Modelling and forecasting the spread tendency of the COVID-19 in China

Deshun Sun
Shenzhen Second People's Hospital

Li Duan
Shenzhen Second People's Hospital

Jianyi Xiong (✉ jianyixiong@126.com)
Shenzhen Second People's Hospital

Daping Wang (✉ wangdp@mail.sustech.edu.cn)
Shenzhen Second People's Hospital   https://orcid.org/0000-0002-2402-8134

Modelling and forecasting the spread tendency of the COVID-19 in China

Deshun Sun[1†]   Li Duan[1†]   Jianyi Xiong[1*]   Daping Wang[1, 2*]

[1]Shenzhen Key Laboratory of Tissue Engineering, Shenzhen Laboratory of Digital Orthopedic Engineering, Guangdong Provincial Research Center for Artificial Intelligence and Digital Orthopedic Technology, Shenzhen Second People's Hospital (The First Hospital Affiliated to Shenzhen University, Health Science Center), Shenzhen, 518035, P. R. China

[2]Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen, 518055, P. R. China

(†) These authors contributed equally to this study.

(*) The correspondence of this work is affiliated to Professor Jianyi Xiong and Professor Daping Wang

Prof. Jianyi Xiong, Shenzhen Second People's Hospital (The First Hospital Affiliated to Shenzhen University, Health Science Center), Shenzhen 518035, P. R. China. E-mail: jianyixiong@126.com ; Tel: 0755-83366338.

Prof. Daping Wang, Shenzhen Second People's Hospital (The First Hospital Affiliated to Shenzhen University, Health Science Center), Shenzhen 518035, P. R. China. E-mail: wangdp@mail.sustech.edu.cn; Tel: 0755-83464301.

Abstract

To forecast the spread tendency of the COVID-19 in China and provide effective strategies to prevent the disease, an improved SEIR model was established. Parameters of our model were estimated based on collected data that issued by the National Health Commission of China (NHCC) from January 10 to March 3. The model was used to forecast the spread tendency of the disease. The key factors influencing the epidemic were explored through modulation of parameters, including the removal rate, the

average number of the infected contacting the susceptible per day and the average number of the exposed contacting the susceptible per day. The correlation of the infected is 99.9% between established model data in this study and issued data by NHCC from January 10 to February 15. The correlation of the removed is 99.8%. The average forecasting error rates of the infected and the removed are 0.78% and 0.75%, respectively, from February 16 to March 3. The peak time of the epidemic forecast by our established model coincided with the issued data by NHCC. Therefore, our study established a mathematical model with high accuracy. The aforementioned parameters significantly affected the trend of epidemic, suggesting that the exposed and the infected population should be strictly isolated. If the removal rate increases to 0.12, the epidemic will come to an end on May 25. In conclusion, the proposed mathematical model accurately forecast the spread tendency of COVID-19 in China and the model can be applied for other countries with appropriate modifications.

Keywords : COVID-19; Mathematical modeling; Parameter estimation; Forecasting; Control strategy

1. Background

COVID-19 is an emergent infectious disease caused by the novel coronavirus 2019-nCoV. It was first detected in Wuhan, China in December of 2019[1]. From COVID-19 outbreak in Wuhan, it has spread across China and beyond. By March 3, 2020, 80270 infection cases have been confirmed and 2981 death cases have been reported. Patients with confirmed COVID-19 are the main sources of infection, and asymptomatic patients are also potential sources of 2019-nCoV infection. Respiratory droplets and contact transmission are considered to be the most important routes of transmission of 2019-nCoV. The population is generally susceptible to the infection disease[2], and the elderly and those with chronic diseases are most susceptible to severe forms of COVID-19 [3, 4]. Multiple strategies have been developed to fight against the spread of COVID-19, including strict isolation, early diagnosis and supporting treatment. Currently, there is no specific medicine to combat the novel coronavirus. Therefore, it is a crucial step to forecast the spread tendency of the acute

infectious disease based on the epidemiology data. Mathematical modeling is one of the most effective methods for forecasting of infectious disease outbreaks and thus yield valuable insights suggest how future efforts may be improved. An important method for epidemiological studies of such acute infectious diseases is mathematical modeling.

Since the epidemic outbreak, some scholars have established mathematical model to forecast the spread of COVID-19 [5-14]. Wu *et al.* [9] calculated that the basic reproductive number of new pneumonia (R_0) is 2.68 based on the established susceptible–exposed–infected–removed (SEIR) model, and forecast that the number of infected people in Wuhan would be 75,815 on January 25, 2020. What's more, the authors also forecast the number of infected people imported from Wuhan to Chongqing, Beijing, Shanghai, Guangzhou and Shenzhen. However, the number of infections reported in this study is inconsistent with the number issued by NHCC (1870 cases), and the difference is large. Zhou *et al.* [10] used the SEIR compartment model to characterize the early spreading of COVID-19, and forecast basic reproduction number (R_0) is between 2.2 and 3.0. However, the model hasn't considered that the susceptible can be infected by confirmed patients during the incubation period that can't be ignored. Besides, Tang *et al.* [5] established a complicated model and forecast that the control reproduction number (R_0) may be as high as 6.47 (95% CI 5.71-7.23). Using their estimated parameter values, the number of the infected will reach the peak around March 10, 2020 and the peak number of the infected is 1.63×10^5. However, the peak number issued by NHCC is 58016 and the time of reaching the peak is February 17. Therefore, a more accurate mathematical model is highly anticipated to forecast the spread tendency of COVID-19.

In this study, we re-established a SEIR model of COVID-19 based on the its transmission mechanism. The aim of this study was to obtain a more accurate mathematical model to forecast the spread tendency of epidemic and provide guidance to control the spread of COVID-19.

2. Methods and data

2.1 Theoretical mathematical model

The most classical model for studying infectious diseases is the SIR compartment model, which was proposed by Kermack and McKendrick in 1927 when studying the black plague[15]. SIR model divides population into three categories: Susceptible, which means the uninfected persons who lack immunity; Infective, which means the persons who are capable of spreading the disease to a susceptible person; Removed, which means the dead or healed [16]. The basic model is as following:

$$
\begin{aligned}
\frac{ds(t)}{dt} &= -\beta i(t)s(t), \\
\frac{di(t)}{dt} &= \beta i(t)s(t) - \gamma i(t), \\
\frac{dr}{dt} &= \gamma i(t).
\end{aligned}
\tag{1}
$$

where $s(t)$, $i(t)$ and $r(t)$ are the susceptible, the infected and the removed, respectively. β is infection rate and γ is removal rate.

Gradually, researchers realized that the incubation period-the time between exposure and the start of symptoms, should be taken into consideration [17-19]. Therefore, the classic SEIR model was modified as following:

$$
\begin{aligned}
\frac{dS(t)}{dt} &= \lambda - \beta S(t)I(t) - dS(t), \\
\frac{dE}{dt} &= \beta S(t)I(t) - (d+\alpha)E(t), \\
\frac{dI}{dt} &= \alpha E(t) - (d+\gamma+\delta)I(t), \\
\frac{dR}{dt} &= \gamma I(t) - dR(t).
\end{aligned}
\tag{2}
$$

According to the classic SEIR model, this study divided the population into four groups: susceptible (S), exposed (E), infected (I) and removed (R). As the susceptible can be infected by the exposed and COVID-19 confirmed infection, we proposed the following transmission mode (Figure 1):
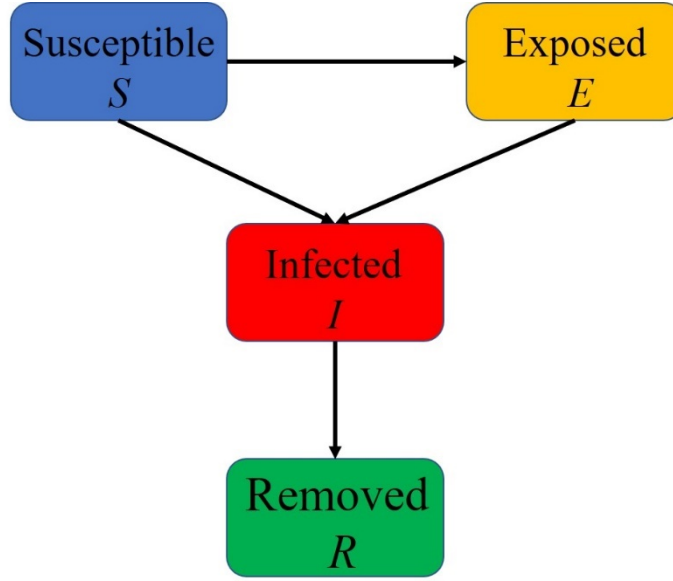
Figure 1 The transmission mechanism of COVID-19.

In this epidemic, the natural mortality of newborns and the natural death of all population were ignored for a short period of time. Because the patients can infect the susceptible during the incubation period, we established the following mathematical model in this study:

$$\frac{dS(t)}{dt} = -\frac{r\beta S(t)I(t)}{N(t)} - \frac{r_2\beta_2 S(t)E(t)}{N(t)},$$

$$\frac{dE}{dt} = \frac{r\beta S(t)I(t)}{N(t)} + \frac{r_2\beta_2 S(t)E(t)}{N(t)} - \delta E(t),$$

$$\frac{dI}{dt} = \delta E(t) - \gamma I(t), \tag{3}$$

$$\frac{dR}{dt} = \gamma I(t).$$

$$N(t) = S(t) + E(t) + I(t) + R(t)$$

where $S(t)$, $E(t)$, $I(t)$ and $R(t)$ are the susceptible, the exposed, the infected and the removed, respectively. r represents the average number of the infected contacting the susceptible per day, β represents the probability of infection by the infected. r_2 represents the average number of the exposed contacting the susceptible per day, β_2 represents the probability of infection by the exposed. δ is the probability of the exposed to become the infected. γ is the removal rate which includes the cure rate and death rate. $N(t)$ is the total population and $N(t) = S(t) + E(t) + I(t) + R(t)$.

2.2 Data

Next, we estimated the parameters based on the number of the infected, the cured and deaths that issued by NHCC every day at 12 pm to obtain an accurate model. Based on the data issued by the NHCC from January 10, 2020 to March 3, 2020, the statistical results are shown in Table 1:

Table 1 The number of the infected, cured, deaths and removed.

Date	Time point	Total infected	Cured	Death	Infected[1]	Removed[2]
2020.01.10	1	41	2	1	38	3
2020.01.11	2	41	6	1	34	7
2020.01.12	3	41	7	1	33	8
2020.01.13	4	41	7	1	33	8
2020.01.14	5	41	7	1	33	8
2020.01.15	6	41	12	2	27	14
2020.01.16	7	41	15	2	24	17
2020.01.17	8	62	19	2	41	21
2020.01.18	9	121	24	3	94	27
2020.01.19	10	198	25	3	170	28
2020.01.20	11	291	25	3	263	28
2020.01.21	12	440	28	9	403	37
2020.01.22	13	571	30	17	524	47
2020.01.23	14	830	34	25	771	59
2020.01.24	15	1287	38	41	1208	79
2020.01.25	16	1975	49	56	1870	105
2020.01.26	17	2744	51	80	2613	131
2020.01.27	18	4515	60	106	4349	166
2020.01.28	19	5974	103	132	5739	235
2020.01.29	20	7711	124	170	7417	294
2020.01.30	21	9692	171	213	9308	384
2020.01.31	22	11791	243	259	11289	502
2020.02.01	23	14380	328	304	13748	632
2020.02.02	24	17205	475	361	16369	836
2020.02.03	25	20438	632	425	19381	1057
2020.02.04	26	24324	892	490	22942	1382
2020.02.05	27	28018	1153	563	26302	1716
2020.02.06	28	31161	1540	636	28985	2176
2020.02.07	29	34546	2050	722	31774	2772
2020.02.08	30	37198	2649	811	33739	3460

2020.02.09	31	40171	3281	908	35982	4189
2020.02.10	32	42638	3996	1016	37626	5012
2020.02.11	33	44653	4740	1113	38800	5853
2020.02.12	34	59804	5911	1367	52526	7278
2020.02.13	35	63851	6723	1380	55748	8103
2020.02.14	36	66492	8096	1523	56873	9619
2020.02.15	37	68500	9419	1665	57416	11084
2020.02.16	38	70548	10844	1770	57934	12614
2020.02.17	**39**	**72436**	**12552**	**1868**	**58016**	**14420**
2020.02.18	40	74185	14376	2004	57805	16380
2020.02.19	41	74576	16155	2118	56303	18273
2020.02.20	42	75465	18264	2236	54965	20500
2020.02.21	43	76288	20659	2345	53284	23004
2020.02.22	44	76936	22888	2442	51606	25330
2020.02.23	45	77150	24734	2592	49824	27326
2020.02.24	46	77658	27323	2663	47672	29986
2020.02.25	47	78064	29745	2715	45604	32460
2020.02.26	48	78497	32495	2744	43258	35239
2020.02.27	49	78824	36117	2788	39919	38905
2020.02.28	50	79251	39002	2835	37414	41837
2020.02.29	51	79824	41625	2870	35329	44495
2020.03.01	52	80026	44462	2912	32652	47374
2020.03.02	53	80151	47204	2943	30004	50147
2020.03.03	54	80270	49856	2981	27433	52837

([1] Infected=Total infected-cured-death; [2] Removed=cured+death)

2.3 Estimation of model parameters

In order to get the value of the parameters in equation (3) based on data issued by NHCC, algorithm (*fmincon* and *lsqnonlin* function) was programmed to estimate the 6 parameters by matlab 2017b. In this study, all the parameters were defined to be nonnegative and bounded, because each parameter has its own significance. Secondly, based on real data, *fmincon* function was employed to estimate the approximate range of each parameter. Estimated parameters by the *fmincon* function were regarded as the initial values. Thereafter, further estimation was performed using the *lsqnonlin* function to achieve the best fitting effect between the simulation curve and the real data curve.

2.4 Average forecasting error rate

The equation of average forecasting error rate (AFER) is as following:

$$average\ error\ rate = \frac{1}{n} \cdot \frac{\sum_{i=1}^{n}|x_i - y_i|}{x_i} \times 100\% \tag{4}$$

where x_i is the real value, y_i is the forecasting value, n is the number of all data which need to be forecast.

3. Results

3.1 Parameters estimation and forecast

Based on Table 1, the method of data-driven modeling was adopted [20, 21]. In order to combat this epidemic, Chinese government has adopted a series of measures, such as the establishment of Huoshenshan Hospital, Leishenshan Hospital, Fangcang Hospital, Wuhan quarantine, home isolation, and sending detachment of medical personnel. The simulation was divided into nine stages. The detailed steps are as follows.

Stage 1: from January 10, 2020 to January 21, 2020 (1th day to 12th day)

From January 10, 2020 to January 21, 2020, there were no quarantine measures. The initial values of the total population, the exposed, the infected, the removed and the susceptible are 1.4×10^9, 0, 41, 2 and ($1.4 \times 10^9 - 41 - 2$), respectively. The estimated average number of the infected contacting the susceptible per day (r) is 20, the estimated infection rate by the infected (β) is 0.043. Therefore, $\beta \cdot r = 0.86$ is similar to the reported literature [22]. The average number of the exposed contacting the susceptible per day (r_2) is 20, and because the probability of infection by the exposed is lower than that of the confirmed infected patients, we set $\beta_2 = 0.025$. The probability that an exposed person turns into a confirmed infected patient (δ) is 0.079, which is similar to the reported literature[21]. The removal rate (γ) is 0.001, and is consistent with the reported literature [14].

Stage 2: from January 22, 2020 to February 3, 2020 (13th day to 25th day)

On January 22, 2020, Nanshan Zhong, a prominent Chinese physician led a

panel of experts to Wuhan. Since he confirmed that the virus can spread from person to person, the government began to perform isolation policy. However, because of the limited number of beds and isolation, r became 2 and r_2 became 7. At the same time, the government has sent some medical teams, and the removal rate (γ) has been improved to 0.0075. δ was 0.079. In the following stages, if not specified, the parameters are the same as the former stage. Measures have been taken on stage 2 was defined as intervention 1.

Stage 3: from February 4, 2020 to February 6, 2020 (26th day to 28th day)

Since February 4, 2020, as the government further improved the diagnosis efficiency and strengthened the isolation of communities, r was 0.1, r_2 was 2, γ was 0.02. Measures have been taken on stage 3 was defined as intervention 2.

Stage 4: from February 7, 2020 to February 11, 2020 (29th day to 33th day)

From February 7, 2020 to February 11, 2020, the parameters were: $r = 0.01$; β=0.0043; $r_2 = 1$; β_2=0.0025; γ=0.025; on February 12, 2020, δ=0.05, γ=0.03.

Stage 5: from February 12, 2020 to February 14, 2020 (34th day to 36th day)

From February 12, 2020, diagnostic criteria have changed, and a positive nucleic acid test is not the only criterion. The government also added characteristic CT imaging patterns to the confirmed cases, so the number of infected cases was 52626 on February 12, 2020. Therefore, we revised the number of infected patients in our model. The parameters were: r=0.01; r_2=2; γ=0.023, δ=0.12.

Stage 6: from February 15, 2020 to February 17, 2020 (37th day to 39th day)

From February 15, 2020, to February 17, 2020, $r = 0.01$; β=0.0043; $r_2 = 1$; β_2=0.0025; γ=0.03; δ=0.1.

Stage 7: February 18, 2020 to February 23, 2020 (40th to 45th)

From February 18, 2020, to February 23, 2020, $r = 0.01$; β=0.0043; $r_2 = 0.5$; β_2=0.0025; γ=0.04; δ=0.04.

Stage 8: February 24, 2020 to February 28, 2020 (46th to 50th)

From February 24, 2020 to February 28, 2020, $r=0.01$; $\beta=0.0043$; $r_2=0.1$; $\beta_2=0.0025$; $\gamma=0.04$; $\delta=0.07$.

Stage 9: February 29, 2020 to March 3, 2020 (51th to 54th)

From February 29, 2020 to March 3, 2020, $r=0.01$; $\beta=0.0043$; $r_2=0.1$; $\beta_2=0.0025$; $\gamma=0.02$; $\delta=0.085$.

From March 4, 2020, with the increase of removal rate and greater isolation strength, the parameter γ will be larger, r and r_2 will be smaller.

3.2 Correlation and average forecasting error rate

Based on the aforementioned nine stages, the date January 10, 2020 was regarded as the starting point, and the data from January 10 to February 15, 2020 was regarded as the training set for model parameter estimation. The R_I between established model data in this study and issued data by NHCC is 99.9% and the R_R is 99.8%. The data from February 16, 2020 to March 3, 2020 were used to forecast and verify the model. The data of the number of the infected of model forecast and issued by NHCC were shown in Table 2. Therefore, the AFER of the infected was 0.78%. Similarly, the AFER of the removed was 0.75% (see Table 2).

Table 2 The forecasting infected and removed from February 16 to March 3.

Time	Day	Infected	Forecasting Infected	Removed	Forecasting Removed
2020.02.16	38	57934	58217	12614	12684
2020.02.17	39	58016	58338	14420	14431
2020.02.18	40	57805	58274	16380	16181
2020.02.19	41	56303	56552	18273	18512
2020.02.20	42	54965	54875	20500	20774
2020.02.21	43	53284	53242	23004	22969
2020.02.22	44	51606	51654	25330	25099
2020.02.23	45	49824	50107	27326	27165
2020.02.24	46	47672	48603	29986	29169
2020.02.25	47	45604	45682	32460	32572
2020.02.26	48	43258	42945	35239	35769
2020.02.27	49	39919	40383	38905	38775
2020.02.28	50	37414	37982	41837	41602
2020.02.29	51	35329	35732	44495	44261

2020.03.01	52	32652	32891	47374	47298
2020.03.02	53	30004	30288	50147	50094
2020.03.03	54	27433	27902	52837	52668

3.3 The dynamic trends of the susceptible, the exposed, the infected and the removed

Based on the above nine stages, the dynamic trends of the susceptible, exposed, infected and removed for 54 days (from January 10, 2020 to March 3, 2020) were simulated. Figure 2 showed the dynamic trends of the susceptible where the initial population is $(1.4\times10^9 - 41 - 2)$ and has a sharp drop from January 19, 2020 to February 5 and remain relatively stable. The dynamic trends of the exposed showed the number of the exposed increased consistently and reached its peak on February 4 and began to gradually decrease (see Figure 3). The dynamic trends of the infected showed that when the government took intervention 2, the growth rate of the infected population decreased significantly. When diagnostic criteria have changed, namely, the government added characteristic CT imaging patterns to the confirmed cases, there was a sudden increase of the number of the infected. The number of people infected reached the peak on February 17 and gradually began to decrease (see Figure 4). The dynamic trends of the removed showed that the number of the removed continued to increase which is consistent with the data issued by NHCC.

Figure 2 Dynamic trends of the susceptible.

Figure 3 Dynamic trends of the exposed.

Figure 4 Dynamic trends of the infected.

The blue line represents the simulation of SEIR model. The red line represents the data issued by NHCC. The green line represents the intervention 1. The light blue line represents the intervention 2. The black line represents the government began to implement new diagnostic criteria.
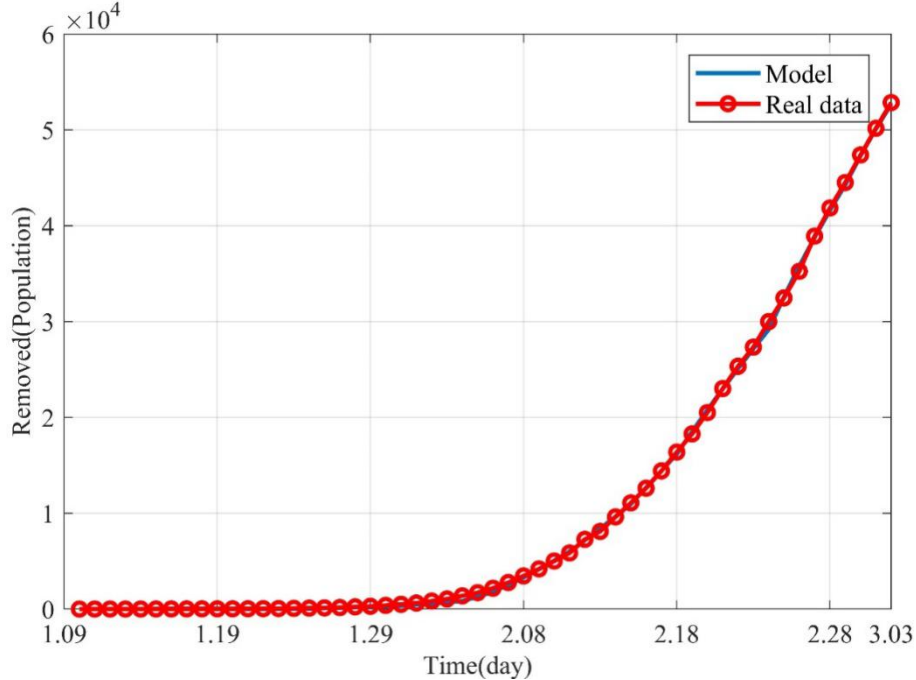

Figure 5 Dynamic trends of the removed.

The removed consists of the cures and the death.The blue line represents the simulation of SEIR model.The red line represents the data issued by NHCC.

3.4 Influence of parameters on epidemic

13

3.4.1 Influence of removal rate on epidemic (γ)

From the 55th day (March 4, 2020), the influence of the removal rate (γ) on the infected and the removed were investigated on condition of changing the removal rate (γ) and fixing other parameters. γ was gradually increased from 0.02 to 0.12. As shown in Figure 6, the dynamic trends of the infected with different removal rates suggested the larger the removal rate is, the better the control effect is. when $\gamma=0.02$, 200 days later (July 29), there are still large numbers of infected individuals, while $\gamma=0.12$, 135 days later (May 25), there will be no infected individuals (see Figure 6). Similarly, if $\gamma=0.02$, 200 days later (July 29), the number of the removed individuals increases, namely, there are still many patients that need to be treated. If $\gamma=0.12$, the number of the removed become stable (Figure 7). Therefore, it indicates that if the removal rate can be improved, the epidemic will be effectively controlled and terminated earlier. From Figure 7, we also get the conclusion that the larger the removal rate is, the shorter the time it takes for the removed to become stable. This also corresponds to the shorter time to reach the stable point in Figure 6. Since the government sent a large number of medical teams to support Wuhan, the epidemic was effectively controlled and the number of the infected decreased, which is consistent with our forecasting results.

Figure 6 Dynamic trends of the infected with different removal rates (γ).
The solid line represents the trends of the infected with the current γ. The purple dotted line represents the trends of the infected when $\gamma=0.12$. The orange dotted li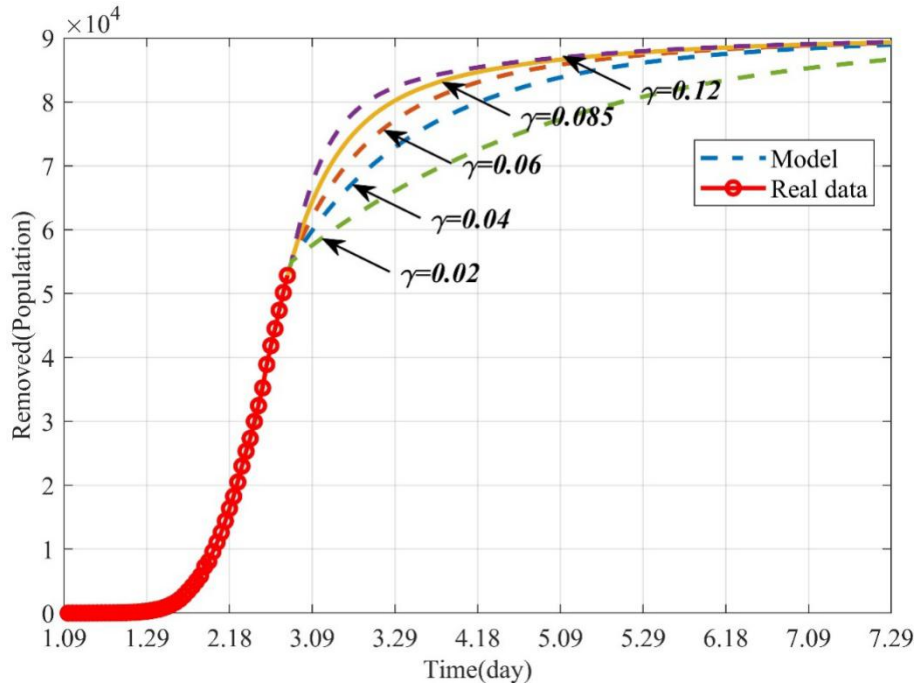ne represents the trends of the infected when $\gamma=0.06$. The blue dotted line represents the trends of the infected when $\gamma=0.04$. The green dotted line represents the trends of the infected when $\gamma=0.02$.

Figure 7 Dynamic trends of the removed with different removal rates (γ).
The solid line represents the trends of the removed with the current γ. The purple

dotted line represents the trends of the removed when $\gamma=0.12$. The orange dotted line represents the trends of the removed when $\gamma=0.06$. The blue dotted line represents the trends of the removed when $\gamma=0.04$. The green dotted line represents the trends of the removed when $\gamma=0.02$.

3.4.2 Influence of the average number of the infected contacting the susceptible per day (r) on the epidemic

This section studied how the average number of the infected contacting the susceptible per day (r) influences the infected and the removed when other parameters keep fixed. The epidemic can be well controlled when $r=0.01$, and finally the infected individuals will disappear. But if the parameter r gradually increases from 0.01 to 5, the number of the infected will increase, and the epidemic will be out of control (Figure 8). Figure 9 showed trends of the removed with different r. If $r=0.01$, 160 days later (June 18), the removed will not increase and remain stable. But if the isolation rate $r \geq 3$, it will take a long time for the epidemic to be effectively controlled, or even out of control. For example, if $r=5$, the number of the infected will increase continually. This is because more persons will be infected and need to be treated. Therefore, timely isolation of the infected and close contacts, and establishment of Huoshenshan Hospital, Leishenshan Hospital and Fangcang Hospital will be very effective.

Figure 8 Dynamic trends of the infected with different r.

The solid line represents the trends of the infected with the current r. The orange dotted line represents the trends of the infected when $r=1$. The yellow dotted line represents the trends of the infected when $r=3$. The purple dotted line represents the trends of the infected when $r=5$.
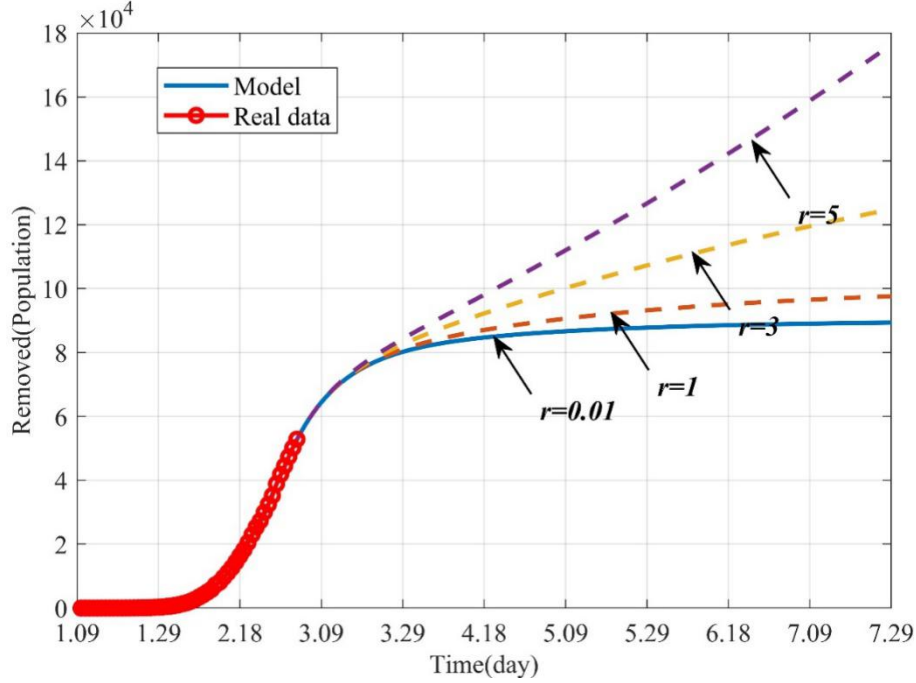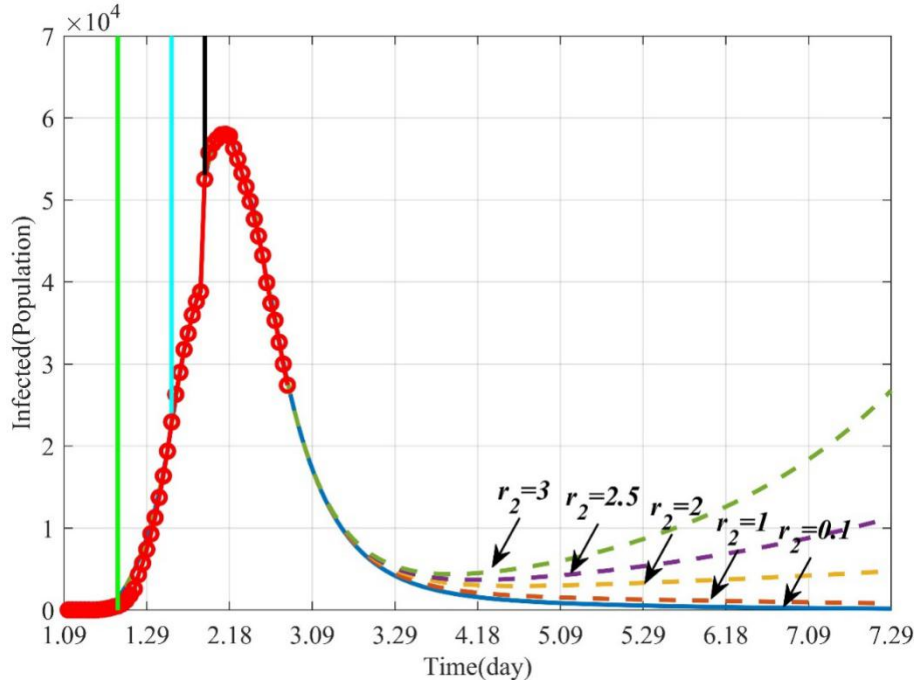

Figure 9 Dynamic trends of the removed with different r.

The solid line represents the trends of the removed with the current r. The orange dotted line represents the trends of the removed when $r=1$. The yellow dotted line represents the trends of the removed when $r=3$. The purple dotted line represents the trends of the removed when $r=5$.

3.4.3 Influence of the average number of the exposed contacting the susceptible per day (r_2) on the epidemic

In this section, the influence of the average number of the exposed contacting the susceptible per day (r_2) on the epidemic was studied by numerical simulation. When $r_2=0.1$ and other parameters keep fixed, 160 days later (June 18), the infected will disappear. If the parameter r_2 increases from 0.1 to 3, the epidemic will become uncontrollable. For example, when $r_2=3$, the number of infections will decrease on the beginning and then increase gradually, and the epidemic will outbreak again (see Figure 10). Similarly, more patients will need to be treated (see Figure 11).

Figure 10 Dynamic trends of the infected with different r_2.

The solid line represents the trends of the infected with the current r_2. The orange dotted line represents the trends of the infected when $r_2=1$. The yellow dotted line represents the trends of the infected when $r_2=2$. The purple dotted line represents the trends of the infected when $r_2=2.5$. The green dotted line represents the trends of the infected when $r_2=3$.
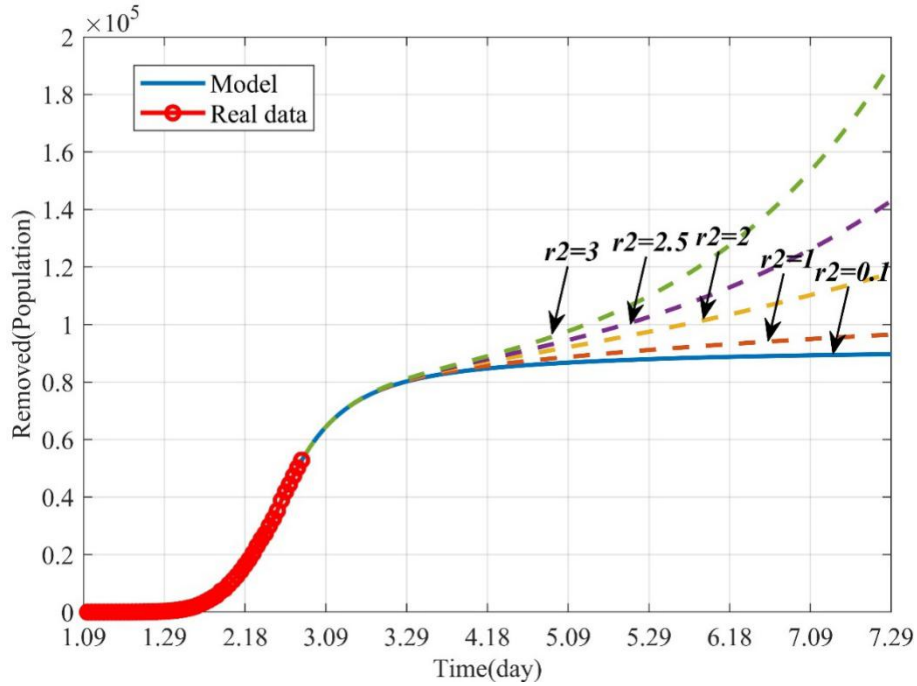

Figure 11 Dynamic trends of the removed with different r_2.

The solid line represents the trends of the removed with the current r_2. The orange dotted line represents the trends of the removed when $r_2=1$. The yellow dotted line represents the trends of the removed when $r_2=2$. The purple dotted line represents the trends of the removed when $r_2=2.5$. The green dotted line represents the trends of the removed when $r_2=3$.

4. Discussion

Due to the rapid spread of COVID-19 and no vaccine or effective treatment available for the epidemic, it has been declared by the World Health Organization as an "international public health emergency". Even the Chinese government has exerted massive efforts to fight against the epidemic COVID-19, there are still new cases of infection every day. In order to scientifically forecast the spread tendency of the disease, we established a mathematical model on COVID-19. The forecasting accuracy of our model has been confirmed by the data issued by NHCC.

In this study, we first modified a classical SEIR model and established a mathematical model according to the possible transmission mechanism of COVID-19. Based on the official data issued by NHCC, the model parameters were estimated and the trends of the infected and the removed populations were forecast in the short term. The average forecasting error rates of the infected and the moved were 0.78% and 0.75%, respectively. Next, series of parameters on this epidemic were studied, including the influences of removal rate (γ), the average number of the infected contacting the susceptible per day (r) and the average number of the exposed contacting the susceptible per day (r_2). Compared with the reported mathematical models, the main advantages of our model are that the correlation of the infected (R_I) between model data and issued data by NHCC is as high as 99.9%, the correlation of the removed (R_R) is as high as 99.8% and the AFER of the infected and the removed are as low as 0.78% and 0.75%, respectively. Therefore, the forecast data derived from our model are approximately validated by the real data issued by NHCC.

We further explored how parameters influence the epidemic by numerical simulations. The results showed that when increase the removal rate (γ) and keep other

parameters fixed, the time to control the epidemic will be greatly shortened. For example, if increase the removal rate (γ) from 0.02 to 0.12 and keep other parameters fixed, the disappeared time of the epidemic will be advanced from July 29 to May 25. This theoretical induction has been proved as the number of the cured increased, since government dispatched medical teams to Hubei province.

Besides, if the average number of the infected contacting the susceptible per day (r) can be effectively reduced, the epidemic can be controlled earlier. For instance, if r is reduced from 5 to 0.01, the epidemic will disappear by May 25, and if $r=5$, the number the infected will be increasing. In addition, if the average number of the exposed contacting the susceptible per day r_2 is 3, the epidemic will outbreak again, and if r_2 is controlled to be less than 0.1, the epidemic will terminate soon. These results have been proved as the number of the infected reached the peak and then decreased, since government has demonstrated an unprecedented level of efforts in dealing with the COVID-19, such as to set up specialized hospitals for nCoV patients, namely Huoshenshan and Leishenshan hospital, Fangcang Hospital.

In conclusion, our established mathematical model can provide theoretical guidance for effective prevention and control of the epidemic COVID-19 in China. With appropriate modifications, it could be applied for other countries currently attacked by the epidemic.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Competing interests

The authors declare there is no conflicts of interest regarding the publication of this paper.

Authors' contributions

Deshun Sun and Li Duan established the model and finished the manuscript, Jianyi Xiong and Daping Wang revised the manuscript.

References :

1.   Huang, C., et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 2020. **395**(10223): p. 497-506.

2.   China, New coronavirus pneumonia diagnosis and treatment plan (trial version 7). 2020.

3.   Chen, N., et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet, 2020. **395**(10223): p. 507-513.

4.   Yang, Y., et al., Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. medRxiv, 2020: p. 2020.02.10.20021675.

5.   Tang, B., et al., Estimation of the Transmission Risk of 2019-nCov and Its Implication for Public Health Interventions. Social Science Electronic Publishing.

6.   Read, J.M., et al., Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. medRxiv, 2020: p. 2020.01.23.20018549.

7.   Li, Q., et al., Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine, 2020.

8.   Shen, M., et al., Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. bioRxiv, 2020: p. 2020.01.23.916726.

9. Wu, J.T., K. Leung, and G.M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet, 2020. **395**(10225): p. 689-697.

10. Zhou, T., et al., Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. Journal of Evidence-Based Medicine. **n/a**(n/a).

11. Liu, T., et al., Transmission dynamics of 2019 novel coronavirus (2019-nCoV). bioRxiv, 2020: p. 2020.01.25.919787.

12. Chen, Y., et al., A Time Delay Dynamical Model for Outbreak of 2019-nCoV and the Parameter Identification. Conference Proceedings, 2020.

13. Zhao, S., et al., Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. bioRxiv, 2020: p. 2020.01.23.916395.

14. Zhang, C. and M. Wang, Origin time and epidemic dynamics of the 2019 novel coronavirus. bioRxiv, 2020: p. 2020.01.25.919688.

15. Jin, Z., Mathematical Modeling and Research of Infectious Disease Dynamics. 2004: Science Press.

16. Tapaswi, P.K. and J. Chattopadhyay, Global stability results of a "susceptible-infective-immune-susceptible" (SIRS) epidemic model. Ecological Modelling, 1996. **87**(1): p. 223-226.

17. Almeida, R., Analysis of a fractional SEIR model with treatment. Applied Mathematics Letters, 2018. **84**: p. 56-62.

18. Diaz, P., et al., A modified SEIR model for the spread of Ebola in Western Africa and metrics for resource allocation. Applied Mathematics and Computation, 2018. **324**: p. 141-155.

19. Sun, C. and Y.-H. Hsieh, Global analysis of an SEIR model with varying population size and vaccination. Applied Mathematical Modelling, 2010. **34**(10): p. 2685-2697.

20. Pei, S., et al., Forecasting the spatial transmission of influenza in the United States. Proceedings of the National Academy of Sciences of the United States of America, 2018. **115**(11): p. 2752.

21. Al, G.C.E., Model Parameters and Outbreak Control for SARS - Volume 10, Number 7—July 2004 - Emerging Infectious Disease journal - CDC.

22. Ai, L., Modelling the epidemic trend of the 2019-nCOV outbreak in Hubei Province, China. medRxiv, 2020: p. 2020.01.30.20019828.
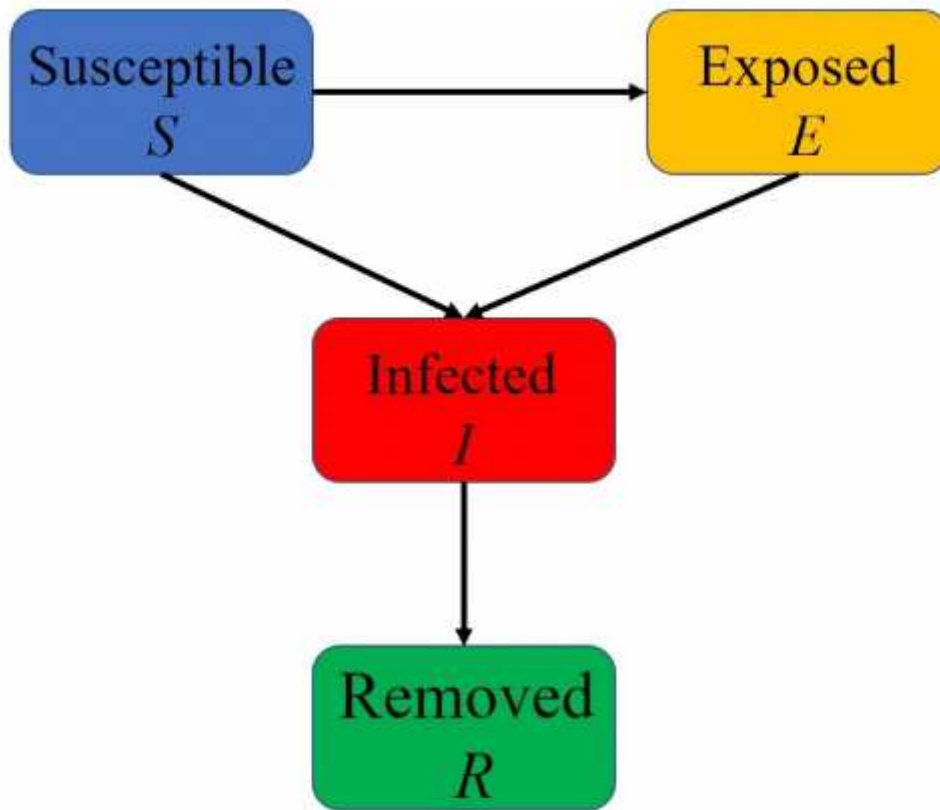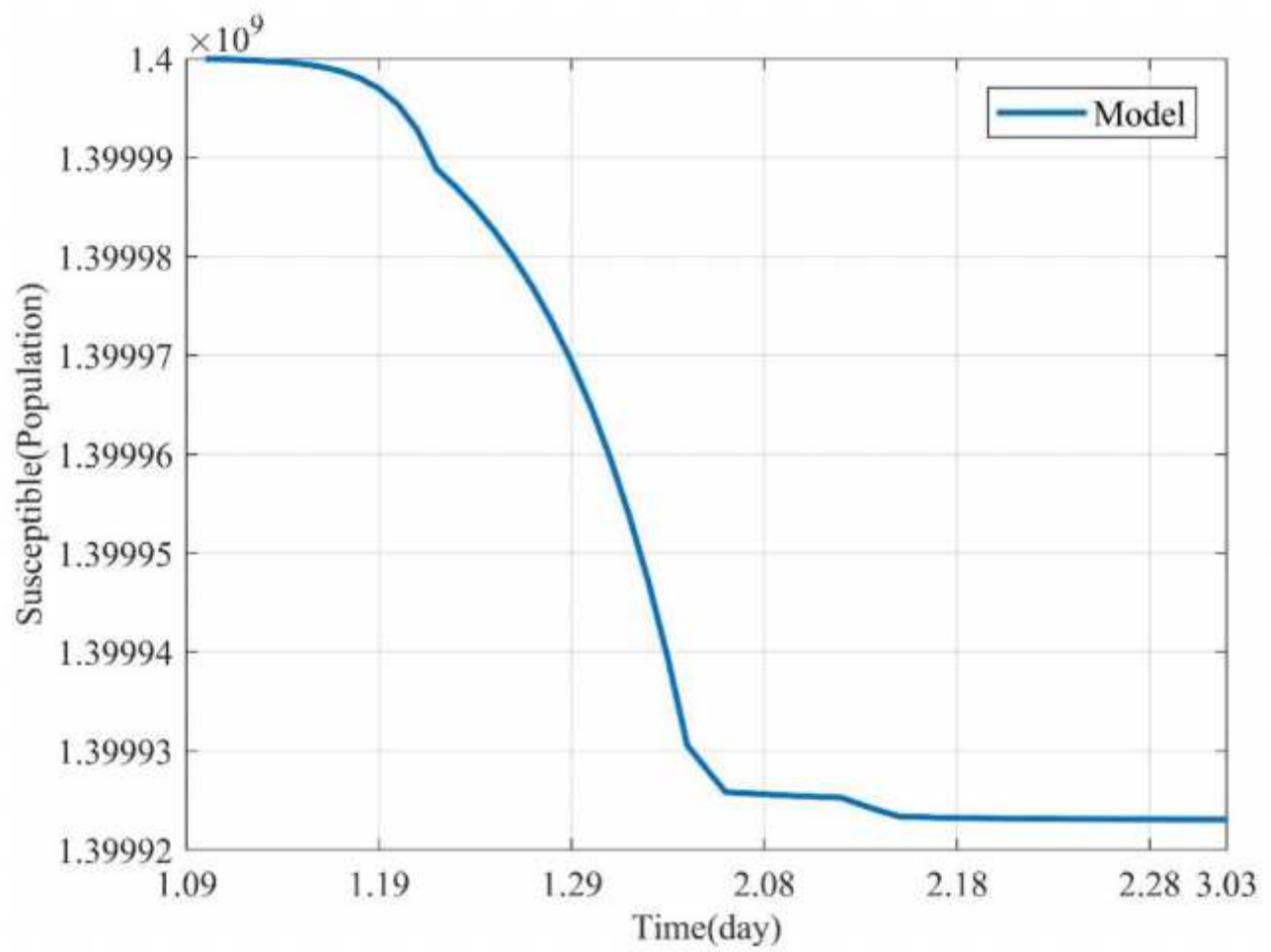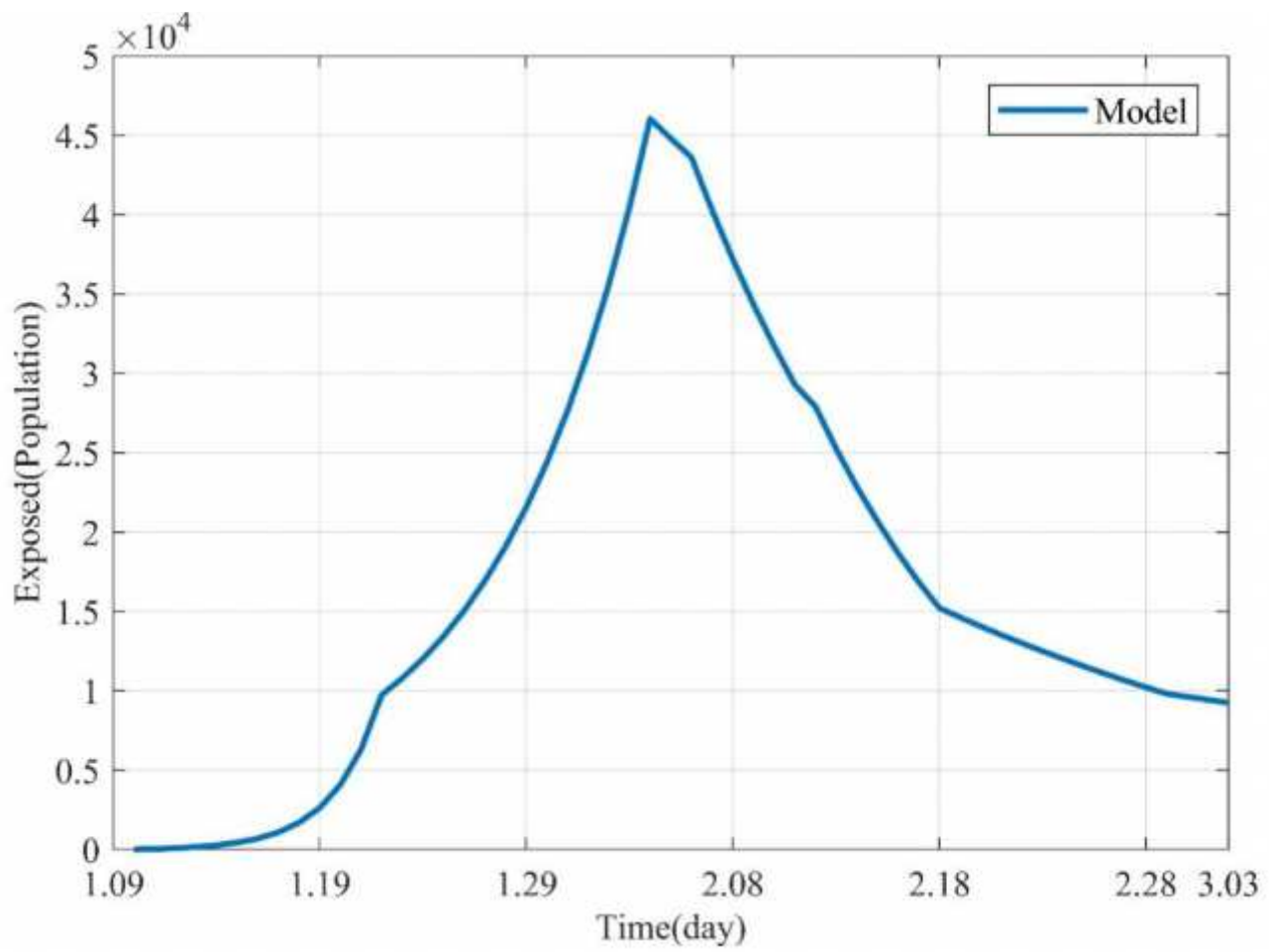
Figures

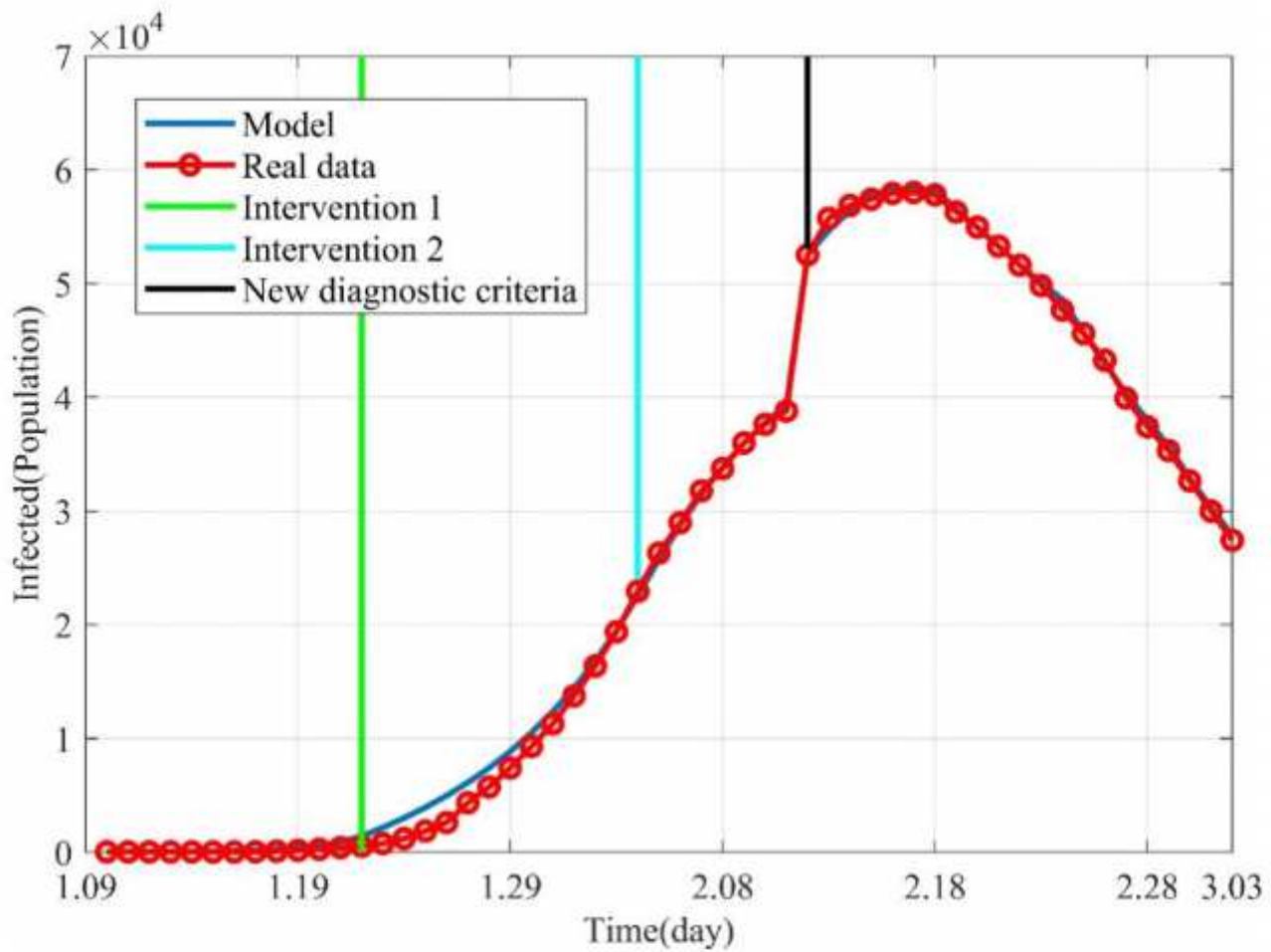**Figure 1**

The transmission mechanism of COVID-19.

Figure 2
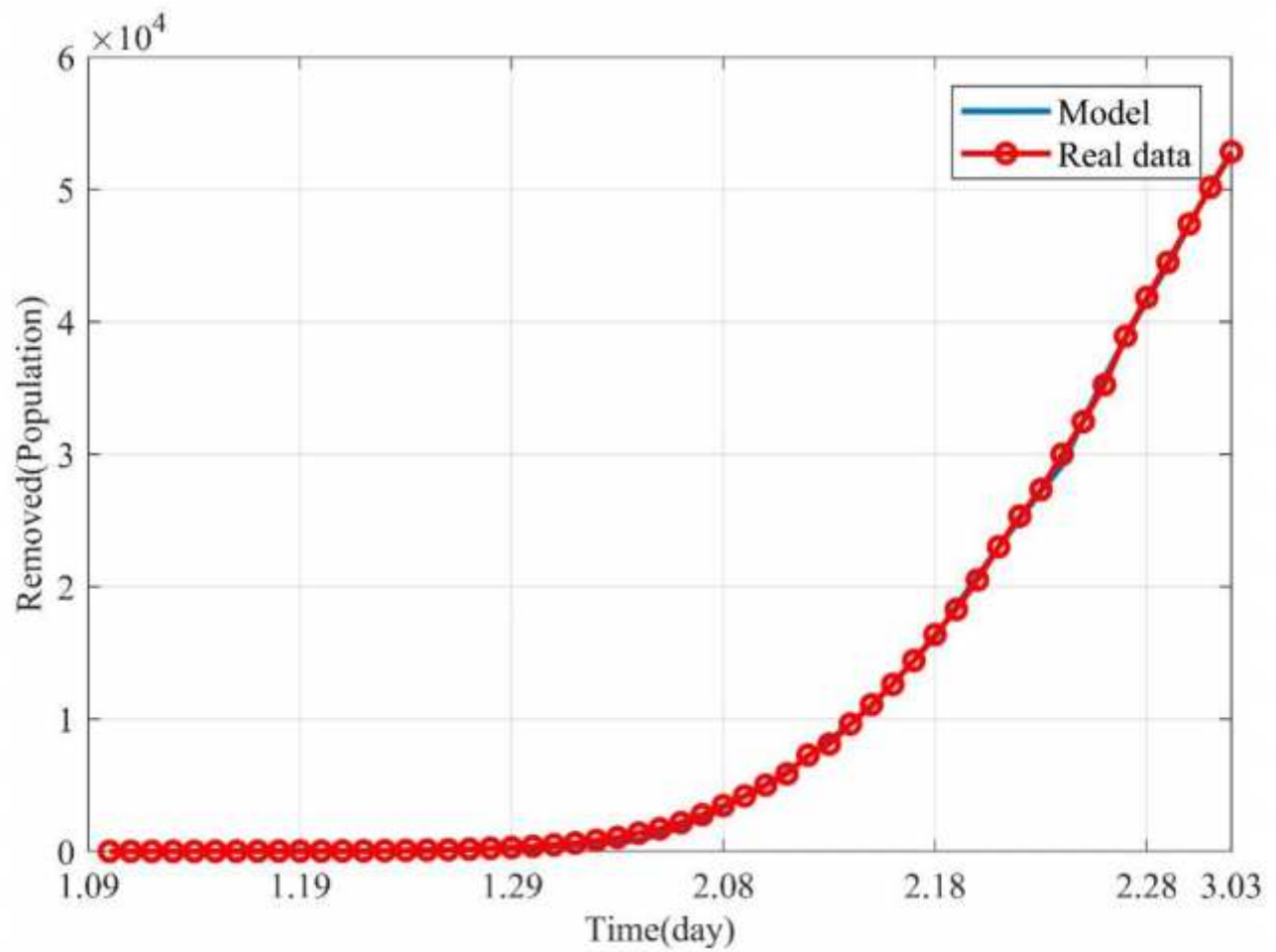
Dynamic trends of the susceptible.

Figure 3

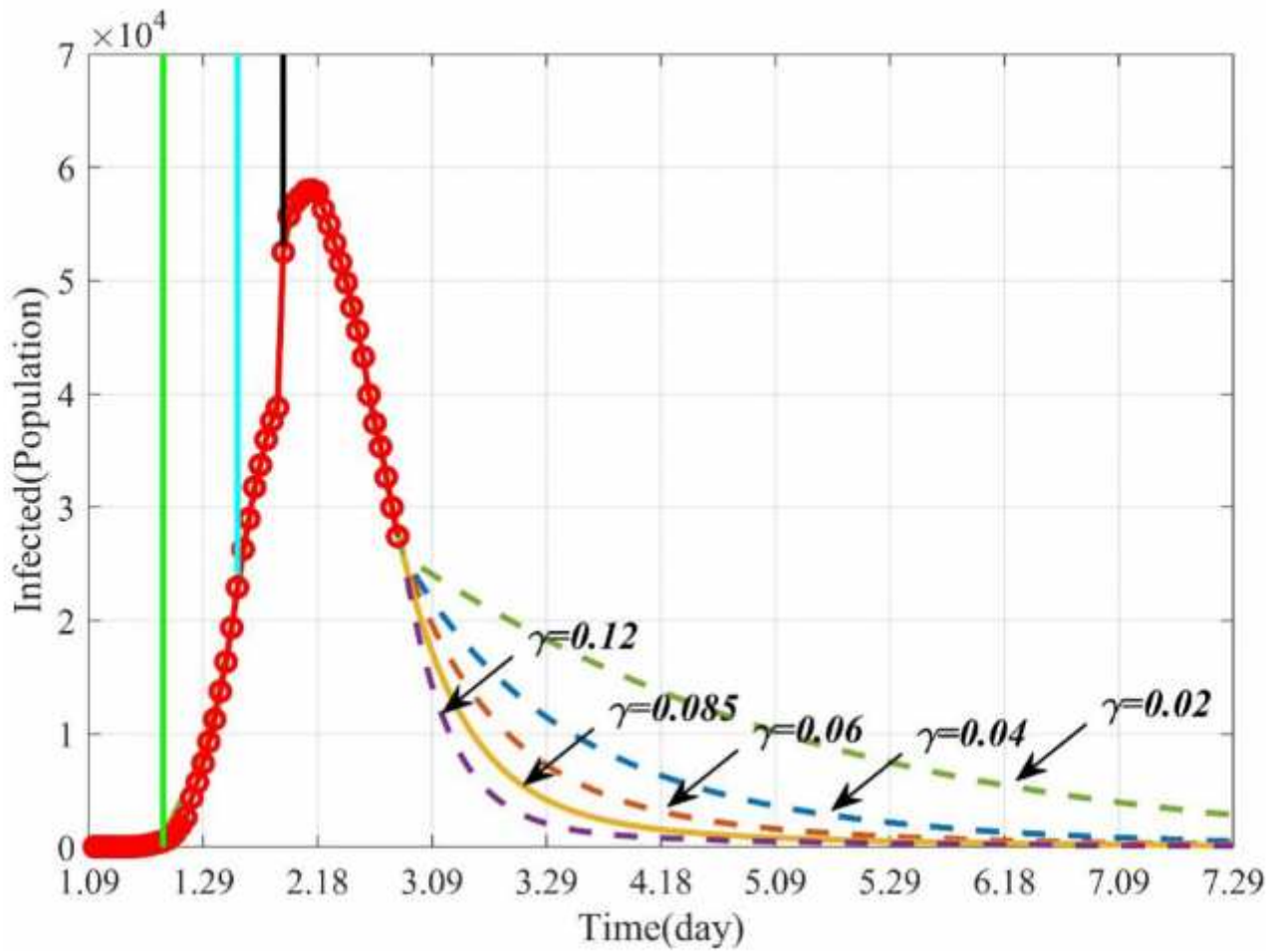Dynamic trends of the exposed.

Figure 4

Dynamic trends of the infected. The blue line represents the simulation of SEIR model. The red line represents the data issued by NHCC. The green line represents the intervention 1. The light blue line represents the intervention 2. The black line represents the government began to implement new diagnostic criteria.
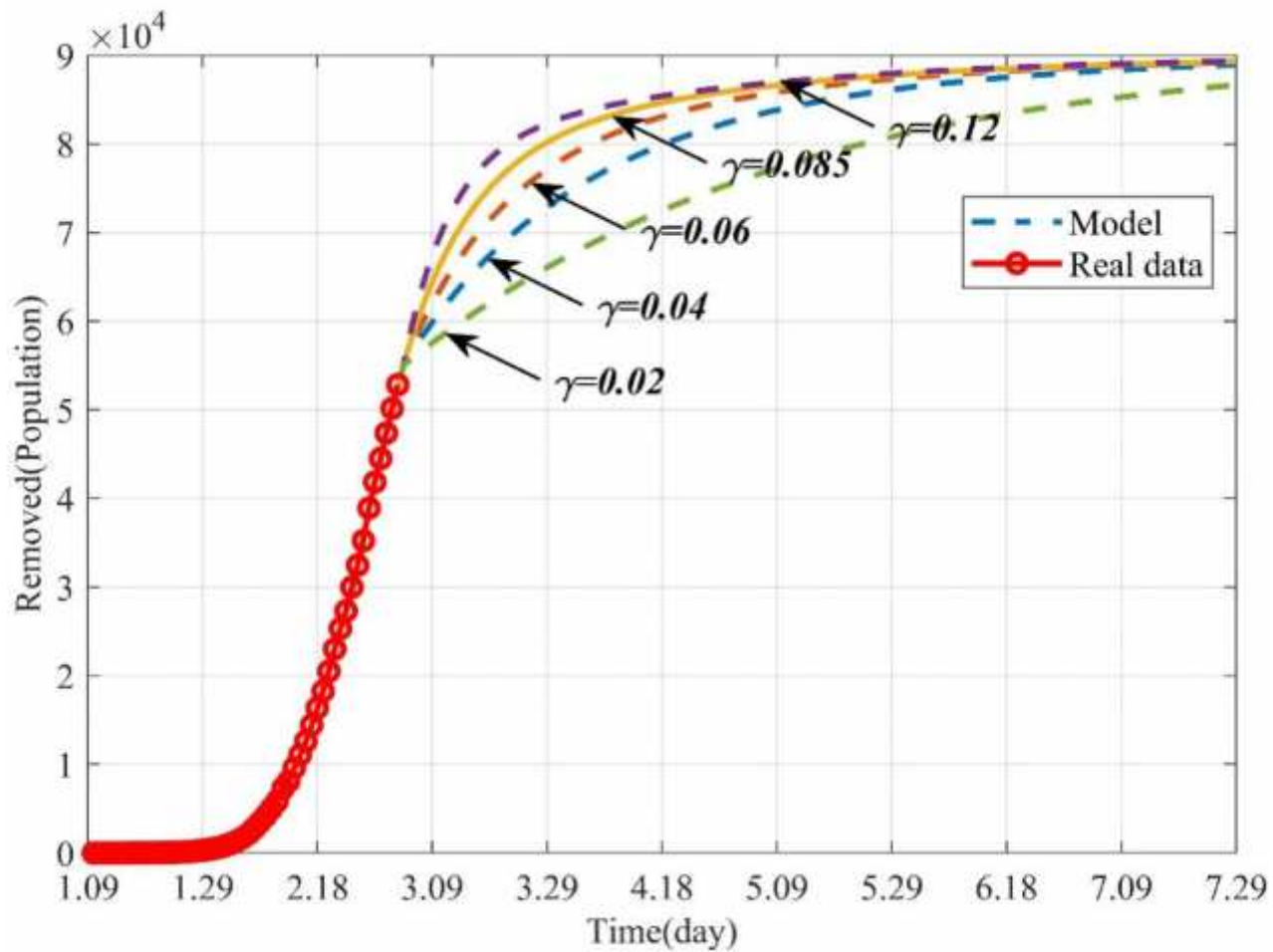
Figure 5

Dynamic trends of the removed. The removed consists of the cures and the death.The blue line represents the simulation of SEIR model.The red line represents the data issued by NHCC.

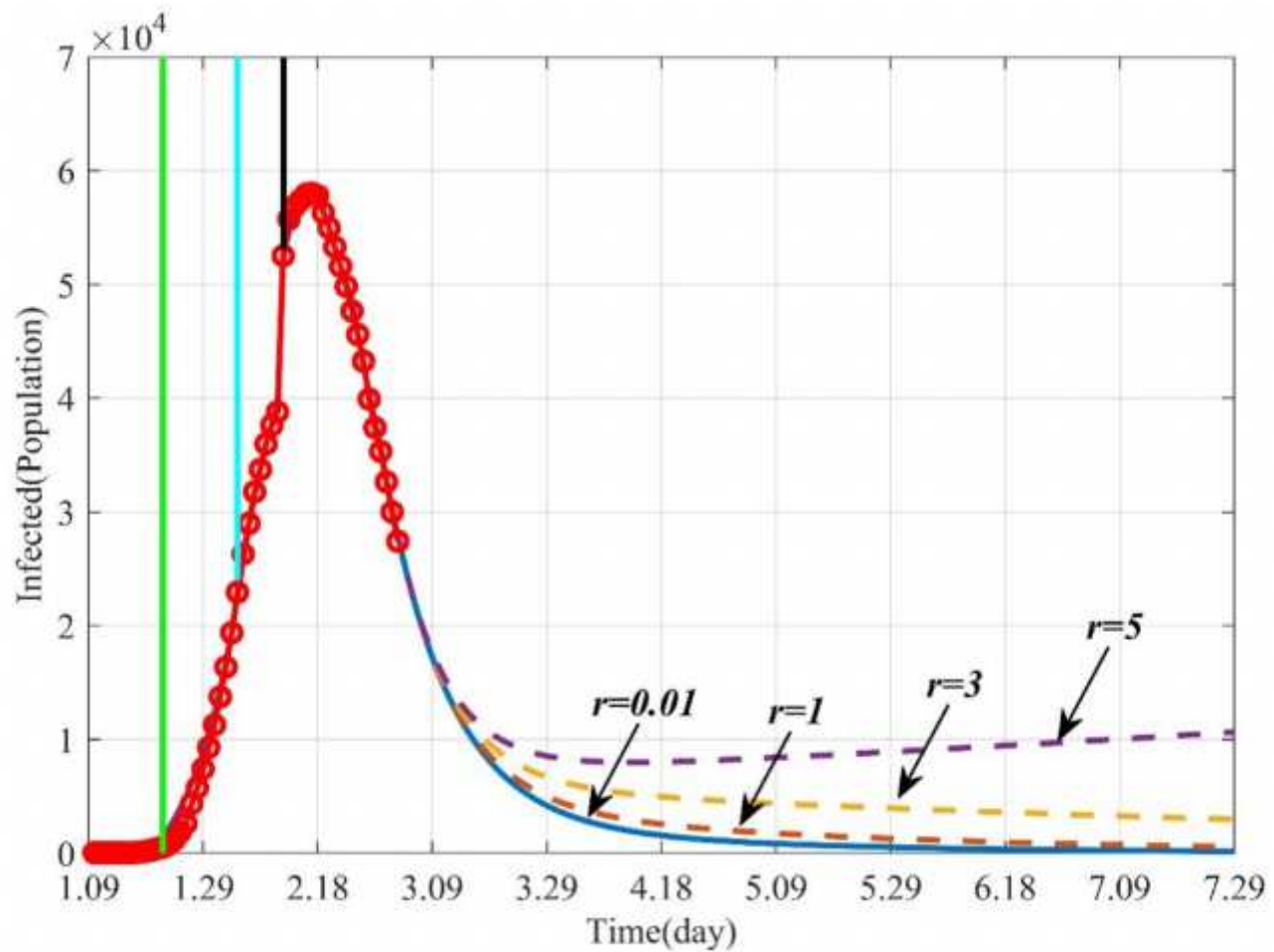**Figure 6**

Dynamic trends of the infected with different removal rates (γ). The solid line represents the trends of the infected with the current γ. The purple dotted line represents the trends of the infected when γ=0.12. The orange dotted line represents the trends of the infected when γ=0.06. The blue dotted line represents the trends of the infected when γ=0.04. The green dotted line represents the trends of the infected when γ=0.02.
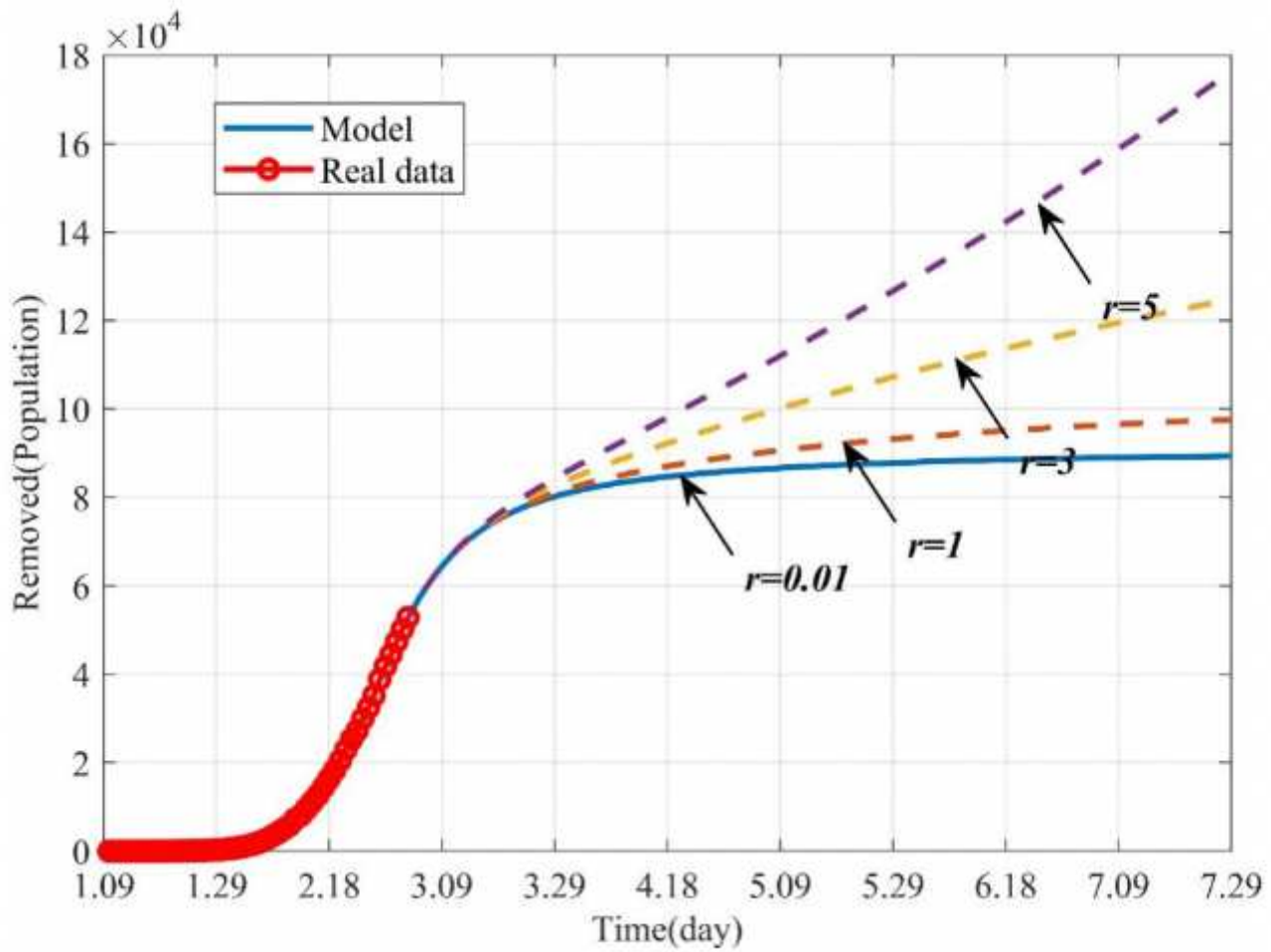
Figure 7

Dynamic trends of the removed with different removal rates (γ) . The solid line represents the trends of the removed with the current γ. The purple dotted line represents the trends of the removed when γ=0.12. The orange dotted line represents the trends of the removed when γ=0.06. The blue dotted line represents the trends of the removed when γ=0.04. The green dotted line represents the trends of the removed when γ=0.02.
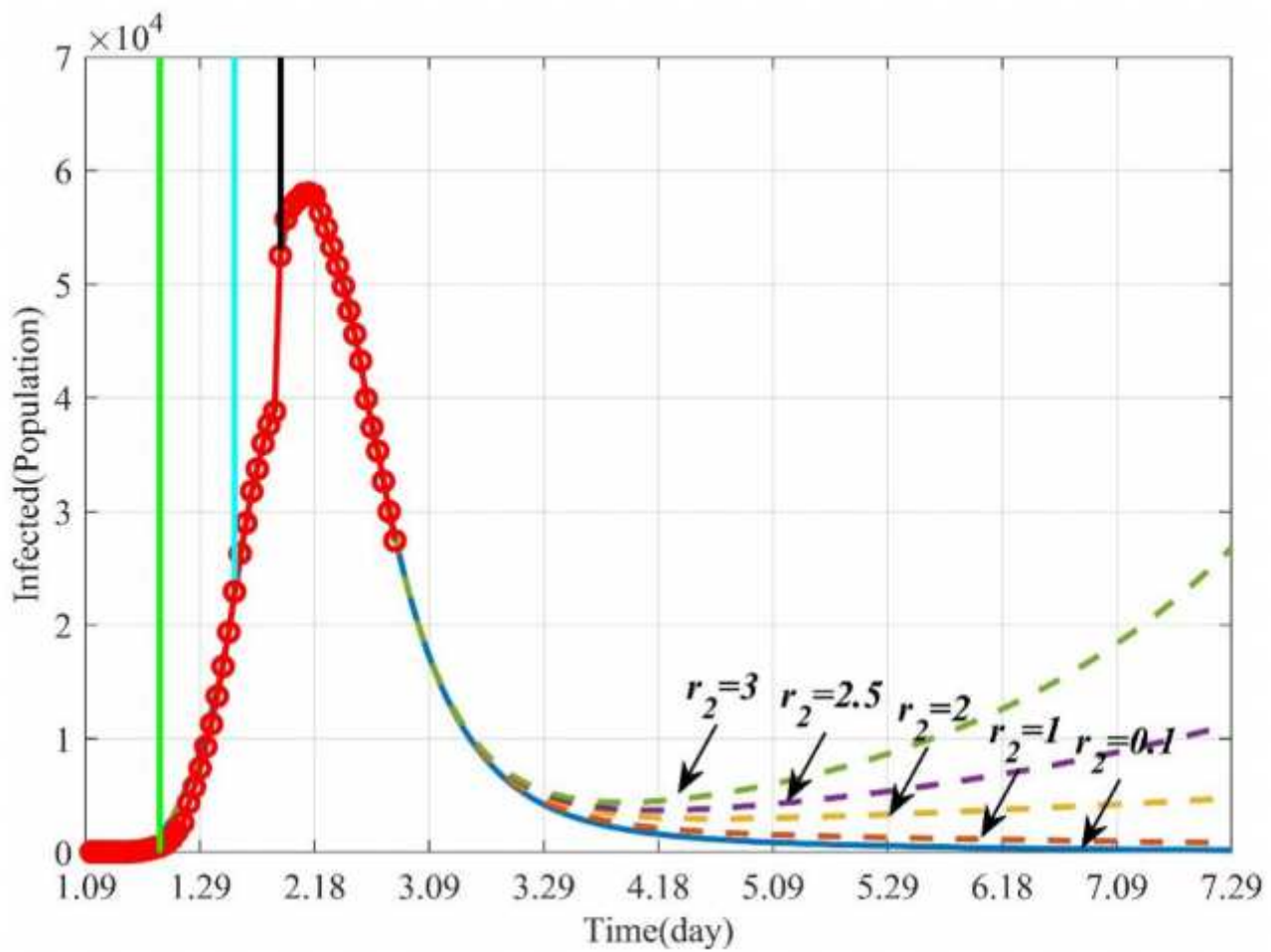
Figure 8

Dynamic trends of the infected with different r. The solid line represents the trends of the infected with the current r. The orange dotted line represents the trends of the infected when r=1. The yellow dotted line represents the trends of the infected when r=3. The purple dotted line represents the trends of the infected when r=5.
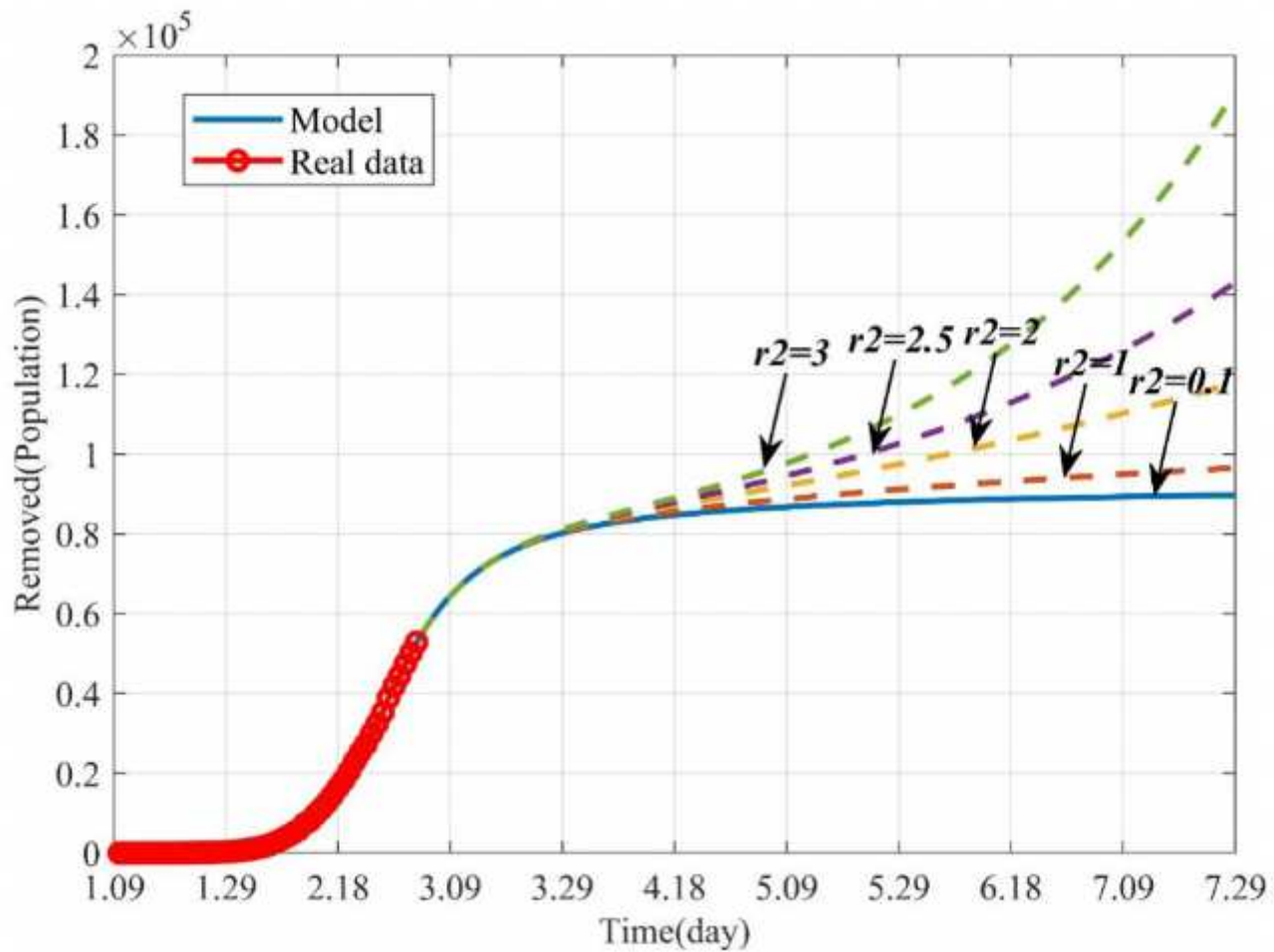
Figure 9

Dynamic trends of the removed with different r. The solid line represents the trends of the removed with the current r. The orange dotted line represents the trends of the removed when r=1. The yellow dotted line represents the trends of the removed when r=3. The purple dotted line represents the trends of the removed when r=5.

Figure 10

Dynamic trends of the infected with different r2. The solid line represents the trends of the infected with the current r2. The orange dotted line represents the trends of the infected when r2=1. The yellow dotted line represents the trends of the infected when r2=2. The purple dotted line represents the trends of the infected when r2=2.5. The green dotted line represents the trends of the infected when r2=3.

Figure 11

Dynamic trends of the removed with different r2. The solid line represents the trends of the removed with the current r2. The orange dotted line represents the trends of the removed when r2=1. The yellow dotted line represents the trends of the removed when r2=2. The purple dotted line represents the trends of the removed when r2=2.5. The green dotted line represents the trends of the removed when r2=3.