medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Short-term forecasts and long-term mitigation evaluations for the COVID-19 epidemic in Hubei Province, China

Qihui Yang[a1*], Chunlin Yi[a1], Aram Vajdi[a], Lee W. Cohnstaedt[b], Hongyu Wu[a], Xiaolong Guo[a], Caterina M. Scoglio[a]

[a] Department of Electrical and Computer Engineering, Kansas State University, Manhattan, Kansas, USA

[b] Arthropod-Borne Animal Diseases Research Unit, Center for Grain and Animal Health Research, USDA ARS, Manhattan, Kansas, USA

[1]These authors equally contributed.

*Corresponding author

Email: qihui@ksu.edu

Abstract

As an emerging infectious disease, the 2019 coronavirus disease (COVID-19) has developed into a global pandemic. During the initial spreading of the virus in China, we demonstrated the ensemble Kalman filter performed well as a short-term predictor of the daily cases reported in Wuhan City. Second, we used an individual-level network-based model to reconstruct the epidemic dynamics in Hubei Province at its early stage and examine the effectiveness of non-pharmaceutical interventions on the epidemic spreading with various scenarios. Our simulation results show that without continued control measures, the epidemic in Hubei Province could have become persistent. Only by continuing to decrease the infection rate through 1) protective measures and 2) social distancing can the actual epidemic trajectory that happened in Hubei Province be reconstructed in simulation. Finally, we simulate the COVID-19 transmission with non-Markovian processes and show how these models produce different epidemic trajectories, compared to those obtained with Markov processes. Since recent studies show that COVID-19 epidemiological parameters do not follow exponential distributions leading to Markov processes, future works need to focus on non-Markovian models to better capture the COVID-19 spreading trajectories. In addition, shortening the infectious period via early case identification and isolation can slow the epidemic spreading significantly.

Keywords: COVID-19; Kalman filtering; network-based model; non-Markovian process

1. Introduction

The 2019 novel coronavirus disease (COVID-19) emerged in Wuhan City, Hubei Province of China, in the late fall of 2019 (Chen et al., 2020). The highly contagious virus outbreak during the Chinese Spring festival period made containment an unprecedented challenge. The risks of human-to-human transmission, long persistence on surfaces, and asymptomatic infections increased spread and made case detection difficult (Chan et al., 2020; Rothe et al., 2020). Increasing case reports saturated hospitals in Hubei Province, overwhelming medical resources. Despite the high case reports, the true infection rate was likely underreported, as infected individuals with only mild symptoms might have recovered without being detected (Nishiura et al., 2020; Sanche et al., 2020).

To stop the viral spread, a series of strict control measures have been implemented in mainland China, including the enforcement of the tourism ban, suspension of the school, and extension of the Spring Festival holiday. From January 23rd to April 8th, 2020, the Chinese government implemented a metropolitan-wide quarantine of Wuhan and nearby cities, suspending all public transportation and advising residents to stay

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

at home (Du et al., n.d.; J. T. Wu, Leung, & Leung, 2020). Since February 8th, a door-to-door check was started in Wuhan to take all confirmed and suspected people into medical care, which ended at around February 19th (X. Li, Zhao, & Sun, 2020).

Understanding the impact of non-pharmaceutical interventions on the epidemic dynamics is crucial for decision-makers to take proper actions to contain the COVID-19. Since the beginning of the outbreak, scientists have been actively retrieving the epidemiological characteristics such as the time from infection to illness onset (incubation period) and the time from illness onset to hospital admission or isolation (infectious period) (Chen et al., 2020; Guan et al., 2020; Huang et al., 2020; Linton et al., 2020). Researchers have used the reported cases during the early stage of the epidemic and predicted the epidemic trajectory, the basic reproductive number R_0 from which ranges from 2.2 to 6.6 (Read, Bridgen, Cummings, Ho, & Jewell, 2020; Sanche et al., 2020; Zhao et al., 2020). Several studies explored the effects of the quarantine of Wuhan (X. Li et al., 2020; Shen, Peng, Guo, Xiao, & Zhang, 2020), travel restrictions (Chinazzi et al., 2020), and other non-pharmaceutical intervention strategies (Lai et al., 2020; Zhu et al., 2020) on the future transmission dynamics at different geographical scales.

China's aggressive and prompt measures have led to a dramatic decrease in the number of new cases. Those measures saved many lives, although at a high economic cost. As of March 18th, 2020, the cumulative number of confirmed cases in China was 81,186. Compared to the previous day, the number of new cases was only 70, most of which were imported from outside the country (T. Wu, Ge, Yu, & Hu, 2020). In the meantime, COVID-19 has further spread around the world, and countries like Italy, South Korea, the United States, and Iran are struggling to contain the epidemic.

In this work, we aim to assess the capability of the ensemble Kalman filter as a good short-term predictor, and test the effectiveness of non-pharmaceutical interventions on the epidemic spreading. First, we provide real-time assessments and forecasts of the case reported in Wuhan City based on the ensemble Kalman filter. Second, we build an individual-level based network model and perform stochastic simulations to reconstruct the epidemics in Hubei Province at its early stage and examine the epidemic dynamics under different scenarios. We consider four scenarios: keeping the early-stage trend without any mitigation, reducing the average node degree within a city through social distancing, reducing the infection rate by adopting protective measures, and decreasing the infection rate through both social distancing and protective measures. Third, we simulate the epidemic spreading in Wuhan City incorporating non-Markovian processes, i.e., non-exponential parameter distributions, and compare the predicted epidemic trajectories obtained from Markov processes.

Mathematical modeling can forecast the size and timing of epidemics, so resource allocations and decision-making can be optimized to maximize intervention strategies. Since most current network-based models are based on Markov processes, this work is probably one of the first to simulate the COVID-19 spreading based on non-Markovian processes. The findings of this work can be a reference for policy recommendations regarding COVID-19 in other countries.

2. Model

In this section, we first introduce the ensemble Kalman filtering method used in the short-term analysis. Then we illustrate epidemic models based on Markov processes and non-Markovian processes built in the long-term analysis.

2.1 Short-term analysis

We aim to provide short-term prediction for the COVID-19 spreading in Wuhan city based on the Kalman filtering, which consists of model state estimation at each observation time based only on the observation up to the present. Considering that there is an unknown time-varying parameter β in the SIR model which cannot be dealt with by widely used classic Kalman filter (Kalman, 1960) and extended

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Kalman filter (Jazwinski, 2007), a dual state-parameter estimation filtering method–ensemble Kalman filter (EnKF) is used for our study (Jensen, 2007; Moradkhani, Sorooshian, Gupta, & Houser, 2005). EnKF is based on Monte Carlo where the approximation of forecast (a priori) state error covariance matrix is made by propagating an ensemble of model states using the updated states (ensemble members) from the previous time step.

It is quite difficult to have a real specific contact network between each person in Wuhan. Therefore, a generalized network is used, which means that every individual in this area is connected with a possibility. To describe the epidemiology of COVID-19 pneumonia, we used the discrete-time S-I-R model which categorizes the hosts within a population as susceptible (S), infectious (I) and removed (R) (Keeling & Rohani, 2011). At any discrete time, every individual has to be in one of these 3 compartments. The susceptible are individuals unexposed to the pathogen but can be infected by contact with the infectious population. Infectious individuals can clear the infection by hospitalization, death and quarantine measures and become removed. The equations for the state transitions are shown as follows.

$$S_{k+1} = S_k - \beta S_k I_k \Delta T \tag{1}$$

$$I_{k+1} = \beta S_k I_k \Delta T + I_k - \gamma I_k \Delta T \tag{2}$$

$$R_{k+1} = R_k + \gamma I_k \Delta T \tag{3}$$

Where β denotes the infection rate and γ denotes the removing rate. In the SIR model, we assume that the infection process of each infectious individual is Poisson and statistically independent and that the removing process for every infected individual is also Poisson. In this case, the infection rate β quantifies the probability that each susceptible individual can be infected by each infectious individual in a time unit, and the recovering rate γ quantifies the average portion of the infectious population that will clear infection in a time unit. Considering that the population flow, prevention measures (like wearing N95 mask) and self-quarantine could affect the infection rate, we take the infection rate β as a time-varying parameter. The removing rate is set as a constant 1/7 based on the estimation from Huang et al. (2020).

Let the state variables (SIR) defined as $x_k \in R^3$, the time-varying parameter $\beta_k \in R$. Consider the following problem

$$x_{k+1} = f(x_k, \beta_k) + q_k \tag{4}$$

$$\beta_{k+1} = h(\beta_k) \tag{5}$$

$$y_k = g(x_k, \beta_k) + v_k \tag{6}$$

where q_k and v_k are model and observation noise which are assumed to be normally distributed and with zero mean and the covariance known.

Because the true states are generally unknown, we can use the ensemble mean as the best estimate of the true state. The error covariance matrices for the predicted and filtered ensemble are given by

$$P_k^- = E[(\hat{x}_k^- - E[\hat{x}_k^-])(\hat{x}_k^- - E[\hat{x}_k^-])^T] \tag{7}$$

$$P_k = E[(\hat{x}_k - E[\hat{x}_k])(\hat{x}_k - E[\hat{x}_k])^T] \tag{8}$$

where P_k^- is the error covariance matrix associated with the forecasted estimate (prior), P_k is the one associated with the updated estimate (posterior), $\hat{x}_k^-$ is the ensemble vector of the forecasted estimate from model equations.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In the ensemble presentation, the model states x_k and the time-varying parameter β_k can be put together in a matrix $A_k \in R^{4\times N}$, holding N ensemble member at time k.

$$A_k = \begin{bmatrix} x_k^1 & x_k^2 & \dots & x_k^N \\ \beta_k^1 & \beta_k^2 & \dots & \beta_k^N \end{bmatrix} \tag{9}$$

The ensemble covariance then can be presented as

$$P_k = \frac{1}{N-1}(\hat{A}_k - \bar{\hat{A}}_k)(\hat{A}_k - \bar{\hat{A}}_k)^T \tag{10}$$

$$\bar{A}_k = \frac{1}{N} A_k \cdot 1_{N\times 1} \tag{11}$$

The measurement can also be formulated in an ensemble presentation

$$Y_k = [y_k^1\ y_k^2\ \dots\ y_k^N\ ] \tag{12}$$

$$\varepsilon_k = [v_k^1\ v_k^2\ \dots\ v_k^N\ ] \tag{13}$$

where $y_k^1 = y_k + v_k^1$, y_k is the true measurement and v_k^i is the perturbation term. The measurement covariance matrix can be estimated as

$$V_k = \frac{\varepsilon_k \cdot \varepsilon_k{}^T}{N-1} \tag{14}$$

Specify to the situation we are dealing with, $A_k = [\,S_k\ I_k\ R_k\ \ \beta_k]^T$ where $S_k = [s_k^1\ s_k^2\ \ \dots\ s_k^N]$ and it is the same with I_k, R_k, and β_k. The measurement is the total number of confirmed infected cases reported by the National Health Commission of China, which can be taken as the sum of the infectious and removed. Therefore, $y_k = C \cdot [x\ \beta]^T$ where $C = [0 \quad 1 \quad 1 \quad 0]$.

2.2 Long-term analysis

In the long-term analysis, several epidemic models are built based on the Susceptible-Exposed-Infectious-Removed (SEIR) compartmental model, which has been frequently used in the study of the COVID-19 (X. Li et al., 2020; Read et al., 2020; J. T. Wu et al., 2020; Zhou et al., 2020). Each node is either in one of the compartments susceptible, exposed, infectious, or removed. Susceptible nodes refer to people free of the virus and can become infected due to contact with infectious nodes. Exposed nodes have been infected by the disease, but have not yet shown symptoms or become infectious. Removed nodes represent those individuals who are recovered, isolated, hospitalized, or dead due to the disease.

2.2.1 Network-based approach with Markovian processes

In this section, we apply the generalized modeling framework (GEMF) proposed by Sahneh, Scoglio, & Van Mieghem (2013) to build an individual-level network model. In GEMF, each node stays in a different state, and the joint state of all nodes follow a Markov process. The node-level description of this Markov process can be expressed as follows:

$$P_r[x_i(t+\Delta t) = 1 | x_i(t) = 0, X(t)] = \beta'(t) Y_i \Delta t \tag{15}$$

$$P_r[x_i(t+\Delta t) = 2 | x_i(t) = 1, X(t)] = \lambda \Delta t + o(\Delta t) \tag{16}$$

$$P_r[X_i(t+\Delta t) = 3 | x_i(t) = 2, X(t)] = \delta \Delta t + o(\Delta t) \tag{17}$$

Where $X(t)$ is the joint state of all individual nodes at time t; $x_i(t) = 0, 1, 2, 3$ means that the state of node i at time t is in susceptible, exposed, infectious and removed compartment, respectively; Y_i is the number

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

of infectious neighbors of node i; $\beta'(t)$ is the rate for one node transitions from susceptible state to the exposed state; λ is the rate from the exposed state to the infectious state, and δ is the removal rate from infectious state to the removed state, which is the inverse of the infectious period.

Right after the COVID-19 outbreak in China, many studies have estimated the epidemiological parameters, which are essential to calibrate the epidemic models. However, we noticed that there exists a variety regarding the epidemiological parameters. Based on 1099 patients in 31 provincial municipalities through January 29th, 2020, Guan et al. (2020) estimated that the median incubation period was 3.0 days, ranging between 0 to 24.0 days. This median incubation period was shorter than the previous studies based on smaller sample sizes, in which the estimated time from infection to illness onset was 5–5.2 days (Gardner, 2020; Q. Li et al., 2020). Du et al. (2020) considered the delay from symptom onset to case detection as 4–5 days. Considering the possibility of asymptomatic infections, we adopt a relatively short mean incubation period as 3 days and mean infectious period as 4 days (Lin et al., 2020).

Hubei Province has 17 prefecture-level cities, and the population size of each city is obtained from the Hubei Statistical Yearbook (Hubei Provincial Bureau of Statistics, 2020). Based on the Baidu location-based service data (https://qianxi.baidu.com/) and the fact that around 5 million people have moved out of Wuhan before the city lockdown on January 23rd (Zhu et al., 2020), the number of people that traveled between the cities in Hubei Province from January 1st to January 29th, 2020 is obtained.

Since data about the actual people contact network inside a city are unavailable, we created contact networks among people within each city following with the homogenous network topology, namely the Erdős-Rényi model (Erdős & Rényi, 1960). More specifically, nodes inside a city are randomly connected with a predefined probability such that each node has around 15 links (the average node degree is 15). Intercity links are created to simulate the interaction among the cities according to the number of people that traveled between the cities. For example, if the daily population flow from city i to city j is K number of people, K number of directed links will be randomly generated between people in city i to people in city j. Considering the implementation of city lockdown on January 23rd, we alter the intercity network topology once during the simulation. In the first network topology, links among cities are created based on the average daily population movement between January 1st–22nd, while the second network reflects the population movement from January 23rd–29th.

To fit the model, we consider the COVID-19 epidemic outbreak starting in Wuhan City from January 19th, because all confirmed cases reported in Hubei Province were in Wuhan City before January 19th, 2020. To reduce the computational cost, the population of each city and population flows among the cities are scaled by 198, which is the number of cumulative confirmed cases in Hubei Province up to January 19th. After scaling, the total number of nodes in the network N equals 298,838. In addition, as the criteria for counting the confirmed cases are changed on February 12th, 2020, there is a jump in the reported number of confirmed cases in Hubei Province on February 13th, 2020. Accordingly, by keeping $\lambda = 0.33$ day^{-1} and $\delta = 0.25$ day^{-1} (average) unchanged, we fitted the transmission model to the cumulative number of confirmed cases in Hubei Province (Health Commission of Hubei Province, 2020), from January 19th to February 12th by tuning the infection rate $\beta'(t)$. We then examine the future course of the epidemic under different scenarios.

GEMF is available in MATLAB, R, Python, and C programming language (Sahneh, Vajdi, Shakeri, Fan, & Scoglio, 2017). We performed extensive stochastic simulations by adapting the GEMF toolbox with MATLAB during the experiment. The number of simulation runs is 500 in all scenarios. We track the number of nodes in each compartment for 365 days and demonstrate the simulation results with confidence intervals.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2.2.2 Network-based approach with non-Markov processes

Since the outbreak of COVID-19, many studies have emerged to learn how this disease spreads. However, an overwhelming majority of the epidemic models have considered Markov processes, in which the transition between compartments follows an exponential distribution. To the best of the authors' knowledge, very few works have been done incorporating non-Markovian processes to epidemic models with respect to COVID-19. In this section, we take Wuhan City as a study area and try to demonstrate how distributions of key epidemiological factors can influence the predicted trajectory of the disease spreading.

More specifically, we incorporate non-exponential interevent time distributions to the SEIR model using the modified Gillespie algorithm proposed by Boguñá, Lafuerza, Toral, & Serrano (2014). Transitions from the susceptible state to the exposed state is a Poisson process that follows the exponential distribution. Following lognormal distributions, the mean and median of the incubation period are estimated at 5.6 days and 5.0 days, respectively, while the infectious period is of mean 3.9 days and median 1.5 days (Linton et al., 2020), which is in line with Sanche et al. (2020). The number of simulation runs is 80.

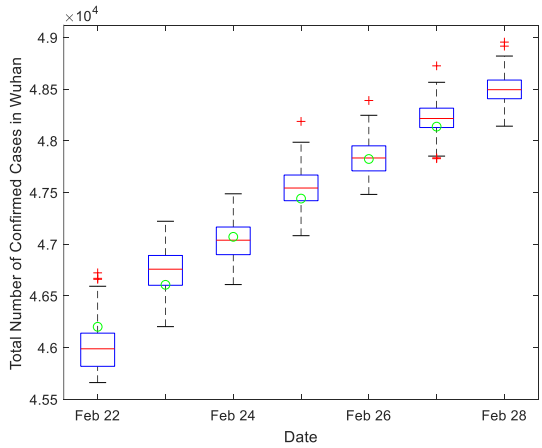
3. Results and discussion

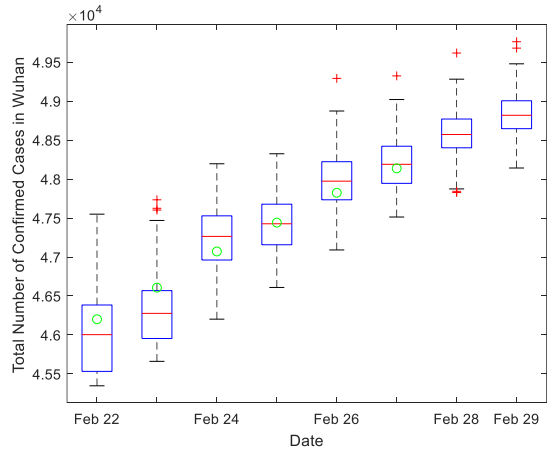
In this section, we provide results regarding the short-term analysis with the Kalman filter. From the long-term perspective, we first present results based on GEMF stochastic simulations and then compare the results based on non-Markovian processes and Markov processes.

3.1 Short-term analysis with the Kalman filter

The predictions of COVID-19 cases in Wuhan city from February 22nd to March 1st were conducted. Every predictive simulation was run 250 iterations. The removing rate we used was 1/7 and the infection rate was estimated by EnKF. The model noise and measurement noise was tuned by state of art.

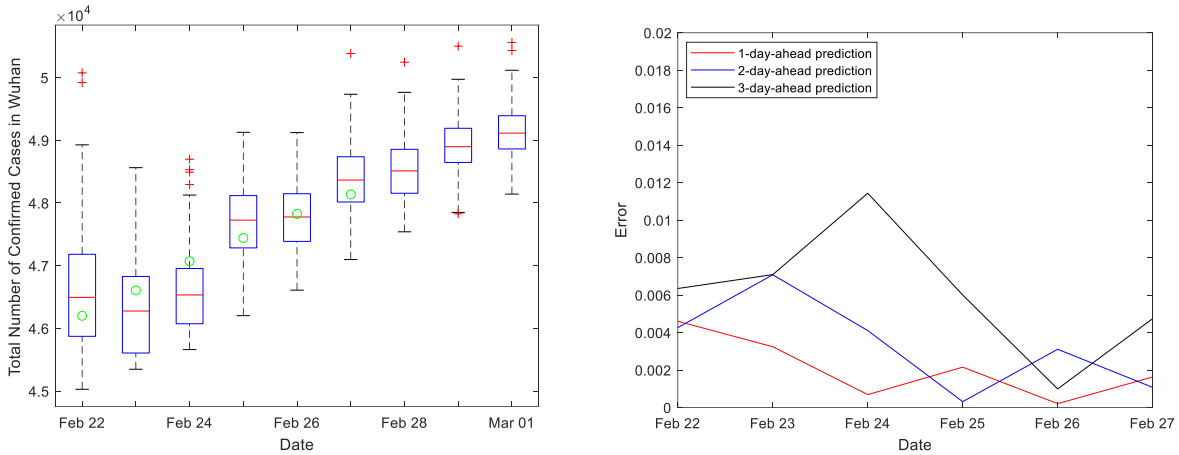
Fig. 1 shows the distributions of the predicted numbers of COVID-19 cases based on data 1, 2 and 3 days before. The results of the EnKF prediction give two ranges – the weak one from minimum to maximum (the solid black line) which can include all the reported real data (the green circle); the stronger one from the lower quartile to the upper quartile (the blue box) which also covers most of the real numbers. The red line inside the box indicates the median of the samples. We can see that most medians are very close to the corresponding real numbers. Taking the median as our predicted number gives the relative error shown in Fig. 1(d). The mean absolute percentage error (MAPE) for 1-day-ahead prediction is 0.19%; the MAPE for 2-day-ahead prediction is 0.32%; the one for 3-day-ahead is 0.62%. The performance of 1-day-ahead prediction is the best and the one of 3-day-ahead is the worst. When *n*-day-ahead prediction is conducted, the results from (*n*-1)-day-ahead is used, so the errors will accumulate.





medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

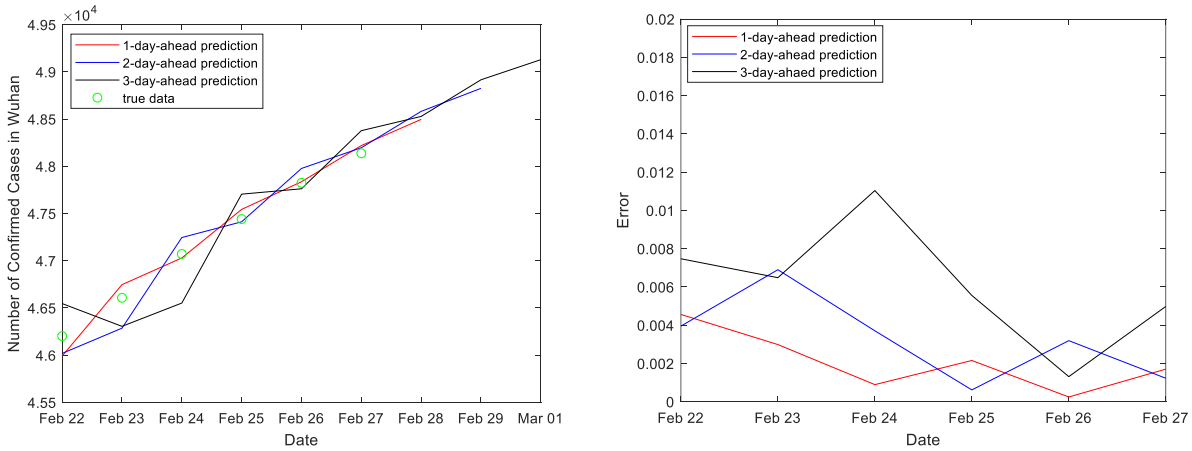
(a) 1-day-ahead prediction (b) 2-day-ahead prediction



(c) 3-day-ahead prediction (d) Relative prediction error

Fig 1. Distribution of short-term prediction.

If we choose the average number of the samples as our predicted number of COVID-19 cases, the results and their relative errors are presented in Fig. 2. The MAPE for 1, 2, 3-day-ahead predictions are 0.2%, 0.31% and 0.62%, which are nearly equal to the ones when the medians are predicted numbers.



(a) Average number of predictions (b) Relative prediction error

Fig 2. Average number as predictions and the error.

The estimation of infection rate β from January 15th to February 27th is presented as the blue line in Fig.3. The orange line in Fig. 3 presents the removing rate. From January 15th to January 27th, the infection rate β is increasing. After January 27th, β starts to decrease. The self-quarantine policy was carried out on January 23rd in Wuhan and a series of administrative actions were taken to prevent the disease from spreading. Considering that there are several days delay from the infection to detection, the estimated β matches with the real situation. On February 12th, because Wuhan authorities changed the case definition to include symptomatic patients in addition to tested ones, there was a huge increase in cases, which results in a huge increase in estimated infection rate β as well. Therefore, the estimated β on February 12th is inaccurate. After February 14th, the infection rate β is less than the removing rate γ, which means the epidemic had peaked in Wuhan.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

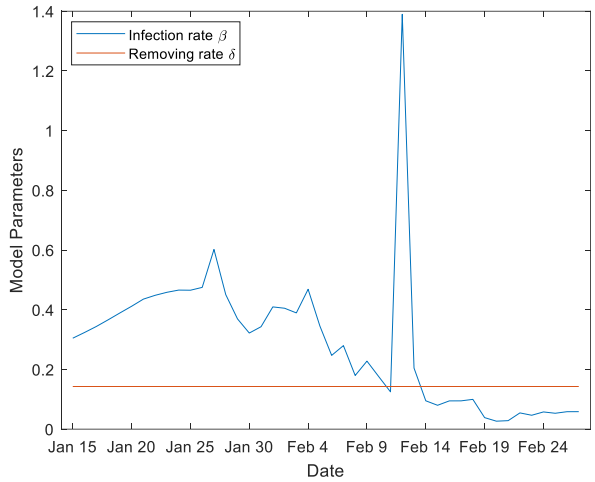


Fig. 3 Estimated infection rate.

3.2 Long-term analysis with the Markovian process

To reconstruct the epidemics in Hubei province at its early stage, we continuously tune the infection rate parameter β' to match the number of nodes in the removed compartment with the number of cumulative confirmed cases from January 19th to February 12th, as shown in Fig. 4.

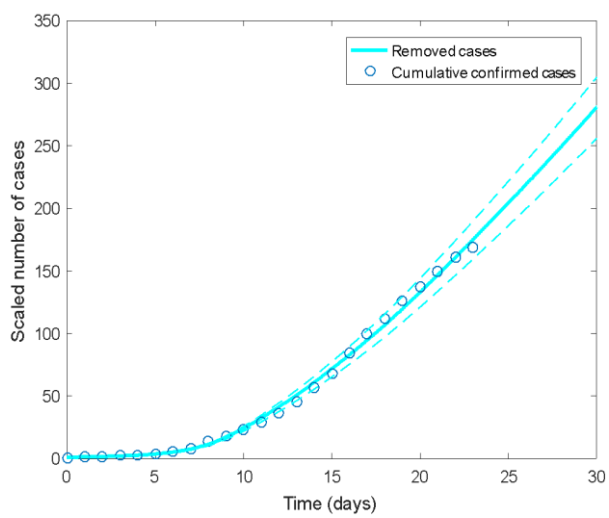


Fig. 4 Fitting the model to the cumulative confirmed cases from January 19th to February 12th.

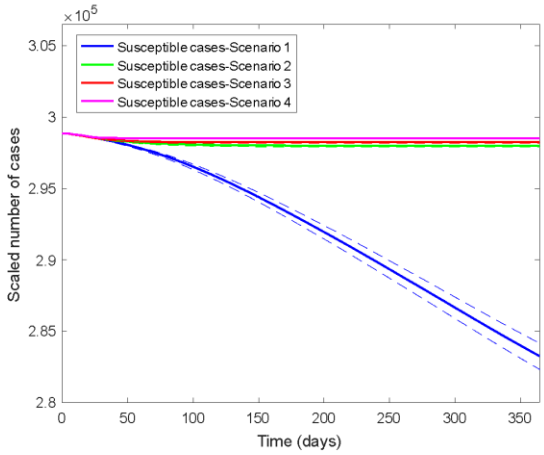
By performing GEMF stochastic simulations, we obtain that β' is equal to 0.12 for day 0–7, and then becomes 0.027 for day 8–10, and 0.024 for day 11–18. Starting from day 19, $\beta' = 0.02$. The infection rate dramatically drops, indicating that public health interventions implemented in Hubei Province have considerably controlled the epidemic development since the nascent stage of the outbreak.

The predicted epidemic dynamics from GEMF stochastic simulations are shown in Fig. 5. We consider four different scenarios: continuing the current trend (Scenario1); decreasing β_2 by 25% (Scenario 2); reducing the average node degree inside a city from 15 to 10 (Scenario 3); reducing β_2 by 25% and average node degree from 15 to 10 (Scenario 4). Preventive measures such as wearing masks can help reduce the infection rate. On the other hand, the average node degree refers to the possible contacts a person may have in daily life, and this can be reduced by contact reduction and social distancing.

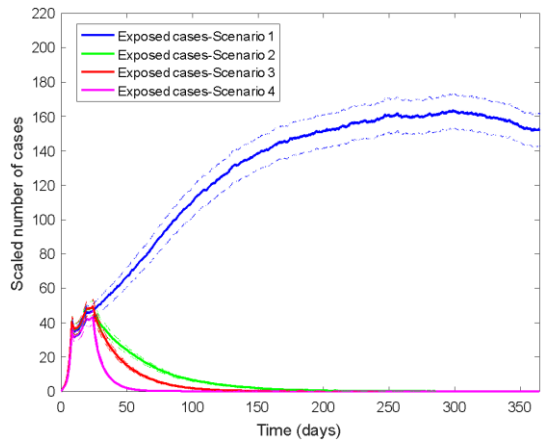
medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In Fig. 5, if the current trend as of February 12th continues, the epidemics in Hubei Province could have become endemic and not been contained in one year. When β' is reduced by 25% in Scenario 2, the number of infectious cases in Hubei Province would reach the peak at day 22 on average (median: 24, IQR: 19-27) and fade out at around day 250 in Fig. 5(c). Reducing the average node degree inside each city from 15 to 10, the epidemic would reach the peak at around day 21 (median: 24, IQR: 19-26) and fade out at day 150 on average. When a combination of reducing β' and average node degree is conducted in scenario 4, the number of infectious cases would reach the peak at around day 19 (median: 23, IQR: 17-25), and the number of removed cases would approximate the actual epidemic trajectory in Hubei Province in the steady-state as shown in Fig. 5 (e).

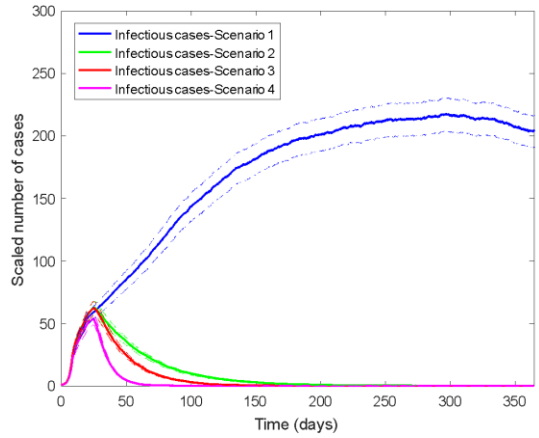
Overall, GEMF simulation results suggest that more strict prevention and control measures have been taken since February 12th to achieve the current cumulative confirmed cases in Hubei Province. Preventive measures and social distancing measures need to be taken in combination by the government to speed up the containment of the epidemic and reduce the peak infection dramatically.



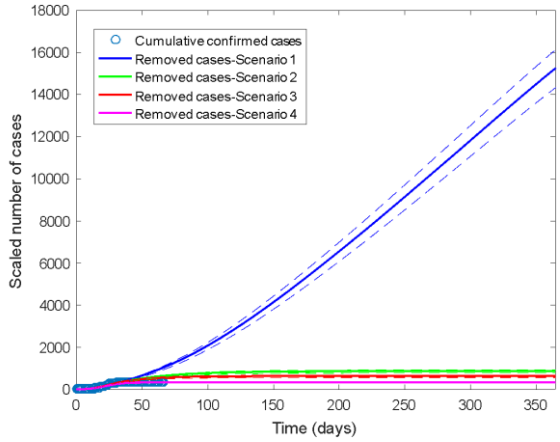
(a) Scaled number of susceptible cases



(b) Scaled number of exposed cases

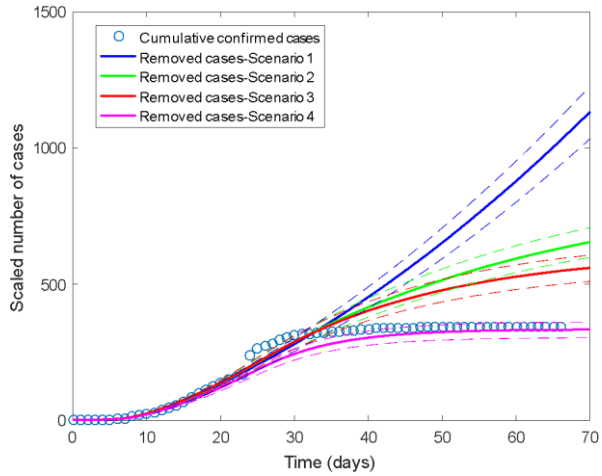


(c) Scaled number of infectious cases



(d) Scaled number of removed cases

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



(e) Scaled number of removed cases compared to the cumulative confirmed cases up to March 24th

Fig 5. Epidemic dynamics under different scenarios from GEMF stochastic simulations.

3.3 Long-term analysis with the non-Markovian process

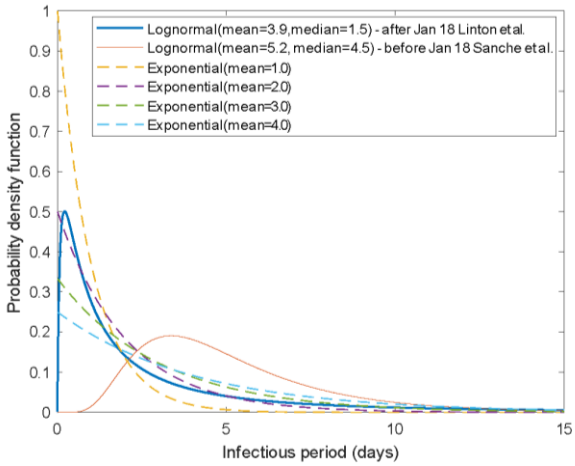
With mean infectious period (from compartments $I \rightarrow R$) set as 3.9 days and mean incubation period (from compartments $E \rightarrow I$) set as 5 days, we first use ordinary differential equations based on Markov processes to estimate the infection rate β from compartments $S \rightarrow E$. Assuming that the epidemic starts from December 1st, β is equal to 0.68 to match the number of cumulative confirmed cases in Wuhan city.

Second, we simulate the disease spread in Wuhan city with transmission between compartments incorporating non-Markovian dynamics. The transition from compartment $S \rightarrow E$ still evolves according to an exponential distribution with $\beta = 0.68$, while transitions from compartments $E \rightarrow I$ and $I \rightarrow R$ follow the lognormal distributions from (Linton et al., 2020; Sanche et al., 2020) in Fig. 6 (a-b). Note that the width of the infectious period becomes much shorter (blue curve) after January 18th, as shown in Fig. 6 (a), because the government has started to trace back and monitor those people who have contact with confirmed cases and built temporary hospitals to hospitalize both patients with mild and severe symptoms since mid-January. In Fig. 6 (a-b), we also show curves of different exponential distributions compared with these lognormal distributions for demonstration purposes.

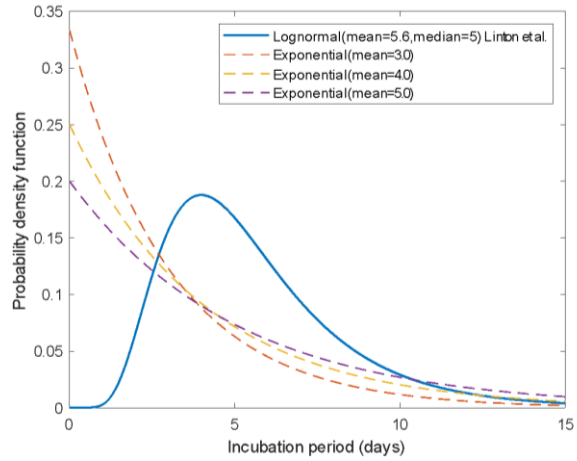
In Fig. 6 (c-d), simulated results incorporating non-Markovian processes deviate from the results obtained from Markov processes. More specifically, the epidemic curve with non-Markovian processes is flattened with a smaller epidemic peak compared with that from Markov processes. This is because the infectious period used in the Markovian model has an exponential distribution (mean =3.9 days) similar to the shape of the exponential distribution (mean=4.0 days) in Fig. 6 (a), but shorter than the lognormal distribution (mean=3.9 days, median=1.5 days) used after January 18th. The reasons above indicate that control measures to reduce the infectious period can effectively mitigate the outbreak.

On the other hand, these two methods with different distributions but with the same mean infectious period lead to inconsistent results. Since the actual infectious period and incubation period do not follow exponential distributions, these simulations suggest that future works need to consider distributions of epidemiological characteristics to better capture the COVID-19 spreading trajectory. Instead of taking mean values derived from clinical studies as input to the epidemic models, researchers need to take into account the empirically derived distribution of the parameters.

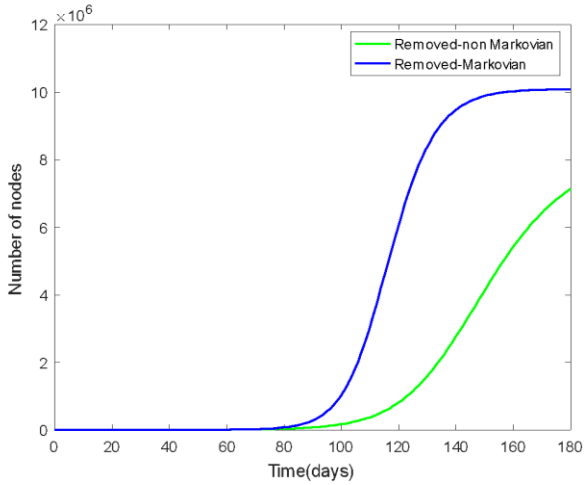
medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



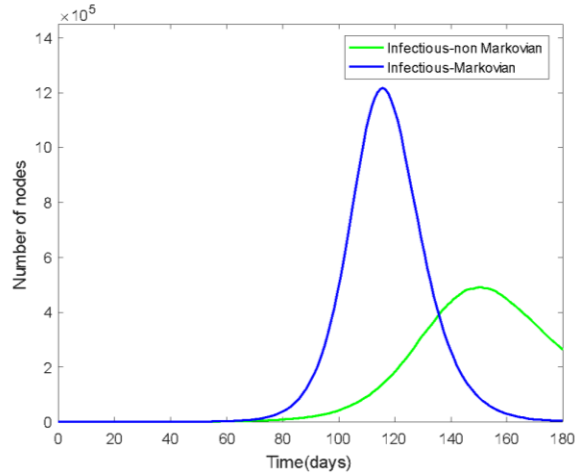
(a) Distribution of the infectious period



(b) Distribution of the incubation period



(c) Number of nodes in the removed compartment



(d) Number of nodes in the infectious compartment

Fig 6. Comparison between results based Markovian process and non-Markovian process.

4. Conclusion

In this work, the ensemble Kalman filter is used to make short-term predictions of the COVID-19 cases in Wuhan City, which resulted in accurate forecasts of daily case reports. The model is able to predict daily cases and the epidemic peak. Knowing the daily cases from forecasts three days in advance allows for the proper resource allocation. The ensemble Kalman filter also allows parameter estimation, which is extremely useful for modeling purposes.

From an epidemiological modeling perspective, we simulate the epidemic spreading in Wuhan city based on non-Markovian processes, and show that different distributions with the same mean infectious period can lead to inconsistency in terms of epidemic dynamics. Since most current network-based approaches are based on Markov processes, in which epidemiological parameters are assumed to follow exponential distributions, non-Markovian process-based models need to be further explored to better predict the COVID-19 spreading. Our results show that reducing the infectious period via measures such as early case identification and isolation can reduce the epidemic size significantly.

In the long-term perspective, we test and confirm the effectiveness of non-pharmaceutical interventions on the containment of the epidemic dynamics by performing GEMF stochastic simulations. The daily cases

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

of the COVID-19 epidemic in Hubei Province peaked at over 15,000 on February 12th and became almost zero in mid-March (Xinhua, 2020). Our simulation results indicate that the containment of this highly contagious disease was achieved by stringent control measures by the Chinese government. Without continued and aggressive control measures, the epidemic in Hubei Province would have become persistent. If the infection rate is reduced through the enhanced protective measures by 25%, the epidemic would reach the peak in mid-February and fade out in late September. Reducing the average node degree inside each city via social distancing measures, the number of infectious cases would peak at mid-February and decrease to zero at mid-June. Only with combined implementation of enhanced protective measures and social distancing measures, the epidemic dynamics would peak at around mid-February and approximate the actual epidemic trajectory in March. This can be an important message for countries going through the exponential growth of the epidemic in the current days.

References

Boguñá, M., Lafuerza, L. F., Toral, R., & Serrano, M. Á. (2014). Simulating non-Markovian stochastic processes. *Physical Review E*, *90*(4), 042108. https://doi.org/10.1103/PhysRevE.90.042108

Chan, J. F.-W., Yuan, S., Kok, K.-H., To, K. K.-W., Chu, H., Yang, J., … Yuen, K.-Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet*, *395*(10223), 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., … Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet*, *395*(10223), 507–513. https://doi.org/10.1016/S0140-6736(20)30211-7

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., … Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. https://doi.org/10.1126/science.aba9757

Du, Z., Wang, L., Cauchemez, S., Xu, X., Wang, X., Cowling, B. J., & Meyers, L. A. (2020). Early Release—Risk for Transportation of 2019 Novel Coronavirus Disease from Wuhan to Other Cities in China. *Emerging Infectious Diseases Journal-CDC*, *26*(5). https://doi.org/10.3201/eid2605.200146

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, *5*(1), 17–60.

Gardner, L. (2020, January 31). Update January 31: Modeling the Spreading Risk of 2019-nCoV. Retrieved February 19, 2020, from https://systems.jhu.edu/research/public-health/ncov-model-2/

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., … Zhong, N. (2020). Clinical characteristics of 2019 novel coronavirus infection in China. *MedRxiv*, 2020.02.06.20020974. https://doi.org/10.1101/2020.02.06.20020974

Health Commission of Hubei Province. (2020). Retrieved March 2, 2020, from http://wjw.hubei.gov.cn/

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Hubei Provincial Bureau of Statistics (2020). Hubei Province Statistical Yearbook 2019. Retrieved March 2, 2020, from http://tjj.hubei.gov.cn/tjsj/sjkscx/tjnj/qstjnj/

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Jazwinski, A. H. (2007). *Stochastic Processes and Filtering Theory*. Courier Corporation.

Jensen, J. P. (2007). Ensemble Kalman Filtering for State and Parameter Estimation on a Reservoir Model. *100*. Retrieved from https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/259624

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552

Keeling, M. J., & Rohani, P. (2011). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.

Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., … Tatem, A. J. (2020). Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. *MedRxiv*, 2020.03.03.20029843. https://doi.org/10.1101/2020.03.03.20029843

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., … Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*, *0*(0), null. https://doi.org/10.1056/NEJMoa2001316

Li, X., Zhao, X., & Sun, Y. (2020). The lockdown of Hubei Province causing different transmission dynamics of the novel coronavirus (2019-nCoV) in Wuhan and Beijing. *MedRxiv*, 2020.02.09.20021477. https://doi.org/10.1101/2020.02.09.20021477

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

Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S., … Nishiura, H. (2020). Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *Journal of Clinical Medicine*, *9*(2), 538. https://doi.org/10.3390/jcm9020538

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, *28*(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Nishiura, H., Jung, S., Linton, N. M., Kinoshita, R., Yang, Y., Hayashi, K., … Akhmetzhanov, A. R. (2020). The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020. *Journal of Clinical Medicine*, *9*(2), 330. https://doi.org/10.3390/jcm9020330

Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A., & Jewell, C. P. (2020). Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. *MedRxiv*, 2020.01.23.20018549. https://doi.org/10.1101/2020.01.23.20018549

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., … Hoelscher, M. (2020). Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *New England Journal of Medicine*, *382*(10), 970–971. https://doi.org/10.1056/NEJMc2001468

Sahneh, F. D., Scoglio, C., & Van Mieghem, P. (2013). Generalized Epidemic Mean-field Model for Spreading Processes over Multilayer Complex Networks. *IEEE/ACM Trans. Netw.*, *21*(5), 1609–1620. https://doi.org/10.1109/TNET.2013.2239658

medRxiv preprint doi: https://doi.org/10.1101/2020.03.27.20045625; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Sahneh, F. D., Vajdi, A., Shakeri, H., Fan, F., & Scoglio, C. (2017). GEMFsim: A stochastic simulator for the generalized epidemic modeling framework. *Journal of Computational Science*, *22*, 36–44. https://doi.org/10.1016/j.jocs.2017.08.014

Sanche, S., Lin, Y. T., Xu, C., Romero-Severson, E., Hengartner, N. W., & Ke, R. (2020). The Novel Coronavirus, 2019-nCoV, is Highly Contagious and More Infectious Than Initially Estimated. *ArXiv:2002.03268 [q-Bio]*. Retrieved from http://arxiv.org/abs/2002.03268

Shen, M., Peng, Z., Guo, Y., Xiao, Y., & Zhang, L. (2020). Lockdown may partially halt the spread of 2019 novel coronavirus in Hubei province, China. *MedRxiv*, 2020.02.11.20022236. https://doi.org/10.1101/2020.02.11.20022236

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *The Lancet*. https://doi.org/10.1016/S0140-6736(20)30260-9

Wu, T., Ge, X., Yu, G., & Hu, E. (2020). Open-source analytics tools for studying the COVID-19 coronavirus outbreak. *MedRxiv*, 2020.02.25.20027433. https://doi.org/10.1101/2020.02.25.20027433

Xinhua (2020). China Focus: China says its COVID-19 peak is over—Xinhua | English.news.cn. Retrieved March 26, 2020, from http://www.xinhuanet.com/english/2020-03/13/c_138871417.htm

Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., … Wang, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*, *92*, 214–217. https://doi.org/10.1016/j.ijid.2020.01.050

Zhou, T., Liu, Q., Yang, Z., Liao, J., Yang, K., Bai, W., … Zhang, W. (2020). Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *ArXiv:2001.10530 [Physics, q-Bio]*. Retrieved from http://arxiv.org/abs/2001.10530

Zhu, X., Zhang, A., Xu, S., Jia, P., Tan, X., Tian, J., … Yu, J. (2020). Spatially Explicit Modeling of 2019-nCoV Epidemic Trend based on Mobile Phone Data in Mainland China. *MedRxiv*, 2020.02.09.20021360. https://doi.org/10.1101/2020.02.09.20021360