PLOS ONE

RESEARCH ARTICLE

Prediction of the COVID-19 epidemic trends based on SEIR and AI models

Shuo Feng[1☯], **Zebang Feng**[1☯], **Chen Ling**[2☯], **Chen Chang**[3], **Zhongke Feng**[3*]

1 School of Software and Microelectronics, Peking University, Beijing, China, **2** School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, China, **3** Beijing Key Laboratory of Precision Forestry, Beijing Forestry University, Beijing, China

☯ These authors contributed equally to this work.
* fengzhongke@126.com

Citation: Feng S, Feng Z, Ling C, Chang C, Feng Z (2021) Prediction of the COVID-19 epidemic trends based on SEIR and AI models. PLoS ONE 16(1): e0245101. https://doi.org/10.1371/journal.pone.0245101

Editor: Jie Zhang, Newcastle University, UNITED KINGDOM

Received: September 30, 2020

Accepted: December 21, 2020

Published: January 8, 2021

Data Availability Statement: All relevant data are within the manuscript and its Supporting information files.

Funding: This study was supported by: Study on the National Natural Science Foundation of China (U1710123).

Competing interests: The authors have declared that no competing interests exist.

Abstract

In December 2019, the outbreak of a new coronavirus-caused pneumonia (COVID-19) in Wuhan attracted close attention in China and the world. The Chinese government took strong national intervention measures on January 23 to control the spread of the epidemic. We are trying to show the impact of these controls on the spread of the epidemic. We proposed an SEIR(Susceptible-Exposed-Infectious-Removed) model to analyze the epidemic trend in Wuhan and use the AI model to analyze the epidemic trend in non-Wuhan areas. We found that if the closure was lifted, the outbreak in non-Wuhan areas of mainland China would double in size. Our SEIR and AI model was effective in predicting the COVID-19 epidemic peaks and sizes. The epidemic control measures taken by the Chinese government, especially the city closure measures, reduced the scale of the COVID-19 epidemic.

1. Introduction

Since the end of the 20th century, new respiratory infections [1] have emerged in many parts of the world [2]. Among them, the genus β-coronavirus of the coronavirus [3] family poses a continuing threat to human health due to its high transmission efficiency, severe infection consequences, and unpredictable timing of epidemics [4]. Over the past few decades, humans have faced many challenges with viral respiratory infections, including SARS-COV in China in 2002 [5,6], H1N1 [7] in Mexico in 2009 [8] and MERS-COV in Saudi Arabia in 2012 [9,10].

In December 2019, the first case of a 2019 coronavirus patient was found in Wuhan, Hubei Province [11]. On January 23,2020, Chinese government closes off Wuhan [12]. The pathogen was named severe acute respiratory syndrome coronavirus 2 (SARS-cov-2) by the international committee for the classification of viruses on February 11, 2020 [13,14]. The name of the disease caused by SARS-cov-2 is COVID-19. Within two months, COVID-19 had spread [15,16] rapidly from Wuhan to all parts of the country. According to China's national health commission, by March 6, the total number of confirmed cases was 80,653.

Scholars from various countries have attempted to study and analyze the epidemic situation of COVID-19 by various means [17–19]. On January 24, British scholars Read [20] et al. used the SEIR(Susceptible-Exposed-Infectious-Removed) model to predict the trend of the

epidemic. They predicted that the number of infections in Wuhan would reach 190,000 by February 4. This estimate clearly overestimated the trend of the outbreak. On January 27, Biao Tang [21] et al. used epidemic data from January 10 to January 22 to predict the epidemic regeneration coefficient of 6.47 (95% confidence interval 5.71–7.23) by the SEIR(Susceptible-Exposed-Infectious-Removed) model and statistical calculation method. Their model estimated that the number of infections would peak on approximately March 10. In early February, Norden E. Huang [22] et al. proposed a simple data-driven model based on natural growth, predicting that the number of infections would peak on approximately February 5, with a cumulative number of confirmed cases between 37,000 and 44,000. On February 24, 2020, Huwen Wang et al. [23] used SEIR(Susceptible-Exposed-Infectious-Removed) model and proposed that the infection coefficient R0 decreased from 2.5 to 0.5 with virus variation and government policy, which had an impact on the prediction. On 19 March 2020, Joseph T. Wu et al. [24] used SEIR(Susceptible-Exposed-Infectious-Removed) model and they study the influence of infection rate, removal rate in different age groups and migration data on prediction model. On March 25, 2020, Moritz U.G.Kraemer et al. [25] used GLM (generalized linear models) model to consider the impact of population migration and age on the number of infected people.

Many scholars began to pay attention to the impact of population migration on the epidemic, trying to find ways to mitigate the spread of the epidemic. On February 28, Zhong Nanshan et al. [26] used the SEIR model and LSTM model of population migration in Wuhan to predict the trend of the epidemic. The model focused on the impact of population migration in Wuhan on epidemic trends. On April 29, Jayson S.jia et al. [27] used a gravity model to prove that the population migration in Wuhan was closely related to epidemic trends in non-Wuhan regions. On May 4, Shengjie Lai et al. [28] used the SEIR model to predict the trend of the epidemic. their findings suggest that the COVID-19 cases would likely have shown a 67-fold increase (interquartile range 44–94) by February 29 without NPIs (non-pharmaceutical interventions). On June 28, Solomon Hsiang et al. [29] used the SIR model and linear regression model to predict the trend of the epidemic. Their models suggest that if the Chinese government had not taken effective measures, the number of infected people in China would have increased 465 times by March 22.

In the early stage of the epidemic, due to the lack of sufficient data, it was difficult for scholars to accurately predict the trend of the epidemic [20–23]. Additionally, a number of recent studies have shown that the development of the epidemic is closely related to population movement [24–26]. Therefore, it is still necessary to study the epidemic trend of COVID-19 at the current stage, which is of practical significance for the analysis, prevention and control of the epidemic [30]. According to the characteristics of the Wuhan and non-Wuhan regions, the SEIR model and deep learning model were established, respectively, considering the population flow. By using the actual data and referring to the existing literature and reports, the parameters of the new model were fitted. Finally, the model was used to estimate and analyze the epidemic trend.

2. Materials and methods

2.1. Data sources

The epidemic data used in this paper were from the latest epidemiological data of COVID-19 reported by the Ding Xiang Yuan [31]. The urban migration index was derived from the Baidu migration project [32], which is based on the users of Baidu and related products to count and calculate the daily number of pedestrian movements in and out of cities by railway, air and highway. The population density, per capita GDP and other urban data of all provinces in

China were obtained from the official website of the National Bureau of Statistics [33]. In this paper, the distance from each province to Wuhan was obtained from the Ovi interactive map [34]. The average temperature of each province in 2019 was from the China Meteorological Administration [35].

2.2. SEIR model

In this section, we briefly discuss the properties of the basic Susceptible-Exposed-Infected-Removed (SEIR) model. The model divided people into four categories: Susceptible (S): individuals not yet infected; Exposed (E): individuals experiencing incubation duration; Infectious (I): confirmed cases; Removed (R): recovered and dead individuals.

Fig 1 shows how individuals move through each compartment in the mode.

The parameters within this model are as follows:

1. Contact rate β controls the rate of spread, which represents the probability of transmitting disease between a susceptible and an infectious individual.
2. Incubation rate σ is the rate of latent individuals becoming infectious.
3. Recovery rate γ is the rate of infected individuals becoming recovered.

The transmission of the virus is then described by the following system of nonlinear ordinary differential equations:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{1-1}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E \tag{1-2}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{1-3}$$

$$\frac{dR}{dt} = \gamma I \tag{1-4}$$

2.3. DNN model

Deep neural networks (DNN) is composed of many parallel and highly correlated computing and processing units, which are similar to the neurons of biological nervous system. Although the structure of a single neuron is simple, the behavior of the neuron system composed of a large number of interconnected neurons is very rich. The neural network has the ability of parallel computing and adaptive learning. Compared with the traditional prediction methods, the

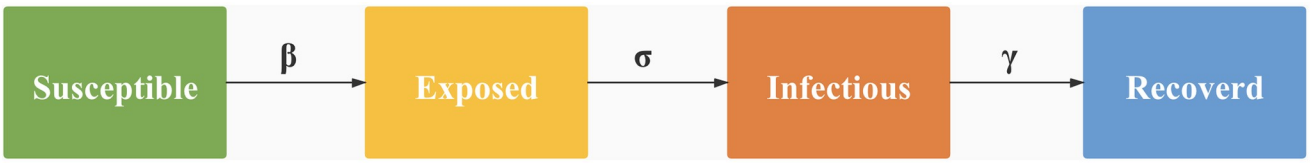

Fig 1. SEIR model with 4 states.

https://doi.org/10.1371/journal.pone.0245101.g001

prediction accuracy of neural network is better. This is mainly due to the characteristics of neural network. Neural network is good at describing the characteristics of complex system with strong nonlinearity and difficult to be expressed by precise mathematical model, and has adaptive ability.

In this section, we briefly discuss the properties of the DNN model. The DNN model first trained a large quantity of data to reduce the loss function and then calculated and updated the parameters in the network to achieve the prediction of new data.

The following equations describe the linear relationship between input and output:

$$Y = \sum_{i=1}^{m} wixi + b \tag{2}$$

In the equations, *xi* represents the input data, *wi* represents the weight parameter, and b represents the bias.

The DNN model in this study consists of a 9-node input layer, a 32-node hidden layer and a 1-node output layer. Tanh was used as the activation function to increase the fitting degree of the model to the nonlinear model. Additionally, dropout was used in this study to alleviate the overfitting phenomenon of the model. The formula of the Tanh activation function is as follows:

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{2x}} \tag{3}$$

The DNN model architecture used in this study is shown in Fig 2.

The loss function can quantitatively determine the quality of the model to select the optimal model. In this study, the mean square error loss function was used. The formula is as follows:

$$SE(y\_, y) = \frac{\sum_{i=1}^{n} (y - y\_)^2}{n} \tag{4}$$

According to the current state for the loss function, we need to update the weight in the direction of the minimum loss to obtain the optimal model. This study uses the Adam (Adaptive Moment Estimation) optimizer to update the weight. Adam is an adaptive learning rate optimization algorithm. By computing the first-moment estimate and the second raw moment estimate of a gradient, an independent adaptive learning rate is designed for different parameters. The calculation formula is as follows:

Update biased first-moment estimate

$$m_{t,i} = \beta_1 \cdot m_{t-1,i} + (1 - \beta_1) \cdot g_{t,i} \tag{5}$$

Update biased second raw moment estimate

$$V_{t,i} = \beta_2 \cdot V_{t-1,i} + (1 - \beta_2) \cdot g_{t,i}^2 \tag{6}$$

Update parameters

$$w_{t+1} = w_t - \eta_t = w_t - lr \cdot \frac{m_{t,i}}{1 - \beta_1{}^t} / \sqrt{\frac{V_{t,i}}{1 - \beta_2{}^t}} \tag{7}$$

2.4. RNN model

Recurrent neural network (RNN) is a kind of neural network for modeling sequence data. It is like a recurrent dynamic system. In this structure, the current output will flow into the next

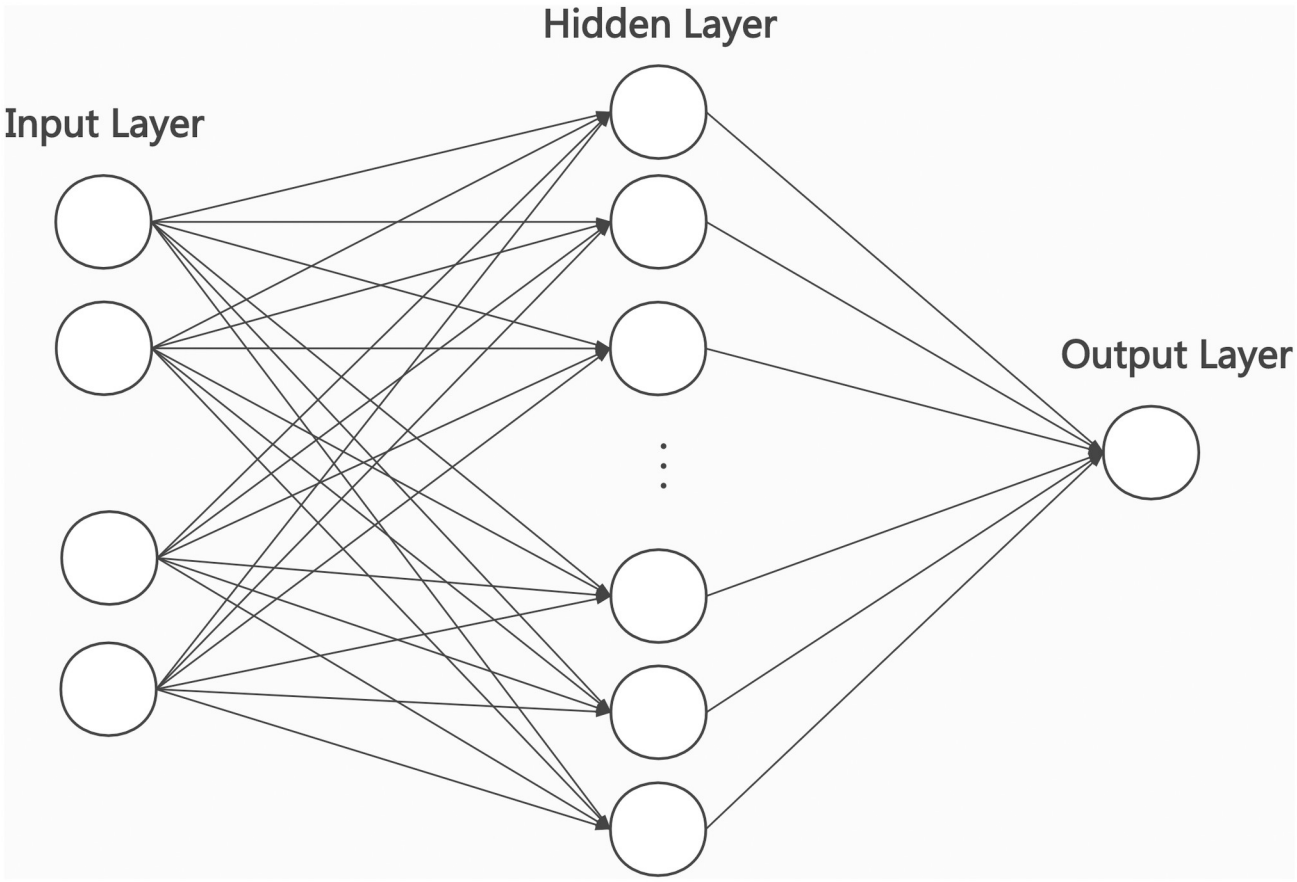

Fig 2. DNN model architecture. The model consists of a 9-node input layer, a 32-node hidden layer and a 1-node output layer.

https://doi.org/10.1371/journal.pone.0245101.g002

input and contribute to the next output. Compared with other neural networks, RNN has some advantages. Fixed input and output is a limitation of general neural networks. The RNN performs well in this aspect, and the sequence can be the input and output of the structure.

In this section, we discuss the use of the RNN is used to predict the number of COVID-19 infections. RNN computes the feature extraction of time-series-based samples by recursion of input sequence data in the direction of sequence evolution.

The RNN consists of input layers, hidden layers, and output layers. The RNN model is shown in Fig 3, and the expansion of Fig 3 is shown in Fig 4.

The calculation formula is as follows:

$$O_t = g(v \cdot S_t) \tag{8}$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1}) \tag{9}$$

f and g are activation functions.

In this study, two layers of the RNN were built to extract the deep features of data, and one layer of the DNN was used to output the results. In both RNN layers, the tanh activation function and dropout were used to increase the fitting degree of the nonlinear model to the data and work for the model overfitting. The complete architecture of our RNN prediction research model is shown in Fig 5:

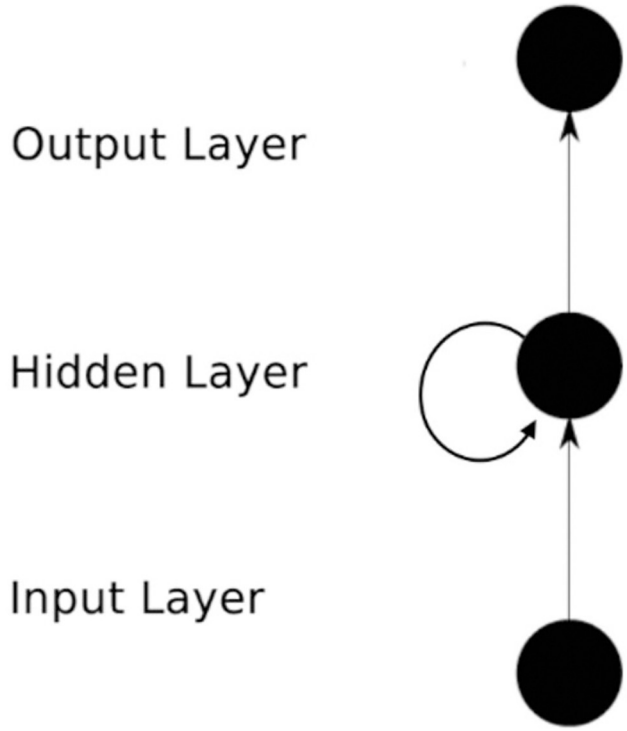

Fig 3. RNN model architecture. The RNN consists of an input layer, a hidden layer, and an output layer.

https://doi.org/10.1371/journal.pone.0245101.g003

3. Results

3.1. Prediction of Wuhan infection trend based on the SEIR model

According to the national health commission, by the end of March 6, the total number of confirmed cases was 80,653 [31]. Among them, 49,871 cases were confirmed in Wuhan, accounting for 60% of the total confirmed cases. The Chinese joint investigation report on novel coronavirus pneumonia (COVID—19) [36] shows that the spread of the epidemic in China has distinct characteristics in the Wuhan and non-Wuhan areas of China. The newly infected

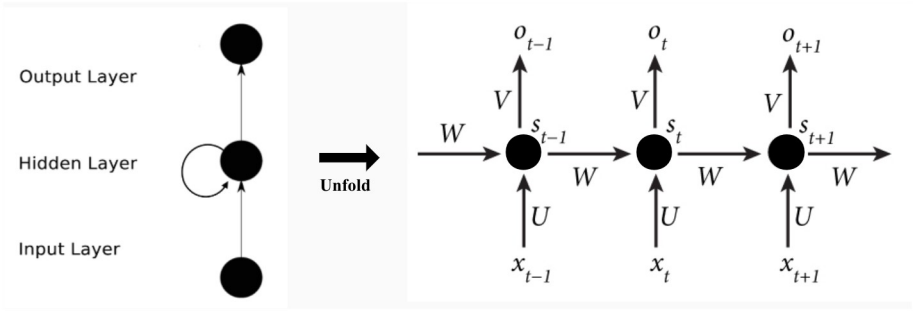

Fig 4. Expansion of the RNN. $t-1$, t, *and* $t+1$ are time series. The input units are $\{x_0, x_1, \ldots, x_t, x_{t+1}, \ldots\}$, and the output units are $\{o_0, o_1, \ldots, o_t, o_{t+1}, \ldots\}$. The hidden units are $\{s_0, s_1, \ldots, s_t, s_{t+1}, \ldots\}$ There is a unidirectional flow of information from the input units to the hidden units and another unidirectional flow of information from the hidden units to the output units. W is the weight of the input, U is the weight of the input units at the moment, and V is the weight of the output units.

https://doi.org/10.1371/journal.pone.0245101.g004

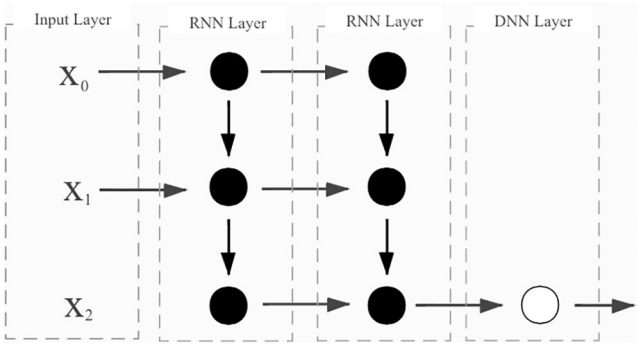

Fig 5. The architecture of the RNN. The number of hidden units weight in the first RNN layer was 100, the number of hidden units weight in the second RNN layer was 50 and the number of node in the DNN layer was 1.

https://doi.org/10.1371/journal.pone.0245101.g005

cases in Wuhan were mainly original cases in Wuhan. As Wuhan is the traffic center of China and the epidemic spread rapidly during the Spring Festival travel rush, the newly infected cases in non-Wuhan areas were mainly imported from Wuhan. Since the newly added cases in Wuhan were mainly infected by original cases, the SEIR model was used to predict the epidemic trends in Wuhan.

To use the SEIR model, the contact rate β, incubation rate σ, recovery rate γ and other parameters needed to be estimated. The initial value of the susceptible population in Wuhan city was similar to that of the permanent resident population in Wuhan city. Because the incubation period of COVID-19 has been reported to be between 2 and 14 days, we chose the midpoint of 7 days. We used a recovery rate of 3% [37]. In the early stage of the outbreak, the number of infected people was small. The susceptible population S in the first 10 days in Wuhan was approximately the same as the population N on the same day in Wuhan. Therefore, S≈N. We obtain the new formula:

$$\frac{dI}{dt} = \beta\frac{IS}{N} - \gamma I \approx (\beta - \gamma)I \tag{10}$$

Finally, it is simplified to:

$$I(t) = e^{(\beta-\gamma)I} \tag{11}$$

Based on the actual number of people infected in the first 10 days in Wuhan, this study estimated the initial β value to be 0.17.

Starting on February 12, the Hubei government changed the way it counted new confirmed cases. The number of new official diagnoses rose sharply as clinically diagnosed cases were included in new cases. Considering the change in the statistical method, this study corrected the initial parameters of the model appropriately to reduce the prediction error.

Based on these estimated parameters and the epidemiological data of Wuhan, the model parameters were fitted and optimized. The number of predicted results is shown in Fig 6, and the results are consistent with the actual situation. In the statistical results, the SEIR model well predicted the infection value of the epidemic situation, but the prediction of the deleted value of the epidemic situation was delayed. This may have something to do with inadequate funeral facilities in the country at the beginning of the epidemic.

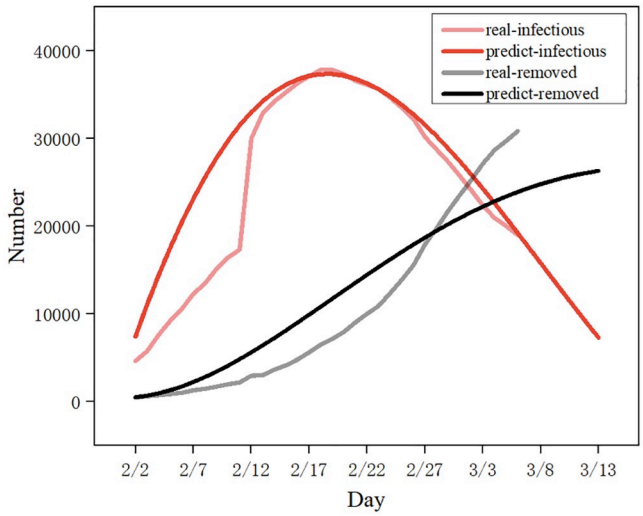

Fig 6. The SEIR model predicted the cumulative number of infections in Wuhan. Data from January 23 to March 3 were used to predict the cumulative number of infections in Wuhan over the next seven days.

https://doi.org/10.1371/journal.pone.0245101.g006

3.2. Prediction of the non-Wuhan infection trend based on the DNN/RNN model

Since the new infection cases in non Wuhan area are affected by so many factors, and Neural network is good at describing the characteristics of complex system with strong nonlinearity and difficult to be expressed by precise mathematical model, and has adaptive ability. Neural network models were used to predict the epidemic trends in non-Wuhan. The number of new infections in non-Wuhan areas has a distinct character. The number of newly infected individuals in different provinces is shown in Fig 7:

The number of newly infected people in different provinces shows two main characteristics of the epidemic in non-Wuhan areas:

1. Wuhan is the first place in China to discover COVID-19 early, and is the capital of Hubei province. The neighboring provinces of Hubei, including Hunan, Sichuan and Anhui, were seriously infected. On January 23, there were 24 new cases in Hunan, 15 new cases in Sichuan and 15 new cases in Anhui.
2. China's supercities were the main focus of the epidemic: by February 2, in addition to the surrounding areas of Hubei, the newly infected cities in Beijing, Shanghai and Guangzhou were also serious. There were 21 new cases in Beijing, 21 new cases in Shanghai and 15 cases in Guangzhou. By February 26, as companies returned to work and provincial workers returned to central cities, new cases were reported in Beijing and Shanghai. There were 10 new cases in Beijing and 1 new cases in Shanghai.

Because of the first point, Wuhan was the radioactive center of the epidemic, and its neighboring provinces are severely infected. Therefore, this paper examined the distance from the target city to Wuhan, the number of existing cases in Wuhan, and the number of newly diagnosed Wuhan as factors. The second point is that since the epidemic situation in supercities were serious, this paper used GDP and population density in target city to characterize whether it was a supercity. In addition, in view of studies have shown that climate can affect the spread of sars-cov-2 [39], so this paper took the average temperature and rainfall

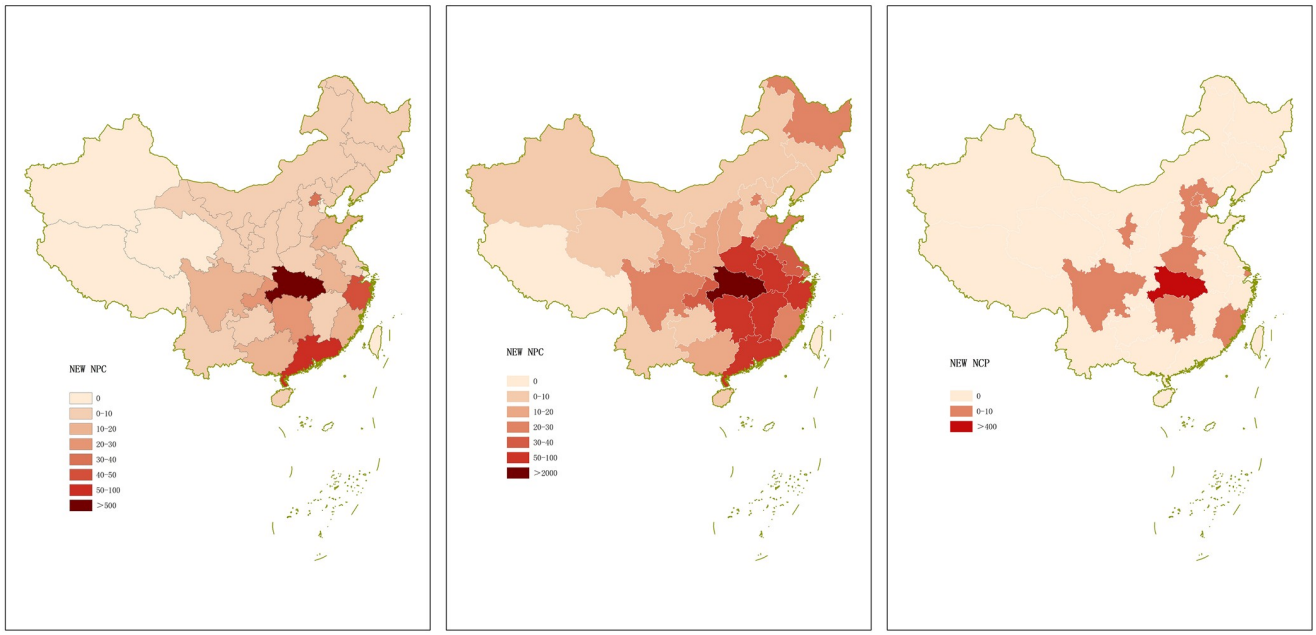

Fig 7. This figure shows the number of new infections in each province [38]: (a) The number of new infections in each province on January 23; (b) The number of new infections in each province on February 2; (c) The number of new infections in each province on February 26.

https://doi.org/10.1371/journal.pone.0245101.g007

mentioned in the paper as factors. Besides, the current situation of the epidemic has a great influence on the future development in the near future. So this paper used the total number of confirmed cases and the number of newly confirmed cases as consideration factors.

Because compared with traditional prediction methods, neural networks have higher prediction accuracy. So, this paper used the DNN and RNN model to predict the number of infected people in non-Wuhan areas of China.

DNN model needs training the weight parameter w and the bias parameter *b*r, the model used following data as input: the total number of confirmed cases in the previous day, the average number of newly confirmed cases in the past 3 days, the number of existing cases in Wuhan, local population density, per capital GDP, distance to Wuhan, average annual temperature, and average annual rainfall the average of the migration population in Wuhan in the past 5 days.

The RNN model used the values of these nine datasets over three consecutive days. Specific data are shown in Table 1:

First, we normalize the data by Min-Max Normalization except the number of epidemic cases. In (12), where X_{max} is the maximum value of the sample data, and X_{min} is the minimum value of the sample data. In additional, we set an appropriate scale to scale the number of cases to match the range of 0 to 1.

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{12}$$

Table 1. Data used by model DNN/RNN.

epidemic data		city data					migration data
confirmed cases	past 3 days cases	local population density	per capital GDP	distance to Wuhan	average annual temperature	average annual rainfall	migration population in Wuhan

https://doi.org/10.1371/journal.pone.0245101.t001

And then the normalized number of epidemic cases X were sent to the model for training, and the loss was calculated by (4). And update the parameters w and b. After 10000 epochs training, loss tends to be stable.

After training, we got the output. The output of the network is the new cases that were scaled down locally on that day. We restored the number of cases by scaling the previous cases. Then we got the result.

Starting February 12, our model continuously predicted new infections in each province over the next three days. On March 6, we summarized the previous data and selected Beijing and Henan provinces with strong representation. The prediction effect of this DNN/RNN model is shown in Fig 8. It can be seen that the AI model with the population migration data can well predict the epidemic situation in non-Wuhan areas.

4. Discussion

Due to the outbreak, the Hubei provincial government launched a level-1 response to the public health emergency on January 23, 2020. The city's buses, subways, ferries, coaches, airports and train stations were suspended. At the same time, all parts of the country also closed the city. we used our model to predict the impact of closure of Wuhan on the the epidemic trends of COVID-19 in China.

To explore the influence of Wuhan sealing on epidemic trends, this study attempted to simulate and predict the epidemic trends of open cities in all provinces. We used the data of Wuhan migration in 2019 to replace the data of Wuhan migration in 2020 to predict the number of people infected in non-Wuhan areas. We selected Beijing and Henan provinces with strong representativeness, and the prediction effect of the DNN model is shown in Fig 9.

Based on the model in this study, it is estimated that if Wuhan did not adopt city closure measures, for the vast majority of provinces such as Beijing, Chongqing and Guangdong, the cumulative number of increased infections in each province would have increased to about 1.5 times within 4 weeks. This shows that the closure of the city had a great effect on inhibiting the further spread of the disease.

Many scholars have come to similar conclusions. Shengjie Lai et al. [28] used the SEIR model to predict the trend of the epidemic. their findings suggest that the COVID-19 cases would likely have shown a 67-fold increase (interquartile range 44–94) by February 29 without NPIs (non-pharmaceutical interventions). Without NPIs, the cumulative number of increased infections would have increased to 18 times within 3 weeks. In their model prediction, NPI contains more than just city closure measures, which make their predictions bigger than ours. Solomon Hsiang et al. [29] used the SIR model and linear regression model to predict the trend of the epidemic. Their models suggest that if the Chinese government had not taken effective measures, the number of infected people in China would have increased 465 times by March 22. To be specific, if China had not implemented policies, the number of infected people would have increased exponentially. The early infection rate in China was 0.31. After China used three policies (1. Emergency declaration 2. Travel ban 3. home isolation), the rate of infection decreased by 0.252. Among them, the travel ban reduced the infection rate by 0.05, the cumulative number of increased infections would have increased to 2.26 times within 3 weeks. Their conclusions are nearly consistent with ours.

Additionally, we can see that the cumulative increase in the number of infections in different provinces did not occur in the early stages of the epidemic. Most provinces experienced a dramatic increase in the number of people infected within 3–7 days without closure. We believe this is due to the incubation period lasting for 2–14 days and the population migration in Wuhan was closely related to epidemic trends in non-Wuhan regions. If the city did not

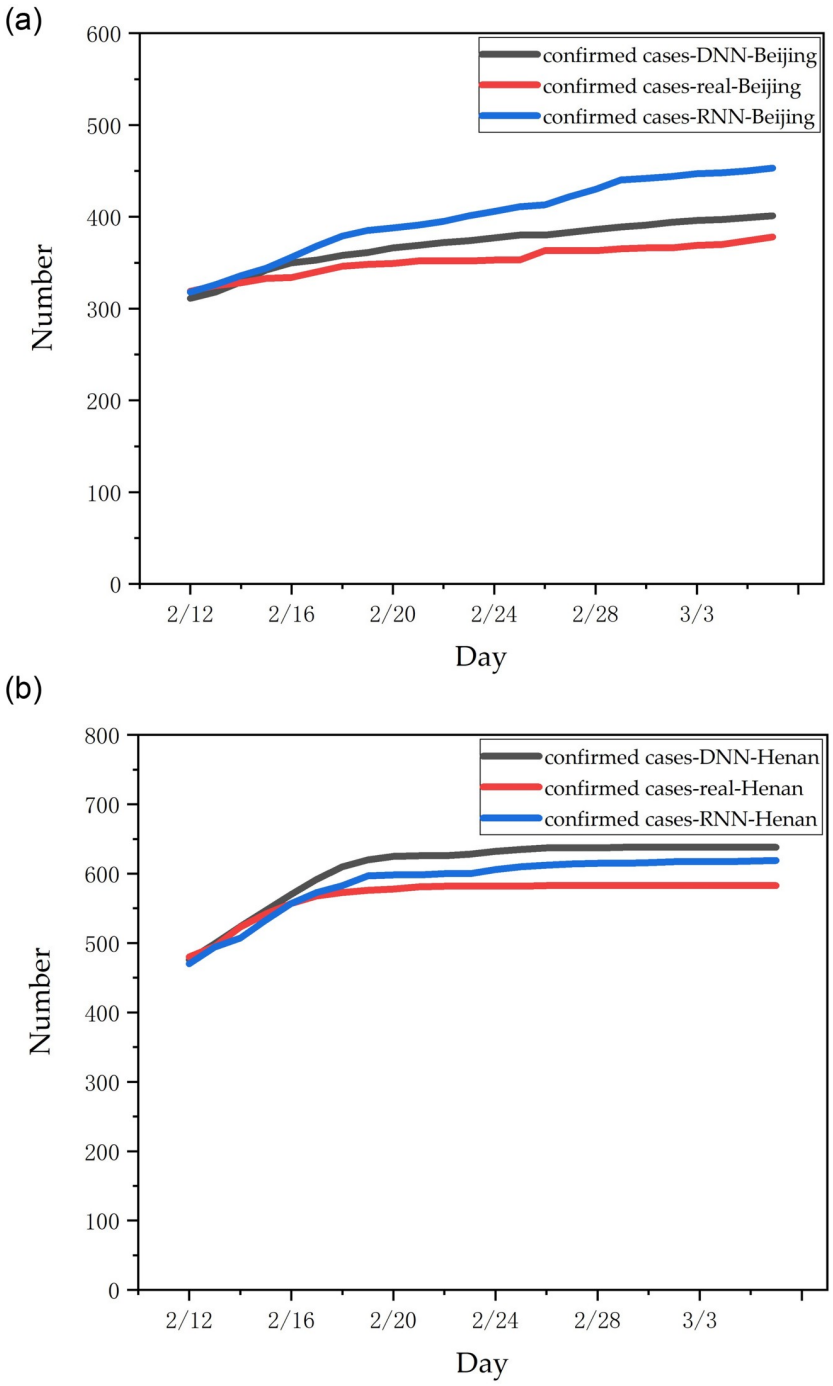

Fig 8. The number of newly confirmed cases predicted by the RNN/DNN model in each province: (a) results in Beijing; (b) results in Henan.

https://doi.org/10.1371/journal.pone.0245101.g008

take measures to close the city, the new sources of infection in the city were mainly due to the latent period of the Wuhan migration population, and the outbreak was delayed by 3–7 days. Jayson S.jia et al. [27] also proved our point. They used a gravity model to prove that the population migration in Wuhan was closely related to epidemic trends in non-Wuhan regions. The

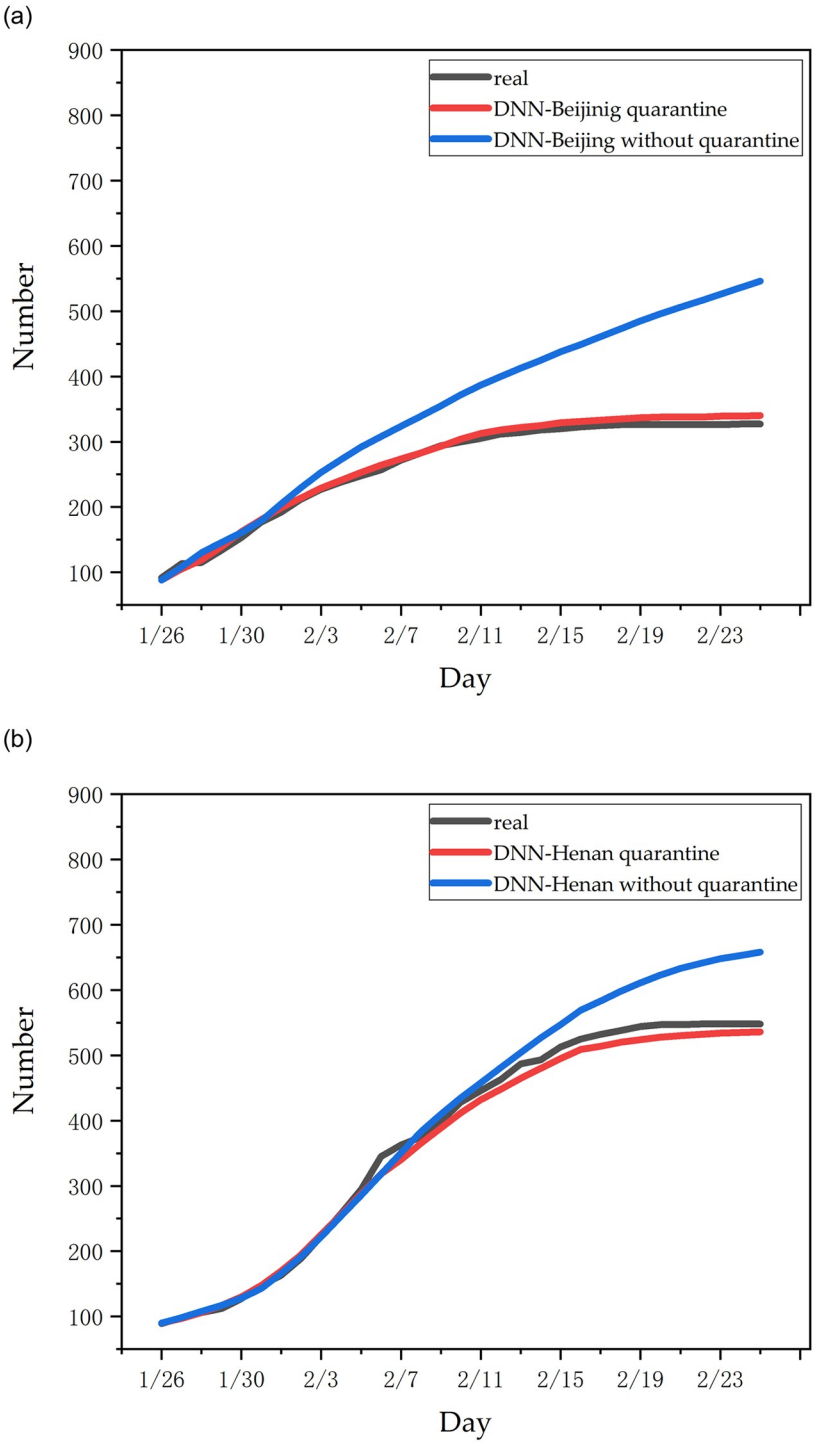

Fig 9. The DNN model was used to predict the total number of confirmed cases in provinces under the condition of city quarantine or not quarantine: (a) results in Beijing; (b) results in Henan.

https://doi.org/10.1371/journal.pone.0245101.g009

correlation coefficient between the cumulative number of increased infections and the population migration in Wuhan increased from 0.522 on January 1 to 0.919 on January 24.

However, Henan Province and some other provinces with a smaller population did not conform to the above rule. In this study, it is believed that before Wuhan's closure on January

23, Henan Province isolated the migrant population from Wuhan in a timely manner, so the impact of the Wuhan migrant population on the cumulative number of confirmed patients in Henan Province was small.

Our study has some limitations. Firstly, we built the model according to the conventional infection model, without considering the parameter fluctuation caused by the possible super disseminator and virus variation in the SEIR model. At the same time, our model does not account for differences in infection rates between the recovered and the general population.

Secondly, we use multiple data including epidemic data, urban data, migration data to predict the epidemic trends of COVID-19 in other parts of China, without considering the potential impact of other factors on COVID-19 in the DNN and RNN model.

Thirdly, in our models, our original parameters are based on previous studies and experience from SARS control. Besides, the data we using is based on the data before March 3, 2020. With the progression of COVID-19, the model parameters will change greatly because of more and more data.

5. Conclusions

In summary, this paper collected the COVID-19 data including the number of confirmed, cured and deaths from January 23 to March 6, 2020, combined with Baidu population migration data and relevant city data of the National Bureau of Statistics, to predict the number of infections in China. The SEIR model could well predict the epidemic situation in Wuhan, which was dominated by primary cases, and the AI model which added population migration data could well predict the epidemic situation in non-Wuhan areas in China with a large number of input infections. Additionally, this study estimated the influence of Wuhan closure on the epidemic trend. The results showed that the closure of Wuhan was an important measure to effectively inhibit the spread of COVID-19 in a large area, and it greatly reduced the number of infections in every part of China.

Supporting information

S1 File. City data.
(XLSX)

S2 File. Confirmed cases.
(XLSX)

S3 File. Migration algebra in Wuhan.
(XLSX)

S4 File. Migration population in Wuhan.
(XLSX)

Author Contributions

Conceptualization: Zhongke Feng.

Data curation: Zebang Feng, Chen Ling, Zhongke Feng.

Formal analysis: Zebang Feng, Chen Ling, Zhongke Feng.

Investigation: Shuo Feng, Zebang Feng, Chen Chang.

Methodology: Shuo Feng, Zebang Feng, Chen Chang, Zhongke Feng.

Project administration: Shuo Feng.

Software: Shuo Feng, Chen Ling.

Supervision: Zhongke Feng.

Validation: Chen Ling.

Visualization: Chen Chang.

Writing – original draft: Shuo Feng, Zebang Feng.

Writing – review & editing: Zhongke Feng.

References

1. Braciale T., Sun J. & Kim T. Regulating the adaptive immune response to respiratory virus infection. Nat Rev Immunol 12, 295–305 (2012). https://doi.org/10.1038/nri3166 PMID: 22402670
2. Weiss R., McMichael A. Social and environmental risk factors in the emergence of infectious diseases. Nat Med 10, S70–S76 (2004). https://doi.org/10.1038/nm1150 PMID: 15577934
3. Falzarano D., de Wit E., Martellaro C. et al. Inhibition of novel β coronavirus replication by a combination of interferon-α2b and ribavirin. Sci Rep 3, 1686 (2013). https://doi.org/10.1038/srep01686 PMID: 23594967
4. Xiong C., Jiang L., Jiang Q. Prevalence and control of human diseases caused by beta coronavirus (β-CoVs). Shanghai Journal of Preventive Medicine 32, 58–66(2020).
5. Zhou P., Yang X., Wang X. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273 (2020). https://doi.org/10.1038/s41586-020-2012-7 PMID: 32015507
6. Ou X., Liu Y., Lei X. et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun 11, 1620 (2020). https://doi.org/10.1038/s41467-020-15562-9 PMID: 32221306
7. Tapia R., Torremorell M., Culhane M. et al. Antigenic characterization of novel H1 influenza A viruses in swine. Sci Rep 10, 4510 (2020). https://doi.org/10.1038/s41598-020-61315-5 PMID: 32161289
8. Liu S., Jiao X., Wang S. et al. Susceptibility of influenza A(H1N1)/pdm2009, seasonal A(H3N2) and B viruses to Oseltamivir in Guangdong, China between 2009 and 2014. Sci Rep 7, 8488 (2017). https://doi.org/10.1038/s41598-017-08282-6 PMID: 28814737
9. Park Y., Walls A.C., Wang Z. et al. Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. Nat Struct Mol Biol 26, 1151–1157 (2019). https://doi.org/10.1038/s41594-019-0334-7 PMID: 31792450
10. Yang C.H., Jung H. Topological dynamics of the 2015 South Korea MERS-CoV spread-on-contact networks. Sci Rep 10, 4327 (2020). https://doi.org/10.1038/s41598-020-61133-9 PMID: 32152361
11. Wu F., Zhao S., Yu B. et al. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269 (2020). https://doi.org/10.1038/s41586-020-2008-3 PMID: 32015508
12. Coronavirus: the first three months as it happened. https://www.nature.com/articles/d41586-020-00154-w (2020).
13. Gorbalenya A.E., Baker S.C., Baric R.S. et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5, 536–544 (2020). https://doi.org/10.1038/s41564-020-0695-z PMID: 32123347
14. Peng X., Xu X., Li Y. et al. Transmission routes of 2019-nCoV and controls in dental practice. Int J Oral Sci 12, 9 (2020). https://doi.org/10.1038/s41368-020-0075-9 PMID: 32127517
15. Yang P., Wang X. COVID-19: a new challenge for human beings. Cell Mol Immunol (2020). https://doi.org/10.1038/s41423-020-0407-x PMID: 32235915
16. Danese S., Cecconi M. & Spinelli A. Management of IBD during the COVID-19 outbreak: resetting clinical priorities. Nat Rev Gastroenterol Hepatol (2020).
17. Gog J.R. How you can help with COVID-19 modelling. Nat Rev Phys (2020).
18. Prevent and predict. Nat Ecol Evol 4, 283 (2020). https://doi.org/10.1038/s41559-020-1150-5 PMID: 32080369
19. Wu J., Leung K., & Leung G., Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in Wuhan, China: a modelling study. The Lancet 395,10225(2020). https://doi.org/10.1016/S0140-6736(20)30260-9 PMID: 32014114

20. Read J. et al. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. Preprint at https://www.medrxiv.org/CONTENT/10.1101/2020.01.23.20018549V2 (2020).
21. Tang B. et al. Estimation of the transmission risk of 2019-nCov and its implication for public health interventions. J Clin Med 9, 462 (2020). https://doi.org/10.3390/jcm9020462 PMID: 32046137
22. Huang N., Qiao F. A data driven time- dependent transmission rate for tracking an epidemic: A case study of 2019-nCoV. Sci Bull (Beijing) 65, 425–427 (2020). https://doi.org/10.1016/j.scib.2020.02.005 PMID: 32288968
23. Wang H., Wang Z., Dong Y. et al. Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China. Cell Discov 6, 10 (2020). https://doi.org/10.1038/s41421-020-0148-0 PMID: 32133152
24. Wu J.T., Leung K., Bushman M. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med 26, 506–510 (2020). https://doi.org/10.1038/s41591-020-0822-7 PMID: 32284616
25. Kraemer M., Yang C. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 368, 493–497 (2020). https://doi.org/10.1126/science.abb4218 PMID: 32213647
26. Yang Z. et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis 12, 165–174 (2020). https://doi.org/10.21037/jtd.2020.02.64 PMID: 32274081
27. Jia J.S., Lu X., Yuan Y. et al. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature 582, 389–394 (2020). https://doi.org/10.1038/s41586-020-2284-y PMID: 32349120
28. Lai S., Ruktanonchai N.W., Zhou L. et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature (2020). https://doi.org/10.1038/s41586-020-2293-x PMID: 32365354
29. Hsiang S., Allen D., Annan-Phan S. et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature (2020).
30. Cuong L., et al. The first Vietnamese case of COVID-19 acquired from China. Lancet Infect Dis 20, 408–409 (2020). https://doi.org/10.1016/S1473-3099(20)30111-0 PMID: 32085849
31. Ding Xiang Yuan (in Chinese). http://www.dxy.cn (2020).
32. Baidu qianxi (in Chinese) https://qianxi.baidu.com (2020).
33. National Bureau of Statistics of China. http://data.stats.gov.cn (2019).
34. Ovi interactive map. http://www.gpsov.com/cn/main.php (2020).
35. China Meteorological Administration. http://www.cma.gov.cn/(2019).
36. China—WHO novel coronavirus pneumonia (COVID-19) joint investigation report. http://www.nhc.gov.cn/jkj/s3578/202002/87fd92510d094e4b9bad597608f5cc2c.shtml (2020).
37. Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. J Med Virol 92, 441–7(2020). https://doi.org/10.1002/jmv.25689 PMID: 31994742
38. ArcGIS Desktop Arcmap 10.7. https://www.esri.com/zh-cn/home.
39. Sobral M., Duarte G. et al. Association between climate variables and global transmission oF SARS-CoV-2. Science of The Total Environment 729, 0048–9697 (2020).