Long-term predictions of current confirmed and dead cases of COVID-19 in China by the non-autonomous delayed epidemic models

Lijun Pei · Mengyu Zhang

Received: 18 April 2021 / Revised: 4 July 2021 / Accepted: 14 July 2021 / Published online: 26 July 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
In this paper, we make long-term predictions based on numbers of current confirmed cases, accumulative dead cases of COVID-19 in different regions in China by modeling approach. Firstly, we use the SIRD epidemic model (S-Susceptible, I-Infected, R-Recovered, D-Dead) which is a non-autonomous dynamic system with incubation time delay to study the evolution of the COVID-19 in Wuhan City, Hubei Province and China Mainland. According to the data in the early stage issued by the National Health Commission of China, we can accurately estimate the parameters of the model, and then accurately predict the evolution of the COVID-19 there. From the analysis of the issued data, we find that the cure rates in Wuhan City, Hubei Province and China Mainland are the approximately linear increasing functions of time $t$ and their death rates are the piecewisely decreasing functions. These can be estimated by finite difference method. Secondly, we use the delayed SIRD epidemic model to study the evolution of the COVID-19 in the Hubei Province outside Wuhan City. We find that its cure rate is an approximately linear increasing function and its death rate is nearly a constant. Thirdly, we use the delayed SIR epidemic model (S-Susceptible, I-Infected, R-Removed) to predict those of Beijing, Shanghai, Zhejiang and Anhui Provinces. We find that their cure rates are the approximately linear increasing functions and their death rates are the small constants. The results indicate that it is possible to make accurate long-term predictions for numbers of current confirmed, accumulative dead cases of COVID-19 by modeling. In this paper the results indicate we can accurately obtain and predict the turning points, the end time and the maximum numbers of the current infected and dead cases of the COVID-19 in China. In spite of our simple method and small data, it is rather effective in the long-term prediction of the COVID-19.

Keywords Long-term prediction · SIRD · SIR · COVID-19 · Maximum numbers · End time · Turning point · Parameter estimation

Introduction
In December 2019, a typical pneumonia occurred in Wuhan City, the capital of Hubei Province, China. On December 31st, the Health Commission of Wuhan reported 27 cases, and the disease quickly spread to other provinces in a short time with the Spring Festival Travel Rush. Initially, this virus was called the SARS-COV-2. On January 30th, 2020, WHO classified the Novel Coronavirus Pneumonia epidemic (NCP) as a public health emergency of international concern, and the NCP outbreak began to attract the attention of countries around the world. On February 11th, 2020, the disease was named as the Corona Virus Disease 2019 (COVID-19) by World Health Organization.

Mathematical modelling can be employed to better understand the dynamics of contagious diseases and simulate different scenarios, providing additional tools for health authorities to better propose adequate policies. Nowadays, a multitude of mathematical models have been employed to describe the dynamics of infectious diseases. The compartmental models especially the SIR and SIRD models have been widely used, due to their mathematical simplicity. First of all, the research works concerning SIR model are analyzed. Dhanwant and Ramanathan (2020)
analyzed the COVID-19 spread in different countries, and identified the main feature of COVID-19 growth by using the SIR model. To forecast the outcome of the COVID-19, some scholars used the SIR model and estimated the epidemic model parameters (Vattay 2020; Bagal et al. 2020). Some research works have improved the SIR model. Martinez-Guerra and Flores-Flores (2021) proposed the Asymptomatic-Susceptible-Infective-Recovered (A-SIR) model. Yang et al. (2021) proposed a SIR model with time fused coefficients but no time delay. Alenezi et al. (2021) used the sensible SIR model to analyze and predict the outbreak of COVID-19 in Kuwait. Turkyilmazoglu (2021) examined the SIR model to propose an analytical approach for providing an explicit formula associated with a straightforward computation of peak time of the outbreak. Furthermore, in some papers, the effect of prevention and control strategies was taken into account in SIR model. Bittihn and Golestanian (2020) proposed a strategy of containment based on fluctuations in the SIR model. Kumar (2020) used an age-structured SIR model with social distance and Bayesian imputation to study the progress of the COVID-19 epidemic in India. Geng et al. (2021) used the SIR model exploring the ramifications of targeted intervention on spatial patterns of new infections in the population agglomeration template. Deng et al. (2021) proposed a non-smooth SIR Filippov system to investigate the impacts of three control strategies (media coverage, vaccination and treatment) on the spread of an infectious disease. Lazebnik et al. (2021) developed an extended mathematical SIR model, allowing a multidimensional analysis of the impact of non-pharmaceutical intervention on the pandemic. Fu et al. (2021) extended the classic SIR model to find optimal decision making to balance economy and public health in the process of vaccination rollout. Depending on the proportions of variants and the public health strategies adopted, including anti-COVID-19 vaccination, Dimeglio et al. (2021) raised a modified SIR epidemiological model to predict how the spread of the virus in regions of France will vary.

Next, the research works involving SIRD model are analyzed. Many scholars have estimated the parameters of the SIRD model and analyzed the spread of the COVID-19. Fernández et al. (2020) used data on death to estimate a standard SIRD model of COVID-19. Sen and Sen (2021) provided a modified SIRD model to analyze data of the pandemic and estimated and compared the key parameters. Gatta et al. (2021) modeled mobility data through a graph series in order to infer the parameters of SIR and SIRD models. Lobato et al. (2021) applied a methodology based on a double loop iteration process to estimate the parameters of SIRD model. Pacheco and Lacerda (2021) proved with the SIRD model that the approach of quantifying the different rates by means of function estimation was very robust and consistent. Gupth et al. (2021) used the SIRD compartment model for parameter estimation and prediction of COVID-19. Ananthi et al. (2021) predicted the infection spread and recovery rate of the epidemic by simulating model and checked the vulnerability. Additionally, a fractional-order SIRD model (Jahanshahi et al. 2021; Nisar et al. 2021) was introduced to predict the development of the COVID-19. Some scholars used SIRD model to estimate the basic reproduction number of the COVID-19 (Lounis and Raeei 2021; Al-Raeei 2021). The impact of prevention and control measures have been taken into account in SIRD models to analyze this epidemic (Zheng et al. 2021; Jason and Gunawan 2021).

In addition to SIR and SIRD, there were also some research works using other models. Susceptible-Exposed-Infective-Recovered (SEIR) (Yang et al. 2020; Peng et al. 2020) was used to analyze this epidemic. Chen et al. (2020) proposed a novel dynamical system with time delay to describe the outbreak of 2019-nCoV in China. Hellewell et al. (2020) used the stochastic transmission model to quantify the potential effectiveness of contact tracing and isolation of cases at controlling an acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-like pathogen. Liu et al. (2020) developed a Susceptible-Infective-Reported-Unreported (SIRU) model to describe this epidemic. Their research focused on the effect of the Chinese imposed public policies designed to contain this epidemic, and the number of reported and unreported cases that have occurred. Lin et al. (2019) proposed conceptual models for the outbreak in Wuhan with the consideration of individual behavioural reaction and governmental actions. Prediction of NACP and the plateau phases of COVID-19 in China were investigated (Pei 2020; Zeng et al. 2020). Transmission potential and severity of COVID-19 (Remuzzi and Remuzzi 2020; Shim et al. 2020; Fanelli and Piazza 2020) were investigated. Model comparison works (Vytla et al. 2021; Chen et al. 2021) have also been done.

Of course, these researches about the COVID-19 were of great significance, but in the process of modeling, these scholars did not consider time delay and time-varying coefficients at the same time. For the adopted SIRD and SIR model, the incubation delay is non-negligible and should be included. Besides, their coefficients are usually not constant but time-dependent. In this paper, we study the development of the COVID-19 through a non-autonomous SIRD (S-Susceptible, I-Infected, R-Recovered, D-Dead) or SIR (S-Susceptible, I-Infected, R-Removed) epidemic models with time delay. They are two different models, mainly because R represents different meanings. In the SIRD model, R represents the cumulative or total number of the recovered groups, while in the SIR model, R represents the cumulative total of the recovered and dead groups. Since in some cities where the scale of the
The subjects of SIRD model are susceptible, infected, recovered, and dead. The population under study is presumed to be invariant. Obviously, \( S(t) + I(t) + R(t) + D(t) = N \). In the model, natural and birth and death rates are not considered. We use the following symbols to mark the numbers of people in each category:

- **S(t)**: Susceptible, representing the number of people who do not have infectious diseases at time \( t \), but are likely to have infectious diseases;
- **I(t)**: Infected, representing the number of people who get infectious diseases at time \( t \);
- **R(t)**: Recovered, representing the cumulative or total number of the recovered groups at time \( t \);
- **D(t)**: Dead, representing the cumulative or total number of the dead groups at time \( t \).

In the above model, \( \beta \) represents the rate of transmission for the susceptible to the infected. In model (1), we take the values of the incubation delay \( \tau \) as the mean value, 4 days. Let’s take \( \gamma(t) = \kappa t \), where \( \kappa \) is a constant and \( \gamma(t) \) represents the cure rate of the infected people. \( \mu(t) \) represents the death rate of infected cases. According to the model, \( \gamma(t) \) is approximately equal to the number of newly cured cases per day divided by the current number of infected cases on the present day. \( \mu(t) \) nearly equals the number of...
Fig. 1 Curves of cure and mortality rates of different regions in China

(a) Cure rate of Wuhan City
(b) Death rate of Wuhan City
(c) Cure rate of Hubei Province
(d) Death rate of Hubei Province
(e) Cure rate of China Mainland
(f) Death rate of China Mainland
(g) Cure rate of Hubei-Non-Wuhan
(h) Death rate of Hubei-Non-Wuhan
newly deaths per day divided by the current number of infected cases on the present day. From Eq.(1), we obtained the following expressions,

\[
\gamma(t) \approx \frac{R(t+1) - R(t)}{I(t)}, \\
\mu(t) \approx \frac{D(t+1) - D(t)}{I(t)}.
\]  

(2)

SIR model

The subjects of SIR model are the susceptible, infected and removed groups. The removed group includes the dead and cured cases. Since in some regions, COVID-19 is not so serious and the number of infected, recovered and dead cases are very small; sometimes the number of dead cases is even 0, so the SIRD model will not work due to the big relative errors of R and D. We combine R (Recovered) and D (Dead) into one category as R (Removed). We use the following symbols to mark the numbers of people in each category:

\[S(t)\] Susceptible, representing the number of people who do not have infectious diseases at time \(t\), but are likely to have infectious diseases;
\[I(t)\] Infected, representing the number of people who develop infectious diseases at time \(t\);
\[R(t)\] Removed, representing the cumulative total of the recovered and dead groups at time \(t\).

Since Beijing, Shanghai and Zhejiang Province are not the outbreak sites, their numbers of deaths will be relatively small. Therefore we use the SIR model rather than the SIRD model for these regions. By the finite difference method, we find that the removed rates are approximately a linear increasing function. The results of removed rates in Beijing and Zhejiang Province are shown in Fig. 2a, b. So, in the model the removed rates are assumed to be linear functions of time \(t\). Through analysis, we can get the non-autonomous time-delayed SIR epidemic model for these regions as the following:

\[
\frac{dS}{dt} = -\frac{\beta S(t - \tau)I(t - \tau)}{N},
\]

\[
\frac{dI}{dt} = -\eta(t)I + \frac{\beta S(t - \tau)I(t - \tau)}{N},
\]

\[
\frac{dR}{dt} = \eta(t)I.
\]  

(3)

In the above model, \(\beta\) represents the rate of transmission from the susceptible to the infected, \(\eta(t)\) represents the removed rate of infected people. Let’s set \(\eta(t) = \kappa t\), where \(\kappa\) is a constant. According to the model, \(\eta(t)\) is equal to the number of newly removed cases on that day divided by the current number of infected patients on that day. From Eq.(3), we obtained the following expression,

\[
\eta(t) \approx \frac{R(t + 1) - R(t)}{I(t)}.
\]  

(4)

Estimation of parameters in the delayed SIRD and SIR models

Due to the adjustment of the standard of diagnosis of COVID-19 in China on February 12th, 2020 and the revision of data of COVID-19 on April 17th, 2020, we use the data of February 14th as the initial functions of \(S(t), I(t), R(t), D(t)\) for the DDEs (1) and use the data from February 15th, 2020 to February 29th, 2020 to estimate the parameters of the delayed SIRD epidemic model in Wuhan City, Hubei Province, China Mainland, Hubei-Non-Wuhan, then make their long-term predictions and compare them with real data from March 1st, 2020 to April 16th, 2020. We use the data of February 7th as the initial functions of \(S(t), I(t), R(t)\) for the DDEs (3) and use the data from February 8th, 2020 to February 22nd, 2020 to estimate parameters of the delayed SIR epidemic model in Beijing and Anhui Province. We use the data of February
5th as the initial functions of \( S(t) \), \( I(t) \), \( R(t) \) for the DDEs (3) and use the data from February 6th, 2020 to February 20th, 2020 to estimate parameters of the delayed SIR epidemic model in Shanghai and Zhejiang Province, then make the long-term predictions and compare them with real data until the end of the epidemic. Since we do not know exactly the real initial functions, here we have to take the constants functions as the initial functions of the non-autonomous DDEs, which will potentially induce errors.

**SIRD model parameters inversion**

In the SIRD model, the parameters to be estimated are \( \beta \) and \( \kappa \). From Fig. 1b, d, f, we see that \( \mu(t) \) is a piecewise function. \( \mu(t) \) is divided into \( t < 10 \) and \( 10 \leq t \leq 15 \), and we get \( \mu(t) \) by calculating the average value separately. From Fig. 1h, we see that the death rate of Hubei-Non-Wuhan is nearly a constant, and \( \mu(t) \) is obtained by calculating the 15-day average. The value of \( \mu(t) \) is shown in Table 1.

Based on the least square method, we use the Isqnonlin function built in Matlab to carry out parameter inversion and get the optimal parameter solutions. We choose the data of the February 14th as the initial functions of \( S(t) \), \( I(t) \), \( R(t) \), \( D(t) \). Here, we use R-squared to evaluate the results of the parameter estimations of the developed models. Results of model evaluation are shown in Table 2. Based on the official data, we get the results of parameter inversion as shown in Table 3.

**SIR model parameter inversion**

In the SIR model, the parameters to be estimated are \( \beta \) and \( \kappa \), too. As in “SIRD model parameters inversion” section, we choose the corresponding initial functions of Beijing, Shanghai, Zhejiang Province and Anhui Province. We present the results of model evaluation in Table 4 and the results of parameter inversion in Table 5.

**Long-term prediction**

The parameters \( \beta \) and \( \kappa \) obtained in Sect. 3 are substituted into Eqs. (1) and (3) respectively, and the numerical method is carried out to simulate the evolutions of \( S(t) \), \( I(t) \), \( R(t) \), \( D(t) \) or \( S(t) \), \( I(t) \), \( R(t) \).

### Table 1 Death rates in different regions in China

| \( \mu \) | Wuhan | Hubei | China Mainland | HNW |
|---|---|---|---|---|
| \( t < 10 \) | 0.002627 | 0.002358 | 0.001956 | 0.001655 |
| \( 10 \leq t \leq 15 \) | 0.001042 | 0.001147 | 0.001178 | 0.001655 |

### Table 2 Model evaluation

| | Wuhan | Hubei | China Mainland | HNW |
|---|---|---|---|---|
| I | 0.84 | 0.98 | 0.99 | 0.98 |
| R | 0.98 | 0.97 | 0.94 | 0.95 |
| D | 0.99 | 0.99 | 0.99 | 0.95 |

### Table 3 SIRD model parameter inversion

| Parameter | Wuhan | Hubei | China Mainland | HNW |
|---|---|---|---|---|
| \( \beta \) | 0.026103 | 0.021221 | 0.019097 | 0.008281 |
| \( \kappa \) | 0.004649 | 0.005994 | 0.007046 | 0.010862 |

### Table 4 Model evaluation

| | Beijing | Shanghai | Zhejiang | Anhui |
|---|---|---|---|---|
| I | 0.98 | 0.90 | 0.98 | 0.94 |
| R | 0.95 | 0.97 | 0.89 | 0.99 |

### Table 5 SIR model parameter inversion

| Parameter | Beijing | Shanghai | Zhejiang | Anhui |
|---|---|---|---|---|
| \( \beta \) | 0.024581 | 0.029473 | 0.022831 | 0.029021 |
| \( \kappa \) | 0.006505 | 0.008825 | 0.008142 | 0.009658 |

### Long-term predictions of Wuhan City, Hubei Province, China Minland and Hubei-Non-Wuhan

In Fig. 3, the predicted data for current infected is compared with its real data. In Fig. 4, the predicted data for cumulative death is compared with its real data. Since the parameter estimation of the data of China Mainland is over fitting, the long-term prediction error is larger. Here, we take a minor disturbance to the estimated parameters of China Mainland, and then use the disturbed parameters to make its long-term prediction. In Fig. 3, we take \( \beta = 0.019097 \), \( \kappa = 0.006246 \). In these figures, solid curves describe the evolution curves of model (1) with the estimated parameters and points represent the officially published data. The blue dots stand for data used to estimate the parameters, and the red dots are the real data for the long-term prediction. Figure 3 indicates that the numbers of current infected will decrease after a peak. The peak is the turning point of the epidemic and the arrival of the turning point shows that the epidemic is under control. Wuhan City, Hubei Province and China Mainland are
expected to end the epidemic at the end of April. Hubei-Non-Wuhan is expected to end the epidemic at the end of March. Figure 4 shows that the predicted numbers of final death are very close to their actual data. And both the total tendencies of the current infected and dead cases agree very well with each other.
Long-term predictions of Beijing, Shanghai, Zhejiang and Anhui Provinces

In Fig. 5, we compare the predicted data for the current infected with their real data. Since the parameter estimation of the data of Beijing also appear over fitting, the long-term prediction error is larger. Here, we take a minor disturbance to the estimated parameters of Beijing, and then use the disturbed parameters to make predictions. In Fig. 5, we take $\beta = 0.025581$, $\kappa = 0.005905$. And the total tendencies agree very well with each other. In these figures, solid curves describe the evolution of model (3) with the estimated parameters and points represent officially published data. The blue dots stand for real data used to estimate the parameters, and the red dots are the real data for the long-term prediction. Figure 5 suggests that the numbers of the current infected decrease after the turning point and the epidemic is gradually under control. Beijing are expected to finish the epidemic at the middle of April. Shanghai, Zhejiang and Anhui Provinces are expected to finish the epidemic at the end of March.

Conclusion

SIRD and SIR epidemic models are classical and effective mathematical models of infectious disease. The SIRD model is suitable for epidemics with a large scale and a large number of death. If the scale of the epidemic is small and the death rate is low, the prediction error with the SIRD model may increase, and it is better to use the SIR model. During the modeling phase, we researched many complex models, such as the SEIR(D) model, where E represents the number of the exposed persons. Since the data of exposed persons has been not accurate statistics, there is no way to predict it precisely. Finally, we find that the simpler the model is the better the prediction effect. It is advisable to employ the SIRD and SIR to predict the COVID-19.

In this paper, SIRD model is used to describe the development of COVID-19 in Wuhan City and other regions and SIR model is used to establish the development of COVID-19 in Beijing and other regions. The simulation results indicate that the long-term predictions are in good agreement with the actual data published by the government. Our dynamic SIRD and SIR models are very effective in predicting the tendencies, especially the turning points, the end time and sizes of the current infected of COVID-19 epidemic and the highest numbers of death.

The recovery rate significantly increases and the death rate gradually decreases. From these figures, we can also estimate the duration and end of the COVID-19. Because the government has imposed the very strict, scientific and effective containment measures, we get a lower and lower rate of transmission and a relatively quick control of the epidemic. As China concentrated all its efforts on curing
the infected and researching treatment options, the cure rate increases and death rate decreases quickly. Our paper may provide us an easy and effective method to predict the long-term evolution of COVID-19.

Acknowledgements
The authors would like to acknowledge the financial support for this research via the National Natural Science Foundation of China (Nos. 11972327, 11372282 and 10702065). They also thank the reviewers for their valuable reviews and kind suggestions.

References
Al-Raei M (2021) The basic reproduction number of the new coronavirus pandemic with mortality for India, the Syrian Arab Republic, the United States, Yemen, China, France, Nigeria and Russia with different rate of cases. Clin Epidemiol Glob Health 9:147–149

Alenezi MN, Al-Anzi FS, Alabdalrazzaq H (2021) Building a sensible SIR estimation model for COVID-19 outbreak in Kuwait. AEJ - Alex Eng J 60(3):3161–3175

Ananthi P, Begum SJ, Jothi VL et al (2021) Survey on Forecasting the vulnerability of COVID-19 in Tamil Nadu. J Phys Conf Ser 1767(1):012006

Bagal DK, Rath A, Barua A et al (2020) Estimating the parameters of susceptible-infected-recovered model of COVID-19 cases in India during lockdown periods. Chaos, Solitons Fractals 140:110154

Bittihn P, Golestanian R (2020) Containment strategy for an epidemic based on fluctuations in the SIR model. arXiv 1–6

Chen Y, Cheng J, Jiang Y et al (2020) A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification. J Inverse Ill-posed Prob 28(2):243–250

Chen Y, Liu F, Yu Q et al (2021) Review of fractional epidemic models. Appl Math Model 97(4):281–307

Deng J, Tang S, Shu H (2021) Joint impacts of media, vaccination and public health interventions policies on pandemic dynamics and the economy: the case of COVID-19. Econ Res-Ekon Istat 1–29

Lin Q, Zhao S, Gao D et al (2020) A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. Int J Infect Dis 93:211–216

Liu Z, Magal P, Seydi O, et al (2020) Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. arXiv:2002.12298

Lobato FS, Platt GM, Libotte GB et al (2021) Formulation and solution of an inverse reliability problem to simulate the dynamic behavior of COVID-19 pandemic. Trends Comput Appl Math 22(1):91–107

Lounis M, Raei MA (2021) Estimation of epidemiological indicators of COVID-19 in algeria with an SIRD model. Eurasian J Med Oncol 5(1):54–58

Martnez-Guerra R, Flores-Flores JP (2021) An algorithm for the robust estimation of the COVID-19 pandemic’s population by considering undetected individuals. Appl Math Comput 405:126273

Niser KS, Ahmad S, Ullah A et al (2021) Mathematical analysis of SIRD model of COVID-19 with Caputo fractional derivative based on real data. Results Phys 21:103772

Pacheco CC, Lacerda C (2021) Function estimation and regularization in the SIRD model applied to the COVID-19 pandemics. Inverse Probl Sci Eng 1–16

Pei L (2020) Prediction of numbers of the accumulated confirmed patients (NACP) and the plateau phase of 2019-nCoV in China. Cogn Neurodyn 14(3):1–14

Peng L, Yang W, Zhang D et al (2020) Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv:2002.06563

Remuzzi A, Remuzzi G (2020) COVID-19 COVID-19 and Italy: what next? Lancet 395(10231):1225–1228

Sen D, Sen D (2021) Use of a modified sird model to analyze COVID-19 data. Ind Eng Chem Res 60(1):4251–4260

Shim E, Tariq A, Choi W et al (2020) Transmission potential and severity of COVID-19 in South Korea. Int J Infect Dis 93:339–344

Turkyilmazoglu M (2021) Explicit formule for the peak time of an epidemic from the SIR model. Phys D 422:132902

Vattay G (2020) Forecasting the outcome and estimating the epidemic model parameters from the fatality time series in COVID-19 outbreaks. Phys Biol 17(6):065002

Vytla V, Ramakuri SK, Peddi A et al (2021) Mathematical models for predicting COVID-19 pandemic: a review. J Phys Conf Ser 1797(1):012009

Yang Z, Zeng Z, Wang K et al (2020) Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis 12(2):165–174

Yang HC, Xue Y, Pan Y, et al (2021) Time fused coefficient SIR model with application to COVID-19 epidemic in the United States. J Appl Stat 1–15
Zeng T, Zhang Y, Li Z, et al (2020) Predictions of 2019-nCoV transmission ending via comprehensive methods. arXiv: 2002.04945

Zheng Z, Xie Z, Qin Y et al (2021) Exploring the influence of human mobility factors and spread prediction on early COVID-19 in the USA. BMC Public Health 21(1):615

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.