

COVID-19 Spreading Prediction with Enhanced SEIR Model

Yixiao Ma, Zixuan Xu, Ziwei Wu, Yong Bai

School of Information and Communication Engineering, Hainan University, Haikou, 570228, China

e-mail address of Yong Bai: bai@hainu.edu.cn

Abstract—The COVID-19 epidemic broke out at the end of 2019 and developed into a global infectious disease in early 2020. In order to understand the spreading trend of the epidemic, we propose an enhanced epidemiology predictive model—eSEIR model by improving the well-mixed SEIR model on the infectious disease dynamics. The eSEIR model incorporates an optimization method to calculate β and γ parameters. Our proposed model is verified using the epidemic data in Italy and China with reduced RMSE (root mean square error) of the predicted curves, and is used to predict the potential epidemic progress in the United States.

Keywords—COVID-19; SIR; SEIR; epidemiological model

I. INTRODUCTION

Since December 2019, there have been many cases of pneumonia infections caused by a new type of coronavirus COVID-19 around the world. Because the coronavirus is extremely pathogenic and has never been discovered by humans, so far there still lacks specific treatment for the diseases it causes. By August 24, 2020, over 23 million cases have been diagnosed globally, and 808,816 deaths have been confirmed. Except for a few countries whose epidemic has been controlled, most countries are still suffering from the epidemic. How to prevent the spread of the epidemic have become the priorities of governments of various countries.

To understand its spreading, some classic epidemiological models have been employed again. For example, Dhanwant and Ramanathan in [1] focused on the traditional Susceptible-Infected-Recovered (SIR) model in this new coronavirus. In the application of viral pneumonia epidemics, another more mature SEIR model that is commonly used to predict epidemics were used in [2] and [3]. It adds a category of Exposed compartment to the SIR model. The well-mixed SEIR model mentioned in [4] takes into account the various situations of multiple groups of people.

The well-mixed SEIR model is potentially a suitable model for predicting the COVID-19 spread as this new coronavirus pneumonia has a certain degree of infectivity during the incubation period. We continue to enhance the well-mixed SEIR model by using the minimize function in [6] to obtain the optimal value of the loss function by calculating the key parameters β and γ . Using the epidemic data in Italy and China, we verify that the proposed enhanced SEIR (eSEIR) model is more accurate for the prediction of COVID-19 spread with reduced RMSE compared with the SIR and SEIR models. We also apply the eSEIR model to forecast the potential spreading trend of United States.

II. SIR AND SEIR MODELS.

In infectious disease dynamics, Kermack and McKendrick developed the SIR infectious disease model in 1927 using a kinetic approach.

The SIR model divides the total population into the three categories: Susceptible, whose number is given as $s(t)$, which represents the number of people who are not sick at time t but are at risk of being infected by that type of disease; Infective, whose number is given as $i(t)$, which represents the number of people who are sick and infectious at time t ; Recovered and its quantity is given as $r(t)$, which represents the number of people who have been removed from the infected population at time t . The total population is $N(t)$, $N(t) = s(t) + i(t) + r(t)$.

The SIR model is based on the following three assumptions.

- 1) The population dynamics such as births, deaths, and mobility are not taken into account. The population always remains a constant, i.e., $N(t) \equiv N = \text{population}$.
- 2) A patient is contagious once he comes into contact with a susceptible person. Assume that the number of susceptible persons that a patient can infect per unit of time at t is proportional to the total number of susceptible persons $s(t)$ with a proportionality factor of β , and hence the number of people infected by all patients per unit of time at t is $\beta s(t)i(t)$.
- 3) At t , the number of people removed from the diseased people per unit time is proportional to the number of patients with a proportionality factor of γ . The number of people removed per unit time is $\gamma i(t)$.

The progression of susceptible people from illness to removal can be represented by a differential equation as follows,

$$\frac{ds(t)}{dt} = -\beta i(t)s(t) \quad (1)$$

$$\frac{di(t)}{dt} = \beta i(t)s(t) - \gamma i(t) \quad (2)$$

$$\frac{dr(t)}{dt} = \gamma i(t) \quad (3)$$

Let S , I and R denote the number of susceptible, infected and removed populations, respectively. As the epidemic progresses, the susceptible population is gradually infected, and the rate of change is represented by

$$\frac{dS}{dt} = -\beta \frac{IS}{N} \quad (4)$$

where β is the product of the average number of infected people exposed to k per day and the probability of transmission b at exposure, i.e., $\beta = kb$.

The rate of change in the infected population is represented by

$$\frac{dI}{dt} = \beta \frac{IS}{N} - \gamma I \quad (5)$$

where γ is the average rate at which the infected population recovers or dies, taken from the average duration of infection.

The rate of change in the removed population is

$$\frac{dR}{dt} = \gamma I \quad (6)$$

The SEIR model is a more mature and commonly used model for epidemic prediction than the SIR model, where the infectious diseases studied have an incubation period and healthy individuals who have contact with the patients do not become sick immediately, but become carriers of the pathogen. In contrast to the SIR model, the SEIR model takes into account the fact that only a fraction of those who have contact with the patients are infectious, resulting in a longer transmission cycle.

The traditional SEIR model divides population into four categories: Susceptible, Exposed, Infective and Recovered.

As the epidemic progresses, the rates of change in the four populations are represented as

$$\frac{dS}{dt} = -\beta \frac{IS}{N} \quad (7)$$

$$\frac{dE}{dt} = \beta \frac{IS}{N} - \alpha E \quad (8)$$

$$\frac{dI}{dt} = \alpha E - \gamma I \quad (9)$$

$$\frac{dR}{dt} = \gamma R \quad (10)$$

The values of the β and γ parameters are the same as those of the SIR model, $\alpha = 1/T$ for the isolation rate. Because the average incubation period of the coronavirus is 2 to 14 days, the median value of 7 days can be taken as the average incubation period of COVID-19, so $\alpha = 1/7$.

III. ENHANCED SEIR MODEL

In the traditional SEIR model, the probability of infection in susceptible and infected populations is not considered separately, and the number of daily contacts between susceptible and infected populations is considered equal. However, it has been shown that the infection rate of COVID-19 infected individuals is significantly higher than the susceptible individuals. On the other hand, the number of daily contacts of infected individuals

is significantly lower than that of susceptible individuals because of the quarantine policy adopted after the outbreak. COVID-19 has been shown to be infectious during the incubation period, and a modification of the SEIR model is therefore necessary.

Based on the epidemiological characteristics of individuals, the clinical progression of COVID-19, and quarantine interventions, we proposed an enhanced SEIR (eSEIR) model to describe the dynamics of the COVID-19 epidemic process.

In the eSEIR model, the population considered is divided into four groups of individuals: Susceptible (susceptible), Exposed (potential and capable of transmitting COVID-19), Infected (symptomatic and capable of transmitting COVID-19), and Recovered (COVID-19 virus immunization). Exposure rates r_1 and r_2 indicate the likelihood of contact between susceptible and infected individuals and latent individuals, respectively. Infection rates β_1 and β_2 control the rate of transmission. In this case, β_1 indicates the probability of infection per exposure when a susceptible individual (Susceptible) comes into contact with an infected individual (Infected) and becomes latently exposed (Exposed). β_2 indicates the probability of infection per exposure when a susceptible individual (Susceptible) comes into contact with an exposed individual (Exposed) and transmits it to another exposed individual (Exposed). There is controversy argument regarding the transmissibility of latent infections. Here we assume that half of these asymptomatic latent infections are infectious, similar to influenza. Therefore, we take $\beta_2 = 1/2 \times \beta_1$ and the latent rate α to be the rate at which a latent individual become infected (i.e., the average latent period is $1/\alpha$).

The eSEIR model is described as follows,

$$\frac{dS}{dt} = -\frac{r_1\beta_1IS}{N} - \frac{r_2\beta_2ES}{N} \quad (11)$$

$$\frac{dE}{dt} = \frac{r_1\beta_1IS}{N} - \alpha E + \frac{r_2\beta_2ES}{N} \quad (12)$$

$$\frac{dI}{dt} = \alpha E - \gamma I \quad (13)$$

$$\frac{dR}{dt} = \gamma I \quad (14)$$

The parameters β and γ are the same whether the SIR, SEIR or eSEIR models are used. The β and γ parameters of the SIR and SEIR models are usually being fitted according to the infected population, but the accuracy of the model is not guaranteed because the data of only one compartment is used to derive the model. Thus, we employ a minimization function to obtain the optimal value of the loss function. In such a method we incorporate the numbers of three categories of people, i.e., the number of infected, the number of recovered and the number of deaths to calculate the β and γ parameters.

For the eSEIR model, we assume that the exposed population is the static rate of confirmed cases, since the exposed population data is not available. Here, we set α to 0.1 to obtain the model fit parameters. The parameters β and γ are solved by using the minimize function to find the extremes of the returned values of

the loss function. The loss function is used as the objective function to find the minimum value by using the L-BFGS-B quasi-Newton method.

IV. MODEL VERIFICATION AND ANALYSIS

The COVID-19 data for verifying the proposed model are data of the confirmed, recovered and deaths in three countries, Italy [7], China [8] and the United States [9]. The time period of the collected data is same for three countries from the onset of the epidemic on January 22 to August 24, 2020.

Firstly, we employ the minimize method to find the optimal value of the loss function and obtain the β and γ parameters of Italy, China and the United States. The obtained β and γ parameters for the three countries are given in Table 1.

TABLE I. COMPARISON OF β AND γ FOR ITALY, CHINA AND THE UNITED STATES

parameter	Country		
	Italy	China	US
β	0.3110	0.3962	0.1941
γ	0.1502	0.2184	0.0958

The β and γ are two proportional coefficients. The total number of susceptible patients is $\beta s(t)i(t)$, and the number of removed people is $\gamma i(t)$ per unit time. β can be viewed as the contact rate and it can be controlled by improving the medical care level or enforcing the staying-at-home policy. β should be interpreted as β_i in (11) and (12). γ is the cure rate, and it reflects a country's advanced medical level.

Since China has a large population, it can be observed that China's β (contact rate) is much higher than other two countries. Compared with China and Italy, the population density of the United States is lower, and its population distribution is relatively small, hence β of the United States is smaller. As the epidemic occurred earlier, China adopted a series of measures and policies relatively early, hence the disease was controlled earlier, so China's γ (that is, cure rate) is relatively high in the recent time, followed by Italy. The outbreak in the United States and the measures to control the epidemic are relatively late, hence the γ value of the United States is obviously low.

The assumed values of r_1 are 0.3, 1.5, 0.5 and r_2 are 1.5, 3.5, 1.5 for Italy, China and the United States, respectively.

To compare the performance of traditional models (SIR, SEIR) and our proposed eSEIR, we employ the metric of RMSE (root mean square error). The RMSE is used to measure the deviation between the observed value and the real value. It can be calculated by

$$RMSE(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x_i) - y_i)^2} \quad (15)$$

We analyze the errors of the three models based on the number of patients. Since the amount of data information per day is different, the amount of information contained in the latest data is assumably higher. We give different weights on the

collected data to optimize the error analysis of evaluation model. Higher weight is assigned to the data of the latest date as

$$weight = \frac{1}{\min(20, d(i))} \quad (16)$$

where $d(i)$ is the number of days from January 22 to the study date.

A. Model analysis on Italy data

The SEIR and eSEIR model curves of the Italy COVID-19 data are shown in Fig. 1 and Fig. 2, respectively.

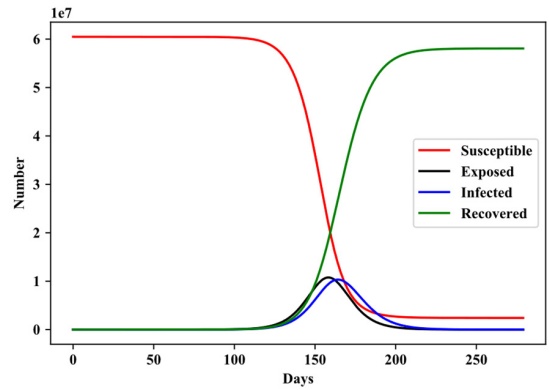


Figure 1. SEIR model curves for Italy data

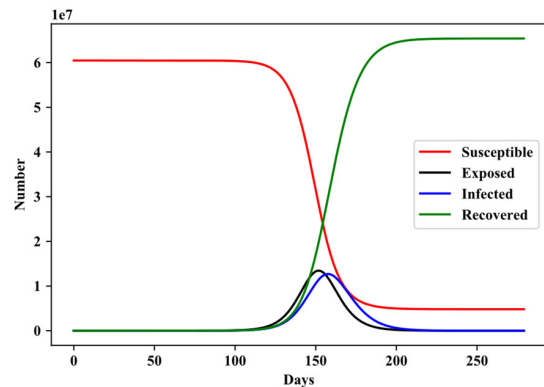


Figure 2. eSEIR model curves for Italy data

The RMSE results of SIR, SEIR, and eSEIR models are 23.81, 9.62, and 8.60, respectively. Thus, eSEIR model results in smallest error and hence is most suitable for the prediction.

B. Model Analysis on China Data

Next, we apply the eSEIR model to study the China data. It can be seen in Fig. 3 that the epidemic occurred relatively early in China. From about 60 to 75 days after the beginning of our research time (January 22, 2020), more serious epidemic began to erupt. The number of people sick every day began to increase violently. In about 100 days, the number of people sick every

day reached the highest. On the other hand, the epidemic ended relatively early. After six or seven months, the epidemic was under control, and the number of people affected every day was small. By comparison, the outbreak in Italy is relatively late, and the number of people reached the maximum daily about 150 days. Its spreading was also controlled in about eight months.

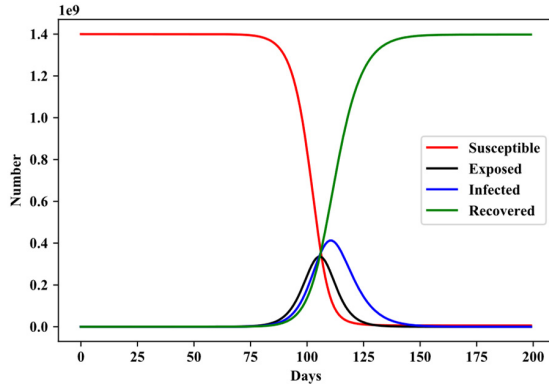


Figure 3. eSEIR model curves for China data

C. Model Analysis on US Data

Lastly, we apply the eSEIR model for forecasting the COVID-19 spreading trend in the United States.

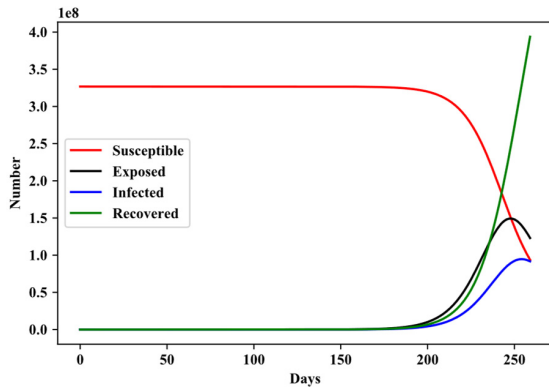


Figure 4. eSEIR model curves for the US data

It can be seen in Fig.4 that the outbreak in the United States is relatively late, large daily change in the number of patients appears after 180 days. Specifically, more serious outbreak happens in August and September. The situation in the United States is supposed to gradually begin to improve after 250 days.

V. CONCLUSIONS

As COVID-19 is a new infectious epidemic, traditional epidemiological models need to be adapted for its spreading prediction. Our proposed eSEIR model respects the properties of COVID-19 epidemic and the control measures. We demonstrated that our proposed eSEIR model is most accurate for the prediction of COVID-19 spreading compared with SIR and SEIR models. Our proposal takes into account different infection rate of latent population and infected population and

the difference of daily contact number between the two groups. The two parameters of β and γ are obtained via optimization method. We verified the proposed model using the COVID-19 data in Italy and China. We also forecasted the future epidemic spreading in the United States. We will observe the progression in the US and may further update our proposed model accordingly.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant (Grant 61561017).

REFERENCES

- [1] Jay Naresh Dhanwant and V. Ramanathan Spain, "Forecasting COVID 19 growth in India using Susceptible-Infected-Recovered (S.I.R) model", 2020, pp.1-6. (references)
- [2] Revanth India. (2020) COVID-19 Forecast : SEIR <https://www.kaggle.com/revanthrex/covid-19-forecast-seir/comments>
- [3] Da Motou China. (2020) Various Mathematical Models of the Transmission of Pneumonia in Wuhan <http://i.dataguru.cn/mportal.php?mod=view&aid=15351>
- [4] Can Hou China. (2020) The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis <https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.25827>
- [5] Patrick Sánchez Spain. (2020) COVID Global Forecast: SIR model + ML regressions <https://www.kaggle.com/saga21/covid-global-forecast-sir-model-ml-regressions>
- [6] Funkyboy Taiwan. (2020) COVID-19 Global Forecast : SEIR + Visualize <https://www.kaggle.com/super13579/covid-19-global-forecast-seir-visualize>
- [7] Lisphilar Japan. (2020) COVID-19 data with SIR model <https://www.kaggle.com/lisphilar/covid-19-data-with-sir-model>
- [8] China. (2020) sa.sogou.com/new-webball/page/sqs/epidemic
- [9] TamaraGoyea US. (2020) CSSEGISandData COVID-19 https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv