JOURNAL OF SOCIAL COMPUTING
ISSN 2688-5255 06/06 pp182−189
Volume 3, Number 2, June 2022
DOI: 10.23919/JSC.2022.0005

Estimating Effective Reproduction Number for SIR Compartmental Model: A Stochastic Evolutionary Approach

W. K. Wong and Filbert H. Juwono*

Abstract: Compartmental pandemic models have become a significant tool in the battle against disease outbreaks. Despite this, pandemic models sometimes require extensive modification to accurately reflect the actual epidemic condition. The Susceptible-Infectious-Removed (SIR) model, in particular, contains two primary parameters: the infectious rate parameter β and the removal rate parameter γ, in addition to additional unknowns such as the initial infectious population. Adding to the complexity, there is an obvious challenge to track the evolution of these parameters, especially β and γ, over time which leads to the estimation of the reproduction number for the particular time window, R_T. This reproduction number may provide better understanding on the effectiveness of isolation or control measures. The changing R_T values (evolving over time window) will lead to even more possible parameter scenarios. Given the present Coronavirus Disease 2019 (COVID-19) pandemic, a stochastic optimization strategy is proposed to fit the model on the basis of parameter changes over time. Solutions are encoded to reflect the changing parameters of β_T and γ_T, allowing the changing R_T to be estimated. In our approach, an Adaptive Differential Evolution (ADE) and Particle Swarm Optimization (PSO) are used to fit the curves into previously recorded data. ADE eliminates the need to tune the parameters of the Differential Evolution (DE) to balance the exploitation and exploration in the solution space. Results show that the proposed optimized model can generally fit the curves well albeit high variance in the solutions.

Key words: Susceptible-Infectious-Removed (SIR) model; adaptive differential evolution; Coronavirus Disease 2019 (COVID-19)

1 Introduction

Compartmental mathematical models have been widely used to predict the spread of the infectious disease in every epidemic circumstance. For example, it has been used for predicting the spread of Severe Acute Respiratory Syndrome (SARS)[1] and the occurred Coronavirus Disease 2019 (COVID-19)[2]. The most common compartmental model is Susceptible-Infectious-Removed (SIR) model. The accuracy of the model largely depends on the parameter tuning of the model. In Ref. [3], conventional approaches, which include deterministic and mathematical approaches, were used to determine basic reproduction number R_0. Note that the basic reproduction number R_0 refers to an average number of secondary cases generated from one primary case, without immunity and vaccination in the community[4]. When $R_0 \geqslant 1$, it means that the disease transmission is exponential while $R_0 < 0$ indicates the controlled virus transmission. In addition, the importance of the fundamental reproduction number was thoroughly discussed.

In order to optimize the data fitting process to determine the model parameters, various stochastic approaches have been used as a feasible means. It is due to its ability to acquire almost optimal solution in

• W. K. Wong and Filbert H. Juwono are with the Department of Electrical and Computer Engineering, Curtin University Malaysia, Miri 98009, Malaysia. E-mail: weikitt.w@curtin.edu.my; filbert@ieee.org.
* To whom correspondence should be addressed.
Manuscript received: 2022-04-21; revised: 2022-06-07; accepted: 2022-06-09

© The author(s) 2022. The articles published in this open access journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

multimodal error fitness surface. For example, in Ref. [5], the authors applied a mixture Gaussian fitting method (singular spectral analysis-Gaussian fitting method) for fitting the model to real data. The model was validated using retrospective data available from China and Republic of Korea, and partially validated by currently available data from Italy, Spain, USA, and the UK. Then, the three model parameters were linked to physical meanings and interpretations related to the COVID-19 pandemic.

In Ref. [6], an SIR curve fitting model was presented using combinational parameter techniques. In Ref. [7], the authors applied Differential Evolution (DE) with a specific parameter for fitting the curve to Susceptible-Exposed-Infectious-Removed (SEIR) model. The fitness of the optimisation used was solely based on infection rate. The outcome of the research was the acquisition of the model parameters. In Ref. [8], researchers applied Particle Swarm Optimization (PSO) for the purpose tuning a model and subsequently predicting growth of pandemic based on the simulation. In Ref. [9], PSO was applied to optimize a neural network for predicting rise in daily cases of COVID-19. In particular, the authors evaluated and compared several approach using R^2 metric. The research highlighted the flexibility of applying stochastic approaches in tuning systems regardless of machine learning prediction model or epidemic model itself. Note that R^2 and root mean square error (RMSE) are two commonly used goodness of fit measurements for fitting COVID-19 predictions against actual scenarios.

In Ref. [10], the authors applied PSO for the purpose of pandemic model fitting for cholera outbreak. Notably, the authors ended with a note stating that tuning the PSO parameters could be a challenging task. The use of neural networks for prediction of rising cases of epidemic infection has been applied in Ref. [11]. Neural network has been used as an efficient method to predict infection rising cases for policy evaluation. In this aspect, various other meta-heuristics approaches should be considered, particularly those with lesser complexity for tuning the parameters. Considering the issues presented in Ref. [10], DE would be an excellent option since it has dominated optimization research for a decade with variations of auto tuning of parameters with varied benchmarking challenges[12].

In this paper, we explore two evolutionary algorithms, i.e., the Adaptive DE (ADE) and PSO for tuning SIR pandemic model parameters to fit the actual recorded data. In specific, this paper deals with acquiring the effective reproduction number value R_T for a specific time window (ΔT) using SIR pandemic model. The model depends on the parameters of transmission rate, β_T and removal rate, γ_T. The research covers the intricate strategies applied, in particular, using COVID-19 data as the test ground of the proposed methodology. To summarize, the main contributions of this research work are as follows:

- A new approach in evaluating β_T and γ_T using stochastic optimization and minimal data (daily cases) with minimalist SIR model and subsequently the reproduction number for the time windows in which $R_T = \frac{\beta_T}{\gamma_T}$ is proposed.
- An encoding approach to acquire SIR parameters values is used.

2 System Model

2.1 SIR model

In this section, we present a brief discussion of the SIR model. SIR remains the most reliable methodology for predicting or evaluating pandemic situations, such as in Refs. [13–17]. In fact, as the COVID−19 pandemic unfolds itself, government strategies and evaluation have almost exclusively been based on pandemic models. The basic model consists of three groups of individuals in a population of N, i.e., susceptible (S), infectious (I), and removed (R).

Individuals who have not been infected but are prone to infection are classified as susceptible. Infectious individuals are those who have been infected with the disease and are capable of transmitting it to others. Individuals who have been removed are those who have been recovered from the disease with immunity or isolated so that they cannot infect others. The transfer diagram of SIR model is shown in Fig. 1.

The SIR model is given by the following equations[18]:

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = -\frac{\beta I(t) S(t)}{N} \tag{1}$$

$$\frac{\mathrm{d}I(t)}{\mathrm{d}t} = \frac{\beta I(t) S(t)}{N} - \gamma I(t) \tag{2}$$

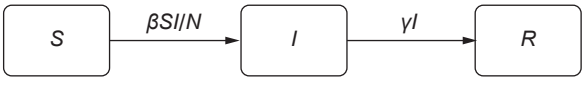

Fig. 1 Transfer diagram of the SIR model.

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = \gamma I(t) \tag{3}$$

where $N = S(t) + I(t) + R(t)$, β and γ are transmission and removal rate constants, respectively, and $S(t)$, $I(t)$, and $R(t)$ are the number of susceptible, infectious, and removed individuals at time t, respectively.

The initial reproduction number R_0 is often used as the R value over an entire period, while, in practice, the R value changes over a period of time. In this paper, we consider the reproductive value for a particular time window (Δt). The windowed effective reproduction number, R_T, therefore, can be expressed by[19]

$$R_T(k) = \beta_T(k)/\gamma_T(k) \tag{4}$$

where $\beta_T(k)$ and $\gamma_T(k)$ are the effective β and γ constants during period k, respectively. According to Ref. [20], the R_T value should be on a frequent basis, such as weekly, in order for the government to implement suitable intervention programs. Here, we assume that the reproduction number changes every one week or two weeks. Therefore, the time step for observation (evaluation window) is either $\Delta t = 1$ or $\Delta t = 2$ (in weeks). This concept is shown in Fig. 2. We assume that the initial reproduction number is given by $R_T(0) = R_0 = \beta/\gamma$. Furthermore, this may be seen as a trade-off to reduce complexity in searching the solution and representation.

We also note that one important parameter, I, is not known. In this case random testing in the population may reveal a rough value, nevertheless this data is often elusive and difficult to ascertain. Similarly, the initial infectious population, I_0, is also unknown, but it can be speculated that this value is not far from the initial value reported on daily basis ($t = 0$). For the sake of this research, we introduce $I_0 = U/\alpha$, where U represents the daily positive case (reported at $t = 0$) and $\alpha > 0$. Some research work has considered U as the number of daily cases of positive infection. In fact, based on the SIR model, U should belong to the removed group as they are

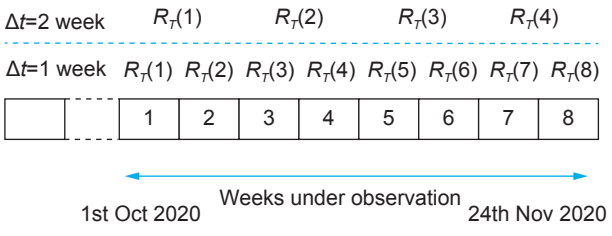

Fig. 2 Evaluation window for estimating R_T.

quarantined or hospitalized, thereby no longer having infectious capability in the population. The accumulated removed population (at $t = 0$) is estimated by considering the daily active cases, those who are still isolated in hospital/isolation center, and those who are recovered.

2.2 Adaptive differential equation

Differential evolution concept has been discussed in Refs. [12, 21, 22]. Basically, differential evolution works by selecting three N-dimensional auxiliary vectors $\{\boldsymbol{X}_{r_1}^G, \boldsymbol{X}_{r_2}^G, \boldsymbol{X}_{r_3}^G\}$ for each i-th solution vector $\boldsymbol{X}_i^G = [x_{i,1}^G, x_{i,2}^G, \ldots, x_{i,j}^G, \ldots, x_{i,N}^G]^{\mathrm{T}}$ for $i = 1, 2, \ldots, P$, where $r_1, r_2, r_3 \in \{1, 2, \ldots, P\}$, $i \neq r_1 \neq r_2 \neq r_3$, P is the population size, and $G = 1, 2, \ldots, G_{max}$ is the maximum generation.

The mutated vector, $\boldsymbol{V}_i^G = [v_{i,1}^G, v_{i,2}^G, \ldots, v_{i,j}^G, \ldots, v_{i,N}^G]^{\mathrm{T}}$, is formed as follows:

$$\boldsymbol{V}_i^G = \boldsymbol{X}_{r_1}^G + F(\boldsymbol{X}_{r_2}^G - \boldsymbol{X}_{r_3}^G) \tag{5}$$

where F is the differential weight. The mutated vector and the principal parent generate a trial vector $\boldsymbol{U}_i^G = [u_{i,1}^G, u_{i,2}^G, \ldots, u_{i,j}^G, \ldots, u_{i,N}^G]^{\mathrm{T}}$ using the following rule:

$$u_{i,j}^G = \begin{cases} u_{i,j}^G, & \text{if } rand_{i,j}[0,1] \leqslant Cr; \\ x_{i,j}^G, & \text{otherwise} \end{cases} \tag{6}$$

where Cr is the crossover rate and $rand_{i,j}[0,1]$ is a random number drawn from standard uniform distribution for each j-th component of the i-th vector. The next generation is calculated as follows:

$$\boldsymbol{X}_i^{G+1} = \begin{cases} \boldsymbol{U}_i^G, & \text{if } g(\boldsymbol{X}_i^G) \geqslant g(\boldsymbol{U}_i^G); \\ \boldsymbol{X}_i^G, & \text{if } g(\boldsymbol{X}_i^G) < g(\boldsymbol{U}_i^G) \end{cases} \tag{7}$$

where $g(\cdot)$ is the fitness function to be optimized.

Differential evolution has been known to be highly sensitive to Cr and F. Proper settings of Cr and F have high complexity. In order to enable proper tuning, we apply an automated tuning approach as stated in Ref. [8], which is as follows:

$$F = \begin{cases} \hat{\alpha} + (1-\hat{\alpha}) \times \sin\left(\frac{\pi \hat{t}}{maxiter} - \frac{\pi}{2}\right), & \text{if } \hat{t} \leqslant \frac{maxiter}{2}; \\ \hat{\alpha} - (1-\hat{\alpha}) \times \cos\left(\frac{\pi}{2} - \frac{\pi \hat{t}}{maxiter}\right), & \text{otherwise} \end{cases} \tag{8}$$

$$Cr = \begin{cases} \hat{\beta} + (1-\hat{\beta}) \times \sin\left(\frac{\pi \hat{t}}{maxiter} - \frac{\pi}{2}\right), & \text{if } \hat{t} \leqslant \frac{maxiter}{2}; \\ \hat{\beta} - (1-\hat{\beta}) \times \cos\left(\frac{\pi}{2} - \frac{\pi \hat{t}}{maxiter}\right), & \text{otherwise} \end{cases} \tag{9}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are constants, $\hat{t}$ is the generation of iteration, and $maxiter$ is the maximum number of iterations.

2.3 PSO

PSO has been discussed thoroughly in many research papers, such as in Ref. [23]. Suppose that we have a population of particles denoted by their positions $\boldsymbol{x}_i$ and velocities $\boldsymbol{v}_i$, where $\boldsymbol{x} \in \mathbb{R}^n$. Each particle has memory that can record its best position, $\mathcal{P}$, i.e., where it gives the best value and also the best position that have been explored by all the particles in the population, $\mathcal{G}$. The update equations of particle p for the d-th dimension at iteration $(i+1)$ are expressed by

$$x_p^d(i+1) = x_p^d(i) + v_p^d(i+1) \tag{10}$$

and

$$\begin{aligned} v_p^d(i+1) = w \cdot v_p^d(i) + c_1 \cdot \rho_1 \cdot (\mathcal{P}_p^d(i) - x_p^d(i)) + \\ c_2 \cdot \rho_2 \cdot (\mathcal{G}_p^d(i) - x_p^d(i)) \end{aligned} \tag{11}$$

where w is the inertia factor, ρ_1 and ρ_2 are random numbers between 0 and 1, and c_1 and c_2 denote self-confidence and swarm confidence coefficients, respectively.

2.4 Implementation

In the implementation, the negative of coefficient of determination, $-R^2$, is chosen as the fitness function as an evaluation of goodness-of-fit. The R^2 is a suitable metric compared to the normalized RMSE. The coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \tag{12}$$

where y_i is the data instance, $\hat{y}_i$ is the predicted data instance, and $\bar{y}$ is the mean of the data. Since this is a minimization approach, the fitness function value is defined in the range of $[\infty, -1]$, in which -1 denotes a perfect fitting.

In this paper, as we observe data for eight weeks, we encode the solution as follows. For 1-week window, there are eight periods, i.e., $k = 1, 2, \ldots, 8$ while for 2-week window, we have $k = 1, 2, \ldots, 4$. Therefore, the solution length for 1-week window is 17, i.e., $[\beta_1, \gamma_1, \ldots, \beta_8, \gamma_8, \alpha]$ while for 2-week window, it is a 9-dimension vector, i.e., $[\beta_1, \gamma_1, \ldots, \beta_4, \gamma_4, \alpha]$.

3 Results and Discussion

3.1 Simulation setting

In this paper, two regions in Malaysia are selected for evaluation of our proposed method. The data sources are obtained from legitimate government sources※ corroborated with daily case reporting. SIR models are deployed to fit into the data of two island states in Malaysia, i.e., Penang island and Labuan island (referred as Location 1 and Location 2 in this paper, respectively). The time frame for the data is from 1st October 2020 to 24th November 2020.

In the SIR model, although there are death cases reported, the number is too insignificant for modelling purposes. In addition, we assume that all the cases are local transmissions. SIR model does not consider imported cases into the model. The initial values are shown in Table 1. For PSO algorithm setting, after some intensive simulations, we choose the following parameter values: $w = 1$ and $c_1 = c_2 = 0.1$.

3.2 Results and analysis

Table 2 shows the R^2 values for both mean and median of the fitness solutions (the difference between median and mean would indicate the variance in the fitness acquired) for $\Delta t = 2$ week using ADE. DE is a population-based optimization approach in which N represents the number of search agents. Each population configuration is run 10 trials to show the average performance. For $\Delta t = 2$ week, we use $N = \{50, 100, 200\}$. The results further solidify the justification to apply a stochastic optimization approach as opposed to the deterministic approach. The results indicate that the fitness surface is highly multimodal with many local minima.

Table 3 shows similar metrics for $\Delta t = 1$ week using ADE. Note that the SIR with $\Delta t = 1$ week requires to double the population due to the fact that the problem complexity has approximately doubled. The implication of these results is that the time window could be better optimized using lesser parameters. Recall the Occam's razor principle which states that the smaller model may be more efficient in representing a particular

Table 1 Initial parameter values for SIR.

Parameter	Initial value	
	Location 1	Location 2
Number of population	1 790 000	99 500
Number of susceptible individuals	1 788 570	99 472
Number of reported daily cases ($t = 0$)	5	1
Number of hospitalized/isolation individuals ($t = 0$)	138	27

※ http://covid-19.moh.gov.my/terkini-negeri/112020/kemaskini-negeri-sehingga-18-november-2020

Table 2 Reproduction number (refer to Eq. (4)), Δt = 2 week, ADE.

Location	Statistics	Mean (N = 50)	Median (N = 50)	Mean (N = 100)	Median (N = 100)	Mean (N = 200)	Median (N = 200)	Best
	R^2	0.9683	0.9681	0.9825	0.9807	0.9812	0.9812	0.9933
	k = 1	2.1476	1.9749	3.0735	1.6853	2.2436	1.4586	1.433
Location 1	k = 2	0.6499	0.6520	0.8286	0.8408	1.1281	0.8492	0.8967
	k = 3	1.1616	1.1374	1.1253	1.0026	0.5598	1.0026	0.9573
	k = 4	2.3310	0.5664	0.7676	0.6887	0.7182	0.7182	1.3782
	R^2	0.9730	0.9731	0.9790	0.9818	0.9846	0.9847	0.9929
	k = 1	2.0900	1.4789	1.7870	1.4540	2.6101	2.0748	0.9690
Location 2	k = 2	1.4502	1.3929	1.4685	1.3155	1.2542	1.2474	1.2791
	k = 3	0.8194	0.8631	0.8851	0.9493	0.9633	0.9755	0.9495
	k = 4	0.6654	0.7187	3.2681	0.6116	0.3669	0.2751	0.6040

Table 3 Reproduction number (refer to Eq. (4)), Δt = 1 week, ADE.

Location	Statistics	Mean (N = 100)	Median (N = 100)	Mean (N = 200)	Median (N = 200)	Mean (N = 400)	Median (N = 400)	Best
	R^2	0.9690	0.9670	0.9744	0.9720	0.9782	0.9790	0.9930
	k = 1	6.2934	4.0270	3.4974	3.7000	12.0936	2.3466	2.6012
	k = 2	1.7927	1.1935	6.6500	1.5137	5.9416	1.3216	0.9011
	k = 3	1.1862	1.0867	0.8747	0.9108	7.4998	1.4240	1.3797
Location 1	k = 4	2.1275	0.7421	1.5863	1.3450	1.3424	0.7796	0.9343
	k = 5	2.1785	1.2690	1.4310	0.8234	0.9309	0.6478	0.4931
	k = 6	1.2547	0.8417	0.9600	0.9682	2.1103	1.3090	1.2370
	k = 7	0.6576	0.3818	0.6932	0.4898	1.0559	0.8094	0.4867
	k = 8	0.8810	0.9365	2.2159	1.3637	4.2264	0.8526	0.5281
	R^2	0.9807	0.9823	0.9855	0.9858	0.9854	0.9855	0.9922
	k = 1	36.6926	2.7802	5.3952	3.5627	3.3423	2.5594	8.1336
	k = 2	2.2009	0.9945	5.2385	1.8750	3.6015	1.4780	1.4676
	k = 3	2.391	1.0355	1.1338	0.5998	1.1473	0.9453	0.9158
Location 2	k = 4	1.7081	1.4657	2.2666	1.6887	1.6442	1.4295	1.5477
	k = 5	1.1123	1.0658	1.2121	1.1859	1.2293	1.2822	1.8707
	k = 6	0.5969	0.5774	0.6056	0.6159	0.5488	0.6090	0.8754
	k = 7	1.2831	0.8412	1.1061	0.4584	1.0200	0.8818	1.2561
	k = 8	5.7382	1.7488	0.9855	0.9974	0.7026	0.4770	1.5663

phenomenon. In this case longer time window yields lesser parameters for optimization and model representation. In addition, the results show that better performing configurations are found with higher population in the optimization problem. Using the obtained R_T, we perform the curve fitting for both locations as shown in Fig. 3. We observe that both configurations give relatively good fitting.

The predicted effective R_T values are also shown in Tables 2 and 3 for both locations. Applying p-value to evaluate the two time windows in consideration, i.e., $\Delta t = 2$ week and $\Delta t = 1$ week, yields the average of 0.9706 and 0.9736 for Location 1 and 0.9789 and 0.9839 for Location 2, respectively. The difference of 0.003 (p-value = 0.2454) and 0.005 (p-value = 0.0323) shows that the difference is statistically significant (> 0.001). Thus, it shows that smaller time window consideration yields slight improvement in the curve fitting. Nevertheless, the difference is marginal and larger time window (Δt = 2 week) may be applied to reduce computation load.

Tables 4 and 5 show the results for PSO algorithm. It can be seen that the results are marginally better than the ADE results. PSO parameters, on the other hand, must be carefully selected, as opposed to the ADE. Moreover, we also compare the R^2 when R_0 is used for the entire eight weeks as seen in Table 6. The β, γ, and α values are also selected using the ADE and PSO algorithms. The equivalent R^2 values are presented showing the mean, median, and the best fitting configurations. As expected, considering changing β and γ values

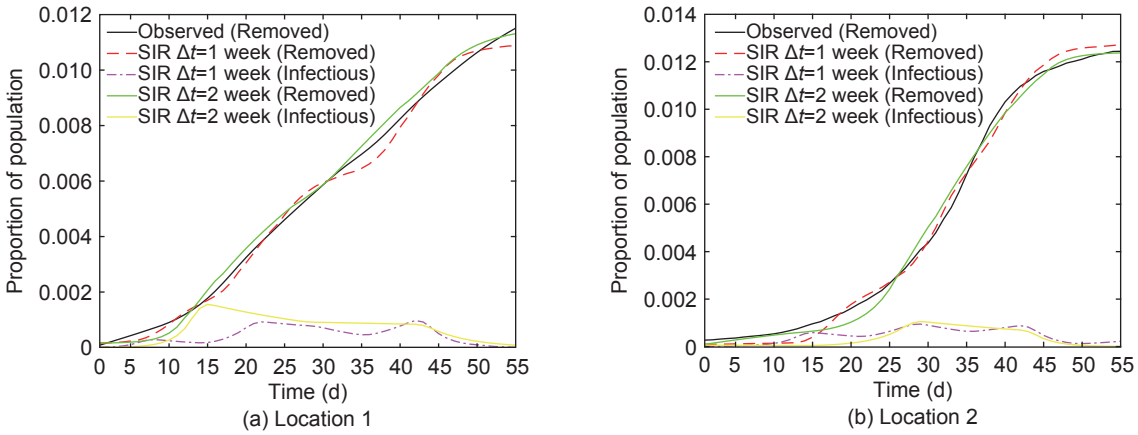

Fig. 3 Curve fitting for Location 1 and Location 2 using ADE.

Table 4 Reproduction number (refer to Eq. (4)), Δ*t* = 2 week, PSO.

Location	Statistics	Mean (N = 50)	Median (N = 50)	Mean (N = 100)	Median (N = 100)	Mean (N = 200)	Median (N = 200)	Best
Location 1	R^2	0.9822	0.9874	0.9902	0.9924	0.9939	0.9947	0.9986
	k=1	2.1179	1.9196	2.1159	2.0532	2.1780	2.0179	1.4835
	k=2	0.9557	0.9751	0.9201	0.9164	0.8735	0.9269	0.9452
	k=3	0.9553	0.9874	1.0300	1.0522	1.0932	1.0862	1.0414
	k=4	0.7044	0.7586	0.9329	0.9478	0.8146	0.8605	0.9321
Location 2	R^2	0.9912	0.9940	0.9897	0.9933	0.9928	0.9936	0.9964
	k=1	2.2371	2.0849	2.0748	1.7933	2.8439	2.2829	1.7418
	k=2	1.3575	1.3113	1.4277	1.4592	1.2766	1.3117	1.1171
	k=3	0.8980	0.9240	0.8831	0.8942	0.9331	0.9181	1.0094
	k=4	0.6338	0.6577	0.7068	0.6538	0.5954	0.6715	0.2982

Table 5 Reproduction number (refer to Eq. (4)), Δ*t* = 1 week, PSO.

Location	Statistics	Mean (N = 100)	Median (N = 100)	Mean (N = 200)	Median (N = 200)	Mean (N = 400)	Median (N = 400)	Best
Location 1	R^2	0.9872	0.9885	0.9938	0.9959	0.9940	0.9946	0.9984
	k=1	3.4207	2.7461	3.8724	3.4887	3.1363	2.9536	1.9073
	k=2	2.3961	2.1817	1.7921	1.3049	1.7025	1.6083	1.1019
	k=3	0.9386	0.9873	1.0833	1.2029	1.0288	0.9333	1.3687
	k=4	0.9866	0.9560	0.8631	0.7258	0.8789	0.8797	0.6196
	k=5	1.0558	1.1078	1.1365	1.2498	1.1507	1.1442	1.0225
	k=6	0.9528	0.9610	1.0737	1.0177	0.9170	0.8284	1.2794
	k=7	1.2682	0.9534	0.8411	0.9110	1.1074	1.1162	0.7647
	k=8	0.9467	0.7704	1.5176	1.3483	0.8972	0.6906	1.3287
Location 2	R^2	0.9912	0.9940	0.9897	0.9933	0.9928	0.9936	0.9964
	k=1	2.6529	1.4316	2.2347	1.8208	4.7703	2.2475	1.0548
	k=2	2.9566	3.2377	1.9441	1.7001	3.3855	2.7042	3.9589
	k=3	1.9804	1.3786	1.6192	1.4230	1.2325	1.1814	1.1545
	k=4	0.8740	0.9692	1.3988	1.4416	1.2417	1.1617	1.0088
	k=5	1.3500	1.2473	1.0528	1.0130	1.2926	1.3006	1.1136
	k=6	0.7184	0.7087	0.6639	0.6938	0.5411	0.5565	1.0415
	k=7	0.5750	0.5201	0.9515	0.9163	0.9771	0.8169	0.9367
	k=8	1.3453	1.2995	1.4330	1.3111	0.9799	0.8771	2.6551

inevitably yields better fitting.

The plots in Fig. 4 show the fitness progression of the best solutions for ADE and PSO methods. The plots show that the optimization method does not converge too

Table 6 R^2 Values for PSO and ADE using R_0.

Location	Algorithm	R^2 value		
		Mean	Median	Best
Location 1	ADE	0.9418	0.9556	0.9767
	PSO	0.8259	0.8485	0.9873
Location 2	ADE	0.8793	0.8960	0.9311
	PSO	0.8644	0.8862	0.9414

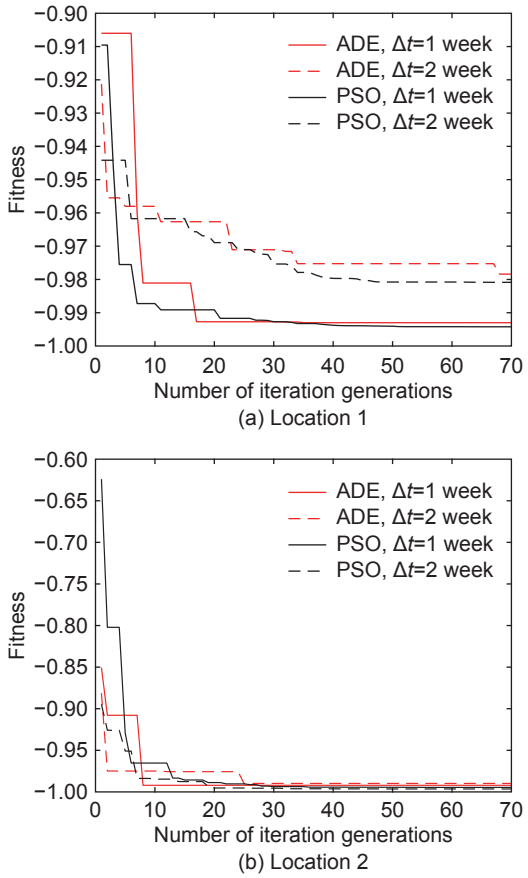

Fig. 4 Fitness vs. iteration generation for best solution.

early, thereby allowing optimal solution finding. Note that the overall best configurations are selected based on the fitness.

4 Conclusion and Future Work

A stochastic curve fitting has been proposed to discover β_T and γ_T and subsequently the reproduction value candidate R_T for the various periods of consideration. Furthermore, we have proposed an encoding method to represent the various parameters with some assumptions of the initial infectious population. The initial infectious number in the community is difficult to find and estimate, it is also considered as a tunable parameter for modelling. When applied to curve fitting on COVID-19 on specific locations, the results show that there are varying solutions available, hence increasing the variability of the possible scenarios. In particular, it must also be highlighted that parameter balance between exploitation and exploration plays an important role given this finding. Due to complexity of this problem, different curve fitting on acquired data may pose different fitness surface. In stating this, stochastic approaches, especially those with automated parameter setting could be high advantageous for this type of problem.

Future works may look into further evaluation on other data to investigate suitability of the proposed approach. Most real time reports often just report the daily occurrence and therefore these may be lumped as removed cases as stated in the paper. In view of vaccination, the approach model may simply include the vaccination population as removed in view of immunity. Again, the vaccination data are often reported by the authorities and would, therefore, be an easy extension to the proposed approach.

References

[1] T. W. Ng, G. Turinici, and A. Danchin, A double epidemic model for the SARS propagation, *BMC Infectious Diseases*, vol. 3, no. 19, pp. 1–16, 2003.

[2] R. U. Din and E. A. Algehyne, Mathematical analysis of COVID-19 by using SIR model with convex incidence rate, *Results in Physics*, vol. 23, p. 103970, 2021.

[3] B. Ridenhour, J. M. Kowalik, and D. K. Shay, Unraveling R0: Considerations for public health applications, *American Journal of Public Health*, vol. 104, no. 2, pp. 32–41, 2014.

[4] I. Locatelli, B. Trächsel, and V. Rousson, Estimating the basic reproduction number for COVID-19 in western Europe, *PLOS ONE*, vol. 16, no. 3, pp. 1–9, 2021.

[5] J. Ren, Y. Yan, H. Zhao, P. Ma, J. Zabalza, Z. Hussain, S. Luo, Q. Dai, S. Zhao, A. Sheikh, et al., A novel intelligent computational approach to model epidemiological trends and assess the impact of non-pharmacological interventions for COVID-19, *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3551–3563, 2020.

[6] E. Severeyn, S. Wong, H. Herrera, A. L. Cruz, J. Velásquez, and M. Huerta, Study of basic reproduction number projection of SARS-CoV-2 epidemic in USA and Brazil, in *Proc. 2020 IEEE ANDESCON*, Quito, Ecuador, 2020, pp. 1–6.

[7] A. Godio, F. Pace, and A. Vergnano, SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence, *International Journal of Environmental Research and Public Health*, vol. 17, no. 10, p. 3535, 2020.

[8] Z. Huang and Y. Chen, An improved differential evolution algorithm based on adaptive parameter, *Journal of Control*

Science and Engineering, vol. 2013, p. 462706, 2013.
[9] A. Riccardi, J. Gemignani, F. Fernández-Navarro, and A. Heffernan, Optimisation of non-pharmaceutical measures in COVID-19 growth via neural networks, *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 1, pp. 79–91, 2021.
[10] D. Akman, O. Akman, and E. Schaefer, Parameter estimation in ordinary differential equations modeling via particle swarm optimization, *Journal of Applied Mathematics*, vol. 2018, p. 9160793, 2018.
[11] C. Zhan, Z. Wu, Q. Wen, Y. Gao, and H. Zhang, Optimizing broad learning system hyper-parameters through particle swarm optimization for predicting COVID-19 in 184 countries, in *Proc. 2020 IEEE International Conference on E-health Networking, Application & Services* (*HEALTHCOM*), Shenzhen, China, 2021, pp. 1–6.
[12] W. Wong and C. I. Ming, A review on metaheuristic algorithms: Recent trends, benchmarking and applications, in *Proc. 2019 7th International Conference on Smart Computing Communications* (*ICSCC*), Sarawak, Malaysia, 2019, pp. 1–5.
[13] I. Cooper, A. Mondal, and C. G. Antonopoulos, A SIR model assumption for the spread of COVID-19 in different communities, *Chaos*, *Solitons* & *Fractals* , vol. 139, p. 110057, 2020.
[14] W. K. Wong, F. H. Juwono, and T. H. Chua, SIR simulation of COVID-19 pandemic in Malaysia: Will the vaccination program be effective? https://arxiv.org/abs/2101.07494, 2021.
[15] W. -J. Zhu and S. -F. Shen, An improved SIR model describing the epidemic dynamics of the COVID-19 in China, *Results in Physics*, vol. 25, p. 104289, 2021.
[16] E. Barbieri, W. E. Fitzgibbon, and J. Morgan, New insights into an epidemic SIR model for control and public health intervention, in *Proc. 2021 IEEE Conference on Control Technology and Applications* (*CCTA*), San Diego, CA, USA, 2021, pp. 150–155.
[17] G. Nakamura, B. Grammaticos, and M. Badoual, Vaccination strategies for a seasonal epidemic: A simple SIR model, *Open Communications in Nonlinear Mathematical Physics*, doi: 10.46298/ocnmp.7463.
[18] H. W. Hethcote, The mathematics of infectious diseases, *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.
[19] Y. -C. Chen, P. -E. Lu, C. -S. Chang, and T. -H. Liu, A time-dependent SIR model for COVID-19 with undetectable infected persons, *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 3279–3294, 2020.
[20] T. V. Inglesby, Public health measures and the reproduction number of SARS-CoV-2, *JAMA* , vol. 323, no. 21, pp. 2186–2187, 2020.
[21] S. Das and P. N. Suganthan, Differential evolution: A survey of the state-of-the-art, *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.
[22] N. Noman, D. Bollegala, and H. Iba, An adaptive differential evolution algorithm, in *Proc. 2011 IEEE Congress of Evolutionary Computation* (*CEC*), New Orleans, LA, USA, 2011, pp. 2229–2236.
[23] S. Das, A. Abraham, and A. Konar, Particle swarm optimization and differential evolution algorithms: Technical analysis, applications and hybridization perspectives, in *Advances of Computational Intelligence in Industrial Systems*, Y. Liu, A. Sun, H. T. Loh, W. F. Lu, and E. -P. Lim, eds. Berlin, Germany: Springer, 2008, pp. 1–38.

W. K. Wong is currently serving as a senior lecturer with the Department of Electrical and Computer Engineering, Curtin University Malaysia. He received the MEng and PhD degrees from University Malaysia Sabah in 2012 and 2016, respectively. Prior to joining academia, he was with the telecommunication and building services industry. His research interests include embedded system development, machine learning applications, and image processing.

Filbert H. Juwono received the BEng degree in electrical engineering and the MEng degree in telecommunication engineering from the University of Indonesia, Depok, Indonesia, in 2007 and 2009, respectively, and the PhD degree in electrical and electronic engineering from University of Western Australia, Perth, WA, Australia, in 2017. He is currently a senior lecturer with the Department of Electrical and Computer Engineering, Curtin University Malaysia. His research interests include signal processing for communications, wireless communications, power-line communications, machine learning applications, and biomedical engineering. He is a senior member of IEEE. He was a recipient of the prestigious Australian Awards Scholarship in 2012. He serves as an associate editor for *IEEE Access*, a review editor for *Frontiers in Signal Processing*, and the editor-in-chief for a newly established journal *Green Intelligent Systems and Applications*.