

# Observing the epidemiological *SIR* model on the COVID-19 data

S. Rojas

*Departamento de Física, Universidad Simón Bolívar, Venezuela.  
e-mai: srojas@usb.ve*

Received 16 June 2020; accepted 10 July 2020

This article shows that in the period January 22–June 28, 2020, the combined data set of cumulative recoveries and deaths from the current coronavirus COVID-19 pandemic falls on the Kermack and McKendrick approximated solution of the epidemiological *SIR* contagious disease model. Then, as an original contribution of this work, based on the knowledge of the infectious period of any epidemic, a methodology is presented in order to find numerical solutions of the full *SIR* model that reproduce the observed data of the epidemic in case it could be described by the *SIR* model. The methodology is first illustrated by finding a solution of the *SIR* model that describes the epidemic data of the Bombay plague of 1905–06 analyzed by Kermack and McKendrick. After that, the methodology is applied to numerically solve the full set of differential equations of the *SIR* contagious model on the above mentioned coronavirus COVID-19 pandemic data set. We also show that the Kermack and McKendrick approximation is observed on the counted combined aggregated recovered and deaths cases from some individual countries and also in some of the cumulative confirmed COVID-19 cases of individual countries.

**Keywords:** Epidemiological *SIR* model; computational physics; physics problem solving; computational modeling; Riccati differential equation.

Este artículo muestra que para el período 22 de enero–28 de junio, 2020, los datos acumulados del conjunto combinado de recuperaciones y muertes producto de la actual pandemia del coronavirus COVID-19 se ajustan a la solución aproximada encontrada por Kermack y McKendrick del modelo epidemiológico para enfermedades contagiosas *SIR*. Luego, como contribución original del presente trabajo, basada en conocimiento del período de infección de alguna epidemia, se presenta una metodología que facilita el obtener soluciones numéricas del modelo completo *SIR* que se ajustan a los datos observados de la epidemia en caso que la misma se pueda describir por tal modelo. La metodología primero se ilustra encontrando una solución del modelo *SIR* que se ajusta a los datos de la plaga de 1905–06 en Bombay estudiada por Kermack y McKendrick. Seguidamente, la metodología se aplica para encontrar soluciones numéricas del modelo completo *SIR* que se ajustan al conjunto de datos de la pandemia coronavirus COVID-19 arriba mencionados. Adicionalmente, se muestra que la aproximación de Kermack y McKendrick se observa en la contabilidad combinada de recuperaciones y muertes agregadas de algunos países y también en la data de casos confirmados contagiados de COVID-19 de algunos países.

**Descriptores:** Modelo epidemiológico *SIR*; física computacional; resolución de problemas en física; modelaje computacional; ecuación diferencial de Ricatti.

PACS: 05.45.Pq; 02.70.Hm; 01.40.gb; 01.40.Ha

DOI: <https://doi.org/10.31349/RevMexFisE.18.35>

## 1. Introduction

The current, social global context dominated by the coronavirus COVID-19 pandemic [1] brings a direct opportunity to consider within Physics the quantitative modeling of infectious diseases like the *SIR* epidemic model which, in spite of its simplicity, has been successfully used in modeling historical data of some epidemics [2–5].

In epidemiology, models can be used to control (and perhaps to eradicate) the infectious disease under consideration by the model. That is, epidemiological models can be used to take action on how to approach an epidemic from the medical point of view. They can, for instance, provide reasoned estimates for the level of vaccination necessary for the control of an infectious disease (for examples see [3] and references therein). Certainly, great care must be exercised before practical use can be made of any epidemic model. Generally, models go through several revisions before they can be exploited with some degree of confidence in the control of an contagious infection. To be illustrative, as referenced in [3], several hypothetical revision scenarios were applied to the

model proposed by Capasso and Paveri-Fontana [6] for the 1973 cholera epidemic in the port city of Bari, in southern Italy.

Following the organization of this work, in the next Sec. 2, we will give a brief description of the *SIR* epidemiological model (referenced from now on as the *SIR* model) and its relevance on the analysis of data from the current coronavirus COVID-19 pandemic. Next, in Sec. 3, the approximated solution of the *SIR* model proposed by Kermack and McKendrick is presented. Then, in Sec. 4, it will be discussed the applicability of the solution found on analyzing observed epidemic data by first presenting a well know epidemic example, namely the Bombay plague (occurred from December 1905 to July 1906). We will provide numerical solutions to the full *SIR* model of differential equations for the Bombay case, which has not been found in the consulted literature. Then, following this discussion, data observed from the coronavirus COVID-19 pandemic will be analyzed in the context of the *SIR* model. Finally, conclusions and extension of this work will be presented in Sec. 6.

## 2. The *SIR* model

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 deceased 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). From a dimensional point of view, one does not assign units to the variables  $S$ ,  $I$ ,  $R$ , and  $N$  whilst the parameters  $\beta$  and  $\gamma$  have units of inverse time (measured typically in days, weeks or months in epidemiological records). Notice that Eq. (1) expresses the interaction between  $S$  and  $I$  (at time  $t$ ) as the product  $SI$  and that a fraction of this product are the individuals that at time  $t$  become infected and removed from  $S$  (which, because of the negative sign in Eq. (1), decreases as time increases). This interaction in the form of the product  $SI$  makes difficult to determine the parameter  $\beta$  from observed epidemiological data. On the other hand, 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 relatively precisely by epidemiologists from epidemiological records (as the inverse of the recovered or infectious period). For the initial conditions, at time  $t = 0$ , we have  $S_0 \equiv S(t = 0) > 0$ ,  $I_0 \equiv I(t = 0) > 0$ , and  $R_0 \equiv R(t = 0) = 0$ . From Eq. (4), this yields  $S_0 + I_0 = N$ .

It is important to keep in mind the basic assumptions of the model. An important assumption in idealizing the model is that, once recovered, individuals become immunized and can not be reinfected. Thus, the transition goes as  $S \rightarrow I \rightarrow R$ . As reported by the world Health Organization (WHO) [7], this assumption is being observed in the evolution of the coronavirus COVID-19 pandemic infectious illness. That is, massive reinfection by the coronavirus has not been reported yet. Also, it is assumed in Eq. (4), that during the disease no people enter or leave the population (including birds and deaths by other reasons). Consequently, it is

expected that the *SIR* model could be observed at the start of an infectious disease, time at which no action has been taken to control the disease and individuals have no restriction to come into contact with one another locally. This assumption is also partially true at the start of the coronavirus COVID-19 pandemic infectious disease in a global context, before each country has blocked their borders. However, it is not easy to have an observed estimated for  $N$ ,  $I_0$ , and  $S_0$  to compare them with the values computed via the *SIR* model. The *SIR* model also assumes that as soon as a susceptible individual catches the disease, they become infected right away. It has been also reported by the World Health Organization [7] that this is the case with the coronavirus COVID-19 pandemic infectious disease: infected individuals start infecting others as soon as they catch the virus and even before presenting any symptoms. Moreover, a study has shown how undocumented infection facilitates the rapid dissemination of the illness [8]. The fact that the assumptions on which the *SIR* model was formulated are partially true in what we are experiencing with the current coronavirus COVID-19 pandemic infectious disease, makes it feasible to observe a *SIR* model behavior on the reported data of the coronavirus COVID-19 pandemic. We will show that this is the case in the studied cases in this work.

Two important aspects need to be mentioned at this point. Equation (2) can be written in the form:

$$\frac{dI}{dt} = \gamma \left( \frac{S}{\rho} - 1 \right) I, \quad (5)$$

where  $\rho = \gamma/\beta$  is sometimes called the *relative removal rate* (which might be called differently by other authors because, unfortunately, the notation and terminology in epidemiological literature is not uniform).

Now, for an epidemic to occur, 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,  $S_0 > S_c = \rho$ . That is,  $\rho$  represents a critical value for an epidemic to occur and the *SIR* model reveals a *threshold phenomenon* [9].

Another important epidemiological parameter which defines how quickly the infectious disease spreads is the *basic reproduction rate* ( $R_0$ ) of the infection, defined as

$$R_0 = \frac{\beta}{\gamma} S_0 = \frac{S_0}{\rho}. \quad (6)$$

This parameter (not to be confused with the initial value  $R_0$ ) measures the number of secondary infections produced by one primary infection in a wholly susceptible population. For instance, if  $R_0 = k$ , then, before recovering, one infected individual will likely infect  $k$  more individuals, each one of which will, in turn, infect  $k$  more individuals and so on. If more than one secondary infections is produced from the primary infection, then  $R_0 > 1$  which gives  $S_0 > \rho$  and, obviously, an epidemic ensues. One should be aware that neither of the parameters  $\beta$  or  $\gamma$  remains constant as the infection evolves. Moreover, the assumptions on which the model are

built are no longer valid as soon as sanitary interventions are applied to control the infection.

Further discussion of the *SIR* model is beyond the scope of this article, and the reader is referred to references [3,4] for a far reaching and enriching discussion of the model.

The solution of the system of equations formed by Eqs. (1)-(3) can be found numerically for given values of  $\beta$ ,  $\gamma$ ,  $S_0$ , and  $N$ . In general, the parameters are unknown and, consequently, a best fit procedure is carried out in order to find a solution, assuming that the *SIR* model is a plausible model for the epidemic. In the reported literature, only a few epidemics have been solved numerically. In this work we provide evidences that the COVID-19 pandemic can be described by this model.

### 3. Approximated solution of the *SIR* model

As discussed in the epidemiological literature [2-4], a straight forward 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.

For not severe epidemics  $R(t)/\rho < 1$ , the landmark work [2] of Kermack and McKendrick (1927) proposes that  $dR/dt$  could be approximated by [2,4]

$$\frac{dR}{dt} = \frac{\gamma}{2} \left( \frac{\alpha\rho}{S_0} \right)^2 S_0 \operatorname{sech}^2 \left( \frac{\alpha}{2} \gamma t - \phi \right), \quad (7)$$

where  $\operatorname{sech}(x)$  is the hyperbolic secant of  $x$  and

$$\alpha = \sqrt{\left( \frac{S_0}{\rho} - 1 \right)^2 + 2 \left( \frac{S_0}{\rho} \right)^2 \left( \frac{N}{S_0} - 1 \right)}, \quad (8)$$

$$\phi = \tanh^{-1} \left( \frac{\frac{S_0}{\rho} - 1}{\alpha} \right). \quad (9)$$

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

An important point to keep in mind is that contrary to Physics, where experiments could be designed to obtain data to test models, in epidemiology data might be available from naturally occurring epidemics (in the Kermack and McKendrick work only one example was presented) and the collected data might be incomplete due to underreporting (as is observed with the coronavirus COVID-19 pandemic [10]). This lack of reliable data makes the estimation of parameters very imprecise. Nevertheless, models can guide the epidemiologists to fine tune (perhaps by trial and error testing) their point of view on the quantitative development of an infectious disease.

### 4. Observing the *SIR* model on epidemic (including COVID-19) data

Summarizing the conceptualizations of previous sections, the set of Eqs. (1)-(4) represents the epidemiological *SIR* model

which, under a few basic conditions, describes the normal time spread in a population of an infectious disease without medical intervention. This set of equations needs to be solved numerically and, in order to do that, in addition to the initial conditions, two parameters ( $\beta$  and  $\gamma$ ) are required.

The initial conditions for the quantities  $S$  and  $I$  are not easy to determine and, according to the consulted literature [3,4], while the parameter  $\gamma$  can be estimated from observing the evolution of the considered epidemic, that is not the case with the parameter  $\beta$ .

The landmark study of Kermack and McKendrick [2] provides an analytical approximated solution (called from now on the KM approximation) of the *SIR* model of Eq. (7) and is supposed to be valid at the start of an epidemic, or in the cases in which the epidemic is not severe. To support his work, Kermack and McKendrick were able to show how nicely the approximated solution of the *SIR* model was able to describe data from a 1905-1906 plague in Bombay, where 80 to 90 percent of the cases reported terminate fatally. As-

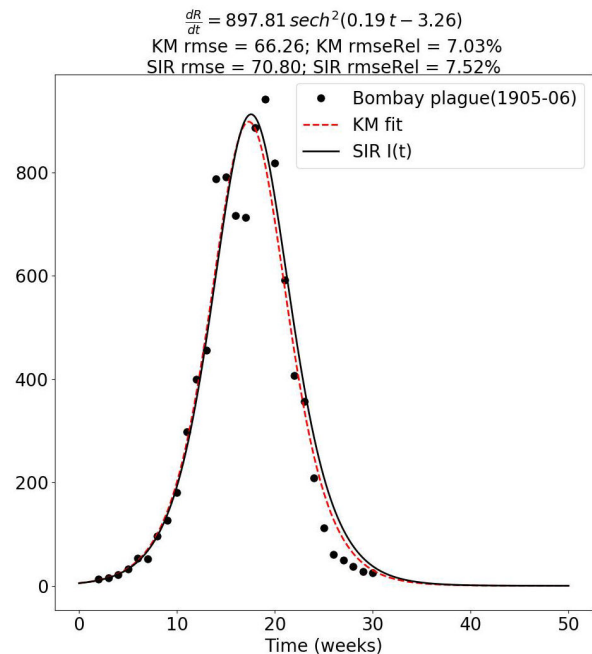


FIGURE 1. This plot shows that the Bombay plague of 1905-06 falls under the KM approximation [2]. The graph also shows one possibility of the  $I(t)$  function obtained by integration (according to the proposed method described in Sec. 5) of the full *SIR* model (Eqs. (1)-(3)), using the initial conditions and parameters  $N = 780030$ ,  $S_0 = 780025$ ,  $I_0 = N - S_0 = 5$ ,  $R_0 = 0$ ,  $\gamma = 7.469$  (given by hand), and  $\beta = \gamma/\rho = 1.0056e - 05$  ( $\rho = 742709.03$ ). Since we were unable to find the actual data used by Kermack and McKendrick [2], the data were approximated (using the GNU Image Manipulation Program *Gimp* [11]) from the Fig. 2.3, shown in page 23 of Ref. 4. The value of  $\gamma$  was chosen to have  $N$  (from Eq. (21)) close to the Bombay population near the year of the epidemic (obtained from a Wikipedia page [12], where it is mentioned that “in the census of 1901, the population had actually fallen to 780,000.”) This figure can easily be updated with more appropriated epidemiological information from the Bombay plague (including a reasoned value of  $\gamma$ ).

suming the fatalities represent the left hand side of Eq. (7), they concluded that the data falls under the curve  $dR/dt = 890\text{sech}^2(0.2t - 3.4)$ . Since then, only a very few infectious diseases have been reported to follow the KM approximation [3].

As a bonus of this work, via the methodology presented in Sec. 5 we can obtain Fig. 1, which shows a numerical solution of the *SIR* model adjusting itself to the data of the Bombay plague studied by Kermack and McKendrick. It is interesting to mention that until now we have been unable to find (other than the KM approximation) a full solution of the *SIR* model adjusted to the Bombay plague. This is indicative of how difficult is to guess appropriate values for the initial conditions and the parameters necessary to perform a successfully integration of the *SIR* model falling on the data. The methodology given in Sec. 5 helps to fill this gap.

**4.1. About the coronavirus COVID-19 pandemic data**

All coronavirus COVID-19 pandemic data used for analysis in this work are publicly available under the *Open Data Commons Public Domain and Dedication License* [13], from

the source listed in [14]. According to the provided description in the website, data is updated daily, starting January 22, 2020. It contains files listing cumulative confirmed cases, cumulative reported deaths and cumulative reported recoveries, disaggregated by country (and sometimes subregions) taken from sources like the *World Health Organization* (WHO) [15] and the *Johns Hopkins University Center for Systems Science and Engineering* [16], among others. The period covered at the moment of writing this works was January 22-June 28, 2020. Other studies [10] have also considered cumulative data on modeling the progression of the coronavirus COVID-19 infectious disease.

**4.2. Analysis of the worldwide aggregated COVID-19 data**

We start our study by first considering worldwide aggregated COVID-19 data [17]. The top left plot on Fig. 2 shows the cumulative reported cases (confirmed, deaths, and recovery) as of June 28, 2020. In the top middle and top right plots on Fig. 2 it is observed how well the data (cumulative combined deaths and recoveries sets, as required by the *SIR* model) falls

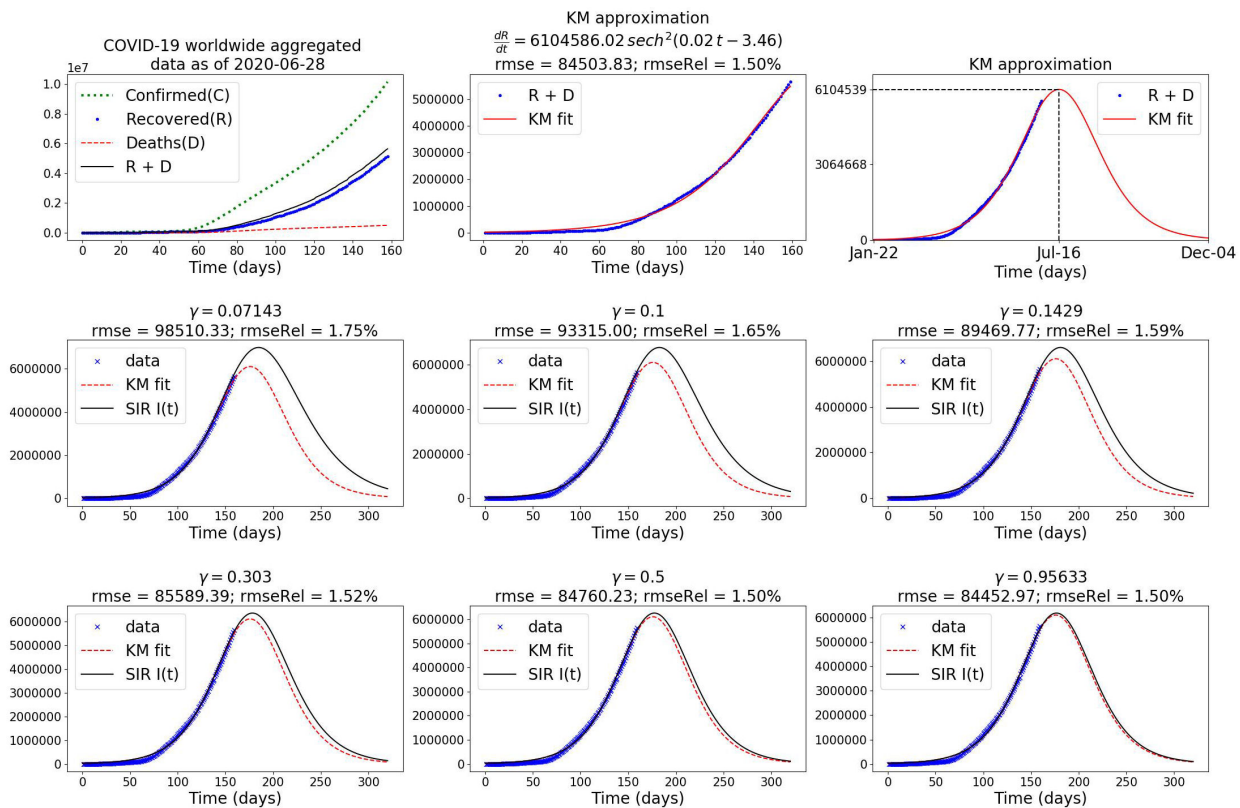


FIGURE 2. The top left plot shows the reported world wide cumulative aggregated COVID-19 data for the period January 22 - June 28, 2020, available in [17]. The top middle and top right plots show that the data points fall under the KM approximation, given by Eq. (7) (the inverse time unit was omitted deliberately from the function). The plots on the middle and bottom rows are examples of different scenarios adjusted to the data by solving the full *SIR* model of Eqs. (1)-(3) via the methodology presented in Sec. 5 which allows the finding of initial conditions and the parameter  $\beta$  if the parameter  $\gamma$  is given (perhaps from observing the evolution of an epidemic that falls under the KM approximation). Solving the *SIR* model could be used to anticipate hypothetical scenarios (as the ones shown here) on how the epidemic would evolve. Notice that the *SIR* solution approaches the KM approximation as  $\gamma$  increases, but the other parameters also change (see Table I for details).

under the KM approximation, defined by the Eq. (7) (which is shown in the title of the top middle plot on Fig. 2 where the inverse time unit was omitted deliberately).

At this point, it is important to internalize that the shape of the KM approximation (the hyperbolic secant squared) was provided by the SIR model. Accordingly, instead of just guessing a best fitting function to a data set, we are finding that the considered coronavirus COVID-19 data falls on a functional shape obtained via the theoretical framework defined by the SIR model. That is, we have shown that the considered data set is well described by the proposed SIR model.

Continuing the analysis, let's mention that in the title of the top middle plot on Fig. 2 it is also shown values for the *Root Mean Square Error* (rmse) and the *Relative Root Mean Square Error* (rmseRel), defined as follows:

$$\text{rmse} = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - KM_i)^2}, \quad (10)$$

$$\text{rmseRel} = \frac{\text{rmse}}{\max(O)}. \quad (11)$$

Here  $O_i$  is the  $i^{\text{th}}$  observation in the considered  $O$  data set,  $KM_i$  is the corresponding value obtained by the KM approximation, and  $\max(O)$  is the maximum value in the considered  $O$  data set. These measures are used as a way to

compare how well different sets of data fall under the SIR model. As the uncertainty in the observed values  $O_i$  is unknown [10], it is unrealistic to emphasize any further statistical measure characterizing any estimated parameters via the KM approximation for the COVID-19 pandemic data set.

Let's mention that every numerical computational work in this article was carried out via the Python scripting programming language and the Numpy/Scipy/Matplotlib libraries described elsewhere [18,19]. In addition, all of the available data from January 22, 2020 until June 28, 2020 were used in the computations. Further information from the epidemiology community is necessary in order to see if any prediction can be achieved by this model in the future. Nevertheless, some insight could be gained by trying to get a well reasoned understanding on why the coronavirus COVID-19 data still (after 6 months) falls under the functional shape anticipated by the SIR model.

Now, considering the fact that contrary to Physics, where experiments could be designed to obtain data to test models, in epidemiology data might be only available from naturally occurring epidemics and the collected data might be incomplete due, for example, to underreporting or any other reason as discussed in [10]. Under such circumstances, it is striking to find out that the KM approximation is observed to be followed by the world wide aggregated data and by data com-

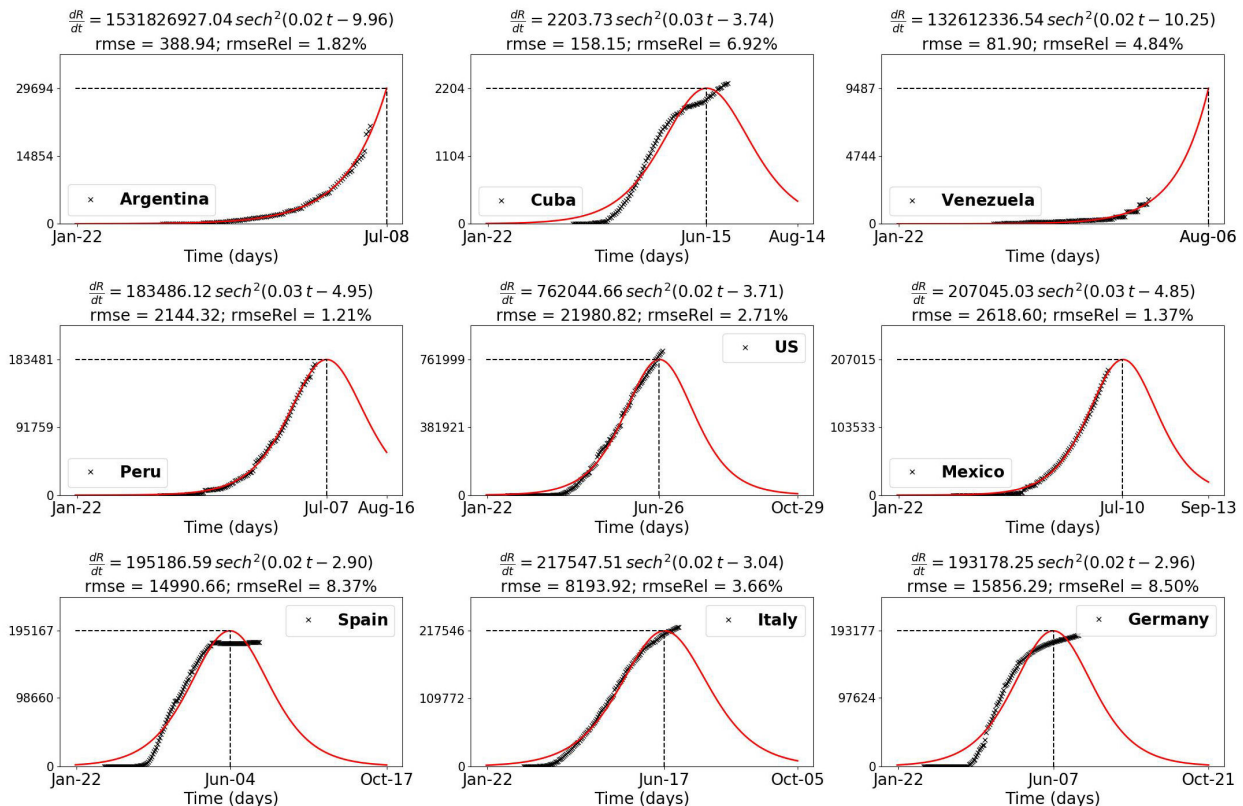


FIGURE 3. The KM approximation is observed in the reported cumulative recoveries and deaths cases of Argentina, Peru, United States of America, and Mexico. Reported data from Cuba (which seems to have started a new wave of contagions), Spain, Italy, and Germany do not fall under the KM approximation. Data source (available in [22]) covers the range January 22 - June 28, 2020. Notice that not every country has reported cases starting from January 22. The inverse time unit was omitted deliberately from the function.

ing from individual countries, as shown in Fig. 3 and 4. From an speculative point of view, this might be a consequence that each cumulative reported data set is a kind of mean value, averaging details of the pandemic evolution at the atomic (city or town) level. In this respect, by knowing how to interpret the reported results, epidemiologists can use them as a way to fine tune (perhaps by trial and error testing) their point of view on the quantitative evolution of the coronavirus COVID-19 infectious disease and offer sound guidance regarding what to expect in terms of infections and deaths for decision making by governors, legislators, mayors, and city council members. Again if the parameter  $\gamma$  is known, the methodology presented in Sec. 5 will help to find many scenarios of the full *SIR* model falling on the data.

## 5. A methodology for solving the full *SIR* model on data falling under the KM approximation

Now, as an original contribution of this work, we will next describe a procedure that can be applied to find numerical solution of the full *SIR* model (Eqs. (1)-(3)) by obtaining, via the KM approximation and knowledge of the parameter  $\gamma$ , necessary information (namely, initial conditions and the parameter  $\beta$ ) to perform the numerical integration of the *SIR* model that adjust itself to the observed data.

As previously mentioned, finding numerical solutions of the *SIR* model, Eqs. (1)-(4), requires, in addition to the initial conditions ( $S_0$  and  $I_0$ , with  $R_0 = 0$ ), knowledge of the parameters  $\beta$  and  $\gamma$ . While  $\beta$  is not easy to determine from epidemiological data [3,4], the removal or recovery rate parameter  $\gamma$ , whose reciprocal ( $1/\gamma$ ) determines the average infectious period of the disease, can be estimated from the observed evolution of an epidemic [4].

In fact, according to the *European Centre for Disease Prevention and Control* regarding the coronavirus COVID-19 pandemic [21] the infectious period is "... estimated to last for 7-12 days in moderate cases and up to two weeks on average in severe cases." Consequently, in this study one can try the set of infectious period ( $1/\gamma$ ) of 7, 10, and 14 days, giving, respectively, in turn,  $\gamma = 0.1429/\text{day}$ ,  $\gamma = 0.1/\text{day}$ , and  $\gamma = 0.07143/\text{day}$ . However, we will consider other values of  $\gamma$  to further explore its effect on the other quantities allowing the integration of the model. This uncertainty in the parameter  $\gamma$  can be rationalized by the variability in responses between different individuals over the infectious period, that in addition that the COVID-19 data by itself contains a high level of uncertainty [10].

The method is as follows. When reporting the observance of the KM approximation in data [2-4], Kermack and McKendrick considered that what is observed in the vertical axis of their graph containing the data from the Bombay plague is the quantity  $dR/dt$  (or  $\gamma I(t)$  on the left hand side of Eq. (7)). However, by studying the case of the influenza on a boarding school presented with some detail in [3], it seems to the author of this work that what is actually observed is the

quantity  $(1/\gamma)(dR/dt)$ . Consequently, considering that such is the case, Eq. (7) is better written in the form:

$$\frac{1}{\gamma} \frac{dR}{dt} = \frac{1}{2} \left( \frac{\alpha \rho}{S_0} \right)^2 S_0 \operatorname{sech}^2 \left( \frac{\alpha}{2} \gamma t - \phi \right). \quad (12)$$

Recalling that the inverse of time is the unit of the parameter  $\gamma$ , this Eq. (12) is dimensionless and represents  $I(t)$  (from Eq. (3)). Using this representation, the constants obtained from an epidemic data set under the KM approximation can be represented in the form:

$$\frac{1}{\gamma} \frac{dR}{dt} = C_1 \operatorname{sech}^2 (C_2 t - \phi), \quad (13)$$

$$C_1 = \frac{1}{2} \frac{(\alpha \rho)^2}{S_0}, \quad (14)$$

$$C_2 = \frac{1}{2} \alpha \gamma, \quad (15)$$

$$C_3 = \tanh(\phi) = \frac{1}{\alpha} \left( \frac{S_0}{\rho} - 1 \right). \quad (16)$$

Equation (16) comes from Eq. (9). From Eqs. (14)-(16) and (8) one can obtain the initial conditions and the parameter  $\beta$  in terms of the parameter  $\gamma$ , all of them necessary to solve numerically the *SIR* model:

$$\alpha = 2C_2 \frac{1}{\gamma}, \quad (17)$$

$$\rho = \frac{C_1}{C_2} \left( C_3 + \frac{1}{\alpha} \right) \gamma, \quad (18)$$

$$\beta = \frac{\gamma}{\rho}, \quad (19)$$

$$S_0 = \alpha \rho \left( C_3 + \frac{1}{\alpha} \right), \quad (20)$$

$$N = S_0 + \frac{S_0}{2} \left( \frac{\rho}{S_0} \right)^2 \left[ \alpha^2 - \left( \frac{S_0}{\rho} - 1 \right)^2 \right]. \quad (21)$$

Next, we set our rationality in action. We have already shown in Fig. 1 that our methodology works on the Bombay plague studied by Kermack and McKendrick [2]. We will now continue by showing that it also works on the considered coronavirus COVID-19 pandemic data by setting a set of  $\gamma$  values (including the ones mentioned above, inferred from the infectious period reported by the *European Centre for Disease Prevention and Control* regarding the coronavirus COVID-19 pandemic [21]) to obtain the set of values compiled on the Table I, which were obtained from Eqs. (17)-(21) using the values of  $C_1 = 6104586.02$ ,  $C_2 = 0.02$ ,  $\phi = 3.46$ , and  $C_3 = 1.00$  corresponding to the KM approximation shown in the top middle plot of Fig. 2. The values compiled in Table I were used to find numerical solutions of the *SIR* model adjusted to the coronavirus COVID-19 data. The results are shown in the plots on the rows at the middle and at the bottom of the Fig. 2 (where for completeness the KM approximation is also included).

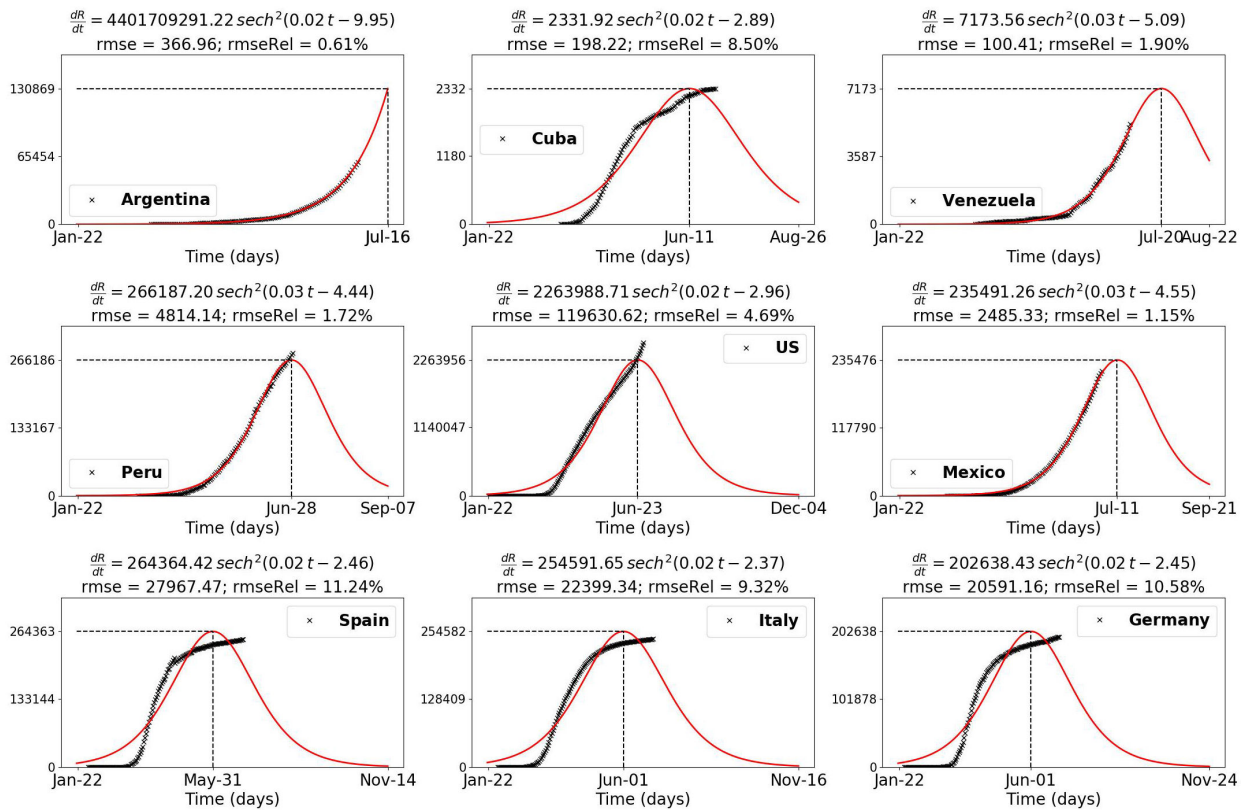


FIGURE 4. The KM approximation is observed in the reported cumulative confirmed cases of Argentina, Venezuela, Peru, and Mexico. Reported data from Cuba (which seems to have started a new wave of contagions), United States of America, Spain, Italy, and Germany do not fall under the KM approximation. Data source (available in [23]) covers the range January 22 - June 28, 2020. Notice that not every country has reported cases starting from January 22. The inverse time unit was omitted deliberately from the function.

Let's note from the values on the Table I that  $I_0 = N - S_0$  is the same on each considered scenario and that the basic reproductive ratio  $R_0$  approaches to one as the numerical solution of the *SIR* model approaches the KM approximation. The values of  $R_0$  might be indicative that the epidemic is not severe (relative to the world population, an scenario corresponding to  $\gamma = 0.95633$ ), as is implied from the KM ap-

proximation. Let's mention that the compiled scenarios in Table I are not fixed, immutable quantities. From there one can start building new ones by trial an error, adjusting the values of  $\gamma$ ,  $\beta$ ,  $I_0$ , and  $S_0$  (with the restriction that  $N = S_0 + I_0$ ). In other words, scenarios from the *SIR* model can easily be improved as information about the coronavirus COVID-19 pandemic from epidemiologists is updated.

TABLE I. The values above are used to solve numerically the epidemiological *SIR* model defined via the Eqs. (1)-(3). The results are shown in the rows at the middle and at the bottom of the Fig. 2. Given the parameter  $\gamma$ , the other quantities were computed using the Eqs. (17)-(21) and the constants  $C_1 = 6104586.02$ ,  $C_2 = 0.02$ ,  $\phi = 3.46$ , and  $C_3 = 1.00$ , obtained from the coronavirus (COVID-19) worldwide aggregated data shown in the top middle plot of the Fig. 2, satisfying the KM approximation, Eq. (7) or Eq. (13). Notice that as  $\gamma$  increases,  $N$  approaches the world population (which by the end of January, 2020, was about 7,794,798,739 individuals [20]) and the value  $R_0$  decreases to one. The range of  $R_0$  values in this table are indicative that the infectious disease is not severe. The scenarios obtained from this values are shown in Fig. 2. They could be improved with better information from epidemiological centers about the coronavirus (COVID-19) pandemic.

$\gamma[1/\text{days}]$	N	$S_0$	$I_0 = N - S_0$	$\rho$	$\beta = \gamma/\rho[1/\text{days}]$	$R_0 = S_0/\rho$
0.07143	96480663	96456822	23841	62206877.11	1.148e-09	1.551
0.10000	152660730	152636890	23840	109552180.95	9.128e-10	1.393
0.14290	261127454	261103614	23840	204752844.72	6.979e-10	1.275
0.30300	921464679	921440838	23841	815581997.17	3.715e-10	1.130
0.50000	2287141587	2287117746	23841	2120340165.98	2.358e-10	1.079
0.95633	7794787378	7794763538	23840	7486874037.44	1.277e-10	1.041

### 5.1. Observing the KM approximation in data by country

To end this section, in Fig. 3 we show that the KM approximation can be observed on the combined aggregated data on deaths and recovered reported by individual countries. The data for analysis comes from [22]. Clearly, Cuba, Spain, Italy, and Germany (Fig. 3) do not follow the KM approximation. We can perform on each data set that follows the KM approximation (Argentina, Peru, US, Mexico, and Venezuela) the proposed method to find parameters and initial conditions that helps to find numerical solutions of the full *SIR* model that falls on the data of each one of them. That is left as an exercise. Additionally, in Fig. 4 we show that the KM approximation can also be observed on the aggregated confirmed cases of Argentina, Peru, US, Mexico, and Venezuela. We can guess that since recovered and deaths come from the population of confirmed cases, then the KM approximation should also work in that data set (confirmed cases) when it does in the former (combined aggregated data on deaths and recovered). The case of the Bombay plague studied by Kermack and McKendrick [2] only includes deceases as the plague was completely lethal.

Let's mention that some authors [24,25] have been able to fit cumulative confirmed cases data from Spain, Italy, Germany (or the alike China) using a logistic function, which is incompatible with the natural initial condition  $R(t=0) = 0$ , as the logistic function is obtained from the Kermack and McKendrick approximation taking  $S(t=0) \approx N$ , leaving (from Eq. (4))  $I(t=0) + R(t=0) = 0$ .

## 6. Concluding remarks and Future work

The following conclusions can be drawn from this work.

Cumulative deaths and recoveries from the coronavirus COVID-19 data [17] reported by countries aggregated worldwide as a whole (Fig. 2) or (in some cases) individual data [22] (Fig. 3) (including cumulative data [23] of confirmed cases as shown in Fig. 4) fall under the Kermack and McKendrick approximation [2] of the *SIR* model, indicative that the coronavirus COVID-19 pandemic is still growing and that it is not severe relative to the population size. Exceptions are countries like Spain, Italy or Germany where the cumulative data seems to have reached a kind of steady state stage. As shown in Figs. 3 and 4, data from the mentioned countries do not fall under the KM approximation.

An important result is that with knowledge of the constants defining the KM approximation (Eqs. (14)-(16)), solution of the full *SIR* model (Eqs. (1)-(3)) which falls on the observed data (*i.e.* Fig. 1 and 2) can easily be obtained following the method of Sec. 5, which only requires knowledge of the epidemic infectious period  $1/\gamma$  (a parameter that epidemiologists could infer from observing the evolution of any

epidemic) from which other necessary quantities to successfully integrate the system of Eqs. (1)-(3) could be obtained via Eqs. (17)-(21). Many other scenarios following the observed data can be constructed as information about the epidemic under consideration from epidemiologists is updated.

Moreover, in case the parameters  $\beta$  and  $\gamma$  are known, the methodology presented in Sec. 5 provides a starting point to initiate an educated trial and error guess work that helps fine tune the usually unknown initial conditions ( $S_0$  and  $I_0$ ), necessary to find numerical solutions of the full system of differential Eqs. ((1)-(3)) (defining the *SIR* model) such that the obtained numerical solution falls under the observed data (as in Figs. 1 and 3).

At this point it is hard to establish for sure which model better describes the evolution of the coronavirus pandemic [10]. 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 inferred from the data.

Consequently, before considering more complex models (requiring much more parameters than the *SIR* model), it is clear that a 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. The fact that the notation and meaning of the terms in epidemiological literature (even in the case of the *SIR* model) is not uniform means that such a task might be elusive for awhile. Nevertheless, that the KM approximation is still being observed for so many days on the coronavirus COVID-19 data is an indication that countries are not taking enough effective epidemiological measures to control it, and this might be the case meanwhile a treatment is available; recoveries and deaths are basically driven by human body individual responses.

This study provides an opportunity to present in the classroom an alternative example of real world data on which physical modeling of complex systems helps to comprehend how nature works outside physical controlled experiments. This is certainly a good example to further motivate Biology, Life Sciences and Medical students in their physics studies quantitatively [26,27].

A natural extension to this work is to consider extensions to the *SIR* model (like SEIR or SIRD) [3,4] to track the progression of the coronavirus COVID-19 infectious disease.

## Acknowledgments

The author is grateful to an anonymous editor who kindly provided useful comments on improving the readability of this article.



1. W. Zhou, *The Coronavirus Prevention Handbook* (Hubei Science and Technology Press, Wuhan, 2020).
2. W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc. Lond. A* **115** (1927) 700, <https://doi.org/10.1098/rspa.1927.0118>.
3. J. D. Murray, *Mathematical Biology*, 3rd ed. (Springer, New York, 2002), <https://doi.org/10.1007/b98868>.
4. M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, 2008).
5. M. Y. Li, *An Introduction to Mathematical Modeling of Infectious Diseases* (Springer, New York, 2018), <https://doi.org/10.1007/978-3-319-72122-4>.
6. V. Capasso and S. L. Paveri-Fontana, A mathematical model for the 1973 cholera epidemic in the European Mediterranean region, *Rev. Epidemiol. Sante Publique* **27** (1979) 121.
7. World Health Organization, Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>.
8. R. Li et al., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), *Science* **368** (2020) 489, <https://doi.org/10.1126/science.abb3221>.
9. H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, Oxford, 1987).
10. D. Kramer, COVID-19 pandemic modeling is fraught with uncertainties, *Phys. Today* **73** (2020) 25, <https://doi.org/10.1063/PT.3.4493>.
11. S. Kimball et al., GNU Image Manipulation Program, <https://www.gimp.org/>.
12. Wikipedia, Bombay Plague Epidemic, <https://en.wikipedia.org/wiki/Bombayplagueepidemic>.
13. Open Data Commons, <https://www.opendatacommons.org/licenses/pddl/1-0/>.
14. Datahub, Coronavirus disease 2019 (COVID-19), <https://datahub.io/>.
15. World Health Organization, <https://www.who.int/>.
16. Johns Hopkins University Center for Systems Science and Engineering, <https://systems.jhu.edu/>.
17. Datahub, Coronavirus disease 2019 (COVID-19) worldwide aggregated data, <https://datahub.io/core/covid-19/r/worldwideaggregated>.
18. S. Rojas, E. A. Chistensen, and F. J. Blanco-Silva, *Learning SciPy for Numerical and Scientific Computing*, 2nd ed. (Packt, Birmingham, 2015).
19. S. Rojas, *Numerical and Scientific Computing with SciPy* (Packt, Birmingham, 2017).
20. Worldometers, Countries in the world by population, <https://www.worldometers.info/world-population/>.
21. European Centre for Disease Prevention and Control, Q&A on COVID-19, <https://www.ecdc.europa.eu/en/covid-19/questionsanswers>.
22. Datahub, Coronavirus disease 2019 (COVID-19) countries aggregated, <https://datahub.io/core/covid-19/r/countriesaggregated>.
23. Datahub, Coronavirus disease 2019 (COVID-19) time series, <https://datahub.io/core/covid-19/r/timeseries-19-covid-combined>.
24. D. Fanelli and F. Piazza, Analysis and forecast of COVID-19 spreading in China, Italy and France, *Chaos Solitons Fractals* **134** (2020) 109761, <https://doi.org/10.1016/j.chaos.2020.109761>.
25. E. B. Postnikov, Estimation of COVID-19 dynamics “on a back-of-envelope”: Does the simplest SIR model provide quantitative parameters and predictions?, *Chaos Solitons Fractals* **135** (2020) 109841, <https://doi.org/10.1016/j.chaos.2020.109841>.
26. S. Rojas, On the need to enhance physical insight via mathematica reasoning, *Rev. Mex. Fis. E* **54** (2008) 75.
27. S. Rojas, Repeated problem solving revisited, *Am. J. Phys.* **77** (2009) 487, <https://doi.org/10.1119/1.3098259>.