Predicting the evolution of the COVID-19 epidemic with the A-SIR model: Lombardy, Italy and São Paulo state, Brazil

Armando G.M. Neves [a,c,1,*], Gustavo Guerrero [b,c,1]

[a] *Universidade Federal de Minas Gerais, Departamento de Matemática, Avenida Antônio Carlos, 6627, Caixa Postal 702, 31270-901 Belo Horizonte, MG, Brazil*
[b] *Universidade Federal de Minas Gerais, Departamento de Física, Brazil*
[c] *COVID-19 Modeling Task Force, Minas Gerais, Brazil*

ABSTRACT

The presence of a large number of infected individuals with few or no symptoms is an important epidemiological difficulty and the main mathematical feature of COVID-19. The A-SIR model, i.e. a SIR (Susceptible–Infected–Removed) model with a compartment for infected individuals with no symptoms or few symptoms was proposed by Gaeta (2020). In this paper we investigate a slightly generalized version of the same model and propose a scheme for fitting the parameters of the model to real data using the time series only of the deceased individuals. The scheme is applied to the concrete cases of Lombardy, Italy and São Paulo state, Brazil, showing different aspects of the epidemic. In both cases we see strong evidence that the adoption of social distancing measures contributed to a slower increase in the number of deceased individuals when compared to the baseline of no reduction in the infection rate. Both for Lombardy and São Paulo we show that we may have good fits to the data up to the present, but with very large differences in the future behavior. The reasons behind such disparate outcomes are the uncertainty on the value of a key parameter, the probability that an infected individual is fully symptomatic, and on the intensity of the social distancing measures adopted. This conclusion enforces the necessity of trying to determine the real number of infected individuals in a population, symptomatic or asymptomatic.

1. Introduction

Although there are good models for predicting the time evolution of an epidemic of diseases such as influenza or measles, models of the same type are not working for the COVID-19. An important feature of the COVID-19 is that it may be asymptomatic or mildly symptomatic in some patients, although causing severe respiratory symptoms in others. As a consequence, there is a large number of undocumented infections [1].

The lack of tests for assessing the health state of large samples of the populations contributes to the spread of the COVID-19, as asymptomatic individuals may not isolate themselves. Although there is no clear distinction between symptomatic and asymptomatic, in this paper we will use the acronym MSA (mildly symptomatic or asymptomatic) to mean a case of COVID-19 weak

enough for not causing death or lead to hospitalization and probably unreported due to the lack of tests.

We consider that a good step towards a good predicting model for COVID-19 has been taken in [2]. In Section 2 we will describe a slight generalization of the A-SIR model proposed in that work and use it in the rest of this paper in predicting the possible evolution of the COVID-19 epidemic in Lombardy, Italy and São Paulo state, Brazil.

Of course, as the pandemic progresses several different attempts are being made to understand it from a mathematical point of view. It is hard to quote and describe all the approaches, many of them still unpublished, so we cite here only some examples. Fanelli and Piazza [3] use the SIR model augmented by a compartment for dead individuals, with parameter estimation by the stochastic differential evolution algorithm [4]. Other papers [5–10] use models with compartmental models other than SIR or A-SIR, or age-structured populations. Examples of papers where the spatial localization of the population is taken into account are [11] and [12].

The A-SIR model is just the traditional SIR (Susceptible–Infected–Removed) model for epidemics, introduced by Kermack and McKendrick [13] almost 100 years ago, with one extra compartment for accounting the MSA infected individuals. The MSA

* Corresponding author at: Universidade Federal de Minas Gerais, Departamento de Matemática, Avenida Antônio Carlos, 6627, Caixa Postal 702, 31270-901 Belo Horizonte, MG, Brazil.
E-mail addresses: aneves@mat.ufmg.br (A.G.M. Neves), guerrero@fisica.ufmg.br (G. Guerrero).
[1] Both authors contributed to all phases of the research.

can still transmit the disease to susceptible individuals, but, as they mostly ignore their condition, it is reasonable that they will remain for larger periods transmitting the disease when compared to fully symptomatic individuals, which will probably isolate themselves after a few days. Of course, when the epidemic is already well developed, a large fraction of the population may have been MSA infected, and, as these individuals are healed, they will contribute to largely decrease the number of susceptible people.

Although the proportion of symptomatic cases has already been estimated to be 16% for the development of the disease in China [1], we will take the liberty to explore the possibility that this proportion may be larger or smaller.

As a support to the possibility that there are less symptomatic cases than previously estimated, we cite [14]. Referring to 11 European countries, the report states that "In all countries, we estimate there are orders of magnitude fewer infections detected than true infections, mostly likely due to mild and asymptomatic infections as well as limited testing capacity". Fig. 1 in that paper illustrates that.

Supporting the other possibility, Lavezzo et al. [15] state that at Vò, Italy, the asymptomatic cases were a fraction of 43.2% of the total. Although clearly casting some doubt, we also cite [16]. The paper states "Among the participants with positive results for SARS-CoV-2, symptoms of Covid-19 were reported (...) by 57% of those in the overall population-screening group. However, 29% of participants who tested negative in the overall population-screening group also reported having symptoms".

One reason for the uncertainty in the outcome of mathematical models for COVID-19 is that the models usually contain parameters for which reasonable values are taken, but sometimes without full scientific support. In particular, the models are extremely sensitive to the infection rate β_0, see 1. We will show in this paper how to ignore the data on the number of currently infected people. These are prone to not only a large uncertainty, because of the MSA cases, but also underreporting of the symptomatic cases due to the lack of tests. We will use only the data on the number of deaths due to the COVID-19. Such a decision has also been taken in [5]. The reason behind it is that we believe that the deaths data are a more faithful indicator than the number of cases. In most instances only the patients in severe conditions are being tested for COVID-19. We are aware, however, that underreporting in deaths is also possible due to the lack of tests, and also that political manipulation – both for increasing and for decreasing deaths numbers – cannot be discarded.

We will restrict for the time being to the study of the development of COVID-19 in Lombardy and in the state of São Paulo. Both cases result in good fits of the model to the data. It will turn out that an important part of the fitting procedure is the way of tuning the value of the infection rate β_0 to the data.

One important conclusion supported by our good fits – both in Lombardy and in São Paulo – is that the adopted social distancing measures taken in both localities did contribute to diminishing the number of deaths due to COVID-19 with respect to the expected behavior if no measures were taken.

An important question is what will happen when the social distancing measures currently in act in most countries are relaxed. One bad possibility is that a second wave of COVID-19 will arise. If not mitigated, the potential number of deaths in the second wave may be larger than the deaths in the first wave. Another possibility is that sufficient herd immunity will have been acquired by the populations after the present epidemic and no large increase of cases should happen after relaxation of the social distancing.

We will show in this paper that neither of the above possibilities can be ruled out for Lombardy. Part of our ignorance is due to the fact that one key parameter of the A-SIR model, the probability that a newly infected individual is symptomatic, is still largely unknown. Another reason for not being able to predict the future of the epidemic is that we do not know how much the social distancing measures adopted were effective in reducing the infection rate of the model.

In the case of São Paulo state, Brazil, the fraction of deaths up to now is much smaller than in Lombardy. Although this is good, it also means that the population is still very susceptible. Strong economic pressure is being exerted on politicians for relaxation of the social distancing measures. We predict that even in the best of the possibilities, the number of infected individuals will steadily grow for a large period and in its peak it will be much larger than present values. Thus, social distancing measures should not be relaxed before the number of infected individuals is small. We see that the increase in the number of cases may be catastrophic if social distancing measures are not strengthened.

The paper is organized as follows. Section 2 starts with a mathematical description of the model and all its parameters. Then we will talk about the linear regime, i.e. the behavior of the solutions of the model for a short time after the beginning of the epidemic. Finally, we will describe conditions for the population fraction of infected individuals to decrease and also the limit behavior after the epidemic is finished. Section 3 describes the procedure for finding values of the parameters such that the output of the A-SIR model fits well the deaths data, both in Lombardy and in São Paulo. The paper is closed by Section 4, in which we draw some conclusions on the results obtained, and by Section 5, in which we account for some changes in the conclusions because of new data released during the time the paper was being written.

2. The A-SIR model

Let $S(t)$, $I(t)$, $A(t)$ and $R(t)$ be the population fractions at time t respectively of susceptible, symptomatic infected, MSA infected and removed individuals. By susceptible, we mean individuals which were not yet infected by the SARS-Cov-2 virus. By symptomatic we mean *fully* symptomatic individuals and by MSA we mean individuals which have either *no* symptoms, or *few* symptoms. By removed we mean individuals which were either healed after infection, isolated (at a hospital or at home) or deceased. The fraction of removed individuals is composed by the sum of symptomatic removed $R_s(t)$ and MSA removed individuals $R_a(t)$, according to whether the individuals were fully symptomatic before removal, or had mild or no symptoms. The time span we are going to consider is of a few months, thus we may ignore births and disregard deaths by reasons other than infection. In particular, we suppose that MSA or susceptible individuals do not die and that all infected individuals do not become susceptible again, at least for the time span we are considering.

The A-SIR model is described by the following set of nonlinear ordinary differential equations:

$$\begin{cases} S'(t) &= -\beta_0 S(I + \mu A) \\ I'(t) &= \beta_0 \xi S(I + \mu A) - \gamma_s I \\ A'(t) &= \beta_0 (1 - \xi) S(I + \mu A) - \gamma_a A \\ R_s'(t) &= \gamma_s I \\ R_a'(t) &= \gamma_a A \end{cases} \quad (1)$$

The latter two equations are not essential for solving the system. In fact, we may calculate R_s and R_a simply by integrating respectively $I(t)$ and $A(t)$. Another simple property of the model, proved by summing all the 5 equations, is that the sum of the fractions S, I, A, R_s and R_a is a constant. If we take $\mu = 1$, this is exactly the same model as in [2], although with different notation.

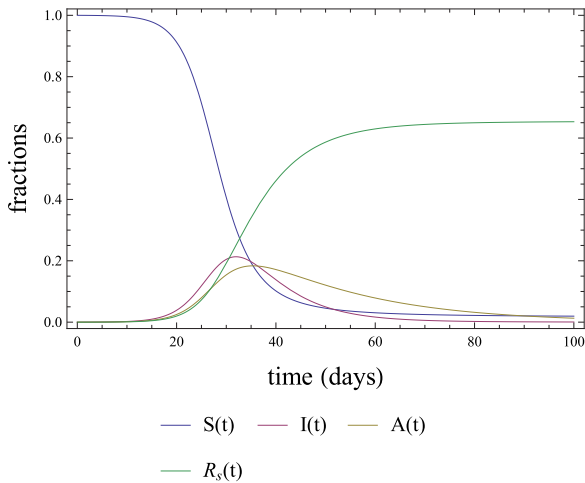All parameters above are considered to be positive and are interpreted as follows:
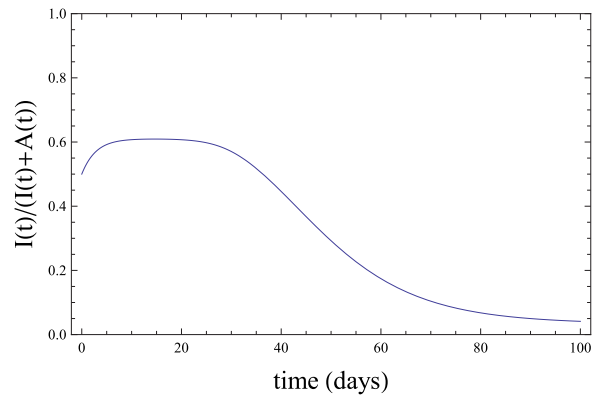
Fig. 2. The fraction of symptomatic cases as a function of time in the A-SIR model for the same parameter values and initial conditions as in Fig. 1.

Fig. 1. Typical behavior in the A-SIR model of the fractions of susceptible, symptomatic infected, MSA infected and symptomatic removed individuals. Parameter values: $\beta_0 = 0.5$, $\mu = 0.5$, $\xi = 2/3$, $\gamma_s = 1/7$, $\gamma_a = 1/21$. The initial conditions are $S(0) = 1$, $I(0) = A(0) = 0.0001$, $R(0) = 0$.

- β_0 is the infection rate of symptomatic individuals;
- $\mu \in (0, 1]$ is a reduction factor such that the infection rate for the MSA is $\mu\beta_0$;
- $\xi \in (0, 1)$ is the probability that a new infection event leads to a symptomatic case;
- γ_s and γ_a are respectively the inverses of the mean time symptomatic and MSA individuals remain infective. We suppose that $\gamma_s > \gamma_a$.

The mean removal time for symptomatic individuals will be considered to be around a week. This does not mean that individuals with symptoms will be healed after one week, but that these individuals, after showing symptoms for some days will either be hospitalized, or stay isolated at home. We will take then $\gamma_s = 1/7$ (days)$^{-1}$. Following [2], we will take $\gamma_a = 1/21$ (days)$^{-1}$, meaning that MSA individuals will remain active in the population for a larger time than symptomatic individuals.

We believe that we may take values such as above for γ_s and γ_a without risk of overestimating or underestimating the size of the epidemic. Other values could have been considered, but would not change the main conclusions. In the following we will explain how to use mortality data to infer the value of the contact rate β_0. It will turn out that the value of the parameter μ will not alter almost anything in the numeric predictions. On the contrary, we will see that the remaining parameter ξ alters drastically the outcome of the model in the future.

As a preparation for understanding things to come, we show in Fig. 1 a typical graph of the fractions S, I, A and R_s as functions of time for a seemingly reasonable choice of parameters. The graphs are obtained by numerically solving Eqs. (1).

For this choice of parameters, note that about 65% of the population are symptomatic removed 100 days after the start of the epidemic. Another thing to notice in the graphs of Fig. 1 is that, although we used $\xi = 2/3$ expecting to obtain that 2/3 of the cases are symptomatic, this does not happen. In fact, by time $t = 40$ the fraction of MSA is larger than the number of symptomatic individuals, even considering that the probability of a case being symptomatic is larger than the probability of a MSA case. The reason for obtaining a large number of MSA individuals is not related to ξ, but to the fact that the mean time γ_a^{-1} that an individual takes as MSA is larger than the mean time γ_s^{-1} taken by a symptomatic individual.

To better understand this important issue, in Fig. 2 we plot, for the same set of parameters and initial conditions, the ratio $I(t)/(I(t) + A(t))$ of symptomatic to total cases. The figure shows that, because of the initial condition $I(0) = A(0)$, the ratio of symptomatic to total infected individuals starts equal to 1/2. After a transient, it becomes almost constant around 0.6, not 2/3, and then decays to 0. Part of this behavior will be explained in the next subsection.

2.1. The linear regime

Another feature of the A-SIR model, already noticed in [2], is that the fraction $S(t)$ is very well approximated by 1 for the initial times. We may use this to approximate the solution of Eqs. (1) for small times. Substituting $S(t)$ by 1 in (1), we get

$$\begin{cases} I'(t) = (\beta_0\xi - \gamma_s)I + \beta_0\xi\mu A \\ A'(t) = \beta_0(1-\xi)I + (\mu\beta_0(1-\xi) - \gamma_a)A \end{cases}, \tag{2}$$

which is a linear system of ordinary differential equations with constant coefficients. Although the exact system (1) cannot be exactly solved, its linear approximation for initial times can be solved in terms of the eigenvalues and eigenvectors of its coefficient matrix

$$M = \begin{pmatrix} \beta_0\xi - \gamma_s & \beta_0\xi\mu \\ \beta_0(1-\xi) & \mu\beta_0(1-\xi) - \gamma_a \end{pmatrix}. \tag{3}$$

As M is 2×2, its eigenvalues can be easily calculated as roots of a quadratic polynomial. It can be shown that the eigenvalues of M are always real and that the smaller of them, denoted λ_-, is negative. The larger eigenvalue of M will be denoted λ_+ and is positive, provided that β_0 is not too small, as will be seen ahead. Although the formula for λ_+ in terms of the parameters of the model is somewhat large, we may invert it and find a rather simple formula for β_0 as a function of λ_+ and the remaining parameters:

$$\beta_0 = \frac{(\lambda_+ + \gamma_s)(\lambda_+ + \gamma_a)}{(\lambda_+ + \gamma_a)\xi + (\lambda_+ + \gamma_s)\mu(1-\xi)}. \tag{4}$$

In Section 3, we will use the above formula along with an estimate of λ_+ derived from the data to fix the parameter β_0.

A straightforward, but lengthy, calculation shows that for any initial conditions the solution of Eqs. (2) satisfies

$$\frac{I(t)}{I(t) + A(t)} \xrightarrow{t \to \infty} \rho, \tag{5}$$

with

$$\rho = 1 - \frac{\beta_0(1-\xi)}{\lambda_+ + \gamma_a + \beta_0(1-\xi)(1-\mu)}. \tag{6}$$

In the above equation $\lambda+$ and β_0 are related through Eq. (4). For the parameter values of Fig. 2 we have $\rho = 0.610411$, which is approximately the height of the plateau in that graph.

This agreement illustrates the more general fact that the exact solution of Eq. (2) is a good approximation to the true solution of Eq. (1) at the beginning of the epidemic. Of course, the exact solution of Eqs. (2) breaks down as an approximation for larger times. This is also shown in Fig. 2, because $\frac{I(t)}{I(t)+A(t)}$ is not always close to the value defined in Eq. (6).

Another consequence of Fig. 2 is that the ratio of symptomatic to total infected individuals is a dynamic quantity. It cannot be included as a parameter in the model's equations as in [1]. Moreover, as shown by Eq. (6), not even for the small interval of time in which the ratio $\frac{I(t)}{I(t)+A(t)}$ is approximately constant, it equals parameter ξ.

We will call *linear regime* the time interval in which the true solution of (1) is well approximated by the solution of (2). For the parameter values in Figs. 1 and 2 the linear regime lasts approximately up to time $t = 25$, in which the fractions of symptomatic and MSA individuals are already quite high.

In the linear regime, any of the quantities $I(t)$, $A(t)$, $R_s(t)$ and $R(t)$ are approximated by exact solutions having the form $c_1 e^{\lambda_+ t} + c_2 e^{\lambda_- t}$, where c_1 and c_2 are constants depending of which quantity we are calculating. As the term $e^{\lambda_- t}$ quickly tends to 0, we see that $I(t)$, $A(t)$, $R_s(t)$ and $R(t)$ are all approximated by $c_1 e^{\lambda_+ t}$, i.e., all of them are exponentially growing in the linear regime whenever $\lambda_+ > 0$. Most importantly, all of them grow at the same rate determined by the largest eigenvalue λ_+. For this reason, λ_+ is called the *Malthusian parameter* of the model [17].

2.2. Conditions for the fraction of infected individuals to decrease

It can be shown [17] that the total fraction of infected individuals $I(t) + A(t)$ in the solutions of (1) will decrease for all $t > 0$ if and only if $\lambda_+ \leq 0$. The same condition is generally written in terms not of λ_+, but of the basic reproduction ratio $\mathcal{R}_0$. $\mathcal{R}_0$ is defined as the mean number of individuals infected by a single infected individual during its whole infective period if the population is entirely susceptible. If $\mathcal{R}_0 \leq 1$, it can be shown that the number of infected individuals will initially decrease and always decrease in the A-SIR model. It is straightforward to calculate $\mathcal{R}_0$ for the A-SIR model following any of the recipes given in [17]. The result is

$$\mathcal{R}_0 = \beta_0 \left[\frac{\xi}{\gamma_s} + \frac{\mu(1-\xi)}{\gamma_a} \right] . \tag{7}$$

As commented before, the Malthusian parameter λ_+ will be positive if β_0 is sufficiently large. The exact condition is exactly that the right-hand side in the above equation is larger than 1, i.e. $\beta_0 > \gamma_s \gamma_a / (\gamma_a \xi + \gamma_s \mu(1-\xi))$.

If $\mathcal{R}_0 > 1$ and the whole population is susceptible, the total number of infected individuals will initially increase, but as the number of susceptible individual decreases, contagion becomes more difficult, and, consequently, the total number of infected will reach a maximum at some time t_*. In the simpler SIR model, it can be shown that t_* is the time such that $S(t_*) = 1/\mathcal{R}_0$. In the A-SIR model, no such simple condition exists, as we have two types of infected individuals and the number of one type may increase at the same time the other decreases. An instance of that is shown in Fig. 1 in the interval between the maximum point of I and the maximum point of A.

Gaeta [2] provided conditions for each of the fractions I and A to decrease. As Fig. 2 illustrates, whenever $\gamma_s > \gamma_a$ and t is large enough, the fraction of MSA individuals is much larger than the fraction of symptomatic infected individuals, so that $I(t) + A(t) \approx$

$A(t)$. We may use this fact to give an *approximate* condition for the fraction of total infected individuals to decrease.

In fact, the third equation in (1) shows that $A(t)$ decreases whenever

$$S(t) < \frac{\gamma_a A}{\beta_0(1-\xi)(I + \mu A)} = \frac{\gamma_a}{\beta_0(1-\xi)\left[\frac{I+A}{A} - (1-\mu)\right]} .$$

Substituting $\frac{I+A}{A}$ in the above formula for its approximate value 1 for large times, we get the *approximate* condition

$$S(t) < \frac{\gamma_a}{\beta_0(1-\xi)\mu} \tag{8}$$

for the decrease of the total number of infected individuals. We stress that the above condition is approximate and holds whenever $\gamma_s > \gamma_a$ and t is large enough.

Notice that the *threshold* on the right-hand side of (8) is higher for smaller β_0. As $S(t)$ starts close to 1 and decreases, the threshold will be easier to attain if β_0 is smaller. In other words, if β_0 is small, less people have to be infected in order that the number of infected individuals starts to decrease.

2.3. The asymptotic equilibrium

All solutions of the A-SIR equations (1) converge as $t \to \infty$ to the disease-free equilibrium in which $S = S_\infty$, I and A are both null and $R = 1 - S_\infty$. For a very contagious virus such as SARS-Cov-2, i.e. for large β_0, S_∞ is close to 0. In other words, almost the entire population is eventually infected in an unmitigated epidemic caused by a sufficiently contagious virus. This is illustrated in Fig. 1, in which S_∞ can be numerically calculated to be 0.0181302.

Individuals exit the susceptible compartment of the model either as symptomatic or MSA infected, and they do so at ratios ξ and $1 - \xi$, respectively. Since they will eventually become either symptomatic removed, or MSA removed, then the symptomatic removed and MSA removed fractions at equilibrium obey

$$\frac{R_s(\infty)}{R_a(\infty)} = \frac{\xi}{1-\xi} .$$

Moreover, $R_s(\infty) + R_a(\infty) = 1 - S_\infty$. Solving the set formed by the latter two equations, we obtain that $R_s(\infty) = (1 - S_\infty)\xi$. In the important case of a very contagious virus,

$$R_s(\infty) \approx \xi . \tag{9}$$

This approximation is also well illustrated in Fig. 1.

3. Fitting the A-SIR model to real COVID-19 epidemic data

According to Crisanti [18] and Fenga [19], the lack of efficient testing has been responsible for a substantial underestimation of the number of cases in the COVID-19 epidemic in Italy. In particular, probably due to testing preferentially the most severe cases, the mortality rate in the regions of Lombardy and Emilia-Romagna was three times larger than in the neighboring region of Veneto, in which testing for COVID-19 was widespread [18]. The same problem caused by lack of tests is reported also in Brazil and elsewhere.

Because of this, we believe that the cumulative number of deaths is a much more faithful indicator of the evolution of the epidemic than the number of confirmed cases, both in Lombardy and in São Paulo. We chose the deaths data in both locations as the sources for our fitting. This approach was also taken e.g. in [14]. The official deaths data, plotted in Fig. 3, were collected respectively in [20] and [21], along with the official numbers of confirmed cases.
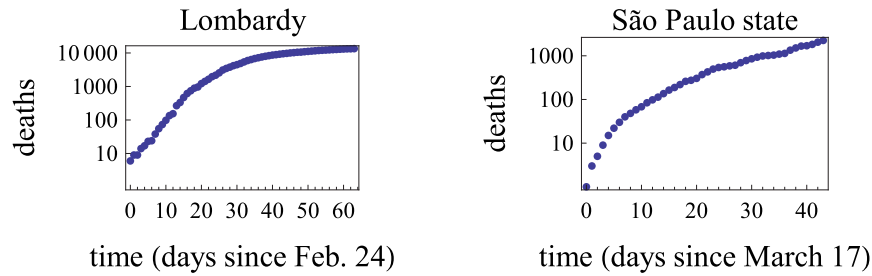
Fig. 3. The number of accumulated deaths due to COVID-19 in Lombardy since Feb. 24, 2020 (left) and in São Paulo state since Mar. 17, 2020 (right). *Source:* Data collected respectively in [20] and [21].

We need one extra parameter, ω, to relate the outcome of the model to the number of reported deaths used here. In fact, the A-SIR model does not make any prediction for the population fraction $D(t)$ of individuals dead due to COVID-19 up to time t. As only the symptomatic cases may die, it is natural to suppose that

$$D(t) = \omega R_s(t) ,\tag{10}$$

where $\omega \in (0, 1)$ is thus interpreted as the *case fatality rate*. The value for ω must also be found. As the lack of tests is a reality, an examination of data for several countries, as in [22], shows that the ratio of deaths to confirmed cases varies broadly among them.

Before entering into details, we describe the fitting procedure in general. Both in Lombardy and in São Paulo state, the epidemic started uncontrolled. Noticing the logarithmic vertical scale, we can see in both panels of Fig. 3 that in the first days the number of deaths seemed to increase exponentially, as expected for the linear regime, see sub Section 2.1. We will call these first days as the phase of *uncontrolled epidemic*. After this phase, in both locations the number of deaths started to increase at a lower rate. One of the important things will be to assess whether this lower rate is a natural consequence of the A-SIR model, or if it is due to the mitigation measures adopted in both locations.

As will be fully explained in the following, the first step will be using the deaths data in the uncontrolled epidemic phase to estimate the values of λ_+ and ω.

In the next steps, for each phase of social distancing we will have another parameter ϵ indicating the intensity of the adopted measures. In Lombardy we will consider two phases of different intensities ϵ_1 and ϵ_2 for the period of social distancing. In São Paulo state, only one phase of social distancing will be considered. We will see that it is possible to use the number of deaths data to obtain estimates for ξ and for the intensities ϵ_i. Although in general an optimal choice for these parameters exists, we will see that many possible choices are almost as good for the purpose of fitting the data with the model. It results that more than one possible good fit of the model to the data exists. We will explore the consequences of this approximate degeneracy in the optimization procedure.

Since parameters γ_a and γ_e are fixed, as already explained, and β_0 will be related to λ_+ by Eq. (4), the parameter μ still remains undetermined. After several experiences we noticed that, *as long as λ_+ is determined and β_0 is related to it by* (4), the results of the model are to large extent independent of μ. This is a consequence of the fact that the initial behavior of the deaths number according to the model is dictated by the linear regime, i.e., λ_+, whereas the final behavior is dictated by ξ, see Eq. (9). The above phenomenon is illustrated in Fig. 4, in which the results of A-SIR numerical solutions with two different values of μ show no noticeable differences. Thus, there will be no problem in adopting for μ any fixed value. We assume $\mu = 0.5$ for the rest of this paper.

The goodness of a fit between the number of deaths reported in the data and the A-SIR model will be quantified by a *cost function* to be minimized. For future reference, our choice of cost function is

$$c_d(\lambda_+, \xi, \mu, \omega) = \frac{N}{d} \sqrt{\sum_{i=0}^{d}(D(i) - \omega R_s(i))^2} .\tag{11}$$

In the above formula, besides other notations already introduced, N is the total population of the location (Lombardy or São Paulo state) and d is the last day up to which the numbers of reported deaths $D(i)$ are being compared to the prediction of the model $\omega R_s(i)$. We do not indicate the dependence of the cost function on the fixed parameters $\gamma_a = 1/21$ and $\gamma_s = 1/7$. The cost function also depends on the initial conditions to be used in the numerical solution of the model's differential equations. The initial conditions will be specified when needed. The value of d depends on which data is used for estimating the parameters. For example, for estimating λ_+ and ω we will take d to be the last day of the uncontrolled epidemic phase. For estimating the intensity of social distancing measures, we will use a large value for d. Of course neither the square root, nor the multiplication by N/d are necessary in the definition of the cost function, but they are included for convenience.

We will describe first the fitting procedure for Lombardy, where more data are available and then proceed to São Paulo state.

3.1. Fitting the epidemic in lombardy

The national lockdown in Italy started on Mar. 9, 2020, day number 14 after Feb. 24, the start date of the Italian time-series. We will define Mar. 9, 2020 to be the end of the uncontrolled epidemic phase for Lombardy.

For the numerical solution of Eqs. (1) in Lombardy we use initial conditions

$$S(0) = 1, \quad I(0) = 1.66 \times 10^{-5},$$
$$A(0) = \frac{1-\rho}{\rho}I(0), \quad R_s(0) = \frac{1}{\omega}6 \times 10^{-7} .\tag{12}$$

Here we used that the population of Lombardy is 10^7 inhabitants and that the number of confirmed COVID-19 cases on day 0 was 166 [20], accounting for the value of $I(0)$. As the initial number of MSA is unknown, our choice for $A(0)$ is a natural one such that the fraction $I(t)/(I(t)+A(t))$ is equal to ρ already on day 0, see (5) and (6). The initial condition for $R_s(0)$ is also the natural one using the number of people dead due to COVID-19 on day 0, obtained from the data [20], and taking into account Eq. (10).

We stress that the initial conditions for $I(0)$ and $A(0)$ are the only places in our fitting procedure where some information on the number of confirmed cases is used. We will not use such information on any other day. It is conceivable that on day 0 the
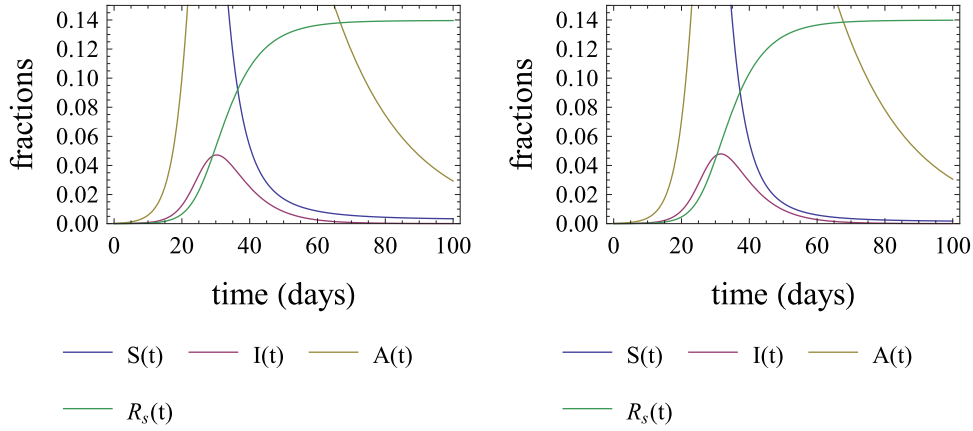
Fig. 4. Two almost indistinguishable solutions of the A-SIR model with $\gamma_s = 1/7$, $\gamma_a = 1/21$, $\xi = 0.16$, $\lambda_+ = 0.3$. The left panel was produced with $\mu = 0.25$ and the right panel with $\mu = 0.6$. Observe that β_0 is calculated by Eq. (4), thus it gets disparate values in the two simulations; 1.06995 for the solution in the left panel, and 0.555397 for the solution in the right panel. The initial conditions were $S(0) = 1$, $I(0) = A(0) = 6.22 \times 10^{-6}$ and $R_s(0) = 0$ for both cases.
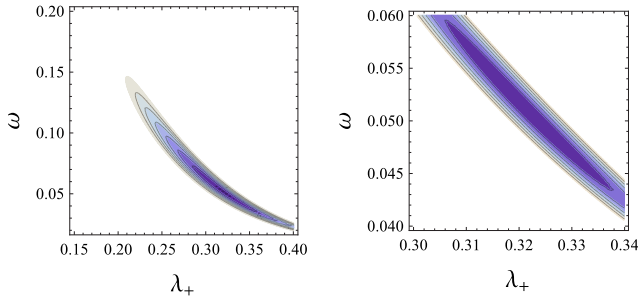

Fig. 5. Left: Contour plots of $c_{14}(\lambda_+, 0.1, 0.5, \omega)$ for a broad range in λ_+ and ω. Right: Zoom-up to the region of interest (darkest area of the left panel). Darker colors are smaller values and lighter colors are higher values of the cost function. In the white area the cost function has larger values than in the colored area. It seems clear that the function has a global minimum in the darker area of the right panel. The numerically determined location of the minimum point is $\lambda_+ = 0.3216$, $\omega = 0.051$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

number of confirmed cases is more reliable than in later days. Moreover, the exact values of $I(0)$ and $A(0)$ do not matter so much, as $I(t)$ and $A(t)$ initially grow exponentially by a rate to be determined by the number of deaths.

In Fig. 5 we show contour plots of the objective function $c_{14}(\lambda_+, 0.1, 0.5, \omega)$. Observe that we are considering only the deaths data up to day 14, i.e. in the uncontrolled epidemic phase. We arbitrarily fixed $\mu = 0.5$ and $\xi = 0.1$. The plots show that, for the chosen values of μ and ξ, in a large region in the (λ_+, ω) plane the cost function has a single local minimum, which is probably a global minimum. The set of parameter values that minimize the cost function, i.e., provide the best fit of the model to the data up to day 14, is $\lambda_+ = 0.3216$, $\omega = 0.051$.

Fig. 6 compares the epidemic data with the solution of Eqs. (1) with initial conditions (12) and parameters $\gamma_s = 1/7$, $\gamma_a = 1/21$ (fixed), $\mu = 0.5$, $\xi = 0.1$ (arbitrarily chosen) and $\lambda_+ = 0.3216$, $\omega = 0.051$ (optimally chosen with respect to the preceding values). The dots in the figure are the data for $D(i)$ divided by ω, and should thus approximate the curve for $R_s(t)$ up to $t = 14$. Numerical experiments (not presented for conciseness) with different values for μ and ξ confirm that the graph $R_s(t)$ up to $t = 14$, to be adjusted to data, almost does not change. Thus, the fit shown in the figure remains equally good independent of the values of μ and ξ. As already mentioned μ is quite irrelevant as far as β_0 is calculated as a function of λ_+. Also, the value of ξ is not relevant for reproducing the first days of the epidemic.
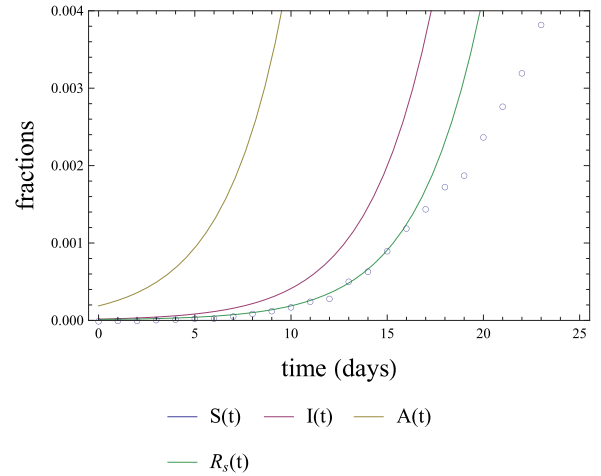

Fig. 6. Results of the A-SIR model, Eqs. (1) with initial conditions (12) and parameters $\gamma_s = 1/7$, $\gamma_a = 1/21$, $\mu = 0.5$, $\xi = 0.1$, $\lambda_+ = 0.3216$, $\omega = 0.051$. The latter two were chosen to minimize the cost function, c_{14}, using the former four parameters. The blue dots correspond to the number of deaths reported in pandemic data divided by ω. The graph of $S(t)$ lies out of the range shown in the figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

It is well clear by Fig. 6 that after day 14 (start of the national lockdown in Italy) the data increase less than the predicted R_s based on an uncontrolled epidemic. This is good evidence that the lockdown was effectively important in reducing the number of deaths due to COVID-19 in Lombardy.

We will incorporate the effect of social distancing measures in our model by introducing a decrease of the infection rate β_0 to a smaller value $\epsilon_1 \beta_0$, where $0 < \epsilon_1 < 1$. More precisely, in order to avoid introducing a discontinuous function into the system of differential equations, we replace β_0 in Eqs. (1) by the smooth function

$$\beta(t) = \beta_0 r(t) \tag{13}$$

with

$$r(t) = 1 - (1 - \epsilon_1)\theta(t - 14) . \tag{14}$$

In the above formula, $\theta(t)$ may be any continuous approximation of the unit step function. We used

$$\theta(t) = \frac{1}{2}(1 + \mathrm{erf}(t)) , \tag{15}$$
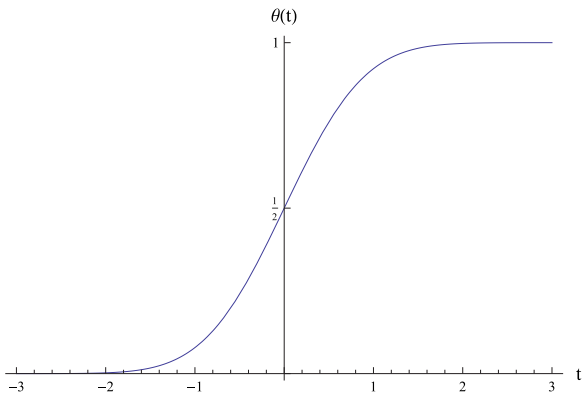
Fig. 7. Graph of the transition function $\theta(t)$ considered in Eq. (15).

where $\mathrm{erf}(z) = 2/\sqrt{\pi} \int_0^z e^{-t^2} dt$ is the integral of the normal distribution. The graph of $\theta(t)$ is shown in Fig. 7. Any similar continuous function switching from values close to 0 to values close to 1 in an interval around 0 of size 1 can be used without relevant changes. The important thing is that $\beta(t) \approx \beta_0$ for $t < 14$ and $\beta(t) \approx \epsilon_1 \beta_0$ for $t > 14$.

In the second step of our fitting procedure, we will fix the values of λ_+ and ω already estimated and evaluate the values for the intensity ϵ_1 of the first phase of the lockdown, and parameter ξ, which was not relevant in the uncontrolled epidemic phase. We referred above to the first phase of the lockdown, because a strengthening of it occurred on Mar. 22, day 27 after Feb. 24. Therefore, to estimate ϵ_1 and ξ we will use $d = 27$ in the cost function, Eq. (11).

Fig. 8 shows contour plots of the cost function c_{27} with parameters γ_s, γ_a, μ, λ_+ and ω having the same values as in Fig. 6, but allowing now variation of ξ and of the new parameter ϵ_1. The left panel shows a larger region in (ϵ_1, ξ) plane and the right panel shows in detail the region in which c_{27} attains its smallest values. We notice that, contrary to the analogous plots in Fig. 5, in which a clear global minimum is evidenced, this time the region where the cost function is close to minimum seems like a large "canyon".

Referring to the darker area in the right panel of Fig. 8, we will call *optimistic* choices for (ϵ_1, ξ) those points in the darker region with smaller values of ξ. In fact, because of Eqs. (9) and (10), such a choice will lead to a small expected number of dead people at

the end of the epidemic. On the contrary, points in the darker region with larger values of ξ will be called *pessimistic choices* for (ϵ_1, ξ).

We show in Fig. 9 plots of two solutions to Eqs. (1) with optimization up to day 27. Both choices for (ϵ_1, ξ) are almost equally good in fitting the data for days 0 to 27, however the results on the left panel were obtained with an optimistic choice for (ϵ_1, ξ), while the results on the right panel were obtained by considering a pessimistic choice.

In both panels of Fig. 8 we also see that after day 27 the green dashed curve of $R_s(t)$ grows faster than the data points. This suggests us that the strengthening of the lockdown in Lombardy after Mar. 22 did produce effects in slowing down the number of deaths.

We will then introduce a further parameter ϵ_2 such that $\beta(t) = \epsilon_2 \beta_0$ for $t > 27$. In order to see the effects of relaxing the social isolation measures, we will also restore $\beta(t)$ smoothly to its value β_0 for $t > 100$. Precisely, we will take

$$r(t) = 1 - (1 - \epsilon_1)\theta(t - 14) - (\epsilon_1 - \epsilon_2)\theta(t - 27)$$
$$+ (1 - \epsilon_2)\theta(t - 100). \tag{16}$$

The value for ϵ_2 will be determined by minimizing the cost function up to day 63, April 27. It turns out that the best choice for ϵ_2 will depend on the choice made for ξ and ϵ_1. In Fig. 10 we show the behavior of the cost function, c_{63}, when varying ϵ_2 for both the optimistic and pessimistic choices for (ϵ_1, ξ) already considered in Fig. 9.

As the final result of the fitting procedure for Lombardy, we show in Fig. 11 the graphs of the solutions to the A-SIR model with two different sets of parameter values such that the model fits rather well the deaths data.

A conclusion to be drawn from the results in Fig. 11 is that the A-SIR model can fit the deaths data for the epidemic in Lombardy extremely well if we consider that the social distancing measures adopted there were less intense from Mar. 9 to Mar. 22, days 14 to 27, and more intense from Mar. 22 to June 3, days 27 to 100. On the other hand, the goodness of the fits does not tell us very much about the future, when the social distancing measures are relaxed. Besides the two parameter choices shown, we have many others almost as good as them. The choices shown here are extreme in the sense that $\xi = 0.04$ is not too much above the minimum value of ξ for the points in the darker region in the right panel of Fig. 8, and $\xi = 0.5$ is the maximum value in the
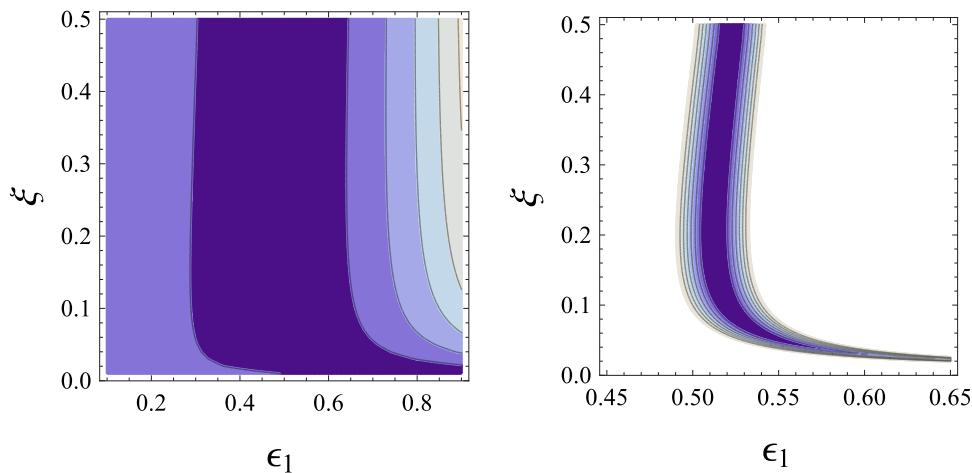

Fig. 8. Left: Contour plot of c_{27} as a function of ϵ_1 and ξ in a large region of parameter choices. Right: Zoom to the smaller region contained in the dark area of the left panel plot. The remaining parameters have the same values as in Fig. 6. Darker colors are smaller values and lighter colors are higher values of the cost function. The global minimum is located at $(0.513, 0.256)$, but the cost function is rather close to the minimum in the large darker region in the right panel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
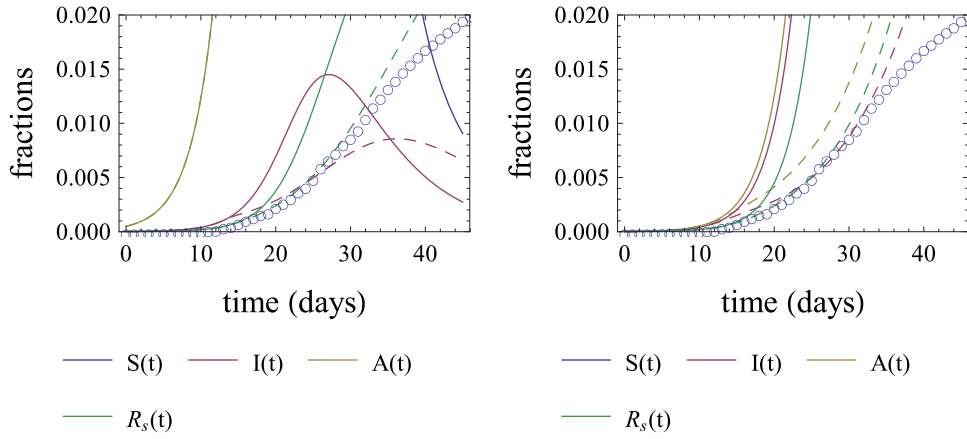
Fig. 9. Plots of two solutions to Eqs. (1) with different possibilities for the effect of the social distancing measures introduced by the lockdown of Mar. 9. In both panels, the full lines show, for comparison, what would be the solutions if the epidemic remained uncontrolled, whereas the dashed lines in corresponding colors show the effect of social distancing measures beginning at $t = 14$. The results on the left panel were produced with $\epsilon_1 = 0.566$, $\xi = 0.04$, and those on the right panel were obtained with $\epsilon_1 = 0.52$, $\xi = 0.5$. The remaining parameters are the optimal ones determined using the results presented in Fig. 6. The choice for the left panel is an example of what we called an optimistic choice. The right panel corresponds to a pessimistic choice for the parameters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
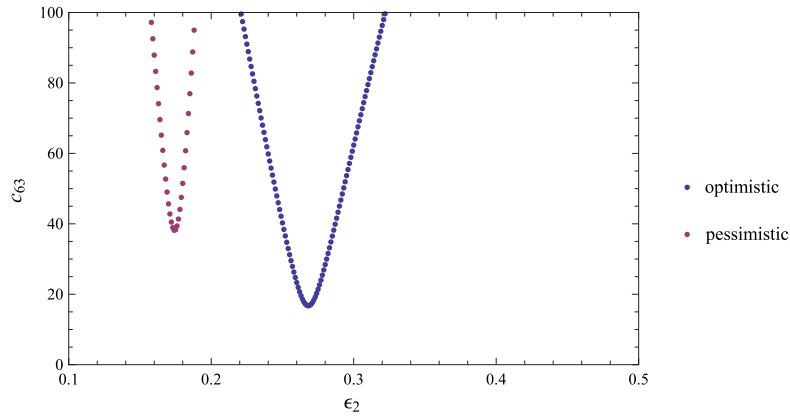

Fig. 10. Plots of the cost function c_{63} for several values of ϵ_2 and two choices of (ϵ_1, ξ), the same optimistic (blue) and pessimistic (red) choices already used in the panels of Fig. 9. The remaining parameter values are the ones used in Fig. 6. The minima of the cost function occur at $\epsilon_2 = 0.268$ (optimistic case), and $\epsilon_2 = 0.186$ (pessimistic case). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

same region. Due to Eqs. (10) and (9), these are also close to the extreme possibilities for the number of deaths at the end of the epidemic.

The optimistic choice, corresponding to the first row of Fig. 11, is such that after complete relaxation of the social distancing, the number of cases continues decreasing and the number of accumulated deaths increases slowly. This happens because the fraction of susceptible individuals since day 38 falls below the threshold on the right-hand side of Eq. (8) and does not increase very much above it after the social isolation measures are removed on day 100.

On the other hand, for the pessimistic choice, depicted in the second row of Fig. 11, the susceptible fraction is below the threshold on the right-hand side of Eq. (8) between days 27 and 100. The red line representing the threshold is not shown in the figure in this time interval, because the threshold is larger than 1. But the susceptible fraction stays high above the threshold after day 100, when social isolation measures are removed. As a consequence, there is a fast increase of the number of cases after day 100, i.e. a second wave of COVID-19 cases, potentially almost as intense as it would have been if the epidemic remained uncontrolled since the beginning.

If we had more knowledge either on the value of ξ or on the values of ϵ_1 and ϵ_2, we might choose among the most optimistic,

the most pessimistic, or some intermediate possibility between them. Due to lack of such knowledge, we present all possibilities.

The value of ξ could be estimated either by clinical research, or a large scale population screening. Such studies are on their way, and their results might help us in removing the uncertainties. In Section 5 we will comment that if the results of a population screening performed in Spain [23] were extrapolated to Lombardy, this would restrict the possibilities for ξ.

As far as ϵ_1 and ϵ_2 are concerned, some attempts have been made to measure the intensity of the social distancing using cell phone localization data, in particular for the regions in Italy [24]. A problem is that we do not know how to relate these measures with the decreases ϵ_1 and ϵ_2 in the infection rate.

3.2. Fitting the epidemic in São Paulo state

In Brazil the first imported case of COVID-19 was identified in the city of São Paulo on Feb. 26, 2020. The first official death, also in São Paulo, occurred on Mar. 17, 2020. For this reason, we chose that date as day 0 for the epidemic in Brazil. Although there has been up to now no nation-wide social distancing measures in Brazil, many state governors and mayors, including the governor of São Paulo state, decreed such measures. It is difficult to identify a clear starting date, as measures were gradual, but it seems
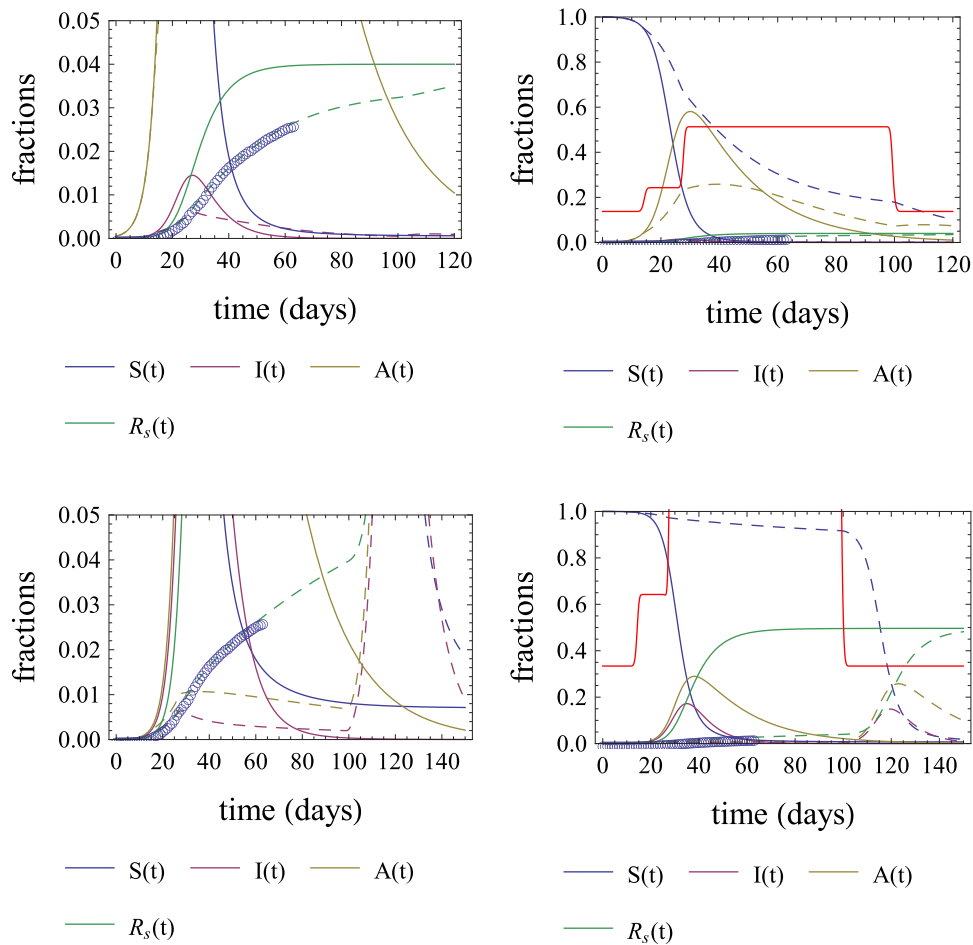
Fig. 11. Expected outcomes for the COVID-19 epidemic in Lombardy for up to 150 days after Feb. 24, compared with the deaths data. We are supposing, as a possibility, that the social isolation measures are completely relaxed on day 100, June 3. An optimistic outcome is presented in the first row, and a pessimistic one in the second row. The plots in the left and right columns differ only in the range of the vertical scale. The parameter values valid for both rows are: $\gamma_s = 1/7$, $\gamma_a = 1/21$, $\lambda_+ = 0.3216$, $\omega = 0.051$, $\mu = 0.5$, see Fig. 6. For the first row only, $\epsilon_1 = 0.566$, $\xi = 0.04$, $\epsilon_2 = 0.268$. For the second row only, $\epsilon_1 = 0.52$, $\xi = 0.5$, $\epsilon_2 = 0.186$. The red lines appearing in the second column are the graphs of $\frac{\gamma_a}{\beta(t)(1-\xi)\mu}$, see Eq. (8). Notice that with a good approximation, when the graph of $S(t)$ is below the red line, the fraction of MSA infected individuals decreases. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reasonable to choose Mar. 23, day number 6 after the first death, as the end of the uncontrolled epidemic phase in the state of São Paulo. Mar. 23 was, in fact, the first day on which all schools were closed in São Paulo state.

After the start of the social distancing measures, cell phone localization data [25] show a tendency of slowly decreasing efficiency of these measures. Although some metropolitan areas in Brazil have already declared lockdown, no such measure has been declared at São Paulo up to the time of the latest data shown in this paper, Apr. 29. As a consequence, we prefer to use a single reduction of infection rates in fitting the epidemic in São Paulo.

The fitting procedure for São Paulo is thus similar to the one for Lombardy, with the exception that we use only one social distancing reduction factor ϵ_1 for the entire period of the data. Taking into account that the population of the state of São Paulo is $N = 4.6 \times 10^7$ people, the initial conditions used for the numerical solution of the ODEs (1) were

$$S(0) = 1, \quad I(0) = 3.56522 \times 10^{-6},$$

$$A(0) = \frac{1-\rho}{\rho} I(0), \quad R_s(0) = \frac{1}{\omega} 2.17391 \times 10^{-8}. \tag{17}$$

The choice of these conditions is analogous to what was done for Lombardy.

The uncontrolled epidemic phase lasted from day 0 (Mar. 17) to day 6 (Mar. 23). We used the deaths during that phase to

obtain estimates for λ_+ and ω. Fig. 12 shows that the cost function $c_6(\lambda_+, 0.1, 0.5, \omega)$ apparently has a global minimum at $(\lambda_+, \omega) = (0.302, 0.074)$.

After determining λ_+ and ω, we used the deaths data up to day 43 to try to determine the parameter ξ and the reducing factor ϵ_1 for the infection rate during the social distancing measures started on day 6. Similar to Lombardy, Fig. 13 shows a canyon shaped region in which the cost function c_{43} is close to its minimum. The parameters used in the figure, besides the optimal values for λ_+ and ω, are specified in its caption.

As in the case of Lombardy, we are faced with the fact that good fits of the A-SIR model with the deaths data can be obtained for a continuum of values for ξ. For small values of ξ we have optimistic outcomes, in the sense that the number of deaths at the end of the epidemic is smaller. For larger values of ξ, we have pessimistic scenarios. Fig. 14 shows results both of an optimistic and a pessimistic solution, with different vertical scales.

One important difference between the cases of São Paulo and Lombardy is that the social distancing period started much earlier in São Paulo. As a consequence, the fraction of deceased people in São Paulo is much smaller up to now. The data points in the graphs of the second column in Fig. 14 are almost invisible. Of course, this is good, but it also has a bad side. In both optimistic and pessimistic cases the fraction of susceptible individuals remains for a long period above the threshold on the right-hand
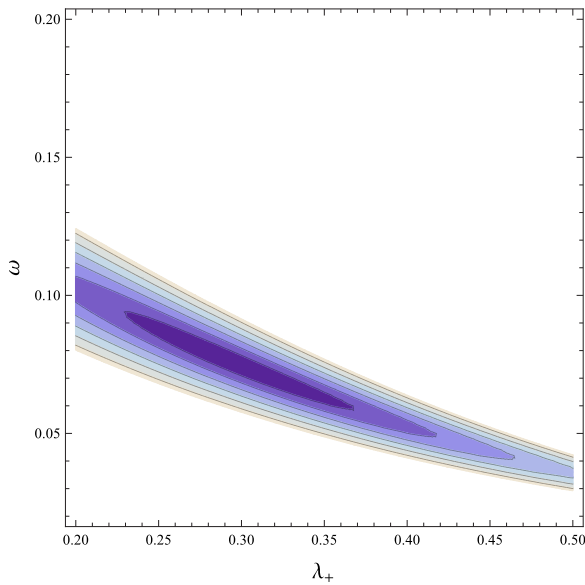
Fig. 12. Contour plot of $c_6(\lambda_+, 0.1, 0.5, \omega)$ for the state of São Paulo. Darker colors are smaller values and lighter colors are higher values of the cost function. In the large white area the cost function has larger values than in the colored area. It seems clear that the function has a global minimum in the darker area. The numerically determined location of the minimum point is $\lambda_+ = 0.302$, $\omega = 0.074$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
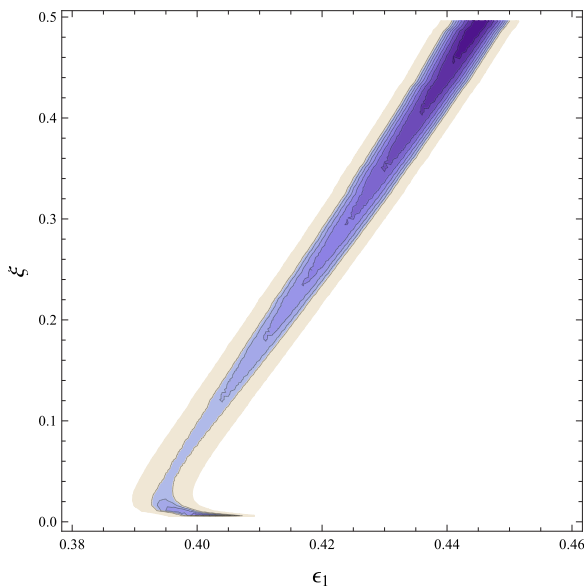

Fig. 13. Contour plot of c_{43} as a function of ϵ_1 and ξ. The remaining parameters are $\gamma_s = 1/7$, $\gamma_a = 1/21$ (fixed), $\mu = 0.5$ (arbitrary) and $\lambda_+ = 0.302$, $\omega = 0.074$ (determined by minimizing c_6). Darker colors are smaller values and lighter colors are higher values of the cost function. The blue canyon shaped region is where c_{43} is close to its minimum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

side of Eq. (8). Consequently, the fraction of infected individuals will be growing for a longer time when compared to Lombardy.

In the optimistic case, we see that if the social distancing measures remain with the present intensity until day 100 (Jun. 23), the fraction of infected individuals will increase up to a maximum by day 80 and start decreasing. If the social distancing measures

are removed on day 100, the fraction of infected individuals will resume growth for some days and then it will decrease again.

In the pessimistic case, the number of infected individuals will grow to a number much larger than the present one. Even if the social distancing measures are relaxed after day 120, there will still be a rapid increase in the number of infected people.

In both cases, we predict that the number of infected individuals at present will still increase considerably before attaining its peak value, which will be by day 90 in the optimistic case, or by day 130 in the pessimistic case. Given that the health services at São Paulo are already operating close to their maximum capacity, in both cases we see that it is necessary that social distancing is intensified in order to raise the threshold for the decrease of the infected fraction and prevent their collapse.

4. Conclusions

The COVID-19 pandemic forced us, scientists, to tackle the difficult task of trying to understand a new disease at the same time it is killing people in our neighborhoods and stressing our health services. As shown in this paper, the lack of solid knowledge on basic questions produces also an ignorance of what may happen in the future, even the present situation being well described by a simple mathematical model. We hope that more basic research may help fill the knowledge gaps, but probably that will take time.

Since the beginning of the pandemic, we have seen many papers with very different predictions on the population fraction that may die as a consequence of COVID-19, most of them too catastrophic. Part of this disparity is due to our ignorance on basic facts about the virus and the disease, as already remarked, and on the number of mildly symptomatic or asymptomatic cases. Another part is a consequence of the difficulty in estimating the many parameters in any realistic mathematical model.

In this paper we used a model as simple as possible in order to have the minimum number of parameters, and devised a procedure to estimate these parameters based only on the more faithful data, the number of deaths. We avoided using information on the number of confirmed cases, or trying to guess underreporting factors. If we had considered more complete models, we would probably have to estimate a larger number of parameters, resulting in a larger uncertainty. We believe that our results, uncertain as they are, may be useful in showing both a worst and a best possible outcome. One important result of our calculations is that the epidemic both in Lombardy and in São Paulo would have been much worse if social distancing measures had not been taken.

We saw that in an optimistic possibility, the number of cases of COVID-19 in Lombardy might not increase after the social distancing measures are relaxed. However, we also saw that it is likely that the number of cases shows a new quick rise, being then necessary either to keep these measures for longer, or use alternative measures. As social distancing is being relaxed at many countries, particularly in Italy, it is possible that the uncertainty in our results will be solved in the next days according to whether the number of cases will grow rapidly, or not.

In São Paulo state, as the fraction of infected individuals is up to now much smaller than in Lombardy, herd immunity is still far. As a consequence, our calculations predict that the fraction of infected individuals will still increase for some time, even in the most optimistic case. As stressed before, strengthening social distance measures could alleviate this situation.

We believe that the two locations in which we fitted the model to real data may serve as examples of what may happen in other locations.
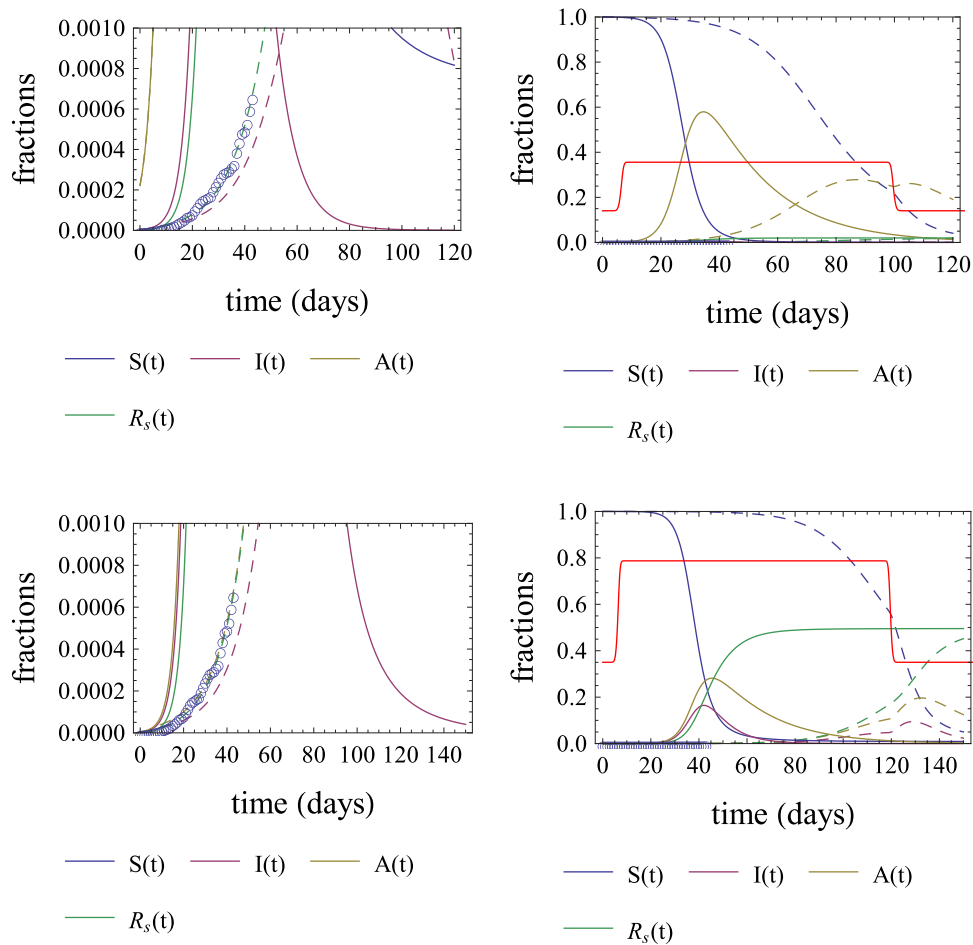
Fig. 14. Plots of expected outcomes for the COVID-19 epidemic in the state of São Paulo for up to 150 days after Mar. 17. The blue dots compare the results with epidemic deaths data. The optimistic and pessimistic outcomes are presented in the first and second rows, respectively. The plots in the different columns are different only in the range of the vertical scale. The parameter values valid for both rows are: $\gamma_s = 1/7$, $\gamma_a = 1/21$, $\lambda_+ = 0.302$, $\omega = 0.074$, $\mu = 0.5$. For the first row only, $\epsilon_1 = 0.395$, $\xi = 0.02$ and the social isolation measures were relaxed on day 100. For the second row only, $\epsilon_1 = 0.445$, $\xi = 0.5$ and the social isolation measures were relaxed only on day 120. The red lines appearing in the second column are the graphs of $\frac{\gamma_a}{\beta(t)(1-\xi)\mu}$, see Eq. (8). Notice that with a good approximation, when the graph of $S(t)$ is below the red line, the fraction of MSA infected individuals decreases. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Post scriptum

As this paper was being written, more data became available both for Lombardy and for São Paulo state. The paper was finished on May 21, but in all figures of Section 3 we decided to keep the data only up to Apr. 29, because the conclusions would not change too much. We report some differences in this section.

The strict lockdown in Italy ended on May 4, but social isolation measures are being slowly relaxed. If we had added the most recent data to Fig. 11, we would still have good fits of the data to the model in both cases, suggesting that the infection rate has not yet increased very much.

If in Eq. (10) we use $t = 85$ and the datum for the population fraction of deceased individuals up to that day, we obtain an estimate $R_s(85) = 0.0306$. Taking into account Eq. (9) and the fact that R_s is an increasing function, this means that although values of ξ smaller than 0.0306 are allowed by Fig. 8, these must be ruled out. The optimistic value $\xi = 0.04$ used in Fig. 11 remains as a possibility.

On May 13 appeared the first results of a serological study for a large random sample of the population in Spain [23]. Although we have not used any data of Spain in this paper, one result in the cited report is interesting to consider here. On page 12 of [23] a map shows the percentages of people having antibodies against

SARS-COV-2 in all provinces of Spain. In the provinces where the epidemic was stronger, these percentages vary between 10.9% and 14.2%. This means that the fraction of susceptible individuals varies between 0.858 and 0.891 in these provinces. Of course these susceptible fractions cannot be blindly extended to any other location, but if we extrapolate them to Lombardy, that would also rule out the optimistic value, $\xi = 0.04$, because it produces values for the susceptible fraction considerably smaller.

Regarding the epidemic in São Paulo, Table 1 shows the ratio of the maximum predicted symptomatic infected individuals to the present value (day 65) of the same quantity in three situations: the pessimistic and optimistic situations considered in Fig. 14, and an intermediate situation (graphs not shown) with $\xi = 0.12$, $\epsilon = 0.405$ and social distance measures with the present intensity up to day 120. The table also shows the predicted date of the maximum.

We see that even in an optimistic situation, an increase of 38% in the number of cases in only 16 days is expected from the results of the A-SIR model. Intensified social distancing measures might help mitigating this situation. In the other two cases, the recommendation for intensifying social distancing is of course still stronger.

Table 1
Predicted day/date for the maximum of symptomatic infected individuals in the state of São Paulo and predicted ratio of the maximum symptomatic infected individuals to the present number of symptomatic infected individuals, day 65 (May 21). We consider three cases: the optimistic and pessimistic cases of Fig. 14 and an intermediate one with $\xi = 0.12$, $\epsilon = 0.405$ and social distancing lasting up to day 120.

	Day/Date	$I_{max}/I(65)$
Optimistic	80 / June 5	1.38
Pessimistic	130 / July 25	39.8
Intermediate	100/ June 25	5.49

Declaration of competing interest

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

Acknowledgments

Funding

References

[1] R. Li, et al., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-cov2), Science (2020) http://dx.doi.org/10.1126/science.abb3221.

[2] G. Gaeta, A simple SIR model with a large set of asymptomatic infectives, Math. Eng. 3 (2) (2020) 1–39, http://dx.doi.org/10.3934/mine.2021013.

[3] D. Fanelli, 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.

[4] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (1997) 341–359.

[5] J.M. Carcione, J.E. Santos, C. Bagaini, J. Ba, A simulation of a COVID-19 epidemic based on a deterministic SEIR model, Front. Public Health 8 (2020) 230, http://dx.doi.org/10.3389/fpubh.2020.00230.

[6] C. Castilho, J.A.M. Gondim, M. Marchesin, M. Sabeti, Assessing the efficiency of different control strategies for the coronavirus (COVID-19) epidemic, 2020, e-print q-bio.PE.

[7] N. Crokidakis, COVID-19 spreading in Rio de Janeiro, Brazil: Do the policies of social isolation really work?, Chaos Solitons Fractals 136 (2020) 109930, https://doi.org/10.1016/j.chaos.2020.109930.

[8] L.H. Duczmal, et al., Vertical social distancing policy is ineffective to contain the COVID-19 pandemic, Cad. Saúde Pública 36 (5) (2020) https://doi.org/10.1590/0102-311X00084420.

[9] P. Magal, G. Webb, Predicting the number of reported and unreported cases for the COVID-19 epidemic in South Korea, Italy, France and Germany, MedRxiv preprint, https://doi.org/10.1101/2020.03.21.20040154.

[10] K. Prem, et al., The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study, Lancet 5 (5) (2020) E261–E270, https://doi.org/10.1016/S2468-2667(20)30073-6.

[11] G.S. Costa, W. Cota, S.C. Ferreira, Metapopulation modeling of COVID-19 advancing into the countryside: an analysis of mitigation strategies for Brazil, MedRxiv preprint https://doi.org/10.1101/2020.05.06.20093492.

[12] C. Gai, D. Iron, T. Kolokolnikov, Localized outbreaks in an S-I-R model with diffusion, J. Math. Biol. 80 (2020) 1389–1411, https://doi.org/10.1007/s00285-020-01466-1.

[13] W.O. Kermack, e A.G. McKendrick, Contributions to the mathematical theory of epidemics, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 138 (1932) 55–83; Proc. R. Soc. Lond. a 141 (1933) 94-122.

[14] S. Flaxman, et al., Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, Nature (2020) https://doi.org/10.1038/s41586-020-2405-7.

[15] E. Lavezzo, et al., Suppression of a SARS-CoV-2 outbreak in the Italian municipality of vo', Nature (2020) https://doi.org/10.1038/s41586-020-2488-1.

[16] D.F. Gudbjartsson, et al., Spread of SARS-CoV-2 in the Icelandic population, N. Engl. J. Med. (2020) http://dx.doi.org/10.1056/NEJMoa2006100.

[17] O. Diekmann, J.A.P. Heesterbeek, M.G. Roberts, The construction of next-generation matrices for compartmental epidemic models, J. R. Soc. Interface 7 (2010) 873–885, http://dx.doi.org/10.1098/rsif.2009.0386.

[18] A. Crisanti, In one Italian town we showed mass testing could eradicate the coronavirus, https://www.theguardian.com/commentisfree/2020/mar/20/eradicated-coronavirus-mass-testing-covid-19-italy-vo (Accessed 14 April 2020.

[19] L. Fenga, CoViD-19: An Automatic, Semiparametric Estimation Method for the Population Infected in Italy, MedRxiv preprint, https://doi.org/10.1101/2020.03.14.20036103.

[20] https://github.com/pcm-dpc/COVID-19/blob/master/dati-regioni/dpc-covid19-ita-regioni.csv (Accessed 14 April 2020).

[21] https://brasil.io/covid19/ (Accessed 14 April 2020).

[22] https://www.worldometers.info/coronavirus/#countries (Accessed 14 April 2020).

[23] Estudio Nacional de Sero-Epidemiología de la Infección por SARS-COV-2 en España, https://portalcne.isciii.es/enecovid19/documentos/ene_covid19_inf_pre.pdf (Accessed 20 May 2020).

[24] E. Pepe, P. Bajardi, L. Gauvin, F. Privitera, C. Cattuto, M. Tizzoni, COVID-19 outbreak response: first assessment of mobility changes in Italy following lockdown, Sci. Data 7 (2020) 230, https://doi.org/10.1038/s41597-020-00575-2.

[25] https://mapabrasileirodacovid.inloco.com.br/pt/?hsCtaTracking=68943485-8e65-4d6f-8ac0-af7c3ce710a2%7C45448575-c1a6-42c8-86d9-c68a42fa3fcc (Accessed 18 May 2020).