

Forecasting the long-term trend of COVID-19 epidemic using a dynamic model

Jichao Sun (✉ jichaosun@tencent.com)

Department of Healthcare and Medicine, Tencent Company

Xi Chen

Department of Healthcare and Medicine, Tencent Company

Ziheng Zhang

Department of Healthcare and Medicine, Tencent Company

Shengzhang Lai

Department of Healthcare and Medicine, Tencent Company

Bo Zhao

Department of Healthcare and Medicine, Tencent Company

Hualuo Liu

Department of Healthcare and Medicine, Tencent Company

Ruihui Zhao

Department of Healthcare and Medicine, Tencent Company

Alexander Ng

Department of Healthcare and Medicine, Tencent Company

Yefeng Zheng

Department of Healthcare and Medicine, Tencent Company

Research Article

Keywords: Coronavirus disease 2019, Forecasting model, machine learning, Dynamic-Susceptible-Exposed-Infective-Quarantined

Posted Date: May 27th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-31770/v1>

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

[Read Full License](#)

Abstract

Background The current outbreak of coronavirus disease 2019 (COVID-19) has recently been declared as a pandemic and spread over 200 countries and territories. Forecasting the long-term trend of the COVID-19 epidemic can help health authorities determine the transmission characteristics of the virus and take appropriate prevention and control strategies beforehand. Previous studies that applied the traditional epidemic models or machine learning models were subject to underfitting or overfitting problems.

Methods We propose a new model named Dynamic-Susceptible-Exposed-Infective-Quarantined (D-SEIQ), by making appropriate modifications of the Susceptible-Exposed-Infective-Recovered (SEIR) model and integrating machine learning based parameter optimization under epidemiological rational constraints. We used the model to predict the long-term reported cumulative numbers of COVID-19 cases in China from 27 January, 2020.

Results We evaluated our model on officially reported confirmed cases from three different regions in China, and the results proved the effectiveness of our model in terms of simulating and predicting the trend of COVID-19 outbreak. In China-Excluding-Hubei area within 7 days after the first public report, our model successfully and accurately predicted the 40 days long trend and the exact date of turning point. The predicted cumulative number (12,506) by 10, March 2020 was only 3.8% different with the actual number (13,005). The parameters obtained by our model proved the effectiveness of prevention and intervention strategies on epidemic control in China.

Conclusions The integrated approach of epidemic and machine learning models could accurately forecast the long-term trend of COVID-19 outbreak. The learned parameters suggested the effectiveness of intervention measures taken in China.

Introduction

Coronavirus disease 2019 (COVID-19) is an infectious pneumonia caused by severe acute respiratory syndrome coronavirus 2 [1]. The disease was first reported in December 2019 in Wuhan city, the capital of Hubei province in China, and has since then spread across China and globally [2]. As of 6 April 2020, a total of 1.27 million COVID-19 cases and 69,400 deaths have been reported in more than 200 countries and territories [3]. The World Health Organization (WHO) has declared the COVID-19 outbreak as a Public Health Emergency of International Concern and a pandemic recently [4].

Forecasting long-term trend of the epidemic can help health authorities determine the transmission characteristics of the virus and take appropriate prevention and control strategies beforehand. Recently, some researchers applied the traditional epidemic models like Susceptible-Exposed-Infective-Recovery (SEIR) or machine learning models like logistic regression to fit the trend of COVID-19 [5, 6]. To the best of our knowledge, most of those researches were performed retrospectively, or subject to overfitting or underfitting problems. The validity of SEIR model depends on accurate estimation of virus transmission characteristics such as the basic reproduction number R_0 , incubation period and infectious period. In a

real scenario, those parameters are not easy to estimate. For example, Wu et al. made estimation of the basic reproduction number using exported cases from China to abroad and estimated that 75 815 individuals had been infected in Wuhan as of Jan 25, 2020 [6], which significantly overestimated the figure. On the other hand, due to scarcity of training data and valid features, machine learning models were subject to overfitting, restricted to retrospective analysis, or only forecasting short-term trends [5, 7-10].

To address the aforementioned issues, we propose a novel model named Dynamic-Susceptible-Exposed-Infective-Quarantined (D-SEIQ), by making appropriate transformations of the SEIR model and integrating machine learning based parameter optimization under reasonable constraints, which improved the performance of long-term trend forecast for COVID-19 in China. In addition, the model parameters could provide insights into the analysis of COVID-19 transmission characteristics and the effectiveness of interventions.

Methods

D-SEIQ model

The primary differences from our D-SEIQ model and SEIR model are 1) replacing recovered individuals R with quarantined individuals Q, and 2) introducing dynamics to the calculation of the effective reproduction number R_t that is dependent on time.

Some previous work employed the traditional SEIR model, which assumes that the exposed individuals (who have been infected but no symptoms yet) are not infectious. However, it is reported that COVID-19 might be transmissible for exposed individuals [11]. Besides, due to lack of specialized treatment, the infectious period should not be interpreted as time between infective (I) and recovered (R) but time between infection (I) and quarantined (Q). Therefore, we proposed to replace the recovered individuals R with the quarantined individuals Q and the model became the SEIQ model. The quarantined individuals Q indicated the confirmed cases who were centrally quarantined. The epidemic spreading model for the SEIQ model is shown in Figure 1.

The transmission dynamics are governed by the following system of equations: (see Equation 1 in the Supplementary Files)

where $N = S(t) + E(t) + I(t) + Q(t)$ is the total population, which is assumed consistent.

Like the SEIR model, parameter β indicates the infectious rate with $\beta = R_t/TE$ where R_t is the dynamic effective reproduction number and TE is the average duration of incubation; parameter σ indicates the incubation rate with $\sigma = 1/TE$. However, in our model, parameter γ indicates the quarantine rate with $\gamma = 1/TI$ (where TI is average duration of an infectious individual to be detected and quarantined). The parameter TI may vary across different regions and the difference reflects the timeliness of patient detection and admission.

The basic reproduction number R_0 is the most important parameter to determine the intrinsic transmissibility of COVID-19, and it is defined as the average number of infections one infectious agent can generate over the course of the infectious period without any interventions. R_0 was assumed to be a constant or arbitrarily modified at specific points for forecasting in previous work [12, 13]. However, in real-world scenarios, with the development of epidemic, more and more interventions are often taken to control the spread, which gradually reduce R_0 . In this work, the basic reproduction number R_0 is generalized to a dynamic value R_t , which is defined as the average number of secondary infectious cases generated by an infectious at time t . After the worldwide outbreak of COVID-19, many governments took considerable measures to contain the spread of the virus. In our preliminary analysis and similar to previous work [14], the infectious rate β was shown to decrease exponentially with time. As parameter TE is constant, the effective reproduction number R_t should follow similar pattern as decreasing exponentially with time. Thus, we introduced time-dependent dynamics to the calculation of R_t for better simulation of the real-world transmission, (see Equation 2 in the Supplementary Files)

where R_∞ is the final reproduction number at the end of the pandemic and θ is the decrease ratio of the reproduction number, which is associated with the corresponding interventions. When $t = 0$, $R_t = R_0$, and it gradually reduces to R_∞ . The epidemic is considered to be under control with $R_t < 1$, and the reasonable range of R_∞ was referred to some previous analysis of coronavirus [15].

Parameter constraints and optimization

The simulation and prediction of the D-SEIQ model requires determination of the parameters R_0 , R_∞ , TE , TI , θ . Although we incorporated machine learning to help us to fit the reported data, the parameter range needs to be set carefully and to conform to epidemiological rationality. For instance, Wu et al. applied an adjusted SEIR model to estimate R_0 ($R_0 = 2.68$) in major cities of China by analyzing the number of cases exported from Wuhan internationally [6]. Some work concluded that the daily reproduction number varied between 2 and 8.[16] Therefore, we set the reasonable range for parameter R_0 to be [2, 7]. Likewise, after reviewing the previous work on the analysis of COVID-19 [2, 11], we summarized the ranges for parameters in our model as Table 1. And, we set $TE > TI$ as additional constraints. Therefore, the parameter optimization process is as follows:

- a. Initialize the number of confirmed cases Q at time $t = 0$ according to the official report.
- b. Initialize the parameters R_0 , R_∞ , TE , TI , θ
- c. Calculate the time-dependent effective reproduction number R_t
- d. Solve ordinary differential equations in Equation (1) to determine $E(t)$, $I(t)$, $Q(t)$
- e. Set loss function as the sum of mean squared errors of daily and cumulative confirmed numbers, and then estimate the parameters R_0 , TE , TI , R_∞ , θ based on grid search with dynamically adapted search steps to obtain the best D-SEIQ model at time t .

Data processing

We obtained the updated data of the cumulative confirmed cases from the National Health Commission (NHC) of the People's Republic of China. The newly confirmed cases were also collected on a daily basis. Considering that medical resources and interventions might vary in different regions, we fitted our model on the data from three different regions: 1) China excluding Hubei, 2) Hubei excluding Wuhan, and 3) Wuhan.

Moreover, we adjusted the number of newly confirmed cases in Wuhan between 12 February and 14 February, due to the inclusion of clinically confirmed cases without coronavirus test. The clinically confirmed cases between 12 February and 14 February were assumed to be suspicious cases in last 7 days. Specifically, we redistributed the clinically confirmed cases according to the distribution of suspected cases over the past 7 days.

Forecasting long-term trends of confirmed case numbers

Because the China's NHC publicly reported case numbers starting from 20 January, we set this date as the starting point of our training data. As of 10 March, the daily increased case numbers declined to single digits across most areas in China, we set this date as the ending point of our model.

We updated our models dynamically from the 7th day following the starting point (i.e., 27 January). In this article, we presented the prediction of our models at the time points of 1st to 5th week, namely 27 January, 4 February, 11 February, 18 February, and 25 February.

For example, the model for the first week (as of 27 January) used the data from 20 January to 26 January for model construction and forecasted the daily increased and cumulative case numbers from 27 January to 10 March.

Results

The simulation and prediction of our D-SEIQ models is illustrated from three different regions: China excluding Hubei, Hubei excluding Wuhan, and Wuhan.

China excluding Hubei

The D-SEIQ model with prediction date of 26 January showed that the cumulative number would reach 65,282 (red dotted line in Figure 2) on 10 March. In retrospect, our model greatly overestimated the development of the epidemic, possibly because at the early stage of the epidemic when intervention had not taken its effect, the number of cases increased sharply and did not show the potential decline of R_t . The overestimation also illustrated the effectiveness of the subsequent control measures.

The D-SEIQ model trained on 27 January showed that the cumulative number would reach 12,506 on 10 March, and the daily number would reach the peak on 1 February. In retrospect, the prediction was quite close to the real scenario. The real cumulative number on 10 March was 13,005 which was only 3.8% different from predicted value. Also, the turning point predicted by our model is exactly the same as the

actual date (around 1 February to 3 February). Therefore, in the region of China excluding Hubei, the D-SEIQ model is shown to successfully estimate the trend for up to 40 days, with one-week data after the first public report.

At the late stage of epidemic spread, the model is capable of fitting on previous data and also predicting the epidemic development. For example, on 11 February, we predicted the cumulative number was 13,006 at the end point while the true value is 13,005.

The parameters learned at the late stage could accurately reflect the intrinsic characteristics of COVID-19. Thus, the parameters on 25 February were used as the estimation of true values. In the region of China excluding Hubei, the basic reproduction number R_0 was estimated to be 6.3; the decrease ratio θ to be 0.2; the incubation period TE to be 3 days; and the infectious period TI to be 2 days. The effective reproduction number R_∞ ultimately dropped to around 0.3.

Hubei excluding Wuhan

The number of confirmed cases grew rapidly in the region of Hubei excluding Wuhan in the first week, which biased our model on 27 January to enormously overestimate the peak value. Our model predicted that the cumulative number would reach 65,763 by 10 March. On the other hand, the overestimation also indicates that, without control, the epidemic would show explosive growth as the influence of control measures remained unseen at the early stage of epidemic.

After the clinically confirmed cases between 12 February and 14 February were adjusted by redistribution, we re-trained our model with adjusted values (Figure 3). The model on 14 February after adjustment showed that the cumulative number would reach 18,844 with an error of 6% compared with the real number.

Similarly, based on the model of the late stage of epidemic (25 February), the transmission parameters of virus were estimated as follows: the basic reproduction number R_0 was 6.3; the decrease ratio θ was 0.15; the final reproduction number R_∞ was 0.2; the incubation period TE was 3 days, and the infectious period TI was 2 days.

Wuhan

In the early days of epidemic outbreak in Wuhan, due to the deficiency of detection capabilities and limited medical resources, the reported numbers were far below the real incidences. During the first week, the daily increased numbers even showed a declining trend, and the D-SEIQ model on 27 January consequently underestimated the epidemic development. There was a large increase in clinically confirmed cases between 12 February and 14 February. We adjusted the numbers on 14 February and the prediction showed that the cumulative number would reach 54,492 at the end point, with an error of 9% from the actual number of 49,980. On 18 February, the D-SEIQ model showed a convincing simulation of

the overall trend, and the overall predicted curve indeed fitted the adjusted values quite well (grey dashed line in Figure 4).

The estimated parameters of the COVID-19 transmission were as follows: the basic reproduction number R_0 was estimated to be 4.63; the decrease ratio θ was 0.1; the final reproduction number R_∞ was 0.15; the incubation period TE was 3 days; and the infectious period TI was 2.5 days.

Analysis of reproduction number R_t

We further analyzed the reproduction number R_t by our D-SEIQ models. We used the R_t learned at the late stage of the simulation. We plotted the R_t curve from 20 January to 10 March as Figure 5 to compare the reproduction numbers in three different regions. At the initial time, R_0 was 6.3 in China excluding Hubei and Hubei excluding Wuhan, both of which was larger than that in Wuhan with $R_0 = 4.63$. However, the decrease ratio θ for R_t was largest in China excluding Hubei (0.20), followed by Hubei excluding Wuhan and then Wuhan. Therefore, R_t in China excluding Hubei dropped below 1 the earliest, meaning that COVID-19 was under control in other provinces sooner than Hubei province. The final R_∞ of three different regions all approached zero, demonstrating a great achievement in epidemic control and interventions.

Discussion

We proposed a new model named D-SEIQ, which takes appropriate modifications of the SEIR model and combines with parameter optimization of machine learning. We evaluated our model on officially reported data from three different regions in China, and the results proved the effectiveness of our model in terms of simulating and predicting the trend of COVID-19 outbreak and regional spread. Specially, in China excluding Hubei area within 7 days after the first public report, our model successfully and accurately predicted the 40 days long trend and the exact date of turning point.

Traditional epidemic transmission models like SEIR need accurate estimation of model parameters such as basic reproduction number, incubation period and infectious period through epidemiological investigation. However, in term of a new epidemic, due to the rapid outbreak, insufficient sample size, and the deviation of investigated data from true data, the traditional epidemic transmission models usually poorly fit to the data. In practice, scholars often made various assumptions for calculation or even used relevant parameters of other virus as substitutions. For example, Wu et al. adopted serial interval estimates for SARS as substitutions and estimated that 75 815 individuals were infected in Wuhan as of Jan 25, 2020 [6], which significantly overestimated the figure. On the other hand, machine learning methods such as logistic regression models were subject to overfitting problems [17], which means they could fit the training data well but fail to predict on unseen data. It was because the data and features were not sufficient, and the models lack epidemic rationality. Deep neural network sequence models like long short-term memory (LSTM) had weak capability to predict the long-term trends and the turning point [18].

Our model takes advantage of both epidemic and machine learning models, which combine the explainability of epidemic model with the data-fitting ability of machine learning. In the process of machine learning, we set the parameters within a reasonable range, and exploit mutual constraints between the parameters.

Meanwhile, we innovatively introduced dynamic R_t , which can reflect the time-dependent influence of intervention measures on basic reproduction number. Overall, our approach could more likely simulate the true scenario of the COVID-19 spread, thus making better fitting and predictions.

Furthermore, the parameters learned by our D-SEIQ model could provide some insights into the assessment of the prevention and control measures on COVID-19. Firstly, the basic reproduction number was relatively large (4 to 6), which was larger than SARS-COV with R_0 ranging from 1.6 to 3.7 [15, 19, 20]. Without strong and effective intervention measures including cities lockdown, travel containment, mask wearing, quarantine, and screening, it could lead to catastrophic consequences to the society. The final reproduction number of different areas of China gradually dropped to around 0.2, illustrating the considerable effect and the significant importance of interventions from governments or public. Secondly, the decrease ratio of R_t was slower in Wuhan which indicates the shortage of medical resources and delayed patient admission in Wuhan. This conclusion is also supported by the parameter of infectious period (TI) with a larger value in Wuhan than other regions of China. Moreover, our model obtained the same incubation period (TE) with 3 days across three regions, which was consistent with that from Chinese CDC official report [11].

The D-SEIQ model is applicable only when the following conditions are satisfied: adequate medical capacities, consistency of control measures and ascertainment criteria, and timely case detection and reporting. This explained the reason why our model performed better in China excluding Hubei region. Therefore, caution need to be taken when applying our model to other countries. The detection and reporting were not timely in some countries like the United States at early phase, and subsequent control measures were introduced at different time points, which might influence the prediction results.

Conclusion

We have proposed a new approach for forecasting the COVID-19 long-term trend. The model has accurately predicted the long-term trend of the epidemic in China, and the parameters learned from the model suggested the effectiveness of the intervention measures that have been taken in China, which can help us analyze and fight against the new epidemic.

Declarations

Acknowledgments

We want to thank Yan Liu from the Chinese Centers for Disease Control and Prevention for providing guidance on data analysis.

Funding

There is no funding associated with the conduction of this study

Authors' contributions

JS contributed to the D-SEIQ model proposal. JS, XC and ZZ contributed to study conception and design, analysis and interpretation of data, drafting of the manuscript, and model construction. SL contributed to acquisition of data. YZ contributed to analysis and interpretation of data, drafting of the manuscript, and critical revision of the manuscript for important intellectual content. BZ, RZ, HL and AN contributed to critical revision of the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

- 1 WHO. Naming the coronavirus disease (COVID-19) and the virus that causes it. 2020. [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it) (accessed Feb 28, 2020).
- 2 Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* 2020; **382**(8): 727-33.
- 3 CSSE at Johns Hopkins University. Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). 2020. <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6> (accessed April 6, 2020).
- 4 WHO. WHO Director-General's opening remarks at the media briefing on COVID-19. March 11, 2020.
- 5 Zhou X, Hong N, Ma Y, et al. Forecasting the Worldwide Spread of COVID-19 based on Logistic Model and SEIR Model. *medRxiv* 2020: 2020.03.26.20044289.
- 6 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. *The Lancet* 2020; **395**(10225): 689-97.
- 7 Tátrai D, Várallyay Z. COVID-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability. *arXiv e-prints*, 2020. <https://ui.adsabs.harvard.edu/abs/2020arXiv200314160T> (accessed March 01, 2020).

- 8 Roosa K, Lee Y, Luo R, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling* 2020; **5**: 256-63.
- 9 Jia L, Li K, Jiang Y, Guo X, Zhao T. Prediction and analysis of Coronavirus Disease 2019. *arXiv e-prints*, 2020. <https://ui.adsabs.harvard.edu/abs/2020arXiv200305447J> (accessed March 01, 2020).
- 10 Yang Z, Zeng Z, Wang K, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease* 2020; **12**(3): 165-74.
- 11 Guan W-j, Ni Z-y, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine* 2020.
- 12 Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* 2020: 2020.01.23.20018549.
- 13 Gupta R, Pandey G, Chaudhary P, Pal SK. SEIR and Regression Model based COVID-19 outbreak predictions in India. *medRxiv* 2020: 2020.04.01.20049825.
- 14 Chen B, Shi M, Ni X, et al. Visual Data Analysis and Simulation Prediction for COVID-19. *arXiv e-prints*, 2020. <https://ui.adsabs.harvard.edu/abs/2020arXiv200207096C> (accessed February 01, 2020).
- 15 Riley S, Fraser C, Donnelly CA, et al. Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. *Science* 2003; **300**(5627): 1961-6.
- 16 Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*.
- 17 Batista M. Estimation of the final size of the COVID-19 epidemic. *medRxiv* 2020: 2020.02.16.20023606.
- 18 Du S, Wang J, Zhang H, et al. Predicting COVID-19 Using Hybrid AI Model. *SSRN Electronic Journal* 2020.
- 19 Chowell G, Castillo-Chavez C, Fenimore PW, Kribs-Zaleta CM, Arriola L, Hyman JM. Model parameters and outbreak control for SARS. *Emerging Infectious Diseases* 2004; **10**: 1258+.
- 20 Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature* 2005; **438**(7066): 355-9.

Table

Table 1 The constrained range for parameters with epidemic rationality

parameter	reasonable range
-----------	------------------

R_0	[2, 7]
TE	[3, 11]
TI	[1, 5]
R_∞	[0.05, 0.35]
θ	[0.05, 0.45]

R_0 basic reproduction number; TE incubation period; TI infectious period; R_∞ the final value of R_t ; θ decrease ratio for R_t

Figures

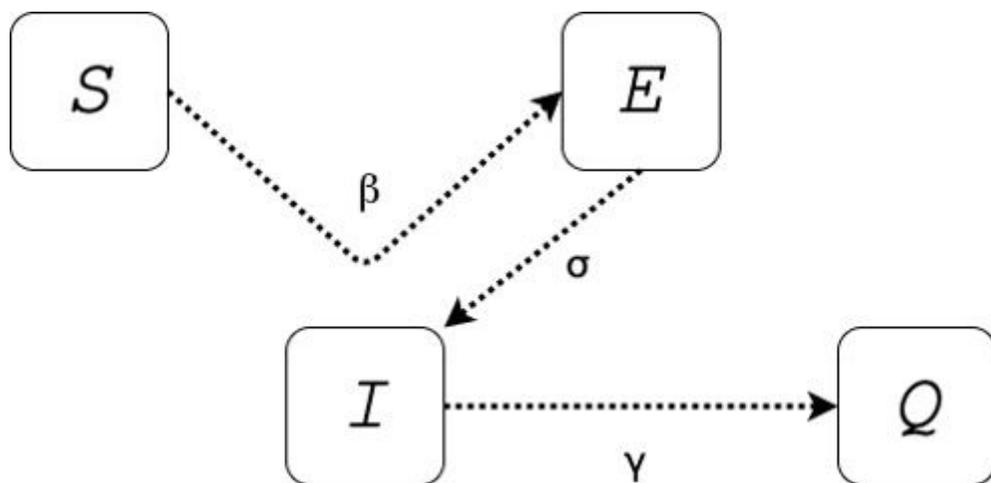


Figure 1

Epidemic spreading diagram for SEIQ model S: susceptible; E: exposed; I: infective; Q: quarantined; β indicates the infectious rate. Parameter σ indicates the incubation rate with $\sigma = 1/TE$ (incubation period). Parameter γ indicates the quarantine rate with $\gamma = 1/TI$ (infectious period).

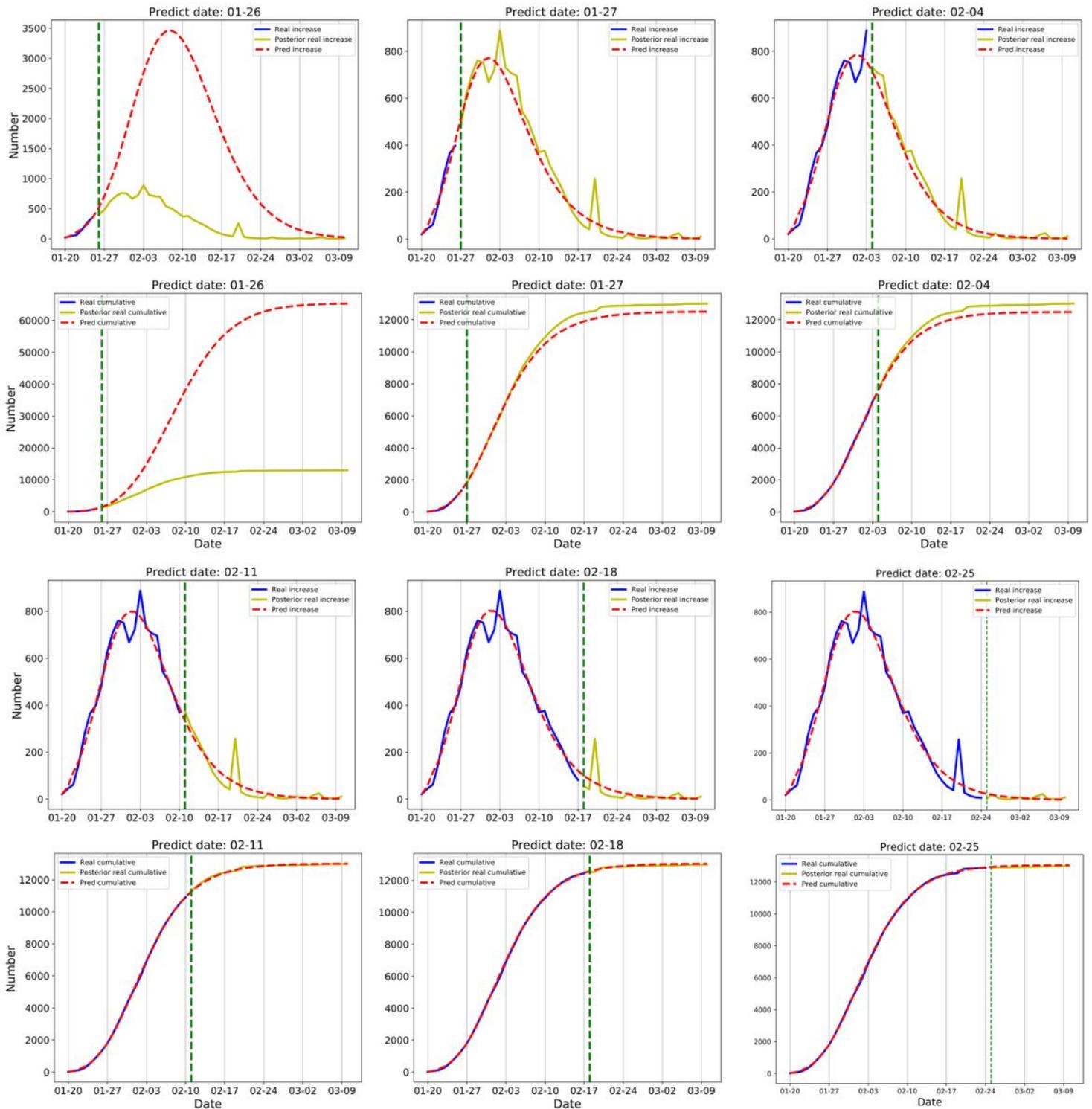


Figure 2

Simulation and prediction results of daily and cumulative confirmed cases in the region of China excluding Hubei Data were shown on each figure with 6 different prediction dates. The results of daily confirmed cases are placed on 1st and 3rd rows and the results of cumulative confirmed cases are placed on 2nd and 4th rows. Green vertical line: the prediction date which separate training data and test data. Solid blue line: the real number of confirmed cases before prediction date, namely training data. Red

dotted line: the number predicted by the D-SEIQ model. Solid yellow line: the retrospective number of confirmed cases, namely test data.

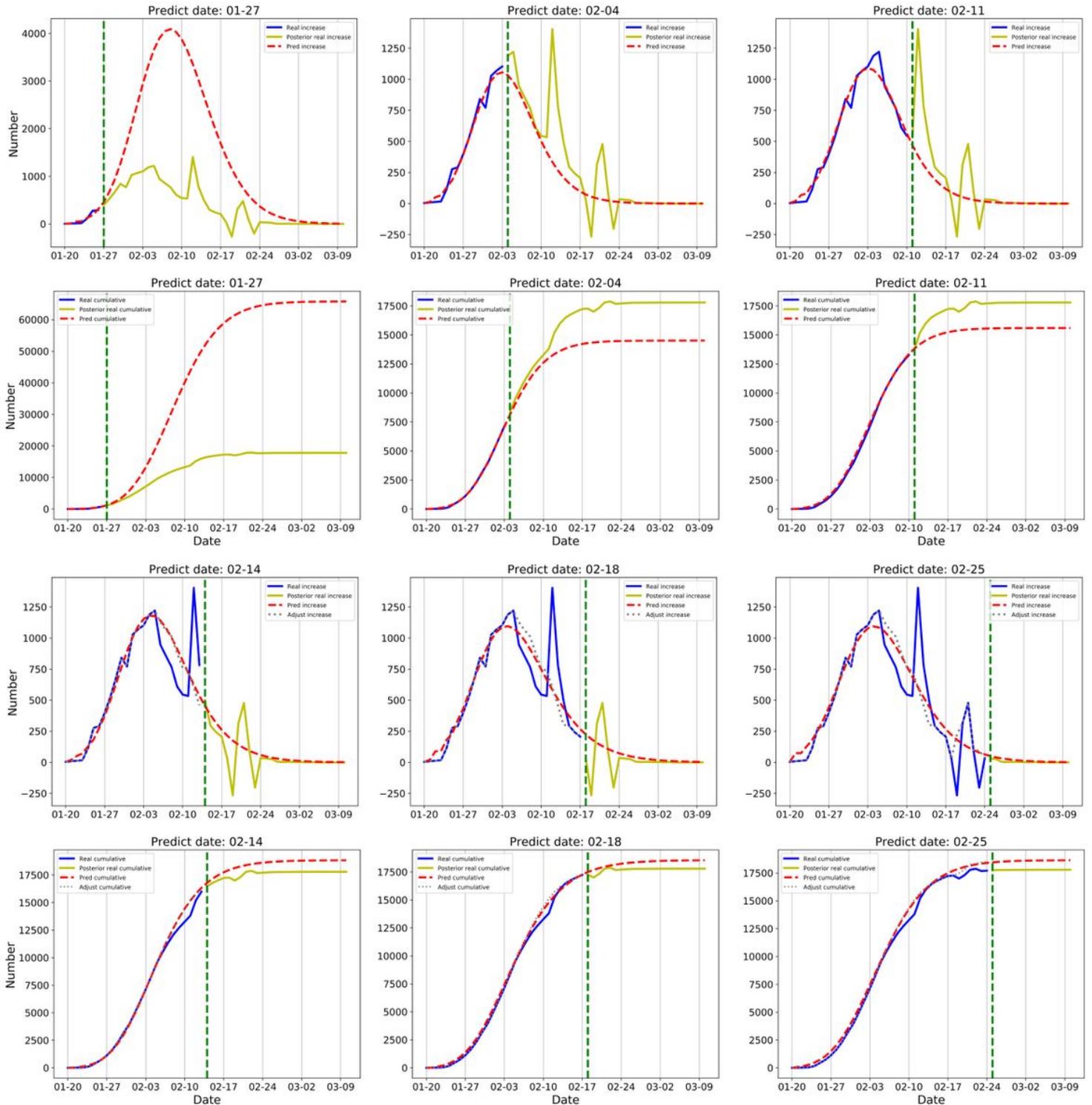


Figure 3

Simulation and prediction results of daily and cumulative confirmed cases in the region of Hubei excluding Wuhan Data were shown on each figure with 6 different prediction dates. The results of daily confirmed cases are placed on 1st and 3rd rows and the results of cumulative confirmed cases are placed on 2nd and 4th rows. Green vertical line: the prediction date which separate training data and test

data. Solid blue line: the real number of confirmed cases before prediction date, namely training data. Red dotted line: the number predicted by the D-SEIQ model. Solid yellow line: the retrospective number of confirmed cases, namely test data. Grey dashed line: the numbers after adjustment of the clinically confirmed cases.

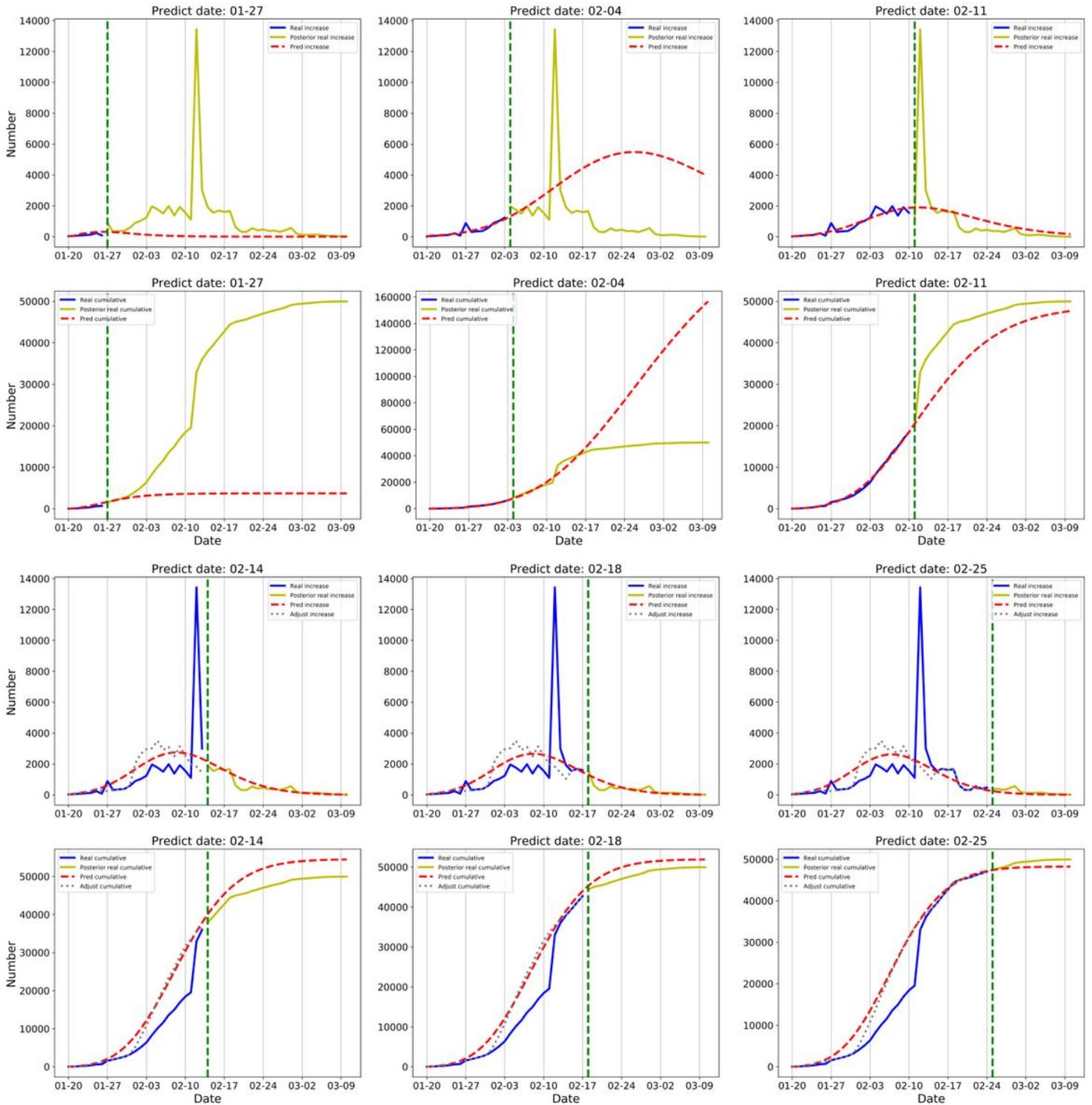


Figure 4

Simulation and prediction results of daily and cumulative confirmed cases in the region of Wuhan Data were shown on each figure with 6 different prediction dates. The results of daily confirmed cases are

placed on 1st and 3rd rows and the results of cumulative confirmed cases are placed on 2nd and 4th rows. Green vertical line: the prediction date which separate training data and test data. Solid blue line: the real number of confirmed cases before prediction date, namely training data. Red dotted line: the number predicted by the D-SEIQ model. Solid yellow line: the retrospective number of confirmed cases, namely test data. Grey dashed line: the numbers after adjustment of the clinically confirmed cases.

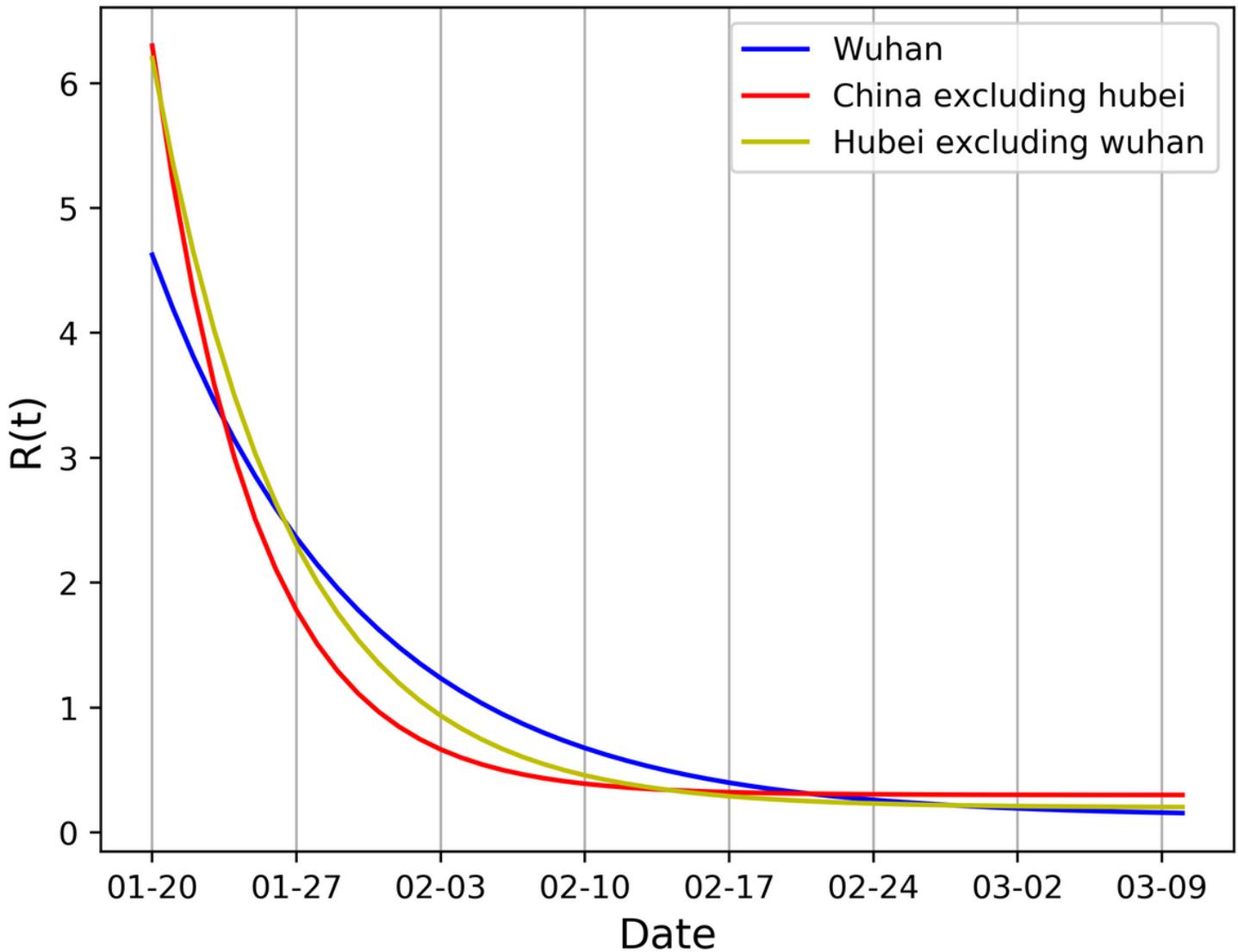


Figure 5

Curve of the change of dynamic effective reproduction number R_t in three different regions of China.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Equations.pdf](#)