



ELSEVIER



COMPUTATIONAL  
AND STRUCTURAL  
BIOTECHNOLOGY  
JOURNAL

journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)

# K-SEIR-Sim: A simple customized software for simulating the spread of infectious diseases

Hongzhi Wang<sup>a,1</sup>, Zhiying Miao<sup>b,1</sup>, Chaobao Zhang<sup>c,1</sup>, Xiaona Wei<sup>a</sup>, Xiangqi Li<sup>d,\*</sup>

<sup>a</sup>Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai 200062, China

<sup>b</sup>School of Optoelectronics and Information Technology, University of Shanghai for Science and Technology, Shanghai, 200090, China

<sup>c</sup>State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai 200031, China

<sup>d</sup>Department of Endocrinology, Shanghai Gongli Hospital, The Second Military Medical University, Shanghai 200135, China

## ARTICLE INFO

### Article history:

Received 9 January 2021

Received in revised form 1 April 2021

Accepted 2 April 2021

Available online 07 April 2021

### Keywords:

Software

Artificial intelligence

Python

COVID-19

SEIR model

Simulation analysis

2019-nCoV

## ABSTRACT

Infectious disease is a great enemy of humankind. The ravages of COVID-19 are leading to profound crises across the world. There is an urgent requirement for analyzing the current pandemic situation, predicting trends over time, and assessing the effectiveness of containment measures. Thus, numerous statistical models, primarily based on the susceptible–exposed–infected–recovered or removed (SEIR) model, have been established. However, these models are highly technical, which are difficult for the public and governing bodies to understand and use. To address this issue, we developed a simple operating software based on our improved K-SEIR model termed as the kernelkernel SEIR simulator (K-SEIR-Sim). This software includes natural propagation parameters, containment measure parameters, and certain characteristic parameters that can deduce the effects of natural propagation and containment measures. Further, the applicability of the proposed software was demonstrated using the example of the COVID-19 outbreak in the United States and the city of Wuhan, China. Operating results verified the potency of the proposed software in evaluating the epidemic situation and human intervention during COVID-19. Importantly, the software can perform real-time, backward-looking, and forward-looking analysis by functioning in data-driven and model-driven ways. All of them have considerable practical values in their applications according to the actual needs of personal use. Conclusively, K-SEIR-Sim is the first simple customized operating software that is highly valuable for the global fight against COVID-19 and other infectious diseases.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Infectious diseases are the common adversaries of humanity and can cause significant social crises [1]. The new infectious disease coronavirus 2019 (COVID-19) has become a pandemic in a short time span [2]. It has affected and continues to widely and profoundly affect the world. Factories and schools have closed, and public transport and travel have been stringently restricted. Until March 19, 2021, a total of 121,464,666 cases and 2,684,093 deaths have been confirmed worldwide (3:44 pm CET, 19 March 2021). Therefore, there is an urgent need to determine ways to analyze the epidemic features, infer key processes, and evaluate strategies for mitigating the impact of this pandemic. Statistical

models are often used to perform such tasks. Numerous attempts have been made to model the spread of COVID-19 in many countries or areas to help scientists, governments, and public fight the crisis [3–7]. For this purpose, the susceptible–exposed–infected–recovered or removal (SEIR) models have been established [4].

The susceptible–infections–recovered (SIR) model was established in 1927, and the SEIR model is derived from the SIR model and additionally includes the incubation period and exposed populations [8]. SEIR has become a classic model in research related to infectious diseases and is also a base for establishing other models. After the outbreak of COVID-19 till March 24, 2020, 31 models have been developed [9]. As of July 1, 2020, more than 40 investigations have been published based on SEIR and modified SEIR models (PubMed and Embase). These model-based researchers primarily wrote their own algorithms to obtain the corresponding experimental results by setting different parameters [10]. They subsequently evaluated the epidemic features and containment

\* Corresponding author.

E-mail address: [lixq@sibs.ac.cn](mailto:lixq@sibs.ac.cn) (X. Li).

<sup>1</sup> These authors contributed equally.

measures and presented suggestions and warnings [11]. However, each adjustment requires professional programming, and flexibly adjusting the parameters in a simulation process is challenging. Moreover, it is nearly impossible to perform real-time regression fitting of parameters for existing data. As these models present data using mathematical language, they are difficult to understand and use for scientists of other academic disciplines as well as general policy makers and staff. None of the abovementioned models can be recommended for clinical use [12].

To address this issue, we developed a kernel SEIR simulator (K-SEIR-Sim), which is a simple customized operating software based on our improved SEIR model. The software functions by setting different parameters. The user only needs to open the software and input the basic parameters to instantly obtain data of the target area regarding the current epidemic situation, trends over time, and effectiveness of control measures and medical treatment, among others. It is important to have such a model that is easily usable, because COVID-19 can bring political, economic, social, environmental, managerial, psychological and other impacts on people [13]. For example, a study documented its large disruptions to physical activity, sleep, time use, and mental health of young adults [14]. Using this model to assess the epidemic features of a city or a region, a psychiatrist can manage to improve mental well-being, a stadium owner can arrange physical activities rationally, an online educator can schedule study time, and so on. Even someone may look at economic impacts of COVID-19 by an estimate of the number of cases. To help users efficiently use this software, we demonstrated the applicability of K-SEIR-Sim using examples of the COVID-19 outbreak in the United States, Wuhan city of China, Diamond Princess, and France. The operating results of the proposed software were found to be excellent.

To the best of our knowledge, this is the first ever introduced simple and customized software based on the widely used SEIR model developed for use by both professional scientists and non-professional ordinary people. Using this software, we can obtain data on the transmission dynamics of infectious diseases, analyze the role of medical interventions, and evaluate the effectiveness of public policy implementations using data available in the public domain. Thus, the proposed software may greatly help in the fight against COVID-19 and other infectious diseases.

## 2. Methods

### 2.1. Developing the K-SEIR model

First, the K-SEIR model divides the population into the following six categories. (1) Susceptible (S): those who are susceptible to the infection. The outbreak has infected people of all ages; thus, all people are generally susceptible. Therefore, in this software, S represents the entire population in a certain area. S has an average daily infection rate of  $\lambda$ . (2) Exposed (E): Asymptomatic carrier of the virus. These individuals are in the incubation period after virus infection. The incubation period increases the difficulty of epidemic prediction and control. We know that the infection rate of the exposed population is typically the same as that of the infected population; thus, the model proposes the same value of  $\lambda$  for E as S. The probability of a carrier turning into an infected case per day is  $\sigma$ , which follows a normal distribution with the average incubation day. (3) Infectious (I): those who showed clinical symptoms after being infected with the virus. Currently, no difference is reported between the infection rates of mild and severe cases; hence, this model does not distinguish between the infection rate of mild and severe patients. Both the exposed (E) and infected (I) can infect S. (4) Removed (R): those who were removed from the infected population. The average daily removal rate for infected patients

is  $\gamma$ , which is divided into two parts: cure rate  $\alpha$  and mortality rate  $\beta$ . Thus,  $\alpha + \beta = \gamma$ . The removal number in the classical SEIR model is decomposed into the self-healing number H (healer) and death number D (death). (5) Healer (H): those who have been healed after infection. Although there were reports of reinfection [15,16], there is no definite conclusion whether COVID-19 patients can be reinfected after recovery [17], to generalize the model, patients who recovered from the illness were set to have an average daily probability  $\mu$  of reinfection. As long as there is reinfection, if the reinfection rate is obtained from an epidemiological survey, the coefficient of reinfection can be set to this value when the software is used; If the healer produces antibodies and is no longer infected, the value of  $\mu$  is 0. (6) Death (D): those who died after infection. This category of population must be properly handled to avoid counting them among people who could infect others.

Hereafter, we set the initial conditions for the K-SEIR model. The number of the population at time  $t$  is  $s, e, i, r, h$ , and  $d$ . At time  $t = 0$ , the corresponding parameters are denoted as  $s_0, e_0, i_0$ , and  $r_0$ . Then,  $s_0 + e_0 + i_0 + r_0 = N$ , where  $N$  is the number of total population in a certain region. The parameter  $K$  was added for human intervention, because it is unrealistic to expect the outbreak of the epidemic to entirely follow natural transmission laws; hence, human intervention is inevitable. First, the government and public often reduce interpersonal communication or increase social distancing owing to executive orders or spontaneous actions, or they can wear masks to prevent the transmission of the virus. These measures can reduce the daily natural infection rate ( $\lambda$ ) by reducing the interpersonal contact rate affected by physical isolation measure ( $k_1$ ). Second, the government can improve the removal rate by expediting the medical treatment. Notably, the removal rate will only accelerate the end of the epidemic, not the increase of cure rate. However, the removal rate will not help in improving the survival rate of the population. Therefore, we divided medical treatment into two parameters: hospital admission capacity ( $k_2$ ) and the ability of medical treatment ( $k_3$ ). Human intervention parameters can be calculated using fitting parameters as follows:

The actual infection rate  $\lambda_k = \lambda \times (1 - k_1)$ , hence,  $k_1 = 1 - \lambda_k/\lambda$  (equation 1, Eq. 1); the actual removal rate  $\gamma_k = \gamma \times k_2$ , hence,  $k_2 = \gamma_k/\gamma$  (equation 2, Eq. 2); and the actual cure rate  $\alpha_k = \alpha \times k_3$ , hence,  $k_3 = \alpha_k/\alpha$  (equation 3, Eq. 3).

### 2.2. Developing the software K-SEIR-Sim

The theoretical K-SEIR model was transformed into a simple software, implemented in PYTHON language, using software engineering. In particular, the tasks of graphical user interface design, control logic, operation logic, precision control, speed control, data visualization, data import and export, parameter fitting, key data display, and other specific contents were completed. Through constant modification and debugging, we developed a stable version of the software: K-SEIR Simulator V2.5. The software and instructions are available online ([http://peiyun.cn/download/seir\\_sim.files/SEIR\\_sim\\_V2.53.exe](http://peiyun.cn/download/seir_sim.files/SEIR_sim_V2.53.exe)) for free download for noncommercial use only.

### 2.3. Practical applications of software K-SEIR-Sim for fighting against COVID-19

#### 2.3.1. Data-driven analysis for the outbreak of COVID-19 in the United States

Data-driven analysis for infectious diseases is a crucial application of the proposed software. First, a matching generalization of the fitting analysis was performed. We collected sample data from March 1 to April 30, 2020 from the official epidemic statistics in the

United States. Thereafter, the corresponding fitting operation was conducted using the collected data. The parameter  $N$  indicates the total population, which is approximately 330 million in the United States. The susceptible population equals the total population minus the nonsusceptible population. The nonsusceptible population corresponds to the total population times the sweep rate, and the sweep rate is generally 5% [18]. Based on the theoretical simulation,  $E$  is 3 for 1 case in the early stage of the epidemic (see Supplementary data S1). Therefore, when the diagnosed number  $I$  on March 1, 2020 was 69, the exposed number was set as 207, i.e.,  $69 \times 3 = 207$ . The parameter of infectious number is  $I$ . The parameter for healer number  $H$  is set as 1, and the parameter for died number  $D$  is also set as 1. After importing the collected sample data from March 1 to April 30, 2020, the average incubation period was manually set, which is generally 4 to 7 days according to the current available data for this epidemic [19]. Herein, the incubation period is set as 5 days, and the function of parameter scanning is used to automatically fit the optimal parameter values of  $\lambda$ ,  $\sigma$ ,  $\gamma$ ,  $\alpha$ , and  $\beta$ . The antibody failure rate  $\mu$  is set as 0. Next, the development of the epidemic is deduced according to these parameters.

In the next step, an error generalization of the fitting analysis was performed. Owing to the lack of early detection ability and detection data for mild patients in the United States, the data published in the early period can be significantly distorted. Therefore, the statistical data from March 1 to April 30, 2020 were intercepted for error analysis.

For the parameter generalization of the fitting analysis, first, the parameters must be obtained using the official data of the United States. As the actual epidemic data are related to human intervention measures, the fitting parameters should include changes in  $k_1$ ,  $k_2$ , and  $k_3$ . Generally,  $k_1$  increases linearly, indicating a gradual increase in social isolation [5];  $k_2$  remains nearly constant, indicating that the removal rate did not change;  $k_3$  increases significantly, indicating that the healing power gradually increased. The actual infection rate depends on the initial infection rate and isolation intensity  $k_1$ . The initial infection rate was calculated to be 0.502, which was obtained by dividing the basic reproduction number  $R_0$  by the infection days. We considered  $R_0 = 3.77$  [20,21] and infection days = incubation period + the time required for diagnosis and treatment, where the incubation period was 5 days and the time required for diagnosis and treatment was 2.5 days. Therefore, the initial infection rate was 0.502. The isolation intensity  $k_1$  can be calculated using the infection rate. Next, the reliability of these parameters should be confirmed using the official data of the United States. After obtaining the optimal parameters by fitting the historical data, a 10-day prediction deduction is conducted. By generalizing errors obtained using the fitting analysis of the 10-day data, both the rationality of the parameters and the reliability of the model are verified. During parameter fitting, the weightage of a parameter can be set based on the requirement of the error of fitting data. When performing fitting analysis, we can sacrifice the current consistency in the pursuit of long-term consistency, which is particularly useful when the actual data do not evidently match the epidemic rule owing to the lack of updated data over a log period. Moreover, as the values of open data are often less than that of the actual data, the weightage of positive and negative parameters can be adjusted to make the fitting data close to the published data. The fitting parameters of the epidemic situation include human intervention parameters.

we also

### 2.3.2. Model-driven analysis for the outbreak of COVID-19 in Wuhan, China

To obtain basic data on infectious diseases, model-driven analysis is required. On January 23, 2020, the number of infections was

830, the number of healers was 34, and the number of deaths was 25 (The data came from China’s National Health Commission). The latent day was set to 5, the removal day was set to 10, and  $R_0$  was set to 3.77. To reflect the effects of lockdown, self-isolation, wearing masks, and cabin hospitals to the maximal possible extent, we set  $k_1 = 0.80$  and  $k_2 = 1.4$ . Because no specific drugs are available, we set  $k_3 = 1.6$ . The total population was set to 100 million.

## 3. Results

### 3.1. K-SEIR model

The proposed K-SEIR model is an improved SEIR model that considers the influence of both natural propagation and human intervention. The parameter  $K$  of human intervention is further divided into three parameters:  $k_1$ ,  $k_2$ , and  $k_3$ . Simultaneously, we decomposed the  $R$  parameter into two parameters:  $H$  and  $D$ . Then, we established the K-SEIR model (Table 1, Fig. 1).

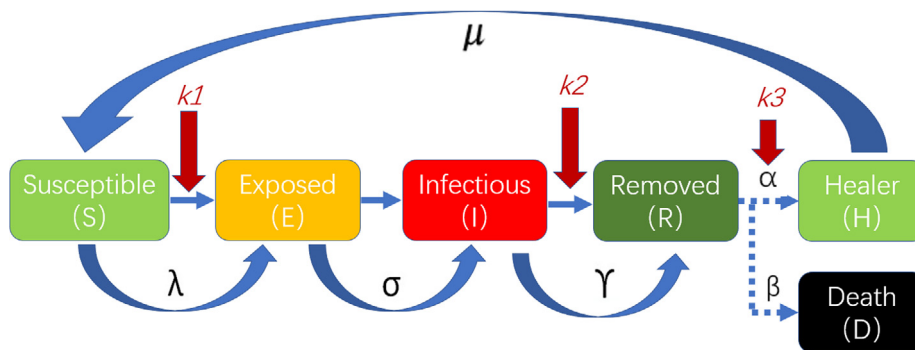
For a total population of 100,000, the parameter of infection rate ( $\lambda$ ) and removal rate ( $\gamma$ ) was simulated as 0.01 and 1, respectively. Fig. 2 displays the results of running the example program. Based on pure numerical and theoretical simulations, we can determine numerous features of this model. The total number of infected cases is very low or very high, and the intermediate state is very narrow. When the total number of infected cases is greater than 90% or less than 10%, the infection rapidly peaks and the epidemic will also rapidly end. When the number of infected cases is between 10% and 90%, the infection slowly peaks and the epidemic lasts for longer periods. To contain the spread of the epidemic, the infection rate should be minimized and the removal rate should be improved. However, the cost-benefit ratio should be considered for this purpose. The mortality rate is the primary factor in evaluating the cost-benefit ratio. For example, influenza has a very low mortality rate and can be allowed to develop regardless of its rate of infection and removal. The changes of  $\lambda$  and  $\gamma$  can also be observed under human intervention; hence, the model is still available under human intervention.

The above simulation analysis of the K-SEIR model suggests that many intervention measures can be employed to contain the spread

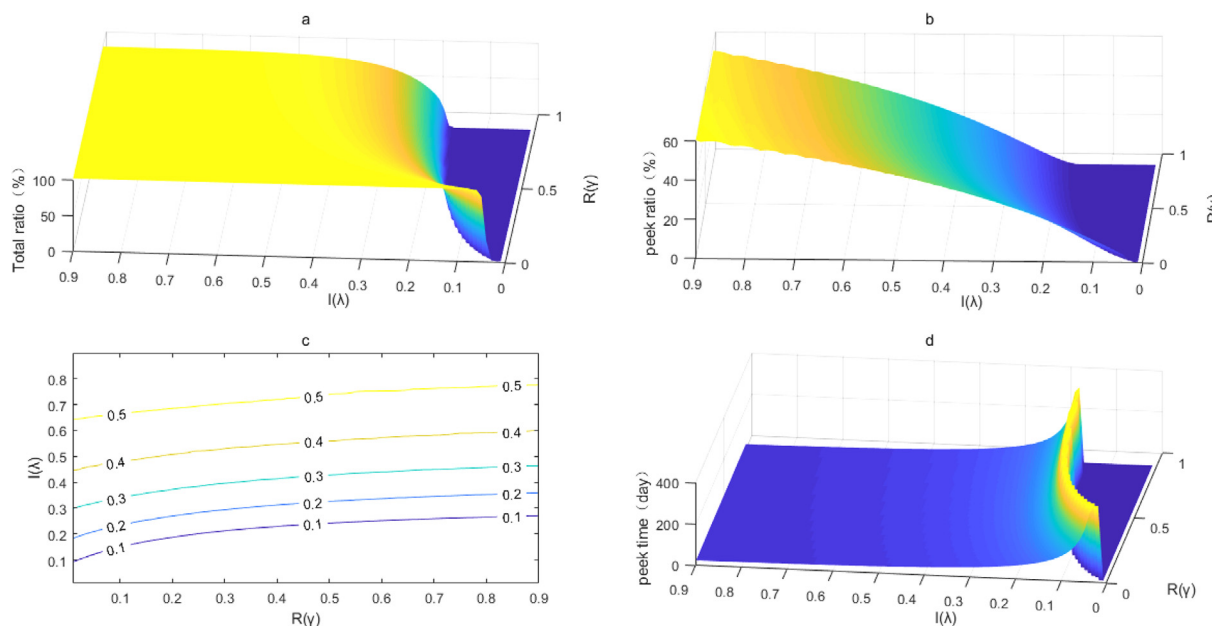
**Table 1**  
Basic information for K-SEIR model.

Population	Formula	Parameter
Susceptible (S)	$\frac{ds}{dt} = -\lambda si/N + \mu h$	$\lambda$ : average daily infection rate s: number of the population (S) at time t i: number of the population (I) at time t $\mu$ : average daily rate of reinfection h: number of the population (H) at time t N: number of total population in a certain region
Exposed (E)	$\frac{de}{dt} = \lambda si - \sigma e$	$\sigma$ : incidence rate per day e: number of the population (E) at time t
Infectious (I)	$\frac{di}{dt} = \sigma e - \gamma i$	$\gamma$ : average daily removal rate for infected patients
Removed (R)	$\frac{dr}{dt} = \gamma i$	It’s the sum of the cured and the dead
Healer (H)	$h = \alpha r$	r: number of the population (R) at time t $\alpha$ : average daily cure rate
Death (D)	$d = \beta r$ $\alpha + \beta = \gamma$ $s_0 + e_0 + i_0 + r_0 = N$ $\lambda_k = (1 - k_1) \lambda$ $\gamma_k = k_2 \gamma$ $\alpha_k = k_3 \alpha$	$\beta$ : average daily mortality rate 0: time $t = 0$ K: kernel for human intervention $k_1$ : physical isolation measure, the coefficient of $\lambda$ $k_2$ : hospital admission capacity, the coefficient of $\gamma$ $k_3$ : the ability of medical treatment, the coefficient of $\alpha$

Note:  $\lambda = R_0/AIP$ .  $R_0$  stands for basic reproductive number, and AIP stands for average infectious period.  $\sigma = 1/IP$ . IP stands for incubation period.



**Fig. 1.** Schematic of the logic structure of the K-SEIR-Sim model. The K-SEIR-Sim model consists of Susceptible (S), Exposed (E), Infectious (I), Removed (R), Healer (H), and Death (D). The parameter of human intervention measures is K, which includes three parameters:  $k_1$  is a measure of physical isolation;  $k_2$  is a measure of hospital admission capacity; and  $k_3$  is a measure of the ability of medical treatment .



**Fig. 2.** Epidemic features by simulation in the natural propagation state. a, Total number of infections. When the infection rate is greater than 0.3, almost all infections eventually occur in the natural state. When the infection rate is less than 0.3, greater removal rate corresponds to a smaller total number of infections. There is a steep climb from a small number of infections at the bottom to almost all infections, with a narrow parameter transition zone. b, Peak number of infected people. The higher the infection rate, the lower the removal rate and the larger the peak number of infected people (up to 60% of the total number of infected people). c, Contour relationship between the proportion of peak infections, infection rate ( $\lambda$ ), and removal rate ( $\gamma$ ). For the contour relationship between the proportion of peak infections of less than 10% occurs only in a very small parameter distribution. d, Peak time. For the peak time, when the number of infected people is high or low, the peak time is short. In areas where the number of confirmed cases was steep, the later the peak occurred and the longer the outbreak lasted, even more than 365 days.

of the epidemic. First, physical isolation measure ( $k_1$ ) can be reduced to reduce the daily natural infection rate ( $\lambda$ ). Second, the removal rate can be improved by increasing the availability of medical treatment. However, the increase in the removal rate and not the increase of the cure rate will accelerate the elimination of the epidemic, which will not help improve the survival rate of the population. Consequently, medical treatment is divided into two parameters,  $k_2$  (hospital admission capacity) and  $k_3$  (the ability of medical treatment), which is highly valuable in the application.

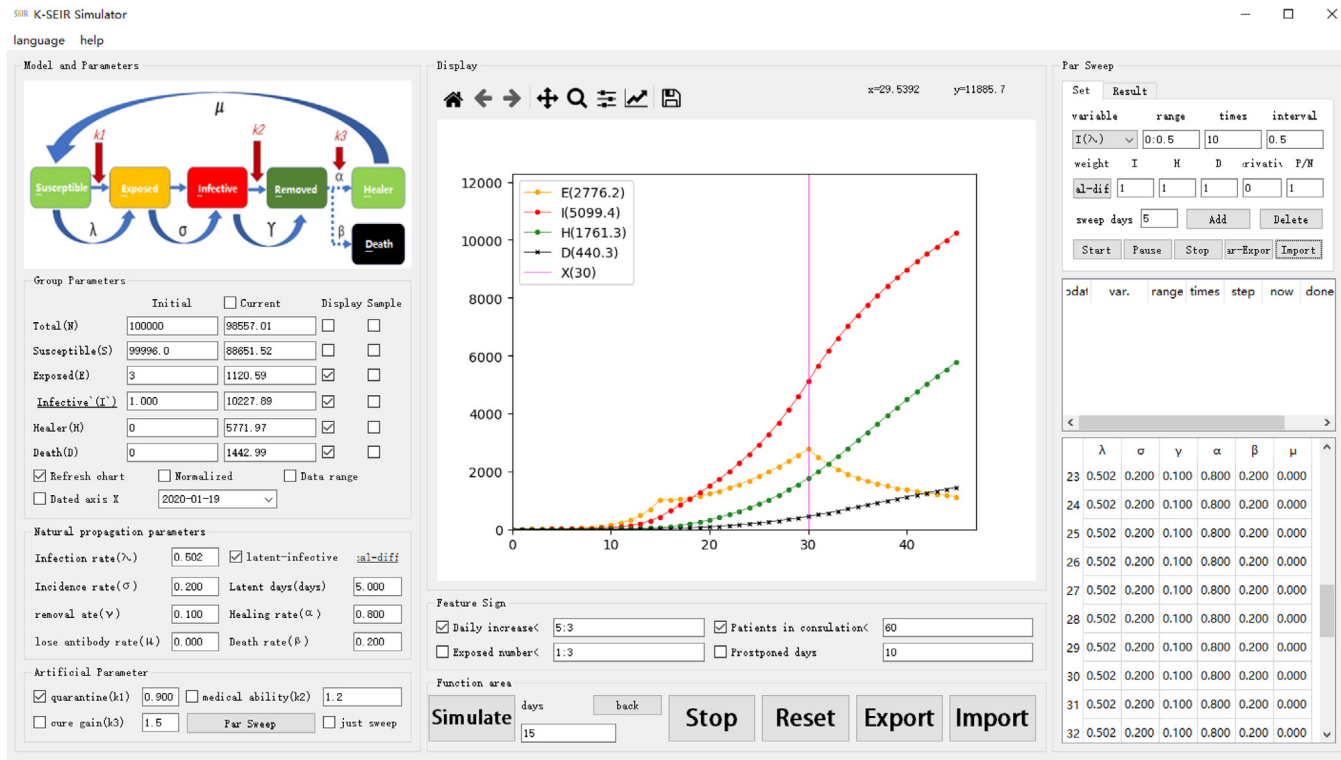
### 3.2. K-SEIR-Sim software

As the current models are based on mathematical programming languages, they are only suitable for professionals. Therefore, universal public welfare software (K-SEIR-Sim V2.53) was established to simulate the development of infectious diseases based on our improved K-SEIR model (Fig. 3).

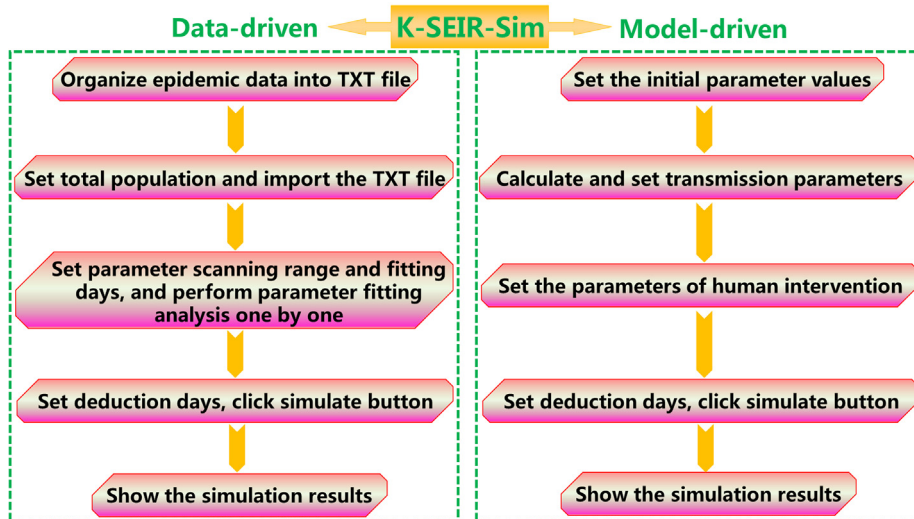
Researchers can manually set the parameter combinations on the software interface to obtain the results of simulation or prediction. They can also arbitrarily adjust the parameters during the running process, without the need for programming skills. Moreover, the software has an automatic parameter-scanning function so that after the actual epidemic data are imported (e.g., confirmed cases, cured cases, and deaths), the parameter-scanning function can be used to obtain the current optimal parameter combination, which makes later prediction. Thus, based on the users' demands, the software can function in two ways: model-driven and data-driven. It can also perform real-time, forward-looking, and backward-looking analysis addressing practical needs.

### 3.3. User's guide of K-SEIR-Sim software

K-SEIR-Sim software works in either data-driven or model-driven way. Both of them are composed of five operation steps (Fig. 4). The data-driven operation consists of the following five



**Fig. 3.** Schematic of the user interface on software startup. The user interface includes three core function modules, namely, parameter settings, simulation of epidemic situations, and parameter scanning. The software can work in three ways: model-driven, data-driven, or mix-driven, based on users' demands.



**Fig. 4.** User's guide of K-SEIR-Sim software. The data-driven and model-driven operation both consist of five step operations, but some operations are different.

steps: create epidemic data TXT file, import the TXT file, perform parameter fitting analysis, carry out the simulation, show the simulation results. The five-step operation of the model-driven way is: set the initial parameters, set the transmission parameters, set the human intervention parameters, carry out the simulation, show the simulation results. TXT file based on epidemic data for data-driven way covers susceptible population, latent infections, confirmed cases, deaths and starting date. For the model-driven way, the initial parameters are involved in total population, susceptible population, latent infections, confirmed cases, deaths and starting date; the transmission parameters based on epidemiological survey

data include infection rate, incidence, and removal rate; and the parameters of human intervention contain isolation intensity, hospital capacity, and healing benefits. The simulation results for both working modes show cumulative number and daily increase number of each population, peak date, remission date, end date and so on.

When the software is run for the first time by an inexperienced person, and when the detailed information is unavailable at the initial stage of an epidemic, how to use this software to simulate the spread? There are a few caveats he/she should pay attention to, please see the [Supplementary data S2](#). For common use, see the examples below.

### 3.4. Practical applications of software K-SEIR-Sim for fighting against COVID-19

#### 3.4.1. Data-driven analysis of the outbreak of COVID-19 in the United States

We performed a fitting analysis to confirm the reliability of the proposed software. During fitting analysis, we must first evaluate whether the simulated and actual data match (Fig. 5). We take the data from France as an example to perform the fitting analysis, and the results show that the software runs well and the simulation results are reliable (see the Supplementary data S3). Further, we collected data from the official epidemic statistical data in the United States from March 1 to April 30, 2020 to conduct the fitting analysis. From the result of this fitting analysis, we obtained the optimal parameters of natural transmission in each stage and the corresponding parameters of human intervention. Based on these parameters, we conducted the comparison between the deduction results and the official data by running the software. The comparison showed that the deduction data agree well with the official data for confirmed cases (I), cured cases (H), and deaths (D).

In the next step, we must examine the errors of the fitting analysis (Fig. 6). From March 1 to March 22, 2020, there were 12 days (accounting for about half of the days) in which the infection and death errors were within 10%, while a recovery error within 10% occurred only within 3 days. From March 23 to April 30 (when the government implemented the quarantine measures), 15 days (37.5%) reflected with an infection error within 4%, more than 40% with a recovery error within 4%, and 25 days (more than 50%) reflected a death error within 4% (Fig. 6). Overall, the accuracy of fitting was higher than 96%. The percentage of days with errors ranging from 4% to 10% were as follows: 50% (infection error), 30% (recovery error), and 37.5% (death error). In conclusion, when the model proposed in this paper is used to fit the existing data, the fitting accuracy can be retained at greater than 90%.

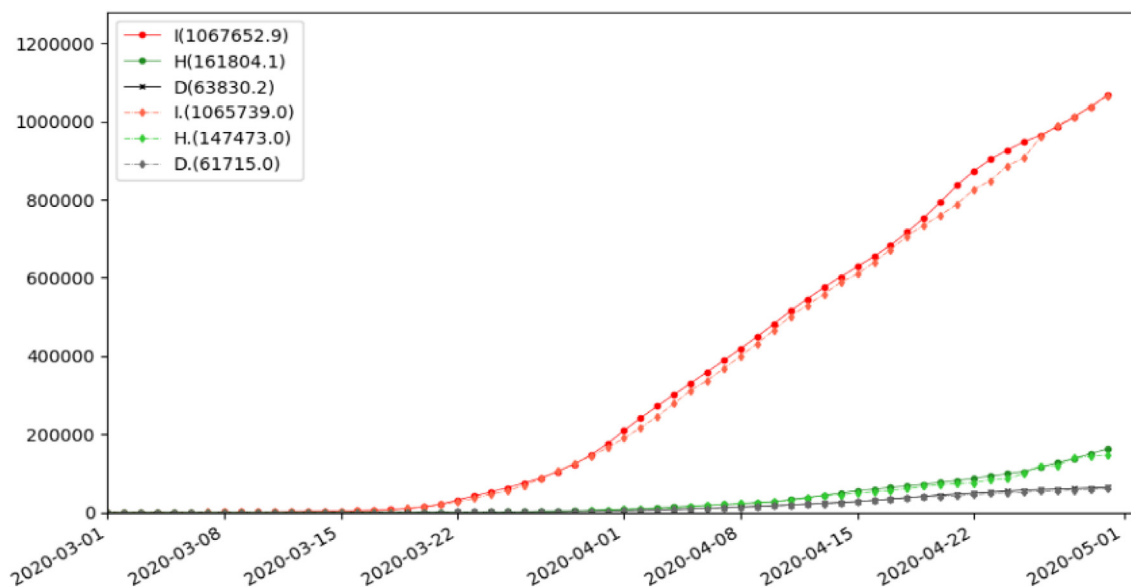
We must also generalize the parameters of the fitting analysis. We first obtain the parameters by running the software (Fig. 7). Fig. 6 shows the optimal transmission dynamics parameters of the outbreak in the United States at different periods starting from

March 1 to April 30, 2020. The parameter of the infection rate is decreasing, indicating that the home quarantine measures adopted by the government are effective and that people’s awareness of prevention and control is gradually improving. Because the government implemented prevention and control measures on March 23, the rate of  $k_1$  increased from 0 to 1 and reached 0.948 on April 30 (according to Eq1). This implies that the average person restricts 95% of their social intensity, showing that the home isolation measures adopted by the government and the public have reached a maximum state and the possibility of a further increase in  $k_1$  is insignificant. If the government resumed work and production, the peak may increase. Another parameter of the removal rate is negligible and relatively stable. Thus, the number of people removed from the confirmed number is very small, indicating that a large number of infected patients are still in the state of diagnosis. The effect of the parameter of medical treatment ( $k_2$ ) is not obvious (the actual removal rate = the initial removal rate  $\times k_2$  (Eq. 2); the removal rate is nearly the same; hence,  $k_2$  is nearly the same, because once hospital admission capacity reaches saturation, it can’t take anymore. However, the mortality rate is not high. The third parameter of the cure rate is increasing. As the actual cure rate is the initial cure rate  $\times k_3$  (Eq. 3), the initial cure rate is approximately 0.33 based on early data analysis. It can be inferred that as of April 30, the increase in the cure rate,  $k_3$ , is 2.23, indicating that the level of medical treatment and the success rate of cure have improved.

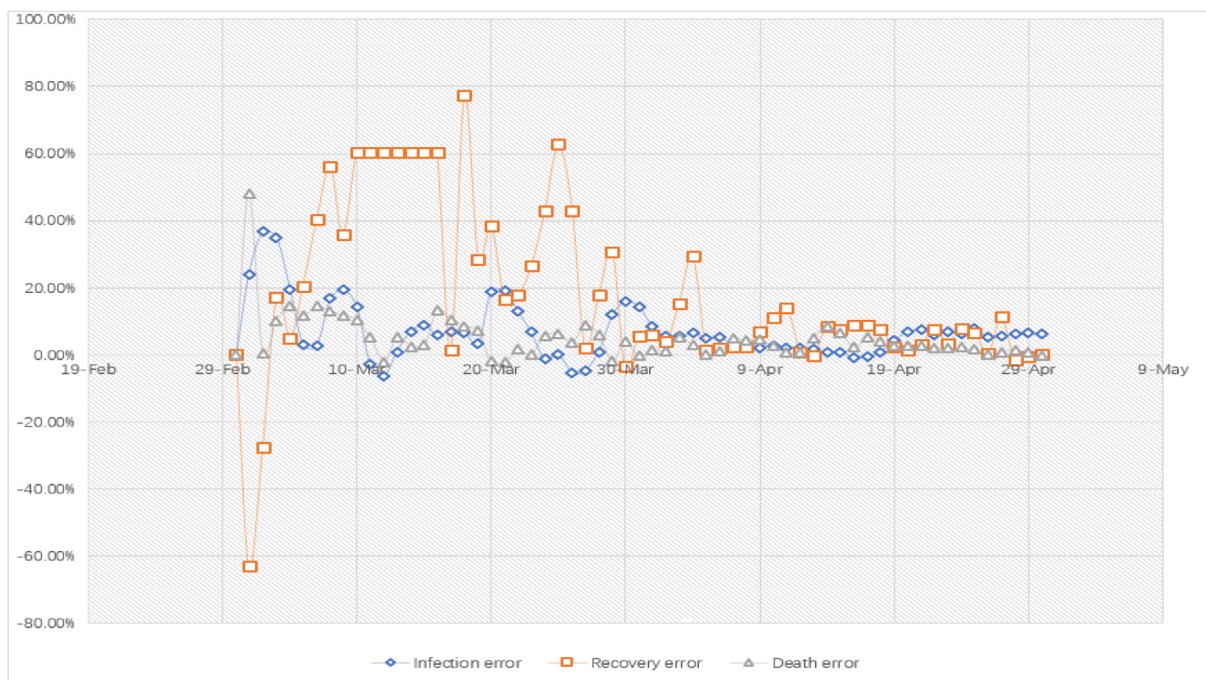
The reliability of the parameters must be verified by running the software (Table 2). Based on the evidence of the optimal parameters obtained by fitting historical data, a 10-day (May 1–10, 2020) prediction deduction analysis was conducted. By comparing the results of this analysis with the published data and from the corresponding error calculation, we can see that the results predicted by the model agree well with the published data and the corresponding error is mostly within 10%.

#### 3.4.2. Model-driven analysis for the outbreak of COVID-19 in Wuhan, China

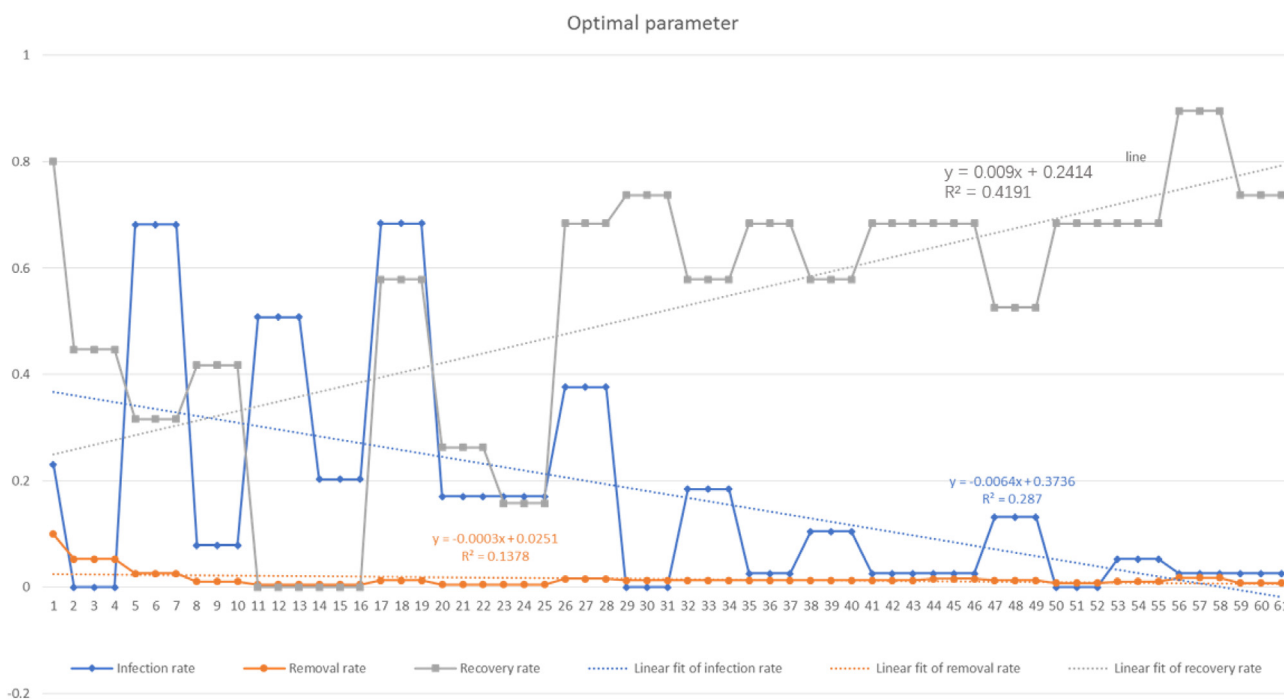
The model-driven analysis for infectious diseases is another crucial application of the proposed software. After entering the



**Fig. 5.** Comparison between deduction results and the official data by simulation analysis. The comparison showed that the deduction data agreed well with the official data, because the corresponding lines are basically overlapped. The abscissa shows time, and the ordinate is population. The above three lines in the icon of the upper left corner of the figure represents the deduction data, while the following three lines shows the official data. I, the number of confirmed cases (deduction); H, healer, i.e., the number of cured cases (deduction); and D, number of the dead (deduction). I., the number of confirmed cases (official); H., healer, i.e., the number of cured cases (official); and D., number of the dead (official).



**Fig. 6.** Error variations of matching between the fitting and official data from the United States from March 1 to April 30. From March 1 to March 22, the infection and death errors were within 10% for 12 days, whereas the recovery error was within 10% for only 3 days. From March 23 to April 30, the infection error was within 4% for 15 days, whereas the recovery and death errors were within 4% for 25 days. All of these fitting accuracies can be maintained at greater than 90%. Negative error indicates that the fitting value is less than the official data.



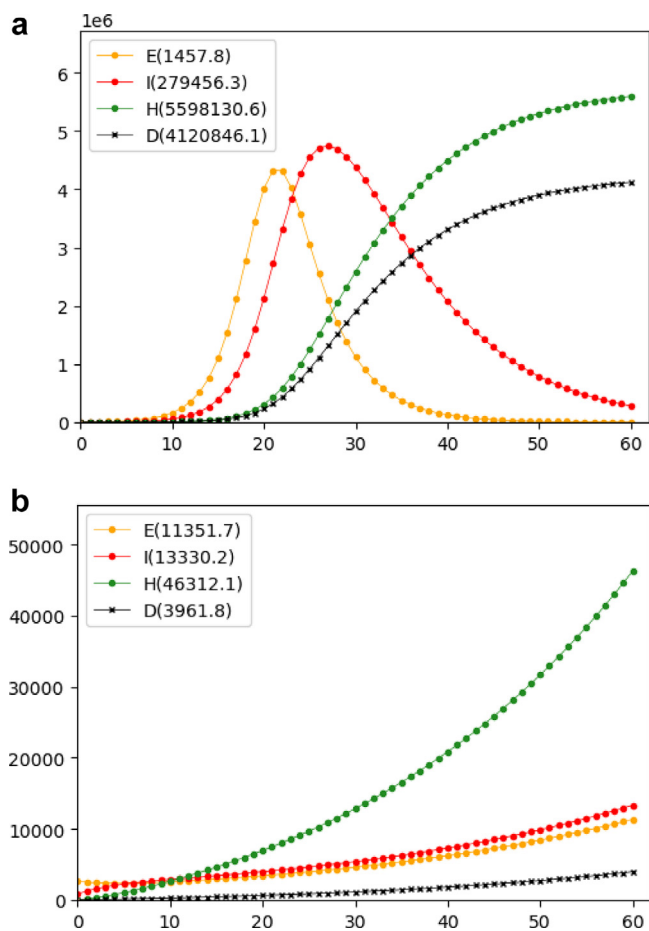
**Fig. 7.** Optimal propagation dynamic parameters in different periods. The infection rate is decreasing. On March 23, the rate of  $k_1$  increased from 0 to 1 and reached 0.948 on April 30. The removal rate is very low and relatively stable. The medical treatment ( $k_2$ ) effect is not obvious, and the mortality rate is not high. The cure rate increased from 0.33 on March 1 to 0.74 on April 30. The abscissa represents the days, counting from March 1.

calculated parameters into the software using the open data, we run the software and obtain the results. We take the data from Diamond Princess as an example to perform model-driven analysis, and the results show that the software runs well and the simulation results are reliable (see the [Supplementary data S4](#)). Further,

The parameters calculated using the data from Wuhan, China, on January 23, the day of the lockdown in Wuhan, are inputted to the software to predict the situation after 60 days ([Fig. 8](#)). There would be more than 5 million healers and more than 4 million deaths in the case of free transmission. However, the lockdown

**Table 2**  
Comparison of prediction results in 10 days (infectious/healer/death).

	5/1	5/2	5/3	5/4
Published data (I/H/D)	https://doi.org/11099275/156089/63972	1134059/164015/65886	1162383/175382/67505	1191854/180152/68702
Simulation data (I/H/D)	1160147/153332/63502	1187453/158894/65487	1214492/164572/67513	1241397/170364/69580
Error (I/H/D)	<b>5.54%</b>  - <b>1.77%</b>  - <b>0.73%</b>	<b>4.71%</b>  - <b>3.12%</b>  - <b>0.61%</b>	<b>4.48%</b>  - <b>6.16%</b> / <b>0.01%</b>	<b>4.16%</b>  - <b>5.43%</b> / <b>1.28%</b>
	5/5	5/6	5/7	5/8
Published data (I/H/D)	1214023/188069/69974	1239848/201152/72381	1263705/213109/74817	1293907/217251/76998
Simulation data (I/H/D)	1268276/176269/71687	1295216/182285/73834	1322288/188411/76020	1349548/194648/78246
Error (I/H/D)	<b>4.47%</b>  - <b>6.27%</b> / <b>2.45%</b>	<b>4.47%</b>  - <b>9.38%</b> / <b>2.01%</b>	<b>4.64%</b>  - <b>11.59%</b> / <b>1.61%</b>	<b>4.30%</b>  - <b>10.40%</b> / <b>1.62%</b>
	5/9	5/10		
Published data (I/H/D)	1324352/223930/78701	1349605/238081/80101		
Simulation data (I/H/D)	1377043/200996/80511	1404811/207456/82816		
Error (I/H/D)	<b>3.98%</b>  - <b>10.24%</b> / <b>2.30%</b>	<b>4.09%</b>  - <b>12.86%</b> / <b>3.39%</b>		



**Fig. 8.** Model-driven analysis for COVID-19 in Wuhan, China. The model-driven analysis for COVID-19 is performed by entering the calculated parameters into the software using the open-access data from Wuhan, China of January 23. The software inferred the situation after 60 days. In the case of free transmission, there would be more than 5 million healers and more than 4 million deaths, whereas the lockdown resulted in 3961.8 deaths and 46,312.1 healers. The abscissa shows time (days), and the ordinate is population.

only caused 3961.8 deaths and 46,312.1 healers, which is close to the corresponding numbers of 3814 and 43,214. These results fully demonstrate the effective precautions of isolation and control measures.

**4. Discussion**

To overcome the health and social crisis caused by infectious diseases such as COVID-19, many decision-support tools have been

proposed that are commonly underpinned by clinical mathematical models [9]. These models are classified into diagnostic and prognostic models [22]. Unfortunately, a recent review demonstrated the poor quality of these models, and none of them can be recommended for clinical use [9]. Methodological shortcomings and the unknown characteristics of new diseases in the urgent situations may be the root cause for this. Developing a clinical analysis model should be a science and an art [22]. To improve the quality of a prediction model, many scientists have introduced the factors of artificial intelligence technology [10,23], R packages [24], intervention measures [25], hospitalization, and demand of intensive care unit [26]. Applications have also been established to assist in this area, including an open electronic health record template to improve interoperability in China [27], a smartphone app for peer-to-peer contact tracing in the United States [28], and an interactive web-based app for complex reporting in the Czech Republic [29]. Herein, we developed generalized and easily operable software, which is important for advancing clinical applications. Notably, when we revised our manuscript, we found an online software named Epidemic Calculator (<http://gabgoh.github.io/COVID/index.html>), which is also developed based on SEIR model. It is very smooth in terms of parameter adjustment and result presentation. It is important to point out that the software we developed is different from the website software in three main aspects. First, our software adds three control parameters,  $k_1$ ,  $k_2$  and  $k_3$ , it can reflect the real situation more. Second, our software can conduct data-driven analysis. After importing existing actual data, we can get transmission parameters, which is more conducive to the analysis of the subsequent development of the epidemic. Third, our software can be downloaded for local use, not affected by network restrictions. Therefore, we think our software is very valuable for epidemiological analysis.

To more closely align with the actual needs, we added artificial parameters and some features to the software. When only natural transmission is simulated, artificial parameters need not be selected. When the transmission under human intervention is simulated, the artificial parameters can be adjusted according to the actual requirements. Moreover, all parameters can be adjusted at any time during the simulation operation process, which can simulate the previous situation and predict the future trend of the disease. The software can also perform real-time analysis and is suitable for COVID-19 and other infectious diseases. Therefore, this software can perform real-time, forward-looking, and backward-looking analysis for infectious diseases. It can work in two ways: data-driven and model-driven. The model-driven method refers to running the software to analyze the development of the disease by entering the values of the parameters, whereas data-driven method indicates running the software to obtain the parameters by entering the open data. We can select a method based on our needs, which complements the special practical value of the application.



In the data-driven method, the actual epidemic data over a certain time period were considered as inputs for the software to perform the fitting analysis with the aim of obtaining running parameters  $R_0$ , infection rate, removal rate, mortality rate, artificial parameters, and so on. The software can evaluate the effectiveness of human control and the situation of the epidemic. Although there may be some statistical-analysis related problems in the authenticity of actual short-term data, the law of epidemic spread can be obtained using comparatively long-term data. For example, in terms of the recovery error, when we performed error analysis based on the example of the data of the United States, we found that there was a 3-day margin of error of 10% from March 1 to March 22, 2020 (Fig. 6). The value of the recovery error was relatively large for remaining days, with some exceeding 40%. A review of official data showed that there were significant data gaps on these days, which were inconsistent with the actual spread of the virus. Considering the recovery number as an example, the healer number per day was 7 from March 2 to March 4, 9 from March 5 to March 8, and 12 from March 9 to March 16. Moreover, official data had not been updated for several days, which is the main reason for the large recovery error. However, infection and death errors remained within a relatively stable range. Further, the recovery errors on March 12 and March 18 increased abnormally because the number of recovery patients released was not updated from March 2 to March 4 and from March 17 to March 18; hence, the distortion of data resulted in the fluctuation of the error. From March 30, the data of the cure error stabilized within a reasonable range. Therefore, in the United States, data-driven analysis of the proposed model has a certain reference value for the simulation and prediction of the epidemic situation. It can also be used in other countries or areas.

The data-driven analysis refers to an attempt to match the actual data. However, the actual data are affected by many statistical aspects; for some days, no data are available. Thus, we also developed the function of model-driven analysis. The model-driven analysis is very simple to use and requires only entering the parameters into the software. Notably, when we simulated the situation 60 days after the lockdown of Wuhan, the number of healers was close to the actual reported number but the number of infections was approximately 3 times higher than the actual reported number (Fig. 8). This is because asymptomatic infections were not counted at that time. According to a large number of literature analyses, asymptomatic infections occupied approximately 10%–60% of the total infections. Thus, the number of our simulated infections is also relatively accurate. Additionally, despite the number of deaths was 2524 as per the official data on February 23, the government added 1290 to the number of deaths when the pandemic ended in Wuhan. Thus, the number of deaths was close to the actual reported number. Of course, when assuming that there is absolutely free transmission and absolutely nothing we can do to combat this disease, the numbers of infections and deaths are staggering. However, that possibility does not exist in current society. Despite this, our simulation analysis showed the usefulness of the lockdown in Wuhan and implemented control measures.

The aim is for this software to be widely used. Using our software, we conducted analysis of a single outbreak wave, simulating the development of the epidemic in different regions of different sizes, ranging from a country (330 million people in the US, Figs. 5–7), a city (10 million people in Wuhan of China, Fig. 8), to a small enclosed space (3711 people on Diamond Princess, see Supplemental data S4). We have obtained the simulation results which are in good agreement with the actual situations. In addition to the simulation of a single epidemic wave the above mentioned, we also used this software to simulate the development of multiple epidemic waves caused by lockdown and lifting of lockdown in France (76.07 million people in France, see Supplemental data S3). This

software simulates by default using day as a unit. France's application shows that it can simulate using week as a unit under the long duration of the epidemic. All we need to do is to change the original parameter values calculated by day to the parameter values calculated by week (equal to the parameter values calculated by day multiplied by 7) and carry out the same simulation operation. We hope it is a useful tool for being widely used.

Theoretically, this software can simulate other epidemics besides COVID-19. For an epidemic that does not require human intervention, we just don't set any K parameters. For epidemics without incubation period, setting incidence rate  $\sigma$  to 0 is sufficient. For epidemics that patients can produce antibodies but antibodies will eventually fail, we only need to set the antibody failure rate  $\mu$  which is obtained by statistics. In a word, our software is of great practical value.

It's also noteworthy that the SEIR model, in fact any model, has an assumption of homogeneity, that is, that people are equally likely to contract the disease and that people follow a homogeneous contact pattern. If homogeneity is not good in a region, it can affect the natural propagation parameters. In different homogeneous environments, the parameters should be different. It should be noted that, no matter how small the region is, the ideal homogenization does not exist, and heterogeneity exists objectively. But, no matter how heterogeneous it is, it will eventually result in statistically homogenized model parameters within a certain region. For example, the epidemiological spread happened in the USA in very heterogeneous geographic and time scale patterns. Although the regions of USA are very different, they are no more different than China and USA, so we can still think of USA as a system. Although the actual situation of USA is not homogeneous, but after a lot of statistical data, it is considered to be homogeneous. Importantly, the simulation results are close to the real situation, so we think that the software is still of practical value. Even so, we cannot ignore it in the interpretation and analysis of simulation results, avoiding being too idealistic.

## 5. Conclusion

There is still no end in sight to combat the COVID-19 crisis. Although many statistical models have been established to analyze the epidemic situation and containment measures, these models are technical and cannot be easily understood and used by management agencies and public. Herein, we developed a simple, customized software based on our improved SEIR model that can be used by both professional scientists and individuals, institutions and social organizations interested in epidemic analysis. This software can deduce both the result of natural propagation and containment effects. It can perform real-time, forward-looking, and backward-looking analysis, and it works in two ways: model-driven and data-driven, which have special practical applications. We demonstrated the applicability of this software using official data from the United States, Wuhan in China, Diamond Princess and France, which also proves the practical value of the software. Conclusively, we developed the first simple and customized operating software K-SEIR-Sim in python language, which can play an important role in the global fight against COVID-19 and other infectious diseases.

## Acknowledgments

Despite our best efforts, many excellent studies cannot be cited because of space constraints. We are grateful to all such scientists for their significant contribution to the containment of this new disease. This work was supported by grants from the Key Discipline Construction Project of Pudong Health and Family Planning

Commission of Shanghai (PWZxk2017-07), Pudong New Area Science and Technology Commission (PKJ2019-Y21), China Postdoctoral Science Foundation (2019M651615), and Shanghai Super-Postdoctoral Science Foundation (2018229).

### Declaration of Competing Interest

The authors declare that they have no conflict of interest.

### Author contributions

Xiangqi Li designed the study and wrote the manuscript. Hongzhi Wang developed the software. Miao Zhi Ying and Chaobao Zhang performed the analyses. Wei Xiao Na collected the data.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.04.004>.

### References

- [1] McCloskey B, Dar O, Zumla A, Heymann DL. Emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread. *Lancet Infect Dis* 2014;14(10):1001–10.
- [2] Remuzzi A, Remuzzi G. COVID-19 and Italy: what next?. *Lancet* 2020;395(10231):1225–8.
- [3] Aleta A, Moreno Y. Evaluation of the potential incidence of COVID-19 and effectiveness of containment measures in Spain: a data-driven approach. *BMC Med* 2020;18(1):157.
- [4] Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, et al. Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *PNAS* 2020;117(19):10484–91.
- [5] Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N. Centre for the mathematical modelling of infectious diseases C-WG, Jit M, Klepac P: The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* 2020;5(5):e261–70.
- [6] López L, Rodó X. The end of social confinement and COVID-19 re-emergence risk. *Nat Hum Behav* 2020;4(7):746–55.
- [7] Truelove S, Abraham O, Altare C, Lauer SA, Grantz KH, Azman AS, et al. The potential impact of COVID-19 in refugee camps in Bangladesh and beyond: a modeling study. *PLoS Med* 2020;17(6):e1003144.
- [8] Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics—II. The problem of endemicity.1932. *Bull Math Biol* 1991;53(1–2):57–87.
- [9] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020;369:m1328.
- [10] Zheng N, Du S, Wang J, Zhang He, Cui W, Kang Z, et al. Predicting COVID-19 in China using hybrid AI model. *IEEE Trans Cybern* 2020;50(7):2891–904.
- [11] Sanchez-Caballero S, Selles MA, Peydro MA, Perez-Bernabeu E. An efficient COVID-19 prediction model validated with the cases of China, Italy and Spain: total or partial lockdowns?. *J Clin Med* 2020;9(5).
- [12] Kuniya T. Prediction of the Epidemic Peak of Coronavirus Disease in Japan. *J Clin Med* 2020;9(3).
- [13] Green ST, Cladi L. COVID-19, politics, economics and how the future pans out are inseparable. *J R Soc Med* 2021;114(2):60–2.
- [14] Giuntella O, Hyde K, Saccardo S, Sadoff S. Lifestyle and mental health disruptions during COVID-19. *Proc he Natl Acad Sci USA* 2021;118(9).
- [15] Vetter P, Cordey S, Schibler M, Vieux L, Despres L, Laubscher F, et al. Clinical, virologic and immunologic features of a mild case of SARS-CoV-2 reinfection. *Clin Microbiol Infect* 2021. <https://doi.org/10.1016/j.cmi.2021.02.010>.
- [16] Adrielle Dos Santos L, Filho PGG, Silva AMF, Santos JVG, Santos DS, Aquino MM, de Jesus RM, Almeida MLD, da Silva JS, Altmann DM, et al. Recurrent COVID-19 including evidence of reinfection and enhanced severity in thirty Brazilian healthcare workers. *J Infect* 2021;82(3):399–406.
- [17] Lee JT, Hesse EM, Paulin HN, Datta D, Katz LS, Talwar A, Chang G, Galang RR, Harcourt JL, Tamin A, et al. Clinical and laboratory findings in patients with potential SARS-CoV-2 Reinfection, May–July 2020. *Clin Infect Dis* 2021.
- [18] Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Biol Sci* 2007;274(1609):599–604.
- [19] Guan W-J, Ni Z-y, Hu Yu, Liang W-H, Ou C-Q, He J-X, et al. Clinical characteristics of coronavirus disease 2019 in China. *New Engl J Med* 2020;382(18):1708–20.
- [20] 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. *New Engl J Med* 2020;382(13):1199–207.
- [21] Lin Y-F, Duan Q, Zhou Y, Yuan T, Li P, Fitzpatrick T, et al. Spread and Impact of COVID-19 in China: a systematic review and synthesis of predictions from transmission-dynamic models. *Front Med (Lausanne)* 2020;7:321.
- [22] Sperrin M, Grant SW, Peek N. Prediction models for diagnosis and prognosis in Covid-19. *BMJ* 2020;369:m1464.
- [23] Vaid S, Cakan C, Bhandari M. Using Machine Learning to Estimate Unobserved COVID-19 Infections in North America. *J Bone Joint Surg Am* 2020;102(13):e70.
- [24] Huang J, Cheng A, Lin Su, Zhu Y, Chen G. Individualized prediction nomograms for disease progression in mild COVID-19. *J Med Virol* 2020;92(10):2074–80.
- [25] Zhao Z, Li X, Liu F, Zhu G, Ma C, Wang L. Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. *Sci Total Environ* 2020;729:138959.
- [26] Chowdhury R, Heng K, Shawon MSR, Goh G, Okonofua D, Ochoa-Rosales C, et al. Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries. *Eur J Epidemiol* 2020;35(5):389–99.
- [27] Li M, Leslie H, Qi B, Nan S, Feng H, Cai H, et al. Development of an openEHR Template for COVID-19 Based on Clinical Guidelines. *J Med Internet Res* 2020;22(6):e20239.
- [28] Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-Peer contact tracing: development of a privacy-preserving smartphone app. *JMIR mHealth uHealth* 2020;8(4):e18936.
- [29] Komenda M, Bulhart V, Karolyi M, Jarkovský J, Mužik J, Májek O, et al. Complex reporting of the COVID-19 epidemic in the Czech Republic: use of an interactive web-based app in practice. *J Med Internet Res* 2020;22(5):e19367.