Research Article

A Method for Parameters Estimation in a Dynamical Model of Ebola Virus Transmission in Sierra Leone

Li Li,1,2 Li-Xia Du,1 Ziheng Yan,3 Jie Zhang,1 and Yong-Ping Wu 4

1School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China
2Science and Technology on Electronic Test and Measurement Laboratory, North University of China, Taiyuan 030051, China
3Chang’an University, Middle Section of Nan’er Huan Road, Xi’an, Shannxi 710064, China
4College of Physics Science and Technology, Yangzhou University, Yangzhou, Jiangsu 225002, China

Correspondence should be addressed to Yong-Ping Wu; ypwu@yzu.edu.cn

Received 24 July 2020; Revised 27 August 2020; Accepted 10 September 2020; Published 18 September 2020

Academic Editor: Xinzhu Meng

Ebola is an infectious virus that causes Ebola hemorrhagic fever in primates and humans, which was first found in 1976. The Ebola virus outbreak in West Africa in 2014 was the largest ever. A lot of researchers use mathematical models to analyze the characteristics of infectious diseases. However, many parameters in the model cannot be estimated completely. To ease the difficulty, we proposed an approach to estimate the parameter based on genetic algorithm (GA). GA uses the natural selection method of the fittest to find the optimal solution of the model. The least residual squares sum is used as fitness function to measure the performance of GA in parameter estimation. Moreover, we used a dynamical model and the real data of Ebola in Sierra Leone to verify the validity of GA. The experimental results indicate that the GA has strong competitiveness compared with the classical method, and it is a feasible method for estimating the parameters of infectious disease models.

1. Introduction

Ebola virus belongs to the family Filoviridae and is considered a prototype pathogen of viral hemorrhagic fever [1]. The virus was first detected in the Ebola river basin in southern Sudan and Congo in 1976 [2–9]. Since the discovery of Ebola virus, only four species of this virus cause human disease, namely, Zaire ebolavirus, Tai Forest ebolavirus, Sudan ebolavirus, and Bundibugyo ebolavirus [10]. The Reston virus causes only animal disease, not human disease. Therefore, the source of the Ebola virus is unknown. Researchers found evidence of asymptomatic infection of Ebola virus in three species of the fruit bats, which suggested that the bats are most likely to be the source of the deadly virus [11]. The bats could carry Ebola virus to other animals and even humans [12–14].

Ebola virus is transmitted through the saliva, the urine, and other body fluids [15, 16]. People can cause infection by direct contact with body fluids which carry the virus, with the virus entering the body through the nose, the mouth, the eyes, and the damaged skin [17]. Humans become infected after contact with the blood, the body fluids, and the infected fruit bats, as well as through the sexual contact [18].

Since there were no good treatments and approved vaccines at the time, the management of Ebola virus was limited to the use of obstacles and palliative care to suppress transmission [19]. A large-scale Ebola outbreak occurred in West Africa in 2014, mainly in Guinea, Liberia, and Sierra Leone. The number of confirmed cases is far greater than that in the past [10]. The lack of effective preventive measures at the time resulted in more people being infected with the Ebola virus [20]. In [21], the authors investigated the effectiveness of small interfering RNAs treatments for Ebola-infected patients. RNA interference can suppress the expression of viral genes; thus it is effective in suppressing Ebola virus replication, and the authors developed monoclonal antibodies against Ebola glycoprotein for the treatment of Ebola-virus-infected people [22]. In addition, some researchers used Sierra Leone’s disease data to study mathematical models of Ebola virus, predict the progress of
the epidemic, and propose preventive control measures and recommendations [5, 23–25].

Research on infectious diseases using dynamic models has become one of the important methods [26–32]. The propagation coefficient of the disease in the model affects the prediction results directly, and, consequently, it is important to estimate the propagation coefficient correctly. Classical parameter estimation methods are the Markov Chain Monte Carlo (MCMC) method, the least-squares method, and so on. The basic principle of the MCMC method is to construct a Markov chain by using the joint posterior probability distribution of the model propagation coefficients and assign any initial value to the simulation until it converges to a stable distribution. This determines the propagation coefficient [33–35]. There are many improved MCMC methods, such as using sequential Monte Carlo (SMC) filter techniques to estimate the propagation coefficients in the model [36]. However, firstly, the traditional method is limited by the calculation cost of the high-dimensional nonlinear model, which may take a lot of calculation time; it is usually not easy to obtain high-precision results, and it is not possible to get all the propagation coefficients at once [37, 38]. Secondly, the numerical estimation of the marginal probability distribution is difficult to achieve in the high-dimensional inversion model [39]. The least-squares method is performed by convolving the simulated data with the real data [40]. Although it has low calculation cost and generality, it does not consider the uncertainty of the inverse problem solution, and the initial value of the propagation coefficient will affect the efficiency of the algorithm. The least-squares method has certain flaws in determining the initial value. If it is set close to the optimal propagation coefficient, the result will be obtained quickly. If it is set far from the optimal propagation coefficient, it will increase the time of the algorithm [41].

In this paper, we present a method to solve inverse problems of differential equations based on GA. The GA is a method that is widely used in parameter estimation and other fields, and it has been proven to be a reliable method for estimating parameters based on nonlinear functions [42, 43]. It has a powerful adaptive search technology and uses the natural selection method of the fittest to simulate the evolution process, and thereby it can effectively solve the optimization problem [44]. When searching in high-dimensional models, GAs are superior to the other traditional search techniques due to their simplicity, effectiveness, versatility, and robustness [45, 46]. We have used the GA of adaptive mutation operator to estimate the parameters of differential equations. The advantage of this method is that the parameters in the high-dimensional model can be completely estimated by a small amount of data and all parameters combinations can be quickly obtained in a limited evolution process. In addition, an effective combination of multiple propagation coefficients can be obtained by GA for reference in studying the propagation dynamics.

The remainder of this paper is organized as follows. In Section 2, we introduced the transmission dynamical model of Ebola in Sierra Leone and the theories and processes of GA. In Section 3, we estimated the values of the parameters in a dynamical model based on GA. What is more, we validate the accuracy of the experiment results. Finally, we give the discussion and conclusions in Section 4.

\[
\begin{align*}
\frac{dS}{dt} &= \left( \beta_1 \frac{SI}{N} + \beta_2 \frac{SI_p}{N} + \beta_{II} \frac{SH}{N} + \beta_f \frac{SF}{N} \right) + \lambda_1 p IS, \\
\frac{dE}{dt} &= \beta_1 \frac{SI}{N} + \beta_2 \frac{SI_p}{N} + \beta_{II} \frac{SH}{N} + \beta_f \frac{SF}{N} - m_1 \eta E - m_2 (1 - \eta) E, \\
\frac{dI_S}{dt} &= m_3 (1 - \eta) E - \lambda_1 p IS - \lambda_2 q (1 - p) I_S - \lambda_3 (1 - q) (1 - p) I_S, \\
\frac{dI_p}{dt} &= \lambda_3 (1 - q) (1 - p) I_S + m_1 \eta E - r_h \theta I_p - \delta_1 (1 - \theta) r_d I_p - r_f(1 - \theta)(1 - \delta_1) I_p, \\
\frac{dH}{dt} &= r_h \theta I_p - r_{dh} \delta_2 H - r_{ih}(1 - \delta_2) H, \\
\frac{dF}{dt} &= \lambda_2 q (1 - p) I_S + \delta_1 (1 - \theta) r_d I_p + r_{dh} \delta_2 H - r_f F, \\
\frac{dR}{dt} &= r_f(1 - \theta)(1 - \delta_1) I_p + r_{ih}(1 - \delta_2) H + r_f F.
\end{align*}
\]
2. Main Method

2.1. A Dynamical Model of Ebola Virus Transmission in Sierra Leone. The time series of Ebola-confirmed case reports were collected from the World Health Organization (WHO) and the Ministry of Health of Sierra Leone. The data include the Ebola outbreak in 14 regions of Sierra Leone, including the suspected cases ($I_s$), the probable cases ($I_p$), and the hospital-confirmed cases ($H$), which are thought to represent the best available data of the Ebola epidemic. Due to the fact that hospitalization is a result of real infections, while suspected and probable cases may not be completely converted into hospitalized cases, it is more accurate to use hospitalization cases to indicate the actual number of the Ebola infections. We collected the newly infected cases for 34 weeks from May 19th, 2014, to January 11th, 2015. More detailed data can be found in [47].

We used a GA based on adaptive mutation operator. Think of the propagation coefficient of the Ebola virus model as a genetic target. It is binary-encoded, and then the genetic operators of random selection with elitism, multipoint crossover, and gene site mutation are used to simulate evolution to find the optimal solution. Set evaluation index for parameter genetic process as fitness function which is an evolution to find the optimal solution. Set evaluation index for parameter genetic process as fitness function which is an evolution to find the optimal solution. Set evaluation index for parameter genetic process as fitness function which is an evolution to find the optimal solution. Set evaluation index for parameter genetic process as fitness function which is an evolution to find the optimal solution.

Based on other literature analyses, this article divides Ebola virus transmission into seven categories, namely, susceptible ($S$), exposed ($E$), suspected individuals may be misdiagnosed ($I_s$), probable individuals ($I_p$), hospital-confirmed cases ($H$), and removed ($R$) [6]. Figure 1 depicted the transmission mechanism of the Ebola virus.

Consequently, we have used the following system of equation (1) to simulate the transmission dynamics of the Ebola virus in Sierra Leone and the biological meanings of parameters can be obtained in Table 1. We quantified the uncertainty of parameter estimates, and we give the 95% confidence intervals in Table 1.

\[
2^{13} < \frac{\text{range}_{\text{max}} - \text{range}_{\text{min}}}{\text{accuracy}} = \frac{1 - 0}{0.0001} = 10000 < 2^{14}. \tag{2}
\]

Therefore, a parameter can be represented by 14 bits of binary, and par consisting of all parameters needs to be encoded with $14 \times 21 = 294$ bits of binary. We collected the disease status of Ebola for 34 weeks, and the detailed data are shown above, so we set the initial population of parameters to 34 parameters set. Afterwards, we need to determine the fitness function; we take the minimum residual sum of squares between the solution of the infected case in the model and the actual infected case:

\[
\text{fitness} = \sum_{t=1}^{34} (\bar{I}(t) - I(t))^2, \quad t = 1, 2, \ldots, 34, \tag{3}
\]

where $\bar{I}(t)$ represents the infected cases in the model at time $t$ and $I(t)$ represents the actual infected cases at time $t$. We use fitness function to evaluate the initial population and give the initial fitness value of parameters.

GA mainly includes three genetic operators: selection, crossover, and mutation. Selection is to apply the selection operator to the group. The purpose of selection is to inherit the optimized parameters directly to the next generation or to generate a new par to the next generation through pairing and crossover. Selection operations are based on the fitness evaluation of parameters in the population. Here we adopt a random selection combined with elitism, which means that we will copy the parameters with higher fitness once, replace the ones with the least fitness, and retain the best parameters of each generation. This is the elite strategy; then we randomly select parameters for crossover and mutation.

Crossover operator plays a key role in GA. It is mainly divided into single-point crossover, two-point crossover, and multiple-point crossover. The commonly used is the single-point crossover; that is, a cross point is randomly set in the parameter string, when the intersection is performed, in front of this point or partial structures of two parameters sets exchange and thus two new parameters sets are generated. An example of a single-point crossover is shown in Figure 2. Because our parameter binary is too long, single-point crossover cannot meet our needs, so we chose multiple-point crossover to increase the diversity of parameters.

The mutation operator is to change certain gene positions of parameter strings in the population; for example, 0 becomes 1, and 1 becomes 0. An example of mutation is shown in Figure 3. We adopted the gene locus mutation with adaptive mutation operator in GA; that is, each gene position was mutated with a certain mutation probability. Moreover, the mutation probability can be adaptively adjusted by the parameters set fitness. When the difference between parameters set fitness and the average fitness of the population is small, it means that parameters are close to each other, which is not conducive to the next crossover. Therefore, it is necessary to increase the mutation probability and reduce the mutation probability when the difference is large, because the mutation probability is usually between 0.001 and 0.1, and the probability of mutation is small, so that it is not easy to destroy the genes of the dominant parameters, and it can jump out when the algorithm falls into the local optimal.
\[
\lambda_3 p I_s \\
\beta_1 (S I_s / N) + \beta_2 (S I_p / N) + \beta_H (S H / N) + \beta_F (S F / N) \\
m_1 (1 - \eta) E \\
\lambda_2 q (1 - p) I_s \\
r_{dH} (1 - \delta_2) H \\
\delta_1 (1 - \theta) r_{dP} \\
r_f F \\
\delta_2 (1 - \theta) (1 - \delta_1) I_p \\
r_I (1 - \delta_1) I_s \\
m_2 (1 - \eta) E \\
r_H \theta I_p \\
\delta_1 (1 - \theta) r_d P \\
r_f F \\
\beta_1 (S I_s / N) + \beta_2 (S I_p / N) + \beta_H (S H / N) + \beta_F (S F / N)
\]

**Figure 1:** Ebola’s transmission mechanism in Sierra Leone.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>Transmission coefficient between communities S and $I_s$</td>
<td>0.9162</td>
<td>[0.1173, 1.1247]</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Transmission coefficient between communities S and $I_p$</td>
<td>0.8929</td>
<td>[0.1341, 1.1803]</td>
</tr>
<tr>
<td>$\beta_H$</td>
<td>Transmission coefficient at the hospital</td>
<td>0.1674</td>
<td>[0.0623, 1.0713]</td>
</tr>
<tr>
<td>$\beta_F$</td>
<td>Transmission coefficient during the funerals</td>
<td>0.2736</td>
<td>[0.0887, 1.0665]</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Misdiagnosed proportion in $I_s$</td>
<td>0.657</td>
<td>[0.1638, 0.6891]</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>Proportion of correctly diagnosed suspected cases that move to $F$</td>
<td>0.1447</td>
<td>[0.0531, 1.1805]</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Proportion of exposed cases entering the $I_p$ compartment</td>
<td>0.619</td>
<td>[0.1008, 1.1189]</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>Time from suspected to susceptible</td>
<td>1.59 days</td>
<td>[0.0641, 1.002]</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>Average time from suspected to death</td>
<td>4.88 days</td>
<td>[0.0418, 1.0139]</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>Time from suspected to probable</td>
<td>2.38 days</td>
<td>[0.0661, 1.0172]</td>
</tr>
<tr>
<td>$m_1$</td>
<td>Time from exposed to probable</td>
<td>1.0007 days</td>
<td>[0.0611, 1.2091]</td>
</tr>
<tr>
<td>$m_2$</td>
<td>Time from exposed to suspected</td>
<td>20.9 days</td>
<td>[0.1053, 1.2612]</td>
</tr>
<tr>
<td>$r_h$</td>
<td>Time from probable to hospitalized</td>
<td>3.56 days</td>
<td>[0.2733, 0.5884]</td>
</tr>
<tr>
<td>$r_{dh}$</td>
<td>Average time from hospitalization to death</td>
<td>1.2 days</td>
<td>[0.1109, 1.187]</td>
</tr>
<tr>
<td>$r_i$</td>
<td>Time from probable to removed</td>
<td>5.69 days</td>
<td>[0.1221, 1.1318]</td>
</tr>
<tr>
<td>$r_{ih}$</td>
<td>Time from hospitalized to removed</td>
<td>2.56 days</td>
<td>[0.0844, 1.1501]</td>
</tr>
<tr>
<td>$r_f$</td>
<td>Average time from death to burial</td>
<td>1.21 days</td>
<td>[0.1323, 0.9899]</td>
</tr>
<tr>
<td>$r_d$</td>
<td>Time from probable to death</td>
<td>1.11 days</td>
<td>[0.0099, 1.183]</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>Mortality rate from probable to death</td>
<td>0.5141</td>
<td>[0.1219, 1.0691]</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>Mortality rate from hospitalized to death</td>
<td>0.6214</td>
<td>[0.0455, 1.1318]</td>
</tr>
<tr>
<td>$N$</td>
<td>Size of the Sierra Leone population</td>
<td>6348350</td>
<td>[47]</td>
</tr>
</tbody>
</table>

**Figure 2:** The two parameters sets A and B exchange the part after the cross point and reconstitute two new parameters sets.
solution. The mutation characteristic of GA can make the solution process randomly search the entire space where the solution may exist and ensure the diversity of the population, so the global optimal solution can be obtained to a certain extent.

Next, we solve model (1), decode the parameters into decimalism, and substitute them into the model solution to get the estimated value of confirmed-infection cases, that is, hospitalization cases. Fitness function (3) is used to fit the real infected cases. Then we find the parameters with the highest fitness in this generation, that is, the parameters with the least error, and keep them in the next-generation genetic process. Afterwards, we cyclically execute selection, crossover, mutation, and evaluation of new parameters set until the maximum number of iterations is reached.

The general steps of using GA to estimate parameters of Ebola model are explained as follows:

Step 1: consider the parameter to be estimated as a gene chromosome, define the parameter using binary coding, and then initialize the population.

Step 2: assign a fitness value to each parameter using equation (3). Starting from the second generation of parameters, parameters are ranked from small to large according to fitness values, and the first two parameters sets with the greatest fitness are duplicated once to replace the two parameters sets with the smallest fitness.

Step 3 (selection process): add randomly initialized parameters to the population to increase population diversity, and then two parameters sets are randomly selected as paternal parameters sets.

Step 4 (crossover): two new offspring parameters are generated by crossing the two parents at multiple points.

Step 5 (mutation): use the gene locus mutation described above in combination with adaptive mutation operator.

Step 6: solve equation (1) using the ode function.

Step 7: convert the types of parameters sets from binary to decimal, substitute it into the solution of equation (1), obtain the predicted value of the diseased cases, and use equation (3) to evaluate the fitness value of the new parameters set to obtain an optimal parameters set.

Step 8: when the fitness of the new offspring produced by genetic manipulation is higher than that of the parent, the new parameters sets replace the parents and are inserted into the parent population for the next genetic manipulation. If the optimal individual remains unchanged for 30 consecutive times, multiple-point crossover and mutation are carried out.

Step 9: save the best parameters set of this generation.

Step 10: if the number of iterations is not reached, proceed with step 2.

The above steps are executed iteratively until the termination condition is reached. Parameter estimation is finished whenever the genetic operation completed. The parameter values to be used in the model are the optimal solution of the last generation of parameters, and the specific values are shown in Table 1. We can study the propagation dynamics and preventive measures of Ebola.

3. Main Results

There are many methods for parameter estimation, including Markov Chain Monte Carlo (MCMC) method and least-squares method. It is not easy to analyze and estimate the parameters in the infectious disease model because there are many parameters in the model which cannot be estimated fully. Consequently, we propose to estimate the parameters in model (1) using GA, which are described in Section 2. Algorithm 1 shows the scheme of GA used on parameter estimation of the Ebola model. In this algorithm, par denotes a set of parameters, constants $a$ and $b$ represent two parameters sets randomly selected from par set and crossed according to the crossover probability $p_c$ to get $a'$ and $b'$. In mutation, $a'$ and $b'$ are changed to $\bar{a}$ and $\bar{b}$ by gene locus mutations, according to mutation probability $p_m$. IC denotes the initial value of each variable in equation (1) and $t$ indicates the time of virus transmission. “count” indicates that successive generations of optimal values have not changed.

In this study, we used different genetic operators and fitness functions (3) to conduct data fitting for the real hospital-diagnosed cases. We conducted 70 experiments and selected a set of parameters which performed well. The results are listed in Table 1. After experimental verification, we chose the crossover probability $p_c$ to be 0.8 and the initial mutation probability $p_m$ to be 0.01. The genetic algebra Constant is 3000 times. The fitting result for the cumulative number of cases is shown in Figure 4.

Figure 5 represents the evolutionary process of the optimal value of each generation in GA. With the increase of genetic algebra, the error between the model solution and the real data is gradually decreasing, which means that the fitness of the model is increasing. Until the maximum genetic algebra is reached, a set of near-optimal parameters are obtained. Since the error persists, we can regard the sub-optimal solution as the optimal solution. The subgraph in Figure 5 is an enlarged view of early inheritance. Although this set of parameters performs well and converges quickly, we can see from the figure that it converges around 200 generations, but, because of the instability of GA, sometimes it takes a long time to converge to the optimal value. Therefore, we unified the genetic algebra to 3000 generations during the experiment. It can be seen from the figure that the
GA can quickly reduce the error in the early stage, and the convergence speed is very fast, which shows the effectiveness of our algorithm.

In [5], the parameters $\beta_1 = 0.0498$, $\beta_H = 0.0225$, and $\beta_F = 0.0013$ are given, and, in [23], the parameters $\beta_I = 0.128, \beta_H = 0.08, \beta_F = 0.111$, and $\delta_1 = \delta_2 = 0.75$ are given.

**Algorithm 1: Parameter estimation of GA.**

```
(1) Parameter length $C_L$ is determined according to parameter range.
(2) Each parameter is represented by a chromosome of $C_L$ length to get a complete parameters set $\text{par} = [\beta_1, \beta_2, \ldots, r_f]$. 
(3) Initialize all parameters set, $\text{par}_c = [\beta_1, \beta_2, \ldots, r_f]$, $c = 34$.
(4) Evaluate the fitness of each $\text{par}$ using $f$.
(5) for $i = 1, 2, \ldots, $, Constant do
(6) if $i = 2$ do
(7) Sort $\text{par}$ allocation fitness in an ascending order.
(8) $\text{par}_{34} = [\beta_1, \beta_2, \ldots, r_f] = \text{par}_i = [\beta_1, \beta_2, \ldots, r_f]$.
(9) $\text{par}_{33} = [\beta_1, \beta_2, \ldots, r_f] = \text{par}_2 = [\beta_1, \beta_2, \ldots, r_f]$.
(10) end if
(11) Randomly add an initialization parameter to replace one of the $\text{par}$
(12) if max $f(\text{par})$ close to average $f(\text{par})$ do
(13) Mutation probability $p_m = 0.1$.
(14) end if
(15) Randomly select $a, b \in \text{par}$ ($a \neq b$).
(16) Cross over $a$ and $b$ according to $p_x, a, b \rightarrow a', b'$.
(17) Perform mutation on $a'$ and $b'$ according to $p_m, a', b' \rightarrow \tilde{a}, \tilde{b}$.
(18) $D = \text{ebolaSol(}\text{par}, IC, t)$. 
(19) Evaluate fitness $f = \text{sum}(D - \text{Data})^2$.
(20) if $f(\tilde{a}) < f(a)$ do
(21) $\text{par}(a) = \text{par}(\tilde{a})$.
(22) end if
(23) if count > 30 do
(24) Multiple crossover and variation.
(25) end if
(26) Find the best parameters set in this generation $\text{par}_p(i)$ that satisfies $\text{best}_p(i) = \text{min} f(\text{par})$.
(27) end for
(28) Find the best parameters set $\text{par}_{best}$ that satisfies $\text{par}_{best} = \text{best}_p(\text{Constant})$.
```

**Figure 4:** The fitting result for actual data of Ebola virus disease in Sierra Leone for 34 weeks with the deterministic model (1).

**Figure 5:** Evolutionary process of model optimal solution. Here, the yellow solid line represents the evolutionary process of the fitness values corresponding to the optimal parameters of each generation, while the blue solid line represents the early amplification part of the model solution.

Meanwhile, the parameters $\beta_F = 0.489, \delta_1 = 0.8, \delta_2 = 0.4$ are given in [6]. Although some of the propagation coefficients are very different from those in other papers, because the deterministic models are different, we have that cabins are
more, and we cannot directly compare them with those in other papers. We are uncertain about patients in the exposed period, so the propagation coefficient associated with it is uncertain. We only use GA to solve each parameter value. Thus, we only perform comparison with some important parameters. In this paper, $\delta_1 = 0.5141$, $\delta_2 = 0.6214$ are close to the above literature to some extent. We obtain the basic reproduction number of Ebola virus in Sierra Leone $R_0 \approx 1.895$ which is calculated in the same way as in Xia et al.’s work [6]; it is basically consistent with $R_0 \approx 1.7$ given in [5] and $R_0 \approx 1.78$ given in [23]. The results show that the GA can accurately estimate all the parameters in the model, and the data are fitted well. Another advantage of the GA is that you are free to set the parameters precision, but you need to consider the length of the parameters set, which is very useful for getting high-precision parameters. It can be seen that GA can be used as a feasible method for parameter estimation.

Due to the fact that we are using a certain mathematical model, we only need to use the actual diseased data as test data and apply them in model (1) to get a set of near-optimal parameters. We set each of the parameter ranges to $[0,1]$, randomly generate the initial parameters as the input of the model, and use the GA to obtain a set of parameters that fit the model better. We can use GA to give a variety of parameter combinations for researchers’ reference. However, some parameters may have overfitting problems, so we conducted 70 experiments to select a set of parameters that are more realistic. We calculated the 95% confidence intervals for the best set of parameters in the 70 experiments, and almost all parameters are within the confidence intervals, which also shows the validity of the GA for parameter estimation. The confidence intervals for all the parameters are shown in Table 1.

### 4. Discussion and Conclusion

This work proposed a GA parameter estimation method based on adaptive mutation operator, which could be applied to biomathematical models and differential equations in other fields. Through GA’s adaptive search parameters, the parameters in various models can be effectively found, and multiple parameter combination schemes can be given, which reduces the process of manual adjustment of parameters by researchers and provides an effective reference for scientific research. In addition, in the study of infectious diseases, due to the complexity of the model, all of the parameters are often difficult to obtain, and the basic reproduction number $R_0$ for evaluating whether there is an outbreak of an infectious disease needs to be obtained by calculating parameters; and the parameters can indicate the transmission dynamics of infectious diseases and the scale of transmission visually. Thus, parameter estimation is the most important, and the GA can be used to effectively find all the parameters in the model, which make up for the shortcomings of the traditional method, such as long calculation time and slow convergence speed. The method we have proposed has been evaluated in the experimental process, where the performance has reached the desired level. Finally, the GA can be applied not only to infectious disease models but also to other mathematical and physical models, and it proposes a new idea for parameter estimation.

Since the initial population of GA needs to assign values randomly, it will lead to output instability; we cannot quantify the uncertainty due to the discrete distribution of the output parameters, which is a common problem of GA and needs to be improved. On the other hand, fitness function is an important factor to determine the pros and cons of genetic evolution, so it is very important to select a suitable fitness function. Finally, a very important step in GA is to find the solution of the model, so it may not be suitable for the unsolvable equation, but it is applicable for most models. There are many improved genetic algorithms, and we can improve the existing algorithms and expand their scope of application.

The GA can solve specific problems with only a small amount of data, and the corresponding fitness function can be used for searching. Therefore, it is a general-purpose algorithm used in many fields. In function optimization, GA can estimate not only the parameters of biomathematical model but also the kinetic parameters of microorganisms. It can also solve the performance parameters of nonlinear physical problems. In addition, GA performs well in path planning, cloud computing task scheduling, communication network design, image feature extraction, and other fields. Furthermore, there are some studies that combine GA with machine learning methods such as neural network.

In this paper, we introduce the basic process of adaptive mutation genetic algorithm, introduce how to use GA to estimate the parameters of Ebola virus model in Sierra Leone, and give the data curve and genetic iterative process for fitting actual infections. In addition, we offer a new idea for parameter estimation in other research fields, such as dynamical model of disease transmission [48–50] or predator-prey interactions [51–53] with spatial effects [32, 54–56] in the form of reaction-diffusion equations.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This work was supported by Program for the Outstanding Innovative Teams (OIT) of Higher Learning Institutions of Shanxi, Natural Science Foundation of Shanxi Province (Grant no. 201801D221003), and China Postdoctoral Science Foundation (Grant no. 2017M621110).
References


[12] Centers for Disease Control and Prevention, Diseases and Conditions.


