

# Forecasting and Analysis of Time Variation of Parameters of COVID-19 Infection in China Using An Improved SEIR Model

**Guoping Zhang**

Public meteorological service center of China Meteorological Administration

**Huaji Pang**

Qingdao Meteorological Bureau

**Yifei Xue** (✉ [ifei.hsueh@outlook.com](mailto:ifei.hsueh@outlook.com))

Jiangxi Provincial Geomatics Center

**Yu Zhou**

Jiangxi Meteorological Service Center

**Ruliang Wang**

Jiangxi Meteorological Service Center

---

## Research Article

**Keywords:** novel coronavirus, COVID-19, compartmental model, risk assessment, epidemiology, mathematical inference, emerging infectious diseases

**Posted Date:** March 6th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-16159/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

RESEARCH

# Forecasting and Analysis of Time Variation of Parameters of COVID-19 Infection in China Using An Improved SEIR Model

Guoping Zhang<sup>1</sup>, Huaji Pang<sup>2</sup>  
, Yifei Xue<sup>3,4\*</sup>  
, Yu Zhou<sup>5</sup>  
and Ruliang Wang<sup>5</sup>

\*Correspondence:

ifei.hsueh@outlook.com

<sup>1</sup>Public meteorological service center of China Meteorological Administration, 100081, Beijing, China

Full list of author information is available at the end of the article

## Abstract

**Background:** Due to the emergency pandemic threat, the COVID-19 has attracted widespread attention around the world. Common symptoms of infection were fever, cough, and myalgia fatigue. On January 31, 2020, the World Health Organization (WHO) declares this outbreak a Public Health Emergency of International Concern (PHEIC).

**Methods:** In order to study the spread of the novel coronavirus pneumonia, this paper proposed an improved SEIR model to simulate the spread of the virus, which includes the effect factor of government intervention. The model parameters are determined based on the daily reported statistical data (up to February 8) of confirmed, suspected, cured, and death. According to utilize the spread rate, the probability of infection of the suspected, the probability of the suspected becoming a confirmed one, the cure rate, the mortality rate, and the quarantine ratio, we performed simulations and parameter calibrations at three region levels, i.e., China, Hubei and non-Hubei respectively. In addition, considering that the government initiated effective prevention and control measures after the outbreak, this paper dynamically estimates all the parameters of the proposed model.

**Results:** The simulation reveals that the parameters of non-Hubei region are not significantly different from Hubei's. Hubei Province has a high transmission rate, low cure rate, high probability of infection, low effective quarantine rate. since January 31, with the continuous strengthening of epidemic prevention and control measures, all parameters of the model have changed significantly. The parameters of Hubei and non-Hubei regions have the same trend. The trend of all parameters is now moving in a direction that is conducive to reducing the number of confirmed, suspected and fatal cases.

**Conclusions:** The number of infections of the virus initially showed a rapid increase in the trend, and the number of infectious case showed a clear downward shape. With the government to take a variety of prevention and control measures and the efforts of the general medical staff, the number of infection curve on February 22 appeared in the top of the arc pattern, indicating that the inflection point began to appear, but the decline in the number of infections slowly.

**Keywords:** novel coronavirus; COVID-19; compartmental model; risk assessment; epidemiology; mathematical inference; emerging infectious diseases

## Introduction

On December 31, 2019, the Wuhan Municipal Health Commission (WMHC) announced that there were several pneumonia cases related to a seafood market in Wuhan, China, and later found that a certain coronavirus may cause this pneumonia. As the statement declared by the World Health Organization (WHO) on January 9, 2020, Chinese authorities preliminary determined a novel coronavirus, by ruling out common respiratory pathogens (SARS-CoV, MERS-CoV, avian influenza, etc.), at the beginning of 2020 [1]. WHO formally designated the name of the novel coronavirus as COVID-19 on February 11th, 2020. The virus responsible for COVID-19 are deemed as highly contagious infectious comparing with SARS, which caused cluster pneumonia cases emergence during the year 2002 and 2003 [2]. According to the Situation Report – 28 released by the WHO, there are 71429 confirmed globally, among which there are 70635 confirmed in China [3].

Previous emerging infectious diseases convinced people that they may result in harmful regional or global disaster [4, 5]. Due to the emergency pandemic threat, the COVID-19 has attracted widespread attention around the world. Common symptoms of infected were fever, cough, myalgia fatigue, etc [6]. Li et al. (2020) investigated 425 confirmed COVID-19 and found that the mean incubation period was 5.2 days, during which the virus have contagious diffusion potential [7].

To trace the transmission process of the pandemic and predict it, researchers tried many ways to model it. It is crucial to take proper prevention and control measures in advance, according to the predicted result of the spatiotemporal outbreak. The SARS epidemic in 2003 leads to a wave of research on the prediction of infectious diseases, the dynamic models represented by the SIR, which is a compartmental model, have been improved. Many scholars tried to forecast the spread of virus based on dynamic models [8, 9, 10, 11, 12, 13]. These models included expanded dynamic model [14], logistic model [15], cellular automata model [16], autoregressive model [17, 18], fly point propagation along the traffic line model [12], among others.

A large number of SARS simulations have also been carried out by scholars [19, 20, 21, 22]. In the 2014 Ebola diffusion simulation in western Africa, scholars began to systematically add intervention factors for human prevention and control measures into research models [23, 24, 25, 26].

The asymptomatic transmission of the COVID-19 makes it difficult for the simulating of virus transmission [27, 28]. Many researchers and research institutes are actively engaging regression models [29, 30] and modified SEIR models [31, 32] or constructing new ones to predict the development of this new coronavirus [33]. Wu et al. (2020), for example, predicts that the turning point of the epidemic in Wuhan will likely come in May using data of the number of people infected from December 31 2019 to January 28, 2020, and concludes that measures of quarantine the city of Wuhan have no significant effect on quick alleviation of the epidemic [8]. Chen et al. (2020) shows that earlier or delayed quarantine of the whole city of Wuhan is unlikely to have significant impact on the arrival of the turning point of the epidemic, but is likely to cause significant change in the number of confirmed cases, which suggests it is necessary to implement the measures of sealing off city as soon as possible [32]. Without the quarantine policy, it is likely to result in the export of more infected people nationwide and even the world, leading to the virus

transmission being more rapid and extensive. Roosa *et al.* (2020) combine the logistic growth model, the Richards model, and the sub-epidemic wave model to give a better forecast [30].

The prediction model of contagious diseases based on artificial intelligence technology with population, transportation, medical, and other big data predict the spatiotemporal spread of diseases. The population OD super network model, for example, can estimate the population flow and therefore predict the spread of infectious diseases based on big data on population migration [34, 35]. It is conducive to guide population flow properly, so as to provide necessary decision-making support for the macro-control of the epidemic and the spatial management of the epidemic. Meanwhile, the model shows that there is a positive correlation between the population input and the number of confirmed cases for Wuhan, which verifies that population flow is one of the main reasons for the early spread of the epidemic [36].

The deep learning method is used to predict the infectivity of COVID-19 [33]. By analyzing the cases tracked by the network search data and using AI technology to analyze the relationship between the network query frequency and ratios of patients, it can be temporally more real-time and spatially more detailed, taking Google Flu Trends (GFT) as an example [37]. Strzelecki (2020) use Google Trends on the COVID-19. The BLUEDOT agency in Canada combines AI technology, big data analysis, and infectious disease expert experiences, the final prediction results are then manually evaluated and revised by experts, and the prediction results can be accurate [38].

The number of samples limits traditional statistical models, and their parameters have no direct physical significance. In contrast, the dynamic models require few parameters, and their parameters have strong physical significance [39, 7, 40]. The parameters involved are directly related to the actual prevention measures of contagious diseases, so it plays a vital role in the actual prediction of infectious diseases [41, 42]. Besides, most of the social and economic statistics can be obtained publicly, which can support the simulation of the dynamic models. In this paper, we use a dynamic model and current data to predict the development of the COVID-19 epidemic.

Of various current predictions, the results predicted by the regression model, the dynamic model, and the AI model are inconsistent with the actual trend in the later period. Except for Hong Kong's forecast that the turning point of the epidemic is likely to be in May, most other forecasts are more optimistic. On the one hand, the prediction model for the maximum number of people infected, referring to Wu *et al.* (2020), was higher [8]. The predicted number of people infected in Wuhan on January 25 is 75815 is inconsistent with the actual data. The main reason might be that the parameters involved in the model are mostly static. However, the Spring Festival in China is around late January to early February 2020, and the spatiotemporal change rate of population migration is the highest in the world. Following Hubei, other provinces gradually took increasingly strict control measures, which means the parameters of the model should be changed over time [42]. This paper makes a dynamic estimation of all parameters of the model according to the daily statistical data on confirmed, suspected, and, cured cases and death, calibrating

daily based on the availability of dynamic data, and analyzes the time changes of model parameters at regional levels: Hubei and non-Hubei provinces.

## Methodology

### Data and analysis

The data used in this paper are all from the official data published by the National Health Commission of People's Republic of China (NHC). Since January 24, real-time records are made available for the whole country, Hubei Province, the more epicene of the epidemic and non-Hubei Province. Non-Hubei data are the total number of cases of all provincial regions in China except Hubei. The data collected are mainly the number of confirmed, suspected, cured case, and deaths per day (Figure 1 and Figure 2).

### Model and parameters

There are four models for novel coronavirus transmission: SI, SIR, SEIR, and SEIS models. The new coronavirus is characterized by strong transmissibility, long incubation period, and in certain severe cases, death. Therefore it is suitable to simulate the virus using the SEIR model. We employed daily statistical data of confirmed, suspected, cured, and death to calibrate parameters in the model. We assume that the birth rate, death rate, and population movement are stable in a short period. Hence we divide the whole population in the research region into six compartments: the number of confirmed patients ( $I$ ); the number of suspected patients ( $E$ ); all population who may contact with suspected patients in the future ( $S$ ); the number of population who will not have contact with suspected or confirmed patients in the future ( $V$ ); the number of cured patients ( $R$ ); the number of death ( $D$ ). The differential equations of these five variables at a certain point of unit time are as follows:

$$\frac{dS}{dt} = -\alpha I(1-p)S - \beta E(1-p)S \quad (1)$$

$$\frac{dI}{dt} = \gamma E - \theta I - \varphi I \quad (2)$$

$$\frac{dE}{dt} = \alpha I(1-p)S - \gamma E + \beta E(1-p)S - \theta E p \quad (3)$$

$$\frac{dR}{dt} = \theta I \quad (4)$$

$$\frac{dD}{dt} = \varphi I \quad (5)$$

$$S + V + I + E + R + D = N \quad (6)$$

where  $\alpha$  is the transmission rate (i.e., the number of people contacted by an infected person each day);  $\beta$  is the infection probability of the suspected patients;  $\gamma$  is the probability of a suspected patient becoming a confirmed case, which is positively related to the inverse of the average incubation period;  $\theta$  is the cure rate of the confirmed patients at a unit time, which is positively related to the multiplicative inverse of the average admission time;  $\varphi$  is the death rate of the confirmed patients;  $p$  is the percentage of quarantine of the confirmed and the suspected patients. Since no vaccine has been developed, there is no vaccination rate in formula 1. The quarantine ratio is introduced, and it is assumed that isolated diagnosed and suspected patients no longer have the ability of transmission, while patients who are not isolated and suspected patients have the ability of transmission. It is further assumed that with treatment, suspected cases have the same cure rate as confirmed patients. Many studies, taking SARS as an example, divide the diffusion of the virus into the early outbreak stage and late prevention and control stage. The research on novel coronavirus focuses on the early stage. The focal point of this paper is the simulation of virus spread since January 24 with the assumption that all parameters are adjustable.

#### Model parameters fitting methods

At the time of  $t = i$  ( $i=1,2,3,\dots, n$ ), the number of suspected cases  $E_i$ , the number of confirmed cases  $I_i$ , the number of cured cases  $R_i$ , the number of deaths  $D_i$  and the death rate are known, while  $\alpha, \beta, \gamma, \theta$ , and  $p$  need to be estimated. Although  $S$  is unknown, we can set a huge number based on the current actual number of confirmed and suspected patients and assume that it will remain unchanged for a certain period. During the process of COVID-19 outbreak, the parameters of the SEIR model remain relatively unchanged. The governments have taken many control measures after the outbreak, so the relevant parameters of the model will have to be changed. As a result, we divide the epidemic into two stages: an early stage and the current stage, and assume that at each stage, the parameters are relatively consistent. The clinical diagnosis of novel coronavirus at the early stage is still insufficient, which leads to a large degree of uncertainty in statistics of both the numbers of confirmed and suspected cases. Therefore, only statistical data since January 31, 2020 is used as the current simulation data. After the differential equations of formula 1-6 are combined, the distribution integral method is used to find the solution at time  $i$ . The day of January 31, which is recorded as the time of  $t = 0$ , marks the beginning of the simulation, and the number of confirmed, suspected, and cured cases, and deaths published on January 31 are recorded as  $I_0, E_0, R_0, D_0$  respectively.  $\varphi$  is known, and  $S$  is set as 300000. Based on the values listed in the literature on coronavirus, the minimum and maximum values intervals are set, and a smaller step is set as well. As shown in Table 1, in total, there are 115200 types of combinations of five parameters. We can take any combination and solve the differential equation together with  $I_0, E_0, R_0, D_0, \varphi$  and  $S$ . That is to say, at  $t = 1$  we can get  $i_1, e_1, r_1, d_1$ , and  $i_n, e_n, r_n, d_n$  at  $t = n$ , and hence the actual statistical values of  $I_1, E_1, R_1, D_1$ , and  $I_n, E_n, R_n, D_n$  at  $t = n$  are all known. The following objective functions are used for optimal solution:

$$o = \sum t = 1^n (|i_t - I_t| + |e_t - E_t| + |r_t - R_t| + |d_t - D_t|) \quad (7)$$

where  $o$  is the target value,  $i_t$ ,  $e_t$ ,  $r_t$ ,  $d_t$  are the estimated values of the number of confirmed, suspected, cured and death cases respectively;  $I_t$ ,  $E_t$ ,  $R_t$ ,  $D_t$  are the actual statistical values at time  $t$  respectively. Among the 115200 possible combinations listed in Table 1,  $o$  values can be obtained by formula 7. The values of  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\theta$ , and  $p$ , which correspond to the minimum of  $o$ , are the estimated results. The COVID-19 can be simulated and predicted by the estimated parameters in formula 1-6, and it also can be used for simulation and comparative analysis between different regions. Considering that the parameters of the model are changing, and the starting time of the model is adjustable, we can obtain the parameters of the model at different time intervals, then contrast and analyze the results at different time and space. After the parameters of the model are optimized, the correlation coefficient between the estimated and the statistical values can be calculated as the evaluation index of the simulation.

## Results and discussion

### Differences of simulation among different regions

Based on the formal public statistical data, we built models at three regions scale (China, Hubei, and non-Hubei) separately. After parameter simulation and prediction using the proposed model, the results are shown in Figure 3. As the numbers of confirmed, suspected, and cured cases and deaths in Hubei Province are much larger than that in non-Hubei Provinces, the simulation results in China are similar to those in Hubei Province as a whole. The number of suspected cases in the non-Hubei region has been in a relatively stable state since January 31, with a downward trend since February 3. Taking formula 7 as the objective function, the parameters of different models at different spatial scales of China, Hubei, and non-Hubei are estimated optimally. The estimation results are shown in Figure 4. Expect the value of  $\alpha$ , the values of  $\beta$ ,  $\gamma$ ,  $\theta$ , and  $p$  at the national level are all between Hubei and non-Hubei provinces. This explains the relatively low number of confirmed and suspected cases in non-Hubei areas. The results also show that efforts of prevention and control measures taken in non-Hubei regions are better than that in Hubei.

### Temporal variation and regional differences of parameters

For the prevention and control of the novel coronavirus pneumonia, the response methods learned previously from SARS are not readily applicable. The government and the public are practicing continuously new prevention and control measures, which necessitates changes in parameters related to the SEIR model. By increasing the simulation time of the SEIR model, the model parameters at different times can be obtained. Figure 5 show the temporal values of  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\theta$  and  $p$ . Both  $\alpha$  and  $\beta$  show a downward trend, while  $\theta$  and  $p$  show an upward trend. This indicates that the epidemic prevention and control measures have been strengthened over time. The quarantine and treatment of confirmed cases, the admission and monitoring of suspected cases, the quarantine and observation of close contacts, and the implementation of the quarantine operation itself have been strengthened, which makes the five parameters change significantly. In the current trend of parameter changes, the number of confirmed, suspected, and cured cases and deaths predicted by the SEIR model tend to decrease over time; on the other hand, the accuracy of

long-term prediction will also reduce dramatically. From the perspective of regional comparison, the values of infection probability  $\beta$  of suspected cases in Hubei and non-Hubei regions are similar, and their trends are similar as well. This indicates that this parameter has no significant regional differences between Hubei and non-Hubei regions, and the probability only related to the characteristics of the virus itself, which is consistent with the physical significance of the parameter. In addition to  $\beta$ , the other four parameters all show that there are significant temporal differences between non-Hubei regions and Hubei. For  $\gamma$ , the value of Hubei Province shows a downward trend over time, while the value of non-Hubei Provinces tends to ascent. As a result, the two values are increasingly converging.

#### Influence of sample size difference on parameter estimation

When using formula 7 to estimate model parameters, for the number of  $n$  samples, 40%, 50%, 60% and 70% of samples are respectively extracted for training, while the other 60%, 50%, 40%, and 30% samples are left for checking. Figure 6(a) shows the effect of the ratio of different training samples on the parameters  $\beta$ ,  $\gamma$ ,  $\theta$ , and  $p$  of SEIR. It indicates that the model parameters tend to remain relatively stable with the increase of training samples. On the contrary,  $\beta$ ,  $\gamma$ , and  $p$  are underestimated, while  $\theta$  is overestimated with the decrease of training samples. In general, based on this method, even if the training sample is reduced by 60%, good parameter estimation can still be achieved. For no training samples, the correlation coefficient between the actual and the estimated values are calculated, and the results are listed in Figure 6(b). As a whole, when the training samples vary from 40% to 70%, the estimated values after training of the model have a very high correlation with the actual values. When the sizes of training samples increases, the correlation coefficient is increasing between the estimated values of  $E$  and  $R$  and their actual values, while the correlation coefficient between  $I$  and  $D$  is decreasing.

#### Influence of predicted $S_0$ value on model prediction results and parameters

In the simulation,  $S_0$  is set at 300,000. In order to analyze the influences of changes of  $S_0$  value on the parameters and prediction results of SEIR, let  $S_0$  change between 300,000-20,000,000 to analyze the variations of prediction results. Figure 7 shows the distribution of prediction results on the 20th day under different  $S_0$  scenarios nationwide. It can be seen from Figure 7 that except Hubei, the prediction results are not affected by the size of  $S_0$ . For the prediction results of Hubei Province, the relationship between the prediction of death number and  $S_0$  is not apparent. When  $S_0$  is more significant than 1,000,000, the number of cured cases has little relation with  $S_0$ . The most influential factor of  $S_0$  is the number of confirmed and suspected cases. When  $S_0$  is smaller than 2,000,000, the predicted number of confirmed and suspected cases approaches the upper limit exponentially. However, when  $S_0$  is greater than 2,000,000, the predicted number of confirmed and suspected cases are not related to  $S_0$ .

#### Model prediction based on dynamic parameters

The fitting of parameters makes all parameters very close to their actual values, and on this basis, the development trend of the number of confirmed, suspected, and

cured cases and deaths can be predicted. With February 7 as the start date, and  $S_0$  set at 2000000, statistical data from February 7 to 11 are used to estimate the parameters of the nationwide model, and the results are shown in Figure 8. Figure 8 illustrates that the number of confirmed cases will peak at around February 22. Since the treatment of the novel coronavirus pneumonia is long, the number of confirmed cases will decrease gradually after February 22. For the simulation parameters, the proportion of quarantine has risen to 0.75, and  $\theta$  has risen to 0.014, while  $\alpha$  has dropped to 0.003, and  $\beta$  has dropped to 0.2. These parameters indicate that the number of suspected cases drops rapidly, prompting the peak number of confirmed cases to come quickly, as strict epidemic prevention and control measures gradually take their effects.

## Conclusions

We utilized the optimized SEIR model to predict and analyze the spread of COVID-19 outbreak in Hubei and other provinces of China. The model parameters are the number of daily contacts per infected person  $\alpha$ , infection probability of suspected patients  $\beta$ , probability of suspected patients becoming confirmed patients  $\gamma$ , the cure rate of confirmed patients at unit time  $\theta$ , quarantine rates of confirmed and suspected patients  $p$ . The input of this model includes the latest statistical data of confirmed, suspected, and cured cases and deaths. The model is then trained to realize the dynamic estimation of parameters. The results show that the correlation coefficient between simulation results and actual values is larger than 0.98, indicating that the proposed optimized SEIR model is suitable for simulating and forecasting the outbreak of novel coronavirus. For the Hubei province and non-Hubei area, the probabilities of suspected cases turning into the confirmed cases are roughly the same. The other four parameters show significant differences, indicating that the epidemic situation in Hubei province is dire. The comparison of the model parameters in different periods shows that the parameters have been continuously changing over time. The COVID-19 exhibits great uncertainty in prediction, diagnosis, cure, and death rates. For China as a whole, the results of early prediction are higher than those of the later rolling prediction. The parameters all turn towards reducing the number of confirmed and suspected cases and death. The parameters from February 7 show the epidemic prevention and control measures nationwide have been quite strict, which makes the number of suspected cases drop rapidly. This will accelerate the peak time for the number of confirmed cases to arrive quickly, and the turning point of the number of confirmed cases is likely to be reached at around February 22.

### Ethical Approval and Consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of supporting data

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This research was funded by the National Natural Science Foundation of China (Grant No. 41871020).

### Author's contributions

Conceptualization, G.Zhang and H.Pang; methodology, G.Zhang, H.Pang and Y.Xue; validation, G.Zhang and H.Pang; formal analysis, G.Zhang and H.Pang; investigation, G.Zhang, H.Pang, Y.Xue, Y.Zhou and R.Wang; resources, Y.Xue and Y.Zhou; data curation, G.Zhang, H.Pang, Y.Xue, Y.Zhou and R.Wang; writing—original draft preparation, G.Zhang; writing—review and editing, Y.Xue; visualization, G.Zhang and H.Pang; project administration, G.Zhang; funding acquisition, G.Zhang. All authors have read and agreed to the published version of the manuscript.

### Acknowledgements

We would like to show our gratitude to the National Health Commission of the People's Republic of China (NHC) for sharing all the daily data about the COVID-19.

### Author details

<sup>1</sup>Public meteorological service center of China Meteorological Administration, 100081, Beijing, China. <sup>2</sup>Qingdao Meteorological Bureau, 266003 Qingdao, China. <sup>3</sup>Jiangxi Provincial Geomatics Center, 330046 Nanchang, China. <sup>4</sup>Jiangxi Province Engineering Research Center of Surveying, Mapping and Geoinformation, 330046 Nanchang, China. <sup>5</sup>Jiangxi Meteorological Service Center, 330046 Nanchang, China.

### References

1. WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China. <http://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china>
2. Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B.S., McLellan, J.S.: Cryo-em structure of the 2019-ncov spike in the prefusion conformation. *bioRxiv* (2020). doi:10.1101/2020.02.11.944462. <https://www.biorxiv.org/content/early/2020/02/15/2020.02.11.944462.full.pdf>
3. WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China. <http://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china>
4. D.Smith, R.: Responding to global infectious disease outbreaks: Lessons from sars on the role of risk perception, communication and management. *Social Science & Medicine* **63**(12), 3113–3123 (2006). doi:10.1016/j.socscimed.2006.08.004
5. Naffakh, N., van der Werf, S.: April 2009: an outbreak of swine-origin influenza a(h1n1) virus with evidence for human-to-human transmission. *Microbes and Infection* **11**(8), 725–728 (2009). doi:10.1016/j.micinf.2009.05.002
6. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B.: Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet* (2020)
7. 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.: Early transmission dynamics in wuhan, china, of novel coronavirus infected pneumonia. *New England Journal of Medicine* (2020). doi:10.1056/NEJMoa2001316. <https://doi.org/10.1056/NEJMoa2001316>
8. Wu, P., Hao, X., Lau, E.H.Y., Wong, J.Y., Leung, K.S.M., Wu, J.T., Cowling, B.J., Leung, G.M.: Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in wuhan, china, as at 22 january 2020. *Eurosurveillance* **25**(3), 2000044 (2020)
9. De-rong, Z., Xiao-hong, C., Ji-an, L.H.: Study on mathematical model for sars prediction and control -a modification based on the model for general infectious diseases. *China Safety Science Journal* (2004)
10. Xiaona, H., Chengyi, L., Liqun, F., Yuchun, C.: Application of infectious disease model in sars prevention. *Chin J Epidemiol* **26**(3), 169–171 (2005)
11. Zhang, J., Lou, J., Ma, Z., Wu, J.: A compartmental model for the analysis of sars transmission patterns and outbreak control measures in china. *Applied Mathematics and Computation* **162**(2), 909–924 (2005)
12. Yang, H., Xiao-Wen, L.I., Shi, H., Zhao, K.G., Han, L.J.: Fly dots spreading model of sars along transportation. *Journal of Remote Sensing* (2003)
13. Tan, X., Liu, Q., He, J., Luo, H.: A study on the transmission dynamic model of sars in guangdong province. *Chinese Journal of Disease Control and Prevention* (2006)
14. Shi, Y.: Stochastic dynamic model of sars spreading. *Chinese Science Bulletin* **48**(13), 1287–1292 (2003)
15. De-sheng, H., peng, G., Bao-sen, Z.: Fitness of morbidity and discussion of epidemic characteristics of sars based on logistic models. *China Public Health* (2003)
16. Mingfeng, H., Chengrui, D.: A cellular automata model for sars spread. *MATHEMATICS IN PRACTICE AND THEORY* **38**(03), 41–46 (2008)
17. Shize, G.: Autoregressive model of transmission of sars. *Journal of Mathematical Medicine* **18**(2), 97–100 (2005)
18. Cui, H.J., Zhong-Lai, L.I., Yang, H., Xiao-Wen, L.I.: Nonlinear regression in sars forecasting. *Journal of Remote Sensing* (2003)
19. Dye, C., Gay, N.: Modeling the sars epidemic. *Science* **300**(5627), 1884–1885 (2003)
20. Small, M., Tse, C.K.: Clustering model for transmission of the sars virus: application to epidemic control and risk assessment. *Physica A-statistical Mechanics and Its Applications* **351**(2), 499–511 (2005)

21. Naheed, A., Singh, M., Lucy, D.: Numerical study of sars epidemic model with the inclusion of diffusion in the system. *Applied Mathematics and Computation* **229**, 480–498 (2014)
22. Okada, M., Okuno, Y., Hashimoto, S., Kita, Y., Kanamaru, N., Nishida, Y., Tsunai, Y., Inoue, R., Nakatani, H., Fukamizu, R., Namie, Y., Yamada, J., Takao, K., Asai, R., Asaki, R., Kase, T., Takemoto, Y., Yoshida, S., Peiris, J.S.M., Chen, P.-J., Yamamoto, N., Nomura, T., Ishida, I., Morikawa, S., Tashiro, M., Sakatani, M.: Development of vaccines and passive immunotherapy against sars corona virus using scid-pbl/hu mouse models. *Vaccine* **25**(16), 3038–3040 (2007)
23. Oestereich, L., Ludtke, A., Wurr, S., Rieger, T., Munoz-Fontela, C., Gunther, S.: Successful treatment of advanced ebola virus infection with t-705 (favipiravir) in a small animal model. *Antiviral Research* **105**, 17–21 (2014)
24. Reid, S.P., Shurtleff, A.C., Costantino, J.A., Tritsch, S.R., Retterer, C., Spurgers, K.B., Bavari, S.: Hspa5 is an essential host factor for ebola virus infection. *Antiviral Research* **109**(1), 171–174 (2014)
25. Boujakjian, H.: Modeling the spread of ebola with seir and optimal control. *SIAM Undergraduate Research Online* **9** (2016)
26. Lawrence, P., Danet, N., Reynard, O., Volchkova, V., Volchkov, V.: Human transmission of ebola virus. *Current Opinion in Virology* **22**, 51–58 (2017)
27. 2019-nCoV Report. <https://www.imperial.ac.uk/search/?q=2019-nCov>
28. Cohen, J.: Scientists are racing to model the next moves of a coronavirus that's still hard to predict. *Science Health* (2020)
29. Liang, Y., Xu, D., Fu, S., Gao, K., Huan, J., Xu, L., Li, J.-d.: A simple prediction model for the development trend of 2019-ncov epidemics based on medical observations. *arXiv: Populations and Evolution* (2020)
30. Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J.M., Yan, P., Chowell, G.: Real-time forecasts of the covid-19 epidemic in china from february 5th to february 24th, 2020. *Infectious Disease Modelling* **5**, 256–263 (2020). doi:10.1016/j.idm.2020.02.002
31. Ming, W.-K., Huang, J., Zhang, C.J.P.: Breaking down of healthcare system: Mathematical modelling for controlling the novel coronavirus (2019-ncov) outbreak in wuhan, china. *bioRxiv* (2020). doi:10.1101/2020.01.27.922443. <https://www.biorxiv.org/content/early/2020/01/30/2020.01.27.922443.full.pdf>
32. Chen, T., Rui, J., Wang, Q., Zhao, Z., Cui, J.-A., Yin, L.: A mathematical model for simulating the transmission of wuhan novel coronavirus. *bioRxiv* (2020). doi:10.1101/2020.01.19.911669. <https://www.biorxiv.org/content/early/2020/01/19/2020.01.19.911669.full.pdf>
33. Guo, Q., Li, M., Wang, C., Wang, P., Fang, Z., tan, J., Wu, S., Xiao, Y., Zhu, H.: Host and infectivity prediction of wuhan 2019 novel coronavirus using deep learning algorithm. *bioRxiv* (2020). doi:10.1101/2020.01.21.914044. <https://www.biorxiv.org/content/early/2020/02/02/2020.01.21.914044.full.pdf>
34. Ya-pin, W., Ying-xia, P., Jin-song, M., Gang, C., Jie-chen, W.: Mechanism analysis of inter-provincial migration flows in china based on spatial od models. *Journal of Northwest Normal University* (2015)
35. Findlater, A., Bogoch, I.I.: Human mobility and the global spread of infectious diseases: A focus on air travel. *Trends in Parasitology* **34**(9), 772–783 (2018)
36. Bogoch, I.I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M.U.G., Khan, K.: Pneumonia of unknown etiology in wuhan, china: Potential for international spread via commercial air travel. *Journal of Travel Medicine* (2020)
37. Dugas, A.S.F., Jalalpour, M., Gel, Y., Levin, S., Torcaso, F., Igusa, T., Rothman, R.: Influenza forecasting with google flu trends. *PLOS ONE* **8**(2), 1–7 (2013)
38. Strzelecki, A.: Infodemiological study using google trends on coronavirus epidemic in wuhan, china. *arXiv preprint arXiv:2001.11021* (2020)
39. Read, J.M., Bridgen, J.R., Cummings, D.A., Ho, A., Jewell, C.P.: Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* (2020). doi:10.1101/2020.01.23.20018549. <https://www.medrxiv.org/content/early/2020/01/28/2020.01.23.20018549.full.pdf>
40. Riou, J., Althaus, C.L.: Pattern of early human-to-human transmission of wuhan 2019-ncov. *bioRxiv* (2020)
41. Kucharski, A.J., Althaus, C.L.: The role of superspreading in middle east respiratory syndrome coronavirus (mers-cov) transmission. *Eurosurveillance* **20**(25), 21167 (2015)
42. Liu, T., Hu, J., Kang, M., Lin, L., Zhong, H., Xiao, J., He, G., Song, T., Huang, Q., Rong, Z., Deng, A., Zeng, W., Tan, X., Zeng, S., Zhu, Z., Li, J., Wan, D., Lu, J., Deng, H., He, J., Ma, W.: Transmission dynamics of 2019 novel coronavirus (2019-ncov). *bioRxiv* (2020)

#### Figures

**Figure 1 Confirmed and suspected.** the number of confirmed and suspected cases in Hubei, non-Hubei provinces, and China level.

**Figure 2 Cured and death.** The number of cured and death cases in Hubei, non-Hubei, and China level.

#### Tables

**Figure 3 Fitting results.** These figures illustrate fitting results of China, Hubei and non-Hubei provinces (simulation on January 31, 2020). (a) simulation and prediction results of China nationwide; (b) simulation and prediction results of Hubei province; (c) simulation and prediction results of non-Hubei provinces.

**Figure 4 Parameters distribution.** The estimation results of parameters distribution of China, Hubei and non-Hubei regions (simulated on January 31, 2020).

**Figure 5 Parameter temporal changes.** The temporal change of the parameters in proposed model.

**Figure 6 Influence of sample size analysis.** Influences of training sample size on model parameters and estimated value.

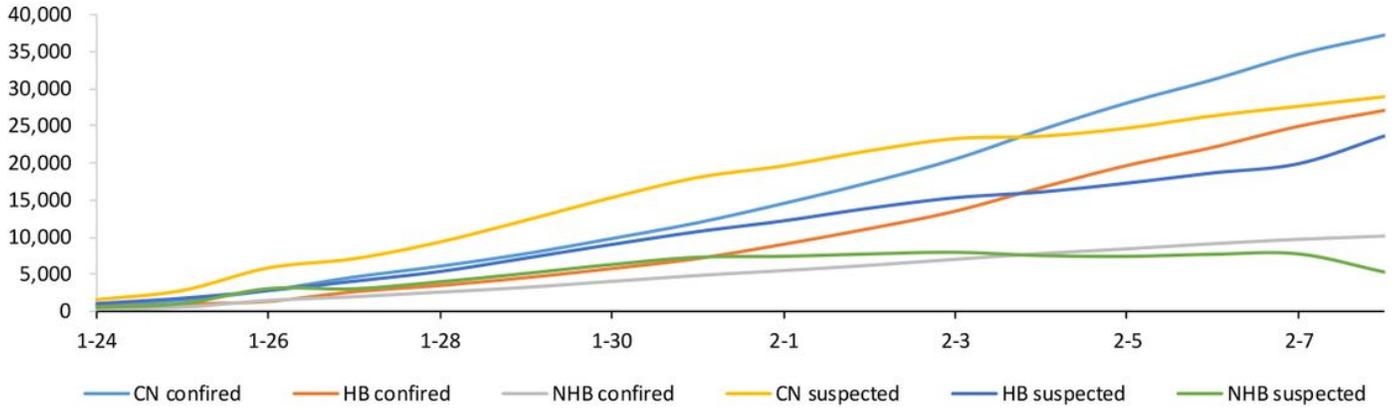
**Figure 7 Influence of  $S_0$  analysis.** Prediction results of different  $S_0$

**Figure 8 Dynamic parameter.** Nationwide forecast results based on dynamic parameter adjustment (forecast from February 7, 2020, peak number of infected cases is likely to reach at around February 22).

**Table 1** Model parameters and value range

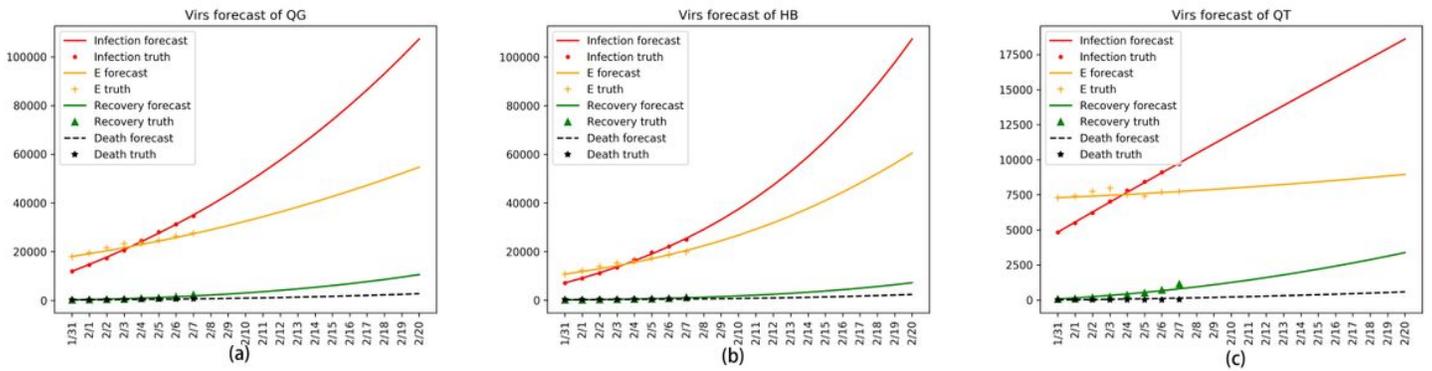
Parameters	Physical meanings	Min	Max	Step	Number
$\alpha$	number of daily contacts per infected person	0.003	0.023	0.002	10
$\beta$	infection probability of suspected patients	0.100	0.700	0.020	30
$\gamma$	probability of suspected patients becoming confirmed patients	0.150	0.170	0.001	20
$\theta$	cure rate of confirmed patients in unit time	0.006	0.030	0.002	12
$p$	quarantine rate of confirmed and suspected patients	0.100	0.900	0.050	16

# Figures



**Figure 1**

Confirmed and suspected: the number of confirmed and suspected cases in Hubei, non-Hubei provinces, and China level.



**Figure 2**

Cured and death: The number of cured and death cases in Hubei, non-Hubei, and China level.

Virs forecast of QG

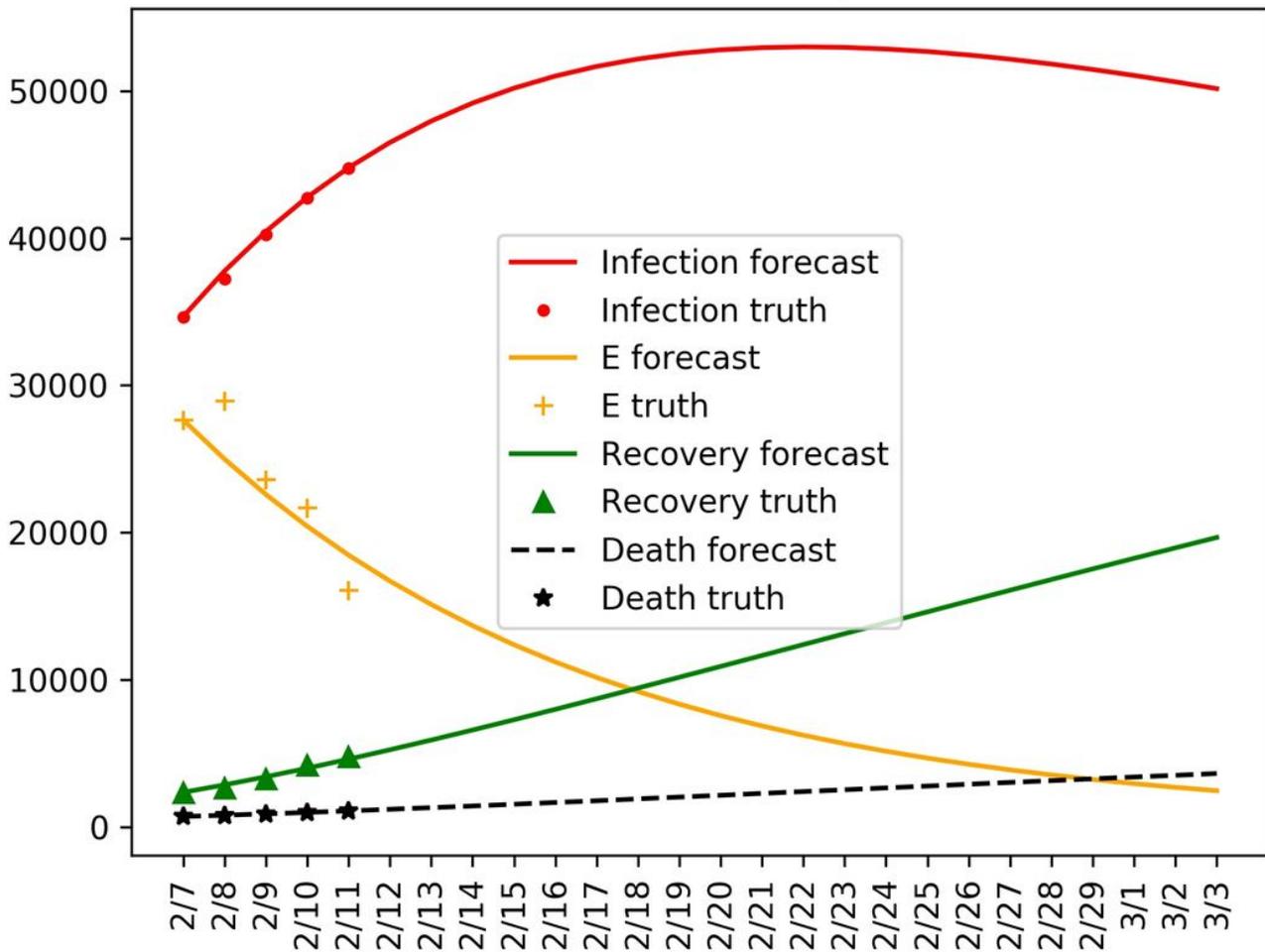


Figure 3

These figures illustrate fitting results of China, Hubei and non-Hubei provinces (simulation on January 31, 2020). (a) simulation and prediction results of China nationwide; (b) simulation and prediction results of Hubei province; (c) simulation and prediction results of non-Hubei provinces.

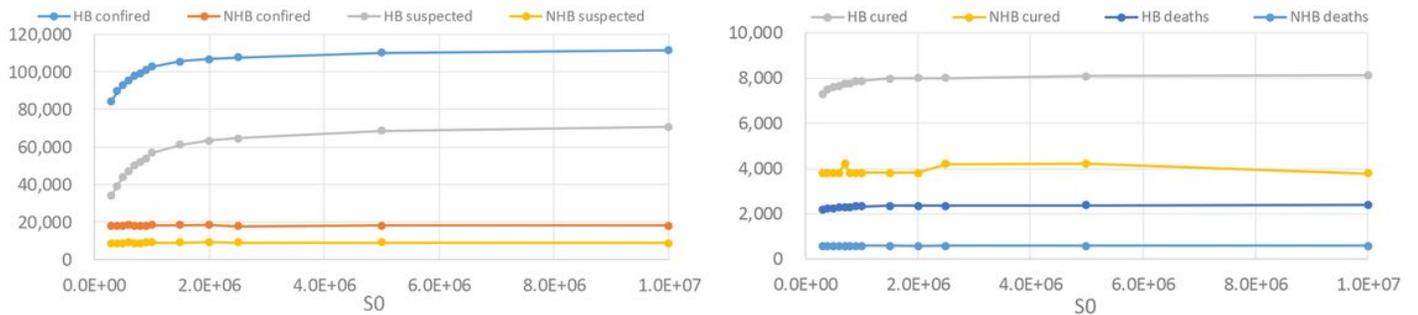


Figure 4

The estimation results of parameters distribution of China, Hubei and non-Hubei regions (simulated on January 31, 2020).

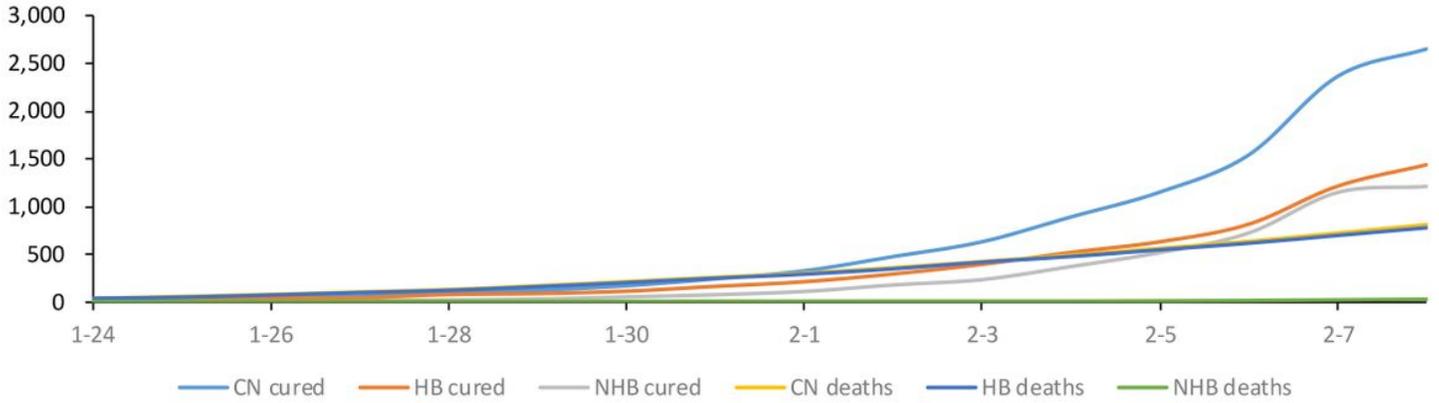


Figure 5

The temporal change of the parameters in proposed model.



Figure 6

Influences of training sample size on model parameters and estimated value.

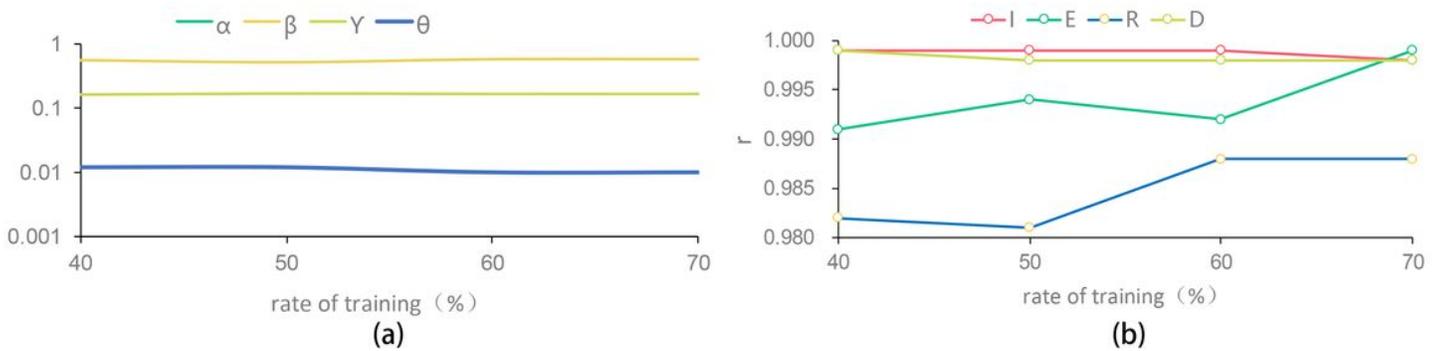
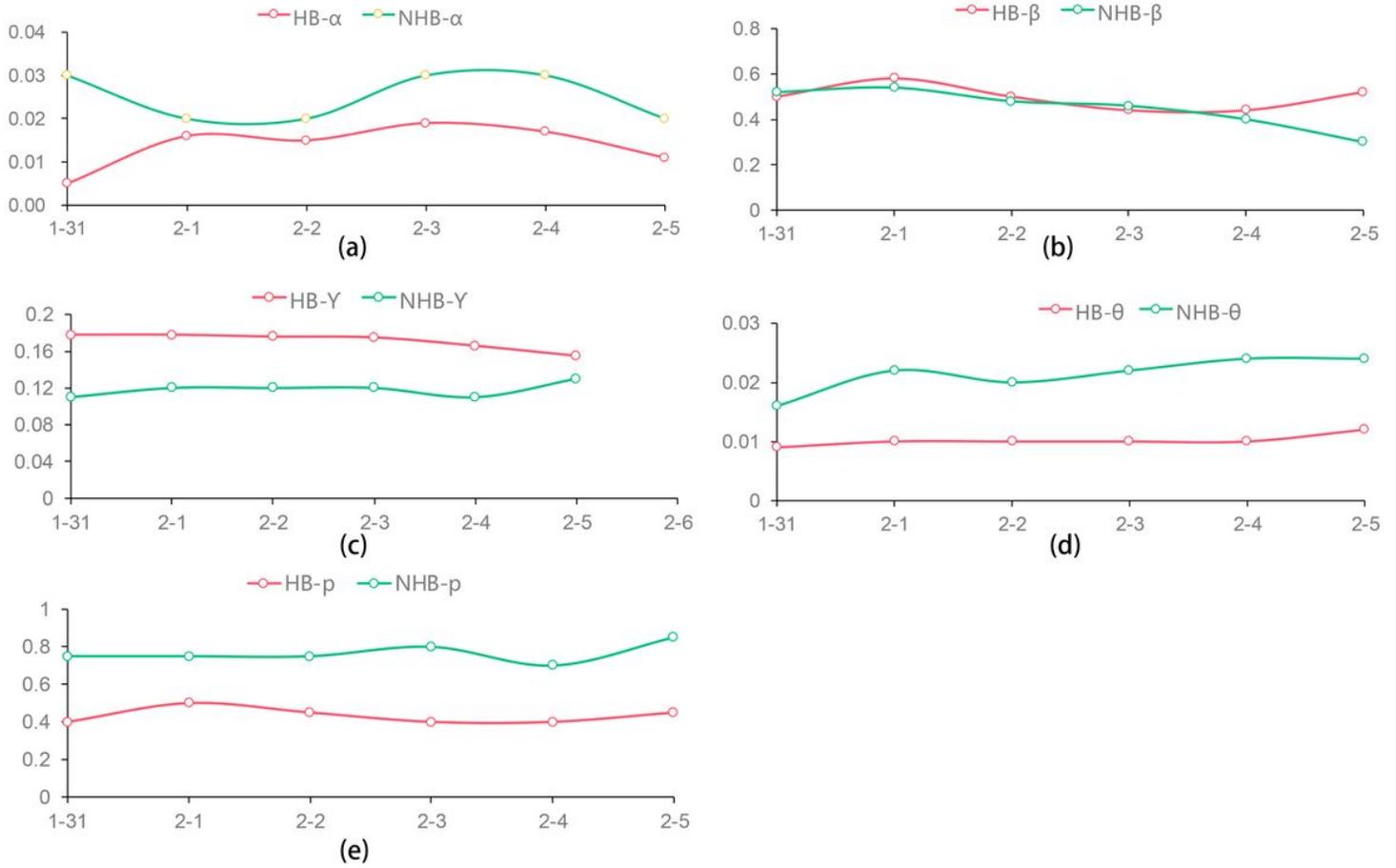


Figure 7

Prediction results of different S0



**Figure 8**

Nationwide forecast results based on dynamic parameter adjustment (forecast from February 7, 2020, peak number of infected cases is likely to reach at around February 22).