

# Time-dependent SI model for epidemiology and applications to Covid-19

L. A. Ureña-López<sup>a</sup> and A. X. González-Morales<sup>a,b</sup>

<sup>a</sup>*Departamento de Física, DCI, Campus León, Universidad de Guanajuato, 37150, León, Guanajuato, México.*

<sup>b</sup>*Consejo Nacional de Ciencia y Tecnología,*

*Av. Insurgentes Sur 1582. Colonia Crédito Constructor, Del. Benito Juárez 03940, Ciudad de México, México.*

*\*e-mail: lurena@ugto.mx; alma.gonzalez@fisica.ugto.mx*

Received 28 January 2021; accepted 8 April 2021

A generalization of the Susceptible-Infectious model is made to include a time-dependent transmission rate, which leads to a close analytical expression in terms of a logistic function. The solution can be applied to any continuous function chosen to describe the evolution of the transmission rate with time. Taking inspiration from real data of the Covid-19, for the case of cumulative confirmed positives and deaths, we propose an exponentially decaying transmission rate with two free parameters, one for its initial amplitude and another one for its decaying rate. The resultant time-dependent SI model, which under extra conditions recovers the standard Gompertz functional form, is then compared with data from selected countries and its parameters fit using Bayesian inference. We make predictions about the asymptotic number of confirmed positives and deaths and discuss the possible evolution of the disease in each country in terms of our parametrization of the transmission rate.

*Keywords:* Epidemiological models; bayesian statistics.

DOI: <https://doi.org/10.31349/RevMexFis.67.050706>

## 1. Introduction

The epidemics of Covid-19 has prompted a lot of researchers from different fields to use their diverse expertise to understand the nature of the Sars-Cov-2 virus and its spreading worldwide. In particular, out of the medical sciences, the epidemics have revealed a rich ground where physicists, mathematicians, and statisticians are eager to apply their knowledge and contribute to its amelioration and containment at the local and global stages.

From the formal point of view, there has been much research in the so-called epidemic models and their mathematical properties, see for instance [1,2], being the compartmental models the most widely used and studied. In general, the models consider the classification of a given population in some parts: Susceptible (S), Infective (I), Recovered (R), among others, and are then dubbed in terms of which of them they consider for the dynamics of the disease: SIR, SEIR, etcetera. There is historical evidence that such models are in good agreement with the dynamics of past epidemics [1], and such past successes have triggered its use in the present crisis, see for instance, [3-11].

However, there is an inherent difficulty in studying the Covid-19 epidemics in real-time: that data collection is not perfect and, in most cases, is certainly incomplete and not very useful to fully characterize the evolution of the epidemics [12,13]. One then must question whether the use of very complex models is convenient, given the scarcity and flaws of the data series. Actually, it seems that simple models are sufficient to understand the epidemics [14,15].

Given the above considerations, here we study the simplest of compartmental models and use them for the description of the current epidemics in different countries. The model only accounts for two parts of the population: Susceptibles and Infectives, and is known as the SI model. The infectives in the model are not supposed to recover, and then its number is an ever-increasing function of time, which is what one sees in the daily reports of cumulative infectives released by health authorities worldwide. In this respect, the SI model seems to be a convenient one to follow the evolution of the real data at hand.

However, the original SI model has one free parameter, the transmission rate, which is assumed to be a constant parameter, but this seems to be an oversimplification that is not in agreement with real data. This has inspired the use of time-dependent transmission rates, which expand the capabilities of the compartmental models to encompass hidden complexities of the data compilations. Here we take this point of view and consider a time-dependent transmission rate in terms of an exponentially decreasing function, as it was done first in [16], and also in Ref. [17] (see also [18,19] for other examples). Additionally, we also take into account the data series of cumulative deaths, using simple assumptions without extending the SI model, which is thought to be at least as reliable as those of cumulative infectives, and probably a better representation of the epidemic's course [14].

A summary of the paper is as follows. In Sec. 2, we present the mathematical description of the standard SI model, which we call the time-independent version, and give a brief account of its main properties. We then explain its generalization to accommodate a time-dependent transmis-

sion rate and find the analytical solution in terms of a logistic function for the general case. It is argued that, because of the time-dependence of the transmission rate, the total population number can be freely chosen, and there is a simpler, approximated solution of the time-dependent SI system in the limit of a large population number in the form of a Gompertz function [20,21]. Using real data, we look for evidence of a time-dependent transmission rate in terms of its definition within the SI system, and then we propose an exponential-like parametrization of the transmission rate.

In Sec. 3, we use the time-dependent models of Sec. 2 and use them to make a fitting of their free parameters to data from some chosen countries. The fit is made using the Bayesian inference, and the results are giving an interpretation in terms of characteristic quantities and times that are intrinsic to our parametrized transmission rate. Finally, the main conclusions of our study and future perspectives are summarized in Sec. 4.

## 2. Mathematical background

Here we present the main mathematical expressions for the time-dependent epidemic model, based on the known compartmental model SI.

### 2.1. Time-independent SI

The SI model is represented by the following set of equations,

$$\dot{S} = -\beta S \frac{I}{N}, \quad \dot{I} = \beta S \frac{I}{N}, \quad (1)$$

Here,  $S(I)$  is the number of susceptible (infectious) people, the total population is  $N = S + I$ ,  $\beta$  is the infection rate (*i.e.*, the probability per unit time that an individual contracts the disease), and a dot means derivative for time.<sup>i</sup>

Because of the conservation equation, the SI system is truly one-dimensional and represented by the equation,

$$\dot{I} = \beta(1 - I/N)I. \quad (2)$$

If  $\beta$  is a constant parameter, the solution of Eq. (2) is the sigmoid, also known as the logistic function [22],

$$I = \frac{N}{1 + e^{\beta(t_0 - t)}}, \quad (3)$$

where  $t_0$  is an integration constant.

From the second derivative of Eq. (3) (the first derivative of Eq. (2)),

$$\ddot{I} = \beta \dot{I} (1 - 2I/N). \quad (4)$$

Then, there is a maximum of the first derivative at  $t = t_0$ , which also corresponds to an inflection point of the logistic function ( $\ddot{I} = 0$ ) at which  $I = N/2$ . As for the logistic function itself, notice that the initial and asymptotic values of the logistic function (3) are, respectively,

$$I_i = \frac{N}{1 + e^{\beta t_0}}, \quad I_\infty = N. \quad (5)$$

In other words, the saturation value of the infectious people is the whole of the available population. If  $N$  is a large number, and for the same  $I_i$ , the only change is the position of the inflection point  $t_0$ , which shifts to larger values.

### 2.2. Time-dependent SI system

Let us now consider the case of a time-dependent infection rate, that is,  $\beta = \beta(t)$ . Notice that we do not need to change the nature of the original SI system (1), and then we can use again Eq. (2) to find a solution to a time-dependent SI system. It can be readily shown that the solution can be written as a generalized logistic function [21,23] (see also [24]),

$$I(t) = \frac{N}{1 + e^{u_0 - u}}, \quad \text{with} \quad u(t) = \int_t^0 \beta(x) dx. \quad (6a)$$

where  $u_0$  is an integration constant. Notice that Eq. (6a) is again the sigmoid function but only now in terms of a new variable  $u$ . The initial and asymptotic values of the infected people are given by

$$I_i = \frac{N}{1 + e^{u_0}}, \quad I_\infty = \frac{N}{1 + e^{u_0 - u_\infty}}. \quad (6b)$$

Here,  $u_\infty = u(t \rightarrow \infty)$ , which may or may not be a finite value, and this depends on the chosen function  $\beta$ , see Eq. (6a).

Notice that the second derivative of Eq. (2) for a time-dependent transmission rate is

$$\ddot{I} = \dot{I} \left[ \frac{\dot{\beta}}{\beta} + \beta(1 - 2I/N) \right]. \quad (6c)$$

Hence, the true inflection point  $\ddot{I} = 0$  does not correspond to  $I = N/2$  anymore as it was in the time-independent case. However, it can be shown that the time-independent SI solution (3) is a particular case of the time-dependent one (6a). In the case  $\beta = \text{const.}$ , one readily obtains that  $u(t) = \beta t$ , and then Eq. (3) is recovered if also  $u_0 = \beta t_0$ .

The exact solution (6a) opens up the possibility to consider the evolution of a disease in which the transmission rate is changing, whether by natural means or by human intervention. This is a more realistic approach, and it has the advantage that we can continue dealing with the accumulated numbers reported by the health systems worldwide.

### 2.3. The large $N$ limit

In the time-independent SI system, the constant of integration is fixed by the initial conditions and the total population number, namely  $e^{\beta t_0} = N/I_i - 1$ , and then the asymptotic value of the infectives is the total population, see Eqs. (5). However, in the time-dependent case, the asymptotic value of infectives also depends on the asymptotic value  $u_\infty$ , which

means that not all the population will get infected, that is,  $I_\infty < N$ .

The ratio between the total and the initial number of infectives can be written as

$$\frac{I_\infty}{I_i} = \frac{1 + e^{u_0}}{1 + C e^{u_0 - u_\infty}}. \quad (7)$$

Here,  $u_\infty$  is an independent constant, and then one can obtain the same ratio by adjusting accordingly the values of  $u_0$  and  $u_\infty$ . In this sense, there is a new degeneracy in the time-dependent system as the total number of infectives is not uniquely determined by the total population number  $N$ . Explicitly, the degeneracy reads

$$e^{u_\infty} = \frac{(I_\infty/I_i)e^{u_0}}{1 + e^{u_0} - (I_\infty/I_i)}. \quad (8)$$

If we keep the ratio  $(I_\infty/I_i)$  fixed, we obtain that

$$\lim_{u_0 \rightarrow \infty} e^{u_\infty} = \frac{I_\infty}{I_i}. \quad (9)$$

The final ratio of infectives, in this limit, can be calculated directly from the asymptotic value of the variable  $u$ . Given that  $e^{u_0} = N/I_i - 1$ , we call this the large  $N$  limit.

Actually, one can do the same exercise in the general expression of infectives. If we write Eq. (6a) (see also Eq. (6b)) in the form,

$$I(t) = \frac{1 + e^{u_0}}{e^u + e^{u_0}} I_i e^u, \quad (10)$$

we find that for large enough values of  $u_0$ , *i.e.*,  $N \rightarrow \infty$ , the function of infectives can be approximated as

$$I(t) \simeq I_i e^u = I_i \exp \left[ \int_0^t \beta(x) dx \right]. \quad (11a)$$

The evolution of the disease is simply driven by the transmission rate, and then the asymptotic limit is obtained if the integral on the rhs of (11a) converges for  $t \rightarrow \infty$ . Actually, the ratio between the asymptotic and initial values of infectives is

$$\frac{I_\infty}{I_i} \simeq e^{u_\infty} = \exp \left[ \int_0^\infty \beta(x) dx \right], \quad (11b)$$

which is in turn the same result as in Eq. (9) above.

The expression (11a) gives a simplified evolution of the disease. For instance, the first derivative is just  $\dot{I} = \beta I$ , whereas the second derivative is  $\ddot{I} = (\dot{\beta} + \beta^2)I$ , and then the inflection point  $t_0$  in the evolution curve of infectives is given by the solution of the equation  $(\dot{\beta} + \beta^2)(t_0) = 0$ .

There is a small warning here about the parametrization of  $\beta$ . If we assume that the transmission rate is given by  $\beta = \beta(t, k_\ell)$ , where  $k_\ell$  are in general constant parameters, one must be aware that their values will depend on  $N$ , and then the values obtained from Eq. (11a) are not the same as those from Eq. (10). They will, in both cases, deliver the same values of  $I_i$  and  $I_\infty$ , but the evolution profile  $I(t)$  will have some differences in the two cases.

## 2.4. Time-dependent $\beta$ from real data

One important question is whether real data suggest a complex evolution of the Covid-19 disease, for which the simple time-independent SI system would be insufficient to describe it. To get an answer directly from the data available, we make use here of an analytical result that can be derived from the general case (6a).

First, we write down the following expression to define the time-dependent function  $\Gamma(t)$ ,

$$\Gamma(t) \equiv \frac{\dot{I}}{I(1 - I/I_\infty)} = \frac{\beta(t)}{1 - e^{u - u_\infty}}, \quad (12a)$$

which is valid for any form of  $\beta$  and is directly derived from Eqs. (6); it is also valid for the large  $N$  approximation (11a). In the time-independent case, for which  $I_\infty = N$ , we readily obtain  $\Gamma(t) = \beta = \text{const}$ . But in the general case, we find at  $t = 0$  that

$$\Gamma(0) = \frac{\beta(0)}{1 - e^{-u_\infty}}. \quad (12b)$$

Another characteristic value is found at late times, as  $u \rightarrow u_\infty$ . In this case, we use the approximation  $1 - e^{u - u_\infty} \simeq u - u_\infty$ , and then from Eq. (6a), we obtain that

$$\lim_{t \rightarrow \infty} \Gamma(t) = - \lim_{t \rightarrow \infty} \left[ \frac{1}{\beta(t)} \int_0^t \beta(x) dx \right]^{-1}, \quad (12c)$$

if such limit exists.

Thus, the function  $\Gamma(t)$  can help us to decide whether to use the time-independent or the time-dependent SI system, as the expression (12a) can be easily calculated from data in countries that already show an asymptotic value of infectives. In Fig. 1, we show data from Mexico,<sup>ii</sup> which is one of the countries that comply with the preceding condition, as seen on the left panel for the number of positives normalized to the latest reported value (the first day in the time series is that of the first registered death). Shown also is the evolution of reported deaths, which is too normalized to the value of the latest reported value.

On the right panel of Fig. 1, we see the daily new cases (with the same mentioned normalization) divided by different combinations of accumulated numbers. In the plots  $P$ , ( $D$ ) refers to confirmed positives (deaths). Notice that the combinations in the plot represent the function (12a), and then data seems to indicate that  $\Gamma(t)$  is a decreasing function that approaches a constant value at late times. In contrast, when we use the expression  $\dot{I}/I$ , the ratio goes to zero. Our interpretation is that data suggest a time-dependent transmission rate  $\beta$ , particularly one that decays with time. Although we only show the case of Mexico, we have found the same trend in all cases in which data already shows an asymptotic value of accumulated positives (*e.g.*, Germany, France, and others).

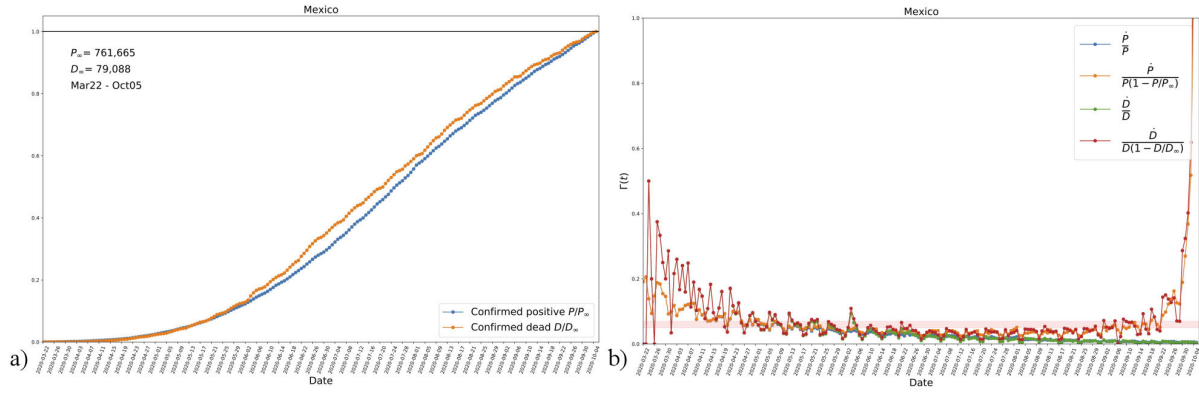


FIGURE 1. a) Data from Mexico showing the evolution of accumulated confirmed positives and deaths, the time series starts at the date of the first registered death (from March 22 until October 5). The values in the plot are normalized with their respective last data point. It can be noticed that both quantities seem to have reached an asymptotic final value. b) The estimated evolution of function  $\Gamma(t)$ , see Eq. (12a), according to data of new daily cases for both confirmed positives and deaths. Shown are three different combinations of each data compilation, where  $P_\infty$  and  $D_\infty$  are the last data points in the corresponding time series. The trends of the three cases suggest a decrease in the transmission rate  $\beta$  with time. The orange-shaded horizontal region (with vertical range 0.05 – 0.07) is just shown for reference. See Sec. 2.4. and the text for more details

It is clearly seen that  $\beta$  is, in general, a decaying function as time proceeds, which in our opinion is a manifestation of governmental intervention to slow the spread of the disease. Although there is not a unique function, given the profile suggested in Fig. 1, we will consider an exponentially decaying function as an approximation to the evolution of  $\beta$ . That is, we write it explicitly as

$$\beta(t) = k_0 e^{-k_1 t}, \quad (13a)$$

where  $k_0, k_1$  are constant parameters. The corresponding time variable in Eq. (6a) is

$$u(t) = \frac{k_0}{k_1} (1 - e^{-k_1 t}), \quad (13b)$$

and then for this case, we see that  $u_\infty = k_0/k_1$ .

In limit  $k_1 t \ll 1$ , we obtain that  $u(t) \simeq k_0 t$ , which is the expected behavior in the time-independent case. Then,  $k_0$  determines the initial growth of the epidemics, whereas  $k_1$  gives the decay time of the transmission rate, with a half-life time given by  $t_{1/2} = \ln 2/k_1$ .

Likewise, the asymptotic value of the infectives is  $I_\infty = I_i e^{k_0/k_1}$ , see Eqs. (9) and (11b), and then the ratio between the asymptotic and initial values will depend on the ratio  $k_0/k_1$ . Here we see the importance of  $k_1$ : the smaller its value, the larger it takes for the epidemics to fade away and the larger the asymptotic value of total infectives.

As for the time-dependent function  $\Gamma(t)$  defined in Eq. (12a), after using Eqs. (13) we find that

$$\Gamma(0) = \frac{k_0}{1 - e^{-k_0/k_1}}, \quad \lim_{t \rightarrow \infty} \Gamma(t) = k_1. \quad (14)$$

Notice that Eq. (14) is also in agreement with the expectation from data in Fig. 1: there are well definite values of  $\Gamma(t)$  at early and late times at late times, where the value at late times

will give us an indication of the decay time of the transmission rate  $\beta$ .

We mentioned before that it is impossible to have an analytical expression for the inflection time of the sigmoid function (6a) for a general  $\beta(t)$ . However, one can instead write an expression for the time at which the infected population  $I(t_{50})$  is one-half (50%) of the asymptotic value, *i.e.*,  $I(t_{50}) = I_\infty/2$ . After some straightforward algebra using Eqs. (6), we find in general for the generalized time variable that

$$u_{50} = u_0 - \ln(1 + 2e^{u_0 - u_\infty}). \quad (15a)$$

Hence, from the particular parametrization (13a), we finally obtain

$$t_{50} = \frac{1}{k_1} \ln \left( \frac{k_0/k_1}{k_0/k_1 - u_{50}} \right). \quad (15b)$$

We will use  $t_{50}$  as a point of reference in the time series of the different data in our analysis below, similar to the inflection point of the standard sigmoid function. Another useful point is that at which the number of infectives is 90% of the asymptotic value,  $I(t_{90}) = 0.9I_\infty$ , as it can be considered as a reference for the upcoming end of the infection period. Following the same calculations that led to Eqs. (15), we find

$$u_{90} = u_0 - \ln[1/9 + (10/9)e^{u_0 - u_\infty}], \quad (16a)$$

and its corresponding time, again for the particular parametrization (13a), is

$$t_{90} = \frac{1}{k_1} \ln \left( \frac{k_0/k_1}{k_0/k_1 - u_{90}} \right). \quad (16b)$$

### 3. Statistical analysis and results

Our main assumption is that the time series of both confirmed positives and deaths have a common origin from the total

number of infected people  $I(t)$ . Formally speaking, we are assuming that

$$P(t) = r_P I(t), \quad D(t) = r_D I(t - t_D), \quad (17)$$

where  $r_P$  is the fraction of infected people that are tested and confirmed as positive, whereas  $r_D$  is the fraction of infected people that is confirmed positive and eventually pass away. Notice that we consider that  $D(t)$  evolves with a time delay  $t_D$  for  $I(t)$  (and also to  $P(t)$ ). Such delay is difficult to measure reliably, and here, we will report the values suggested by the data itself.

In the following sections, we do the fitting to data using two different models. The first one, which we call model A, considers the generalized logistic function (6a) and the parametrization (13a) of the transmission rate. The second one, which we dub model B, uses the large  $N$  approximation represented by Eq. (11a) and the same parametrization of the transmission rate. As discussed previously in Sec. 2.4, data seems to discard the time-independent SI system, but for completeness, we also report its fitting to data in Appendix C.

### 3.1. Model A and fitting to data

We will consider the following parametrization of the logistic function of the confirmed positive people, in cumulative numbers,

$$P(t) = P_i \frac{1 + e^{u_0}}{1 + e^{u_0 - u(t)}}, \quad (18a)$$

Following the discussion at the beginning of this section, we will also consider the reports of accumulated deaths, which we assume to follow a similar logistic function as those of the confirmed positive, except for a different amplitude and inflection point,

$$D(t) = D_i \frac{1 + e^{\hat{u}_0}}{1 + e^{\hat{u}_0 - u(t - t_D)}}, \quad (18b)$$

Our model then has seven free parameters:  $P_0, D_0, t_D, u_0, \hat{u}_0, k_0$  and  $k_1$ , and then one does not hope for tight constraints on them given the scarcity of data about Covid-19 infections in general. Additionally, one must remember that data is not generated in a systematic way as in a laboratory experiment, and there may be many sources of error in the data management and processing of new cases.

We will assume that the data provided follows the trend of the real number of infected people and deaths and that both of these numbers will eventually reach a saturation value soon as has happened in past diseases. Moreover, as we do not know the systematic errors in the processing of the data, we will assume some level of intrinsic error by using a Poisson distribution. To ease the fitting of data from different countries, we normalize the data so that the first point in the time series is of the order of unity. Given this, we consider flat priors in the form:  $0 < P_i < 10, 0 < D_i < 10, 0 < t_D < 100, 0 < u_0, \hat{u}_0 < 30$ , and  $0 < k_0, k_1 < 3$ .

To fit the data, we will use the `emcee` code [25], a python code, to random sample a probability distribution that has

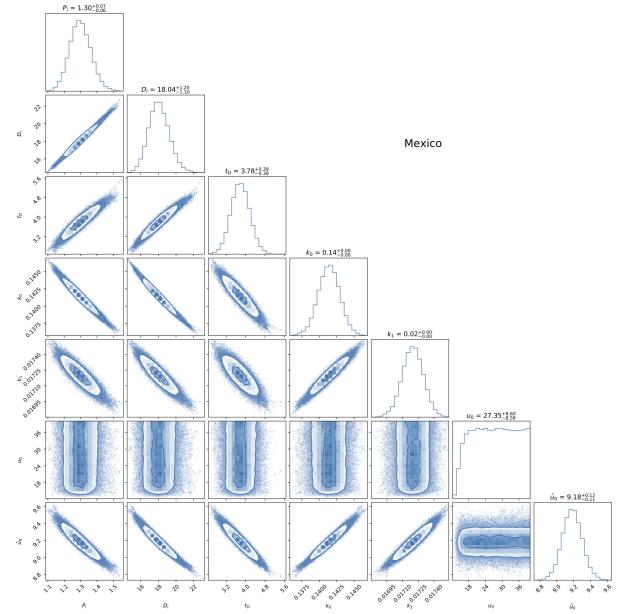


FIGURE 2. Triangle plot of the fitting to data from Mexico (see also Fig. 1) of the parameters  $P_i, D_i, t_D, k_0, k_1, u_0$ , and  $\hat{u}_0$  of model A described in Sec. 3.1. In general, all parameters are well constrained by the combination of accumulated confirmed positives and deaths. See the text for more details.

been extensively used for a variety of applications. See Appendix D for a wider explanation of how the likelihood for this model was prepared (as well as for the others presented in this work). For the analysis, a set of 100 chains with 30000 steps in each one were run. The results are shown in Fig. 2 for the free parameters of our model. A common feature of all the countries we considered, not just the one shown in Fig. 2, is that the values of  $P_i, D_i, t_D, k_0, k_1, u_0$ , and  $\hat{u}_0$ , are all well constrained by the data, which is consistent with our assumption that both confirmed positives and deaths follow the same trend of evolution.

We have applied our model to the data of 9 other countries, which at the moment of writing are the countries with the highest number in cumulative positives and deaths in terms of their population size, and the fitting results are summarized in Table I. As said before,  $P_\infty$  and  $D_\infty$  are the expected final numbers for the cumulative positives and deaths in each case, even for countries that have not yet reached an asymptotic value.

The delay time between positives and deaths,  $t_D$ , is for all cases lower than 20 days, which is consistent with the general fact that all deaths were first confirmed as positive, and then  $t_D$  will represent the delay time between a positive test and the occurrence of death.

Next are the parameters of the transmission rate  $\beta$ , where we note a strong similarity in the values of the different countries. First, we recall that  $\beta(0) = k_0$ , and then this parameter is the value of the transmission rate at the start of the time series. Likewise, parameter  $k_1$  is the decay rate of the transmission rate. The last two columns in Table I show the values

TABLE I. Fitted values of parameters in model A, see Eqs. (18), as obtained from the data of different countries. The confidence regions for the parameters in the case of Mexico are shown in Fig. 2.

Country	$P_\infty$	$D_\infty$	$t_D$	$k_0$	$k_1$ (1/days)	$u_0$	$\hat{u}_0$
Mexico	$980,256 \pm_{16,645}^{17,593}$	$97,084 \pm_{484}^{487}$	$3.78 \pm_{0.98}^{1.02}$	$0.14 \pm_{0.00}^{0.00}$	$0.017 \pm_{0.000}^{0.000}$	$27.35 \pm_{13.50}^{12.52}$	$9.18 \pm_{0.30}^{0.31}$
Peru	$1,075,559 \pm_{12,868}^{13,086}$	$51,066 \pm_{524}^{537}$	$12.00 \pm_{0.58}^{0.58}$	$0.10 \pm_{0.00}^{0.00}$	$0.015 \pm_{0.000}^{0.000}$	$26.19 \pm_{14.75}^{13.67}$	$26.19 \pm_{14.83}^{13.66}$
Belgium	$77,073 \pm_{3,321}^{1,464}$	$9,802 \pm_{22}^{23}$	$9.82 \pm_{8.00}^{0.27}$	$0.30 \pm_{0.01}^{0.10}$	$0.046 \pm_{0.009}^{0.004}$	$6.47 \pm_{0.29}^{0.93}$	$4.78 \pm_{0.06}^{2.16}$
Bolivia	$187,445 \pm_{6,992}^{7,699}$	$15,807 \pm_{1,323}^{1,547}$	$0.06 \pm_{0.06}^{0.42}$	$0.09 \pm_{0.01}^{0.01}$	$0.009 \pm_{0.001}^{0.001}$	$7.13 \pm_{0.21}^{0.24}$	$8.27 \pm_{0.30}^{0.38}$
Brazil	$6,968,037 \pm_{51,706}^{50,851}$	$166,676 \pm_{362}^{356}$	$0.00 \pm_{0.00}^{0.01}$	$0.13 \pm_{0.00}^{0.00}$	$0.016 \pm_{0.000}^{0.000}$	$28.77 \pm_{12.35}^{11.12}$	$8.03 \pm_{0.02}^{0.03}$
Chile	$442,318 \pm_{14,407}^{7,888}$	$12,488 \pm_{87}^{98}$	$19.47 \pm_{18.54}^{4.32}$	$0.09 \pm_{0.00}^{0.01}$	$0.006 \pm_{0.000}^{0.000}$	$6.19 \pm_{0.13}^{0.56}$	$5.87 \pm_{0.37}^{1.71}$
Ecuador	$138,598 \pm_{1,101}^{1,108}$	$10,568 \pm_{62}^{61}$	$0.01 \pm_{0.01}^{0.07}$	$0.11 \pm_{0.00}^{0.00}$	$0.019 \pm_{0.000}^{0.000}$	$8.46 \pm_{0.27}^{0.39}$	$26.70 \pm_{14.24}^{13.17}$
United States	$6,559,124 \pm_{9,778}^{9,639}$	$182,774 \pm_{141}^{140}$	$0.00 \pm_{0.00}^{0.00}$	$0.19 \pm_{0.00}^{0.00}$	$0.022 \pm_{0.000}^{0.000}$	$29.97 \pm_{10.72}^{9.93}$	$7.88 \pm_{0.00}^{0.00}$
United Kingdom	$329,828 \pm_{2,401}^{2,345}$	$41,328 \pm_{49}^{49}$	$10.74 \pm_{1.54}^{0.12}$	$0.32 \pm_{0.00}^{0.02}$	$0.044 \pm_{0.000}^{0.001}$	$25.12 \pm_{15.13}^{14.74}$	$6.37 \pm_{0.02}^{0.52}$

TABLE II. Fitted values of extra parameters in the case of model A, for the same countries as in Table I. Shown are the total population number  $N$ , the half-life time  $t_{1/2}$  of the transmission rate  $\beta$ , see Eq. (13b), the crossing times of half the asymptotic values of cumulative positives and deaths,  $t_{50}$  and  $t_{D50}$ , respectively, see Eqs. (15), in number of days after the start of the time series.

Country	$N$	$t_{1/2}$	$t_{50}$	$t_{D50}$
Mexico	$199,748,406,667,825 \pm_{199,748,131,793,760}^{54,555,581,054,363,484,160}$	$40.30 \pm_{0.60}^{0.64}$	$143.81 \pm_{1.52}^{1.60}$	$134.48 \pm_{0.45}^{0.44}$
Peru	$454,369,877,704,612 \pm_{454,369,701,437,638}^{391,684,701,709,530,300,416}$	$45.68 \pm_{0.37}^{0.38}$	$145.80 \pm_{1.04}^{1.06}$	$157.80 \pm_{0.96}^{0.98}$
Belgium	$150,462 \pm_{73,199}^{61,584}$	$15.09 \pm_{1.19}^{3.65}$	$38.43 \pm_{8.32}^{1.51}$	$35.40 \pm_{0.15}^{0.15}$
Bolivia	$196,415 \pm_{11,032}^{14,886}$	$81.28 \pm_{7.83}^{8.64}$	$138.06 \pm_{2.92}^{3.24}$	$181.65 \pm_{7.61}^{8.53}$
Brazil	$5,743,375,136,031,503 \pm_{5,743,350,307,608,080}^{388,424,216,445,922,902,016}$	$44.13 \pm_{0.18}^{0.17}$	$157.58 \pm_{0.62}^{0.60}$	$124.14 \pm_{0.20}^{0.20}$
Chile	$442,424 \pm_{14,470}^{7,961}$	$115.51 \pm_{7.65}^{8.39}$	$92.42 \pm_{2.89}^{1.30}$	$105.67 \pm_{0.38}^{0.40}$
Ecuador	$2,172,244 \pm_{519,254}^{1,031,514}$	$37.38 \pm_{0.15}^{0.15}$	$111.76 \pm_{0.79}^{0.79}$	$114.32 \pm_{0.54}^{0.53}$
United States	$11,573,101,571,769,328 \pm_{11,572,844,606,558,488}^{237,514,150,381,025,689,600}$	$31.53 \pm_{0.03}^{0.03}$	$115.00 \pm_{0.13}^{0.13}$	$81.67 \pm_{0.07}^{0.07}$
United Kingdom	$18,562,540,734,729 \pm_{18,562,537,349,892}^{47,014,334,860,169,912,320}$	$15.79 \pm_{0.36}^{0.14}$	$53.53 \pm_{0.55}^{0.53}$	$46.80 \pm_{0.09}^{0.09}$

of the integration constants  $u_0$  and  $\hat{u}_0$ , which are directly related to the total population number  $N$ .

Other quantities of interest, which are derived from the basic parameters in Table I, are also shown in Table II. The first one is the total population number  $N$ , which within model A is understood to be the total population in susceptible form for the spreading of the disease. We can see that this number is less than the total population in the case of countries that seem to have the epidemics under control, but in some others, it can be much larger than the whole world population. This only signals the lack of convergence of model A for the parameters  $u_0$  and  $\hat{u}_0$ , whose fitted values are close to the upper limit in their priors.

More meaningful is the half-lifetime of the transmission rate represented by  $t_{1/2}$ , which is directly calculated from parameter  $k_1$ . The lowest value corresponds to Belgium, with the disease decreasing by half every 15.09 days, whereas the highest corresponds to Chile, for which the disease decreases by half every 115.51 days. It is interesting to note the relation between  $t_{1/2}$  and the measures taken to control the epidemics, as the lower values are for countries with the strictest lockdown measurements.

Almost as meaningful as  $t_{1/2}$  are the values of the half-crossing times  $t_{50}$  and  $t_{D50}$ , calculated from Eq. (15b). The lowest values correspond to Belgium, for which the half-crossing occurred just 38 and 35 days after the report of the first deaths. The largest values are for Brazil, resulting in 157 and 124 days, respectively.

In the top and middle panels of Fig. 3, we show the comparison of data with the estimated evolution curves from our fitting, where the latter are represented by 500 instances of the model using a sample of values around those of maximum likelihood. In general we see a better a very good agreement with the data, which suggests that Eqs. (18) represent well the evolution of real data. For reference in the plots, we also show the times at half-crossing, around which the number of confirmed positives and deaths were half the asymptotic values  $P_\infty$  and  $D_\infty$ , respectively.

Also, in the bottom panel of Fig. 3, we show the time evolution of the transmission rate  $\beta(t)$  and its comparison with data, represented here by the new daily cases of both confirmed positives and deaths. We see a very good agreement with the combination of data that, in principle, represents the transmission rate. Likewise, there is good agreement with the

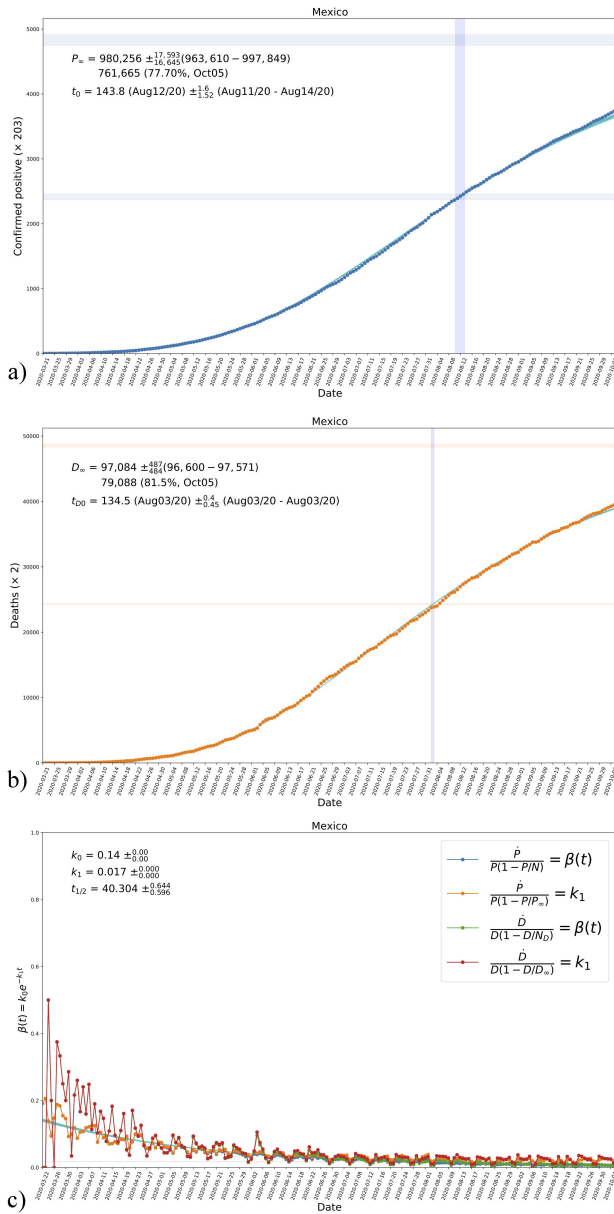


FIGURE 3. The resultant evolution curves of accumulated confirmed positives a) and deaths b) in the case of Mexico, according to the values in the triangle plot in Fig. 2. Shown in the figures are the obtained asymptotic values  $P_\infty$  and  $D_\infty$ , see Eq. (7) and Table I, in the top blue-shaded horizontal regions. The vertical blue-shaded regions mark the time of half-crossing in each case; the region for  $I_\infty/2$  is also shown for reference. c) The resultant evolution of the transmission rate  $\beta$  according to the parametrization in Eq. (13a) and its comparison with data, see Fig. 1 and Tables I and II. Shown are also the obtained values of the parameters  $k_0$  and  $k_1$  and the corresponding half-life time of the disease  $t_{1/2}$ . The horizontal red-shaded region represents the obtained value of  $k_1$ , see Eq. (12b). See the text for more details.

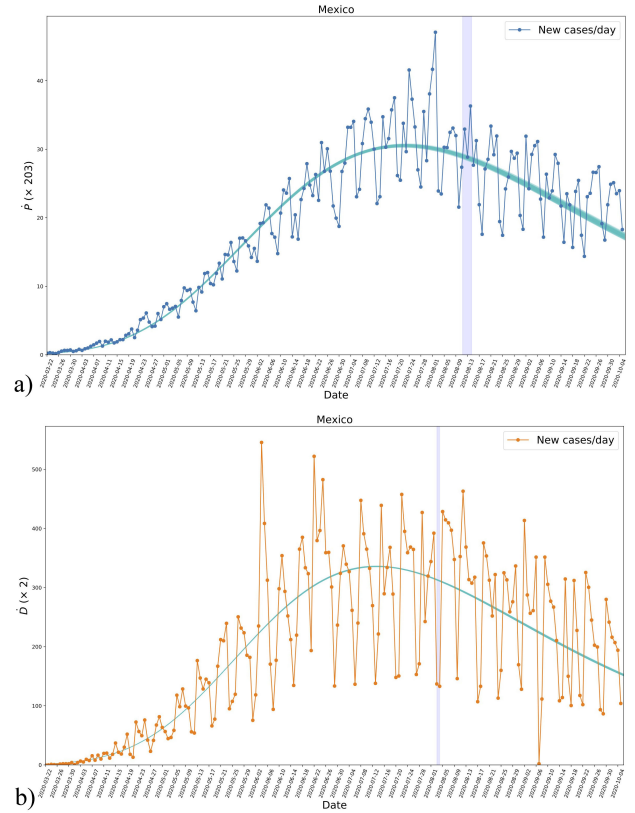


FIGURE 4. The derivatives  $\dot{P}$  and  $\dot{D}$  for positives a) and deaths b), respectively, obtained from the data of new daily cases and the analytical expression (2) using the parameters fitted to the data, see Fig. 2. Even though the new daily cases were not used for the fitting, we see a consistent agreement with the results. The blue-shaded vertical regions mark the crossing of half the asymptotic values in each case as in Fig. 3. See the text for more details.

combination that represents the parameter  $k_1$ , which in the plot is represented by the horizontal red-shaded region.

As an extra comparison with data, we show in Fig. 4 the derivative of both confirmed positives and deaths as obtained from the fitting to data and using the analytical formula (2) for each case. It must be recalled that the trend of new cases was not used for the fitting to data, and then the mentioned comparison is useful as an extra validation of the fitting, even if it does not appear to be as good as for the cumulative cases in Fig. 3. Additionally, the results show that the derivative of the model can follow the daily evolution of the disease and not just the global trend.

As the last feature, we show in Fig. 4 the time of half-crossing with a vertical blue-shaded region, which indicates that the maximum of new daily cases is reached some days before and then the presence of such maximum seems to be a necessary condition for the inflection of the cumulative cases.

### 3.2. Model B and fitting to data

As explained before, for model B, we will consider the following parametrisation of confirmed positives and deaths, in cumulative numbers,

$$P(t) = P_i e^u, \quad D(t) = D_i e^{u(t-t_D)}, \quad (19)$$

where the variable  $u$  has the same form as in Eq. (13b). For this parametrization, we obtain directly that  $\dot{P}/P = \beta$  and  $\dot{D}/D = \beta e^{k_1 t_D}$ , and also obtain the same limit results as in Eqs. (14). Explicitly, the functional form of model B (19) is

$$F(t) = F_i \exp \left[ \frac{k_0}{k_1} (1 - e^{-k_1 t}) \right], \quad (20)$$

which is no other but the Gompertz function [20,21,26]. In this way, one can see that there is a direct connection between the SI model and the Gompertz function, mediated by an exponentially decaying transmission rate and what we called the large- $N$  limit.

As said before, for this simplified model, see Sec. 2.3 above, one can find analytical expressions for different quantities of interest, which are actually the same one obtains from the Gompertz function (20) [21]. The first ones are the asymptotic values at  $t \rightarrow \infty$ , which are given by  $P_\infty = P_i e^{k_0/k_1}$  and  $D_\infty = D_i e^{k_0/k_1}$ , for confirmed positives and deaths, respectively.

Another analytical result is the time for the inflection of the curve, which we denote by  $t_0$  following the nomenclature of the time-independent SI system. The expression is

$$t_0 = \frac{1}{k_1} \ln \left( \frac{k_0}{k_1} \right) (+t_D), \quad (21a)$$

where we have included the time shift  $t_D$  of the function  $D(t)$ . Likewise, the numbers of confirmed positives and deaths at the inflection time are

$$P_0 = P_i e^{k_0/k_1 - 1}, \quad D_0 = D_i e^{k_0/k_1 - 1}, \quad (21b)$$

which is the same functional form for both of them. One can easily see that  $P_0 = P_\infty/e$  and  $D_0 = D_\infty/e$ .

The inflection point also corresponds to a maximum in the derivative of Eqs. (19), and then we find

$$\dot{P}_0 = k_1 P_i e^{k_0/k_1 - 1}, \quad \dot{D}_0 = k_1 D_i e^{k_0/k_1 - 1}. \quad (21c)$$

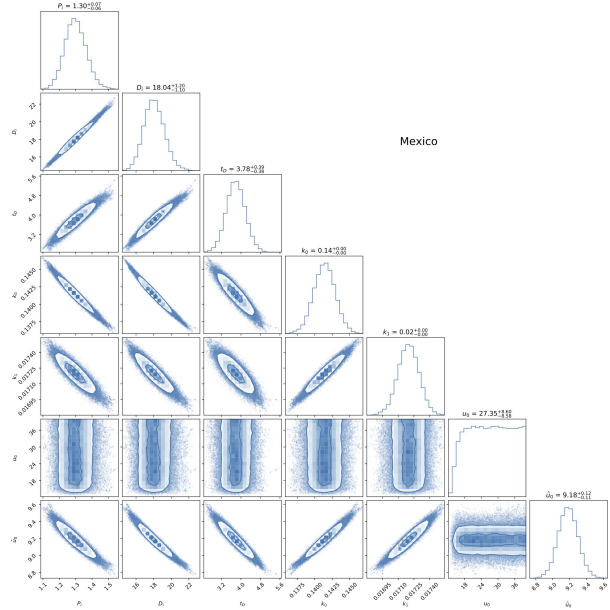


FIGURE 5. Triangle plot of the fitting to data from Mexico of the parameters  $P_0$ ,  $D_0$ ,  $t_D$ ,  $k_0$ , and  $k_1$  of model B described in Sec. 3.2. In general, and similarly to model A in Fig. 2, all parameters are well constrained by the combination of cumulative confirmed positives and deaths. See the text for more details.

This time the model has five free parameters:  $P_i$ ,  $D_i$ ,  $t_D$ ,  $k_0$  and  $k_1$ , and as in model A the last two of them are common to both confirmed positives and deaths. We will again assume some level of intrinsic error by using a Poisson distribution and the same normalization of the data so that the first point is of order of an unity. Given this, we consider flat priors in the form:  $0 < P_i < 10$ ,  $0 < D_i < 10$ ,  $0 < t_D < 100$  and  $0 < k_0, k_1 < 3$ .

We again took the emcee algorithm, see Appendix D, using 100 chains with 20,000 steps in each one. The results are shown in Fig. 5 for the free parameters of our model. As happened before in model A, the values of  $P_0$ ,  $D_0$ ,  $t_0$ ,  $k_0$ , and  $k_1$  are well constrained by the data, and their values are similar and of the same order of magnitude as those of model A. This is seen from a quick comparison of the common parameters in Figs. 2 and 5.

TABLE III. Fitted values of parameters in model B, see Eqs. (19), as obtained from the data of different countries. The confidence regions for the parameters in the case of Mexico are shown in Fig. 5.

Country	$P_\infty$	$D_\infty$	$t_D$	$k_0$	$k_1$ (1/days)
Mexico	$885, 589 \pm_{6,075}^{6,084}$	$96, 848 \pm_{485}^{486}$	$0.02 \pm_{0.02}^{0.10}$	$0.16 \pm_{0.00}^{0.00}$	$0.019 \pm_{0.000}^{0.000}$
Peru	$1, 074, 659 \pm_{12,693}^{13,147}$	$51, 009 \pm_{526}^{530}$	$11.98 \pm_{0.58}^{0.58}$	$0.10 \pm_{0.00}^{0.00}$	$0.015 \pm_{0.000}^{0.000}$
Belgium	$73, 582 \pm_{1,264}^{1,300}$	$9, 748 \pm_{21}^{22}$	$3.45 \pm_{2.39}^{0.72}$	$0.55 \pm_{0.03}^{0.12}$	$0.074 \pm_{0.001}^{0.001}$
Bolivia	$279, 451 \pm_{9,223}^{9,928}$	$16, 866 \pm_{633}^{674}$	$15.14 \pm_{1.13}^{1.13}$	$0.11 \pm_{0.00}^{0.00}$	$0.014 \pm_{0.000}^{0.000}$
Brazil	$4, 907, 718 \pm_{15,750}^{15,697}$	$169, 413 \pm_{390}^{387}$	$0.00 \pm_{0.00}^{0.00}$	$0.15 \pm_{0.00}^{0.00}$	$0.019 \pm_{0.000}^{0.000}$
Chile	$497, 634 \pm_{16,172}^{56,953}$	$14, 470 \pm_{1,571}^{89}$	$14.83 \pm_{14.61}^{0.81}$	$0.20 \pm_{0.01}^{0.28}$	$0.025 \pm_{0.000}^{0.007}$
Ecuador	$140, 948 \pm_{857}^{830}$	$10, 502 \pm_{58}^{58}$	$0.05 \pm_{0.05}^{0.23}$	$0.11 \pm_{0.00}^{0.00}$	$0.019 \pm_{0.000}^{0.000}$
United States	$5, 186, 207 \pm_{5,084}^{5,080}$	$192, 021 \pm_{142}^{141}$	$0.00 \pm_{0.00}^{0.00}$	$0.23 \pm_{0.00}^{0.00}$	$0.028 \pm_{0.000}^{0.000}$
United Kingdom	$312, 737 \pm_{2,307}^{2,051}$	$41, 127 \pm_{49}^{50}$	$0.95 \pm_{0.36}^{1.21}$	$0.66 \pm_{0.06}^{0.02}$	$0.060 \pm_{0.000}^{0.000}$



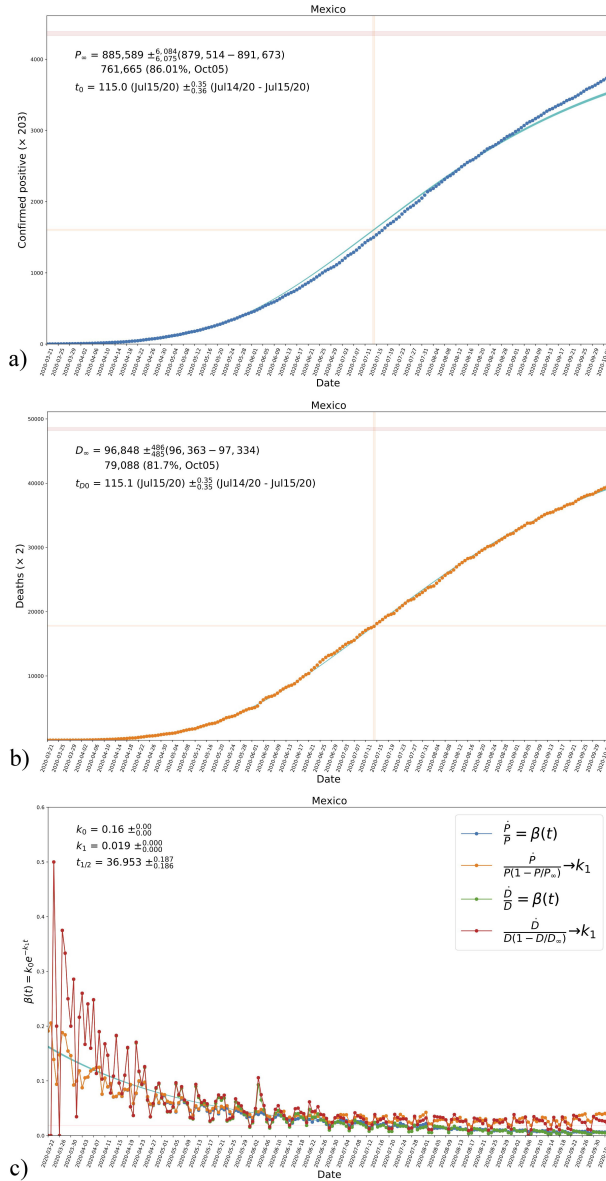


FIGURE 6. The resultant evolution curves of cumulative confirmed positives a) and deaths b) for model B in the case of Mexico, according to the values in the triangle plot in Fig. 5. Shown in the figures are the obtained asymptotic values  $P_\infty$  and  $D_\infty$ , see Eq. (7) and Table IV, in the top blue-shaded horizontal regions in each plot. The vertical blue-shaded regions mark the time inflection in each case; the region for  $P(t_0)$  is also shown for reference. c) The resultant evolution of the transmission rate  $\beta$  according to the parametrization in Eq. (13a) and its comparison with data, see Fig. 1 and Tables III and IV. Shown are also the obtained values of the parameters  $k_0$  and  $k_1$  and the corresponding half-life time of the disease  $t_{1/2}$ . The horizontal red-shaded region represents the obtained value of  $k_1$  see Eq. (12b). See the text for more details.

The fitting values of the parameters of model B for the same countries as in model A, are shown in Table III. Firstly, we notice again that the asymptotic values  $P_\infty$  and  $D_\infty$  are quite similar to those obtained for model A. This indicates

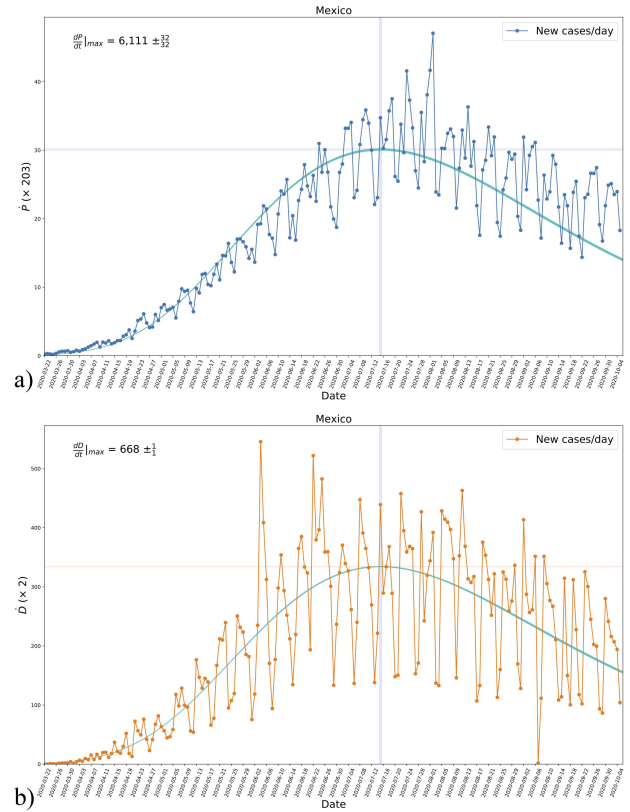


FIGURE 7. The derivatives  $\dot{P}$  and  $\dot{D}$  for positives a) and deaths b), respectively, obtained from the data of new daily cases and the analytical expression of model B, see Eq. (19), with the parameters fitted to the data, see Fig. 5. Even though the new daily cases were not used for the fitting, we see a consistent agreement with the results. The blue-shaded vertical regions mark the inflection time in each case, as in Fig. 6, whereas the blue-shaded horizontal regions mark the maximum values of each case. See the text for more details.

that model B, although simpler than model A, is also a good model to follow the evolution of the cumulative cases. Secondly, the similarity in the results extends to the other common parameters between the models, as is the case of  $t_D$ ,  $k_0$ , and  $k_1$ , which again supports the validity of model B to fit the data under simpler assumptions.

The same happens when one compares the fitting results of the half-life time  $t_{1/2}$  with those in Table II; they are very similar to each other for the respective countries. The similarity extends for the case of the inflection times  $t_0$  and  $t_{D0}$ , which are lower than the half-crossing times of model A. This is as expected, given that in model A, the half-crossing should happen after the inflection of the resultant evolution curve. As in Table II, we find that the lowest characteristic times correspond to Belgium, whereas the highest correspond to Bolivia.

We repeated the comparisons between the data and model B in the same form as in Fig. 3. The new results are shown in Fig. 6. As anticipated in Sec. 2.4, model B is also good at following the trend of the data and the only changes are in

TABLE IV. Fitted values of extra parameters in the case of model B, for the same countries as in Table I. Shown are the half-life time  $t_{1/2}$  of the transmission rate  $\beta$ , see Eq. (13b), and the inflection times of cumulative positives and deaths,  $t_0$  and  $t_{D0}$ , respectively, see Eqs. (21a).

Country	$t_{1/2}$	$t_0$	$t_{D0}$
Mexico	$36.95 \pm_{0.19}^{0.19}$	$115.03 \pm_{0.36}^{0.35}$	$115.06 \pm_{0.35}^{0.35}$
Peru	$45.63 \pm_{0.36}^{0.37}$	$121.59 \pm_{0.87}^{0.89}$	$133.56 \pm_{0.78}^{0.79}$
Belgium	$9.40 \pm_{0.10}^{0.09}$	$27.38 \pm_{0.71}^{2.40}$	$30.83 \pm_{0.13}^{0.12}$
Bolivia	$50.31 \pm_{1.03}^{1.08}$	$147.92 \pm_{2.40}^{2.54}$	$163.06 \pm_{2.61}^{2.73}$
Brazil	$36.12 \pm_{0.09}^{0.09}$	$107.45 \pm_{0.17}^{0.16}$	$107.45 \pm_{0.16}^{0.16}$
Chile	$28.18 \pm_{6.55}^{0.22}$	$86.18 \pm_{1.75}^{10.85}$	$100.99 \pm_{6.52}^{0.38}$
Ecuador	$37.30 \pm_{0.17}^{0.16}$	$93.90 \pm_{0.46}^{0.44}$	$93.96 \pm_{0.42}^{0.42}$
United States	$25.04 \pm_{0.02}^{0.02}$	$76.14 \pm_{0.05}^{0.05}$	$76.14 \pm_{0.05}^{0.05}$
United Kingdom	$11.59 \pm_{0.05}^{0.08}$	$40.24 \pm_{1.16}^{0.38}$	$41.19 \pm_{0.07}^{0.07}$

values of the fitted parameters, which also result in changes of the final quantities  $P_\infty$  and  $D_\infty$ , although the obtained values are consistent one to each other in their order of magnitude in the two models.

Finally, in Fig. 7, we show the comparison of the derivatives of model B for both the confirmed positives and deaths with their corresponding data. We also see a good agreement in both the time profile and the location and height of the maximum points in the two plots. This time model B is more easily manageable, and we show the maximum daily new cases as expected from the theoretical expectations. Notice that the time at the location of the maximum is the same as that of the inflection time  $t_0$  in the evolution of the cumulative cases, see Fig. 6.

#### 4. Final comments

We have presented a generalization of the well-known SI system to include the possibility of a time-dependent transmission rate and used a particular exponential-like parametrization of it to fit and describe the evolution of real data for the epidemics of Covid-19.

Although the use of time-dependent transmission rates is not new in the literature of epidemic models, this is the first time that such an approach have been applied to the SI system, for which there exists an analytical form of the general solution. The latter is the standard logistic function, but now with a generalized time variable that results from the direct integration of the transmission rate. These features simplify the handling of the model and ease its comparison with data.

For the functional form of the time-dependent transmission rate, we chose a decaying exponential with two free parameters, the first one for the initial value of the transmission rate and the second one for its decay rate. In this form, our generalized model has enough complexity to fit the data reliably, whereas at the same time provides meaningful quan-

ties to describe the evolution of cumulative positives and deaths.

One of those meaningful quantities is the decay rate, and the related one is the half-life time of the transmission rate. It was clear from our results that the half-life time is shorter for countries that have taken the strictest measures of public containment. Other countries seem to have an almost three times larger half-life time, which means that it will take longer for them to tame the epidemics. Even though in medical terms one may wish to have slow epidemics to avoid saturation of hospital services, this also means that containment measures will have to take place for longer too, which may result in a general public being tired of the governmental intervention.

Our model also suggests an interesting correlation between the initial value of the transmission rate and its decay rate: countries that experienced faster dissemination at the beginning are the ones that report a larger decay rate, as is the case of Belgium and the United Kingdom. Likewise, countries with a slow initial spreading seem to be the ones with also a lower decay rate. These last countries were not hit as badly as others at the start of their epidemics, but all so far indicates, according to our models, that they will have some of the highest death tolls in the world.

A surprising result was the connection of the SI system with the Gompertz function. As explained in the text, such connection requires some assumptions in between, mainly that the transmission rate is time-dependent with an exponential form, and that we consider a largely susceptible population. Although the Gompertz function is quite useful for a plethora of growth phenomena, ours is the first study that shows a derivation of it from an infectious model.

One final note on the fittings we obtained for the models. The Bayesian inference is an appropriate method to fit data if one faces a unique realization of the natural phenomenon under study, which is the case of the present epidemics of Covid-19, as the data reported by each country is not at all the result of repeated experiments under controlled conditions. In this respect, the Bayesian analysis allows us to do a sampling of values of the free parameters around the point of maximum likelihood. This does not mean that one finds the best and only fitting to data, but the best possible fit given the proposed model. This helps to explain the well-defined confidence regions in the triangle plots of the fitted parameters, even though the resultant curves may not even look good by eye when compared to data (see, for instance, model C in Appendix C).

The numbers reported here are not definitive, and they may change considerably if the epidemics follow a different evolution soon. However, we believe that our models may help characterize the present evolution of the disease and can be considered to decide about further public measures to handle the epidemics.

## Appendix

### A. Parametric form of the SI system

For completeness, we also show here an alternative derivation of the solution of the SI system in a parametric form. Considering a change of time-variable by the relation  $d\tau = (I/N)dt$ , the solution is

$$S(\tau) = S_i e^{-\beta\tau}, \quad I(\tau) = N - S(\tau), \quad (\text{A.1a})$$

where the explicit relation between  $t$  and  $\tau$  is

$$t = \int_{\tau}^0 \frac{d\tau}{1 - S(\tau)/N}. \quad (\text{A.1b})$$

The above solution can also be generalized in the case of a time-dependent transmission rate  $\beta = \beta(\tau)$ , and the only change is the solution for the susceptibles,

$$S(\tau) = S_i \exp \left[ - \int_{\tau}^0 \beta(x) dx \right]. \quad (\text{A.2})$$

The solutions for the number of infectives  $I(\tau)$  and the relation between the time variables  $t$  and  $\tau$  is the same as in Eqs. (A.1).

### B. Time-dependent SI model

Here we present an alternative method to solve the time-dependent SI system (B.1). If we take a general function  $\xi(t) = \beta S/N$ , then  $\dot{I} = \xi I$  and the number of infected is given by

$$I(t) = I_i \exp \left[ \int_t^0 \xi(x) dx \right], \quad (\text{B.1a})$$

whereas that of susceptibles, from the direct integration of the equation  $\dot{S} = -\xi(t)I(t)$ , is

$$S(t) = S_i + I_i - I(t). \quad (\text{B.1b})$$

Equation (B.1b) clearly shows the consistency of the solution as from it one recovers the constraint equation  $S + I = S_i + I_i = N$ .

For the case in which  $\beta = \text{const.}$ , the evolution equation for the susceptibles becomes a constraint equation for the functional form of  $\xi(t)$ , namely,

$$\xi(t) = \beta \left[ 1 - \frac{I_i}{N} \exp \left( \int_t^0 \xi(x) dx \right) \right]. \quad (\text{B.2})$$

The exact solution of Eq. (B.2) exists and can be obtained from the exact solution of the SI system, see Eq. (2). Hence, we find from  $\xi(t) = \dot{I}/I$ , or from a direct substitution in Eq. (B.2), that

$$\xi(t) = \beta \left[ 1 + e^{\beta(t-t_0)} \right]^{-1}. \quad (\text{B.3})$$

Notice that  $\xi(t \rightarrow -\infty) = -\beta$  and  $\xi(t \rightarrow \infty) = 0$ , and in this sense  $\xi(t)$  can be considered a kind of time-dependent transmission rate, in which the time dependence is clearly inherited from the evolution of the factor  $S(t)/N$ . Actually, the solutions obtained from Eqs. (B.1) are

$$\begin{aligned} \frac{S(t)}{N} &= \left[ 1 + e^{\beta(t-t_0)} \right]^{-1}, \\ I(t) &= \frac{I_i (1 + e^{\beta t_0})}{1 + e^{-\beta(t-t_0)}}, \end{aligned} \quad (\text{B.4})$$

and then one can also see, after a quick comparison with Eq. (3), that  $N = I_i(1 + e^{\beta t_0})$ .

The inflection point  $t_*$  in the evolution of infectives is found from the condition  $\ddot{I}(t_*) = 0$ , which explicitly reads

$$\left( \dot{\xi} + \xi^2 \right)_{t_*} = 0. \quad (\text{B.5})$$

We can see that a necessary condition to satisfy the preceding equation is  $\dot{\xi} < 0$ , ie, that  $\xi$  is a decaying function of time.

Another property of the general solution (B.1a) is that  $N$ , the total population number, is a free parameter. This is quite convenient, as it means that

$$\lim_{N \rightarrow \infty} \frac{I(t)}{N} = 0, \quad (\text{B.6})$$

except for the case  $\beta = \text{const.}$ , for which

$$\lim_{N \rightarrow \infty} \frac{I(t)}{N} = \left[ 1 + e^{\beta(t_0-t)} \right]^{-1}. \quad (\text{B.7})$$

The reason behind these results is the mechanism that puts an end to the epidemics. In the case  $\beta = \text{const.}$  the epidemics end because of the exhaustion of the susceptible population, whereas in any other time-dependent case, it ends because the transmission rate decays. In the first case, we then need to know the total population beforehand, whereas in the second one that number is irrelevant as long as  $I(t) < N$ .

Moreover, we can calculate the equivalent time-dependent transmission rate as

$$\xi(t) = \beta(t)S(t)/N = \beta(t)(1 - I(t)/N), \quad (\text{B.8})$$

and then  $\xi(t) \rightarrow \beta(t)$  in the limit  $N \rightarrow \infty$ . Thus, the solution of the infected population in the same limit can simply be written as

$$I(t) = I_i \exp \left[ \int_t^0 \beta(x) dx \right], \quad (\text{B.9})$$

which is the same functional form as model B in Sec. 3.2; see also Eq. (11b).

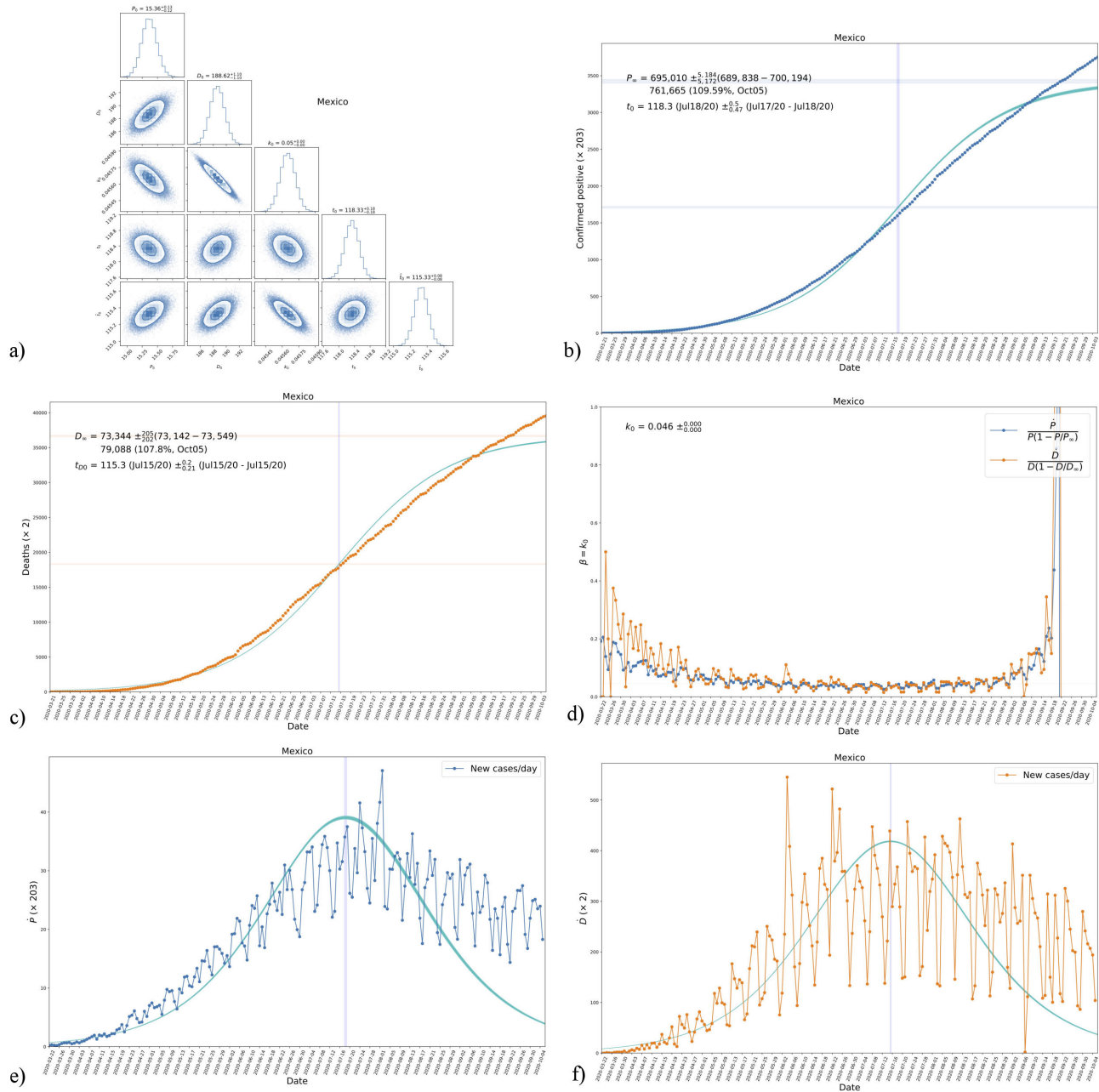


FIGURE 8. a) Triangle plot of the fitting to data from Mexico of the parameters  $P_0$ ,  $D_0$ ,  $t_P$ ,  $t_D$ , and  $k_0$  of model C. All parameters are well constrained by the combination of accumulated confirmed positives and deaths. The resultant evolution curves of cumulative confirmed positives b) and deaths c), according to model C with the values of its parameters as shown in the triangle plot. Shown in the figures are the obtained asymptotic values  $P_\infty$  and  $D_\infty$  in the top blue-shaded horizontal regions. The vertical blue-shaded regions mark the inflection time in each case,  $t_P$  and  $t_D$ , respectively; the region for  $P(t_P)$ , and  $D(t_D)$  are also shown for reference. d) The resultant transmission rate  $\beta$  and its comparison with data. The horizontal red-shaded region represents the obtained value of  $k_0$ . e) The derivatives  $\dot{P}$  and  $\dot{D}$  for positives f) and deaths g), respectively, were obtained from the data of new daily cases and the analytical expression (2) using the parameters fitted in the triangle plot. The blue-shaded vertical regions mark the inflection time in each case, as in Fig. 8. See the text for more details.

### C. Fitting to data with the time-independent SI system

For completeness, we show in this appendix the fitting to data using the time-independent logistic function (3). We use the

data from Mexico for easy comparison with the results of models A and B in the main text. The explicit functions for cumulative positives and deaths are

$$P(t) = P_0 \frac{1 + e^{k_0 t_P}}{1 + e^{k_0(t_P - t)}},$$

$$D(t) = D_0 \frac{1 + e^{k_0 t_D}}{1 + e^{k_0(t_D - t)}}. \quad (\text{C.1})$$

Figure 8 is a compilation of all results related to model C. The triangle plot (top left panel) shows the confidence regions for the free parameters in the model, namely  $P_0$ ,  $D_0$ ,  $t_0$ , and  $t_{D_0}$ . All parameters are well constrained, and their confidence regions are well defined too. However, when model C itself is compared directly with data, in the top right and middle left panels of Fig. 8, we see that the agreement is not good enough, as in both cases of cumulative positives and deaths, the asymptotic values are lower than the value of the last data point.

Likewise, in comparison with new daily cases, in the middle right panel of Fig. 8, we see that the fit is not good either, which confirms again that a constant transmission rate is not appropriate for the complexity of the data.

Finally, in the bottom panels of Fig. 8, we also compare the derivative of model C for a direct comparison with new daily cases. Here we see a clear offset of the predicted maximum with respect to that of the data, which leads us to conclude, with all results together, that the simple logistic function (3) is not appropriate to describe the evolution of real data from the epidemics of Covid-19.

## D. Montecarlo Analysis

To fit models A, B, and C, described in Secs. 3.1., 3.2. and Appendix , we used the Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler code: emcee [25,27], to explore the parameter space of the aforementioned models and estimate their values together with their corresponding uncertainties (for other sampler codes in the cosmology and astrophysical literature see [28]).

In general, an MCMC algorithm allows to randomly sample probability distributions. For instance, such distribution could correspond to a likelihood  $\mathcal{L}(D|\vec{\theta})$ , which is the conditional probability of having the data  $D$  given a model with a vector of free parameters  $\vec{\theta}$ . Another instance would be the posterior distribution  $P(\vec{\theta}|D)$ , which is the conditional probability of having a preferred vector of free parameters  $\vec{\theta}$ , assuming a parametric model  $M$ , given the data  $D$ .

The two preceding distributions are related by Bayes theorem through the prior distribution, itself a probability distribution, of the parameters in the model  $\pi(\vec{\theta})$ , and then

$$P(\vec{\theta}|D) = \frac{\mathcal{L}(D|\vec{\theta})\pi(\vec{\theta})}{P(D)}, \quad (\text{D.1})$$

where we have also added the probability of the data  $P(D)$ . All the probability functions have to be normalized to unity as a function of the parameters in the model, e.g.  $\int \pi(\vec{\theta})d^n\theta = 1$ , where  $n$  is the dimension of the parameter space  $\vec{\theta}$ .

In this sense, the also called marginal likelihood  $P(D)$  is just the normalization factor of the posterior  $P(\vec{\theta}|D)$ :

$\int \mathcal{L}(D|\vec{\theta})\pi(\vec{\theta})d^n\theta = P(D)$ . In practice, being the quantity  $P(D)$  and overall constant factor, to sample the posterior  $P(\vec{\theta}|D)$  one typically only works with the quantity  $\mathcal{L}(D|\vec{\theta})\pi(\vec{\theta})$ .

Hence, the MCMC algorithm is used to find the maximum of the posterior, and in turn, the maximum likelihood estimators  $\hat{\vec{\theta}}$ , together with the confidence regions around the maximum to estimate the uncertainty in the free parameters in the model. As explained in Ref. [28] for emcee, many *walkers* move through the parameter space with trial moves, where the latter is calculated on the positions of each of the other walkers to take into account information about the underlying distribution. The algorithm is then run until the mean position of the walkers makes only small oscillations around the mean parameters.

Although one may think that Bayesian statistics, as explained above, simply leads to the best fitting model (so-called goodness-of-fit), the inclusion of priors helps to avoid the danger of overfitting, that is, to avoid having more parameters than strictly needed to describe the data (for more details see for instance [29-31] and references therein).

For our purposes, the data  $D$  corresponds to either the number of confirmed positive cases or the number of deaths, and model  $M$  is one of those described in each of the mentioned sections. Given the nature of the data, i.e. independent counts of confirmed cases of Covid-19, we use a Poisson likelihood distribution, instead of the common and the simple assumption of a Gaussian one, in the form,

$$\mathcal{L}(D|\vec{\theta}) = \prod_i \frac{e^{-\lambda_i} \lambda_i^{d_i}}{d_i!}, \quad (\text{D.2})$$

where the product is over the number of days  $i$ , or days bins, in the time series,  $d_i$  is the number of cases observed each day, and  $\lambda_i = \lambda_i(\vec{\theta})$  is the corresponding model prediction.

As part of the procedure, we consider the two samples, positive and death counts, as independent, which is a simple approximation as the data sets must necessarily be correlated. The only way in which we incorporate such correlation is through the parameter  $t_D$ , the delay time between positives and deaths. Finally, the total likelihood is simply the product of the likelihoods for positive and death cases:  $\mathcal{L}(D|\vec{\theta}) = \mathcal{L}(\text{Positive}|\vec{\theta}) \times \mathcal{L}(\text{Death}|\vec{\theta})$ , where  $\vec{\theta}$  is the full vector of free parameters in the model.

We have assumed flat priors for all the parameters, which in our case are uninformative enough, and with ranges that allow us to explore the parameter space. To ease the exploration of the parameter space, the initial point in the MCMC chains is chosen to correspond with the maximum of the posterior found by a simple maximization routine.

As mentioned before, we used emcee for the MCMC sampling, for which the the length of the chain of 20,000 (30,000 in some cases) steps and 100 independent walkers. Those numbers were chosen after testing different setups to ensure convergence and a good sampling of the posterior.

Finally, we checked that the length of the chain was

longer was than the auto-correlation time. The code was run using multithreading to allow us to use multiple processors and increase the calculation speed. A *Jupyter* notebook with a sample code to run similar analyses as those presented in this work can be found in [32].

## Acknowledgments

We are thankful to Argelia Bernal, Juan Barranco, Nana Cabo, Gustavo Niz, and Demián Mayorga for enlighten-

ing conversations and comments. AXG-M acknowledges support from Cátedras CONACYT. This work was partially supported by Programa para el Desarrollo Profesional Docente; Dirección de Apoyo a la Investigación y al Posgrado, Universidad de Guanajuato, research Grant No. 036/2020; CONACYT México under Grants No. A1-S-17899, 286897, 297771, 304001; and the Instituto Avanzado de Cosmología Collaboration.

- 
- i.* It is possible to define new normalized variables in the form  $\hat{S} = \beta S/N$  and  $\hat{I} = \beta I/N$ , and then Eqs. (1) are simply written as  $\dot{\hat{S}} = -\hat{S}\hat{I}$  and  $\dot{\hat{I}} = \hat{S}\hat{I}$ . The constraint equation also becomes  $\hat{S} + \hat{I} = \beta$ , and then the transmission rate only appears for the initial conditions.
- ii.* All data considered in this work have been taken from: <https://ourworldindata.org/coronavirus-source-data>
1. P. V. D. D. W. J. Brauer, F., Seasonal and pandemic influenza: Strategic and tactical models, *Can. Appl. Math. Q.* **19** (2010) 139.
  2. A. Ahmada, S. Garhwal, S. K. Ray, G. Kumar, S. J. Malebary, and O. M. O. Barukab, *The number of confirmed cases of covid-19 by using machine learning: Methods and challenges* (2020), arXiv:2006.09184[cs.SI]
  3. A. G. M. Neves and G. Guerrero, *Predicting the evolution of the covid-19 epidemic with the a-sir model: Lombardy, italy and são paulo state, brazil* (2020), arXiv:2005.11182[q-bio.PE]
  4. S. Afonso, J. Azevedo, and M. Pinheiro, *Epidemic analysis of covid-19 in brazil by a generalized seir model* (2020), arXiv:2005.11420 [q-bio.PE]
  5. H. Hult and M. Favero, *Estimates of the proportion of sars-cov-2 infected individuals in sweden* (2020), arXiv:2005.13519[q-bio.PE]
  6. T. Piasecki, P. B. Mucha, and M. Rosińska, *A new seir type model including quarantine effects and its application to analysis of covid-19 pandemia in poland in march/april 2020* (2020), arXiv:2005.14532[q-bio.PE]
  7. D. Faranda and T. Alberti, *Modelling the second wave of covid-19 infections in france and italy via a stochastic seir model* (2020), arXiv:2006.05081[q-bio.PE]
  8. Z. S. Khan, F. V. Bussel, and F. Hussain, *A predictive model for covid-19 spread applied to six us states* (2020), arXiv:2006.05955[q-bio.PE]
  9. D. Pulido, D. Basurto, M. Cándido, and J. Salas, *Geospatial spread of the covid-19 pandemic in mexico* (2020), arXiv:2006.07784[physics.soc-ph]
  10. M. A. Capistran, A. Capella, and J. A. Christen, *Forecasting hospital demand during covid-19 pandemic outbreaks* (2020), arXiv:2006.01873[q-bio.PE]
  11. R. A. Barrio, K. K. Kaski, G. G. Haraldsson, T. Aspelund, and T. Govezensky, *Modelling covid-19 epidemic in mexico, finland and iceland* (2020), arXiv:2007.10806 [physics.soc-ph]
  12. J. Zelner, J. Riou, R. Etzioni, and A. Gelman, *Accounting for uncertainty during a pandemic* (2020), arXiv:2006.08745 [stat.AP]
  13. A. Altmejd, J. Rocklöv, and J. Wallin, *Nowcasting covid-19 statistics reported withdelay: a case-study of sweden* (2020), arXiv:2006.06840 [q-bio.PE]
  14. T. Carletti, D. Fanelli, and F. Piazza, *Covid-19: The unreasonable effectiveness of simple models* (2020), arXiv:2005.11085[q-bio.PE]
  15. M. Fiacchini and M. Alamir, *The ockham's razor applied to covid-19 model fitting french data* (2020), arXiv:2007.11379[stat.AP]
  16. N. C. Bizet and A. C. M. de Oca, *Modelos sir modificados para la evolución del covid19* (2020), arXiv:2004.11352[q-bio.PE]
  17. A. Palladino *et al.*, *Modelling the spread of covid19 in italy using a revised version of the sir model* (2020), arXiv:2005.08724 [physics.soc-ph]
  18. S.-P. Chao, *Simplified model on the timing of easing the lockdown* (2020), arXiv:2007.14072 [physics.soc-ph]
  19. Y. Nesterov, *Online analysis of epidemics with variable infection rate* (2020), arXiv:2007.11429 [physics.soc-ph]
  20. B. Gompertz, *On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies*, *Philosophical Transactions of the Royal Society of London B* **182** (1825) 513.
  21. C. P. Winsor, *The Gompertz Curve as a Growth Curve*, *Proceedings of the National Academy of Sciences* **18** (1932) 1.
  22. P.-F. Verhulst, *Notice sur la loi que la population suit dans son accroissements*, *Corr. Math. Physics* **10** (1838) 113.
  23. R. Pearl and L. J. Reed, *On the Mathematical Theory of Population Growth*, *Metron* **3** (1923) 6.
  24. A. Tsoularis and J. Wallace, *Analysis of logistic growth models*, *Mathematical Biosciences* **179** (2002) 21.
  25. D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *emcee: The MCMC Hammer*, **125** (2013) 306. arXiv:1202.3665 [astro-ph.IM]
  26. K. M. Tjørve and E. Tjørve, *The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family*, *PLoS ONE* **12** (2017) 1. <https://doi.org/10.1371/journal.pone.0178691>

27. emcee, <https://emcee.readthedocs.io/en/stable/>.
28. R. Allison and J. Dunkley, Comparison of sampling techniques for Bayesian parameter estimation, *Monthly Notices of the Royal Astronomical Society* **437** (2013) 3918. arXiv:1308.2675.
29. R. Trotta, Bayes in the sky: Bayesian inference and model selection in cosmology, *Contemporary Physics* **49** (2008) 71, arXiv:0803.4089.
30. C. R. Jenkins and J. A. Peacock, The power of Bayesian evidence in astronomy, *Monthly Notices of the Royal Astronomical Society* **413** (2011) 2895. <https://doi.org/10.1111/j.1365-2966.2011.18361.x>
31. G. Efstathiou, Limitations of Bayesian Evidence applied to cosmology, *Monthly Notices of the Royal Astronomical Society* **1320** (2008) 1314. <https://doi.org/10.1111/j.1365-2966.2008.13498.x>
32. L. A. Ureña Lopez and A. X. González-Morales, *Data Lab Covid-19 notebook*, <https://colab.research.google.com/drive/17RzCjkhXn9-jLcn6P70XnkReJsx52Npy#scrollTo=ncYYSNT-UVa3>.