



Some Studies and Applications of Infectious Disease Model SIR

Yi Peng, Chengci Zhang

College of Applied Mathematics, Chengdu University of Information Technology
Chengdu 610025, Sichuan, P. R. China

ABSTRACT: Infectious diseases have always been the public enemy of mankind, and the spread of some malignant diseases even poses a serious threat to national security. Therefore, the prevention and control of infectious diseases have become a major issue related to human health, national economy and people's livelihood. Infectious disease dynamics is an important method to theoretically and quantitatively study the epidemic law of infectious diseases, Mathematicians have done a lot of research work in this area. The basic mathematical model of infectious diseases, to study the transmission speed, spatial range, transmission route, dynamic mechanism and other issues of infectious diseases, so as to guide the effective prevention and control of infectious diseases. In this paper, we mainly study the SIR model of infectious diseases. The SIR model is applied to the epidemic caused by the new coronavirus in Wuhan in 2020, to carry out numerical simulation and prediction of the epidemic, analyze the development trend of the epidemic, and provide some help in epidemic prevention and anti-epidemic.

KEYWORDS: Infectious Diseases, SIR Models, Outbreaks, Coronavirus

Received 13 August, 2022; Revised 26 August, 2022; Accepted 28 August, 2022 © The author(s) 2022. Published with open access at www.questjournals.org

I. INTRODUCTION

1.1 RESEARCH BACKGROUND

Since epidemiological mathematics is not only a branch of medical mathematics, but also a branch with more applied mathematics knowledge, in order to establish a more accurate model for describing infectious diseases and then popularize the infectious disease dynamic model, that is, epidemic disease mathematical model. Infectious diseases have the characteristics of epidemic, seasonality and locality^[1]. Common infectious disease models are divided into SI^[2], SIR^[3], SIRS^[4], SEIR^[5] models according to the types of infectious diseases, etc. , according to the propagation mechanism, it is divided into different types based on ordinary differential equations, partial differential equations, and network dynamics. With the in-depth study of infectious disease models, it is found that the problems involved cover many disciplines such as medicine and biology, so it has a strong application background and practical significance.

1.2 RESEARCH STATUS AT HOME AND ABROAD

In 1760, Daniel Bernouilli established the first mathematical model to study the vaccination of healthy people against vaccinia to prevent smallpox, which is also the earliest known infectious disease and epidemic model so far. But the earliest model is still due to the research work of En'Ko from 1873 to 1894, and the real developed and confirmed model was in the early 20th century, W. h. Hamer et al in J. A series of studies with the help of Brownlee. By 1906, Hamer had developed a discrete model for studying measles, which for the first time assumed that the incidence was derived from multiplying the number of susceptible and sick. In 1911, a model for the study of continuous differential equations for malaria was established by Ross, for which he also won the second Nobel Prize in Medicine. By 1927, the work of Kermack and Mekendrick laid the foundations for infectious disease models. They divided the total population of a certain area into three categories: Susceptible (S), Infected (I), and Removed (R), and then established a SIR infectious disease model based on the dynamic research method, and analyzed the spread of The laws and epidemic trends were studied, and the threshold theory was proposed: if the number of susceptible individuals in the population is above the threshold, the infectious disease will maintain; below the threshold, the infectious disease will tend to become extinct^[6]. Most of these models are suitable for the study of general laws of various infectious diseases.

During the SARS epidemic that occurred in 2003, researchers at home and abroad established many dynamic models, that is, infectious disease dynamic models. Infectious disease dynamics^[7] is an important method for theoretical quantitative research. Through this method, we can study the transmission characteristics and trends of this disease, and study the effect of various preventive measures on the control of this disease. Decision-making departments provide a reference. Most of the relevant SARS transmission studies have adopted the infectious disease SIR model. The spread of SARS disease has the following characteristics: hyper-transmissibility, isolation stops further spread and cured individuals are not reinfected^[8].

Until the pneumonia virus that broke out from Wuhan in recent years^[9], it spread to the whole country and even the world in a short time. Its spread is fast. This virus is an unknown new type of coronavirus (COVID-19), which has the characteristics of a pandemic. It has the characteristics of complex transmission routes and rapid transmission. Since the beginning of the outbreak, there have been many This kind of mutation has greatly increased the difficulty of preventing and fighting the epidemic. Therefore, researchers from many countries across the country and even the world have begun to join in the research on this virus, and proposed and designed many infectious disease models, and analyzed them through the models to provide the greatest help in the fight against the epidemic.

1.3 SIGNIFICANCE OF RESEARCH

The study of infectious diseases is a very complex issue. In recent years, with the outbreak of the new coronavirus, which has seriously affected the economic and social development and exchanges in various countries and regions, the issue of infectious diseases has become a very important research topic. The mathematical model is established through a large amount of actual data and based on the existing mathematical knowledge, which can well reflect the data changes. Therefore, it is very important to establish a mathematical model of infectious diseases. He can analyze the existing situation based on the existing infection data, so that he can draw lessons from it and take a series of effective measures for the control of future infectious diseases, thereby reducing the number of infected people. In addition, the use of infectious disease mathematical models can also predict future epidemics, and then take precautions to better control the epidemic.

The infectious disease model is based on the characteristics of population growth, the occurrence of disease, the law of spread and development within the population, and the social factors related to it, to establish a mathematical model that can reflect the dynamic characteristics of infectious diseases. Through the qualitative, quantitative analysis and numerical simulation of the dynamic behavior of the model, we can analyze the development process of the disease, reveal the epidemic law, predict the changing trend, and analyze the cause and key of the disease epidemic.

II. MODELS AND METHODS

The SIR model (in the dynamics of infectious diseases, the classic SIR infectious disease model established by Kermack and McKendrick in 1927 with the method of dynamics^[10]) is the most classic model, as follows:

$$\begin{cases} \frac{dS(t)}{dt} = \frac{-\beta I(t)S(t)}{N}, S(0) = S_0 \geq 0 \\ \frac{dI(t)}{dt} = \frac{\beta I(t)S(t)}{N} - \gamma I(t), I(0) = I_0 \geq 0 \\ \frac{dR(t)}{dt} = \gamma I(t), R(0) = R_0 \geq 0 \end{cases} \quad (1)$$

Among them, S represents the susceptible population, I represents the infected population, R represents the removed population (including those who have been cured and died), β represents the infection rate, γ represents the removal rate, and there are:

$$S(t) + I(t) + R(t) = N \quad (2)$$

N represents the total number of people, and the total number of people does not change.

The numerical simulation method used in this paper is the least squares method. The least squares method is to simulate the numerical value, and to find the optimal parameter value in the data by finding the sum of the smallest error, that is, to find the difference between the estimated value and the actual value. The sum of squares is the smallest.

III. DISCRETIZATION AND NUMERICAL SIMULATION OF MODELS

3.1 DISCRETIZATION OF SIR MODELS FOR INFECTIOUS DISEASES

The model (1) is discretized (written in difference format) as:

$$\begin{cases} \frac{S(t + \Delta t) - S(t)}{\Delta t} = -\beta S(t)I(t) \\ \frac{I(t + \Delta t) - I(t)}{\Delta t} = \beta S(t)I(t) - \gamma I(t) \\ \frac{R(t + \Delta t) - R(t)}{\Delta t} = \gamma I(t) \end{cases} \quad (3)$$

Therefore

$$\begin{cases} S(t + \Delta t) = S(t) - \beta S(t)I(t)\Delta t \\ I(t + \Delta t) = I(t) + \beta S(t)I(t)\Delta t - \gamma I(t)\Delta t \\ R(t + \Delta t) = R(t) + \gamma I(t)\Delta t \end{cases} \quad (4)$$

Making $\Delta t = (t + 1) - t = 1$, we have

$$\begin{cases} S(t + 1) - S(t) = -\beta S(t)I(t) \\ I(t + 1) - I(t) = \beta S(t)I(t) - \gamma I(t) \\ R(t + 1) - R(t) = \gamma I(t) \end{cases} \quad (5)$$

3.2 NUMERICAL SIMULATION OF THE WUHAN EPIDEMIC

3.2.1 Numerical Simulation

Check relevant information online (<http://www.sy72.com/covid/list.asp?id=433&s1=0&s2=2022-4-18>), we obtained some epidemic data in Wuhan in January and February 2020, which are compiled in the form of a table below:

Date	Total Population	The Current Number of Confirmed Cases	The Cumulative Number of Confirmed Case	Number of People Cure	Death Toll
2020/1/14	12326500	34	41	7	1
2020/1/15	12326500	27	41	12	2
2020/1/16	12326500	28	45	15	2
2020/1/17	12326500	41	62	19	2
2020/1/18	12326500	94	121	24	3
2020/1/19	12326500	169	198	25	4
2020/1/20	12326500	227	258	25	6
2020/1/21	12326500	326	363	28	9
2020/1/22	12326500	380	425	28	17
2020/1/23	12326500	441	495	31	23
2020/1/24	12326500	502	572	32	38
2020/1/25	12326500	533	618	40	45
2020/1/26	12326500	593	698	42	63
2020/1/27	12326500	1463	1590	42	85
2020/1/28	12326500	1739	1905	62	104
2020/1/29	12326500	2063	2261	69	129
2020/1/30	12326500	2390	2639	90	159
2020/1/31	12326500	2897	3215	126	192
2020/2/1	12326500	3779	4109	106	224
2020/2/2	12326500	4687	5142	190	265
2020/2/3	12326500	5902	6384	269	313
2020/2/4	12326500	7686	8351	303	362
2020/2/5	12326500	9255	10117	448	414
2020/2/6	12326500	10613	11618	527	478
2020/2/7	12326500	12360	13603	698	545
2020/2/8	12326500	13497	14982	877	608
2020/2/9	12326500	15175	16902	1046	681
2020/2/10	12326500	16500	18454	1206	748
2020/2/11	12326500	17361	19558	1377	820
2020/2/12	12326500	30043	32994	1915	1036
2020/2/13	12326500	33033	35991	1922	1036
2020/2/14	12326500	34289	37914	2502	1123
2020/2/15	12326500	35314	39462	2915	1233
2020/2/16	12326500	36385	41152	3458	1309
2020/2/17	12326500	37152	42752	4219	1381
2020/2/18	12326500	38007	44412	4908	1497

Figure 1: 1.14-2.18 Wuhan Epidemic Data

We calculate the obtained data separately $S(t)$, $I(t)$, $R(t)$, namely:

Susceptible number = total number - cumulative number of confirmed cases

The number of infected = the current number of confirmed cases

Removed = healed + dead

So, we get the required data:

Date	Susceptible Population $S(t)$	Infected Population $I(t)$	Remove Crowd $R(t)$
2020/1/14	12326459	34	8
2020/1/15	12326459	27	14
2020/1/16	12326455	28	17
2020/1/17	12326438	41	21
2020/1/18	12326379	94	27
2020/1/19	12326302	169	29
2020/1/20	12326242	227	31
2020/1/21	12326137	326	37
2020/1/22	12326075	380	45
2020/1/23	12326005	441	54
2020/1/24	12325928	502	70
2020/1/25	12325882	533	85
2020/1/26	12325802	593	105
2020/1/27	12324910	1463	127
2020/1/28	12324595	1739	166
2020/1/29	12324239	2063	198
2020/1/30	12323861	2390	249
2020/1/31	12323285	2897	318
2020/2/1	12322391	3779	330
2020/2/2	12321358	4687	455
2020/2/3	12320116	5802	582
2020/2/4	12318149	7686	665
2020/2/5	12316383	9255	862
2020/2/6	12314882	10613	1005
2020/2/7	12312897	12360	1243
2020/2/8	12311518	13497	1485
2020/2/9	12309598	15175	1727
2020/2/10	12308046	16500	1954
2020/2/11	12306942	17361	2197
2020/2/12	12293506	30043	2951
2020/2/13	12290509	33033	2958
2020/2/14	12288586	34289	3625
2020/2/15	12287038	35314	4148
2020/2/16	12285348	36385	4767
2020/2/17	12283748	37152	5600
2020/2/18	12282088	38007	6405

Figure 2: Epidemic data after 1.14-2.18

Plot the obtained data to obtain the image trends of three categories: $S(t)$, $I(t)$, and $R(t)$:

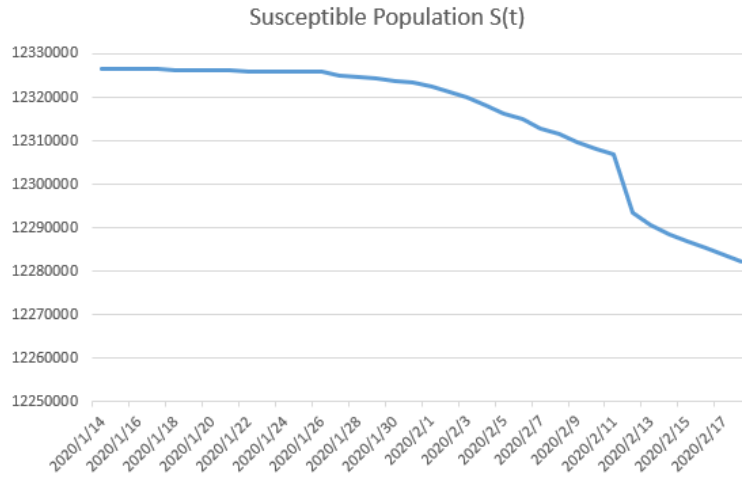


Figure 3: 2020.1.14-2.18 S(t) trend chart of susceptible population

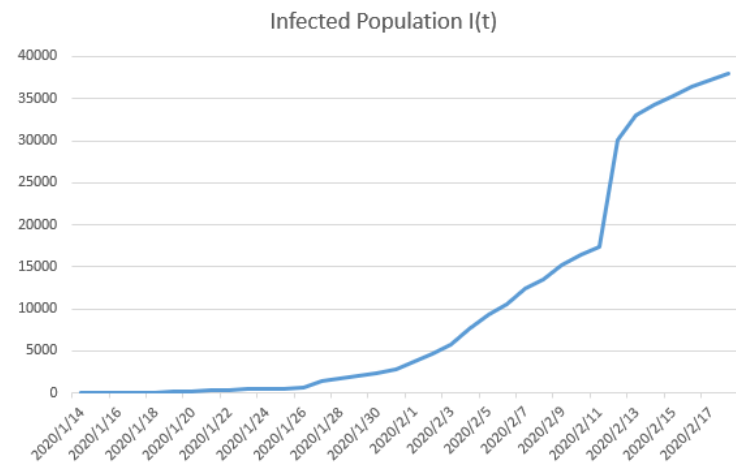


Figure 4: 2020.1.14-2.18 Infected population I(t) trend chart

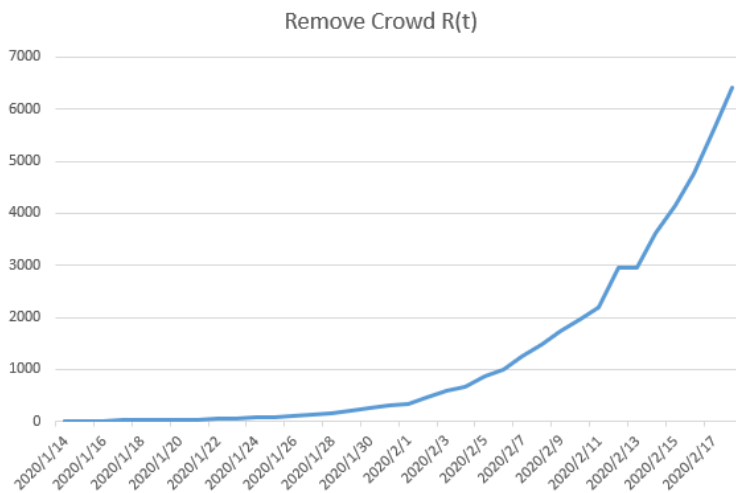


Figure 5: 2020.1.14-2.18 Removal Crowd R(t) Trend Chart

We put the data in Figure 2 into the objective function of the least squares method:

$$L_1 = \sum_{t=1}^T [\bar{S}(t) - S(t)]^2 \quad (6)$$

and

$$L_2 = \sum_{t=1}^T [\bar{R}(t) - R(t)]^2 \quad (7)$$

Among them, $\bar{S}(t)$ represents the actual value of the number of susceptible people, that is, the observed value, $S(t)$ represents the estimated value of the number of susceptible people, that is, the predicted value, then L_1 represents the error between the observed value of the susceptible population and the predicted value. Similarly, $\bar{R}(t)$ represents the observed value of the number of people removed, $R(t)$ represents the predicted value of the number of people removed, and L_2 represents the error between the observed value and the predicted value of the number of people removed.

Because the SIR model after discretization is shown in (5), we need to reorganize the table in Figure 2, as shown in Figure 6:

data	S(t)	I(t)	R(t)	S(t+1)-S(t)	R(t+1)-R(t)	S(t)I(t)
2020/1/14	12326459	34	8	0	6	419099606
2020/1/15	12326459	27	14	-4	3	332814393
2020/1/16	12326455	28	17	-17	4	345140740
2020/1/17	12326438	41	21	-59	6	505383958
2020/1/18	12326379	94	27	-77	2	1158679626
2020/1/19	12326302	169	29	-60	2	2083145038
2020/1/20	12326242	227	31	-105	6	2798056934
2020/1/21	12326137	326	37	-62	8	4018320662
2020/1/22	12326075	380	45	-70	9	4683908500
2020/1/23	12326005	441	54	-77	16	5435768205
2020/1/24	12325928	502	70	-46	15	6187615856
2020/1/25	12325882	533	85	-80	20	6569695106
2020/1/26	12325802	593	105	-892	22	7309200586
2020/1/27	12324910	1463	127	-315	39	18031343330
2020/1/28	12324595	1739	166	-356	32	21432470705
2020/1/29	12324239	2063	198	-378	51	25424905057
2020/1/30	12323861	2390	249	-576	69	29454027790
2020/1/31	12323285	2897	318	-894	12	35700556645
2020/2/1	12322391	3779	330	-1033	125	46566315589
2020/2/2	12321358	4687	455	-1242	127	57750204946
2020/2/3	12320116	5902	582	-1967	83	71481313032
2020/2/4	12318149	7686	665	-1766	197	94677293214
2020/2/5	12316383	9255	862	-1501	143	113988124665
2020/2/6	12314882	10613	1005	-1985	238	130697842666
2020/2/7	12312897	12360	1243	-1379	242	152187406920
2020/2/8	12311518	13497	1485	-1920	242	166168558446
2020/2/9	12309598	15175	1727	-1552	227	186798149650
2020/2/10	12308046	16500	1954	-1104	243	203082759000
2020/2/11	12306942	17361	2197	-13436	754	213660820062
2020/2/12	12293506	30043	2951	-2997	7	369333800758
2020/2/13	12290509	33033	2958	-1923	667	405992383797
2020/2/14	12288586	34289	3625	-1548	523	421363325354
2020/2/15	12287038	35314	4148	-1690	619	433904459932
2020/2/16	12285348	36385	4767	-1600	833	447002386980
2020/2/17	12283748	37152	5600	-1660	805	456365805696
2020/2/18	12282088	38007	6405	-615	1493	466805318616

Figure 6: Epidemic data reorganized from 1.14-2.18

From formula (5), we can get:

$$\begin{cases} S(t+1) = \bar{S}(t) - \beta \bar{S}(t) \bar{I}(t) \\ R(t+1) = \bar{R}(t) + \gamma \bar{I}(t) \end{cases} \quad (8)$$

where $\bar{S}(t)$, $\bar{I}(t)$ and $\bar{R}(t)$ represent the actual data (i.e. observations).

Therefore, it can be seen from Figure 6 that the number of susceptible people $S(t)$ and $S(t)I(t)$ ($S(t)$ and $I(t)$ in the table are the actual data of the susceptible population, namely $\bar{S}(t)$ and $\bar{I}(t)$) data in two columns. Therefore, by substituting the number of susceptible people $\bar{S}(t)$ for each day into the first formula in Equation (8), the predicted value of the number of susceptible people on the next day can be obtained (that is, $\bar{S}(t)$ in Equation (5)). Take time $t = 0, 1, 2, \dots, 17$ ($t = 0$ means January 14th, $t = 1$ means January 15th, and so on), when $t = 0$, substitute $\bar{S}(0)$ into the first formula in formula (8), the predicted value $S(1)$ on January 15 can be obtained, and all the data are substituted in turn, that is:

$$L_1 = \sum_{t=1}^T [\bar{S}(t) - S(t)]^2 = \sum_{t=1}^T [\bar{S}(t) - (\bar{S}(t-1) - \beta \bar{S}(t-1) \bar{I}(t-1))]^2 \quad (9)$$

Find the partial derivative with respect to β , that is:

$$\begin{aligned} \frac{dL_1}{d\beta} = \sum_{t=1}^T [\bar{S}^2(t) - 2\bar{S}(t)\bar{S}(t-1) + 2\bar{S}(t)\bar{S}(t-1)\bar{I}(t-1) + \bar{S}^2(t-1) \\ - 2\bar{S}^2(t-1)\bar{I}(t-1) + 2\beta\bar{S}^2(t-1)\bar{I}^2(t-1)] = 0 \end{aligned} \quad (10)$$

Similarly, it can be seen from Figure 6 that the number of people removed $R(t)$ and the number of infected people $I(t)$ ($R(t)$ in the table is also the observed value of the number of people removed, i.e. $\bar{R}(t)$ in formula (5)) are two columns data. Therefore, by substituting the number of people removed $\bar{R}(t)$ for each day into the second equation in equation (8), the predicted value of the number of susceptible people on the next day (that is, $R(t)$ in equation (5) can be obtained). Take the time $t = 0, 1, 2, \dots, 17$ ($t = 0$ means January 14th, $t = 1$ means January 15th, and so on), when $t = 0$, set $\bar{R}(0)$ substitute into the first formula in formula (8), the predicted value $R(1)$ on January 15 can be obtained, and all the data are substituted in sequence, namely:

$$L_2 = \sum_{t=1}^T [\bar{R}(t) - R(t)]^2 = \sum_{t=1}^T [\bar{R}(t) - (\bar{R}(t-1) + \gamma\bar{I}(t-1))]^2 \quad (11)$$

Find the partial derivative with respect to γ , that is:

$$\begin{aligned} \frac{dL_2}{d\gamma} = \sum_{t=1}^T [2\bar{R}(t)\bar{I}(t-1) - 2\bar{R}(t)\bar{R}(t-1) - 2\bar{R}(t)\bar{I}(t-1) + \bar{R}^2(t-1) \\ + 2\bar{R}(t-1)\bar{I}(t-1) + 2\gamma\bar{I}^2(t-1)] = 0 \end{aligned} \quad (12)$$

Put $S(t)$ in Figure 6 (that is, $\bar{S}(t)$ in the formula), $I(t)$ (that is, $\bar{I}(t)$ in the formula), and $R(t)$ (that is, the $\bar{R}(t)$) and $S(t)I(t)$ (that is, $\bar{S}(t)\bar{I}(t)$ in the formula) data are substituted into formula (10) and formula (12) respectively. And data fitting is realized through Python programming (see the appendix for the code), so the parameters β and γ are obtained, and the parameter values from January 15th to February 18th are obtained:

$$\begin{cases} \beta \approx 4.43 \times 10^{-9} \\ \gamma \approx 0.021 \end{cases} \quad (\text{Retain 3 digits after the decimal point}) \quad (13)$$

Then bring the obtained parameters into (5) for calculation, that is, to obtain the predicted value, we will compare the observed value (actual value) with the predicted value, as shown in Figure 6:

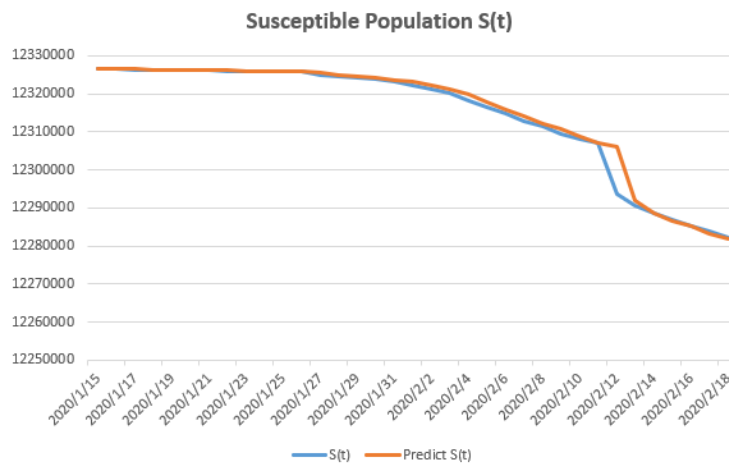


Figure 6: 2020.1.15-2.18 Susceptible population $S(t)$ comparison chart

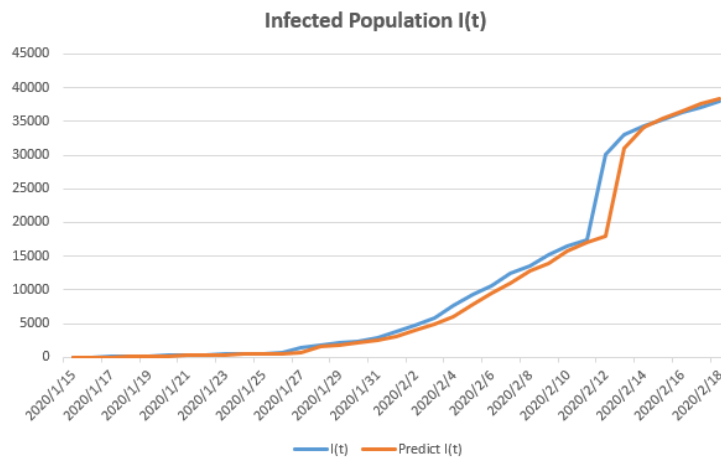


Figure 7: 2020.1.15-2.18 Infected population $I(t)$ comparison chart

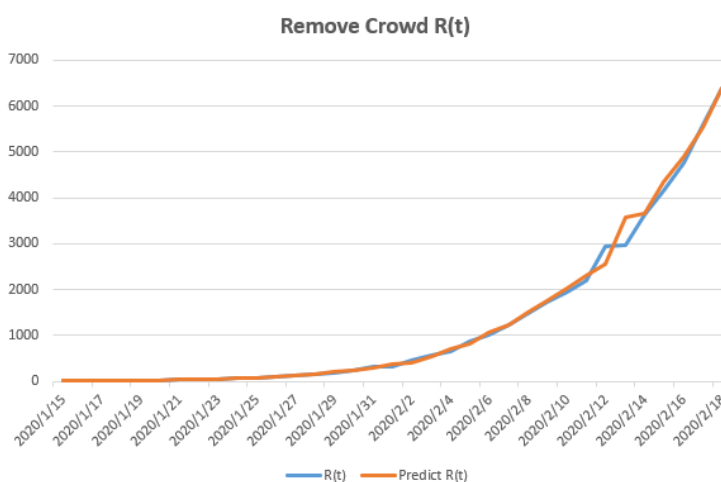


Figure 8: 2020.1.15-2.18 Removal Crowd $R(t)$ Comparison Chart

First of all, it can be seen from the comparison chart of the susceptible population $S(t)$ (Figure 6) that the actual susceptible population is very close to the predicted susceptible population, and the predicted results from February 11 to February 13 appeared slightly. There is a large error, but the overall forecast basically achieves the expected results. From the comparison chart of the infected population $I(t)$ (Figure 7), it can also be seen that the actual number of infections is very close to the predicted number of infections, and the prediction results from February 11th to February 13th have a slightly larger error. The overall forecast is good. The prediction of the comparison chart (Figure 8) of removing the population $R(t)$ also basically achieved the expected results, and the place with a large error also appeared in the prediction from February 11 to February 13.

3.2.2 Error Analysis

In order to understand the reason for the large error from February 11 to February 13, we checked the relevant information and learned the reason (<https://new.qq.com/rain/a/20200213A0C8WM00>). The reason for the sharp increase in the number of infections from February 11 to February 12 was that after the 11th, the clinically diagnosed cases were also included in the confirmed cases, that is, "false negatives" were also counted among the confirmed cases. The number of confirmed cases has soared, as has the number of deaths. Among them, clinical cases refer to patients who meet the two conditions of "fever, fever or respiratory symptoms", "the number of leukocytes and lymphocytes in the early stage of illness is reduced", and the symptoms of pneumonia are detected by CT.

The above is the comparison and error analysis of the actual and predicted values of Wuhan epidemic data from January 15 to February 18, 2020.

3.2.3 Epidemic Forecast

Next, we will predict and simulate the future development of the epidemic based on the parameter values obtained from January 15, 2020 to February 18, 2020, and draw and compare the obtained data, as follows:

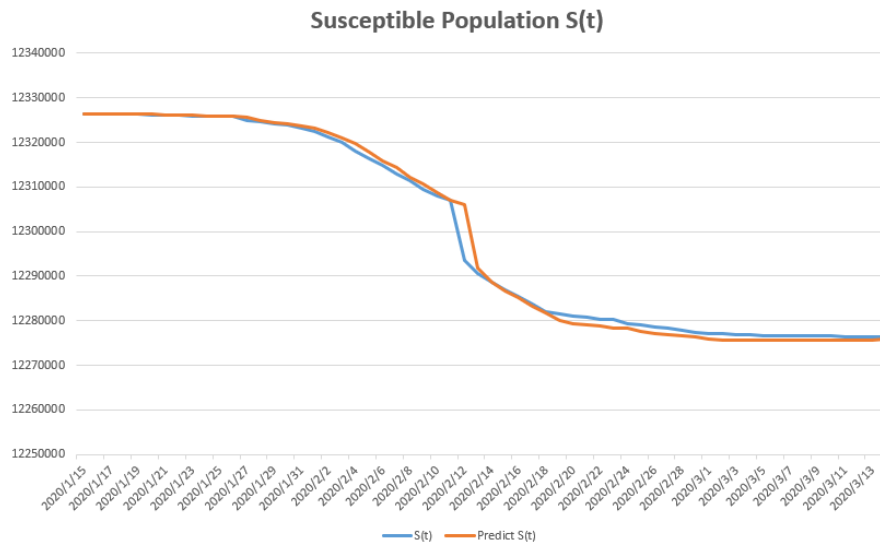


Figure 9: Prediction of $S(t)$ for the susceptible population in 2020.2.19-3.14

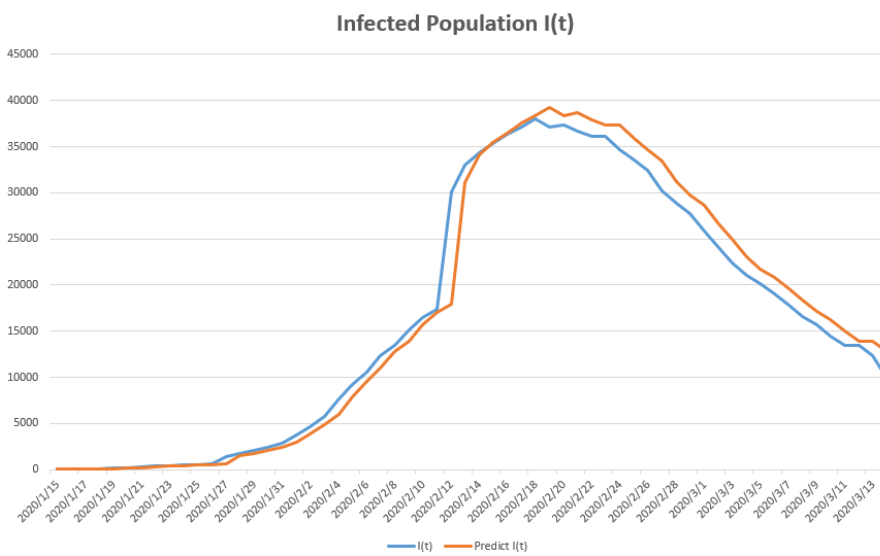


Figure 10: Prediction of the infected population $I(t)$ in 2020.2.19-3.14

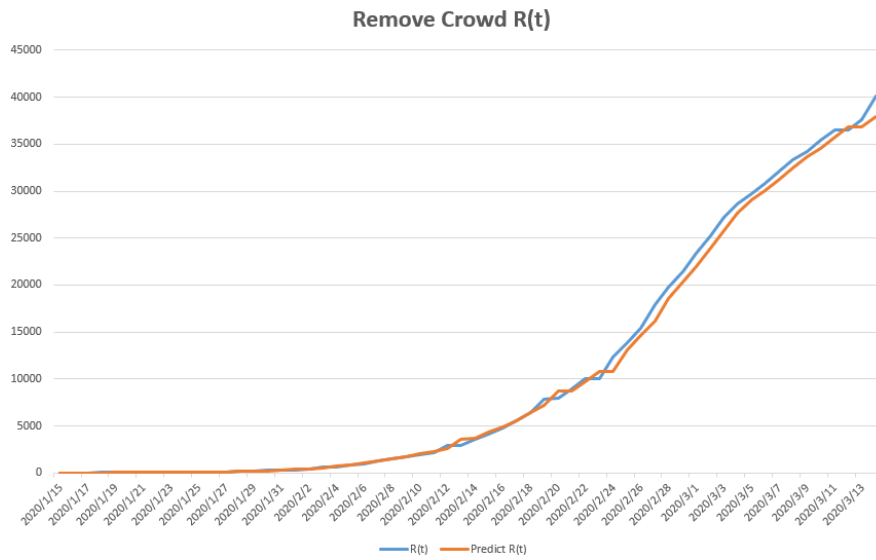


Figure 11: 2020.2.19-3.14 Prediction of removal of crowd $R(t)$

From Figure 9, we can see that the predicted value of the number of susceptible people is close to the actual number of susceptible people. From Figure 10, it can be seen that the predicted value of the number of infected people is close to the actual number of infected people. From Figure 11, it can be seen that the predicted value of the number of people removed is close to Actual removals.

Below is a data table of actual and predicted values:

data	S(t)	Predict S(t)	I(t)	Predict I(t)	R(t)	Predict R(t)
2020/1/15	12326459	12326457.143	27	35.143	14	8.714
2020/1/16	12326455	12326457.526	28	27.907	17	14.567
2020/1/17	12326438	12326453.471	41	28.941	21	17.588
2020/1/18	12326379	12326435.761	94	42.378	27	21.861
2020/1/19	12326302	12326373.867	169	97.159	29	28.974
2020/1/20	12326242	12326292.772	227	174.679	31	32.549
2020/1/21	12326137	12326229.605	326	234.628	37	35.767
2020/1/22	12326075	12326119.199	380	336.955	45	43.846
2020/1/23	12326005	12326054.250	441	392.770	54	52.980
2020/1/24	12325928	12325980.920	502	455.819	70	63.261
2020/1/25	12325882	12325900.589	533	518.869	85	80.542
2020/1/26	12325802	12325852.896	593	550.911	105	96.193
2020/1/27	12324910	12325769.620	1463	612.927	127	117.453
2020/1/28	12324595	12324830.121	1739	1512.156	166	157.723
2020/1/29	12324239	12324500.054	2063	1797.427	198	202.519
2020/1/30	12323881	12324126.368	2390	2132.309	249	241.323
2020/1/31	12323285	12323730.519	2897	2470.291	318	299.190
2020/2/1	12322391	12323126.847	3779	2994.316	330	378.837
2020/2/2	12321358	12322184.711	4687	3905.930	455	409.359
2020/2/3	12320116	12321102.187	5802	4844.406	582	553.427
2020/2/4	12318149	12319799.338	7686	5996.820	665	703.842
2020/2/5	12316383	12317729.580	9255	7944.014	862	826.406
2020/2/6	12314882	12315878.033	10613	9565.612	1005	1056.355
2020/2/7	12312897	12314303.009	12360	10969.118	1243	1227.873
2020/2/8	12311518	12312222.810	13497	12774.630	1485	1502.560
2020/2/9	12309598	12310781.873	15175	13949.690	1727	1768.437
2020/2/10	12308046	12308770.484	16500	15683.841	1954	2045.675
2020/2/11	12306942	12307146.343	17361	17053.157	2197	2300.500
2020/2/12	12293506	12305895.483	30043	17942.936	2951	2561.581
2020/2/13	12290509	12291869.851	33033	31048.246	2958	3581.903
2020/2/14	12288586	12288710.454	34289	34137.853	3625	3651.693
2020/2/15	12287038	12286719.360	35314	35435.571	4148	4345.069
2020/2/16	12285348	12285115.803	36385	36494.603	4767	4689.594
2020/2/17	12283748	12283367.779	37152	37601.136	5600	5531.085
2020/2/18	12282088	12281726.299	38007	38393.509	6405	6380.192
2020/2/19	12281473	12280020.052	37129	39276.801	7898	7203.147
2020/2/20	12281154	12279452.925	37398	38369.366	7948	8677.709
2020/2/21	12280840	12279119.343	36680	38547.299	8980	8733.358
2020/2/22	12280299	12278844.457	36174	37905.263	10027	9750.280
2020/2/23	12280299	12278331.072	36163	37382.274	10038	10786.654
2020/2/24	12279429	12278331.670	34691	37370.907	12380	10797.423
2020/2/25	12279059	12277541.883	33563	35849.606	13878	13108.511
2020/2/26	12278676	12277233.299	32392	34683.878	15432	14582.823
2020/2/27	12278363	12276914.052	30179	33473.716	17958	16112.232
2020/2/28	12277943	12276721.469	28836	31186.772	19721	18591.759
2020/2/29	12277378	12276374.573	27700	29798.871	21422	20326.556
2020/3/1	12277185	12275871.431	25903	28624.869	23412	22003.700
2020/3/2	12277074	12275776.189	24144	26767.848	25282	23955.963
2020/3/3	12276960	12275760.870	22368	24950.106	27172	25789.024
2020/3/4	12276829	12275743.473	21050	23114.799	28621	27641.728
2020/3/5	12276703	12275684.167	20115	21752.783	29682	29063.050
2020/3/6	12276629	12275609.030	19011	20786.555	30860	30104.415
2020/3/7	12276588	12275595.078	17772	19645.691	32140	31259.231
2020/3/8	12276552	12275621.465	16627	18365.323	33321	32513.212
2020/3/9	12276535	12275647.739	15732	17182.094	34233	33670.167
2020/3/10	12276522	12275679.414	14514	16257.214	35464	34563.372
2020/3/11	12276514	12275732.656	13462	14998.550	36524	35768.794
2020/3/12	12276514	12275781.870	13462	13911.428	36524	36806.702
2020/3/13	12276509	12275781.870	12358	13911.428	37633	36806.702
2020/3/14	12276501	12275836.911	9911	12770.571	40088	37892.518

Figure 12: Data for observed and predicted values of $S(t)$, $I(t)$ and $R(t)$, respectively

So, we can get: that average error between the actual value of $S(t)$ and the predicted value

$$\frac{\left| \text{total actual value} - \text{total predictive value} \right|}{\text{total actual value}} \approx 4.747 \times 10^{-6}.$$

Similarly, we can get(Reserved to three decimal places) that

the average error between the actual value of $I(t)$ and the predicted value ≈ 0.0231 ;

the mean error between the actual value of $R(t)$ and the predicted value ≈ 0.030 .

3.2.4 Result Analysis

From the predicted results of the susceptible population, we can see that the number of susceptible people is decreasing, and the number of susceptible people began to level off around February 19, 2020, which means that the number of infected people gradually peaked. Therefore, correspondingly, we can see from the predicted number of infections in Figure 10 that the predicted peak of the number of infections is also around

the 19th, after which the number of infections gradually decreases, while the number of cures and deaths gradually increases, and the number of infections will gradually tend to 0. Furthermore, we can see from Figure 11 that the number of cured and dead people is increasing, and when the number of infected people gradually tends to 0, the number of cured and dead people will also gradually stabilize.

IV. CONCLUSION

The infectious disease SIR model used in this paper is the most classic model among infectious disease models, but this model also has certain limitations. It is suitable for numerical simulation and prediction of infectious diseases in a closed area, and variables are also limited. However, this model is relatively applicable to the research and prediction after the Wuhan epidemic is closed in 2020. It provides a reference for measures such as isolation intensity and treatment to control and reduce the spread of the new coronavirus. The development and treatment of sick patients has provided assistance for better control of the epidemic.

REFERENCES

- [1]. Epidemiological characteristics. Terminology online.
- [2]. Cunde Yuan, Baoan Hu. SI infectious disease model with stage structure [J]. Chinese Journal of Applied Mathematics. 2002(2):193-203.
- [3]. Ronggui Luo, Tao Jiang. Research on technology diffusion model based on SIR infectious disease model [J]. Chinese Journal of Management Engineering. 2006(1):32-35.
- [4]. Kaiyuan Liu, Lansun Chen. Global analysis of a SEIR infectious disease model with vertical infectious disease and impulse immunity [J]. System Science and Mathematics. 2010(3):323-333.
- [5]. Zhen Jin, Zhien Ma. SIRS infectious disease model with continuous and pulse vaccination [J]. Journal of North China Institute of Technology. 2003(4):235-243.
- [6]. Zhengyi Lu, Yicang Zhou. Advances in Mathematical Biology [M]. Beijing: Science Press. 2006.
- [7]. Jiang Qiyuan edited the tutoring course (9) Instructor: Lei Deng.
- [8]. Wenbing Xu, Houbao Xu, Jingyuan Yu, Guangtian Zhu. Existence, Uniqueness and Stability of SARS Mathematical Model Solutions [J]. Journal of Applied Functional Analysis. 2004(2):140-145.
- [9]. Yue Yan, Yu Chen, Keji Liu, Xinyue Luo, Boxi Xu, Yu Jiang, Jin Cheng. Modeling and prediction of novel light-like virus pneumonia epidemic based on a class of time-delay dynamical systems [J]. Science in China: Mathematics. China Science Journal. 2020(3):1-8.
- [10]. Zhien Ma. Mathematical modeling and research of infectious disease dynamics [M]. Science Press. 2004.

APPENDIX: RELATED OPERATION CODES

```
from sklearn import linear_model
import numpy as np
import pandas
data = pandas.read_excel('D:/pycharm/PyCharm Community Edition 2021.3.2/2020.xlsx', sheet_name='Sheet1',
header=0)
data2=data['R(t+1)-R(t)']
data3=data['I(t)']
y=data2.values
x=data3.values
x=x.reshape(30,1)
reg = linear_model.LinearRegression()
reg.fit(x,y)
print('Intercept',reg.intercept_)
print('Regression parameters',reg.coef_)
res = reg.predict([[22282]])
print('Forecast result',res)
```