**R₀ estimation for COVID-19 pandemic through exponential fit**

Zheng Mingliang¹ | Theodore E. Simos²,³,⁴,⁵,⁶ | Charalampos Tsitouras⁷

1 College of Mechanical and Electrical Engineering, Taihu University of Wuxi, Wuxi 214064, China
2 College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610225, China
3 Scientific and Educational Center “Digital Industry”, South Ural State University, 76 Lenin Ave., 454 080, Chelyabinsk, Russia
4 Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan
5 Data Recovery Key Laboratory of Sichun Province, Neijing Normal University, Neijiang, China
6 Section of Mathematics, Department of Civil Engineering, Democritus University of Thrace, Xanthi, Greece
7 General Department, National and Kapodistrian University of Athens, GR 34400, Euripus campus, Greece

**Correspondence**
Theodore E. Simos, 10 Konitsis St., GR-175 64 Athens, Greece.
Email: tsimos.conf@gmail.com

Communicated by: T. Monovasilis

**Funding information**
Natural Science Foundation of Universities in Jiangsu Province,
Grant/Award Number: 20KJD460001

**1 | INTRODUCTION**

The SIR epidemic model is defined by a system of initial value problem equations. It was introduced by Kermack and McKendrick in their hallmark work.¹ The population P is divided in three parts, namely, the susceptible S, the infectious I, and the recovered R. These numbers change in time z. Thus, we consider them as functions S(z), I(z), and R(z). People are infected and are transferred from compartment S to compartment I with rate β(z) according to

$$\frac{dS(z)}{dz} = -\frac{\beta(z)S(z)I(z)}{P^2}.$$
Analogously, people that recover are transferred from compartment $I$ to compartment $R$ with rate $\gamma(z)$. No back transfers are allowed, that is, there is no movement from compartment $R$ to compartments $I$ or $S$ nor movements from compartment $I$ to $S$.

In this simplified consideration, we assume that there are no exists group behaviors and the population does not divided in age intervals. We also presume absence of births and deaths within the population which remains constant. Thus, $P = S(z) + I(z) + R(z)$. In the following, we use the corresponding percentages setting

$$r(z) = \frac{R(z)}{P}, s(z) = \frac{S(z)}{P}, i(z) = \frac{I(z)}{P}.$$  

Since the population remains constant, we also notice

$$\frac{ds(z)}{dz} + \frac{di(z)}{dz} + \frac{dr(z)}{dz} = 0,$$

and in consequence

$$i(z) + s(z) + r(z) = 1 = i(0) + s(0) + r(0). \tag{1}$$

The actual SIR epidemic model is described then by the following system of three differential equations$^2$–$^4$

$$\begin{align*}
\frac{ds(z)}{dz} &= -\beta(z)s(z)i(z), \\
\frac{di(z)}{dz} &= \beta(z)s(z)i(z) - \gamma(z)i(z), \\
\frac{dr(z)}{dz} &= \gamma(z)i(z).
\end{align*} \tag{2}$$

The system above along with starting values

$$s_0 = s(0), i_0 = i(0) \text { and } r_0 = r(0),$$

form an initial value problem which in view of Equation (1) is reduced to

$$\begin{align*}
\frac{ds(z)}{dz} &= -\beta(z)s(z)(1 - s(z) - r(z)), \\
\frac{dr(z)}{dz} &= \gamma(z)(1 - s(z) - r(z)). \tag{3}
\end{align*}$$

Initial value problems (2) and (3) can’t be solved analytically. Thus, we are usually approximate their solutions numerically through Runge–Kutta methods.$^5$–$^8$

Our main concern here is the reproduction number which is defined as

$$R_0 = \frac{\beta(z)}{\gamma(z)}.$$  

But solving Equation (3) is not the issue in our present study. We actually know $s(z)$ and $r(z)$ for various points in time, and we are interested in estimating the varying parameters $\beta(z)$ and $\gamma(z)$ in order to get $R_0$.

2 \  | \ **EXPONENTIAL FITTING OF THE DATA**

After setting as $z_n$ the current time, we also define as $r_n = r(z_n)$ and $s_n = s(z_n)$ the current observed values of recovered and susceptible people. Then, the previous time step is named $z_{n-1}$. The time step is usually the past day. We assume that $z_{j} \in \mathbb{Z}$ for all $j = n, n - 1, n - 2, \ldots \$. Thus, $z_n = z_{n-1} + 1$, $z_{n-1} = z_{n-2} + 1$, and so forth. The values
are given. The population is also known. Then, from Equation (3), we may derive the formula

$$\frac{ds}{s} = -\frac{\beta(z)}{\gamma(z)} dr = -R_0 dr.$$

Integrating the results, we get

$$s(z) = s(z_0) e^{-R_0 (r(z) - r(z_0))}, \quad (4)$$

and in consequence after having knowledge of various past values of $s$ and $r$, we arrive at $R_0$ given by the formula

$$R_0 = \frac{\log s_n - \log s_{n-k}}{r_n - r_{n-k}}. \quad (5)$$

### 2.1 Constant $R_0$

In case of $\beta(z)$ and $\gamma(z)$ being constants $\beta$ and $\gamma$, respectively, the formula (4) is exact. We will verify this in MATLAB, by choosing $\beta(z) = \beta = 0.3$ and $\gamma(z) = \gamma = 0.12$ and initial values

$$s(0) = s(z_0) = 0.999, r(0) = r(z_0) = 0.$$

Then, we may estimate the values of $s$ and $r$ for $z_{1} = 1, z_{2} = 2, \ldots, z_{50} = 50$, writing the lines below in MATLAB and copy the screen here

```matlab
>> be=0.3; ga=0.12;
>> fcn=@(z,y) [-be*y(1)*(1-y(1)-y(2))
               ga*(1-y(1)-y(2))];
>> [zout,yout]=ode113(fcn,(0:1:50),[0.999 0], ...
                     odeset(abstol,100*eps,reltol,100*eps));
>> % set the artificial data
>> s=yout(:,1);r=yout(:,2);
>>$\approx$ estimate Ro by using time steps $z_{21}=21$ & $z_{30}=30$
>> format long
>> -(log(s(31))-log(s(22)))/(r(31)-r(22))
ans =
2.499999999999999
>>
>> % estimate Ro using any random couple of steps
>> -(log(s(17))-log(s(13)))/(r(17)-r(13))
ans =
2.500000000000005
```

Thus, using $s$ and $r$ from any two distinct time points, we may recover by Equation (5) the exact constant value of $R_0$. Thus, our new approach clearly outperforms older ones where four to five digits of error were accepted.

### 2.2 Variable $R_0$

The above is not true for the case of parameters varying with time. Then, we may use Equation (4) as model for applying an exponential fitting to the data. We consider knowing the data $s_n, s_{n-1}, \ldots, s_{n-k}$ and $r_n, r_{n-1}, \ldots, r_{n-k}$. Then, considering $s$ as function with respect to $r$, we fit a function of the form

$$s(r) = e^{-R_0(r-R)}.$$
This is a decaying function in accordance to the monotonicity of the data since \( r \) is ascending and \( s \) is decaying. We actually ask for a linear fitting to the data \( r_n, r_{n-1}, \ldots, r_{n-k} \) and \( \log s_n, \log s_{n-1}, \ldots, \log s_{n-k} \). \( R_0 \) is the slope of this linear least squares fit. As conclusion

\[
R_0 \approx \frac{(k + 1) \sum_{j=0}^{k} r_{n-j} \log s_{n-j} - \sum_{j=0}^{k} r_{n-j} \sum_{j=0}^{k} \log s_{n-j}}{(k + 1) \sum_{j=0}^{k} r_{n-j} - \left( \sum_{j=0}^{k} r_{n-j} \right)^2},
\]

(6)

Since \( r \) is clearly ascending and \( s \) is descending, we may easily verify the reliability of the formula above, that is, \( R_0 \geq 0 \) always.

The question risen now is the magnitude of \( k \), that is, how many data to use. We circumvent this by an iterative approach. We start by setting \( k = 1 \) and get the estimation

\[
R_0^{(1)} \approx \frac{\log s_n - \log s_{n-1}}{r_n - r_{n-1}}.
\]

We proceed with

\[
R_0^{(2)} \approx \frac{3 \cdot (r_n \log s_n + r_{n-1} \log s_{n-1} + r_{n-2} \log s_{n-2}) - (r_n + r_{n-1} + r_{n-2}) (\log s_n + \log s_{n-1} + \log s_{n-2})}{3 \cdot (r_n^2 + r_{n-1}^2 + r_{n-2}^2) - (r_n + r_{n-1} + r_{n-2})^2}.
\]

and in case that \(|R_0^{(2)} - R_0^{(1)}| < 0.01\), we accept \( R_0^{(2)} \) as the value on demand. Otherwise, we proceed with \( k = 3 \) in Equation (6) and finding \( R_0^{(3)} \). We again ask if \(|R_0^{(2)} - R_0^{(3)}| < 0.01\) and accept \( R_0^{(3)} \) as the value on demand. We continue this procedure until we find some distance lower than 1/100. In case of repeated failures, we stop for \( k = 20 \) and set as \( R_0 \). We programmed this in MATLAB, and the corresponding listing is given in Appendix A as function \texttt{r0exp}.

3 | DATA FOR COVID-19

We may find data for the COVID-19 outbreak online.\(^{12}\) There we may retrieve daily numbers of confirmed, recovered, and deaths. The data are shown in the following unformatted display

| Date       | Country   | Confirmed | Recovered | Deaths |
|------------|-----------|-----------|-----------|--------|
| 2020-02-04 | China     | 23707     | 843       | 491    |
| 2020-02-05 | China     | 27440     | 1115      | 563    |
| 2020-02-06 | China     | 30587     | 1477      | 633    |
| 2020-02-07 | China     | 34110     | 1999      | 718    |
| 2020-02-08 | China     | 36814     | 2596      | 805    |
| 2020-02-09 | China     | 39829     | 3219      | 905    |
| ...        |           |           |           |        |

The vector \( r \) is formed after calculating over each line

\[
r(\text{date}) = \frac{\text{recovered}+\text{deaths}}{N},
\]

while the vector \( s \) is formed as

\[
s(\text{date}) = \frac{\text{confirmed}}{N},
\]

where the population for various countries was retrieved from Wikipedia.\(^{13}\) The data online\(^{12}\) are not reliable in many cases. We experience missing or inaccurately reported data for some countries. This can be disastrous if there are constantly fault reports. See the example in Table 1, where 10 deaths and no recoveries were reported for 21 consecutive days while there were about 350 infections. Little can be done to treat such cases. We arrive at \( R_0 \approx 30 \) then. Something that certainly wasn’t the case in Greece in June 2020.
At the starting phase, when pandemic outbreaks, we may observe numbers like those presented in Table 2. It does not seem to exist any method that can extract reliable results from the data above only. A single country’s data cannot be used to alarm an emerge. It needs to have a global view.

It is obvious that data have to be reported correctly and have some visible fluctuation.

| Date  | Country | Confirmed | Recovered | Deaths |
|-------|---------|-----------|-----------|--------|
| ...   | ...     | ...       | ...       | ...    |
| 4/6/2020 | Greece | 2952 | 1374 | 180 |
| 5/6/2020 | Greece | 2967 | 1374 | 180 |
| ...   | ...     | ...       | ...       | ...    |
| 23/6/2020 | Greece | 3302 | 1374 | 190 |
| 24/6/2020 | Greece | 3310 | 1374 | 190 |
| ...   | ...     | ...       | ...       | ...    |

**TABLE 1** Data for Greece in June 2020

| Date  | Country | Confirmed | Recovered | Deaths |
|-------|---------|-----------|-----------|--------|
| ...   | ...     | ...       | ...       | ...    |
| 30/1/2020 | Italy | 0 | 0 | 0 |
| 31/1/2020 | Italy | 2 | 0 | 0 |
| ...   | ...     | ...       | ...       | ...    |
| 19/2/2020 | Italy | 3 | 0 | 0 |
| 20/2/2020 | Italy | 3 | 0 | 0 |
| ...   | ...     | ...       | ...       | ...    |

**TABLE 2** Data for Italy at the start of the COVID-19 outbreak

**FIGURE 1** Brazil: $R_0$ estimation between January 1, 2021 until April 4, 2021

**FIGURE 2** Canada: $R_0$ estimation between January 1, 2021 until April 4, 2021
We present diagrams with the estimation of $R_0$ for various countries for 2021. The actual dates are from January 1 to April 4. Thus, in Figure 1, we present results for Brazil. In Figures 2–6, we present results for Canada, Germany, Israel, Italy, and Russia, respectively.

Interpreting the results, we verify that at the beginning of April 2021 only Israel and Russia experience $R_0 < 1$. Particularly for Israel, we observe a decisive drop of $R_0$ below 1/2 which came after extensive vaccination of the population. $R_0$ for Russia remained below 1 for almost all 2021. $R_0$ for Brazil was above 1 for almost all 2021. R0 for Canada,
Germany, and Italy climbed above 1 after March. For smoothing the curves, we chose $k = 10$ in the fifth line of the program in Appendix A.

5 | CONCLUSION

Fitting the data available for COVID-19 outbreak (i.e., susceptible and recovered population) with a simple decaying exponential model is proposed. The fit is exact if there is a constant rate of infection and recovery. Otherwise, we get reliable results since the monotonicity of the fit to the data is retained.

ACKNOWLEDGEMENT

The first author (Zheng Mingliang) has received external funding: Natural Science Foundation of Universities in Jiangsu Province (20KJD460001).

CONFLICTS OF INTEREST

There are no conflicts of interest to this work.

ORCID

Zheng Mingliang https://orcid.org/0000-0001-5681-0051
Theodore E. Simos https://orcid.org/0000-0002-9220-6924
Charalampos Tsitouras https://orcid.org/0000-0001-6801-8117

REFERENCES

1. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics—I. Proc R Soc A. 1927;115:700-721.
2. Beckett SJ, Dominguez-Mirazo M, Lee S, Andris C, Weitz JS. Spread of COVID-19 through Georgia, USA. Near-term projections and impacts of social distancing via a meta population model. https://doi.org/10.1101/2020.05.28.20115642; 2020.
3. Korsbo N, Jonsson H. Its about time: analysing simplifying assumptions for modelling multi-step pathways in systems biology. PLoS Comput Biol. 2020;16:e1007982.
4. Weitz JS, Beckett SJ, Coenen AR, et al. Modeling shield immunity to reduce COVID-19 epidemic spread. Nat Med. 2020;26:849-854.
5. Medvedev MA, Simos TE, Tsitouras C. Fitted modifications of Runge-Kutta pairs of orders 6(5). Math Meths Appl Sci. 2018;41:6184-6194.
6. Tsitouras C. Optimized explicit Runge-Kutta pair of orders 9(8). Appl Numer Math. 2001;38:123-134.
7. Tsitouras C. Runge–Kutta pairs of order 5(4) satisfying only the first column simplifying assumption. Comput Maths Appl. 2011;62:770-775.
8. Tsitouras C. Explicit Runge-Kutta methods for starting integration of Lane-Emden problem. Appl Math Comput. 2019;354:353-364.
9. MATLAB version. The Mathworks, Inc., Natick, MA; 2019b.
10. Medvedeva MA, Simos TE, Tsitouras C, Katsikis VN. Direct estimation of SIR model parameters through second order finite differences. Math Meths Appl Sci. 2021;44:3819-3826.
11. Simos TE, Tsitouras C, Kovalnogov VN, Fedorov RV, Generalov DA. Real-time estimation of R0 for COVID-19 spread. Mathematics. 2021;9:664.
APPENDIX: A

The MATLAB program \( r0_{\text{exp}} \) is presented below. We avoided the explicit use of Equation (6). The built-in function \( \text{polyfit} \) is used instead.

```matlab
function r0=r0exp(s,r);
% give 21 last percentage values of s (susceptible) \& r (recovered)
% embed successive exponential fits based on various past points
r0prev=-1e16;r0=1e16;
k=2;r=r(:);s=s(:);
while abs(r0prev-r0)>0.01 & k<21,
    r0prev=r0;
    r0=polyfit(r(end-k:end),log(s(end-k:end)),1);
    r0=-r0(1);
    k=k+1;
end
```

How to cite this article: Mingliang Z, Simos TE, Tsitouras C. \( R_0 \) estimation for COVID-19 pandemic through exponential fit. *Math Meth Appl Sci*. 2022;45:1632–1639. https://doi.org/10.1002/mma.7878