

# 1 **Trend prediction of COVID-19 based on ARIMA model in mainland** 2 **of China**

3

4 Chuqiao Han<sup>a,b,1</sup>, Xifeng Ju<sup>a,b,1</sup>, Jianghua Zheng<sup>a,b,\*</sup>

5 a, College of Resources and Environment Sciences, Xinjiang University, Urumqi 830046, China

6 b, Key Laboratory of City Intellectualizing and Environment Modelling, Xinjiang University,  
7 Urumqi 830046, China

8

## 9 **Highlights:**

10 1. We predicted future COVID-19 occurrences in mainland of China based on ARIMA model.

11 2. We validated the model based on the previous outbreak data with actual data for June, 2020.

12 3. The measures taken by the government have contained spread of the epidemic

13 4. The combination of multiple models may improve the robustness of the model

14

15 **Abstract:** The ongoing pandemic of COVID-19 has aroused widespread concern around the  
16 world and poses a severe threat to public health worldwide. In this paper, the autoregressive  
17 integrated moving average (ARIMA) model was used to predict the epidemic trend of COVID-19  
18 in mainland of China. We collected the cumulative cases, cumulative deaths, and cumulative  
19 recovery in mainland of China from January 20 to June 30, 2020, and divided the data into  
20 experimental group and test group. The ARIMA model was fitted with the experimental group

---

<sup>1</sup>Chuqiao Hana and Xifeng Ju are co-first authors of the article.

\*Corresponding author at: College of Resources & Environment Science, Xinjiang University, Urumqi 830046, China; Institute of Arid Ecology and Environment, Xinjiang University, Urumqi 830046, China; Key Laboratory for Oasis Ecology, Xinjiang University, Urumqi 830046, China.

E-mail address: zheng\_jianghua@126.com (J. Zheng).

21 data, and the optimal model was selected for prediction analysis. The predicted data were  
22 compared with the test group. The average relative errors of actual cumulative cases, deaths,  
23 recovery and predicted values in each province are between -22.32%-22.66%, -9.52%-0.08%,  
24 -8.84%-1.16, the results of the comprehensive experimental group and test group show The error  
25 of fitting and prediction is small, the degree of fitting is good, the model supports and is suitable  
26 for the prediction of the epidemic situation, which has practical guiding significance for the  
27 prevention and control of the epidemic situation.

28 **Keywords:** ARIMA model; COVID-19; optimal model; prediction

29

## 30 **1. Introduction**

31 COVID-19 has spread worldwide and has caused tens of thousands of human deaths. At the  
32 same time[1], it has a huge impact on the world's medical, public health, economic, and other  
33 aspects [2]. In order to effectively suppress the spread of COVID-19, the Chinese government has  
34 successively announced the launch of a first-level response mechanism for major public health  
35 emergencies and implemented strict prevention and control measures[3]. By the end of June 2020,  
36 mainland of China had 85000 confirmed COVID-19 cases and 4648 deaths. Although the situation  
37 of prevention and control is getting better, the global situation is still not optimistic.

38 ARIMA model considers the law that the historical data of the research object itself changes  
39 with time, and uses this to predict future values, that is, to replace various influence factors with  
40 time. At present, the advantage of ARIMA model in predicting infectious diseases has been  
41 confirmed in many studies[4-5].

42 This study is based on the autoregressive integrated moving average model to predict the  
43 cumulative cases, cumulative deaths, and cumulative recovery of COVID-19 in mainland of China,  
44 and select the most appropriate model to simulate the epidemic pattern by simulating multiple  
45 models. A preliminary explanation is given to evaluate the effects of the epidemic prevention and  
46 control measures at this stage.

47

## 48 **2. Data and Methods**

### 49 **2.1 Data Sources**

50 The COVID-19 case data comes from the National and local health and construction

51 commission daily information release (<https://ncov.dxy.cn/>). We collected the cumulative cases,  
52 deaths, recovery in mainland of China from January 20, 2020 to June 30, 2020.

53

## 54 2.2 Methods

55 ARMA model is a stationary time series model, but COVID-19 daily cases, deaths and  
56 recovery changes with volatility, uncertainty, is a non-stationary random process, so this article  
57 introduces the difference operator I (The original non-stationary time series can be improved into  
58 a stationary time series after d-order difference), establish the COVID-19 cumulative cases, deaths,  
59 and recovery trend prediction ARIMA model. In recent years, the ARIMA model has become one  
60 of the most commonly used methods in the prediction of many epidemics. This method is  
61 particularly suitable for short-term prediction of infectious diseases, and its prediction accuracy  
62 has been widely recognized, at the same time, it can also provide effective help for disease  
63 prevention and policy-making[6].

64 ARIMA model is composed of autoregressive model (AR model (P)), moving average  
65 model( MA model (q) ) and difference operator I (d), where p, d and q are autoregressive order,  
66 difference order and moving average order of time series[7]. The expression of the p-order AR  
67 model is as follows:

$$68 \quad Y_t = C_1 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + e_t \quad ( 1 )$$

69 Where  $\alpha_i$  (  $i=1,2,\dots,p$  ) is the AR model coefficient,  $e_t$  is the random interference term, and  
70  $C_1$  is the constant.

71 The mathematical expression of the q order MA model is:

$$72 \quad Y_t = G(B)e_t + c_2 \quad ( 2 )$$

73 
$$G ( B ) = ( 1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_q B^q ) \quad (3)$$

74 Where  $\beta_j$  ( $j=1,2,\dots,q$ ) is the MA model coefficient,  $e_t$  is the random interference term, and  $c_2$   
75 is the constant.

76 Combining the time series after the difference of order  $d$ , the final expression of the ARIMA  
77 model is:

78 
$$F ( x ) = c + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \beta_1 e_{t-1} + \dots + \beta_q e_{t-q} \quad (4)$$

79 Where  $\alpha$  represents the coefficient of AR,  $\beta$  represents the coefficient of MA, and  $c$  is the  
80 constant term.

81 ARMA model construction ( Fig.1 The flowchart of ARIMA model Fig.1 ) :

82 (1) Data difference: the collected COVID-19 data are divided for several times to make it  
83 stable, and then the autocorrelation and partial autocorrelation tests are performed on the  
84 differential data.

85 (2) The choice of modeling parameters: according to an information criterion (AIC) and  
86 Bayesian information criterions (BIC), combing with the autocorrelation function (ACF) and  
87 partial autocorrelation function (PACF) of the residual sequence to determine the order of the  
88 model, and finally select the model with the highest fitting degree[8].

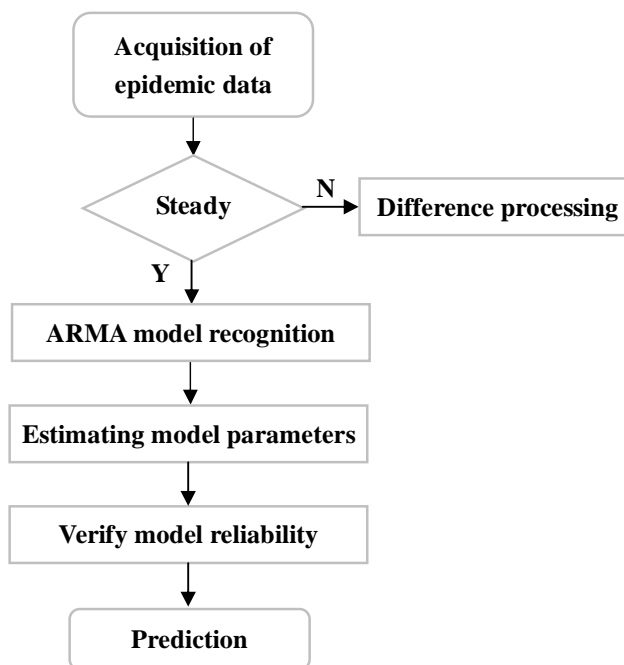
89 
$$AIC = -2 \ln ( L ) + 2m \quad (5)$$

90 
$$BIC = -2 \ln ( L ) + m \ln ( n ) \quad (6)$$

91 Where:  $L$  is the maximum likelihood function of the model,  $m$  is the number of estimated  
92 parameters, and  $n$  is the sample size.

93 (3) Verify the reliability of the model: we use the model to predict the cumulative cases,  
94 deaths, recovery in Chinese provinces (excluding Hong Kong, Macao, and Taiwan) from June

95 1-30, 2020, calculate the relative error from the actual cases, and verify the model predictive  
96 effect.  
97



98  
99

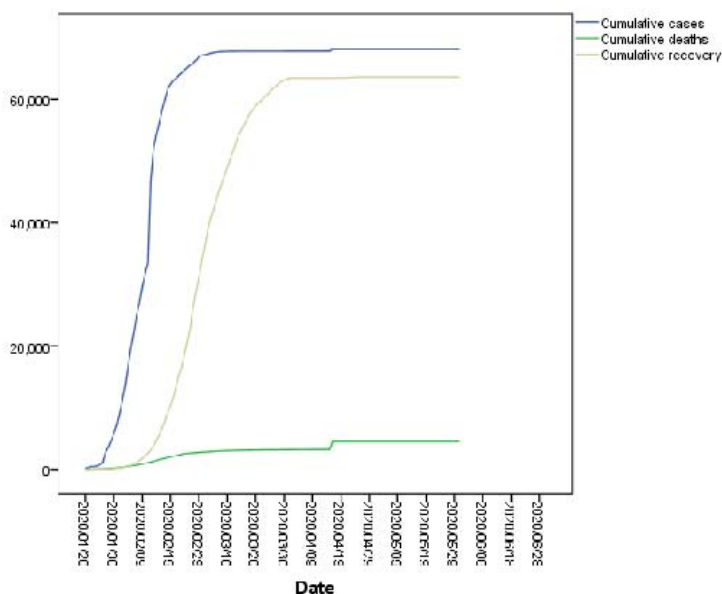
Fig.1 The flowchart of ARIMA model

### 100 3. Results

101 We build a model based on the cumulative cases, deaths, and recovery data in mainland of  
102 China from January 20 to May 30, 2020. Here we take Hubei Province as an example to describe  
103 the recognition process of the model.

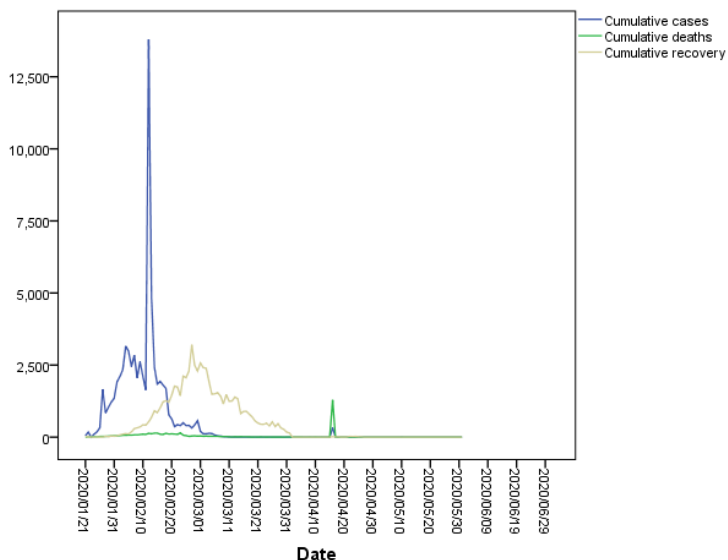
104 First, we build the original time series figure (Fig.2), from the figure, we can see that the  
105 initial outbreak of COVID-19 increased exponentially. The Chinese government took timely  
106 measures to control the epidemic. It can be seen from Fig.2 that the original data time series are  
107 not stationary, so we need to carry out first-order differential processing on the original data. The  
108 processed time series is shown in Fig.3. Since the time series in Fig.3 is still unstable, we need to

109 make two difference,  $d = 2$ , as shown in Fig.4. It can be seen from the figure that the data floats up  
110 and down in "0", which can judge that the data is stable after the second-order difference, so  $d = 2$   
111 is determined.



112

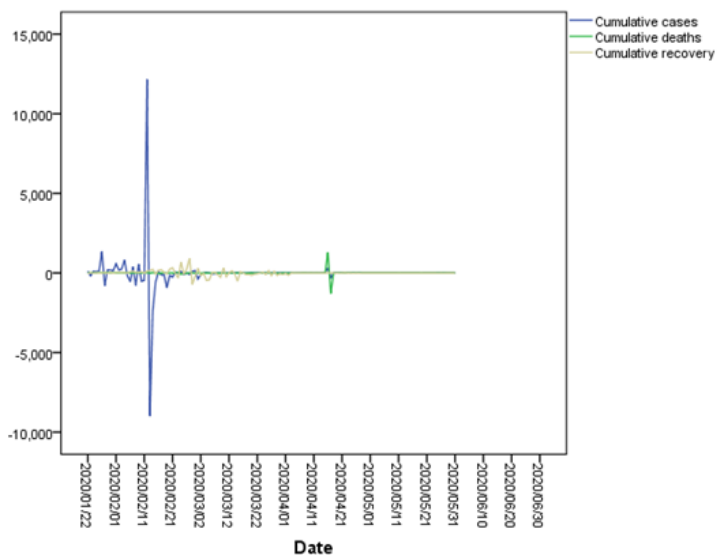
113 **Fig.2** The cumulative cases, deaths, recovery original time series figure



114

115 **Fig.3** First-order difference processing time series figure ( $d=1$ )

116



117

118

**Fig.4** Second-order difference processing time series figure (d=2)

119

120

After the d value is determined to be 2 by differential processing, we need to select different

121

combinations of p and q values to test the goodness of fit, and use the models with relatively small

122

AIC and BIC values as the best model, and then verify the model.

123

In the end, we obtained the models for predicting the cumulative cases, deaths, and recovery

124

in Hubei Province as ARIMA(0,2,1), ARIMA(1,2,1), ARIMA(1,2,4). The residual of the model is

125

tested by white noise sequence. As shown in Fig.5, the autocorrelation function and partial

126

autocorrelation function of the residual sequence are basically within 95% confidence interval,

127

indicating that there is no autocorrelation in the residual sequence. Then the model through the

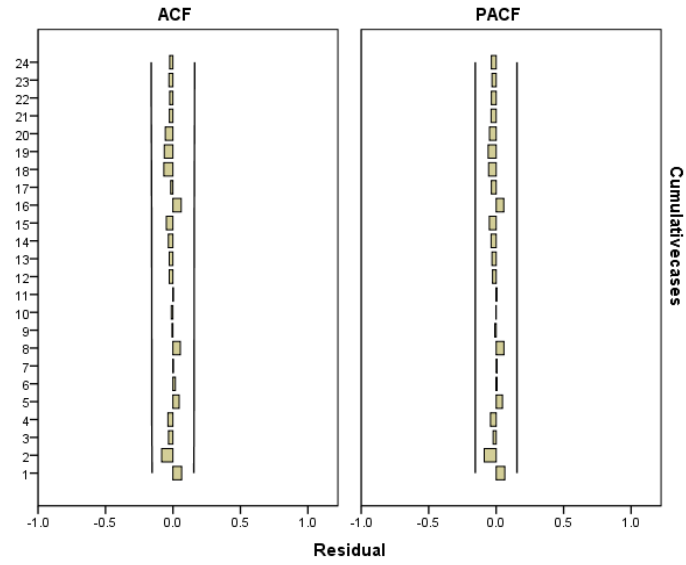
128

white noise test, the model can be used to predict the cumulative cases, deaths, and recovery in

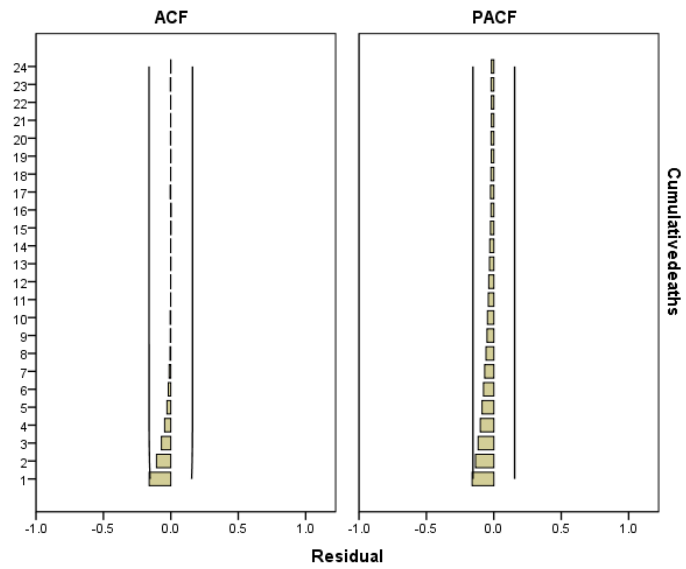
129

Hubei Province.

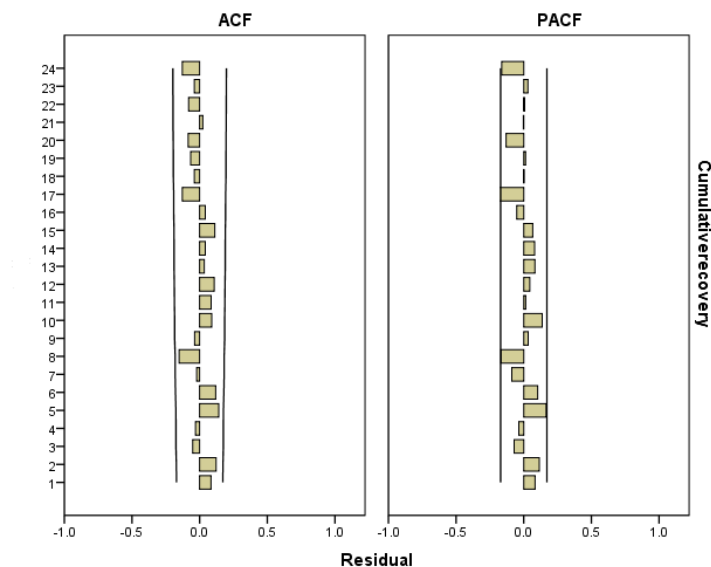




130



131



132

133 **Fig.5** Residual sequence figure of ARIMA model for the cumulative cases, deaths, recovery in  
 134 Hubei Province

135

136 Model test and error analysis : We divided the data into experimental group (January 20 -  
 137 May 31, 2020) and test group (June 1 - June 30, 2020), First of all, we bring the data of the  
 138 experimental group into the ARIMA model of the cumulative cases, deaths, recovery in each  
 139 province, and calculate the average relative error between the actual number and the fitted value.

140 The results are shown in Table 1.

141

142 **Table 1**

143 the average relative error table of actual cumulative cases, deaths, recovery and fitted values in each province  
 144 during January 20, 2020 to May 31, 2020

Province	Cumulative cases model	Cumulative cases	Cumulative deaths model	Cumulative deaths	Cumulative recovery model	Cumulative recovery
----------	------------------------	------------------	-------------------------	-------------------	---------------------------	---------------------

	type	average	type	average	type	average
		relative		relative		relative
		error		error		error
Beijing	ARIMA(0,2,1)	0.60%	ARIMA(1,2,17)	1.92%	ARIMA(0,2,1)	3.20%
Shanghai	ARIMA(2,2,1)	0.68%	ARIMA(0,1,0)	1.94%	ARIMA(0,2,4)	3.01%
Gansu	ARIMA(1,1,0)	0.92%	ARIMA(0,1,2)	0.00%	ARIMA(0,2,1)	0.71%
Sichuan	ARIMA(0,2,0)	1.75%	ARIMA(0,1,0)	1.63%	ARIMA(0,2,11)	3.47%
Hebei	ARIMA(1,2,8)	2.09%	ARIMA(0,1,4)	1.24%	ARIMA(0,2,7)	0.79%
Shaanxi	ARIMA(3,2,0)	1.78%	ARIMA(0,1,5)	1.49%	ARIMA(0,2,1)	1.65%
Guangdong	ARIMA(0,2,1)	19.64%	ARIMA(6,1,0)	2.23%	ARIMA(2,2,2)	1.63%
Liaoning	ARIMA(2,2,0)	1.12%	ARIMA(0,1,0)	0.93%	ARIMA(1,2,2)	2.36%
Chongqing	ARIMA(1,2,0)	1.01%	ARIMA(0,2,2)	0.65%	ARIMA(2,2,2)	1.16%
Fujian	ARIMA(1,2,0)	1.14%	ARIMA(0,1,0)	0.00%	ARIMA(0,2,1)	2.58%
Tianjin	ARIMA(0,2,5)	0.48%	ARIMA(0,2,1)	1.29%	ARIMA(0,2,1)	3.04%
Jiangsu	ARIMA(1,2,0)	2.15%	ARIMA(0,0,0)	*	ARIMA(0,2,13)	2.61%
Hubei	ARIMA(0,2,1)	1.99%	ARIMA(1,2,1)	1.69%	ARIMA(1,2,4)	0.98%
Zhejiang	ARIMA(3,2,0)	0.51%	ARIMA(0,1,0)	0.00%	ARIMA(0,2,7)	2.64%
Henan	ARIMA(2,2,0)	2.80%	ARIMA(1,2,2)	1.69%	ARIMA(2,2,2)	1.89%
Hainan	ARIMA(0,2,1)	0.62%	ARIMA(3,1,11)	1.83%	ARIMA(3,2,0)	3.12%
Heilongjiang	ARIMA(0,2,1)	1.30%	ARIMA(0,2,7)	1.39%	ARIMA(1,2,0)	1.21%
Jilin	ARIMA(0,2,1)	2.43%	ARIMA(0,1,0)	0.87%	ARIMA(0,2,2)	2.23%
Hunan	ARIMA(0,2,1)	2.01%	ARIMA(0,1,2)	1.62%	ARIMA(0,2,1)	1.86%

Shandong	ARIMA(1,2,0)	2.00%	ARIMA(0,1,6)	2.21%	ARIMA(4,2,0)	3.28%
Tibet	ARIMA(0,1,0)	0.00%	ARIMA(0,0,0)	*	ARIMA(0,1,0)	0.00%
Guangxi	ARIMA(0,2,8)	1.41%	ARIMA(0,1,4)	0.00%	ARIMA(0,2,3)	2.06%
Shanxi	ARIMA(6,2,0)	0.79%	ARIMA(0,0,0)	*	ARIMA(0,2,1)	2.54%
Ningxia	ARIMA(0,2,12)	0.49%	ARIMA(0,0,0)	*	ARIMA(2,2,1)	4.13%
Xinjiang	ARIMA(1,2,2)	1.15%	ARIMA(0,1,9)	1.38%	ARIMA(0,2,1)	0.41%
Jiangxi	ARIMA(0,2,5)	2.23%	ARIMA(0,1,0)	0.00%	ARIMA(0,2,3)	0.85%
Guizhou	ARIMA(1,2,7)	0.80%	ARIMA(0,1,0)	0.86%	ARIMA(0,2,1)	2.79%
Inner Mongolia	ARIMA(0,1,2)	-1.23%	ARIMA(0,1,0)	0.00%	ARIMA(1,1,1)	-8.57%
Anhui	ARIMA(1,2,0)	3.82%	ARIMA(0,1,0)	2.46%	ARIMA(0,2,10)	1.37%
Qinghai	ARIMA(0,1,6)	-2.92%	ARIMA(0,0,0)	*	ARIMA(0,2,6)	0.00%
Yunnan	ARIMA(1,2,0)	2.23%	ARIMA(0,1,0)	0.98%	ARIMA(0,2,6)	0.59%

145 Note: \* means that the cumulative death is 0, which cannot be calculated.

146

147 It can be seen from Table 1 that the average relative error of the cumulative cases in each  
 148 province is between -2.92% and 3.82% (except for Guangdong Province); the average relative  
 149 error of the cumulative cases in Guangdong Province is 19.64%. By analyzing the data of  
 150 Guangdong Province, we found that on January 23, 2020, the actual cumulative cases were 26,  
 151 and the fitting value was 1. This is mainly since at the early stage of the epidemic, the  
 152 transmission route of COVID-19 was not known, and the government did not take effective  
 153 measures, so the number of cases increased exponentially, which led to a large error at that time.  
 154 The average relative errors of cumulative deaths and recovery in various provinces are between

155 0-2.46% and -8.57%-4.13%, the errors are small and the fit is good, so the model is suitable. We

156 further analyzed the average relative error between the actual cases and the predicted values in

157 each province of the test group (June 1 - June 30, 2020). The results are shown in Table 2.

158

159 **Table 2**

160 the average relative error table of actual cumulative cases, deaths, recovery and fitted values in each province from

161 June 1st to 30th

Province	Cumulative cases	Cumulative deaths	Cumulative recovery
	average relative error	average relative error	average relative error
Beijing	22.66%	0.00%	0.45%
Shanghai	3.00%	-9.40%	-0.50%
Gansu	6.47%	0.00%	0.66%
Sichuan	-21.32%	-6.67%	-0.34%
Hebei	2.76%	0.00%	0.60%
Shaanxi	-0.32%	0.00%	0.19%
Guangdong	0.44%	0.00%	1.16%
Liaoning	1.74%	0.00%	-0.98%
Chongqing	0.30%	0.00%	0.00%
Fujian	0.88%	0.00%	-2.07%
Tianjin	1.93%	0.00%	1.10%
Jiangsu	0.05%	*	0.26%
Hubei	0.56%	0.08%	0.05%

Zhejiang	0.04%	0.00%	0.00%
Henan	0.00%	0.00%	-0.28%
Hainan	0.87%	0.00%	0.60%
Heilongjiang	0.21%	-8.70%	0.21%
Jilin	-0.15%	0.00%	-8.84%
Hunan	0.00%	0.00%	0.00%
Shandong	-1.30%	0.00%	0.35%
Tibet	0.00%	*	0.00%
Guangxi	-0.21%	0.00%	0.00%
Shanxi	0.00%	*	0.00%
Ningxia	0.00%	*	0.00%
Xinjiang	0.00%	0.00%	0.00%
Jiangxi	0.00%	0.00%	0.00%
Guizhou	0.00%	0.00%	0.00%
Inner Mongolia	-9.13%	0.00%	0.62%
Anhui	0.00%	-9.52%	0.00%
Qinghai	0.00%	*	0.00%
Yunnan	0.00%	0.00%	0.44%

---

162 Note: \* means that the cumulative death is 0, which cannot be predicted.

163

164 It can be seen from Table 2 that the average relative errors of actual cumulative cases, deaths,

165 recovery and predicted values in each province are between -22.32%-22.66%, -9.52%-0.08%,

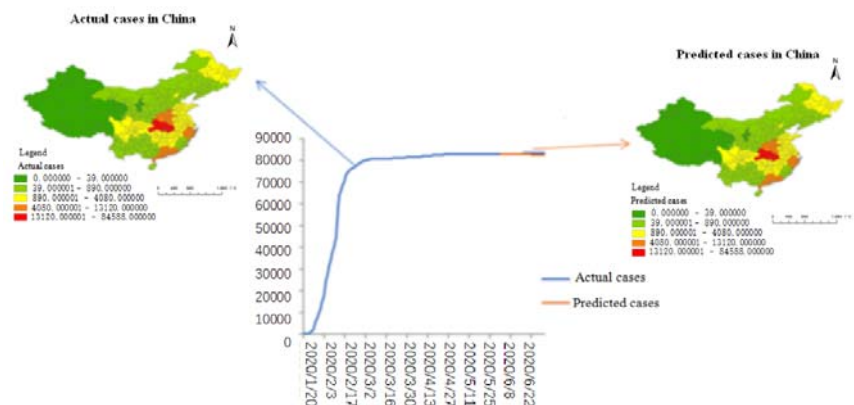
166 -8.84%-1.16%. The results of comprehensive experimental group and test group show that the  
167 error between fitting and prediction is small, the fitting effect is good, the model supports and is  
168 suitable for epidemic situation prediction.

169 COVID-19 trend in mainland of China : According to each provincial model, we get the  
170 overall trend of COVID-19 in mainland of China, as shown in Fig.6. The trend of the fitting line  
171 between the predicted value and the actual value of the cumulative cases, deaths and recovery was  
172 consistent (6a, 6b, 6c). At the beginning of the epidemic, the cases showed an exponential upward  
173 trend. At the end of January, the Chinese government took measures to seal the city, isolate it from  
174 the source of infection, and reduce the possibility of infection, the epidemic was controlled in  
175 mainland of China. The epidemic situation in Hubei, Guangdong, Zhejiang, Henan and other  
176 places in mainland of China is relatively serious (6a). Residents in this area still need to do  
177 epidemic prevention and control work to prevent the possibility of a second rebound of the  
178 epidemic situation.

179 With the increase of confirmed cases, the number of deaths initially gradually increased. It  
180 can be seen from 6b that the number of deaths in mid-April showed a linear upward trend. The  
181 reason for this data surge is that the preliminary data statistics are incomplete, the statistical  
182 standards are inconsistent and there are varying degrees of delay in the statistical process [9]. The  
183 state has proposed efforts to reduce deaths, requiring provinces to increase the number of  
184 ventilators to ensure adequate medical facilities. The number of deaths in mainland of China has  
185 remained stable around May. In this fight against the epidemic, many volunteers have appeared in  
186 many places, and medical personnel throughout the country have conducted targeted research on  
187 the epidemic, and their experience and level have been improved, resulting in the continuous

188 improvement of the number of cures (6c).

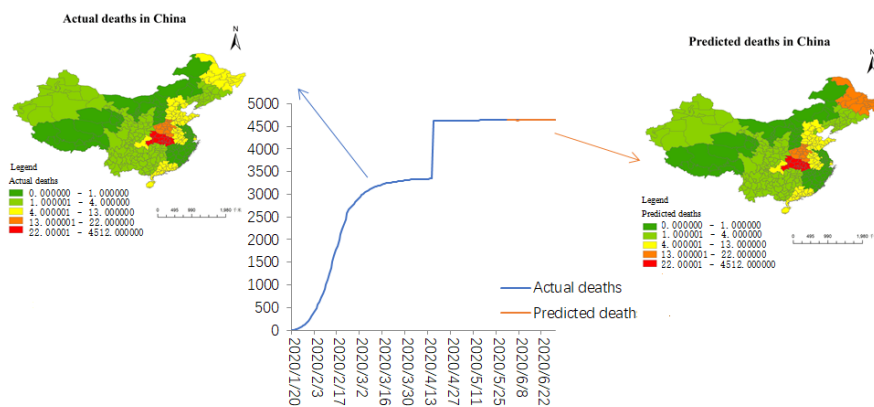
189



190

191

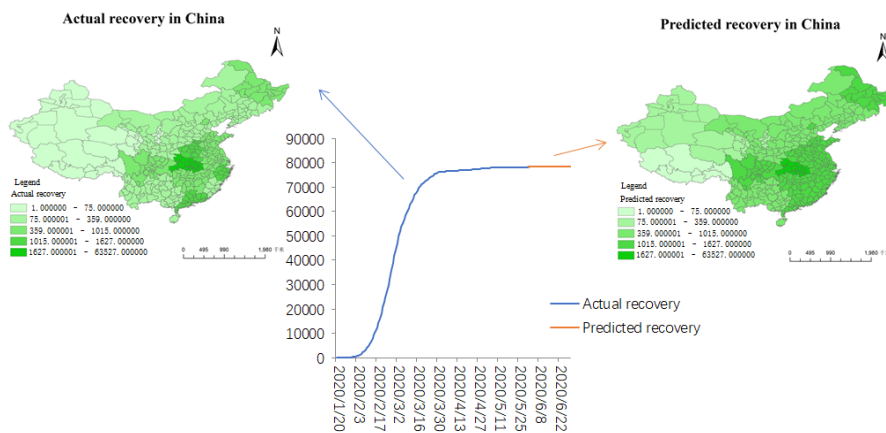
(a)



192

193

(b)



194

195

(c)

196

Fig.6 Cumulative cases, deaths, recovery figure in mainland of China



#### 197 **4.Discussion**

198 Data from January to June 2020 shows that the epidemic in mainland of China has peaked at  
199 the end of February and has been steadily declining since then. This is mainly due to the closed  
200 management measures taken by the Chinese government in the face of the COVID-19 pandemic  
201 to stop population-intensive Place, and achieved good results. By the end of March 2020, all  
202 regions had lifted the closure measures and adjusted the primary response to major public health  
203 emergencies to secondary, and the cases in the epidemic have not changed significantly since then.  
204 This may be related to climate factors such as continuous temperature increase in various places.  
205 The studies have shown that changes in temperature and humidity may be important factors  
206 affecting the spread of COVID-19[10].

207 In recent years, ARIMA model, neural network and so on are more active in the field of  
208 epidemic prediction because they can relatively accurately explore the occurrence and  
209 development of epidemics[11-15].

210 The reason why ARIMA model can be widely favored in the field of epidemiological  
211 prediction is that it not only absorbs the advantages of traditional regression analysis but also takes  
212 advantage of the moving average, and can affect many factors affecting the development of the  
213 epidemic (such as temperature, humidity, aerosol, population migration, etc.) integrated into time  
214 variables for quantitative expression, is a method with strong practicability and high prediction  
215 accuracy. However, ARIMA model also has certain limitations. It is a mathematical model built  
216 on past historical data. Therefore, ARIMA model is only suitable for short-term forecasting. If the  
217 forecasting time is too long, it will increase the forecasting error and affect the forecasting  
218 accuracy[16-18].

219 At present, COVID-19 has become a global epidemic, and the epidemic situation is still  
220 accelerating and has not yet reached its peak. The prevention and control of the epidemic situation  
221 must not be delayed. Based on the COVID-19 data of mainland of China, we found that the model  
222 has a high degree of fitting, which can predict the development trend of the epidemic situation in  
223 the future. However, in practice, other factors[19-22] will also affect the development trend of the  
224 epidemic situation, causing predictions to deviate, this needs further study. In practical  
225 applications, we can make the model have better prediction effect and accuracy by combining  
226 with multiple models.

## 227 **5. Conclusion**

228 We used the epidemic data of Chinese provinces from January 20 to June 31, 2020, in which  
229 the data before May 30 were involved in model fitting, the data of June were tested in the model,  
230 and the ARIMA model was used to fit the data, so as to obtain the epidemic prediction of different  
231 periods in mainland of China. Through the verification of the existing data, the prediction effect is  
232 good. The model can be used to predict the epidemic situation in mainland of China in the future  
233 and make positive contributions to policy makers' prevention and control of the epidemic and  
234 protection of people's lives. On the other hand, the applicability and robustness of the model in  
235 mainland of China can also be studied in other countries and regions, so as to verify the accuracy  
236 and improve the performance of the model and provide assistance for epidemic prevention and  
237 control in this region.

238

239 **Funding:** spatial data analysis and geoscientific calculation are used to study urban vulnerability  
240 based on microgeographic units(grant numbers: 41461035)

241

242 **Acknowledgements:** This work was financially supported by spatial data analysis and  
243 geoscientific calculation are used to study urban vulnerability based on microgeographic  
244 units(grant numbers: 41461035). Thanks to the National and local health and construction  
245 commission for providing data on the daily cumulative cases, deaths and recovery.

246

247 **Declaration of Interest Statement:**No conflict of interest exists in the submission of this  
248 manuscript, and manuscript is approved by all authors for publication. I would like to declare on  
249 behalf of my co-authors that the work described was original research that has not been published  
250 previously, and not under consideration for publication elsewhere, in whole or in part. All the  
251 authors listed have approved the manuscript that is enclosed.

252

253

## 254 **Reference**

- 255 1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan,  
256 China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine* 2020;doi:  
257 10.1056/NEJMoa2001316
- 258 2. Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and  
259 pathogenesis. *Journal of Medical Virology* 2020;doi: 10.1002/jmv.25681
- 260 3. Cyranoski D. What China's coronavirus response can teach the rest of the world. *Nature*  
261 2020;doi: 10.1038/d41586-020-00741-x
- 262 4. Wei W, Jiang J, Liang H, Gao L, Liang B, Huang J, et al. Application of a combined model  
263 with autoregressive integrated moving average (arima) and generalized regression neural

- 264 network (grnn) in forecasting hepatitis incidence in heng county, China. PLoS ONE 2016;doi:  
265 10.1371/journal.pone.0156768
- 266 5. Liu L, Luan RS, Yin F, Zhu XP, Lü Q. Predicting the incidence of hand, foot and mouth  
267 disease in Sichuan province, China using the ARIMA model. *Epidemiology and Infection*  
268 2016;doi: 10.1017/S0950268815001144
- 269 6. Huang L, Li XX, Abe EM, Xu L, Ruan Y, Cao CL, et al. Spatial-temporal analysis of  
270 pulmonary tuberculosis in the northeast of the Yunnan province, People's Republic of China.  
271 *Infectious Diseases of Poverty* 2017;doi: 10.1186/s40249-017-0268-4
- 272 7. Liu Q, Liu X, Jiang B, Yang W. Forecasting incidence of hemorrhagic fever with renal  
273 syndrome in China using ARIMA model. *BMC Infectious Diseases* 2011;doi:  
274 10.1186/1471-2334-11-218
- 275 8. Niemann KO, Quinn G, Goodenough DG, Visintini F, Loos R. Addressing the effects of  
276 canopy structure on the remote sensing of foliar chemistry of a 3-dimensional, radiometrically  
277 porous surface. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote*  
278 *Sensing* 2012;doi: 10.1109/JSTARS.2011.2179637
- 279 9. AHMADI, Mohsen, et al. Investigation of effective climatology parameters on COVID-19  
280 outbreak in Iran. *Science of The Total Environment*,  
281 2020,138705.<https://doi.org/10.1016/j.scitotenv.2020.138705>.
- 282 10. Ahmadi M, Sharifi A, Dorosti S, Jafarzadeh Ghouschi S, Ghanbari N. Investigation of  
283 effective climatology parameters on COVID-19 outbreak in Iran. *Science of the Total*  
284 *Environment* 2020;doi: 10.1016/j.scitotenv.2020.138705
- 285 11. Wah W, Das S, Earnest A, Lim LKY, Chee CBE, Cook AR, et al. Time series analysis of

- 286 demographic and temporal trends of tuberculosis in Singapore. BMC Public Health 2014;doi:  
287 10.1186/1471-2458-14-1121
- 288 12. Du Z, Xu L, Zhang W, Zhang D, Yu S, Hao Y. Predicting the hand, foot, and mouth disease  
289 incidence using search engine query data and climate variables: An ecological study in  
290 Guangdong, China. BMJ Open 2017;doi: 10.1136/bmjopen-2017-016263
- 291 13. Mollalo A, Mao L, Rashidi P, Glass GE. A gis-based artificial neural network model for spatial  
292 distribution of tuberculosis across the continental united states. International Journal of  
293 Environmental Research and Public Health 2019;doi: 10.3390/ijerph16010157
- 294 14. Zou JJ, Jiang GF, Xie XX, Huang J, Yang XB, Lu M. Application of a combined model with  
295 seasonal autoregressive integrated moving average and support vector regression in forecasting  
296 hand-foot-mouth disease incidence in Wuhan, China. Medicine (United States) 2019;doi:  
297 10.1097/MD.00000000000014195
- 298 15. Li Z, Wang Z, Song H, Liu Q, He B, Shi P, et al. Application of a hybrid model in predicting  
299 the incidence of tuberculosis in a Chinese population. Infection and Drug Resistance 2019;doi:  
300 10.2147/IDR.S190418
- 301 16. Wang Y wen, Shen Z zhou, Jiang Y. Comparison of ARIMA and GM(1,1) models for  
302 prediction of hepatitis B in China. PLoS ONE 2018;doi: 10.1371/journal.pone.0201987
- 303 17. Petukhova T, Ojkic D, McEwen B, Deardon R, Poljak Z. Assessment of autoregressive  
304 integrated moving average (ARIMA), generalized linear autoregressive moving average  
305 (GLARMA), and random forest (RF) time series regression models for predicting influenza A  
306 virus frequency in swine in Ontario, Canada. PLoS ONE 2018;doi:  
307 10.1371/journal.pone.0198313

- 308 18. Jackson ML, Chung JR, Jackson LA, Phillips CH, Benoit J, Monto AS, et al. Influenza vaccine  
309 effectiveness in the United States during the 2015–2016 season. *New England Journal of*  
310 *Medicine* 2017;doi: 10.1056/NEJMoa1700153
- 311 19. Cai Y, Huang T, Liu X, Xu G. The effects of “Fangcang, Huoshenshan, and Leishenshan”  
312 hospitals and environmental factors on the mortality of COVID-19. *PeerJ* 2020;doi:  
313 10.7717/peerj.9578
- 314 20. C. Luo et al., Possible Transmission of Severe Acute Respiratory Syndrome Coronavirus 2  
315 (SARS-CoV-2) in a Public Bath Center in Huai'an, Jiangsu Province, China. *JAMA Netw Open*  
316 3, e204583
- 317 21. Lipsitch M. Seasonality of SARS-CoV-2: Will COVID-19 go away on its own in warmer  
318 weather? *Center for Communicable Disease Dynamics Harvard Blog* 2020;doi:  
319 10.1017/CBO9781107415324.004
- 320 22. Oliveiros B, Caramelo L, Ferreira NC, Caramelo F. Oliveiros et al 2020 Temperature and  
321 humidity in doubling time of cases. 2020. doi: 10.1101/2020.03.05.20031872