Journal of the Royal Statistical Society Series C: Applied Statistics, 2023, **72**, 368–386
https://doi.org/10.1093/jrsssc/qlad013
Advance access publication 28 March 2023
Original Article

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series C
Applied Statistics

Non-stationary spatio-temporal point process modeling for high-resolution COVID-19 data

Zheng Dong[1], Shixiang Zhu[2], Yao Xie[1], Jorge Mateu[3] and Francisco J. Rodríguez-Cortés[4]

[1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2]Heinz College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[3]Department of Mathematics, Universitat Jaume I, Castelló de la Plana, Valencia 12071, Spain
[4]Escuela de Estadística, Universidad Nacional de Colombia, Medellín, Colombia

Address for correspondence: Yao Xie, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. Email: yao.xie@isye.gatech.edu

Abstract

Most COVID-19 studies commonly report figures of the overall infection at a state- or county-level. This aggregation tends to miss out on fine details of virus propagation. In this paper, we analyze a high-resolution COVID-19 dataset in Cali, Colombia, that records the precise time and location of every confirmed case. We develop a non-stationary spatio-temporal point process equipped with a neural network-based kernel to capture the heterogeneous correlations among COVID-19 cases. The kernel is carefully crafted to enhance expressiveness while maintaining model interpretability. We also incorporate some exogenous influences imposed by city landmarks. Our approach outperforms the state-of-the-art in forecasting new COVID-19 cases with the capability to offer vital insights into the spatio-temporal interaction between individuals concerning the disease spread in a metropolis.

Keywords: spatio-temporal point processes, self-exciting models, nonstationary kernel, neural networks, COVID-19

1 Introduction

The outbreak of coronavirus disease 2019 (COVID-19) since 2020 has swept the world and is still developing. It causes a dramatic loss of human lives (Chriscaden, 2020) and presents an unprecedented challenge to public health, food systems, and the world. Tracking the dynamics of COVID-19 enables the human being to take target protecting measures to curb the pandemic's spread and design health surveillance systems. However, limited and biased information about local COVID-19 cases makes it extremely difficult to effectively control strategies against the pandemic.

There is a large amount of *aggregated* data consistently collected and publicly available, which contains rich information about COVID-19 cases. For instance, Johns Hopkins Center for Systems Science and Engineering (JHU CSSE) establishes an interactive COVID-19 dashboard to track the global coronavirus development (John Hopkins University, 2020) which reports the daily confirmed cases and deaths worldwide up to the state level. New York Times (2020) also tracks daily the county-level counts of confirmed cases and deaths in the USA. Such data help the scientific researcher model the disease transmission on an aggregated level and play a pivotal role in tracking the propagation patterns of the virus and helping policymakers act effectively to revitalize economic and social development.

Received: September 18, 2021. **Accepted:** December 22, 2022

However, such aggregated data lack precise information about individual cases and present a significant challenge in modeling the spatio-temporal dynamics of human-to-human disease transmission when capturing the fine spatial heterogeneity of case distribution in a small region. Aggregated data may lose fine-grained spatio-temporal information, which will lead the administrative officials to make biased decisions. For example, it is reported in Bizzarri et al. (2020) that the unreliable preliminary data, as well as inaccurate models, significantly affected the political decisions of the Italian administration. Another example in Guenther et al. (2020) documents a superspreader event in Germany in a meat processing plant; modeling such an event requires accounting for the precise plant location information, and aggregated data may miss such crucial local information.

In this paper, we consider a unique high-resolution dataset for individual cases of COVID-19 in Cali, Colombia, the second-largest city in the country. This data records individual confirmed cases for six months, from 15 March to 30 September of 2020, with each case's time and location information. To take full advantage of the fine-grained dataset, we develop a nonstationary spatio-temporal point process model, assuming that previously infected events trigger the newly confirmed cases. We assume the triggering effect is nonstationary (Hendry & Pretis, 2016) since the virus is likely to spread more slowly in sparsely populated rural than densely populated areas. This fact entails stationary point processes nonapplicable: the stationary kernel is 'shift-invariant' and only depends on the temporal and location differences between events. Moreover, we consider the exogenous promotion of densely populated city landmarks in the model since the COVID-19 virus proves to spread quickly through respiratory droplets (Jasper et al., 2020), and aerosol transmission in crowded and inadequately ventilated spaces (Leclerc et al., 2020). We represent parameters of the nonstationary kernel by neural networks to enhance model flexibility while maintaining interpretability. The model is estimated by solving a maximum-likelihood problem via a computationally efficient strategy to tackle the intractable numerical integration in the log-likelihood function. We conduct an extensive real-data study, which reveals the unique transmission dynamics of COVID-19 and confirms that a few landmarks in the city play an essential role in spreading the virus. The model and results will help policymakers monitor coronavirus dynamics and provide a template for tracking real-time data for future epidemics and implementing health surveillance systems. Since similar high-resolution datasets will not be so rare in the future, the need for such an approach is not limited to the situation of Cali.

The paper is organized as follows. The rest of this section discusses some relevant literature on COVID-19 modeling and spatio-temporal point processes (STPPs). We then introduce our motivating (and unique) dataset in Section 2. In Section 3, we review basic knowledge about point processes, propose our framework with a nonstationary spatio-temporal kernel, and illustrate our fine-crafted parameterization scheme with a simple neural network. Section 4 presents the computational strategies for model estimation with an approximation to the likelihood. Section 5 interprets the results from real data and compares them with several benchmark models. Lastly, Section 6 concludes the paper.

Related work

Compartmental models are widely developed to describe the overall COVID-19 infection in a region. The simplest SIR compartmental model (Harko et al., 2014) assigns the population into three compartments with labels $\boldsymbol{S}$ (susceptible), $\boldsymbol{I}$ (infectious), and $\boldsymbol{R}$ (recovered), respectively. Deterministic differential equations fit the transition rates between each kind of compartment. Advanced compartmental models are further designed by reframing the basic one with different compartments (Lin et al., 2020; Nande et al., 2020). SEIRD in Korolev (2021) and forced SEIRD models in Loli Piccolomini and Zama (2020) are adopted in various epidemic scenarios by introducing compartments of exposed and deceased populations into the system. Other extensions, such as splitting the infected population according to infection severity (Nande et al., 2020) and introducing unreported infected population (Lin et al., 2020) are also considered. Compartmental models assume a stable population of the inspected region, thus perform well when applied to large regions such as a country or state. However, they usually do not consider detailed spatial information such as population migration across regions.

Much work has been done on predicting the number of COVID-19 cases and deaths. Kraemer (2020) and Woody et al. (2020) adopted Generalized linear models to predict the number of daily cases and deaths during the first-wave COVID-19 in China and the USA, respectively. Autoregressive models are also widely used to forecast confirmed cases at a state-level (Agosto & Giudici, 2020; Mamode Khan et al., 2020; Triaccaa & Triacca, 2021). There are also several studies (Institute for Health Metrics and Evaluation, 2020; Northeastern University, Laboratory for the Modeling of Biological and Socio-technical Systems, 2021) adopted by the Centers for Disease Control and Prevention (CDC) for COVID-19 case forecast in the USA. Our approach differs from these methods in two ways: (a) Our model provides finer-grained predictions based on the unique data, and (b) we focus on capturing the spatio-temporal correlation between confirmed cases and emphasize the interpretability of the proposed model.

Spatio-temporal analysis of the COVID-19 plays a pivotal role in understanding the dynamics of the spread of COVID-19. Angulo et al. (2013) introduces a spatio-temporal BME-SIR model integrating the disease representation at different locations to generate disease predictions. Bai et al. (2020) divides the regional-level COVID-19 time-series data in the USA into several periods and develops a piecewise stationary SIR model coupled with spatio-temporal dependence. In addition, a vector autoregressive model developed by Zhu, Bukharin, et al. (2021) considers local spatio-temporal correlations, mobility, and demographic factors, aiming to estimate COVID-19 cases and deaths at a county level in the USA. In Chiang et al. (2020), a multivariate Hawkes process is adopted to model the occurrence of confirmed cases across the US counties by incorporating social and health covariates. However, most of these methods use spatially or temporally aggregated data, which hinders us from understanding the spread of COVID-19 at an individual level.

A few studies attempt to model the dynamics of COVID-19 using point processes. Gajardo and Müller (2021) proposes a point process regression framework of COVID-19 cases and deaths conditioned on mobility and economic covariates. Giudici et al. (2023) focuses on country-level case prediction in 27 European countries by augmenting spatio-temporal point process model with mobility network covariates. Li et al. (2021) introduces a generative and intensity-free point process model based on an imitation learning framework to track the spread of COVID-19 and forecast county-level cases in the USA. Compared with the previous methods, our approach is more flexible by considering nonstationarity in the spatial correlation, which is highly interpretable and expressive in representing the spread of the disease.

Recent works Berry et al. (2020), Brandén et al. (2020), López-Feldman et al. (2021) investigate the relationship between various factors and COVID-19 by collecting confirmed COVID-19 cases and mortality data at the individual level. However, these data are aggregated and reported at the county or state level. Other works Fu et al. (2020), Guo et al. (2020) use individual patient data, including peripheral blood samples, observed monocytes, and T-cell data from patients with severe COVID-19 symptoms to demonstrate the effectiveness of COVID-19 treatments, which is different from our objective in this paper.

Last but not least, some related studies use similar techniques developed in this paper. Du et al. (2016), Mei and Eisner (2016), Zhang et al. (2020) model discrete events using neural-network-based point process models. However, most of these works aim to enhance the representative power by taking advantage of the recurrent neural structure (Hochreiter & Schmidhuber, 1997) or the attention mechanism (Vaswani et al., 2017) to represent the historical information, which lacks interpretability and is unable to capture long-term effects. A wide array of research focuses on characterizing the triggering effects between events using a fine-crafted kernel function. Original works Ogata (1988, 1998) introduce a parametric kernel in Epidemic Type Aftershock Sequence (ETAS) to capture the triggering effects between earthquakes. There are a few prior works in studying nonstationary kernels in Gaussian processes. Some studies (Higdon et al., 1998; Lang et al., 2007; Vasudevan et al., 2011) aim to model spatial correlation based on a squared exponential kernel whose covariance structure is location-dependent. Another paper Remes et al. (2017) develops a nonstationary spectral kernel based on a generalized Fourier transform, whose spectral density is represented as a Gaussian mixture. Another line of research focuses on developing nonstationary kernels in point processes (Zhu, Bukharin, et al., 2022; Zhu, Li, et al., 2021; Zhu, Wang, et al., 2022; Zhu, Zhang, et al., 2021). These studies capture intricate spatio-temporal dependence between discrete events by introducing neural networks to kernel designs. Our work is a significant extension of the previous approaches, which greatly enhances the expressiveness of the nonstationary kernel by considering both inter- and intra-influences between spatial kernel feature functions while still being highly interpretable and computationally efficient.

A few works incorporate the exogenous effects into point process models by adding terms in the conditional intensity function (Farajtabar et al., 2017; Rizoiu et al., 2017; Zhu, Ding, et al., 2021).

2 Data description and preliminary analysis

The COVID-19 dataset provided by the Municipal Public Health Secretary of Cali[1] documents the individual-level confirmed COVID-19 cases collected from Cali, one of the major cities in Colombia, the capital of the Valle del Cauca department and the most populated city in southwest Colombia, with 2,227,642 residents according to the 2018 census. As shown in Figure 1a, more than half of the population concentrates in neighborhoods of low socio-economic strata located mainly in the east, northeast, and west. Almost a 10th of the population under the line of poverty agglomerates in the city's eastern neighbourhoods. The population with higher socio-economic strata distributes in the other city areas, concentrating the wealthiest population in the city's south. The city spans 560.3 km^2 (216.3 square miles) with 120.9 km^2 (46.7 square miles) of the urban area, making it the second-largest and the third most populated city in the country. As the only major Colombian city with access to the Pacific coast, Cali is the leading industrial and economic center in the country's south, with one of Colombia's fastest-growing economies. Cali's international airport is located in the northeast part of the city, and it is Colombia's third-largest airport in terms of passengers (Wikipedia, 2021).

The dataset records 38,611 cases from 15 March to 30 September 2020, including 28 weeks. Specifically, a COVID-19 case was recorded once confirmed, with the diagnosed date of the patient and the geographical location (measured in longitude and latitude) of their residence. The testing procedures were carried out across the entire urban area, with similar testing rates in each comuna. Unlike other commonly seen COVID-19 datasets that only report the aggregated number of cases or deaths at a state or county level, this dataset records each confirmed case's exact location and time. In practice, we observe periodic weekly oscillations in daily reported cases and deaths, which may be caused by testing bias (higher testing rates on certain days of the week). To reduce such bias, we aggregate the number of cases and deaths of each county by weeks. Figure 2 presents the spatial distribution of confirmed cases at four particular weeks in Cali. We note that the first confirmed case of COVID-19 in Colombia appeared on 6 March 2020. On 12 March, the country soon declared a state of emergency. On 15 March, Cali reported the first positive person. Then the authorities announced the mandatory isolation of the entire city for just eight days (Presidency of the Republic of Colombia, 2020). The first case reported in the city based on people who went to health services occurred in high socio-economic strata. However, the disease quickly spread and concentrated in the most vulnerable areas with low socio-economic strata. After early efforts of the government to contain the pandemic, inevitably, the virus spread throughout the city, affecting a large part of the population. The above public health decisions are known not significantly to affect the dynamics of the virus spreading. Thus, we do not consider the impact of these decisions in our model for simplicity.

Besides COVID-19 events, we also collect the location of three kinds of landmarks in Cali, including churches, schools, and town halls, from the Administrative Department of Municipal Planning.[2] These locations play an important role in understanding the wide and rapid spreading of the virus. According to James et al. (2021), there is a high COVID-19 positive rate among attendees to events at places such as churches. As a clear note in this line, among 92 attendees at a rural Arkansas church during 6–11 March, 35 (38%) developed laboratory-confirmed COVID-19, and three persons died (James et al., 2021). The landmark dataset has three town halls, 49 small and large churches, and 77 schools. Figure 1b shows the exact locations of these collected landmarks.

Our preliminary study suggests that the confirmed cases are unevenly distributed across the city and correlated in time and space. In Figure 3, we show the spatial distribution of all the confirmed cases. As we can see, most of the reported cases concentrate in the city's centre, particularly in Comuna 11. More cases are reported in the eastern Cali than in the western Cali, which presents a heterogeneous spatial profile of the COVID-19 cases in Cali. The first three panels in Figure 4

---

[1] https://www.cali.gov.co/salud/

[2] https://www.cali.gov.co/planeacion/

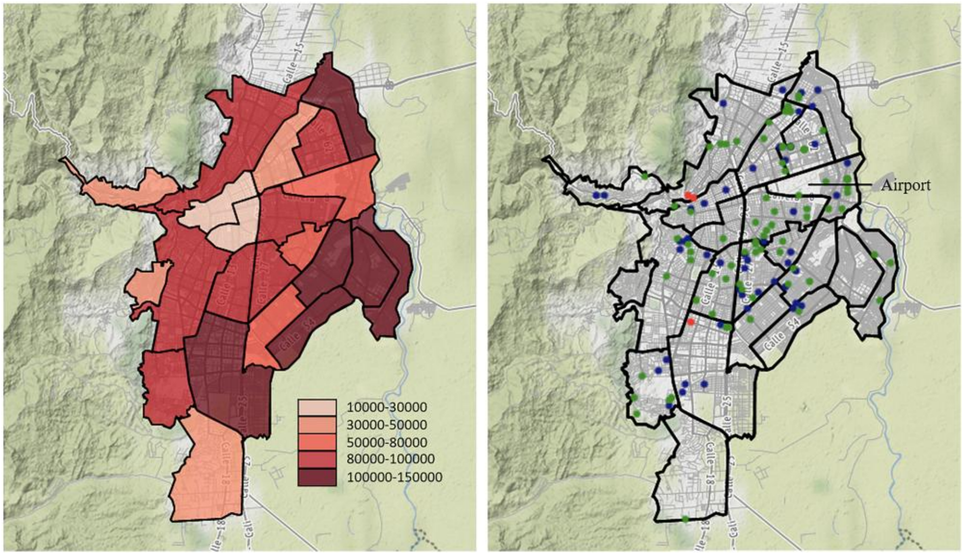

Figure 1. *Left panel*: Population distribution in Cali. Each polygon bounded by bold lines represents a comuna (a municipality-level subdivision in Cali); there are 22 comunas in the city of Cali. *Right panel*: Landmarks in Cali. Each dot represents the landmark's location, and its colour indicates the type of the landmark, where the red dot is a town hall, the blue dot is a church, and the green dot is a school.

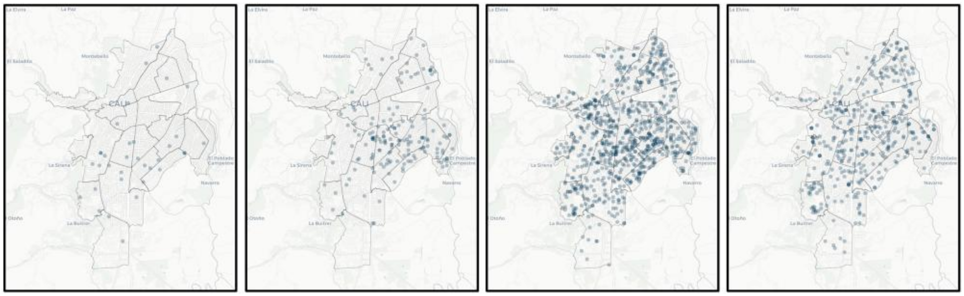

Figure 2. Snapshots of confirmed COVID-19 cases at four particular weeks in 2020: 29 March, 17 May, 12 July, and 23 August. Each dot represents the location of a confirmed case. Note that darker dots indicate multiple dots being overlapped.

show the partial autocorrelation functions (PACF) (Brockwell & Davis, 1991) of daily confirmed cases for three comunas in Cali. Short lags (less than one week) appear to be highly relevant to the current confirmed cases at each comuna, highlighting a significant temporal dependence. The last panel of Figure 4 shows the spatial correlation versus the distance between different locations in Cali. Specifically, we investigate the time series of cases occurrence rate (estimated by KDE) at 1,000 arbitrary locations. As we can see, a strong spatial correlation is observed in the vicinity of an arbitrary location, while the correlation between two locations weakens with their distance.

3 Methodology

This section presents our nonstationary spatio-temporal point process model for COVID-19 cases. We first revisit some essential background of STPPs in the following. Then we propose a novel point process model with a nonstationary kernel function, which captures complex triggering effects between events in time and space. Lastly, we characterize the influence of city landmarks as an exogenous promotion.

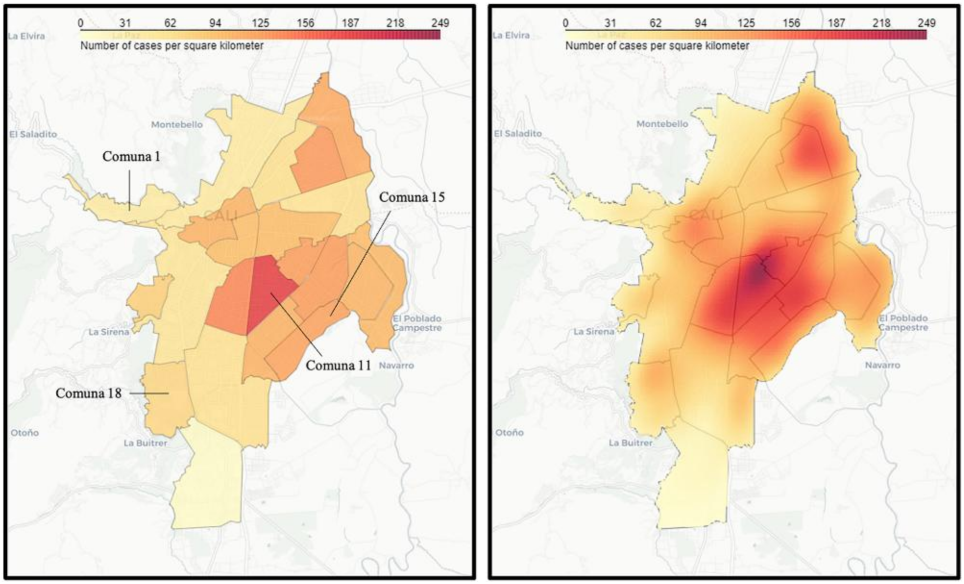

Figure 3. Spatial distribution of the confirmed cases at two spatial resolutions. The colour depth indicates the number of confirmed cases in 1 km^2; *Left panel* shows the case density per comuna. *Right panel* shows the spatially continuous case density estimated by KDE.

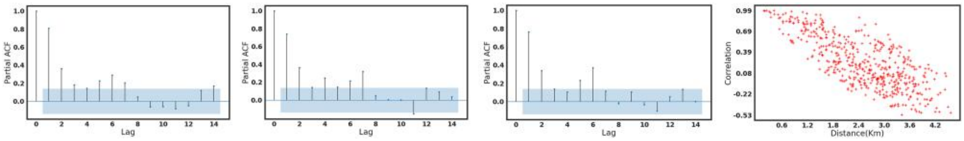

Figure 4. *Panel 1, 2, and 3*: partial autocorrelation function (PACFs) of the time series of confirmed cases in three comunas. The *x*-axis is the time lag in days. The shaded area represents nonsignificant PACFs. *Panel 4*: Correlation coefficients between the series of confirmed cases density at five arbitrarily chosen locations and those at other 199 locations in their neighbourhoods against the separating distances measured in kilometres.

3.1 Background: spatio-temporal point processes

STPPs is a popular model for discrete events data that occur in space and time González et al. (2016), Reinhart (2017). Denote the observation space as $\mathcal{X} = [0, T] \times \mathcal{S} \subseteq \mathbb{R}^{+} \times \mathbb{R}^{2}$, where T is the time horizon and $\mathcal{S}$ represents the space of geographic coordinate system (GCS). Each confirmed case is a *discrete event* defined by a data tuple $x := (t, s)$, where $t \in [0, T]$ is the time when the individual was diagnosed with COVID-19 and $s \in \mathcal{S}$ represents the location of residence of confirmed case. Let $\mathcal{H}_t := \{x_i = (t_i, s_i) \mid t_i < t\}$ denote the events' history before time t. Let $\mathbb{N}$ be a counting measure on $[0, T] \times \mathcal{S}$ corresponding to $\mathcal{H}_T$, i.e., for any $S \subset [0, T] \times \mathcal{S}$, $\mathbb{N}(S) = |S \cap \mathcal{H}_T|$, the number of occurred events in the set S. For any function $f : [0, T] \times \mathcal{S} \rightarrow \mathbb{R}$, the integral w.r.t. the counting measure is defined as

$$\int_S f(\tau, r)\, \mathrm{d}\mathbb{N}(\tau, r) = \sum_{(t_i, s_i) \in S \cap \mathcal{H}_T} f(t_i, s_i).$$

Given the observed history $\mathcal{H}_t$, the probability structure of the point process is characterized by the conditional intensity function $\lambda(t, s)$ (for notational simplicity, we omit the dependence on $\mathcal{H}_t$), which is defined as:

$$\lambda(t, s)\,\mathrm{d}t \cdot |B(s, \mathrm{d}s)| = \mathbb{E}[\mathrm{d}\mathbb{N}(t, s)|\mathcal{H}_t]. \tag{1}$$

Here, $B(s, \mathrm{d}s)$ is a ball centred at s in the space $\mathcal{S}$ with radius ds, and $|B(s, \mathrm{d}s)|$ is the Lebesgue measure.

Hawkes processes (Hawkes, 1971) is a type of self-exciting point process that captures the triggering effects between events. Assuming that influences from past events are linearly additive to the current event, the conditional intensity for a Hawkes point process takes the form of

$$\lambda(t, s) = \lambda_0 + \int_0^t \int_{\mathcal{S}} k(t, \tau, s, u)\,\mathrm{d}\mathbb{N}(\tau, u), \tag{2}$$

where $\lambda_0 > 0$ denotes the background intensity, and $k(t, t', s, s')$ is a triggering kernel function that captures the influence of past events on the likelihood of event occurrence at the current time. In this work, we do not assume the kernel function to be positive or shift-invariant (to capture the nonstationary process as we will define later on).

The parameters can be estimated by maximum-likelihood estimation (MLE). Given the observed point pattern $\boldsymbol{x}$, we can write the log-likelihood as

$$\ell(\boldsymbol{x}) = \sum_{i=1}^{\mathbb{N}([0,T]\times\mathcal{S})} \log \lambda(t_i, s_i) - \int_0^T \int_{\mathcal{S}} \lambda(\tau, u)\,\mathrm{d}u\,\mathrm{d}\tau, \tag{3}$$

where $\mathbb{N}([0, T] \times \mathcal{S})$ is the number of observed events (see the derivation of the log-likelihood in Online Supplementary Material, Appendix B).

The ETAS model is one of the most common STPPs (Ogata, 1988, 1998), which has been widely adopted in modeling typical spatio-temporal datasets such as earthquakes. ETAS model uses a Gaussian diffusion kernel

$$k(t, t', s, s') = \frac{C\,\mathrm{e}^{-\beta(t-t')}}{2\pi\sqrt{|\Sigma|}(t-t')} \cdot \exp\left\{-\frac{(s - s' - \mu)^{\top}\Sigma^{-1}(s - s' - \mu)}{2(t-t')}\right\},$$

where $\Sigma \equiv \mathrm{diag}(\sigma_x^2, \sigma_y^2)$ is a two-dimensional diagonal matrix representing the covariance of the spatial correlation, β is the decaying rate, μ is the mean shift, and C is a constant. However, the diffusion kernel is stationary and only depends on the spatio-temporal distance between two events. In addition, the kernel assumes the spatial correlation is isotropic and unable to capture complex spatial dependence.

3.2 A nonstationary triggering Gaussian kernel

We introduce a nonstationary triggering kernel, which can vary continuously over space and plays a vital role in modeling the heterogeneous spatial correlation across different regions. For model simplicity and computational efficiency, we adopt the commonly used assumption that the triggering effect of a past event is separable in space and time:

$$k(t, t', s, s') = \nu(t, t') \cdot \upsilon(s, s'),$$

where $\nu(t, t')$ is a kernel that captures the dependence between time t and t', and $\upsilon(s, s')$ is a spatial kernel that captures the nonstationary correlation between location s and s'.

3.2.1 A stationary temporal kernel

As the virus spreads and affects a significant portion of the population in a short period, we can assume temporal virus transmission is through a shift-invariant kernel with exponential decay:

$$\nu(t, t') = C\,\mathrm{e}^{-\frac{1}{2\sigma_0^2}(t-t')^2}, \quad t > t'.$$

Here, $C > 0$ is a parameter that controls the magnitude of the kernel, and $\sigma_0 > 0$ is a parameter that controls the decaying rate of the event's temporal influences. We assume $t > t'$ to capture the fact that a historical event at time t' impacts the current time t but not vice versa.

3.2.2 A nonstationary spatial kernel

The complex nature of the spatial spread of COVID-19 requires a nonhomogeneous and nonstationary spatial kernel function in the point process. Given two arbitrary locations $s, s' \in \mathcal{S}$, we define the spatial kernel $\upsilon(s, s')$ as a inner product between two feature mappings ϕ_s and $\phi_{s'}$, i.e.,

$$\upsilon(s, s') = \langle \phi_s, \phi_{s'} \rangle, \quad s, s' \in \mathcal{S},$$

where the inner product for functions $\langle f, g \rangle := \int_{\mathbb{R}^2} f(u)g(u)\,\mathrm{d}u$. We represent the feature mapping ϕ_s as a weighted sum of a set of R independent kernel-induced feature functions $\{\kappa_s^{(r)} := \kappa^{(r)}(s, \cdot)\}_{r=1}^R$:

$$\phi_s = \sum_{r=1}^{R} w_s^{(r)} \kappa_s^{(r)},$$

where $\kappa^{(r)} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_+$ is a general kernel and $w_s^{(r)}$ is the corresponding weight of that feature function at location s. The location-dependent weight satisfies $\sum_{r=1}^{R} w_s^{(r)} = 1$ at any arbitrary location s. Hence the spatial kernel can be re-written as

$$\upsilon(s, s') = \sum_{1 \le r_1, r_2 \le R} w_s^{(r_1)} w_{s'}^{(r_2)} \langle \kappa_s^{(r_1)}, \kappa_{s'}^{(r_2)} \rangle.$$

The rationale of this design is twofold: (a) Using a linear combination of the product of feature functions enhances the representative power of the spatial kernel. Note that when $r_1 = r_2$, the kernel captures self-correlation (self-similarity of feature functions) and otherwise captures the cross-correlation (similarity between two feature functions). (b) The spatial kernel can also be highly interpretable if $\kappa^{(r)}$ takes a specific parametric form; following the idea in Higdon et al. (1998), Zhu, Bukharin, et al. (2022), we choose κ_s to be a Gaussian function centred at s with covariance matrix Σ_s since the spatial correlation between two events decays as their distance increases in general. The spatial kernel is specified to be:

$$\upsilon(s, s') = \sum_{1 \le r_1, r_2 \le R} \frac{w_s^{(r_1)} w_{s'}^{(r_2)}}{2\pi |\Sigma_s^{(r_1)} + \Sigma_{s'}^{(r_2)}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (s - s')^\top (\Sigma_s^{(r_1)} + \Sigma_{s'}^{(r_2)})^{-1} (s - s') \right\}. \tag{4}$$

See detailed derivation of (4) in Online Supplementary Material, Appendix C. Figure 5 gives an example of the spatial kernel with two feature functions.

Now we specify the kernel-induced feature function κ_s. According to Higdon et al. (1998), there exists a one-to-one mapping between a bivariate normal distribution specified by Σ_s and its one standard deviation ellipse. Note that κ_s is centred at s, so the ellipse's centre is fixed at s. Thus we can specify the ellipse by a pair of focus points and the fixed area A. The focus points are denoted by $\boldsymbol{\psi}_s = (\boldsymbol{\psi}_x(s), \boldsymbol{\psi}_y(s))$ and $-\boldsymbol{\psi}_s = (-\boldsymbol{\psi}_x(s), -\boldsymbol{\psi}_y(s))$, where $\boldsymbol{\psi}_s \in \Psi \subset \mathbb{R}^2$. Hence, given $\boldsymbol{\psi}_s$ and A, the corresponding Σ_s can be written as

$$\Sigma_s = \tau_z^2 \begin{pmatrix} Q + \dfrac{\|\boldsymbol{\psi}_s\|^2}{2} \cos 2\alpha & \dfrac{\|\boldsymbol{\psi}_s\|^2}{2} \sin 2\alpha \\ \dfrac{\|\boldsymbol{\psi}_s\|^2}{2} \sin 2\alpha & Q - \dfrac{\|\boldsymbol{\psi}_s\|^2}{2} \cos 2\alpha \end{pmatrix},$$

where $Q = \sqrt{4A^2 + \|\boldsymbol{\psi}_s\|^4 \pi^2} / 2\pi$, $\alpha = \tan^{-1}(\boldsymbol{\psi}_y(s) / \boldsymbol{\psi}_x(s))$, $\tau_z > 0$ is a scaling parameter that controls

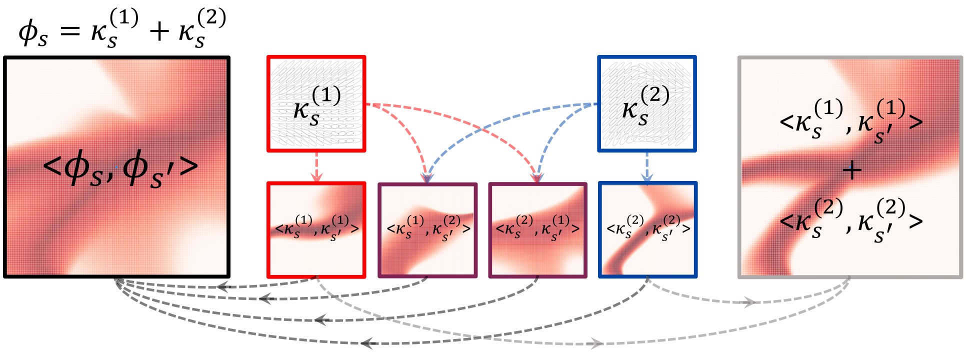

Figure 5. An example of the nonstationary spatial kernel with two feature functions evaluating at location s (the centre of the box), i.e., $\upsilon(s, s') = \langle \phi_s, \phi_{s'} \rangle$, $\forall s' \in \mathcal{S}$, where $\phi_s = \kappa_s^{(1)} + \kappa_s^{(2)}$. The two boxes in the middle indicate the cross-correlated terms ($\kappa_s^{(1)} \cdot \kappa_{s'}^{(2)}$ and $\kappa_s^{(2)} \cdot \kappa_{s'}^{(1)}$); the red and blue boxes indicate the self-correlated terms ($\kappa_s^{(1)} \cdot \kappa_{s'}^{(1)}$ and $\kappa_s^{(2)} \cdot \kappa_{s'}^{(2)}$).

the overall level of the covariance (see the derivation in Online Supplementary Material, Appendix D). We consider A as a hyper-parameter.

3.2.3 *Neural network-based kernel representation*

We develop a neural network-based representation for the kernel-induced feature function similar to the idea in Zhu, Bukharin, et al. (2022) and Zhu, Wang, et al. (2022). A key feature of our nonstationary spatial kernel is that for any location $s \in \mathcal{S}$, we can estimate a mapping that obtains the focus point $\boldsymbol{\psi}_s$ and the corresponding location-dependent weight w_s. To this end, we represent the mapping $\varphi : \mathcal{S} \to \Psi \times [0, 1]$ from the location to the space of focus points Ψ and the weights $[0, 1]$ using a fully connected multi-layer neural network. The input of the neural network is the two-dimensional location vector s, and the output is the concatenation of the corresponding focus point $\boldsymbol{\psi}_s$ and its weight w_s. Here, each hidden layer is equipped with a softplus activating function $f(x) = \log(1 + e^x)$ (see the detailed specification of the neural network in Section 5). Neural networks allow a flexible representation of the covariance and the corresponding kernel-induced feature function due to their well-known universal approximation power. In our implementation, we adopt the same network architecture for all R kernel-induced feature functions, as illustrated in Figure 6.

3.3 Exogenous promotion of city landmarks

To incorporate the influence of city landmarks, we assume each landmark has a constant exogenous promotion to the virus spread at their locations. To achieve this, we adopt an idea similar to Zhu, Ding, et al. (2021) and introduce an additional term to the conditional intensity function $\lambda(t, s)$ (2):

$$\lambda(t, s) = \lambda_0 + \sum_{l=1}^{L} \gamma_l g(s \mid s_l, \Sigma_l) + \sum_{(t', s') \in \mathcal{H}_t} k(t, t', s, s'). \tag{5}$$

The second and third terms represent the exogenous promotion at location s and the endogenous excitation at location s and time t, respectively. We use L to denote the number of landmarks, and γ_l indicates the significance of landmark l. We assume that the exogenous effect induced by landmarks decays with distance to them. Hence, the influence of landmark l located at s_l is modelled by a Gaussian function $g(s \mid s_l, \Sigma_l)$ centred at location $s_l \in \mathcal{S}$ with covariance Σ_l. Here, we define $\Sigma_l := \sigma_l^2 \mathbf{I}$, where $\mathbf{I}$ is an identity matrix.

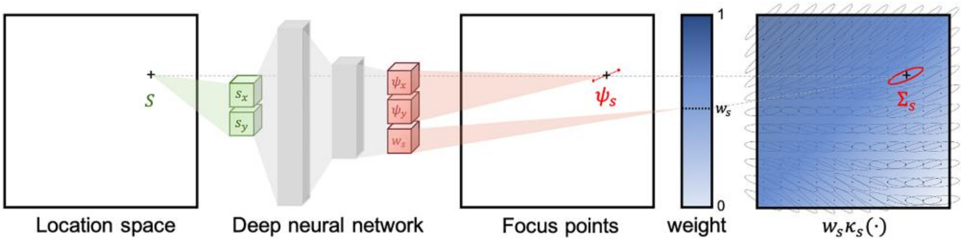

Figure 6. An illustration of a deep neural network that maps an arbitrary location s to a spatial kernel, consisting of a feature function κ_s (represented through focus points) and weight w_s.

4 Efficient computation of the log-likelihood function

The log-likelihood of the spatio-temporal point process defined in (3) is often intractable due to the double integral term. Numerical integral can also be expensive: if the number of randomly sampled points in a three-dimensional space is K and the total number of events is N, the computational complexity is $\mathcal{O}(KN)$ $(K \gg N)$ using commonly used numerical integration techniques. In our case, we can write the integral term as

$$\begin{aligned}\int_0^T\int_{\mathcal{S}} \lambda(\tau, u)\,\mathrm{d}u\,\mathrm{d}\tau = \lambda_0|\mathcal{S}|T + \int_0^T \sum_{l=1}^{L} \gamma_l \underbrace{\int_{\mathcal{S}} g(u \mid s_l, \Sigma_l)\,\mathrm{d}u}_{(i)}\,\mathrm{d}\tau \\ + \int_0^T \sum_{t_i<\tau} C\,\mathrm{e}^{-\frac{1}{2\sigma_0^2}(\tau-t_i)^2}\,\mathrm{d}\tau \cdot \underbrace{\int_{\mathcal{S}} \upsilon(u, s_i)\,\mathrm{d}u}_{(ii)},\end{aligned} \tag{6}$$

where $|\mathcal{S}|$ is the Euclidean area of the city, and evaluating (i), (ii) are difficult in general because (a) Both (i) and (ii) require the integral over the geographical space of Cali $\mathcal{S}$, which has an irregular shape; (b) In (ii), $\upsilon\,(u, s_i)$ is location-dependent and parameterized by a neural network.

We circumvent these two difficulties by simplifying the calculation of the integral without significantly impacting the model's accuracy: (a) We expand the integration region $\mathcal{S}$ to the entire geographical space $\mathbb{R}^2$ and account for the boundary effect error by ϵ_1. Note that the kernel g or κ_s are Gaussian concentrated around s and most events are located in the interior of $\mathcal{S}$ when choosing sufficiently large $\mathcal{S}$. As suggested by Ogata (1998), such boundary effect is usually negligible due to the decreased activity in the region's edges. (b) We assume the distance between two focus points ($\boldsymbol{\psi}_s$ and $-\boldsymbol{\psi}_s$) at an arbitrary location s is bounded by a threshold $2c$ (which can be obtained by rescaling the output of neural networks); a large distance between focus points leads to an overstretched ellipse, which is unrealistic in practice. Therefore, when performing numerical integration, we approximate the kernel-induced feature function κ_s by a standard Gaussian function denoted by κ_s^0, which corresponds to a standard deviation ellipse centred at s with area A. The relative error of the integral approximation is denoted by ϵ_2. In short, these two assumptions reduce the double integral (6) to an analytical form that can be evaluated directly without numerical integration.

Proposition 1 (Approximation of the integral in the likelihood function). Assume the area of the corresponding ellipse of κ_s is A and the distance between its focus points is restricted to be smaller than $2c$, then the integral in (6) can be approximated by

$$\begin{aligned}\int_0^T\int_{\mathcal{S}} \lambda(\tau, r)\,\mathrm{d}r\,\mathrm{d}\tau = (1+\epsilon_2)\Bigg[\lambda_0|\mathcal{S}|T + T\sum_{l=1}^{L}\gamma_l \\ + \sqrt{2\pi}C\sigma_0 \sum_{i=1}^{\mathbb{N}([0,T]\times\mathcal{S})} \left\{h\left(\frac{T-t_i}{\sigma_0}\right) - \frac{1}{2}\right\}\Bigg] - \epsilon_1,\end{aligned} \tag{7}$$

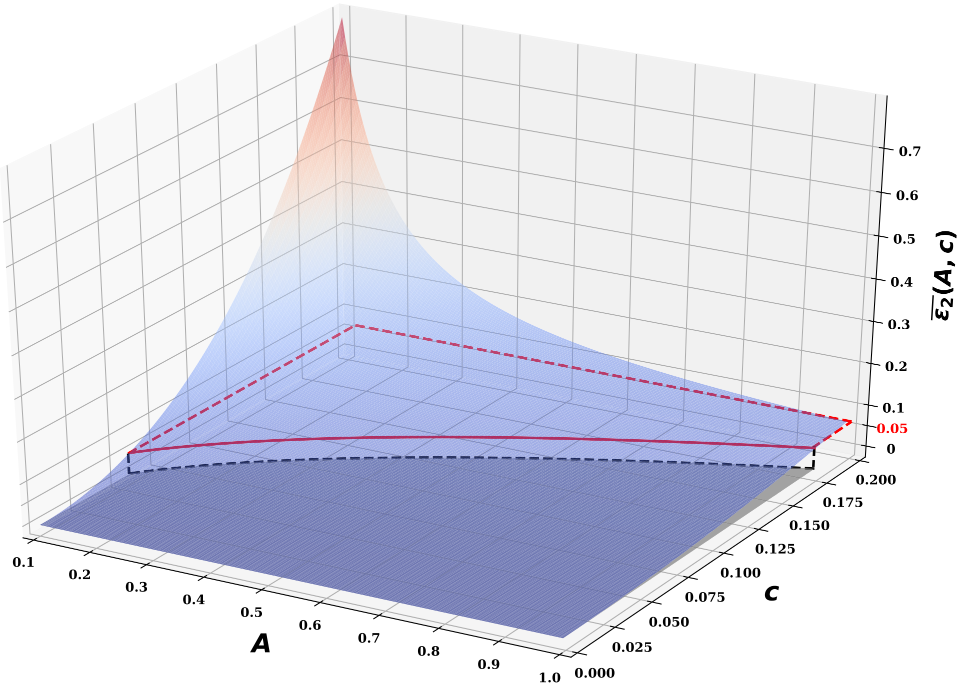

Figure 7. Surface plot for the upper bound of the relative error $\overline{\epsilon_2}$ with regards to hyper-parameters *A* and *c*. The horizontal coordinates represent the value of *A* and *c*, respectively, and the vertical coordinate represents the value of $\overline{\epsilon_2}(A,\ c)$. The red solid line is a surface contour valued at 0.05. The grey shaded area in the horizontal plane represents the set of (*A*, *c*) that satisfies $\overline{\epsilon_2}(A,\ c) < 0.05$. We can observe that the higher the value of *A* and the lower the value of *c*, the smaller the upper bound of the relative error.

where the function h is the cumulative density function of the standard normal distribution and ϵ_1, ϵ_2 are the boundary effect error and the relative error of the integral approximation, respectively. Ignoring the boundary effect error ϵ_1, the relative error $\epsilon_2 \in (-1,\ +\infty)$ can be bounded by:

$$|\epsilon_2| < \max\left\{U - 1,\ 1 - \frac{1}{U}\right\},$$

where $U = (\sqrt{4A^2 + c^4\pi^2} + c^2\pi)/2A$. (see the proof in Online Supplementary Material, Appendix E)

Remark 1 Proposition 1 leads to a computationally efficient calculation of the integral with complexity $\mathcal{O}(N)$. We denote the upper bound of the relative error as $\overline{\epsilon_2}$ and its dependence on hyper-parameters A and c is illustrated in Figure 7. In general, a larger c results in a more expressive spatial kernel but requires a larger A to control the approximation error. In practice, we select $c = 0.1$ and $A = 0.35$ to limit the relative error ϵ_2 under 0.05 and ensure a certain level of expressiveness for the spatial kernel.

5 COVID-19 data case study in Cali

In this section, we present the numerical results for studying the real COVID-19 data in Cali, which is described in Section 2. We first investigate the model's explanatory power by evaluating the in-sample performance and visualizing the estimated kernel-induced feature functions and

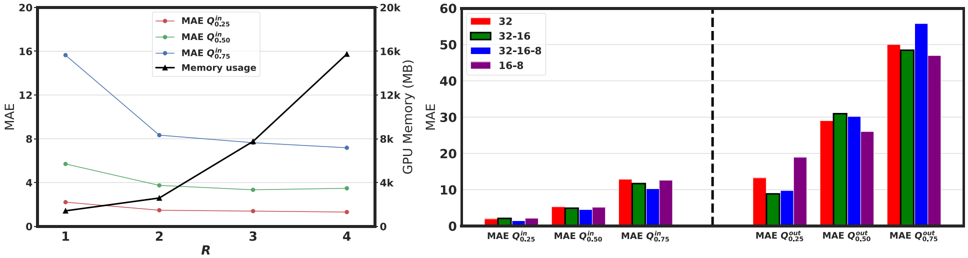

Figure 8. Performance of the proposed model with different numbers of components R in the feature mapping ϕ_S or different neural network architectures. *Left panel*: mean absolute error (MAE) and GPU memory usage of the in-sample estimations with $R = 1, 2, 3, 4$. The red, green, and blue lines represent three different quartiles of MAE for the in-sample estimations, respectively, and the black line represents the increase of GPU memory usage when R grows. *Right panel*: MAE of the in-sample and out-of-sample estimations with four different neural network architectures. The colour code and the corresponding number series represent different neural network structures; for example, '32-16' indicates a two-hidden-layer neural network, and there are 32 and 16 nodes for each layer. The left three groups show the MAEs of in-sample estimation. The right three groups show the MAEs of out-of-sample estimation. In the following experimental results, we adopt the architecture 32-16.

their corresponding spatial kernel. We also study the exogenous effects of the city landmarks. Finally, we compare the out-of-sample predictive performance of the proposed method with four baseline approaches. In this section, $\{\text{MAEQ}_q^{\text{in}}, \text{MAEQ}_q^{\text{out}}\}$ denote to the lower q-quantile of the mean absolute error (MAE) (Willmott & Matsuura, 2005) for the in-sample and out-of-sample estimation, respectively.

Our experimental settings are as follows. We consider a mixture kernel with $R = 3$ components, which achieves the balance between the predictive performance and the computational efficiency according to the results shown in Figure 8a. Figure 8b compares the out-of-sample performance for four network architectures; we choose a network architecture that achieves good performance for our data: a two-hidden-layer neural network with 32 and 16 nodes in each hidden layer for each kernel-induced feature function. We select the hyper-parameters $A = 0.35$ and $c = 0.1$ based on actual needs, and estimate model's parameters $\{\lambda_0, C, \sigma_0, \tau_z, \{\gamma_l\}_{l=1}^{L}, \{\sigma_l\}_{l=1}^{L}, \{\varphi^{(r)}\}_{r=1}^{R}\}$ by solving the maximum likelihood problem via gradient descent We train the model with the entire training set in each epoch. The initial learning rate is 1 and will decay to 0.1 of its last value when there is no likelihood increment for 10 epochs. The algorithm stops when the likelihood oscillation is less than 1 for 30 epochs. We use Adam optimizer (Kingma & Ba, 2017) for all experiments. In the following, we refer to the proposed framework as a NonStationary Spatio-Temporal Point Process (NSSTPP).

5.1 Model interpretation

To evaluate our model's goodness-of-fit, we compare the in-sample estimations of different models on the one-week-ahead number of cases, performed as follows. We first fit the model using the entire 28 weeks of data. The in-sample estimation can then be obtained by feeding the same data into the fitted model and finding an empirical expectation of the conditional intensity at a given week according to the equation (5). We compare our model with five baselines that are commonly adopted in modeling infectious epidemics: (a) homogeneous Poisson process (as a sanity check); (b) susceptible-infectious-recovered (SIR) model; (c) autoregressive (AR) time-series model; (d) epidemic-type aftershock sequence (ETAS) model; (e) Our model without exogenous effects (NSSTPP-Exo). See Online Supplementary Material, Appendix A for a detailed review of the baseline methods and their hyper-parameter choices. Here, we focus on predicting the number of cases for each comuna, which is a subregion of the city (instead of predicting the occurrence of an event), to compare the predictive accuracy with other discrete methods. In practice, it can be calculated by $\int_{s \in S^*} \hat{\lambda}(t^*, s)\, \mathrm{d}s$ using numerical integration, where t^* denotes the time and S^* denotes the region of comuna that we want to predict in. Figure 9 shows the estimated number of cases by different models in each comuna of Cali. Table 1 summarizes the in-sample estimation performance measured by two commonly used performance metrics, the log-likelihood and MAE. In Online Supplementary Material, Appendix A, we first present the results of prediction uncertainty

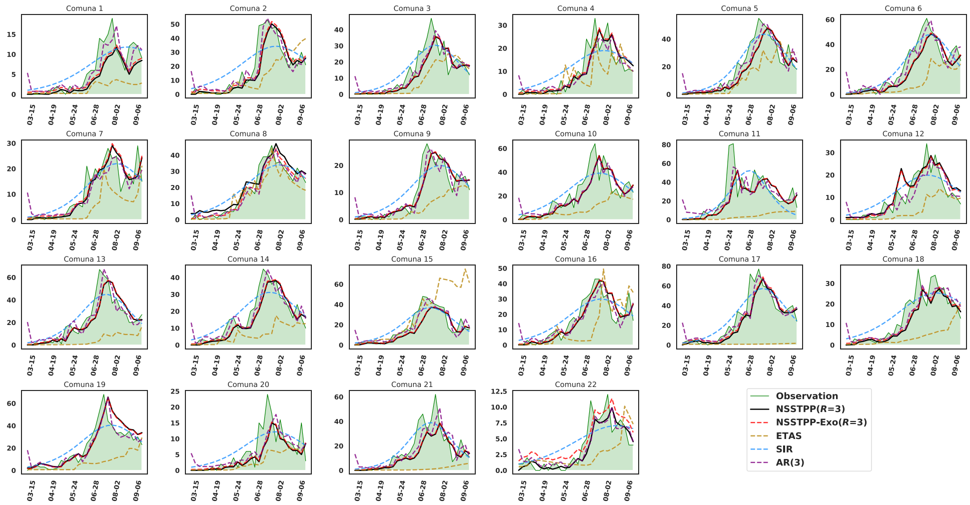

Figure 9. Comparison of the proposed model with baseline models. The green lines and shaded areas represent the ground truth. The black and red lines indicate the in-sample estimation of our nonstationary point process model. The yellow, blue, and purple lines represent the in-sample estimation of the ETAS model, SIR model, and AR(3) model, respectively.

Table 1. Performance of in-sample estimation. The bold values mark the best performance among all models regarding different metrics.

Models	Log-likelihood ($\times 10^4$)	MAE $Q^{\text{in}}_{0.25}$	MAE $Q^{\text{in}}_{0.5}$	MAE $Q^{\text{in}}_{0.75}$
Random	/	5.000	11.000	18.000
SIR	/	1.862	3.759	7.391
AR(3)	/	1.307	2.880	**6.496**
ETAS	$4.868_{(0.0058)}$	$1.486_{(0.102)}$	$4.737_{(0.240)}$	$14.895_{(0.420)}$
NSSTPP-Exo ($R = 1$)	$8.671_{(0.0772)}$	$0.834_{(0.087)}$	$3.145_{(0.186)}$	$7.922_{(0.374)}$
NSSTPP-Exo ($R = 2$)	$9.138_{(0.0886)}$	$0.806_{(0.088)}$	$2.728_{(0.178)}$	$7.119_{(0.361)}$
NSSTPP-Exo ($R = 3$)	$9.190_{(0.0906)}$	$0.853_{(0.090)}$	$\mathbf{2.613}_{(0.165)}$	$7.000_{(0.339)}$
NSSTPP ($R = 3$)	$\mathbf{9.331}_{(0.0937)}$	$\mathbf{0.797}_{(0.085)}$	$2.620_{(0.161)}$	$6.757_{(0.330)}$

Note. Numbers in brackets are standard deviation.

quantification (Sahai & Khurshid, 1993), and then compare the performance using different temporal kernels. The results show that our method outperforms other baseline approaches in both log-likelihood and MAE. Besides, we observe a significant performance gain compared to the ETAS model, which emphasizes the importance of the nonstationarity of the spatial kernel in capturing complex spatio-temporal patterns.

We study the in-sample explanatory power of our model and interpret the estimation results on the data in Cali. First, we visualize three learned spatial kernel-induced feature functions, which reveal the underlying spatio-temporal transmission dynamics of COVID-19 in Cali, as shown in Figure 10. Recall that at any location s, $\kappa_s^{(r)}$ is a Gaussian kernel with a spatially varying covariance matrix represented by two focus points of its one standard deviation ellipse. Therefore, we connect two focus points of each sampled covariance matrix over space using a red line segment for visualization. The angle and length of each red line can be interpreted as the direction and strength of influence at the particular location. The colour depth of the background represents the value of the corresponding weight $w_s^{(r)}$ at location s of each $\kappa_s^{(r)}$, indicating the significance of $\kappa_s^{(r)}$ at that

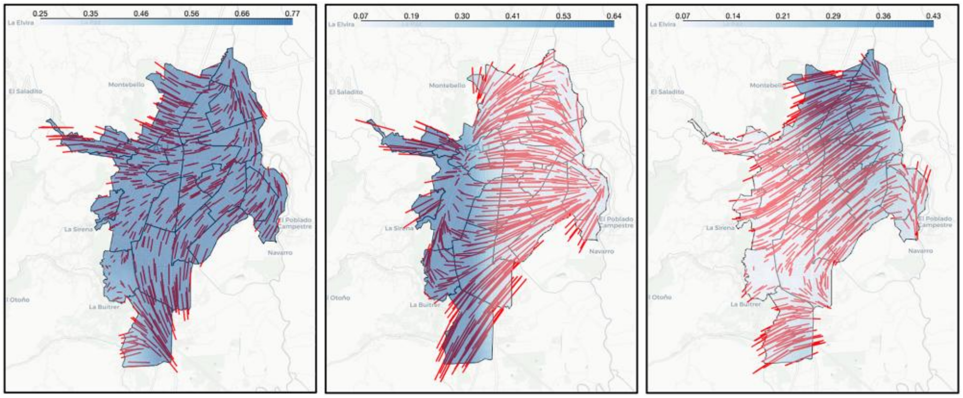

Figure 10. Visualization of three learned kernel-induced feature functions over Cali. Each panel shows $\kappa_s^{(1)}$, $\kappa_s^{(2)}$, and $\kappa_s^{(3)}$ over space, respectively. The line segments plotted above the polygons are edges that connect two focus points of location s. The shaded area shows the intensity of weight $w_s^{(r)}$ of each $\kappa_s^{(r)}$ over space. Darker colours mean larger weights.

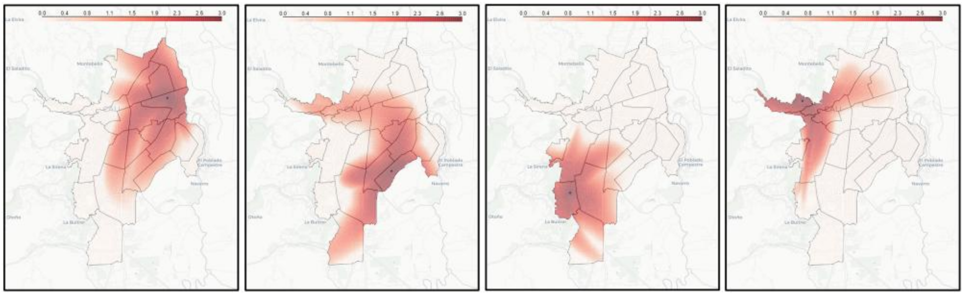

Figure 11. Evaluation of the spatial kernel $\upsilon(s, \cdot)$ with s fixed at four typical locations over space: city airport, centre of Comuna 15, centre of Comuna 18, and centre of Comuna 1. These panels intuitively show the spatial influence of the regional hubs located in different parts of the city. The dots represent the fixed location. The colour depth indicates the intensity of the kernel value, and the darker colour represents a higher kernel value.

location. These results suggest that the virus spreads rapidly across the region, following the diagonal direction from the Southwestern to the city's Northeastern. We can also observe a more subtle but complicated spreading pattern near the city's border.

Figure 11 visualizes the estimated spatial kernel $\upsilon(s, \cdot)$ given one of its input s, which can be treated as the influence of the location s. Here we present four examples, including the airport, the centre of Comuna 1, Comuna 15, and Comuna 18. Each example demonstrates that each location radiates the influence to its surrounding region in a different manner. The results show that the airport significantly influences the other city region as most northern areas have relatively high kernel values. As the most populated community in Cali, Comuna 15 also casts its influences on the city's Southeastern side. In addition, the impact of the location in Comuna 1 extends narrowly to two different directions, which correspond to two major routes in Cali. We note that these examples also emphasize the significance of the nonstationarity of the proposed method.

We also visually and quantitatively examine the exogenous effect of the city landmarks, as shown in Figure 12. Recall that the exogenous effect of each landmark is assumed to be an isotropic bivariate normal distribution, where γ_l and σ_l can be interpreted as the *intensity* and the *sphere of influence* of the exogenous effects of landmarks l, respectively. We visualize the learned σ_l and γ_l on the map of Cali in Figure 12. We also report the distributions of these two learned parameters for different categories of landmarks. As we can observe, the exogenous effects of the landmarks located in the centre of the city (the most severely affected areas) tend to have smaller intensities (γ_l)

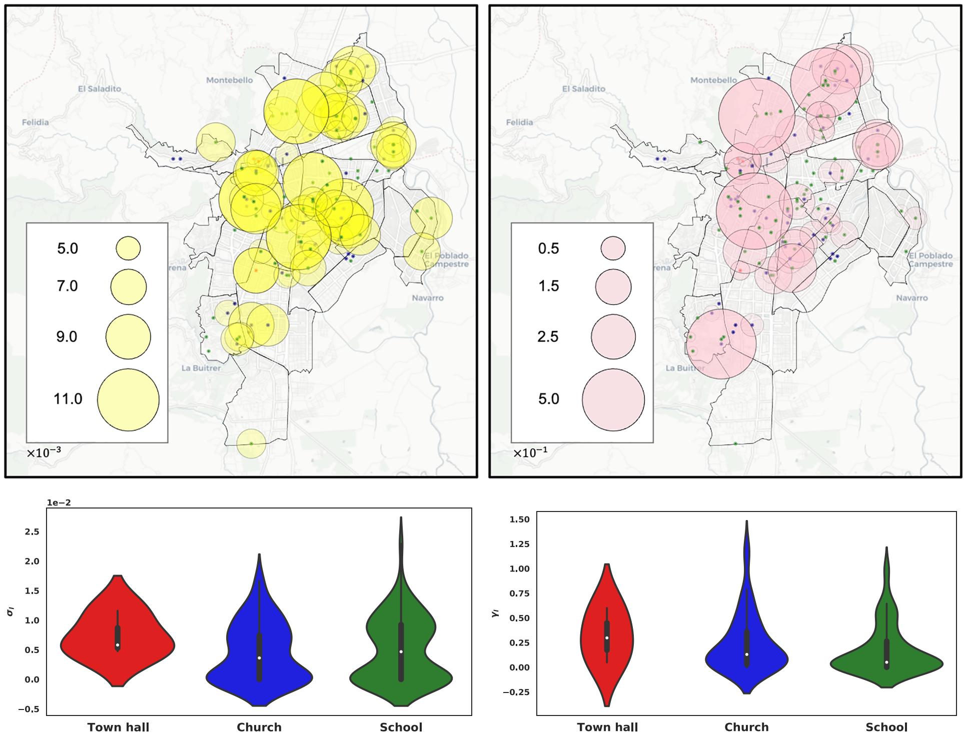

Figure 12. Estimated exogenous effects of landmarks in Cali. *Two upper panels* visualize the spatial distribution of learned $\{\sigma_l\}_{l=1}^{L}$ and $\{\gamma_l\}_{l=1}^{L}$, respectively. *Two lower panels* show the distributions of learned parameters for different categories of landmark.

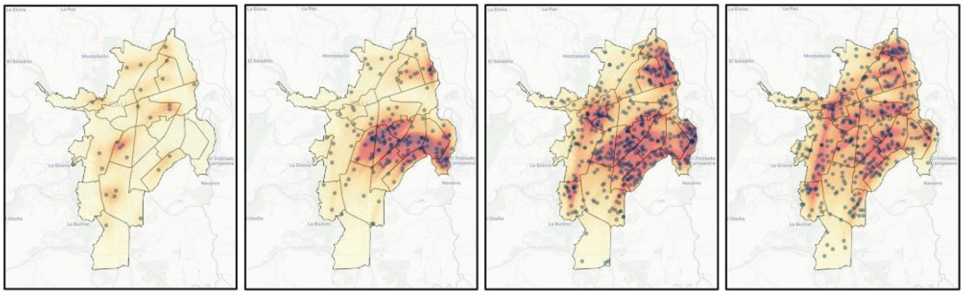

Figure 13. Predicted conditional intensity at four weeks: 22 March, 17 May, 28 June, and 30 August. The black dot represents an actual case reported in that week. The colour depth indicates the conditional intensity at the corresponding location, and a darker colour means a higher risk for citizens to be infected.

but larger spheres of influence (σ_l) on their neighbourhood. This result indicates that town halls may have a more significant influence than other landmarks. We see a natural explanation here: The landmarks located at the centre receive more people during the day. They act as super-spreaders of the virus, indicated by larger spheres of influence.

5.2 Predictive performance

We assess the model's predictive power by performing the one-week-ahead out-of-sample prediction of the number of cases. The out-of-sample prediction is a one-week-ahead prediction. We randomly hold out the data of one certain week $t*$ as the testing data and train the model using the

Table 2. Performance of out-of-sample estimation for last four weeks. The bold values mark the best performance among all models regarding different metrics.

Models	MAE $Q^{\text{out}}_{0.25}$	MAE $Q^{\text{out}}_{0.5}$	MAE $Q^{\text{out}}_{0.75}$
Random	10.000	19.000	28.000
SIR	7.101	12.937	18.809
AR(3)	3.332	6.571	12.433
ETAS	$4.908_{(0.820)}$	$12.143_{(0.848)}$	$22.238_{(0.852)}$
NSSTPP-Exo ($R = 1$)	$4.783_{(0.661)}$	$8.944_{(0.748)}$	$15.834_{(0.843)}$
NSSTPP-Exo ($R = 2$)	$4.710_{(0.643)}$	$9.038_{(0.722)}$	$15.804_{(0.816)}$
NSSTPP-Exo ($R = 3$)	$4.548_{(0.648)}$	$8.229_{(0.734)}$	$15.750_{(0.809)}$
NSSTPP ($R = 3$)	$\mathbf{2.843}_{(0.585)}$	$\mathbf{6.398}_{(0.751)}$	$\mathbf{10.495}_{(0.892)}$

data before this week. Then we use the fitted model to predict the conditional intensity function $\hat{\lambda}(t^*, s)$ for the week t^* over space. Figure 13 shows the predicted conditional intensity at four particular weeks, which represent four different stages of the pandemic: (a) the early stage, (b) the week before the first outbreak, (c) the week before the second outbreak, and (d) the week in the stabilized plateau of the pandemic development. As we can observe, our method can capture the spatial occurrences of these cases, detect regions with sparsely distributed cases by showing a lower intensity, and show a higher intensity in other regions with densely distributed cases. To further validate the effectiveness of our model, we implement the out-of-sample prediction for the last four weeks in our data. We report the prediction MAE and corresponding confidence interval for our proposed model (NSSTPP) and other baselines in Table 2. The results validate that our model significantly outperforms other baseline methods for these four weeks, verifying that our model does not overfit data.

6 Conclusions

Based on an unprecedented fine-grained COVID-19 dataset in Cali, Colombia, we propose a spatio-temporal point process framework equipped with a nonstationary kernel to model the epidemic transmission at an individual level. The kernel is composed of a set of kernel-induced feature functions. Each feature function is represented by a neural network aiming to enhance the model flexibility while being interpretable. We also develop an efficient log-likelihood estimation by approximating the double integral using an analytical expression. Our numerical study in Cali has shown that the proposed approach achieves promising predicting performance and the learned model is highly interpretable.

We believe our methodology combines in a natural while novel way theory of point processes with machine learning methods (such as neural networks), providing a unified framework for dealing with highly nonstationary spatio-temporal point patterns. The method is more general than the focused application and can be used, extended, and adapted to several natural phenomena represented by locations in space and time.

There are many ways the proposed method can be extended, and one possibility could be considering nonGaussian kernels and alternative neural network methods. In any case, the data should always guide these ad-hoc adaptations.

The global results finding in this work for the city of Cali show an increased risk of contracting COVID-19 in the centre, northeast, and northeast of the city, which are located the communes with more unsatisfied basic needs. On the other hand, in the south of the city, the risk of contagion is lower, and it is an area where people with greater purchasing power live. Considering the locations of the city landmarks in the model as transmission sources is undoubtedly an indispensable tool for predicting the spread of the virus. These outputs are very close to the reality experienced in that region of Colombia during the pandemic. The current official figures for Cali city show significant progress in the fight against the COVID-19.

Acknowledgments

The authors are grateful to the Municipal Public Health Secretary of Cali, Valle del Cauca, Colombia for providing the COVID-19 data used in this paper.

Conflict of interest: The authors declare no conflicts of interest.

Funding

The work of Z.D., S.Z., and Y.X. is partially supported by National Science Foundation CCF-1650913, CMMI-2015787, DMS-1938106, DMS-1830210. F.J.R.C. has been partially supported by Universidad Nacional de Colombia, HERMES projects, grant/award number: 51279.

Data availability

The COVID-19 dataset and landmark information in the city of Cali are available at https://github.com/McDaniel7/COVID-Cali-Colombia. The code is available upon request.

Supplementary material

Supplementary material are available at *Journal of the Royal Statistical Society: Series C* online.

References

Agosto A., & Giudici P. (2020). A poisson autoregressive model to understand COVID-19 contagion dynamics. *Risks*, *8*(3), 77. https://doi.org/10.3390/risks8030077

Angulo J., Yu H.-L., Langousis A., Kolovos A., Wang J., Madrid A. E., & Christakos G. (2013). Spatiotemporal infectious disease modeling: A BME-SIR approach. *PLoS ONE*, *8*(10), e72168-e72168. https://doi.org/10.1371/journal.pone.0072168

Bai Y., Safikhani A., & Michailidis G. (2020). 'Non-stationary spatio-temporal modeling of COVID-19 progression in the US', medRXiv, preprint: not peer reviewed. https://doi.org/10.1101/2020.09.14.20194548

Berry I., Soucy J.-P. R., Tuite A., & Fisman D. (2020). Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *Canadian Medical Association Journal*, *192*(15), E420–E420. https://doi.org/10.1503/cmaj.75262

Bizzarri M., Di Traglia M., Giuliani A., Vestri A., Fedeli V., & Prestininzi A. (2020). New statistical RI index allow to better track the dynamics of COVID-19 outbreak in Italy. *Scientific Reports*, *10*(1), 22365. https://doi.org/10.1038/s41598-020-79039-x

Brandén M., Aradhya S., Kolk M., Härkönen J., Drefahl S., Malmberg B., Rostila M., Cederström A., Andersson G., & Mussino E. (2020). Residential context and COVID-19 mortality among adults aged 70 years and older in Stockholm: A population-based, observational study using individual-level data. *The Lancet Healthy Longevity*, *1*(2), e80–e88.https://doi.org/10.1016/S2666-7568(20)30016-7

Brockwell P. J., & Davis R. A. (1991). Estimation of the mean and the autocovariance function. In *Springer Series in Statistics* (pp. 218–237). Springer. https://doi.org/10.1007/978-1-4419-0320-4_7

Chiang W.-H., Liu X., & Mohler G. (2020). 'Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates', medRxiv. https://www.medrxiv.org/content/early/2020/12/20/2020.06.06.20124149

Chriscaden K. (2020). Impact of COVID-19 on people's livelihoods, their health and our food systems. https://www.who.int/news/item/13-10-2020-impact-of-COVID-19-on-people's-livelihoods-their-health-and-our-food-systems

Du N., Dai H., Trivedi R., Upadhyay U., Gomez-Rodriguez M., & Song L. (2016). Recurrent marked temporal point processes. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. https://doi.org/10.1145/2939672.2939875

Farajtabar M., Yang J., Ye X., Xu H., Trivedi R., Khalil E., Li S., Song L., & Zha H. (2017). Fake news mitigation via point process based intervention. In *Proceedings of the 34th International conference on machine learning,* 1097-1106. PMLR. https://doi.org/10.48550/arXiv.1703.07823

Fu B., Xu X., & Wei H. (2020). Why tocilizumab could be an effective treatment for severe COVID-19? *Journal of Translational Medicine*, *18*(1), 1–5. https://doi.org/10.1186/s12967-020-02339-3

Gajardo Á, & Müller H.-G. (2021). Point process models for COVID-19 cases and deaths. *Journal of Applied Statistics*, 1–16. https://doi.org/10.1080/02664763.2021.1907839

Giudici P., Pagnottoni P., & Spelta A. (2023). Network self-exciting point processes to measure health impacts of COVID-19. *Journal of the Royal Statistical Society Series A: Statistics in Society*, qnac006. https://doi.org/10.1093/jrsssa/qnac006

González J. A., Rodríguez-Cortés F. J., Cronie O., & Mateu J. (2016). Spatio-temporal point process statistics: A review. *Spatial Statistics*, *18*(Part B), 505–544.

Guenther T., Czech-Sioli M., Indenbirken D., Robitailles A., Tenhaken P., Exner M., Ottinger M., Fischer N., Grundhoff A., & Brinkmann M. (2020). Investigation of a superspreading event preceding the largest meat processing plant-related SARS-Coronavirus 2 outbreak in Germany. *SSRN Journal*, 1-30. https://ssrn.com/abstract=3654517

Guo C., *et al.* (2020). Tocilizumab treatment in severe COVID-19 patients attenuates the inflammatory storm incited by monocyte centric immune interactions revealed by single-cell analysis. bioRxiv.

Harko T., Lobo F. S. N., & Mak M. K. (2014). Exact analytical solutions of the susceptible-infected-recovered (SIR) epidemic model and of the sir model with equal death and birth rates. *Applied Mathematics and Computation*, *236*(2014). 184–194. https://doi.org/10.1016/j.amc.2014.03.030

Hawkes A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90. https://doi.org/10.1093/biomet/58.1.83

Hendry D. F., & Pretis F. (2016). All change! The implications of non-stationarity for empirical modelling, forecasting and policy. *SSRN Journal*, 1-28. https://ssrn.com/abstract=2898761

Higdon D., Swall J. L., & Kern J. (1998). Non-stationary spatial modeling. *Bayesian Statistics*, *6*. https://doi.org/10.48550/arXiv.2212.08043

Hochreiter S., & Schmidhuber J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Institute for Health Metrics and Evaluation (2020). Modeling COVID-19 scenarios for the United States. *Nature Medicine*, *27*(1), 94–105.https://doi.org/10.1038/s41591-020-1132-9

James A., Eagle L., Phillips C., Hedges D. S., Bodenhamer C., Brown R., Wheeler J. G., & Kirking H. (2021). High COVID-19 attack rate among attendees at events at a church—Arkansas, March 2020. https://www.cdc.gov/mmwr/volumes/69/wr/mm6920e2.htm

Jasper F.-W. C., *et al.* (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet*, *395*(10223), 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9

John Hopkins University (2020). Coronavirus resource center. https://coronavirus.jhu.edu

Kingma D. P., & Ba J. (2017). Adam: A method for stochastic optimization. https://doi.org/10.48550/arXiv.1412.6980

Korolev I. (2021). Identification and estimation of the SEIRD epidemic model for COVID-19. *Journal of Econometrics*, *220*(1), 63–85. https://doi.org/10.1016/j.jeconom.2020.07.038

Kraemer M. U. G., Yang C.-H., Gutierrez B., Wu C.-H., Klein B., Pigott D. M., du Plessis L., Faria N. R., Li R., Hanage W. P., Brownstein J. S., Layan M., Vespignani A., Tian H., Dye C., Pybus O. G., & Scarpino S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497. https://doi.org/10.1126/science.abb4218

Lang T., Plagemann C., & Burgard W. (2007). Adaptive non-stationary kernel regression for terrain modeling. In *Robotics: Science and systems* (Vol. 6). Citeseer.

Leclerc Q. J., Fuller N. M., Knight L. E., Funk S., & Knight G. M. (2020). What settings have been linked to SARS-COV-2 transmission clusters? *Wellcome Open Research*, *5*, 83. https://doi.org/10.12688/wellcomeopenres.15889.2

Li S., Wang L., Chen X., Fang Y., & Song Y. (2021). Understanding the spread of COVID-19 epidemic: A spatio-temporal point process view. *CoRR*. https://arxiv.org/abs/2106.13097

Lin Q., Zhao S., Gao D., Lou Y., Yang S., Musa S. S., Wang M. H., Cai Y., Wang W., Yang L., & He D. (2020). A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International Journal of Infectious Diseases*, *93*, 211–216. https://doi.org/10.1016/j.ijid.2020.02.058

Loli Piccolomini E., & Zama F. (2020). Monitoring Italian COVID-19 spread by a forced SEIRD model. *PLoS ONE*, *15*(8), e0237417. https://doi.org/10.1371/journal.pone.0237417

López-Feldman A., Heres D., & Marquez-Padilla F. (2021). Air pollution exposure and COVID-19: A look at mortality in Mexico City using individual-level data. *Science of the Total Environment*, *756*, 143929. https://doi.org/10.1016/j.scitotenv.2020.143929

Mamode Khan N., Soobhug A. D., & Heenaye-Mamode Khan M. (2020). Studying the trend of the novel coronavirus series in mauritius and its implications. *PLoS ONE*, *15*(7), e0235730. https://doi.org/10.1371/journal.pone.0235730

Mei H., & Eisner J. (2016). The neural Hawkes process: A neurally self-modulating multivariate point process. *CoRR*. http://arxiv.org/abs/1612.09328

Nande A., Adlam B., Sheen J., Levy M. Z., & Hill A. L. (2020). 'Dynamics of COVID-19 under social distancing measures are driven by transmission network structure', medRxiv. https://doi.org/10.1101/2020.06.04.20121673

New York Times (2020). Coronavirus data in the united states. https://www.nytimes.com/article/coronavirus-county-data-us.html

Northeastern University, Laboratory for the Modeling of Biological and Socio-technical Systems (2021). COVID-19 modeling. https://covid19.gleamproject.org/

Ogata Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27. https://doi.org/10.2307/2288914

Ogata Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402. https://doi.org/10.1023/A:1003403601725

Presidency of the Republic of Colombia (2020). Decrees during the COVID-19 pandemic. https://coronaviruscolombia.gov.co/Covid19/decretos.html

Reinhart A. (2017). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, *33*(3), 299-318. https://doi.org/10.1214/17-STS629

Remes S., Heinonen M., & Kaski S. (2017). Non-stationary spectral kernels. In *Advances in neural information processing systems*, 30. https://doi.org/10.48550/arXiv.1811.10978

Rizoiu M.-A., Xie L., Sanner S., Cebrian M., Yu H., & Van Hentenryck P. (2017). Expecting to be hip: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th International Conference on World Wide Web*. https://doi.org/10.1145/3038912.3052650

Sahai H., & Khurshid A. (1993). Confidence intervals for the mean of a poisson distribution: A review. *Biometrical Journal*, *35*(7), 857–867. https://doi.org/10.1002/bimj.4710350716

Triaccaa M., & Triacca U. (2021). Forecasting the number of confirmed new cases of COVID-19 in Italy for the period from 19 May to 2 June 2020. *Infectious Disease Modelling*, *6*(2), 362–369. https://doi.org/10.1016/j.idm.2021.01.003

Vasudevan S., Ramos F., Nettleton E., & Durrant-Whyte H. (2011). Non-stationary dependent gaussian processes for data fusion in large-scale terrain modeling. In *2011 IEEE International Conference on Robotics and Automation* (pp. 1875–1882). IEEE.

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., & Polosukhin I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (pp. 5998–6008).

Wikipedia (2021). Cali. https://en.wikipedia.org/wiki/Cali

Willmott C., & Matsuura K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79–82. https://doi.org/10.3354/cr030079

Woody S., Tec M. G., Dahan M., Gaither K., Lachmann M., Fox S., Meyers L. A., & Scott J. G. (2020). 'Projections for first-wave COVID-19 deaths across the us using social-distancing measures derived from mobile phones', medRxiv.

Zhang Q., Lipani A., Kirnap O., & Yilmaz E. (2020). Self-attentive Hawkes process. In H. D. III & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning*, vol. 119 of *Proceedings of Machine learning research* (pp. 11183–11193). PMLR. http://proceedings.mlr.press/v119/zhang20q.html

Zhu S., Bukharin A., Xie L., Santillana M., Yang S., & Xie Y. (2021). High-resolution spatio-temporal model for county-level COVID-19 activity in the US. *ACM Transactions on Management Information Systems (TMIS)*, *12*(4), 1–20. https://doi.org/10.1145/3468876

Zhu S., Bukharin A., Xie L., Yamin K., Yang S., Keskinocak P., & Xie Y. (2022). Early detection of COVID-19 hotspots using spatio-temporal data. *IEEE Journal of Selected Topics in Signal Processing*, *16*(2), 250-260. https://doi.org/10.1109/JSTSP.2022.3154972

Zhu S., Ding R., Zhang M., Van Hentenryck P., & Xie Y. (2021). Spatio-temporal point processes with attention for traffic congestion event modeling. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 7298-7309. https://doi.org/10.1109/TITS.2021.3068139

Zhu S., Li S., Peng Z., & Xie Y. (2021). Imitation learning of neural spatio-temporal point processes. *IEEE Transactions on Knowledge and Data Engineering*, *34*(11), 5391–5402. https://doi.org/10.1109/TKDE.2021.3054787

Zhu S., Wang H., Dong Z., Cheng X., & Xie Y. (2022). Neural spectral marked point processes. In *International Conference on Learning Representations*. https://openreview.net/forum?id=0rcbOaoBXbg

Zhu S., Zhang M., Ding R., & Xie Y. (2021). Deep fourier kernel for self-attentive point processes. In *International Conference on Artificial Intelligence and Statistics* (pp. 856–864). PMLR.