Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Comment on “Estimation of COVID-19 dynamics “on a back-of-envelope”: Does the simplest SIR model provide quantitative parameters and predictions?”

S. Rojas
Departamento de Física, Universidad Simón Bolívar, PO BOX 85000, Caracas 1080-A, Venezuela

1. Introduction

In a recent article published in this journal [1], after some (unnecessary) considerations, the author presents the logistic function (Eq. (8) in [1]) as an alternative solution of the differential equation known as the Kermack and McKendrick 1927 approximation [2] of the SIR epidemiological model [3,4] in order to fit data regarding the cumulative confirmed cases of COVID-19 infected cases from some countries. Clearly, the proposed logistic function (Eq. (8) in [1]) does not accomplish the initial condition \( R(0) = 0 \).

In this note we show that the data of the countries discussed in reference [1] and of a few other countries (see Fig. 1) can also be fitted using the \( R(t) \) solution obtained from the Kermack and McKendrick 1927 approximation of the SIR model, making it unnecessary the use of the Verhulst (logistic) Eq. (8) in [1].

2. Methods

2.1. Brief analytical analysis

The SIR model considers a population of size \( N \) on which, at time \( t \), \( S(t) \) individuals are susceptible of being infected as a consequence that \( I(t) \) individuals are already infected and can transmit or spread the disease to the susceptible population. The number of individuals \( R(t) \) represents those who have recovered from the disease (which, if lethal, also includes death individuals) and can not be reinfected. Thus, the dynamics of the disease, introduced in 1927 by Kermack and McKendrick [2], is modeled by the set of differential equations:

\[
\frac{dS}{dt} = -\beta IS \quad (1)
\]

\[
\frac{dI}{dt} = \beta IS - \gamma I \quad (2)
\]

\[
\frac{dR}{dt} = \gamma I \quad (3)
\]

\[
\frac{d}{dt}(S + I + R) = 0 \rightarrow S + I + R = N \quad (4)
\]

In these equations, the parameters \( \beta \) (the infection rate) and \( \gamma \) (the recovery or removal rate of infectives) are constants: \( \beta \) controls the transition between \( S \) and \( I \), Eq. (1), while \( \gamma \) controls the transition between \( I \) and \( R \), Eq. (3).

For an epidemic to occur [2–5], the number of infected individuals needs to increase from the initial number of infected individuals \( I_0 \). This condition will happen if at time zero, \( S_0 = \rho = \gamma / \beta \). That is, \( \rho \) represents a critical value for an epidemic to occur and the SIR model reveals a threshold phenomenon [6].

From a dimensional point of view, assigning no units to \( S \), \( I \), \( R \), and \( N \) the parameters \( \beta \) and \( \gamma \) have units of inverse of time (measured typically in days, weeks or months in epidemiological records).

Quantitatively, while the interaction in the form of the product \( SI \) makes it difficult to determine the parameter \( \beta \) from observed
epidemiological data, from Eq. (3) the inverse of the parameter \( \gamma \) gives a measure of the time spent by individuals in the infectious stage. Consequently, by carefully observing the development of an infectious disease, the parameter \( \gamma \) can be estimated (as the inverse of the recovered or infectious period) by epidemiologists from epidemiological records. One should be aware that neither of the parameters \( \beta \) or \( \gamma \) remains constant as the infection evolves [2–5]. Moreover, the assumptions on which the model are built are no longer valid as soon as sanitary interventions are applied to control the infection.

As discussed in the epidemiological literature [2–4], a straightforward combination of the SIR model Eqs. (1)–(3) leads to a non-linear differential equation for \( dR/dt \), interpreted as the properly counted individuals removed (either because they have recovered or death) from medical units:

\[
\frac{dR}{dt} = \gamma \left[N - R - S0 \exp\left(-\frac{R}{\rho}\right)\right] \tag{5}
\]

For not severe epidemics, Kermack and McKendrick [2] considered \( R(0)/\rho < 1 \) and proposes that \( dR/dt \) could be approximated by:

\[
\frac{dR}{dt} = \gamma \left[N - S0 + R\left(\frac{S0}{\rho} - 1\right) - \frac{S0}{2} \left(\frac{1}{\rho}\right)^2 R^2\right]. \tag{6}
\]

Considering that \( S0 \left(\frac{1}{\rho}\right)^2 > 0 \), \( \left(\frac{S0}{\rho} - 1\right) > 0 \), and \( N - S0 = l0 > 0 \), the solution of Eq. (6) can be written in the form Kermack and McKendrick [2], Keeling and Rohani [4]

\[
R(t) = \alpha \left(\frac{S0}{\rho}\right)^2 \left[\frac{1}{\alpha} \left(\frac{S0}{\rho} - 1\right) + \tanh \left(\frac{\alpha}{2} \sqrt{\gamma t - \phi}\right)\right]. \tag{7}
\]

where \( \tanh(x) \) is the hyperbolic tangent of \( x \), and

\[
\alpha = \sqrt{\left(\frac{S0}{\rho} - 1\right)^2 + 2 \left(\frac{S0}{\rho}\right)^2 \left(\frac{N}{S0} - 1\right)}, \tag{8}
\]

\[
\phi = \tanh^{-1} \left(\frac{S0}{\alpha} - \frac{1}{\alpha}\right). \tag{9}
\]

Here \( \tanh^{-1}(x) \) is the inverse of the hyperbolic tangent of \( x \).

From Eq. (7), we also obtain the Kermack and McKendrick approximated solution (or the KM approximation) of the SIR model [2]

\[
\frac{dR}{dt} = \frac{\gamma}{2} \sqrt{\frac{\alpha \rho}{S0}} S0 \sech^2 \left(\frac{\alpha}{2} \sqrt{\gamma t - \phi}\right). \tag{10}
\]

where \( \sech(x) \) is the hyperbolic secant of \( x \). Kermack and McKendrick were able to study a Bombay 1905–1906 plague using Eq. (10)
2.2. Brief numerical analysis

Using cumulative confirmed cases data reported by the European Centre for Disease Prevention and Control [7] regarding the coronavirus COVID-19 pandemic outbreak, we used computing routines to fit data using Eq. (7) written in the form:

\[ R(t) = C_0 + C_1 \tanh(C_2 t - C_3), \quad (11) \]

setting \( C_0 = C_1 \tanh(C_2) \) to meet the initial condition \( R(t = 0) = 0 \).

To find numerical solution of the SIR model, Eqs. (1)–(3), in addition to the parameters \( \beta \) and \( \gamma \) we also need to know initial conditions \( S_0 = S(t = 0), I_0 = I(t = 0), \) and \( R_0 = R(t = 0) \). As required by the SIR model, we set \( R_0 = 0 \). As already mentioned, an estimated for \( \gamma \) could be obtained from epidemiological records as its inverse \( (1/\gamma) \) determines the average infectious period of the disease [4]. According to the European Centre for Disease Prevention and Control regarding the coronavirus COVID-19 pandemic [8] the infectious period is “...estimated to last for 7–12 days in moderate cases and up to two weeks on average in severe cases.” Accordingly, \( \gamma \) for computation used in this comment were set to yield an infectious period in that range.

For \( \beta, S_0, \) and \( I_0 \) it is not easy to have observed estimated values. To find reasonable starting values for them we applied a heuristic approach which turns out to be helpful in order to find numerical solution of the full SIR model adjusting itself to data fitted by the Kermack and McKendrick solution in Eq. (10). Then, by a standard trial and error approach we were able to find suitable parameters for solving the full SIR model adjusting itself to the COVID-19 cases reported in this comment.

To have an idea of how well each one of the fit adjust itself to the data, we use the Root Mean Square Error (rmse) and the Relative Root Mean Square Error (rmseRel), defined as follows:

\[ \text{rmse} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - F_i)^2} \quad (12) \]

\[ \text{rmseRel} = \text{rmse} / \max(O) \quad (13) \]

Here \( O_i \) is the \( i \)th observation in the considered data set; \( F_i \) is the corresponding value obtained by the corresponding fitting method; and \( \max(O) \) is the maximum value in the considered data set. As the uncertainty in the observed values \( O_i \) is unknown [9], it is unrealistic to emphasize any further statistical measure characterizing the estimated parameters used in the analysis for the COVID-19 pandemic data set.

At this point it should be mentioned that the numerical computational work in this comment was carried out via the Python scripting programming language and the Numpy/SciPy/Matplotlib libraries described elsewhere [10,11].

3. Results

The data for analysis comes from European Centre for Disease Prevention and Control [7], and the period covered at the moment of start writing this note was (for most countries) December 31, 2019–June 29, 2020.

Compiled in Table 1 are the parameters that best fit the data (from each studied country) to the function \( R(t) \) of Eq. (7) expressed in the form of Eq. (11). The results are shown in Fig. 1. The reported rmse and relative rmse values are indicative that a reasonable fit has been attained. The corresponding fit is also indicative that it is unnecessary to use the logistic function (i.e. Eq. (8) in [1]) as an alternative to the Kermack and McKendrick approximated Eq. (6) to fit the data.

Similarly, in Table 2 we compiled values of the quantities required to find numerical solution of the full SIR epidemiological model defined by Eqs. (1)–(3) for each country whose results is given in Fig. 1. Also, the reported rmse and relative rmse values are indicative that a reasonable match has been attained. The results are indicative that the SIR model is a good choice to get a better understanding of COVID-19 data.

4. Concluding remarks and future work

Cumulative confirmed cases from the coronavirus COVID-19 data reported by the European Centre for Disease Prevention and
Control [8] of a number of countries were found to fall under the Kermack and McKendrick 1927 [2] approximated solution (7) of the SIR epidemiological model. We were also able to show that the full SIR model could be solved numerically adjusting itself to the analyzed data.

Since other, more complex, alternative approaches to the problem has been proposed [12], at this point it is hard to establish for sure which model better describe the evolution of the coronavirus COVID-19 pandemic [9]. Consequently, given that the SIR model captures some of the COVID-19 data behavior, it could provide guidance to get better insight on the evolution of the pandemic as the only two parameters ($\beta$ and $\gamma$) entering in the model are more or less well understood by epidemiologists and can be guessed from the data.

Consequently, before considering more complex models (requiring much more parameters than the SIR model), it is clear that a better qualitatively understanding of the parameters $\beta$ and $\gamma$ in addition to the initial condition $I_0, S_0$ (restricted to $N = I_0 + S_0$) is necessary to give an appropriated quantitative account of an epidemic.

We are confident that the methodology applied in the development of this comment could also be extended to analyze other sets of data.

Declaration of Competing Interest

The author has no competing interest to declare.

Acknowledgments

The author is grateful to an anonymous referee who kindly provided useful comments to improving this article.

References

[1] Postnikov EB. Estimation of COVID-19 dynamics “on a back-of-envelope”: does the simplest sir model provide quantitative parameters and predictions? Chaos Solitons Fractals 2020;135(109841):1–6.
[2] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc Roy Soc Lond A 1927;115:700–21.
[3] Murray JD. Mathematical biology. I. An introduction. 3rd ed. Springer; 2002.
[4] Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2008.
[5] Li MY. An introduction to mathematical modeling of infectious diseases. Springer; 2018.
[6] Stanley HE. Introduction to phase transitions and critical phenomena. Oxford University Press; 1987.
[7] ECDC. Geographic distribution of COVID-19 cases worldwide. 2020. https://opendata.ecdc.europa.eu/covid19/casedistribution/csv [last accessed June 29, 2020].
[8] European Centre for Disease Prevention and Control, Q & A on COVID-19. https://www.ecdc.europa.eu/en/covid-19/questions-answers.
[9] Kramer D. COVID-19 pandemic modeling is fraught with uncertainties. Phys Today 2020;73(6):25–7.
[10] Rojas S, Christensen EA, Blanco-Silva FJ. Learning SciPy for numerical and scientific computing. Packt Publishing Ltd.; 2015.
[11] Rojas S. Numerical and scientific computing with SciPy (Book-Video). Packt Publishing Ltd.; 2017.
[12] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals 2020;134(109761):1–5.