medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Spatially Explicit Modeling of 2019-nCoV Epidemic Trend based on Mobile Phone Data in Mainland China

Xiaolin Zhu[1,*], Aiyin Zhang[1], Shuai Xu[1], Pengfei Jia[2], Xiaoyue Tan[1], Jiaqi Tian[1], Tao Wei[3], Zhenxian Quan[4], Jiali Yu[2]

[1]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China
[2]China Academy of Urban Planning and Design, Beijing, 100044, China
[3]School of Psychology, Shenzhen University, Shenzhen, 518060, China
[4]Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

*Corresponding author: xiaolin.zhu@polyu.edu.hk

Abstract

In December 2019, Wuhan, China reported an outbreak of atypical pneumonia caused by the 2019 novel coronavirus (2019-nCoV). As of February 7, 2020, the total number of the confirmed cases in mainland China reached to 34,546 of whom 722 have died and 2,050 recovered. While most Chinese cities have confirmed cases, the city-level epidemical dynamics is unknown. The aim of this study is to model the dynamics of 2019-nCoV at city level and predict the trend under different scenarios in mainland China. We used mobile phone data and modified the classic epidemiological Susceptible - Infectious - Recovered (SIR) model to consider several unique characteristics of the outbreak of 2019-nCoV in mainland China. The modified SIR model was trained using the confirmed cases from January 25 to February 1 and validated by the data collected on February 2, 2020. The prediction accuracy of new infected cases on February 2 (R^2 = 0.94, RMSE = 18.24) is higher than using the classic SIR model (R^2 = 0.69, RMSE = 40.18). We used the trained model to predict the trend in the next 30 days (up to March 2, 2020) under different scenarios: keeping the early-stage trend, controlling the disease as successfully as SARS in 2003, and increasing person-to-person contacts due to work/school resuming. Results show that the total infected population in mainland China will be 10.53, 0.15, and 0.41 million and 67%, 100%, 91% Chinese cities will control the virus spreading by March 2, 2020 under the above three scenarios. Our study also provides the city-level spatial pattern of the epidemic trend for decision makers to allocate resources for controlling virus spreading.

Keywords: epidemiology, novel coronavirus, 2019-nCoV, SIR model, disease control

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/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

1. Introduction

Wuhan, a large city with 14 million residents and a major air and train transportation hub of central China, identified a cluster of unexplained cases of pneumonia on December 29, 2019 (Li et al., 2020). Four patients were initially reported and all these initial cases were linked to the Huanan Seafood Wholesale Market (Zhu et al., 2020). Chinese health authorities and scientists did immediate investigation and isolated a novel coronavirus from these patients by January 7, 2020, which is then named as 2019-nCoV by the World Health Organization (Chen et al., 2020; Wang et al., 2020). 2019-nCoV can cause acute respiratory diseases that progress to severe pneumonia (Huang et al., 2020). The infection fatality risk is around 3% estimated from the data of early outbreak (Perlman, 2020; Wang et al., 2020). Information on new cases strongly indicates human-to-human spread (Fuk-Woo Chan et al., 2020; Li et al., 2020; Riou and Althaus, 2020). Infection of 2019-nCoV quickly spread to other cities in China and other countries (Figure 1). It becomes an event of global health concern (Hui et al., 2020). Up to February 7, 2020, according to the reports published by the Chinese Center for Disease Control and Prevention, all provinces of mainland China have confirmed cases and the total number reaches to 3,4546, of whom 722 have died and 2050 recovered; 24 oversea countries have 285 confirmed cases (1 died). Chinese government took immediate actions to control the spread of disease, including closing the public transportation from and to Wuhan on January 23, extending the Spring Festival holiday, postponing the school-back day, and suspending all domestic and international group tours.

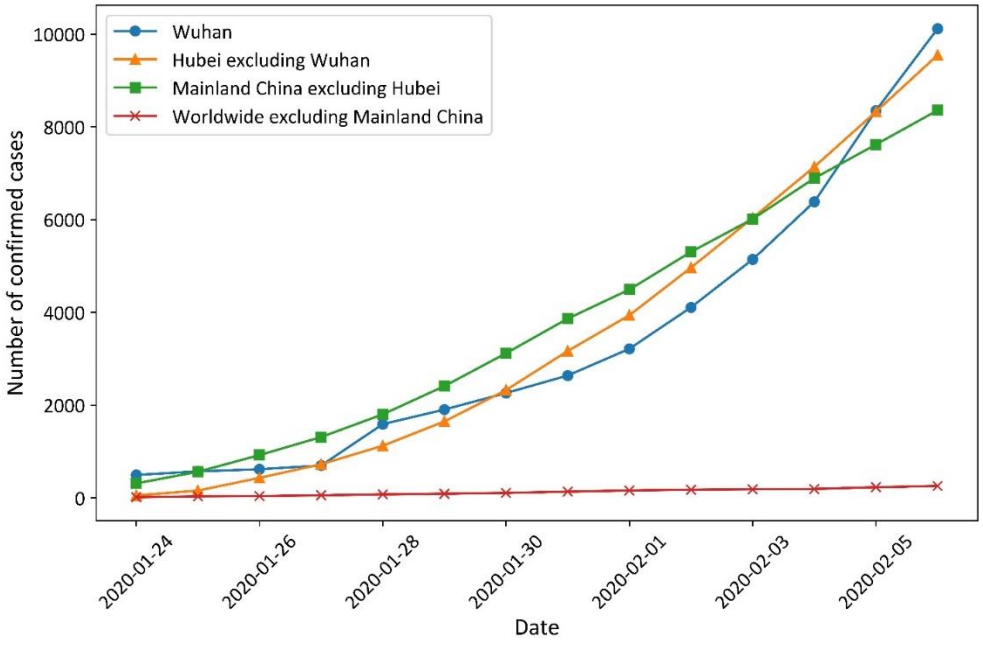

Figure 1. Cumulative number of confirmed cases of 2019-nCoV as of February 6, 2020, in Wuhan, Hubei province excluding Wuhan, mainland China excluding Hubei province, and outside mainland China.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

Unfortunately, many external factors bring a challenge to control virus spreading. First, it might be already late to stop the migration of infected cases to places outside Wuhan. All initial cases lived in Wuhan and many confirmed cases in other cities have traveled to Wuhan, suggesting that Wuhan is the center of 2019-nCoV outbreak. However, around 5 million Wuhan residents left Wuhan in January 2020 due to the Spring Festival (January 24, 2020). It is a Chinese tradition that people come back to hometown to celebrate the Spring Festival with their parents and visit relatives. It is very likely that a considerable number of infected cases have moved from Wuhan to other cities before the Spring Festival (Wu et al., 2020), although Wuhan government implemented border control on January 23. Second, because many infected cases have symptoms similar with the cold and flu, such as fever and cough, people may not be aware of their infection of 2019-nCoV, especially before the Chinese central government announced measures to the general public for preventing the transmission of disease on January 20, 2020. A study based on 425 patients at the early stage of outbreak revealed that the time from infection to illness onset is 5.2 days (Li et al., 2020). As a result, presymptomatic cases who have left Wuhan may not be isolated themselves from their family and relatives (Munster et al., 2020). It is highly possible that these infected cases spread the virus to their family members or relatives (Fuk-Woo Chan et al., 2020). Third, due to the sudden outbreak of virus, the preparation and resources for preventing virus transmission are limited. It was reported that the protective equipment in many hospitals in Wuhan was in short supply so that it is difficult to maintain strict personal hygiene. With the quick increase of infected cases, Wuhan and other cities in Hubei Province have large pressure to isolate and give medical treatment to infected people. All above factors can make preventing the spread of 2019-nCoV even more difficult than the severe acute respiratory syndrome (SARS), another coronavirus outbreak in China 17 years ago that caused more than 8000 infections and 800 deaths.

Projecting the epidemic trend of 2019-nCoV outbreak is critical for the decision makers to allocate resources and take appropriate actions to control virus transmission. Right after the outbreak, several studies have retrieved the epidemiological parameters and predicted the future situation (Nishiura et al., 2020; Read et al., 2020; Shen et al., 2020; Zhao et al., 2020). These studies used the reported cases at the early stage of outbreak and modelled epidemic dynamics in Wuhan or nation-wide. The basic reproductive number (R_0) from these studies ranges from 2.68 to 5.47. A recent study (Wu et al., 2020) used air passenger data and social medium data to forecast the spread of 2019-nCoV in Wuhan and other major Chinese cities. They estimated that 75,815 individuals have been infected in Greater Wuhan as of January 25, 2020 and epidemics are already growing exponentially in major cities of China with a 1-2 weeks lag time behind Wuhan outbreak. Although these studies at the early stage of outbreak help us to understand some important epidemic information of 2019-nCoV, the fine-scale epidemic trend in all Chinese cities remains unknown, which is more helpful for allocating medical resources to achieve the optimal result for preventing disease spreading.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

To model the fine-scale epidemic dynamics of all individual cities in mainland China, we proposed a spatially explicit approach. We first used mobile phone data to obtain the number of people that traveled from Wuhan to each individual city, then modified the classic epidemiological Susceptible - Infectious – Recovered (SIR) model to fit the dynamics of 2019-nCoV at the city level and finally used this modified model to predict the trend under different scenarios. Particularly, our model considers each individual city with two groups of susceptible population, i.e., local residents and those from Wuhan, because they may have different transmission rate (β). Parameters of our model were retrieved by training the model with the daily-confirmed cases at the city level and the calibrated model was then used to predict the trend in the next 30 days under three conditions: keeping the early-stage trend, successfully controlling the spread as SARS, and increasing person-to-person contacts during the transportation post to the Spring Festival.

2. Data

We collected the daily outbreak data of 2019-nCoV Pneumonia in 334 prefecture-level cities in mainland China from January 11 to February 2, 2020, including confirmed, dead and cured cases from an online platform reporting real-time statistics of 2019-nCoV (https://ncov.dxy.cn/ncovh5/view/pneumonia). These daily reported data were used to train and validate our epidemic model. We employed China Unicom mobile phone database (https://www.cubigdata.cn) to obtain the inter-city human mobility. China Unicom is one of three largest mobile service providers in China. It has 0.32 billion users. Considering 2019-nCoV originating from Wuhan and the initial cases were reported around January 1, 2020, we inferred the number of people who have Wuhan travel history during January 1-31, 2020 in each city based on the mobile phone dataset (Figure 2 and Appendix Table 1). In addition, Household Registered Population at 2017 year-end derived from census data was used to approximate the number of local residents in each city during 2020 Spring Festival (Appendix Table 1).

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

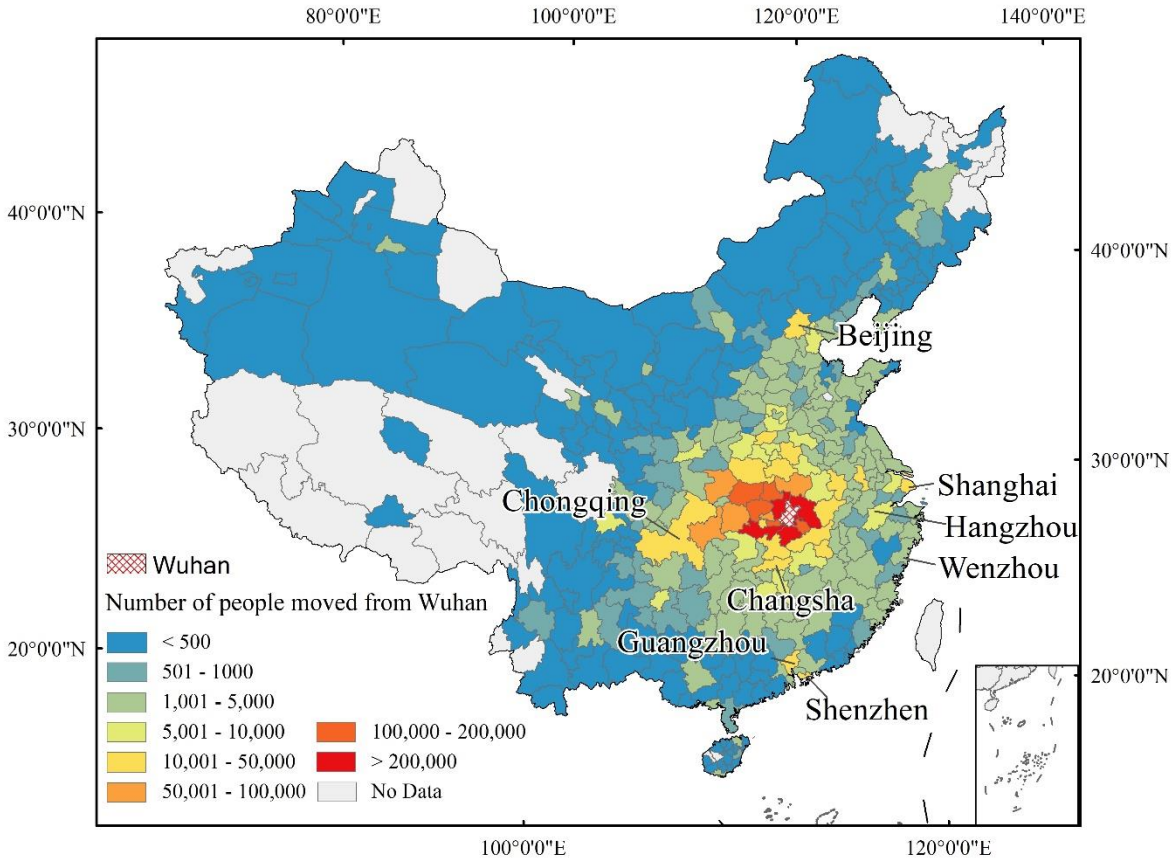

Figure 2. Population from Wuhan to other cities in mainland China during January 1-24, 2020

3. Methodology

3.1 Modification of SIR model

Our proposed model stems from the SIR model, a classic approach to simulate epidemiological dynamics. The model comprises three variables: S for the susceptible population, I for the infectious, and R for the recovered/removed. This model explicitly quantified the full cycle of diseases spreading in human-to-human. The above three variables change over time through the following differential equations (Equation 1).

$$\begin{cases} \frac{dS}{t} = -\frac{\beta SI}{N} \\ \frac{dI}{t} = \frac{\beta SI}{N} - \gamma I \\ \frac{dR}{t} = \gamma I \end{cases} \tag{1}$$

In Equation (1), each variable was regarded as a function of time (t): $S = S(t)$, $I = I(t)$, $R = R(t)$, and N denoted the total population number ($N = S + I + R$); β and γ denote the daily transmission rate and daily recovery rate respectively. Susceptible-Exposed-Infectious-

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

Recovered (SEIR) model is another widely used epidemic model that includes exposed population. A recent study found that the latent/exposed cases can also spread the disease (Rothe et al., 2020). Therefore, we treated exposed cases as infected cases in the classic SIR model rather than using SEIR model.

We further modified the SIR structure based on the unique characteristics of the outbreak of 2019-nCoV. First, in all cities other than Wuhan, the initial infectious cases are most likely imported from Wuhan. Second, many Wuhan residents moved to other cities due to the Spring Festival and this mobility was closed after the border control on January 23. Third, those people from Wuhan have low contacts with local residents because Chinese government required them to implement self-quarantine. Accordingly, in the modified SIR model, the susceptible variable S was divided into two groups: S_1, the number of susceptible without Wuhan travel history, and S_2, the number of susceptible traveling from Wuhan. These Wuhan-inbound groups (S_2), potentially carrying virus with higher transmission rate (β_2), differed from local residents (S_1) with relatively lower transmission rate (β_1) as Chinese government took measures to reduce the person-to-person contacts ($\beta_2 \geq \beta_1$). In our modified SIR model, recovered population I was extended to include those cured, died, and isolated in hospital because they cannot transmit the virus. The differentiate equations of our modified SIR model is as follows:

$$\begin{aligned} \frac{dS_1}{t} &= -\frac{\beta_1 S_1 I}{N_o} \\ \frac{dS_2}{t} &= -\frac{\beta_2 S_1 I}{N_w} \\ \frac{dI}{t} &= \left(\frac{\beta_1 S_1}{N_o} + \frac{\beta_2 S_1}{N_w} - \gamma\right) I \\ \frac{dR}{t} &= \gamma I \end{aligned} \tag{2}$$

Where N_o is the total local population derived from the census data, and N_w represent the total inflow population from Wuhan to each city during January 1 to January 24, estimated from mobile phone data provided by China Unicom.

In the modified SIR model, four variables need to be initialized: (1) initial number of infectious I_0, treated as a parameter to be estimated (see section 3.2); (2) initial number of local susceptible S_{01}, equal to the total number of the local population of the City $S_{01} = N_o$; (3) initial number of susceptible traveling from Wuhan S_{02}, equal to the population from Wuhan excluding the initial infectious from Wuhan $S_{02} = N_w - I_0$; and (4) initial number of removed R_0, assuming no recovered, hospitalized, or death at initial state $R_0 = 0$.

3.2 Estimation of model parameters

Our modified SIR model has three important parameters: transmission rate β_1 among local residents, β_2 among people from Wuhan, and recovery rate γ. For γ, we assume that once an infected individual is hospitalized, the person will be segregated and therefore no longer infectious. According to a recent study using the first 425 patients (Li et al., 2020), the mean

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

incubation period of 2019-nCoV is 5.2 days, and the mean duration from illness onset to hospital admission is 9.1 days. We assume that the incubation period and duration from illness onset to first medical visit is similar with these 425 infected cases. Therefore, the estimated infectious period is $5.2 + 9.1 = 14.3$ days and γ equals $1/14.3 = 0.0699$.

For parameters β_1 and β_2, we used the daily cumulative confirmed cases at early outbreak stage (up to February 1, 2020) to retrieve their optimal values, because early-stage dynamics is less affected by prevention interventions. We first estimated the optimal value of β_2 using reported data of Wuhan since the epidemic model for Wuhan only has parameter β_2. Then, the estimated β_2 was used as a prior parameter for the estimation of β_1 for individual prefecture-level cities. The Nelder-Mead algorithm (Nelder and Mead, 1965) was employed by this study to estimate parameters through minimizing the sum of squared differences between the simulated and actual daily cumulative confirmed cases. Since the epidemic model is highly sensitive to the initial infectious number, and the reported confirmed cases of the first few days may be much fewer than actual cases because the public has just raised awareness of the virus. Therefore, using the number of reported cases on the first onset day as initial infectious number in the model can cause large uncertainty for the estimation of β_1 and β_2. To reduce the impact of the initial inputs, we did not use the reported data of the first few days and treated the initial infectious number I_0 as another parameter to be estimated together with β_1 and β_2. Specifically, estimation of β_2 used daily cumulative confirmed cases of Wuhan from January 17 to February 1, 2020 and assumed January 1, 2020 as the start point. Estimation of β_1 used reported data from January 25 to February 1, 2020 and assumed January 20 as the start point when massive inter-city mobility happened before the Spring Festival. Comparison between the result of model simulation using the estimated parameters and the reported cases was used to evaluate the goodness of model fitting (see some examples in Figure 3).

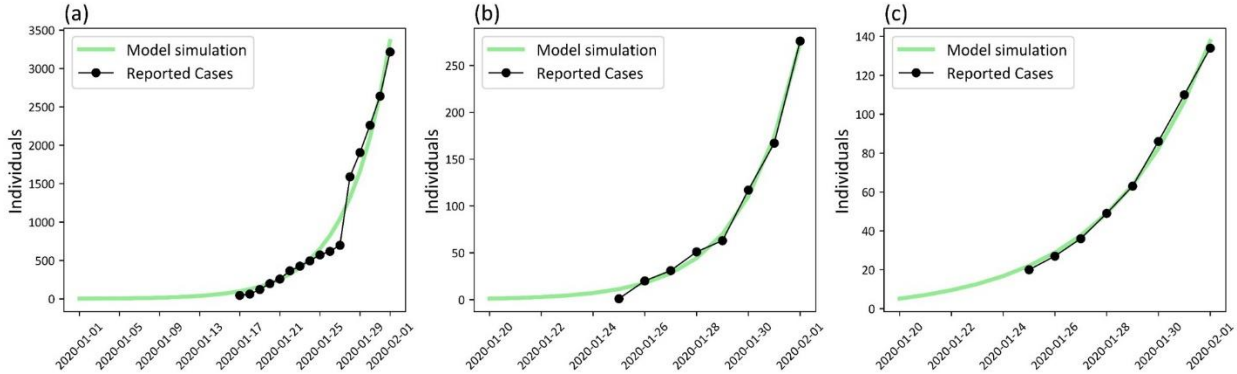

Figure 3. Comparison between result of model fitting and reported confirmed cases of (a) Wuhan (β_2 = 0.303), (b) Yichang (β_1 = 0.221) and (c) Shenzhen (β_1 = 0.022)

To examine the goodness of model fitting, we calculated R-squared and root mean square error (RMSE) between the fitting result and confirmed cases for each city, the median value of R-squared and RMSE is 0.89 and 1.39 respectively, indicating our model can well fit the early-stage spreading trend in majority of cities. However, for some cities with small number of

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

confirmed cases or with data errors, the fitted model is unsatisfactory. Therefore, for these cities with R-squared less than 0.70, the β_1 value of the city was treated as failure of estimation. 63 out of 237 cities were eliminated from parameter estimation in this process. Abnormally high β_1 estimation (> 3rd quantile+1.5 IQR) of four cities outside Hubei province was substituted by 90% quantile values of all cities. In addition, to ensure enough data points for fitting the model, β_1 was only estimated for cities with daily reported confirmed cases of 5 consecutive days or longer. For cities cannot estimate β_1 directly from report data, their β_1 values were spatially interpolated by the inverse distance weighted interpolation from their neighboring cities (i.e., cities with common borders), or assigned to the lower quartile of β_1 from all cities if their neighboring cities also do not have a valid β_1 estimation.

To test the accuracy of our proposed model, we used the reported confirmed cases on February 2 to validate our simulation result. We compared the differences between both the cumulative and daily increase of the predicted infected cases of February 2 with the reported confirmed cases of the same day. The coefficient of determination, R-squared and the RMSE were calculated to examine the model accuracy. To demonstrate the effectiveness of our modified SIR model, the results was also compared with the prediction from the classic SIR model that uses nationwide uniform parameters.

3.3 Prediction of different scenarios

The epidemic trend is the joint effect of virus transmissibility and outbreak control (Anderson et al., 2004). The control mechanism of diagnosis and isolation was successfully applied in many megacities during SARS period and proved to contain the spread of the virus (Chowell et al., 2003). Beneficial from the SARS experience, the government at all levels have responded quickly to limit the movement of people at the beginning of outbreak period. However, different future scenarios of virus spread may occur due to the influence of other factors such as the massive transportation post to the Spring Festival and the following work resuming. Meanwhile, the epidemiologic feature of novel coronavirus, a shorter double period and infectious incubation period compared with SARS, adds to the uncertainty (Li et al., 2020). Hence, we estimated the different epidemic trends in a 30-day period (up to March 2, 2020) through the model parameter manipulation, namely, transmission rate β to reveal the effectiveness of government control and removal rate γ to represent the promptness of medical treatment. Three scenarios were designed, and their parameter's temporal variations of one example were shown in Figure 4. The adjustment of model parameters starts from February 2, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

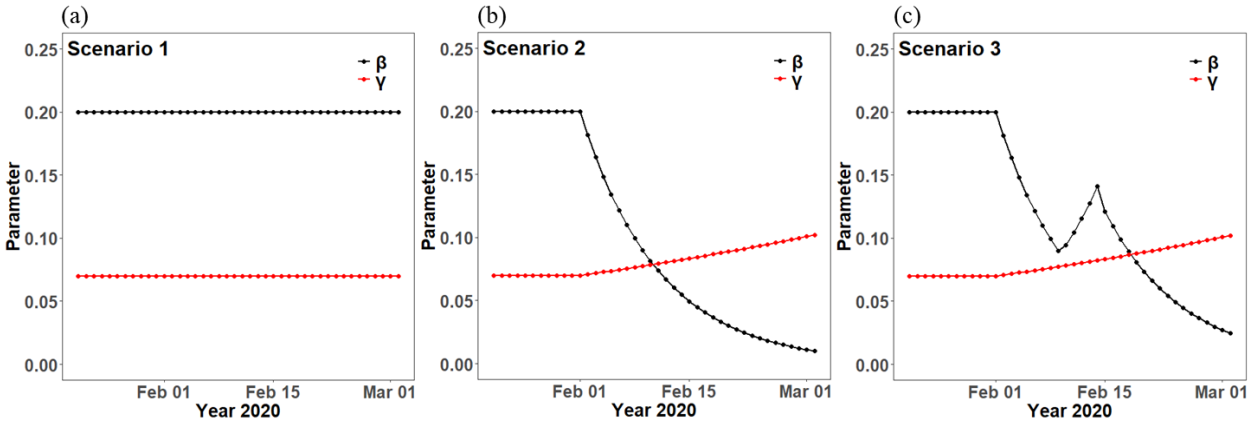

Figure 4. An example of model parameters' temporal variation under three scenarios: (a) keeping the trend of early stage, (b) controlling the spread as successfully as SARS, and (c) increasing person-to-person contacts due to work/school resuming after Spring Festival

Scenario 1 denoted the pessimistic view about the virus spread trend. The stable value of β and γ revealed the failure of outbreak control and implied the free propagation of virus, which is similar with the exponential increase of original SIR model. On the contrary, scenario 2 estimated the future trend under the successful control, with continuously decreasing transmissibility and increasing medical care capacity. According to the history of controlling SARS 17 years ago, we assumed that this novel coronavirus could be fully controlled within a month, which would push down the transmissibility of each city to very low levels (e.g., β equals to 0.01) (Lai, 2005). The exponential function was used to describe the attenuation of β in our study (Fisman et al., 2014), which has been employed to simulate the control process of other infectious disease such as the MERS-coronavirus. For those cities that have two susceptible groups, their β_1 and β_2 values would decrease concurrently. The change of γ value illustrated more available hospital beds in the future so that the average diagnostic isolation time reduced from 14.3 to 9.8 days. In scenario 3, the interference of work/school resuming was considered, so a short rebound was introduced to the transmission rate β (i.e. rate of February 10-14 gradually increases to that of February 5). The γ in scenario 3 would keep the same increasing trend as scenario 2.

4. Results

4.1 Results of parameter estimation

Four parameters were estimated in our modified SIR model, namely transmission rates β_1 and β_2, removal rate γ, and initial infectious population I_0. As aforementioned in section 3.2, we estimated the transmission rate β_2 and the recovery rate γ to be 0.303 and 0.0699 respectively, and the transmission rate β_1 and initial infectious population I_0 vary from city to city (Appendix Table 2). Figure 5 (a) illustrated estimation results of β_1 in cities with sufficient daily reported

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

confirmed cases (namely direct estimation). β_1 values of other cities without direct estimation were spatially interpolated (Figure 5 (b)). As the transmission rate among local residents in each city, β_1 reflects the intensity of control measures adopted by each local government, as well as the awareness of citizens to take protective measures. For example, β_1 in megacities such as Beijing, Shanghai, Guangzhou and Shenzhen are surprisingly low considering their intensive traffic and population mobility, which may attribute to higher health literacy of their citizens (Shen et al., 2015). Similar to the estimation process of β_1, we estimated initial infectious population I_0 by employing the direct estimation and interpolation, results are shown in Figure 5 (c) and (d) respectively. It shows that I_0 is highly spatially clustered, which is similar to the spatial distribution of confirmed cases. Cities with higher I_0 are mainly concentrated in eastern China, especially in Hubei Province and its surrounding areas, Pearl and Yangtze River delta region, as well as highly populated cities such as Beijing and Chongqing. These places may import many infected cases from Wuhan before Spring Festival (Wu et al., 2020).

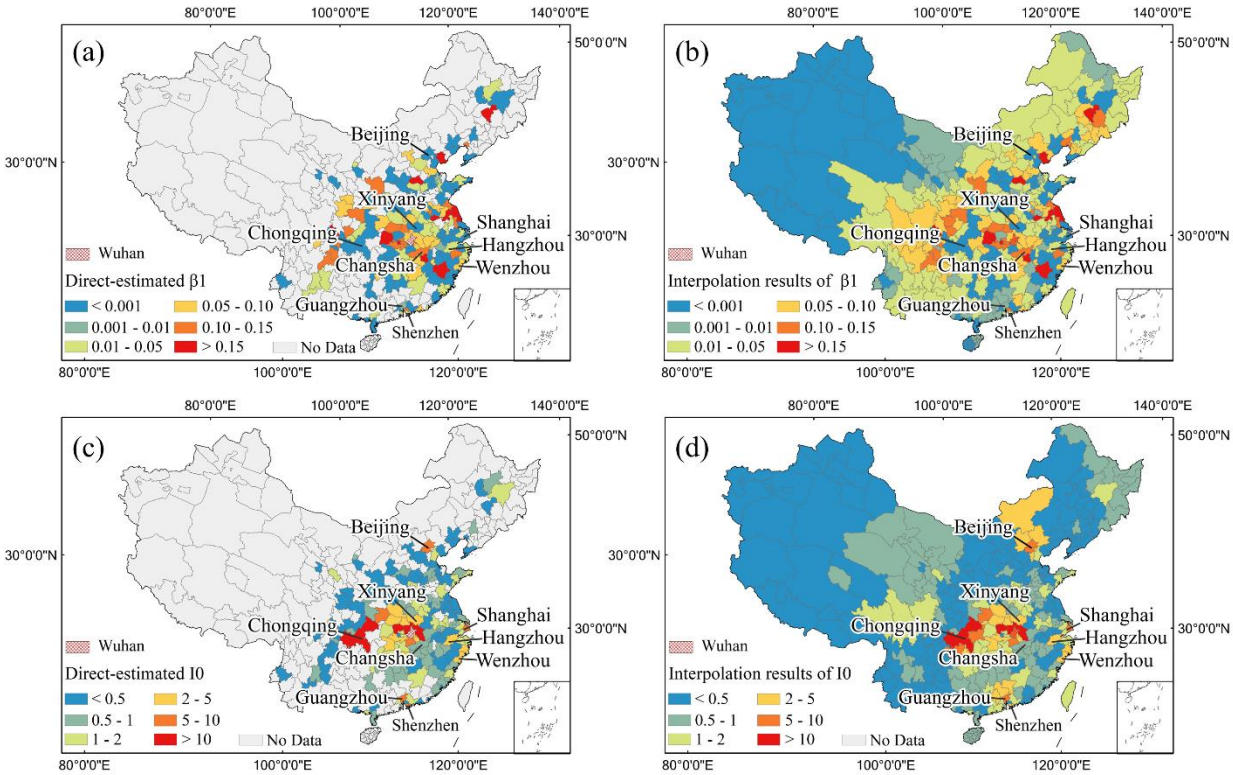

Figure 5. Results of parameter estimation: β_1 (a) and I_0 (c) of cities estimated using daily confirmed cases; β_1 (b) and I_0 (d) of other cities spatially interpolated.

4.2 Results of model validation

The modified SIR model with estimated parameters of each individual city was used to predict the number of cumulative infected cases (including removed cases) and the number of new infected cases on February 2, 2020. The predicted values by the modified SIR model well

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

matched the reported cumulative infected cases (R-squared: 0.99, RMSE: 27.34, P < 0.0001, Figure 6.a). The capacity of the modified SIR model for predicting daily increase of new infected cases is also acceptable (R-squared: 0.94, RMSE: 18.24, P < 0.0001, Figure 6.b). The fitted regression line is close to the 1:1 reference line in Figures 6 (a) and (b), indicating no bias in prediction. To demonstrate the effectiveness of our modified SIR model, the results was compared with the prediction from the classic SIR model that uses nationwide uniform parameters (β= 0.332, γ=0.0699) (Figures 6.c and d). Compared with the modified SIR model, the predicted values by the classic SIR model does not match very well with the cumulative infected cases (R-squared: 0.77, RMSE: 158.73, P < 0.0001, Figure 6.c) or the daily new infected cases (R-squared: 0.69, RMSE: 40.18, P < 0.0001, Figure 6.d). The predicted values by the classic SIR model significantly overestimated the infected number than the reported cases, where the fitted regression line is clearly below the 1:1 reference line.

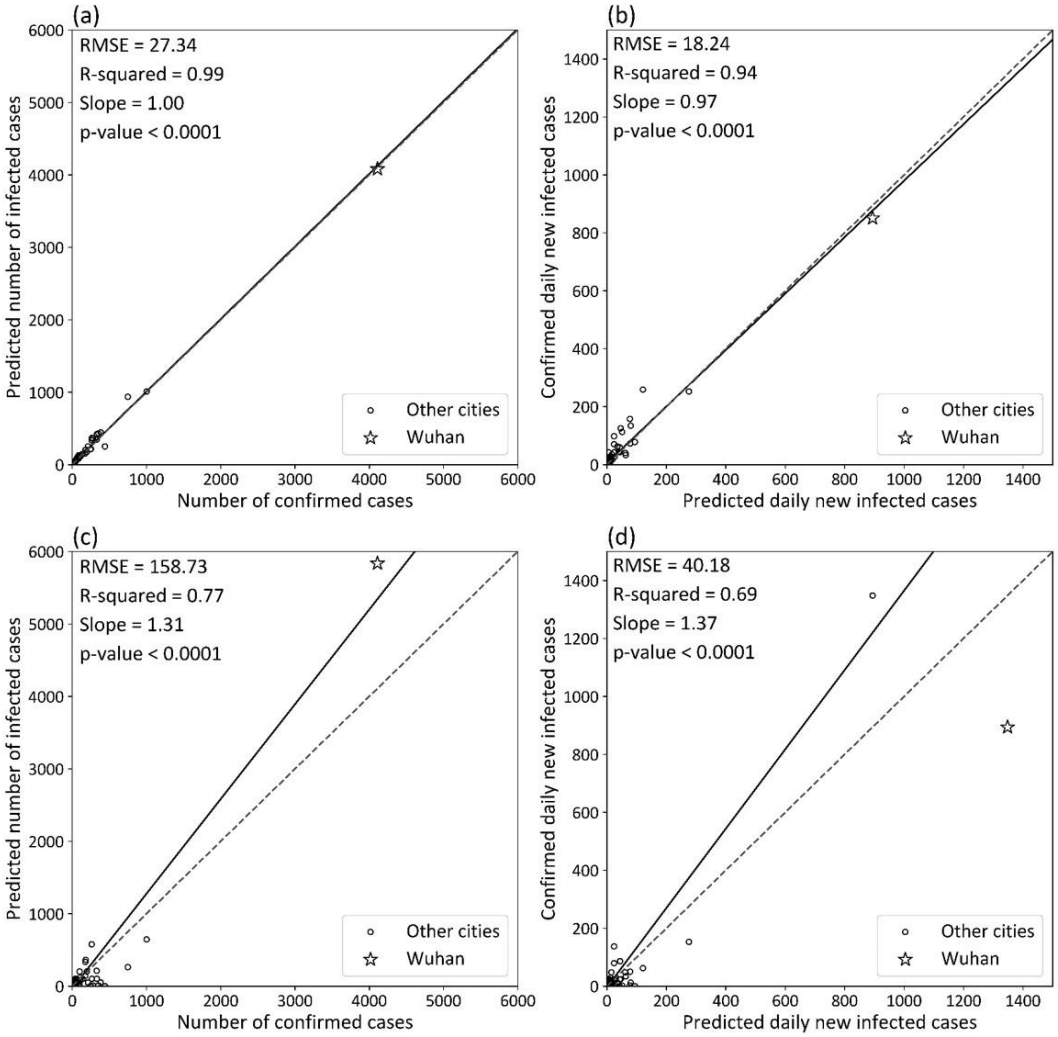

Figure 6. Model validation of the modified SIR model (cumulative case (a) and daily new case (b)) and classic SIR model (cumulative case (c) and daily new case (d)) based on data of February 2, 2020. The solid lines represent the fitted linear regression curve and dashed lines represent 1:1 lines for reference.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

4.3 Three scenarios of epidemic dynamics

We predicted changes of infected cases and removed cases of each city up to March 2, 2020 under three different scenarios (scenario 1 - keeping the trend; scenario 2 - controlling the disease as successfully as SARS in 2003; and scenario 3 - increasing person-to-person contacts due to work/school resuming). Our prediction shows that the whole mainland China will have 10,529,530, 148,137, and 411,082 people infected up to March 2, 2020 under the above three scenarios respectively. To provide an intuitive picture about epidemic dynamics in different scenarios, we showed in Figure 7 the number of active infections (*S* variable in our model, Appendix Table 3) in each city on February 15 and March 2 respectively. The infected people will mainly distribute in the central and eastern provinces, the number of western cities at a relatively low level under all scenarios. Different scenarios lead to different spatial and temporal changes. A significant increase between February 15 and March 2 is observed in nearly all cities under scenario 1, especially for Hubei and adjacent provinces (Figures 7.a and d). In scenario 2, active infected cases in many cities will be lower than 100 by March 2, although the number of infected people in Hubei province is still quite high (Figure 7.b and e). In scenario 3, if the government does not restrict increasing transportation post to the Spring Festival holiday, the cities close to Wuhan will face a severe situation that the active infections will be over 5000 on March 2 (Figure 7.c and f). Other regions such as the Pearl and Yangtze River delta will also be influenced by the increasing person-to-person contact.

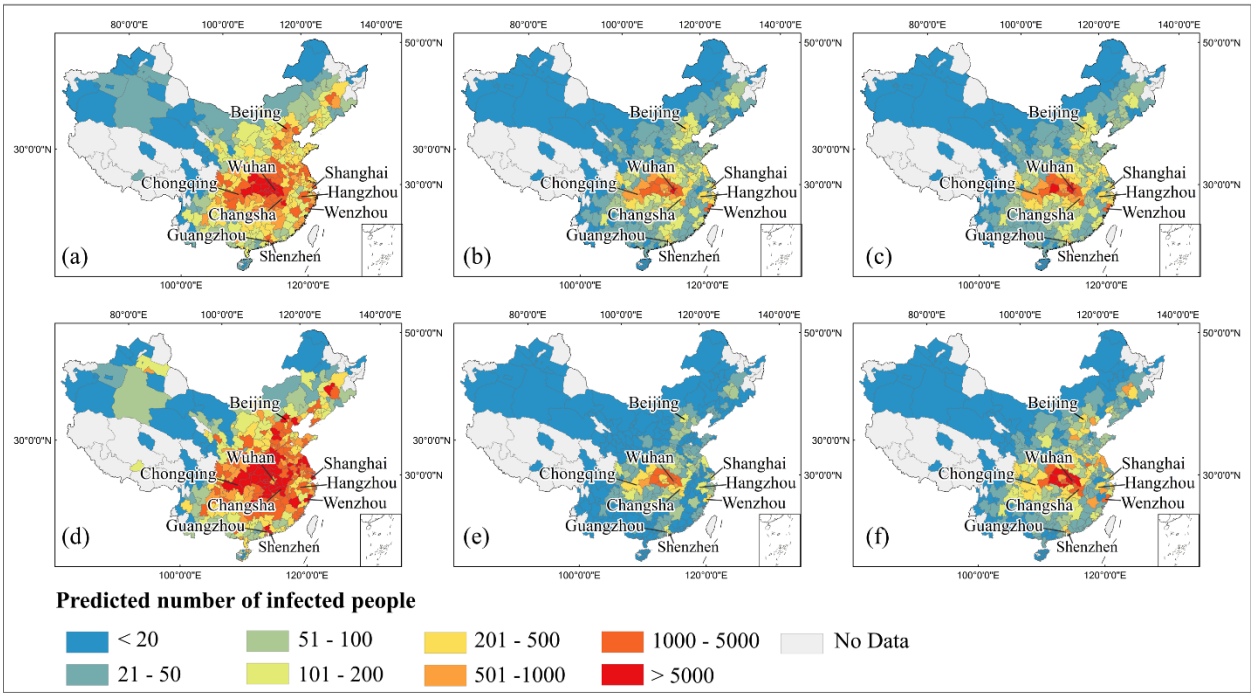

Figure 7. Mapping of predicted number of infected people on February 15: (a) scenario 1, (b) scenario 2, (c) scenario 3 and March 2: (d) scenario 1, (e) scenario 2, and (f) scenario 3

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

To better understand the specific attributes of epidemic dynamics under different scenarios, we investigated the temporal changes of active infections across all cities in mainland China. In Figure 8, we show the results in provinces with large number of confirmed cases, first-tier cities, and cities within Wuhan one-hour economic circle. If there is no reduction in transmissibility (scenario 1), the active infections in Wuhan will grow exponentially during the whole period of prediction, while in other provinces and cities the epidemic peak will emerge in late February. With strict restriction on the movement of people and isolation (scenario 2), in all these provinces and cities, the active infections will decrease around middle February. The comparison between results of scenario 1 and 2 demonstrates the urgency and effectiveness of city-level quarantine to Hubei province. In scenario 3 where transmissibility rebounds after the public holiday in all cities, the peak of active infections will postpone ten days and the magnitude will increase by about 50% compared with scenario 2. Our simulation suggested that strict quarantine of inner- and inter-city population movement during February would have a significant effect on the suppression of virus spreading.

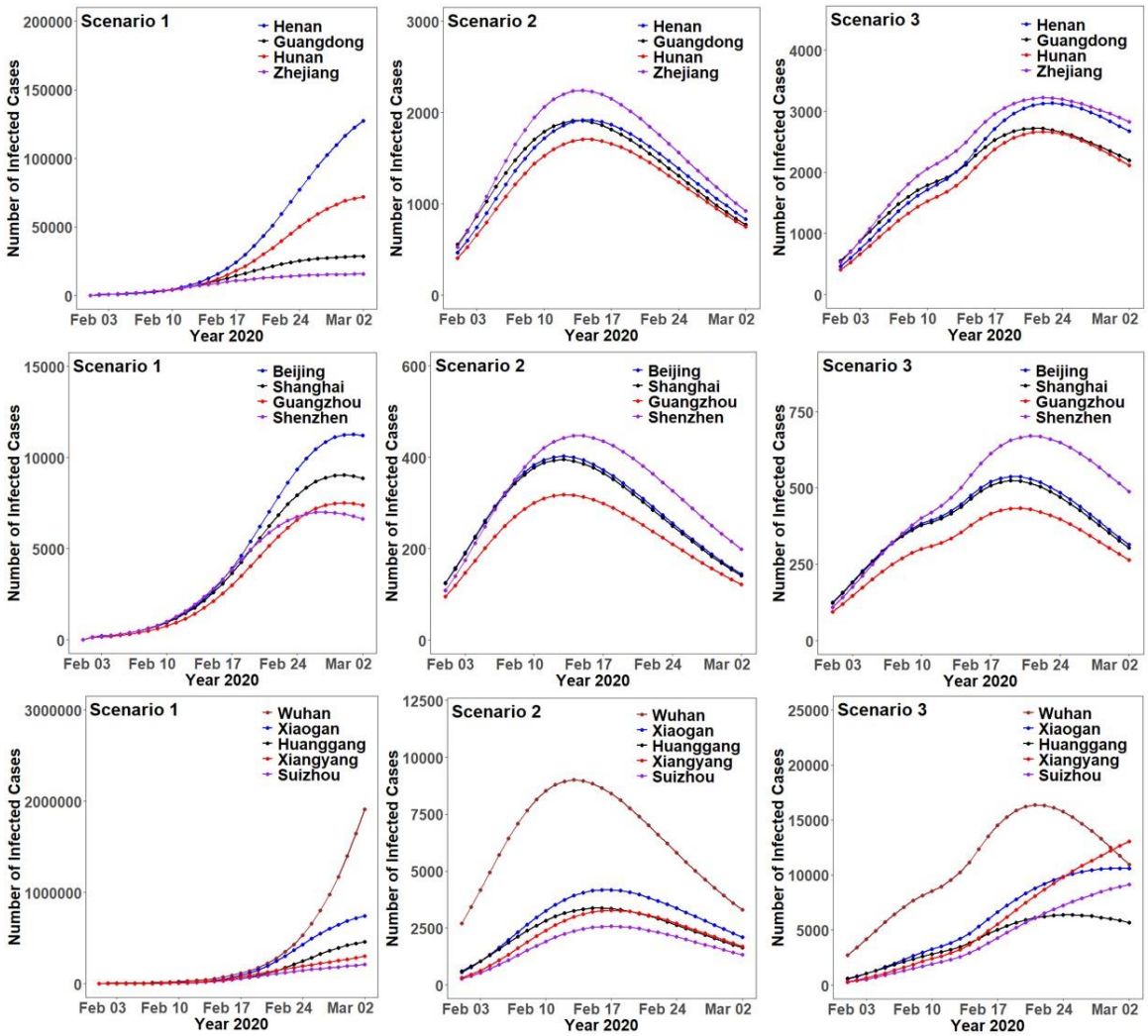

Figure 8. Temporal change of active infections in representative provinces and cities based on the modified SIR model

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

Another critical attribute of epidemic dynamics is the net daily increase of infections, i.e., the difference between new infections and removal in one day. Figure 9 shows the temporal changes of net daily increase of infections in three scenarios. Among all three scenarios, Hubei provinces will contribute the majority of increments in mainland China. Under different scenarios, the peak of net daily increase of infections emerges at different time. With no strict restriction (scenario 1), the peak point was not observed during the prediction period. The peak point will appear around February 10 with strict restriction (scenario 2), but it will shift to middle February or later with work/school resuming (scenario 3).

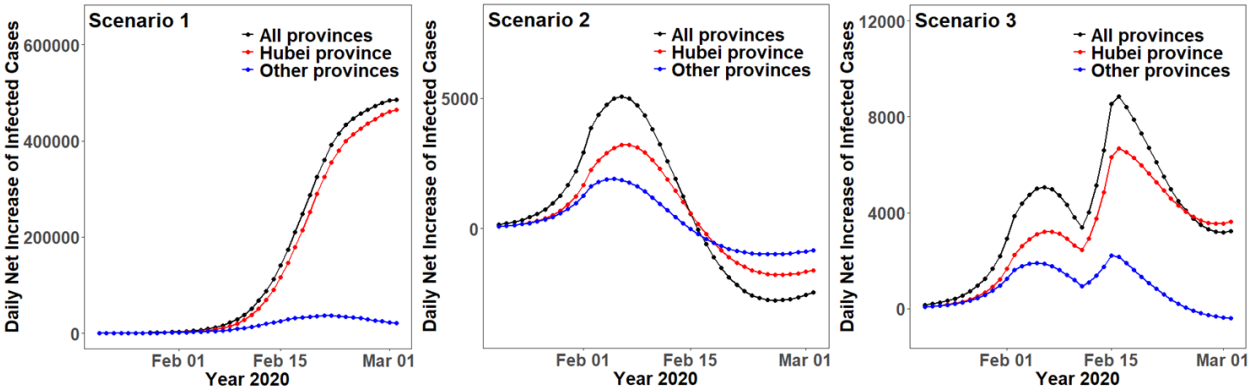

Figure 9. Net daily increase of infections in Hubei and other provinces under three scenarios

We consider the net daily increase of infections lower than one case as the signal of effective control of coronavirus outbreak. Among all 333 cities in mainland China, our modelling shows that 67%, 100%, and 91% will successfully control the disease by March 2 under the three scenarios respectively (Appendix Table 3). For those 17 cities within Hubei province, 0%, 100%, and 53% will control the disease by March 2 under the three scenarios respectively. We mapped the date when a city successfully controls the coronavirus outbreak in Figure 10. The results show that some western cities will have no possibility of virus outbreak, as the predicted net daily increase is always lower than one in all scenarios. The spatial pattern of successful control date is similar across three scenarios, a distinguished gradient decrease from central region (Hubei, Henan, Hunan and Chongqing) to the peripheral regions. The urban agglomeration along the coastline would have late dates of successful control, especially in scenario 3 with work/school resuming after the Spring Festival.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

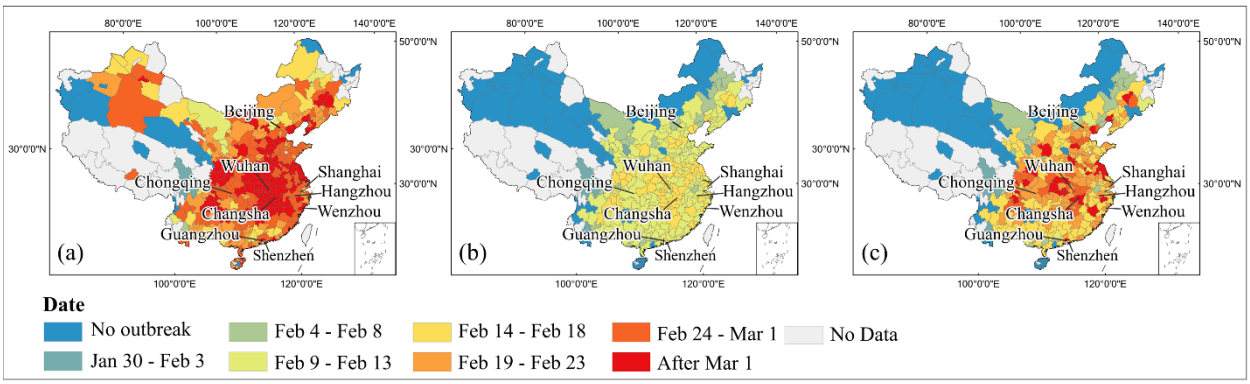

Figure 10. Dates when the coronavirus outbreak will be controlled in each city of mainland China under scenarios: (a) keeping the trend of early stage, (b) controlling the spread as successfully as SARS, and (c) increasing person-to-person contacts due to work/school resuming after Spring Festival

5. Discussion

In this study, we modeled the epidemic trend of 2019-nCoV for each individual city in mainland China and used the model to predict the future scenarios under different conditions. Compared with other recent modeling studies, our study has some strength. First, the unique characteristics of the novel coronavirus were taken into account, and a modified version of the original SIR epidemic model was established for each individual city. Considering that the new coronavirus originating from Wuhan and the majority of the local infected patients outside Wuhan had Wuhan travel history before the Spring Festival, we extended the SIR model to make it suitable for the transmission characteristics of the novel coronavirus in mainland China. The proposed model estimated the transmission rate β for the Wuhan inflows and local residents separately to describe the spreading pattern of the coronavirus in a more realistic way. Second, our proposed spatially explicit model is able to obtain fine scale prediction result. Different cities have different transmission rates due to their own conditions (for instance, population density and human mobility characteristics). If the transmission rate in national scale is used to model the epidemics of different cities (Wu et al., 2020), the prediction of epidemic trend of all major cities is similar to Wuhan (see Figure 4 in Wu et al., 2020). In this study, different transmission rates of different cities were estimated using the city-level migration data and reported confirmed cases, then the epidemic prediction results at the city scale were obtained to provide references for undertaking more balanced and efficient control measures.

Our predictions of three future scenarios, namely failure of outbreak control, successful control, and considering work/school resuming, provide information for decision makers to allocate resources for stopping the disease spread. Generally speaking, densely populated cities and cities

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

in central China will face severe pressures to control the epidemic, since the increase of infected people is inevitable in all three scenarios in the near future. By comparing predictions of the three scenarios, it is obvious that reducing the transmissibility is a critical approach to decline the net daily increase and controlling the magnitude of epidemics. Fortunately, the latest number of confirmed diagnoses (Figure 1) is lower than that of free propagation of the virus assumed in scenario 1, indicating current control measures implemented by Chinese government are effective, including controlling traffic between Wuhan and other regions, isolating suspected patients, canceling mass gatherings, and requiring people to implement protective measures. However, once the Spring Festival travel rush appears recently as scenario 3 designed (as most provinces plan to resume work on February 9), it will inevitably cause considerable growth in transmissibility and further re-increase of epidemic. In addition, current insufficient supply of protective equipment may exacerbate this situation. Therefore, public health interventions should be performed continuously to obtain the best results of epidemic control. The following measures are recommended to implement continuously in the near future, such as, postponing work/school resuming, arranging work-from-home, instructing enterprises to implement epidemic prevention measures. Essentially, all measures are for reducing population mobility and person-to-person contact, and there is no panacea for all conditions, hence interventions in different regions should be adapted according to local epidemics.

Our modeling work has several limitations. First, due to the limited prior knowledge for this sudden Wuhan 2019-nCoV outbreak, the infection rate and recovery rate in this study are regarded as the same for different age groups, which may result in errors of predication for cities with different age structures. Second, the model parameters were estimated using the reported confirmed cases that may be lower than the actual number of infections, so the parameter estimation may not represent the real situation. Third, besides the transmission between Wuhan and other cities, we do not consider other inter-city transmissions. Although the Chinese government strictly controlled the traffic between cities, the inter-city transmission may contribute to the epidemic dynamics in the future days, especially during days of work and school resuming.

Author Contributions: XZ designed the experiments. PJ, ZQ, and JY collected and processed data. AZ and SX analyzed data. All authors interpreted the results and wrote the manuscript.

Funding: This research received no external funding.

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

Reference

Anderson, R.M., Fraser, C., Ghani, A.C., Donnelly, C.A., Riley, S., Ferguson, N.M., Leung, G.M., Lam, T.H., Hedley, A.J., 2004. Epidemiology, transmission dynamics and control of

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

SARS: The 2002-2003 epidemic. Philos. Trans. R. Soc. B Biol. Sci. 359, 1091–1105. https://doi.org/10.1098/rstb.2004.1490

Chen, Y., Liu, Q., Guo, D., 2020. Coronaviruses: genome structure, replication, and pathogenesis. J. Med. Virol. https://doi.org/10.1002/jmv.25681

Chowell, G., Fenimore, P.W., Castillo-Garsow, M.A., Castillo-Chavez, C., 2003. SARS outbreaks in Ontario, Hong Kong and Singapore: The role of diagnosis and isolation as a control mechanism. J. Theor. Biol. 224, 1–8. https://doi.org/10.1016/S0022-5193(03)00228-5

Fisman, D., Rivers, C., Lofgren, E., Majumder, M.S., 2014. Estimation of MERS-Coronavirus Reproductive Number and Case Fatality Rate for the Spring 2014 Saudi Arabia Outbreak: Insights from Publicly Available Data. PLoS Curr. 1–25. https://doi.org/10.1371/currents.outbreaks.98d2f8f3382d84f390736cd5f5fe133c

Fuk-Woo Chan, J., Yuan, S., Kok, K.-H., Kai-Wang To, K., Chu, H., Yang, J., Xing, F., Liu, J., Chik-Yan Yip, C., Wing-Shan Poon, R., Tsoi, H.-W., Kam-Fai Lo, S., Chan, K.-H., Kwok-Man Poon, V., Chan, W.-M., Daniel Ip, J., Cai, J.-P., Chi-Chung Cheng, V., Chen, H., Kim-Ming Hui, C., 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. Lancet. https://doi.org/10.1016/S0140-6736(20)30154-9

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan , China. Lancet. https://doi.org/10.1016/S0140-6736(20)30183-5

Hui, D.S., I Azhar, E., Madani, T.A., Ntoumi, F., Kock, R., Dar, O., Ippolito, G., Mchugh, T.D., Memish, Z.A., Drosten, C., Zumla, A., Petersen, E., 2020. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — The latest 2019 novel coronavirus outbreak in Wuhan, China. Int. J. Infect. Dis. https://doi.org/10.1016/j.ijid.2020.01.009

Lai, D., 2005. Monitoring the SARS Epidemic in China: A Time Series Analysis. J. Data Sci. 3, 279–293. https://doi.org/10.6339/JDS.2005.03(3).229

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T.Y., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z., 2020. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N. Engl. J. Med. 0, 1–9. https://doi.org/10.1056/NEJMoa2001316

Munster, V.J., Koopmans, M., van Doremalen, N., van Riel, D., de Wit, E., 2020. A Novel Coronavirus Emerging in China — Key Questions for Impact Assessment. N. Engl. J. Med. https://doi.org/10.1056/NEJMp2000929

Nelder, J.A., Mead, R., 1965. A Simplex Method for Function Minimization. Comput. J. 7, 308–313. https://doi.org/10.1093/comjnl/7.4.308

medRxiv preprint doi: https://doi.org/10.1101/2020.02.09.20021360; this version posted February 11, 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. All rights reserved. No reuse allowed without permission.

Nishiura, Jung, Linton, Kinoshita, Yang, Hayashi, Kobayashi, Yuan, Akhmetzhanov, 2020. The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020. J. Clin. Med. https://doi.org/10.3390/jcm9020330

Perlman, S., 2020. Another Decade, Another Coronavirus. N. Engl. J. Med. https://doi.org/10.1056/NEJMe2001126

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. https://doi.org/10.1101/2020.01.23.20018549

Riou, J., Althaus, C.L., 2020. Pattern of early human-to-human transmission of Wuhan 2019-nCoV. bioRxiv. https://doi.org/10.1101/2020.01.23.917351

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., Seilmaier, M., Drosten, C., Vollmar, P., Zwirglmaier, K., Zange, S., Wölfel, R., Hoelscher, M., 2020. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N. Engl. J. Med. https://doi.org/10.1056/NEJMc2001468

Shen, M., Hu, M., Liu, S., Chang, Y., Sun, Z., 2015. Assessment of the Chinese Resident Health Literacy Scale in a population-based sample in South China. BMC Public Health 15, 1–11. https://doi.org/10.1186/s12889-015-1958-0

Shen, M., Peng, Z., Xiao, Y., Zhang, L., 2020. Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. bioRxiv. https://doi.org/10.1101/2020.01.23.916726

Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F., 2020. A novel coronavirus outbreak of global health concern. Lancet. https://doi.org/10.1016/S0140-6736(20)30185-9

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. Lancet. https://doi.org/10.1016/S0140-6736(20)30260-9

Zhao, S., Ran, J., MUSA, S.S., Yang, G., Lou, Y., Gao, D., Yang, L., He, D., 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. bioRxiv. https://doi.org/10.1101/2020.01.23.916395

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. https://doi.org/10.1056/NEJMoa2001017