



Review

A Review of Mathematical Models of COVID-19 Transmission

Huan Yang¹, Xiaolin Lin^{1*}, Jianquan Li¹, Yijiang Zhai², Jing Wu¹

¹School of Mathematics and Data Science, Shaanxi University of Science and Technology, Xi'an, China

²School of General Education, Shaanxi College of Communication Technology, Xi'an, China
E-mail: linxl@sust.edu.cn

Received: 24 October 2022; **Revised:** 2 February 2023; **Accepted:** 16 February 2023

Abstract: Corona Virus Disease 2019 (COVID-19) is a disease caused by the novel coronavirus, and due to its rapid spread, the World Health Organization has declared it a worldwide outbreak. Since the first case was detected in December 2019, the fight against this deadly disease has begun. Numerous studies have been done on the COVID-19 outbreak's model of dissemination and epidemic patterns. It mainly focuses on the selection of methods, the determination of parameters, and the evaluation of prevention and control measures. Taking the spread of the COVID-19 epidemic as the research object, combined with the current development of the COVID-19 epidemic, this paper sorts out the relevant mathematical models for the study of the spread of COVID-19, among which the models based on Susceptible-Infected-Recovered (SIR) model and Susceptible-Exposed-Infectious-Removed (SEIR) model and the mathematical models combined with these two models are mainly selected. Finally, the importance of reasonable and effective control of parameters and multi-model combined modeling is pointed out in the future.

Keywords: COVID-19, epidemiology, mathematical model

MSC: 92B05, 34C60

1. Introduction

In early December 2019, there was a rapid global spread of Corona Virus Disease 2019 (COVID-19), a highly contagious respiratory infection with typical symptoms including fever, cough, and dyspnea, which in severe cases can lead to pneumonia, acute respiratory syndrome, renal failure, and death [1-2]. The outbreak of COVID-19 has had a huge impact on countries around the world, with countless lives and properties lost, and is still raging in many countries and regions today. According to the latest data, as of June 6, 2022, there were 53.89 million confirmed cases and 6.29 million deaths worldwide, of which more than 880,000 cases had been reported in the United States. The epidemic's prevention and control are facing a serious situation of complexity, enormity, and recurrence. In addition to the biomedical and health care disciplines and industries that directly serve epidemic control, researchers from other disciplines are also actively participating in this global epidemic control sniper war with their expertise.

Prevention, detection, and control are the main methods to reduce the spread of viruses. Mathematical models and computer simulations, despite their limitations and shortcomings, are still one of the best methods to analyze the spread of viruses and control their prevalence. Modeling is very important in epidemiology because, in most cases, we cannot

perform biological experiments and there are no pharmaceutical solutions. Mathematical modeling must be used to develop and understand epidemiological phenomena in a relevant way as well as to quantitatively simulate the possible effects of different intervention strategies [3]. Therefore, it is very important to use suitable mathematical models to simulate, analyze, and predict the COVID-19 epidemic and make recommendations for prevention and control during the virus transmission stage.

This paper tries to make a brief summary of the mathematical models in the study of the law of the epidemic of COVID-19, focusing on reviewing the modeling methods of the transmission mechanism and the basic data sources. For the combination of the two models, it mainly introduces the specific methods and steps of prediction. These reviews provide important reference value for us to analyze the epidemic situation of novel coronavirus pneumonia scientifically, take effective prevention and control measures, and establish an early warning mechanism for infectious diseases in the future.

The transmission models of COVID-19 generally include the mathematical model based on SIR and SEIR, the discrete stochastic model, the multi-stage dynamic time-delay model, the micro-scale model of coinfection of SARS-CoV-2 and bacteria, the system dynamics model, the model based on space-time attention mechanisms, as well as the dual model combined with the SEIR model and the time series prediction model. The SIR Model can describe the trend of epidemic development and the general law of disease transmission as a whole, but it ignores some details, the most important of which is that it does not consider the incubation period. The SEIR model, which takes the incubation period into account, is also the most widely used in this COVID-19 outbreak. Therefore, most of the mathematical models summarized in this paper are based on SEIR. In the process of epidemic development, many external factors will affect the SEIR model, and the model cannot self-regulate. Therefore, some researchers combine the SEIR model and time series prediction so as to predict the development trend of the epidemic more accurately and provide reliable suggestions for epidemic prevention and control. Finally, we conclude with a conclusion and a vision for the future.

2. COVID-19 infectious disease model

2.1 SIR model

SIR is a classical mathematical model of infectious diseases, whose basic idea is to divide people in the outbreak population into a susceptible population (S), infected population (I), and removed population (R). The differential equations describing the classical SIR model are as follows:

$$\begin{cases} S' = -\beta SI, \\ I' = \beta SI - \gamma I, \\ R' = \gamma I. \end{cases} \quad (1)$$

Zhang L et al. [4] assumed that the total population will not be infected twice in the epidemic cycle, and that the total population will be N . $S(t)$, $I(t)$ and $R(t)$ denote the number of susceptible, infected and removed populations at time t , respectively, β denotes the probability of infected people infecting others, and γ denotes the probability of infected people recovering.

Based on the above model, they conducted a simulation using Matlab R2018b based on the data published by Harbin City and relevant parameters in the article by Yang Y et al. [5]. According to the population density of Harbin City and the intensity of epidemic prevention, the development of the epidemic situation was analyzed, and the turning point and development trend of the epidemic were predicted.

Zhang Y et al. [6] added a COVID-19 death population and improved the classical SIR model by dividing the population into four categories: uninfected S , infected I , cured R , and dead D , but without considering birth and natural death factors, and their improved SIR model is as follows:

$$\begin{cases} S' = \lambda IS - \mu S - mS, \\ I' = -\lambda IS, \\ R' = \mu S, \\ D' = mS, \end{cases} \quad (2)$$

where λ represents the daily exposure rate of patients, that is, the average number of effective contacts per patient per day; μ represents the daily cure rate; and m represents the infectious disease mortality rate. According to the improved SIR model, the COVID-19 epidemic data released on the official website of the National Health Commission on January 23, 2020, was taken as the initial value, and the Runge-Kutta method was used to simulate the spread of the epidemic numerically. The number of infected patients at three time points ($t = 10, 20,$ and $30 d$) was calculated by simulating different daily contact rates λ (taken as 0.30, 0.20, 0.15, 0.10, and 0.05, respectively) and compared with the actual data, thus verifying that the isolation method can effectively reduce the daily contact rate.

Yin N [7] carried out simulations in R data analysis software based on the SIR model in three major aspects: cross infection, contact infection, and average number of infected persons. Through the simulation results, he predicted the state of the spread of COVID-19 in a limited area and found out the methods to prevent and control the occurrence or spread of the epidemic.

In order to study the combined impact of different epidemiological factors (such as the shortage of medical resources and the possibility of reinfection) on the outbreak of COVID-19, Salman AM et al. [8] constructed the SIRS model as follows:

$$\begin{cases} S' = \eta N - \delta SI / N + \varepsilon R - \nu S, \\ I' = \delta SI / N - (\psi + \nu) I - \rho I / (\Phi + I), \\ R' = \psi I + \rho I / (\Phi + I) - (\varepsilon + \nu) R, \\ N = S + I + R, \end{cases} \quad (3)$$

where N is the total population, η is the birth rate, ν is the death rate, δ is the spread rate, and ψ is the population recovery rate, ε is the rate of reinfection, ρ indicates the unit of time to provide medical resources, Φ is half full and constant.

Salman AM et al. investigated the transmission dynamics of COVID-19 according to the daily confirmed case data released by the Ministry of Health of Malaysia. Based on the above model, neural network technology was used for parameter estimation, and MATLAB ode15s was used to simulate the development trend of COVID-19. The results showed that the model could well reflect the infection trajectory trend of the COVID-19 pandemic in Malaysia, and the transmission dynamics may be affected by reinfection and limited medical resources. Therefore, the use of mathematical modeling methods to understand the interactions between different epidemiological factors may help support the authorities in making informed decisions to effectively control the spread of the epidemic.

2.2 SEIR model and its improvement

Since the susceptible population generally has an incubation period and does not immediately show symptoms, it becomes a carrier of the pathogen and has a certain probability of infection. This transmission mode is more consistent with the spread of COVID-19. Therefore, the SEIR model is formed by introducing the lurker on the basis of the SIR model.

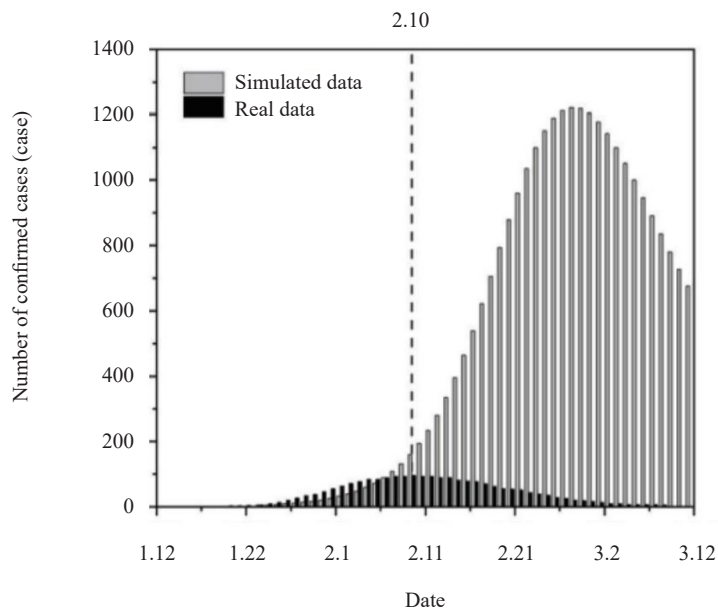
The SEIR model is currently the most representative mathematical model for studying the dynamics of infectious diseases in fixed populations. The general SEIR model classifies the study population into four types: susceptible, latent, infected, and emigrated states. However, at the beginning of the epidemic, Wan S et al. [9] found that SARS-CoV-2 was infectious during the incubation period, and the classical SEIR model could not simulate the transmission trend of COVID-19 well.

2.2.1 Introduction of a SEIR model with latent infection factors

Since COVID-19 is different from severe acute respiratory syndrome in 2013, latent patients are also infectious. Therefore, Chen L et al. [10] added the latent infectiousness factor to the standard SEIR model and constructed the following mathematical model:

$$\begin{cases} S' = -\beta_I I \cdot \frac{S}{N} - \beta_E E \cdot \frac{S}{N}, \\ E' = \beta_I I \cdot \frac{S}{N} + \beta_E E \cdot \frac{S}{N} - \alpha E, \\ I' = \alpha E - \gamma I, \\ R' = \gamma I, \end{cases} \quad (4)$$

where β_I is the transmission rate of the infected state I , i.e., the probability that a susceptible individual is infected by the infected state and enters the latent phase; β_E is the transmission rate of the latent state E , i.e. the probability that a susceptible individual is infected by the latent state and enters the latent phase; α is the probability that the latent state E is converted to the infected state I ; and γ is the removal rate of the infected state I , i.e., the sum of the recovery and mortality rates.



Source: Extracted from [10]

Figure 1. Comparison of SEIR simulations and real confirmed cases

The authors compared the SEIR model simulation with real COVID-19 data collected from the official website of the Shanxi Provincial Health Commission. The data of confirmed cases before February 10, 2020, were obtained by simulation, and the data of confirmed cases simulated by the model after February 10 were obtained under the same virus transmission speed and control effect (see Figure 1). Before February 10, the graph trend of simulated data was consistent with that of real data due to the fitting parameters of real data in the same period, but after February 10, there was a big difference between them. The real confirmed case data reached its peak on February 10 (96 cases), while the simulated data would reach its peak on March 2 (2091 cases). The simulated peak was 21 times higher than the real peak, indicating that the original transmission chain of the epidemic was gradually cut off under the strict epidemic

prevention and control measures of the whole society, and the gradual decline of the confirmed cases in Shanxi Province also showed the remarkable effect of strengthening the prevention and control measures.

2.2.2 SEIR model with logistic curve (LSEIR)

Due to the lack of timely and effective prevention and control measures in the early stages of the epidemic, the daily number of new infections increased rapidly. Later, with the implementation of intervention measures such as requiring the wearing of masks and closing the community, the daily number of new infections gradually decreased, which mainly affected the effective contact rate of the population. Shi C et al. [11] introduced a logistic curve to simulate a time-varying effective contact rate on the basis of the SEIR model and also considered the influence of imported cases to establish an improved Logistic-SEIR (LSEIR) model. Contact between quarantined individuals and susceptible individuals is not permitted under isolation measures. Assuming that the inflow and flow travel values of susceptible individuals do not change much, the following discrete dynamic model is established:

$$\begin{cases} S(t+1) = S(t) - \beta(t) \frac{S(t) \cdot I(t)}{N}, \\ E(t+1) = E(t) + \beta(t) \frac{S(t) \cdot I(t)}{N} - \sigma \cdot E(t) + E_m(t), \\ I(t+1) = I(t) + \sigma \cdot E(t) - \gamma \cdot I(t) + I_{in}(t), \\ R(t+1) = R(t) + \gamma \cdot I(t), \\ \beta(t) = a + \frac{b}{1 + e^{\lambda(t-d)}}, \\ N(1) = S(1) + E(1) + I(1) + R(1), \end{cases} \quad (5)$$

where $E_m(t)$ and $I_{in}(t)$ represent the number of imported latent cases and imported cases on day t , parameter d represents the time at the inflection point of the logistic curve, parameter $1/\sigma$ represents the average incubation period (days), and parameter $1/\gamma$ represents the time from onset to isolation or recovery (days).

Based on the individual data of Guangzhou, the above model was used to simulate the spread of COVID-19 under different interventions so as to quantitatively analyze the effects of different prevention and control measures, provide guidance for epidemic prevention and control, and also provide a methodological reference for research in other regions. However, there are still some limitations in the model, which is only based on the study of one city in the time dimension. If the specific geographic information of the infected person or the data of multiple cities can be obtained and the analysis can be carried out in the space-time dimension, the feasibility of the method can be further verified, and targeted prevention and control can be provided for different regions.

2.3 SEIR model for introducing prevention and control measures

2.3.1 SEIR model with different control measures and their intensities

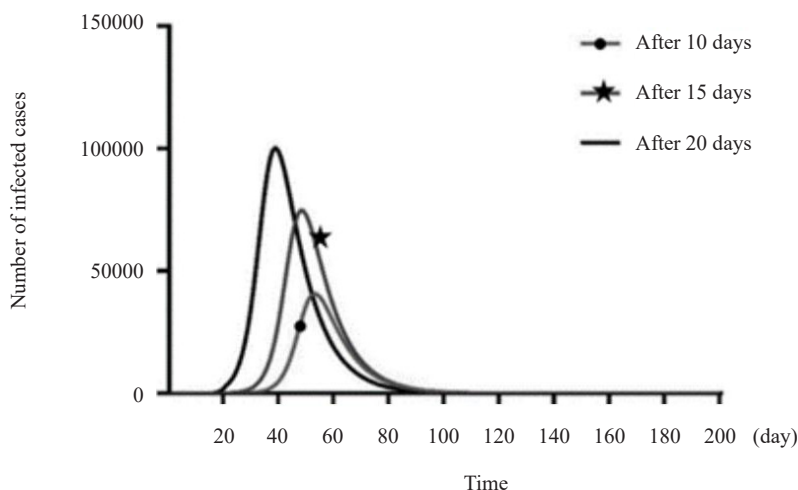
COVID-19 has a long incubation period and is highly contagious. In order to curb the spread of COVID-19 in a timely manner, a number of national and local policies have been introduced, such as extending holidays, traffic control, community prevention and control, delaying the start of school and work, urging people to wear masks, and going out less [12]. Therefore, Cai J et al. [13] established the SEIR model to analyze the impact of different control start times and different control intensities on the COVID-19 epidemic situation (Wuhan as an example). The differential equation of infectious disease dynamics of the specific SEIR model is as follows:

$$\begin{cases} S_i = S_{i-1} - \gamma\beta_1 I_{i-1} S_{i-1} / N - \gamma_1 \beta_2 E_{i-1} S_{i-1} / N, \\ E_i = E_{i-1} - \gamma_1 \beta_1 I_{i-1} S_{i-1} / N - \alpha E_{i-1} + \gamma_1 \beta_2 E_{i-1} S_{i-1} / N, \\ I_i = I_{i-1} + \alpha E_{i-1} - \gamma I_{i-1}, \\ R_i = R_{i-1} + \gamma I_{i-1}, \end{cases} \quad (6)$$

where S is the number of susceptible cases. Since all people are generally susceptible to COVID-19, S is the total number of cases. $\alpha = 0.1$ is the probability of the latent person turning into the infected person, β_1, β_2 respectively represent the probability of the infected person and the latent person transmitting to the susceptible group, and $\beta_1 = \beta_2 \approx 0.048$ is calculated according to the data published by the National Health Commission. γ_1, γ_2 are the number of cases where the infected person and the latent person contact the susceptible person. Because the government implements relatively strict control and quarantine at home, so $\gamma_1 = \gamma_2 = 3$, that is, only 3 cases were contacted at most; γ means the probability of recovery is 0.1.

Cai J et al. collected the official data of the COVID-19 epidemic in Wuhan from January 20 to February 18, 2020, built a SEIR model using MATLAB to simulate the development trend of the COVID-19 epidemic, and analyzed the correlation between the simulated COVID-19 epidemic in Wuhan and the actual epidemic using the Pearson correlation test. The influence of the prevention and control strategy on the development of COVID-19 in Wuhan was analyzed by adding different parameters, such as different durations of control (i.e., 10 days, 15 days, and 20 days after the lockdown of Wuhan on January 23) and different intensities of control (i.e., 10, 7, and 3 cases were exposed to susceptible persons, respectively).

The impact of different control start times on the COVID-19 outbreak in Wuhan is shown in Figure 2. The simulated peak number of COVID-19 infections was 40,850 cases, 74,734 cases, and 100,358 cases, respectively, after the control started on the 10th, 15th, and 20th day of Wuhan lockdown. Scenario 1 was 160.95% lower than Scenario 3, and scenario 1 was 98.67% lower than Scenario 2.

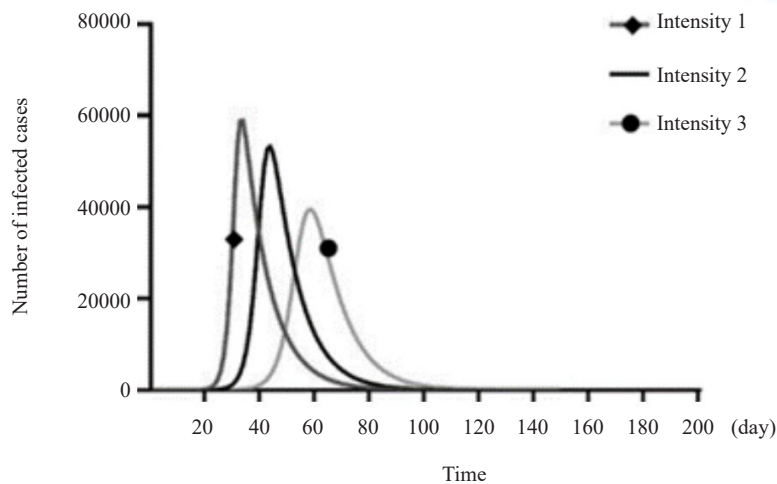


Source: Extracted from [13]

Figure 2. Comparison of the number of COVID-19 cases at different starting times of control

The influence of different control intensities on the epidemic situation of COVID-19 is shown in Figure 3. The results showed that the peak numbers of COVID-19 infections under different control intensities were 58,959, 53,279, and 39,485, respectively. The peak number of infection cases in intensity 3 was reduced by 49.21% compared with intensity 1, and the peak number of infection cases in intensity 3 was reduced by 34.8% compared with intensity 2.

According to the simulation results of the SEIR model above, local governments should take strict control measures more quickly to control the development of COVID-19. The earlier the control starts and the stronger the control intensity, the more effective the intervention will be. Governments at all levels should strictly control the spread of the epidemic and cut off the transmission route as soon as possible.



Source: Extracted from [13]

Figure 3. Comparison of the number of COVID-19 cases with different control intensities

2.3.2 SEIR model considering tracking isolation interventions

Due to the existence of patients with an incubation period who are not easy to effectively isolate and have a strong virus transmission ability, it is still necessary for researchers to study the COVID-19 transmission law, which has important practical significance for the situation analysis and prevention and control guidance of the epidemic. Therefore, Cao S et al. [14] established a modified SEIR infectious disease dynamics model that considered the virus transmission ability of patients with a latent period and the effect of tracking and isolation interventions on the epidemic situation.

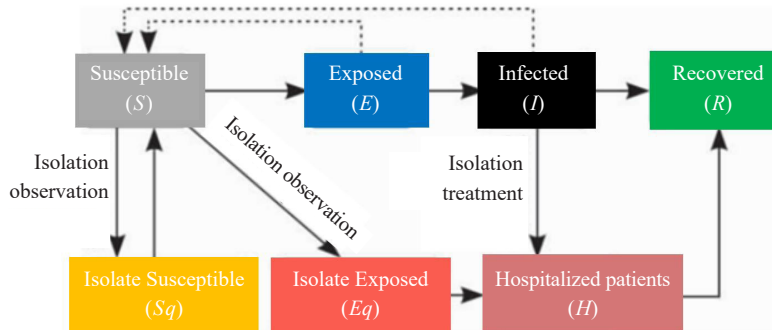
Based on the following assumptions: all individuals in the population have the probability of being infected, and when the infected individuals recover, they will produce antibodies; that is, the recovered population R will not be infected again. Considering the isolation measures to prevent and control infectious diseases, three groups of people were added to the population group in the classical model: isolated susceptible persons (S_q), isolated contacts (E_q), and isolated infected persons (I_q). In view of the fact that quarantined infected patients will be immediately sent to designated hospitals for isolation and treatment, all of these people are transformed into hospitalized patients H in this model. Thus, S , I and E in the revised model refer to susceptible, infected, and contacts who were missed by quarantine measures, respectively. Quarantined susceptible persons become susceptible again after release from quarantine, while infected persons and contacts have different degrees of infection ability to susceptible persons, so that they can transform into contacts. The transformation relationship of the crowd is shown in Figure 4.

The modified SEIR equation for COVID-19 was constructed as follows:

$$\begin{cases} S' = -[\rho c \beta + \rho c q(1 - \beta)]S(I + \theta E) + \lambda S_q, \\ E' = \rho c \beta(1 - q)S(I + \theta E) - \sigma E, \\ I' = \sigma E - (\delta_1 + \alpha + \gamma_1)I, \\ S'_q = \rho c q(1 - \beta)S(I + \theta E) - \lambda S_q, \\ E'_q = \rho c \beta q S(I + \theta E) - \delta_q E_q, \\ H' = \delta_1 I + \delta_q E_q - (\alpha + \gamma_H)H, \\ R' = \gamma_1 I + \gamma_H H, \end{cases} \quad (7)$$

where q is the isolation ratio, β is the probability of infection, c is the contact rate, ρ is the effective contact coefficient, and the reference value of the effective contact coefficient is set as 1, ρc is the effective contact rate. σ is the conversion rate of contacts to infected persons, and $\sigma = 1/7$, α is the case fatality rate, δ_1 is the isolation rate of infected persons, and

γ_1 is the recovery rate of infected persons. δ_q is the conversion rate from isolated contacts to isolated infected persons, and γ_H is the recovery rate of isolated infected persons.



Source: Extracted from [14]

Figure 4. Population transformation in the modified SEIR infectious disease dynamics model

Cao Shengli et al. developed a revised SEIR infectious disease dynamics model based on Hubei Province epidemic data from January 23 to February 24, 2020, and used the model to estimate the impact of prevention, control, isolation, centralized treatment, and other measures on the development of the epidemic. The results show that prevention, control, isolation, and medical tracking are important measures to effectively curb the spread of COVID-19. The home-based quarantine measures initiated by the government have played an important role in curbing the spread of the epidemic. Strict medical tracking and the isolation of people who have been in contact with infected people have also effectively prevented the rapid growth of the epidemic. Important measures such as centralized reception and stratified treatment played a key role in the rapid decline of the peak number of infected people and had an important control effect on the development of the epidemic. The improvement of personal safety protection measures will greatly curb the development of the epidemic, and strict self-protection measures will play a role in curbing the rapid growth of the number of infected people.

The results confirm that the model is reliable in the situation analysis of infectious disease transmission, which can provide some theoretical support for the formulation of future epidemic intervention decisions. However, there will inevitably be some differences between the model and reality, which will lead to some deviations in the analysis and prediction results.

2.4 SEIR model of different population was introduced

2.4.1 Introduction of SEIR model of close contacts (SEIRC)

Considering the current epidemic management program in China, anyone who has close contact with a confirmed or asymptomatic infected person will be isolated and observed, so Su Z et al. [15] built the following model based on the classical SEIR model population group with the addition of close contacts (including close contacts of close contacts):

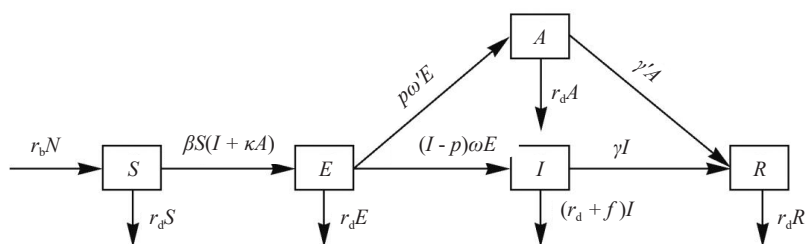
$$\begin{cases} S' = -\rho\omega\beta(1 + \rho_c)S(I + E) + \gamma_c C, \\ E' = \rho\omega\beta S(I + E) + \delta C - \sigma E, \\ I' = \sigma E + \sigma_c C - (\alpha + \gamma_1)I, \\ R' = \gamma_1 I, \\ C' = \rho\omega\beta S \rho_c - (\sigma_c + \delta + \gamma_c)C, \end{cases} \quad (8)$$

where C is the number of people in close contact, the probability of infection is β , the probability of susceptible contact is ω , the effective exposure factor for latents is ρ and the exposure factor for close contacts is ρ_c , σ is the rate of transition from latents to infected, σ_c is the rate of transition from close contacts to infected, δ is the rate of transition from close contacts to latents, α is the mortality rate, and γ is the recovery rate.

Su Z et al. used the Python language environment for code writing and simulation analysis, as well as data from the Harbin Municipal Government released in 2021. By comparing the number of infections simulated by the SEIRC model with the actual number of infections, it can be found that the epidemic in Harbin in early 2021 was well controlled and there was no large-scale explosive infection. According to the simulation of the number of infected people by the proportion of vaccination, the vaccine is still the best method to control the epidemic at the present stage, and the spread of the epidemic can be effectively prevented and controlled after the population has reached a certain proportion of vaccination.

2.4.2 Introduction of the SEIR model for latent infections (SEIAR)

Because most mathematical models of COVID-19 transmission do not take into account the issue of latent transmission capacity, that is, people who carry the virus without being quarantined or under medical observation but can still transmit it to susceptible people. For this reason, Chen T et al. [16] planned to adopt the Susceptible-Exposed-Infectious-Asymptomatic-Recovered (SEIAR) model for COVID-19 cases in Xiamen. The model makes relevant assumptions as follows: recessive infected person A is infectious, and its transmission ability is κ times that of dominant infected person I ($0 \leq \kappa \leq 1$); The proportion of latent infection was p , the latent and latent periods were $1/\omega$ and $1/\omega'$, respectively. The time interval from onset to first diagnosis in case I was $1/\gamma$. The case fatality rate was f , and the infectious period of recessive infection A was $1/\gamma'$. Thus, a schematic diagram of SEIAR is obtained, as shown in Figure 5.



Source: Extracted from [16]

Figure 5. SEIAR model for COVID-19 fitting

Establish the corresponding differential equation as follows:

$$\begin{cases} S' = r_b N - \beta S(I + \kappa A) - r_d S, \\ E' = \beta S(I + \kappa A) - p\omega' E - (1-p)\omega E - r_d E, \\ I' = (1-p)\omega E - \gamma I - (r_d + f)I, \\ A' = p\omega' E - \gamma' A - r_d A, \\ R' = \gamma I + \gamma' A - r_d R, \end{cases} \quad (9)$$

The Berkeley Madonna 8.3.1 software was used to simulate the above model, and the fourth-order Runge-Kutta method was used to solve the differential equation. The judgment basis of the output of the model fitting data was the minimum root mean square, and the chi-square test was used to test the goodness of fit between the output of the optimal result and the actual data.

Chen T et al. collected data on COVID-19 cases in Xiamen up to March 7, 2022, established a database, and divided the epidemic data in Xiamen into imported cases and secondary cases for separate modeling. Imported cases

were used as the source of infection, and the model's calculated secondary cases were fitted with actual secondary cases. The model fit the epidemic data well, and the goodness-of-fit results showed that the differences between the model and the actual data were not statistically significant. Overall, the SEIR model with the introduction of hidden transmitters was superior to the traditional SEIR model for studying the spread of the epidemic.

2.4.3 Introduction of the SEIR model for the dead population (SEIRD)

With the continuous deepening of research, scholars further introduced the death population to construct the SEIRD model. Ferjouchia H et al. [17] used the SEIRD model to provide a theoretical framework for the prediction of the continued epidemic of COVID-19 in Morocco, and the structure and parameters of the proposed model provided insights and ideas for the development of virus dynamics. Kliestik T et al. [18] used the SEIRD model to consider the measures taken by Azerbaijan to contain the epidemic and their effectiveness. Hu T et al. [19] used the Runge Kutta method to construct a prediction model based on improved SIR infectious disease infection and a SEIRD model based on causality, aiming at the transmission of infectious diseases in closed systems. The findings revealed that the number of infectious diseases infected was proportional to the daily population contact rate of latent individuals. After taking prevention and control measures, the infection rate and the number of infected people decreased. Viguerie A et al. [20] proposed a SEIRD mathematical model based on partial differential equations coupled with a non-uniform diffusion model, which describes the spatiotemporal spread of the COVID-19 pandemic, aims to capture dynamic data based on human habits and geographic features, and compares it with actual data in Lombardy, Italy. Strong consistency was found in the data, and simulations of scenarios exploring the relaxation of lockdown restrictions suggest that reopening lockdown restrictions should take into account the specific dynamics of local population density and contagion.

Based on the analysis of COVID-19 epidemic data and transmission mechanisms, Wang J et al. [21] divided infected people into asymptomatic infected persons (I_0), mildly infected persons (I_1), severely infected persons (I_2) and critically ill infected persons (I_3) based on the SEIRD model, and constructed a SEIRD model more consistent with the spread of COVID-19. In this transmission model, the basic assumptions are as follows: the latent population is the infected person with no obvious symptoms during the latent period [22], so the latent population is assumed not to be infectious. Recovered people are immune and no longer infectious; Patients must undergo stage I_1 when they reach stage I_2 and stage I_1, I_2 when they reach stage I_3 , and only patients in stage I_3 have a chance of dying.

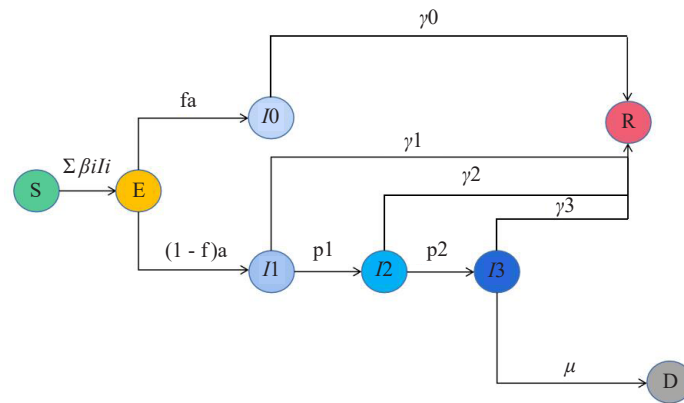
The specific model diagram and dynamic model are shown in Figure 6.

$$\begin{cases} S' = -(\beta_0 I_0 + \beta_1 I_1 - \beta_2 I_2 - \beta_3 I_3)S, \\ E' = (\beta_0 I_0 + \beta_1 I_1 - \beta_2 I_2 - \beta_3 I_3)S - aE, \\ I_0' = faE - \gamma_0 I_0, \\ I_1' = (1-f)aE - (\gamma_1 + p_1)I_1, \\ I_2' = p_1 I_1 - (\gamma_2 + p_2)I_2, \\ I_3' = p_2 I_2 - (\gamma_3 + \mu)I_3, \\ R' = \gamma_0 I_0 + \gamma_1 I_1 + \gamma_2 I_2 + \gamma_3 I_3, \\ D' = \mu I_3, \end{cases} \quad (10)$$

where a represents the rate of progression from stage E to I (where part f progresses to asymptomatic infection and part $1 - f$ to symptomatic infection), β_i represents the rate of recovery and immunization at stage I_i , p_i represents the rate of progression from stage I_i to stage I_{i+1} , μ represents the death rate at the most severe stage of the disease, all of which are in days.

Wang Jing used the coronavirus data in the United States from March to May 2020 provided on the official website of the coronavirus app, fitted the model parameters using the accumulated infection data collected, and simulated the epidemic trend under different transmission rates ($\beta_{0,1} = 0.2, 0.4, 0.6, 0.8$; $\beta_{2,3} = 0.02, 0.04, 0.06, 0.08$, respectively), showing that the main transmission rate affecting the development of the epidemic was the single-day transmission rate of asymptomatic and mildly infected patients, while the single-day transmission rate of severely and critically infected patients had minimal effect on the development of the epidemic, holding other influencing factors constant. The effect

of single-day transmission rates on the development of the epidemic was negligible for severely and critically infected individuals. Based on the changes in the number of infections and deaths resulting from simulated manual interventions (including the degree and duration of interventions), the earlier the intervention is implemented, the more it reduces the negative impact of the epidemic, reduces the rate of infection and mortality, and to some extent slows down the development of the epidemic and reduces social fear.

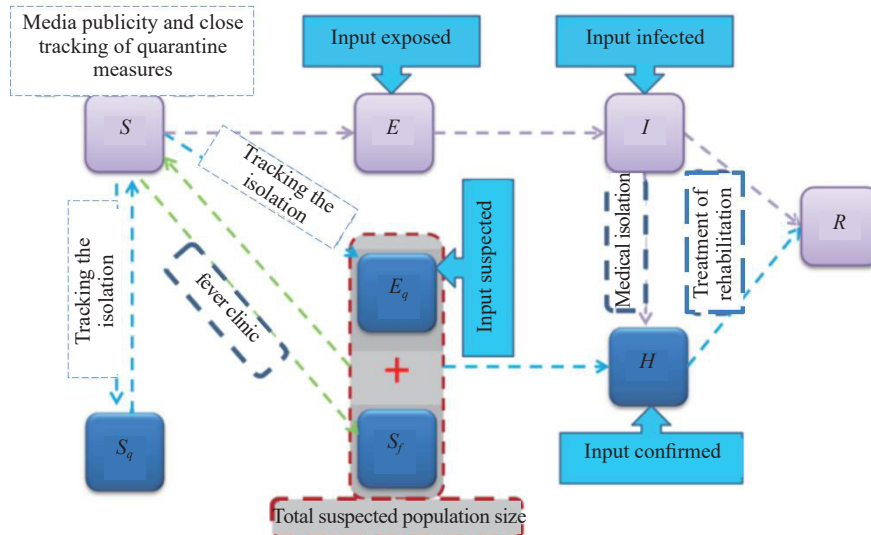


Source: Extracted from [21]

Figure 6. The COVID-19 epidemic SEIIRD

2.5 Other infectious disease models

2.5.1 Discrete stochastic model



Source: Extracted from [23]

Figure 7. Schematic diagram of COVID-19 epidemic and prevention and control strategy model

Based on the transmission mechanism of the disease, the epidemiological status of individuals, and prevention and control measures, and considering the small number of imported cases in Shaanxi Province in the early years, a differential system with random inputs is more suitable to portray this situation than a deterministic model, and can well

describe the randomness of the population statistics. Thus, Tang S et al. [23] divided the total population into seven bins, including susceptible (S), exposed (E), infected (I), hospitalized (H), cured (R), isolated susceptible (Sq) and isolated suspected (B), with the flow shown in Figure 7, and developed a new model as follows [24].

Prevention and control strategies include close follow-up and hospital isolation; grey areas indicate suspected cases and febrile outpatients.

$$\begin{cases} S_{t+1} = S_t - D_{11}(t) - D_{12}(t) - D_{13}(t) + D_{51}(t) + (1-f)D_{41}(t), \\ E_{t+1} = E_t - (1-q(t))D_{11}(t) - D_{21}(t) + P_E(\lambda_E), \\ I_{t+1} = I_t + D_{21}(t) - D_{31}(t) - D_{32}(t) - D_{33}(t) + P_I(\lambda_I), \\ B_{t+1} = B_t + q(t)D_{11}(t) + D_{13}(t) - D_{41}(t) + P_B(\lambda_B), \\ S_{qt+1} = S_{qt} + D_{12}(t) - D_{51}(t), \\ H_{t+1} = H_t + D_{31}(t) + fD_{41}(t) - D_{61}(t) - D_{62}(t) + P_H(\lambda_H), \\ R_{t+1} = R_t + D_{33}(t) + D_{61}(t). \end{cases} \quad (11)$$

They used detailed data released by the Shaanxi Provincial Health Commission to find the infection chain of COVID-19 transmission in Shaanxi Province and calculate the average number of control regrowth under strict control strategies. The effective regeneration number was estimated using the inverse convolution method and the distribution of successive generations. The conclusion showed that the effective regeneration number had been declining since January 23, 2020, and was stable and almost zero until February 23, which was in good agreement with the number of no new confirmed cases or very few confirmed cases at that time. Secondly, in order to describe the relatively small number of cases (about half of the cases are imported) and the resulting randomness, they established a discrete system with random input. This model framework addresses the significant limitation that deterministic models cannot describe imported cases and randomness [25-26]. Finally, they used the data on population movement in Shaanxi Province and other cities to study the impact of population movement (return to work) on the second outbreak of the epidemic. The main results show that, with a small proportion of imported cases, intermittent population inflows are maintained, with a low risk of transmission despite small secondary outbreaks. The proposed stochastic discrete model and its analysis method can be used as an important model framework to study the development trend of the epidemic in Shaanxi Province or other similar provinces (the number of cases is small and the proportion of imported cases is large). However, one disadvantage of this type of model framework is that conducting a systematic theoretical study of the threshold theory is difficult.

2.5.2 Multi-stage dynamic time-delay dynamic model

At the beginning of the outbreak, people did not pay enough attention to the epidemic in foreign countries, and the diagnosis and testing were not timely, resulting in a big difference between the recorded data and the actual number of infected people, which is not conducive to the analysis of the epidemic. In view of this situation, Zhang L et al. [27] improved the classical epidemic dynamics model and established a discrete-time multi-stage dynamic time-delay dynamics model by comprehensively considering the development characteristics of the epidemic, the influence of intervention, medical conditions, experience transmission, and other factors, combined with the development process of the epidemic in China. Based on the classical SEIR model, this model divides the virus transmission cycle into six stages, and uses the instantaneous value at a certain moment to process each stage, so it can be used to analyze the epidemic situation in foreign countries. Assuming that the number of susceptible persons exposed to confirmed infected persons within a day is $Q_1(t)$, and the number of susceptible persons exposed to latently infected persons within a day is $Q_2(t)$, both of which are decreasing functions of time and have exponential form, the time segments of $Q_1(t)$ and $Q_2(t)$ are expressed as follows:

$$Q_1(t) = \begin{cases} n_1, & \text{stage 1} \\ (n_1 - n_2)e^{-\kappa_1 t - 15} + n_2, & \text{stage 2} \\ (n_2 - n_3)e^{-\kappa_2 t - 22} + n_3, & \text{stage 3-6} \end{cases} \quad (12)$$

$$Q_2(t) = \begin{cases} n'_1, & \text{stage 1} \\ (n'_1 - n'_2)e^{-\kappa'_1 t - 15} + n'_2, & \text{stage 2} \\ (n'_2 - n'_3)e^{-\kappa'_2 t - 22} + n'_3, & \text{stage 3-6} \end{cases} \quad (13)$$

among them, n_1 and n'_1 representing the first stage confirmed infections and latent infections every day come into contact with the number of susceptible people, n_2 and n_3 represent the second phase, the second phase in the late diagnosis of infection of susceptible people contacts every day. n'_2 and n'_3 respectively represent the minimum number of susceptible persons exposed to the infected persons in the second stage and the late stage of the second stage, κ_1 and κ_2 respectively represent the exponential decline rate of the confirmed persons in the second stage and the third stage, and κ'_1 , κ'_2 respectively represent the exponential decline rate of the infected persons in the second stage and the third stage.

The transmission status and policy implementation at different stages determine the effective contact rate of infected persons. According to the transmission characteristic parameters at different periods above, the following model is established:

$$\begin{cases} S(t+1) - S(t) = -\alpha_{1i}\beta_{1j}I(t) - \alpha_{2m}\beta_{2n}E(t), \\ E(t+1) - E(t) = -\alpha_{1i}\beta_{1j}I(t) - \alpha_{2m}\beta_{2n}E(t) - \\ \quad \alpha_{2i}\beta_{1j}I(t-t_1) - \alpha_{2m}\beta_{2n}E(t-t_1), \\ I(t+1) - I(t) = -\alpha_{1i}\beta_{1j}I(t-t_1) - \alpha_{2m}\beta_{2n}E(t-t_1) - \\ \quad \delta\alpha_{1i}\beta_{1j}I(t-t_1-t_2) - \alpha_{2m}\beta_{2n}E(t-t_1-t_2), \\ R(t+1) - R(t) = \delta\alpha_{1i}\beta_{1j}I(t-t_1-t_2) - \alpha_{2m}\beta_{2n}E(t-t_1-t_2), \end{cases} \quad (14)$$

where α_{1i} and α_{2m} ($i, m = 1, 2, 3$) respectively refer to the daily contact number of susceptible persons with confirmed infected persons and latently infected persons, and β_{1j} and β_{2n} ($j, n = 1, 2, 3$) respectively refer to the contact probability of susceptible persons with confirmed infected persons and latently infected persons in different periods.

The difference between equations (12) and (13) is that the corresponding coefficients α_{1i} , α_{2m} , β_{1j} and β_{2n} are different in different stages of epidemic prevention and control: the first stage is the initial stage of the outbreak, β_{11} is the infection rate of the confirmed infected person, and α_{11} is the effective contact number. β_{21} is the infection rate of the infected person during the incubation period, α_{21} is the number of effective contacts. The second stage was the transmission period under the quarantine policy. Due to the effective implementation of the epidemic prevention policy, the probability of infection was reduced when the infected persons contacted healthy people, and the effective contact rates of latent and confirmed infected persons were β_{12} and β_{22} , respectively. The third stage is the emergency prevention stage. Due to the implementation of the policy in the last stage, the isolation efforts are increased, and the contact rate of infected persons is reduced. The effective contact rate of latent infected persons and confirmed infected persons is β_{13} and β_{23} , respectively. In the fourth stage, the intensity of isolation did not change, but with the transition of the incubation period, the number of infected people reached its peak, and the coefficient of infection remained unchanged. The fifth stage was the palliative period, in which medical resources were guaranteed to a certain extent. The decrease in the number of infected persons and the increase in the cure rate improved the situation of epidemic transmission, prevention, and control. At this time, the number of susceptible persons who had contact with confirmed infected persons and latent infected persons was α_{12} and α_{22} respectively. In the sixth stage, the number of infected people decreased rapidly, and all aspects showed the optimization of the palliative period. The numbers of susceptible people exposed to confirmed infected people and latently infected people were α_{13} and α_{23} , respectively.

According to the epidemic data released by the World Health Organization, the above model adopts the discrete time method, improves the multi-time delay dynamic model, describes the transmission of the epidemic in different

stages of epidemic prevention and control, and is suitable for the simulation of the transmission process of COVID-19 in various countries. However, this model does not consider the flow of people from abroad. After the epidemic enters the sixth stage or the end of one cycle, the next cycle may come, and the virus at this time will often mutate, such as the delta virus. Therefore, the model needs to be continuously optimized while considering external factors.

2.5.3 Introducing microscale mathematical models of bacteria

Studies have shown that the number of inflammatory monocytes and neutrophils increases in patients with COVID-19, while the number of lymphocytes decreases sharply [28-30]. This phenomenon can also be observed in transgenic mouse models [31-32]. These various available data point to a significant increase in morbidity and mortality if SARS-CoV-2 infected individuals become infected with a mix of bacterial pathogens. Understanding the mechanisms by which these pathogens interact with the immune system is key to developing effective treatments for COVID-19. However, macroscopic mathematical models can only be used to study the influence of crowd behavior on the development of epidemics, and it is difficult to penetrate into the pathogenesis of the disease. Meanwhile, existing clinical data show that patients infected with SARS-CoV-2 have the possibility of co-infection with bacteria.

In order to elucidate the mechanism of coinfection of SARS-CoV-2 and bacteria and to find the reason for the abnormal number of lymphocytes and neutrophils, Zhou Yu et al. [33] proposed a mathematical model and quantitative analysis to understand how lymphocytes and neutrophils affect the dynamics of coinfection.

Firstly, by analyzing the clinical data of patients who died in Wuhan University People's Hospital and combining it with clinical data from other hospitals, they found that lymphocyte count and neutrophil count could be used as indicators of the severity of the patients' disease. In order to describe the dynamic process of COVID-19, the sample data were transformed into time series data by interpolation. Secondly, bacteria were introduced into the model, with [LYM], [Neu], [Virus], and [Bac] representing the concentrations of lymphocytes, neutrophils, SARS-Cov-2, and bacteria, respectively. The corresponding mathematical models were based on the law of mass action, the Michaelis-Menten equation, and the Hill equation, whereupon the following interaction diagrams and models were available for these four variables:

$$\begin{cases} \frac{d[Neu]}{dt} = r_N \left(1 - \frac{[Neu]}{N_m (1 + f_1([Bac]))} \right) - f_5([Virus])[Neu][Bac], \\ \frac{d[LYM]}{dt} = r_L \left(1 - \frac{[LYM]}{L_m (1 + f_2([Virus]))} \right) - f_3([Virus])[LYM], \\ \frac{d[Virus]}{dt} = \frac{r_v [Virus]}{1 + \alpha_6 [Virus]} - f_4[LYM][Virus] - d_v [Virus], \\ \frac{d[Bac]}{dt} = r_B [Bac] \left[1 - \frac{[Bac]}{B_m} \right] - f_5 [Virus][Neu][Bac], \end{cases} \quad (15)$$

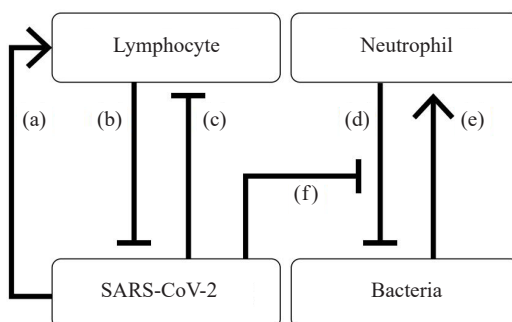
where

$$f_i(x) = \eta_i \frac{(\alpha_i x)^{n_i}}{1 + (\alpha_i x)^{n_i}}, \quad i = 1, 2, 3, \quad f_4(x) = \alpha_4 \frac{(\gamma_1 x)^{n_4}}{1 + (\gamma_1 x)^{n_4}}, \quad f_5(x) = nB \frac{1}{1 + (\alpha_5 x)^{n_5}},$$

$$\text{and } n_1 = 3, n_2 = 2, n_3 = 4, n_4 = 3, n_5 = 3.$$

In the first equation [Neu], the first term indicates that neutrophil production is linearly proportional to proliferation itself, and is activated by bacterial infection, which is described by a non-linear Hill function f_1 based on process (e) in Figure 8. The second term indicates that neutrophil apoptosis is linearly proportional to neutrophil phagocytosis and

bacterial infection, but non-linearly proportional to viral infection, which is represented by a non-linear Hill function f_5 based on process (f) in Figure 8, because neutrophils are weakened when viral infection is sufficiently severe. The derivation of the second equation [LYM] is almost identical to that of the first, with the two Hill functions f_2 and f_3 and the two processes (a) and (c), respectively, used to represent the two processes in Figure 8. The first term in the third equation [Virus] and the first term in the fourth equation [Bac] are the growth of viruses according to the Michaelis-Menten function and the growth of bacteria according to the logistic growth function, respectively, which are often used to describe the growth of viruses and bacteria because viruses and bacteria are not exponential and cannot grow infinitely in the human body. In the second term of the third equation and the second term of the fourth equation, the processes (b) and (d) in Figure 8 are represented by the two Hill functions f_4 and f_5 , respectively. The third term in the third equation is the degradation of the virus according to the law of mass action.



Source: Extracted from [33]

Figure 8. Illustration of a SARS-CoV-2 and bacterial co-infection system

The above model is dimensionless:

$$N = \frac{[Neu]}{N_m}, L = \frac{[LYM]}{L_m}, V = \alpha_6[Virus], B = \frac{[Bac]}{B_m}, T = r_B t.$$

Then the following model can be obtained:

$$\begin{cases} \frac{dN}{dT} = k_1 \left[1 - \frac{N}{1 + g_1(B)} \right] - k_3 g_5(V) NB, \\ \frac{dL}{dT} = k_5 \left[1 - \frac{N}{1 + g_2(V)} \right] - g_3(V) L, \\ \frac{dV}{dT} = \frac{k_9 V}{1 + V} - g_4(L) V - k_{12} V, \\ \frac{dB}{dT} = B(1 - B) - k_{13} g_5(V) NB, \end{cases} \quad (16)$$

where

$$g_1(B) = \frac{\eta_1(B)^{n_1}}{k_2^{n_1} + (B)^{n_1}}, g_2(V) = \frac{\eta_2(V)^{n_2}}{k_6^{n_2} + (V)^{n_2}}, g_3(V) = \frac{k_8(V)^{n_3}}{k_7^{n_3} + (V)^{n_3}},$$

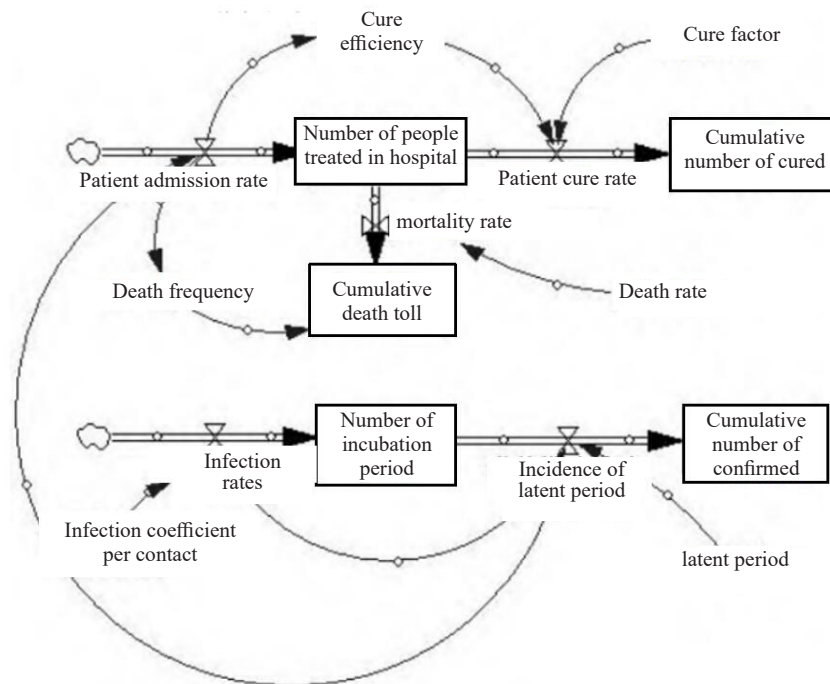
$$g_4(L) = \frac{k_{11}(L)^{n_4}}{k_{10}^{n_4} + (L)^{n_4}}, g_5(V) = \frac{k_4^{n_5}}{k_4^{n_5} + (V)^{n_5}},$$

and

$$k_1 = \frac{r_N}{N_m r_B}, k_2 = \frac{1}{\alpha_1 B_m}, k_3 = \frac{n_B B_m}{r_B}, k_4 = \frac{\alpha_6}{\alpha_5}, k_5 = \frac{r_L}{L_m r_B}, k_6 = \frac{\alpha_6}{\alpha_2},$$

$$k_7 = \frac{\alpha_6}{\alpha_3}, k_8 = \frac{\eta_3}{r_B}, k_9 = \frac{r_V}{r_B}, k_{10} = \frac{1}{\gamma_1 L_m}, k_{11} = \frac{\alpha_4}{r_B}, k_{12} = \frac{d_V}{r_B}, k_{13} = \frac{n_B N_m}{r_B}.$$

The model was further validated by theoretical analysis and the fitting of different clinical data. Finally, through branching analysis and numerical simulations, it was found that bacterial infection and immune-related parameters within a certain range lead to a shift in the system between three steady states, namely mild, severe, and death. Furthermore, they analyzed the impact of the time it takes for a patient to switch from a high-risk to a low-risk zone on the recovery time. The results suggest that co-infection with viruses and bacteria can explain changes in neutrophils and lymphocytes and that initial bacterial infection and immune-related parameters have a significant impact on the severity and mortality of patients with neocolonial pneumonia. The model and quantitative analysis suggest that prevention of bacterial infection and immune enhancement in the early stages of infection may be a potential therapeutic option for patients at high risk of COVID-19.



Source: Extracted from [41]

Figure 9. System dynamics flow diagram of COVID-19

2.5.4 System dynamics models

Yang S et al. [34] established a system dynamics simulation model based on the characteristics of COVID-19 infection, combined with different economic and social environments, and comprehensively considered the prevention and control effects and economic inhibition effects of three public response programs: “economic operation protection”, “public health protection”, and “balanced consideration”. Liu H et al. [35] constructed an analysis model of COVID-19 patient change based on system dynamics to reasonably estimate and predict the number change of confirmed patients in Gansu Province. Wang Y et al. [36] established a comprehensive prediction model for epidemic transmission in Italy based on the analysis results of the transmission system dynamics model in the early stage of COVID-19 epidemic.

The system dynamics model of COVID-19 mainly involves decision variables such as cumulative number of confirmed cases, cumulative number of cured persons, cumulative number of cured persons, contact infection rate and admission rate of patients [37-40]. According to the transmission characteristics of novel coronavirus and the incidence process of novel coronavirus, Yi J [41] studied and analyzed the epidemic scale in Beijing since 2020-01-21 based on the data released by Beijing Municipal Health Commission, and built a novel coronavirus transmission system dynamics model based on the system dynamics theory. The model includes 12 variables, 4 constants, and 13 equations, In hospital treatment, the number of state variables, (the cumulative number of cure, accumulative total number of confirmed deaths, the number of the incubation period, the cumulative), rate variable (patients hospitalized rate and cure rate, mortality rate, infection rate, incidence of latent), auxiliary variables (cure efficiency, frequency of death), constant (cure factor, death factor, the incubation period, each contact transmission coefficient). The system dynamics flow diagram of COVID-19 is shown in Figure 9.

The corresponding model equations and parameters are set as follows:

$$IP = INTEG(MR, 5), \tag{17}$$

$$IPR = INTEG(IR - MR, 21). \tag{18}$$

In (17)-(18): IP is the cumulative number of confirmed cases; $INTEG$ function is an integral function; MR is the incidence rate of patients; IPR is the number of latent cases; IR is the transmission rate.

The initial prediction date for this model was January 21, 2020, with a cumulative total of 5 confirmed cases and 21 close contacts as of that date.

$$CP = INTEG(CR, 0), \tag{19}$$

$$HP = INTEG(HR - DR - CR, 5). \tag{20}$$

In equations (19)-(20): CP is the cumulative number of cures; CR is the patient cure rate; 0 is the initial value of the cumulative number of cures at the initial time; HP is the number of patients in hospital; HR is the patient admission rate; DR is the mortality rate.

$$DP = INTEG(DR, 0). \tag{21}$$

In (21): DP is the cumulative number of deaths; 0 is the initial value of the cumulative number of deaths.

$$HR = MR, \tag{22}$$

$$IR = I_1 * IPR, \tag{23}$$

$$CR = Ce * Cf, \tag{24}$$

$$DR = Df * df. \tag{25}$$

In equations (22)-(25): I_1 is the transmission coefficient; Ce , Cf , Df , df is the cure efficiency, cure factor, death factor and death frequency respectively, where the cure factor = 1 - death factor, and the Beijing death factor is set at 0.02.

$$df = DRLAY \cdot FIXED(HR, 6, 0), \tag{26}$$

$$Ce = DRLAY \cdot FIXED(HR, 16, 0), \tag{27}$$

$$MR = DRLAY \cdot FIXED(IR, IP, 0). \tag{28}$$

Equations (26)-(28) are both delay functions, with 6 representing the period from diagnosis to death for deceased patients and 16 representing the period during which patients treated in hospital are discharged.

$$I_1 = 0.2 - STEP(0.11, 20). \tag{29}$$

Equation (29) is a step function, and the transmission coefficient of each contact was 0.2 at the early stage of the epidemic in Beijing and decreased to 0.09 after 20 days, so the coefficient of each contact in Beijing is a function of time change.

Table 1. Comparison between simulation results and actual values of the COVID-19 System Dynamics Model (Unit: person)

Date	2020-02-14	2020-02-15	2020-02-16	2020-02-17	2020-02-18	2020-02-19
Confirmed predicted value	337	351	365	377	387	396
Confirmed actual value	372	375	378	379	387	393

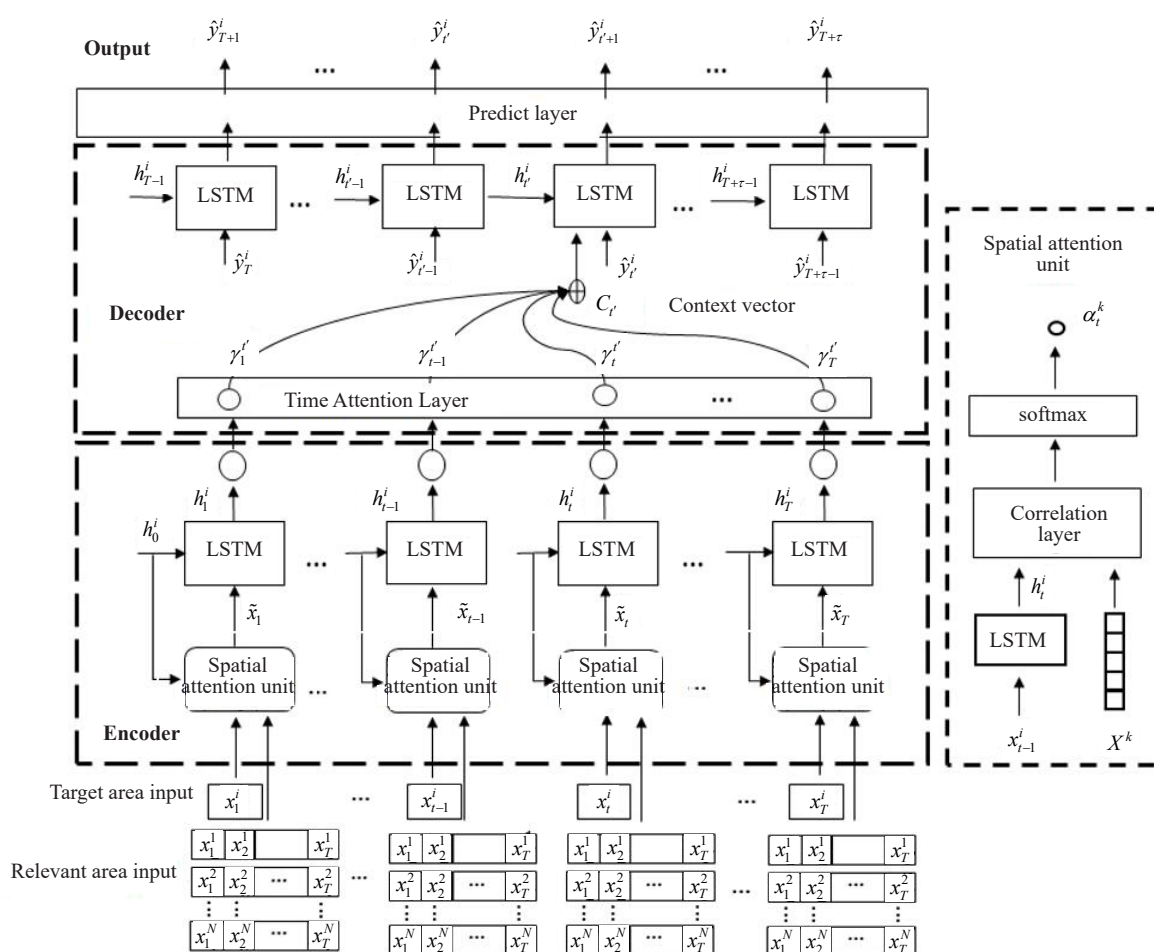
Source: Jingyu Yi (2022)

The results (see Table 1) showed that the system dynamics model of COVID-19 could well predict and fit the number of confirmed cases, cured cases, and deaths during the one-month epidemic's development in Beijing. The one-time test of the model reflects that the model has good consistency in predicting COVID-19 at a certain stage, and the model can play a good role in predicting the development trend of the epidemic.

2.5.5 A model based on spatio-temporal attention mechanisms

The transmission of infectious diseases is multi-dimensional. In addition to the distribution characteristics in time, another important aspect is the distribution characteristics in space. Therefore, through the application of spatial analysis theory and technology, the establishment of a spatio-temporal transmission model can more intuitively simulate and predict the spatio-temporal spread process of infectious diseases. Therefore, in view of the characteristics of complex regional association and strong temporal dependence during the spread of COVID-19 [42], Bao Xin et al. [43] proposed a spatio-temporal attention-driven autoencoder framework to capture the dynamic spatial correlation between infection sequences by introducing a spatial attention mechanism. Moreover, the temporal attention mechanism was used to mine the complex temporal dependence in the epidemic sequence so as to accurately predict the spread trend of SARS-CoV-2 in different regions.

The framework of the COVID-19 epidemic trend prediction method driven by spatio-temporal attention is shown in Figure 10. The overall architecture of the model adopts the structure of an autoencoder, and each codec contains an Long Short-Term Memory network (LSTM), which is used to extract the temporal characteristics of epidemic sequence data and predict the future epidemic trend, respectively. The model consists of two key parts, namely the spatial attention mechanism at the encoder and the temporal attention mechanism at the decoder. On the encoder side of the model, the spatial attention mechanism is first introduced to learn the association between the target region and the epidemic sequence in its related region. Then, based on the correlation information, the epidemic time series data of relevant regions were introduced and fused with the epidemic time series data of target regions. The enhanced time series features for predicting the epidemic situation in target regions were extracted by the LSTM network. At the decoder end of the model, the temporal attention mechanism and LSTM network are fused to effectively capture the dynamic temporal dependence in the epidemic time series, so as to predict the epidemic dynamics in the future period more accurately.



Source: Extracted from [43]

Figure 10. COVID-19 epidemic prediction framework based on spatio-temporal attention mechanism

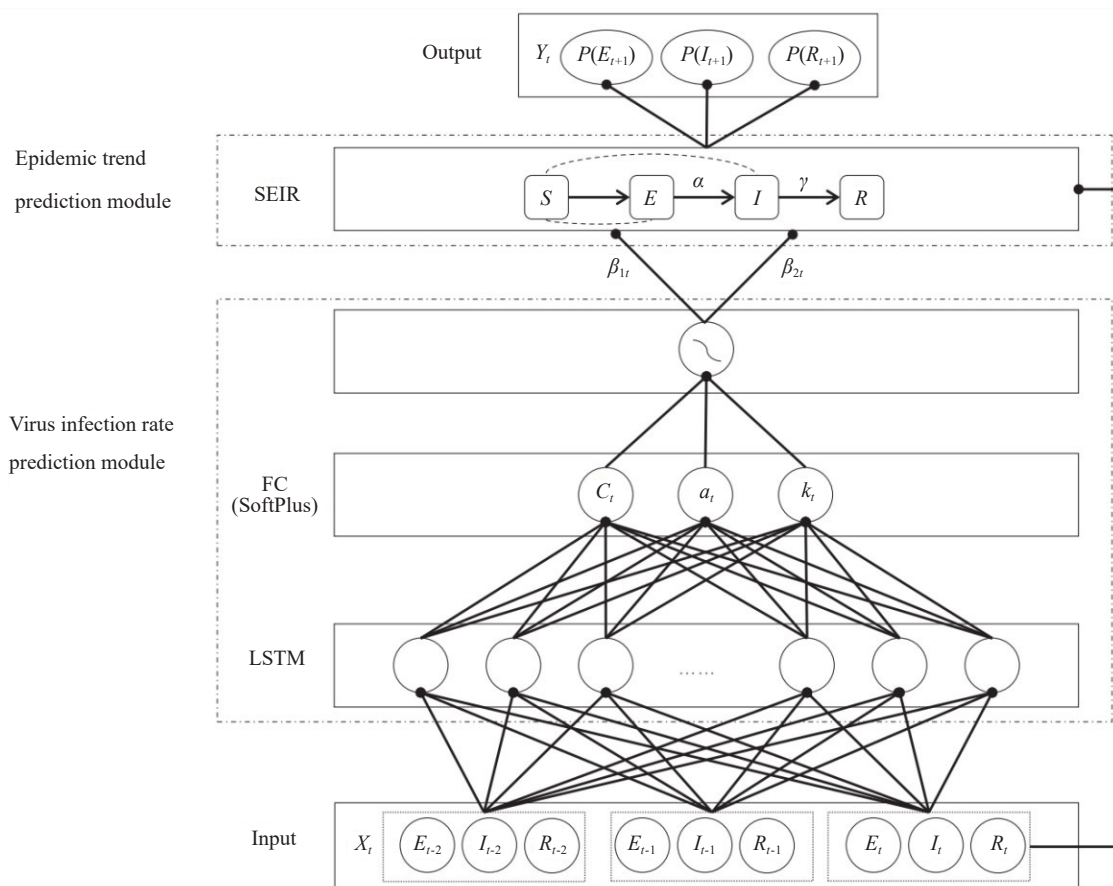
A large number of experiments were conducted on the publicly available COVID-19 data sets, and the results verified the effectiveness of the model. The prediction error indices root mean squared error and mean absolute error on the European COVID-19 dataset were reduced by 21.7% and 24.5%, respectively, and root mean squared error and mean absolute error on the Chinese COVID-19 dataset were reduced by 22.6% and 26.7%, respectively.

2.6 A mathematical model combining two models

2.6.1 LSTM combined with SEIR model (LS-Net)

With the continuous progress and development of deep learning in recent years, it stands out in many fields for its excellent experimental performance. In the temporal prediction problem, the most common recurrent neural network can learn the correlation information between the historical input and the current input, so as to make a temporal prediction. However, if the time series is too long, the traditional recurrent neural network is prone to the phenomenon of gradient disappearance and cannot deal with problems with long-term dependence. The later LSTM can solve this problem well. In the current COVID-19 outbreak, the change and development of the epidemic will be affected by various measures in the early stage, so the trend prediction of the epidemic can be regarded as a time-series prediction problem with long-term dependence.

Therefore, Gan Yu et al. [44] made use of the advantages of the traditional mathematical method of the SEIR model and the artificial intelligence method in timing prediction and used the PyTorch deep learning framework and Python language to build a prediction model. Based on the parameters of the SEIR model and the characteristics of COVID-19 transmission, this paper proposes a prediction method of virus infection rate based on LSTM and, combining it with the SEIR model, proposes a prediction model of the COVID-19 epidemic trend, the LSTM-SEIR Network (LS-Net). In addition, the epidemic data released by many provinces and cities in China from December 8, 2019 to March 28, 2020 was collected for experiments.



Source: Extracted from [44]

Figure 11. LS-Net overall structure

The overall structure of LS-Net is shown in Figure 11. The epidemic trend prediction module includes a SEIR model layer, which mainly realizes the prediction of the COVID-19 epidemic trend, while the virus infection rate prediction module includes an LSTM layer, a fully connected layer, and a nonlinear transformation layer, mainly to achieve the prediction of the COVID-19 infection rate.

LS-Net uses LSTM's excellent time sequence information learning ability to predict the virus infection rate in real time. Finally, the real-time updated virus infection rate was applied to the SEIR model to realize the intelligent dynamic prediction of the epidemic development trend. The experimental results show that LS-Net can eliminate the tedious process of manually splitting time periods for dynamic calculation, and use a small amount of training data to predict the peak value of existing confirmed cases in the epidemic. Therefore, this model can provide a reference for people to predict the development trend of the epidemic under the current prevention and control measures to a certain extent.

2.6.2 A combined SEIR and ARIMA model (SEIR-ARIMA hybrid model)

The Autoregressive Integrated Moving Average (ARIMA) model is a time series prediction method that mainly captures the fluctuations of historical data and describes the relationship between current values and historical values. It is suitable for COVID-19 prediction with time series. Moreover, its own model is constructed with endogenous variables and is not affected by exogenous variables, so it can well solve the interference of external factors.

Although the SEIR model can predict the development of the epidemic situation well, the initial probability of infection is only estimated, and the probability of infection will be reduced under the implementation of epidemic prevention and control, so the SEIR model cannot self-adjust according to the historical fluctuation data. The advantages of the ARIMA model, which is determined by endogenous variables and not affected by external factors, can make up for this shortcoming. Therefore, Dong Z et al. [45] constructed the SEIR-ARIMA hybrid model, and established a special COVID-19 epidemic map by collecting and analyzing statistical data from the official website of the National Health Commission on COVID-19 in all provinces, so as to predict and analyze the COVID-19 epidemic in different time periods and locations.

The model is based on linear regression of the SEIR and ARIMA prediction results. The prediction steps are mainly divided into the following three points: First, the epidemic data is analyzed by the SEIR model, and the predicted value f is obtained by combining relevant research and parameter optimization; Secondly, ARIMA model prediction was carried out, and the historical epidemic data was used as the input value to test whether the time series data were stationary, and an augmented dickey-fuller was used to verify the hypothesis. If it is not stationary, the method of unary difference and binary difference is adopted until the data is stationary. The values of p and q were obtained by calculating ACF and PACF, and the ARIMA model was established. Training the sample set, optimizing the model, and getting the predicted value; Finally, the predicted values f and e of the above two models were taken as the input values, and the real data were taken as the target values. Linear regression modeling was used to calculate the predicted values of the SEIR-ARIMA hybrid model.

The calculation results showed that the model was relatively reliable for the analysis of the development trend of COVID-19. In order to truly reflect the effectiveness of the experiment, Dong Zhanggong et al. also implemented the SEIR-Logistic mixed model and the SEIR-LSTM mixed model and compared them with SEIR-ARIMA, concluding that SEIR-ARIMA prediction has achieved better prediction results. Therefore, based on the mixture of SEIR-ARIMA, it is conducive to the scientific decision-making of the country in the face of the epidemic and has a good application value for the prevention of other types of infectious diseases in our country in the future.

3. Conclusion and future prospects

The spread of COVID-19 is a very complex process, and the interweaving of human and natural factors greatly increases the complexity of modeling and simulation prediction. Therefore, based on a comprehensive understanding of the characteristics and important influencing factors of disease transmission, we need to put forward reasonable hypotheses and construct scientific mathematical models and verification methods so as to effectively guide epidemic prevention and treatment.

3.1 Conclusion

From the existing research results, most of the mathematical models of COVID-19 transmission are based on the SEIR model. The main research focuses on the prediction of inflection points and peaks, the dynamics of epidemic transmission, and the impact of epidemic prevention and control measures. Such models are easy to analyze and apply. If the underlying assumptions and parameter estimates are not very sound, there may be problems both early and late in the development of an infectious disease, and it does not reveal the impact of human individual interactions on disease transmission. At the same time, with the rapid development of computers today, it is more practical to make full use of computing and data capabilities, such as considering micro-scale models of bacteria. However, such models have extremely high requirements for data, and the calculation process is relatively complex. In addition, linking classical infectious disease dynamics models with artificial intelligence or statistical methods will make it easier for us to analyze and predict the epidemic.

3.2 Vision for the future

On the one hand, it is worthwhile to combine the actual epidemic data to better estimate the parameters of the transmission model. Only by fully exploring the inherent patterns of the epidemic development data and effectively controlling the identifiability and moderate sensitivity of the model parameters can we make scientific and reasonable epidemic development predictions to enable humans to better combat the disease.

On the other hand, due to the complexity of the spread of the epidemic, it is difficult to analyze every factor affecting the development of the new coronary pneumonia epidemic. In future research, it is necessary to make full use of the latest developments in computer science and intelligent technology to combine multiple models, combining classical mathematical models of infectious diseases with emerging models such as neural network dynamics, in order to improve the adaptive nature of the models and the scientific rationality of the simulation predictions.

We believe that by establishing accurate and reliable mathematical models, we can correctly explore the epidemic law of the disease so as to formulate scientific, accurate, and efficient prevention and control measures and clear the haze of the COVID-19 epidemic at an early date.

Acknowledgments

The author would like to thank all those who assisted in this work, as well as the reviewers for their valuable comments.

Conflict of interest

The authors declare that there is no personal or organizational conflict of interests with this work.

References

- [1] Rudrapal M, Khairnar SJ, Borse LB, Jadhav AG. Coronavirus disease-2019 (COVID-19): an updated review. *Drug Research (Stuttg)*. 2020; 70(9): 389-400.
- [2] Xiao Y, Qian K, Luo Y, Chen S, Lu M, Wang G, et al. Severe acute respiratory syndrome coronavirus 2 infection in renal failure patients: a potential covert source of infection. *European Urology*. 2020; 78(2): 298-299.
- [3] Diallo O, Kone Y, Sanogo C, Pousin J. A mathematical model of COVID-19: Analysis and identification of parameters for better decision making. *Applied Mathematics*. 2022; 13(02): 205-214.
- [4] Zhang L, Ma R, Han W, Xu J. Analysis of new coronavirus pneumonia in Harbin and suggestions for resumption of work under the epidemic. *Journal of Chongqing Normal University (Natural Science)*. 2020; 37(02): 107-111.
- [5] Yang Y, Sun Q, Wang Y, Yan X, Le T. Analysis and trend prediction of new coronavirus pneumonia (NCP) in Chongqing city. *Journal of Chongqing Normal University (Natural Science)*. 2020; 37(01): 135-140.

- [6] Zhang Y, Li J. Prediction and analysis of propagation of novel coronavirus pneumonia epidemic based on SIR model. *Journal of Anhui University of Technology (Natural Science)*. 2020; 37(01): 94-101.
- [7] Salman AM, Ahmed I, Mohd MH, Jamiluddin MS, Dheyab MA. Scenario analysis of COVID-19 transmission dynamics in Malaysia with the possibility of reinfection and limited medical resources scenarios. *Computers in Biology and Medicine*. 2021; 133: 104372.
- [8] Yin N. A simulation of COVID-19 epidemic propagation in limited area based on SIR model. *Statistics and Decision*. 2020; 36(5): 15-20.
- [9] Wan S, Liu Y, Liu M. Progress on the basic reproduction number of SARS-CoV-2. *Chinese Science Bulletin*. 2020; 65(22): 2334-2341.
- [10] Chen L, Wang K. Analysis of the characteristics of the COVID-19 in Shanxi province based on SEIR model. *Journal of Shanxi Datong University (Natural Science Edition)*. 2021; 37(06): 40-45.
- [11] Shi C, Yang L, Qian J, Luo L, Ou C. Quantitative evaluation of the public health interventions of COVID-19 epidemic in Guangzhou based on LSEIR model. *Chinese Journal of Health Statistics*. 2021; 38(06): 833-836.
- [12] Geng H, Xu A, Wang X, Zhang Y, Yin X, Ma M, et al. Analysis of the role of current prevention and control measures in the epidemic of corona virus disease 2019 based on SEIR model. *Journal of Jinan University (Natural Science Edition)*. 2020; 41(02): 175-180.
- [13] Cai J, Jia H, Wang K. Prediction of development trend of COVID-19 in Wuhan based on SEIR model. *Shandong Medical Journal*. 2020; 60(06): 1-4.
- [14] Cao S, Feng P, Shi P. Study on the epidemic development of COVID-19 in Hubei province by a modified SEIR model. *Journal of Zhejiang University (Medical Sciences)*. 2020; 49(02): 178-184.
- [15] Su Z, Wan L. Prevention and control evaluation of the COVID-19 epidemic based on the SEIR model - A Case of Harbin City. *Natural Science Journal of Harbin Normal University*. 2022; 38(01): 85-94.
- [16] Chen T, Zhao Z, Rui J, Yu S, Zhu Y, Xu J, et al. Estimating the transmissibility of coronavirus disease 2019 and assessing the effectiveness of the countermeasures to control the disease in Xiamen City. *Journal of Xiamen University (Natural Science)*. 2020; 59(03): 298-303.
- [17] Ferjouchia H, Kouidere A, Zakary O, Rachik M. Optimal control strategy of COVID-19 spread in Morocco using SEIRD model. *Moroccan Journal of Pure and Applied Analysis*. 2020; 7(1): 66-79.
- [18] Aliyeva T, Rzayeva U, Azizova R. A SEIRD model for control of COVID-19: case of Azerbaijan. *SHS Web of Conferences*. 2021; 92(7): 01001.
- [19] Hu T, Gao N, Zhu H, Wang J, Hu C. Research on the spread of infectious diseases in closed systems based on SIR and SEIRD models. *Frontiers in Medical Science Research*. 2021; 3(1): 40-44.
- [20] Viguier A, Lorenzo G, Auricchio F, Baroli D, Hughes T, Patton A, et al. Simulating the spread of COVID-19 via a spatially-resolved susceptible-exposed-infected-recovered-deceased (SEIRD) model with heterogeneous diffusion. *Applied Mathematics Letters*. 2021; 111: 106617.
- [21] Wang J, Dong Y. Simulation Research on COVID-19 Transmission and control measures based on SEIRD model. *Journal of System Simulation*. 2022; 111(1): 1532-1546.
- [22] Sang M, Ding Y, Bao M, Fang Y, Lu B. A transmission dynamics model based on the characteristics of SARS-CoV-2 and preventive and control measures. *Systems Engineering-Theory & Practice*. 2021; 41(1): 124-133.
- [23] Tang S, Tang B, Bragazzi NL, Xia F, Li T, He S, et al. Analysis of COVID-19 epidemic traced data and stochastic discrete transmission dynamic model (in Chinese). *Scientia Sinica Mathematica*. 2020; 50: 1-16.
- [24] Wang X, Tang S, Chen Y, Feng X, Xiao Y, Xu Z. When will be the resumption of work in Wuhan and its surrounding areas during COVID-19 epidemic? A data-driven network modelig analysis (in Chinese). *Scientia Sinica Mathematica*. 2020; 50: 969-978.
- [25] Tang B, Bragazzi NL, Li Q, Tang S, Xiao Y, Wu J. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). *Infectious Disease Modelling*. 2020; 5: 248-255.
- [26] Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *Journal of Clinical Medicine*. 2020; 9(2): 462.
- [27] Zhang L, Li D, Ren J. Analysis of COVID-19 by discrete multi-stage dynamics system with time delay. *Geomatics and Information Science of Wuhan University*. 2020; 45(05): 658-666.
- [28] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020; 395(10223): 497-506.
- [29] Carolina L, Patrick W, Jon K, Tiago BRC, Julio S, Maria S, et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature*. 2020; 584: 463-469.
- [30] Zhang B, Zhou X, Qiu Y, Song Y, Feng F, Feng J, et al. Clinical characteristics of 82 cases of death from

COVID-19. *PloS One*. 2020; 15(7): 2-8.

- [31] Winkler ES, Bailey AL, Kafai NM, Nair S, McCune BT, Yu J, et al. SARS-CoV-2 infection of human ACE2-transgenic mice causes severe lung inflammation and impaired function. *Nature Immunology*. 2020; 21: 1327-1335.
- [32] Zheng J, Wong L, Li K, Verma AK, Ortiz ME, Wohlford-Lenane C, et al. COVID-19 treatments and pathogenesis including anosmia in K18-hACE2 mice. *Nature*. 2021; 589(7843): 603-607.
- [33] Zhou Y, Huang M, Jiang Y, Zou X. Data-Driven mathematical modeling and dynamical analysis for SARS-CoV-2 coinfection with bacteria. *International Journal of Bifurcation and Chaos*. 2021; 31(11): 1-21.
- [34] Yang S, Chen W. Study on the control and economic inhibition effect of short-term public response Plan for sudden epidemic - A system dynamics simulation analysis based on COVID-19. *Soft Science*. 2020; 34(11): 130-136.
- [35] Liu H, Jia H, Wang Y, Liu B, Yao J, Yan X. Estimation method and evaluation of system dynamics model for the Initial spread of COVID-19 - A Case Study of Gansu Province. *Journal of University of Electronic Science and Technology of China (Social Sciences Edition)*. 2020; 22(03): 36-45.
- [36] Wang Y, Wu J, Liu H, Xu X, Liu F. Modeling and analysis of novel coronavirus pneumonia transmission scale growth: Research based on Italian epidemic data. *Northwest Population Journal*. 2020; 41(04): 114-126.
- [37] Bai L, Guo P, Fan J. Modeling and prediction of the number of confirmed cases of new coronavirus pneumonia in Hubei province. *Quality Safety Inspection and Testing*. 2020; 30(2): 10-12.
- [38] Huang H, Chen J, Xie Z, Tang J. Dynamic simulation study on the evolution of COVID-19 epidemic. *China Storage & Transport*. 2020; 6: 136-138.
- [39] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *The New England Journal of Medicine*. 2020; 382(13): 1199-1207.
- [40] Peirlinck M, Linka K, Costabal FS, Kuhl E. Outbreak dynamics of COVID-19 in China and the united states. *Biomechanics and Modeling in Mechanobiology*. 2020; 4: 2179-2193.
- [41] Yi Y. Simulation of SARS-CoV-2 based on system dynamics model: A case study of Beijing. *Science and Technology & Innovation*. 2022; 6: 113-115.
- [42] Lei S, Jiang F, Su W, Chen C, Chen J, Mei W, et al. Clinical characteristics and outcomes of patients undergoing surgeries during the incubation period of COVID-19infection. *EClinical Medicine*. 2020; 21: 100331.
- [43] Bao X, Tan Z, Bao B, Xu C. Prediction model of COVID-19 based on spatiotemporal attention mechanism. *Journal of Beijing University of Aeronautics and Astronautics*. 2022; 48(8): 1495-1504.
- [44] Gan Y, Wu Y, Wang J. Epidemics trend prediction model of COVID-19. *CAAI Transactions on Intelligent Systems*. 2021; 16(3): 528-536.
- [45] Dong Z, Song B, Meng Y. Prediction of COVID-19 based on mixed SEIR-ARIMA Model. *Computer and Modernization*. 2022; 2: 1-6.