Enhanced Knowledge in Sciences and Technology Vol. 3 No. 1 (2023) 118-127
© Universiti Tun Hussein Onn Malaysia Publisher's Office

EKST

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/ekst
e-ISSN : 2773-6385

Optimal Parameter Estimation of Epidemic Model for Coronavirus Disease Based on Gauss-Newton Computational Approach

Kai Wen Chua[1], Sie Long Kek[2*]

[1, 2] Department of Mathematics and Statistics,
Faculty of Applied Sciences and Technology,
Universiti Tun Hussein Onn Malaysia (Pagoh Campus),
84600 Pagoh, Muar, Johor, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/ekst.2023.03.01.014
Received 15 January 2023; Accepted 20 March 2023; Available online 3 August 2023

Abstract: Mathematical modelling of coronavirus disease (COVID-19) is eagerly required in public health to understand the evolution of COVID-19. However, unknown parameters in the model prevent the progress of the modelling. This report aims to study the spread of COVID-19 in Malaysia through the susceptible-infected-recovered (SIR) model. By using the real data of COVID-19 from 1 January to 30 September 2022 to estimate the model parameters, which are the transmission rate and recovery rate, a loss function was introduced. The Gauss-Newton recursion equation was derived and the value of parameters was updated iteratively until convergence was achieved. With these optimal parameter estimates, the SIR model was solved numerically. The model solution revealed that the spread of COVID-19 increased exponentially for the following 2 months. In addition, the prediction results for the coming 16 years show that the number of infected cases will reach a peak of 8.1493 million in 2.5 years. After 10 years, the spread of COVID-19 will stay at a total cumulative of 2.7794, 0.02757 and 29.857 million for susceptible, infected and recovered cases, respectively. In conclusion, parameter estimation in the SIR model is satisfactorily performed and prediction results of the spread of COVID-19 in Malaysia are clearly interpreted.

Keywords: Coronavirus Disease, Susceptible-Infected-Recovered Model, Parameter Estimation, Gauss-Newton Computational Approach

1. Introduction

Coronavirus disease 2019 (COVID-19), an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first discovered in December 2019 in Wuhan, China [1]. The disease spreads globally and became an outbreak. Thus, World Health Organization (WHO) declared COVID-19 a pandemic on March 11, 2020 [2]. The first case of COVID-19 in Malaysia was an imported case from China that was reported on 25 January 2020 [3]. Since then, the outbreak has become serious in Malaysia. To slow down the transmission of the disease, people have been urged to follow the standard operating procedures (SOPs) such as wearing masks, washing hands, using alcohol-

*Corresponding author: slkek@uthm.edu.my

based hand sanitizer, practicing social distancing, and being vaccinated. However, the virus can spread in small liquid particles from an infected person's mouth or nose into the air. Infected individuals with mild symptoms will recover without special treatment, while those are underlying medical conditions and elderly citizens are at a higher risk of developing serious illnesses and requiring medical attention. Hence, mathematical modelling in epidemics is required to analyse and predict the evolution of this outbreak [4].

In fact, epidemics modelling can be dated back to 1662 for the work done by John Graunt [5]. Daniel Bernoulli also created the earliest mathematical model to defeat the progress of smallpox in 1760 [6]. Later, epidemic modelling has evolved rapidly since the early 20th century and become a useful tool to study the spreading of disease, predict the future course of an outbreak and evaluate strategies to control the disease. The Susceptible-Infected-Recovered (SIR) model is a fundamental compartmental model that has been widely used in epidemic modelling. Previous studies showed that this model has been applied to calculate the infection parameter and reproduction number of COVID-19 as well as predict the expected number of infected cases [7, 8]. Although applying mathematical models to epidemics provides information for understanding the dynamics of infectious disease transmission, exact parameters in an existing SIR mathematical model, which are the transmission and recovery rates, are unknown. Their value can only be measured through collected observation data and modern big data technology [9].

This paper mainly focuses on mathematical modelling for the epidemics of COVID-19 in Malaysia using the SIR model. The real data on COVID-19 in Malaysia were obtained from the GitHub website (https://github.com/MoH-Malaysia/covid19-public) that is contributed by the Ministry of Health (MOH) Malaysia, and the daily reported numbers of confirmed, recovered and death cases are considered [10, 11,12]. The Gauss-Newton computational approach is applied to estimate parameters in the SIR model, namely transmission and recovered rates. For this aim, a least-square optimization problem, which minimizes the measurement of differences between the real data and the estimated output, is defined. Then, the Gauss-Newton recursion equation is derived. Through updating the parameters repeatedly, the optimal parameter estimates are obtained when convergence is achieved, and the loss function is minimized. With these parameter estimates, the SIR model is solved by using the Runge-Kutta fourth-order method. So, the numerical results are interpreted as the prediction results for the spread of COVID-19.

The rest of the paper is organized as follows. In Section 2, the problem of the general model to be solved is described, the Gauss-Newton computational approach is discussed, and the Runge-Kutta numerical method is expressed. In Section 3, the data visualization of COVID-19 is presented. The SIR model is discussed for modelling COVID-19. The optimal parameter estimates, which are the effective transmission rate and the recovered rate, are obtained using the Gauss-Newton computational approach. Prediction results of COVID-19 using the SIR model are interpreted. Finally, a conclusion is made.

2. Materials and Methods

Consider a general model for COVID-19 [13] given by

$$\dot{x}(t) = f(x(t), t, \theta), \qquad Eq.\ 1$$

where $x \in \Re^n$ is the state vector and t is time; while $\dot{x}$ is the rate of change of the state over time, and $f : \Re^n \times \Re \times \Re^m \to \Re^n$ is the function dynamics. Here, $\theta \in \Re^m$ is the unknown parameter in the model. Suppose the initial value of the state is

$$x(0) = x_0,$$

and the solution of the system (Eq.1) is assumed to be measured by the output equation

$$y(t) = h(x(t), t) \qquad Eq.\ 2$$

where $y \in \Re^p$ is the output vector and $h : \Re^n \times \Re \times \Re^m \to \Re^p$ is the measurement function. Because the state equation (Eq. 3) is nonlinear and complex, obtaining the solution of the state equation (Eq. 1) is difficult, especially with the unknown model parameter θ.

Therefore, we propose an estimated output model as follows,

$$\hat{y}(t, \theta) = h(x(t), t, \theta)\,. \qquad Eq.\ 3$$

Here, our aim is to solve the state equation (Eq. 1) given the parameter estimate $\hat{\theta}$. For doing so, the model parameter θ will be estimated iteratively so that the state trajectory $x(t)$ can be determined, in turn, to approximate the solution of the output measurement (Eq. 2). This problem is known as the parameter estimation problem for a nonlinear system based on the output residual and is referred to as Problem (*P*).

2.1 Gauss-Newton Computational Approach

To solve Problem (*P*), let us define an optimization problem [14] as follows,

$$\text{Minimize} \qquad J(\theta) = \frac{1}{2} r(\theta)^{\mathrm{T}} r(\theta)\,, \qquad Eq.\ 4$$

where *J* is the loss function, which represents the sum of square error, and $r(\theta)$ is the residual function, which is defined by

$$r(\theta) = y(t) - \hat{y}(t, \theta)\,. \qquad Eq.\ 5$$

The gradient of the loss function *J* is

$$\nabla J(\theta) = -\nabla r(\theta)^{\mathrm{T}} (y(t) - \hat{y}(t, \theta))\,, \qquad Eq.\ 6$$

and the Hessian is

$$\nabla^2 J(\theta) = -\nabla^2 r(\theta)^{\mathrm{T}} (y(t) - \hat{y}(t, \theta)) + \nabla r(\theta)^{\mathrm{T}} \nabla r(\theta)\,, \qquad Eq.\ 7$$

where $\nabla r(\theta)$ is the Jacobian matrix of the residual function $r(\theta)$.

Now, write the loss function as the second order Taylor series expansion,

$$J(\theta) \approx J(\theta^{(i)}) + \nabla J(\theta^{(i)})(\theta - \theta^{(i)}) + \frac{1}{2}(\theta - \theta^{(i)})^{\mathrm{T}} \nabla^2 J(\theta^{(i)})(\theta - \theta^{(i)})\,. \qquad Eq.\ 8$$

Taking the first-order necessary condition, we have

$$\nabla J(\theta) \approx \nabla J(\theta^{(i)}) + \nabla^2 J(\theta^{(i)})(\theta - \theta^{(i)}) = 0\,. \qquad Eq.\ 9$$

After rearranging (Eq. 9), the normal equation [15]

$$\nabla^2 J(\theta^{(i)})(\theta - \theta^{(i)}) = -\nabla J(\theta^{(i)})\,, \qquad Eq.\ 10$$

is obtained, and assuming the inverse of the Hessian exists, we have the following equation,

$$(\theta - \theta^{(i)}) = -(\nabla^2 J(\theta^{(i)}))^{-1} \nabla J(\theta^{(i)})\,. \qquad Eq.\ 11$$

By letting $\theta = \theta^{(i+1)}$, the updating equation

$$\theta^{(i+1)} - \theta^{(i)} = -(\nabla^2 J(\theta^{(i)}))^{-1} \nabla J(\theta^{(i)}) \qquad Eq.\ 12$$

is addressed. Then, substitute (Eq. 6) and (Eq. 7) into (Eq. 12) yielding

$$\theta^{(i+1)} - \theta^{(i)} = -(-\nabla^2 r(\theta^{(i)})^{\mathrm{T}} (y(t) - \hat{y}(t, \theta^{(i)})) + \nabla r(\theta^{(i)})^{\mathrm{T}} \nabla r(\theta^{(i)}))^{-1}$$

$$\times \nabla r(\theta^{(i)})^{\mathrm{T}} (y(t) - \hat{y}(t, \theta^{(i)})) . \qquad Eq.\ 13$$

For simplification, we ignore the first term at the right-side in (Eq. 13) to obtain the recursion equation,

$$\theta^{(i+1)} = \theta^{(i)} - (\nabla r(\theta^{(i)})^{\mathrm{T}} \nabla r(\theta^{(i)}))^{-1} \nabla r(\theta^{(i)})^{\mathrm{T}} (y(t) - \hat{y}(t, \theta^{(i)})) , \qquad Eq.\ 14$$

with the initial value $\theta^{(0)}$ is given. Hence, (Eq. 14) is known as the Gauss-Newton approach [16].

During the iteration, the model parameter $\theta^{(i)}$ is updated through using (Eq. 14). When the convergence is achieved, the optimal parameter estimate $\theta^* = \theta^{(i+1)} \approx \theta^{(i)}$ will minimize the loss function (Eq. 4) and the trajectory of the state equation (Eq. 1) can be approximated through numerical methods.

2.2 Runge-Kutta Numerical Method

The numerical solution of the model (Eq. 1) is computed using the Runge-Kutta fourth order (RK4) method,

$$x_{n+1} = x_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \qquad Eq.\ 15$$

$$t_{n+1} = t_n + h \qquad Eq.\ 16$$

for n = 0, 1, 2, 3, ..., with

$$k_1 = f(x_n, t_n, \theta^{(i)}), \qquad Eq.\ 17$$

$$k_2 = f\left(x_n + \frac{h}{2} k_1, t_n + \frac{h}{2}, \theta^{(i)} \right), \qquad Eq.\ 18$$

$$k_3 = f\left(x_n + \frac{h}{2} k_2, t_n + \frac{h}{2}, \theta^{(i)} \right), \qquad Eq.\ 19$$

$$k_4 = f(x_n + hk_3, t_n + h, \theta^{(i)}), \qquad Eq.\ 20$$

and the initial conditions t_0, y_0 are given. Here, $x_n = x(t_n)$, $h > 0$ is the step size, and the parameter estimate $\theta^{(i)}$ at the iteration number i is considered. Finally, the numerical solution of the model is obtained and interpreted as the prediction results for the spread of COVID-19.

3. Results and Discussion

From the real data on COVID-19 in Malaysia (https://github.com/MoH-Malaysia/covid19-public), the daily reported numbers of confirmed, recovered and death cases were considered from 1 January 2022 until 30 September 2022, which are 273 data for each daily number of confirmed, recovered and death cases in the cumulative form. Then, these data are used to estimate the unknown parameters in the SIR model, which are transmission and recovered rates, and the solution of the SIR model is applied to predict the spread of COVID-19 that may happen in the following 61 days. Also, the trend of COVID-19 spreading is interpreted. For the simulation, the GNU Octave is used.

3.1 The Classical SIR Model

Consider the classical SIR model given as follows.

$$\frac{dS(t)}{dt} = -\beta \frac{S(t)I(t)}{N} \qquad Eq.\ 21$$

$$\frac{dI(t)}{dt} = \beta \frac{S(t)I(t)}{N} - \gamma I(t) \qquad Eq.\ 22$$

$$\frac{dR(t)}{dt} = \gamma I(t) \qquad Eq.\ 23$$

where $S(t)$ is the number of susceptible people at time *t*, $I(t)$ is the number of people infected at time *t*, $R(t)$ is the number of people who have recovered or removed at time *t*, and *N* is the total population given by

$$N = S(t) + I(t) + R(t)\,. \qquad Eq.\ 24$$

In the classical SIR model, there are three classes, which are susceptible, infected and removed individuals. These three classes are also known as three compartments. The susceptible individuals, denoted by *S*, are the class of individuals who are healthy but can contract the disease. The infected individuals, denoted by *I*, are the class of individuals who have contracted the disease and are now sick with it. The recovered or removed individuals, denoted by *R*, are the class of individuals who have recovered and cannot contract the disease again. In this study, the total population is assumed as the total population of Malaysia.

According to [17], the relationships between compartment models presented are based on four fundamental assumptions. First, the vital dynamics are neglected and the size of the population is constant. Second, the population is assumed to be a homogeneous continuum, meaning that all people have an equal number of contacts, the probability of the transmission of the disease between a susceptible and an infectious people during their contact remains constant and the infected people are equally distributed among the population. Third, the rate of flow between compartments *I* and *R* is directly proportional to the size of compartment I. Fourth, the recovered people acquire immunity and cannot spread the infection. Those who fall victim to the disease are treated as recovered.

Here, β represents the effective transmission rate and γ represents the recovered rate. Using these parameters, the basic reproduction ratio

$$R_0 = \frac{\beta}{\gamma} \qquad Eq.\ 25$$

can be used to calculate the average number of new infections from a single infection. In particular, when $R_0 > 1$, the infection will spread in a population.

The dynamics of the SIR model in (Eq. 21- Eq. 23) are interpreted as follows. The rate of change in susceptible is negative because susceptible individuals will be infected when in contact with the infected individual. So, the number of susceptible individuals is decreased with time. While the rate of change in infected depends on differences between the number of susceptible individuals to be infected and the number of infected individuals who recovered from the disease. The transition from susceptible to infected increases the number of infected individuals, whereas the transition from infected to recovered decreases the number of infected individuals. On the other hand, the rate of change in recovery is positive since infected individuals who recovered from the disease increase the number of recovered individuals.

3.2 Parameter Estimation of SIR Model with COVID-19 Real Data

Using the real data of COVID-19 from the GitHub website (https://github.com/MoH-Malaysia/covid19-public), the model parameters, which are the effective transmission rate β and the recovered rate γ, were estimated through the Gauss-Newton computation approach. The result of the parameter estimation is shown in Table 1.

Table 1: Optimal parameter estimates

Model Parameter	Symbol	Value
Effective transmission rate	β	4.6945×10^{-3}
Recovered rate	γ	1.7021×10^{-3}

The transmission rate of $\beta = 4.6945\times10^{-3}$ indicates that the susceptible individuals are infected on an average of 4.6945×10^{-3} in the period from 1 January 2022 until 30 September 2022. While the recovered rate of $\gamma = 1.7021\times10^{-3}$ reveals that the individuals are recovered from the infection on an average of 1.7021×10^{-3} in the period mentioned. Hence, the SIR model with these parameter estimates is given as follows,

$$\frac{dS(t)}{dt} = -4.6945\times10^{-3}\frac{S(t)I(t)}{N} \qquad Eq.\ 26$$

$$\frac{dI(t)}{dt} = 4.6945\times10^{-3}\frac{S(t)I(t)}{N} - 1.7021\times10^{-3}I(t) \qquad Eq.\ 27$$

$$\frac{dR(t)}{dt} = 1.7021\times10^{-3}I(t) \qquad Eq.\ 28$$

Figure 1 shows the solution of the SIR model in (Eq. 26 - Eq. 28). This solution could predict closely the real data of COVID-19 with the root-mean-square error (RMSE) is 6448.3. In addition, Figure 2 shows the estimation error for the infected and recovered cases from 1 January 2022 until 30 September 2022. Here, our aim is to provide a full visualization between the estimated and actual values. Hence, the accuracy of the solution from the SIR model is satisfactorily accepted since the real data of COVID-19 involves large values. Moreover, the performance of the Gauss-Newton computational approach was verified by the RMSE value. The basic reproductive number R_0 is 2.7581. This value indicates that COVID-19 in Malaysia will highly spread in the following days and the number of infected cases will be increasing.

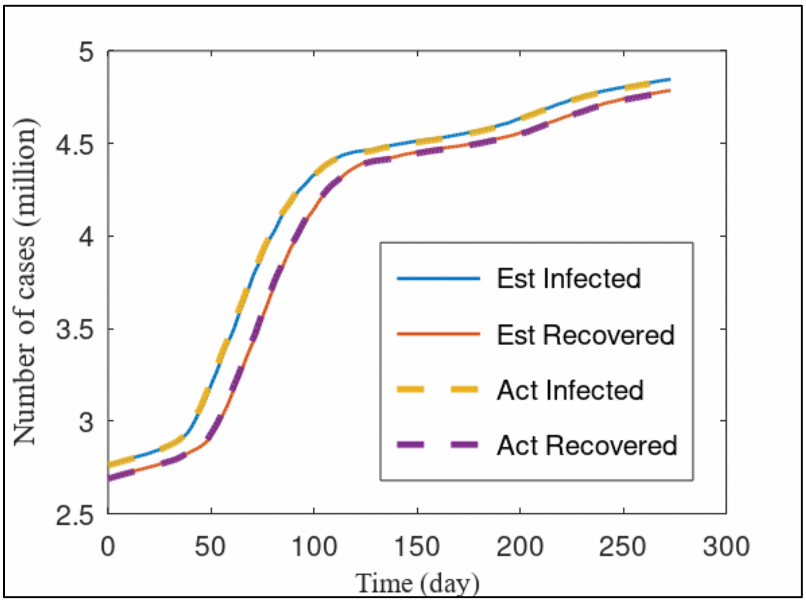

Figure 1: Prediction of infected and recovered cases, 01 Jan-30 Sep 2022

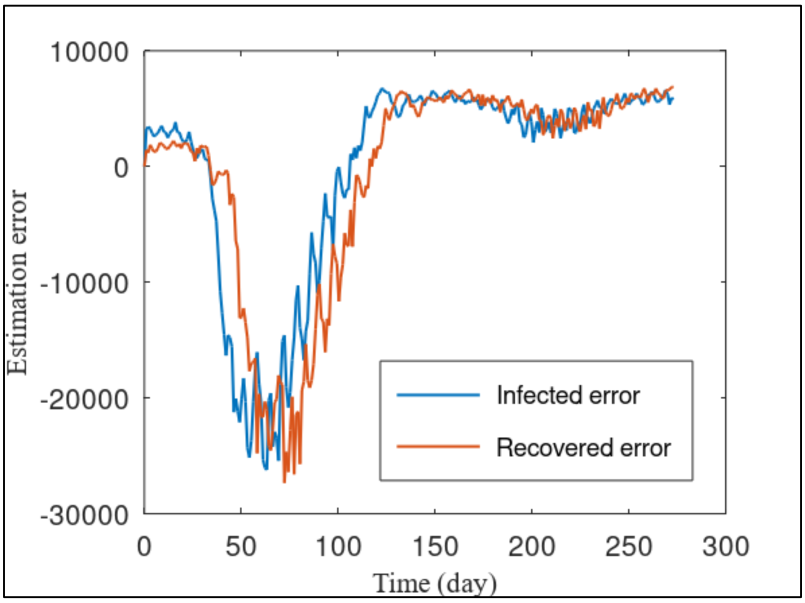

Figure 2: Estimation error of infected and recovered cases, 01 Jan-30 Sep 2022

3.3 Prediction of COVID-19 with SIR Model

Now, apply the SIR model with parameter estimates as given in (Eq. 26 - Eq. 28) to predict the total cumulative infected cases for the following 61 days, which are from 01 October 2022 to 30 November 2022. For this aim, the effective transmission rate β and the recovered rate γ that are given in Table 1 are not changed. Figure 3 shows the prediction result of infected cases and is compared with the real data of COVID-19 for these two months. On the 334th day, the predicted total cumulative infected cases were 5.3176 million, and the real total cumulative infected cases were 4.9922 million. The prediction result was higher than the real COVID-19 cases due to the SIR model does not consider the policies that have been taken into action by the government, for example, the vaccination plan, social distancing, and wearing a face mask.

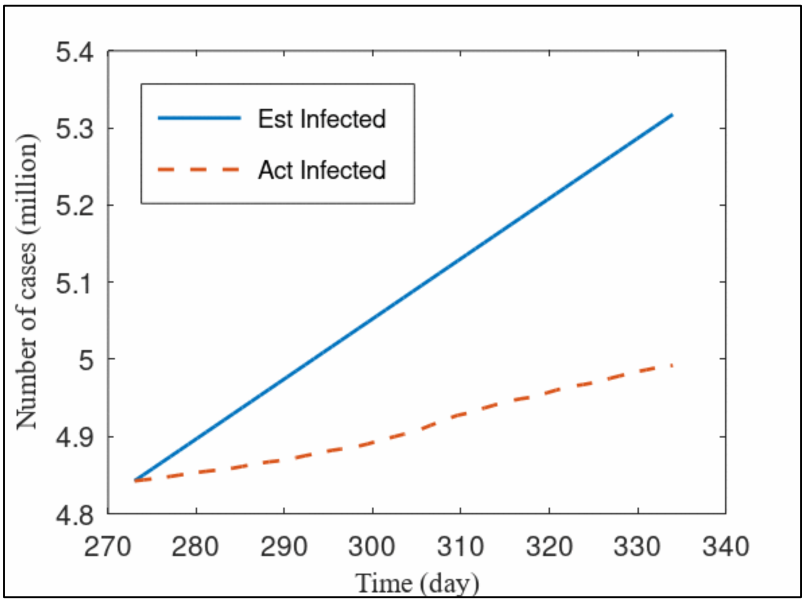

Figure 3: Prediction for infected cases in the following 61 days

In addition, the prediction of the spread of COVID-19 in the following 16 years is shown in Figure 4. Our aim is to have a full visualization of the spreading trend of disease through observation of the graphical result of the SIR model. It is noticed that in the following 937 days, which is within 2.5 years, the total cumulative infected cases are predicted to reach a peak of 8.1493 million. After that, the infected cases will reduce gradually due to the total cumulative of the susceptible cases being decreased and the recovered cases being increased. There are expected to have a total cumulative of 2.7794, 0.02757 and 29.857 million for susceptible, infected and recovered cases after 10 years. These numbers of cases indicate the steady state of COVID-19 in Malaysia.

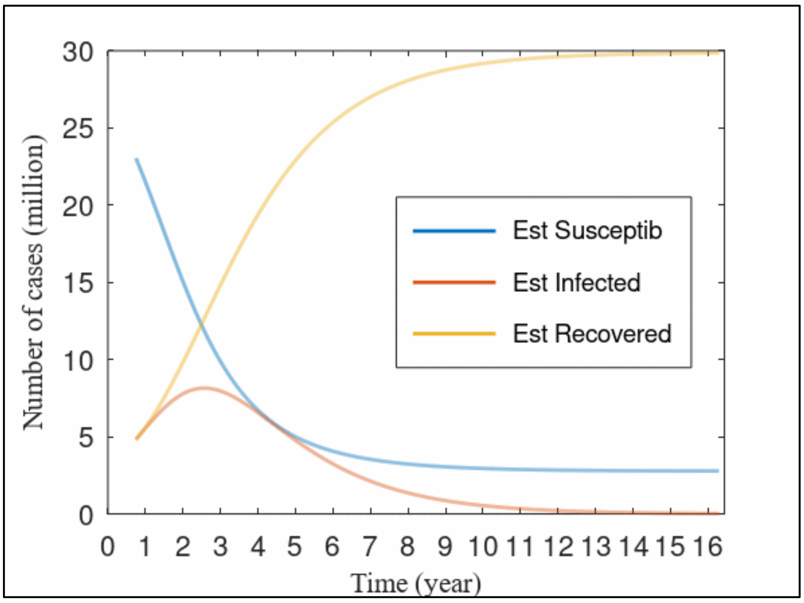

Figure 4: Prediction result from SIR model

4. Conclusion

In this paper, the classical SIR model was applied to predict the spreading trend of COVID-19 in Malaysia. Using the real data of COVID-19 from 1 January 2022 until 30 September 2022, the transmission and recovery rates were estimated through the Gauss-Newton computation approach, and the SIR model with these parameter estimates was solved using the Runge-Kutta fourth-order method. Prediction results showed that the infected cases of COVID-19 in Malaysia will reach a peak of 8.1493 million in 2.5 years (937 days) and tend to stay at steady states of 2.7794, 0.02757 and 29.857 million, respectively, for the expected total cumulative for susceptible, infected and recovered cases after 10 years. For future study, it is recommended to use complex mathematical models to study the spread of COVID-19 in Malaysia, whereby the number of compartments in the epidemic modelling can be increased and more possible influence factors should be considered to present the nearly real outbreak of COVID-19. The computational method used should be able to give a smaller root-mean-square error, and any ill conditions during the computation procedure ought to be resolved. Thus, more accuracy of prediction results could be given.

Acknowledgement

The authors would like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support.

References

[1] World Health Organization (WHO). (2020). *Coronavirus disease (COVID-19)*. Retrieved on May 20, 2022, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[2] Ndaïrou, F., Area, I., Nieto, J.J., Torres, D.F.M. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solitons Fractals*. 2020. 135:109846.

[3] Abdullah, N.H. (2020). *Pengesanan Kes Baharu yang Disahkan Dijangkiti 2019 Novel Coronavirus (2019-nCoV) di Malaysia*. Retrieved on May 10, 2022, from https://www.moh.gov.my/index.php/database_stores/store_view_page/21/1301)

[4] Prodanov, D. Analytical parameter estimation of the SIR epidemic model. Applications to the COVID-19 pandemic. *Entropy*. 2021. 23(1): 1-20.

[5] Daley, D. J., & Gani, J. *Epidemic Modelling: An Introduction*. Cambridge University Press. 2005.

[6] Hethcote, H. W. The mathematical models of diseases. *SIAM Review*. 2000. 42(4): 599-653.

[7] Srivastava, V., Srivastava, S., & Chaudhary, G. A systematic approach for COVID-19 predictions and parameter estimation. *Personal and Ubiquitous Computing*. 2020.

[8] Nesteruk, I. Estimations of the coronavirus epidemic dynamics in South Korea with the use of SIR model. *Researchgate.Net*. 2020.

[9] Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., Cao, Z., Wang, J., Yuan, W., Zhu, Y., Song, C., Chen, J., Xu, J., Li, F., Ma, T., Jiang, L., Yan, F., Yi, J., Hu, Y., Liao, Y. & Xiao, H. COVID-19: Challenges to GIS with big data. *Geography and Sustainability*. 2020. 1(1): 77-87

[10] Moualeu, D.P., Bowong, S. & Kurths, J. Parameter estimation of a tuberculosis model in a patchy environment: Case of Cameroon. *BIOMAT Conference 2013*. 2014. pp. 352-373.

[11] Amaral, M. T. R., Conceição, K. S., de Andrade, M. G., & Padovani, C. R. Generalized growth curve model for COVID-19 in Brazilian states. *Revista Brasileira de Biometria*. 2020. 38(3): 125-146.

[12] Bärwolff, G. *A Contribution to the Mathematical Modeling of the Corona / COVID-19 Pandemic*. medRxiv. 2020.

[13] Hale, J. K. *Ordinary Differential Equations*. Dover Publications. 2009.

[14] Chong, E. K. P., & Zak, S. H. *An Introduction to Optimization*. 4th Ed. John Wiley & Sons, Inc. 2013.

[15] Deuflhard, P. *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*. Springer Series in Computational Mathematics 35. 2004.

[16] Argyros, I. K., & Hilout, S. *Computational Methods in Nonlinear Analysis: Efficient Algorithms, Fixed Point Theory and Applications*. World Scientifc. 2013.

[17] Mart, J. *Parameter Optimization in COVID-19 Epidemiological Models*. Bachelor Thesis. Charles University; 2021.