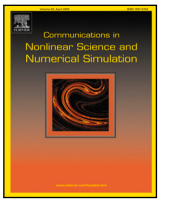
Commun Nonlinear Sci Numer Simulat 136 (2024) 108100

Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

journal homepage: www.elsevier.com/locate/cnsns



Research paper

A stochastic model for the early stages of highly contagious epidemics by using a state-dependent point process

Jonathan A. Chávez Casillas

Department of Mathematics and Applied Mathematical Sciences, University of Rhode Island, 5 Lippit Rd, Kingston, 02881, RI, USA

ARTICLE INFO

Keywords:
SIRQ-model
Overdispersion
Self-exciting process
Quarantine distribution
Stochastic-intensity
State-dependent

ABSTRACT

The recent COVID-19 pandemic has shown that when the reproduction number is high and there are no proper measurements in place, the number of infected people can increase dramatically in a short time, producing a phenomenon that many stochastic SIR-like models cannot describe: overdispersion of the number of infected people (i.e., the variance of the number of infected people during any interval is very high compared to the average). To address this issue, in this paper we explore the possibility of modeling the total number of infections as a state-dependent self-exciting point process. In this way, infections are not independent among themselves, but any infection will increase the likelihood of a new infection while also the number of currently infected and recovered individuals are included into determining the likelihood of new infections, Since long term simulation is extremely computationally intensive, exact expressions for the moments of the processes determining the number of infected and recovered individuals are computed, while also simulation algorithms for these state-dependent processes are provided.

1. Introduction

As the COVID-19 pandemic unfolded, we found the need to develop models that capture the salient features of the epidemics. However, as noted in [1], different types of models are appropriate at different stages, and for addressing different kinds of questions. For example, some statistical methods based on machine learning techniques are very useful in predicting the behavior on the short term. However, they are not very effective for the long term predictions nor for describing evolving circumstances. In this sense, stochastic models could be seen as a more robust and effective tool because they allow for a more systematic representation of complex social interactions, individual and collective behavioral adaptation, and public policies. The recent pandemic had returned epidemic modeling to the forefront of worldwide public policy-making [2]. Many scientists from different fields, particularly medicine, biology, mathematics, physics and chemistry (for a few, non-exhaustive, list of examples, see [3–9]) tried to contribute in the development of new models.

Because of their versatility and capabilities to model properties that regular stochastic SIR models cannot, point process models have been recently explored as a viable option. Models using these processes are data-driven and flexible enough to allow for parametric or nonparametric estimation of the reproduction number and the time scale at which the contagions occur (see [2]). Furthermore, at some level, they can also be viewed as stochastic versions of the popular compartmental models used in epidemiology. Because these processes have just been recently introduced, a very brief introduction to point processes is provided as well as how they have recently been used to model the COVID-19 epidemics.

E-mail address: jchavezc@uri.edu.

https://doi.org/10.1016/j.cnsns.2024.108100
Received 27 January 2023; Received in revised form 22 May 2024; Accepted 23 May 2024
Available online 25 May 2024

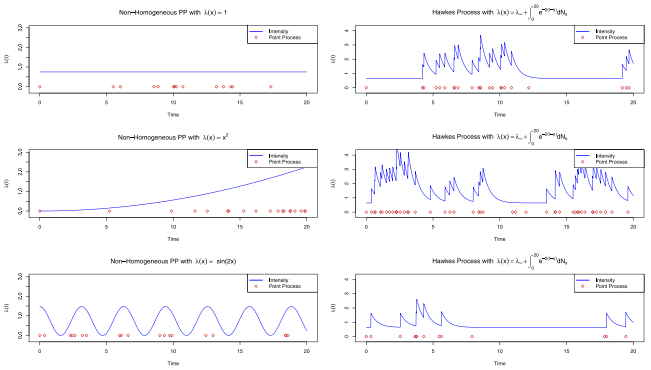


Fig. 1. Six realizations of point process with $T = 20$ and $\mathbb{E}[N_T] = 25$. On the left, three different inhomogeneous Poisson processes with intensities proportional to a constant, x^2 and $\sin(2x)$ are shown, whereas on the right three realizations of *the same* Hawkes process are displayed.

1.1. An introduction to point processes

Point processes can be roughly understood as a random set of points in a space $\mathcal{X}$. The topology and properties of the space can be very general, making them a versatile tool for modeling different phenomena. One of the most common usages of point processes are occurrences of events in time (called temporal processes), location of objects (called spatial processes) in space, or a mixture of the previous two (called spatio-temporal processes).

One way to characterize point processes is through its *stochastic intensity*, which provides a summary, at any given time t, of the likelihood that some new future event arrives in the time interval $[t, t+h)$ given its history, i.e., given the times of all past events up until time t.

Exhaustive treatments containing all the important probabilistic and statistical features of Point processes can be found in, e.g. [10–12].

A realization of a point process over $[0,\infty)$ is a sequence $\{T_0\}_{n\geq 1}$ in $[0,\infty]$ such that $T_0 = 0$, $T_n < \infty$ and $T_n < T_{n+1}$. For each realization of the point process, there is a counting function, or counting process, N_t defined as

$$N_t = \begin{cases} n & \text{if} \quad t \in [T_n, T_{n+1}),\ n \geq 0 \\ \infty & \text{if} \quad t \geq \lim_{n\to\infty} T_n \end{cases}$$

The counting process N_t indicates how many events have occurred up to time t. The inter-arrival times between events is $\tau_n := T_n - T_{n-1}$ for $n \geq 1$ and for $\tau_0 := 0$, N_t can be written as $N_t = \max_{n\geq 0}\{\tau_0 + \tau_1 + \cdots + \tau_n \leq t\}$. Many of the point processes found in the literature are simple, which means that only one event arrives each time. That is, $\lim_{h\to 0} \mathbb{P}\left[\frac{N(t+h)-N(t)>1}{h}\right] = 0$.

To explain the above definitions, we can look at the simplest and most common point process: the Poisson process. In this case, the intensity is completely deterministic, i.e. $\lambda(t) : [0,\infty) \to [0,\infty)$ and

$$\mathbb{P}[N(t) - N(s) = k] = e^{-\int_s^t \lambda(u)du} \frac{1}{k!} \left[\int_s^t \lambda(u)du\right]^k.$$

Since the intensity is deterministic we know that the *infinitesimal* probability of a new event in the interval $[t, t+h)$ is roughly $\approx h\lambda(t)$, with this probability being independent of the process' history. In this case, it can be shown that $\mathbb{E}[N_t] = \int_0^t \lambda(s)ds$.

The Poisson process is the canonical model for epidemic processes and most of the classic stochastic SIR models can be viewed as Poisson processes. However, sometimes the phenomena that need to be analyzed exhibit a "clustering effect", where the appearance of one new event will trigger the occurrence of more new events, which the Poisson process is incapable of capturing. In the Hawkes Processes family, the (stochastic) intensity function, $\Lambda(t)$, "feeds" itself from the point process N_t. That is, the intensity increases when a point arrives, which in turn will trigger more points to arrive. In the most classic example, the intensity can be characterized by

$$\Lambda(t) = \lambda_0 + \int_0^t \phi(t-s)dN_s = \lambda_0 + \sum_{T_i<t} \phi(t - T_i), \tag{1}$$

where $N_t = \{T_1, T_2, T_3, \ldots, T_{N_t}\}$ is itself the random point process and T_i is the time of the $i-$th occurrence.

Moreover, if $\phi(t) = Qe^{-rt}$, then the process (N_t, λ_t) is a Markov Process (a process whose evolution depends solely on the current state of the system and not in the history of *how* it arrived to such current state), and the intensity function decays exponentially between events. As before, we can also compute the expected number of events in the interval $[0, T]$ which turns out to be $\mathbb{E}[N_t] = \frac{\lambda_0}{1-\int_0^\infty \phi(s)ds}$.

Fig. 1, tries to illustrate the concepts and processes described above. In all the plots, the intensity function and a realization of the point process are shown. At the same time, all the processes below were "standardized" by setting the *expected* number of points to be the same in all cases. The left column shows 3 different Poisson processes with different intensities. In the right

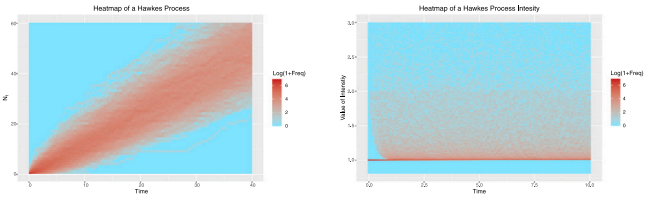


Fig. 2. Heatmap of the Hawkes process and its intensity. The frequency is given in a logarithmic scale.

column, 3 different realizations of the same Hawkes process are plotted. Because the intensity is stochastic, it will change between realizations, but more importantly the *clustering* phenomena can be observed. Whenever a new event arrives, the intensity spikes and thus the likelihood of more events happening increases. However, in between arrivals of events, the intensity decreases exponentially reducing the likelihood of more events coming but as there is always a baseline at which the intensity process cannot go below, it guarantees the occurrence of a new event at some point. To illustrate this, we present, in Fig. 2, a heatmap of the Hawkes process N_t and one of its intensity λ_t.

1.2. Linking point processes to epidemiology

Epidemic models have been trying to model, understand and predict different features and generalities of the pandemics that humanity has lived with. However, it was not until around 1930 that the first stochastic epidemic model was created. Before that, all the models were deterministic and were based on creating different systems of ODEs, but some desirable random effects were missing. Nonetheless, those deterministic models provided the skeleton upon which the behavior of corresponding stochastic systems is built. Furthermore, many times, when populations are large, the limiting behavior of such stochastic systems converges to a corresponding system of ODEs. The main models discussed in this proposal will be variations of the famous stochastic susceptible–infected–recovered (SIR) models and belong to the class of the so-called compartmental models, where each individual is placed into one of the compartments and the individuals move between the compartments under predetermined rules. For a survey on the description of many stochastic models, see [13].

The most common stochastic epidemic model, and the building block for many of the more detailed and precise models produced nowadays, is the so-called *general stochastic model*,

- There are three compartments where an individual can be: Susceptible (S_t), Infected (I_t) and Recovered (R_t), satisfying $N = S_t + I_t + R_t$, with N being the population size. Moreover, each individual belongs to one and only one compartment.
- A recovered individual cannot contract the disease again.
- During the interval $[t, t+\Delta t]$, one of the following things has to happen:

1. A susceptible individual gets infected with probability $\beta \frac{I_t}{N}$. Since there are S_t susceptible people at that moment,
$$\mathbb{P}\Big[(S_{t+\Delta t}, I_{t+\Delta t}) - (S_t, I_t) = (-1, 1)\Big] = \beta S_t \frac{I_t}{N} \Delta_t + o(\Delta t) \tag{2}$$
2. An infected individual recovers from the virus at any time with likelihood γ. Thus,
$$\mathbb{P}\Big[(S_{t+\Delta t}, I_{t+\Delta t}) - (S_t, I_t) = (0, -1)\Big] = \gamma I_t \Delta_t + o(\Delta t) \tag{3}$$
3. Nothing happens and the system remains the same.
$$\mathbb{P}\Big[(S_{t+\Delta t}, I_{t+\Delta t}) - (S_t, I_t) = (0, 0)\Big] = 1 - \left(\beta \frac{S_t}{N} + \gamma\right) I_t \Delta_t + o(\Delta t) \tag{4}$$

Many of the compartmental models that are available nowadays, to allow for an easy simulation and mathematical tractability, remain within the framework of the Markovian world. As such, those models have deterministic limits and diffusion approximations. The main results and techniques to obtain such limits are detailed in [14,15].

1.3. Motivation and features of the proposed model

In a more abstract form, the model (2)–(4) above can be written in terms of the difference between some Poisson processes and, at their core, many stochastic SIR-type models can be established as functions of Poisson processes or as limits of them, which leads to an immediate question of *whether the point process leading the epidemic model can be generalized to include more features observed empirically in the data*. Besides, there have been few models that try to explore the dynamics of the frequency at which individuals get infected and its effects on the likelihood of future people getting infected. That is, many of the current models either try to elaborate on the number of compartments or on the finesse of the conditions required for a certain individual to transition from one compartment to another but there are few models that try to get a more realistic description of *when* such transitions occur

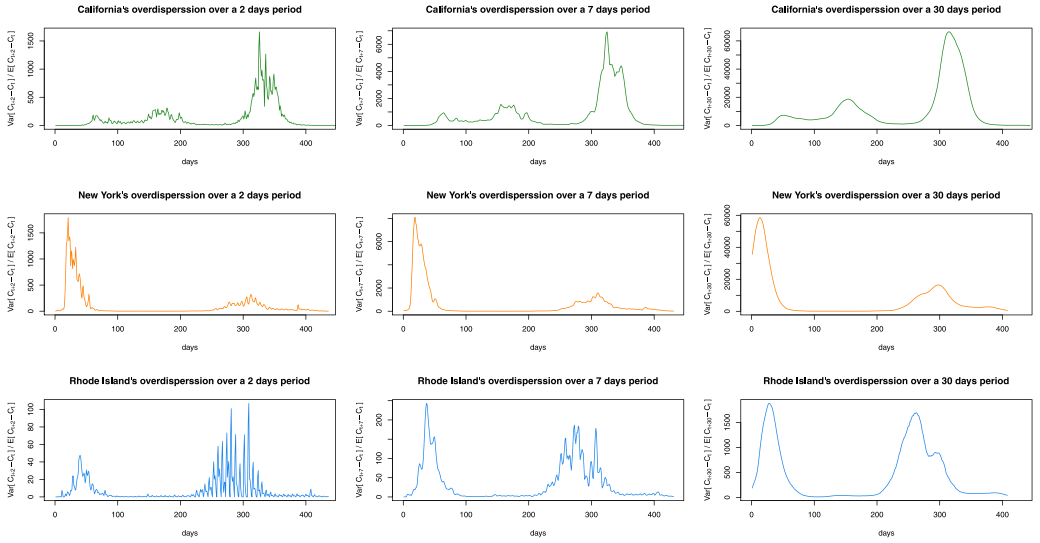


Fig. 3. Plots showing the over-dispersion in a 2,7 and 30-day window of the number of newly infected people during COVID-19 in 3 US states: New York, California and Rhode Island. The x axis shows the time in days, starting from February 2020, while the y axis shows the quotient between the variance and the average of the new cases within that period. A quotient greatly larger than 1, indicates high over-dispersion in the "arrival process" of new infected people.

and *how* the frequency of such transitions affects the future evolution of the system. Indeed, in most of the current compartmental models such as (2)–(4), each susceptible individual is equally likely to get infected and that probability is solely dependent on two factors: the proportion of infected individuals and the number of susceptible individuals at time t.

When the arrival processes of infectious individuals display *over-dispersion* (i.e. the variance of the number of arrivals in a given interval exceeds the corresponding expected value), the standard assumption of having a Poisson process driving the number of infections is not valid and new models are required. As Fig. 3 suggests, COVID-19 proved to be a very contagious virus with a superb level of over-dispersion. This implies that rather than having a stochastic epidemic model featuring Poisson processes, a Hawkes point process might be a much more suitable candidate to model it, which is the main reason for the model presented in this work.

Our first step will be the creation of a model that generalizes the Poissonian flow of infected individuals. In contrast to deterministic models, rather than modeling the number of infected people, I_t, we need to model the contagion process C_t counting the number of infected people up to time t. The reason for this is that I_t can increase (with a new infection) and decrease (with a recovery) but the counting process associated with a point process must be non-decreasing.

The proposed model has different salient features that distinguish it from many of the existing point processes-based epidemic models. The most important of these is the inclusion of a state-dependent intensity process for the number of future newly infected individuals, allowing us to model the likelihood of a new infection being dependent on the current state of the system via the number of susceptible and already recovered individuals. This is an important feature that has been incorporated in almost all of the deterministic models, where the rate of change of infected individuals depends on the number of susceptible and recovered individuals. However, it is not a feature in many of the point-process-based epidemic models. One of the main drawbacks of having a state-dependent intensity is that simulation of the system becomes computationally burdensome which is why in this paper we also characterize the moments of the random vector (I_t, R_t) as the solution of a system of differential equations. In this way, the average behavior of the system can be analyzed as well as its volatility. These important features distinguish this model with respect to previous stochastic epidemic models. In fact, as it can be seen from Fig. 8, one of the main features of this model is that you can get SIR-type curves for the process I_t simply by modulating correctly the parameters α and β in the state-dependent baseline intensity function (Eq. (A1) below) without the need of an external or exogenous factor, which is a desired characteristic.

Finally, it is important to mention that although there are many simulation algorithms available for Hawkes processes, there is no pseudo-code available for simulating state-dependent Hawkes processes as the one presented in this work (see Appendix B). In this work an adaptation to the algorithm presented in [16] is given to complement the work here. In this work, it was also verified that the solution to the system of differential equations satisfied by the moments of the number of infected and recovered people closely followed the simulated quantities.

How does this model relate to existing models in the literature? One of the first attempts of incorporating Hawkes processes into a SIR model is provided in [17]. In there, the authors explain a relationship between the general stochastic SIR model and a finite population Hawkes-SIR model (they account for the fact that the population is capped at N and therefore $C_t \leq N$), but this model has some limitations. In particular, they only explore a model where the *conditional expectation* of the intensity of C_t conditioned on *all*, past and future, times at which there is a recovery coincides with a finite Hawkes process with a baseline intensity $\lambda_0(t) = 0$. These restrictions cripple the model in the sense that a true Hawkes-SIR model would require the unconditional intensity of C_t to be

of the form (1) with $\lambda_0(t) \neq 0$ and also not be the result of a conditional expectation on past and, more importantly, future unknown times. However, the authors therein included a term that modulates the intensity to be proportional to the proportion of susceptible individuals is a good way to mimic the classic SEIR model and get a glimpse of the long-run behavior of the process.

Another model that has tried to make use of point processes to model the number of individuals getting infected is [18], where the author considers a discrete time Hawkes process with also an infinite amount of initial susceptible individuals and compartmentalizes individuals. However, the intensity process has constant coefficients and there is no clear characterization of the joint behavior of the moments for I_t and R_t, becoming in a broad sense a particular case of the model presented here. Another important model is found in [19], where the authors formulate a time-modulated point process with the stochastic intensity of the process (found in Eq. (5) below) has a purely time-dependent modulating function $\mu(t)$ multiplying the stochastic integral therein. This function will serve as a way to control the future likelihood of new infections depending on the time t. However, contrary to our model, that modulation is exogenous to the system in the sense that μ does not depend on the current state of the system but rather on the "global environment". Finally, it is worth mentioning the paper by [20] as it is one of the most complete stochastic models for epidemics using point processes. In there, the authors model the intensity function as the sum of a baseline background rate and incorporate a (exponential) function that relates to the mobility between different regions or counties and the demographic features. However, as they report on the reported infections rather than at the moment of the contagion process, they introduce a delay in the equation to account for this.

Nonetheless, there is missing a model that becomes a natural extension of a stochastic SIR model. That is, there is no model in the literature where people are divided into the S, I, R compartments and people start arriving at the "infected" compartment according to a Hawkes process. Recovered individuals should arrive via a Poisson or renewal process, since contrary to infections, the recovery of a person does not affect how other people recover. One of the aims of this proposal is to create and analyze such a model as well as find conditions under which a disease introduced into a community will develop into a large outbreak, and if it does, conditions under which the disease may become endemic. This condition is linked to the so-called basic reproductive number, R_0, defined as the expected number of secondary infective cases per primary cases in a susceptible population.

As it can be noted, many of the models that have started to incorporate Hawkes processes are very recent and many were inspired by the highly infectious rate of COVID-19, making this a novel area to propose, analyze and create models that can shed some light into the question of how fast these type of diseases propagates, hopefully contributing to the decision process on policy-making by providing more accurate estimates when incorporating some of the salient feature of these types of models.

Scope and Limitations of the current model: In its current form, the model presented in this article provides a way to model the overdispersion and high variability observed in the data. However, due to the limiting ability to incorporate a stronger dependency on the current state of the epidemic and the different rates at which different variants have been shown to be transmitted, this model is more suitable for keeping track of the pandemic at the beginning stage when not so many social and political reactions have occurred. To account for this, the author is currently investigating a regime-switching alternative. Further, as it is, these types of models should only be used to forecast a short to medium-term prediction.

Organization of this article: In Section 2, a brief description of the model is presented while in Section 3 we compute the moments of the process which can be used to predict the medium-term dynamics of the current number of infected people given the probabilistic distribution of the quarantined people (which in terms can be used for public policy). Section 4 provides a brief survey and ideas of how to perform efficient estimation on a state-dependent self-exciting process. However, since estimation is out of the scope of this paper, it will be treated in subsequent research. In Section 5, we provide different numerical examples to explore how the different parameters of the model play a role in it and their sensitivity. The conclusions and elements of further research are presented in Section 6. Finally, two appendices are also provided. In Appendix A all the proofs to the technical lemmas and theorems of Section 3 are provided while in Appendix B the pseudo-code of the different algorithms used to simulate the different processes are provided.

2. A brief description of the model

In the model presented in this article, an important assumption will be that the population is rather large (technically, it is assumed that the possible number of susceptible people is infinite to facilitate the analysis of the moments of the number of infected people I_t). This assumption provides a tractable background for the problem and is not extremely unrealistic given that the model is specifically designed to simulate a pandemic in its early stages. Considering that up to May 1st, 2022, only 6.6% of the population has been reportedly infected[1] (although it is of general consensus that this number is very underestimated, especially now with the development of at-home test kits), we can think of the number of susceptible individuals as a rather large number and the model will still be accurate. A precise finite model can be created, but there is little to gain and the closed formulas obtained in Section 3 become very cumbersome. This compartmental model follows a similar SIR dynamics to the models in [13], where S_t, R_t and I_t represent the number of susceptible, recovered and infected people at time t. Since, as discussed above, the number of susceptible people is assumed to be infinite, we cannot have a classic relationship such as $S_t + I_t + R_t = N$, but we will rather only focus on the number of infected individuals I_t and the number of recovered individuals R_t.

[1] Information taken from https://covid19.who.int/.

The main modeling assumption in this paper will be that the (historical) number of infected people up to time t, denoted by C_t, will be driven by a (Hawkes-like) counting process with stochastic intensity

$$\Lambda_t = \lambda^{\infty}(I_t, R_t) + \int_0^t Q_s \phi(s,t) dC_s, \tag{5}$$

where $\{Q_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables independent among themselves and from every other process. As described below, these random variables Q will be interpreted as the level of quarantine that the $n-$th infected person will have, modulating the probability that this individual will produce future infections.

In this paper, we will assume that the baseline intensity $\lambda^{\infty}(i,\rho)$ of the Hawkes-like process C_t given in Eq. (5) reacts to the state of the epidemics at time t. Indeed, we will assume that if $I_t = i$ and $R_t = \rho$. Then,

$$\lambda^{\infty}(i,\rho) = \lambda_0 + i\log(\alpha) + \rho\log(\beta), \tag{A1}$$

where $\alpha > 1$ and β, λ_0 are positive numbers.

The form of the baseline intensity given in assumption (A1) provides great flexibility to consider multiple situations. We describe some of them next.

- *A pure Hawkes approach.* In this setting, the intensity takes the form of a classical Hawkes process. This could be a basic model to use for explaining the over-dispersion observed in the data. This can be achieved by setting $\alpha = \beta = 1$. In this sense, the baseline intensity will be independent of the number of infected and recovered and the general intensity of the point process will only depend on the cumulative number of infected people and times of infection of these. In this case, the intensity becomes

  $$\Lambda_t = \lambda_0 + \int_0^t Q_s \phi(s,t) dC_s. \tag{6}$$

  This case is considered in [21] where the author analyzes, simulates and fits a classical Hawkes process for the total number of infected people in France but to signify a change in public policy modifies the intensity to mimic a lock-down as in China. As another example, in [18] where the author discretizes the intensity to analyze and fit the epidemic model to the gathered data of several countries, including Mexico. This model is very popular because there are well-established algorithms to simulate and calibrate the model to existing data.
- *A proportional-to-infections approach.* In this case, the baseline intensity will be proportional to the current number of infected people. Here, not only can over-dispersion be captured but the intensity, and thus the likelihood, of observing a new infection will increase or decrease according to the current number of infected people. This can be achieved by setting $\lambda_0 = 0$ and $\beta = 1$ and the intensity becomes

  $$\Lambda_t = \lambda_0 + i\log(\alpha) + \int_0^t Q_s \phi(s,t) dC_s, \tag{7}$$

- *A state dependent Hawkes baseline intensity*. This is the more general case and the one that will be considered in this paper. In this case, assumption (A1) can be used to focus on different aspects of the model depending on whether $0 < \beta < 1$, $\beta = 1$ or if $\beta > 1$. Indeed, if $0 < \beta < 1$, then the quantity $\rho\log(\beta)$ is negative and in this case, we can think of a situation where the recovered individuals suppress the epidemics. This is more in line with the classical assumptions of an SIR model where by increasing the number of recovered people the infectivity of the virus decays and new cases are less likely. However, this behavior hinges on a critical assumption: recovered people cannot contribute more to the epidemics, and they become isolated from it. This assumption is not completely true for the COVID-19 epidemics where many recovered people become reinfected and can contribute to the spread of the disease. One should be careful here though, if the number of recovered individuals becomes very large and $\beta < 1$, the baseline intensity $\lambda^{\infty}(i,\rho)$ might become negative and this is not permitted. Several solutions could be considered at this point, the two more relevant being to stop the process before $\lambda^{\infty}(i,\rho) < 0$ or change the form Assumption (A1) to $\lambda^{\infty}(i,\rho) = \left(\lambda_0 + i\log(\alpha) + \rho\log(\beta)\right)_+$, where $f(x) = (x)_+$ represents the positive part of x, being this latter option a very interesting case left out for further research. If $\beta = 1$, then the number of recovered people does not affect the model and the process R_t becomes a homogeneous Poisson process independent of C_t. Finally, whenever $\beta > 1$ the recoveries can also contribute to the infection of new individuals (rather than suppressing it). As discussed above, in the current pandemic many individuals can become reinfected and keep transmitting the virus to other people. This feature might come in handy specifically in the current situation with COVID-19 as universities and other workplaces treat recovered individuals as non-infectious but indeed they might become infectious again but at a different rate than that of susceptible or currently infected individuals.

It is also important to determine the parameters of the model:

- The baseline infective rate λ_0, which is associated with the likelihood that someone new gets infected due to exogenous factors to the model such as migrations, population dynamics, etc. This is not affected by the amount of infected or recovered people.
- The infective rate $\alpha > 1$. This parameter (or more precisely, the logarithm of it) measures how the likelihood at time t of a new infection will increase due to the number of infected people at that particular instant regardless of the total number of people that have been infected by the virus. The higher α, the more likely a new infection will occur at the level of current infections I_t.

- The situational rate $\beta > 1$. This parameter (or more precisely, the logarithm of it) measures how the likelihood at time t of a new infection will increase or decrease due to the number of recovered people *up to* that particular instant. As $\beta \to 0$ the likelihood that a new infection will occur at the level of current of R_t decreases while if $\beta \to \infty$ the likelihood will increase.
- The mean recovery time μ. This parameter is the expected recovery time period from the infection. There are several studies about this quantity and it depends on several factors, but according to [22] a good approximation could be around 9 days.
- The historical influence parameter r (also called the reversion coefficient r). This parameter will measure the influence of previous (historical) infections in the arrival of a new infection and indicates the speed at which the likelihood of a new infection decays to the current level of the baseline intensity in the absence of a new infection. In fact, for our model, the increase in the likelihood of a new infection due to a previous infection at time $s < t$ is proportional to $e^{-r(t-s)}$. This means that the more recent an infection, the more it will contribute to the likelihood of a further infection. This behavior is the key difference concerning the classical epidemiological models and what allows for the clustering and *over-dispersion* observed in the data for the current pandemic.
- The probability distribution of the "quarantine effect" Q. This random variable specifies the level of "quarantine" each individual will have. Indeed, our model specifies some stochastic dynamics where the probability of a new infection at time t is driven by the number of people that have been infected up to that time and how recent their infections have been. This quarantine factor will determine the proportion at which an infected person will contribute to a new infection. From the modeling perspective, this random variable can be thought of as a measure of quarantining. The lower the (random variable) Q, the lower the contribution of such infected individuals to a new infection. This random variable can be discrete or continuous, but its support has to be over the positive numbers. In other words, an infected individual must have a positive contribution to the general likelihood of generating a new infection even if it is small That is, there cannot be a "perfect quarantine", which is consistent because people have to go to the groceries or buy basic services and even interact with delivery services by receiving goods at home. For this work, we will impose the mild restriction that the Moment Generating Function (MGF) of Q exists on a neighborhood of 0.

Remark 1. It is also known (see Chapter 3.3 in [23]) than when the random variable Q is constant, and the intensity of the point process is the classical Hawkes process in Eq. (1) with $\phi(t) = Qe^{-rt}$, then the *branching ratio* of the process would be given by

$$n = \frac{Q}{r}$$

where as above, r is the reversion coefficient. In the SIR process where the number total (historical) number of people that have been infected by the virus is driven by a Hawkes process the branching ratio n has the interpretation that when $0 < n < 1$, it becomes the ratio of the number of people that one individual will infect relative to the entire population; that is, it is related to the epidemiological basic reproduction number $\mathcal{R}_0$.

Remark 2. Assumption (A1) specifies the form of the so-called baseline intensity. This quantity will not depend on the past number of infections but solely on the present number of infected and recovered individuals. In fact, this baseline intensity remains constant between events, that is, between new infections or recoveries, the probability of a new infection is exponentially distributed with rate $\lambda^{\infty}(i, \rho)$ and thus the process dictating the arrival of a new infection is equal in distribution to a homogeneous Poisson process with rate $\lambda^{\infty}(i, \rho)$.

Also, notice that α and β represent the factors that drive the baseline intensity according to the state of the system. The more infected people there are at the moment, the more likely a new infection will occur, and the more recovered individuals there are in the present moment, the less likely a new infection will occur.

Further, since we do not want a "degenerate" Hawkes process, we will assume that Q does not have an atom at 0. That is,

$$\mathbb{P}[Q \leq 0] = 0 \tag{A2}$$

3. Derivation of the moments of the number of infected individuals I_t

As mentioned in the previous section, one of the features that the epidemic model presented here has is that every person is assumed to have a random level of quarantine, whose law is given by Q. As an easy example, assume that $supp(Q) = \{0.25, 0.75, 1\}$. In this case, $Q = 0.25$ will imply a higher level of quarantine (contributing less to new infections) while $Q = 1$ would mean a low level of quarantine (so that this person will increase the likelihood of a new infection happening). Of course, $Q = 0.5$ would be an intermediate case.

To compare how these different levels of quarantine and other parameters affect the model, it might be worth looking at the average behavior of I_t and R_t. This is because depending on $\mathrm{Var}[I_t]$, some comparisons might not be depicted accurately by plotting some trajectories of the process.

Recall from Eq. (5) that the total number of infected people up to time t, denoted by C_t, is determined by a counting process with stochastic intensity given by

$$\Lambda_t = \lambda^{\infty}(I_t, R_t) + \int_0^t Q_s \phi(s, t) dC_s, \tag{8}$$

where $\{Q_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables independent among themselves and from every other process denoting the level of quarantine that the $n-$th infected person will have, modulating the probability that this individual will produce future infections.

We are interested in computing the generating function of the triplet $(I_t, R_t, \lambda(t))$

$$\mathbb{E}\left[z^{I_t} w^{R_t} e^{-s\lambda_t}\right]$$

Here, we will assume that all the stochastic processes are Markovian, which can be attained if

- The recovery time is exponentially distributed with rate parameter μ. That is, if τ is the recovery time for an infected individual, then

$$\mathbb{P}[\tau \leq t] = 1 - e^{-\mu t}. \tag{A3}$$

  This implies in particular that if $I_t = k$, then

$$\mathbb{P}[R_{t+\Delta t} - R_t = 1] = \mathbb{P}[\min\{\tau_1, \dots, \tau_k\} \leq \Delta t] = 1 - e^{k\mu\Delta t} = k\mu\Delta t + o(\Delta t) \tag{9}$$

- The self-exciting kernel $\phi(\bullet)$ is an exponential function. That is,

$$\phi(s,t) = e^{-r(t-s)} \tag{A4}$$

These assumptions can be relaxed, but the analytical tractability will be lost and different techniques would have to be employed. The analysis presented here is inspired by the one presented in [24] but here is generalized to allow a state dependent baseline intensity.

In order to compute the joint distribution of $(I_t, R_t, \Lambda(t))$, we need to solve a system of differential equations presented next.

Theorem 3. *Let $F(t,i,\rho,\lambda) = \mathbb{P}[I_t = i, R_t = \rho, \Lambda_t \leq \lambda]$ and $f(t,i,\rho,\lambda) = \dfrac{\partial F}{\partial \lambda}(t,i,\rho,\lambda)$. Then, under assumptions (A3)–(A4), $f(t,i,\rho,\lambda)$ satisfy the Partial difference differential equation (PDDE):*

$$\begin{aligned}\frac{\partial}{\partial t} f(t,i,\rho,\lambda) - \frac{\partial}{\partial \lambda}\left(r\lambda f(t,i,\rho,\lambda)\right) + r\lambda^{\infty}(i,\rho)\frac{\partial}{\partial \lambda} f(t,i,\rho,\lambda) = \\ \int_0^{\lambda - \Delta\lambda_{i,\rho}^{\infty,1}} y f(t,i-1,\rho,y)\frac{\partial}{\partial \lambda}\mathbb{P}[Q \leq \lambda - \Delta\lambda_{i,\rho}^{\infty,1} - y]dy + \\ (i+1)\mu f(t,i+1,\rho-1,\lambda - \Delta\lambda_{i,\rho}^{\infty,2}) - (i\mu + \lambda) f(t,i,\rho,\lambda)\end{aligned} \tag{10}$$

where

$$\Delta\lambda_{i,\rho}^{\infty,1} = \lambda^{\infty}(i,\rho) - \lambda^{\infty}(i-1,\rho) \qquad \text{and} \qquad \Delta\lambda_{i,\rho}^{\infty,2} = \lambda^{\infty}(i,\rho) - \lambda^{\infty}(i+1,\rho-1)$$

The next objective is to compute the generating function of the triple $(I_t, R_t, \Lambda(t))$, which can then be used to compute the moments and other quantities of interest. However, since I_t and R_t are discrete and $\Lambda(t)$ is continuous, we will have to compute the transformations of those random variables separately.

Theorem 4. *Let $\varphi(t,z,w,s) = \mathbb{E}\left[z^{I_t} w^{R_t} e^{-s\Lambda(t)}\right]$ for $|z| \leq 1$ and $|w| \leq 1$. Then, $\varphi := \varphi(t,z,w,s)$ satisfies the PDE*

$$\begin{aligned}\frac{\partial}{\partial t}\varphi + \left[rs - 1 + \left(\frac{1}{\alpha}\right)^s M_Q(s) z\right]\frac{\partial}{\partial s}\varphi + \left[(\mu + rs\log(\alpha))z - \mu\left(\frac{\alpha}{\beta}\right)^s w\right]\frac{\partial}{\partial z}\varphi \\ + \left[\log(\beta) rsw\right]\frac{\partial}{\partial w}\varphi = -\lambda_0 rs\varphi\end{aligned} \tag{11}$$

with initial condition

$$\varphi(0,z,w,s) = z^{i_0} e^{-s(\lambda_0 + i_0\log(\alpha))}$$

Since the resulting PDE is linear of first order, we can apply the method of characteristics to simplify the problem to a system of 2 by 2 ODEs. However, the solution can be quite messy and numerical methods are very likely to be needed to solve such system which we just point out for the sake of completeness.

Corollary 5. *Let (z,w,s) be a fixed point in $[0,\infty)^3$ and let $\Xi(t) := \varphi(t,z,w,s) = \mathbb{E}\left[z^{I_t} w^{R_t} e^{-s\Lambda(t)}\right]$. Then, $\Xi(t)$ is the solution of the system of ODEs*

$$\begin{cases} \dfrac{d\vartheta}{dt} &= -1 + r\vartheta(t) + \left(\dfrac{1}{\alpha}\right)^{\vartheta(t)} M_Q(\vartheta(t))\left[z\exp\left(\mu t + r\log(\alpha)\displaystyle\int_0^t \vartheta(u)du\right)\right. \\ & \qquad \left. -w\mu\displaystyle\int_0^t \left(\frac{\alpha}{\beta}\right)^{\vartheta(u)} \exp\left(-\mu u + r\log(\beta)\int_0^u \vartheta(y)dy\right)du\right] \\ \dfrac{d\Xi}{dt} &= -\lambda_0 r\vartheta(t) \\ \vartheta(0) &= s \\ \Xi(0) &= z^{i_0}\exp\left(-\vartheta(\lambda_0 + i_0\log(\alpha))\right) \end{cases} \tag{12}$$

Our objective is to be able to compute the moments of the random variables I_t and R_t, however the system of ODEs (12) might not prove that useful for this task, therefore we turn our attention to the PDE (11) again. To obtain the joint $(i,j,k)-$th moment of I_t, R_t and $\Lambda(t)$ we take the $i-$th derivative of the PDE (11) with respect to z, the $j-$th derivative with respect to w and the $k-$th derivative with respect to s plug in the values $z = w = 1$ and $s = 0$. However, for simplicity, we will proceed to give an explicit formula for the first and second moments of the processes $I_t, R_t, \Lambda(t)$ since these are the ones that might be used the most.

Unfortunately, the moments of I_t and R_t are not independent of those of $\Lambda(t)$, therefore to compute the first moment, we will need to solve a system of 3 linear differential equations. Each equation will be obtained by differentiating the PDE (11) with respect to one parameter z, w or s and plugging the values mentioned before. We show the procedure in the next two lemmas.

Lemma 6. *For any time $t > 0$, let $\overline{\mathbf{X}}_t = \left[\mathbb{E}[\Lambda(t)], \mathbb{E}[I_t], \mathbb{E}[R_t]\right]^T$. Then, $\overline{\mathbf{X}}_t$ is the solution to the system of differential equations*

$$\frac{d}{dt}\overline{\mathbf{X}}_t = \mathbf{A}_1\overline{\mathbf{X}}_t + \mathbf{b}_1 \tag{13}$$

$$\overline{\mathbf{X}}_0 = \begin{bmatrix} \lambda_0 + i_0\log(\alpha) \\ i_0 \\ 0 \end{bmatrix} \tag{14}$$

where

$$\mathbf{A}_1 = \begin{bmatrix} \log(\alpha) + \mathbb{E}[Q] - r & (r-\mu)\log(\alpha) + \mu\log(\beta) & r\log(\beta) \\ 1 & -\mu & 0 \\ 0 & \mu & 0 \end{bmatrix} \text{ and } \mathbf{b}_1 = \begin{bmatrix} \lambda_0 r \\ 0 \\ 0 \end{bmatrix}$$

That is,

$$\overline{\mathbf{X}}_t = \overline{\mathbf{X}}_0 e^{\mathbf{A}_1 t} + \int_0^t e^{\mathbf{A}_1(t-s)}\mathbf{b}_1\, ds \tag{15}$$

We can use the same approach as in Lemma 6 to compute the second moments. Besides, we will need to consider all 6 possible double products of the random variables I_t, R_t and $\Lambda(t)$.

An important remark is that without too much work we can transition from the models presented in Section 2 by just modifying the parameters of the original model. For example, if a pure-Hawkes (pH) model with intensity given by Eq. (6) wants to be considered; then by setting $\alpha = \beta = 1$ in the previous result it is obtained that the following:

$$\frac{d}{dt}\overline{\mathbf{X}}_t^{pH} = \mathbf{A}_1^{pH}\overline{\mathbf{X}}_t + \mathbf{b}_1^{pH} \tag{16}$$

$$\overline{\mathbf{X}}_0^{pH} = \begin{bmatrix} \lambda_0 \\ i_0 \\ 0 \end{bmatrix} \tag{17}$$

where

$$\mathbf{A}_1^{pH} = \begin{bmatrix} \mathbb{E}[Q] - r & 0 & 0 \\ 1 & -\mu & 0 \\ 0 & \mu & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_1 = \begin{bmatrix} \lambda_0 r \\ 0 \\ 0 \end{bmatrix}$$

which agrees with the result provided in [24]. Further, if a model whose infection rate increases or decreases proportional to the number of infected people in its baseline intensity (pI) as given by Eq. (7) is sought, then it should be set $\beta = 1$ and the resulting system of ODEs is

$$\frac{d}{dt}\overline{\mathbf{X}}_t^{pI} = \mathbf{A}_1\overline{\mathbf{X}}_t^{pI} + \mathbf{b}_1^{pI} \tag{18}$$

$$\overline{\mathbf{X}}_0^{pI} = \begin{bmatrix} \lambda_0 + i_0\log(\alpha) \\ i_0 \\ 0 \end{bmatrix} \tag{19}$$

where

$$\mathbf{A}_1^{pI} = \begin{bmatrix} \log(\alpha) + \mathbb{E}[Q] - r & (r-\mu)\log(\alpha) & 0 \\ 1 & -\mu & 0 \\ 0 & \mu & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_1 = \begin{bmatrix} \lambda_0 r \\ 0 \\ 0 \end{bmatrix}$$

Remark 7. Notice that in the two particular cases above given by systems of ODEs in Eqs. (16)–(19) since $\beta = 1$, then the intensity becomes independent of R_t and thus the entire system becomes independent of R_t which is reflected by the third column of the matrices $\mathbf{A}_1^{pH}$ and $\mathbf{A}_1^{pI}$ being 0.

Finally, a third important differentiation from the base case $\alpha, \beta > 1$ is when the number of recovered individuals actually *decreases* the likelihood of a new infection, such as in a classical SIR model and which is also described in Section 2. In this case, and as mentioned earlier, provided that the process is stopped whenever the intensity becomes negative, the corresponding system

of differential equations for the first moments can be obtained from Lemma 6 by setting $\beta = 1/\tilde{\beta}$ with $\tilde{\beta} > 1$ so that, in this case, $\beta < 1$ and the baseline intensity given by Assumption (A1) becomes

$$\lambda^{\infty}(i,\rho) = \lambda_0 + i\log(\alpha) + \rho\log(\beta) = \lambda_0 + i\log(\alpha) - \rho\log(\tilde{\beta}). \tag{20}$$

In this case, the corresponding system of differential equations governing the first moments of this system is given by

$$\frac{d}{dt}\overline{\mathbf{X}}_t^{SIR} = \mathbf{A}_1\overline{\mathbf{X}}_t^{SIR} + \mathbf{b}_1^{SIR} \tag{21}$$

$$\overline{\mathbf{X}}_0^{SIR} = \begin{bmatrix} \lambda_0 + i_0\log(\alpha) \\ i_0 \\ 0 \end{bmatrix} \tag{22}$$

where

$$\mathbf{A}_1^{SIR} = \begin{bmatrix} \log(\alpha) + \mathbb{E}[Q] - r & (r-\mu)\log(\alpha) - \mu\log(\tilde{\beta}) & -r\log(\tilde{\beta}) \\ 1 & -\mu & 0 \\ 0 & \mu & 0 \end{bmatrix}, \mathbf{b}_1 = \begin{bmatrix} \lambda_0 r \\ 0 \\ 0 \end{bmatrix}$$

Next, a characterization of the second moments for the general state-dependent intensity is provided. The techniques and methods to obtain them follow from the ones used in Lemma 6 but to solve this system it is necessary to obtain the solution of the system (13)–(14), since its solution its dependent (as expected) on the first moments of $(\Lambda(t), I_t, R_t)$. For the following Lemma, it will be useful to simplify the notation so that the matrix used to compute second moments is displayed nicely. Indeed, let

$$\begin{aligned} C_1 &:= \log(\alpha) + \mathbb{E}[Q] - r \\ C_2 &:= (r-\mu)\log(\alpha) + \mu\log(\beta) \\ C_3 &:= r\log(\beta) \\ C_4 &:= r\log(\alpha) \end{aligned}$$

Lemma 8. *For any $t > 0$, let $\overline{\mathbf{Y}}_t = \left[\mathbb{E}[\Lambda^2(t)], \mathbb{E}[I_t(I_t-1)], \mathbb{E}[R_t(R_t-1)], \mathbb{E}[\Lambda(t)I_t], \mathbb{E}[\Lambda(t)R_t], \mathbb{E}[I_tR_t]\right]^T$. Then, $\overline{\mathbf{Y}}_t$ is the solution to the system of equations*

$$\frac{d}{dt}\overline{\mathbf{Y}}_t = \mathbf{A}_2\overline{\mathbf{Y}}_t + \mathbf{b}_2 \tag{23}$$

$$\overline{\mathbf{Y}}_0 = \left[(\lambda_0 + i_0\log(\alpha))^2, i_0^2, 0, i_0(\lambda_0 + i_0\log(\alpha)), 0, 0\right]^T \tag{24}$$

where,

$$\mathbf{A}_2 = \begin{bmatrix} 2C_1 & 0 & 0 & 2C_2 & 2C_3 & 0 \\ 0 & -2\mu & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\mu \\ 1 & C_2 & 0 & C_1 & 0 & C_3 \\ 0 & 0 & C_3 & -\mu & C_1 & C_2 \\ 0 & \mu & 0 & 0 & 1 & -\mu \end{bmatrix}$$

and

$$\mathbf{b}_2 = \begin{bmatrix} \left[\log^2(\alpha) + 2\log(\alpha) + \mathbb{E}[Q^2] + 2\lambda_0 r\right]\mathbb{E}[\Lambda(t)] + \mu\log^2\left(\frac{\alpha}{\beta}\right)\mathbb{E}[I_t] \\ 0 \\ 0 \\ (C_1 + r)\mathbb{E}[\Lambda(t)] + r\lambda_0\mathbb{E}[I_t] + C_4\mathbb{E}[R_t] \\ (r\lambda_0 + C_3)\mathbb{E}[R_t] + (C_2 - C_4)\mathbb{E}[I_t] \\ 0 \end{bmatrix}$$

That is,

$$\overline{\mathbf{Y}}_t = \overline{\mathbf{Y}}_0 e^{\mathbf{A}_2 t} + \int_0^t e^{\mathbf{A}_2(t-s)}\mathbf{b}_2 ds \tag{25}$$

4. Some considerations regarding the estimation of the state dependent Hawkes process

As with many point processes, there are mainly three types of estimation procedures: MLE, (generalized)MoM and LSE-type methods. Below, a discussion on each of these three processes, their advantages and disadvantages is presented. It is important to emphasize that there is not a definite answer to which method provides an advantage when doing inference over the parameters of the model and this intricate topic is left out for future research.

- **Maximum Likelihood Estimation (MLE).** This method is one of the most used and regarded within the academic literature. This is due in part because many theoretical results guarantee that the MLE methods will yield an optimal solution.

Unfortunately, for many Markovian point processes the evaluation of the log-likelihood function is of order $O(N_t^2)$. Several algorithms such as EM are used to improve this (see [25]), but in general, the biggest problem of MLE is that the likelihood curve is very flat, with many local maxima and without a clear way to decide which is the global maximum. An important step in improving this is discussed in [25,26]. Many of these methods apply to the exponential kernel form used in this paper (see Assumption (A4)), but also have a constant baseline intensity. It is still an ongoing research topic for this author to generalize such methods to a state-dependent intensity.

- **(Generalized) Method of Moments (gMoM)**. These methods are derived under the assumption that the process is in its limiting stationary state, and as with any Method of Moments, the idea is to form a system of equations of $n \times n$, where n is the number of parameters to be determined and n linear independent equations relating the moments are used. However, for Hawkes processes, usually, there are not enough linear independent conditions on the moments and thus the Autocovariance function is introduced to generate other equations. A prime example is given in the work by [27]. Further, these methods are inefficient when the number of observations is "small" and they usually do not work properly on higher dimensions. However, it is important to mention that in our case, these methods might not be the best since we are trying to model the early stages of the epidemic where the processes are far away from the stationary state.
- **Least Square Estimation (LSE) type methods**. These methods have not been explored until recently, mainly because the order of these methods is similar to the ones of MLE. However, the recent work by [28] shows that, unlike MLE methods, LSE methods can possess certain algebraic properties that help with the stochastic approximation of the kernels to then maximize or minimize the LSE functional.

The nature of the process we are dealing with requires special care in the estimation of parameters and as such, this paper will not try to just follow the MLE method presented in [10, Section 7.2], but rather this sensitive topic is left as an object of further research. Especially if Assumption (A4) is relaxed and we allow for a more general kernel and also a state-dependent self-exciting kernel.

5. Empirical and numerical examples

The objective of this section is to provide numerical evidence of the behavior that the model has under the different scenarios proposed as well as the implications of increasing or decreasing the level of quarantine provided by the random variable Q and other changes in the parameters.

For all the experiments there will be four levels of Quarantine: high (Q_H), medium-high (Q_{M_H}), medium-low (Q_{M_L}) and low (Q_L). To completely specify the levels of quarantine provided by these random variables, their distribution is presented next.

$$Q_H = \begin{cases} 0.05 & \text{w.p. } 0.92 \\ 0.45 & \text{w.p. } 0.03 \\ 0.60 & \text{w.p. } 0.02 \\ 0.95 & \text{w.p. } 0.03 \end{cases}, \qquad Q_{M_H} = \begin{cases} 0.05 & \text{w.p. } 0.11 \\ 0.45 & \text{w.p. } 0.18 \\ 0.60 & \text{w.p. } 0.46 \\ 0.95 & \text{w.p. } 0.25 \end{cases},$$

$$Q_{M_L} = \begin{cases} 0.05 & \text{w.p. } 0.12 \\ 0.45 & \text{w.p. } 0.46 \\ 0.60 & \text{w.p. } 0.32 \\ 0.95 & \text{w.p. } 0.10 \end{cases}, \qquad Q_L = \begin{cases} 0.05 & \text{w.p. } 0.02 \\ 0.45 & \text{w.p. } 0.02 \\ 0.60 & \text{w.p. } 0.02 \\ 0.95 & \text{w.p. } 0.91 \end{cases}$$

As can be seen, the random variables take the same values in all cases but their probabilities change. They are specified so that $\mathbb{E}[Q_L] = 0.9$, $\mathbb{E}[Q_{M_L}] = 0.6$, $\mathbb{E}[Q_{M_H}] = 0.5$ and $\mathbb{E}[Q_H] = 0.1$.

Next, we provide several figures with various simulations of the different cases mentioned in Section 3. In particular, it is important to notice how the difference in the parameters affects the speed at which the number of infected individuals grows. Also, it is important to remark that although the simulation algorithm 2 is a variation of the thinning algorithm by Ogata described in Appendix B, to the author's knowledge, an explicit algorithm to simulate a Hawkes process with a state-dependent intensity is not readily available. Thus, as part of this work, a detailed pseudo-code for simulating this kind of processes is provided in Appendix B. Furthermore, due to the immense amount of simulations and computations required to simulate these processes –in the end, the thinning algorithm is a variant of an acceptance–rejection method and as such, many simulations are rejected– the process cannot be easily simulated for large time intervals.

In Figs. 4–7 below, the number of infected people is simulated, under different parameters, for the small time interval $[0, 4]$ together with the corresponding solution of the differential equation under different scenarios, where either α, β or the Quarantine distribution changes. This exercise has two purposes: to verify that the simulation algorithm and the differential equations yield similar results (performing a cross-verification) and to show how the change in different parameters yields logical conclusions as well as exploring how such changes affect the speed at which the number of infected people grows.

To understand the behavior of the process counting the number of infected people under a general purely Hawkes process (i.e. we set $\alpha = \beta = 1$), in Fig. 4, I_t is plotted under 12 different scenarios and each scenario under the 4 different Quarantine scenarios. In all the scenarios, the parameters were set to be the same except for the decaying parameter r (see Eq. (A4)) and the constant baseline intensity λ_0. As expected, the number of infected people decreases as r gets larger, meaning that an infectious person will contribute to a new infection significantly only during a short time span.

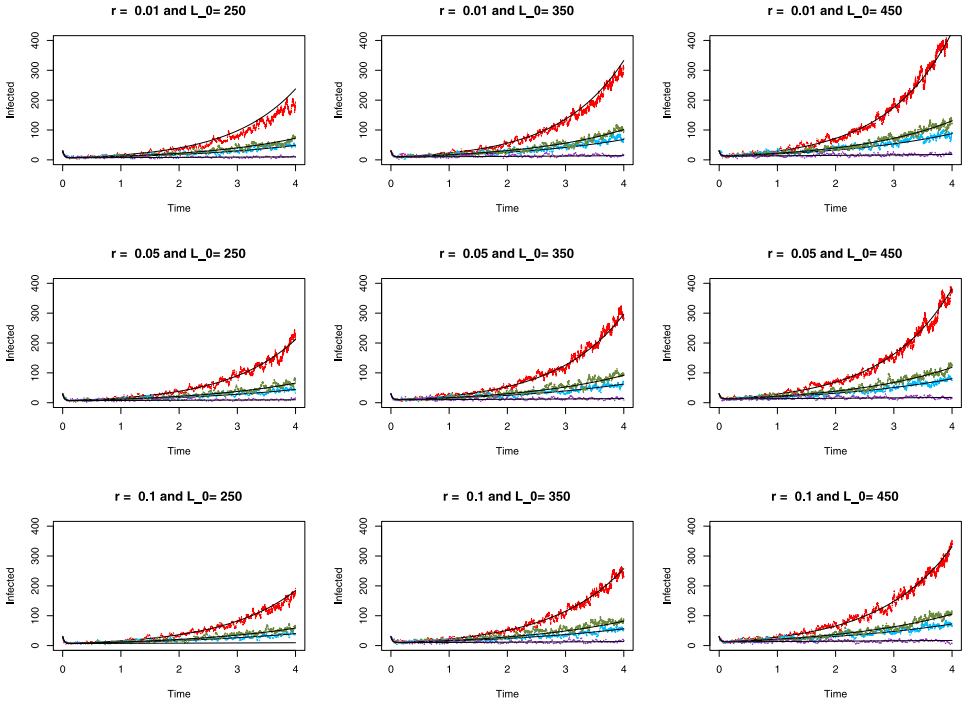


Fig. 4. Figure showing 12 simulations and their behavior for different parameters under a purely Hawkes approach. Plots in the same row share the same value of r while plots in the same column share the same value of λ_0. The rest of the parameters, as well as the scale, were kept the same for comparison purposes. The Quarantine level is also represented, with High Quarantine (Q_H) in purple, Medium-High Quarantine (Q_{M_H}) in blue, Medium-Low Quarantine (Q_{M_L}) in green and Low Quarantine (Q_L) in red. Furthermore, the mean (deterministic) behavior of the system is plotted with the same color as the simulated paths.

Two interesting cases arrive in the general state-dependent Hawkes baseline intensity case. First, as can be seen in Fig. 5, when $\alpha > \beta$ (in this case $\alpha = 1.4$ while $\beta = 1.2$), we see that there are more infections as compared to the purely Hawkes case reflected in Fig. 4. However, when $\beta > \alpha > 1$, we see that the infections also grow as compared within the baseline case of the purely Hawkes model but they are actually even higher than in the previous case where $\alpha > \beta$. The reason for this is that the number of infected people becomes, as time progresses, comparatively smaller than the number of people that have recovered and continue propagating the virus. This is a classical behavior of viruses that do not create immunity as is the case of COVID-19. Finally, in Fig. 7 we plot the case where $\alpha > 1$ but $\beta < 1$. This is to illustrate how a β less than 1 can mitigate greatly the size of the epidemics. Indeed, as a good comparison, note that, with all the parameters save α and β kept constant, the maximum range up to time 4 for the number of infected individuals in the Pure Hawkes case is roughly 400, while on the state-dependent Hawkes cases with $\alpha > \beta > 1$ is of 600; when $\beta > \alpha > 1$ is of 1200 while on the case where $\beta < 1 < \alpha$ it is of 200.

Finally, in Fig. 8, the long-run behavior of the state-dependent Hawkes process with $\beta < 1$ is displayed. However, as mentioned above the simulation of the process for larger time windows is not feasible and as such only the solution of the differential equation is provided. The purpose of this figure is to show that in this case, we can observe the typical behavior of a SIR model. Since this was computed in the long run, we only plot t vs $\mathbb{E}[I_t]$.

6. Conclusions and further research direction

A phenomenon that has been observed in the COVID-19 epidemic is the overdispersion of processes describing the number of contagions C_t. That is, for different time windows, the variance of the number of new contagions within that time interval is larger than their average. I.e., that for any real number $\tau > 0$,

$$\mathrm{Var}[C_{t+\tau} - C_t] \gg \mathbb{E}[C_{t+\tau} - C_t]$$

This work tries to construct a model that takes into account the overdispersion observed in epidemics with a high control reproduction number that the classic stochastic-SIR models fail to capture due to their Poissonian nature (where the expectation and variance are roughly the same for any time interval).

One possible solution to model the overdispersion is through the usage of regular point processes, in particular self-exciting point processes. Not only can they capture this phenomena but the rationale behind them makes sense. Every new infection will increase the likelihood of a new infection occurring. However, one of the main difficulties, and a very active area of research currently, is the efficient estimation of its parameters. While it is part of the ultimate goal of this work to research efficient estimation methods,

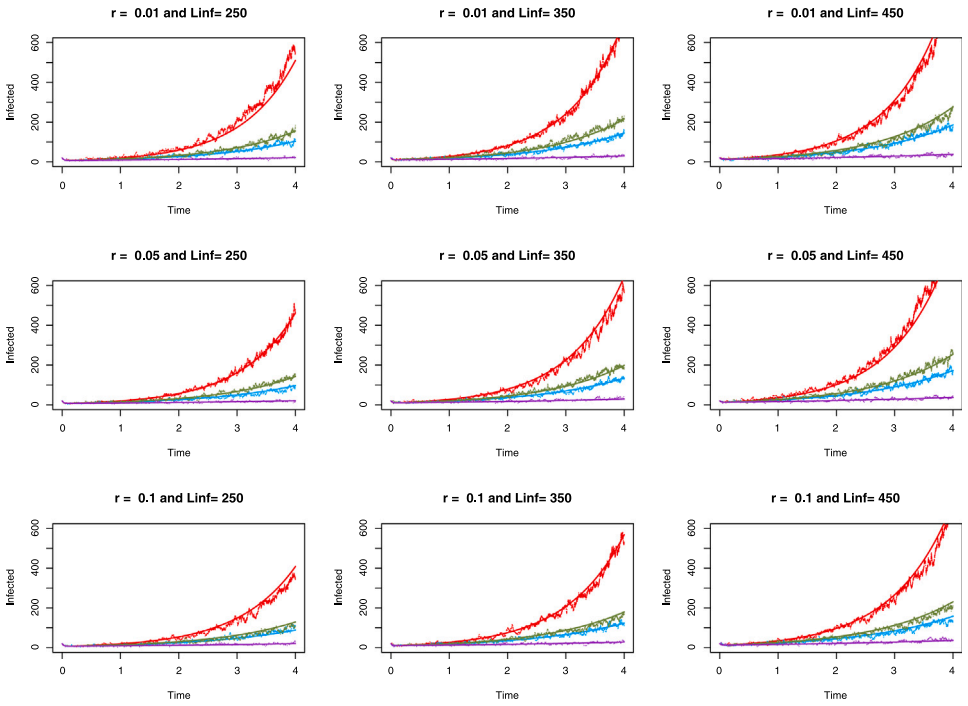


Fig. 5. Figure showing 12 simulations and their behavior for different parameters under a general state-dependent Hawkes intensity with $\alpha = 1.4$ and $\beta = 1.2$. Plots in the same row share the same value of r while plots in the same column share the same value of λ_0. The rest of the parameters, as well as the scale, were kept the same for comparison purposes. Furthermore, the Quarantine level is also represented, with High Quarantine (Q_H) in purple, Medium-High Quarantine (Q_{M_H}) in blue, Medium-Low Quarantine (Q_{M_L}) in green and Low Quarantine (Q_L) in red.

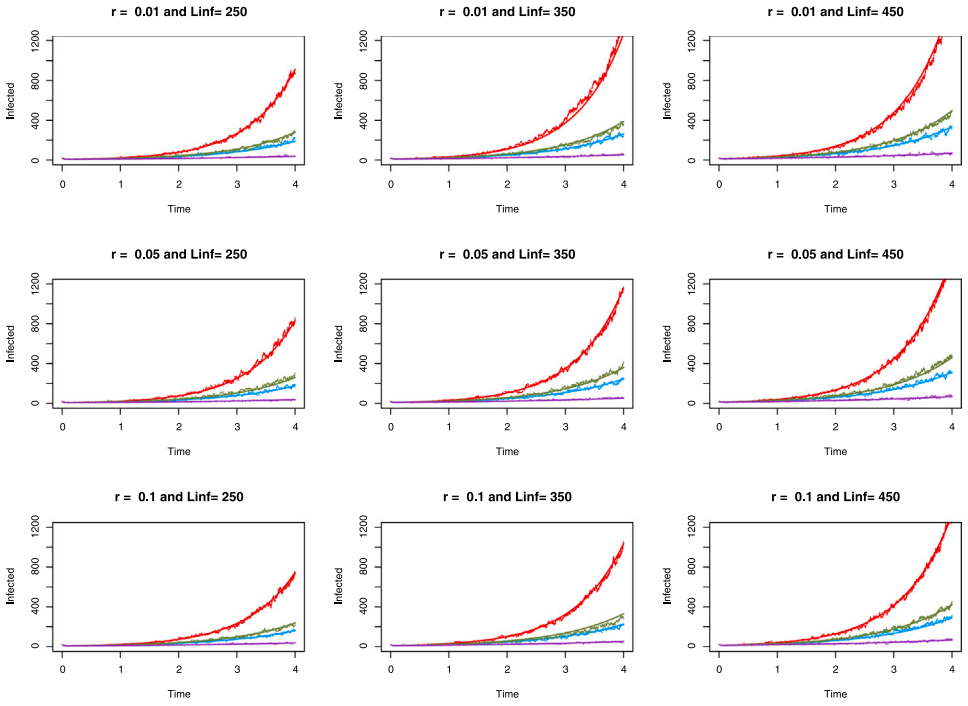


Fig. 6. Figure showing 12 simulations and their behavior for different parameters under a general state-dependent Hawkes intensity with $\alpha = 1.2$ and $\beta = 1.4$. Plots in the same row share the same value of r while plots in the same column share the same value of λ_0. The rest of the parameters, as well as the scale, were kept the same for comparison purposes. Furthermore, the Quarantine level is also represented, with High Quarantine (Q_H) in purple, Medium-High Quarantine (Q_{M_H}) in blue, Medium-Low Quarantine (Q_{M_L}) in green and Low Quarantine (Q_L) in red.

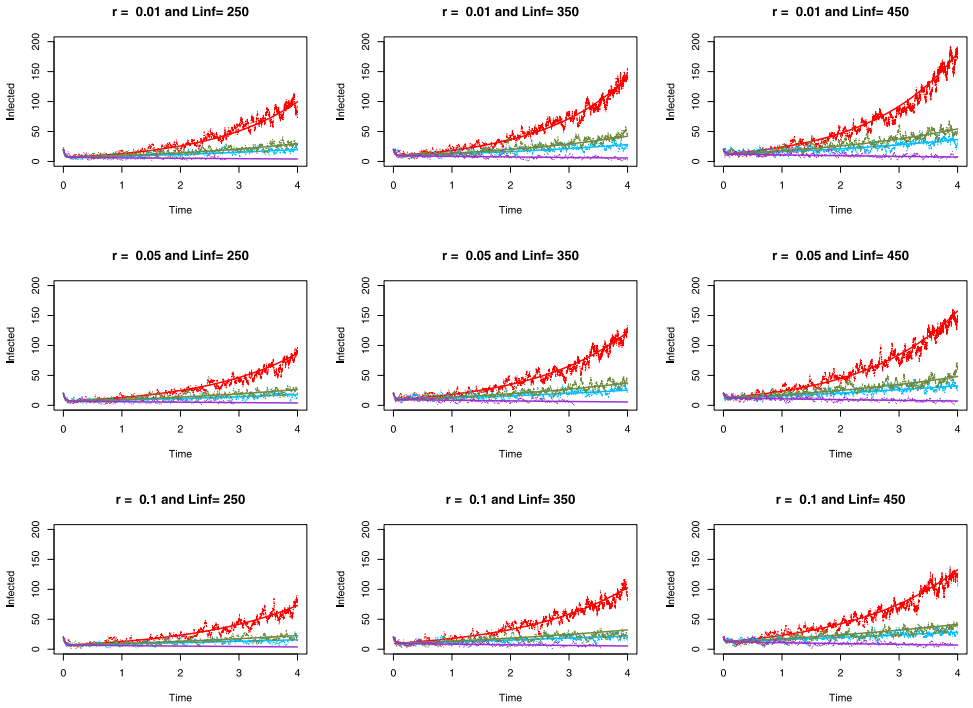


Fig. 7. Figure showing 12 simulations and their behavior for different parameters under a general state-dependent Hawkes intensity with $\alpha = 1.2$ and $\beta = 0.8$. Plots in the same row share the same value of r while plots in the same column share the same value of λ_0. The rest of the parameters, as well as the scale, were kept the same for comparison purposes. Furthermore, the Quarantine level is also represented, with High Quarantine (Q_H) in purple, Medium-High Quarantine (Q_{M_H}) in blue, Medium-Low Quarantine (Q_{M_L}) in green and Low Quarantine (Q_L) in red.

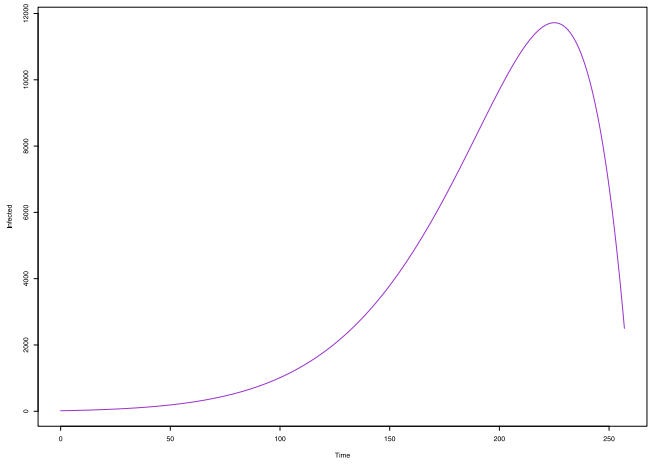


Fig. 8. Figure showing the long term behavior of the mean for the number of infected people under a general state dependent Hawkes intensity.

it is not the primary intent of this paper. The basic estimation method is in terms of the likelihood function, but there is no clear way to determine whether the method found a local maxima or a global one. Furthermore, the likelihood curve is very flat and although theoretical convergence is guaranteed, in practice it is hard to achieve.

To continue this research, there are two interesting directions to pursue. First to explore a regime-switching model where the parameters of the model change or switch randomly to different scenarios considering possible social, economic and political factors. Particularly of interest is the level of quarantining for the general population, where a stochastic optimization problem can be devised for minimizing the economic impact of a disease like COVID-19. Another direction of research is creating a full compartmental model where the transitions to different compartments are guided by point processes. This approach to classical stochastic compartmental models is of interest due to the flexibility and the statistical properties of point processes.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Jonathan A. Chávez Casillas: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

none

Data availability

Data was public.

Appendix A. Proofs of Section 3

Proof of Theorem 3. Let $\Delta t > 0$ denote an infinitesimal time step. Due to the dynamics of the SIR model, from time t to time $t+\Delta t$, and since we assume that the point processes defining the arrivals of events is simple, only one of following four (disjoint) events can happen:

- Event of type (I): There is a new infection but no recoveries. That is,
$$C_{t+\Delta t} - C_t = 1 \qquad \text{and} \qquad R_{t+\Delta t} - R_t = 0$$
- Event of type (II): There is no new infection but a recovery. Then,
$$C_{t+\Delta t} - C_t = 0 \qquad \text{and} \qquad R_{t+\Delta t} - R_t = 1$$
- Event of type (III): There is a new infection and a recovery. Thus,
$$C_{t+\Delta t} - C_t = 1 \qquad \text{and} \qquad R_{t+\Delta t} - R_t = 1$$
- Event of type (IV): There are no new infections neither recoveries. Therefore,
$$C_{t+\Delta t} - C_t = 0 \qquad \text{and} \qquad R_{t+\Delta t} - R_t = 0$$

Recall that $\{t_i\}_{i=1}^{\infty}$ are the jump times of the counting process $(C_t)_{t\geq 0}$. Then, define the function

$$J(t) = \int_0^t Q_s e^{-r(t-s)} dC_s = \sum_{t_i \leq t} Q_i e^{-r(t-t_i)}.$$

Fix the terminal state at time $t+\Delta t$, and computing

$$\begin{aligned}
F\Big(t+\Delta t, i, \rho, \lambda - r(\lambda - \lambda^{\infty}(i,\rho))\Delta t\Big) &= \mathbb{P}\Big[I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, \Lambda(t+\Delta t) \leq \lambda - r(\lambda - \lambda^{\infty}(i,\rho))\Delta t\Big] \\
&= \mathbb{P}\left[I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, \lambda^{\infty}(I_{t+\Delta t}, R_{t+\Delta t}) + \sum_{t_i \leq t+\Delta t} Q_i e^{-r(t+\Delta t - t_i)} \leq \lambda - r(\lambda - \lambda^{\infty}(i,\rho))\Delta t\right] \\
&= \mathbb{P}\Big[I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, \lambda^{\infty}(i,\rho) + J(t+\Delta t) \leq \lambda(1 - r\Delta t) + r\Delta t \lambda^{\infty}(i,\rho)\Big] \\
&= \mathbb{P}\Big[I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, (1 - r\Delta t)\lambda^{\infty}(i,\rho) + J(t+\Delta t) \leq \lambda(1 - r\Delta t)\Big] \\
&= \mathbb{P}\Big[I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, (1 - r\Delta t)\big[\lambda^{\infty}(I_t, R_t) + J(t)\big] + (1 - r\Delta t)\big[\lambda^{\infty}(i,\rho) - \lambda^{\infty}(I_t, R_t)\big]
\end{aligned}$$

$$
\begin{aligned}
&\quad + \Big(J(t+\Delta t) - (1-r\Delta t)J(t)\Big) \le \lambda(1-r\Delta t)\Big] \\
&= \mathbb{P}\Big[I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, (1-r\Delta t)\Lambda(t) + (1-r\Delta t)\big[\lambda^{\infty}(i,\rho) - \lambda^{\infty}(I_t,R_t)\big] \\
&\qquad + \Big(\sum_{t_i \le t+\Delta t} \big\{(1-r\Delta t)Q_i e^{-r(t-t_i)} + o(\Delta t)\big\} - (1-r\Delta t)J(t)\Big) \le \lambda(1-r\Delta t)\Big] \\
&= \mathbb{P}\left[I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, \Lambda(t) + \big[\lambda^{\infty}(i,\rho) - \lambda^{\infty}(I_t,R_t)\big] + \sum_{t < t_i \le t+\Delta t} Q_i e^{-r(t-t_i)} + o(\Delta t) \le \lambda\right] \\
&= \mathbb{P}\left[I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, \Lambda(t) \le \lambda - \big[\lambda^{\infty}(i,\rho) - \lambda^{\infty}(I_t,R_t)\big] - \sum_{t < t_i \le t+\Delta t} Q_i + o(\Delta t)\right]
\end{aligned}
$$

Since we are later going to take a difference quotient, divide by Δt and take the limit as $dt \to 0$, for ease of notation, we will disregard all the terms of lower order after displaying them once and we will introduce the following notation

$$
\begin{aligned}
\Delta\lambda^{\infty,1}_{i,\rho} &= \lambda^{\infty}(i,\rho) - \lambda^{\infty}(i-1,\rho) \\
\Delta\lambda^{\infty,2}_{i,\rho} &= \lambda^{\infty}(i,\rho) - \lambda^{\infty}(i+1,\rho-1) \\
\Delta\lambda^{\infty,3}_{i,\rho} &= \lambda^{\infty}(i,\rho) - \lambda^{\infty}(i,\rho-1) \\
\Delta\lambda^{\infty,4}_{i,\rho} &= \lambda^{\infty}(i,\rho) - \lambda^{\infty}(i,\rho) = 0
\end{aligned}
$$

Define

$$
E(t+\Delta t; t, i, \rho, \lambda) = \left\{ I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, \Lambda(t) \le \lambda - \big[\lambda^{\infty}(i,\rho) - \lambda^{\infty}(I_t,R_t)\big] - \sum_{t < t_i \le t+\Delta t} Q_i \right\}
$$

Then, by recalling that the point processes are simple and that $\{Q, Q_i\}_{i=1}^{\infty}$ is a sequence of i.i.d. random variables, we have that

$$
E(t+\Delta t; t, i, \rho, \lambda) = \big\{ I_{t+\Delta t} = i, R_{t+\Delta t} = \rho, \Lambda(t) \le \lambda - \big[\lambda^{\infty}(i,\rho) - \lambda^{\infty}(I_t,R_t)\big] - Q \big\}
$$

and thus we can defined the events

$$
\begin{aligned}
E_I(t,i,\rho,\lambda) = \Big\{ & I_t = i-1, R_t = \rho, \Lambda(t) \le \lambda - \Delta\lambda^{\infty}_1(i,\rho) - Q,\ C_{t+\Delta t} - C_t = 1, \\
& R_{t+\Delta t} - R_t = 0 \Big\} \\
E_{II}(t,i,\rho,\lambda) = \Big\{ & I_t = i+1, R_t = \rho-1, \Lambda(t) \le \lambda - \Delta\lambda^{\infty}_2(i,\rho),\ C_{t+\Delta t} - C_t = 0, \\
& R_{t+\Delta t} - R_t = 1 \Big\} \\
E_{III}(t,i,\rho,\lambda) = \Big\{ & I_t = i, R_t = \rho-1, \Lambda(t) \le \lambda - \Delta\lambda^{\infty}_3(i,\rho) - Q,\ C_{t+\Delta t} - C_t = 1, \\
& R_{t+\Delta t} - R_t = 1 \Big\} \\
E_{IV}(t,i,\rho,\lambda) = \Big\{ & I_t = i, R_t = \rho, \Lambda(t) \le \lambda - \Delta\lambda^{\infty}_4(i,\rho),\ C_{t+\Delta t} - C_t = 0, \\
& R_{t+\Delta t} - R_t = 0 \Big\}
\end{aligned}
$$

By considering the likelihood of the 4 types of events happening and the dynamics of the SIR model, it is clear that

$$
\begin{aligned}
\mathbb{P}\Big[E(t+\Delta t; t, i, \rho, \lambda)\Big] &= \mathbb{P}\Big[E_I(t,i,\rho,\lambda)\Big] + \mathbb{P}\Big[E_{II}(t,i,\rho,\lambda)\Big] \\
&\quad + \mathbb{P}\Big[E_{III}(t,i,\rho,\lambda)\Big] + \mathbb{P}\Big[E_{IV}(t,i,\rho,\lambda)\Big]
\end{aligned} \tag{A.1}
$$

In the following, for exposition purposes, we will compute each of the fours summands separately, and we will only consider operations up to the order Δt. Further, Then, it follows that for the vector $\mathbf{x_t} = (t,i,\rho,\lambda) \in \mathbb{R}^4$,

$$
\begin{aligned}
\mathbb{P}\Big[E_I(\mathbf{x_t})\Big] &= \int_0^{\lambda - \Delta\lambda^{\infty,1}_{i,\rho}} f(t,i-1,\rho,y)\mathbb{P}[C_{t+\Delta t} - C_t = 1]\mathbb{P}[R_{t+\Delta t} - R_t = 0]\mathbb{P}\Big[Q \le \lambda - \Delta\lambda^{\infty,1}_{i,\rho} - y\Big]dy \\
&= \int_0^{\lambda - \Delta\lambda^{\infty,1}_{i,\rho}} f(t,i-1,\rho,y)\Big[y\Delta t + o(\Delta t)\Big]e^{-(i-1)\mu\Delta t}\mathbb{P}\Big[Q \le \lambda - \Delta\lambda^{\infty,1}_{i,\rho} - y\Big]dy
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,1}} f(t,i-1,\rho,y)\Big[y\Delta t + o(\Delta t)\Big]\Big[1-(i-1)\mu\Delta t + o(\Delta t)\Big]\mathbb{P}\Big[Q \le \lambda - \Delta\lambda_{i,\rho}^{\infty,1} - y\Big]dy \\
&= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,1}} y\Delta t f(t,i-1,\rho,y)\mathbb{P}\Big[Q \le \lambda - \Delta\lambda_{i,\rho}^{\infty,1} - y\Big]dy + o(\Delta t),
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}\Big[E_{II}(\mathbf{x_t})\Big] &= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,2}} f(t,i+1,\rho-1,y)\mathbb{P}[C_{t+\Delta t}-C_t=0]\mathbb{P}[R_{t+\Delta t}-R_t=1]dy \\
&= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,2}} f(t,i+1,\rho-1,y)\Big[1-y\Delta t+o(\Delta t)\Big]\left[1-e^{(i+1)\mu\Delta t}\right]dy \\
&= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,2}} f(t,i+1,\rho-1,y)\Big[1-y\Delta t+o(\Delta t)\Big]\left[1-(1-(i+1)\mu\Delta t+o(\Delta t))\right]dy \\
&= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,2}} f(t,i+1,\rho-1,y)\Big[1-y\Delta t+o(\Delta t)\Big](i+1)\mu\Delta t dy + o(\Delta t) \\
&= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,2}} (i+1)\mu\Delta t f(t,i+1,\rho-1,y)dy + o(\Delta t),
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}\Big[E_{III}(\mathbf{x_t})\Big] &= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,3}} f(t,i,\rho-1,y)\mathbb{P}[C_{t+\Delta t}-C_t=1]\mathbb{P}[R_{t+\Delta t}-R_t=1]\mathbb{P}\Big[Q \le \lambda - \Delta\lambda_{i,\rho}^{\infty,3} - y\Big]dy \\
&= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,3}} f(t,i,\rho-1,y)\Big[y\Delta t+o(\Delta t)\Big]\left[1-e^{-i\mu\Delta t}\right]\mathbb{P}\Big[Q \le \lambda - \Delta\lambda_{i,\rho}^{\infty,3} - y\Big]dy \\
&= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,3}} f(t,i,\rho-1,y)\Big[y\Delta t+o(\Delta t)\Big]\Big[i\mu\Delta t+o(\Delta t)\Big]\mathbb{P}\Big[Q \le \lambda - \Delta\lambda_{i,\rho}^{\infty,3} - y\Big]dy \\
&= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,3}} yi\mu(\Delta t)^2 f(t,i,\rho-1,y)\mathbb{P}\Big[Q \le \lambda - \Delta\lambda_{i,\rho}^{\infty,3} - y\Big]dy + o(\Delta t) \\
&= o(\Delta t)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{P}\Big[E_{IV}(\mathbf{x_t})\Big] &= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,4}} f(t,i,\rho,y)\mathbb{P}[C_{t+\Delta t}-C_t=0]\mathbb{P}[R_{t+\Delta t}-R_t=0]dy \\
&= \int_0^{\lambda} f(t,i,\rho,y)\Big[1-y\Delta t+o(\Delta t)\Big]e^{i\mu\Delta t}dy \\
&= \int_0^{\lambda} f(t,i+1,\rho-1,y)\Big[1-y\Delta t+o(\Delta t)\Big]\Big[1-i\mu\Delta t+o(\Delta t)\Big]dy \\
&= (1-i\mu\Delta t)\int_0^{\lambda} f(t,i+1,\rho-1,y)\Big[1-y\Delta t+o(\Delta t)\Big]dy + o(\Delta t) \\
&= (1-i\mu\Delta t)F(t,i,\rho,y) - \int_0^{\lambda} y\Delta t f(t,i,\rho,y)dy + o(\Delta t)
\end{aligned}
$$

Therefore, by using Eq. (A.1) and recalling that $F\Big(t+\Delta t,i,\rho,\lambda-r(\lambda-\lambda^{\infty}(i,\rho))\Delta t\Big) = \mathbb{P}\Big[E(t+\Delta t;t,i,\rho,\lambda)\Big]$, we have that

$$
\begin{aligned}
F\Big(t+\Delta t,i,\rho,\lambda-r(\lambda-\lambda^{\infty}(i,\rho))\Delta t\Big) &= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,1}} y\Delta t f(t,i-1,\rho,y)\mathbb{P}\Big[Q \le \lambda - \Delta\lambda_{i,\rho}^{\infty,1} - y\Big]dy \\
&+ \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,2}} (i+1)\mu\Delta t f(t,i+1,\rho-1,y)dy \\
&+ (1-i\mu\Delta t)F(t,i,\rho,y) - \int_0^{\lambda} y\Delta t f(t,i,\rho,y)dy + o(\Delta t)
\end{aligned}
$$

Dividing by Δt and letting $\Delta t \to 0$,

$$
\begin{aligned}
\frac{\partial}{\partial t}F\Big(t,i,\rho,\lambda\Big) - r(\lambda-\lambda^{\infty}(i,\rho))\frac{\partial}{\partial\lambda}F(t,i,\rho,\lambda) &= \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,1}} y f(t,i-1,\rho,y)\mathbb{P}\Big[Q \le \lambda - \Delta\lambda_{i,\rho}^{\infty,1} - y\Big]dy \\
&+ \int_0^{\lambda-\Delta\lambda_{i,\rho}^{\infty,2}} (i+1)\mu f(t,i+1,\rho-1,y)dy \\
&- i\mu F(t,i,\rho,y) - \int_0^{\lambda} y f(t,i,\rho,y)dy
\end{aligned}
$$

Differentiating with respect to λ; applying Leibniz Integral Rule where appropriate; and using Assumption (A2), we get the result. □

Proof of Theorem 4. Define

$$\psi(t,i,\rho,s) := \int_0^\infty e^{-s\lambda} f(t,i,r,\lambda) d\lambda \tag{A.2}$$

$$\varphi(t,z,w,s) := \mathbb{E}\left[z^{I_t} w^{R_t} e^{-s\Lambda(t)}\right]. \tag{A.3}$$

By definition,

$$\begin{aligned}
\varphi(t,z,w,s) &= \mathbb{E}\left[z^{I_t} w^{R_t} e^{-s\Lambda(t)}\right] \\
&= \sum_{i=0}^\infty \sum_{\rho=0}^\infty \int_0^\infty z^i w^\rho e^{-s\lambda} f(t,i,\rho,\lambda) d\lambda \\
&= \sum_{i=0}^\infty \sum_{\rho=0}^\infty z^i w^\rho \int_0^\infty e^{-s\lambda} f(t,i,\rho,\lambda) d\lambda \\
&= \sum_{i=0}^\infty \sum_{\rho=0}^\infty z^i w^\rho \psi(t,i,\rho,s)
\end{aligned}$$

Thus, the strategy is to first use Theorem 3 to find a PDDE equation that characterizes the function $\psi(\bullet)$ and then transform that PDDE as the corresponding one for $\varphi(\bullet)$.

Multiplying Eq. (10) by $e^{-s\lambda}$ and integrating over $(0,\infty)$ we get:

$$\begin{aligned}
&\frac{\partial}{\partial t}\int_0^\infty f(t,i,\rho,\lambda) e^{s\lambda} d\lambda - \int_0^\infty \frac{\partial}{\partial\lambda}(r\lambda f(t,i,\rho,\lambda)) e^{-s\lambda} d\lambda \\
&+ r\lambda^\infty(i,\rho)\int_0^\infty e^{-s\lambda}\frac{\partial}{\partial\lambda} f(t,i,\rho,\lambda) d\lambda = \\
&\quad \int_0^\infty \int_0^{\lambda-\Delta\lambda_1^\infty(i,\rho)} y f(t,i-1,\rho,y) d\mathbb{P}[Q \le \lambda - \Delta\lambda_1^\infty(i,\rho) - y] dy e^{-s\lambda} d\lambda \\
&\quad + (i+1)\mu \int_0^\infty f(t,i+1,\rho-1,\lambda-\Delta\lambda_2^\infty(i,\rho)) e^{s\lambda} d\lambda \\
&\quad - \int_0^\infty (i\mu+\lambda) f(t,i,\rho,\lambda) e^{-s\lambda} d\lambda
\end{aligned}$$

Using integration by parts and Fubini's theorem, and relation (A.2)

$$\begin{aligned}
&\frac{\partial}{\partial t}\psi(t,i,\rho,s) - \int_0^\infty rs\lambda f(t,i,\rho,\lambda) e^{-s\lambda} d\lambda + r\lambda^\infty(i,\rho)\int_0^\infty s f(t,i,\rho,\lambda) e^{-s\lambda} d\lambda = \\
&\quad \int_0^\infty \int_{y+\Delta\lambda_1^\infty(i,\rho)}^\infty y f(t,i-1,\rho,y) d\mathbb{P}[Q \le \lambda - \Delta\lambda_1^\infty(i,\rho) - y] e^{-s\lambda} d\lambda dy + \\
&\quad (i+1)\mu e^{-s\Delta\lambda_2^\infty(i,\rho)} \int_{-\Delta\lambda_2^\infty(i,\rho)}^\infty f(t,i+1,\rho-1,\lambda) e^{s\lambda} d\lambda - \\
&\quad i\mu\psi(t,i,\rho,s) - \int_0^\infty \lambda f(t,i,\rho,\lambda) e^{-s\lambda} d\lambda
\end{aligned}$$

Let $M_Q(s) = \mathbb{E}[e^{-sQ}]$ denote the moment generating function of Q, which by assumption it exists on a neighborhood of 0. Then, by using the fact that $f(t,i,\rho,\lambda) = 0$ if $\lambda < 0$, definition (A.2) and Fubini's Theorem we can further simplify the above equation as

$$\begin{aligned}
&\frac{\partial}{\partial t}\psi(t,i,\rho,s) + rs\frac{\partial}{\partial s}\psi(t,i,\rho,s) + rs\lambda^\infty(i,\rho)\psi(t,i,\rho,s) = \\
&\qquad e^{-s\Delta\lambda_1^\infty(i,\rho)} M_Q(s)\frac{\partial}{\partial s}\psi(t,i-1,\rho,s) \\
&\qquad + (i+1)\mu e^{-s\Delta\lambda_2^\infty(i,\rho)}\psi(t,i+1,\rho-1,s) \\
&\qquad - i\mu\psi(t,i,\rho,s) + \frac{\partial}{\partial s}\psi(t,i,\rho,s)
\end{aligned}$$

Finally, rearranging terms we have

$$\begin{aligned}
&\frac{\partial}{\partial t}\psi(t,i,\rho,s) + (rs-1)\frac{\partial}{\partial s}\psi(t,i,\rho,s) + rs\lambda^\infty(i,\rho)\psi(t,i,\rho,s) \\
&\qquad + e^{-s\Delta\lambda_1^\infty(i,\rho)} M_Q(s)\frac{\partial}{\partial s}\psi(t,i-1,\rho,s) = \\
&\qquad (i+1)\mu e^{-s\Delta\lambda_2^\infty(i,\rho)}\psi(t,i+1,\rho-1,s) - i\mu\psi(t,i,\rho,s)
\end{aligned} \tag{A.4}$$

At this point, we use the particular form of the baseline intensity of the Hawkes process C_t, $\lambda^\infty(i,\rho)$. By using Assumption (A1), we have that

$$\lambda^\infty(i,\rho) = \lambda_0 + i\log(\alpha) + \rho\log(\beta)$$

and substituting this into the PDDE (A.4) we have

$$\begin{aligned}\frac{\partial}{\partial t}\psi(t,i,\rho,s)+(rs-1)\frac{\partial}{\partial s}\psi(t,i,\rho,s)+rs\lambda^{\infty}(i,\rho)\psi(t,i,\rho,s)\\ +\left(\frac{1}{\alpha}\right)^{s}M_Q(s)\frac{\partial}{\partial s}\psi(t,i-1,\rho,s)=\\ (i+1)\mu\,(\alpha\beta)^{s}\,\psi(t,i+1,\rho-1,s)-i\mu\psi(t,i,\rho,s)\end{aligned} \tag{A.5}$$

By doing some algebra and using the fact $\psi(t,i,\rho,s)=0$ if $i<0$ or $\rho<0$, we have the following relations

$$\sum_{i=0}^{\infty}\sum_{\rho=0}^{\infty} iz^{i}w^{\rho}\psi(t,i,\rho,s)=z\frac{\partial}{\partial z}\varphi(t,z,w,s) \tag{A.6}$$

$$\sum_{i=0}^{\infty}\sum_{\rho=0}^{\infty} \rho z^{i}w^{\rho}\psi(t,i,\rho,s)=w\frac{\partial}{\partial w}\varphi(t,z,w,s) \tag{A.7}$$

$$\sum_{i=0}^{\infty}\sum_{\rho=0}^{\infty} z^{i}w^{\rho}\psi(t,i-1,\rho,s)=z\varphi(t,z,w,s) \tag{A.8}$$

$$\sum_{i=0}^{\infty}\sum_{\rho=0}^{\infty}(i+1)z^{i}w^{\rho}\psi(t,i+1,\rho-1,s)=w\frac{\partial}{\partial z}\varphi(t,z,w,s) \tag{A.9}$$

Multiplying both sides by $z^{i}w^{\rho}$; adding those terms as a series; using Assumption (A1); and using relation (A.3), we obtain the result. □

Proof of Lemma 6. Taking Eq. (11), substituting the value function $\varphi(t,z,w,s)$ and computing some of the derivatives within,

$$\begin{aligned}&\frac{\partial}{\partial t}\mathbb{E}\left[z^{I_t}w^{R_t}e^{-s\Lambda(t)}\right]+\left[rs-1+\left(\frac{1}{\alpha}\right)^{s}M_Q(s)z\right]\mathbb{E}\left[-\Lambda(t)z^{I_t}w^{R_t}e^{-s\Lambda(t)}\right]\\ &\quad+\left[(\mu+rs\log(\alpha))\,z-\mu\,(\alpha\beta)^{s}\,w\right]\mathbb{E}\left[I_t z^{I_t-1}w^{R_t}e^{-s\Lambda(t)}\right]\\ &\quad-\left[rs\log(\beta)w\right]\mathbb{E}\left[R_t z^{I_t}w^{R_t-1}e^{-s\Lambda(t)}\right]=-rs\lambda_0\mathbb{E}\left[z^{I_t}w^{R_t}e^{-s\Lambda(t)}\right]\end{aligned} \tag{A.10}$$

Taking partial derivatives with respect to s from Eq. (A.10) and plugging in $(z,w,s)=(1,1,0)$,

$$\begin{aligned}-\frac{\partial}{\partial t}\mathbb{E}\left[\Lambda(t)\right]-\left[r-\log(\alpha)-\mathbb{E}[Q]\right]\mathbb{E}\left[\Lambda(t)\right]+\left[r\log(\alpha)-\mu\log(\alpha\beta)\right]\mathbb{E}[I_t]\\ -\left[r\log(\beta)\right]\mathbb{E}\left[R_t\right]=-r\lambda_0\end{aligned} \tag{A.11}$$

where we have used that $M_Q(0)=1$ and $\frac{d}{ds}M_Q(s)\Big|_{s=0}=-\mathbb{E}[Q]$. Similarly, taking partial derivatives with respect to z from Eq. (A.10) and plugging in $(z,w,s)=(1,1,0)$,

$$\frac{\partial}{\partial t}\mathbb{E}\left[I_t\right]-\mathbb{E}\left[\Lambda(t)\right]+\mu\mathbb{E}\left[I_t\right]=0 \tag{A.12}$$

Finally, taking partial derivatives with respect to w from Eq. (A.10) and plugging in $(z,w,s)=(1,1,0)$,

$$\frac{\partial}{\partial t}\mathbb{E}\left[R_t\right]-\mu\mathbb{E}\left[I_t\right]=0 \tag{A.13}$$

Arranging Eqs. (A.11)–(A.13) in matrix form, we get the result. □

Appendix B. Simulation algorithms

When the baseline intensity $\lambda^{\infty}(i,\rho)$ is constant, several simulation algorithms are available. However, it is rare to find simulation algorithms for more general intensities. All these algorithms were coded using R as the base programming language but they are easy to translate to any other programming language.

However, by using the classical thinning algorithm by Ogata presented in [16], we can simulate our process, but some care and considerations need to be taken.

Before showing the main algorithm, we need to be able to evaluate the intensity at any time t given the history. In this case, the history is provided in an array `tIRarray` which is a $3\times n$ array with Row 1 having all the times at which I_t or R_t changed. Rows 2 and 3 have the value of I_t and R_t, respectively, at the times on row 1.

Next, we will show the algorithm to simulate the state dependent Hawkes process (SDHP). Notice how the intensity function needs to be updated between new infections as recoveries affect the intensity function also.

Algorithm 1: Computation of the Intensity function at time t.

Result: Compute $\Lambda(t) = \lambda^{\infty}(i,\rho) + \int_0^t Q_s e^{-r(t-s)} dN_s$, for $t > 0$;
Inputs: C_t,Q_t, `tIRarray`,t,...;
Output: `Intensity`=function();
Initialization;
$n \leftarrow \text{length}(C_t) = ||C_t||$;
$aux \leftarrow sum(C_t < t) = \int_0^t dC_s$;
Computation of the intensity function;
if $(n = 0 \;||\; aux = 0)$ **then**
  return $\lambda^{\infty}(i_0, \rho_0)$;
else
  $Qs \leftarrow Q_t[1 : aux]$;
  $Cs \leftarrow C_t[1 : aux]$;
  $exps \leftarrow \exp(-r(t - Cs))$;
  $integ \leftarrow sum(Qs * exps)$;
  $lam \leftarrow \lambda^{\infty}(I_t, R_t)$;
  return $lam + integ$;
end

Algorithm 2: Simulation of a trajectory of the State-dependent Hawkes SIR model

Result: Simulation of State-Dependent Hawkes Process with intensity function $\Lambda(t) = \lambda^{\infty}(i,\rho) + \int_0^t Q_s e^{-r(t-s)} dN_s$
Inputs: T, Qvals, Qprobs, $i_0, \rho_0, \ldots$;
Output: SDHP=function();
Initialization;
`tIRarray` $\leftarrow [0, i_0, \rho_0]^T$;
$t \leftarrow 0$;
while $t \leq T$ **do**
  $M \leftarrow$ `Intensity` $(t, \ldots)$;
  if $M \leq 0$ **then**
    break;
  end
  $E \sim \text{Exp}(M)$;
  $t \leftarrow t + E$;
  $U \sim \text{Unif}(0, M)$;
end
As $\lambda^{\infty}(i,\rho)$ can change between increments of C_t (via only a recovery), we need to update our intensity function;
Raux $\leftarrow$ 0;
while $Raux < E$ **do**
  PossibleR $\sim$ Exp($\mu \cdot$ Icurrent);
  if $(PossibleR + Raux) > E$ **then**
    break
  end
  Raux $\leftarrow$ Raux + PossibleR;
  Update `tIRarray`;
end
if $t < E$ **then**
  ## Perform a thinning routine to accept new Infections;
  $\Lambda_N \leftarrow$ `Intensity`$(t, \ldots)$;
  if $U < \Lambda_N$ **then**
    Accept t as time of Infection;
    Update all parameters;
  end
end

References

[1] Adiga A, Chen J, Marathe M, Mortveit H, Venkatramanan S, Vullikanti A. Data-driven modeling for different stages of pandemic response. J Indian Inst Sci 2020;100(4):901–15.
[2] Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of COVID-19. Proc Natl Acad Sci 2020;117(29):16732–8.
[3] Cheng X, Han Z, Abba B, Wang H. Regional infectious risk prediction of COVID-19 based on geo-spatial data. PeerJ 2020;8:e10139.
[4] Chiang W-H, Liu X, Mohler G. Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. Int J Forecast 2022;38(2):505–20.
[5] dos Santos Gomes DC, de Oliveira Serra GL. Machine learning model for computational tracking and forecasting the COVID-19 dynamic propagation. IEEE J Biomed Health Inf 2021;25(3):615–22.
[6] Franco E. A feedback SIR (fSIR) model highlights advantages and limitations of infection-dependent mitigation strategies. 2020, arXiv preprint arXiv:2004.13216.
[7] Hazarika BB, Gupta D. Modelling and forecasting of COVID-19 spread using wavelet-coupled random vector functional link networks. Appl Soft Comput 2020;96:106626.
[8] Liu Q-H, Zhang J, Peng C, Litvinova M, Huang S, Poletti P, et al. Model-based evaluation of alternative reactive class closure strategies against COVID-19. Nat Commun 2022;13(1):1–10.
[9] Van den Driessche P. Reproduction numbers of infectious disease models. Infect Dis Model 2017;2(3):288–303.
[10] Daley DJ, Vere-Jones D. An introduction to the theory of point processes: volume I: elementary theory and methods. Springer; 2003.
[11] Daley DJ, Vere-Jones D. An introduction to the theory of point processes: volume II: general theory and structure. Springer New York; 2008.
[12] Brémaud P. Point Process Calculus in Time and Space: An Introduction with Applications, vol. 98, Springer Nature; 2020.
[13] Greenwood PE, Gordillo LF. Stochastic epidemic modeling. In: Mathematical and statistical estimation approaches in epidemiology. Springer; 2009, p. 31–52.
[14] Kurtz TG. Approximation of population processes. SIAM; 1981.
[15] Meyn SP, Tweedie RL. Markov chains and stochastic stability. Springer Science & Business Media; 2012.
[16] Ogata Y. On Lewis' simulation method for point processes. IEEE Trans Inf Theory 1981;27(1):23–31.
[17] Rizoiu M-A, Mishra S, Kong Q, Carman M, Xie L. SIR-Hawkes: Linking epidemic models and Hawkes processes to model diffusions in finite populations. In: Proceedings of the 2018 world wide web conference. 2018, p. 419–28.
[18] Escobar JV. A Hawkes process model for the propagation of COVID-19: Simple analytical results. Europhys Lett 2020;131(6):68005.
[19] Garetto M, Leonardi E, Torrisi GL. A time-modulated Hawkes process to model the spread of COVID-19 and the impact of countermeasures. 2021, arXiv preprint arXiv:2101.00405.
[20] Chiang W-H, Liu X, Mohler G. Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. 2020, medRxiv.
[21] Lesage L. A Hawkes process to make aware people of the severity of COVID-19 outbreak: application to cases in France (Ph.D. thesis), Université de Lorraine; University of Luxembourg; 2020.
[22] Chowdhury A, Shahbaz M, Karim R, Islam J, Dan G, Shuixiang H. A comparative study on ivermectin-doxycycline and hydroxychloroquine-azithromycin therapy on COVID-19 patients. Eurasian J Med Oncol 2021;5:63–70.
[23] Laub PJ, Lee Y, Taimre T. The elements of Hawkes processes. Springer; 2022.
[24] Koops DT, Saxena M, Boxma OJ, Mandjes M. Infinite-server queues with Hawkes input. J Appl Probab 2018;55(3):920–43.
[25] Lewis E, Mohler G, Brantingham PJ, Bertozzi AL. Self-exciting point process models of civilian deaths in Iraq. Secur. J. 2012;25(3):244–64.
[26] Veen A, Schoenberg FP. Estimation of space–time branching process models in seismology using an em–type algorithm. J Amer Statist Assoc 2008;103(482):614–24.
[27] Foschi R, Lilla F, Mancini C. Warnings about future jumps: properties of the exponential Hawkes model. 2020, Available at SSRN 3639050.
[28] Cartea Á, Cohen SN, Labyad S. Gradient-based estimation of linear Hawkes processes with general kernels. 2021, arXiv preprint arXiv:2111.10637.