EPIDEMIC TRENDS ANALYSIS AND RISK ESTIMATION OF 2019-NCOV OUTBREAK

Qinghe Liu[1], Deqiang Li[1], Zhicheng Liu[1], Zefei Gao[1], Junkai Zhu[1],Junyan Yang[2], Qiao Wang[1]*

1 School of Information Science and Engineering, Southeast University, Nanjing, 210096, China
{liuqinghe, deqiang_li, zhichengliu, zf_gao, junkai_zhu, qiaowang} @seu.edu.cn

2 School of Architecture, Southeast University, Nanjing, 210096, China
yjy-2@163.com

Corresponding Author: Q. Wang (Email: qiaowang@seu.edu.cn)

Date: Feb. 9th, 2020

ABSTRACT.

Backgroud: Since the outbreak of coronaviruses on December 8, 2019, the spread of new coronaviruses is increasing. With no specific medicine, analysis of epidemic trends is critical for disease control and epidemic risk management.

Methods: We collected information on demographic characteristics, exposure history, and illness timelines of laboratory-confirmed cases of NCIP that had been reported by Feburary 6, 2020. Considering the huge population migration during the spring festival of China, a Flow-SEIR model is proposed to perform empirical analysis. Assuming that most of people will go back to work from February 10 to the end of February, the risk of the coming migration from home to workplace is defined and estimated, based on the confirmed cases and the Spring Festival migration data. We use data from Baidu Migration to estimate the coming migration for returning to the workplace after vacation from February 10 to 28, 2020. The risk of migration for returning was then estimated.

Main findings:

(1). The time when the number of daily new cases has reached a peak level, i.e. the national epidemic inflection point, is expected to be on February 7 - February 9. After the inflection point, the number of infected people will grow slowly and the epidemic will be gradually controlled. We define the temporal number of infected cases as the net cumulative number of cases. And the peak value represents the maximum of the net cumulative number, after this peak point, the current number of cases will gradually decrease and the epidemic will gradually subside. The predicted peak value of confirmed cases in Hubei can reach 62800 (56900 - 70300, 0.95 CI). The peak time in Hubei is on February 29 (Feb. 25 - Mar. 08, 0.95 CI). The peak arrival time of areas except Hubei will be in the interval of Mar. 8-Mar. 15, 2020. It will take about 1.5-2 months from the peak to the end of the epidemic. Most provinces have the predicted peak values less than 1000. The potential epidemic situation will be stable in March 2020 – June 2020.

(2). Rather than larger outbreak, the estimated number of infective patients in Hubei (in which Wuhan is located) presents a front and steep feature, manifesting the overall trend that the epidemic tends to be consistent.

(3). The necessary consciousness of self-protection and isolation is very effective for epidemic prevention and control. It can reduce the number of patients by nearly 90%. The effect of self-protection and isolation is stronger than the policy of traffic blockage. Provincial level traffic blockage can only alleviate 21.06% - 22.38% of the peak number of patients.

(4). Early warning of epidemic situation is very important. In the experimental environment, the results show that if there is 1 day in advance of the timely urban blockage, about 3600 people will eventually be reduced in the country. If there is 1 day of delay, about 1800 people will finally be at risk.

(5). Assuming that the false negative rate is 50% and the potential transmission rate is 1-4 times higher than that of the confirmed patients, the real peak number of patients is likely to be between 214400-472500. Suppose that the transmission rate of the false negative patient is twice that of the confirmed patient, if the false negative rate decreased by 5% in one week, the peak number of patients will slow down by 11.62%. If the proportion decreased by 10% in one week, the peak number of patients will decreased by 21.91%, which shows timely detection of false negative population is more effective for epidemic control.

(6). Under the natural condition, the return of Spring Festival transportation will aggravate the

epidemic situation, especially for Guangdong, Zhejiang, Jiangsu, Hunan, Henan, Shanghai, Fujian and Beijing.

We did not perform the global propagation of 2019-nCoV because that the available initial data is not sufficient to make valuable conclusions.

Keywords： coronavirus; 2019-nCoV; Flow-SEIR Model; SEIR Model; SARS; Mathematical Model of Infectious Disease

1. INTRODUCTION

Coronavirus [1] is a single stranded RNA virus, which is spherical in shape and has petal-like spines. It usually causes mild respiratory infections and can sometimes be fatal. Since its first discovery and identification in 1965, there have been three large-scale outbreaks of highly treatable infectious diseases caused by coronavirus, namely, SARS in mainland China in 2003, MERS in Saudi Arabia in 2012 and MERS in South Korea in 2015 [2].

At the beginning of December 2019, the first patient infected by a new type of coronavirus was found in Wuhan, Hubei, which is located in the central part of China. Then similar symptoms were found in the same city. After being infected, the main symptoms of patients in the early stage include low fever and fatigue, which were similar to the symptoms caused by general influenza, so it spreads in a large area without people's awareness. On January 12, 2020, the World Health Organization officially named the new coronavirus causing pneumonia in Wuhan as "2019 new coronavirus [3] (2019-nCoV)". On January 22, 2020, researchers speculated that the new coronavirus might come from wild bats[4]. During the Spring Festival of 2018 (from February 1 to March 12, 2018), the total estimated number of passenger trips in mainland China is about 2.97 billion [5]. As of 9 am on February 7, 2020, the cumulative reports of 31 provinces (districts and municipalities) received by the National Health Commission of the People's Republic of China [6] are shown in Table 1. In the table, Hubei Province's data of confirmed cases is temporarily excluded since Hubei Province, as a severe epidemic area, has the possibility of a large backlog of confirmed cases, combined with the shortage of medical resources and the uncertainty of the diagnosis method, which makes the data in Hubei more unreliable than other provinces. Among these reports, 22112 cases of pneumonia with new coronavirus infection were found in Hubei Province[7].

Table 1. Real time epidemic situation in China by 9am February 7

(*Data source: data released by the National Health Commission of the People's Republic of China and the Health Commissions of all provinces and autonomous regions***)**

Cumulative total	Confirmed cases (except Hubei)	Death cases	Cured cases	Suspected cases
31 provinces of China	9049	637	1543	27657

The serious epidemic has brought many urgent problems. How is the effect of policy intervention and self isolation? When the epidemic can be stable? When the growth of the number of confirmed cases can stop? To answer these questions, we preform an empirical analysis of the epidemic trend, the effect of the policy of traffic blockage and self isolation, and the early warning of Spring Festival return rush.

In section 2 of this paper, the most up-to-date literature by 7 Feb 2020 related to 2019-nCoV epidemic situation will be discussed and briefly analyzed. Section 3 and 4 presents SEIR model and Markov chain model in details. Section 5 presents the experiments settings and empirical analysis. Finally, section 6 and 7 are discussion and conclusion.

2 RELATED WORKS

On January 24, 2020, J. M. Read *et al* [8] applied the transmission model and found that the reproduction number of infection is 3.8, and 72% - 75% of the transmission needed to be restricted in order to inhibit the growing infection rate; Only 5.1% of Wuhan patients are confirmed cases, while as of January 21, 11341 people should have been infected. In addition, they found that traffic restrictions only slowed the spread of the epidemic by 24.9%. Inspired by them, our group will

further analyze the impact of self isolation and traffic interruption in this paper.

On January 24, 2020, S. Zhao *et al* [9] used index function and the data from January 10 to January 21, 2020, in mainland China and continuous time interval (SI) of SARS and MERS as the SI estimation of new coronavirus, then estimated the reproduction number of infection as 3.3 – 5.7, and the law of early outbreak was exponential distribution.

On January 26, 2020, S. Jung *et al* [10] estimated the various time periods of the epidemic and found that the incubation period is 2-9 days, with 95% confidence, the average incubation period is 4-5 days, the average delay from onset to death is 13.8 days and the average time from onset to hospitalization is 3 days. They suggested that recommended isolation should be at least 9 days.

On January 27, 2020, B. Tang *et al* [11] established a transmission dynamics model based on the transmission mechanism of Wuhan new coronavirus, isolation and other strategies. They thought the reproduction number of Infection is 3.8 and the intervention measures can reduce the reproduction number of Infection and transmission risk. Based on the epidemic data from Jan. 10 to Jan. 22, they predicted the basic regeneration number of new coronavirus pneumonia transmission in Wuhan is 6.47 (95% confidence interval is 5.71 – 7.23). If the control measures before Jan. 22 were continued, the epidemic would reach the peak value around Mar. 10. As an extension, our works carries on peak prediction and arrival time estimation to simulate two kinds of ecology in Hubei Province and other areas.

On January 27, 2020, J. A. Backer *et al* [12] estimated that the average incubation period should be 5.8 days (4.6-7.9 days, 95% confidence interval) and the range of variation was 1.3-11.3 days, according to the travel records and symptoms of 34 confirmed cases found outside Wuhan. From which they considered that 14 days of isolation will ensure that no symptoms will occur. In addition, they compared the 2019-nCoV with SARS and MERS.

On January 31, 2020, J. Wu *et al* [13] used data from Dec 31, 2019, to Jan 28, 2020 thought that the basic reproduction number of infection was estimated by using Markov Chain Monte Carlo methods and presented by the the resulting posterior mean and 95% confidence interval. They estimated that the basic reproduction number of infection in Wuhan in 2019 was 2.68 (95% confidence interval is 2.47-2.86). In addition, as of January 25, the number of infections was 75815 (9-5% confidence interval is 37304-130330). The epidemic doubling time was 6.4 days (95% confidence interval is 5.8-7.1).

On February 2, 2020, Adam J Kucharski *et al* [14] from London School of Hygiene & Tropical Medicine used a stochastic transmission model, based on confirmed cases in Wuhan and exported cases originating in Wuhan and estimated that the reproduction number of infection from mid-December to mid-January 2020 fluctuated between 1.6-2.9.

Epidemiological study is often an afterwards research, such as our previous work[15]. However, there are many uncertainties and risks in the development of 2019-nCoV, so we need to provide some suggestions from the data analysis work in time.

Since the outbreak of the epidemic, many scholars have concentrated in Wuhan and other areas, and studied the impact of basic regeneration number and traffic interruption. Some also estimated the real number of people infected in Wuhan. However, the outbreak of the epidemic has a huge impact on the economy and people's livelihood. Predicting the peak date and evaluating the effectiveness of the control measures will play a decisive role in ensuring the smooth operation of the social economy. In addition, our work has broken through the geographical boundary of Wuhan, trying to focus on Hubei and other regions outside the province for larger scale analysis. At the same

time, the risk assessment and early warning of the Spring Festival return migration phenomenon which few people mentioned are carried out in different regions.

3 FLOW-SEIR MODEL

3.1. SEIR Model.

The SEIR model is introduced in 1984 by Aron, J. and I. Schwartz [16], where S, E, I and R referring to the Susceptibles, Exposed, Infective, and Recovered, respectively. Compared with the standard SIR model, the SEIR model emphasized the importance of the Exposed. In the epidemic transmission, when susceptible ones are infected, most of them will neither immediately get sick nor display symptoms, but patients in the exposed period are also infectious, and during the recovery they will possibly be immune. For the details of SEIR model, please refer to as [17][18]. Finally, the incubation period and the time-varying propagation rate are investigated in [19] and [20] respectively.

Among various mathematical models concerning epidemic, the advantages of SEIR are as follows:

(1). Neither SI nor SIR models are acceptable in our situation. In the SI model, the population is divided into Susceptible and Infective. This model assumes that the population remains unchanged, and then the exponential growth rate and the infection rate are readily available from the observation. However, due to the dynamic population flows in and out during the Chinese Spring Festival, this SI model is not applicable current situation. On the other hand, although the SIR model includes Recovered ones, the diseases with exposed period is not in consideration.

(2). For SEIR model, The Exposed is added to improve the prediction performance, which is in accordance with the clinical findings on 2019-nCoV infection cases. Therefore, the number of people who are not affected by the disease and the number of people who recovered from the disease can be estimated.

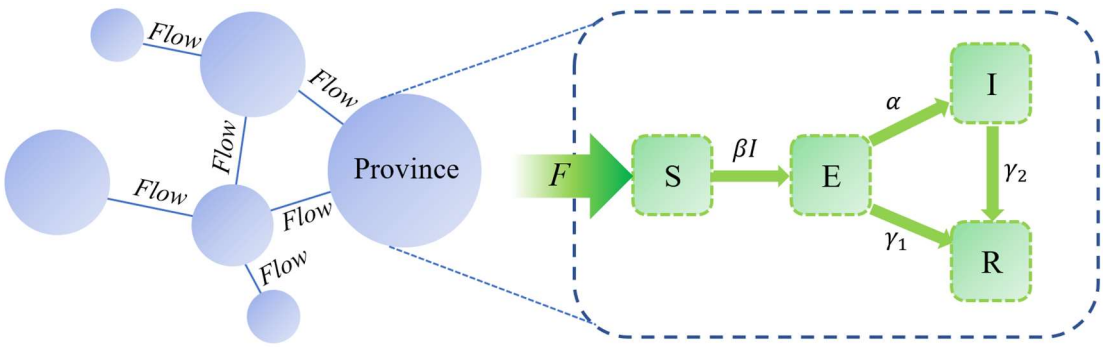


FIGURE. 1 Model Diagram

3.2. Flow-SEIR Model.

We proposed a Flow-SEIR model in this work, here is the model hypothesis：

(1). The interaction of people flow between domestic and foreign areas is not considered.

(2). The recovery probability increases linearly with the development of epidemic situation.

(3). In the course of transmission, the probability of infection of the virus in contact with each person is equal.

Different from standard SEIR model, we should further consider the special communication feature in Chinese Springer Festival, and then we add the factor of flow which reflects the mobility

of the population. This forms our Flow-SEIR model:

$$\begin{cases} \dfrac{dS}{dt} = F - \beta IS \\ \dfrac{dE}{dt} = \beta IS - (\alpha + \gamma_1)E \\ \dfrac{dI}{dt} = \alpha E - \gamma_2 I \\ \dfrac{dR}{dt} = \gamma_1 E + \gamma_2 I \end{cases} \tag{3.1}$$

Table 2. List of Symbols

S	Number of people susceptible
E	Number of people exposed
I	Number of people infective
R	Number of people recovered
β	Rate of those who are susceptible and being infected.
α	Rate of those who are exposed and being infectious.
γ_2	Those who are infectious and being recovered.
γ_1	Those who are exposed and being recovered.

Clearly, the flow F(t) rewrites the standard SEIR model by the first equation of (3.1). In practice, this term F(t) is estimated from the Baidu flow API, here we omit the details of data processing. Definitely this time-varying term must affects the behavior of dynamical system (3.1), we still omit all the mathematical details here in this work.

3.3. Practical Consideration.

It is worth noting that the transmission process of the 2019-nCoV can be divided into the following four main stages:

(1). The outbreak is potentially spreading, but was not been aware of.

(2). The first case has been reported since December 8, 2019, but the effect is seriously underrated;

(3). Offcial confirmation (including transmission control in key cities such as Wuhan);

(4). Future: The epidemic is under control and the city commutes return to normal since the patients are completely cured.

The proposed model simulates phase 2 and 3. We assume that in each stage, the recovery rates are relatively consistent at all stages, in terms of the natural recovery of self-recovered and medical treatment. But at each stage, the infection rate and the population can vary dramatically. For the first two stages, the spread of the virus satisfies the exponential distribution. For the third stage, it is found that there is a large number of population flows in and out because of the transportation during Spring Festival.

Assuming that the suspected and infected population have been isolated and do not have the ability to transmit to the outside, starting with the number of potential susceptible $\boldsymbol{S}$, then the transmission might increase by:

$$\beta = kR_0(1 - \varepsilon) \tag{3.2}$$

in which β is the rate of those who are susceptible and being infected shown in table 2, and k is a constant which will be estimated, R_0 is basic reproduction number. For the third stage, there are a

large number of population flows from Wuhan, because of the Spring Festival as well as the panic. As a result, F will increase sharply and x will also rises. After then, multiple cities began to restrict traffic, F will decrease, because of self-protection and isolation policy, and the self-segregation factor ε will increases, then because of self protection and isolation policy, the self-segregation factor ε increases, then β values is in decline gradually.

For the cure rate parameters, we collect the empirical data related to pneumonia as the prior distributiona, and take the infection of 2019-nCoV as the likelihood, and maximum the posterior as the iteration rule.

3.4. Flow and the Network of Cities.

Next, we use the network model to describe the nationwide prediction regarding each province (and municipality) as an independent node, and utilize the city-specific Flow-SEIR model correspondingly. The flow intensity of each edge of the network is *F*, represented potential impact on population flow. To perform this approach, we collected population migration data from Baidu API which provides daily intra-province population migration strength. For the i-th city or province, this floating population can be mathematically written as follows,

$$F_i = \sum_{j=1,2,\ldots,N_j}(1-\tau)\delta_{ji}S_{ji} \tag{3.3}$$

in which $(1-\tau)\delta_i S_i$ is the flow factor of the corresponding node, δ_{ji} is the natural flow factor of the corresponding node independent of control policy of traffic blockage. It represents the ratio of floating population to potential population. In the third stage, δ_{ji} of Wuhan and some cities is very large, after the intervention of policy forces, the flow factor will decrease rapidly. The decline ratio with different values can reflect the impact degree caused by different policy. We introduce traffic blockage factor τ which is between 0 and 1. The higher the degree of traffic blockage over traffic flow implemented, the greater the flow of restriction follows. Conversely, when τ is taken as 0, the traffic flow is in the uncontrolled state.

3.5. Parameter Estimation.

In practical process, equation (3.1) is implemented by following iterations with prescribed steps:

$$\begin{cases} S \leftarrow S + F - \beta IS \\ E \leftarrow E + \beta IS - (\alpha + \gamma_1)E \\ I \leftarrow I + \alpha E - \gamma_2 I \\ R \leftarrow R + \gamma_1 E + \gamma_2 I \end{cases} \tag{3.4}$$

In the process of parameter iteration, in order to get convergence parameters faster, we use Gradient Descent method to build BP network. In this model, the loss function f is defined by the least square method:

$$f = \frac{1}{M}\sum_{m=1}^{M}(z_m - \hat{z}_m)^2 \tag{3.5}$$

in which *M* means the length of the real data, z_m is the real data on the *m*th time tag, *m*=1,2,…,*M* and $\hat{z}_m$ is the estimated value.

4 RISK ESTIMATION OF MIGRATION

Now we estimate the risk of the migration after vacation. The legal work day is on Feb 10, and we estimate the average exposed period to be seven days, according to the released data by the National Health Commission of China.

Assuming that most of people who return home during Spring Festival from Jan 10th to Jan 24th will go back to their former workplace from Feb 10 to Feb 28, 2020. We can estimate the risk

of migration from Feb 10 to Feb 20 after vacation among all provinces, based on the migration data from Jan 10th to Jan 24th, which represents the migration that most people went home. We should estimate the potential cases number of each province for the past seven days, based on our proposed model, in fact, we use the total number of confirmed cases in the last seven days to estimate the potential cases number of each province during the migration time for simplicity. Then the migration risk can be expressed as follows:

$$\boldsymbol{R_{\pi i}} = \sum_{i \neq j} \boldsymbol{F_{j,i} S_i X_j} \tag{4.1}$$

where $\boldsymbol{R_{\pi i}}$, $\boldsymbol{F_{j,i}}$, $\boldsymbol{S_i}$, $\boldsymbol{X_j}$ were the migration risk index, the migration proportion from the *j*th province to the *i*th province, the migration scale of the *i*th province and the potential number of infections of *j*th province.

In practical process, equation(4.1) can be prescribed as equation (4.2).

$$R_{\pi i} = [X_1 \quad X_2 \quad \cdots \quad X_N] \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1N} \\ \vdots & & \ddots & \vdots \\ F_{N1} & & \cdots & F_{NN} \end{bmatrix} \begin{bmatrix} S_1 \\ \vdots \\ S_N \end{bmatrix} \tag{4.2}$$

in which **X** is the confirmed cases of the origin province, **F** the migration proportion from origin province to the target province, **S** the migration scale of the target province, and $\boldsymbol{R_{\pi i}}$ the coming risk index of the target province.

We get the above migration data from Baidu Migration, that **S** was the intra-province population migration strength, **F** was the intra-province migration proportion. And **X** was the confirmed cases of the origin province from Feb 2 to Feb 8.

The local risk index is based on the factors of the local basic population, the confirmed cases, the transmission and medical policy of the *i*th province.

5. MODEL IMPROVEMENT BASED ON DATA

There are some deficiencies in our research, the most important of which is that the data used for modeling is not accurate. This may lead to a deviation or error in our prediction results.

To be specific,we use the data released by the National Health Commission of China and each province, it is obtained by the government through certain rules. In practical analysis, we found the number of confirmed cases increases dramatically on some days. Following we will talk about some limitations of the data and its possible causes.

5.1. Limitations of data and possible causes

The death rate in Hubei is much higher than other provinces. Due to the large number of confirmed cases in Hubei, some problems such as the shortage of medical resources, and the lack of doctors and hospital beds began to come up. As a result, many patients cannot be treated in time and died.

Data Missing: Diagnostic reagents are in short supply in Hubei, which makes some patients unable to get timely diagnosis and treatment. As a result, the confirmed number of patients is less than the actual number. Many patients may have recovered or died in the process of waiting for diagnosis, but this part is not shown in the statistical data. We cannot obtain statistics of these patients, but they are really infected.

Data Latency: There are many reasons for data latency:

(1) The virus has an incubation period, about 7-14 [21]days and patients will have symptoms several days after being infected. As a result, the diagnosis of the patient will be delayed.

(2) In the earlier stage (before Jan 25th), due to the lack of promotion of diagnosis methods

and insufficient medical conditions of some small cities, many patients cannot be diagnosed in time.

(3) In Hubei Province, the number of infected people is too large, there are not enough diagnostic reagents, so many patients need to queue for reagents and wait for a few days.

(4) The diagnostic method is real-time PCR, which has not yet been satisfactory with some limitations of false-positive and false-negative results[22], false-negative result is mainly caused by irregularities in the procedure that can result in DNA damage. Due to testing rights transferred to the hospital, there may be many improper operation problems leading to the occurrence of false negative, so that patients need more than two tests to be diagnosed.

And in the other provinces except for Hubei, patients can almost get timely diagnosis and treatment due to the small number of patients and the popularization of diagnostic technology. However, with the increasing number of patients, other provinces with more patients should also pay attention to the problem of medical resources.

5.2. Process in practical case

For the problem of high death rate in Hubei Province, we reduced the ratio of γ_1 and γ_2 of Hubei's model according the real data that $\gamma_{1real} = q_1\gamma_1, \gamma_{2real} = q_2\gamma_2$. For the data missing problem, we enlarged the ratio of α of Hubei's model, that $\alpha_{real} = p\alpha$.

For the data latency problem, it can be assumed that the number of confirmed cases published on one day is the number of confirmed cases n_i days ago, and n_i is the number of delayed days in the ith province, expressed as

$$I_{real}(t) = I_{data}(t - \mathrm{n}_i) \tag{5.1}$$

For Hubei, due to the above factors, the latency will be relatively long, which may reach 5-7 days. For other provinces, if patients can get treatment in time, the delay should only include the time waiting for the result of the diagnostic reagents, is about 1 day.

6. MAIN RESULTS OF FLOW-SEIR MODEL

6.1. Estimate of the Peak of Confirmed Cases .

We define peak value represents the peak of I(t), which is the cumulative number of infected cases minus the cumulative number of cures, i.e. the number of net infected cases. After this peak point, the temporal number of cases will gradually decrease and the epidemic will gradually subside.

Assuming that our model takes no accout of the impact of return trip of Spring Festival temporarily, we first simulated the epidemic trend of 34 provinces of China, and predicted the peak number of confirmed cases and its arrival time in each area, as shown in figure1, 2. Parameters such as the prevalence rate and cure rate obeyed the Gaussian distribution in the model and we took the mean value.

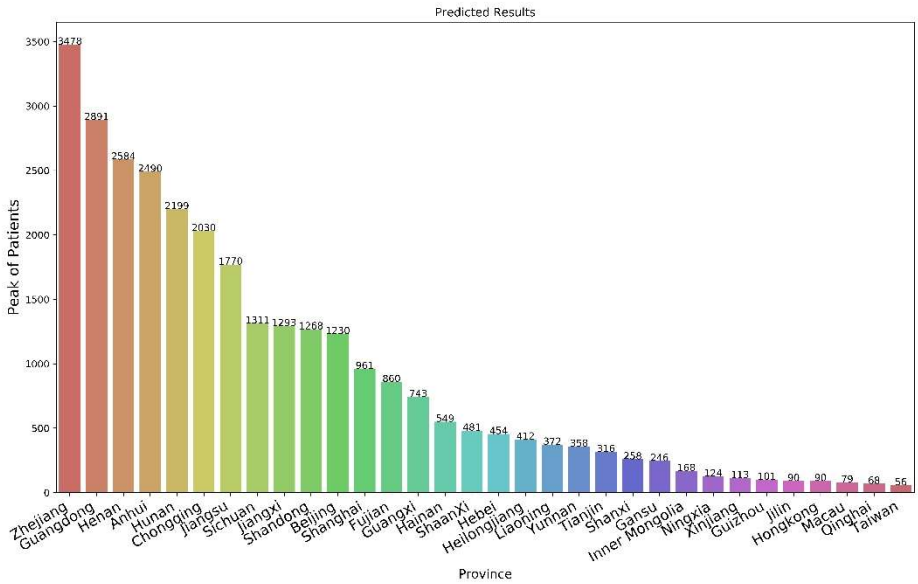


FIGURE 2. Predicted peak number of patients in each province of China, the x-axis is the logarithmcalculation of the predicted number of people based on 10, and the y-axis corresponds to different geographical areas.

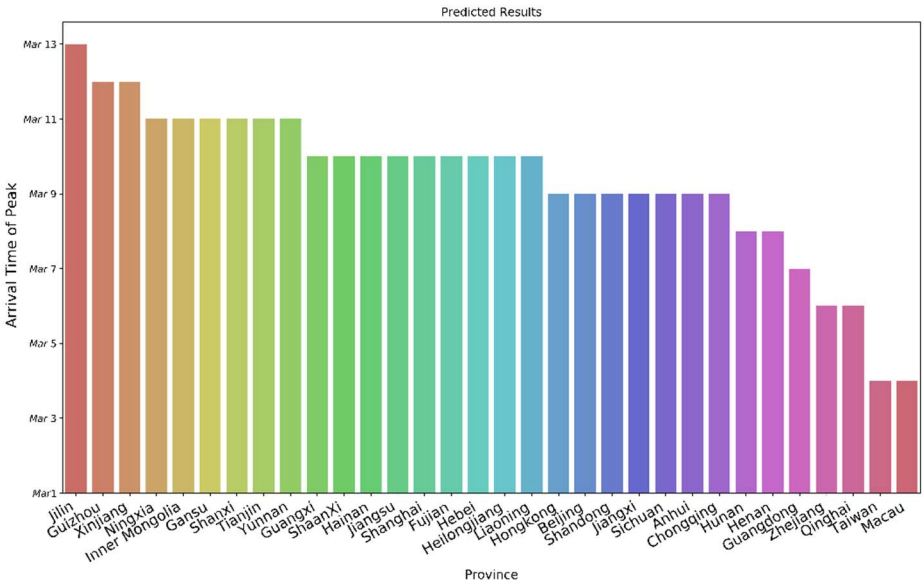


FIGURE 3. Predicted arrival time of peak number of confirmed cases in each province of China, the x-axis is the prediction of peak arrival time at average level, and the y-axis corresponds to different geographical areas.

On average, the peak number of confirmed cases of Hubei is predicted to reach 62800(only 30%-50% of the real number of cases considering the limited medical resources), and the peak arrival time is on February 29, 2020. Tibet has both the minimum peak value and the earliest peak arrival time. According to the data up to Feburary 7, 2020, the number of confirmed cases in Tibet is only 1, and there is no suspected patient. The predicted total peak number of confirmed cases of

new coronavirus in China will reach 85500 (76700 - 97500, 0.95 CI), is about 13 times of that of SARS virus, according to the statistics released by the World Health Organization is 5327. On the one hand, this shows that when the transmission of new coronavirus is very high, on the other hand, it is also the rapid development of traffic caused by the rapid development of cross transmission speed.

The predicted peak value is combined with different interval, which is listed in table 3 below. Except for Hubei, there are no other provinces that the peak value is more than 3000 but Zhejiang.

Table 3 Predicted peak number of confirmed cases of each province based on $\tau = 0.9$, $\varepsilon = 0.25$ and other parameters are on average.

Provinces (autonomous region, or special administrative region)	Peak value interval
Hubei	>40000
Zhejiang，Guangdong，Henan，Hunan，Chongqing	[2000, 4000]
Anhui，Sichuan，Jiangxi，Shandong，Jiangsu，Beijing	[1000, 2000]
Shanghai，Fujian，Guangxi，Hainan	[500, 1000]
Shaanxi，Hebei，Heilongjiang，Liaoning，Yunnan，Tianjin，Shanxi，Gansu	[200, 500]
Inner Mongolia，Ningxia，Xinjiang，Guizhou，Hong Kong，Jilin，Macau，Qinghai， Taiwan，Tibet	<200

Table 3 distinguishs Hubei Province as a single category to reflect the corresponding relationships and differences. In terms of peak population, Hubei is far higher than other regions. Zhejiang, Guangdong, Henan, Hunan and Chongqing are the places most likely to become the next worst-hit area at present, and the relevant departments should pay attention to them. The peak value of most parts of the country is predicted to be stable within 1000.

It is predicted that the peak arrival time of Hubei will be at the end of February, which is much earlier than other provinces. Therefore, it is very urgent to analyze the progress of the disease in Hubei. Then, comparative analysis and discussion on more details of Hubei are conducted.

6.2. Simulation of Confirmed Cases.

Figure 4 and 5 show the simulation trend of Hubei Province and other regions, respectively. According to the results, the potential population affected by the epidemic will be stable in March 2020. Potentially affected population base is more than 100000. As of Feburary 5, 2020, the number of confirmed cases in Hubei is 16678, only 16.7% of the population base of the potential population. As a result, there are still a large number of infected cases that have not been tracked or confirmed, so only 30%-50% confirmed cases were estimated in our model, in other words, our model underestimates the real number of confirmed cases. The epidemic situation in Hubei will not be completely controlled and eliminated until June 2020. The peak arrival time of except-Hubei is about 1-2 week later. Except that the trend of except-Hubei is basically consistent with that of Hubei province, the change of the epidemic situation of except-Hubei is higher than that in Hubei. It can be seen that the trend of the confirmed cases' number in Hubei is steeper. This is consistent with the characteristics of a diffusion center, of which the change time is earlier and drastic, while comparing with the center, the periphery presents the characteristics of postposition and slowness. It can be found that the trend is usually steeper and ahead of time in the center of the epidemic compared with other regions. On the one hand, it is related to the timely protection measures that have taken to other areas. On the other hand, it is related to the information lag of the central location. In the early stage of the epidemic, the spread in the outbreak area almost shows an unhindered exponential

growth, which is one of the reasons for the steeper trend of the central position. Similarly, according to the data, compared with other regions in Hubei Province, Wuhan also shows obvious prepositiveness and steepness.

The error between the predicted data and the real data is shown in Figure 6. Here we difine the inflection point as the peak value of new infections number per day, and after the inflection point, the number of infected people will grow slowly and the epidemic will be gradually controlled. We can get the inflection point after first-order difference calculation of the cumulative cases trend, which is expected to be on February 7 - February 9.

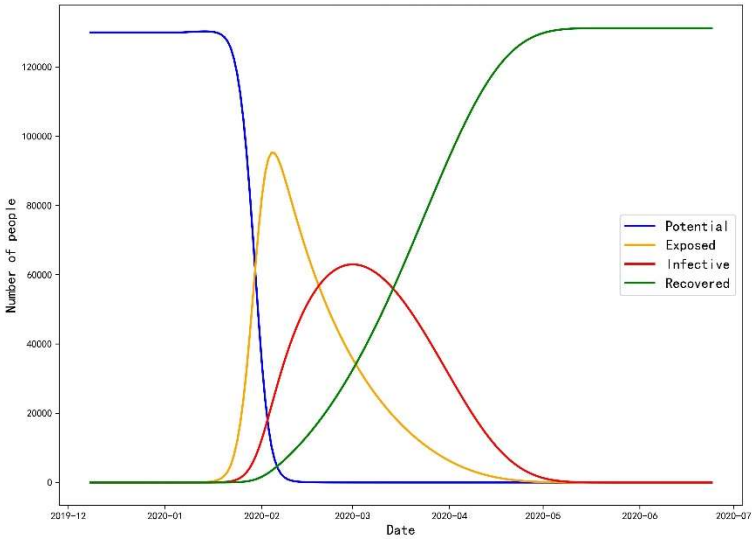


FIGURE 4. Simulated epidemic trend of Hubei based on $\tau = 0.9$ and $\varepsilon = 0.25$

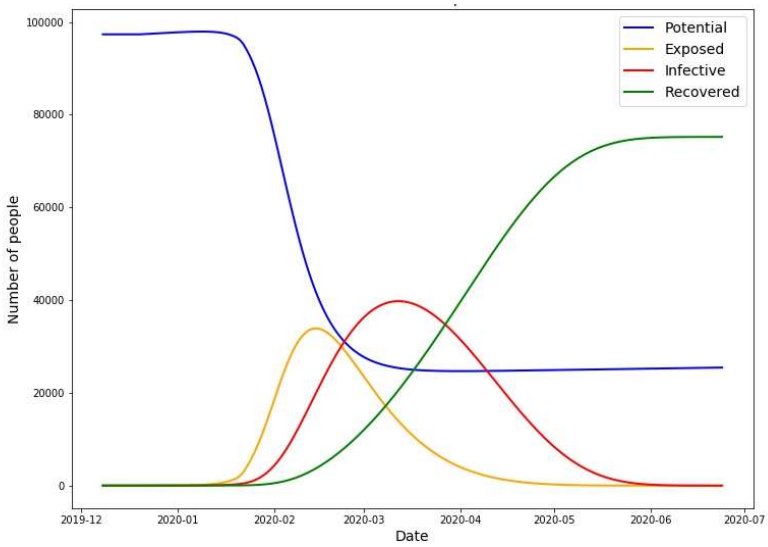
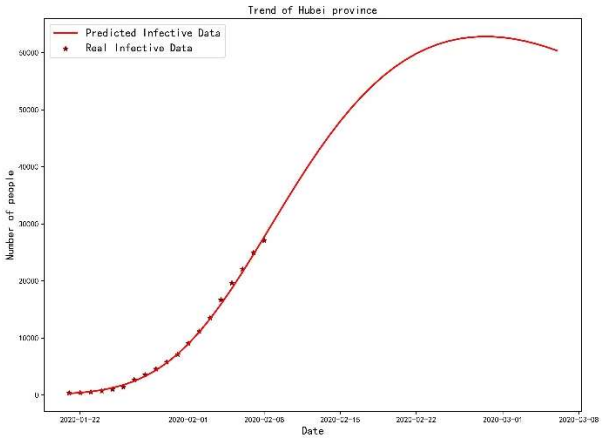
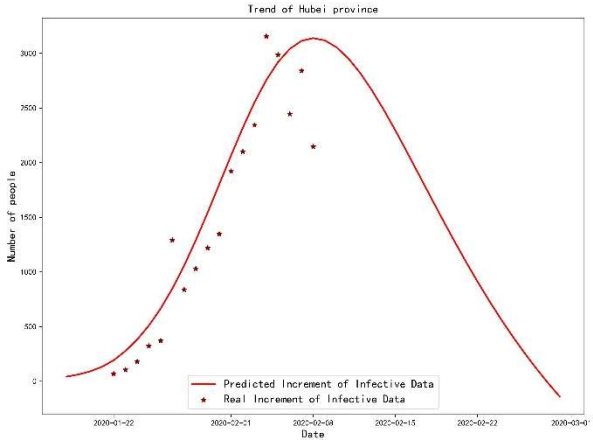


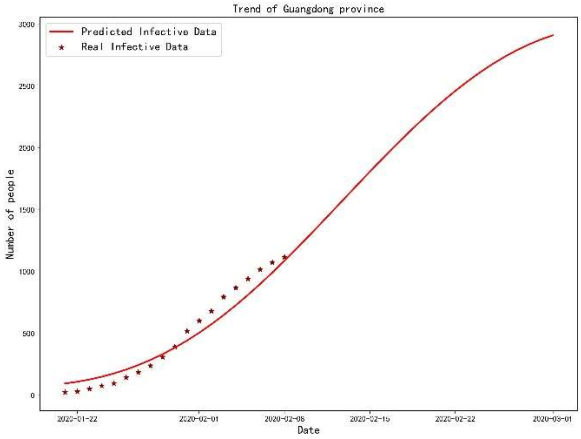
FIGURE 5. Simulated epidemic trend of China except Hubei based on $\tau = 0.9$ and $\varepsilon = 0.25$



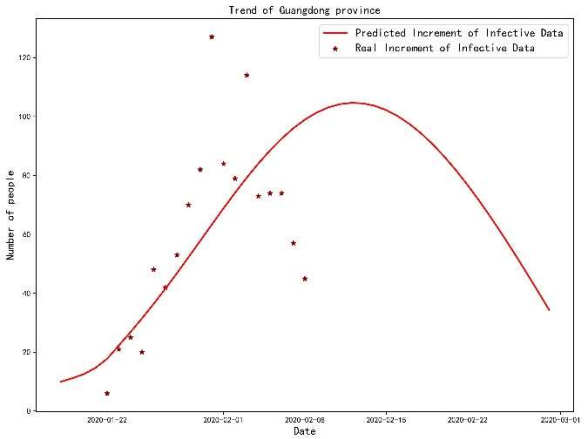
(a) Cumulants of Hubei



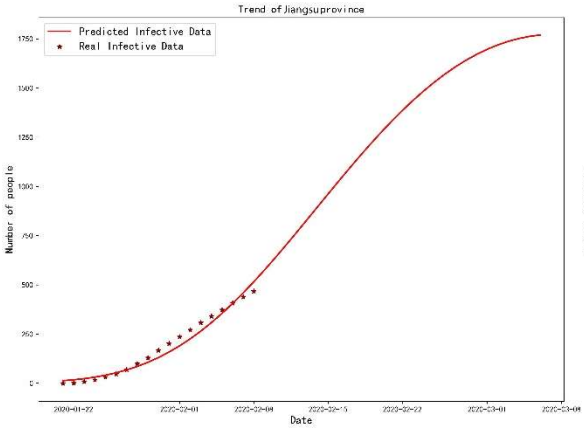
(b) Increment of Hubei



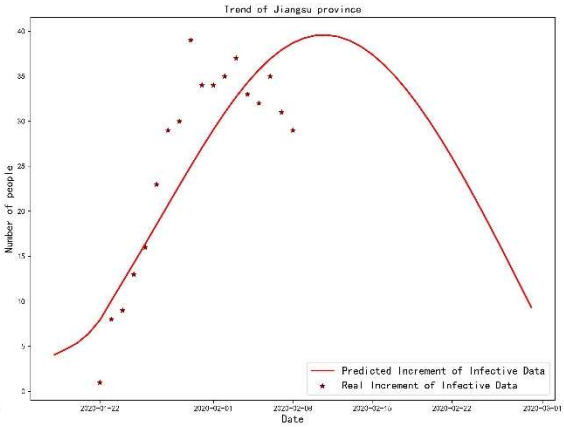
(c) Cumulants of Guangdong



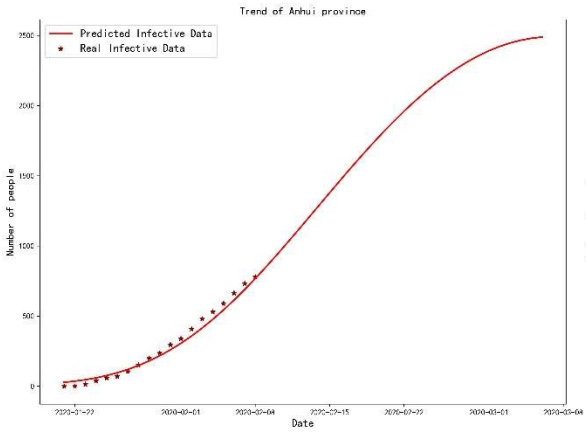
(d) Increment of Guangdong

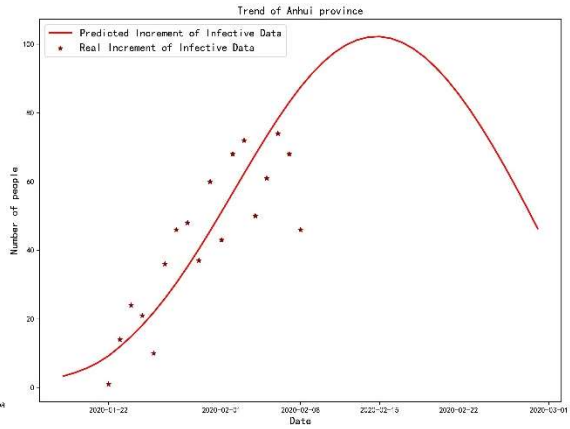


(e) Cumulants of Jiangsu



(f) Increment of Jiangsu





(g) Cumulants of Anhui **(h) Increment of Anhui**

FIGURE 6. Comparisons between the predicted infective data and the real infective data

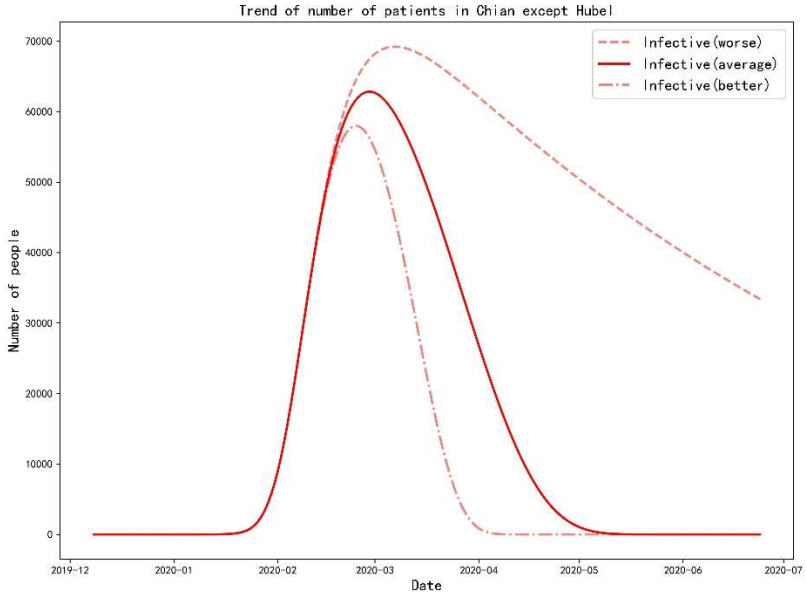


FIGURE 7. Iteration results of the number of confirmed cases in Hubei Province. Worse condition uses recovery rate of 0.05 confidence interval, better condition uses recovery rate of 0.95 confidence interval.

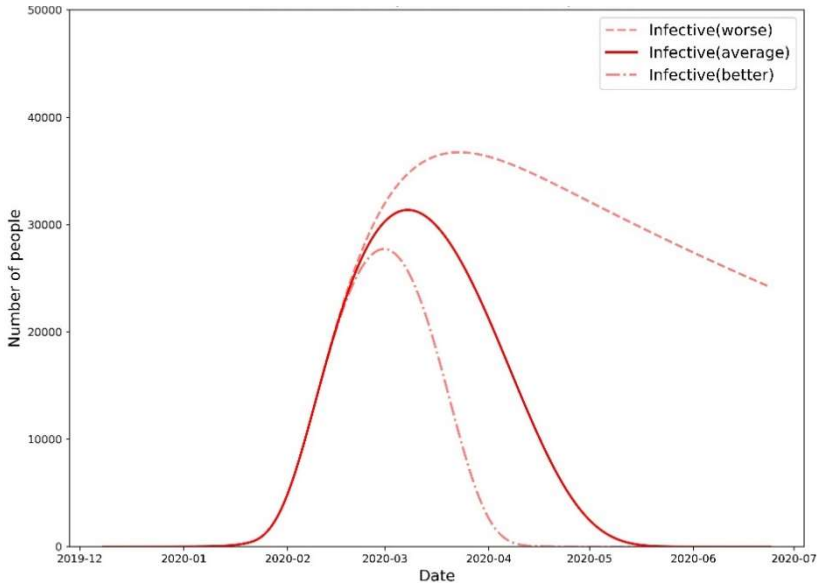


FIGURE 8. Iteration results of the number of confirmed cases in China except Hubei. Worse condition uses recovery rate of 0.05 confidence interval, better condition uses recovery rate of 0.95 confidence interval.

In Figure 7 and 8, we took the upper quantile of 0.9 as the cure rate, and took the upper quantile of 0.1 as the prevalence rate, to represent a better case for simulation; and we took the upper quantile of 0.1 as the cure rate, took the upper quantile of 0.95 as the prevalence rate, to represent a worse case for simulation, and the results are shown in Table 4 and 5.

The peak time fluctuations of Hubei and other regions are respectively about 2020.02.29 (02.25 - 03.08, 0.95 confidence interval) and 2020.03.09 (2020.03.02 – 2020.03.24, 0.95 confidence interval). On baseline condition, the daily growth of confirmed cases will reach the peak in the next week and gradually stabilize, and then begin to decline.

Table 4 Simulation results of epidemic situation in Hubei

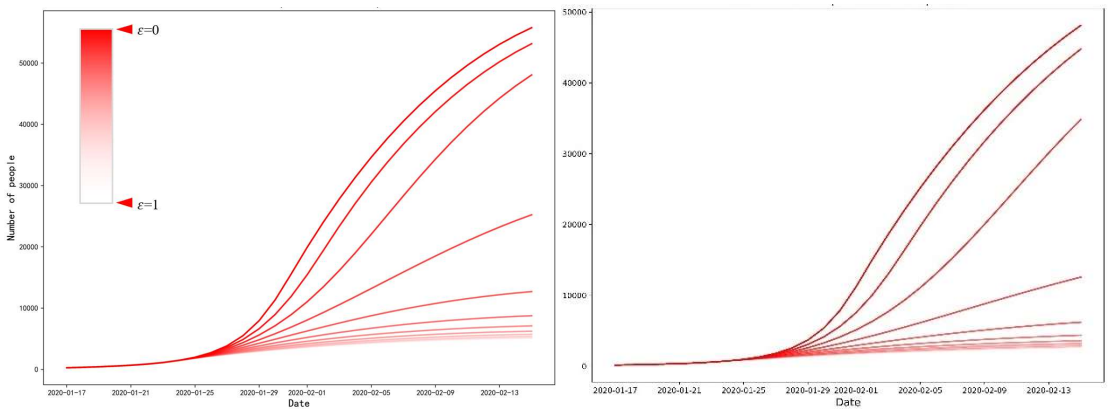
	Peak value time	Peak value
Worse case	2020. 03.11	57000
Average case	2020. 03.01	49000
Better case	2020. 02.25	43800

Table 5 Simulation results of epidemic situation in China except Hubei

	Peak value time	Peak value
Worse case	2020. 03.24	36800
Average case	2020. 03.09	31300
Better case	2020. 03.02	27700

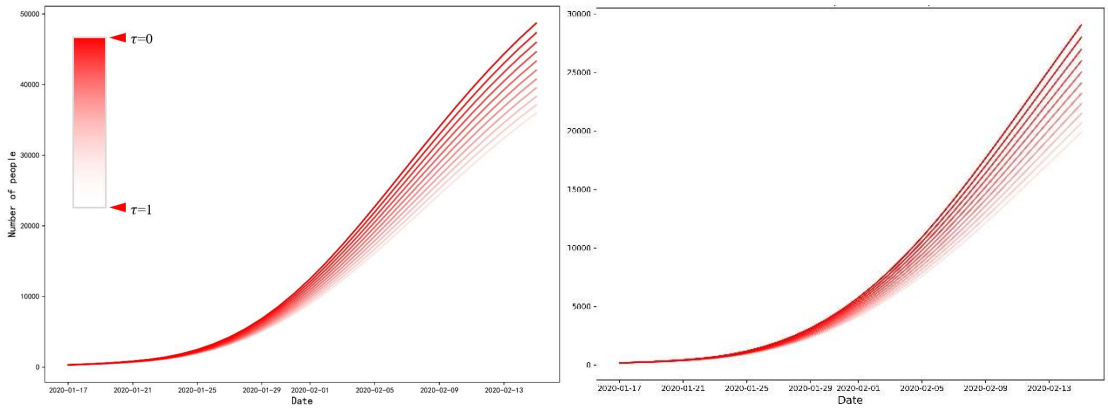
6.3. The Efficiency of Traffic Blockage and Self-Segregation

Compared with the larger area of Hubei, the epidemic situation in Wuhan presents a front and steep feature, but the overall trend tends to be consistent. Since traffic blockage factors and self-segregation factors are difficult to quantify, we take different values of them in [0,1] to obtain experimental results, as shown in Fig 9 and Fig 10. The self-segregation factors reflect the importance the public attaches to the outbreak and the extent to which measures are taken to protect and isolate themselves. With the increase of self-segregation factor, the peak of confirmed cases' number is significantly decreased. When ε=1, that is, all people were in a completely ideal state of self-protection and isolation, the peak would decrease by 89.68% on average, while ε=0, the peak will increase by 20.4%. Traffic blockage factor of τ reflects the policy intervention to limit the extent of the city traffic flow. If τ is close to 1, the target area is almost completely blocked foreign flows, the peak number will drop by 21.06%-22.39% than that with τ=0. In contrast, people's voluntary cooperation is more important than traffic blockage, so we once again encourage people to do a good job of self-protection, both to protect themselves and others.



(a) Hubei (b) Except Hubei

FIGURE 9. Confirmed cases forecast under different ε, the x-axis is date, the y-axis is number of confirmed cases.



(a) Hubei (b) Except Hubei

FIGURE 10. Confirmed cases forecast under different τ, the x-axis is date, the y-axis is number of confirmed cases.

6.4. What are the effects of earlier or later warning of 2019 SARS-CoV?

In figure 11, we tested the trend of confirmed cases with self-segregation measures at different time points. The results showed that under the fluctuation of 0.01 of self-segregation factor ε, the peak number of confirmed cases in Hubei would arrive 0-2 days ahead of time, and that in areas outside Hubei Province would arrive 1-3 days ahead of time. For the peak population in Hubei Province, 1 week of delay will lead to an increase of 8.50% - 10.95%, while 1 week of ahead will slow down by 24.26% - 25.33%, and 2 weeks of ahead will slow down by 57.10% - 57.46%. For the peak population outside Hubei, 1 week of delay will lead to an increase of 20.92% - 22.77%, while 1 week of ahead will slow down 32.78% - 35.60%, and 2 weeks of ahead will slow down 63.75% - 67.13%. For the time translations of self-segregation impact, the improvement rate of peak scale shows nonlinear change, and it is greatly improved with the advance of impact time. With the increase of time delay to more than 15 days, serious epidemic centers such as Hubei are likely to appear in various regions.

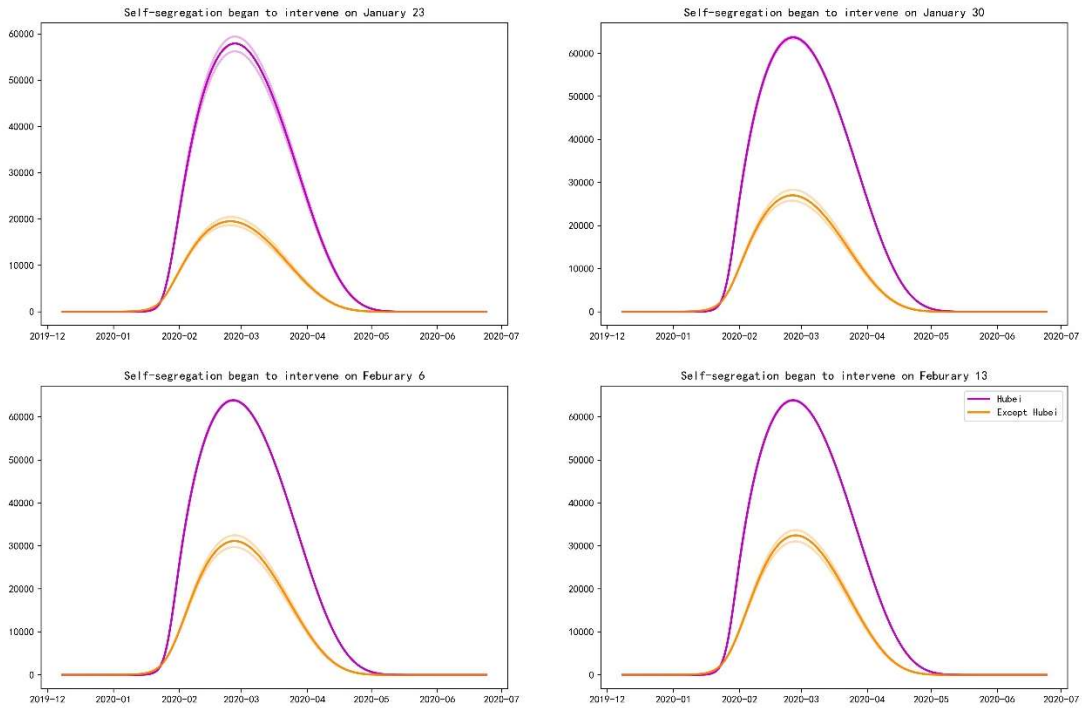


FIGURE 11. The self-segregation factor ε is set to take effect on four time: January 23, January 30, Feburary 6, Feburary 13. The x-axis is the date and the y-axis is the number of patients.

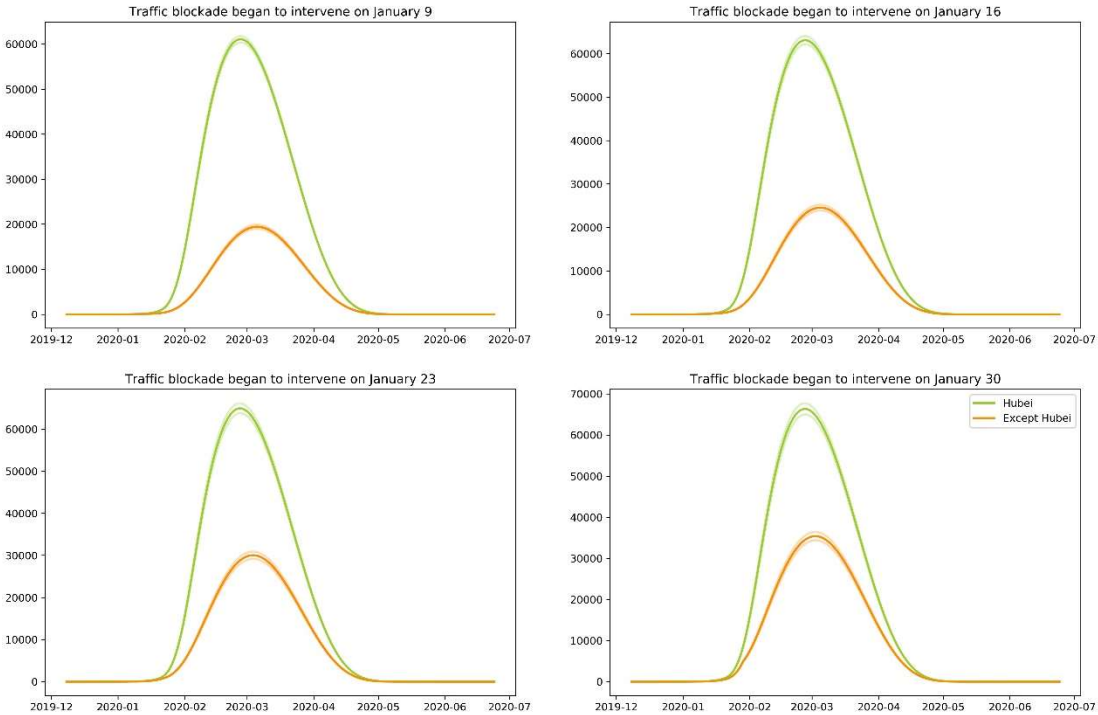


FIGURE 12. The policy control of traffic blockage factor τ is set to take effect on four time: January 9, January 16, January 23, January 30. The x-axis is the date and the y-axis is the number of patients.

The self-segregation factor ε in Figure 11 is changed to the traffic blockage control factor τ to represent the impact of traffic interruption, and four different effective times are also analyzed to get Figure 12. In the experiment, it is assumed that after traffic blockage, the traffic flow between provinces would be completely cut off and according to the study [23,24] on the number of basic

infections, it is conservatively assumed that the transmission rate of the virus in the case of natural transmission is 2.5 times of that after the traffic blockage. The results showed that under the fluctuation of 0.1 of the policy control of traffic blockage factor τ. The peak arrival time and the peak number of people in Hubei Province are earlier and smaller when the traffic is blocked. One week ahead of time, the peak number of people was only relieved by 3.10% - 3.79%, and even two weeks ahead of time, it was only relieved by 6.64% - 8.08%. This also shows that when the traffic blockage of Wuhan and other areas began on January 23, the epidemic situation in Hubei Province has been very serious, and the potential transmission level has been quite high. However, in domestic areas outside Hubei, the peak number will drop by 15.30% - 16.70% under the condition of one week ahead of schedule regulation, and will drop by 31.74% - 33.09% two weeks ahead of schedule regulation.

Considering an average daily rough estimate based on experimental conditions, every day in advance of the timely urban blockade, about 3600 people will eventually be reduced in the country, and every day of delay, about 1800 people will finally be at risk.

6.5. Risk Estimation for Migration

We can draw a conclusion from figure 13 with the method of risk estimation, that Guangdong, Zhejiang, Jiangsu, Hunan, Henan, Shanghai, Fujian and Beijing were the top-risk provinces. We used normalization method to measure the rank of risks of all provinces, instead of the origin migration risks. For origin risk, see Appendix 2. Guangdong, Zhejiang, Jiangsu, Shanghai and Beijing are the main places that people get their work, which may account for the great risks. For Hunan and Henan, the two provinces adjoin with Hubei, the most damaged province, which may get more potential number of infections, so the risks are also high. For Beijing, the capital of China, the scale of Spring Festival migration is not larger than Guangzhou, and the returning migration will not be so strong because of the traffic blockage, thus the risk is lower than Guangzhou.

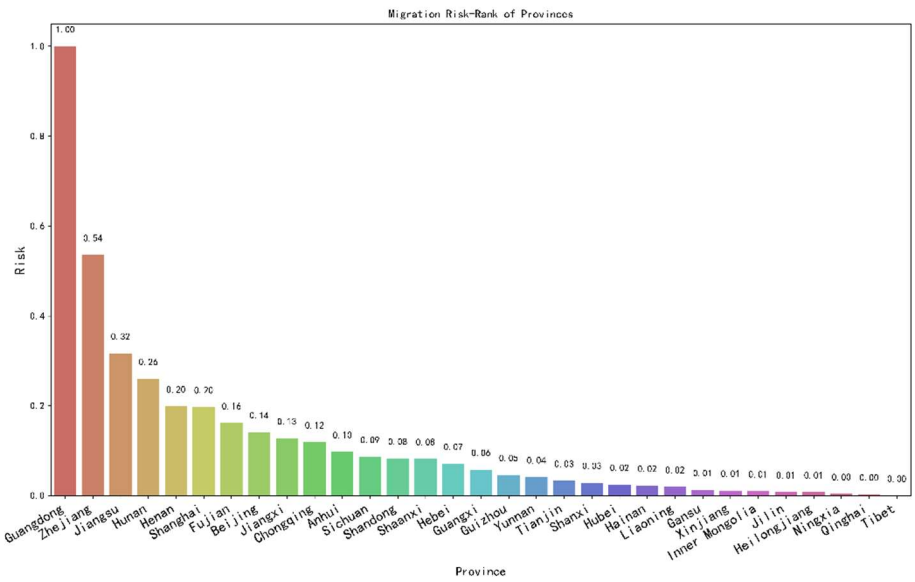


Figure 13 Risks-rank of all provinces for the returning migration

6.6. Estimation of the real number of patients in Hubei

In this part, we consider that there are a large number of false negative cases. Because these

people have not received hospital treatment, the transmission is stronger than the confirmed cases. Therefore, we assume that the potential transmission rate of false negative is 1 times of that of confirmed cases.

In figure 14, assuming *l* is 2, the peak number of patients is between 49000 and 268000. If *l* is 3, the maximum number of patients is 380000. Assuming *l* is 4, the peak value reaches 490000. If *l* reaches 5, the peak value will exceed 600000.

In figure 15, the results show that if the false negative rate decreased by 5% in one week, and the peak number of patients slowed down by 11.62%. If the proportion decreased by 10% in one week, the peak number of patients decreased by 21.91%. It can be seen that timely detection of false negative population is more effective for epidemic control.

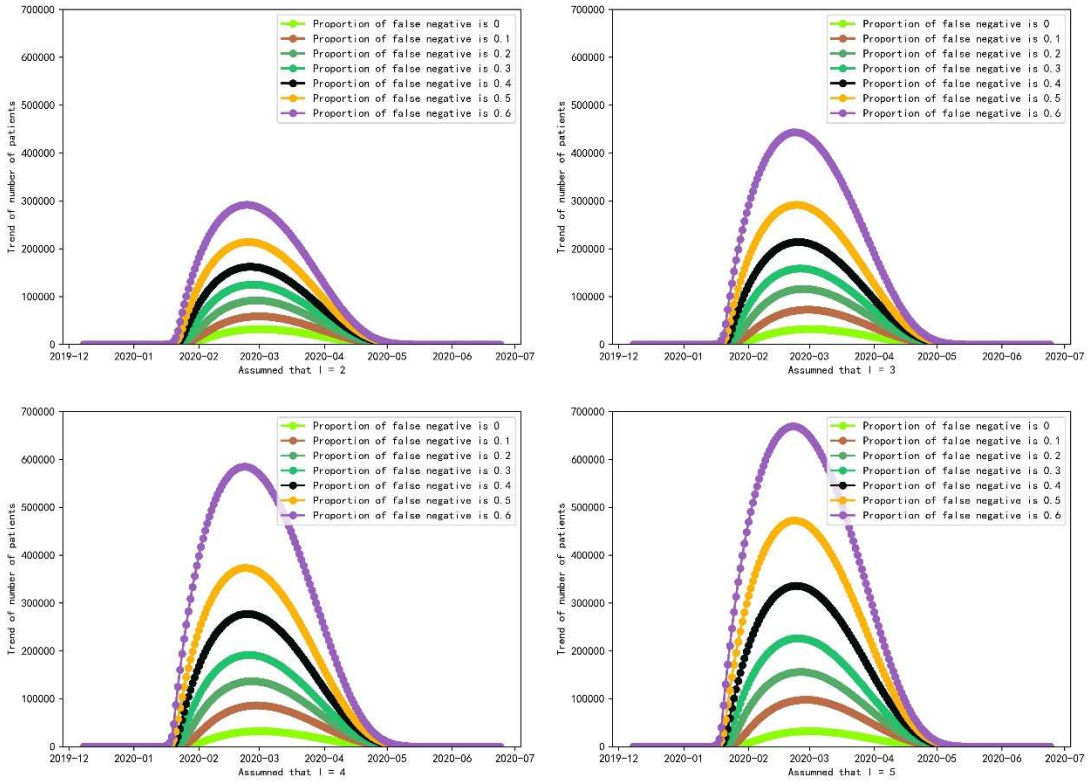


Figure 14. Estimation of the real number of patients based on different l (2, 3, 4, 5)

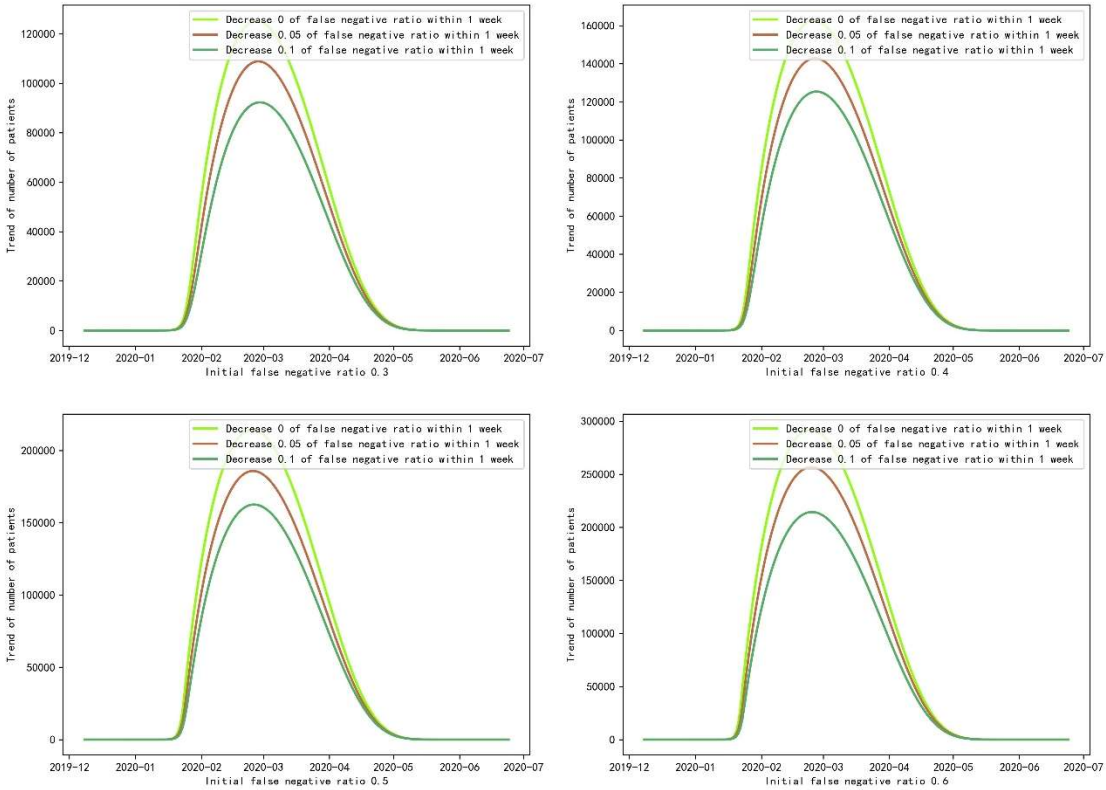


Figure 15. Estimation of the real number of patients based on different decrease on false negative ratio (0%, 5%, 10%) within a week.

7. CONCLUSION

We perform simulation analysis of Flow-SEIR model, which is proposed for epidemic trends analysis and risk estimation. From analysis, we draw the following conclusion:

(1). On average, the predicted peak value can reach 62800 (56900 - 70300, 0.95 confidence interval), and the peak time is on 2020.02.29(2020.02.25 – 2020.03.08, 0.95 confidence interval) and peak of areas except Hubei will come at 2020.03.09 (2020.03.02 – 2020.03.24, 0.95 confidence interval). Peak arrival time in other regions except Hubei will occur 1-2 weeks behind that of Hubei. The peak value of the whole country is predicted to be about 85500 (76700 - 97500, 0.95 confidence interval), which will be 13 times of the SARS virus in 2003. It will take about 1.5-2 months from the peak to the end of the epidemic.

(2). Except for Hubei and Zhejiang, there are no serious disease cases in other provinces of which the peak value is almost lower than 3000. Most provinces have the predicted peak values ranging from 0 to 1000. In recent days, the number of confirmed cases in Zhejiang has increased rapidly, and there is the possibility of exponential growth trend.

(3). The epidemic situation in the south is more serious than that in the north of China, and more serious in the East than that in the West. Tibet in the southwest of China is the least affected by the epidemic situation mainly due to the defense of sparse population in the plateau area.

(4). Rather than larger outbreak, the estimated number of infective patinets in Hubei(in which Wuhan is situated) presents a front and steep feature, manifesting the overall trend that the epidemic tends to be consistent.

(5). If the masses take protective measures, the peak number of cases will fall greatly by 89.68%. However, if the masses completely ignored self-Segregation factor, the peak is predicted to rise by 20.40%. If the external population flow in the target area is almost completely blocked, the peak of inaction will decrease by 21.06%-22.38%. In short, when the masses raise their cooperate awareness of self-protection and follow certain control policies regarding preventive measures, the outbreak will probably be slowing down.

(6). Early warning for patients, timely traffic blockage and self isolation measures are extremely effective, especially in areas with less epidemic situation. Based on experimental conditions, every day in advance of the timely urban blockage, about 3600 people will eventually be reduced in the country, and every day of delay, about 1800 people will finally be at risk.

(7). A large amount of people flow from home to workplace may incease the risk of bringing more infections to the destinations, especially for Guangdong, Zhejiang, Jiangsu, Hunan, Henan, Shanghai, Fujian and Beijing.

Reference

[1] Chen Y, Liu Q, Guo D. Coronaviruses: genome structure, replication, and pathogenesis[J]. Journal of Medical Virology, 2020:doi:10.1002/jmv.25681

[2] Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU, Khan K. Pneumonia of unknown etiology in Wuhan, China: potential for international spread via commercial air travel[J]. Journal of Travel Medicine 2020:doi:10.1093/jtm/taaa1008

[3] 'Definition of suspected cases of unexplained pneumonia', the National Health Commission of the People's Republic of China (in Chinese). [http://www.nhc.gov.cn/]

[4] Li W, Shi Z, Yu M, *et al*. Bats are natural reservoirs of SARS-like coronaviruses[J]. Science, 2005, 310(5748): 676-679.

[5] Hu M. Visualizing the largest annual human migration during the Spring Festival travel season in China[J]. Environment and Planning A: Economy and Space, 2019, 51(8): 1618-1621.

[6] National Health Commission of People's Republic of China: http://www.nhc.gov.cn/ [cited 7 Feb 2020]

[7] Health Commission of Hubei Province: http://wjw.hubei.gov.cn/ [cited 7 Feb 2020]

[8] Read JM, Bridgen JRE, Cummings DAT, *et al*. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions[J]. medRxiv, 2020.

[9] Zhao S, Ran J, MUSA SS, *et al*. 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[J]. bioRxiv, 2020:doi:10.1016/j.ijid.2020.01.050

[10] Jung S, Kinoshita R, Hayashi K, *et al*. Epidemiological identification of a novel infectious disease in real time: Analysis of the atypical pneumonia outbreak in Wuhan, China, 2019-20 [J]. medRxiv, 2020.doi:10.1101/2020.01.26.20018887

[11] Tang B, Wang X, Li Q *et al.* Estimation of the transmission risk of 2019-nCov and its implication for public health interventions. Available at SSRN 3525558, 2020

[12] Backer JA, Klinkenberg D, Wallinga J. The incubation period of 2019-nCoV infections among travellers from Wuhan, China[J]. medRxiv, 2020:doi: 10.1101/2020.01.27.20018986

[13] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study[J]. the Lancet, 2020: doi:10.1016/S0140-6736(20)30260-9

[14] Kucharski AJ, Russell TW, Diamond C *et al.* . Early dynamics of transmission and control of 2019-nCoV: a mathematical modelling study[J]. medRxiv, 2020:doi:10.1101/2020.01.3120019901

[15] Lu J, Liu W, Xia R, *et al*. Effects of closing and reopening live poultry markets on the epidemic of human infection with avian influenza A virus[J]. Journal of biomedical research, 2016, 30(2): 112.

[16] Aron JL, Schwartz IB. Seasonality and period-doubling bifurcations in an epidemic model[J].Journal of theoretical biology, 1984, 110(4): 665-679.

[17] Li MY, Muldowney JS. Global stability for the SEIR model in epidemiology[J]. Mathematical biosciences, 1995, 125(2): 155-164.

[18] Li MY, Smith HL, Wang L. Global dynamics of an SEIR epidemic model with vertical transmission[J]. SIAM Journal on Applied Mathematics, 2001, 62(1): 58-69.

[19] Kim S, Byun JH, Jung IH. An SEIR Epidemic Model where Empirical Distribution of Incubation Period has Approximated by Coxian Distribution[J]. arXiv preprint

arXiv:1906.05556, 2019.
[20] Smirnova A, deCamp L, Chowell G. Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the SEIR model[J]. Bulletin of mathematical biology, 2019, 81(11): 4343-4365.
[21] Pediatric Committee. Emergency response plan for the neonatal intensive care unit during epidemic of 2019 novel coronavirus [J]. Chin J Contemp Pediatr, 22(2): 91-95.
[22] Lin C , Ye R , Xia YL. A meta-analysis to evaluate the effectiveness of real-time PCR for diagnosing novel coronavirus infections[J]. Genetics and molecular research: GMR, 2015, 14(4):15634-15641.
[23] Shao P, Shan Y. Beware of asymptomatic transmission: Study on 2019-nCoV prevention and control measures based on extended SEIR model[J]. bioRxiv, 2020:doi: 10.1101/2020.01.28.923169
[24] Lodge EK, Schatz AM, Drake JM. Protective population behavior change in outbreaks of emerging infectious disease[J]. bioRxiv, 2020:doi:10.1101/2020.01.27.921536

Appendix1 Data Source

Reported Cases Data:

http://www.nhc.gov.cn/
http://wjw.beijing.gov.cn/
http://wjw.fujian.gov.cn/
http://wsjk.tj.gov.cn/
http://wsjkw.hebei.gov.cn/
http://www.xjhfpc.gov.cn/
http://wsjkw.nx.gov.cn/
http://ynswsjkw.yn.gov.cn/wjwWebsite/web/index
http://www.gzhfpc.gov.cn/
http://wsjkw.sc.gov.cn/
http://wsjkw.cq.gov.cn/
http://wst.hainan.gov.cn/swjw/index.html
http://wsjkw.gxzf.gov.cn/
http://wsjkw.gd.gov.cn/
http://wjw.hunan.gov.cn/
http://wjw.hubei.gov.cn/
http://www.hnwsjsw.gov.cn/
http://wsjkw.shandong.gov.cn/
http://hc.jiangxi.gov.cn/
http://wjw.ah.gov.cn/
http://www.zjwjw.gov.cn/
http://wjw.jiangsu.gov.cn/
http://wsjkw.sh.gov.cn/
http://wsjkw.jl.gov.cn/
http://wjw.nmg.gov.cn/
http://wjw.beijing.gov.cn/
http://sxwjw.shaanxi.gov.cn/
http://wjw.shanxi.gov.cn/
http://wsjk.ln.gov.cn/
http://wsjk.gansu.gov.cn/
http://wsjkw.hlj.gov.cn/
https://wsjkw.qinghai.gov.cn/
http://wjw.xizang.gov.cn/

Provincial-Level Migration Data:

http://qianxi.baidu.com/

Appendix2 The migration risk of each province during the coming migration for returning to the workplace after vacation before February 10th, 2020

Province	Migration Risk
Guangdong	1.0000
Zhejiang	0.5369
Jiangsu	0.3184
Hunan	0.2619
Henan	0.2007
Shanghai	0.1990
Fujian	0.1634
Beijing	0.1435
Jiangxi	0.1297
Chongqing	0.1210
Anhui	0.0996
Sichuan	0.0887
Shandong	0.0848
Shaanxi	0.0838
Hebei	0.0731
Guangxi	0.0600
Guizhou	0.0483
Yunnan	0.0436
Tianjin	0.0356
Shanxi	0.0311
Hubei	0.0268
Hainan	0.0245
Liaoning	0.0217
Gansu	0.0151
Xinjiang	0.0137
Inner Mongolia	0.0125
Jilin	0.0104
Heilongjiang	0.0103
Ningxia	0.0061
Qinghai	0.0051
Tibet	0.0022