

# Modified SEIQHRDP and Machine Learning Prediction for the Epidemics

Li Yike and Elena Gubar

*St. Petersburg State University,  
7/9, Universitetskaya nab., St. Petersburg, 198504, Russia  
E-mail: st102941@student.spbu.ru; e.gubar@spbu.ru*

**Abstract** This paper is dedicated to investigating the transmission and prediction of viruses within human society. In the first phase, we augment the classical Susceptible-Exposed-Infectious-Recovered (SEIR) model by incorporating four novel states: protected status ( $P$ ), quarantine status ( $Q$ ), self-home status ( $H$ ), and death status ( $D$ ). The numerical solution of this extended model is obtained using the well-established fourth-order Runge-Kutta algorithm. Subsequently, we employ the next matrix method to calculate the basic reproduction number ( $R_0$ ) of the infectious disease model. We substantiate the stability of the basic reproductive number through an analysis grounded in Routh-Hurwitz theory. Lastly, we turn to the application and comparison of statistical models, specifically the Autoregressive Integrated Moving Average (ARIMA) and Bidirectional Long Short-Term Memory (Bi-LSTM) models, for time series prediction.

**Keywords:** dynamics model, Runge-Kutta, ARIMA, Bi-LSTM model.

## 1. Introduction

In the course of human societal development, infectious diseases have consistently posed a significant global challenge. Modern epidemiology truly began when biochemist Kermack and physician Mckendrick (Kermack and Mckendrick, 1927) introduced a mathematical model known as the SIR (Susceptible-Infectious-Recovered) model. Jia W. P. et al. based on the classic SIR model, analyzed and forecasted the Italian epidemic using a dynamically extended SIR model, estimating the duration of the epidemic (Jia et al., 2020). Gubar E. et al. combined the SIR model with replicative dynamics to describe virus mutations (Gubar et al., 2018). They formulated optimal control problems to investigate the optimal strategies for healthcare and quarantine decisions.

In addition to the classical SIR model, Sun H. C. et al. combined campus COVID-19 data with the traditional SEIR model to validate the effectiveness of digital tracking and prevention strategies (Sun et al., 2021). Gubar E. et al. proposed a hierarchical model for virus transmission, dividing the transmission process into three levels: city, region, and country (Gubar et al., 2023). They utilized an improved Susceptible-Exposed-Infectious-Recovered-Death (SEIRD) model to define virus spread at each level and analyzed the balance between the cost of active virus transmission and the implementation of appropriate quarantine measures. These models simulate the evaluation of different intervention measures' effectiveness and reveal hidden characteristics of disease transmission. Therefore, process-based simulation models do not solely aim to accurately predict the future scale of infections. Instead, they serve as critical planning tools for deploying intervention strategies and regulatory decisions during epidemics.

<https://doi.org/10.21638/11701/spbu31.2023.08>

## 2. Dynamic model of epidemics

The emergence of Coronavirus disease 2019 (COVID-19) in Wuhan has attracted widespread global interest. In this section, we introduce a modified Susceptible-Exposed-Infectious-Removed (SEIR) model as the primary tool for epidemic analysis. To provide a solid foundation for our subsequent derivations, we begin by elucidating fundamental concepts and background knowledge integral to our proof. Following this, we present our revised SEIR model, subsequently calculating its basic reproduction number ( $R_0$ ) and establishing its stability through rigorous proof. Finally, we investigate the impact of parameter variations on the model, analyzing diverse strategies for epidemic management.

### 2.1. Dynamic model

Epidemic models frequently adopt a compartmental approach, wherein the host population (in the context of this article, human beings) is stratified into a limited number of compartments, each comprising individuals sharing identical disease status characteristics. To augment our capacity to characterize the global spread of COVID-19 comprehensively, we introduce the SEIQHRDP (Susceptible-Exposed-Infectious-Quarantined-Home-Recovered-Death-Protected) model. The SEIQHRDP model incorporates eight distinct compartments:

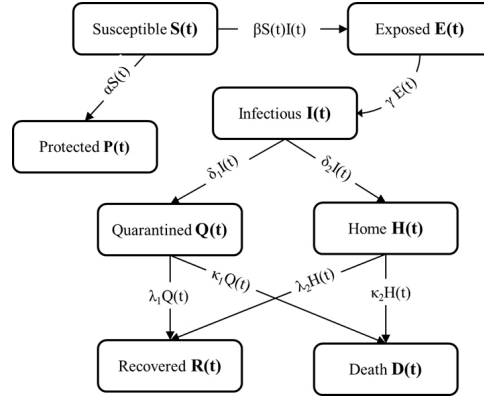
- *Susceptible* $\{n_S\}$ : Individuals who lack immunity to the infectious agent are susceptible and may become infected upon exposure.
- *Exposed* $\{n_E\}$ : Individuals in the incubation period are those who have been exposed to an infected person while they were susceptible and are now carrying the virus themselves.
- *Infectious* $\{n_I\}$ : Individuals who are currently infected represent those who are actively carrying the infection and have the potential to transmit it to susceptible individuals they come into contact with.
- *Quarantined* $\{n_Q\}$ : Individuals with an infected person isolated in a hospital, in times of medical shortage, are often referred to as seriously ill infected individuals.
- *Home* $\{n_H\}$ : Individuals with an infected person isolated at home, also referred to as asymptomatic infected individuals during times of medical shortage.
- *Recovered* $\{n_R\}$ : Individuals with immunity (having contracted and recovered from the disease) do not impact transmission dynamics upon contact with others.
- *Death* $\{n_D\}$ : Individuals who have died from the virus, including those who passed away in hospitals and at home.
- *Protected* $\{n_P\}$ : Individuals who are not previously infected with the virus adopt preventive measures, such as mask-wearing and maintaining a safe distance.

For all  $t$ ,  $N = n_S + n_E + n_I + n_Q + n_H + n_R + n_D + n_P$  represents the total population within the enclosed area. We define  $S(t) = \frac{n_S}{N}$ ,  $E(t) = \frac{n_E}{N}$ ,  $I(t) = \frac{n_I}{N}$ ,  $Q(t) = \frac{n_Q}{N}$ ,  $H(t) = \frac{n_H}{N}$ ,  $R(t) = \frac{n_R}{N}$ ,  $D(t) = \frac{n_D}{N}$ , and  $P(t) = \frac{n_P}{N}$  as the proportions of individuals in the categories of *Susceptible*, *Exposed*, *Infectious*, *Quarantined*, *Home*, *Recovered*, *Death*, and *Protected*, respectively. The three assumptions of the model are as follows:

1. The model considers a constant population size, denoted as "N."

2. Susceptible individuals transition to an infected state and become contagious upon contact with an infectious individual. Infected individuals who are hospitalized or in home isolation have significantly reduced transmission rates.
3. Susceptible individuals share the same effective exposure rate. And after the infected person recovers, they will not be infected again within a short period of time.

The flowchart depicting the SEIQHRDP model is presented below:



**Fig. 1.** Flow chart for the SEIQHRDP model

The differential equation governing the dynamics of the SEIQHRDP epidemic model is expressed as follows:

$$\begin{aligned}
 \frac{dS(t)}{dt} &= -\beta S(t)I(t) - \alpha S(t), \\
 \frac{dE(t)}{dt} &= \beta S(t)I(t) - \gamma E(t), \\
 \frac{dI(t)}{dt} &= \gamma E(t) - (\delta_1 + \delta_2) I(t), \\
 \frac{dQ(t)}{dt} &= \delta_1 I(t) - \lambda_1 Q(t) - \kappa_1 Q(t), \\
 \frac{dH(t)}{dt} &= \delta_2 I(t) - \lambda_2 H(t) - \kappa_2 H(t), \\
 \frac{dR(t)}{dt} &= \lambda_1 Q(t) + \lambda_2 H(t), \\
 \frac{dD(t)}{dt} &= \kappa_1 Q(t) + \kappa_2 H(t), \\
 \frac{dP(t)}{dt} &= \alpha S(t).
 \end{aligned} \tag{1}$$

Where  $\alpha \geq 0$  represents the average protection rate and  $\beta \geq 0$  is explicitly split in two terms as  $\beta_0 \cdot k$ . Where  $\beta_0$  stands for the rate of infection per effective contact, and  $k$  represents the number of contacts with other individuals (Pastor et al., 2015). Parameter  $\gamma \geq 0$  is the reciprocal of the average latency time. Rates  $\delta_1 \geq 0, \delta_2 \geq 0$  denote the average isolation rate of infected people in the hospital and at home, respectively. Parameters  $\lambda_1 \geq 0, \lambda_2 \geq 0$  stand for the recovery rate in hospital and at home, respectively. Rates  $\kappa_1 \geq 0, \kappa_2 \geq 0$  represent the death rate in hospital and at home, respectively.

**Theorem 1.** *The system (1) gives only positive solutions for positive initial conditions.*

**Remark 1.** *Theorem1* implies that if the initial conditions satisfy  $S(0) > 0, E(0) \geq 0, I(0) \geq 0, Q(0) \geq 0, H(0) \geq 0, R(0) \geq 0, D(0) \geq 0, P(0) \geq 0$ , then the solutions of the system (1), denoted as  $S(t), E(t), I(t), Q(t), H(t), R(t), D(t), P(t)$ , remain positive for all  $t \geq 0$ .

*Proof.* Starting from the differential equation  $\frac{dS(t)}{dt} = -\beta S(t)I(t) - \alpha S(t)$ , we obtain the following expression:

$$\frac{dS(t)}{dt} + (\beta I(t) + \alpha)S(t) = 0.$$

By separating variables, we obtain:

$$\frac{dS(t)}{S(t)} = -(\beta I(t) + \alpha) dt.$$

Solving the above equation we get:

$$S(t) = C e^{-\int_0^t \beta I(t) + \alpha dt}.$$

Where  $C$  is an arbitrary constant. At time  $t = 0$ , we have  $C = S(0) > 0$ , thus leading to:

$$S(t) = S(0) e^{-\int_0^t \beta I(t) + \alpha dt}.$$

This implies  $S(t) > 0$  for all  $t$ . Similarly, we can demonstrate that  $E(t) \geq 0, I(t) \geq 0, Q(t) \geq 0, H(t) \geq 0, R(t) \geq 0, D(t) \geq 0$ , and  $P(t) \geq 0$ , ensuring the non-negativity of these variables throughout the analysis.

The initial conditions for the population distribution within each compartment are defined as follows:  $n_S > 0, n_E \geq 0, n_I > 0, n_Q \geq 0, n_H \geq 0, n_R \geq 0, n_D \geq 0$ , and  $n_P \geq 0$ . Additionally, we assume a stable total population size, where the natural birth rate equals the natural death rate, expressed as  $dN/dt = 0$ . The interplay of these factors is crucially influenced by the basic reproduction number. As the epidemic unfolds, heightened awareness of self-protection among the population, in conjunction with government-imposed epidemic prevention policies, results in a steady decline in the susceptible population. This decline is characterized by the introduction of a positive protection rate, denoted as  $\alpha$ , within the model. Specifically, individuals transitioning into the protector compartment  $P$  are induced by susceptibles within compartment  $S$  at a rate of  $\alpha S$ .

New infections within compartment  $E$  result from contact between susceptible individuals in compartment  $S$  and infected individuals in compartment  $I$  at a rate denoted as  $\beta SI$ . The infection process within compartment  $I$  is initiated by compartment  $E$  at a rate of  $\gamma E$  during the incubation period. Subsequently, individuals who become infected make choices regarding isolation, with some opting for home isolation, represented by a rate of  $\delta_2 I$ , while others choose hospital treatment, represented by a rate of  $\delta_1 I$ . Transitions from compartment  $Q$  to compartment  $R$  occur at a rate of  $\lambda_1 Q$ , and from compartment  $H$  to compartment  $R$  at a rate of  $\lambda_2 H$ , contributing to the population of recoverers in compartment  $R$ . Similarly, the compartment for deceased individuals,  $D$ , is replenished from compartment  $Q$  at a rate of  $\kappa_1 Q$  and from compartment  $H$  at a rate of  $\kappa_2 H$ .

## 2.2. Basic reproduction number and stability analysis

The basic reproductive number, denoted as  $R_0$ , serves as a crucial epidemiological parameter for quantifying the infectivity and transmissibility of infectious agents. Specifically, it represents the number of secondary cases that a single case would generate within a population entirely susceptible to the disease. In mathematical terms,  $R_0$  acts as a threshold for the stability of the disease-free equilibrium, and its value correlates with both the peak and final size of an epidemic outbreak. When  $R_0 < 1$ , it indicates that, on average, a limited number of infected individuals introduced to a fully susceptible population will fail to propagate the disease further, resulting in its containment. Conversely, when  $R_0 > 1$ , each successive generation will witness an increase in the number of infected individuals, facilitating the widespread transmission of the disease (Van den and Watmough, 2002).

The computation of the basic reproduction number ( $R_0$ ) has reached a relatively mature stage, particularly in the context of COVID-19. In our analysis, we employ the next-generation method to estimate  $R_0$  for COVID-19 across diverse geographical regions. As per the previously defined parameters, the total population size, denoted as  $N$ , remains a constant throughout the system's evolution, ensuring its invariance. This is formally expressed as the sum of individuals in all compartments equating to 1, i.e.,  $S(t) + E(t) + I(t) + Q(t) + H(t) + R(t) + D(t) + P(t) = 1$ . The disease-free equilibrium (DFE) represents the state of equilibrium within the system in the absence of infected individuals. At the initial time point, the susceptible compartment is initialized as  $S(0) = 1$ , while the remaining compartments' ratios are set to zero. Consequently, we establish the DFE point as  $x_0 = (1, 0, 0, 0, 0, 0, 0, 0)$  at the initial time. Subsequently, we derive the matrices  $F$  and  $V$ .

$$F = \begin{bmatrix} \beta SI \\ 0 \end{bmatrix}, V = \begin{bmatrix} \gamma E \\ -\gamma E + (\delta_1 + \delta_2)I \end{bmatrix}.$$

Differentiate  $F$  and  $V$  with respect to  $E$  and  $I$  separately, and then substitute the results into the disease-free equilibrium point:

$$F = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix}, V = \begin{bmatrix} \gamma & 0 \\ -\gamma & \delta_1 + \delta_2 \end{bmatrix}.$$

The next generation matrix of the system is defined as follows:

$$F \cdot V^{-1} = \begin{bmatrix} \frac{\beta}{\delta_1 + \delta_2} & \frac{\beta}{\delta_1 + \delta_2} \\ 0 & 0 \end{bmatrix}.$$

$R_0$  is determined as the maximum eigenvalue or spectral radius of the next generation matrix, thus obtaining the expression for  $R_0$  as:

$$R_0 = \frac{\beta}{\delta_1 + \delta_2}. \quad (2)$$

**Theorem 2.** *The model around DFE is locally asymptotically stable if  $R_0 < 1$ .*

*Proof.* Begin by computing the Jacobian matrix corresponding to system (1) as follows:

$$J = \begin{pmatrix} -\beta I(t) - \alpha & 0 & -\beta S(t) & 0 & 0 \\ \beta I(t) & -\gamma & \beta S(t) & 0 & 0 \\ 0 & \gamma & -(\delta_1 + \delta_2) & 0 & 0 \\ 0 & 0 & \delta_1 & -(\lambda_1 + \kappa_1) & 0 \\ 0 & 0 & \delta_2 & -(\lambda_2 + \kappa_2) & 0 \end{pmatrix}.$$

The terms  $\frac{dR(t)}{dt}$ ,  $\frac{dD(t)}{dt}$ , and  $\frac{dP(t)}{dt}$  are not included because they decouple from the system (1). In other words, the  $8 \times 8$  matrix that includes  $\frac{dR(t)}{dt}$ ,  $\frac{dD(t)}{dt}$ , and  $\frac{dP(t)}{dt}$  has three entire columns of zeros. Then, by substituting the DFE  $x_0 = (1, 0, 0, 0, 0)$  into the Jacobian matrix, we obtain  $J(x_0)$ :

$$J(x_0) = \begin{pmatrix} -\alpha & 0 & -\beta & 0 & 0 \\ 0 & -\gamma & \beta & 0 & 0 \\ 0 & \gamma & -(\delta_1 + \delta_2) & 0 & 0 \\ 0 & 0 & \delta_1 & -(\lambda_1 + \kappa_1) & 0 \\ 0 & 0 & \delta_2 & -(\lambda_2 + \kappa_2) & 0 \end{pmatrix}.$$

The characteristic roots for  $J(x_0)$  are:

$$\begin{aligned} \rho_1 &= 0, \\ \rho_2 &= -\alpha, \\ \rho_3 &= -(\lambda_1 + \kappa_1), \\ \rho_4 &= \frac{-(\delta_1 + \delta_2 + \gamma) - \sqrt{[(\delta_1 + \delta_2) - \gamma]^2 + 4\beta\gamma}}{2}, \\ \rho_5 &= \frac{-(\delta_1 + \delta_2 + \gamma) + \sqrt{[(\delta_1 + \delta_2) - \gamma]^2 + 4\beta\gamma}}{2}. \end{aligned}$$

For values of  $R_0$  less than 1, specifically when  $R_0 < 1$ , this implies that  $\beta < \delta_1 + \delta_2$ , with all parameters being greater than zero. Under these conditions, the characteristic roots, denoted as  $\rho_2$ ,  $\rho_3$ ,  $\rho_4$ , and  $\rho_5$ , exhibit negativity. Furthermore, when the largest eigenvalue of the Jacobian matrix  $J$  is zero, this signifies, in biological terms, a disease growth rate of zero (Wonham and Lewis, 2008). In accordance with the Routh-Hurwitz criteria(?), we conclude that the Disease-Free Equilibrium (DFE) is locally asymptotically stable under these circumstances.

### 2.3. Numerical Simulation

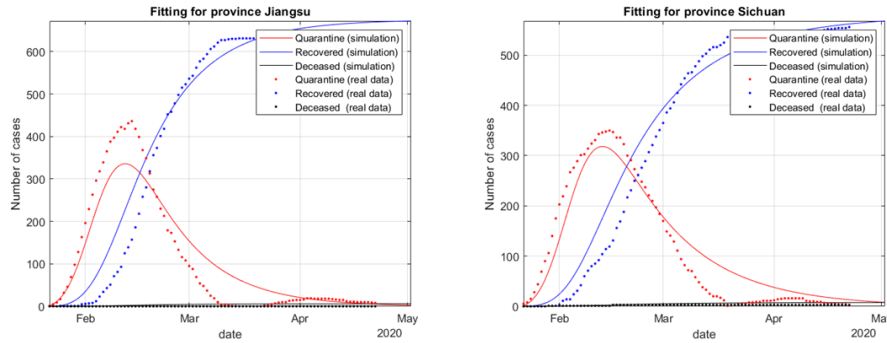
In this section, we validate our findings through a series of numerical simulations. Initially, we curated data from the ten provinces in China where the disease first emerged to conduct a thorough verification analysis. The challenge in numerically simulating the model primarily revolves around parameter selection. Some parameters were acquired through pathological research, while the remainder were determined by fitting real-world data using the Levenberg-Marquardt algorithm (Gavin, 2019). Subsequently, with the model parameters in hand, we applied the classical fourth-order Runge-Kutta (RK4) algorithm to solve and analyze the non-linear ordinary differential equations.

We utilized the COVID-19 database provided by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (Dong et al., 2020). The dataset offers a reliable time series for confirmed cases, deaths, and recoveries. In our model, nine parameters require estimation, all of which are initially set to values within the  $[0,1]$  boundary range. Regarding the initial values for compartments, we assume a small number of infected individuals at the beginning of the epidemic, establishing initial conditions as follows:  $Q_0 = 1$ ,  $I_0 = 10$ ,  $E_0 = 10$ ,  $H_0 = 10$ , and  $P_0 = 0$ . This configuration denotes an initial state with 1 quarantined individual, 10 infected individuals, 10 exposed individuals, 10 home-quarantined individuals, and 0 protected individuals. The initial values for recoverers and fatalities are already known. The total population size is set to the average provincial population of China,  $N = 30,000,000$ . The initial value for compartment  $S$  is calculated as the

difference between the total population and the sum of the remaining seven compartments. Additionally, we make the assumption that the number of individuals in hospital isolation equals the reported cases count minus the number of recoveries and deaths. Below, we outline the algorithmic steps for parameter fitting and the application of the RK4 method.

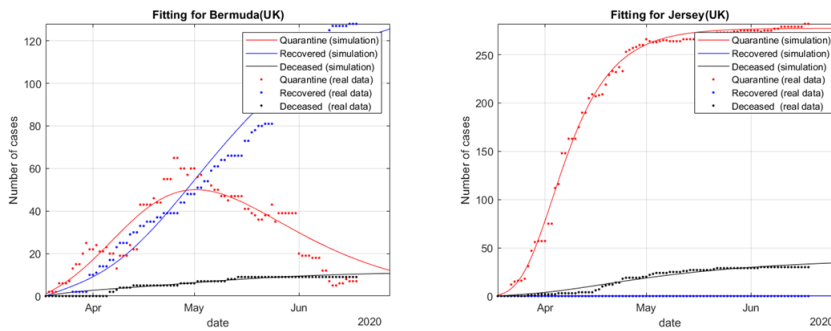
- (a) **Data Preparation:** Begin by acquiring the necessary data. Data can be sourced online or from local files. Follow this step with necessary data processing.
- (b) **Initialization:** Set the initial values and define the start and end times for the analysis. Develop the PtoODE function, responsible for computing the solution of an Ordinary Differential Equation (ODE) using the model parameters. In this context, the RK4 algorithm, either self-defined or the `ode45` function, can be employed to solve the ODE.
- (c) **Objective Function Creation:** Modify the PtoODE function to return only the ODE output for specific variables, such as  $Q$ ,  $R$ , and  $D$ . Create an objective optimization problem using the output from the modified PtoODE function and compare it to the observed data. The objective function is typically formulated as  $obj = \sum(\sum((myfcn - yvals)^2))$ , where  $myfcn$  represents the output of the modified PtoODE function, and  $yvals$  denotes the observed data. Solve this problem based on the limited set of observations, typically employing the `lsqrsolve` function to minimize the objective equation.
- (d) **Termination Criteria:** Determine the termination conditions for the optimization process. Two common methods are employed: firstly, termination occurs when the objective function value is less than a predefined tolerance value. Secondly, termination happens when the number of iterations exceeds a predefined maximum limit. If the first method is used, it indicates that the fitted parameter values have been obtained. Conversely, if the second method is invoked, it signifies a deviation between the initially selected parameter values and the true values, necessitating a re-adjustment of the parameter initialization for recalibration.
- (e) **Final Solution:** Use the parameter values obtained in the fourth step to call the PtoODE function once more to complete the solution of the differential equations, thereby finalizing the parameter fitting process.

Having established the algorithm and initial conditions, the subsequent step involves programming implementation. To execute parameter fitting, we employ the `lsqrsolve` function. It's essential to note that our parameter fitting relies on a restricted set of observations, specifically the values associated with compartments  $Q$ ,  $R$ , and  $D$ . For validation purposes, we have selected ten Chinese provinces with documented outbreaks. These provinces include Anhui, Beijing, Chongqing, Fujian, Guangdong, Heilongjiang, Henan, Jiangsu, Jiangxi, and Sichuan. For programming convenience, and to maintain uniform initial conditions, these ten cities were chosen due to their respective outbreaks commencing on January 22. Consequently, we set the start date as January 22, 2020, and designate the end date as April 22, 2020. The results of the parameter fitting for China are depicted in the figure below.



**Fig. 2.** The parameter fitting outcomes for Jiangsu province are displayed in the left image, while the right image illustrates the parameter fitting results for Sichuan province. Detailed parameter fitting results for the remaining eight Chinese provinces are available in Appendix A. In this context, the term "*simulate*" refers to the model-generated results, while the "*real data*" is sourced from Johns Hopkins University (Dong et al., 2020). These visual representations provide a comprehensive overview of the parameter fitting performance for various regions.

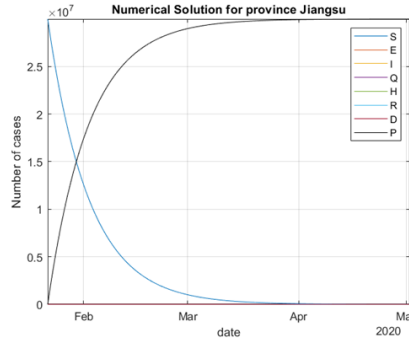
To evaluate the generalizability of our model beyond the context of COVID-19 outbreaks in China, we conducted an experiment using the United Kingdom (UK) as an illustrative example (selected arbitrarily). For consistency in programming and to maintain uniform initial conditions, we chose Bermuda, Gibraltar, Isle of Man, and Jersey as additional test locations. These four regions were selected because their respective outbreaks commenced simultaneously on March 18, 2020, and concluded on June 18, 2020, covering a three-month period. Specifically, we set  $Q_0 = 1, I_0 = 10, E_0 = 10, H_0 = 10, P_0 = 0$ . The average total population for each location in the UK is  $N = 100,000$ . The results of the parameter fitting for the UK are illustrated in the Figure 3, highlighting the model’s adaptability to diverse regions.



**Fig. 3.** The left image depicts the parameter fitting outcomes for Bermuda, while the right image displays the parameter fitting results for Jersey. In this context, "*simulate*" signifies the results generated by the model, while the "*real data*" is sourced from Johns Hopkins University (Dong et al., 2020). These visual representations provide a comparative assessment of the model’s performance for different regions, reaffirming its versatility and applicability beyond the Chinese context.

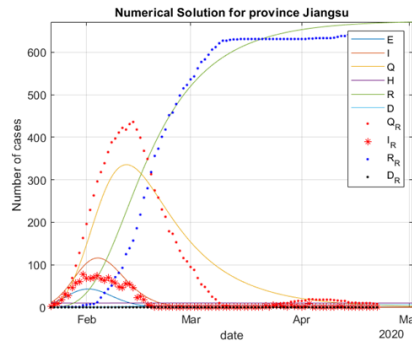


Following the execution of the *lsqrsolve* function, we obtain the parameter values derived from fitting for the ten Chinese provinces and the four locations within the UK. The detailed parameter values are available in Appendix C.. As illustrated in Figure 4, upon substituting the fitted parameters, we generate the results of the SEIQRDP model, providing insights into the disease dynamics in Jiangsu province.



**Fig. 4.** The outcomes derived from the SEIQRDP model for Jiangsu province. Each curve represents the population count within specific compartments over time, with time units divided into 24 segments per day.

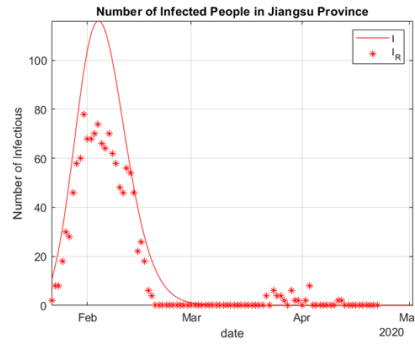
Typically, in the early stages of an outbreak, susceptible individuals constitute the majority of the province’s population. Despite the relatively low protection rate, the sheer magnitude of the  $S$  compartment leads to a substantial population within the  $P$  compartment. This explains the significant population within compartments  $S$  and  $P$ . Consequently, we exclude compartments  $S$  and  $P$  from analysis and focus our attention on the remaining compartments. The figure below presents the results for Jiangsu province concerning these remaining compartments, namely  $(E, I, Q, H, R, D)$ .



**Fig. 5.** The numerical solution of the model for Jiangsu province. Within this context,  $E, I, Q, H, R,$  and  $D$  denote the population counts within their respective compartments across time intervals. Furthermore,  $Q_R, I_R, R_R, D_R$  correspond to the daily real data pertaining to compartments  $Q, I, R, D,$  respectively. We operate under the assumption that the actual number of infected individuals is twice the number of confirmed cases.

Upon comparing the outcomes presented in Appendix A, a consistent pattern in the dynamics model becomes evident. During the initial phase of the epidemic, the model underestimates the number of quarantined individuals compared to the actual count. This discrepancy arises because, at the outset, medical resources are readily available. However, the fitting parameter utilized in our model remains constant, resulting in an estimate that represents the average number of quarantined individuals during this period. Similarly, at the commencement of the epidemic, the model's estimate for the number of recovered individuals surpasses the actual value. This discrepancy is attributed to the fixed recovery rate employed in our fitting, which does not account for the dynamic nature of the actual recovery rate.

To explore the impact of crucial parameters on virus transmission, we conducted separate analyses to assess the disparities between the number of infected individuals and real-world data under varying parameter configurations.



**Fig. 6.** We conducted a comparison of the number of infected individuals in Jiangsu province. Here,  $I$  signifies the count of infected individuals simulated by the model, while  $I_R$  denotes the actual number of reported infected individuals.

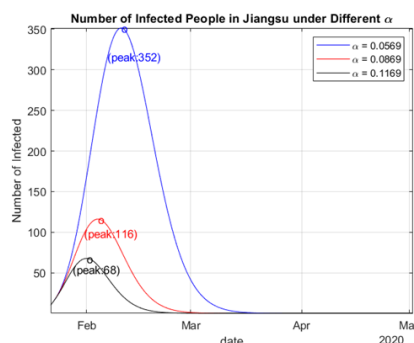
In Figure 6, the term  $I_R$  can be denoted as 'ascertainment,' which refers to the process of discerning the characteristics, status, or events within a population or study group. Associated with this concept is the noun 'ascertainment bias,' signifying a systematic deficiency in equally representing all categories of cases or individuals that should ideally be included in a sample. This bias can stem from various factors such as the medical environment or economic considerations (Porta, 2014).

Next, we proceed to analyze the influence of critical parameters within the model on the number of infected individuals. The parameter values acquired through fitting for Jiangsu province are as follows:  $\alpha = 0.0869$ ,  $\beta = 1$ ,  $\gamma = 1$ ,  $\delta_1 = 0.3418$ ,  $\delta_2 = 0.000101$ ,  $\lambda_1 = 0.064392$ ,  $\lambda_2 = 0.0001$ ,  $\kappa_1 = 0.00057$ ,  $\kappa_2 = 0.0001$ .

The protection rate  $\alpha = 0.0869$  indicates that at the onset of the epidemic, the population had a relatively low rate of protection, suggesting a lack of awareness and preventive measures. The value of  $\beta = 1$  implies that the product of the average number of contacts per person and the probability of disease transmission between a susceptible and an infectious individual is equal to one. This signifies a high transmission potential in Jiangsu province. The rate  $\gamma = 1$  signifies that the reciprocal of the virus's incubation period is one, indicating a short incubation period of one day at the outbreak's initiation. Parameter values  $\delta_1 = 0.3418$  and  $\delta_2 = 0.000101$  represent the isolation rates for infected individuals in hospital and

at home, respectively. Notably, the isolation rate in hospitals is notably higher, suggesting that people's inclination towards home isolation is relatively limited. Values  $\lambda_1 = 0.064392$  and  $\lambda_2 = 0.0001$  respectively signify the recovery rates for infected individuals in hospital isolation and home isolation. Evidently, the recovery rate in hospital isolation is higher due to the availability of better medical resources. Finally, rates  $\kappa_1 = 0.00057$  and  $\kappa_2 = 0.0001$  represent the mortality rates in hospital and home isolation, respectively, with the mortality rate in hospitals being slightly higher. This discrepancy arises because individuals isolated in hospitals typically include severely ill patients, resulting in a location with both a high recovery rate and mortality rate under normal circumstances.

We initiated our analysis by examining the influence of the protection rate on the number of infections. Specifically, we varied the value of  $\alpha$  (protection rate) while maintaining the other parameter values constant. We considered parameter  $\alpha$  with values of 0.0569, 0.0869, and 0.1169, with  $\alpha = 0.0869$  serving as the initial value. The resulting changes in the number of infected individuals are as follows:

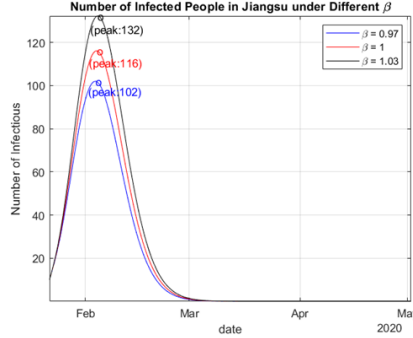


**Fig. 7.** We observed changes in the number of infected individuals in Jiangsu province across various protection rates. The remaining parameters were kept constant:  $\beta = 1, \gamma = 1, \delta_1 = 0.3418, \delta_2 = 0.000101, \lambda_1 = 0.064392, \lambda_2 = 0.0001, \kappa_1 = 0.00057, \kappa_2 = 0.0001$ .

Figure 7 illustrates that higher protection rates among the population lead to a decrease in the peak number of infections and an earlier time to reach that peak. This suggests that a higher protection rate can lead to an earlier end to the epidemic. Initially, the estimated protection rate for susceptible individuals in the system was 0.0869. If strict government control measures are implemented at the outset, and protection rates are increased to 0.1169, the peak number of infected individuals in Jiangsu province would decrease from 116 to 68. Conversely, if protection measures are neglected and the protection rate drops to 0.0569, the peak number of infections in Jiangsu province would rise from 116 to 352, and the time to reach the peak would be extended, prolonging the infectious disease process. These findings underscore the importance of increasing protection rates, which can be achieved through measures such as mask-wearing, maintaining safe distances in crowded places, government-imposed controls.

We next investigate the impact of the parameter  $\beta$  on the number of infected individuals. In this context,  $\beta$  represents the product of the average number of contacts per person per hour, denoted as  $k$ , and the probability of disease transmission in a contact between a susceptible and an infectious compartment, represented as

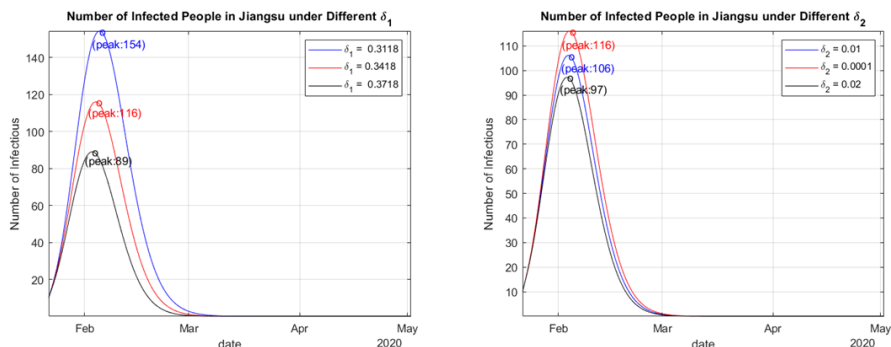
$\beta_0$ . The value range of  $\beta_0$  lies within the interval  $[0, 1]$ . Assuming a consistent average number of contacts per person per unit of time, the probability of infection is directly proportional to  $\beta$ . Consequently, we vary the value of  $\beta$  while keeping other parameter values constant. Specifically, we consider  $\beta$  values of 0.97, 1, and 1.03, with the initial value being  $\beta = 1$ . The changes in the number of infected individuals are as follows:



**Fig. 8.** The alterations in the count of infected individuals in Jiangsu province are displayed in Figure 8. In this analysis, all parameters remain constant except for  $\beta$ , which is varied. The parameter values held constant are as follows:  $\alpha = 0.0869, \gamma = 1, \delta_1 = 0.3418, \delta_2 = 0.000101, \lambda_1 = 0.064392, \lambda_2 = 0.0001, \kappa_1 = 0.00057, \kappa_2 = 0.0001$ .

The observations in Figure 8 illustrate the impact of varying  $\beta$ , which represents the average number of contacts per person per time multiplied by the probability of disease transmission in a contact between a susceptible and an infectious compartment. Parameter  $\beta$  encompasses both the average contact frequency and infection probability for each contact. Assuming a constant average contact frequency, changes in  $\beta$  directly influence the infection probability. Likewise, when keeping the virus’s infection probability constant, variations in  $\beta$  reflect changes in the average contact frequency per person per unit of time. Increasing  $\beta$  from 1 to 1.03 results in an elevation of the peak number of infected individuals in Jiangsu province from 116 to 132. Conversely, decreasing  $\beta$  from 1 to 0.97 reduces the peak number of infected individuals from 116 to 102. These findings emphasize the importance of controlling both the average contact frequency and the virus’s inherent infection probability to manage the epidemic effectively.

We proceeded to investigate the influence of isolation rates on the number of infections. Specifically, we varied the values of  $\delta_1$  (hospital isolation rate) and  $\delta_2$  (home isolation rate), keeping all other parameter values constant. The parameter  $\delta_1$  was assigned values of 0.3118, 0.3418, and 0.3718, with an initial value of  $\delta_1 = 0.3418$ . Additionally, the parameter  $\delta_2$  was set to 0.01, 0.0001, and 0.02, with an initial value of  $\delta_2 = 0.0001$ . The resulting changes in the number of infected individuals are outlined below:



**Fig. 9.** The image on the left illustrates the effect of  $\delta_1$  (hospital isolation rate) on the number of infected individuals in Jiangsu province. Meanwhile, the image on the right demonstrates the impact of  $\delta_2$  (home isolation rate) on the number of infected individuals in Jiangsu province. The remaining parameters are held constant at the following values:  $\alpha = 0.0869$ ,  $\beta = 1$ ,  $\gamma = 1$ ,  $\lambda_1 = 0.064392$ ,  $\lambda_2 = 0.0001$ ,  $\kappa_1 = 0.00057$ ,  $\kappa_2 = 0.0001$ .

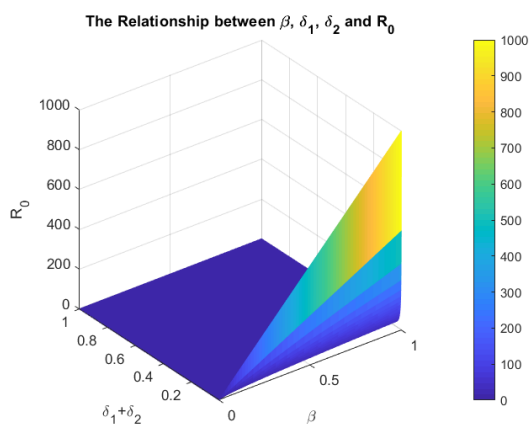
As illustrated in Figure 9, an increase in either  $\delta_1$  (hospital isolation rate) or  $\delta_2$  (home isolation rate) leads to a reduction in the epidemic's peak. The spread of the virus is contingent upon the isolation rate, and an elevated isolation rate corresponds to a diminished viral spread. This substantiates our hypothesis that individuals under hospital care or home isolation exhibit markedly lower transmission rates. Thus, enhancing public awareness of isolation measures can effectively curb virus propagation.

After scrutinizing the impact of crucial parameters on infection numbers in Jiangsu province, we proceeded to investigate the correlation between these parameters and the basic reproductive number. By substituting the fitted parameters (Appendix C) into formula (2), we derived the basic reproductive number for each province, as presented in the table below:

**Table 1.** Basic Reproductive Numbers of Different Provinces in China

Province	$\beta$	$\delta_1$	$\delta_2$	$R_0$
Anhui(CN)	1	0.323654	0.000106	3.088707685
Beijing(CN)	0.999999	0.227907	0.141219	2.709099332
Chongqing(CN)	1	0.357595	0.0001	2.795677882
Fujian(CN)	1	0.311163	0.056692	2.718462438
Guangdong(CN)	0.999997	0.307476	0.0001	3.251219211
Heilongjiang(CN)	1	0.200023	0.130433	3.026121481
Henan(CN)	0.999988	0.312456	0.000101	3.199378033
Jiangsu(CN)	1	0.341768	0.000101	2.92509704
Jiangxi(CN)	1	0.325987	0.0001	3.066666258
Sichuan(CN)	1	0.303308	0.050794	2.824045049
Bermuda(UK)	0.369	0.0912	0.0001	4.04162103
Gibraltar(UK)	1	0.1099	0.2713	2.623294858
Isle of Man(UK)	1	0.0466	0.1889	4.246284501
Jersey(UK)	1	0.1847	0.0092	5.157297576

In accordance with Table 1, it is evident that the basic reproduction number ( $R_0$ ) of COVID-19 in various provinces across China falls within the range of [2.7, 3.3]. In contrast, the basic reproduction number for COVID-19 in the United Kingdom (UK) lies within the range of [2.6, 5.2], with an average  $R_0$  surpassing that of China. This suggests a higher transmission potential of the virus in the UK compared to China. As an illustration, consider Jiangsu province, with a basic reproductive number standing at approximately 2.93. This implies that, without any intervention in epidemic control and assuming no immunity, an infected individual in Jiangsu province would infect nearly 3 people on average. The figure below illustrates the relationship between  $\beta$ ,  $\delta_1$ ,  $\delta_2$ , and  $R_0$ .



**Fig. 10.** The relationship between  $\beta$ ,  $\delta_1$ ,  $\delta_2$ , and  $R_0$  is depicted graphically, with a colorbar on the right side of the figure. The intensity of yellow in the colorbar corresponds to higher values of  $R_0$ .

Figure 10 illustrates that the virus exhibits high contagiousness. Without any isolation measures, the basic reproductive number can attain substantial values, implying that, theoretically, a single infected individual can initiate the infection of numerous susceptible individuals. With time, a large portion of the population could become infected throughout the system. In cases where the virus has a relatively high fatality rate, this scenario might even pose a threat to human survival.

### 3. Time series forecasting

The time series forecasting method, a form of quantitative forecasting, essentially operates as a regression forecasting technique. Its fundamental principle involves recognizing the continuity in the evolution of phenomena, utilizing historical time series data for statistical analysis, and deducing the future trends of these phenomena.

#### 3.1. Forecasting model

Over time, the daily increase in the number of infected individuals contributes to the creation of a comprehensive time series dataset. In this study, we opted for two distinct models, namely the Autoregressive Integrated Moving Average (ARIMA) and Bidirectional Long-Short Term Memory (Bi-LSTM), for prediction. The epi-

demic SEIQHRDP dynamics model can predict the number of infections. But at the same time, we have also observed that the epidemic may have multiple waves. In order to assess the predictive capabilities of the SEIQHRDP model, we have established SEIQHRDP1 for single-wave predictions using specified initial values. Additionally, we have employed the model values at the corresponding time point of the previous wave's outbreak as the initial conditions to establish the SEIQHRDP2 model to accommodate the forecasting requirements for multiple waves of the epidemic.

**ARIMA model** The Autoregressive Integrated Moving Average (ARIMA) model, also known as the differential integrated moving average autoregressive model, is a prominent technique for time series prediction and has been applied in forecasting virus outbreaks. This model operates on the premise of transforming non-stationary time series into stationary time series by regressing the dependent variable solely on its lag value, the present value, and the lag value of the random error component. The ARIMA model is a combination of the  $AR(p)$  model and the  $MA(q)$  model. After obtaining stationary data through  $d$ -order differencing, predictions can be made. The ARIMA (p, d, q) model can be expressed as (Ho and Xie, 1998):

$$y'_t = \alpha_0 + \sum_{i=1}^p \alpha_i y'_{t-i} + \varepsilon_t + \sum_{i=1}^q \beta_i \varepsilon_{t-i}.$$

In the ARIMA model, we have the expression  $y'_t = \Delta^d y_t = (1 - L)^d y_t$  ( $d$ -order difference expression, where  $L$  is the lag operator). When  $d = 1$ ,  $\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$ . Here,  $y_t$  represents the current value,  $\alpha_i$  stands for the autocorrelation coefficient,  $\beta_i$  represents the error coefficient, and  $\varepsilon_t$  denotes white noise. Constructing an ARIMA model involves six main steps.

- (a) **Acquisition of Time Series Data:** This step involves obtaining the time series data, which includes handling missing values and checking for any singular values.
- (b) **Preprocessing of Time Series:** In this phase, several tasks are performed, such as conducting stationarity tests and white noise tests. Simultaneously, it is assessed whether the original data requires differencing.
- (c) **Model Identification:** Identifying an appropriate model from known models that suits the given time series process is essential. This step also involves determining if the data exhibits any seasonal trends, etc., and selecting the corresponding ARIMA model through various data representations.
- (d) **Determination of Model Order:** Once the model type is identified, determining its order is crucial. The selection of the model order can be guided by criteria like AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion).
- (e) **Parameter Estimation:** In this step, the parameters of the model are estimated using the maximum likelihood estimation method.
- (f) **Model Verification:** The primary goal here is to verify the goodness-of-fit of the model. If the initially derived model fails certain tests, it may need to be refitted until it passes self-noise tests.

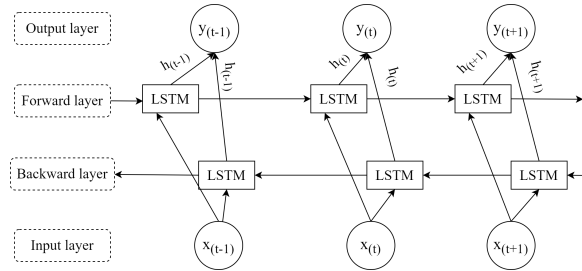
The Akaike Information Criterion (AIC), introduced by the Japanese statistician Akaike in 1973, serves as a minimum information criterion used in statistical

modeling(Bozdogan, 1987). Generally, a larger value of the likelihood function indicates a better fit for the model. The AIC criterion is mathematically expressed as:

$$AIC = 2k - 2 \ln(L).$$

Where,  $k$  represents the number of parameters, and  $L$  denotes the likelihood function. The optimal model selection involves choosing the model with the lowest AIC value. The essence of Akaike’s information criterion is to identify the model that provides the best data explanation while minimizing the inclusion of free parameters.

**Bi-LSTM model** The Long Short-Term Memory (LSTM) model is an artificial neural network model utilized in artificial intelligence and deep learning. Due to its distinctive architectural design, LSTM is well-suited for processing and forecasting time series. By incorporating both historical and future data, the model training process becomes more precise. At each time step, the input is concurrently fed into two LSTM networks running in opposite directions, and the output is collectively determined by the outputs of these two unidirectional LSTM networks. The network architecture of the Bi-LSTM model is illustrated below (Shahid et al., 2020).



**Fig. 11.** Architecture of Bi-LSTM

As depicted in Figure 11, the Bi-LSTM architecture consists of a two-layer neural network. The first layer takes the series input from the left, while the second layer begins with the series input from the right, performing a reverse processing akin to the first layer. Ultimately, the results from both layers are combined and processed to produce the output.

**3.2. Forecast comparison**

In this study, we take Jiangsu province as a representative case. Following the procedures delineated in Section 3.1, we conducted training for the ARIMA model. By selecting the model with the lowest AIC value, we obtained the model parameters:  $p = 2$  and  $q = 1$ . To enforce stationarity, we applied a first-order difference to the initial data, setting  $d = 1$  accordingly.

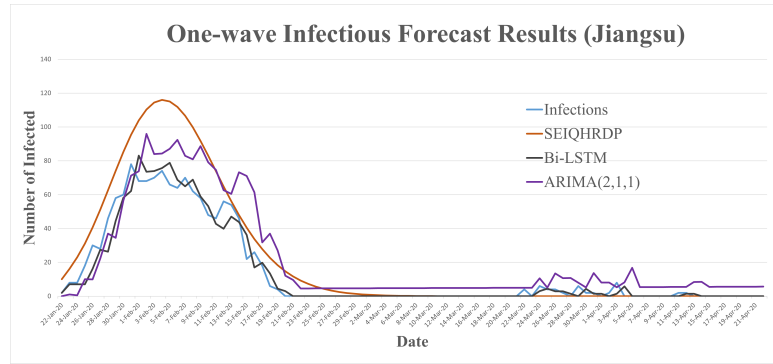
Concerning the hyperparameter configuration for the Bi-LSTM model, we adopted the following table as our hyperparameter settings:



**Table 2.** Bi-LSTM hyperparameter setting table

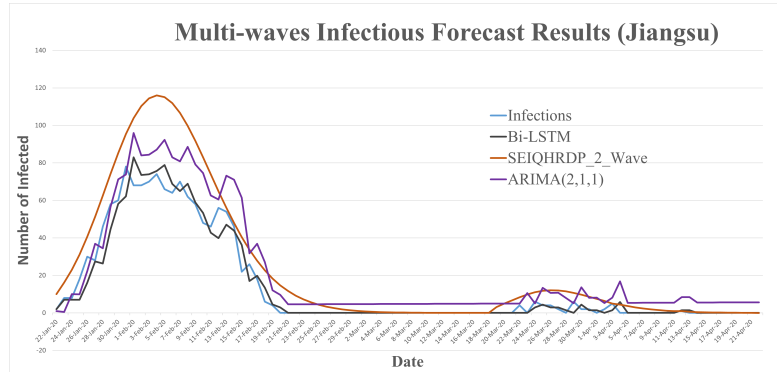
Hyperparameter	Values	Parameter Description
batch_size	32	Size of the sample to be processed each time
epochs	500	Number of training times
units	32	Number of hidden neurons
activation	relu	Model activation function
dropout	0.1	Proportion of hidden layer neurons discarded
optimizer	adam	Model optimizer selection
loss	RMSE	Model loss function selection

In this section, we commence by conducting a comparative analysis of various models on the training dataset. The parameter settings for the SEIQRDP model in Jiangsu province remain consistent with our previous values:  $\alpha = 0.0869$ ,  $\beta = 1$ ,  $\gamma = 1$ ,  $\delta_1 = 0.3418$ ,  $\delta_2 = 0.000101$ ,  $\lambda_1 = 0.064392$ ,  $\lambda_2 = 0.0001$ ,  $\kappa_1 = 0.00057$ ,  $\kappa_2 = 0.0001$ .



**Fig. 12.** The SEIQRDP model employs a single-wave prediction approach. This signifies that its parameters remain constant throughout the specified timeframe. In the graphical representation, the purple line illustrates the daily count of infected individuals predicted by the ARIMA(2,1,1) model. Conversely, the black line corresponds to the daily predictions of infected individuals generated by the Bi-LSTM model. Finally, the blue line represents the actual number of infected individuals in Jiangsu province.

According to data from Worldometer (<https://www.worldometers.info/coronavirus/>), it has come to our attention that numerous regions have experienced multiple waves of epidemic outbreaks. In this scenario, our approach entails setting the initial infected count back to zero and identifying the timing of the second wave. Subsequently, we employ the previously estimated parameters and initial values to invoke the algorithm once more, allowing for predictions in a multi-wave context. In this regard, we initialize each parameter with the values obtained from the prior model fitting, setting them as follows:  $Q_0 = 0$ ;  $I_0 = 0$ ;  $E_0 = 10$ ;  $R_0 = 632$ ;  $H_0 = 10$ ;  $D_0 = 6$ . The time selection begins with the first three days of the second wave, starting on March 18, 2020. The ensuing comparison highlights the prediction outcomes for the multi-wave scenario, contrasting the results of Bi-LSTM and ARIMA(2,1,1). It is worth noting that SEIQRDP exhibits suboptimal predictive performance in the context of multiple waves. However, due to its inherent complexity, the Bi-LSTM model surpasses many alternatives in terms of fitting accuracy.



**Fig. 13.** Forecasting the number of infected in Jiangsu province for two waves. In contrast, the predictions generated by the other models remain unchanged

The obtained parameter results after fitting the second wave are as follows:  $\alpha = 0.1729, \beta = 0.9340, \gamma = 1, \delta_1 = 0.0706, \delta_2 = 0.1900, \lambda_1 = 0.0078, \lambda_2 = 0.0001, \kappa_1 = 0.00032, \kappa_2 = 0.0001$ . Comparing these parameters to the ones obtained during the fitting of the first wave, a significant improvement is observed in  $\alpha$ . This suggests that as the epidemic progresses, the general public’s self-protection awareness increases, naturally leading to a decrease in the average contact rate  $\beta$ . The reduction in  $\delta_1$  may be attributed to insufficient medical resources, prompting mildly symptomatic patients to opt for home isolation, consequently increasing the probability of  $\delta_2$  for home isolation. As public awareness about the virus grows, the mortality rate decreases compared to the initial stages of the outbreak. To assess the quality and fitting effectiveness of the model’s prediction results, we adopted the root mean square error (RMSE) and coefficient of determination ( $R^2$ ) as measurement standards (Chicco et al., 2021). The table below presents the RMSE and  $R^2$  values for the different models depicted in Figure 13.

**Table 3.** Here are the RMSE and  $R^2$  values for the predictive models. SEIQRDP1 corresponds to a single-wave prediction, while SEIQRDP2 denotes a two-wave prediction.

	ARIMA(2,1,1)	Bi-LSTM	SEIQRDP1	SEIQRDP2
RMSE	10.94510873	5.254492166	15.43665338	15.65904331
R-squared	0.793147929	0.952326042	0.58854152	0.576600673

As shown in Table 3, the Bi-LSTM model exhibits the highest  $R^2$  value and the lowest RMSE value, indicating its superior predictive capability compared to other models in Jiangsu province. "SEIQRDP1" signifies single-wave infectious disease prediction, while "SEIQRDP2" represents two waves of infectious disease predictions. In theory, multi-wave prediction should yield a smaller RMSE value than single-wave prediction. However, in this case, the second wave of the epidemic in Jiangsu province has not yet fully developed, with only sporadic cases. If the second wave intensifies or leads to a significant increase in infected individuals, the multi-wave prediction results of the SEIQRDP model are anticipated to improve.

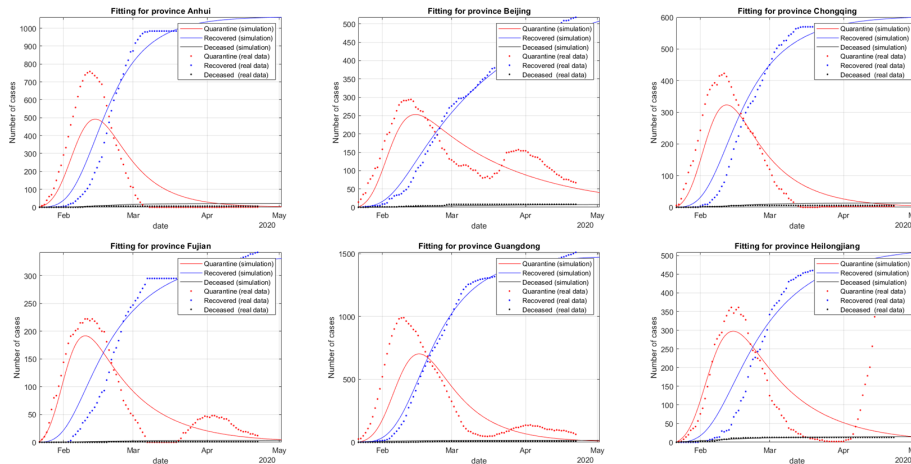
#### 4. Conclusion

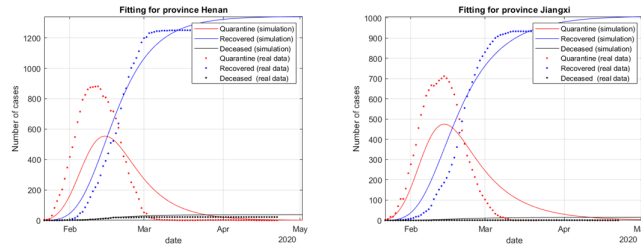
In our study on the spread of COVID-19, we initially developed and examined the SEIQHRDP infectious disease dynamics model, incorporating various compartments. We calculated the model's basic reproduction number using the next generation matrix method and demonstrated its stability. Subsequently, we conducted an analysis of the fluctuations in COVID-19 infection rates, considering three key factors: the protection rate, virus infection rate, and isolation rate. Our findings indicate that implementing increased self-protection measures (such as mask-wearing, maintaining physical distancing, and government-imposed restrictions) at the onset of the epidemic can effectively curb the spread of COVID-19. Additionally, it is imperative to reduce contact rates, which entails implementing stricter government controls.

Regarding infectious disease prediction, our approach commences with the fitting of actual data to the SEIQHRDP model, thereby obtaining the relevant parameter estimates. Subsequently, we conduct a comparative assessment of predictive performance between the kinetic model, ARIMA, and Bi-LSTM models. Our findings reveal that as the dataset for training expands, both the Bi-LSTM and ARIMA models exhibit superior predictive capabilities when contrasted with the kinetic model. Particularly, the neural network model, owing to its intricacy, demonstrates enhanced predictive accuracy in forecasting the number of infected individuals compared to the ARIMA model.

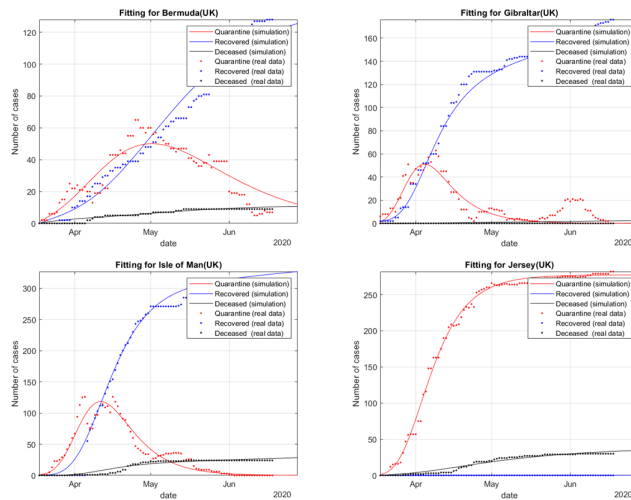
### Appendix

#### 1. Appendix A





**Fig. A1.** The eight figures presented above illustrate the parameter fitting outcomes for eight distinct provinces in China. In these figures, "simulate" corresponds to the model-generated results, while the "real data" is sourced from Johns Hopkins University (Dong et al., 2020). For a comprehensive analysis of these fitting results, please refer to Section 2.3 (Numerical Simulation).



**Fig. A2.** The four figures presented above depict the parameter fitting outcomes for four locations in the UK, with "simulate" representing the model-generated results. The dataset used is sourced from Johns Hopkins University (Dong et al., 2020). Comparing these results with the fitting outcomes for the eight provinces in China (Appendix A), it becomes evident that the SEIQRDP model performs better in fitting the data from the UK. This difference could potentially be attributed to variations in quarantine practices between the two regions.

## 2. Appendix B

Location	$\alpha$	$\beta$	$\gamma$	$\delta_1$	$\delta_2$	$\lambda_1$	$\lambda_2$	$\kappa_1$	$\kappa_2$	$R_0$
Anhui(CN)	0.077578	1	0.999984	0.323654	0.000106	0.069738	0.0001	0.00137	0.000101	3.088707685
Beijing(CN)	0.085329	0.999999	0.999992	0.227907	0.141219	0.024835	0.020473	0.000503	0.0001	2.709099332
Chongqing(CN)	0.086198	1	1	0.357595	0.0001	0.056781	0.0001	0.001259	0.000104	2.795677882
Fujian(CN)	0.10275	1	0.999945	0.311163	0.056692	0.048186	0.001882	0.000465	0.0001	2.718462438
Guangdong(CN)	0.073676	0.999997	0.999847	0.307476	0.0001	0.062245	0.000103	0.000381	0.020062	3.251219211
Heilongjiang(CN)	0.08186	1	1	0.200023	0.130433	0.04309	0.000474	0.001195	0.0001	3.026121481
Henan(CN)	0.074324	0.999988	0.999976	0.312456	0.000101	0.083289	0.000164	0.002315	0.0001	3.199378033
Jiangsu(CN)	0.086861	1	0.999996	0.341768	0.000101	0.064392	0.0001	0.00057	0.0001	2.92509704
Jiangxi(CN)	0.0785	1	1	0.3262	0.0001	0.0678	0.0001	0.001	0.0001	3.064664419
Sichuan(CN)	0.084263	1	0.999994	0.303308	0.050794	0.049651	0.000812	0.000605	0.0001	2.824045049
Bermuda(UK)	0.1697	0.369	0.1206	0.0912	0.0001	0.0384	0.0581	0.0023	0.0308	4.04162103
Gibraltar(UK)	0.0951	1	0.9268	0.1099	0.2713	0.0889	0.0021	0.0003	0.0001	2.623294858
Isle of Man(UK)	0.0894	1	1	0.0466	0.1889	0.0859	0.0003	0.0052	0.0001	4.246284501
Jersey(UK)	0.1697	1	0.3717	0.1847	0.0092	0.0009	0.0001	0.0002	0.0249	5.157297576

The table provided above presents the fitted parameter values acquired through the methodology elucidated in Section 4.3. The initial ten rows display the parameter fitting outcomes for ten provinces in China, while the final four rows depict the fitting results for the four local regions in the UK, serving as the control group. For a detailed explanation of these parameters, please refer to Section 2.1 Dynamic Model. Upon comparing the calculated basic reproductive numbers, it becomes evident that, on average, the basic reproductive number in the United Kingdom surpasses that of China. Consequently, the virus exhibited a higher degree of spread in the UK compared to China.

## References

- Bozdogan, H. (1987). *Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions*. Psychometrika, **52(3)**, 345–370.
- Chicco, D., Warrens, M. J. and Jurman, G. (2021). *The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation*. PeerJ Computer Science, **7**, e623.
- Dong, E., Du, H. and Gardner, L. (2020). *An interactive web-based dashboard to track COVID-19 in real time*. The Lancet infectious diseases, **20(5)**, 533–534.
- Gavin, H. P. (2019). *The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems*. Department of Civil and Environmental Engineering, Duke University, 19.
- Gubar, E., Taynitskiy, V., Fedyanin, D. and Petrov, I. (2023). *Quarantine and Vaccination in Hierarchical Epidemic Model*. Mathematics, **11(6)**, 1450.
- Gubar, E., Taynitskiy, V. and Zhu, Q. (2018). *Optimal control of heterogeneous mutating viruses*. Games, **9(4)**, 103.
- Ho, S. L. and Xie, M. (1998). *The use of ARIMA models for reliability forecasting and analysis*. Computers and industrial engineering, **35(1–2)**, 213–216.
- Jia, W. P., Han, K., Song, Y., Cao, W. Z., Wang, S. S., Yang, S. S., Wang, J. W., Kou, F. Y., Tai, P. G. and Li, J. (2020). *Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China*. Frontiers in medicine, **7**, 169.
- Kermack, W. O. and Mckendrick, A. G. (1927). *A contribution to the mathematical theory of epidemics*. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, **115(772)**, 700–721.
- Porta, M. (2014). *A dictionary of epidemiology*. Oxford university press.s
- Pastor, S. R., Castellano, C., Van, M. P. and Vespignani, A. (2015). *Epidemic processes in complex networks*. Reviews of modern physics, **87(3)**, 925.
- Shahid, F., Zameer, A. and Muneeb, M. (2020). *Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM*. Chaos, Solitons and Fractals, **140**, 110212.
- Sun, H. C., Liu, X. F., Du, Z. W., Xu, X. K. and Wu, Y. (2021). *Optimal control of heterogeneous mutating viruses*. IEEE Transactions on Computational Social Systems, **8(6)**, 1302–1310.
- Van den, D. P., and Watmough, J. (2002). *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*. Mathematical biosciences, **180(1–2)**, 29–48.
- Wonham, M. J. and Lewis, M. A. (2008). *A comparative analysis of models for West Nile virus*. Mathematical epidemiology, 365–390.