An Improved SEIR Model for Reconstructing the Dynamic Transmission of COVID-19

Xiaohan Yuan[1], Shuyu Chen[1*], Lu Yuwen[1], Shuqi An[1], Sha Mei[1], Tianqiang Chen[1]

[1]School of Big Data & Software Engineering, Chongqing University, Chongqing, China

*Corresponding author: Shuyu Chen (sychen@cqu.edu.cn)

Email: {xhyuan, sychen, luyuwen, shuqian, shamei, tqchen}@cqu.edu.cn

Abstract—With the recent outbreak of coronavirus disease 2019 (COVID-19), human life and the world economy have been severely affected, the propagation and scale of COVID-19 is top of mind for everyone. To reconstruct the development trend of COVID-19, we investigate the issue of the epidemic spreading process under vigorous non-pharmaceutical interventions. Here, an improved Susceptible-Exposed-Infectious-Recovered (SEIR) model with dynamic variables (i.e., health exposure individuals and close contacts) is proposed to predict the scale of COVID-19 and its dynamic evolution. We assume that the number of contacts and the reproduction number of COVID-19 changes dynamically over time. Then a gradient descent method is applied to estimate the effective reproduction number. We use the proposed model to reconstruct the dynamic transmission of COVID-19 in Chongqing between 14 January and 24 March 2020. The results show a similar development trend with a real-world epidemic. Our work has important implications when considering strategies for continuing surveillance and interventions to eventually contain outbreaks of COVID-19.

1. Introduction

In December 2019, the Wuhan Municipal Health Commission reported a group of novel coronavirus disease (COVID-19) of infecting the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1], [2] in China. As the COVID-19 outbreak continues and spreads rapidly all over the world due to massive human movement, governmental bodies began to adopt a series of non-pharmaceutical interventions to contain the geographical spread of the disease, consisting of suspension of all intra- and inter-city transportation, compulsory mask wearing in public places, cancellation of social gatherings and the home quarantine of individuals with presumed infections [3]. These interventions can effectively crush the epidemic curve and reduce the attack rate, which can potentially shed light on global efforts to control the outbreak of the COVID-19 epidemic.

To prevent another outbreak of the COVID-19 epidemic, it is necessary to interpret transmission dynamics and epidemic trajectories of the COVID-19. In this scenario, epidemiological models can be utilized to predict the future course of the disease. Meanwhile, the impact of non-pharmaceutical interventions and related control measures used to slow the contagion can be estimated via epidemiological models [4]. Therefore, the investigation of epidemiological models can provide precious time and potential knowledge to enhance health care resources and develop effective immunological defenses [5]. However, there are many bottlenecks to reconstruct and predict the dynamic transmission of the COVID-19 via epidemiological models. First, individuals have mobility. All individuals (i.e., infectious individuals and healthy individuals) are moving frequently and randomly, especially, some social activities such as Chinese New Year will catalyze individuals movements [6]. Second, the location of the individuals is uncertain, making it difficult to predict the number of new infectious individuals. Last, asymptomatic infectious individuals have high covertness.

The Susceptible-Exposed-Infectious-Recovered (SEIR) model has been widely used to predict the dynamic transmission of the COVID-19 [7], [8]. The SEIR model can estimate the number of infections in the next few days and can derive the evolution of urban epidemics and provide support for prevention and control decisions. He et al. [8] built an SEIR epidemic model for COVID-19 and discussed the nonlinear dynamics of the system by introducing the seasonality and stochastic of the parameters. Zhong et al. [9] proposed an early prediction of the COVID-19 outbreak based on a simple SEIR model and limited epidemiological data in China. They predicted the number of cumulative cases between 76000 and 230000, and about 22,000 to 74,000 unrecovered cases of infection until early March. Wu et al. [10] combined monthly flight booking information and mobile data provided by Tencent to trace the contact of confirmed cases, inferring the international spread of COVID-19 based on the SEIR model. Yang et al. [11] combined the SEIR and multi-agents simulation model to explain that the exposed state is the key to spreading the epidemic. Bhati et al. [12] adopted the SEIR model to simulate COVID-19 spread under the effect of the Bhilwara model of containment policies [13]. Although the above works have achieved good results via the SEIR model with the contacts of confirmed cases, they use a fixed basic reproduction number and set the number of contact (or contact rate) as the fixed value which is contrary to the real-world scenario for COVID-19.

In epidemiology, the basic reproduction number refers to the average number of secondary cases produced by an infected case without external intervention and all individuals have no immunity, written as R_0. Some scholars have proposed methods for the estimation of R_0 of COVID-19. They estimate R_0 as a fixed constant value. Wu et al. [10] estimated that the R_0 is 2.68 (95% CrI 2.47–2.86) by Stochastic Markov Chain Monte Carlo method based on the data of the COVID-19 epidemic from Dec 31, 2019 to Jan 28, 2020. An investigation has indicated that the estimated value of R_0 ranges from 1.4 to 6.49, with a median of 2.79, an average of 3.28, and an interquartile range of 1.16 [14]. In fact, R_0 is mainly affected by the duration of the infection period, the probability of infecting susceptible individuals in one contact, and the number of new susceptible individuals contacted per unit time [15]. Therefore, the effective reproduction number similar to R_0 is more reasonable to be used in the SEIR model due to its dynamic characteristics.

The reasons why the contact changes dynamically over time are as follows. First, the infectious individuals are all generated from contacts due to close contact. As the number of infectious cases increases, the possibility of contact repeating will be greater. If the contact rate has been maintained at the same fixed value as before, it may cause the predictive number of patients to be higher. Second, following the cordon sanitarian of Wuhan, other provinces and cities in China have successively initiated primary response to major public health emergencies to control the flow of population. Except for family members who live together, everyone basically does not touch other people (but indirect contact through objects cannot be ignored, such as elevators, corridors, etc.). This will lead to a rapid reduction in the number of contacts with COVID-19 cases, but not immediately.

In this paper, we are motivated to exploit an improved SEIR model for reconstructing the dynamic transmission of COVID-19. The contacts are divided into two classes the health exposure individuals and the close contacts which conform to the real-world scenario. We propose a dynamic health-close-SEIR (HC-SEIR) model with the health exposure individuals and the close contacts, where two classes change dynamically over time. We utilize a non-linear fitting method with a least squares method to estimate the value of two classes. A gradient descent method is proposed to estimate the effective reproduction number. A dataset from Chongqing between 14 January and 24 March 2020 is divided into a training set and a testing set for parameter estimation and performance evaluation. Experiments and results demonstrate the effectiveness of the proposed model in the dataset from Chongqing.

The rest of this paper is organized as follows. In Section 2, the Dynamic-HC-SEIR model is built. In Section 3, we use a COVID-19 dataset from Chongqing to train the proposed model and estimate the parameters of COVID-19. In Section 4, results are illustrated and conclusions are drawn in Section 5.

2. Reconstruction Method of the COVID-19

The COVID-19 has been effectively controlled in some areas such as China. To prevent the epidemic from breaking out again, we propose an improved SEIR model to reconstruct the outbreak of the epidemic, analyze the regular pattern of the spread of infectious diseases, and eventually put forward suggestions to contain the spread of the COVID-19.

2.1. The Dynamic-HC-SEIR Models

In this paper, we propose a Dynamic-HC-SEIR model via introducing health exposure individuals (H), close contacts individuals (C), and adjusting the definition of variables. The total population of a certain area is divided into six distinct epidemic classes: Susceptible individuals (S), Health-exposure individuals (H), Close-contacts individuals (C), Exposed individuals (E), Infectious individuals (I), and Recovered individuals (R). We denote the number of population in the different six classes at time slot t by $S(t)$, $H(t)$, $C(t)$, $E(t)$, $I(t)$, and $R(t)$, respectively.

Note that the population migration data of a certain area should also be taken into account, because the outbreak of COVID-19 epidemic coincides with Chinese New Year, which have high population mobility. Therefore, we introduce the number of move-in populations $In(t)$ and the number of move-out population $Out(t)$ in the proposed model. Particularly, in our model, the latent infectious individuals and asymptomatic infectious individuals of the COVID-19 [1] form E-class. According to the report by the Chinese Center for Disease Control and Prevention, asymptomatic infections refer to no relevant clinical symptoms, such as fever, cough, sore throat, and other self-perceived or clinically identifiable symptoms and signs, but respiratory tract and other specimens show a positive test for the pathogen of COVID-19.

For the COVID-19, the S-class consists of people who are at risk of becoming infected by the SARS-CoV-2. The close contacts have closely contacted with pathogen carriers, mainly from relatives and friends and colleagues who have dinner together. However, the H-class refers to individuals who have unconsciously contacted with pathogen carriers in public, but do not know each other. The pathogen carriers are those who carry the SARS-CoV-2. They are mainly composed of patients who have been infected with the SARS-CoV-2, possible virus carriers, etc. The latter mainly come from people in epidemic areas, personnel in areas with severe outbreaks of infectious diseases. The individuals in the exposed class are in latent period and these are asymptomatic but possibly infective. It is worth noting that some of the infected cases may not have symptoms at all. Unless they are actively checked, they cannot confirm whether they have been infected until they recover. The I-class consists of highly infectious individuals who have obvious symptoms and have been medically diagnosed. We consider these individuals have been isolated due to the

---

1. In the paper, we call it asymptomatic infection for short.

current clinical diagnosis and treatment process in China, and therefore cannot transmit the disease to susceptible individuals. The removed individuals are those who have recovered from COVID-19 with acquired immunity or died. Hence, we propose the Dynamic-HC-SEIR model shown as in Fig. 1.
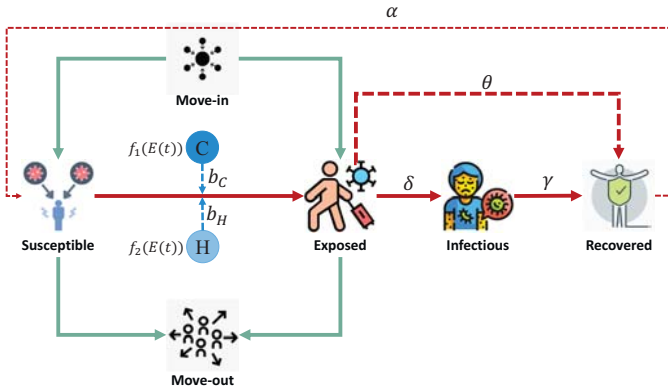

Figure 1. Structure of The Improved SEIR Model

The traditional SEIR model is expressed by

$$
\begin{cases}
\frac{dS(t)}{dt} = -\frac{\beta S(t) I(t)}{N}, \\
\frac{dE(t)}{dt} = \frac{\beta S(t) I(t)}{N} - \delta E(t), \\
\frac{dI(t)}{dt} = \delta E(t) - \gamma I(t), \\
\frac{dR(t)}{dt} = \gamma I(t),
\end{cases}
\tag{1}
$$

where the total population N is a constant and it is assumed that birth rate is the same as death rate. However, in the real-world situation, the outbreak of COVID-19 coincides with Chinese New Year, which promotes the movement of people. Therefore, the N is considered as a dynamic value with the change of the number of migrant population. Due to the non-pharmaceutical intervention control measures taken by the government, the contact transmission rate β that represents the probability of the spreading disease from an infectious to a susceptible individual will be a dynamic parameter over time t, which is expressed by the number of the contact. Thus, the differential equation of S is modified as follow by introducing move data, dynamic number of the contact, and considering the probability of COVID-19 re-infection

$$
\frac{dS(t)}{dt} = S_{in}(t) - S_{out}(t) - \frac{S(t) * C(t) * b_C}{N(t)}
- \frac{S(t) * H(t) * b_H}{N(t)} + \alpha * R(t),
\tag{2}
$$

where $C(t), H(t)$ are related to the number of infections and the range of movement, and they can be fitted by the dataset. The number of susceptible population moving into an area from all other areas is expressed as

$$
S_{in}(t) = In(t) * (1 - \lambda).
\tag{3}
$$

Here, the migrants are considered to have no symptoms of COVID-19. The λ is regard as the probability of latent infected individual moving out a area. So the number of susceptible population moving out an area is expressed as

$$
S_{out}(t) = Out(t) * (1 - \lambda).
\tag{4}
$$

Similar to the above equation, in addition to increasing the number of latent infected individuals from S-class, the E-class also subtracts the number of translating to confirmed infection cases and recovered in the current unit time. So, the difference between the number of latent infection at time slot $t + 1$ and slot t is

$$
\frac{dE(t)}{dt} = E_{in}(t) - E_{out}(t) + \frac{S(t) * C(t) * b_C}{N(t)} +
\frac{S(t) * H(t) * b_H}{N(t)} - \delta * E(t) - \theta * (1 - \delta) * E(t).
\tag{5}
$$

The number of latent infection population moving into an area from all other areas is calculated by

$$
E_{in}(t) = In(t) * \lambda,
\tag{6}
$$

and the number of moving out is as follow

$$
E_{out}(t) = Out(t) * \lambda.
\tag{7}
$$

The number of newly confirmed patients based on time t can be expressed by the following differential equation

$$
\frac{dI(t)}{dt} = \delta * E(t) - \gamma * I(t).
\tag{8}
$$

Naturally, the differential equation of the number of recovered individuals is written by

$$
\frac{dR(t)}{dt} = \theta * (1 - \delta) * E(t) + \gamma * I(t) - \alpha * R(t).
\tag{9}
$$

In this model, it is obvious that N is not the total population of the area.

$$
N(t) = S(t) + E(t) + I(t) + R(t).
\tag{10}
$$

Therefore, our model can be summarized as

$$
\begin{cases}
\frac{dS(t)}{dt} = S_{in}(t) - S_{out}(t) - \frac{S(t)*C(t)*b_C}{N(t)} \\
\quad - \frac{S(t)*H(t)*b_H}{N(t)} + \alpha * R(t), \\
\frac{dE(t)}{dt} = E_{in}(t) - E_{out}(t) + \frac{S(t)*C(t)*b_C}{N(t)} \\
\quad + \frac{S(t)*H(t)*b_H}{N(t)} - \delta * E(t) - \theta * (1 - \delta) * E(t), \\
\frac{dI(t)}{dt} = \delta * E(t) - \gamma * I(t), \\
\frac{dR(t)}{dt} = \theta * (1 - \delta) * E(t) + \gamma * I(t) - \alpha * R(t).
\end{cases}
\tag{11}
$$

2.2. Dynamic Parameters with Statistical Model

Contact tracing of COVID-19 can predict and control the spread of the epidemic. According to the *Centers for Disease Control and Prevention* for the determination of close contacts of COVID-19 cases, the close contact C is defined as any individual who has not taken effective protection within 6 feet of an infected person for at least 15 minutes. Usually, the close contact consists of people who live in the same house, eat together, or have frequent social interactions with infected individuals in the 14 days before confirmed. Hence, the close contact and the confirmed cases are related. People unconsciously contacted pathogen carriers are not included in this group (eg. contact in public places, but do not know each other).

As the COVID-19 appeared before the Chinese New Year, the effective prevention and control measures were not taken, residents can go to work, go shopping, travel at home and abroad, etc. as usual until Wuhan is closed on January 23, 2020. The flow of population is large, and causes the epidemic to spread ragingly. The Health-exposure individuals are those who have unconsciously contacted pathogen carriers. They mainly refer to contact with pathogen carriers in public places, such as supermarkets, parks, bars, cinema, public transportation, etc, but do not know each other. It can be determined by tracking the source of infection of COVID-19 cases.

Through the personal and family kinship (spouse, parents, children, etc.) information and other related information of family members in the household data of the infected individuals, the potential close contacts can be determined. Combined with the dynamic analysis of the mobile phone location data of pathogen carriers by communication operators, the trajectory and regularity of the staged activities of pathogen carriers are found, and health exposure individuals can be determined. The purpose of tracing contacts is to shorten the transmission time of the case in the community, so as to reduce the R_0 which is the daily reproduction number of SARS-CoV-2. In our model, the number of close contacts and health exposure individuals are regarded as dynamic variable $C(t)$, $H(t)$ that changes over time t, and are related to the daily number of asymptomatic infections. The daily number of asymptomatic infections can be determined by the daily number of newly confirmed infections announced by the Health Commission and the incubation period of the disease.

We fit the number of close contact by Non-linear function

$$\frac{C(t)}{E(t)} = f_1(t). \tag{12}$$

Similar to $C(t)$, we also set the number of health exposure individuals as

$$\frac{H(t)}{E(t)} = f_2(t). \tag{13}$$

The fitting degree of trend line (coefficient of determination) is determined by

$$\aleph^2 = \frac{SS_{tot} - SS_{res}}{SS_{tot}}, \tag{14}$$

which is used in statistics to measure the proportion of the independent variable explained by the change of the dependent variable. Supposing a data set includes n observation values of $y_1, \ldots, y_n$, and the corresponding model prediction values are $f_1, \ldots, f_n$. We define the residual $e_i = y_i - f_i$ and the average observation value is $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. So the sum of total squares is obtained by

$$SS_{\text{tot}} = \sum_{i=1}^{n} (y_i - \bar{y})^2, \tag{15}$$

and the sum of residual squares is

$$SS_{\text{res}} = \sum_{i=1}^{n} (y_i - f_i)^2 = \sum_i e_i^2. \tag{16}$$

We collect demographic information about the COVID-19 cases reported by the Chongqing Municipal Health Commission, and inspired by some scholars on estimating the transmission rate of COVID-19, we assume R_0 corresponds to the slope of a logistic curve

$$R_0(t) = \frac{Ae^{-k(t+m)}}{\left(1 + e^{-k(t+m)}\right)^2}, \tag{17}$$

where t is the number of days following the time point of the data, A, k, m are unknown parameters to be estimated. To obtain the R_0 over time t, we combine the formula (19) to minimize the loss function (18) to estimate the parameters A, k and m.

$$Loss = \frac{1}{n} \sum_{t=1}^{n} \left(I_{\text{esti},t} - I_{\text{real},t}\right)^2, \tag{18}$$

where $I_{\text{real},t}$ is the real value of confirmed cases published on the official website and $I_{\text{esti},t}$ is calculated by Dynamic-HC-SEIR model. The parameters is acquired by an optimization algorithm of the gradient descent.

$$R_0(t) \propto \frac{b_H * H(t) + b_C * C(t)}{u_1 * E(t) + u_2 * I(t)}. \tag{19}$$

3. Estimation of the Proposed Model Parameters

We use the COVID-19 dataset of Chongqing about the number of reported confirmed, curative, and death cases from January 21 to March 17, 2020. The real data before February 14 is selected as training data to estimate the parameters of system. The training set contains the number of close-contact and health-exposure individuals provided by Chongqing Nan'an District Health Commission, and part of the data is listed in Table 1. For the sake of simplicity when writing the form, we use the field name *CAI* to indicate

TABLE 1. THE COVID-19 DATA FROM NAN'AN

Date	CAI	NAI	NCC	NHE
Jan 14,2020	1	1	3	56
Jan 15,2020	1	0	1	36
Jan 16,2020	2	1	1	62
Jan 17,2020	3	1	2	123
Jan 18,2020	3	0	2	122
Jan 19,2020	4	1	4	136
Jan 20,2020	7	3	13	423
Jan 21,2020	7	0	4	387
Jan 22,2020	8	1	5	368
Jan 23,2020	8	0	5	364
Jan 24,2020	8	0	7	365
⋮				

TABLE 2. DESCRIPTION OF THE MODEL PARAMETERS

Symbol	Description
δ	The transition rate from the exposed to infected class
θ	The removed rate of asymptomatic infections
γ	Average recovery or death rate of symptomatic infection
α	Temporary immunization rate of recovered individuals
b_C	Probability of close contacts carrying the virus
b_H	Probability of health-exposure persons carrying the virus
λ	Probability of latent infected individuals moving out a area.

Figure 2. Migration index for Chongqing during Chinese New Year, 2020.

cumulative asymptomatic infections cases, *NAI* to indicate new asymptomatic infections cases, *NCC* to indicate new close contacts, and *NHE* to indicate new health exposure.

To make the Dynamic-HC-SEIR model more universal, it is necessary to discuss the parameter identification. The description of the model parameters is illustrated in Table 2. In our model, the parameters b_C and b_H represent the infection probability of close-contact individuals and health-exposure individuals, respectively. They can be estimated by the actual COVID-19 epidemic data provided by Nan'an District, Chongqing. Actlly, the b_C is equal to the average of the ratio that is the number of new latent infections (Exposed individuals) on day t to the number of new close contacts on day $t-1$. The b_H is equal to the average of the ratio that is the number of new latent infections on day t to the number of new health exposure on day $t-1$. Therefore, during the spread of the COVID-19, the parameter b_C is calculated by

$$b_C = \frac{1}{n} \sum_{t=1}^{n} \frac{\Delta E(t)}{C(t-1)}, \qquad (20)$$

and b_H is equal to

$$b_H = \frac{1}{n} \sum_{t=1}^{n} \frac{\Delta E(t)}{H(t-1)}. \qquad (21)$$

The values for b_C and b_H are 0.1784 [95% confidence interval (CI),0.0546 to 0.3023] and 0.0034 (95% CI, 0.0010 to 0.0059), respectively.

The parameter δ is the transition rate from the exposed to infected class. Its initial value is calculated by $\delta = \frac{1}{T}$, where T is the incubation period of the SARS-CoV-2, which has been reported to be generally between 2 to 14 days, and the
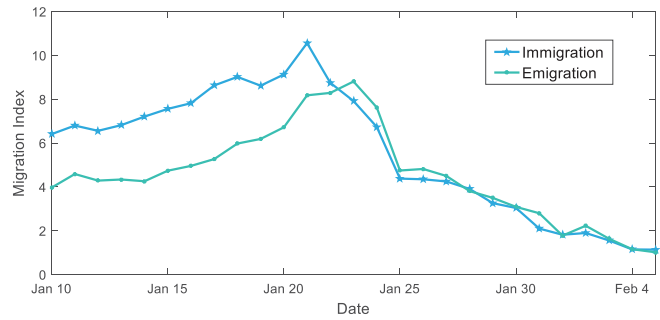
longer one can even be 28 days. one study has pointed out that the median incubation period of the new coronavirus is estimated to be 5.1 days, and 97.5% of infected people will develop symptoms within 11.5 days of infection [16]. We choose the mode value of 14 days.

Asymptomatic infections are as infectious as symptomatic infections. A study found that in 24 asymptomatic cases, the median period from positive to negative viral nucleic acid in asymptomatic patients is 9.5 days with the interquartile ranges from 3.5 to 13.0 days, and the longest is 21 days [17]. However, according to China's epidemic prevention and control measures, asymptomatic infections should be isolated for 14 days. At present, the duration of infection in asymptomatic cases is still being followed with interest. Here, we utilize the above information to set the initial value of θ, the removed rate of asymptomatic infections, to 0.1646 (95% CI, 0.0608 to 0.2684).

We consider confirmed patients to be isolated in the hospital for treatment. When they are confirmed, patients with COVID-19 have a median 10.7 day hospitalizations [18]. For γ, the average recovery or death rate of the confirmed patient, we set $\gamma = \frac{1}{D}$ as an initial value, where D is the average length of time from the date of diagnosis to discharge from the hospital, and $D = 10.7$. The case fatality rate of symptomatic patients is approximately 4.38% [19], which illustrates that nearly 95% of the confirmed patients in the hospital can recover. However, there have been reports that the recovered patients may be infected again [20]. According to the time node of the first case of re-infection of the new crown in Hong Kong, we take the initial value of α to 0.0067 (95% CI, 0.0013 to 0.0317), which is regarded as the temporary immunization rate of recovered individuals.

For Nan'an District of Chongqing City, we utilize the estimated data from the 14th day before the first confirmed patient for simulation, so the data involved in the modeling starts from January 14, 2020. To successfully reconstruct the spread of the COVID-19 epidemic in Chongqing, in the model training phase, the initial value of the model variables select the data on January 14.
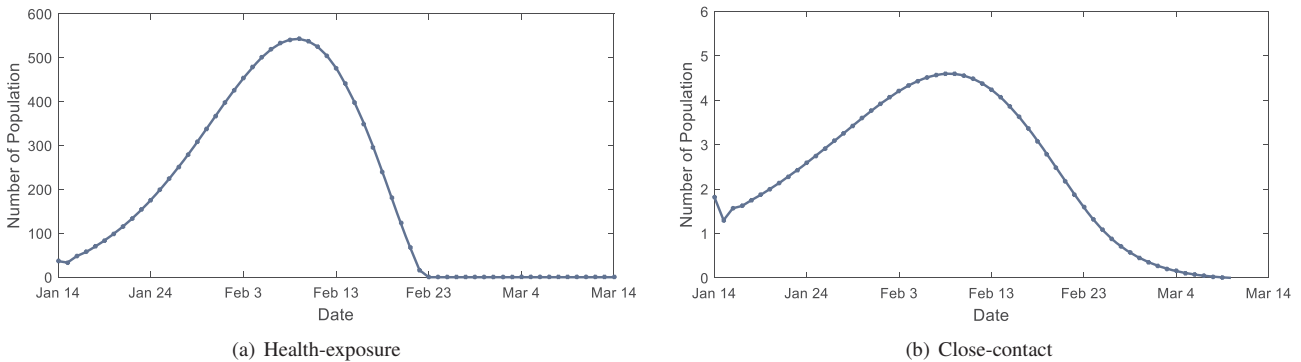
(a) Health-exposure

(b) Close-contact

Figure 3. Contact of COVID-19 cases from Nan'an District, Chongqing Municipality

4. Experiments and Results

We research the COVID-19 epidemic in Chongqing Municipality, a neighboring city of Wuhan where the epidemic first broke out. As one of the cities closest to the outbreak, Chongqing is once in a very dangerous situation. On the evening of January 21, 2020, the National Health Commission confirmed the first confirmed case of COVID-19 with an imported case in Chongqing. Its complete information on confirmed cases of COVID-19 and the data of the urban migration population during Chinese New Year has been obtained before the time of writing. It can be seen from Figure 2 that before January 22, the number of immigration in Chongqing is much larger than the number of emigration and after the lockdown of Wuhan on January 23, the immigrant population is slightly lower than the emigration with gradually towards 0 and flatten. Because the local public health intervention measures have been strictly implemented and the residents' awareness of self-isolation at home has been significantly improved follow Wuhan's intervention policy.

We also obtained the daily number of health exposures $H(t)$ and close contact individuals $C(t)$ provided by the Chongqing Nan'an District Health Committee. The fitted curve of health exposure $H(t)$ and close contacts $C(t)$ is shown in Figure 3, and their coefficients of determination are 0.2879 and 0.4323, respectively. The changes of $H(t)$ and $C(t)$ are closely related to the number of $E(t)$ in Figure 4. As of February 13, the increase in the values of $H(t)$ and $C(t)$ is consistent with the change in the number of exposed class, which reached the maximum on February 13. Later, with the strengthening of intervention measures, most of $E(t)$ is found and isolated, causing $H(t)$ and $C(t)$ to drop and gradually tend to 0.

The estimated trajectories of effective reproduction number R_0 is shown in Figure 5. The fitting results of R_0 introduce that the maximum value of R_0 is 2.25, and $R_0 < 1$ after 10 days of isolation measures, which indicate that the epidemic in the Chongqing municipality has been well controlled and the transmissibility dropped from the initial 1.93 to less than 0.9.
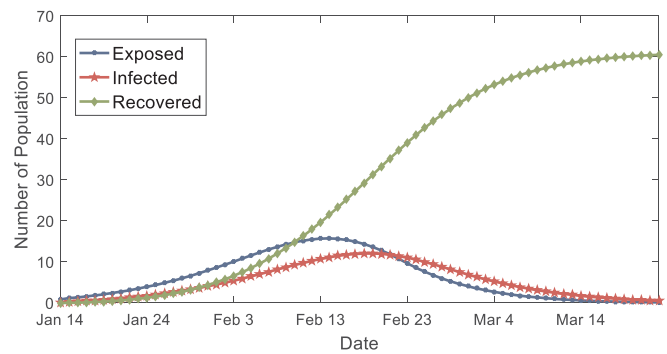

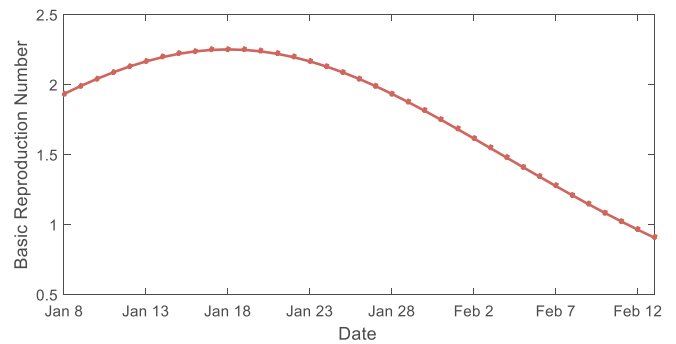Figure 4. The estimated cases of COVID-19 in Nan'an

Figure 5. The fitting results of reproduction number R_0 of COVID-19 in Chongqing

The Dynamic-HC-SEIR is an epidemic transmission model based on the traditional SEIR, which is used to estimate the spreading trend of the current cases and the epidemic dynamics. We consider the migration data $In/Out(t)$ into the model and assume that the basic reproductive number R_0 of COVID-19 changes dynamically with time t. Our model uses the epidemic data reported from January 8, 2020 to February 12, 2020 as the training set to calculate the parameters. In this training phase, the stochastic gradient descent method is used to obtain the non-statistical param-

eters of the model. Compared with the traditional SEIR model with Root Mean Square Error (RMSE) of 42.9802, our model has an error only about 14.5481 for the infected cases. The fitting results of the estimated curve and the actual reported data are shown in Figure 6. We can see that our training model fits the change of the actual number of cases well from the subgraph (*Infected*). Particularly, the number of population about the estimated E is equal to the number of population about the actual E on January 27, and the estimated E and the actual E reach their peaks at the same time for our proposed model in the subgraph (*Exposed*). Furthermore, it can be seen that the Dynamic-HC-SEIR model follows the direction of changes in the real data well.
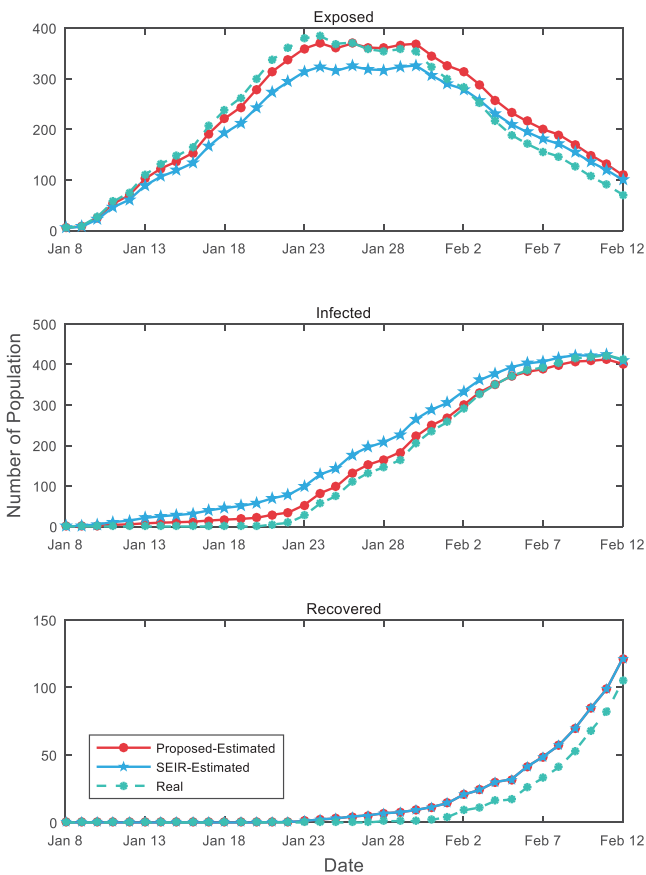

Figure 6. The fitting results in the training phase

This result is also verified by the actual reported infectious populations illustrated in Figure 7. In the testing phase, although there are some deviations between the number of estimated cases (our proposed and traditional SEIR) and the number of actual cases, it can be found that the estimated values of our model fit better with the real values. Moreover, our model predicts the spread trend of the COVID-19 better than the traditional SEIR model, and the test error of 21.7974 is obviously lower than that of 139.2414 in the traditional SEIR model. From the predicted results shown

in Figure 7 (*Infected* and *Recovered*), we find our predicted values are more in line with the actual datas, and the values from February 13 to February 20 are very close to the actual values for the two variables I and R. We observe that the infected cases peak is 413 to be reached on Feb 11, and the COVID-19 gradually ends after Mar 15.
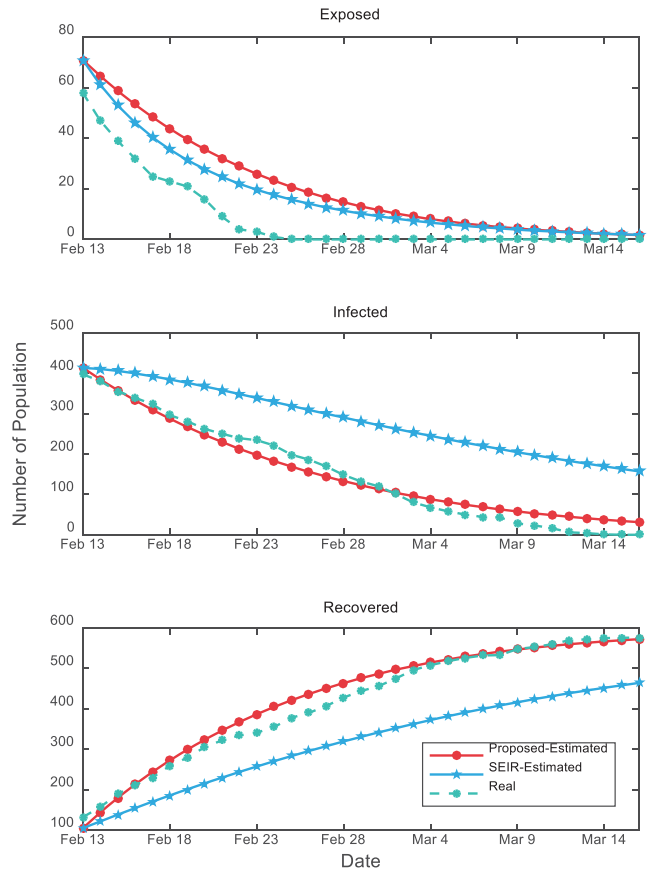

Figure 7. The fitting results in the testing phase

5. Conclusion

In this paper, a dynamic HC-SEIR model has been proposed to reconstruct the transmission of the COVID-19, where the contacts H, C and the effective reproductive number R_0 are considered into this model as dynamic parameters over time. The epidemic has been reconstructed by introducing migrant population data based on the proposed model in Chongqing. We have divided the data into a training set and a test set according to a certain proportion, and have estimated the parameters of the proposed model by the gradient descent method. Our model has simulated the spread of the COVID-19 epidemic well, and has obtained a same development trend with real-world scenario.

Acknowledgments

References

[1] A. J. Kucharski et al., "Early dynamics of transmission and control of covid-19: a mathematical modelling study," *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 553 – 558, 2020.

[2] Q. Li et al., "Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia," *New England Journal of Medicine*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] A. Kuzdeuov et al., "A network-based stochastic epidemic simulator: Controlling covid-19 with region-specific policies," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2020.

[4] W.-j. Guan et al., "Clinical characteristics of coronavirus disease 2019 in china," *New England Journal of Medicine*, vol. 382, no. 18, pp. 1708–1720, 2020.

[5] J. Yawney and S. A. Gadsden, "A study of the covid-19 impacts on the canadian population," *IEEE Access*, vol. 8, pp. 128240–128249, 2020.

[6] P. Di Giamberardino, D. Iacoviello, F. Papa, and C. Sinisgalli, "Dynamical evolution of covid-19 in italy with an evaluation of the size of the asymptomatic infective population," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2020.

[7] Y. Cai, Y. Kang, and W. Wang, "A stochastic sirs epidemic model with nonlinear incidence rate," *Applied Mathematics and Computation*, vol. 305, pp. 221 – 240, 2017.

[8] S. He, Y. Peng, and K. Sun, "SEIR modeling of the COVID-19 and its dynamics," *NONLINEAR DYNAMICS*, pp. 1–4, 2020.

[9] L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, and D. Liu, "Early prediction of the 2019 novel coronavirus outbreak in the mainland china based on simple mathematical model," *IEEE Access*, vol. 8, pp. 51761–51769, 2020.

[10] J. T. Wu, 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*, vol. 395, no. 10225, pp. 689 – 697, 2020.

[11] F. Yang, H. Liu, H. Qi, and X. Liu, "Seir evolutionary simulation model of the infectious disease emergency," in *Proc. ICIICII*, pp. 315–318, 2016.

[12] A. Bhati and A. Jagetiya, "Prediction of covid-19 outbreak in india adopting bhilwara model of containment," in *5th International Conference on Communication and Electronics Systems*, pp. 951–956, 2020.

[13] J. Hellewell et al., "Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts," *The Lancet Global Health*, vol. 8, no. 4, pp. e488 – e496, 2020.

[14] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of COVID-19 is higher compared to SARS coronavirus," *Journal of Travel Medicine*, vol. 27, 2020.

[15] K. Dietz, "The estimation of the basic reproduction number for infectious diseases," *Statistical Methods in Medical Research*, vol. 2, no. 1, pp. 23–41, 1993.

[16] S. A. Lauer et al., "The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application," *Annals of Internal Medicine*, vol. 172, no. 9, pp. 577–582, 2020.

[17] Z. Hu, C. Song, C. Xu, G. Jin, and H. Shen, "Clinical characteristics of 24 asymptomatic infections with covid-19 screened among close contacts in nanjing, china," *SSRN Electronic Journal*, 2020.

[18] S. J. Shah et al., "Clinical features, diagnostics, and outcomes of patients presenting with acute respiratory illness: A retrospective cohort study of patients with and without covid-19," *EClinicalMedicine*, pp. 100518, 2020.

[19] Q. Pham, D. C. Nguyen, T. Huynh-The, W. Hwang, and P. N. Pathirana, "Artificial intelligence (ai) and big data for coronavirus (covid-19) pandemic: A survey on the state-of-the-arts," *IEEE Access*, vol. 8, pp. 130820–130839, 2020.

[20] K. K. W. To et al., "COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing," *Clinical Infectious Diseases*, 2020.