Journal of
Clinical Medicine

Article

Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13 – 23, 2020

Kimberlyn Roosa[1*], Yiseul Lee[1], Ruiyan Luo[1], Alexander Kirpich[1], Richard Rothenberg[1], James M. Hyman[2], Ping Yan[3], and Gerardo Chowell[1]

[1] Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, USA
[2] Department of Mathematics, Center for Computational Science, Tulane University, New Orleans, LA, USA
[3] Infectious Disease Prevention and Control Branch, Public Health Agency of Canada, Ottawa, Canada
* Correspondence: kroosa1@student.gsu.edu

Received: 14 February 2020; Accepted: ; Published:

Abstract: The ongoing COVID-19 epidemic continues to spread within and outside of China, despite several social distancing measures implemented by the Chinese government. Limited epidemiological data are available, and recent changes in case definition and reporting further complicate our understanding of the impact of the epidemic, particularly in the epidemic's epicenter. Here we use previously validated phenomenological models to generate short-term forecasts of cumulative reported cases in Guangdong and Zhejiang, China. Using daily reported cumulative case data up until February 13, 2020 from the National Health Commission of China, we report 5- and 10-day ahead forecasts of cumulative case reports. Specifically, we generate forecasts using a generalized logistic growth model, the Richards growth model, and a sub-epidemic wave model, which have each been previously used to forecast outbreaks due to different infectious diseases. Forecasts from each of the models suggest the outbreaks may be nearing extinction in both Guangdong and Zhejiang; however, the sub-epidemic model predictions also include the potential for further sustained transmission, particularly in Zhejiang. Our 10-day forecasts across the three models predict an additional 65 – 81 cases (upper bounds: 169 – 507) in Guangdong and an additional 44 – 354 (upper bounds: 141 – 875) cases in Zhejiang by February 23, 2020. In the best-case scenario, current data suggest that transmission in both provinces is slowing down.

Keywords: COVID-19; coronavirus; China; real-time forecasts; phenomenological models; sub-epidemic model

1. Introduction

The ongoing epidemic of a novel coronavirus illness (COVID-19) began in Hubei Province, China, in December 2019 and continues to cause infections in multiple countries, threatening to become a pandemic. However, the bulk of the associated morbidity and mortality is still concentrated within the province of Hubei, China. As of February 13, 2020, there have been 59,907 cumulative cases, including 1,368 deaths, reported globally with 48,206 cases reported in Hubei alone [1]. To control the epidemic, the Chinese government has enacted a range of social distancing strategies, such as city-wide lockdowns, screening measures at train stations and airports, active case finding, and isolation of suspected cases. The numbers of cases and deaths continue to accumulate every day. However, the transmission appears to be slowing down outside Hubei due to strict lockdowns combined with isolation and quarantine measures [1-3].

While the transmission potential of this novel coronavirus can reach high values [4, 5], the epidemiological features of COVID-19 are still unclear, and changes in reporting of cases and deaths complicate the analysis of the epidemic. For instance, the case definition has been revised over time, and, as of February 12, 2020, reported cases of the disease incorporate clinically suspected cases in addition to laboratory-confirmed cases, which led to a noticeable increase in cases in the province of Hubei on February 13, 2020. These additional cases are likely historical cases that occurred days or weeks earlier. In the absence of additional information, this change in reporting obscures the true underlying epidemic trajectory, especially in Hubei, and complicates the inference of epidemiological parameters, such as the effective reproduction number, and the calibration of mechanistic transmission models that rely on data after February 12, 2020.

Earlier work has shown that phenomenological growth models, including the sub-epidemic growth model, can capture the empirical patterns of past epidemics and are useful to generate short-term forecasts of the epidemic trajectory in real time. These approaches are especially useful when the epidemiological data are limited [6-9]. Real-time short-term forecasts generated from such models can be useful to guide the allocation of resources that are critical to bring the epidemic under control. In this paper, we use dynamic models to generate 5-day and 10-day ahead forecasts of the cumulative reported cases in the provinces of Guangdong and Zhejiang, China.

2. Methods

2.1. Data

We use data from the National Health Commission of China which reports the cumulative cases for 34 provinces, including municipalities, autonomous regions, and special administrative regions [1]. We collected reported case data each day at 12 pm (GMT-5) from the initial date of reporting, January 22, 2020, to February 13, 2020. We then forecasted the trajectory of the epidemic in the provinces of Guangdong and Zhejiang, which have exhibited a high burden of COVID-19. We did not fit the data or forecast the epidemic in Hubei, as the recent change in the reporting criteria resulted in a significant jump in reported cases on February 13, 2020. The impact of the new reporting criteria in this province will require a detailed analysis before the data can be fit by our current models.

2.2. Models

We use three phenomenological models that have been previously applied to various infectious disease outbreaks, including other respiratory illnesses, such as SARS and pandemic influenza [10, 11], and to this current outbreak [12]. The generalized logistic growth model (GLM) and the Richards model extend the simple logistic growth model with an additional scaling parameter [9, 11, 13]. We also apply a sub-epidemic model, which accommodates complex epidemic trajectories, such as multiple peaks and sustained or damped oscillations, by assembling the contribution of inferred overlapping sub-epidemics [10]. Appendix A includes a detailed description of the models and their parameters.

2.3. Short-term forecasts

We calibrate each of the models to the daily case counts reported for Guangdong and Zhejiang provinces. We fit the model to the "incidence" curve while presenting the cumulative curves. Reported data are available beginning January 22, 2020, so the calibration period includes daily data from January 22 – February 13, 2020. We estimate the best-fit solution for each model using nonlinear least squares fitting, a process that yields the set of model parameters that minimizes the sum of squared errors between the model and the data. The initial conditions are set to the first data point, and initial parameter estimates can be found in Supplemental Table 1.

We use a parametric bootstrap approach to generate uncertainty bounds around the best-fit solution assuming a Poisson error structure; detailed descriptions of this method are provided in references [9, 14]. We refit the models to each of the M = 200 datasets generated by the bootstrap approach, resulting in M best-fit parameter sets that are used to construct the 95% confidence intervals for each parameter. Further, each model solution is used to generate m = 30 additional simulations extended through a 10-day forecasting period. We construct the 95% prediction intervals for forecasts with these 6,000 (M × m) curves.

3. Results

We present results for 5- and 10-day forecasts generated on February 13, 2020 for the provinces of Guangdong and Zhejiang, China. Figures 1 and 2 contain the estimated ranges of cumulative case counts from 5- and 10-day forecasts for Guangdong and Zhejiang, respectively. Ten-day ahead forecasts from each model with the calibration data are shown in Figures 3 – 5.

3.1. Guangdong

Our 5-day average forecasts for Guangdong are nearly equivalent across the three models, ranging from 1,290 – 1,304 cumulative reported cases (Figure 1). As of February 13, 2020, Guangdong has 1,241 reported cases [1], so forecasts predict an additional 49 – 63 cases in the next 5 days. Upper bounds (UB) of 95% prediction intervals for both the GLM and Richards model suggest that up to 1,392 cases could accumulate, while the sub-epidemic prediction intervals are substantially wider and include up to 1,699 cases; this translates to an additional 151 – 458 additional cases by February 18, 2020.

10-day forecasts suggest very little increase from the 5-day forecasts, especially for those predicted by the GLM and Richards model (Figure 1). Average 10-day forecasts predict between 1,306 – 1,322 cumulative cases with upper bounds ranging from 1,410 – 1,748 cases. This forecast suggests that an additional 65 – 81 cases (UB: 169 – 507) will be reported by February 23, 2020.

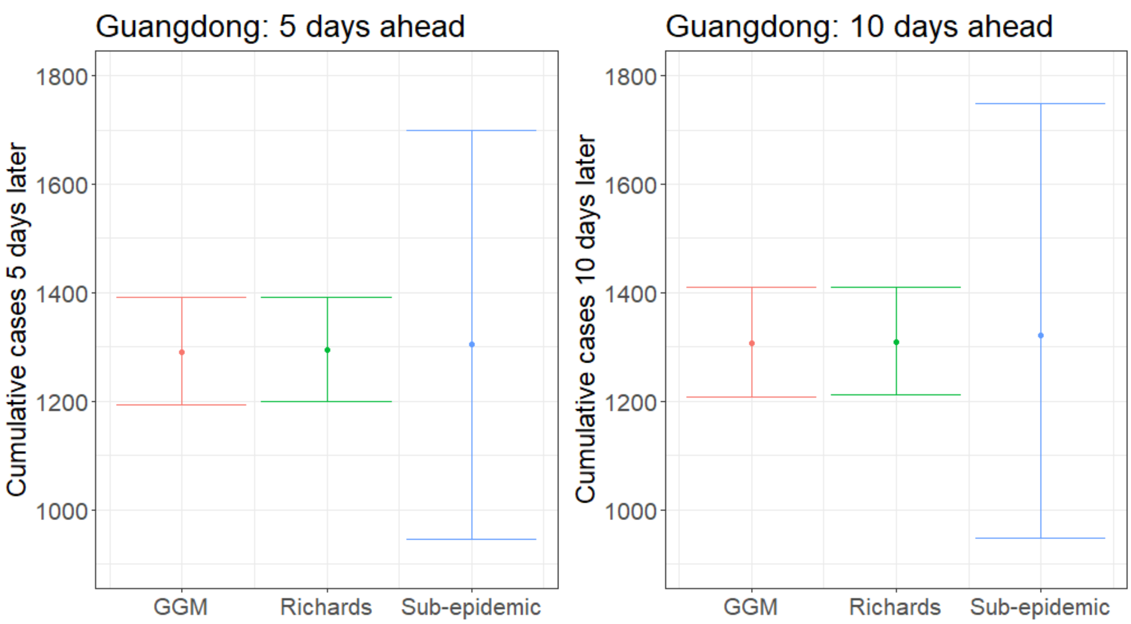

Figure 1. Forecasting results of 5- and 10-days ahead estimates of cumulative reported case counts for Guangdong, China, generated on February 13, 2020. The dots are the mean estimates for each model, and the hinge lines represent the 95% prediction intervals.

3.2. Zhejiang

Average 5-day forecasts from the GLM and Richards model are nearly equivalent for Zhejiang (1,181 and 1,186, respectively), while the sub-epidemic model predicts an average of 1,405 cumulative cases (Figure 2). The sub-epidemic model also has significantly higher upper bounds, suggesting the possibility of up to 1,853 cases, while the GLM and Richards only predict up to 1,276 and 1,279, respectively. As of February 13, 2020, Zhejiang has a total cumulative reported case count of 1,145 [1]; therefore, the models are predicting an additional 36 – 260 cases in the next five days (UB: 131 – 708).

Our 10-day forecasts from the GLM and Richards model show little increase in cases from 5 to 10 days ahead; however, the sub-epidemic model forecasts increase significantly during this time (Figure 2). 10-day forecasts across the models predict 1,189 – 1,499 cumulative cases, on average, with upper bounds ranging from 1,286 – 2,020 cases. This corresponds to an additional 44 – 354 (UB: 141 – 875) cases in Zhejiang by February 23, 2020.

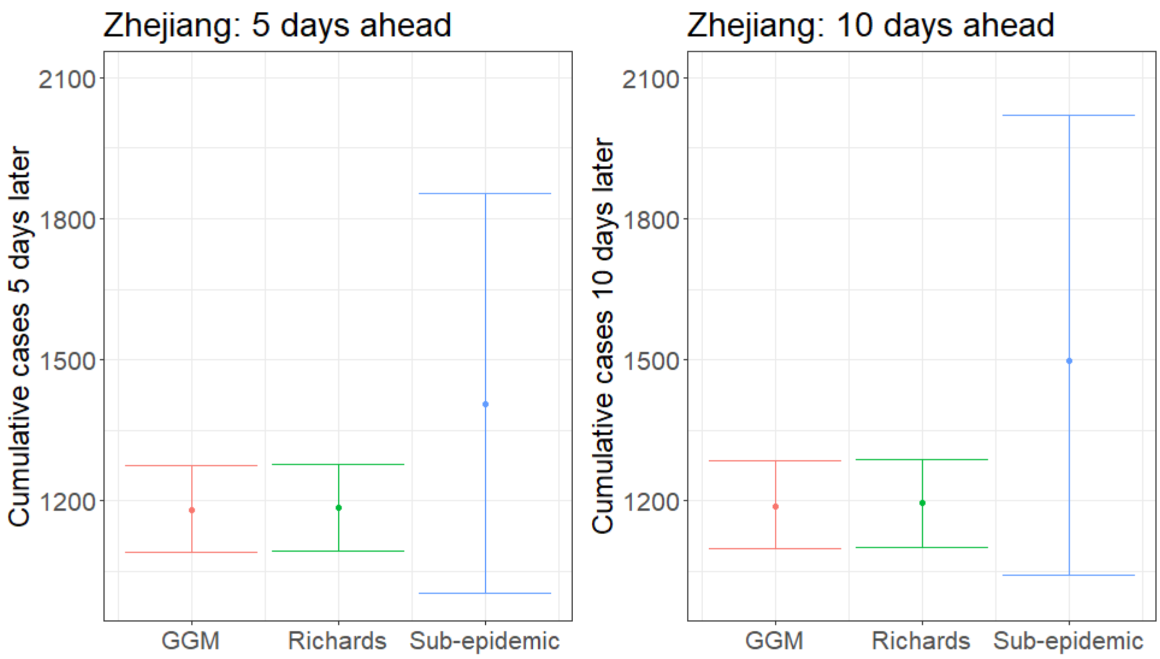

Figure 2. Forecasting results of 5- and 10-days ahead estimates of cumulative reported case counts for Zhejiang, China, generated on February 13, 2020. The dots are for mean estimates for each model, and the hinge lines represent the 95% prediction intervals.

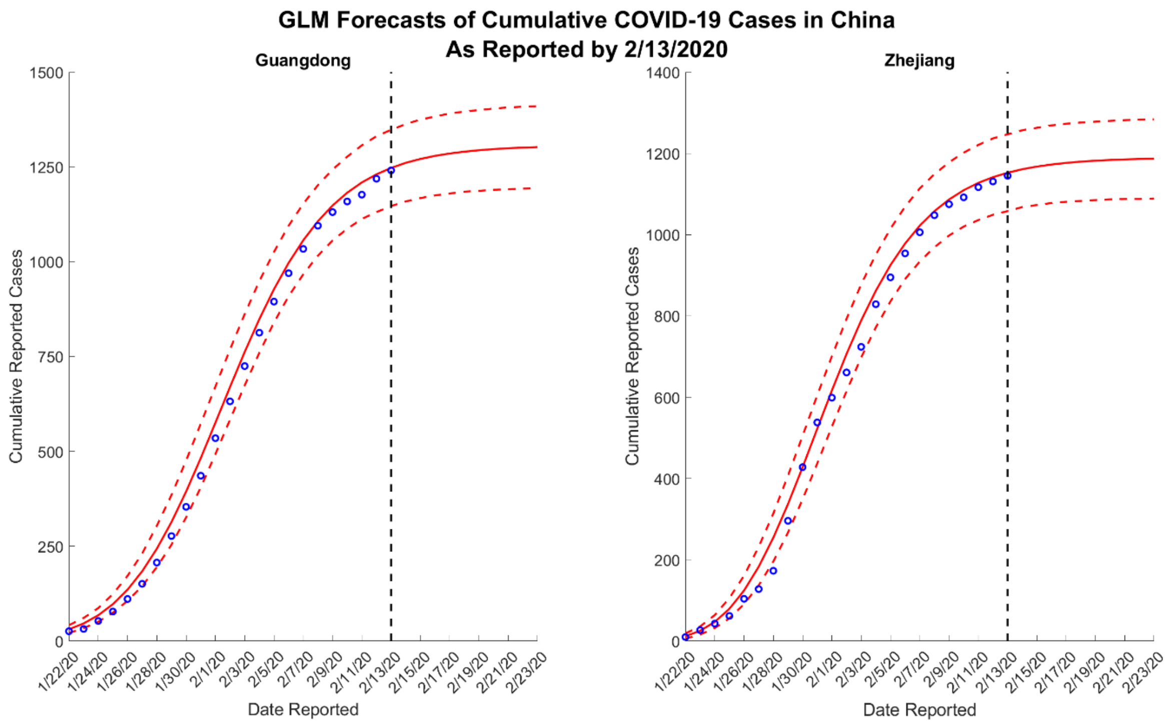

Figure 3. 10-day ahead GLM forecasts of cumulative reported COVID-19 cases in Guangdong and Zhejiang, China – generated on February 13, 2020. The blue circles correspond to the cumulative cases reported up until February 13, 2020; the solid red lines correspond to the mean model solution; the dashed red lines depict the 95% prediction intervals; and the black vertical dashed line separates the calibration and forecasting periods.

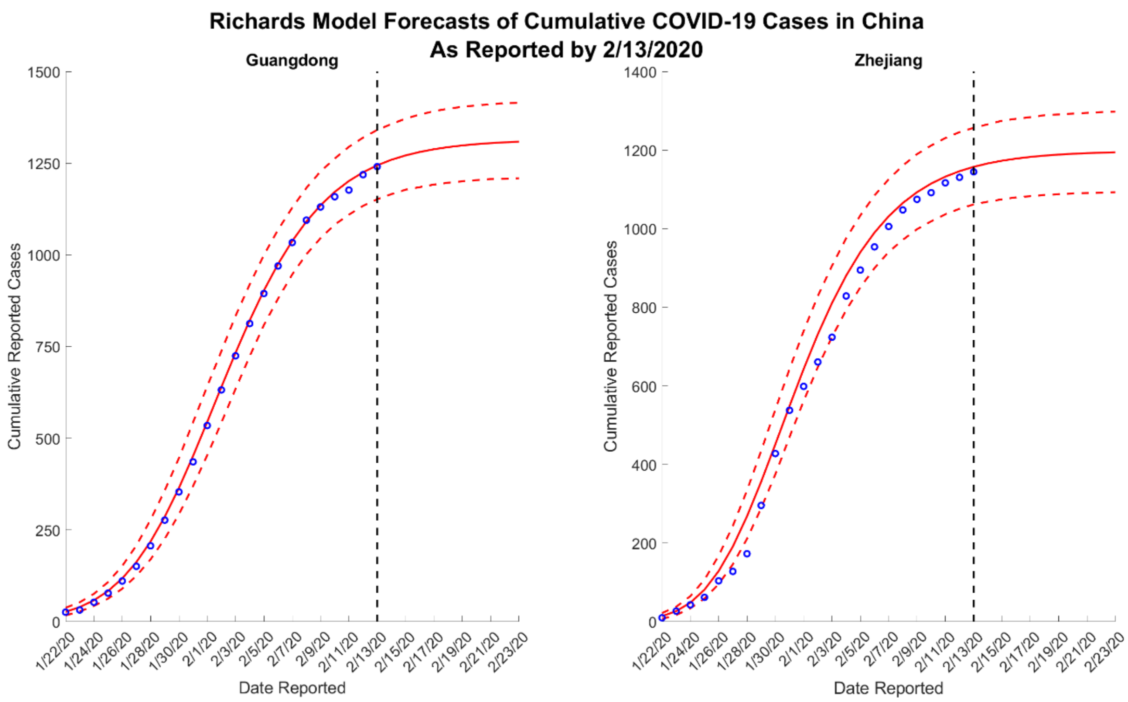

Figure 4. 10-day ahead Richards model forecasts of cumulative reported COVID-19 cases in Guangdong and Zhejiang, China – generated on February 13, 2020. The blue circles correspond to the cumulative cases reported up until February 13, 2020; the solid red lines correspond to the mean model solution; the dashed red lines depict the 95% prediction intervals; and the black vertical dashed line separates the calibration and forecasting periods.

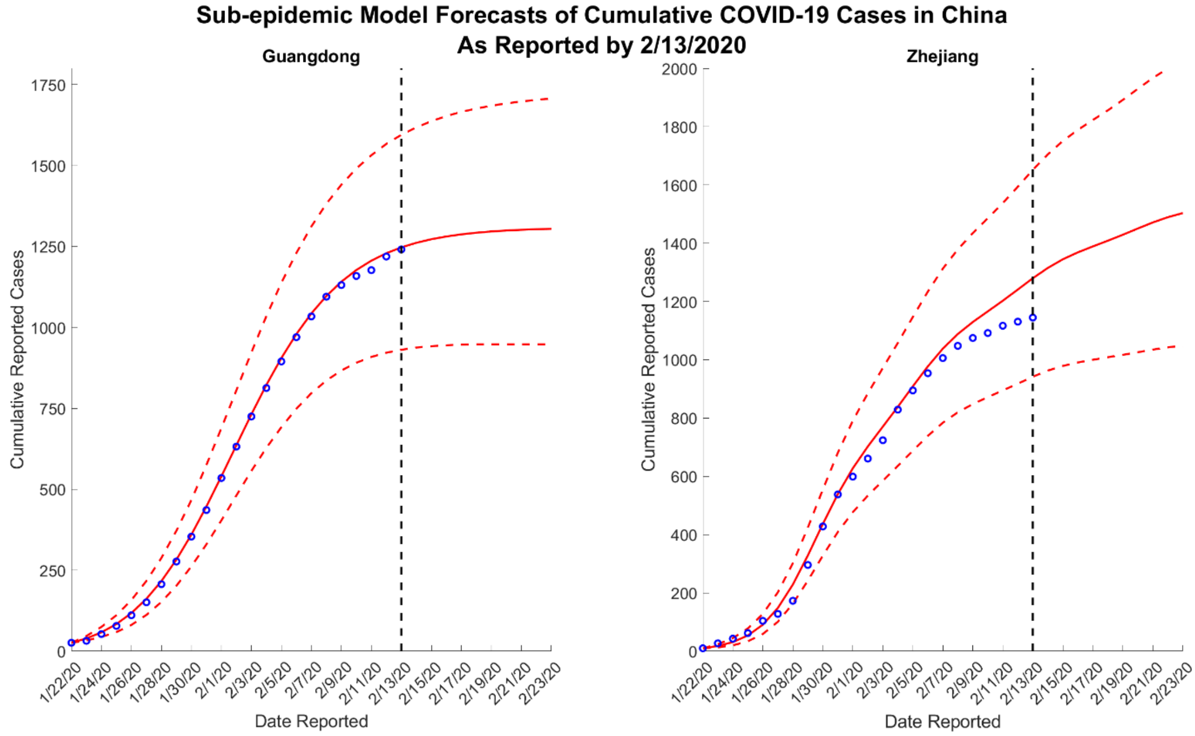

Figure 5. 10-day ahead sub-epidemic model forecasts of cumulative reported COVID-19 cases in Guangdong and Zhejiang, China – generated on February 13, 2020. The blue circles correspond to the cumulative cases reported up until February 13, 2020; the solid red lines correspond to the mean model solution; the dashed red lines depict the 95% prediction intervals; and the black vertical dashed line separates the calibration and forecasting periods.

4. Discussion

We present timely short-term forecasts for reported cases of COVID-19 in Guangdong and Zhejiang, China. Based on data reported up to February 13, 2020, the models predict 65 – 81 additional cases (UB: 169 – 507) in Guangdong and 44 – 354 (UB: 141 – 875) additional cases in Zhejiang by February 23, 2020. Overall, our forecasts suggest that the epidemics in these two provinces continue to slow down.

Across all forecasts, the GLM and Richards model provide comparable mean estimates and prediction intervals, while the sub-epidemic model forecasts exhibit significantly greater uncertainty (Figures 1 & 2). While the mean estimates for Guangdong are nearly equivalent across all three models, the mean estimates generated by the sub-epidemic model are significantly higher for Zhejiang. Both the GLM and Richards models predict that the provinces are nearing the end of the epidemic (Figures 3 & 4). However, forecasts from the sub-epidemic model, which accommodates more complex trajectories, suggest a longer epidemic wave (Figure 5). Specifically, the sub-epidemic model forecasts for Zhejiang suggest additional smaller sub-epidemics have yet to occur, resulting in higher estimates of cumulative case counts.

While we do not know the true underlying epidemic trajectory, it is reasonable to assume that the sub-epidemic forecasts better capture the uncertainty for the next 10 days. The fluctuating case definition may partially explain the slowing down observed in the data that result in the GLM and Richards model predicting extinction. The kink in the Zhejiang data may suggest a case definition change around February 6, 2020, which would partially explain a decrease in the new daily cases reported. The slowing in cases after February 6th is apparent in both Guangdong and Zhejiang. This pattern must be interpreted with caution; it is not entirely clear whether this is a true decline in transmission, or if it is an artificial decline due to the changing case definition. Therefore, the sub-epidemic model forecasts likely better capture both possibilities. Additionally, on February 14, 2020, China officially reported that 1,716 healthcare workers have been infected with the disease, suggesting disease amplification in healthcare settings as it has been previously documented for past SARS and MERS outbreaks [15]. This analysis does not account for the greater potential for transmission in healthcare settings. Finally, we note that our forecasts are not sensitive to the last data point used to calibrate the models (February 13, 2020), despite the most recent change in case reporting (Supplemental Tables 2 & 3).

In conclusion, while our models predict the outbreaks in Guangdong and Zhejiang have nearly reached extinction, our forecasts need to be interpreted with caution given the unstable case definition and reporting patterns. Thus, we point readers to the sub-epidemic model predictions specifically, which suggest that another smaller wave of cases is in process. If the observed decline in case incidence is true, the predictions likely reflect the impact of the social distancing measures implemented by the Chinese government. In the best-case scenario, the model forecasts based on the current data suggest that transmission in both provinces is slowing down.

Author Contributions: KR and GC conducted forecasts and data analysis; YL retrieved and managed data; all authors contributed to writing and revising subsequent versions of the manuscript. All authors read and approved the final manuscript.

Funding: GC is supported by NSF grants 1610429 and 1633381.

Acknowledgments: We thank Homma Rafi (Director of Communications, School of Public Health, Georgia State University) for creating and maintaining an online record of daily short-term forecasts.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

GLM

The generalized logistic growth model (GLM) extends the simple logistic growth model with a scaling of growth parameter p that accommodates sub-exponential growth patterns [16-19]. The GLM is defined by the differential equation

$$C'(t) = rC(t)^p(1 - \frac{C(t)}{K})$$

where $C(t)$ is the cumulative cases at time t, r is the early growth rate, p is the scaling of growth parameter, and K is the carrying capacity and final epidemic size. Values of $p = 1$ correspond to exponential growth rate, where $p = 0$ represents constant growth, and $0 < p < 1$ defines sub-exponential growth.

Richards model

The Richards model also extends the simple logistic growth model through a scaling parameter, a that measures the deviation from the symmetric simple logistic curve [9, 11, 13]. The Richards model is defined by the differential equation

$$C'(t) = rC(t)\left(1 - (\frac{C(t)}{K})^{a}\right)$$

where *C(t)* represents the cumulative case count at time *t*, *r* is the growth rate, *K* is the final epidemic size, and *a* is a scaling parameter.

Sub-epidemic model

While the GLM and Richards model only accommodate s-shaped dynamics, the sub-epidemic wave model supports complex epidemic trajectories by aggregating asynchronous sub-epidemics. For this approach, we assume that the observed curve is the aggregate of multiple overlapping sub-epidemics, where each sub-epidemic is modeled using the GLM [10]. An epidemic wave composed of n overlapping sub-epidemics is modeled as

$$C_i{}'(t) = rA_{i-1}(t)C_i(t)^{p}\left(1 - \frac{C_i(t)}{K_i}\right)$$

where $C_i(t)$ is the cumulative number of infections in sub-epidemic *i* ($i = 1, \ldots, n$), K_i is the size of the i^{th} sub-epidemic, and the growth rate *r* and scaling parameter *p* are the same across sub-epidemics [10]. Further, when $n = 1$, the model returns to the single-equation GLM as presented above.

The timing of onset for each consecutive sub-epidemic is modeled with a regular structure, such that the $(i+1)^{th}$ sub-epidemic is triggered when the cumulative case count of sub-epidemic *i*, $C_i(t)$, exceeds the threshold C_{thr}. The $(i+1)^{th}$ sub-epidemic begins before the i^{th} sub-epidemic reaches extinction. The size of consecutive sub-epidemics (K_i) is modeled such that the size declines exponentially for each subsequent sub-epidemic, where

$$K_i = K_0 e^{-q(i-1)}$$

and K_0 is the size of the first sub-epidemic ($K_1 = K_0$), and *q* is the rate of decline, where $q = 0$ corresponds to no decline. The total final epidemic size is

$$K_{tot} = \sum_{i=1}^{n_{tot}} K_0 e^{-q(i-1)} = \frac{K_0(1 - e^{-qn_{tot}})}{1 - e^{-q}}$$

where n_{tot} is the finite number of overlapping sub-epidemics, calculated as

$$n_{tot} = \left\lfloor -\frac{1}{q}\ln\left(\frac{C_{thr}}{K_0}\right) + 1\right\rfloor$$

References

1. Chinese National Health Commission. *Reported Cases of 2019-nCoV*. 02/02/2020 - 02/13/2020]; Available from: https://ncov.dxy.cn/ncovh5/view/pneumonia?from=groupmessage&isappinstalled=0.
2. World Health Organization, *Novel Coronavirus (2019-nCoV) Situation Reports*. 2020: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
3. Li, Q., et al., *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.* The New England Journal Of Medicine, 2020.

4. Mizumoto, K., K. Kagaya, and G. Chowell, *Early epidemiological assessment of the transmission potential and virulence of 2019 Novel Coronavirus in Wuhan City: China, 2019-2020*. 2020: https://doi.org/10.1101/2020.02.12.20022434.
5. Sanche, S., et al., *The Novel Coronavirus, 2019-nCoV, is Highly Contagious and More Infectious Than Initially Estimated*. 2020: https://doi.org/10.1101/2020.02.07.20021154
6. Pell, B., et al., *Using phenomenological models for forecasting the 2015 Ebola challenge.* Epidemics, 2018. **22**: p. 62-70.
7. Chowell, G., et al., *Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics*. 2016, Public Library of Science, 2016-05-31.
8. Roosa, K., et al., *Multi-model forecasts of the ongoing Ebola epidemic in the Democratic Republic of Congo, March – October 2019*. 2020: Submitted to Royal Society Interface.
9. Chowell, G., *Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts.* Infectious Disease Modelling, 2017. **2**: p. 379-398.
10. Chowell, G., A. Tariq, and J.M. Hyman, *A novel sub-epidemic modeling framework for short-term forecasting epidemic waves.* BMC Medicine, 2019. **17**(1): p. 164-164.
11. Wang, X.-S., J. Wu, and Y. Yang, *Richards model revisited: Validation by and application to infection dynamics.* Journal of Theoretical Biology, 2012. **313**: p. 12-19.
12. Roosa, K., et al., *Real-time forecasts of the 2019-nCoV epidemic in China from February 5th to February 24th, 2020.* Infectious Disease Modeling, 2020.
13. Richards, F., *A flexible growth function for empirical use.* Journal of Experimental Botany, 1959. **10**(2): p. 290-301.
14. Roosa, K. and G. Chowell, *Assessing parameter identifiability in compartmental dynamic models using a computational approach: Application to infectious disease transmission models.* Theoretical Biology and Medical Modelling, 2019. **16**(1).
15. Chowell, G., et al., *Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study.* BMC Medicine, 2015. **13**(210).
16. Viboud, C., L. Simonsen, and G. Chowell, *A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks.* Epidemics, 2016. **15**: p. 27-37.
17. Chowell, G., et al., *Real-time forecasting of epidemic trajectories using computational dynamic ensembles.* Epidemics, 2020. **30**.
18. Roosa, K., R. Luo, and G. Chowell, *Comparative assessment of parameter estimation methods in the presence of overdispersion: a simulation study.* Mathematical Biosciences and Engineering, 2019. **16**(5): p. 4299-4313.
19. Ganyani, T., et al., *Assessing the relationship between epidemic growth scaling and epidemic size: The 2014-16 Ebola epidemic in West Africa.* Epidemiology and Infection, 2018.