



OPEN

## Application of genetic algorithm combined with improved SEIR model in predicting the epidemic trend of COVID-19, China

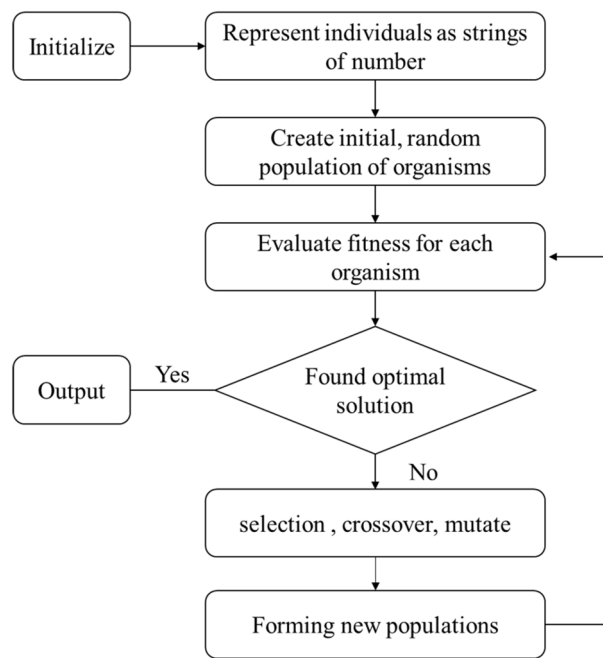
Zhenzhen Qiu, Youyi Sun, Xuan He, Jing Wei, Rui Zhou✉, Jie Bai & Shouying Du

Since the outbreak of the 2019 Coronavirus disease (COVID-19) at the end of 2019, it has caused great adverse effects on the whole world, and it has been hindering the global economy. It is urgent to establish an infectious disease model for the current COVID-19 epidemic to predict the trend of the epidemic. Based on the SEIR model, the improved SEIR models were established with considering the incubation period, the isolated population, and genetic algorithm (GA) parameter optimization method. The improved SEIR models can predict the trend of the epidemic situation better and obtain the more accurate epidemic-related parameters. Comparing some key parameters, it is capable to evaluate the impact of different epidemic prevention measures and the implementation of different epidemic prevention levels on the COVID-19, which has significant guidance for further epidemic prevention measures.

Since the outbreak of the 2019 Coronavirus disease (COVID-19), it has been threatening people's lives, livelihoods, economic and social development, etc.<sup>1</sup>. By the end of February 2021, COVID-19 had spread to more than 230 countries, with more than 100 million confirmed cases and 2.5 million deaths<sup>2</sup>. China was the first country to be affected. In the early stage of the COVID-19 outbreak, the Chinese government quickly adopted strict epidemic prevention measures, including lockdown of cities<sup>3</sup>, the establishment of Fangcang shelter hospitals<sup>4</sup>, self-isolation<sup>5</sup>, compulsory isolation<sup>6</sup> and real-time epidemic tracking<sup>7</sup>, which promptly and effectively contained the large-scale trend of the epidemic in China in a timely manner. However, there are still occasional local small-scale outbreaks of COVID-19. COVID-19 is mainly transmitted by respiratory droplets and contact, but so far there is no cure<sup>8</sup>. Therefore, it is of great significance for future epidemic prevention and control and resource allocation to fit the epidemic situation in many places, to obtain the accurate epidemic model, and to predict the trend of COVID-19.

At the beginning of the twentieth century, Kermack and McKendrick<sup>9</sup> established the SIR (Susceptible Infected Recovered) model using the nonlinear dynamics method, and began to have the establishment of the dynamic model of infectious diseases and the analysis and research of its dynamics. Then according to the transmission characteristics of different infectious diseases, people developed the SIRS model that considers the secondary disease. Further, the SEIR model that considers the incubation period of the disease was improved. These models opened a new chapter for the study of infectious disease dynamics, and greatly promoted the research progress of infectious diseases<sup>10,11</sup>. In addition, applying other algorithms to the prediction and modeling of infectious disease dynamics can optimize parameters and fit the real infectious disease propagation process better<sup>12,13</sup>. For example, Godio et al.<sup>14</sup> used a particle swarm optimization (PSO) solver to apply stochastic methods to fit epidemic data to improve the accuracy of SEIR model prediction. With the occurrence of the Novel Coronavirus Pneumonia epidemic, Rezapour et al.<sup>15</sup> used the Caputo fractional derivative combined with the SEIR mathematical model to predict the spread of COVID-19 in Iran and the world; Marinca et al.<sup>16</sup> used the optimal auxiliary function method (OAFM) to obtain an approximate solution of the COVID-19 SEIR model, and developed a SEIR model suitable for the number of deaths from COVID-19. Although these models can predict the popular trend of COVID-19 to some extent, there are still some shortages. For example, most scholars do not pre-process the existing data before prediction and analysis, and the prediction accuracy of the models needs to be improved. Compared with other infectious diseases, although COVID-19 has a lower mortality rate, it is more infectious and has a certain infectivity during the incubation period. Besides, objective factors such

School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 102488, China. ✉email: r.zhou-bucm@qq.com



**Figure 1.** Flow chart of Genetic Algorithm.

as isolation measures have a great impact on it, which leads to the inaccuracy of the existing infectious disease prediction model to obtain the true epidemic parameters<sup>17–20</sup>.

Genetic Algorithm (GA) was first proposed by Professor John Holland<sup>21</sup>. It is a method to find the optimal solution or approximate optimal solution to a complex problem by simulating the natural evolution process. It has been used in neural networks<sup>22</sup>, combinatorial optimization<sup>23</sup>, artificial intelligence<sup>24,25</sup>, genetic programming<sup>26</sup>, data mining<sup>27</sup> and other fields. The optimization principle is shown in Fig. 1. When usually using GA to optimize parameters, we need to set each parameter that needs to be estimated as an individual and set the possible value range of each parameter in the application. GA can encode all parameters to form the initial population, and define the fitness function at the same time. Generally, the root mean square error (RMSE) between the test results and the real results after the parameters are substituted is the standard to measure the quality of the individual. The smaller the error, the better the performance of the population. The optimal solution can be obtained. Finally, the optimal parameter population is assigned to the target model to obtain the final results.

Compared with the traditional optimization algorithm, GA utilizes probabilistic rules instead of certain rule. Therefore, GA has the characteristics of global optimization, simple operation, and it is suitable for solving complex optimization problems. In this research, we used GA to analyze the influence of social and government behaviors on disease dynamics to optimize the model parameters, and took into account the changes of objective factors such as the gradual improvement of isolation measures in the early and late stages of the epidemic. Using simulation software Matlab, two improved SEIR epidemic prediction models were constructed, including the SEIR model. One is considering incubation period infectivity, and the other is considering both incubation period infectivity and isolation measures. The trend of epidemic in different periods was simulated and the prediction results were obtained. The results provide reference for the parameters of epidemic prediction and the improvement and optimization of epidemic prevention measures in the future.

## Methods

**Traditional SEIR infectious disease prediction model.** Because COVID-19 has an incubation period, and its average incubation period is 7 days, we chose to consider the SEIR infectious disease model of the latent person as the basis for modeling. In the traditional SEIR model, the population is divided into the following four categories<sup>28</sup>:

Susceptible (S), healthy people who may be infected.

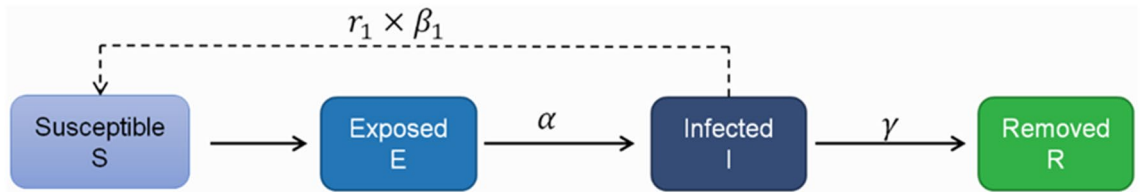
Exposed (E), people who have been infected and have not shown pathological features.

Infected (I), people who have been infected and show pathological features.

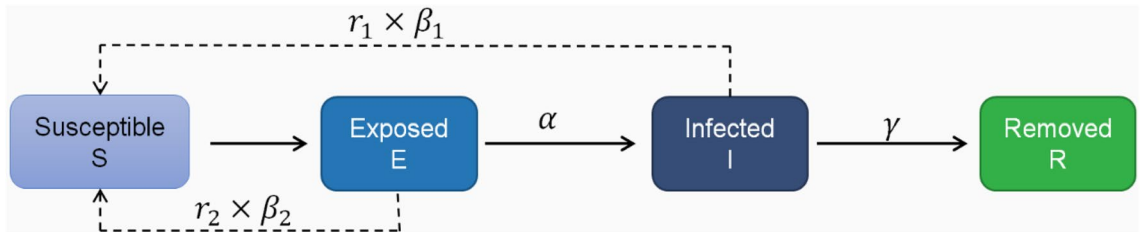
Removed (R), people who have died or cured that are no longer contagious and will not be infected.

At the same time, the following assumptions are often made<sup>29,30</sup>:

1. It is assumed that the total population in a certain area is constant, and the natural birth rate and natural death rate are not considered, and the movement of people between regions is not considered.
2. Recovered persons can develop antibodies and will not be infected again recently.
3. The exposed (E) is not contagious.



**Figure 2.** Schematic diagram of the traditional SEIR model.



**Figure 3.** Schematic diagram of SEIR model considering latent infectivity.

Therefore, based on the traditional SEIR model, the mutual conversion relationship among the four populations is shown in Fig. 2. The susceptible have a certain probability of being infected into the exposed after contact with the infected, and exposed persons will be transformed into infected persons after a period of incubation period, and the infected persons will be cured or die and become removers.  $r_1$  is the number of effective contacts of infected persons, and  $\beta_1$  is the probability of infection of infected persons each time they come into contact with susceptible persons.  $\alpha$  is the conversion rate of exposed to infected persons, which is the reciprocal of the treatment cycle. At the same time, set the total population in the area to  $N$ , and  $S + E + I + R = N$ . Establishing differential equations for the above relationship can be obtained:

$$\begin{cases} \frac{dS}{dt} = -r_1\beta_1 I \frac{S}{N} \\ \frac{dE}{dt} = r_1\beta_1 I \frac{S}{N} - \alpha E \\ \frac{dI}{dt} = \alpha E - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (1)$$

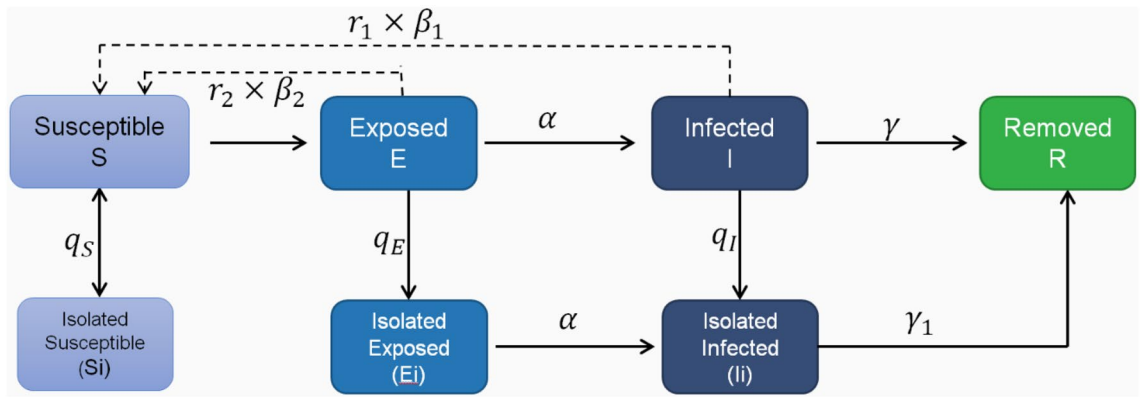
By (1) is modified to iterative forms available:

$$\begin{cases} S_{t+1} = S_t - r_1\beta_1 I_t \frac{S_t}{N} \\ E_{t+1} = E_t + r_1\beta_1 I_t \frac{S_t}{N} - \alpha E_t \\ I_{t+1} = I_t + \alpha E_t - \gamma I_t \\ R_{t+1} = R_t + \gamma I_t \end{cases} \quad (2)$$

**Improved SEIR model A: infectious disease prediction model considering the infectivity of incubation period.** According to the “COVID-19 Diagnosis and Treatment Program (Trial Eighth Edition)” issued by the National Health Commission on August 18, 2020, the source of infection is not only the patients infected by novel coronavirus, but also the asymptomatic infection. Namely, COVID-19 is infectious in the incubation period. Therefore, the traditional SEIR model needs to be changed as follows:

As shown in Fig. 3, based on the traditional SEIR model, the infectivity of the latent person to the susceptible person is increased, and the susceptible person may be infected and become a new latent person after contact with the exposed or the infected. Among them,  $r_2$  is the number of effective contacts of the latent, and  $\beta_2$  is the infection probability of each contact of the latent with a susceptible person, which can be concluded that the new differential equation:

$$\begin{cases} \frac{dS}{dt} = -r_1\beta_1 I \frac{S}{N} - r_2\beta_2 E \frac{S}{N} \\ \frac{dE}{dt} = r_1\beta_1 I \frac{S}{N} + r_2\beta_2 E \frac{S}{N} - \alpha E \\ \frac{dI}{dt} = \alpha E - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (3)$$



**Figure 4.** Schematic diagram of SEIR model considering incubation infectivity and isolation measures.

By (3) is modified to iterative forms available:

$$\begin{cases} S_{t+1} = S_t - r_1\beta_1 I_t \frac{S_t}{N} - r_2\beta_2 E_t \frac{S_t}{N} \\ E_{t+1} = E_t + r_1\beta_1 I_t \frac{S_t}{N} + r_2\beta_2 E_t \frac{S_t}{N} - \alpha E_t \\ I_{t+1} = I_t + \alpha E_t - \gamma I_t \\ R_{t+1} = R_t + \gamma I_t \end{cases} \quad (4)$$

**Improved SEIR model B: predicting infectious diseases considering the infectivity of incubation period and isolation measures.** In the early stage of the epidemic, due to the active actions of the government, many isolation measures were taken, such as the isolation treatment of infected people in Fangcang shelter hospitals, the centralized observation and home isolation. Isolation treatments are depending on the close contacts of COVID-19. The policy provides for different levels of isolation of the susceptible, the exposed and the infected, which plays a key role in the trend of the epidemic situation. Based on the existing epidemic prevention and control measures in China, the SEIR model B was further improved in order to simulate the trend of the epidemic accurately:

As shown in Fig. 4, on the basis of the previous model, measures to isolate various groups of people are considered. First infected after diagnosis will be quarantined as isolators  $I_g$  no longer contagious infection<sup>31</sup>, and after exposed (E), infected (I) contact with susceptible people (S), all in close contact with infected people (I) became the exposed (E), while the uninfected close contacts remain in the susceptible crowd (S). The exposed (E) will be quarantined after the nucleic acid test result is positive, which is called exposed isolator  $E_g$ . The relevant uninfected contacts will be isolated as susceptible isolators  $S_g$ . Among them,  $S_g$  will not be infected during the isolation period, and will return to susceptible crowd (S) after the end of the isolation period (set the isolation period as  $\mu$ ). Susceptible crowd (S) have the possibility of being infected again. After the incubation period is over, if the exposed (E) who have been the isolated will be confirmed cases of COVID-19 and become infected isolator  $I_g$ ,  $E_g$  and  $I_g$  are no longer infectious because they are isolated. In the end, the infected and infected isolators will be cured or die as the remover. Where,  $q_s$  is the proportion of close contacts isolated among the susceptible,  $q_E$  is the probability of the exposed being isolated, and  $q_I$  is the probability of the infected being isolated. Based on the above population relations, the original differential equation can be extended as follows:

$$\begin{cases} \frac{dS}{dt} = -\frac{S[r_1I(\beta_1+q_s-q_s\beta_1)+r_2E(\beta_2+q_s-q_s\beta_2)]}{N-S_g-E_g-I_g} \\ \frac{dS_g}{dt} = \frac{Sg_s[r_1I(1-\beta_1)+r_2E(1-\beta_2)]}{N-S_g-E_g-I_g} \\ \frac{dE}{dt} = \frac{S(r_1\beta_1I+r_2\beta_2E)}{N-S_g-E_g-I_g} - \alpha E(1-q_E) - Eq_E \\ \frac{dE_g}{dt} = Eq_E - \alpha E_g \\ \frac{dI}{dt} = \alpha E(1-q_E) - Iq_I - I\gamma(1-q_I) \\ \frac{dI_g}{dt} = E_g\alpha + Iq_I - I_g\gamma_1 \\ \frac{dR}{dt} = I\gamma(1-q_I) + I_g\gamma_1 \end{cases} \quad (5)$$

| Parameters | $r_1$  | $r_2$  | $\beta_1$ | $\beta_2$ | $\gamma$ | $\gamma_1$ | $q_S$  | E   |
|------------|--------|--------|-----------|-----------|----------|------------|--------|-----|
| Assignment | 4.6701 | 5.1396 | 0.4896    | 0.0244    | 0.7401   | 0.0631     | 0.0046 | 606 |

**Table 1.** Estimated results of epidemic parameters in Wuhan.

For the convenience of calculation, let  $\frac{dA}{dt} = \frac{S}{N-S_g-E_g-I_g}$ , in (5) and modified to iterative forms available:

$$\begin{cases} A_t = \frac{S_t}{N-S_{gt}-E_{gt}-I_{gt}} \\ S_{t+1} = S_t - A_t r_1 I_t (\beta_1 + q_S - q_S \beta_1) - A_t r_2 E_t (\beta_2 + q_S - q_S \beta_2) + S_{g(t-\mu)} \\ S_{g(t+1)} = S_{gt} + A_t q_S [r_1 I_t (1 - \beta_1) + r_2 E_t (1 - \beta_2)] - S_{g(t-\mu)} \\ E_{t+1} = E_t + A_t (r_1 \beta_1 I_t + r_2 \beta_2 E_t) - \alpha E_t (1 - q_E) - E_t q_E \\ E_{g(t+1)} = E_{gt} + E_t q_E - \alpha E_{gt} \\ I_{t+1} = I_t + E_t \alpha (1 - q_E) - I_t q_I - I_t \gamma (1 - q_I) \\ I_{g(t+1)} = I_{gt} + E_{gt} \alpha + I_t q_I - I_{gt} \gamma_1 \\ R_{t+1} = R_t + I_t \gamma (1 - q_I) + I_{gt} \gamma_1 \end{cases} \quad (6)$$

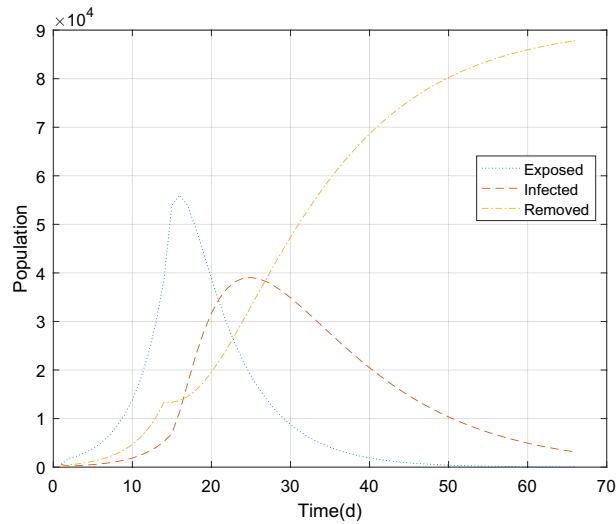
**Parameter estimation and model fitting.** The results of improved SEIR models, A and B, are greatly affected by the initial parameters. In order to establish the SEIR infectious disease model, appropriate values of the key parameters should be selected, which are  $E, q_S, q_E, q_I, r_1, \beta_1, r_2, \beta_2, \alpha, \gamma, \gamma_1, \mu$ . Among them, the incidence probability ( $\alpha$ ) of the exposed is taken as the inverse of the incubation period, and the incubation period is taken as 7 days in line with most reports<sup>32,33</sup>, namely  $\alpha = \frac{1}{7} \approx 0.1429$ . According to the isolation policy of Wuhan and Beijing, the isolation period is 14 days, namely  $\mu = 14$ .  $q_E$  is the isolation rate of the exposed and  $q_I$  is the isolation probability of the infected. According to the current epidemic prevention and control strategy, all confirmed patients will be isolated, so  $q_E$  is the accuracy rate of confirmed patients and  $q_I = 1$ . According to a report on April 18, 2020, the accuracy of nucleic acid detection is about 50% to 70%, and with the epidemic under control, the number of existing cases is decreasing, and nucleic acid kits are sufficient. The accuracy rate should be improved compared with the initial stage of the epidemic, so its maximum value is set, namely  $q_E = 0.7$ . Other parameters are estimated independently according to different conditions in different regions.

## Result

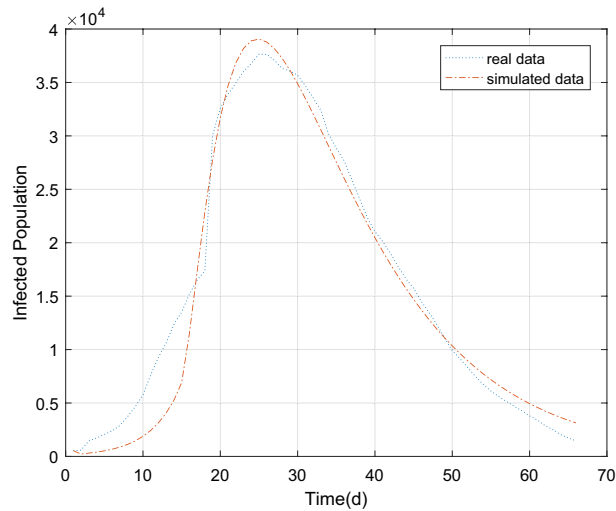
**Predicting the trend of COVID-19 in Wuhan with the improved SEIR models.** According to the above mentioned improved SEIR models and GA, the epidemic situation in Wuhan was simulated. Wuhan epidemic is the earliest COVID-19 epidemic in China. Due to insufficient nucleic acid detection kits in the early stage and insufficient cognition of COVID-19, there may be some deviation between the earlier official reported data and the actual data<sup>34,35</sup>. Wuhan has been closed since January 23, 2020. In order to ensure reliable data, we selected the epidemic related data of January 25, 2020 solstice and March 20 in Wuhan for collation and simulation. According to the dynamics of the epidemic situation in Wuhan, piecewise function can be used for modeling<sup>36</sup>. According to the first report of Wuhan Health Commission on February 8, 2020, the number of people receiving isolation treatment was divided into sections. During the period of January 25th and February 7th, Fangcang shelter hospitals, Huo shenshan hospital and Lei shenshan hospital under construction were not fully put into use, and the medical resources were relatively limited and could not provide perfect isolation measures. Therefore, the model in Eq. (4), which only considered the infectivity in the incubation period but did not consider the isolation measures, was used for data simulating. During the period of 8 February and 10 March, the isolation measures were guaranteed as the isolation hospitals were put into operation. Therefore, the improved SEIR model in Eq. (6) considering the infectivity of incubation period and isolation measures was used for data simulation. The data of epidemic in Wuhan was taken as the data collection area. Due to the strict city closure measures adopted by Wuhan since January 23, inter-regional personnel flow was not considered. The total population in Wuhan is set as  $N = 11,081,000$  according to Wuhan Statistical Yearbook 2019 released by Wuhan Statistical Bureau<sup>37</sup>. The remaining unknown parameters were optimized by GA, and the RMSE of I and R were used as the fitness function values to estimate the parameters. Table 1 shows the estimated results of various parameters of the epidemic in Wuhan.

The parameter estimation results were substituted into the piecewise differential equation to obtain the fitting results of the epidemic in Wuhan, as shown in Figs. 5 and 6 below. There was a good correlation between the model simulating results and the real value ( $r = 0.9849, P < 0.05$ ). It is proved that the combination of GA and improved SEIR models can fit the existing epidemic data better and restore the unknown parameters.

**Predicting the trend of COVID-19 in Beijing with the improved SEIR model.** In the same method, we simulating the epidemic outbreak of COVID-19 in Beijing Xinfadi. Unlike the initial outbreak in Wuhan, a complete epidemic prevention and control system had been established in Beijing at the first time of the outbreak in Xinfadi. Therefore, the improved SEIR model B for predicting the trend of COVID-19 was used for simulation. The premise is to take Beijing as a fixed region, without considering inter-regional mobility. According to the 2019 Beijing Statistical Yearbook released by the Beijing Municipal Bureau of Statistics, the total population



**Figure 5.** The population of exposed, infected and removed in Wuhan over time.



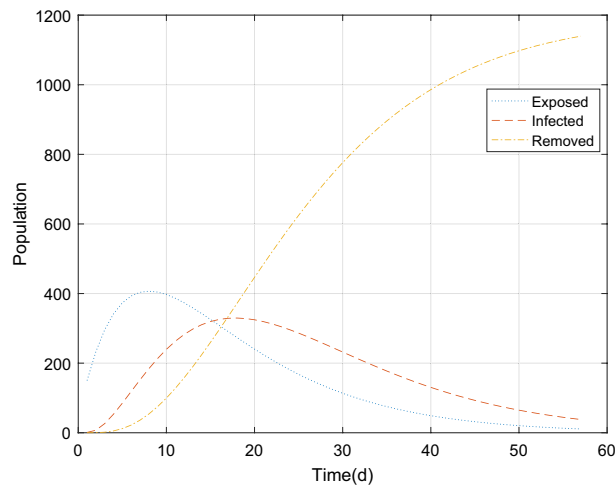
**Figure 6.** Comparison between simulated data and real data of the existing infected in Wuhan.

| Parameters | $r_1$ | $r_2$ | $\beta_1$ | $\beta_2$ | $\gamma$ | $\gamma_1$ | $q_s$ | E   |
|------------|-------|-------|-----------|-----------|----------|------------|-------|-----|
| Assignment | 1.373 | 1.676 | 0.959     | 0.350     | 0.901    | 0.114      | 0.007 | 150 |

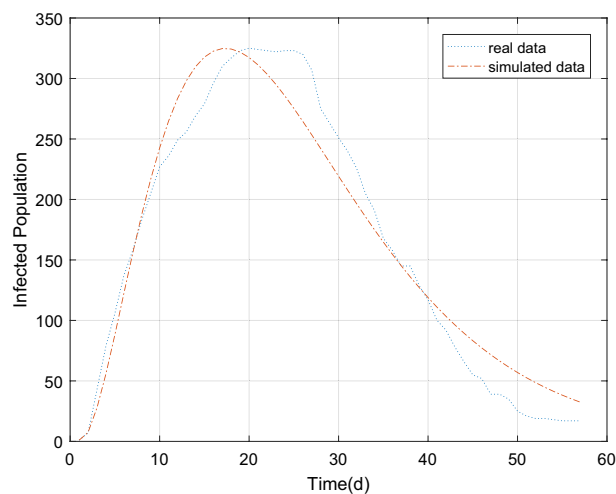
**Table 2.** Estimated results of epidemic parameters in Beijing Xinfadi.

of Beijing is set as  $N=21,536,000$ <sup>38</sup>. The remaining unknown parameters were optimized by GA, and the RMSE of I and R was used as the fitness function values to estimate the parameters. Table 2 shows the estimated results of the epidemic parameters in Beijing Xinfadi.

The estimated parameters in Table 2 were substituted into the differential equation in Eq. (6) to obtain the simulating results of the epidemic situation in Beijing Xinfadi, as shown in Figs. 7 and 8 below. Based on the comparison of the estimated parameters between Wuhan and Beijing, it can be seen that with the trend of the epidemic situation, the parameters will vary greatly. Comparison with  $r_1$  and  $r_2$ , it can be seen that differences on the initial stage of the epidemic, individual protection has been improved, and the effective contact rate of the exposed (E) and the infected (I) is still reduced even after the resumption of work and production. At the same time, the comparison of  $\gamma_1$  and  $\gamma$  shows that because the number of COVID-19 cases in Beijing Xinfadi is far less than that in Wuhan and the medical resources are abundant, the cure rate of COVID-19 in Beijing Xinfadi is greatly improved.



**Figure 7.** The population of exposed, infected and removed in Beijing Xinfadi over time.



**Figure 8.** Comparison between simulated data and real data of the existing infected in Beijing Xinfadi.

## Discussion

On the one hand, the parameters optimized by GA can intuitively express the effectiveness of various epidemic prevention measures in different regions at different times, which provide reference for epidemic prediction in the future. On the other hand, compared with the traditional SEIR model, the improved SEIR model considered the latent infectivity and human intervention measures to establish differential equations including the new population and the original four groups of people, so as to show the influence of the latent infectivity and human intervention on the trend of COVID-19 in the process of epidemic transmission. Therefore, the use of the improved SEIR model can fit the trend of COVID-19 epidemic better. Meanwhile, epidemic prevention measures have great impact on epidemic parameters. Then the improved SEIR model can quickly analyze the effectiveness of existing epidemic prevention measures.

According to the estimated epidemic parameters in Wuhan and Beijing Xinfadi, the severity of the epidemic and the prevention and control degree in different places at different times can be compared. The values of  $r_1$  and  $r_2$  are influenced by individual protection, such as wearing masks and disinfection frequency, as well as policy regulation, such as entertainment places where people gather are not open to the public, home isolation, and crowd control in indoor places<sup>39,40</sup>. From the results of the numerical comparison, it can be seen that the  $r_1$  and  $r_2$  values of the Wuhan epidemic are much larger than the values of the Beijing Xinfadi epidemic. This may be due to the shortage of masks and protective clothing and other materials in the early stage of the Wuhan epidemic, and the lack of awareness of the people to wear masks correctly, which caused the people to be unable to achieve effective individual protection. With the improvement of prevention and control policies and the popularization of epidemic prevention knowledge, it can be seen that in the half a year (approx. 2020.7–2020.12),  $r_1$  and  $r_2$  decreased significantly in Beijing Xinfadi.

The simulation results of existing cases in Wuhan and Beijing Xinfadi showed a good correlation with the true values (Wuhan:  $r = 0.9849$ ,  $P < 0.05$ ; Beijing Xinfadi:  $r = 0.9764$ ,  $P < 0.05$ ). As can be seen from the Figs. 6 and 8, the simulation results of COVID-19 in Wuhan were even better. The simulation results at the initial stage of the epidemic are different from the true value of the trend of the epidemic, which may be due to the lack of nucleic acid testing capabilities in Wuhan at the initial stage, and the number of confirmed and tested daily is related to the amount of testing. It is also possible that the simulation results of the peak of the epidemic case are higher than the true value because the initial detection ability is insufficient. What's more, some infected (I) are not diagnosed, and this error has to be accepted. Compared with the simulating results in Wuhan, the simulating results in Beijing are not very ideal. The reason is that a small cluster of cases was generated in the early stage of the new outbreak, and the effectiveness of the epidemic prevention and control policy cannot be reflected in the simulation. After the first batch of clustered cases were isolated and treated, the epidemic was effectively controlled. The strict isolation measures are effective, resulting in deviation of simulation results from the real epidemic data. In order to improve the accuracy of the simulating, piecewise simulating can be performed with the time points of occurrence and elimination. In fact, the piecewise function function that was used to predict the trend of COVID-19 in Wuhan performed greatly.

## Conclusion

In this research, GA as a common optimization algorithm is appropriate for estimating unknown parameters in the improved SEIR models A and B, and the simulated parameters can also be used to predict unknown epidemic trends. Compared with the short-term prediction of the number of infected cases of the ANFIS model, GA combined with the improved SEIR models in this study were able to predict the entire epidemic period better. It can be seen from the comparison diagrams that the overall trend from the initial stage, peak, turning point and zero clearance of the COVID-19 epidemic is accurate ( $r = 0.9849$ ). In addition, compared with Caputo derivative of fractional order, using GA to estimate the optimal values is very simple. This method in our research is also considering the following impact factors, such as different prevention and control strategies to evaluate the impact of different prevention measures and levels on COVID-19. Based on our research, the improved SEIR models can accurately predict epidemic trend in epidemic regions and countries. Because the real data is affected by many aspects, piecewise function performed greatly and can be used for the further optimization. These results show that GA combined with the improved SEIR models are highly adaptable and suitable for the prediction of the trend of COVID-19. The estimation of propagation parameters plays an important role in inflection points and peak prediction, epidemic trend calculations, and simulated transmission dynamics. This research will provide a certain reference for the advance deployment of medical resources during the epidemic and the arrangements for resuming work and production after the epidemic (Supplementary Information).

## Data availability

Corresponding databases were established based on the epidemic data reported by the official website of Wuhan Municipal Health Commission (<http://wjw.wuhan.gov.cn/>) from June 11 to August 6, 2020 and the official website of Beijing Municipal Health Commission (<http://wjw.beijing.gov.cn/>) from June 11 to August 6, 2020.

Received: 9 October 2021; Accepted: 17 May 2022

Published online: 26 May 2022

## References

- Di Giamberardino, P., Iacoviello, D., Papa, F. & Sinisgalli, C. A data-driven model of the COVID-19 spread among interconnected populations: epidemiological and mobility aspects following the lockdown in Italy. *Nonlinear Dyn.* **106**, 1239–1266. <https://doi.org/10.1007/s11071-021-06840-2> (2021).
- Zhang, N. *et al.* Analysis of efficacy of intervention strategies for COVID-19 transmission: A case study of Hong Kong. *Environ. Int.* **156**, 106723. <https://doi.org/10.1016/j.envint.2021.106723> (2021).
- Ye, T. T. *et al.* Health and related economic benefits associated with reduction in air pollution during COVID-19 outbreak in 367 cities in China. *Ecotox. Environ. Safe.* **222**, 112481. <https://doi.org/10.1016/j.ecoenv.2021.112481> (2021).
- Jiang, H. *et al.* Quantitative assessment of the effectiveness of joint measures led by Fangcang shelter hospitals in response to COVID-19 epidemic in Wuhan China. *BMC Infect. Dis.* **21**, 626. <https://doi.org/10.1186/s12879-021-06165-w> (2021).
- Yuan, Z. M. *et al.* Modelling the effects of Wuhan's lockdown during COVID-19 China. *Bull. World Health Organ.* **98**, 484–494 (2020).
- Jang, H., Polgreen, P. M., Segre, A. M. & Pemmaraju, S. V. COVID-19 modeling and non-pharmaceutical interventions in an outpatient dialysis unit. *PLoS Comput. Biol.* **17**, e1009177. <https://doi.org/10.1371/journal.pcbi.1009177> (2021).
- Stange, M. *et al.* SARS-CoV-2 outbreak in a tri-national urban area is dominated by a B.1 lineage variant linked to a mass gathering event. *PLoS Pathog.* **17**, e1009374. <https://doi.org/10.1371/journal.ppat.1009374> (2021).
- Danane, J., Hammouch, Z., Allali, K., Rashid, S. & Singh, J. A fractional-order model of coronavirus disease 2019 (COVID-19) with governmental action and individual reaction. *Math. Meth. Appl. Sci.* <https://doi.org/10.1002/mma.7759> (2021).
- Ucakan, Y., Gulen, S. & Koklu, K. Analysing of tuberculosis in Turkey through SIR, SEIR and BSEIR Mathematical Models. *Math. Comput. Model. Dyn. Syst.* **27**, 179–202 (2021).
- Fan, X. M. Global stability of multigroup SIRS epidemic model with varying population sizes and stochastic perturbation around equilibrium. *Abstr. Appl. Anal.* **2014**, 154725 (2014).
- Wang, Y. M. & Liu, G. R. Stability of a nonlinear stochastic epidemic model with transfer from infectious to susceptible. *Complexity* **2020**, 9614670 (2020).
- Rajasekar, S. P. & Pitchaimani, M. Ergodic stationary distribution and extinction of a stochastic SIRS epidemic model with logistic growth and nonlinear incidence. *Appl. Math. Comput.* **377**, 125143 (2020).
- Alota, C. P., Pilar-Arceo, C. P. C. & de Reyes, V. A. A. An edge-based model of SEIR epidemics on static random networks. *Bull. Math. Biol.* **82**, 96 (2020).
- Godio, A., Pace, F. & Vergnano, A. SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence. *Int. J. Environ. Res. Public Health.* **17**, 3535 (2020).



15. Rezapour, S., Mohammadi, H. & Samei, M. E. SEIR epidemic model for COVID-19 transmission by Caputo derivative of fractional order. *Adv. Differ. Equ.* **2020**, 490 (2020).
16. Marinca, B., Marinca, V. & Bogdan, C. Dynamics of SEIR epidemic model by optimal auxiliary functions method. *Chaos Solitons Fract.* **147**, 110949 (2021).
17. Chen, D. X., Yang, Y. F., Zhang, Y. F. & Yu, W. Prediction of COVID-19 spread by sliding mSEIR observer. *Sci. China-Inf. Sci.* **63**, 166–178 (2020).
18. Khan, M. A., Khan, Y. & Islam, S. Complex dynamics of an SEIR epidemic model with saturated incidence rate and treatment. *Phys. Stat. Mech. Appl.* **493**, 210–227 (2018).
19. Chimmula, V. K. R. & Zhang, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fract.* **135**, 109864 (2020).
20. Younsi, F. Z., Bounnekar, A. & Hamdadou, D. SEIR-SW, Simulation model of influenza spread based on the small world network. *Tsinghua Sci. Technol.* **20**, 460–473 (2015).
21. Katoch, S., Chauhan, S. S. & Kumar, V. A review on genetic algorithm: past, present, and future. *Multimed. Tools Appl.* **80**, 8091–8126 (2020).
22. Chen, N. C., Xiong, C., Du, W. Y., Wang, C. & Lin, X. An improved Genetic Algorithm coupling a back-propagation neural network model (IGA-BPNN) for water-level predictions. *Water* **11**, 1795 (2019).
23. Misevicius, A., Kuznecovaite, D. & Platuziene, J. Some further experiments with crossover operators for Genetic Algorithms. *Informatica* **29**, 499–516 (2018).
24. Chen, Z. Q., Zhou, J. H., Sun, R. Z. & Kang, L. A new evolving mechanism of genetic algorithm for multi-constraint intelligent camera path planning. *Soft Comput.* **25**, 5073–5092 (2021).
25. Yan, Y. Error recognition of robot kinematics parameters based on genetic algorithms. *J. Ambient. Intell. Humaniz. Comput.* **11**, 6167–6176 (2020).
26. Ragalo, A. & Pillay, N. Evolving dynamic fitness measures for genetic programming. *Expert Syst. Appl.* **109**, 162–187 (2018).
27. Ting, C. K., Wang, T. C., Liaw, R. T. & Hong, T. P. Genetic algorithm with a structure-based representation for genetic-fuzzy data mining. *Soft Comput.* **21**, 2871–2882 (2017).
28. Almeida, R. Analysis of a fractional SEIR model with treatment. *Appl. Math. Lett.* **84**, 56–62 (2018).
29. Wang, Y., Cao, J. D., Alsaedi, A. & Ahmad, B. Edge-based SEIR dynamics with or without infectious force in latent period on random networks. *Commun. Nonlinear Sci. Numer. Simul.* **45**, 35–54 (2017).
30. Kim, S., Byun, J. H. & Jung, I. H. Global stability of an SEIR epidemic model where empirical distribution of incubation period is approximated by Coxian distribution. *Adv. Differ. Equ.* **2019**, 469 (2019).
31. Sun, H. C., Liu, X. F., Xu, X. K. & Wu, Y. Analysis of COVID-19 spreading and prevention strategy in schools based on continuous infection model. *Acta Phys. Sinica.* **69**, 74–83 (2020).
32. Yang, Z. F. *et al.* Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.* **12**, 165–174 (2020).
33. Cai, Y. Y. *et al.* Association between incubation period and clinical characteristics of patients with COVID-19. *J. Int. Med. Res.* **48**, 0300060520956834 (2020).
34. Alenezi, M. N., Al-Anzi, F. S., Alabdulrazzaq, H., Alhusaini, A. & Al-Anzi, A. F. A study on the efficiency of the estimation models of COVID-19. *Results Phys.* **2021**, 104370 (2021).
35. Tang, B., Wang, X. & Li, Q. Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions. *J. Clin. Med.* **9**, 462 (2020).
36. Wan, K. K. *et al.* When will the battle against novel coronavirus end in Wuhan: A SEIR modeling analysis. *J. Glob. Health.* **10**, 011002 (2020).
37. Zhang, L. Y. & Zhou, J. Q. *Hubei Statistical Yearbook* 60–61 (China Statistical Publishing House, 2020).
38. Pang, J. Q. *Beijing Statistical Yearbook* 59+61–85 (Beijing: China Statistical Publishing House, 2020).
39. Leontitis, A. *et al.* SEAHIR: A Specialized Compartmental Model for COVID-19. *Int. J. Environ. Res. Public Health.* **18**, 2667 (2021).
40. Pang, L. Y., Liu, S. H., Zhang, X. N., Tian, T. H. & Zhao, Z. Transmission dynamics and control strategies of COVID-19 in Wuhan China. *J. Biol. Syst.* **28**, 543–560 (2020).

## Author contributions

Z.Q. and Y.S. contributed to research design, data analysis, and manuscript writing; X.H. and J.W. contributed to data acquisition and manuscript drafting; R.Z. and J.B. contributed to the critical revision of the manuscript's key content. S.D. gave key guidance to the revised manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-12958-z>.

**Correspondence** and requests for materials should be addressed to R.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022