Forecast and analysis of COVID-19 epidemic based on improved SEIR model

Borui Gu

School of Electrical and Electronic Engineering, North China Electric Power University, Baoding, China

*Corresponding author: boruigu@ncepu.edu.cn

Abstract. Since the outbreak of novel coronavirus pneumonia (COVID-19), China has taken effective measures to control the spread of the epidemic. By the end of April 2020, the global situation of COVID-19 (COVID-19) prevention and control is still very serious. In this paper, a mathematical model was established to predict the future trend of COVID-19 by studying its transmission routes, propagation speed, spatial range and dynamic mechanism. This article first analyzes the changes in the transmission route in chronological order. After understanding the transmission route of COVID-19, we started to study the transmission speed and used MATLAB to fit the data we found. Secondly, consider the spatial scope of the spread of new coronary pneumonia, and use FineBI software to draw static maps to reflect global medical record data. Finally, we further consider the immigration and emigration of the population, and use Euler's numerical method to solve the established partial differential equation model to study the dynamic mechanism of the new coronavirus pneumonia (COVID-19). Through the above analysis, a modified SEIR transmission dynamics model considering the virus transmission ability of patients in the incubation period and the effect of tracking isolation intervention measures on the epidemic was established. At the end of this article, based on actual data, the revised SEIR model is used to predict the future trend of COVID-19.

1. Introduction
Since December 2019, some medical institutions in Wuhan City, Hubei Province have successively reported patients with unexplained pneumonia, and the city has continued to monitor the development of their disease. On February 11, 2020, the World Health Organization officially named the disease caused by the virus as COVID-19.

Because COVID-19 is highly contagious and it is difficult to detect patients in the incubation period, the number of infected population has increased explosively and the scope of transmission has been expanding. At present, 114 countries, territories or regions in six continents have been affected by the epidemic to varying degrees. The growth trend of the epidemic abroad is not optimistic. Although the epidemic in China is under control, the population movement caused by the active resumption of work and classes makes it face greater challenges.

Therefore, it is necessary for us to establish a mathematical model to predict the development trend of COVID-19 based on its transmission characteristics, so as to take effective measures for prevention and control.

2. Assumption
Based on the actual situation, the following reasonable assumptions are made to simplify the problem:
1. Assuming that all individuals in the population have a probability of being infected, when the infected individual recovers, they will produce antibodies, that is, the recovered population will not be infected again; 
2. Hubei province is a closed node, which can basically ignore the provincial population mobility; 
3. Only consider the communication from person to person. 
4. At present, there are no specific drugs against the COVID-19 virus, so it is assumed that the daily recovery rate (daily cure rate) of patients diagnosed with COVID-19 remains unchanged; 
5. Since the exact relationship between COVID-19 transmission and the external environment, especially temperature changes, has not been determined, the effect of temperature on COVID-19 transmission is not considered in this study.

Table 1. Model symbol description

| NOTESATIONS | Symbols | Descriptions |
|-------------|---------|--------------|
| S(t)        | Number of susceptible people |
| E(t)        | The number of people exposed |
| I(t)        | The number of infections |
| R(t)        | Number of cures and deaths |
| K(t)        | Total number of people |
| R_0         | The Basic Reproductive Number |
| β           | Infection rate |
| c(t)        | Contact rate |
| γ           | Exit rate |
| σ           | Incidence during the incubation period |
| μ           | Average migrant population ratio |

3. Model establishment and solution

3.1. Establishment of infectious disease dynamics model for covid-19

The SEIR model generally divides the population within the epidemic range of infectious diseases into the following four categories:

1. Susceptible people (S), refers to those who have not been infected, but lack of immunity, and are susceptible to infection after contact with infected persons;
2. Exposed people (E) refers to a person who has been in contact with an infected person but is temporarily unable to infect other people. It is applicable to infectious diseases with a long incubation period;
3. Symptomatic infected people (I) refers to a person who has contracted an infectious disease, can spread to S class members, and turn them into class E or class I members;
4. Recovered people (R) refers to a person who is isolated or has immunity due to illness. If the immunization period is limited, R class members can revert to S class.

Taking into account the isolation measures for the prevention and treatment of infectious diseases, the model adds the isolation of susceptible persons (S_q), isolation of exposed persons (E_q) and isolation of infected persons (I_q). Since people who are isolated and infected will be immediately sent to designated hospitals for isolation and treatment, this part of the population is all transformed into hospitalized patients in this model (H). Isolation susceptible persons are transformed into susceptible persons again after being released from quarantine, while the infected and latent persons have different degrees of ability to infect susceptible persons and turn them into latent persons.

The conversion rates of susceptible persons (S) to isolated susceptible persons (S_q), isolation latent persons (E_q), isolated latent persons (E_q), and latent persons (E) are respectively \( ρc_0q(1 − β) \), \( ρc_βq \) and \( ρc_β(1 − q) \). At the same
time, considering the impact of the non-isolated infected persons (I) and latent persons (E) on the susceptible population, the susceptible persons (S) who have been released from isolation are transformed into susceptible persons (S), so the governing equation of the number of susceptible persons is

\[
\frac{dS}{dt} = -\left[\rho c \beta + \rho c q (1 - \beta)\right] S(I + \theta E) + \lambda S_q
\]

(1)

Where \( \theta \) is the ratio of the transmission capacity of the latent person to the infected person. For simplicity, suppose that the infectivity of patients in the incubation period is the same as that of patients who have shown symptoms, that is \( \theta = 1 \). \( \lambda \) is the isolation release rate, taking \( \lambda = \frac{1}{14} \) (isolation duration is 14 days). The differential equation is as follows:

\[
\begin{align*}
\frac{dS}{dt} &= -[\rho c \beta + \rho c q (1 - \beta)] S(I + \theta E) + \lambda S_q \\
\frac{dE}{dt} &= \rho c \beta (1 - q) S(I + \theta E) - \sigma E \\
\frac{dI}{dt} &= \sigma E - (\beta_I + \alpha + \gamma_I) I \\
\frac{dS_q}{dt} &= \rho c q (1 - \beta) S(I + \theta E) - \lambda S_q \\
\frac{dE_q}{dt} &= \rho c \beta q S(I + \theta E) - \delta_q E_q \\
\frac{dH}{dt} &= \beta_I I + \delta_q E_q - (\alpha + \gamma_H) H \\
\frac{dR}{dt} &= \gamma_I I + \gamma_H H
\end{align*}
\]

(2)

Among them, \( \sigma \) is the conversion rate of latent persons to infected persons, taking \( \sigma = \frac{1}{7} \) (the incubation period is 7 days). \( \delta_I \) is the isolation rate of the infected person. \( \gamma_I \) is the recovery rate of the infected. \( \delta_q \) is the rate of conversion from isolated latent persons to isolated infected persons. \( \gamma_H \) is the recovery rate of isolated infected persons.

3.2. The Basic Reproductive Number \( R_0 \)
(1) Model Derivation

The number of people infected during the early free transmission period is negligible compared with the total population. Therefore, when \( t \) approaches 0, \( S(t) \) approaches \( N \). At this time, the basic reproductive number \( R_0 \) can be expressed as

\[
R_0 = \left(1 + \frac{\lambda}{\sigma}\right) \left(1 + \frac{\lambda}{\gamma}\right)
\]

(3)
Where \( \lambda = \frac{\ln Y(t)}{t} \) is the growth rate during the early exponential growth, \( Y(t) \) is the number of symptomatic infections at time \( t \). The incubation period and the infection period can be expressed as \( T_i = \frac{1}{\sigma} \) and \( T_i = \frac{1}{\gamma} \). The generation time can be approximated as sequence interval, which is \( T_g = T_i + T_i \). Taking \( \rho = \frac{T_i}{T_s} \) as the proportion of the incubation period in the generation time, then the basic reproductive number can be expressed as:

\[
R_0 = 1 + \lambda T_i + \rho (1 - \rho) (\lambda T_i)^2
\]

(2) Model Parameters

To estimate \( R_0 \), the key parameters are \( \lambda \), \( \rho \) and \( T_g \). The source and calculation process are given in the following three points:

1. \( \lambda \)
   
   \( \lambda \) can be determined by \( Y(t) \), that is, the approximate value of \( \lambda \) is 0.123 by matlab fitting.

2. \( \rho \)
   
   Existing research results show that the value of \( \rho \) for SARS transmission is between 0.5 and 0.8. Taking the average incubation period as 5.1 days, \( \gamma = 0.1 \), and calculating \( \rho = 0.34 \). In the parameter sensitivity analysis, we consider the case of \( \rho \) from 0.5 to 1. Because \( \rho (1 - \rho) \) has symmetry, this actually covers all possible values of \( \rho \).

3. \( T_g \)
   
   Lipsitch et al.’s analysis of SARS transmission data showed that the average \( T_g \) was 8.4 days. They believe that \( T_g \) may be higher in the early stage of disease outbreak (mean value is 10.0). The \( T_g \) interval for sensitivity analysis is \( T_g \in (8, 12) \). In this paper, \( T_g \) value of incubation period and infection period was calculated as 15.1.

(3) Calculation results of basic regeneration number

After setting all the key parameters, the final value of \( R_0 \) is 3.63. This result is higher than the average, but a new study published in the US journal CDC 《Emerging Infectious Diseases》 believes that the \( R_0 \) value of new coronary pneumonia is as high as 5.7, but this is only one of the scenarios based on the research data in the paper. It is still unclear what the \( R_0 \) value of COVID-19 is, so it can be considered that the calculation from the model given in this article is reasonable.

3.3. Considering population migration

Considering that work will resume after the epidemic has been relatively eased, we have added a population migration factor based on the province as a unit to predict the development trend of the epidemic.

(1) Model establishment

During the resumption of work, the susceptible (S) and the latent (E) can undergo population migration. We consider the inflow and outflow of these people in a certain province and get the following process:
(2) SEIR model between cities

\[
\begin{align*}
\partial_t S_n &= -\frac{\beta c I_n S_n}{K_n} + \sum_{m \neq n} (w_{mn} S_m - w_{mn} S_n) \\
\partial_t E_n &= \frac{\beta c I_n S_n}{K_n} - \sigma E_n + \sum_{m \neq n} (w_{mn} E_m - w_{mn} E_n) \\
\partial_t I_n &= \sigma E_n - \gamma I_n + \sum_{m \neq n} (w_{mn} I_m - w_{mn} I_n) \\
\partial_t R_n &= \gamma I_n + \sum_{m \neq n} (w_{mn} R_m - w_{mn} R_n)
\end{align*}
\]

3.4. Spatial scope analysis

By using FineBI software, the daily global cumulative diagnosis data collected can be imported, and the static map of corresponding date can be obtained. By selecting different dates, the specific data of the number of people diagnosed in different countries around the world can be found. Figure 2 and 3 represent cumulative global diagnosis on 20 and 28 April respectively.

Figure 1. The flow chart of the model

Figure 2. Global epidemic distribution map on April 20
3.5. Calculate the remaining parameters of SEIR model

The model is solved by Euler's numerical method, and the time step is 0.01. I in this model is the current number of infected people, defined as the confirmed cases announced that day minus cured cases and deaths.

The above-mentioned SEIR model parameter assignment is shown in the table. The probability of exposure to infection has been adjusted, and the parameters q and α are fitted and optimized based on more current raw data to improve the accuracy of model prediction.

Table 2. The initial value of the SEIR model

| Variable | Variable Description         | Value    | Value description                              |
|----------|------------------------------|----------|-----------------------------------------------|
| S        | Susceptible people           | 59170000 | Population of Hubei Province                  |
| E        | Exposed people               | 4007     | Number of people in incubation period on January 23, 2020 |
| I        | Infected people              | 786      | Number of infected people on January 23, 2020  |
| Sq       | Quarantined susceptible      | 2776     | Number of people still under medical observation |
| Eq       | Quarantined exposed          | 400      | Estimated value, less than Sq                  |
| H        | Hospital patient             | I+Eq     | Isolated infected and potential patients       |
| R        | Recovered people             | 31       | Number of recovered people on January 23, 2020 |
Table 3. Parameter assignment of SEIR model

| Parameter | Parameter Description                                      | Value  |
|-----------|------------------------------------------------------------|--------|
| c         | Contact rate                                              | 2      |
| $\delta_I$| Isolation rate of infected persons                        | 0.13   |
| $\delta_q$| Conversion rate from quarantined latent persons to quarantined infected persons | 0.13   |
| $\gamma_I$| Recovery rate of infected persons                         | 0.007  |
| $\rho$    | Effective contact coefficient                              | 1      |
| $\theta$  | The ratio of the infectious ability of the latent person to the infected person | 1      |
| $\lambda$ | Reciprocal of isolated contact speed                       | 1/14   |

Use matlab to find the parameter values that best match the actual epidemic data, and the results are shown in Table 4.

Table 4. Parameter fitting results of SEIR model

| Parameter | Parameter Description                  | Value     |
|-----------|----------------------------------------|-----------|
| $\gamma_H$| Recovery rate of quarantined infected persons | 0.014     |
| $\beta$   | Probability of infection               | $2.24 \times 10^{-9}$ |
| $q$       | Isolation ratio                        | $1.1 \times 10^{-6}$ |
| $\alpha$  | Case fatality rate                     | $2.7 \times 10^{-4}$ |

3.6. Forecast of epidemic situation in Hubei Province

![Figure 4. Forecast of epidemic situation in Hubei Province](image)

Based on actual data, use Matlab to find the parameter values that best match the actual epidemic data, and establish a revised SEIR model. We forecast the development trend of the epidemic in Hubei
Province from February 22 to March 23. (The prediction result is shown in the figure). The prediction results show that: The number of COVID-19 infections in Hubei Province reached its peak around February 26. After that, the number of infections was controlled to a certain extent, but still maintained a high level, which is basically consistent with the actual situation of the epidemic.

4. Conclusion
The predicted results are in good agreement with the actual data, which proves that the revised SEIR model is reliable in evaluating the epidemic trend.

5. References
[1] Jianjun Li, Shan He. Population flow, information dissemination efficiency and epidemic prevention and control – Based on the new type of coronary pneumonia(covid-19) [J]. Journal of Central University of Finance and economics, 2020 (04): 116-128
[2] Liren Yang, Zixuan Yang, Li Yang, Guilan Xie, Ruiqi Wang, Shanshan Wang, Wenfang Yang. The factors affecting the transmission and development of new coronavirus pneumonia in the provinces: Based on data analysis of 30 provinces and cities, [J/OL]. Journal of Xi'an Jiao Tong University (Medical Science): 1-13[2020-05-10].
[3] Shengli Cao, Peihua Feng, Pengpeng Shi. Application of modified SEIR infectious disease dynamics model to the prediction and evaluation of COVID-19 epidemic situation in Hubei Province [J/OL]. Journal of Zhejiang University (Medical Edition): 1-13[2020-05-10].
[4] TANG B, WANG X, LI Q, et al. Estimation of the transmission risk of 2019-nCov and its implication for public health interventions[J]. J Clin Med, 2020, 9(2):462. DOI: 10.3390/jcm9020462.
[5] Lauer SA, Grantz KH, Bi Q, etal. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020; 172:577–582. [Epub ahead of print 10 March 2020]. doi: https://doi.org/10.7326/M20-0504