

# A simple transmission dynamics model for predicting the evolution of COVID-19 under control measures in China

**Chenjing Shang**

College of Life Science and Oceanography, Shenzhen University

**Yang Yang** (✉ [yang.yang0513@gmail.com](mailto:yang.yang0513@gmail.com))

Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences

**Gui-Ying Chen** (✉ [gychen@scsio.ac.cn](mailto:gychen@scsio.ac.cn))

State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences and Southern Marine Science and Engineering Guangdong Laboratory

**Xiao-Dong Shang** (✉ [xdshang@scsio.ac.cn](mailto:xdshang@scsio.ac.cn))

State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences and Southern Marine Science and Engineering Guangdong Laboratory

---

## Research Article

**Keywords:** COVID-19, prediction model, coronavirus, growth rate

**Posted Date:** April 2nd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-20994/v1>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at *Epidemiology and Infection* on January 1st, 2021. See the published version at <https://doi.org/10.1017/S0950268821000339>.

# A simple transmission dynamics model for predicting the evolution of COVID-19 under control measures in China

Chenjing Shang<sup>1</sup>, Yang Yang<sup>2,\*</sup>, Gui-Ying Chen<sup>3,4,\*</sup>, Xiao-Dong Shang<sup>3,4,\*</sup>

<sup>1</sup> College of Life Science and Oceanography, Shenzhen University, Shenzhen 518060, China;

<sup>2</sup> Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, 572000, China;

<sup>3</sup> State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China;

<sup>4</sup> Southern Marine Science and Engineering Guangdong Laboratory, Guangzhou 511458;

\* Correspondence: xdshang@scsio.ac.cn, yang.yang0513@gmail.com, gychen@scsio.ac.cn

## Abstract:

Epidemic forecasting provides an opportunity to predict geographic disease spread and counts when an outbreak occurs and plays a key role in preventing or controlling their adverse impact. However, conventional prediction models based on complex mathematical modeling rely on the estimation of model parameters, which yields unreliable and unsustainable results. Herein, we proposed a simple model for predicting the epidemic transmission dynamics based on nonlinear regression of the epidemic growth rate and iterative methods, which is applicable to the progression of the COVID-19 outbreak under the strict control measures of the Chinese government. Our model yields reliable and accurate results as confirmed by the available data: we predicted that the total number of infections in mainland China would be 91,253, and the maximum number of beds required for hospitalized patients would be 62,794. We inferred that the inflection point (when the growth rate turns from positive to negative) of the epidemic across China would be mid-February, and the end of the epidemic would be in late March. This model is expected to contribute to resource allocation and planning in the health sector while providing a theoretical basis for governments to respond to future global health crises or epidemics.

**Keywords:** COVID-19; prediction model; coronavirus; growth rate

## 1. Introduction

In December 2019, the outbreak of COVID-19 that could cause severe respiratory symptoms and even deaths emerged in Wuhan, Hubei Province, China<sup>1,2</sup>. This new coronavirus was confirmed to be able to transmit between humans on January 20, 2020, significantly increased the risk of international spread<sup>3</sup>. To mitigate the spreading of the epidemic, the Chinese central government progressively implemented the highest-level metropolitan-wide quarantine control in Wuhan city and 31 provinces since January 23-24, 2020<sup>4</sup>. In the meantime, scientists are racing to characterize the virus, model epidemics, and develop diagnostic reagents and vaccines<sup>5-9</sup>.

As of February 26, the spreading of COVID-19 has been reported in 52 countries, and the number of confirmed cases worldwide has reached 83,389. The progression of the epidemic has aroused widespread concerns of scientists<sup>10,11</sup>. The epidemiological community has long used the basic reproduction number  $R_0$  to describe the spread of epidemics<sup>12</sup>. It can be thought of the expected number of cases directly generated by one case where all individuals are susceptible to infections in the case of natural transmission of the virus. At present, many groups have estimated the  $R_0$  value of COVID-19 (1.4-6.47, average 2-3) through different models, but the results differ greatly and are far from reality<sup>13,14</sup>. On January 16, Neil Ferguson's team at Imperial College predicted that there would be 1,723 infected people in Wuhan on January 12, and the number of infected people worldwide would reach 100,000 by January 26<sup>15</sup>. Wu predicts that the number of infected people in Wuhan will reach 75,815 on January 25<sup>13</sup>. Huang and Qiao proposed a data-driven model to predict the peak of the outbreak<sup>16</sup>. As the Chinese government has implemented stringent highest-level health intervention, none of these models are suitable for predicting the development of epidemics under this circumstance. Finding an accurate and simple predictive dynamic model is the key to predicting the evolution of the epidemic.

Herein, we proposed a simple model based on non-linear regression and iterative methods to predict the progression of the epidemic under stringent governmental control. Using this model, we predicted the inflection point and end time of the epidemic, the expected number of infected patients, and the maximum number of beds required for regions including Wuhan, Hubei province, Guangdong province, and mainland China. This model can help the government to prepare in advance in the allocation of medical resources and the deployment of medical staff in the event of an epidemic crisis.

## 2. Results and Discussion

The methods we use are the non-linear regression and iterative methods commonly used in the study of Natural science<sup>17-20</sup>. Compared with the Susceptible-Infected-Removed (SIR) and Susceptible-Exposed-Infectious-Recovered (SEIR) models used in epidemiology, our method is simple, accurate, and reliable.

The daily growth rate of confirmed cases is calculated as follows:

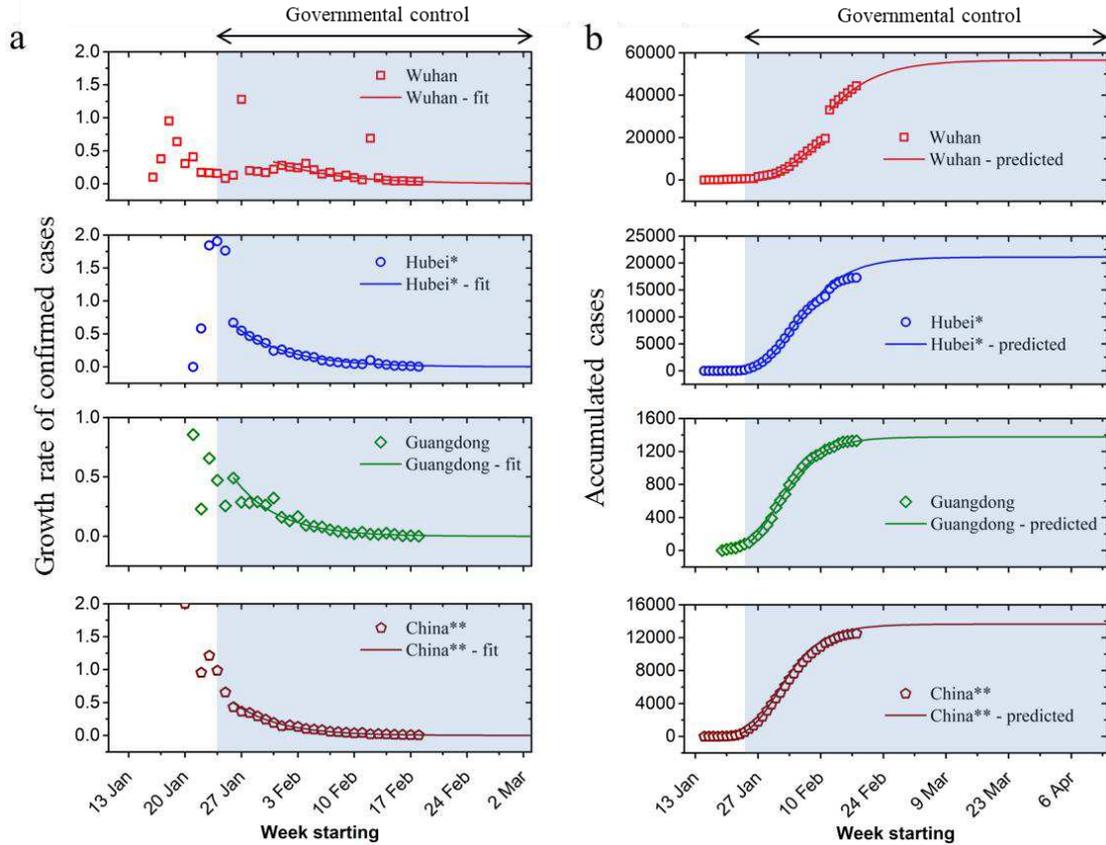
$$T = \frac{N_n - N_{n-1}}{N_{n-1}} \quad (1)$$

where  $N_n$  and  $N_{n-1}$  are the number of confirmed cases on day  $n$  and  $n-1$ , respectively.

The daily growth rate of confirmed cases is obtained by formula (1) and plotted in **Figure 1a**, where the hollow markers are calculated from real data and the solid lines are fitted curve. The growth rate  $T$  fluctuates during the early stage of the outbreak and shows a large value because of a lack of governmental health intervention as well as the small sample size has a greater impact on  $T$ .  $T$  stabilizes and decreases rapidly after January 23-24, 2020 due to nationwide quarantine policy. This growth rate curve decays exponentially and is obtained by fitting:

$$T = ae^{-\beta t} \quad (2)$$

Here  $a$  is a constant that represents the growth rate at  $t = 0$ ,  $\beta$  is an attenuation coefficient that indicates the efficiency of government isolation and quarantine, and  $t$  is the time representing the evolution of the epidemic. Formula (2) shows the evolution of the epidemic under the government's stringent isolation and quarantine of patients. The attenuation coefficient  $\beta$  is a parameter that measures the efficiency of isolation and quarantine.



**Figure 1.** (a) Growth rate of confirmed cases, and (b) accumulated cases for Wuhan, Hubei province, Guangdong province, and mainland China. Asterisks denote Wuhan or Hubei is excluded (\*: Wuhan excluded; \*\*: Mainland China excluded Hubei province). Red square for Wuhan, blue circle for Hubei, green diamond for Guangzhou, and brown pentagon for mainland China.

Note that the Chinese central government implemented nationwide quarantine policy from January 23-24, 2020. Since there were delays for the control measures to take effect, except for Wuhan, we used the data starting from January 26 for model fitting (**Figure 1a**). On the other hand, due to the shortage of monitoring tools and medical resources, the data of Wuhan during the early stage cannot truly reflect its real epidemic situation. We therefore used data of Wuhan starting from January 31 for model fitting. Furthermore, the jump in confirmed cases reported in Wuhan on February 12 is due to a change in the criteria for counting diagnoses of the virus.

The summary of fitted parameters is displayed in **Table 1**. The trends of the four curves are comparable (**Figure 1**), which indicates that under the mandatory quarantine by the Chinese government, the transmission and evolution of the epidemic are artificially limited. The values of fitted attenuation coefficient  $\beta$  of the four curve show an increasing trend (Wuhan 0.13; Hubei 0.15; Guangdong 0.18; mainland China 0.18). We speculate that the faster the growth rate decays, the greater the government's impact over the epidemic. Note that the outbreak in Wuhan was not well controlled at the early phase of the outbreak, which led to the rapid spread of the outbreak and the rapid increase in the number of infected patients. Therefore, the attenuation coefficient of the growth rate is the smallest. The attenuation coefficient for mainland China (except Hubei Province) is similar to that of Guangdong Province, indicating that the extent of public health intervention is comparable. The smaller attenuation coefficient of Hubei Province compared to the rest of China is possibly due to the challenge of implementing the control measures because of the overwhelmed medical system in Hubei.

**Table 1.** List of fitted parameters for formula (2).

Parameter	Wuhan	Hubei* province	Guangdong province	China**
$a$	2.503	3.420	3.657	3.553
$\beta$	0.126	0.153	0.182	0.179
$R^2$	0.905	0.981	0.929	0.984

\*: Wuhan excluded; \*\*: Mainland China excluded Hubei province.

Based on formula (2), we use nonlinear regression and iterative methods to predict cumulative confirmed cases:

$$N = N_0 \prod_{t=1}^n (1 + ae^{-\beta t}) \quad (3)$$

where  $N$ ,  $N_0$  is the number of confirmed cases, and the expected number of cases at  $t = 0$ , respectively.

Using the formula (3), the cumulative numbers of confirmed cases over time for Wuhan, Hubei (except Wuhan), Guangdong province, and mainland China (except Hubei) are predicted. **Figure 1b** shows the up-to-date confirmed cases (hollow markers) and predictions (solid lines) using our model for the four datasets. The model predictions are in good agreement with the actual numbers. Based on our model, the total predictions of infected people in Wuhan, Hubei (except Wuhan), Guangdong province, and mainland China (except Hubei) during this outbreak are 56,519, 21,093, 1,377 and 13,641 respectively (**Table 3**). The death toll can also be predicted in the same way. Furthermore, we used this model to predict the further development and evolution of the epidemic. Although the number of confirmed cases is increasing (**Figure 1b**), the daily growth rate of confirmed cases is decreasing (**Figure 1a**). It indicates that the move of the epidemic goes better. When the confirmed cases

reach the peak, the epidemic tends to the end. Using this method, the estimated end time of the epidemic in Wuhan, Hubei (except Wuhan), Guangdong province, and mainland China (except Hubei) will be May 5, April 13, March 15, and March 27, 2020 respectively.

Forecasting the number of hospitalized patients is conducive to the accurate preparation of beds and other resource allocation for the hospital, and provides significant guidance in response to the outbreak. The number of hospitalized patients, or the number of required hospital beds, can be obtained by subtracting the number of confirmed cases by the number of recoveries and deaths as follows:

$$M = M_1 - M_2 - M_3 \quad (4)$$

where  $M$ ,  $M_1$ ,  $M_2$ , and  $M_3$  are the number of hospitalized patients, confirmed patients, recoveries, and deaths, respectively. Substituting formula (1) we obtain the growth rate of hospitalized patients for these four regions and the results are plotted in **Figure 2a**. The choices of symbols and colors correspond to those in **Figure 1**. This growth rate curve also follows an exponentially decaying trend and is obtained by fitting:

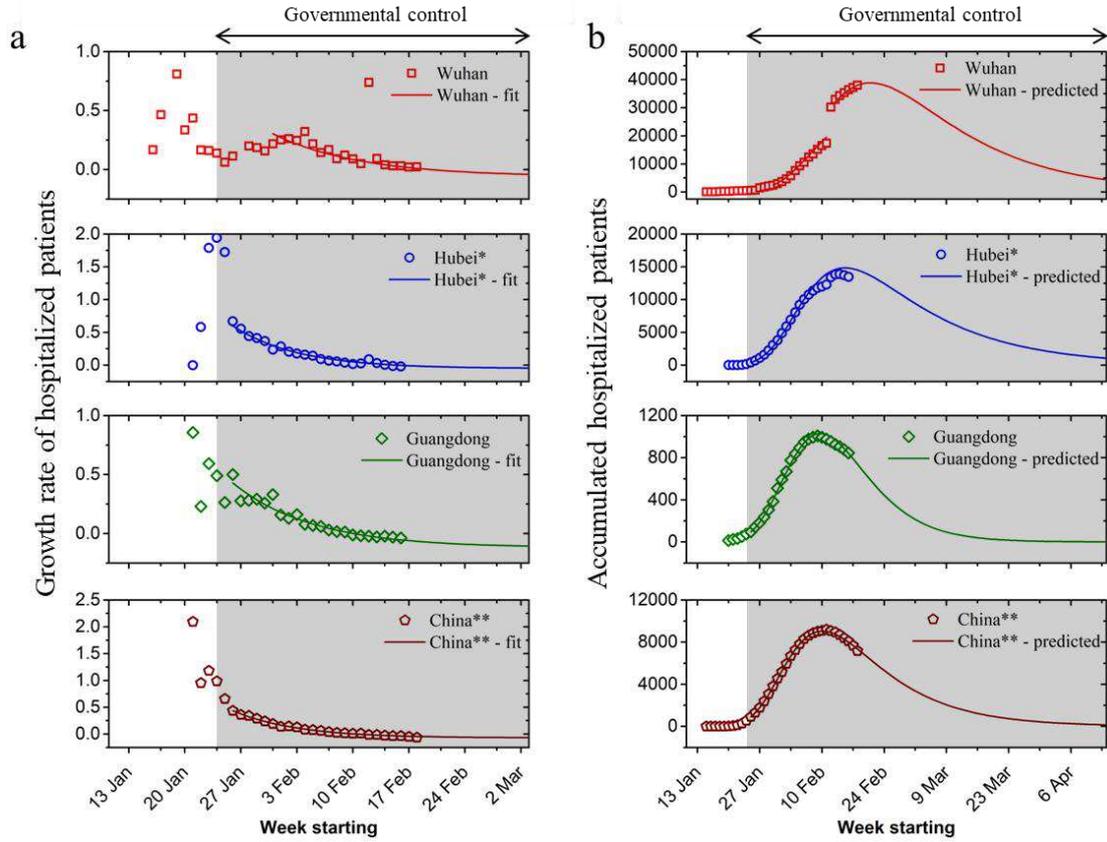
$$K = K_0 + be^{-\gamma t} \quad (5)$$

where  $K$  represents the growth rate of hospitalized patients, the sum of  $K_0$  and  $b$  represents the growth rate at  $t = 0$ , and  $\gamma$  represents the attenuation coefficient of the growth rate. The detail of fitted parameters for formula (5) is available in **Table 2**.

**Table 2.** List of fitted parameters for formula (5).

Parameter	Wuhan	Hubei* province	Guangdong province	China**
$K_0$	-0.050	-0.051	-0.120	-0.071
$b$	1.811	2.520	1.658	2.024
$\gamma$	0.096	1.231	0.101	0.126
$R^2$	0.838	0.973	0.914	0.989

\*: Wuhan excluded; \*\*: Mainland China excluded Hubei province.



**Figure 2.** (a) Daily growth rate, and (b) accumulated hospitalized patients for Wuhan, Hubei province, Guangdong province, and mainland China. Asterisks denote Wuhan or Hubei is excluded (\*: Wuhan excluded; \*\*: Mainland China excluded Hubei province).

Then, we apply iterative methods to predict the number of hospitalized patients:

$$M = M_0 \prod_{t=1}^n (1 + K_0 + be^{-\gamma t}) \quad (6)$$

where  $M_0$  is the number of hospitalized patients at  $t = 0$ . The results are shown in solid lines in **Figure 2b**. Our predictions are in good agreement with existing data. The continuous increase in the number of hospitalized patients during the early phase (**Figure 2b**) indicates that the number of newly infected people is greater than the sum of recoveries and deaths. At this stage, the development and spreading of the epidemic continues. When we reach the maximum value, the number of newly confirmed cases is equal to the sum of recoveries and deaths. When the number of patients in the hospital starts to decrease, it means that the number of newly confirmed patients is less than the sum of recoveries and deaths. When this curve reaches the maximum and begins to turn around, we call this point the inflection point. This point can also be seen in **Figure 2a**, that is, the point at which the growth rate changes from positive to negative. The peak value is the maximum number of beds required by hospitalized patients (**Table 3**). The maximum accumulated hospitalized patients for these four regions will be 38,888 (Wuhan), 14,849 (Hubei except Wuhan), 1,029 (Guangdong) and 9,057 (Mainland China except Hubei). The outbreak in Wuhan overwhelmed the health

systems. The patients have been forced to turn away due to a lack of beds and medical supplies. Accurately estimating the number of beds needed will assist the government in resource allocation and planning during a public health crisis.

**Table 3.** Summaries of predictions using the proposed model.

	Wuhan	Hubei* province	Guangdong province	China**
Inflection point	2020/02/21	2020/02/15	2020/02/10	2020/02/10
Total required beds	38,888	14,849	1,029	9,057
Total cases	56,519	21,093	1,377	13,641
End time	2020/5/5	2020/04/13	2020/03/15	2020/03/27

\*: Wuhan excluded; \*\*: Mainland China Hubei province excluded.

### 3. Conclusions

Several models for the prediction of the COVID-19 outbreak existed, but none could well predict the transmission of the epidemics under the stringent public health intervention. Based on the growth rates  $T$  and  $K$ , we proposed a simple model for studying and predicting the progression of the epidemics. We evaluated the role of government control policies based on this model and predicted the inflection point and end time of the epidemic, the maximum number of hospitalized patients, and the expected number of infected people (**Table 3**). Our predictions are in excellent agreement with the existing data. This model is expected to contribute to resource allocation and planning in the health sector while providing a theoretical foundation for governments to respond to future global health crises or epidemics.

### 4. Methods

Four representative data sets were selected for the analysis: Wuhan, Hubei province (except Wuhan), Guangdong Province, and mainland China (except Hubei). Hubei province (except Wuhan), which surrounds Wuhan city, was selected for the study as it is the hardest-hit area nearest to the epidemic source, Wuhan. Guangdong Province was chosen as it has the largest number of confirmed cases after Hubei Province, with two large floating population and economically developed cities Guangzhou and Shenzhen. Mainland China (except Hubei) reflects the overall development of epidemics across the country. The data were from the official website of the National Health Commission of the People's Republic of China from January 15 to February 18, 2020.

Nonlinear regression and iterative methods were used for predicting the cumulative number and growth rate for confirmed cases and required hospital beds, respectively.

The daily growth rate of confirmed cases is calculated as follows:

$$T = \frac{N_n - N_{n-1}}{N_{n-1}} \quad (1)$$

where  $N_n$  and  $N_{n-1}$  are the number of confirmed cases on day  $n$  and  $n-1$ , respectively. This growth rate curve decays exponentially. The model parameters are obtained by least-square fitting of the nonlinear curve:

$$T = ae^{-\beta t} \quad (2)$$

where  $a$  is a constant that represents the growth rate at  $t = 0$ ,  $\beta$  is an attenuation coefficient, and  $t$  is the time. Based on formula (2), we use iterative methods to predict cumulative confirmed cases:

$$N = N_0 \prod_{t=1}^n (1 + ae^{-\beta t}) \quad (3)$$

where  $N$ ,  $N_0$  is the number of confirmed cases and the expected number of cases at  $t = 0$ , respectively.

The predictions for hospitalized patients were made in a similar way. First, the number of hospitalized patients, or the number of required hospital beds, can be obtained by subtracting the number of confirmed cases by the number of recoveries and deaths as follows:

$$M = M_1 - M_2 - M_3 \quad (4)$$

where  $M$ ,  $M_1$ ,  $M_2$ , and  $M_3$  are the number of hospitalized patients, confirmed patients, recoveries, and deaths, respectively. Substituting formula (1) we obtain the growth rate of hospitalized patients for each region. This growth rate curve also follows an exponentially decaying trend and is obtained by fitting:

$$K = K_0 + be^{-\gamma t} \quad (5)$$

where  $K$  represents the growth rate of hospitalized patients, the sum of  $K_0$  and  $b$  represents the growth rate at  $t = 0$ , and  $\gamma$  represents the attenuation coefficient of the growth rate.

Then, we apply iterative methods to predict the number of hospitalized patients:

$$M = M_0 \prod_{t=1}^n (1 + K_0 + be^{-\gamma t}) \quad (6)$$

where  $M_0$  is the number of hospitalized patients at  $t = 0$ .

### Author Contributions:

X-D. S. initiated the research. X-D. S. and G-Y. C. conceived and designed the model. C. S. and Y. Y. performed the data processing model fitting and drafted the manuscript. All authors contributed to the discussion and revision of the manuscript.

### **Funding:**

This project was supported by the National Key R&D Program of China (2018YFA0902500), the National Natural Science Foundation of China (41706137, 41630979, 41670622), the Natural Science Foundation of Guangdong Province, China (2017A030310332), State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences (Project No. LTO1909), the Natural Science Foundation of SZU (2019078).

### **Conflicts of Interest:**

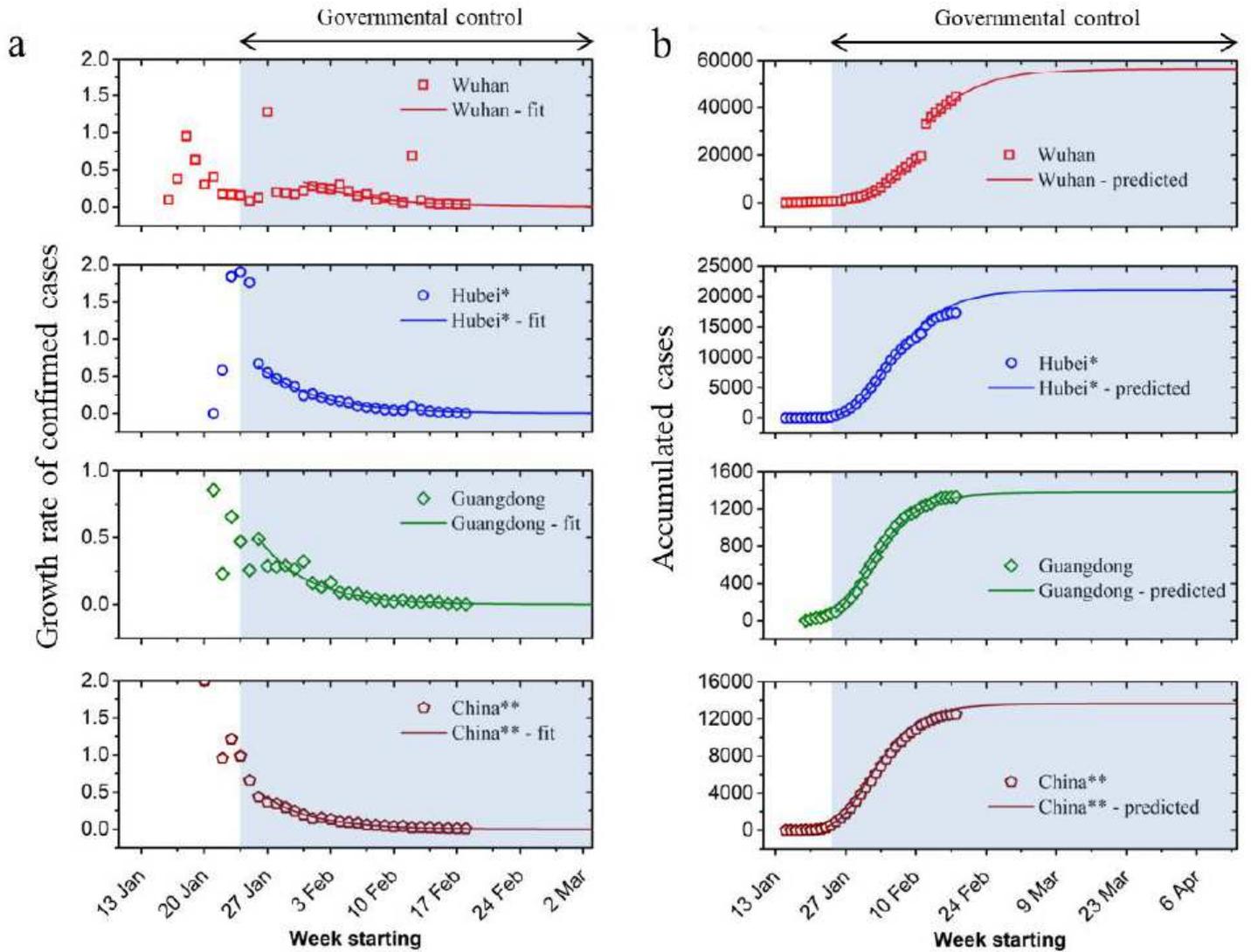
The authors declare no conflict of interest.

### **References:**

- 1 Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *The Lancet* (2020).
- 2 Zhang, N. *et al.* Recent advances in the detection of respiratory virus infection in humans. *Journal of Medical Virology*.
- 3 Chan, J. F.-W. *et al.* A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* (2020).
- 4 Hellewell, J. *et al.* Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health* (2020).
- 5 Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* (2020).
- 6 Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine* (2020).
- 7 Wang, D. *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *Jama* (2020).
- 8 Cohen, J. Chinese researchers reveal draft genome of virus implicated in Wuhan pneumonia outbreak. *Science* (2020).
- 9 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 1-4 (2020).
- 10 Rocklöv, J., Sjödin, H. & Wilder-Smith, A. COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. *Journal of Travel Medicine* (2020).
- 11 Ebrahim, S. H. & Memish, Z. A. Saudi Arabias measures to curb the COVID-19 outbreak: temporary suspension of the Umrah pilgrimage. *Journal of Travel Medicine*, taaa029 (2020).
- 12 Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine* (2020).

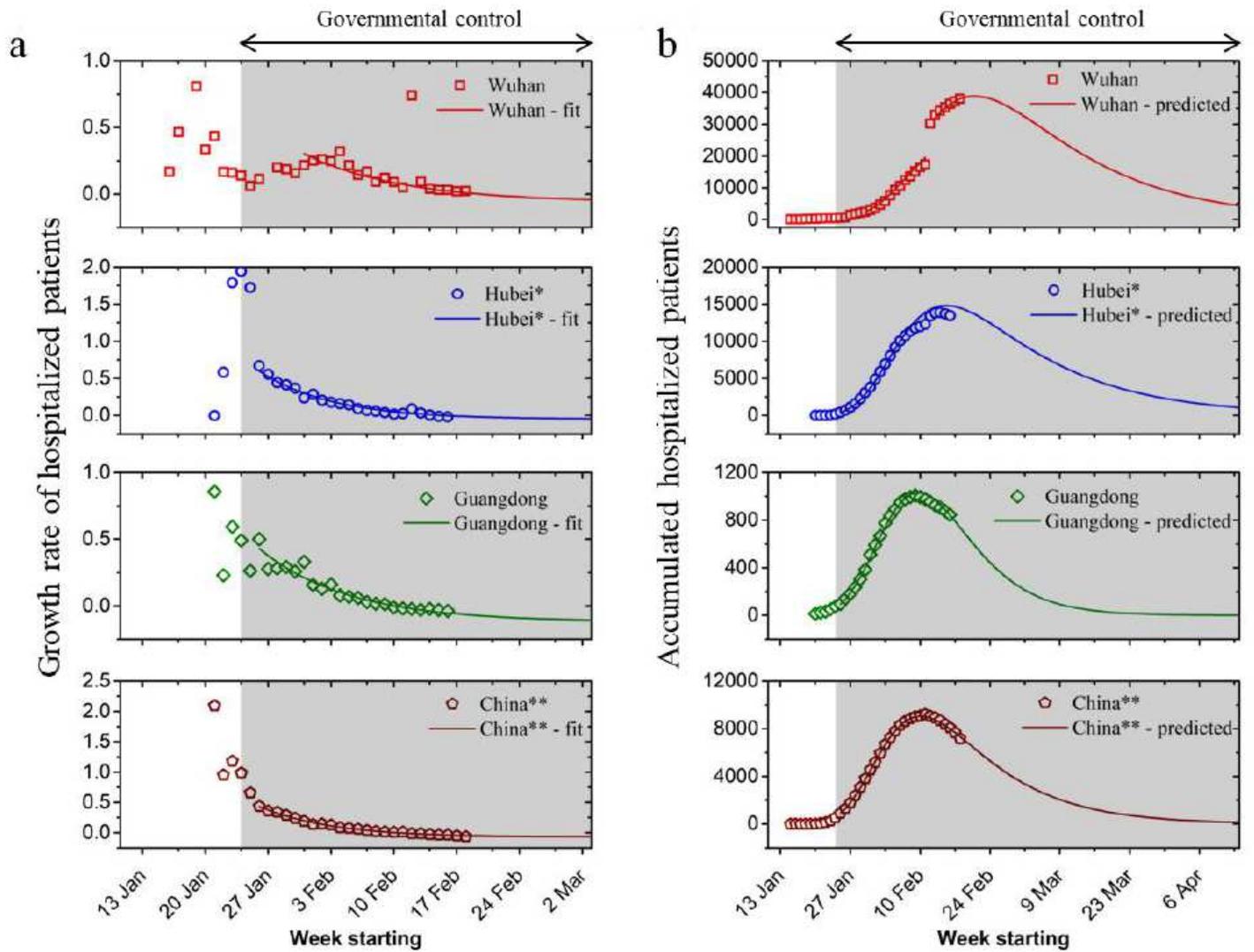
- 13 Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* (2020).
- 14 Chen, J. Pathogenicity and Transmissibility of 2019-nCoV—A Quick Overview and Comparison with Other Emerging Viruses. *Microbes and infection* (2020).
- 15 Imai, N., Dorigatti, I., Cori, A., Riley, S. & Ferguson, N. M. (2020).
- 16 Huang, N. E. & Qiao, F. A data driven time-dependent transmission rate for tracking an epidemic: a case study of 2019-nCoV. *Science Bulletin* (2020).
- 17 Shang, X.-D., Tong, P. & Xia, K.-Q. Scaling of the local convective heat flux in turbulent Rayleigh-Bénard convection. *Physical review letters* **100**, 244503 (2008).
- 18 Shang, X.-D., Liang, C.-R. & Chen, G.-Y. Spatial distribution of turbulent mixing in the upper ocean of the South China Sea. *Ocean Science* **13** (2017).
- 19 Shang, X.-D., Qiu, X.-L., Tong, P. & Xia, K.-Q. Measured local heat transport in turbulent Rayleigh-Bénard convection. *Physical review letters* **90**, 074501 (2003).
- 20 Xie, X., Shang, X., van Haren, H. & Chen, G. Observations of enhanced nonlinear instability in the surface reflection of internal tides. *Geophysical Research Letters* **40**, 1580-1586 (2013).

# Figures



**Figure 1**

(a) Growth rate of confirmed cases, and (b) accumulated cases for Wuhan, Hubei province, Guangdong province, and mainland China. Asterisks denote Wuhan or Hubei is excluded (\*: Wuhan excluded; \*\*: Mainland China excluded Hubei province). Red square for Wuhan, blue circle for Hubei, green diamond for Guangzhou, and brown pentagon for mainland China.



**Figure 2**

(a) Daily growth rate, and (b) accumulated hospitalized patients for Wuhan, Hubei province, Guangdong province, and mainland China. Asterisks denote Wuhan or Hubei is excluded (\*: Wuhan excluded; \*\*: Mainland China excluded Hubei province).