

Detection of surges of SARS-Cov-2 using nonparametric Hawkes models

Sophie Phillips¹, George Mohler², and Frederic Schoenberg^{1,3}.

¹ Department of Statistics, 8142 Math-Science Building, UCLA, Los Angeles, CA 90095-1554, USA.

² Department of Computer Science, Boston College, 245 Beacon Street, Boston, MA 02116.

³ Corresponding author, frederic@stat.UCLA.edu .

Abstract. Hawkes point process models have been shown to forecast the number of daily new cases of epidemic diseases, including SARS-CoV-2 (Covid-19), with high accuracy. Here, we explore how accurately Hawkes models forecast surges of Covid-19 in the United States. We use Hawkes models to estimate the effective reproduction rate R_t and transmission density parameters for Covid-19 case counts in each of the 50 United States, then forecast R_t in future weeks with simple exponential smoothing. A classifier based on $R_t > x$ is applied to predict upcoming surges in cases each week from August 2020 to December 2021, using only data available up to that week. At false alarm rates below 5%, the forecasts based on R_t are correct more often than forecasts based on smoothing the raw case count data, achieving a maximum accuracy of 88% with $R_t > 1.75$. The optimal decision boundary uses a combination of R_t and observed data.

Keywords: Covid-19, effective reproduction rate, epidemiological forecasting, forecasting, Hawkes model, point process.

1 Introduction

Hawkes point process models and their slight variants such as the epidemic-type aftershock sequence (ETAS) model (Ogata 1988, Ogata 1998), HawkesN (Rizoiu et al. 2018, Bertozzi et al. 2020, Chiang et al. 2022), and the recursive model (Schoenberg et al. 2019) have proven to be useful in modeling the spread of a variety of epidemic diseases such as Ebola (Harrigan et al. 2019, Park et al. 2022), Chlamydia (Schoenberg 2022), SARS (Wallinga and Teunis 2004,

Cauchemez et al. 2006), measles (Farrington et al. 2003), Meningococcal disease (Meyer et al. 2012), Rocky Mountain Spotted Fever (Schoenberg et al. 2019), and SARS-CoV-2 (Covid-19) (Mohler et al. 2020, Schoenberg 2023, Phillips and Schoenberg 2024). Kresin et al. (2022) compared the accuracy of Hawkes models to that of compartmental models, such as SEIR, across various infectious diseases, and found that Hawkes models offered substantially higher accuracy than compartmental models, with errors in forecast case counts approximately 20-30% smaller on average than compartmental models in most cases.

Here, we explore how accurately Hawkes models can forecast surges of Covid-19 in the United States. One way to evaluate the model’s forecasting efficacy is to consider its accuracy in estimating the effective reproduction rate, which is an important and widely used indicator for the transmission intensity of an epidemic and an early warning signal for disease emergence (Anderson 1991, Southall et al. 2020). This parameter, denoted R_t , represents the mean number of secondary cases generated by a new infection at time t (Cori et al. 2013). In theory, R_t is as an early warning statistic for outbreaks, with values above 1 indicating an emerging disease (Southall et al. 2020). Accurate estimates of R_t can be used to inform surveillance efforts and assess the impact of interventions (Eichner et al. 2003, Ferguson et al. 2006, Domenech et al. 2018, Bertozzi et al. 2020, Chiang et al. 2022).

R_t can be estimated retrospectively or in real time. Wallinga and Teunis (2004) use an expectation maximization method to reconstruct the transmission chains and count the estimated number of secondary cases per individual. This method is only suitable for retrospective analysis as the estimates of R_t at a given time rely on subsequent case records, although modifications for real-time estimation have been proposed (Cauchemez et al. 2006). Alternatively, Cori et al. (2013) use Bayesian inference to estimate a posterior distribution for the reproduction rate, assuming a gamma distribution on R_t to obtain an analytical expression for the posterior. Hawkes models represent the rate of infections as a branching process, allowing for non-parametric estimation of the transmission density and reproduction number (Bertozzi et al. 2020). While Wallinga and Teunis (2004) and Cori et al. (2013) also use point processes to estimate R_t , we avoid distributional assumptions on the parameter, and use a simpler least-squares estimation method. We also present an alternating least-squares procedure for estimating the transmission parameters.

In this paper, we examine the use of the estimated reproduction rate for forecasting outbreaks in incidence, using transmission parameters estimated from a discrete Hawkes model with state-level Covid-19 incidence data from the first six months of 2020. Case count data are incrementally added week by week and R_t is sequentially re-estimated. We assess the forecast accuracy by observing how frequently the estimated reproduction rate exceeding a threshold precedes surges in cases, and compare with forecasts based solely on prior case counts.

The structure of the paper is outlined as follows. Statewide data on Covid-19 case counts data are described in Section 2, followed by a description of the model and estimation methods in section 3. These methods are applied to estimate the transmission parameters and reproduction rate for all 50 states, and the resulting estimates are further analyzed to see how well they could have forecast surges in incidence. Results are detailed in Section 4, and a discussion is given in Section 5.

2 Data Description

Daily counts of new Covid-19 cases were obtained for each state from the Johns Hopkins Covid-19 Data Repository which sourced case data from the CDC and state public health departments (Dong et al. 2020, JHU CSSE 2023). The case counts include both cases confirmed by a positive PCR test and probable cases defined by a combination of antigen testing and epidemiological criteria (CDC 2021). Positive tests are dated either according to the date of report or date of sample collection (CDC 2021). Daily case counts for each state were obtained for the period 2/1/2020 to 12/31/2021, covering a total of 100 weeks. We ended the study period at the end of 2021 as many states shifted to less frequent reporting schedules in 2022, causing irregularities in the case count data (Dong et al. 2022). We model here only confirmed and suspected cases, though many Covid-19 infections are asymptomatic or unrecorded (Ma et al. 2021). The data used are publicly available from <https://github.com/CSSEGISandData/COVID-19>.

During the study period, most states experienced two major surges in reported cases from November to March in both 2020 and 2021, though the timing, duration, and magnitude of these surges varied substantially between states. The reported case counts for California, Texas, Vermont, and Hawaii are shown in Figure 1 as examples. California, Texas, and Florida experienced the highest total number of new cases during the study period, with 5.3, 4.5, and 3.9 million cases respectively. North Dakota and Alaska experienced the highest per capita incidence during this period, while Hawaii and Vermont experienced both the lowest per capita incidence as well as the least total incidence, with approximately 99,900 and 60,200 total cases, respectively, during this study period. Hawaii and Vermont experienced modest increases in reported case counts during 2020, whereas California and Texas had significant increases in both the Summer and Winter of 2020. Sudden spikes in case count records, such as those in Texas in early 2021 and in Vermont in late 2021, appear in several states throughout the study period.

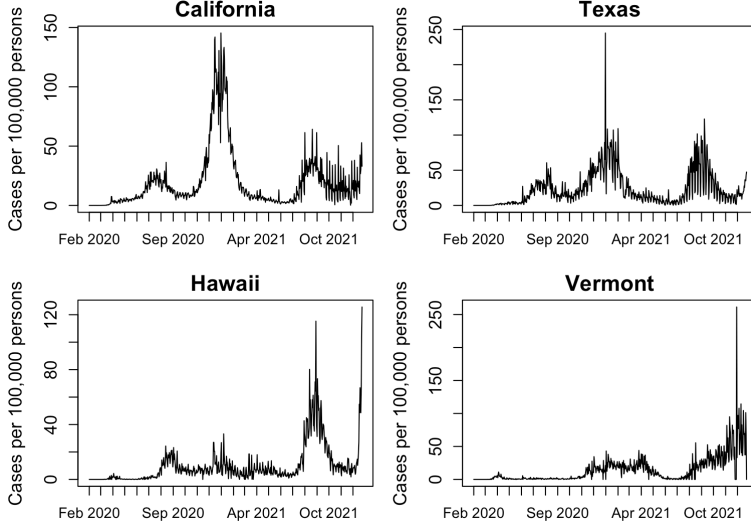


Figure 1: Daily reported cases per 100,000 persons in California, Texas, Hawaii, and Vermont. Population estimates are from the 2020 decennial census.

3 Methods

3.1 Discrete Hawkes model

The Hawkes self-exciting point process (Hawkes 1971) models the rate of infections $\lambda(t)$ conditional on previous infection times t_i as

$$\lambda(t) = \mu + \sum_{t_i < t} R(t_i)g(t - t_i) \quad (1)$$

The transmission density $g(t)$ and productivity function $R(t)$ are directly analogous to those in epidemiological models, and accommodate a variety of parametric and non-parametric forms. The parameter μ represents the rate of new cases immigrating into the location of interest, and is typically assumed to remain constant over time, though it can also be modeled as a function of time and/or space.

In epidemiological applications, Hawkes models have demonstrated superior forecasting accuracy and flexibility relative to traditional compartmental models (Bertozzi et al. 2020, Mohler et al. 2020, Chiang et al. 2022, Kresin et al. 2022). Following Phillips and Schoenberg (2024), we use a discrete approximation of the continuous time Hawkes model (1) to model the case count rather than rate, as cases are reported as daily counts rather than precise event times. Kirchner (2016) shows this binned form of Hawkes model corresponds to an autoregressive model. We fit the model by finding parameters that minimize

the squared error between the observed and modeled daily case counts, as in Schoenberg (2023).

3.2 Non-parametric estimation of $R(t)$

We fit the reproduction rate non-parametrically as a step function over weekly intervals, thus estimating a total of 100 parameters for each state over the study period. Each estimate represents the expected number of infections directly transmitted by each case reported that week. The least-squares estimates for $R = (R_1 \dots R_{99})$ can be obtained analytically using the method proposed for continuous-time Hawkes processes in Schoenberg (2022) and for discrete case counts in Phillips and Schoenberg (2024). We estimate the final estimate R_{100} with exponential smoothing. Below, we describe the method for the general case of estimating $R(t)$ over T days using a window length τ , yielding T/τ parameters to estimate.

We seek to minimize the error in estimating the case count $N(t)$ for day t :

$$\sum_{t=2}^T [N(t) - \mu - \sum_{s=1}^{t-1} R(s)g(t-s)N(s)]^2. \quad (2)$$

Let N be the vector of observed daily counts excluding the first day, and G be the $(T-1) \times (T/\tau-1)$ matrix with entries $G_{ij} = \sum_{s=(j-1)\tau+1}^{j\tau} g(i-s)N(s)$. G_{ij} represents the total expected number of new cases on day i directly infected from cases occurring in time interval j . With this notation, (2) can be expressed as:

$$GR = N \quad (3)$$

A column of 1s can be appended to G to estimate the baseline intensity μ as an intercept. When G is invertible, the estimates $\hat{R}$ can be computed quickly and directly using standard least-squares solvers. However, the matrix G is typically sparse and ill-conditioned as the transmission $g(t)$ decays over relatively short intervals. As in Phillips and Schoenberg (2024), to avoid issues with matrix singularity and improve the stability of estimates, we regularize the system in (3) by adding a penalty on the solution norm:

$$\hat{R} = \min_{R_1 \dots R_{T/\tau-1}} \|N - \mu - GR\|_2^2 + \rho \|R\|_2^2 \quad (4)$$

The regularization parameter ρ controls the degree of smoothing. We incrementally increase ρ until all estimates are non-negative, searching over values of $10^{-k}\sigma_1(G)$, where σ_1 is the largest singular value of G . The standard errors for the regularized least-squares estimates can be calculated as:

$$\begin{aligned} \text{Var}(\hat{R}) &= s^2(G^T G + \rho I)^{-1}, \\ s^2 &= \frac{1}{T-p} \|N(t) - \hat{N}(t)\|^2, \end{aligned} \quad (5)$$

where $p = T/\tau$. The covariance matrix (5) can be calculated efficiently using the Cholesky factorization of $G^T G + \rho I$.

In order to emulate real-time forecasting, we apply simple exponential smoothing to forecast $R_{T/\tau}$ from the least-squares estimates $R_1 \dots R_{T/\tau-1}$ calculated using (4). This method forecasts values by applying a weighted average of previous estimates, with weights that decay exponentially at a rate of α . We use $\alpha = 0.5$ here. The variance of the forecast value $R_{T/\tau}$ is calculated based on the residual variance from earlier forecasts (Brown 1963).

3.3 Estimation of transmission density parameters

Transmission parameters are estimated for each state using data from 2/1/2020 to 7/31/2020. Since the transmission time density for statewide Covid-19 data was shown in Schoenberg (2023) to be approximately normal, we model the transmission time density parametrically using a truncated normal distribution: $g \sim N_{\geq 0}(\nu, \sigma)$, and estimate the mean ν and standard deviation σ for each state.

The reproduction rate $R(t)$ is fit simultaneously with the transmission parameters as follows. Starting from initial estimates $R^{(0)} = (R_1^{(0)} \dots R_{26}^{(0)})$ and $\mu^{(0)}$, a non-linear optimizer determines ν and σ that minimize the squared error in daily cases. For simplicity, we initialize $R^{(0)}$ as a constant vector of 0.5 and $\mu^{(0)} = N(t)$, and use the BFGS routine in R's `optim`. R and μ are then calculated analytically based on (4), with the matrix G formed using the current estimates of ν and σ and augmented to include an intercept. The final estimate of R is estimated with exponential smoothing as detailed in the previous section. We iterate between these steps until the estimates of ν , σ , μ , and R converge.

A penalty is added to the objective function in the non-linear optimization step to constrain the support of the transmission density $g(t)$ to a pre-specified interval $[0, s]$. This penalty was found to improve convergence and stability of estimates, resembling the effect of Bayesian priors. Here, we set the upper bound on the time between reported cases to $s = 21$ days. We additionally penalize values of the transmission standard deviation σ below 1 to prevent overfitting. Although the L-BFGS-B algorithm allows for constrained optimization, estimates often converge to the boundary, perhaps due to the irregular loss function. Instead, the unconstrained BFGS algorithm is used with penalty $C(\nu, \sigma)$:

$$C(\nu, \sigma) = \left[\sum_{i=1}^s g(i; \nu, \sigma) \right]^{-2} + \sigma^{-2}$$

$$\nu^{(i)}, \sigma^{(i)} = \min_{\nu, \sigma} C(\nu, \sigma) RMSE(\nu, \sigma, R^{(i-1)}),$$

so that the loss is inflated for $\hat{\sigma} < 1$ or for parameter estimates placing mass outside $[0, s]$.

3.4 Forecasting surges using R_t

To mimic estimation in real-time, we sequentially incorporate case data week by week from 8/1/2020 to 12/25/2021, re-estimating $R(t)$ using only data up to the latest available date for a total of 72 forecast periods. For any time t , the estimate of $R(t)$ is used to forecast upcoming surges in cases just after time t . We define a surge as the times when the smoothed weekly case count increases by at least 10% or by at least 1,000 cases for more than three consecutive weeks. This criterion effectively captures major outbreaks.

We use $R_t > x$, where x is some threshold, as a classifier to predict whether a surge, as defined above, will occur either one or two weeks after time t . This method is compared to an “incidence-based” classifier based on the percent change in weekly incidence: $\frac{\sum_{t \in w} N(t) - N(t-7)}{\sum_{t \in w} N(t-7)} > y$, where $N(t)$ is the case count at time t and w represents a given week. It is also compared to a “random guesser” null model that simply predicts surge times completely at random. We evaluate the performance of each classifier using a receiver operator characteristic (ROC) curve which reports the true positive and false rate across a range of thresholds x, y .

A linear support vector machine (SVM) is used to determine the line that maximally separates weeks during surges from non-surge weeks based on the previous week’s predicted R_t and change in incidence (Boser et al. 1992). The SVM is trained to predict whether the 72 forecast periods in each state are during a surge using the observed and estimated measures from the previous week.

4 Results

4.1 Transmission density estimates

Estimates for transmission time parameters converged for all states in an average of 9 iterations using alternating least-squares, as described in Section 3.3, applied to the first six months of Covid-19 case data for each state. Across the 50 states, the root mean squared error (RMSE) in predicted cases ranged from 6.5 cases/day in Vermont to 958 cases/day in Texas, with mean 184 cases. Louisiana had the highest population-adjusted RMSE of 9.4 cases/day per 100,000 persons.

As examples, the observed and modeled case counts for Texas and Louisiana are shown in Figure 2. The high RMSE in these states is due to several spikes in cases that are smoothed over by the model. Initially, there is significant uncertainty in $R(t)$, with estimates exceeding 4 in the early stage of the pandemic. Standard errors are large in both Texas and Louisiana when fewer than 1,200 cases had been reported, but decrease substantially later in the pandemic. Eventually, $R(t)$ fluctuates around 1, with estimates above 1 corresponding to periods of increasing incidence.

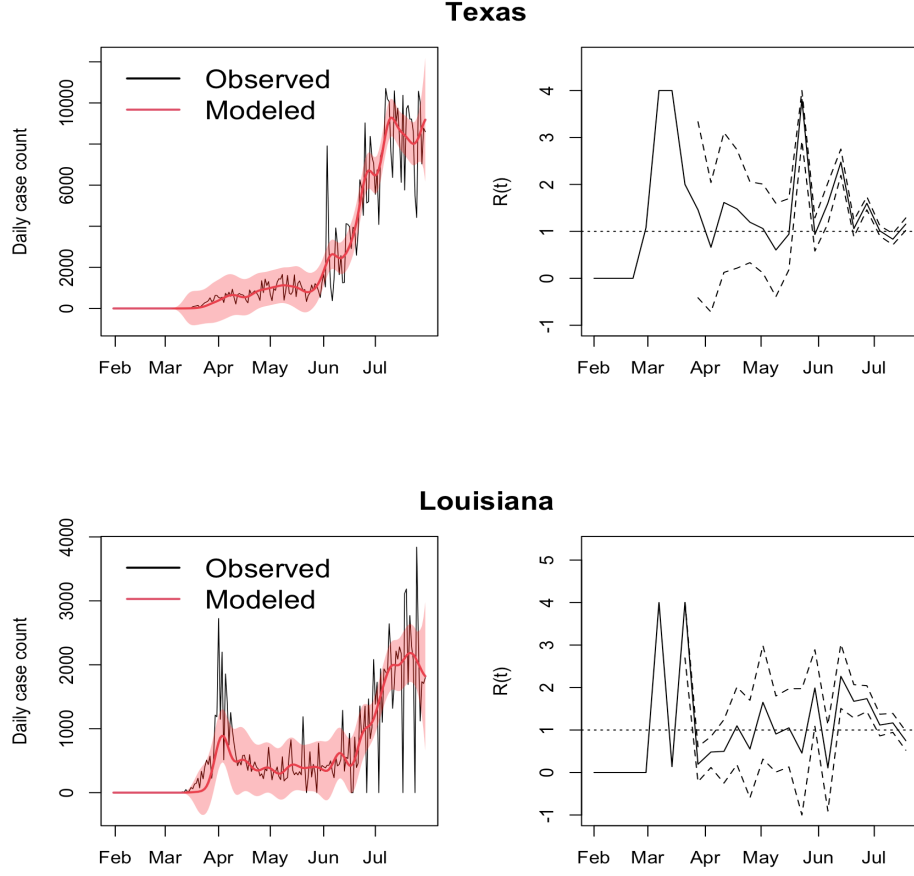


Figure 2: Left panel: Observed (black) and modeled (red) daily case counts in 2020. The ribbon indicates the range of estimated cases using the estimated reproduction rate $\pm$ one standard deviation, calculated from (5). Right panel: The reproduction rate $R(t)$ estimated based on (4), with dashed lines indicating the interval $\pm$ one standard deviation around the estimates. Uncertainty estimates are not displayed until the reported case count exceeds 1,200. For reference, a dotted line indicates $R = 1$.

The estimated truncated normal transmission densities for all 50 states are shown in Figure 3. The estimates for most states are centered at a mean of 10.5 days with standard deviation 3.75 days. New York, New Jersey, and Pennsylvania have noticeably lower transmission means than the other states, with means of 6.5, 6.7, and 7.0 days respectively. These states experienced initial outbreaks in April 2020, which is several months earlier than the other states.

The estimated background rate of immigration μ was less than 5 cases/day in 38 of the states. New York is a notable exception with estimated rate of 158 cases/day, followed by Colorado and Ohio with rates of 67 and 45 cases/day, respectively.

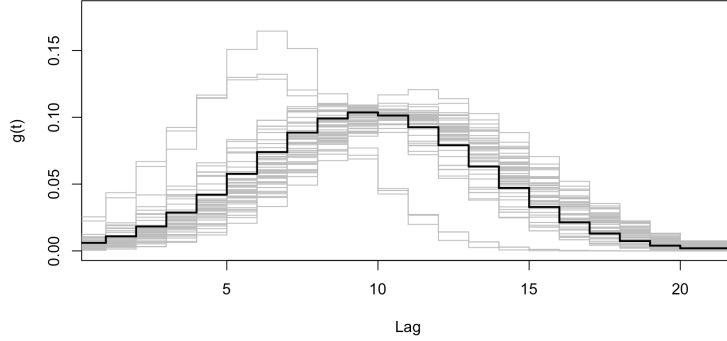


Figure 3: Estimated transmission densities for all 50 states (gray lines), and averaged across all states (thick black line). The transmission density here reflects the time between reported cases and is assumed to follow a truncated normal distribution.

4.2 Forecasting surges in cases

Figure 4 shows ROC curves for forecasts of surges, averaged over all 50 states, using 3 methods: Hawkes estimates $\hat{R}_t > x$, an “incidence-based” model that uses the percent change in weekly incidence, and a random guesser that predicts surge weeks at random. The ROC curves for individual states using the Hawkes model are also shown in Figure 4. Based on the averaged ROC curves, at a false positive rate of 1%, the Hawkes model achieves a true positive rate of 37%, compared to 16% for the incidence-based model and 2.8% for the random guesser. These rates correspond to a decision boundary of $R_W > 1.75$, and change in incidence $> 59\%$. With this false positive rate, the Hawkes model correctly predicts an average of 4 more surge weeks per state than the incidence-based model. The incidence-based model captures more true positives than the Hawkes model at false positive rates above 5%. The Hawkes model achieves its highest accuracy across all states of 88% using the decision boundary $R_W > 1.75$ while the highest average accuracy from incidence based model is 82% using percent change $> 55\%$.

Four states stand out with significantly lower true positive rates in Figure 4. These states are Kansas, South Dakota, Vermont, and Washington, which experienced sustained periods of high incidence that did not meet our surge

criteria, unlike most states where case counts quickly decreased after peaking. To illustrate this, Figure 5 shows the predicted surge weeks for Kansas and Vermont, which had the lowest true positive rates among all 50 states. Vermont had a long period of sustained incidence in the summer of 2020 that was not correctly forecast by the Hawkes model, while Kansas had a similar period in the summer of 2021. For the other 46 states, the Hawkes model forecast very accurately: at a false positive rate of 5%, the average true positive rate of the Hawkes model among these states is 65%.

Figure 6 shows the predicted R_t and change in incidence for each of the 72 forecast weeks across all 50 states. A linear SVM is used to determine the optimal separating line between the surge and non-surge weeks based on the previous week's change in incidence and predicted R_t . This optimal boundary achieves 81% accuracy at forecasting surge weeks, with a 5.3% false positive and 55% true positive rate.

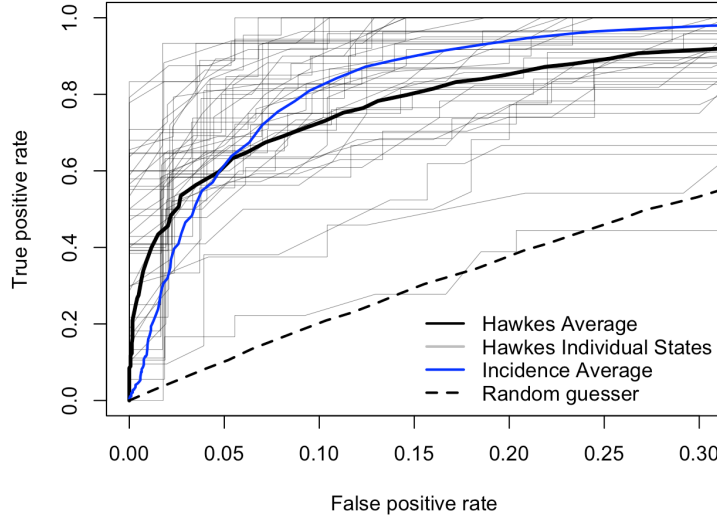


Figure 4: ROC curves comparing three methods for forecasting surges in cases: R_t estimates from a Hawkes model (solid black), observed changes in incidence (blue), and a random predictor (dashed black). The true and false positive rates are averaged across all 50 states, with individual state curves shown for the Hawkes model (gray).

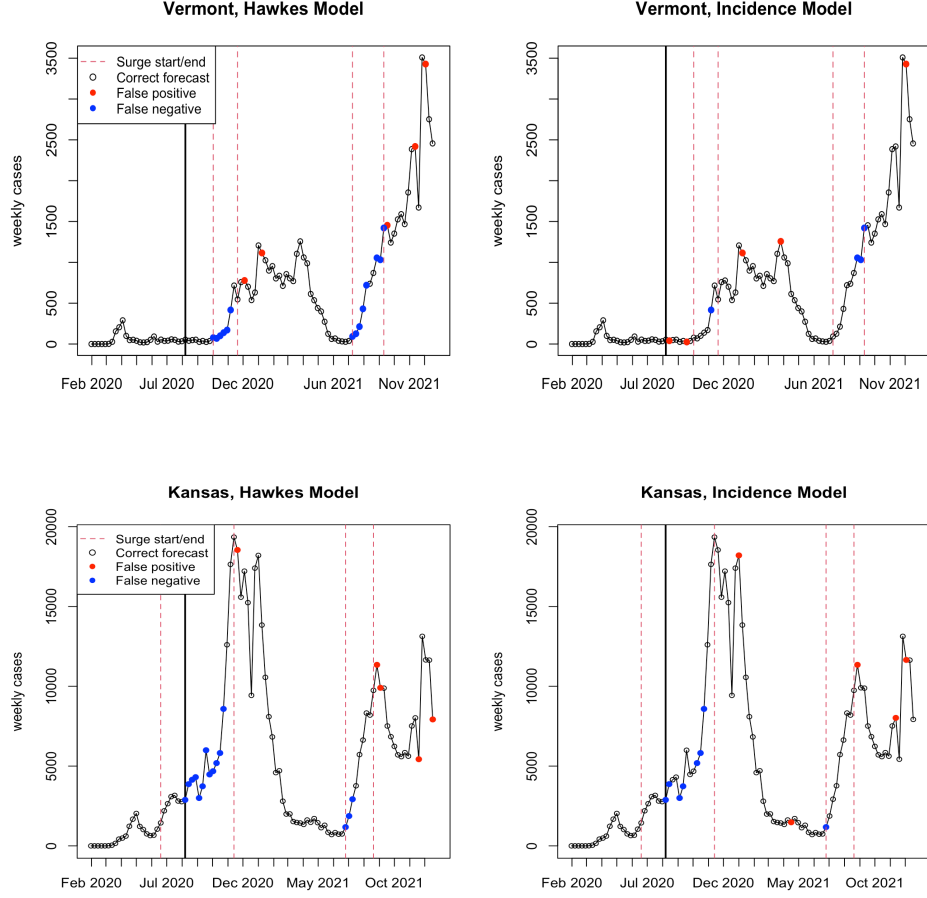


Figure 5: Predicted surge weeks for the Hawkes model (left) and incidence-based model (right) at a false alarm rate of 10%. The corresponding decision boundaries are $R_t > 1.4$ and change in incidence $> 41\%$ for Vermont and $R_t > 1.67$ and change in incidence $> 19\%$ for Kansas. Red dots indicate false alarms while blue dots indicate surge weeks that were missed by the model. The training period used to fit transmission parameters is indicated by the thick black line, while red dashed lines indicate true surge periods.

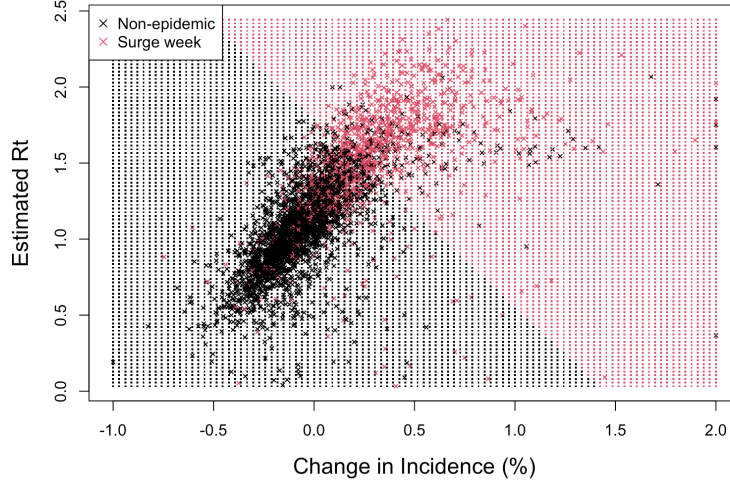


Figure 6: Linear decision boundary between surge weeks (red) and non-surge weeks (black) determined by a linear SVM using data from all 50 states. Each marked point (x) indicates the percent change in incidence and the estimated R_t from the previous week, colored by whether the following week was during a surge.

5 Discussion

The Hawkes model provides a simple analytic formula for both real-time and retrospective estimation of the effective reproduction number R_t . Estimates can be calculated without distributional assumptions on R_t and accommodate a wide range of transmission densities. At a false positive rate of 1%, the real-time Hawkes model had a 37% true positive rate over all 50 states, compared to a 16% true positive rate for an incidence-based model. This represents the correct detection of 4 more surges per state, on average, compared to forecasts based solely on observed incidence during the 72 forecast periods in this study (Figure 4).

We estimate considerable variation in R_t over the study period (Figure 2). Many factors such as seasonality, changes in social contacts, emergence of new viral variants with higher transmissibility, control measures, vaccinations, and growing herd immunity could influence R_t . Future work should consider inference methods to disentangle these influences. Estimates of R_t are also sensitive to changes in case definitions and testing rates, highlighting the need for structured surveillance programs for accurate estimation and inference.

The estimates of mean transmission time here (Figure 3) are somewhat higher than in some other studies of Covid-19 transmission (Lauer et al. 2020, Schoenberg 2023). The transmission time distribution estimated here reflects the intervals between reported cases, delayed from the actual infection times by the incubation period, which Lauer et al. (2020) estimated to have a median of 5.1 days, and potentially further lagged by delays in testing after exposure. For simplicity, we assumed the transmission time distribution was fixed over the study period, but future work could extend this by using different transmission parameters in different time periods.

The least-squares uncertainty estimates (5) are unstable at the start of the pandemic when fewer than 1,200 cases were reported, but decrease significantly once more data are available. This instability is due to the sparsity in the initial columns of the transmission matrix whose inverse is used to calculate standard errors. Future work should aim to somehow adjust standard errors at the beginning of an outbreak when case counts are low and at the end of the observation window to address uncertainty from right censoring. Additionally, methods to propagate uncertainty in the transmission parameters should be considered in future work.

Acknowledgement.

This material is based upon work supported by the National Science Foundation under grant number DMS-2124433.

6 References

- Anderson, R. (1991). *Infectious diseases of humans: dynamics and control*. Cambridge University Press.
- Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B., & Sledge, D. (2020). The challenges of modeling and forecasting the spread of Covid-19. *Proceedings of the National Academy of Sciences*, 117(29), 16732–16738. DOI: 10.1073/pnas.2006520117.
- Boser, B.E., Guyon, I.M., & Vapnik, V.N. (1992). A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, Pittsburgh, 27-29 July 1992, 144-152. DOI: 10.1145/130385.130401.
- Brown, R.G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series*. Prentice-Hall.
- Cauchemez S., Boelle P.Y., Donnelly C.A., Ferguson N.M., Thomas G., Leung G.M., Hedley A.J., Anderson R.M., & Valleron A.J. (2006). Real-time

- estimates in early detection of SARS. *Emerging infectious diseases*, 12(1), 110. DOI: 10.3201/eid1201.050593.
- Centers for Disease Control (CDC) (2021). 2021 case definition. <https://ndc.services.cdc.gov/case-definitions/coronavirus-disease-2019-2021/>, last accessed 9/12/24.
- Chiang W.H., Liu X., & Mohler G. (2022). Hawkes process modeling of Covid-19 with mobility leading indicators and spatial covariates. *Int J Forecast*, 38(2), 505-520. DOI: 10.1016/j.ijforecast.2021.07.001.
- Cori A., Ferguson N.M., Fraser C., & Cauchemez S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol*, 178(9), 1505-12. DOI: 10.1093/aje/kwt133.
- Domenech de Cells, M., Magpantay, F., King, A. & Rohani, P. (2018). The impact of past vaccination coverage and immunity on pertussis resurgence. *Sci. Transl. Med*, 10(434). DOI: 10.1126/scitranslmed.aaj1748.
- Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track Covid-19 in real time. *The Lancet Infectious Diseases*, 20(5), 533-534. DOI: 10.1016/S1473-3099(20)30120-1.
- Dong, E., Ratcliff, J., Goyea, T.D., Katz, A., Lau, R., Ng, T.K., Garcia, B., Bolt, E., Prata, S., Zhang, D., Murray, R.C., Blake, M.R., Du, H., Ganjikhloo, F., Ahmadi, F., Williams, J., Choudhury, S., & Gardner, L.M. (2022). The Johns Hopkins University Center for Systems Science and Engineering Covid-19 Dashboard: data collection process, challenges faced, and lessons learned. *The Lancet Infectious Diseases*. 22(12), e370-e376. DOI: 10.1016/S1473-3099(22)00434-0.
- Eichner, M., & Dietz, K. (2003). Transmission Potential of Smallpox: Estimates Based on Detailed Data from an Outbreak. *American Journal of Epidemiology*, 158(2), 110-117. DOI: 10.1093/aje/kwg103.
- Farrington C.P., Kanaan M.N., & Gay N.J. (2003). Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics*, 4(2), 279-95. DOI: 10.1093/biostatistics/4.2.279.
- Ferguson N.M., Cummings D.A., Fraser C., Cajka J.C., Cooley P.C., & Burke D.S. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442(7101), 448-52. DOI: 10.1038/nature04795.
- Harrigan, R., Mossoko, M., Okitolonda-Wemakoy, E., Schoenberg, F.P., Hoff, N., Mbala, P., Wannier, S.R., Lee, S.D., Ahuka-Mundeke, S., Smith, T.B., Selo, B., Njokolo, B., Rutherford, G., Rimoin, A.W., Tamfum, J.J.M., & Park, J. (2019). Real-time predictions of the 2018-2019 Ebola virus disease outbreak in the Democratic Republic of Congo using Hawkes point process models. *Epidemics*, 28, 100354. DOI: 10.1016/j.epidem.2019.100354.
- Hawkes, A. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1), 83-90. DOI: 10.1093/biomet/58.1.83.
- Johns Hopkins Center for Systems Science and Engineering (JHU CSSE). (2023). <https://github.com/CSSEGISandData>, last accessed 9/9/24.
- Kirchner, M. (2016). Hawkes and INAR(∞) processes. *Stochastic Processes*

- and their Applications, 26(8), 2494–2525. DOI: 10.1016/j.spa.2016.02.008.
- Kresin, C., Schoenberg, F.P., & Mohler, G. (2022). Comparison of Hawkes and SEIR Models for the Spread of Covid-19. *Advances and Applications in Statistics*, 74, 83-106. DOI: 10.17654/0972361722019.
- Lauer S.A., Grantz K.H., Bi Q., Jones F.K., Zheng Q., Meredith H.R., Azman A.S., Reich N.G., & Lessler J. (2020). The Incubation Period of Coronavirus Disease 2019 (Covid-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med.*, 172(9), 577-582. DOI: 10.7326/M20-0504.
- Ma, Q., Liu, J., Liu, Q., Kang, L., Liu, R., Jing, W., Wu, Y., & Liu, M. (2021). Global Percentage of Asymptomatic SARS-CoV-2 Infections Among the Tested Population and Individuals With Confirmed Covid-19 Diagnosis: A Systematic Review and Meta-analysis. *JAMA Netw Open*, 4(12), e2137257. DOI: 10.1001/jamanetworkopen.2021.37257.
- Meyer, S., Elias, J., & Hohle, M. (2012). A space-time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics*, 68(2), 607616. DOI: 10.1111/j.1541-0420.2011.01684.
- Mohler, G. O., Schoenberg, F.P., Short M.B., & Sledge D. (2020). Analyzing the world-wide impact of public health interventions on the transmission dynamics of Covid-19. DOI: arXiv:2004.01714.
- Ogata, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Association*, 83(401), 927. DOI: 10.2307/2288914.
- Ogata, Y. (1998). Space-Time Point-Process Models for Earthquake Occurrences. *Annals of the Institute of Statistical Mathematics*, 50, 379402. DOI: 10.1023/A:1003403601725.
- Park, J., Chaffee, A.W., Harrigan, R.J., & Schoenberg, F.P. (2022). A non-parametric Hawkes model of the spread of Ebola in west Africa. *Journal of Applied Statistics*, 49, 621–637. DOI: 10.1080/02664763.2020.1825646.
- Phillips S., & Schoenberg F. Efficient nonparametric estimation of variable productivity Hawkes processes. *Journal of Applied Statistics*, in review.
- Rizoiu, S., Mishra, S., Kong, Q., Carman, M., & Xie, L. (2018). SIR-Hawkes: Linking epidemic models and Hawkes processes to model diffusions in finite populations. *Proceedings of the 2018 World Wide Web Conference*, 419-428. DOI: 10.1145/3178876.3186108.
- Schoenberg, F.P., Hoffmann, M., & Harrigan, R.J. (2019). A recursive point process model for infectious diseases. *Annals of the Institute of Statistical Mathematics*, 71(5), 1271-1287. DOI: 10.1007/s10463-018-0690-9.
- Schoenberg, F. (2022). Nonparametric estimation of variable productivity Hawkes processes. *Environmetrics*, 33(6). DOI: 10.1002/env.2747.
- Schoenberg, F. (2023). Estimating Covid-19 transmission time using Hawkes point processes. *Ann. Appl. Stat.*, 17(4), 3349–3362. DOI: 10.1214/23-AOAS1765.
- Southall, E., Tildesley, M.J., & Dyson, L. (2020). Prospects for detecting

early warning signals in discrete event sequence data: Application to epidemiological incidence data. *PLoS computational biology*, 16(9). DOI: 10.1371/journal.pcbi.1007836.

Wallinga, J., & Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol*, 160, 509-516. DOI: 10.1093/aje/kwh255.