Comparison of the Hawkes and SEIR models for the spread of Covid-19

Conor Kresin*, Frederic Paik Schoenberg*, and George Mohler†

Abstract

Two models that capture the spread of infectious diseases, the Hawkes point process model and the SEIR compartmental model, are compared with regard to their use in modeling the Covid-19 pandemic. The physical plausibility of the SEIR model is weighed against the parsimony and flexibility of the Hawkes model. The mathematical connection between Hawkes and SEIR models is described.

*Department of Statistics, University of California Los Angeles

†Department of Computer and Information Science, Indiana University - Purdue University Indianapolis

MSC 2020 subject classifications. Primary-60G55; secondary-92D30

Key words: compartmental models, Coronavirus, Hawkes models, model evaluation, point process models, self-exciting point process models

1 Introduction

The SARS-CoV2 (Covid-19) pandemic spread from China to at least 188 countries or regions in the first six months of 2020 [15]. The characteristics of the Covid-19 virus have been estimated and forecasted by numerous researchers with highly variable results. Estimates of properties such as reproduction rate (or time-varying reproduction number), numbers of individuals infected, hospitalization rates, fatality rates, and efficacy of containment measures have varied widely [4, 22]. Accurate real-time estimates of the spread of Covid-19 are difficult to achieve without population-wide testing [18, 6]. Nevertheless, it is important for researchers to accurately estimate and forecast the dynamics of Covid-19 so that optimal public policy measures and other responses can be adopted.

Several different frameworks have been proposed for modeling the spread of Covid-19, including compartmental models such as the SEIR (Susceptible $\rightarrow$ Exposed $\rightarrow$ Infectious $\rightarrow$ Removed) differential equation model, and branching point process models such as the Hawkes point process model [70, 37, 6, 10]. This paper compares these two approaches for forecasting Covid-19. Relative to Hawkes models, SEIR models and their variants have been used far more widely to describe the Covid-19 pandemic [7, 24, 31, 42] as well as other infectious diseases such as Ebola [43] and SARS [19]. However, recent studies have suggested that Hawkes models may be more accurate [87]. For general discussion of mathematical and statistical models of epidemiological phenomena, see [60, 28].

This paper is structured as follows. Following a review of Hawkes and SEIR models in Section 1, we compare their advantages and disadvantages, especially with respect to forecasting Covid-19 cases or deaths in Section 2. In Section 3, we detail the mathematical connection between Hawkes processes and SEIR models, and in Section 4 we provide concluding remarks.

1.1 The Hawkes Model

The Hawkes model or *self-exciting* point process model is commonly used to model clustered point patterns in applications such as seismology, finance, crime, and infectious diseases [16, 69, 57, 9]. A spatial-temporal Hawkes process is specified by the model

$$\begin{aligned} \lambda(s,t) &= \mu(s) + K \int_{t'<t} g(s-s', t-t') dN(s',t') \\ &= \mu(s) + K \sum_{(s',t'):t'<t} g(s-s', t-t'), \end{aligned} \tag{1}$$

for $s \in X \subseteq \mathbb{R}^2$ and $t \in [0, T)$, where $\lambda(s,t|\mathcal{H}_t)$ is the conditional rate at which points (events) are expected to accumulate around spatial-temporal location (s,t), given information on all previous events. The conditional intensity uniquely characterizes the finite-dimensional distribution of any simple point process (see Prop. 7.2.IV of [16]), and thus equation (1) fully specifies the model.

The function g is typically assumed to be a density, i.e. to be nonnegative and to integrate to 1 over all time and space, and is called the *triggering density*. Common choices for g are the exponential or Pareto densities in time, and the Gaussian or Pareto densities in space [69]. The constant K is called the *productivity*. Provided g is a density function, K is the expected number of points triggered directly by each point, and is thus closely connected to the reproduction number in compartmental models such as SEIR. Each background point, associated with $\mu(s)$, is expected to generate $K + K^2 + K^3 + ... = 1/(1-K) - 1 = K/(1-K)$ triggered points. As a result, in a Hawkes process, the expected fraction of background points is $1 - K$.

Given a dataset consisting of n points within a space-time observation region B, the parameters in Hawkes processes are typically fit by maximum likelihood estimation (MLE), where one obtains parameter estimates $\hat{\Theta}$ maximizing

$$L(\Theta) = \sum_{i=1}^{n} \log(\lambda(s_i, t_i)) - \int_B \lambda(s,t) dt ds.$$

The resulting estimates have desirable properties. For instance, Ogata (1978) showed that the MLE $\hat{\Theta}$, is, under standard conditions, asymptotically unbiased, consistent, asymptotically normal, and asymptotically efficient, with standard errors readily constructed using the diagonal elements of the inverse of the Hessian of L evaluated at $\hat{\Theta}$ [59]. Further, if the fitted model is missing some relevant covariates, under general conditions the MLE will nevertheless be consistent, provided the effect of the missing covariates is small [74]. The triggering function can also be estimated non-parametrically [49], and some authors have also estimated the background rate $\mu(s)$ nonparametrically, *e.g.* [91, 65]. Bayesian methods can also be used to estimate parameters and quantify uncertainty in Hawkes process models [68, 53].

A host of variations of the Hawkes model have been proposed [10, 70]. The HawkesN model, as defined in [70], has a Hawkes conditional intensity scaled by the proportion of events which can still occur after time t, in order to account for the dynamic decrease in the number of susceptible individuals in a given location [70]:

$$\lambda(t) = (1 - I_c(t)/N)(\mu + K \sum_{t' < t} g(t - t')). \tag{2}$$

In the context of a Hawkes process modeling the spread of an infectious disease, $I_c(t)$ is the cumulative number of infections that have been recorded up to time t and N is the total population size.

Hawkes models and their slight variants such as the epidemic-type aftershock sequence (ETAS) model [57, 58], HawkesN [70, 6, 54], and the recursive model [76] have been shown to be useful in modeling infectious diseases such as Ebola [39, 64], chlamydia [75], SARS [82, 9], measles [21], meningococcal disease [51], and Rocky Mountain Spotted Fever [76]. Hawkes models have also been shown to be the best fitting models for forecasting seismicity in rigorous, purely prospective earthquake forecasting studies such as the Collaboratory for the Study of Earthquake Predictability (CSEP) [13, 14, 90, 8, 27, 77].

1.2 The SEIR Model

SEIR models and their variants have been widely used to model and forecast the spread of many contagious diseases including Covid-19 [81, 86, 7, 29, 44, 32, 30, 66, 52, 31]. Such models employ a wide variety of modifications to the classic SEIR model, including using Bayesian inference [52], machine learning [31], mobility networks and ensemble approaches [32], and mixed-effects curve fitting [81] to fit parameters, as well as slight compartmental variants like SuIER which account for unreported cases [30]. Other models explicitly define scenarios for government interventions or enforcement of public health policies in specific populations [44, 7, 29].

SEIR models assume that individuals within each category, or *compartment* (susceptible, exposed, infectious, and recovered), share pertinent characteristics, and the size of the population of interest N is equal to the total number of individuals in the compartments [40]. SEIR models are a slight extension of SIR (Susceptible $\rightarrow$ Infectious $\rightarrow$ Removed) models, generalized to account for the fact that there is an incubation time for some infectious diseases like Covid-19, during which the exposed host may be asymptomatic and thus not recorded as infected. SEIR models can be either deterministic, in which case they are comprised of a system of differential equations, or stochastic, in which case they are based on a Markov chain framework. Given large populations, sufficient initial spread, and enough time, the deterministic framework should resemble the stochastic framework in expectation, assuming properly specified models [70].

Deterministic SEIR models, such as that described in Figure 1, can provide a reasonable approximation of the characteristics of a contagious disease such as Covid-19. There are numerous variations, but the basic idea conveyed in Figure 1 common to compartmental models is that there is some rate at which people shift from one portion of the population to another, e.g. from the susceptible population to the exposed population, and these rates may be fixed or allowed to vary over time subject to certain constraints. Deterministic models such as that shown in Figure 1 can be extended to allow parameters governing the force of infection, number of cases by symptom onset, and death rate, with movement between compartments commonly specified as binomial random variables. Such a model has been suggested for the transmission of Ebola, for instance [43]. Number of cases or deaths are commonly specified as a negative binomial random variable [45, 35].

Perhaps the most common method for estimating compartmental infectious disease model parameters is by using Bayesian estimation [62, 12]. Prior parameters are often decided on using subject matter experts [52] or parameters fit to prior outbreaks [25]. Bayesian SEIR models have been employed to model infectious diseases such as Ebola [25], Visceral Leishmaniasis [62] and Covid-19 [52]. Prior distributions are typically specified by compartment. For instance, Frasso et al. specified number of deaths as beta-distributed, duration of incubation as normal, and observed cases as negative binomial [25]. Disease characteristics such as reproduction number has been modelled within the context of SIR models with a gamma prior [12]. Joint posteriors are then solved for using a

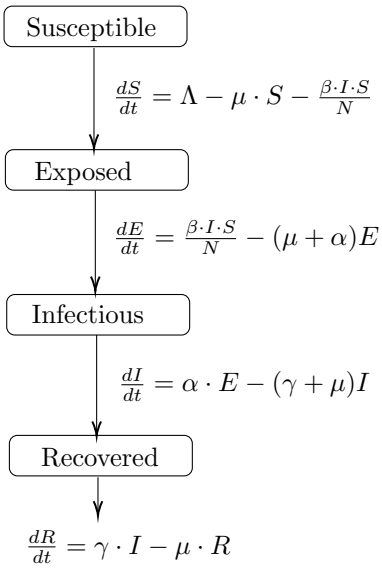

Figure 1: Diagram of the deterministic SEIR model. Definitions: N is a constant number of individuals in a susceptible population, $\beta \cdot I$ is equal to the force of infection, Λ equal to birth rate, μ equal to death rate, γ equal to mortality rate, α^{-1} equal to the average incubation period. Such a model has reproduction number $R_0 = \frac{\alpha \cdot \beta}{(\mu+\alpha)(\mu+\gamma)}$.

MCMC approach such as Metropolis-Hastings or Gibbs sampling [12, 43].

Stochastic versions of the SIR and SEIR models allow researchers to include the effect of networks of individuals, but specification and parameter estimation can be more challenging [3]. Various stochastic SEIR models have been developed to model Covid-19 data. A stochastic SEIR model with parameters fit using grid search, which may be viewed as a relatively agnostic machine learning approach, was implemented in [31], and a stochastic SEIR model hybrid with agent-based simulation was suggested in [42]. The compartmental approach of the SEIR model is slightly modified to accommodate under-detection and differentiated government intervention in the DELPHI model [7]. Their flexibility notwithstanding, the difficulty in estimating time-varying parameters in real time for stochastic SEIR models is well known, especially for large populations [55].

Parameters in SEIR models are often estimated using opinions of expert epidemiologists or using data from other locations or past epidemics [11]. This is attractive in the sense that expert opinion is integrated, but there is ample opportunity for bias as well as mis-specification, and the parameter estimates have a covariance structure that can be difficult to estimate. Further, non-identifiability is a known problem for compartmental models [26]. Although there exists algebraic approaches for testing identifiability such as exhaustive modeling [83], such methods are not implemented in any of the above referenced Covid-19 SEIR models. Crucially, estimated SEIR parameters in the early stages of a epidemic (before peak infection) have been shown to be struc-

turally nonidentifiable [73].

2 Comparison of Point Process and Compartmental Models

Hawkes and SEIR models both offer flexible (and somewhat complementary) frameworks for modeling infectious diseases. Hawkes models allow for nonparametric estimation of the triggering function g, as well as spatial covariates, and an intrinsic network-effect. SEIR models offer a far more physically plausible framework for describing Covid-19 relative to the Hawkes model. Specifically, SEIR models allow for specification of stochastic movement between compartments based on previous epidemics and expert opinions. The compartmental model framework allows for natural implementation of known networks within the population of interest [55]. Further, quantities of interest to epidemiologists and policy makers such as infection rate within a population can be imputed using SEIR models [43].

Within the context of a SEIR model, the spread or transmission of an infectious disease such as Covid-19 occurs via Markovian diffusion which, under certain regularity conditions, ultimately converges to a stationary distribution. In the context of a Hawkes model, background events trigger future events, and these trigger subsequent events, ultimately resolving due to the decay of the chosen triggering function if the productivity is less than one. In general, a point process is considered to be stationary when for all bounded Borel subsets $A_1, \ldots, A_r$ of the real line, the joint distribution of $\{N(A_1 + t), \ldots, N(A_r + t)\}$ is independent of $t \in \mathbb{R}$ [16]. A Hawkes process with $K > 1$ is not stationary [80]. The HawkesN model is a stationary process for $K > 1$.

The link between Hawkes-like and stochastic SIR models is explored in detail in [70], where it is shown that an exponentially decaying triggering function chosen for a finite population Hawkes model (HawkesN) coincides in expectation with the number of individuals infected in a stochastic SIR model as it approaches stationarity. This connection between SIR and Hawkes models was explored in particular in the context of Covid-19 [6], where it was shown that the HawkesN and SIR models converge if the triggering function is exponential and the reproduction number in the SIR model is constant [6]. SIR and HawkesN models are shown to provide similar fit to Twitter re-tweet diffusions in [70].

One may also compare features such as the doubling time for both Hawkes and SIR/SEIR models. In the early exponential growth stage, the doubling time for SIR is $\tau = \log(2)/(\gamma(R_0 - 1))$ [1]. The relationship between estimates of R_0 and doubling time for simulations of compartmental models is summarized in [50]. The parameter K is intuitively similar to R_0, as it represents the expected number of events triggered by a previous event. The doubling time for HawkesN models as a function of K is shown next to the doubling time of a SIR model as a function of R_0 in Figure 2. It should be noted that doubling time for Hawkes models quickly approaches zero for $K > 1$, justifying the finite population correction present in the HawkesN in the context of modeling infectious diseases

such as Covid-19 [70].

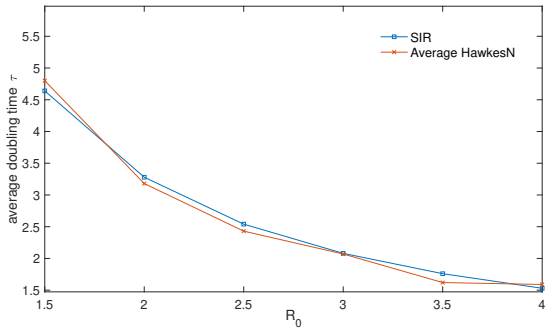

Figure 2: Average doubling time for HawkesN model with $\beta = \frac{1}{4}$, $I_0 = 10$, population size $N = 10^6$, and using mean intensity over 100 simulations per K (notated R_0 for SIR). Doubling time is defined as t such that $N(t) = 20$.

Hawkes models offer computationally inexpensive parametric and non-parametric estimates for important characteristics of infectious diseases such as Covid-19. Due to computational difficulty, and model-specification convenience, both SEIR and Hawkes models often make assumptions such as fixed population size, or homogeneity individuals within compartments. Despite this, the difficulty of specifying large population size stochastic SEIR models in real time is not trivial. In general, Hawkes models seem to be far simpler to implement than SEIR-type models, and in a pandemic such as the spread of Covid-19, where resources can be scarce and policies and health-allocations must be made in real-time, quick and accurate short term forecasts are highly valuable [85].

Relative to Hawkes processes, SEIR models are more natural mathematical representations of the spread of contagious diseases. However, in implementation SEIR models often require estimation of more parameters and structural modifications. With complexity, there is more opportunity for bias and random errors in parameter estimates, as well as large covariances between pairs of parameter estimates, and in some cases problems of identifiability [20, 71]. More pressingly, each component of the model is susceptible to mis-specification, which can result in highly variable estimates and large forecasting errors [34, 61].

Problems such as these can be particularly severe in the case of Covid-19, where available data used to fit parameters can rely can have substantial errors, due to undercounted infected populations and testing policies that vary over time and space [41]. Both Hawkes and SEIR models assume a homogenous population and do not explicitly account for testing errors, but Hawkes and HawkesN models appear to perform better than their SEIR equivalents for modeling the spread of infectious diseases (see Table 1 below).

2.1 Comparison of Covid-19 Results in the Current Literature

SEIR models appear to be far more widely used by State and Federal agencies for forecasting Covid-19 cases and deaths, with a notable exception being the State of New Jersey which is primarily using a multivariate Hawkes model [36]. Table 1 summarizes results comparing the accuracy of Hawkes models and their variants with SEIR models and their variants for forecasting infectious diseases. Point process models have been found to forecast incidence of mumps in Pennsylvania better than compartmental SVEILR models [38]. Further, point process models have been found to improve fit and forecasting performance relative to SEIR models when applied to incidence of pertussis in [87]. Yuan et al. find substantially improved accuracy of Hawkes models over SEIR models for forecasting Covid-19 in the European Union, California, New York, and for the United States as a whole [89].

Data	**Better Fit**	**Worse Fit**	**Reduction RSME**	**Authors**
Pertusis in NV	Recursive Hawkes	SEIR	19%	[87]
Mumps in PA	Recursive Hawkes	SVEILR	38%	[38]
	Hawkes	SVEILR	26%	
Covid-19 in CA	SEIR	Hawkes	(∗)	[6]
Covid-19 in IN				
Covid-19 in NY	Hawkes	SEIR		
Covid-19 in CA	Hawkes	SEIR	63%	[89]
Covid-19 in NY			21%	
Covid-19 in US			31%	
Covid-19 in EU			27%	
Covid-19 in US	Hawkes Variants	SEIR	(∗∗)	[10]
Ebola in W. Africa	Hawkes	SEIR	38%	[64]

Table 1: Prior results comparing the forecasting accuracy of point process and compartmental models for infectious diseases. Errors reported are the root mean squared error (RMSE) and (∗∗) mean absolute error of daily forecasts. Model selection using (∗) Akaike Information Criterion (AIC) and (∗∗) Normalized Discounted Cumulative Gain.

Hawkes models are directly compared to SIR and SEIR models to explain the spread of Covid-19 in California, Indiana, and New York in [6]. The Akaike information criterion (AIC) is used to evaluate the candidate models, and by this metric, HawkesN performs more poorly relative to its compartmental counterparts for Covid-19 death data, and with mixed results for Covid-19 case data. However, fitted parameters are found to vary materially across locations, and relative fit of parameters across models is concluded to not be strongly indicated. Rather than concluding on the merits of either type of model, the authors note the difficulty of using limited data at the beginning of an epidemic such as that of Covid-19 [6].

In the context of the Covid-19 pandemic, compartmental models such as SIR and SEIR have been noted to generally have low accuracy for long-term forecasts, and machine learning models have been proposed as a superior alternative [2]. Compartmental models also may be poorly calibrated for forecasting more than five days out: forecast numbers of Covid-19 cases in Italy six days in the future based on the SEIR model were 14% too low on average [2]. Various compartmental models for forecasting Covid-19 yield slightly different projections of future cases or future deaths [72]. However, estimates of variability vary widely, with prediction interval widths often varying by a factor of 3 [7, 24].

Some variation is to be expected in both mean predicted deaths and size of prediction interval between the models as each are designed differently, and with varying assumptions. Estimates of the initial reproduction number R_t for COVID-19 vary around 3.28 (1.4, 6.5) [63]. Of course, values of R_t are observed to vary substantially depending on social distance policies. In China, estimates of R_t decreased from 2 to 1 when public health measures were put in place [88]. Estimates of R_t in Singapore correspondingly decreased over time by between 78.2% and 99.3% [37]. Similar results were observed in Europe as a result of public health measures [23].

SEIR forecasts of future confirmed cases or deaths depend critically on estimates of the total numbers of asymptomatic or mildly symptomatic cases, which are highly uncertain [78, 5] and extremely difficult to estimate accurately [48, 79]. Jewell et al. [37] note that more detailed and complex models may be more sensitive to assumptions regarding the incubation and infectious periods and other estimates of transmission characteristics. Further, SIR and SEIR models are highly sensitive to assumptions regarding social movement and the estimated impacts of containment policies [67]. SIR and SEIR models are known to be particularly sensitive to assumptions about the distribution of latent and infectious periods [47, 84]. Further, as discussed above, nonidentifiable parameters can be an issue for compartmental models, and methods for dealing with nonidentifiability of parameters tend to work better for simpler models than for more complex compartmental models [71].

3 Further Connections Between Hawkes and SEIR Models

The productivity constant K in the Hawkes model is the obvious analogue of the reproduction rate R_0 in SEIR, with both interpretable as the expected number of direct transmissions per infected individual. Further, several variations of the Hawkes process in Equation 1 have deeper connections to SEIR-type compartmental models. The point process governed by Equation 2 is a continuous time analog of a discrete stochastic SIR model when $g(t)$ is specified as exponential [70]. When $g(t)$ is chosen to be gamma distributed, the Hawkes process also can approximate staged compartment models, like SEIR, if the average waiting time in each compartment is equal [47]. More complex parametric (or non-parametric) inter-infection time distributions $g(t)$ may be employed within the Hawkes process framework in situations where disease dynamics cannot be

captured by a SIR or SEIR model. In the early exponential growth stage of an epidemic, before finite population and social distancing effects play a role, the linear Hawkes process in Equation 1 can readily be used to model new infections (see Figure 3).

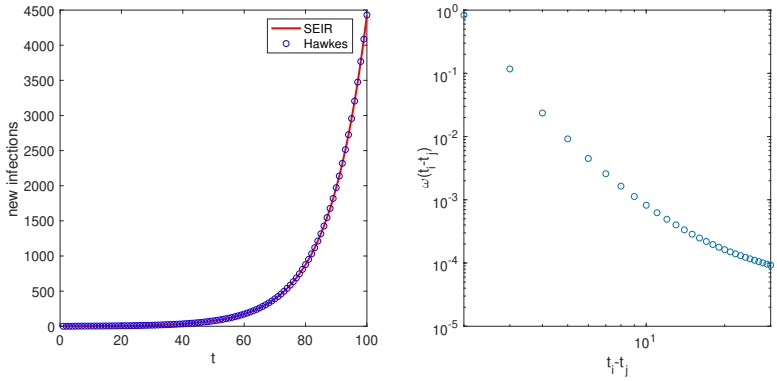

Figure 3: Left: (Red) SEIR differential equation $dS/dt = -\beta SI/N$, $dE/dt = \beta SI/N - \mu E$, $dI/dt = \mu E - \gamma I$, $dR/dt = \gamma I$, where $\beta = \gamma R_0$, $\gamma = .1$, $R_0 = 2$, $\mu = 1$, and $N = 5 \cdot 10^8$. (Blue) linear Hawkes process $\lambda_t = \mu + \sum_{t>t_i} Kg(t - t_i)$ fit to the SEIR curve of new infections using non-parametric expectation-maximization [54]. Right: Non-parametric histogram estimate for $g(t)$ corresponding to the Hawkes process fit.

While the Hawkes process can approximate SEIR in some situations with an appropriately chosen kernel $g(t)$, queue-Hawkes processes [17] can also be used to model an exposed latent class of events. Let N be population size, N_t^E be the cumulative sum of infections (whether recovered or not) up to time t. Then we may define a hybrid model incorporating features of both SEIR and Hawkes, which we call a SEIR-Hawkes process, where the intensity of newly exposed cases is given by

$$\lambda^E(t) = \left(1 - \frac{N_t^E}{N}\right) \sum_{t>t_j^I} R_0 \gamma \exp\Big(- \gamma(t - t_j^I)\Big), \tag{3}$$

and the times of infection are generated via

$$P(t_j^I > t_j^E + c) = \int_c^\infty \mu \exp\Big(- \mu(s - t_j^E)\Big) ds. \tag{4}$$

Realizations of the SEIR-Hawkes process can be generated via Lewis' thinning method for simulation [56, 46]. We first simulate an upper-bounding Hawkes process with intensity

$$\nu^E(t) = \sum_{t>s_j^I} R_0 \gamma \exp\Big(- \gamma(t - s_j^I)\Big). \tag{5}$$

$$P(s_j^I > s_j^E + c) = \int_c^\infty \mu \exp\Big(- \mu(s - s_j^E)\Big) ds. \tag{6}$$

Because the Hawkes process in Equation 5 has a branching process representation [33], the process can be simulated iteratively; for each event pair (s_j^I, s_j^E), by

1. Generating a Poisson random variable M with mean R_0 .
2. Generating $l = 1, ..., M$ events with inter-event times $s_l^E - s_j^I$ given by an exponential random variable with parameter γ.
3. Generating $l = 1, ..., M$ events with inter-event times $s_l^I - s_l^E$ given by an exponential random variable with parameter μ.

Thinning then proceeds sequentially by accepting each event pair (s_j^I, s_j^E) with probability $\lambda^E(s_j^E)/\nu^E(s_j^E)$ where λ^E is computed using only accepted events in the history and ν^E is computed using all simulated events. In Figure 4 we simulate the SEIR-Hawkes process with parameters $\mu = 1$, $\gamma = .1$, $R_0 = 2$, $N = 1000$ and $N_0^E = 10$ ($t_1^E = ...t_{10}^E = 0$) and compare to the forward-Euler approximate solution ($dt = .01$) of a SEIR differential equation $dS/dt = -\beta SI/N$, $dE/dt = \beta SI/N - \mu E$, $dI/dt = \mu E - \gamma I$, $dR/dt = \gamma I$, where $\beta = \gamma R_0$.

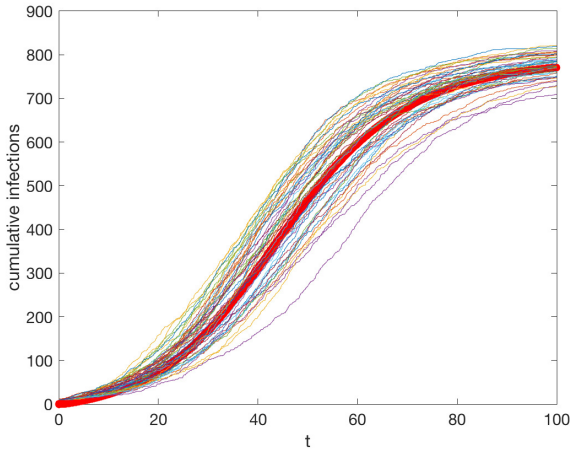

Figure 4: SEIR differential equation simulation (red) and 50 realizations of the SEIR-Hawkes process. Parameters for the SEIR model are $R_0 = 2$, $\mu = 1$ for the $E \to I$ rate, $\gamma = .1$ for the $I \to R$ rate, and population size $N = 1000$.

4 Conclusion

The SEIR model is currently far more widely used to model epidemic diseases such as Covid-19 than the Hawkes model, and its parameterization is physically plausible, with parameters that are readily interpeted in the epidemiological community. The SEIR model also appears to forecast epidemics adequately in most cases, especially in the early spread of the disease. However, the Hawkes model seems to offer more accurate forecasts, with approximately 20-30% smaller errors on average in most cases. Among the several reasons listed in Section 2 for this discrepancy, the most significant seem to be mis-specification

in the SEIR model and its sensitivity to errors in estimates of latent quantities such as the number of asymptomatic individuals and the distribution of incubation times. In general, when maximal accuracy is desired, models for forecasting observations should typically be only as complex as necessary to represent the main features of interest in the data, with minimal dependence on unobserved or noisy data [39].

There are close connections between SEIR and Hawkes models, and indeed the two types of models can be constructed to be equivalent or to converge to one another in special cases. The SEIR-Hawkes model described here may provide further linkage between the two paradigms in cases where one seeks the accuracy of point process modeling without sacrificing the physical plausibility and interpretation of SEIR parameters, and the model is shown here to emulate characteristics of SEIR models closely.

5 Acknowledgements

This research was supported by NSF grants SCC-1737585 and ATD-1737996.

References

[1] Linda JS Allen. Some discrete-time si, sir, and sis epidemic models. *Mathematical biosciences*, 124(1):83–105, 1994.

[2] Sina F Ardabili, Amir Mosavi, Pedram Ghamisi, Filip Ferdinand, Annamaria R Varkonyi-Koczy, Uwe Reuter, Timon Rabczuk, and Peter M Atkinson. Covid-19 outbreak prediction with machine learning. *Available at SSRN 3580188*, 2020.

[3] Jesus R Artalejo, Antonis Economou, and Maria Jesus Lopez-Herrero. The stochastic seir model before extinction: Computational approaches. *Applied mathematics and computation*, 265:1026–1043, 2015.

[4] David Baud, Xiaolong Qi, Karin Nielsen-Saines, Didier Musso, Léo Pomar, and Guillaume Favre. Real estimates of mortality following covid-19 infection. *The Lancet infectious diseases*, 2020.

[5] Eran Bendavid, Bianca Mulaney, Neeraj Sood, Soleil Shah, Emilia Ling, Rebecca Bromley-Dulfano, Cara Lai, Zoe Weissberg, Rodrigo Saavedra, James Tedrow, et al. Covid-19 antibody seroprevalence in santa clara county, california. *MedRxiv*, 2020.

[6] Andrea L Bertozzi, Elisa Franco, George Mohler, Martin B Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of covid-19. *arXiv preprint arXiv:2004.04741*, 2020.

[7] Dimitris Bertsimas. Mit covidanalytics, May 2020. covidanalytics.io [Online; accessed 24-May-2020].

[8] Andrew Bray, Ka Wong, Christopher D Barr, Frederic Paik Schoenberg, et al. Voronoi residual analysis of spatial point process models with applications to california earthquake forecasts. *The Annals of Applied Statistics*, 8(4):2247–2267, 2014.

[9] Simon Cauchemez, Pierre-Yves Boëlle, Christl A Donnelly, Neil M Ferguson, Guy Thomas, Gabriel M Leung, Anthony J Hedley, Roy M Anderson, and Alain-Jacques Valleron. Real-time estimates in early detection of sars. *Emerging infectious diseases*, 12(1):110, 2006.

[10] Wen-Hao Chiang, Xueying Liu, and George Mohler. Hawkes process modeling of covid-19 with mobility leading indicators and spatial covariates. *medRxiv*, 2020.

[11] Gerardo Chowell and Hiroshi Nishiura. Transmission dynamics and control of ebola virus disease (evd): a review. *BMC medicine*, 12(1):196, 2014.

[12] Damian Clancy, Philip D O'Neill, et al. Bayesian estimation of the basic reproduction number in stochastic epidemic models. *Bayesian Analysis*, 3(4):737–757, 2008.

[13] Robert Alan Clements, Frederic Paik Schoenberg, and Danijel Schorlemmer. Residual analysis methods for space-time point processes with applications to earthquake forecast models in california. *The Annals of applied statistics*, pages 2549–2571, 2011.

[14] Robert Alan Clements, Frederic Paik Schoenberg, and Alejandro Veen. Evaluation of space–time point process models using super-thinning. *Environmetrics*, 23(7):606–616, 2012.

[15] Coronavirus COVID. Global cases by the center for systems science and engineering (csse) at johns hopkins university. *Coronavirus Resource Center. March*, 17:2020, 19.

[16] Daryl J Daley and D Vere Jones. *An Introduction to the Theory of Point Processes: Elementary Theory of Point Processes*. Springer, 2003.

[17] Andrew Daw and Jamol Pender. The queue-hawkes process: Ephemeral self-excitement. *arXiv preprint arXiv:1811.04282*, 2018.

[18] Michael Day. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in italian village. *BMJ*, 368:m1165, 2020.

[19] Chris Dye and Nigel Gay. Modeling the sars epidemic. *Science*, 300(5627):1884–1885, 2003.

[20] Neil D Evans, Lisa J White, Michael J Chapman, Keith R Godfrey, and Michael J Chappell. The structural identifiability of the susceptible infected recovered model with seasonal forcing. *Mathematical biosciences*, 194(2):175–197, 2005.

[21] CP Farrington, MN Kanaan, and NJ Gay. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics*, 4(2):279–295, 2003.

[22] Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. 2020.

[23] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Helen Coupland, Thomas A Mellan, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, Pablo NP Guzman, et al. Estimating the number of infections and the impact of non-pharmaceutical interventions on covid-19 in european countries: technical description update. *arXiv preprint arXiv:2004.11342*, 2020.

[24] Institute for Health Metrics and Evaluation (IHME). Imhe covid-19 predictions, May 2020. covid19.healthdata.org [Online; accessed 24-May-2020].

[25] Gianluca Frasso and Philippe Lambert. Bayesian inference in an extended seir model with nonparametric disease transmission rate: an application to the ebola epidemic in sierra leone. *Biostatistics*, 17(4):779–792, 2016.

[26] Keith R Godfrey and Michael J Chapman. Identifiability and indistinguishability of linear compartmental models. *Mathematics and Computers in Simulation*, 32(3):273–295, 1990.

[27] Joshua Seth Gordon, Robert Alan Clements, Frederic Paik Schoenberg, and Danijel Schorlemmer. Voronoi residuals and other residual analyses applied to csep earthquake forecasts. *Spatial Statistics*, 14:133–150, 2015.

[28] Nicholas C Grassly and Christophe Fraser. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology*, 6(6):477–487, 2008.

[29] Jia Gu, Han Yan, Ya Huang, Yu Zhu, Hao Sun, Xin Zhang, Yu Wang, Yumou Qiu, and Song Chen. Better strategies for containing covid-19 epidemics—a study of 25 countries via an extended varying coefficient seir model. *medRxiv*, 2020.

[30] Quanquan Gu. Ucla statistical machine learning lab, June 2020. covid19.uclaml.org [Online; accessed 19-June-2020].

[31] Youyang Gu. Covid-19 projections using machine learning, May 2020. covid19-projections.com [Online; accessed 24-May-2020].

[32] Alex Perkins Guido Espana. Notre dame mobility, June 2020. github.com/TAlexPerkins/covid19_NDmobility_forecasting [Online; accessed 19-June-2020].

[33] Alan G Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[34] Nicolas Hengartner and Paul Fenimore. Quantifying model form uncertainty of epidemic forecasting models from incidence data. *Online Journal of Public Health Informatics*, 10(1), 2018.

[35] P Hernández, C Pena, A Ramos, and JJ Gómez-Cadenas. A simple formulation of non-markovian seir. *arXiv preprint arXiv:2005.09975*, 2020.

[36] New Jersey Covid-19 Information Hub. How is the state using data to make decisions and slow thespread of covid-19, May 2020. covid19.nj.gov [Online; accessed 16-May-2020].

[37] Nicholas P Jewell, Joseph A Lewnard, and Britta L Jewell. Predictive mathematical models of the covid-19 pandemic: underlying principles and value of projections. *Jama*, 323(19):1893–1894, 2020.

[38] Park J. Kaplan A.M. and Schoenberg F. Nonparametric estimation of recursive point processes with application to mumps in pennsylvania, 2020. Submitted May 20, 2020.

[39] J Daniel Kelly, Junhyung Park, Ryan J Harrigan, Nicole A Hoff, Sarita D Lee, Rae Wannier, Bernice Selo, Mathias Mossoko, Bathe Njoloko, Emile Okitolonda-Wemakoy, et al. Real-time predictions of the 2018–2019 ebola virus disease outbreak in the democratic republic of the congo using hawkes point process models. *Epidemics*, 28:100354, 2019.

[40] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[41] Simas Kucinskas. Tracking r of covid-19. *Available at SSRN 3581633*, 2020.

[42] Los Alamos National Laboratory. Covid-19 confirmed and forecasted case data, May 2020. covid-19.bsvgateway.org [Online; accessed 24-May-2020].

[43] Phenyo E Lekone and Bärbel F Finkenstädt. Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics*, 62(4):1170–1177, 2006.

[44] Joseph Chadi Lemaitre, Kyra H Grantz, Joshua Kaminsky, Hannah R Meredith, Shaun A Truelove, Stephen A Lauer, Lindsay T Keegan, Sam Shah, Josh Wills, Kathryn Kaminsky, et al. A scenario modeling pipeline for covid-19 emergency planning. *medRxiv*, 2020.

[45] Simon A Levin and Viggo Andreasen. Mathematical models of infectious diseases. *Forefronts*, 2(8):4–6, 1986.

[46] PA W Lewis and Gerald S Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval research logistics quarterly*, 26(3):403–413, 1979.

[47] Alun L Lloyd. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1470):985–993, 2001.

[48] Thomas Lumley. Counting rare things is hard, April 2020. statschat.org.nz [Online; accessed 16-May-2020].

[49] David Marsan and Olivier Lengline. Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079, 2008.

[50] Stefano Merler, Marco Ajelli, Laura Fumanelli, and Alessandro Vespignani. Containing the accidental laboratory escape of potential pandemic influenza viruses. *BMC medicine*, 11(1):252, 2013.

[51] Sebastian Meyer, Leonhard Held, and Michael Höhle. Spatio-temporal analysis of epidemic phenomena using the r package surveillance. *arXiv preprint arXiv:1411.0416*, 2014.

[52] Matthew D. Parno Ian D. Detwiller Matthew W. Farthing William P. England Glover E. George Michael L. Mayo, Michael A. Rowland. Us army engineer research and development center, June 2020. github.com/reichlab/covid19-forecast-hub [Online; accessed 19-June-2020].

[53] George Mohler et al. Modeling and estimation of multi-source clustering in crime and security data. *The Annals of Applied Statistics*, 7(3):1525–1539, 2013.

[54] George Mohler, Frederic Schoenberg, Martin B Short, and Daniel Sledge. Analyzing the impacts of public policy on covid-19 transmission in indiana: The role of model and dataset selection.

[55] Pierre Montagnon. A stochastic sir model on a graph with epidemiological and population dynamics occurring over the same time scale. *Journal of mathematical biology*, 79(1):31–62, 2019.

[56] Yosihiko Ogata. On lewis' simulation method for point processes. *IEEE transactions on information theory*, 27(1):23–31, 1981.

[57] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27, 1988.

[58] Yosihiko Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

[59] Yosihiko Ogata et al. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(1):243–261, 1978.

[60] Philip D O'Neill. Introduction and snapshot review: relating infectious disease transmission models to data. *Statistics in medicine*, 29(20):2069–2077, 2010.

[61] Dave Osthus, Kyle S Hickmann, Petruţa C Caragea, Dave Higdon, and Sara Y Del Valle. Forecasting seasonal influenza with a state-space sir model. *The annals of applied statistics*, 11(1):202, 2017.

[62] Marie V Ozanne, Grant D Brown, Jacob J Oleson, Iraci D Lima, Jose W Queiroz, Selma MB Jeronimo, Christine A Petersen, and Mary E Wilson. Bayesian compartmental model for an infectious disease with dynamic states of infection. *Journal of applied statistics*, 46(6):1043–1065, 2019.

[63] An Pan, Li Liu, Chaolong Wang, Huan Guo, Xingjie Hao, Qi Wang, Jiao Huang, Na He, Hongjie Yu, Xihong Lin, et al. Association of public health interventions with the epidemiology of the covid-19 outbreak in wuhan, china. *Jama*, 2020.

[64] Junhyung Park, Adam W Chaffee, Ryan J Harrigan, and Frederic Paik Schoenberg. A non-parametric hawkes model of the spread of ebola in west africa, 2018.

[65] Junhyung Park, Frederic Paik Schoenberg, Andrea L Bertozzi, and P Jeffrey Brantingham. Investigating clustering and violence interruption in gang-related violent crime data using spatial-temporal point processes with covariates. 2019.

[66] Spencer Carran Sarah Cobey Katelyn Gostic Lauren McGough Sylvia Ranjeva Frank Wen Phil Arevalo, Ed Baskerville. Forecasting sars-cov-2 dynamics for the state of illinois, June 2020. github.com/cobeylab/covid_IL [Online; accessed 19-June-2020].

[67] Gergo Pinter, Imre Felde, Amir Mosavi, Pedram Ghamisi, and Richard Gloaguen. Covid-19 pandemic prediction for hungary; a hybrid machine learning approach. *A Hybrid Machine Learning Approach (May 2, 2020)*, 2020.

[68] Jakob Gulddahl Rasmussen. Bayesian inference for hawkes processes. *Methodology and Computing in Applied Probability*, 15(3):623–642, 2013.

[69] Alex Reinhart et al. A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, 33(3):299–318, 2018.

[70] Marian-Andrei Rizoiu, Swapnil Mishra, Quyu Kong, Mark Carman, and Lexing Xie. Sir-hawkes: linking epidemic models and hawkes processes to model diffusions in finite populations. In *Proceedings of the 2018 World Wide Web Conference*, pages 419–428, 2018.

[71] Kimberlyn Roosa and Gerardo Chowell. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theoretical Biology and Medical Modelling*, 16(1):1, 2019.

[72] Jay Boice Ryan Best. Where the latest covid-19 models think we're headed — and why they disagree, May 2020. fivethirtyeight.com [Online; posted 24-May-2020].

[73] Timothy Sauer, Tyrus Berry, Donald Ebeigbe, Michael M Norton, Andrew Whalen, and Steven J Schiff. Identifiability of infection model parameters early in an epidemic. *medRxiv*, 2020.

[74] Frederic Paik Schoenberg. A note on the consistent estimation of spatial-temporal point process parameters. *Statistica Sinica*, pages 861–879, 2016.

[75] Frederic Paik Schoenberg. Nonparametric estimation of variable productivity hawkes processes. *arXiv preprint arXiv:2003.08858*, 2020.

[76] Frederic Paik Schoenberg, Marc Hoffmann, and Ryan J Harrigan. A recursive point process model for infectious diseases. *Annals of the Institute of Statistical Mathematics*, 71(5):1271–1287, 2019.

[77] Danijel Schorlemmer, Maximilian J Werner, Warner Marzocchi, Thomas H Jordan, Yosihiko Ogata, David D Jackson, Sum Mak, David A Rhoades, Matthew C Gerstenberger, Naoshi Hirata, et al. The collaboratory for the study of earthquake predictability: achievements and priorities. *Seismological Research Letters*, 89(4):1305–1313, 2018.

[78] Neeraj Sood, Paul Simon, Peggy Ebner, Daniel Eichner, Jeffrey Reynolds, Eran Bendavid, and Jay Bhattacharya. Seroprevalence of sars-cov-2–specific antibodies among adults in los angeles county, california, on april 10-11, 2020. *JAMA*, 2020.

[79] Balaji Srinivasan. Peer review of "covid-19 antibodyseroprevalence in santa clara county, california", April 2020. medium.com/@balajis [Online; accessed 16-May-2020].

[80] Gabriele Stabile and Giovanni Luca Torrisi. Risk processes with non-stationary hawkes claims arrivals. *Methodology and Computing in Applied Probability*, 12(3):415–429, 2010.

[81] Chandini Jain Vishal Tomar. Auquan data science, June 2020. covid19-infection-model.auquan.com [Online; accessed 19-June-2020].

[82] Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*, 160(6):509–516, 2004.

[83] Eric Walter and Yves Lecourtier. Unidentifiable compartmental models: what to do? *Mathematical biosciences*, 56(1-2):1–25, 1981.

[84] Helen J Wearing, Pejman Rohani, and Matt J Keeling. Appropriate models for the management of infectious diseases. *PLoS medicine*, 2(7), 2005.

[85] Lee Worden, Rae Wannier, Nicole A Hoff, Kamy Musene, Bernice Selo, Mathias Mossoko, Emile Okitolonda-Wemakoy, Jean Jacques Muyembe-Tamfum, George W Rutherford, Thomas M Lietman, et al. Real-time projections of epidemic transmission and estimation of vaccination impact during an ebola virus disease outbreak in the eastern region of the democratic republic of congo. *arXiv preprint arXiv:1811.01175*, 2018.

[86] Teresa Yamana, Sen Pei, and Jeffrey Shaman. Projection of covid-19 cases and deaths in the us as individual states re-open may 4, 2020. *medRxiv*, 2020.

[87] Ah Sung Yang. *Modeling the Transmission Dynamics of Pertussis Using Recursive Point Process and SEIR model*. PhD thesis, UCLA, 2019.

[88] Chong You, Yuhao Deng, Wenjie Hu, Jiarui Sun, Qiushi Lin, Feng Zhou, Cheng Heng Pang, Yuan Zhang, Zhengchao Chen, and Xiao-Hua Zhou. Estimation of the time-varying reproduction number of covid-19 outbreak in china. *International Journal of Hygiene and Environmental Health*, page 113555, 2020.

[89] Baichuan Yuan. Multivariate hawkes processes for real-time covid-19 death forecasting, 2020. Preprint submitted to Journal of LATEX Templates.

[90] J Douglas Zechar, Danijel Schorlemmer, Maximilian J Werner, Matthew C Gerstenberger, David A Rhoades, and Thomas H Jordan. Regional earthquake likelihood models i: First-order results. *Bulletin of the Seismological Society of America*, 103(2A):787–798, 2013.

[91] Jiancang Zhuang, Yosihiko Ogata, and David Vere-Jones. Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth*, 109(B5), 2004.