple domains and show that **ConvHawkes** consistently outperforms existing methods in event prediction tasks. The experiments also demonstrate that **ConvHawkes** provides a better understanding of the underlying mechanisms by which various external factors influence the triggering processes.

The main contributions of this paper are as follows:

- We propose a novel Hawkes process model, **ConvHawkes** (Convolutional Hawkes process) for modeling diffusion processes and predicting spatio-temporal events. It accurately and effectively predicts spatio-temporal events by leveraging the external features contained in georeferenced images (e.g., satellite images and map images), that impact triggering processes.
- We present an extension of the neural network model and integrate it into the Hawkes process framework. This formulation allows us to utilize the external features present in the unstructured image data, and to automatically discover their complex effects on the triggering process, while at the same time yielding tractable optimization.
- We conduct extensive experiments on real-world datasets from different domains. With regard to event occurrence, the proposed method achieves better predictive performance than several existing methods on all datasets (Sect. 6).

2 Preliminaries

This section starts by providing the theoretical background to spatio-temporal Hawkes processes.

Point process is a random sequence of event occurrences over a domain. We assume here a sequence of events with known times and locations. Let (t, \mathbf{s}) be the event written as

the pair of time $t \in \mathbb{T}$ and location $s \in \mathbb{S}$, where $\mathbb{T} \times \mathbb{S}$ is a subset of $\mathbb{R} \times \mathbb{R}^2$. We denote the number of events falling in subset A of $\mathbb{T} \times \mathbb{S}$ as $N(A)$. The general approach to identifying a point process is to estimate “intensity” function $\lambda(t, s)$. Intensity $\lambda(t, s)$ represents the rate of event occurrence in a small region. Given the history $\mathcal{H}(t)$ up to t , intensity is defined as

$$\lambda(t, s|\mathcal{H}(t)) \equiv \lim_{|dt| \rightarrow 0, |ds| \rightarrow 0} \frac{\mathbb{E}[N(dt \times ds)|\mathcal{H}(t)]}{|dt||ds|}, \tag{1}$$

where dt is a small interval around time t , $|dt|$ is its length and ds is a small region containing location s , $|ds|$ is its area. \mathbb{E} is an expectation term. The functional form of intensity is designed to appropriately capture the underlying dynamics of event occurrence.

The Hawkes process is an important class of point process models, and its intensity is modeled as the cumulative effects from all the past events $\mathcal{H}(t)$, represented by

$$\lambda(t, s|\mathcal{H}(t)) = \mu + \sum_{i:t_i < t} \alpha_i g(t - t_i, s - s_i), \tag{2}$$

where μ is a base intensity independent of the preceding events. t_i and s_i is the time and location of the i -th event; α_i is a constant that represents the strength of the influence of the i -th event; $g(\cdot) \geq 0$ is a triggering kernel that specifies the decaying effect of the i -th event. For computational simplicity, the triggering kernel function is often factorized into temporal and spatial components as follows:

$$g(t - t_i, s - s_i) = g_1(t - t_i)g_2(s - s_i), \tag{3}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are temporal and spatial decay functions, respectively. Typical choices for the temporal decay function include power-law, exponential, and Rayleigh functions (Mishra et al., 2016). Gaussian kernel is commonly used as the spatial decay function.

Given a sequence of events, $\mathcal{D} = \{(t_n, s_n)\}_{n=1}^N$, $t_n \in \mathbb{T}$ and $s_n \in \mathbb{S}$, the likelihood is given by

$$p(\mathcal{D}|\lambda(t, s)) = \prod_{n=1}^N \lambda(t_n, s_n) \cdot \exp\left(-\int_{\mathbb{T} \times \mathbb{S}} \lambda(t, s) dt ds\right). \tag{4}$$

3 Problem definition

This subsection formally defines the problem of spatio-temporal event prediction.

Event Sequence Each event is represented by the tuple (t, s) , where $t \in \mathbb{T} \subseteq \mathbb{R}$ denotes its time and $s \in \mathbb{S} \subseteq \mathbb{R}^2$ is its location (i.e., latitude and longitude). We assume that we have a sequence of N events up to time T , denoted by $\mathcal{D} = \{(t_n, s_n)\}_{n=1}^N$.

Image Sequence Additionally, we have an image dataset (e.g., satellite image, night light image, weather map). The image dataset is represented as a sequence of images, e.g., a collection of satellite images acquired at different times covering the area of interest \mathbb{S} . An image dataset example is presented on the left in Fig. 2. Formally, we denote $I \in \mathbb{R}^{C \times H \times W}$ as the image, where H and W are image height and width, respectively; C is the number of channels. Each image is annotated with time τ when the observation was made. Each pixel of image $I[h, w]$ is georeferenced and corresponds to a fixed geospatial area (e.g., 500 m by 500 m). The corresponding latitude/longitude coordinates of the geospatial area for the (h, w) -th pixel are represented by $\mathbf{x}_{h,w}$, where $\mathbf{x}_{h,w}$ is the coordinates of the pixel center. For

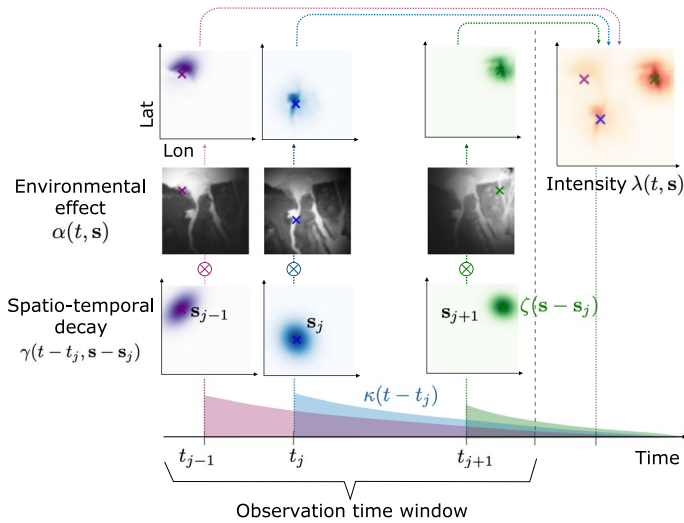


Fig. 1 Illustration of the proposed method

specific kinds of images (e.g., weather map), besides historical sequence, future sequence of the images (e.g., weather forecast maps) is available. Let $\mathcal{I} = \{(I_l, \tau_l)\}_{l=1}^L$ be the sequence of images over the time window $[0, T + \Delta T)$, where L is the number of observations.

Event Prediction Problem Given the event sequence \mathcal{D} in the observation time window $[0, T)$, and the image dataset \mathcal{I} in the time period $[0, T + \Delta T]$, we aim to

- predict the number of events within any given spatial area and time period in $[T, T + \Delta T]$
- predict times and locations of events in the future time window $[T, T + \Delta T]$,

by leveraging \mathcal{D} and \mathcal{I} .

4 Convolutional Hawkes processes

This section presents the proposed method for spatio-temporal event prediction, referred to as **ConvHawkes** (Convolutional Hawkes process). We provide the model formulation of ConvHawkes followed by parameter learning and prediction.

4.1 Model overview

We propose a novel extension of Hawkes process for modeling triggering processes and predicting spatio-temporal events. The triggering processes are significantly influenced by the external factors such as population, weather, road network and terrain.

The idea of this work is to leverage image data (e.g., satellite image and weather map) to capture such heterogeneity in the external factors and determine their effect on the triggering process. To this end, we incorporate the neural network model into the Hawkes process formulation. We illustrate our method in Fig. 1. Specifically, we extend the neural network

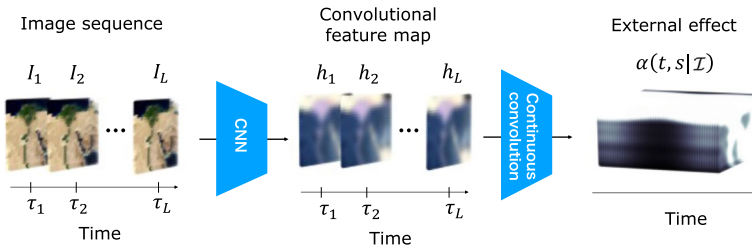


Fig. 2 Overall architecture of the external effect module

that learns the influence of the external factors by incorporating continuous kernel convolution, and parameterize the Hawkes process intensity based on the extended model. The proposed model learns latent external features from georeferenced images; and also learns external effects at each location, while at the same time providing tractable learning.

4.2 Model formulation

We develop a flexible and tractable framework based on Hawkes process to learn the underlying external effects and spatio-temporal relationships between events from image data, e.g., satellite image, map image and weather map. Formally, ConvHawkes designs the conditional intensity as follows:

$$\lambda(t, \mathbf{s} | \mathcal{H}(t)) = \mu + \sum_{j: t_j < t} \underbrace{\alpha(t, \mathbf{s} | \mathcal{I})}_{\text{external effect}} \underbrace{\gamma(t - t_j, \mathbf{s} - \mathbf{s}_j)}_{\text{Spatio-tempora decay}}, \tag{5}$$

where μ is the background rate of event occurrence. As seen in Eq. (5), our model consists of two components: *external effect* and *spatio-temporal decay*. The external effect $\alpha(\cdot)$ is specified by a neural network function, which captures the influence of the external factors. The spatio-temporal decay $\gamma(\cdot)$ is designed by a triggering kernel function over space and time that describes the decay in the influence of past events with spatio-temporal distance. In the following, we describe the formulation of each component and the rationale behind them.

External effect We model the external effect $\alpha(\cdot)$ based on a neural network model. The architecture of the external effect module is given in Fig. 2.

For each image dataset, the image sequence is first processed by a convolutional neural network (CNN). The CNN is designed such that its output has the same size of the input image sequence, which makes it straightforward to utilize the time stamps, and location information of the images in the subsequent continuous convolution layer. We can use the encoder-decoder-based CNN (Yasrab, 2017; Yasrab et al., 2017), CNN-RNN encoder–decoder (Attia et al., 2017), or other deep neural networks that are suitable for the given image data. In this paper, we choose a simple CNN with N_l layers. As shown in the experimental section (Sect. 6.5), our proposed method produces satisfactory prediction performance even with this simple neural architecture. Each image of the image sequence I_l is fed into the CNN architecture and transformed into the latent feature map \mathbf{h}_l , where $\mathbf{h}_l \in \mathbb{R}^{H \times W \times d}$. Here d is the dimension size of the latent feature map. For the sake of simplicity, we fix $d = 1$ in the experiments.

Next we apply continuous kernel convolution to these latent feature map to expand the learned latent feature map over discrete pixel space onto the continuous spatio-temporal space. Formally, given the latent feature map \mathbf{h}_l and their associated time τ_l and latitude/longitude coordinates for each pixel $\mathbf{x}^{h,w}$, the output of the convolutional layer at time t and location \mathbf{s} is written by

$$\alpha((t, \mathbf{s})|\mathcal{I}) = \sum_t \sum_{h,w} \mathbf{h}_l[h, w] f(t - \tau_l, \mathbf{s} - \mathbf{x}^{h,w}), \quad (6)$$

where $f(\cdot)$ is a convolution kernel defined as continuous functions over the temporal and spatial plane. The definition for the continuous convolution kernel $f(\cdot)$ is provided later in this subsection. $\mathbf{h}_l[h, w] \in \mathbb{R}^d$ denotes the (h, w) -th pixel of latent feature map \mathbf{h}_l . $\alpha(\cdot)$ is a scalar function that quantifies the external effects at time t and location \mathbf{s} . Intuitively, the external feature map $\alpha((t, \mathbf{s})|\mathcal{I})$ indicates how likely an event is to occur at time t and location \mathbf{s} given preceding events that trigger it. This procedure is inspired by the work of Schütt et al. (2017) and Wang et al. (2018), which generalizes the discrete convolution used in standard CNNs to a continuous one. Our method is unique in that it does not require any discrete approximation. The above formulation enables the neural network model to be directly injected in the end-to-end framework of Hawkes process. At the same time, it yields tractable optimization (as discussed in Sect. 4.3).

Continuous convolution kernel To ensure computation simplicity, we factorize the continuous convolution kernel $f(\cdot)$ into temporal and spatial components such that: $f(t - \tau, \mathbf{s} - \mathbf{x}) = h(t - \tau)k(\mathbf{s} - \mathbf{x})$, where $h(\cdot)$ and $k(\cdot)$ are the kernel functions for temporal and spatial convolutions, respectively. In our case, we use the uniform kernel for the temporal convolution, which is defined by

$$h(t - \tau) = \mathbb{1}[\tau - \Delta < t < \tau + \Delta], \quad (7)$$

where $\mathbb{1}[\cdot]$ is an indicator function that indicates 1 when the condition holds, and 0 otherwise; Δ is the binwidth parameter. Without loss of generality, in our experiment, we fix Δ as the time interval between the observations. This is equivalent to piece-wise approximation. If we have no future observations or predictions of the images, the last image in the image sequence is used for prediction. For the spatial convolution, we can select a Gaussian kernel:

$$k(\mathbf{s} - \mathbf{x}) = \exp(-(\mathbf{s} - \mathbf{x})^\top \Sigma_k^{-1}(\mathbf{s} - \mathbf{x})), \quad (8)$$

where Σ_k is a 2×2 covariance matrix (bandwidth) of the kernel. We can use other convolution kernel functions, such as uniform and Rayleigh.

Spatio-temporal decay Following previous work (Rocque et al., 2011; Pratiwi et al., 2017), the spatio-temporal decay kernel functions are taken to be separable in space and time such that: $\gamma(t - t_j, \mathbf{s} - \mathbf{s}_j) = \kappa(t - t_j)\zeta(\mathbf{s} - \mathbf{s}_j)$. Regarding the temporal decay function $\gamma(\cdot)$, the exponential decay function is the standard choice:

$$\kappa(t - t_j) = \exp(-\beta(t - t_j)), \quad (9)$$

where $\beta > 0$ is the decay factor. This implies that the occurrence of an event grows when events occur but their influence decreases exponentially at the rate of β over time.

A typical form of the spatial decay function is based on a Gaussian distribution as follows:

$$\zeta(\mathbf{s} - \mathbf{s}_j) = \exp\left(-(\mathbf{s} - \mathbf{s}_j)^\top \Sigma_\zeta^{-1}(\mathbf{s} - \mathbf{s}_j)\right), \tag{10}$$

where Σ_ζ is a 2×2 covariance matrix (bandwidth) of the kernel. Intuitively, when the j th event occurs, the probability of the next event occurring is higher in the neighborhood of location \mathbf{s}_j . The bandwidth parameter Σ_ζ quantifies how strongly the influence from each past event decays over space. Other kernel functions, such as uniform and Rayleigh are viable alternatives.

4.3 Parameter learning

Given a list of observed events up to time T (total of N events) \mathcal{D} and the image dataset \mathcal{I} , the logarithm of the likelihood function is written as

$$\begin{aligned} \mathcal{L} = & \sum_{n=1}^N \log \left[\mu + \alpha(t_n, \mathbf{s}_n) \sum_{j:t_j < t_n} \gamma(t_n - t_j, \mathbf{s}_n - \mathbf{s}_j) \right] \\ & - \left[\mu T |\mathbb{S}| + \underbrace{\sum_{n=1}^N \int_{t_n}^T \int_{\mathbb{S}} \alpha(t, \mathbf{s}) \gamma(t - t_n, \mathbf{s} - \mathbf{s}_n) dt ds}_{\text{call this } A_n} \right], \end{aligned} \tag{11}$$

where $|\mathbb{S}|$ denotes the area of spatial region \mathbb{S} . The computation difficulty comes from the integral of the neural network function (i.e., CNN) in the external effect $\alpha(\cdot)$ of term A_n . With our formulation, the neural network function \mathbf{h}_l can be moved outside the integral, and A_n is rewritten as

$$A_n = \sum_l \sum_{h,w} \mathbf{h}_l[h, w] \int_{t_n}^T \kappa(t - t_n) h(t - \tau_l) dt \int_{\mathbb{S}} \zeta(\mathbf{s} - \mathbf{s}_n) k(\mathbf{s} - \mathbf{x}^{h,w}) ds. \tag{12}$$

Consequently, we can obtain closed form solutions of the integral in term A_n for standard decay and convolution kernel functions, as shown in Appendix A.2. In the case of the Gaussian kernel pair defined by Eqs. (8) and (10), it is given by an error function (the derivation is provided in Appendix A.2). The resulting log-likelihood is fully tractable, permitting the use of gradient-based algorithms. We apply simple back-propagation for training ConvHawkes. During the training phase, we adopt mini-batch optimization.

4.4 Event number prediction

The point process model can be used to predict the expected number of events by integrating the estimated intensity over specific time period $W_T = [T_p, T_q]$ and the area of interest $W_S \subset \mathbb{S}$ such that

$$\begin{aligned} N(W_T \times W_S) &= \int_{W_T} \int_{W_S} \lambda(t, \mathbf{s}) dt ds \\ &= \sum_l \sum_{h,w} \mathbf{h}_l[h, w] \int_{T_p}^{T_q} \zeta(t - t_n) h(t - \tau_l) dt \int_{W_S} \zeta(\mathbf{s} - \mathbf{s}_n) k(\mathbf{s} - \mathbf{x}^{h,w}) ds, \end{aligned} \tag{13}$$

where $N(A)$ is the number of events that fall into subset A . As mentioned in Sect. 4.3, we can obtain closed form solutions of the above integral.

Moreover, the **ConvHawkes** model can simulate the occurrence time of the next event and its location by adopting the thinning algorithm Reinhart (2018).

5 Related work

Spatio-temporal prediction constitutes an important problem with various applications such as public safety, transportation, health care, and environment. The conventional approach to this problem is regression. Early works are based on traditional machine learning methods, including classical time-series models like vector autoregression (VAR) (Chandra & Al-Deek, 2009; Zivot & Wang, 2006) and autoregressive integrated moving average (ARIMA) (Van Der Voort et al., 1996), and support vector regression (SVR) (Zhang & Liu, 2009). Recently, deep learning models have been successfully applied to this problem. For example, Ma et al. (2015) and Zhao et al. (2017) employ long short-term memory (LSTM) networks for traffic prediction, which captures the long-term temporal dependencies. Several studies (Kang & Kang, 2017; Zhang et al., 2016, 2017) use convolutional neural networks (CNNs) to capture the non-linear spatial dependencies. Yao et al. (2018) combine LSTM and CNN to jointly model both spatial and temporal dependencies in traffic data. In recent literature, graph neural networks (GNNs) have been adopted for spatio-temporal traffic graphs (Guo et al., 2019; Yu et al., 2017; Zhao et al., 2019) and epidemic forecasting (Kapoor et al., 2020) to handle the complex spatio-temporal correlations. However, all the aforementioned methods focus on predicting the aggregated number of events within a predefined spatial region and time interval. This task is fundamentally different from ours. In this paper, we aim to directly model a sequence of events in continuous time and space, without aggregation, by using explicit information about location and/or time.

Point process is a powerful mathematical framework for modeling a sequence of events that occur in a continuous space and/or time domain. Hawkes processes (Hawkes, 1971) have been proven effective in describing the phenomenon of mutual excitation between events (i.e., triggering process); examples include earthquakes and aftershocks (Musmeci & Vere-Jones, 1992; Ogata, 1998; Zhuang et al., 2002), gang-on-gang violence (Louie et al., 2010), terrorist attacks (Porter & White, 2012), near repeat crimes (Mohler et al., 2011; Zhu & Xie, 2019), disease transmission (Choi et al., 2015; Reinhart, 2018), financial transactions (Bacry et al., 2015), and social activities (Blundell et al., 2012; Farajtabar et al., 2015). Early work made fixed parametric assumptions regarding the functional form of the conditional intensity, which is often too restrictive to depict real triggering process. Recent studies employ neural networks to enhance the expressiveness of point processes. For example, Xiao et al. (2017) present a generative adversarial network-based framework for estimating the intensity of an inhomogeneous Poisson process. Chen et al. (2020) leverage neural ODEs to parameterize marked temporal point processes. These models are based on inhomogeneous Poisson processes; they do not directly consider the influence of past events. Some other works (Du et al., 2016; Mei & Eisner, 2017) propose to parameterize the intensity of temporal Hawkes processes by a recurrent neural network (RNN) to learn the non-linear influence from past events. Omi et al. (2019) generalize the RNN-based Hawkes process model to further improve its expressive power. Transformer Hawkes process (Zuo et al., 2020) and self-attentive Hawkes process (Zhang et al., 2020) employ a self-attention mechanism to capture the non-linear temporal correlation between events.

Table 1 Statistics of datasets used in this paper

	Area	Time span	Source	# of events
Conflict	Africa	1 Mar, 2018–31 Mar, 2020	ACLED ¹	16,801
Protest	Middle East	1 Mar, 2018–31 Mar, 2020	ACLED ¹	34,243
Disease	Europe	1 Mar, 2020–31 Aug, 2020	EMPRES-i ²	21,529

These models focus on learning the temporal dependencies between events, and cannot be easily extended to account for the spatial aspect. More recent work (Zhu et al., 2019) extends this approach to spatio-temporal Hawkes processes to consider both spatial and temporal domains. Despite the advances, all the above methods ignore the effects of external factors on the triggering processes.

Some efforts have been made to incorporate external features into Hawkes processes. For instance, several studies have proposed temporal Hawkes process methods that take account of external features such as population density (Meyer, 2018), transportation networks (Aldrin et al., 2015; Wilder-Smith & Gubler, 2008), human mobility patterns (Kim et al., 2017), weather (Mohler et al., 2011; Servadio et al., 2018), fault structure (Musmeci & Vere-Jones, 1992). However, it is still challenging to effectively utilize complex unstructured data like images.

Another line of work (Maya et al., 2019) takes account of the external features represented in images and texts by combining Poisson process modeling and deep neural network. However, the method of Okawa et al. (2019) assumes that events occur independently of one another, and thus does not adequately describe the triggering phenomena in which there exists strong interaction between events. We focus on the triggering process, and aim at capturing history-dependent and self-exciting phenomena such as diseases, armed conflicts and earthquakes.

6 Experiments

We used real-world event datasets from different domains to evaluate the predictive performance of ConvHawkes.

6.1 Datasets

We used three real-world event datasets and five image datasets. All the datasets are publicly available. The statistics of these datasets are given in Table 1.

6.1.1 Event data

We conducted experiments on three event datasets from different domains.

- **Conflict** Conflict dataset, which is provided by ACLED project,¹ consists of roughly 17,000 armed conflicts in Africa dated from April 1, 2018 to March 31, 2020. Every event is recorded in the form of time and location (latitude and longitude coordinates).
- **Protest** Protest dataset, which was gathered by ACLED project,¹ contains over 34,000 demonstration events in Middle East over a four year period from April 1, 2018 to March 31, 2020. Each record contains time and location of the protest.
- **Disease** Disease dataset is a collection of reported incidents of animal disease outbreaks that occurred in Europe, provided by EMPRES-i² it contains 21,529 records, each of which shows time, latitude and longitude.

The procedure of data preprocessing is detailed in Appendix B.1.

6.1.2 Georeferenced image

We incorporated five image datasets as the external features: **nightlight**, **landcover**, **weather**, **population** and **road**. These georeferenced images were all sourced from open GIS databases.

- The source of **nightlight** image is the Night time Lights of the World data processed and distributed by the NGDC,³ we used the $16,801 \times 43,201$ tiles that cover the entire world.
- For **landcover** image, the data source is the world map image file, at scale of 1:10 m, provided within the Natural Earth⁴ package.
- The world map files for **weather** and **population** were taken from GeoNetwork website.⁵ with a spatial resolution of 5 arc minutes.
- For **road**, the shapefile of roads was downloaded from gROADS⁶ The shapefile was converted into a GeoTIFF file.

The input images were saved in GeoTIFF format. As preprocessing, we cropped GeoTIFF images for the three areas of interest (i.e., Africa, Middle East, Europe) and resized them to 120×114 pixels for Africa, 120×147 for Middle East, 120×127 for Europe. The examples of a **population** image is given in Figs. 5a and 6a, and **landcover** image in Fig. 7a. In the experiment, we only used static images which not contain time information. Thus, the number of observations L is fixed to 1. Details of the data collection procedure are shown in Appendix B.2.

¹ Armed Conflict Location and Event Dataset (ACLED). <https://www.acledata.com>. Accessed on December 10, 2021.

² EMPRES Global Animal Disease Information System (EMPRES-i). <http://empres-i.fao.org/eipws3g/>. Accessed on April 1, 2021.

³ Image and Data processing by the National Oceanic and Atmospheric Administration's (NOAA) National Geophysical Data Center (NGDC). <https://ngdc.noaa.gov/ngdc.html>. Accessed on April 1, 2021.

⁴ Natural Earth. <https://www.naturalearthdata.com>. Accessed on April 1, 2021.

⁵ Food and Agriculture Organization (FAO), GeoNetwork. <http://www.fao.org/geonetwork>. Accessed on April 1, 2021.

⁶ NASA Socioeconomic Data and Applications Center (SEDAC), Global Roads Open Access Data Set, Version 1 (gROADSv1). <http://dx.doi.org/10.7927/H4VD6WCT>. Accessed on April 15, 2021.

6.2 Comparison methods

We compared the proposed **ConvHawkes** against four widely used point process methods.

- **HPP** (Spatio-temporal homogeneous Poisson Process): The intensity is assumed to be constant over time and space: $\lambda(t, \mathbf{s}) = \lambda_0$. This optimization can be solved in closed form.
- **RMTTP** (Recurrent Marked Temporal Point Process) (Du et al., 2016): RMTTP uses RNN to describe the intensity of the marked temporal point process. RMTTP is primarily intended to model event timing and categorical event feature (marker). To allow comparison, we partitioned the area of interest using a pre-defined rectangular grid; and mapped latitude and longitude values of event data into particular grids (hereafter referred to as *regions*). Then the latitude and longitude coordinates were replaced by a region index. The region indices are regarded as marks.
- **Hawkes** (Spatio-temporal Hawkes Process) (Reinhart, 2018): Intensity is given by Eq. (3), which does not accept any additional features. We choose an exponential decay function, see Eq. (9), as the temporal decay function $h(\cdot)$, and Gaussian kernel shown as Eq. (10) for the spatial decay function $k(\cdot)$.
- **DMPP** (Deep Mixture Point Process) (Maya et al., 2019): This method incorporates the external features represented in images and texts by combining Poisson process modeling and deep neural networks. We used the same image datasets used in **ConvHawkes** as the external features for DMPP.

6.3 Experimental settings

For the experiments, we divided each dataset into training, validation and test sets in chronological order with the ratios of 80%, 10%, and 10%. The model parameters were trained using the ADAM optimizer (Kingma & Ba, 2014) with a learning rate of 0.002. We tuned all the models using early stopping based on the log-likelihood performance on the validation set with a maximum of 200 epochs and a patience of 10 epochs. Batch size was set to 256 for all methods. The hyperparameters of each model were optimized via grid search. For the neural networks-based models (i.e., RMTTP, DMPP and **ConvHawkes**), we chose the number of layers N_l from $\{1, 2, 3, 4, 5\}$, and the number of units per layer N_u from $\{1, 3, 5, 8\}$. For CNN-based methods (i.e., DMPP and **ConvHawkes**), we searched the filter size N_k in the CNN over $\{1, 3, 5\}$. The uniform kernel function was selected for the temporal and spatial convolution. The mathematical definitions are given in Appendix A.1. The chosen hyperparameters are presented in Sect. 6.5.1. The pixel intensities of color channels were normalized to $[0, 1]$, and then used as input of our model.

6.4 Evaluation metrics

Our experiments use the following two metrics in evaluating all models. For both metrics, lower values indicate better performance.

- **NLL** (Negative Log-Likelihood) is used to assess the likelihood of the occurrence of the events over the test period; it is calculated as

Table 2 Negative log-likelihood (NLL)

	Conflict	Protest	Disease
HPP	– 8.872	– 9.130	– 9.081
Hawkes	– 10.156	– 10.525	– 10.806
DMPP	– 9.531	– 9.465	– 9.902
Proposed	– 11.548	– 11.583	– 11.988

Lower is better. The best performance is shown in bold. Our proposal, ConvHawkes, outperforms four existing methods

$$\sum_{n=N}^{N+N_t} \left[-\log \lambda(t_n, \mathbf{s}_n) + \int_{t_{i-1}}^{t_n} \int_{\mathbb{S}} \lambda(t, \mathbf{s}) dt d\mathbf{s} \right], \quad (14)$$

where N_t is the number of events in the test period.

- **NMAE** (Normalized Mean Absolute Error) evaluates the discrepancies between the predicted number of events in small time intervals and pre-defined regions and the ground truth. We first split the test time period $[T, T + \Delta T]$ into S successive small time intervals. Also, we partitioned the area of interest \mathbb{S} into R uniform grid regions. For each time interval $(t_s, t_{s+1}]$ and each region $(\mathbf{s}_r, \mathbf{s}_{r+1})$, given the history of events up to t_s , we predicted the number of events in $(t_s, t_{s+1}]$ and $(\mathbf{s}_r, \mathbf{s}_{r+1})$, $\hat{N}((t_s, t_{s+1}], (\mathbf{s}_r, \mathbf{s}_{r+1}))$, described in Eq. (13). Then, we measured the average normalized difference between the predicted and observed number of events over all the time intervals and the pre-defined regions as follows:

$$\text{NMAE} = \frac{\sum_{r=1}^R \sum_{s=1}^S |\hat{N}((t_s, t_{s+1}], (\mathbf{s}_r, \mathbf{s}_{r+1})) - N((t_s, t_{s+1}], (\mathbf{s}_r, \mathbf{s}_{r+1}))|}{\sum_{r=1}^R \sum_{s=1}^S N((t_s, t_{s+1}], (\mathbf{s}_r, \mathbf{s}_{r+1}))}, \quad (15)$$

where $\hat{N}((t_{s+1}, t_s], (\mathbf{s}_r, \mathbf{s}_{r+1}))$ is the predicted number of events in the small time interval $(t_{s+1}, t_s]$ and the grid region $(\mathbf{s}_r, \mathbf{s}_{r+1})$ and $N(\cdot)$ is the ground truth at the s -th time interval and r th region. In our experiment, we partitioned the spatial area of interest using a 5×5 uniform grid, and divided the test period into 20 time intervals. Therefore $S = 20$ and $R = 25$.

6.5 Performance comparison

In this section, we compare ConvHawkes with existing point process methods for event prediction.

Table 2 shows the negative log-likelihood (NLL) of the test data for the three event datasets. Note that since the temporal point processes (i.e., RMTPP) cannot calculate spatial likelihood, the NLL results of these methods are not reported on this table. We trained the proposed method with each of the five image datasets (i.e., nightlight, landcover, weather, population, road) and reported the best performance among the different image datasets in Tables 2 and 3. The population image yields the best prediction performance for Conflict and Protest datasets; the landcover produces the best result for Disease dataset. We can see that the proposal, ConvHawkes, outperforms all existing methods examined across all the datasets. HPP delivers the worst prediction accuracy since it fails to account for temporal or spatial dependencies between events. DMPP performs worse than Hawkes on all the datasets. This is expected, because DMPP does not explicitly model the mutual excitation

Table 3 Normalized mean absolute error (NMAE) with standard deviation (in the bracket)

	Conflict	Protest	Disease
HPP	1.144 (0.055)	1.116 (0.096)	1.277 (0.230)
RMTTP	0.876 (0.094)	0.925 (0.159)	0.940 (0.300)
Hawkes	0.464 (0.023)	0.520 (0.064)	0.481 (0.087)
DMPP	0.685 (0.041)	0.867 (0.119)	0.865 (0.231)
Proposed	0.344 (0.015)	0.466 (0.043)	0.423 (0.073)

Lower is better. The best performance is shown in bold. Our proposal, ConvHawkes, outperforms four existing methods

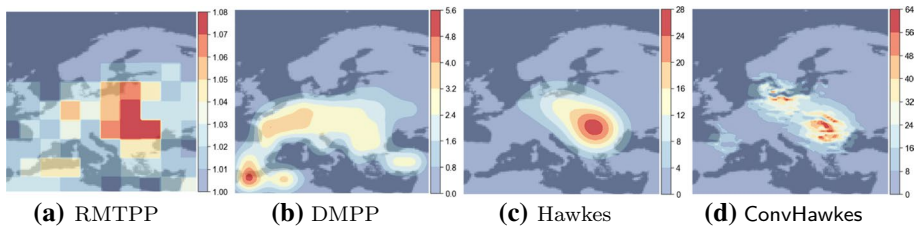


Fig. 3 **a** RMTTP, **b** DMPP, **c** Hawkes, **d** ConvHawkes. Conditional intensity of diseases in Europe estimated by each method at March 1st 2020. The x-axis and y-axis represent longitude and latitude respectively

between events and thus cannot capture triggering patterns. For all the datasets, **Hawkes** outperformed the other existing methods. This is possibly because **Hawkes** models the mutual excitation between events with decay over spatio-temporal distances, while **DMPP** does not explicitly consider the spatial dependencies between events. **ConvHawkes** produces even better performance than **Hawkes**. The results suggest that our method can extract the meaningful features from the images, and effectively learn their impact on the triggering processes.

Table 3 reports the Normalized Mean Absolute Error (NMAE) of five different methods on the three event datasets. The result again demonstrates the effectiveness of our approach. Compared to the strongest baseline, **ConvHawkes** offers a NMAE improvement of 34.9% for the Conflict data ($p < 0.001$; paired t-test), 11.6% NMAE improvement for the Protest data ($p < 0.1$), 13.7% NMAE improvement for the Disease data ($p < 0.001$). This supports the above conclusion.

Our **ConvHawkes** demonstrated improvements in all evaluation metrics used. This is probably because **ConvHawkes** can capture the spatial heterogeneity of the triggering process as well as the spatio-temporal decay effects. We can see this in Fig. 3, which depicts the conditional intensity of diseases learned by four different methods on March 1, 2020. In Fig. 3c, the spatial influences seem to be evenly distributed for **Hawkes**. **ConvHawkes** intensity (Fig. 3d) is more unevenly distributed along the densely populated urban areas.

6.5.1 Sensitivity analysis

In this section, we analyze the impact of hyperparameters and experimental settings. We report the prediction performance of **ConvHawkes** under different settings for the three event datasets.

Table 4 Performance comparison of the proposed method with different images on three event datasets

	Conflict	Protest	Disease
nightlight	– 11.207	– 11.379	– 11.272
landcover	– 11.021	– 10.814	– 11.115
weather	– 11.111	– <i>11.336</i>	– <i>11.149</i>
population	– 11.548	– 10.918	– 10.959
road	– 10.937	– 11.050	– 11.088

The number indicates NLL. Lower is better. The best performance is in boldface and second best is italicised

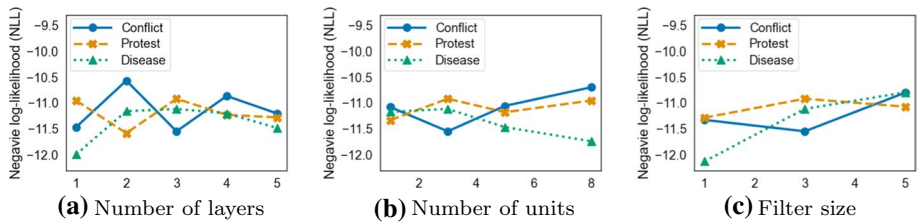


Fig. 4 **a** Number of layers, **b** number of units, **c** filter size impact of hyper-parameters on NLL performance

Impact of Different Images Table 4 examines the importance of different images for event prediction by individually incorporating each of the image datasets into the proposed model. For Conflict data, NLL is improved when adding population image. This is consistent with the prior observation: unrest spreads among densely populated areas. We can see that incorporating nightlight images improves the prediction performance for Protest and Disease datasets. This is probably because nightlight is correlated to population density. We can observe that the weather image is important for Disease data. This finding matches the previous study: weather change affects on disease transmission (Parham & Michael, 2010). In general, ConvHawkes can achieve stable performance across different image datasets. ConvHawkes with different image datasets is consistently better than all the comparison methods (Table 2), which ensures all the image datasets used in this paper are important for event prediction, and that ConvHawkes can effectively utilize these images.

Network Structure We show the impact of network structures in Fig. 4a–c. Except for the parameters being tested, all other parameters were held to default values. The NLL performance tends to be stable for all datasets. The prediction performance slightly improves when layer size N_l is 3 for Conflict data, 2 for Protest data, and 1 for Disease data. As shown in Fig. 4b, ConvHawkes performs robustly for different number of units, N_u , across all data sets. The prediction performance saturates as filter size N_k in the CNN increases. The proposed method yields similar results for the other metrics (i.e., NMAE). Throughout the experiment, we set $N_l = 3, N_u = 3, N_k = 3$ for Conflict dataset; $N_l = 2, N_u = 3, N_k = 3$ for Protest dataset; and $N_l = 1, N_u = 3, N_k = 3$ for Disease dataset.

6.6 Analysis of feature learning

To further verify the above conclusion, we qualitatively explore the estimated intensity and the latent feature maps learned from the input image by our method.

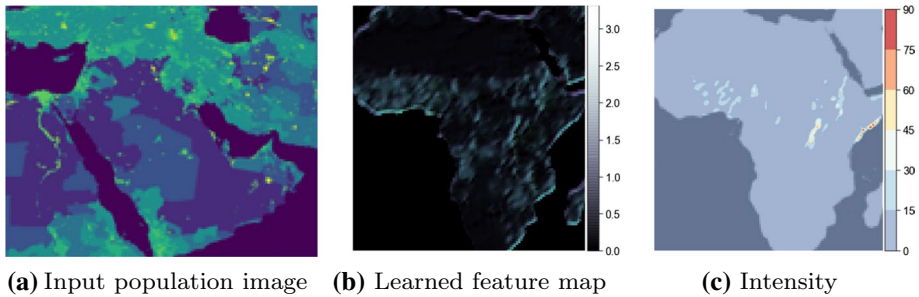


Fig. 5 **a** Input population image, **b** learned feature map, **c** intensity learned feature map and intensity for Conflict dataset

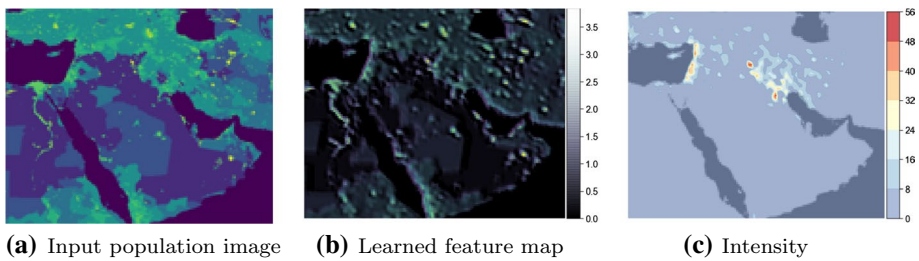


Fig. 6 **a** Input population image, **b** learned feature map, **c** intensity learned feature map and intensity for Protest dataset

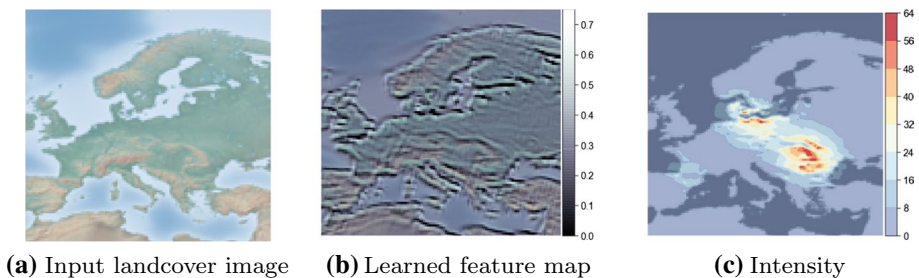


Fig. 7 **a** Input landcover image, **b** learned feature map, **c** intensity Learned feature map and intensity for Disease dataset

Figures 5–7 show the input image, the learned latent feature map and intensity for Conflict, Protest, Disease datasets. The x-axis and y-axis represent longitude and latitude respectively. Figures 5a and 6a show the input population image for Africa and Middle East, respectively. Figure 5a is the input landcover image for Europe. In the learned feature maps (Figs. 5b, 6b, 7b), the lighter shades are higher feature values and the darker shades indicate lower feature values. In Figs. 5b and 6b, we can observe that ConvHawkes highlights coastal areas for Conflict and Protest datasets. This is expected, since the unrest events are strengthened in densely populated coastal areas. ConvHawkes (Figs. 5c and 6c) exhibits heterogeneous intensity, in which the spatial influence is spread along the coastal areas. As shown in Fig. 7, the landcover

image serves as an important feature for Disease dataset. This may be because landcover is associated with other characteristics including weather and population. The proposed method can automatically discover discriminative features from the images, providing insights about the effects the underlying external factors have on the triggering process.

7 Conclusion

In this paper, we tackled the problem of spatio-temporal event prediction. Our proposal, **ConvHawkes** (Convolutional Hawkes Process), is a novel Hawkes process model based on a deep learning approach. Specifically, we combine CNN with continuous kernel convolution and model the Hawkes process intensity parameter by using an extended neural network model. The key advantage of **ConvHawkes** over existing methods is that it can utilize the rich contexts present in image data, including satellite images, map images and weather maps, and automatically discover their complex effects on the event triggering processes. At the same time, this formulation makes analytical integration over the intensity, which is required for Hawkes process estimation, tractable. Using three real-world datasets from different domains (i.e., armed conflicts, protests, diseases), we demonstrated that the proposed method is able to provide higher event prediction accuracy than existing methods.

To the best of our knowledge, this work is the first attempt towards incorporating image data into self-exciting spatio-temporal point process models. For future work, we plan to extend the proposed approach by combining it with recent point process models.

A. Appendix

A.1 Convolutional kernel

We factorize the convolutional kernel function $f(\cdot)$ into temporal and spatial components, and model each component by the uniform kernel:

$$f(t - \tau, \mathbf{s} - \mathbf{x}) = \mathbb{1}[|t - \tau| < \Delta] \mathbb{1}[\|\mathbf{s} - \mathbf{x}\| < w], \quad (16)$$

where $\mathbb{1}[\cdot]$ is an indicator function, and Δ and w are positive parameters that threshold the kernels to zeros. In our experiment, we fix Δ as the time interval between the observations; w is the pixel size of the image. This is equivalent to a piece-wise approximation. Here we consider the simplest case for the implementation simplicity; but note that our method can be easily generalized to other forms.

A.2 Likelihood computation

With the uniform kernel function (Eq. 7), the integral over time can be performed analytically as follows:

$$\int_{t_n}^T \mathbb{1}[\tau - \Delta\tau < t < \tau] g(t - t_n) dt = \left[G(t - t_n) \right]_{\max(t_n, \tau - \Delta\tau)}^{\max(\tau, \max(t_n, \tau - \Delta\tau))}, \quad (17)$$

where $G(\cdot)$ is the derivative of the temporal decay kernel $g(\cdot)$. For the exponential decay defined by Eq. (9), $G(t - t_n) = -\exp(-\beta(t - t_n))$.

For the pair of the Gaussian convolutional kernel (Eq. 8) and Gaussian decay function (Eq. 10), the integral over space \mathbb{S} is described as the sum of error functions:

$$\begin{aligned} & \int_{\mathbb{S}} \zeta(\mathbf{s} - \mathbf{s}_j)k(\mathbf{s} - \mathbf{x}_{h,w})d\mathbf{s} \\ &= \int_{\mathbb{S}} \exp\left(-(\mathbf{s} - \mathbf{x})^T \Sigma_k^{-1}(\mathbf{s} - \mathbf{x})\right) \exp\left(-\frac{1}{2}(\mathbf{s} - \mathbf{s}_j)^T \Sigma_{\zeta}^{-1}(\mathbf{s} - \mathbf{s}_j)\right) d\mathbf{s} \\ &= \frac{1}{\sqrt{\det(2\pi(\Sigma_k + \Sigma_{\zeta}))}} \exp\left[-\frac{1}{2}(\mathbf{s}_j - \mathbf{x})^T (\Sigma_k + \Sigma_{\zeta})^{-1}(\mathbf{s}_j - \mathbf{x})\right] \\ &\quad \times \int_{\mathbb{S}} \exp\left(-\frac{1}{2}(\mathbf{s} - \mathbf{x}_c)^T \Sigma_c^{-1}(\mathbf{s} - \mathbf{x}_c)\right) d\mathbf{s}, \end{aligned} \tag{18}$$

where

$$\mathbf{x}_c = (\Sigma_k^{-1} + \Sigma_{\zeta}^{-1})^{-1} (\Sigma_k^{-1} \mathbf{s}_j + \Sigma_{\zeta}^{-1} \mathbf{x}) \tag{19}$$

$$\Sigma_c = (\Sigma_k^{-1} + \Sigma_{\zeta}^{-1})^{-1} \tag{20}$$

The Gaussian integral in the above equation can be expressed in terms of the error function for the diagonal covariance matrices Σ_k and Σ_{ζ} . The integral in the likelihood (Eq. 11) has analytic form for many other kernels including Rayleigh and power-law.

B. Experiment

B.1 Event dataset

Conflict and Protest datasets were acquired from ACLED¹ website at <https://acleddata.com/data-export-tool/> ACLED is an event-level data on political violence and protest that includes date, location and type of event. For Conflict dataset, we extracted violent conflict events (i.e., battles, remote violence and bombings, and violence against civilians) occurred in Africa from the ACLED database. For Protest dataset, we collected civil demonstrations (i.e., protests and riots) in Middle East.

Disease data were obtained through the online EMPRES-i² system (<https://empres-i.review.fao.org/#/>.) The disease outbreak data for domestic poultry from 1 January 2004 to 31 December 2009 were obtained from the EMPRES-i database. Each confirmed outbreak contains outbreak location (latitude and longitude) and date of observation.

B.2 Georeferenced image

- The source of **nightlight** image is the Night time Lights of the World data processed and distributed by NGDC³: [F182013.v4c_web.stable_lights.avg_vis.tif], from <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>.
- For **landcover** image, the data source is the world map image file found in the Natural Earth⁴ package: [HYP_HR_SR_W_DR.tif] at https://github.com/nvkelso/natural-earth-raster/tree/master/10m_rasters.

Table 5 Execution times of different methods in seconds

	Conflict	Protest	Disease
HPP	3.585	3.189	4.370
RMTTP	136.473	34.711	61.410
Hawkes	29.387	30.414	33.943
DMPP	20.054	151.486	27.120
Proposed	715.874	992.047	885.619

- The world map files for **weather**, **population**, **livestock** and **terrain** were taken from GeoNetwork website⁵ at <http://www.fao.org/geonetwork/srv/en/main.home?uuiid=fc32c5de-440c-46aa-9cad-81f4c8b84c6a>, namely,
 - weather: [clim.tif]
 - population: [popd.tif]
 - livestock: [lvstd.tif]
 - terrain: [slp.tif]
- For **road**, the shapefile of the roads is downloaded from gROADS⁶ at [https://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1/data-download:\[gROADS-v1-africa.shp\]](https://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1/data-download:[gROADS-v1-africa.shp]). We use an open source GIS platform, Quantum GIS (QGIS), to convert the shapefile to a GeoTIFF file.

B.3 Implementation details

All code was implemented using Python 3.9 and Keras Chollet (2015) with a TensorFlow backend Abadi et al. (2016). We conducted all experiments on a machine with four 2.8GHz Intel Cores and 16GB memory. We compare the training times of the different methods on three datasets in Table 5.

Author contributions All authors revised the manuscript and approved the final manuscript, and agree to be accountable for all aspects of the work related to its accuracy or integrity.

Funding Not applicable.

Code availability Source code is not available.

Material availability All the datasets are publicly available.

Declarations

Conflict of interest The author's declare that they have no conflict of interest.

Ethics approval Not applicable.

Consent to participate Yes.

Consent for publication Yes.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., & Devin, M, et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint [arXiv:160304467](https://arxiv.org/abs/160304467)
- Aldrin, M., Huseby, R., & Jansen, P. (2015). Space-time modelling of the spread of pancreas disease (pd) within and between Norwegian marine salmonid farms. *Preventive Veterinary Medicine*, *121*(1–2), 132–141.
- Attia, M., Hossny, M., Nahavandi, S., & Asadi, H. (2017). Surgical tool segmentation using a hybrid deep cnn-rnn auto encoder-decoder. In: *International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, pp 3373–3378
- Bacry, E., Mastromatteo, I., & Muzy, J. F. (2015). Hawkes processes in finance. *Market Microstructure and Liquidity*, *1*(01), 1550005.
- Blundell, C., Beck, J., & Heller, K.A. (2012). Modelling reciprocating relationships with hawkes processes. In: *Proceedings of NeurIPS*, pp 2600–2608
- Chandra, S. R., & Al-Deek, H. (2009). Predictions of freeway traffic speeds and volumes using vector autoregressive models. *Journal of Intelligent Transportation Systems*, *13*(2), 53–72.
- Chen, R.T., Amos, B., & Nickel, M. (2020). Neural spatio-temporal point processes. arXiv preprint [arXiv:201104583](https://arxiv.org/abs/201104583)
- Choi, E., Du, N., Chen, R., Song, L., & Sun, J. (2015). Constructing disease network and temporal progression model via context-sensitive Hawkes process. In: *Proceedings of ICDM*, IEEE, pp 721–726
- Chollet, F, et al. (2015). keras. <https://github.com/fchollet/keras>
- De La Rocque, S., Balenghien, T., Halos, L., Dietze, K., Claes, F., Ferrari, G., Guberti, V., & Slingenbergh, J. (2011). A review of trends in the distribution of vector-borne diseases: is international trade contributing to their spread? *Rev Sci Tech*
- Du, N., Dai, H., Trivedi, R., Upadhyay, U., Gomez-Rodriguez, M., & Song, L. (2016). Recurrent marked temporal point processes: Embedding event history to vector. In: *Proceedings of the 22nd KDD*, ACM, pp 1555–1564
- Farajtabar, M., Wang, Y., Rodriguez, M.G., Li, S., Zha, H., & Song, L. (2015). Coevolve: A joint point process model for information diffusion and network co-evolution. In: *Proceedings of NeurIPS*, pp 1954–1962
- Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 922–929.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90.
- Kang, H. W., & Kang, H. B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *PLoS ONE**12*(4), e0176244.
- Kapoor, A., Ben, X., Liu, L., Perozzi, B., Barnes, M., Blais, M., & O'Banion, S. (2020). Examining covid-19 forecasting using spatio-temporal graph neural networks. arXiv preprint [arXiv:200703113](https://arxiv.org/abs/200703113)
- Kim, M., Jurdak, R., & Paini, D. (2017). Modeling reflexivity of social systems in disease spread. arXiv preprint [arXiv:171106359](https://arxiv.org/abs/171106359)
- Kingma, D.P., Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint [arXiv:14126980](https://arxiv.org/abs/1412.6980)
- Lee, E.D., Daniels, B.C., Myers, C.R., Krakauer, D.C., & Flack, J.C. (2019). Emergent regularities and scaling in armed conflict data. arXiv preprint [arXiv:190307762](https://arxiv.org/abs/190307762)
- Louie, K., Masaki, M., & Allenby, M. (2010). A point process model for simulating gang-on-gang violence. Project Report
- Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, *54*, 187–197.
- Maya, O., Iwata, T., Kurashima, T., Tanaka, Y., Toda, H., & Ueda, N. (2019). Deep mixture point processes: Spatio-temporal event prediction with rich contextual information. In: *Proceedings of KDD*, ACM
- Mei, H., & Eisner, J.M. (2017). The neural hawkes process: A neurally self-modulating multivariate point process. In: *Proceedings of NeurIPS*, pp 6754–6764
- Meyer, S., et al. (2018). Self-exciting point processes: Infections and implementations. *Statistical Science*, *33*(3), 327–329.
- Mishra, S., Rizoiu, M.A., & Xie, L. (2016). Feature driven and point process approaches for popularity prediction. In: *Proceedings of CIKM*, ACM, pp 1069–1078
- Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100–108.

- Morse, S.S. (2001). Factors in the emergence of infectious diseases. *Plagues and politics*, pp 8–26
- Musmeci, F., & Vere-Jones, D. (1992). A space-time clustering model for historical earthquakes. *Annals of the Institute of Statistical Mathematics*, 44(1), 1–11.
- Nicolas, G., Durand, B., Duboz, R., Rakotondravao, R., & Chevalier, V. (2013). Description and analysis of the cattle trade network in the Madagascar highlands: potential role in the diffusion of rift valley fever virus. *Acta Tropica*, 126(1), 19–27.
- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2), 379–402.
- Ogata, Y., Katsura, K., & Tanemura, M. (2003). Modelling heterogeneous space-time occurrences of earthquakes and its residual analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4), 499–509.
- Omi, T., Ueda, N., & Aihara, K. (2019). Fully neural network based model for general temporal point processes. arXiv preprint [arXiv:190509690](https://arxiv.org/abs/190509690)
- Parham, P. E., & Michael, E. (2010). Modeling the effects of weather and climate change on malaria transmission. *Environmental Health Perspectives*, 118(5), 620–626.
- Patz, J. A., Daszak, P., Tabor, G. M., Aguirre, A. A., Pearl, M., Epstein, J., et al. (2004). Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence. *Environmental Health Perspectives*, 112(10), 1092–1098.
- Porter, M. D., White, G., et al. (2012). Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics*, 6(1), 106–124.
- Pratiwi, H., Slamet, I., Saputro, D., et al. (2017). Self-exciting point process in modeling earthquake occurrences. *Journal of Physics: Conference Series*, 855(1), 012033.
- Reinhart, A., et al. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, 33(3), 299–318.
- Schütt, K., Kindermans, P.J., Felix, H.E.S., Chmiela, S., Tkatchenko, A., & Müller, K.R. (2017). Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In: *Proceedings of NeurIPS*, pp 991–1001.
- Servadio, J. L., Rosenthal, S. R., Carlson, L., & Bauer, C. (2018). Climate patterns and mosquito-borne disease outbreaks in South and Southeast Asia. *Journal of Infection and Public Health*, 11(4), 566–571.
- Van Der Voort, M., Dougherty, M., & Watson, S. (1996). Combining Kohonen maps with Arima time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, 4(5), 307–318.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In: *Proceedings of NeurIPS*, pp 5998–6008.
- Wagner, M., Tsui, F., Cooper, G., Espino, J.U., Harkema, H., Levander, J., Villamarin, R., Voorhees, R., Millett, N., & Keane, C., et al. (2011). Probabilistic, decision-theoretic disease surveillance and control. *Online Journal of Public Health Informatics*, 3(3).
- Wang, S., Suo, S., Ma, W.C., Pokrovsky, A., & Urtasun, R. (2018). Deep parametric continuous convolutional neural networks. In: *Proceedings of CVPR*, pp 2589–2597.
- Wilder-Smith, A., & Gubler, D. J. (2008). Geographic expansion of dengue: The impact of international travel. *Medical Clinics of North America*, 92(6), 1377–1390.
- Xiao, S., Farajtabar, M., Ye, X., Yan, J., Song, L., & Zha, H. (2017). Wasserstein learning of deep generative point process models. In: *Proceedings of NeurIPS*, pp 3247–3257.
- Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., & Li, Z. (2018). Deep multi-view spatial-temporal network for taxi demand prediction. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol 32
- Yasrab, R. (2017). Dcseg: Decoupled cnn for classification and semantic segmentation. In: *Proceedings of the International Conference on Knowledge and Smart Technologies*.
- Yasrab, R., Gu, N., & Zhang, X. (2017). An encoder–decoder based convolution neural network for future advanced driver assistance system. *Applied Sciences*, 7(4), 312.
- Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint [arXiv:170904875](https://arxiv.org/abs/170904875).
- Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2018). Self-attention generative adversarial networks. arXiv preprint [arXiv:180508318](https://arxiv.org/abs/180508318).
- Zhang, Q., Lipani, A., Kirnap, O., & Yilmaz, E. (2020). Self-attentive Hawkes process. In: *International Conference on Machine Learning, PMLR*, pp 11183–11193.
- Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. In: *Thirty-first AAAI conference on artificial intelligence*.

- Zhang, J., Zheng, Y., Qi, D., Li, R., & Yi, X. (2016). Dnn-based prediction model for spatio-temporal data. In: *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp 1–4
- Zhang, Y., & Liu, Y. (2009). Traffic forecasting using least squares support vector machines. *Transportmetrica*, 5(3), 193–213.
- Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). Lstm network: A deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2), 68–75.
- Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., et al. (2019). T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9), 3848–3858.
- Zhu, S., & Xie, Y. (2019). Crime linkage detection by spatial-temporal-textual point processes. arXiv preprint [arXiv:190200440](https://arxiv.org/abs/190200440).
- Zhu, S., Li, S., & Xie, Y. (2019). Reinforcement learning of spatio-temporal point processes. arXiv preprint [arXiv:190605467](https://arxiv.org/abs/190605467).
- Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458), 369–380.
- Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. *Modeling Financial Time Series with S-Plus®* pp 385–429.
- Zuo, S., Jiang, H., Li, Z., Zhao, T., & Zha, H. (2020). Transformer Hawkes process. In: *International Conference on Machine Learning*, PMLR, pp 11692–11702.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.