Chaos, Solitons and Fractals 153 (2021) 111556

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Robust optimal parameter estimation for the susceptible-unidentified infected-confirmed model

Chaeyoung Lee[a], Soobin Kwak[a], Sangkwon Kim[a], Youngjin Hwang[a], Yongho Choi[b], Junseok Kim[a,*]

[a] Department of Mathematics, Korea University, Seoul 02841, Republic of Korea
[b] Department of Mathematics and Big Data, Daegu University, Gyeongsan-si, Gyeongsangbuk-do 38453, Republic of Korea

ARTICLE INFO

Article history:
Received 23 June 2021
Revised 3 October 2021
Accepted 19 October 2021
Available online 3 November 2021

Keywords:
COVID-19
SUC model
Optimal parameter estimation
Least-squares fitting

ABSTRACT

In this study, we consider a robust optimal parameter estimation method for the Susceptible-Unidentified infected-Confirmed (SUC) epidemic dynamics model. One of the problems in determining parameter values associated with epidemic mathematical models is that the optimal parameter values are very sensitive to the initial guess of parameter values. To resolve this problem, we fix the value of one parameter and solve an optimization problem of finding the other parameter values which best fit the confirmed population. The fixed parameter value can be obtained using data from epidemiological surveillance systems. To demonstrate the robustness and accuracy of the proposed method, we perform various numerical experiments with synthetic and real-world data from South Korea, the United States of America, India, and Brazil. The computational results confirm the potential practical application of the proposed method.

1. Introduction

Recently, the coronavirus disease 2019 (COVID-19) pandemic has caused unprecedented global social and economic crisis [1]. The novel coronavirus, which acutely causes the severe respiratory syndrome, is more contagious and more survivable than any other virus [2]. To overcome the infectious disease, numerous researchers in various fields started virological and biological studies; and have developed treatment methods and vaccines [3–6]. Some studies have identified the current states of the COVID-19 spread using mathematical methods. Badr et al. defined a daily mobility ratio after analyzing mobility patterns using mobile phone data and studied the statistical correlation between social distancing, mobility patterns, and COVID-19 cases [7]. In [8], the author analyzed the epidemiological trend in India using mathematical methods.

Mathematical modeling of COVID-19 has also been extensively studied to predict the spread trend in order to prevent or prepare for further damage in the serious pandemic situation that is not easily suppressed. Many existing epidemic models such as susceptible-infected (SI), susceptible-infected-removed (SIR), and susceptible-exposed-infectious-recovered (SEIR) models have been utilized [9]. The SI model is used for understanding complex behaviors and prediction of the spread of COVID-19 disease [10]. In [11], Cooper et al. studied the spread of COVID-19 in different communities using a SIR model. Shaobo et al. studied the SEIR model for COVID-19 depending on control strategies [12]. Korolev [13] forecasted the future number of deaths using the SEIR model with deaths for COVID-19. In [14,15], the authors presented the prediction of COVID-19 transmission dynamics using the SEIR model with consideration for hospital-quarantined cases, and quantified the potential effects of strategies on school opening.

There have also been studies on mathematical models that take into account the quarantine or isolation by appending more compartments to the SEIR model. Peter et al. [16] formulated a mathematical model under consideration of five compartments: susceptible, exposed, infected, quarantined, and recovered. Using the model and real data in Pakistan, the authors estimated parameters and performed numerical simulations to observe various features. Memon et al. [17] included both the quarantined and isolated into the SEIR framework. In [18], Musa et al. proposed a mathematical model with the effect of health education knowledge on COVID-19. Based on the SEIR model, they incorporated awareness programs and various hospitalization plans for mild and severe cases. More complicated models were presented. In [19], Khan and Atangana proposed a model representing the interaction between bats and unknown hosts in the early stages of the spread of COVID-19. They also extended it considering both the quarantined and the isolated; and predicted the dynamics of COVID-19 in China in Khan et al. [20].

* Corresponding author.
E-mail address: cfdkim@korea.ac.kr (J. Kim).
URL: https://mathematicians.korea.ac.kr/cfdkim/ (J. Kim)

https://doi.org/10.1016/j.chaos.2021.111556
0960-0779/

Table 1
Description of the variables and parameters for the SUC model.

Variable/parameter	Description
N	Total population. Constant value.
t	Time
$S(t)$	Susceptible population at time t
$U(t)$	Unidentified infected population at time t
$C(t)$	Confirmed population at time t
β	Average number of contacts made by an infected individual per time
γ	Inverse of the days taken for confirming the unidentified infected
Δt	Time step size
S_n	Susceptible population at time $n\Delta t$ for integer $n = 0,\ 1,\ 2, \ldots$
U_n	Unidentified infected population at time $n\Delta t$ for integer $n = 0,\ 1,\ 2, \ldots$
C_n	Confirmed population at time $n\Delta t$ for integer $n = 0,\ 1,\ 2, \ldots$
β^0	Initial guess value of β for the fitting function **lsqcurvefit**
γ^0	Initial guess value of γ for the fitting function **lsqcurvefit**
U_0^0	Initial guess value of U_0 for the fitting function **lsqcurvefit**

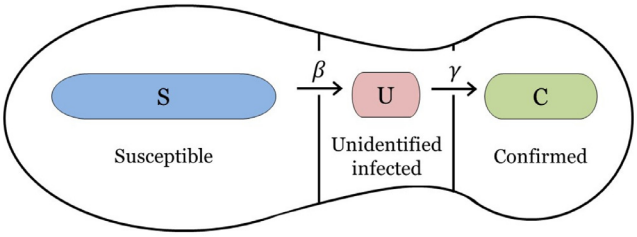

Fig. 1. Compartmental diagram of the SUC model.

Many mathematical models are focused on predicting the number of confirmed cases. However, above all, in the case of a novel infectious disease that has no cure and is highly contagious, such as COVID-19, it is the most important thing to find people who are infected but not confirmed in the unprecedented situation where community transmission and clusters of cases are the main issues of COVID-19. Because people who have already been confirmed are isolated, the chances of transmission from the confirmed population are rare. On the other hand, the unidentified infected people are not isolated so that they can transmit COVID-19 to susceptible people. Accordingly, mathematical modelings for the unidentified infected cases were developed [21–23]. For estimation of future situations using mathematical models, past and present data are necessary. Although the numbers of the confirmed, deaths, and recovered cases are known in the real world, the numbers of the unidentified infected cases are unknown. Furthermore, there is no uniqueness of the optimal parameter values for the models owing to an underdetermined system of the epidemic equations with limited case data.

In this study, we consider a method of robust optimal parameter estimation for the Susceptible-Unidentified infected-Confirmed (SUC) epidemic mathematical model for COVID-19 [23]:

$$\frac{dS(t)}{dt} = -\beta\frac{S(t)U(t)}{N}, \tag{1}$$

$$\frac{dU(t)}{dt} = \beta\frac{S(t)U(t)}{N} - \gamma U(t), \tag{2}$$

$$\frac{dC(t)}{dt} = \gamma U(t). \tag{3}$$

The variables and parameters are tabulated in Table 1. The unidentified infected population can spread the disease and has not yet been confirmed. We assume the total population N is constant. The values of $\beta S(t)/N$ and γ are growth and decay rates of the unidentified infected population $U(t)$ at time t, respectively. Fig. 1 shows a compartmental diagram of the SUC model.

The SUC model is a fundamental building block equation. Although the model is simple, it can capture the main epidemic dynamics of COVID-19 using as little information as possible. The simple and basic model would be useful to utilize the estimated U in health policies by policymakers who are not experts in mathematical modeling. However, in the cases of complex models with many variables and parameters, the models require generally many fitted parameters and it is not easy to validate the robustness of the models.

In [23], a nonlinear curve-fitting function **lsqcurvefit** was employed to find optimal parameters β, γ, and $U(0)$. The **lsqcurvefit** solver is a popular built-in function in MATLAB optimization toolbox, which has been widely applied to find the best-fitted coefficients in epidemic modeling [15,24,25]. The function depends on the initial guess values for the parameters to estimate because it is based on an iterative method with local convergence [26].

The main purpose of this paper is to propose a robust method for parameter estimation using the SUC model. The proposed method is not sensitive to the initial guess values for the parameters. It is expected that unknown parameters are well fitted regardless of initial guesses and the accuracy of the estimation for the unidentified infected population will be improved. Furthermore, the estimated unidentified infected cases are potentially helpful for health authorities to establish social distancing policies or vaccine distribution strategies.

The contents of this paper are as follows: in Section 2, we present the numerical solution algorithm for the robust optimal parameter estimation. Using the proposed algorithm, computational experiments are performed in Section 3. We first apply the proposed method to an example situation and verify it, and then we estimate the number of unidentified infected cases with real-world confirmed data from several countries. We discuss the results and the basic production number in Section 4. In Section 5, we provide our conclusions. Note that the MATLAB code for the computational implementation is given in Appendix for interested readers.

2. Numerical solution algorithm

Let $S_n = S(n\Delta t)$, $U_n = U(n\Delta t)$, and $C_n = C(n\Delta t)$, where Δt is a time step. The SUC model is solved using a finite difference scheme. For $n = 0,\ 1,\ 2, \ldots$, we have the following discrete equations:

$$S_{n+1} = S_n - \Delta t\beta\frac{S_nU_n}{N}, \tag{4}$$

$$U_{n+1} = U_n + \Delta t\left(\beta\frac{S_nU_n}{N} - \gamma U_n\right), \tag{5}$$

$$C_{n+1} = C_n + \Delta t \gamma U_n, \tag{6}$$

where β, γ, U_0 are the unknown parameters. To solve the discrete system of Eqs. (4)–(6), we need to know these parameter values. However, in the real-world population, the number of the unidentified infected cases U is unknown and only the cumulative numbers of the confirmed cases C are known. To estimate the unknown unidentified infected cases U, we use the SUC model (4)–(6) and the fitting function **lsqcurvefit** in MATLAB R2020b, which is a nonlinear curve-fitting function in a least-squares sense [27]. That is, we obtain optimal parameters β, γ, U_0 which minimize the following cost function:

$$\mathcal{E}(\beta,\ \gamma,\ U_0) = \frac{1}{2}\sum_{i=1}^{p}(\widehat{C}_i - C_{n_i})^2, \tag{7}$$

where p is the number of the given real cumulative confirmed case data $\widehat{C}_i$ $(i = 1, 2, \ldots, p)$ and C_{n_i} $(i = 1, 2, \ldots, p)$ are the numerical solutions from Eqs. (4)–(6) at the corresponding times. However, the optimal parameter values (β, γ, U_0) estimated by using the fitting function strongly depend on the initial guess values of $(\beta^0, \gamma^0, U_0^0)$, which implies the non-uniqueness of the optimal parameter values.

To prove the existence of the non-uniqueness of the optimal parameter values if only the cumulative confirmed case data is used, let us consider the following equations with $\gamma' \neq \gamma$:

$$\frac{d\bar{S}(t)}{dt} = -\beta'\frac{\bar{S}(t)\bar{U}(t)}{N}, \tag{8}$$

$$\frac{d\bar{U}(t)}{dt} = \beta'\frac{\bar{S}(t)\bar{U}(t)}{N} - \gamma'\bar{U}(t), \tag{9}$$

$$\frac{dC(t)}{dt} = \gamma'\bar{U}(t). \tag{10}$$

Let us assume both Eqs. (4)–(6) and (8)–(10) best fit the cost function (7). By the assumption of the best fitting of the cumulative confirmed case data, $\gamma U(t) = \gamma'\bar{U}(t)$ holds from Eqs. (6) and (10). By substituting $\bar{U}(t) = (\gamma/\gamma')U(t)$ into Eq. (9), we have

$$\frac{\gamma}{\gamma'}\frac{dU(t)}{dt} = \beta'\frac{\bar{S}(t)}{N}\frac{\gamma}{\gamma'}U(t) - \gamma'\frac{\gamma}{\gamma'}U(t).$$

After simplifying, we have

$$\frac{dU(t)}{dt} = \left(\beta'\frac{\bar{S}(t)}{N} - \gamma'\right)U(t).$$

Therefore, from Eq. (5), it holds that $\beta'\bar{S}(t)/N - \gamma' = \beta S(t)/N - \gamma$. From this reason, there is no guarantee that the solution of the governing equations is unique.

To resolve this non-uniqueness problem, we fix the value of γ and compute the optimal parameter values of $(\beta,\ U_0)$ which minimize the cost function (7):

$$[\beta,\ U_0] = \textbf{lsqcurvefit}(\text{'}SUCmodel\text{'},\ [\beta^0,\ U_0^0], \textbf{Tdata},\ \textbf{Cdata}), \tag{11}$$

where $[\beta,\ U_0]$ are the optimized parameters and $[\beta^0,\ U_0^0]$ are the initial guess of parameters for *SUCmodel*; and **Cdata** is the real cumulative confirmed case data at times **Tdata**. Because γ is the inverse of the average time until an unidentified infected individual is confirmed, we can approximately estimate γ value through epidemiological investigation.

3. Numerical tests

For all numerical computations, the time step size is taken as $\Delta t = 0.001$.

3.1. Comparison with the previous work

The proposed method in this paper estimates two parameters β and U_0 whereas three parameters, β, U_0, and γ, were estimated in the previous work [23]. In this section, using an example case, simulations are performed to demonstrate the robustness of the proposed method by comparing the computational results between the proposed and previous methods.

We first generate 7-day data of confirmed and unidentified infected population, listed in Table 2, using the SUC model with the initial parameters tabulated in Table 3. The generated values of the unidentified infected and confirmed population are called as U_{ref} and C_{ref}, respectively. The reason why we set $\gamma = 1/4$ is that it takes approximately four days before being confirmed with COVID-19 [28,29].

First, we fix one of the two parameters, β and U_0, as a reference value, and try to calculate the optimal values by changing the initial setting of the other value. Here, $\gamma = 1/4.5\,1/4,\ 1/3.5$ are used for the proposed method as shown in Fig. 2(a) and (b); and $\gamma^0 = 1/4.5\,1/4,\ 1/3.5$ are used for the previous method as shown in Fig. 2(c) and (d). Fig. 2(a) and (c) show the optimal β with the initial guess β^0 from 1/30 to 10/3 and fixed $U_0^0 = 3000$, respectively; and Fig. 2(b) and (d) show the optimal U_0 with the initial U_0^0 from 300 to 30,000 and fixed $\beta^0 = 1/3$, respectively. While the results of the previous method imply a large difference depending on the initial value, the results computed by the proposed one mean that the optimal values are not sensitive to the initial guess values.

More specifically, to examine the changes according to the initial guess values β^0 and U_0^0, we compare the relative error E_{err} defined as follows:

$$E_{err} = |\max(P) - \min(P)|/\bar{P},$$

where P is the optimal values β or U_0 for each initial guess value and $\bar{P}$ is the average value of P. Table 4 shows the relative errors of the corresponding optimal values estimated by each initial guess value when $\gamma = 1/4.5, 1/4,\ 1/3.5$ are used. The results indicate that the estimated parameters are almost constant.

Next, we compute the number of unidentified infected patients using arbitrary values of β, U_0 and the generated confirmed data in Table 2. The parameters are taken as listed in Table 5.

We compare the estimation of unidentified infected population depending on the fixed value γ. Fig. 3 represents the computational results of the unidentified infected and confirmed populations. When γ is taken as the same as the reference value, the estimated results of the unidentified infected and confirmed population are the same as the reference values. When different values of γ are used, only the estimated confirmed population are the same as the reference value. In these cases, the estimated unidentified infected population are not the same owing to the non-uniqueness of the optimal parameters in the given model, nonetheless, those dynamics are considerably similar. Because the unidentified infected population in the real world is unknown, it is meaningful to estimate its dynamics using an appropriate fixed value of γ that is the inverse of the average time for confirming infected individuals.

Table 6 lists the estimated values of U and γU with respect to γ. Depending on the γ value used, U takes different values, however, γU is almost constant.

Table 7 shows the estimated values of β, γ, $\beta S_0/N - \gamma$, U_0, and γU_0. Note that the values of $\beta S_0/N - \gamma$ are approximately constant despite using different values of γ.

3.2. Estimation with real-world data

In this section, we estimate optimal parameters using the proposed method, and then analyze the results. Here, we use the con-

Table 2
Generated data for the unidentified infected and confirmed patients.

Days	0	1	2	3	4	5	6
U_{ref}	3000.0	3260.2	3542.9	3850.1	4184.0	4546.7	4940.8
C_{ref}	20000.0	20782.0	21631.9	22555.5	23559.1	24649.8	25835.0

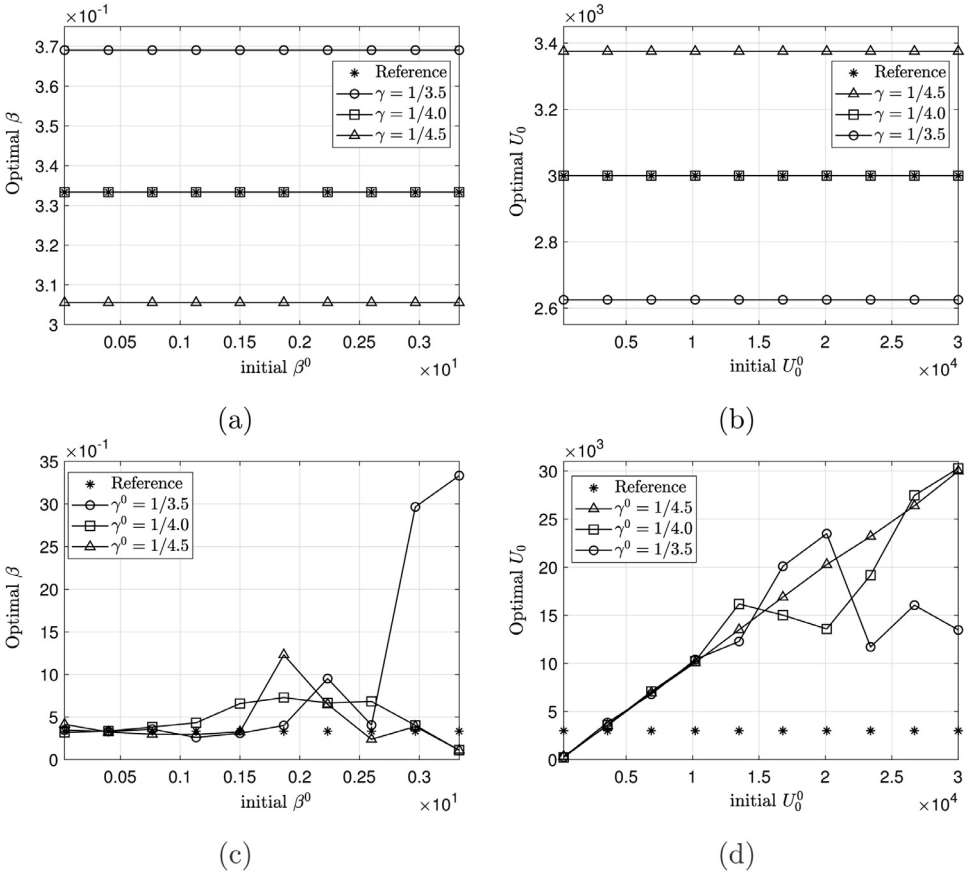

Fig. 2. Comparison of the optimal values between the (a) and (b) proposed and (c) and (d) previous methods depending on the initial guess values: (a) and (c) β; (b) and (d) U_0.

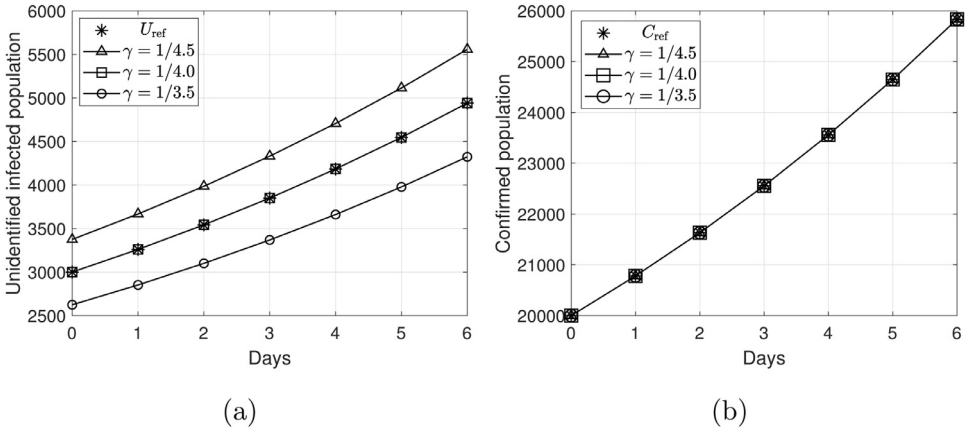

Fig. 3. Computational results using the generated confirmed data: (a) unidentified infected and (b) confirmed populations.

Table 3
Initial conditions for generating the data listed in Table 2.

Parameter	Value
N	5×10^7
U_0	3000
C_0	20, 000
S_0	$N - U_0 - C_0$
β	1/3
γ	1/4

Table 4
Relative error of the estimated U_0 and β.

Case	U_0	β
$\gamma = 1/4.5$	1.7e−10	2.0e−10
$\gamma = 1/4.0$	3.1e−13	7.9e−13
$\gamma = 1/3.5$	1.1e−10	1.9e−11

Table 5
Initial parameters for the SUC model.

Parameter	Value
N	5×10^7
U_0^0	$2(C_{\text{ref}}(1) - C_{\text{ref}}(0)) = 1564.1$
C_0	$C_{\text{ref}}(0) = 20,000$
S_0	$N - U_0 - C_0$
β^0	0.5
γ	1/3.5, 1/4, 1/4.5

Table 7
Estimated values according to the fixed parameter γ.

Case	β	γ	$\beta S_0/N - \gamma$	U_0	γU_0
Reference	0.3333	0.2500	0.0832	3000.0	750.0
$\gamma = 1/4.5$	0.3055	0.2222	0.0832	3375.0	750.0
$\gamma = 1/4.0$	0.3333	0.2500	0.0832	3000.0	750.0
$\gamma = 1/3.5$	0.3691	0.2857	0.0832	2625.0	750.0

Table 8
Confirmed cases data C_{ref} used for computation.

Days	Date	S. Korea	USA	India	Brazil
0	4/24	116033.6	31451885.1	15656055.4	13997593.7
1	4/25	116714.3	31511586.7	15966350.1	14055397.0
2	4/26	117390.3	31569361.0	16287956.4	14113700.4
3	4/27	118061.0	31625391.1	16618701.9	14170517.0
4	4/28	118737.1	31679185.1	16958864.3	14227049.6
5	4/29	119405.3	31731814.9	17308229.9	14283976.3
6	4/30	120054.0	31783169.9	17665270.0	14340904.0
7	5/1	120681.3	31833766.0	18030196.9	14401290.4
8	5/2	121303.1	31883181.4	18401237.6	14461566.6
9	5/3	121923.4	31931706.3	18774443.4	14521246.6
10	5/4	122547.9	31979742.9	19152518.6	14580407.0
11	5/5	123159.0	32028297.4	19533644.4	14638993.6
12	5/6	123755.1	32075840.6	19919485.3	14698325.7
13	5/7	124331.9	32121991.6	20309288.4	14756739.1
14	5/8	124919.1	32166005.3	20698960.9	14815722.7
15	5/9	125500.4	32208815.1	21090240.4	14876213.9
16	5/10	126078.1	32249867.1	21481236.3	14936200.1
17	5/11	126651.4	32289473.9	21868334.0	14997611.6
18	5/12	127218.7	32327645.1	22250589.7	15059106.0
19	5/13	127806.1	32364411.0	22625769.0	15119937.0
20	5/14	128425.3	32399782.3	22990799.1	15181253.3

Table 9
Initial values used in this section.

Country	N	U_0^0	C_0	S_0	β^0
S. Korea	5×10^7	1361.4	116033.6	$N - U_0 - C_0$	1/3
USA	3.3×10^8	119403.1	31451885.1	$N - U_0 - C_0$	1/3
India	13.6×10^8	620589.4	15656055.4	$N - U_0 - C_0$	1/3
Brazil	2.1×10^8	115606.6	13997593.7	$N - U_0 - C_0$	1/3

firmed case data in the South Korea (S. Korea), the United States of America (USA), India, and Brazil obtained from the data of the WHO Coronavirus dashboard from April 18, to May 14, 2021, as of June 1, 2021 [30]. Note that the average values of seven days of the confirmed data including each observation date, which are rounded to the second place after the decimal point, are used instead of the reported numbers of confirmed patients each day because of the difference in the number of people being tested for COVID-19 on weekdays and weekends. Refer to Table 8 for 21 days of the confirmed cases data C_{ref} for each country used in the experiments.

We compute the unidentified infected population on the current date, i.e., April 30, and both the confirmed and unidentified infected cases in the next two weeks using the first 7 days in Table 8. The confirmed case data of the last 14 days in Table 8 is used to verify the prediction. In all subsequent computations, we set $U_0 = 2(C_{\text{ref}}(1) - C_{\text{ref}}(0))$ and $C_0 = C_{\text{ref}}(0)$ unless otherwise stated.

First, we estimate and compare the unidentified infections for each value of γ using the confirmed population data that is the first 7-day data in Table 8. We compare the estimation of the unidentified infected population depending on the fixed value γ when 7-day data is used. We use the following initial parameters referred to in Table 9 and $\gamma = 1/3.5,\ 1/4,\ 1/4.5$.

Fig. 4 represents the computational results of the unidentified infected and confirmed populations. The estimated unidentified infected population is observed according to the fixed values of γ as shown in Fig. 4(a) while the estimated and real confirmed populations are consistent as shown in Fig. 4(b).

Table 6
Numbers of the unidentified infected population and those multiplied by γ with respect to different fixed values of γ.

Days		0	1	2	3	4	5	6
U	Reference	3000.0	3260.2	3542.9	3850.1	4184.0	4546.7	4940.8
	$\gamma = 1/4.5$	3375.0	3667.7	3985.8	4331.4	4707.0	5115.0	5558.4
	$\gamma = 1/4.0$	3000.0	3260.2	3542.9	3850.1	4184.0	4546.7	4940.8
	$\gamma = 1/3.5$	2625.0	2852.7	3100.1	3368.9	3661.0	3978.3	4323.2
γU	Reference	750.0	815.0	885.7	962.5	1046.0	1136.7	1235.2
	$\gamma = 1/4.5$	750.0	815.0	885.7	962.5	1046.0	1136.7	1235.2
	$\gamma = 1/4.0$	750.0	815.0	885.7	962.5	1046.0	1136.7	1235.2
	$\gamma = 1/3.5$	750.0	815.0	885.7	962.5	1046.0	1136.7	1235.2

Table 10
Seven-day data of each country: numbers of unidentified infected cases and those multiplied by γ.

Country	Days		0	1	2	3	4	5	6
S.	U	$\gamma = 1/4.5$	3076.5	3057.1	3037.9	3018.8	2999.7	2980.8	2962.0
Korea		$\gamma = 1/4.0$	2734.7	2717.5	2700.3	2683.3	2666.4	2649.6	2632.9
		$\gamma = 1/3.5$	2392.8	2377.8	2362.8	2347.9	2333.1	2318.4	2303.8
	γU	$\gamma = 1/4.5$	683.7	679.4	675.1	670.8	666.6	662.4	658.2
		$\gamma = 1/4.0$	683.7	679.4	675.1	670.8	666.6	662.4	658.2
		$\gamma = 1/3.5$	683.7	679.4	675.1	670.8	666.6	662.4	658.2
USA	U	$\gamma = 1/4.5$	272401.6	264001.7	255852.6	247947.3	240279.0	232841.0	225626.9
		$\gamma = 1/4.0$	242131.5	234667.8	227425.7	220399.1	213582.0	206968.6	200553.4
		$\gamma = 1/3.5$	211861.4	205334.0	198998.8	192850.8	186884.9	181096.2	175479.9
	γU	$\gamma = 1/4.5$	60533.7	58667.0	56856.1	55099.4	53395.3	51742.5	50139.3
		$\gamma = 1/4.0$	60532.9	58667.0	56856.4	55099.8	53395.5	51742.2	50138.4
		$\gamma = 1/3.5$	60531.8	58666.9	56856.8	55100.2	53395.7	51741.8	50137.1
India	U	$\gamma = 1/4.5$	1383330.3	1423169.6	1464058.7	1506019.5	1549073.9	1593243.9	1638551.3
		$\gamma = 1/4.0$	1229603.6	1265035.6	1301393.6	1338696.1	1376961.4	1416208.1	1456454.5
		$\gamma = 1/3.5$	1075876.8	1106901.7	1138728.6	1171372.7	1204849.0	1239172.3	1274357.6
	γU	$\gamma = 1/4.5$	307406.7	316259.9	325346.4	334671.0	344238.7	354054.2	364122.5
		$\gamma = 1/4.0$	307400.9	316258.9	325348.4	334674.0	344240.4	354052.0	364113.6
		$\gamma = 1/3.5$	307393.4	316257.6	325351.0	334677.9	344242.6	354049.2	364102.2
Brazil	U	$\gamma = 1/4.5$	261279.2	259958.4	258628.0	257288.5	255940.1	254583.0	253217.6
		$\gamma = 1/4.0$	232243.0	231073.4	229893.4	228703.3	227503.4	226294.1	225075.6
		$\gamma = 1/3.5$	203206.9	202188.5	201158.8	200118.1	199066.8	198005.2	196933.6
	γU	$\gamma = 1/4.5$	58062.0	57768.5	57472.9	57175.2	56875.6	56574.0	56270.6
		$\gamma = 1/4.0$	58060.8	57768.4	57473.4	57175.8	56875.9	56573.5	56268.9
		$\gamma = 1/3.5$	58059.1	57768.1	57473.9	57176.6	56876.2	56572.9	56266.7

Table 11
When using the first 7-day data of each country, the estimated values according to the fixed parameter γ.

Country	Case	β	γ	$\beta S_0/N - \gamma$	U_0	γU_0
S.	$\gamma = 1/4.5$	0.2164	0.2222	−0.0063	3076.5	683.7
Korea	$\gamma = 1/4.0$	0.2443	0.2500	−0.0063	2734.7	683.7
	$\gamma = 1/3.5$	0.2801	0.2857	−0.0063	2392.8	683.7
USA	$\gamma = 1/4.5$	0.2112	0.2222	−0.0313	272401.6	60533.7
	$\gamma = 1/4.0$	0.2419	0.2500	−0.0313	242131.5	60532.9
	$\gamma = 1/3.5$	0.2814	0.2857	−0.0313	211861.4	60531.8
India	$\gamma = 1/4.5$	0.2538	0.2222	0.0284	1383330.3	307406.7
	$\gamma = 1/4.0$	0.2819	0.2500	0.0284	1229603.6	307400.9
	$\gamma = 1/3.5$	0.3181	0.2857	0.0285	1075876.8	307393.4
Brazil	$\gamma = 1/4.5$	0.2330	0.2222	−0.0050	261279.2	58062.0
	$\gamma = 1/4.0$	0.2628	0.2500	−0.0050	232243.0	58060.8
	$\gamma = 1/3.5$	0.3011	0.2857	−0.0050	203206.9	58059.1

Table 12
Description of parameters for the basic reproduction number.

Model	Parameters	Description
SIR	S_{SIR}	Susceptible population
	I_{SIR}	Infected population
	R_{SIR}	Removed population
	β_{SIR}	Average number of contacts made by an infected individual per time
	γ_{SIR}	Recovery rate
	$\mathcal{R}_0$	Basic reproduction number in the SIR model
	D_R	Average period of time until an infected person is recovered ($=1/\gamma_{SIR}$)
SUC	S	Susceptible population
	U	Unidentified infected population
	C	Confirmed population
	β	Average number of contacts made by an infected individual per time
	γ	Rate that the unidentified infected individuals become confirmed
	D_C	Average period of time until an unidentified infected person is confirmed ($=1/\gamma$)

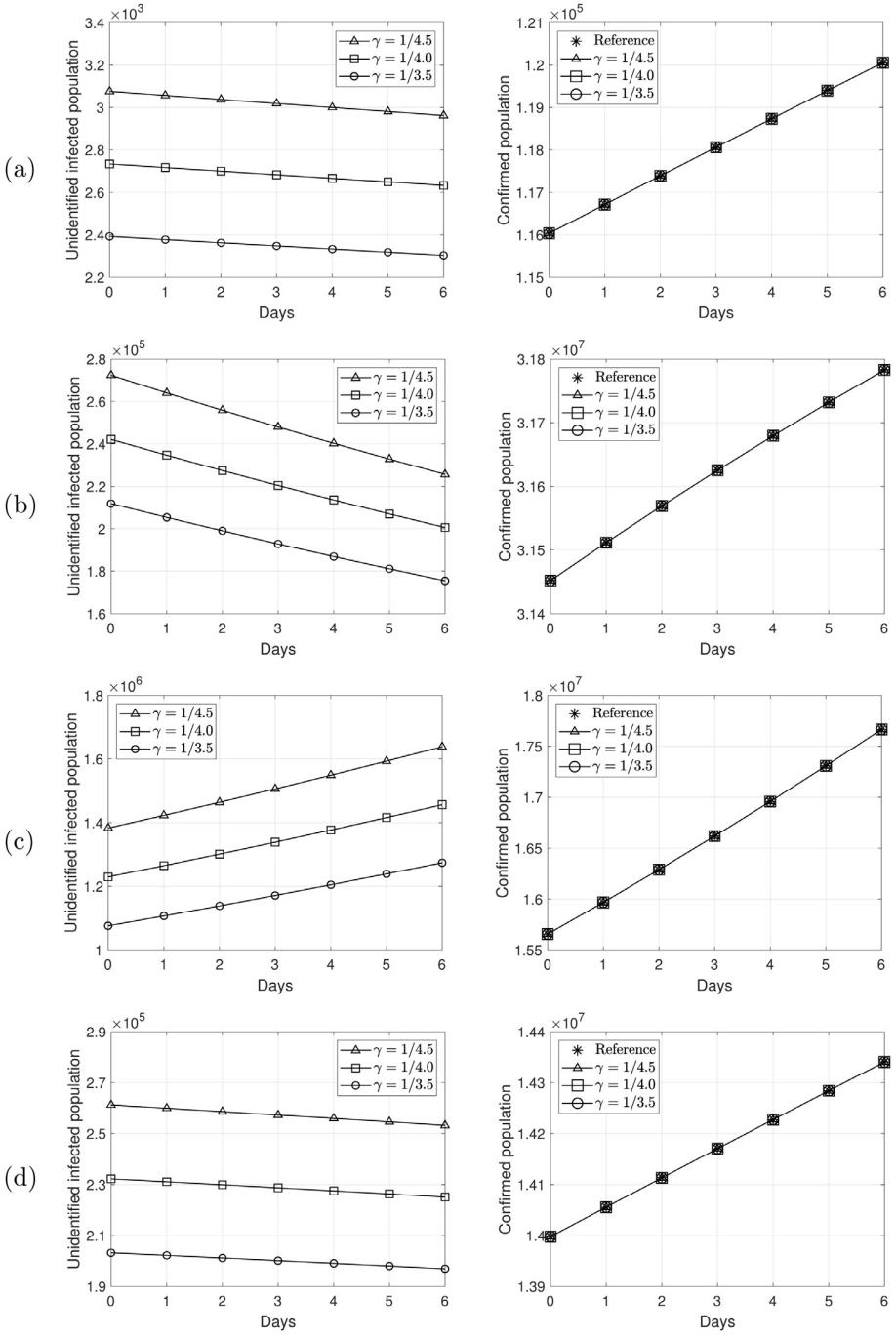

Fig. 4. Computational results from the first 7-day confirmed data of (a) S. Korea, (b) USA, (c) India, and (d) Brazil in Table 8: the unidentified infected (left) and confirmed (right) populations.

Table 10 lists the estimated values of U and γU with respect to γ from the confirmed data of four countries. Depending on the γ value used, U_0 takes different values, however, γU_0 is almost constant.

Table 11 shows the estimated values of β, γ, $\beta S_0/N - \gamma$, U_0, and γU_0 of the four countries using the first 7-day data. Note that the values of $\beta S_0/N - \gamma$ and γU_0 are approximately constant despite using different values of γ.

Now, we predict the confirmed and unidentified infected population for the next two weeks from April 30, using the same initial settings and the optimal parameters $(\beta,\ U_0)$ obtained from the above test, that is, from the recent 7-day confirmed case data. Fig. 5(a) shows the computational results using the confirmed data of each country. Since we obtain the optimal parameters for fitting the predicted and actual numbers of confirmed cases from our proposed method, the confirmed data for the next two weeks is well predicted as shown in the left column of Fig. 5(a). In addition, when $\gamma = 1/4$, the predicted data of U is shown in the right column of Fig. 5(b).

4. Discussion

We presented the robust method for parameter estimation using the SUC model that is a building block model for evaluating

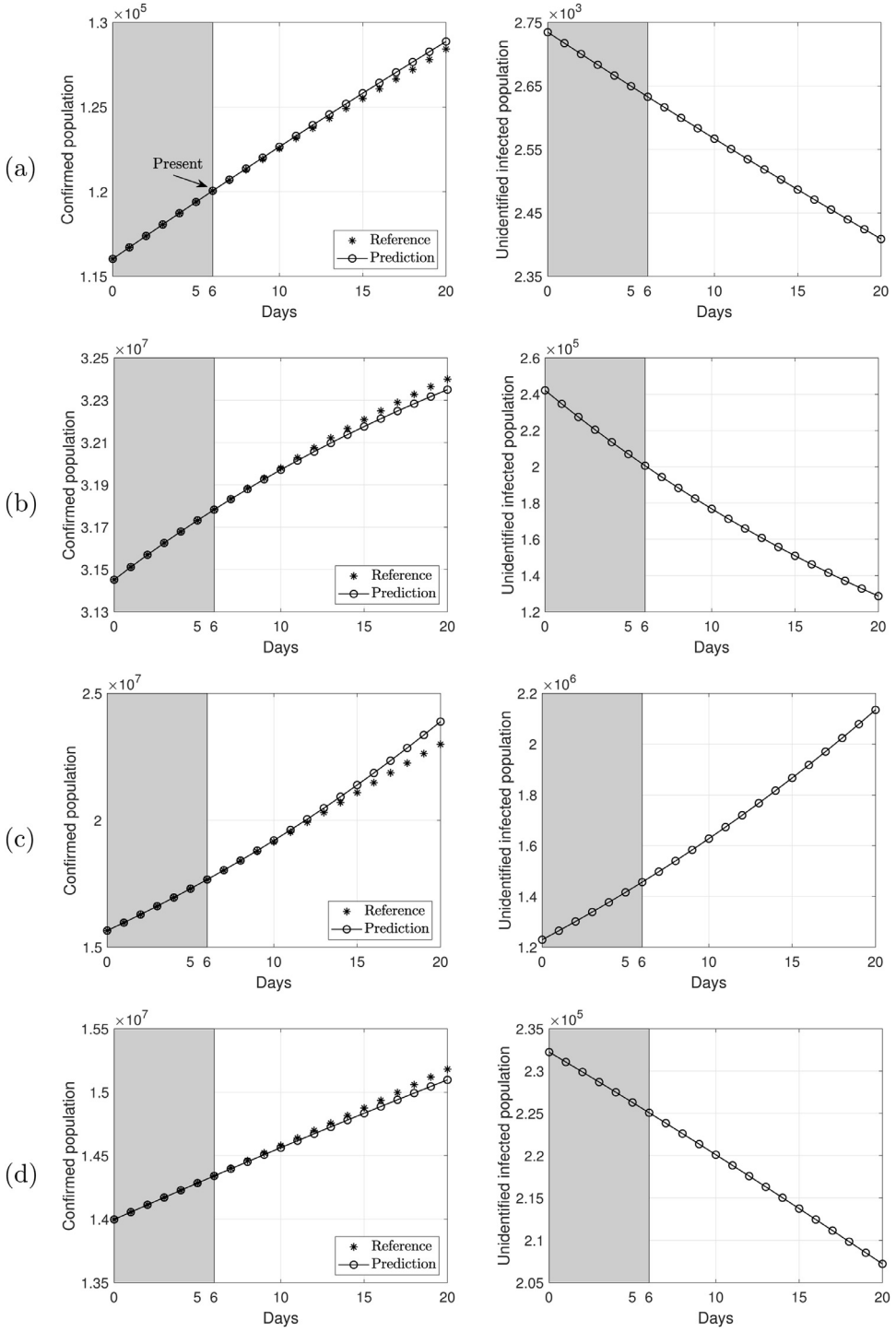

Fig. 5. Prediction of the confirmed population (left column) and the unidentified infected population (right column) when $\gamma = 1/4$ in the four countries: (a) S. Korea, (b) USA, (c) India, and (d) Brazil.

the unidentified infected population U. The basic and simple SUC model captures the main dynamics of COVID-19 instead of complex models which consider detailed information or divide the entire populations into multiple subgroups by specific criteria. A complex model that considers more various variables may reflect the real world better, however, it is difficult to collect data suitable for the model, and there are difficulties in that there are many parameters to be fitted. On the other hand, using the SUC model that contains only essential variables, the published data is sufficient as the input data and the fitted parameters are minimized.

In the COVID-19 pandemic situation, where community transmission and clusters of cases are dominant, quickly isolating people who are infected but not confirmed (i.e., the unidentified infected cases) is of paramount importance to prevent the spread

Table 13
Basic reproduction numbers of four countries.

Country	β	γ	β/γ	$\mathcal{R}_0$
S. Korea	0.2443	0.2500	0.9771	3.4198
USA	0.2419	0.2500	0.9678	3.3873
India	0.2821	0.2500	1.1285	3.9497
Brazil	0.2628	0.2500	1.0512	3.6791

of the infectious disease. Estimating the number of U can provide guidelines for government policies such as the scale of hospital beds, medical staff, and medicine and medical supplies to be secured. In addition, the estimated trend of U can be used as a reference for social distancing policies or vaccine distribution strategies. Consequently, the SUC model is so intuitive and persuasive that can be understood by public health authorities and health professionals who are not experts in mathematical modeling.

Although the SUC model has such advantages, there is no uniqueness of the optimal parameter values for the model owing to an underdetermined system of the epidemic equations with limited case data. Our focus concentrated on developing the robust method for fitting the unknown parameters and estimating U. Hence, the proposed method is not sensitive to the initial guesses for the parameters. It was verified through comparison with the previous work [23] and the computational experiments performed with data from four countries where the spread of COVID-19 was seriously progressing. In future work, we present theoretical analysis such as a backward bifurcation analysis [6] for the advanced SUC model which introduces the fractional derivatives or includes more variables for reflection of the real-world situation [19,20].

Next, let us consider the basic reproduction number $\mathcal{R}_0$, one of the important indicators to quantify the transmission potential for infectious diseases. $\mathcal{R}_0$ is defined as the number of secondary infections caused by one infectious individual in a completely susceptible population during the infectious period [31,32]. We consider the basic reproduction number in the SUC model using the definition of $\mathcal{R}_0$ in the SIR model [33]. Table 12 lists the description of parameters for the basic reproduction number in the SIR and SUC models.

The basic reproduction number in the SIR model is $\mathcal{R}_0 = \beta_{\text{SIR}}/\gamma_{\text{SIR}}$, and it can be rewritten as follows because β implies the same concept in both SIR and SUC models:

$$\mathcal{R}_0 = \frac{\beta_{\text{SIR}}}{\gamma_{\text{SIR}}} = \frac{\beta}{\gamma} \cdot \frac{\gamma}{\gamma_{\text{SIR}}}.$$

Note that $\gamma_{\text{SIR}} = 1/D_R$ and $\gamma = 1/D_C$. Thus, it holds that

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \cdot \frac{D_R}{D_C}. \tag{12}$$

Here, D_R is known to be around 14 days [33] and D_C is approximately 4 days [28,29]. Therefore, using Eq. (12) and $\gamma = 1/4$, we compute and list the basic reproduction numbers $\mathcal{R}_0$ of four countries, S. Korea, USA, India, and Brazil, in Table 13. The countries were investigated in Section 3.2.

The computed $\mathcal{R}_0$ from Eq. (12) may be different from the value from other models because the SUC model is one of the most simple and basic and the unidentified infected cases are unknown in the real world.

5. Conclusion

We presented a robust optimal parameter estimation for the SUC epidemic model. In order to predict the future dynamics of infectious disease, it is first necessary to obtain the optimal values of the parameters required for the epidemic model. However, the uniqueness of the optimal parameters cannot be guaranteed in the model due to the limitation of the given information, and then there is a problem that the predicted result varies depending on the initial setting. To overcome a problem in finding optimal parameters, we solved an optimization problem of determining the other parameter values which best fit the reported confirmed population while one parameter was fixed. We demonstrated that the effectiveness of the proposed method through several computational experiments with the manufactured and real-world data. From the numerical results, the sensitivity to the initial conditions was greatly reduced. Therefore, it is able to estimate the number of the unidentified infected population more reasonably than the method used in the previous study of the SUC model. In future work, we will introduce the fractional derivatives into the SUC model such as Ref. [19,20]. Depending on the fractional order, it is expected that the estimated values are fitted better. Moreover, we will extend the SUC model by taking into account more variables related to the real trend of the pandemic. The SUC model currently in use assumes many strict restrictions, that is, there are no reinfections, births and deaths, vaccines, etc. For more accurate estimation, we plan to develop a model with additional appropriate variables.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Chaeyoung Lee: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Soobin Kwak:** Data curation, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Sangkwon Kim:** Data curation, Investigation, Validation, Visualization, Writing – original draft. **Youngjin Hwang:** Data curation, Investigation, Validation, Visualization, Writing – original draft. **Yongho Choi:** Funding acquisition, Investigation, Validation, Writing – original draft, Writing – review & editing. **Junseok Kim:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing.

Acknowledgments

The first author (C. Lee) was supported by the National Research Foundation (NRF), Korea, under project BK21 FOUR. The author (Y. Choi) was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF2020R1C1C1A0101153712). The corresponding author (J.S. Kim) was supported by Korea University Research Grant. The authors appreciate the reviewers for their constructive comments, which have improved the quality of this paper.

Appendix

The following code is the main program, which is available from the corresponding author's webpage: https://mathematicians.korea.ac.kr/cfdkim/open-source-codes/.

```
clear all; format long; global N g C0;
% Make data
N=5.0e+7; bi=1/3; gi=1/4; C(1)=20000; U(1)=3000; S(1)=N-C(1)-U(1);
p=7; Nt=1000*(p-1); dt=0.001; t=linspace(0,p-1,p);
for i = 1:Nt
    S(i+1) = S(i) - dt*bi*S(i)*U(i)/N;
    U(i+1) = U(i) + dt*(bi*S(i)*U(i)/N - gi*U(i));
    C(i+1) = C(i) + dt*gi*U(i);
end
refR0=bi/gi; refS=S(1:1000:end); refU=U(1:1000:end); refC=C(1:1000:end);
% Estimation
clear S U C;
Tdata=[0:p-1]; g=1/4.5; b0=0.5; U0=2*(refC(2)-refC(1)); C0=refC(1);
param = lsqcurvefit(@SUCmodel,[b0 U0],Tdata,refC);
b=param(1);U(1)=param(2);C(1)=C0;S(1)=N-U(1)-C(1);
t=linspace(0,dt*Nt,Nt+1);
for i = 1:Nt
    S(i+1) = S(i) - dt*b*S(i)*U(i)/N;
    U(i+1) = U(i) + dt*(b*S(i)*U(i)/N - g*U(i));
    C(i+1) = C(i) + dt*g*U(i);
end
figure(1); plot(Tdata,U(1:1000:end),'ko-'); grid on; box on
xlabel('Days'); ylabel('Unidentified infected population');
figure(2); plot(Tdata,refC,'k*-',Tdata,C(1:1000:end),'ko-');
legend('Data','Numerical solution'); grid on; box on
xlabel('Days'); ylabel('Confirmed population')

function f = SUCmodel(Parameter, Tdata)
global N g C0;
b = Parameter(1); U(1) = Parameter(2); C(1) = C0; S(1) = N-C(1)-U(1);
Nt = 1000*Tdata(end); dt = 0.001;
for i = 1:Nt
    S(i+1) = S(i) - dt*b*S(i)*U(i)/N;
    U(i+1) = U(i) + dt*(b*S(i)*U(i)/N - g*U(i));
    C(i+1) = C(i) + dt*g*U(i);
end
t = linspace(Tdata(1),Tdata(end),Nt+1); f = interp1(t,C,Tdata);
end
```

References

[1] Headey D, Heidkamp R, Osendarp S, Ruel M, Scott N, Black R. Impacts of COVID-19 on childhood malnutrition and nutrition-related mortality. Lancet 2020;396(10250):519–21.
[2] Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol 2020;19:141–54.
[3] Calabró L, Peters S, Soria JC, Di Giacomo AM, Barlesi F, Covre A. Challenges in lung cancer therapy during the COVID-19 pandemic. Lancet Resp Med 2020;8(6):542–4.
[4] Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC. Remdesivir for the treatment of COVID-19. N Engl J Med 2020;383(19):1813–26.
[5] Graham BS. Rapid COVID-19 vaccine development. Science 2020;368(6494):945–6.
[6] Agusto FB, Khan MA. Optimal control strategies for dengue transmission in Pakistan. Math Biosci 2018;305:102–21.
[7] Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. Lancet Infect Dis 2020;20(11):1247–54.
[8] Dauji S. Sen's innovative method for trend analysis of epidemic: a case study of COVID-19 pandemic in India. Trans Indian Eng 2020;6(2):507–21.
[9] Kasinathan P, Montoya OD, Gil-González W, Arul R, Moovendan M, Dhivya S. Application of soft computing techniques in the analysis of COVID-19: a review. Eur J Mol Clin Med 2020;7(6):2480–503.
[10] Matouk AE. Complex dynamics in susceptible-infected models for COVID-19 with multi-drug resistance. Chaos Solitons Fractals 2020;140:110257.
[11] Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. Chaos Solitons Fractals 2020;139:110057.
[12] He S, Peng Y, Sun K. SEIR modeling of the COVID-19 and its dynamics. Nonlinear Dyn 2020;101(3):1667–80.
[13] Korolev I. Identification and estimation of the SEIRD epidemic model for COVID-19. J Econ 2021;220(1):63–85.
[14] Kim S, Seo YB, Jung E. Prediction of COVID-19 transmission dynamics using a mathematical model considering behavior changes in Korea. Epidemiol Health 2020;42:e2020026.
[15] Kim S, Kim YJ, Peck KR, Jung E. School opening delay effect on transmission dynamics of coronavirus disease 2019 in Korea: based on mathematical modeling and simulation study. J Korean Med Sci 2020;35(13):e143.
[16] Peter OJ, Qureshi S, Yusuf A, Al-Shomrani M, Idowu AA. A new mathematical model of COVID-19 using real data from Pakistan. Results Phys 2021;24:104098.
[17] Memon Z, Qureshi S, Memon BR. Assessing the role of quarantine and isolation as control strategies for COVID-19 outbreak: a case study. Chaos Solitons Fractals 2021;144:110655.

[18] Musa SS, Qureshi S, Zhao S, Yusuf A, Mustapha UT, He D. Mathematical modeling of COVID-19 epidemic with effect of awareness programs. Infect Dis Model 2021;6:448–60.
[19] Khan MA, Atangana A. Modeling the dynamics of novel coronavirus (2019-nCoV) with fractional derivative. Alexandria Eng J 2020;59(4):2379–89.
[20] Khan MA, Atangana A, Alzahrani E. The dynamics of COVID-19 with quarantined and isolation. Adv Differ Equ 2020;2020(1):1–22.
[21] Ivorra B, Ferrández MR, Vela-Pérez M, Ramos AM. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. Commun Nonlinear Sci Numer Simul 2020;88:105303.
[22] Melis M, Littera R. Undetected infectives in the COVID-19 pandemic. Int J Infect Dis 2021;104:262–8.
[23] Lee C, Li Y, Kim J. The susceptible-unidentified infected-confirmed (SUC) epidemic model for estimating unidentified infected population for COVID-19. Chaos Solitons Fractals 2020;139:110090.
[24] Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts. Infect Dis Model 2017;2(3):379–98.
[25] López L, Rodo X. A modified SEIR model to predict the COVID-19 outbreak in spain and Italy: simulating control scenarios and multi-scale epidemics. Results Phys 2021;21:103746.
[26] Almeida R, Bastos NR, Monteiro MTT. Modeling some real phenomena by fractional differential equations. Math Meth Appl Sci 2016;39(16):4846–55.
[27] López CP. Optimization techniques via the optimization toolbox. In: MATLAB optimization techniques. Berkeley, CA: Apress; 2014. p. 85–108.
[28] Krishna MV. Mathematical modelling on diffusion and control of COVID-19. Infect Dis Model 2020;5:588–97.
[29] Abu-Raddad LJ, Chemaitelly H, Malek JA, Ahmed AA, Mohamoud YA, Younuskunju S. Two prolonged viremic SARS-CoV-2 infections with conserved viral genome for two months. Infect Genet Evol 2021;88:104684.
[30] Daily cases and deaths by date reported to WHO. WHO Coronavirus (COVID-19) Dashboard, available online: https://covid19.who.int/WHO-COVID-19-global-data.csv.
[31] Allen LJ. An introduction to mathematical biology. Upper Saddle River. New Jersey: Pearson Education; 2007.
[32] Ganyani T, Faes C, Chowell G, Hens N. Assessing inference of the basic reproduction number in an SIR model incorporating a growth-scaling parameter. Stat Med 2018;37(29):4490–506.
[33] Katul GG, Mrad A, Bonetti S, Manoli G, Parolari AJ. Global convergence of COVID-19 basic reproduction number and estimation from early-time SIR dynamics. PLoS One 2020;15(9):e0239800.