

Transmission potential and forecasting of the number of Coronavirus disease 2019 cases in Hubei Province, China

Ke-wei Wang

Jiangnan University

Jie Gao

Shangdong University

Hua Wang

Jiangnan University

Xiao-long Wu

Jiangnan University

Qin-fang Yuan

Jiangnan University

Yang Cheng (✉ woerseng@126.com)

Jiangnan University

Research article

Keywords: COVID-19, transmission, SARS-CoV-2, ARIMA–SVM, R(t)

Posted Date: November 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-36755/v2>

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

[Read Full License](#)

Transmission potential and forecasting of the number of Coronavirus disease 2019 cases in Hubei Province, China

Ke-wei Wang^{1,2,#}, Jie Gao^{3#}, Hua Wang², Xiao-long Wu¹, Qin-fang Yuan¹, Yang Cheng²

¹ Department of hospital infection, Affiliated Hospital of Jiangnan University, Wuxi, China

² Wuxi School of Medicine, Jiangnan University, Wuxi, China

³ Medicine, Shandong Province Maternal and Child Health Care Hospital Affiliated to Shandong University, Jinan, China

[#]Ke-Wei Wang and Jie Gao contributed equally to this work

*Correspondence: Yang Cheng, MD, Department of public health and Preventive Medicine, Wuxi School of Medicine, Jiangnan University, no. 1800, lihu avenue, wuxi, jiangsu, 214122, China. woerseng@126.com

Background: Coronavirus disease 2019 (COVID-19) was first reported in Wuhan, Hubei Province, China. We aimed to describe the temporal and spatial distribution and the transmission dynamics of COVID-19 and to assess whether a hybrid model can forecast the trend of COVID-19 in Hubei Province.

Method: The data of COVID-19 cases were obtained from the website of the Chinese Center for Disease Control and Prevention, whereas the data on the resident population were obtained from the website of the Hubei Provincial Bureau of Statistics. The temporal and spatial distribution and the transmission dynamics of COVID-19 were described. A combination of an autoregressive integrated moving average (ARIMA) and a support vector machine (SVM) was constructed to forecast the trend of COVID-19.

Results: A total of 56,062 confirmed COVID-19 cases, which were mainly concentrated in Wuhan, were reported from 16 January to 16 March 2020 in Hubei Province. The daily number of confirmed cases exponentially increased to 3,156 before 4 February 2020, fluctuated on an upward trend to 4,823 before 13 February 2020, and then markedly decreased to one case after 16 March 2020. The highest mean reproduction number $R(t)$ of 9.48 was recorded on 16 January 2020, after which it decreased to 2.15 on 2 February 2020 and further dropped to less than one on 13 February 2020. In the modelling stage, the mean square error, mean absolute error and mean absolute percentage error of the hybrid ARIMA–SVM model decreased by 98.59%, 89.19% and 89.68%, and those of SVM decreased by 98.58%, 87.71% and 88.94% compared with the ARIMA model. Similar results were obtained in the forecasting stage.

Conclusion: Public health interventions resulted in the terminal phase of COVID-19 in Hubei Province. The hybrid ARIMA–SVM model may be a reliable tool for forecasting the trend of the COVID-19 epidemic.

Keywords COVID-19, transmission, SARS-CoV-2, ARIMA–SVM, $R(t)$

Background

In December 2019, coronavirus disease 2019 (COVID-19) was detected in Wuhan (the capital of Hubei Province, China) and then quickly spread to other cities outside Wuhan [1, 2]. As of 12 November 2020, a total of 52,041,441 confirmed COVID-19 cases had been reported in 191 countries or regions globally [3]. Although recent studies have illustrated the epidemiology of COVID-19 in Wuhan, the COVID-19 transmission intensity from Wuhan to other cities in Hubei Province has not been estimated. Therefore, a retrospective study of the temporal and spatial distribution and the transmission dynamics of COVID-19 in Hubei Province was needed.

With the miscalculation of the early epidemic trend, the COVID-19 pandemic has become a large economic burden globally. Many statistical methods have been used to predict the trend of the disease [4, 5]. The autoregressive integrated moving average (ARIMA) is the most popular time series used to forecast infectious diseases. The ARIMA facilitates a linear relationship between present and past data points and errors in a time series [6]. However, infectious diseases are characterised by several challenges, such as uncertainty, complexity and a nonlinear pattern [7, 8]. The artificial neural network is ideal for forecasting via complex nonlinear mapping without prior knowledge of problem-solving [9, 11].

Therefore, this study was designed to describe the temporal and spatial distribution and the transmission dynamics of COVID-19 between 16 January and 13 March 2020 in Hubei Province to identify implications for the rapid intervention of infectious diseases, especially COVID-19. We also created a hybrid ARIMA–support vector machine (SVM) model to forecast the trend of COVID-19 in Hubei Province. We further assessed whether three models

(i.e. ARIMA, SVM and ARIMA–SVM) could forecast the trend of COVID-19 in Hubei Province. We aimed to select the optimal model using only the data from the original areas during the early outbreak period. This study provides a methodological reference for the early prediction of the epidemic.

Methods

Data collection

The time series of the observations of confirmed COVID-19 cases from 16 January to 16 March 2020 was obtained from the website of the Chinese Center for Disease Control and Prevention (<http://2019ncov.chinacdc.cn/2019-nCoV/>). The information collected included the number of confirmed cases in Hubei Province. Data of the resident population were obtained from the website of the Hubei Provincial Bureau of Statistics (<http://tjj.hubei.gov.cn/>).

A confirmed case was defined as a suspected case with a positive result for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) according to real-time reverse transcriptase–polymerase chain reaction assay or high-throughput sequencing of nasal and pharyngeal swab specimens [12]. The real-time reproduction number, $R(t)$, is the expected number of secondary cases that each infected individual would infect if the conditions remained as they were at time t . A Bayesian statistical framework was used to calculate $R(t)$ based on the number of COVID-19 cases, the number of secondary cases, the serial interval, and the five-day moving average [12, 13]. $R(t)$ and its 95% credible interval (CI) for the entire period were calculated.

Three models, namely, ARIMA, SVM and ARIMA–SVM, were constructed to forecast the trend of COVID-19. We compared the performance of these models in the train dataset and the

test dataset. Initially, 58 days of data between 16 January and 13 March 2020 were analysed to construct the single ARIMA model. Considering the small sample size, the predictive validity of the models was calculated using three days of data from 14 to 16 March 2020.

The ARIMA model

The ARIMA model was first developed in 1970 [14]. ARIMA was calculated using the following formula: ARIMA (p, d, q), where p denotes the number of parameters in the autoregressive (AR) model, d represents the degree of difference and q stands for the number of parameters in the moving average (MA) model. A logarithmic transformation ARIMA model (p, d, q) was constructed by autonomous identification in R version 3.6.2. Several ARIMA models were identified. The optimum model was selected using Akaike's information criterion (AIC). White noise tests were used to determine the independence and normal distribution of the residual data.

The SVM model

SVM is a machine learning tool that is used to carry nonlinear classification and regression of data [15]. For the classification of subjects, SVM constructs an optimal classification hyperplane by mapping data to a high-dimensional space. In this study, the steps for establishing the SVM model for nonlinear regression were as follows:

- 1) The number of daily new confirmed COVID-19 cases in Hubei Province was y_t , and the sliding window was used to train SVM for y_t . The value of $y_{(t+1)}$ was predicted using SVM constructed by $y_t, y_{(t-1)}, \dots, y_{(t-k+1)}$. Considering the small sample size, four loops of data were selected, and the test dataset only included the data of cases in the last three days.
- 2) These optimal parameters of SVM were determined by gridding search to minimise the

mean square error (MSE) between the predicted and actual values for the training model.

3) SVM was used to predict the daily number of confirmed cases in Hubei Province.

The ARIMA–SVM model

The hybrid model developed in this study combined ARIMA (linear approach) and SVM (nonlinear approach). A three-step methodology was followed. In the first step, an ARIMA model was used to analyse the linear dimension. In the second step, the SVM model was used to model the residuals derived from the ARIMA model, whereby e_t represented the residual parameter at time t in the ARIMA model, in which $e_t = y_t - \hat{y}_t$, where \hat{y}_t denotes the forecast value and y_t represents the daily number at time t . In the third step, the predicted error (\hat{e}_t), denoting the estimation of e_t , was obtained using the SVM model (nonlinear approach). The ARIMA–SVM model yields the predicted value $\hat{\mu}_t = \hat{y}_t + \hat{e}_t$.

Forecast evaluation

The performance of the model was determined based on the similarity between the forecast values and test results and the observed data. Three different measures were used to compare the performance of the ARIMA, SVM and ARIMA–SVM models, including MSE, mean absolute error (MAE) and mean absolute percentage error (MAPE). Lower values suggest higher prediction accuracy. In other words, an optimal prediction model was obtained by minimising the MSE, MAE and MAPE.

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}$$

Data processing and analysis

ArcGIS software version 10.6 (Environmental Systems Research Institute Inc.) was used to plot the geographical locations of daily new confirmed COVID-19 cases across Hubei Province from 16 January to 16 March 2020, and we carried out descriptive analyses of the dates of illness onset. Other analyses were conducted in R version 3.6.2. In all analyses, a two-tailed significance level of 0.05 was used.

Ethical review

The study protocol and utilisation of COVID-19 case data were obtained from the websites of the Chinese Center for Disease Control and Prevention (<http://2019ncov.chinacdc.cn/2019-nCoV/>) and the Hubei Provincial Bureau of Statistics (<http://tjj.hubei.gov.cn/>). No ethical issues were identified.

Results

Temporal patterns and reproduction number ($R[t]$) of COVID-19 cases

A total of 56,062 confirmed COVID-19 cases were reported from 16 January to 16 March 2020 in Hubei Province, China. The daily confirmed case number exponentially increased to 3,156 before 4 February 2020, fluctuated to 4,823 before 13 February 2020, and then markedly decreased to one case after 16 March 2020 (Figure 1). The occurrence of COVID-19 was initially concentrated in Wuhan before 23 January 2020, and it then spread rapidly to all areas in Hubei Province by 29 January (Figure 2). Although Wuhan had the highest rate, other areas in Hubei Province (except Enshi) had a rate of above 0.5 cases per 10,000 people between 29 January and 18 February. After 19 February, the daily new COVID-19 cases gradually

decreased in Hubei Province and were concentrated in Wuhan. The highest mean reproduction number $R(t)$ of 9.48 was recorded on 16 January 2020, it then dropped to 2.15 on 2 February 2020. This further decreased to less than one case on 13 February 2020 (Figure 3).

The optimised ARIMA model

After the logarithmic transformation for the number of new confirmed COVID-19 cases, autonomous identification in R version 3.6.2 was used to construct ARIMA. Based on the results of the goodness-of-fit test statistics, we confirmed ARIMA (0, 2, 1) as the optimal model, with the lowest AIC (83.10) of the six candidate models. The parameter estimates of the ARIMA model are shown in Table 1. The new confirmed COVID-19 cases from 14 to 16 March 2020 were predicted using the constructed ARIMA model (0, 2, 1). The QQ graph and Ljung-Box test ($\chi^2 = 1.0268$, $p = 0.3109$) of the residuals suggested that the residuals were normal and were not correlated (Figure 4). Hence, the model fitting data was sufficient and can be used for prediction.

Table 1 Model estimation of the ARIMA (0, 2, 1) model

Variable	Coefficient	Standard error	t statistics	P values
MA (1)	-0.8506	0.0603	-14.106	<0.0001

The single SVM model

The SVM model facilitated the prediction of the number of confirmed cases in Hubei Province by representing nonlinear and complex data. According to the steps of establishing the SVM model, the previous four-time data were used as input parameters into the model. Considering that the input data had a small sample size, we did not use cross-validation to determine the model. We adopted the eps-regression model and used different kernel functions. The radial was used as the kernel function, and we obtained the best delay predicting result, in which the

number of support vectors was eight, the kernel parameter was $\gamma = 1e-04$, and the insensitive loss coefficient was $\varepsilon = 0.1$. The prediction results were compared with the initial results, as shown in Figures 1 and 5.

The hybrid ARIMA–SVM model

The residual series from ARIMA was used as the target series for the SVM. Similar to the single SVM model, the SVM model for residual series was constructed with five support vectors, a kernel parameter of $\gamma = 0.001$ and an insensitive loss coefficient of $\varepsilon = 0.1$. The prediction results were compared with the initial results, as shown in Figures 1 and 5.

Prediction accuracy

The parameters of MAE, MSE and MAPE in the ARIMA, SVM and hybrid ARIMA–SVM models in the train dataset and the test dataset are shown in Table 2. In the modelling stage, using the train dataset, the MSE (7,144.944), MAE (37.333) and MAPE (0.042) of the ARIMA–SVM were lower than those of the ARIMA and SVM models. Further, in the modelling stage, the MSE, MAE and MAPE of the hybrid ARIMA–SVM model decreased by 98.59%, 89.19% and 89.68%, whereas those of the SVM decreased by 98.58%, 87.71% and 88.94%, respectively, compared with the ARIMA model. In the forecasting stage, using the test dataset, the hybrid ARIMA–SVM model had lower MSE, MAE and MAPE than the ARIMA model, whereas the SVM had higher MSE, MAE and MAPE than the ARIMA and hybrid ARIMA–SVM models. Therefore, the ARIMA–SVM model was superior to the ARIMA and SVM models in accurately forecasting the number of COVID-19 cases in Hubei Province.

Table 2 Comparison of the modeling and forecasting performance of ARIMA, SVM and ARIMA-SVM model

Model	train dataset			test dataset		
	MSE	MAE	MAPE	MSE	MAE	MAPE
ARIMA	505840.134	345.324	0.407	2.282	1.418	0.892
SVM	7165.682	42.431	0.045	30.355	5.045	2.194
ARIMA-SVM	7144.944	37.333	0.042	2.071	1.417	0.776

Discussion

In this study, 56,062 confirmed COVID-19 cases, which were mainly concentrated in Wuhan, were reported from 16 January to 16 March 2020 in Hubei Province, China. The temporal and spatial distribution and the transmission dynamics of COVID-19 were also described. Three models, namely, ARIMA, SVM and ARIMA–SVM, were constructed using time series to predict the number of COVID-19 cases. The values of MSE, MAE and MAPE of the ARIMA–SVM model were the least among the three models, suggesting the superiority of this model in predicting the daily number of COVID-19 cases in Hubei Province.

Considering the outbreak of COVID-19 in Wuhan, Hubei Province’s response was immediate and emphatic. To stop the spread of the disease, the government imposed public health interventions, including city lockdown, traffic suspension, home quarantine, compulsory mask-wearing in public places and cancellation of social gatherings after 23 January 2020. A stay-at-home policy was implemented, and Fang Cang Shelter hospitals were used after 4 February 2020. Door-to-door and individual-to-individual symptom screening was performed for all residents of Hubei Province after 17 February 2020 [12, 16, 17]. In Hubei Province, the COVID-19 epidemic was of approximately 60 days’ duration. It started with four cases reported

on 16 January 2020 to the peak value of a mean of 2201.76 cases daily between 29 January and 18 February 2020 and then to a mean of 233.78 between 19 February and 16 March 2020. The end of the COVID-19 epidemic was about to come true in Hubei Province [12, 16]. These changes in the time trend of new COVID-19 cases demonstrated the effectiveness of the public health interventions implemented in Hubei Province in preventing the spread of COVID-19 [16–19]. Additionally, with the implementation of public health interventions, the $R(t)$ gradually decreased and eventually reached less than 1 after 13 February 2020. This result further emphasised that public health interventions effectively stopped the spread of COVID-19.

Although most countries implemented various preventive and control strategies for COVID-19, the strict containment strategy implemented in China could not be followed in other countries, herd immunity was even advocated in several countries. Nowadays, the world has been suffering from the COVID-19 pandemic for about a year, and it is sweeping through 191 countries/regions (52,041,441 confirmed cases with 1,282,046 deaths) [3]. Therefore, accurately forecasting the trend of the second wave of COVID-19 is a great challenge. In this study, compared to ARIMA and SVM, the hybrid ARIMA–SVM model was a more accurate forecasting tool for the trend of COVID-19 in Hubei Province. This result provides a theoretical reference for initial and short-term prediction of infectious diseases.

The early and accurate prediction of COVID-19 can provide an important reference for outbreak control. ARIMA, which explores the dynamic development and change of infectious disease over time, has been widely used for early prediction of some infectious diseases, such as tuberculosis, schistosoma, malaria and bacterial dysentery [20–23]. However, ARIMA cannot adjust some confounding factors that influence infectious disease, and it cannot estimate

some parameters when the data are complex. Hence, it results in the inaccurate prediction of outbreaks in small samples [14]. Since COVID-19 was first detected in Wuhan, various public health interventions have been used to help control the outbreak [12]. These interventions inevitably disrupted the original onset state of the disease and have led to the instability and discontinuity of the time series data, thus making ARIMA unsuitable. Furthermore, ARIMA based on a linear model cannot capture information about nonlinear data of time series. With the development of machine learning, the prediction accuracy of the disease can be improved substantially through machine learning [9]. Among the machine learning methods, the SVM has excellent generalisation ability and can effectively help solve many unpredictable problems. SVM can perform classification and regression by training [15]. Nowadays, SVM is widely used in time series by fitting small sample data, high-dimensional data, nonlinear data and other complex data with high prediction accuracy.

To utilise the advantages of various models, many studies have constructed a combination of kinds of models to carry out a time series for infectious diseases, such as tuberculosis, schistosoma, malaria and bacterial dysentery [5, 8, 20, 24]. Compared with the single model, the hybrid model can fully utilise all kinds of sample information more systematically and comprehensively. A study has reported the application of the ARIMA model on the COVID-19 epidemic dataset from the website of the Johns Hopkins University [25]. However, this study only visualises the predicted epidemiological trend of the prevalence and incidence of COVID-19 and lacks the values of prediction accuracy, including MSE, MAE and MAPE. This suggests the need for further comparison and data collection in real-time. In contrast, our study, which included 56,062 confirmed COVID-19 cases (the laboratory confirmation of SARS-CoV-2 infection in the biosamples) from 16 January to 16 March 2020 in Hubei Province, can meet the requirements for further comparison. Another study that evaluates the spatial

dependency and temporal dynamics of COVID-19 demonstrates that incidence rates are concentrated in the Wuhan metropolitan area, but the prediction of the trend of COVID-19 is lacking [26]. In this study, ARIMA was used to fit the linear part of the time series of daily new confirmed cases of COVID-19 in Hubei Province, whereas the SVM model was used to fit the nonlinear part of the time series. The hybrid ARIMA–SVM model, with the advantages of ARIMA and SVM, was established to predict the number of COVID-19 cases.

COVID-19 is a new, infectious disease that may be influenced by population density, environmental changes, public health interventions and climate [27, 28]. To comprehend how multiple factors influence COVID-19, the collection and analyses of data on the epidemic situation and related influencing factors from different provinces and countries after a substantial amount of time are required. As the COVID-19 pandemic continues and the data on this disease gradually become available, collaborative research across regions or countries may have a ready answer for the dilemma we are facing today: the virus can invade through any area at any time. However, during the early outbreak period, the combined model using only the data from the original areas may be a reliable forecasting tool based on the predicted effect of the hybrid ARIMA–SVM model, which was better than those of ARIMA and SVM alone.

Conclusions

With the implementation of public health interventions, the number of daily new COVID-19 cases and the $R(t)$ first increased rapidly and then decreased gradually from 16 January to 16 March 2020 in Hubei Province. This change emphasised that public health interventions effectively stopped the spread of COVID-19. Further, the hybrid ARIMA–SVM model may be a reliable forecasting tool for the trend of COVID-19 in Hubei Province. Our findings may

have public health implications for the prevention and control of infectious diseases, especially COVID-19.

Declarations

Abbreviations

COVID-19: coronavirus disease 2019; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; RT-PCR: Real-time fluorescent reverse transcription polymerase chain reaction; R(t): real-time reproduction number; CI: credible intervals; SARS: Severe Acute Respiratory Syndrome; MERS: Middle East Respiratory Syndrome.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by The National Natural Science Foundation of China (81871681), The National First-class Discipline Programme of Food Science and Technology (JUFSTR20180101), The Fundamental Research Funds for the Central Universities funded by the Ministry of Education of China (JUSRP51710A), the project of Wuxi Science and Technology Supporting Plan (N2020X004 and N2020X011), Wuxi Project of Double-Hundred Talent Plan (No. HB2020040 and No. BJ2020100), Mega-project of Wuxi Commission of Health (No. Z202007 and No. Z202019) and Wuxi Key Medical Disciplines (No. ZDXK009),

Wuxi Project of Youth Talent (No. QNRC008). The funders had no role in study plan, design, data collection and analysis, interpretation of the results, decision to publish, or preparation of the manuscript.

Authors' contributions

YC had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. KW and JG conceived the study. HW, XW and QY analyzed data and made the figures. KW and YC wrote the first draft of the manuscript. All authors interpreted the results, and approved the final version for submission.

Acknowledgements

Not applicable.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020;382(8):727-733.
2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* 2020;395(10223):470-473.
3. Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU), <https://coronavirus.jhu.edu/map.html>. Accessed: 12 November 2020
4. Ray EL, Reich NG. Prediction of infectious disease epidemics via weighted density ensembles. *PLoS Comput Biol* 2018;14(2):e1005910.
5. Su TL, Jaki T, Hickey GL, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models. *Stat Methods Med Res* 2018;27(1):185-197.
6. Zheng A, Fang Q, Zhu Y, Jiang C, Jin F, Wang X. An application of ARIMA model for predicting total health expenditure in China from 1978-2022. *J Glob Health* 2020;10(1):010803.

7. Xie J, Kanghuai Z, Cai Y, Wang N, Wang H, Zhao Y, *et al.* Application of ARIMA model in monitoring the use rate of antibiotics in outpatients in 2010 - 2018. *Int J Clin Pharmacol Ther* 2020;58(5):282-288.
8. Fang X, Liu W, Ai J, He M, Wu Y, Shi Y, *et al.* Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. *BMC Infect Dis* 2020;20(1):222.
9. Guo Z, He K, Xiao D. Early warning of some notifiable infectious diseases in China by the artificial neural network. *R Soc Open Sci* 2020;7(2):191420.
10. Dakappa PH, Prasad K, Rao SB, Bolumbu G, Bhat GK, Mahabala C. Classification of Infectious and Noninfectious Diseases Using Artificial Neural Networks from 24-Hour Continuous Tympanic Temperature Data of Patients with Undifferentiated Fever. *Crit Rev Biomed Eng* 2018;46(2):173-183.
11. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, *et al.* A Deep Learning Approach to Antibiotic Discovery. *Cell* 2020;180(4):688-702.e13.
12. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, *et al.* Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA* 2020; 323(19):1-9.
13. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol* 2013;178(9):1505-12.
14. Sato RC. Disease management with ARIMA model in time series. *Einstein (Sao Paulo)* 2013;11(1):128-31.
15. Li Y, Zhang T. Deep neural mapping support vector machines. *Neural Netw* 2017;93:185-194.
16. Leung K, Wu JT, Liu D, Leung GM. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a

- modelling impact assessment. *The Lancet* 2020;395(10233):1382-1393.
17. Adhikari SP, Meng S, Wu Y, Mao Y, Ye R, Wang Q, *et al.* Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. *Infectious diseases of poverty* 2020;9(1):29-12.
 18. Deng S, Peng H. Characteristics of and Public Health Responses to the Coronavirus Disease 2019 Outbreak in China. *Journal of clinical medicine* 2020;9(2):575.
 19. Ye Q, Wang B, Mao J, Fu J, Shang S, Shu Q, *et al.* Epidemiological analysis of COVID-19 and practical experience from China. *J Med Virol* 2020; doi:10.1002/jmv.25813.
 20. Wang KW, Deng C, Li JP, Zhang YY, Li XY, Wu MC. Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiol Infect* 2017;145(6):1118-1129.
 21. Yan L, Wang H, Zhang X, Li MY, He J. Impact of meteorological factors on the incidence of bacillary dysentery in Beijing, China: A time series analysis (1970-2012). *PLoS One* 2017;12(8):e0182937.
 22. Wu JY, Zhou YB, Chen Y, Liang S, Li LH, Zheng SB, *et al.* Three Gorges Dam: Impact of Water Level Changes on the Density of Schistosome-Transmitting Snail *Oncomelania hupensis* in Dongting Lake Area, China. *PLoS Negl Trop Dis* 2015;9(6):e0003882.
 23. Ouedraogo B, Inoue Y, Kambire A, Sallah K, Dieng S, Tine R, *et al.* Spatio-temporal dynamic of malaria in Ouagadougou, Burkina Faso, 2011-2015. *Malar J* 2018;17(1):138.
 24. Liu W, Bao C, Zhou Y, Ji H, Wu Y, Shi Y, *et al.* Forecasting incidence of hand, foot and mouth disease using BP neural networks in Jiangsu province, China. *BMC Infect Dis* 2019;19(1):828.
 25. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data Brief* 2020;29:105340.

26. Yang W, Deng M, Li C, Huang J. Spatio-Temporal Patterns of the 2019-nCoV Epidemic at the County Level in Hubei Province, China. *Int J Environ Res Public Health* 2020;17(7).
27. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020;323(13):1239-1242.
28. E GA, C BS, S BR, de Groot RJ, Christian D, A GA, *et al.* The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature microbiology* 2020;5(4):536-544.

Figure titles and legends

Figure 1 Daily number of laboratory-confirmed COVID-19 cases in Hubei Province, China from 16 January to 13 March 2020.

Figure 2 Geographical distribution of 56,062 laboratory-confirmed COVID-19 cases in Hubei Province, China, with dates of illness onset from 16 January to 16 March 2020. Provinces are shaded according to confirmed case rate per 10,000 people during different periods. According to the mean COVID-19 cases per day and the public health interventions implemented, the outbreak was divided into five periods as follows: (A) a mean of 63.5 between 16 and 23 January; (B) city lockdown with traffic suspension and home quarantine and a mean of 672.83 between 23 and 29 January; (C) a mean of 2,154.00 between 29 January and 5 February; (D) Fang Cang Shelter hospitals and a mean of 2,317.46 between 6 and 18 February; (E) the number of COVID-19 cases per day under 1,000 and a mean of 233.78 between 19 February and 16 March.

Figure 3 Mean reproduction number ($R[t]$) with 95% CI based on confirmed COVID-19 cases

in Hubei Province, China. The effective reproduction number $R(t)$ is defined as the mean number of secondary cases generated by a typical primary case at time t in a population for the whole period over a five-day moving average. The blue horizontal line indicates $R(t) = 1$, below which sustained transmission is unlikely when anti-transmission measures are sustained, indicating that the outbreak is under control. A black dotted line indicates the 95% credible intervals of the mean reproduction number.

Figure 4 Normal QQ plot of standardised residuals for the ARIMA (0, 2, 1) model.

Figure 5 Daily number of COVID-19 cases in Hubei Province, China from 14 to 16 March 2020.

Figures

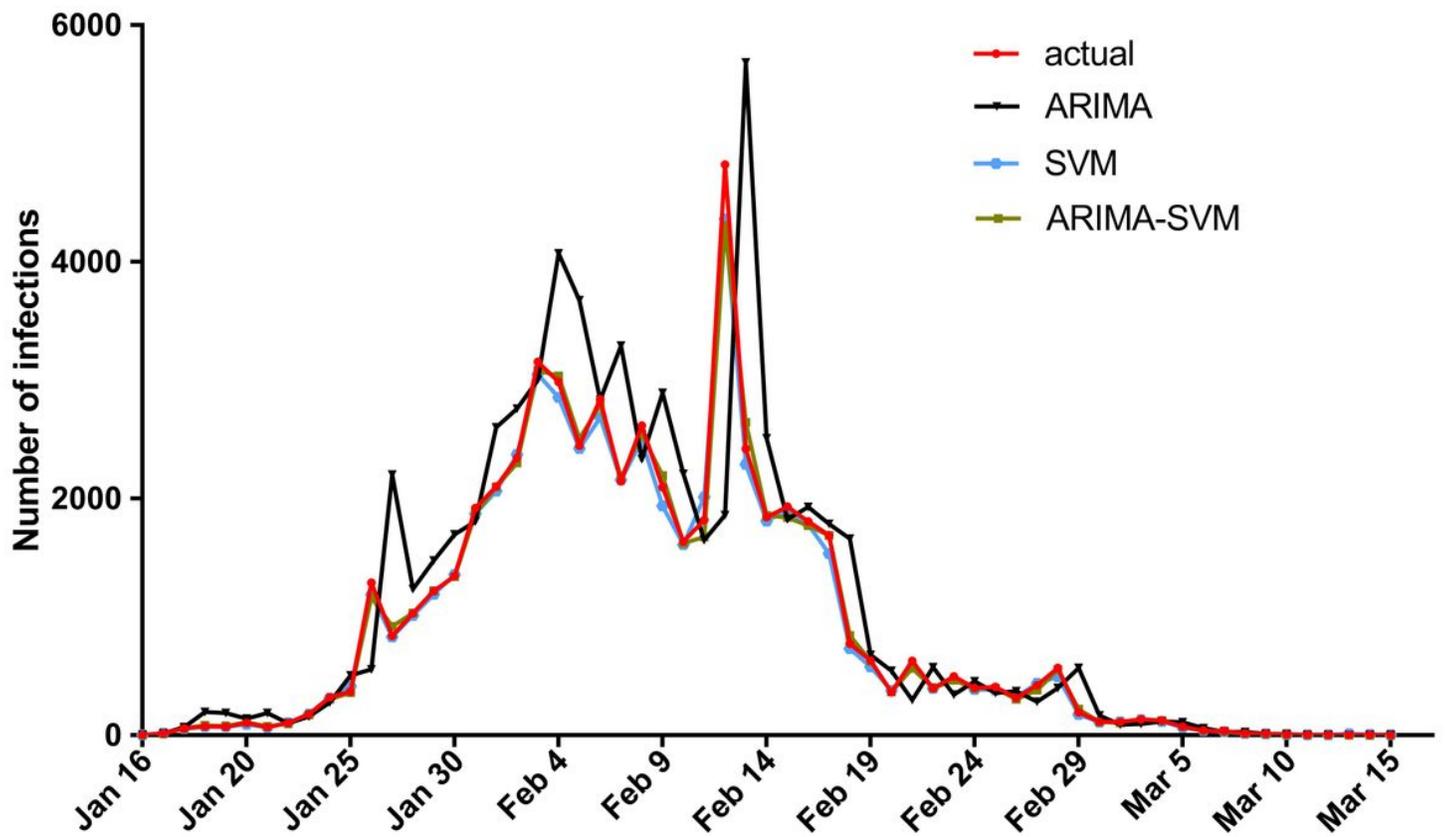


Figure 1

Daily number of laboratory-confirmed COVID-19 cases in Hubei Province, China from 16 January 2020 to 13 March 2020.

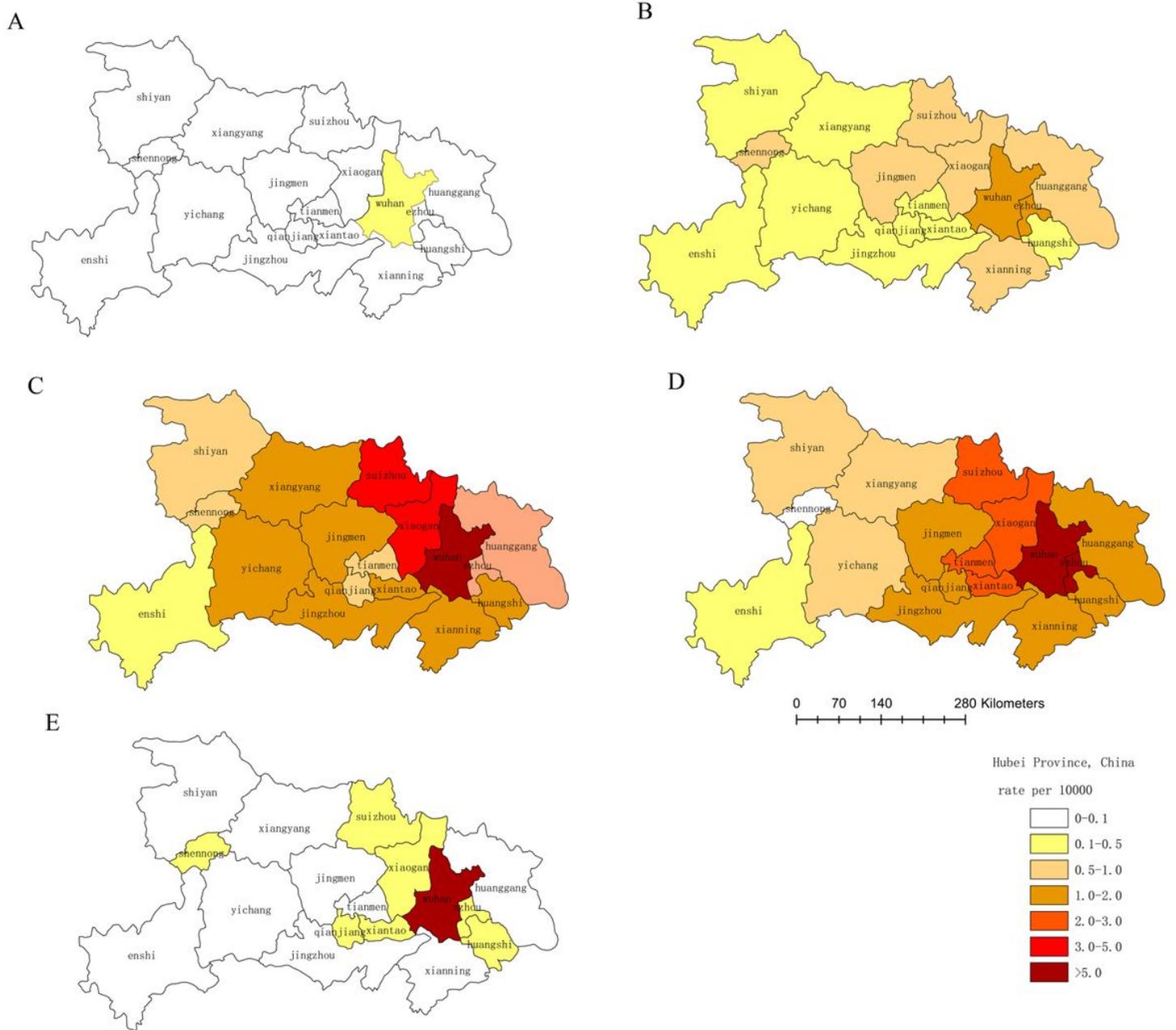


Figure 2

Geographical distribution of 56,062 laboratory-confirmed COVID-2019 cases in Hubei province, China, with dates of illness onset from 16 January 2020 to 16 March 2020. Provinces are shaded according to confirmed case rate per 10,000 people during different periods. According to the mean of COVID-19 cases per day and public health interventions implemented, the outbreak was divided into five periods as follows: (A) mean of 63.5 in 16 and January 23; (B) city lockdown with traffic suspension and home quarantine and the mean of 672.83 between January 14 and January 29; (C) mean of 2,154.00 between January 29 and February 5; (D) Fang Cang Shelter hospitals and the mean of 2,317.46 between February 6 and February 18; (E) number of COVID-19 cases per day under 1,000 and the mean of 233.78 between February 19 and March 16. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning

the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

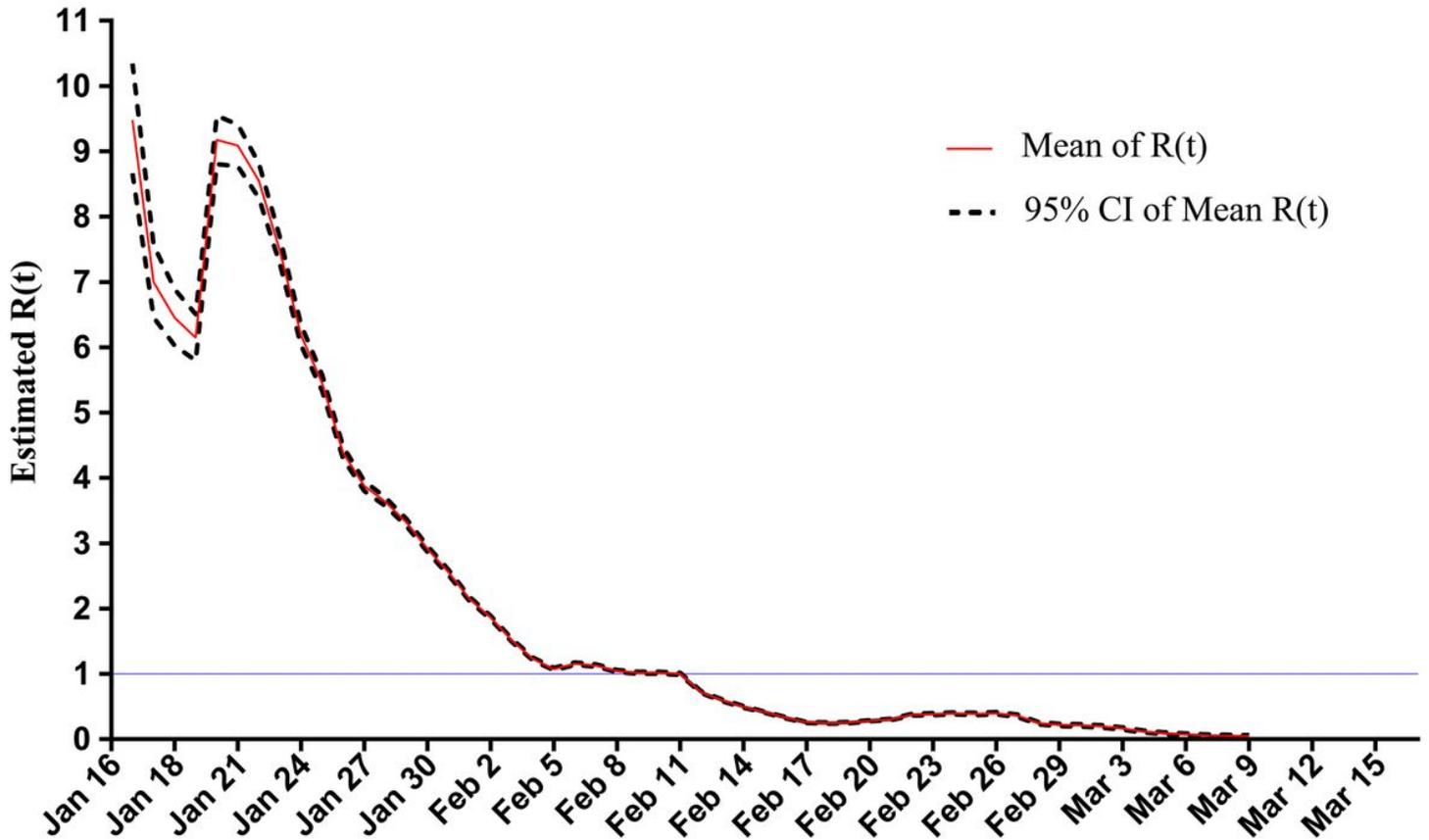


Figure 3

Mean reproduction number ($R[t]$) with 95% CI based on confirmed COVID-19 cases in Hubei Province, China. The effective reproduction number $R(t)$ is defined as the mean number of secondary cases generated by a typical primary case at time t in a population for the whole period over a five-day moving average. The blue horizontal line indicates $R(t) = 1$, below which sustained transmission is unlikely when anti-transmission measures are sustained, indicating that the outbreak is under control. A black dotted line indicates the 95% credible intervals of the mean reproduction number.

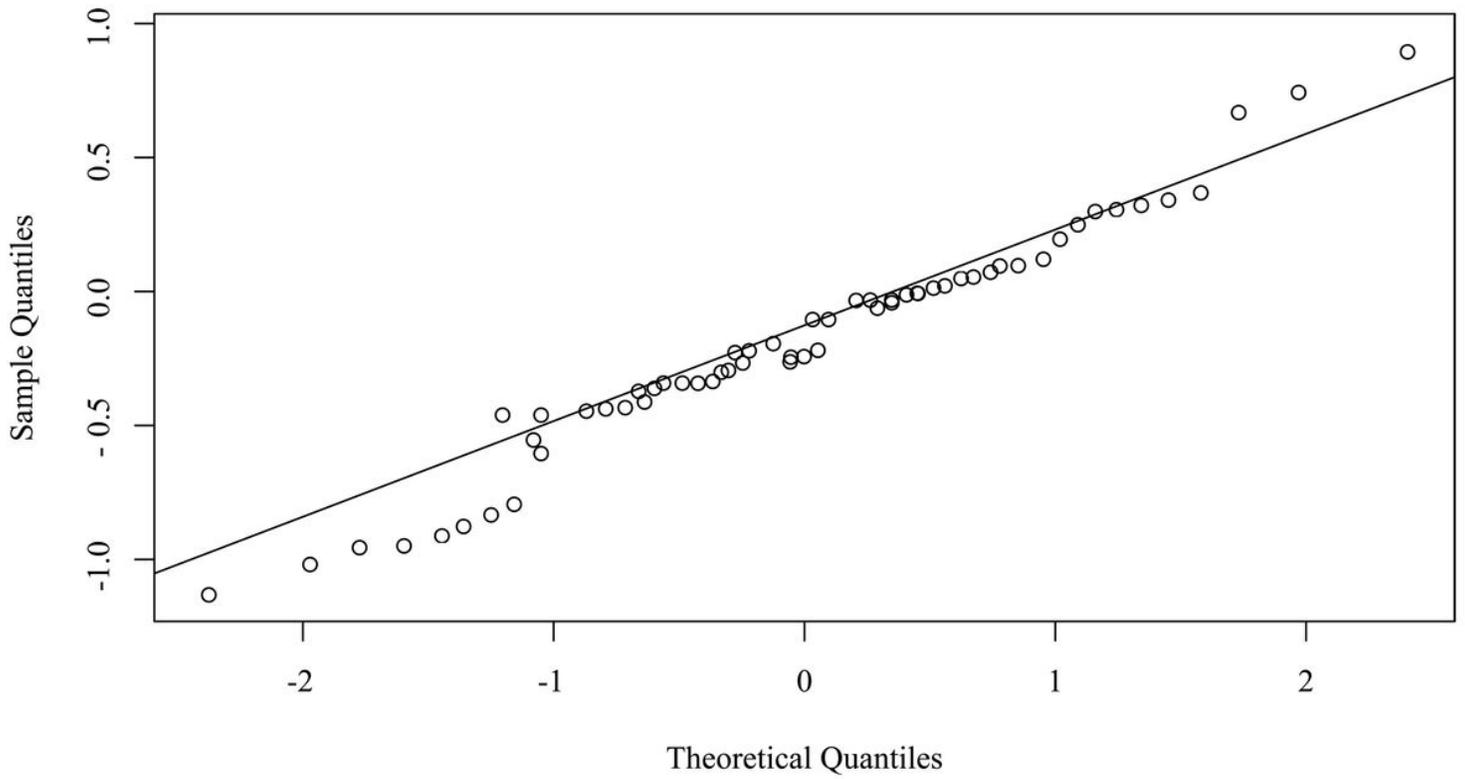


Figure 4

Normal QQ plot of standardized residuals for the ARIMA (0, 2, 1) model.

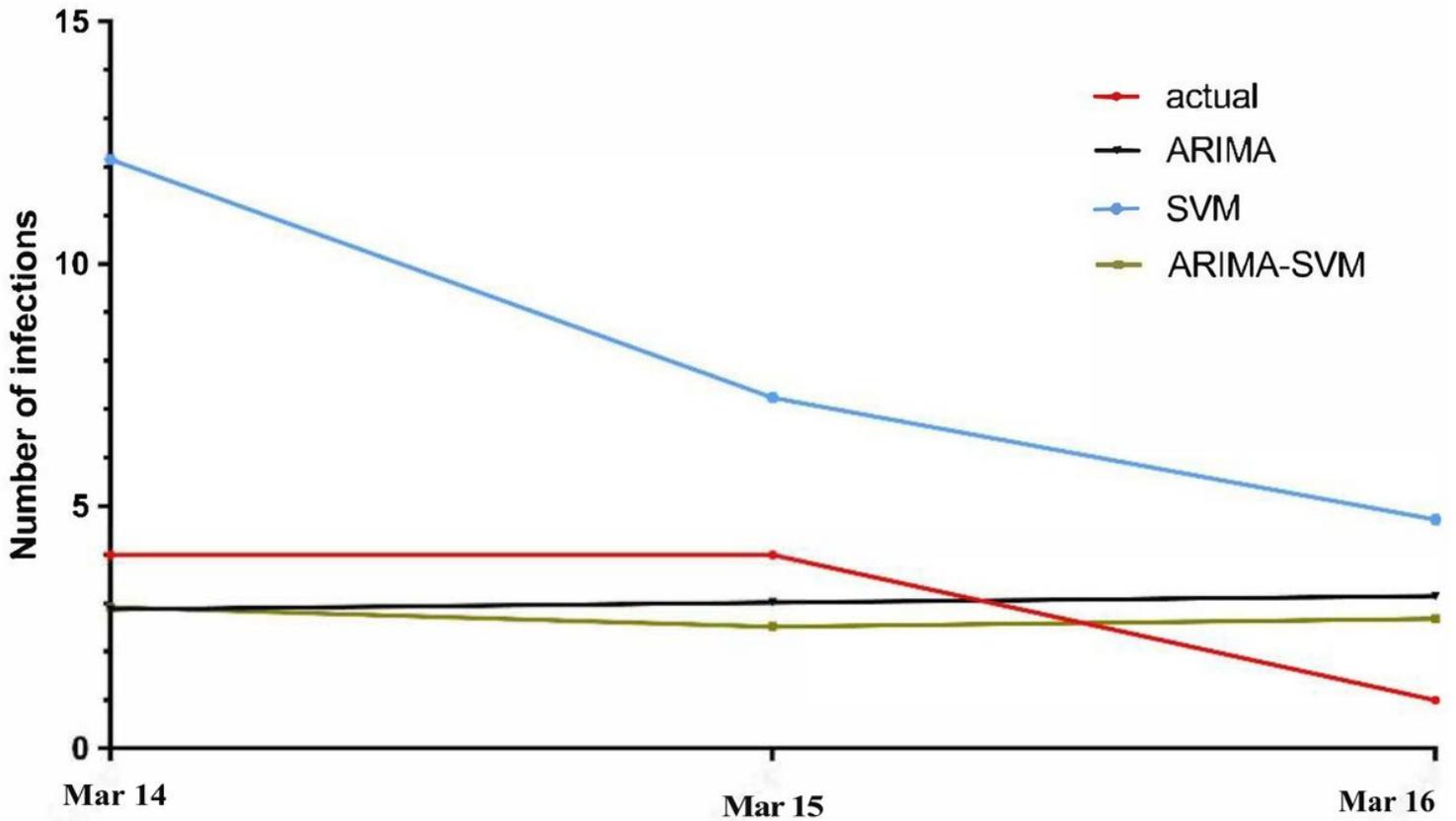


Figure 5

Daily number of COVID-19 cases in Hubei Province, China from March 14 to March

Supplementary Files

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

- [Originalmaps.rar](#)