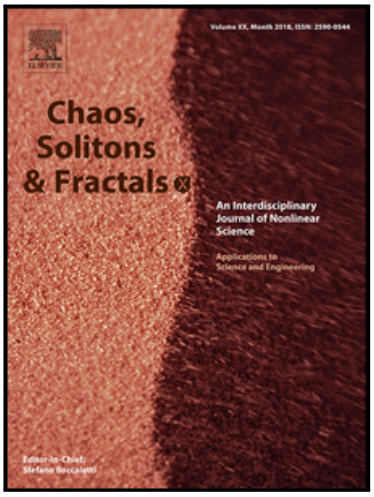
Journal Pre-proof



The first 100 days: modeling the evolution of the COVID-19 pandemic

Efthimios Kaxiras, Georgios Neofotistos, Eleni Angelaki

PII: S0960-0779(20)30511-7
DOI: https://doi.org/10.1016/j.chaos.2020.110114
Reference: CHAOS 110114

To appear in: *Chaos, Solitons and Fractals*

Received date: 3 May 2020
Revised date: 8 July 2020
Accepted date: 9 July 2020

Please cite this article as: Efthimios Kaxiras, Georgios Neofotistos, Eleni Angelaki, The first 100 days: modeling the evolution of the COVID-19 pandemic, *Chaos, Solitons and Fractals* (2020), doi: https://doi.org/10.1016/j.chaos.2020.110114

MANUSCRIPT ENTITLED:

The first 100 days: modeling the evolution of the COVID-19 pandemic

Re-Submitted for publication to CHAOS, SOLITONS and FRACTALS Journal, on June 24, 2020.

Authors:

Efthimios Kaxiras [1,2] (first author), **Georgios Neofotistos** [1,4], **Eleni Angelaki** [3,4]

Affiliation:

[1] Institute for Applied Computational Science,
Harvard J.A. Paulson School of Engineering and Applied Sciences,
Harvard University, Cambridge, MA, USA

[2] Physics Department, Harvard University, Cambridge, MA, USA

[3] Harvard J.A. Paulson School of Engineering and Applied Sciences,
Harvard University, Cambridge, MA, USA

[4] Physics Department, University of Crete, Heraklion, Greece

HIGHLIGHTS:

- A new compartmental model, forced-SIR, quantifies COVID-19 pandemic's impact
- Effectiveness of COVID-19 intervention measures is linked to model's parameters
- Application of the model to 10 countries reveals a wide range of COVID-19 impacts

The first 100 days: modeling the evolution of the COVID-19 pandemic

Efthimios Kaxiras[a,b,*], Georgios Neofotistos[a,d], Eleni Angelaki[c,d]

[a] *Institute for Applied Computational Science, Harvard J.A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA*
[b] *Department of Physics, Harvard University, Cambridge, MA, USA*
[c] *Harvard J.A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA*
[d] *Department of Physics, University of Crete, Heraklion, Greece*

Abstract

A simple analytical model for modeling the evolution of the 2020 COVID-19 pandemic is presented. The model is based on the numerical solution of the widely used Susceptible-Infectious-Removed (SIR) populations model for describing epidemics. We consider an expanded version of the original Kermack-McKendrick model, which includes a decaying value of the parameter β (the effective contact rate) due to externally imposed conditions, to which we refer as the forced-SIR (FSIR) model. We introduce an approximate analytical solution to the differential equations that represent the FSIR model which gives very reasonable fits to real data for a number of countries over a period of 100 days (from the first onset of exponential increase, in China). The proposed model contains 3 adjustable parameters which are obtained by fitting actual data (up to April 28, 2020). We analyze these results to infer the physical meaning of the parameters involved. We use the model to make predictions about the total expected number of infections in each country as well as the date when the number of infections will have reached 99% of this total. We also compare key findings of the model with recently reported results on the high contagiousness

*Corresponding author
Email addresses: kaxiras@seas.harvard.com (Efthimios Kaxiras), neofotistos@seas.harvard.edu (Georgios Neofotistos), eleni@seas.harvard.edu (Eleni Angelaki)

and rapid spread of the disease.

Keywords: COVID-19, compartmental model, modeling pandemic evolution

1. Introduction

The recent pandemic due to the COVID-19 virus has created unprecedented turmoil and changed the daily life of people over the entire planet. It has also yielded a grim toll of victims that succumb to its attack. While there is great expertise in the medical community and the community of statisticians in dealing with epidemics, less is known about this particular disease to make reliable predictions for the evolution of the current pandemic.

In studying past epidemics, scientists have systematically applied "random mixing" models which assume that an infectious individual may spread the disease to any susceptible member of the population, as originally considered by Kermack and McKendrick [1]. More recent modeling approaches considered contact networks in which the epidemic spreads only across the edges of a contact network within a population ([2] [3] [4]), Bayesian inference models [5], models of spatial contacts in real cities or countries or in large-scale artificial cities and synthetic populations ([6] [7] [8]), and computational predictions of protein structures [9], to name just a few of the modeling efforts.

In the case of COVID-19, there is considerable uncertainty in the data collected about infected individuals due to the difficulty of testing large numbers of suspected cases. Although an avalanche of research studies are currently investigating the COVID-19 epidemiological characteristics ([10] [11] [12] [13] [14]), it appears that a simple model which can capture the basic behavior of the pandemic phenomenon, in spite of the large uncertainty in the data, can possibly offer useful guidance for its near-term and longer-term evolution. This paper aims to provide such a simple model with very few adjustable parameters.

2. The model

2.1. Derivation of the model

The original mathematical description of the spread of an infectious disease in a population is the so-called SIR model, due to Kermack and McKendrick [1] which divides the (fixed) population of N individuals into three compartments (groups, classes):

- $S(t)$ the number of individuals susceptible but not yet infected with the disease;
- $I(t)$ the number of infected individuals;
- $R(t)$ the number of individuals removed (recovered) from the infected group, either by becoming healthy again with long-term immunity or by passing away.

Models that involve additional compartments can be constructed by considering more detailed stages of the infection or flows among the various compartments. The choice of the compartments is related to the disease that is being studied. For example, in the SIS model, susceptible individuals can become infected/infectious but when they are recovered/removed they can become susceptible again, that is, no permanent immunity is acquired. In the SEIR model, individuals are grouped in four compartments, one more than in the SIR model: the additional compartment is the group of "Exposed" (E) individuals, who have become infected but not yet infectious themselves. A recent model [15] considers eight stages of infection: susceptible (S), infected (I), diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H) and extinct (E), collectively termed SIDARTHE. In the present study we have chosen to focus on the SIR model, which is the simplest model incorporating the Covid-19 epidemic dynamic. We are thus effectively combining the exposed and the infectious compartments of the SEIR model into one group; $I(t)$ = infected and/or infectious individuals (referred hereafter as infected individuals). Although this combination of compartments misses some details of the disease evolution, in light of the

approximate analytical solution we introduce later, it should not substantially affect the conclusions to be drawn from the approximate model.

The SIR model involves two positive parameters, β and γ which have the following meaning:

- β describes the effective contact rate of the disease: an infected individual comes into contact with β other individuals per unit time (the fraction that are susceptible to contracting the disease is S/N);
- γ is the mean removal (recovery) rate, that is, $\frac{1}{\gamma}$ is the mean period of time during which an infected individual can pass it on before being removed from the group of the infected individuals.

This model obeys the following differential equations:

$$\frac{dI}{dt} = \beta I \frac{S}{N} - \gamma I \tag{1a}$$

$$\frac{dS}{dt} = -\beta I \frac{S}{N} \tag{1b}$$

$$\frac{dR}{dt} = \gamma I \tag{1c}$$

Many recent studies (e.g. [16], [17], [18]), have attempted to model the data of the COVID-19 pandemic by imposing time-dependence conditions (or by data assimilation) on the rates β and γ involved in the original SIR model, in order to account for the imposition of social-distancing measures, quarantine of infected individuals, and other interventions designed to slow down the spread of the disease. Motivated by such considerations, we will introduce a variation of the original model in which the parameter β is a time-dependent, *monotonically decreasing* function. This change can drastically affect the evolution of the populations. We give below a specific example to illustrate this point. Since the presence of time-dependence in β introduces a forcing term, which for reasonable parameter values lowers the number of infected individuals ("flattens the curve").

The system of equations that describe the SIR, with or without the time-dependence in the parameter β, can be easily solved numerically, as shown in

Fig. 1, giving the three group populations (S, I, R) as a function of time. Kermack and McKendrick pointed out that "it is impossible from these equations to obtain $I(t)$ as an explicit function of t" (p. 713, [1]), but provided approximations valid under certain conditions. Here we aim to give a simple approximate analytical solution inspired by the numerical solution.

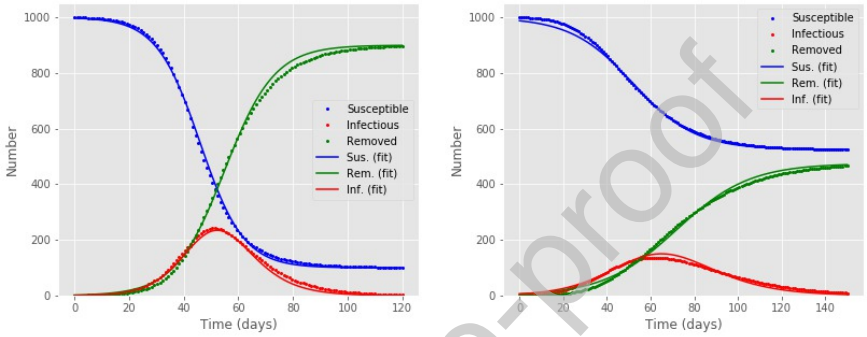


Figure 1: Numerical solution of the models, giving the susceptible $S(t)$ (blue points), infected $I(t)$ (red points) and removed $R(t)$ (green points) populations as functions of time t in days; the corresponding colored lines give the approximate solutions obtained by the analytical expressions, Eq. (3). **Left**: The SIR model, with parameter values $\beta = 0.25$ and $\gamma = 1/10$; the parameter β is constant. **Right**: The SIR model with a time-dependent parameter β with exponential decay and parameter values $\beta_0 = 0.25$, $\gamma = 1/20$, $\lambda = 50$ (see text for details).

We observe from the numerical solution shown of the SIR model, shown in Fig. 1, that both the susceptible and the removed populations (S and R, respectively) behave like sigmoids, which is the typical behavior of solutions to differential equations that involve exponential growth and decay. Moreover, the infected population is always given by the following expression

$$I(t) = N - S(t) - R(t) \tag{2}$$

From these observations, we take the approximate solutions to be given by:

$$\tilde{S} = N - \frac{N'}{1+\mathrm{e}^{-\alpha_1(t-t_1)}} \tag{3a}$$

$$\tilde{R} = \frac{N'}{1+\mathrm{e}^{-\alpha_2(t-t_2)}} \tag{3b}$$

$$\tilde{I} = N - \tilde{S}(t) - \tilde{R}(t) = \frac{N'}{1+\mathrm{e}^{-\alpha_1(t-t_1)}} - \frac{N'}{1+\mathrm{e}^{-\alpha_2(t-t_2)}} \tag{3c}$$

where N', α_1, α_2, t_1, t_2 are treated as adjustable parameters, with t_1 and t_2 representing the times at which the $\tilde{S}$ and $\tilde{R}$ populations reach their sigmoid midpoint values, respectively. Interestingly, the analytical expressions introduced above fit even better the numerical solution of the model with a time-dependent β parameter. In Fig. 1 we give examples of how well the approximate analytical expressions fit the "exact" numerical ones. In these examples, for the model with time-dependent β we assumed $\beta(t) = \beta_0 \exp(-t/\lambda)$, although we emphasize that this assumption is for illustrative purposes only and does not affect the general behavior of the model. Indeed, as we show below, $\beta(t)$ takes in our solution the behavior of a sigmoid. For the SIR model in the example of Fig. 1, the fit to the analytical expression of Eq. (3) has an RMSE value of 9.4 and the integral of the $I(t)$ values differs from the exact result by -9.5%. For the model with time-dependent β in the example we considered, the fit to the analytical model of Eq. (3) has an RMSE value of 8.3 and the integral of $I(t)$ differs from the exact result by 0.3%.

Since the analytical model of Eq. (3) can capture the behavior of the SIR model including a time-dependent β, which represent the "forcing" or "flattening" of the curve of infected individuals, we refer to it as the "FSIR" model.

2.2. Analysis of the model

Here we derive relations between the parameters used in the model of Eq. (3), and the parameters of the original set of differential equations, Eq. (1). To keep the expressions simple, we will assume $\alpha_1 = \alpha_2 = \alpha$ and define $\Delta t = t_2 - t_1$. By inserting the expressions for $\tilde{S}$ and $\tilde{I}$ in Eq. (1b) we find:

$$\beta(t) = \frac{\alpha}{\mathrm{e}^{\alpha\Delta t}-1}\left(\frac{1+\mathrm{e}^{-\alpha(t-t_1)}\mathrm{e}^{\alpha\Delta t}}{1-n'+\mathrm{e}^{-\alpha(t-t_1)}}\right) \tag{4}$$

where we have defined $n' = N'/N$. Similarly, by inserting the expressions for $\tilde{R}$ and $\tilde{I}$ in Eq. (1c) we obtain:

$$\gamma(t) = \frac{\alpha}{1 - \mathrm{e}^{-\alpha\Delta t}} \left(\frac{1 + \mathrm{e}^{-\alpha(t-t_1)}}{1 + \mathrm{e}^{-\alpha(t-t_1)}\mathrm{e}^{\alpha\Delta t}} \right) \tag{5}$$

Thus, in the approximate model described by Eq. (3), the parameters β and γ of the original SIR model become time-dependent, if we treat α as constant to be determined by fitting the data (see next section). In the FSIR model the effect of interventions and measures can be inferred from the values of the adjustable parameters t_1, Δt and N', as will be explained in the next section, so that there is no need to impose specific time-dependent conditions on the model parameters themselves. It should be noted that the time-dependence of the coefficients $\beta(t)$ and $\gamma(t)$ is an inherent property of the proposed analytical model described by Eq.s (3); $\beta(t)$, as a monotonically decreasing function of time, can relate to the population's decreased mobility, lock-downs and other non-pharmaceutical interventions, which decrease the contact rate among susceptible individuals ([19], [20]).

The quantity $n' = N'/N$ we defined in the expression of $\beta(t)$ is the fraction of the original susceptible population that was infected, and thus does become part of the removed population. There are two possible limiting values for this quantity: $n' \to 1$, the limit in which the entire susceptible population was exposed and eventually becomes removed population, and $n' \to 0$, the limit in which only a tiny fraction of the susceptible population was exposed. In the first limit we obtain:

$$n' \to 1 \Rightarrow \beta_1(t) = \frac{\alpha}{\mathrm{e}^{\alpha\Delta t} - 1} \left(\mathrm{e}^{\alpha(t-t_1)} + \mathrm{e}^{\alpha\Delta t} \right) \tag{6}$$

while in the second limit we obtain:

$$n' \to 0 \Rightarrow \beta_2(t) = \frac{\alpha}{\mathrm{e}^{\alpha\Delta t} - 1} \left(\frac{1 + \mathrm{e}^{-\alpha(t-t_1)}\mathrm{e}^{\alpha\Delta t}}{1 + \mathrm{e}^{-\alpha(t-t_1)}} \right) \tag{7}$$

From the first expression we see that for $t \gg t_1$ the value of $\beta_1(t)$ increases exponentially, which is an unphysical result. From the second expression, we see that β is a *monotonically decreasing* function of time and for $t \gg t_1$ tends

to the constant value

$$\lim_{t \gg t_1} \beta_2(t) = \alpha/(\mathrm{e}^{\alpha \Delta t} - 1),$$

which is the expected behavior in the FSIR model.

For $t \ll t_1$ and assuming that $\alpha \Delta t > 1$ we find that

$$\beta_2 \approx \alpha \frac{\mathrm{e}^{\alpha \Delta t}}{\mathrm{e}^{\alpha \Delta t} - 1} \approx \alpha,$$

which relates the adjustable parameter α of the analytical model to the value of the parameter β appearing in the original SIR model.

The quantity $R_0 = \beta/\gamma$ of the SIR model is used to estimate the value of the basic reproduction number of an epidemic. From our analytical model, in the limit $n' \to 0$, the quantity β/γ takes the form

$$\beta/\gamma = \mathrm{e}^{-\alpha \Delta t} \left(\frac{1 + \mathrm{e}^{-\alpha(t - t_1)} \mathrm{e}^{\alpha \Delta t}}{1 + \mathrm{e}^{-\alpha(t - t_1)}} \right)^2 \tag{8}$$

For $t = 0$, and assuming that $\alpha t_1 \gg 1$ (as is the case for the fits to reported data discussed in the next section), this quantity becomes

$$t = 0: \quad \beta/\gamma = \mathrm{e}^{-\alpha \Delta t} \left(\frac{1 + \mathrm{e}^{\alpha t_1} \mathrm{e}^{\alpha \Delta t}}{1 + \mathrm{e}^{\alpha t_1}} \right)^2 \approx \mathrm{e}^{\alpha \Delta t}.$$

For $t \gg t_2$, the quantity β/γ becomes

$$t \gg t_2: \quad \beta/\gamma \approx \mathrm{e}^{-\alpha \Delta t}.$$

The first number is very large for typical values of the parameters in the FSIR model obtained from fits to reported data, while the second value is very small, close to zero. Neither result is realistic. In the important range $t_1 < t < t_2$, we find from numerical results that this quantity is approximately described by a decaying exponential in time

$$t_1 < t < t_2: \quad \beta/\gamma \sim \mathrm{e}^{-t/\lambda},$$

with $\lambda \approx 1/2\alpha$. This result implies that in this range we would expect $\beta \sim \mathrm{e}^{-\alpha t}$ (the functional form we assumed for illustrative purposes in Fig. 1), and

$\gamma \sim \mathrm{e}^{\alpha t}$. From this last expression, taking the time average of γ in the range $t_1 < t < t_2 = t_1 + \Delta t$, which we call $\overline{\gamma}$, we find

$$\overline{\gamma} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \gamma_2 \mathrm{e}^{\alpha(t-t_2)} dt = \frac{\gamma_2}{\alpha \Delta t} \left[1 - \mathrm{e}^{-\alpha \Delta t}\right] \approx \frac{1}{\Delta t}$$

where we have used

$$\gamma_2 \approx \frac{\alpha}{1 - \mathrm{e}^{-\alpha \Delta t}}$$

from the expression of Eq. (5), a reasonable approximation for $t \gtrsim t_2$. The last relation for the average value of γ is obeyed to a good approximation for each case of reported data we examined.

Using the preceding analysis that led to the relations $\beta \approx \alpha$ for the initial value of β and $\overline{\gamma} \sim 1/\Delta t$ for the average value of γ, we suggest that a reasonable representation of the quantity β/γ is given by the value of $\alpha \Delta t$. Thus, we will use this value as a proxy for R_0, and will refer to it as $\overline{R}_0$. The parameters estimated from the fit of our analytical model to reported data give a value of $\overline{R}_0$ which is in agreement with the recently reported median value of R_0 for the pandemic.

3. Application to reported data

We use our analytical FSIR model to fit the behavior of infected populations of different countries, as obtained from the European Centre for Disease Prevention and Control (ECDC) [21], for a period ending on April 28, 2020 which corresponds to 100 days from the onset of the exponential growth of reported cases in China.

In order to obtain a meaningful fit, we had to consider data for each country that show a monotonic increase at the beginning. This means that a few data points in each case were excluded, as they corresponded to sporadic reports of very few isolated cases, typically 1 to 10 in a given day, interspersed by several days of zero cases. In practice this means that the fitting begins at a certain cutoff day denoted as t_0. In order to make the fit more robust and simpler, we chose $\alpha_1 = \alpha_2 = \alpha$. Moreover, we found by trial-and-error that the value

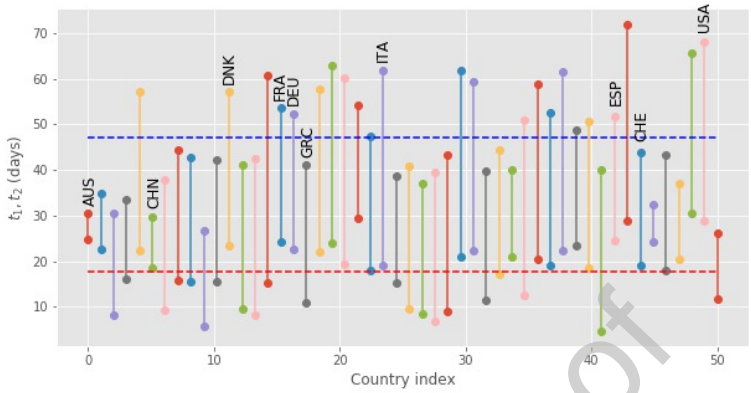


Figure 2: The values of the parameters t_1 and $t_2 = t_1 + \Delta t$ for 50 countries, obtained by fitting the FSIR model with data up to 28 April 2020. The dashed lines give the average values of the parameters, $\langle t_1 \rangle$ (red) and $\langle t_2 \rangle = \langle t_1 \rangle + \langle \Delta t \rangle$ (blue) of the FSIR model defined by Eq. (3). The countries with labels are used to examine the behavior of the model in more detail.

$\alpha = 0.25$ is the optimal choice for all the countries we considered. This leaves three adjustable parameters in the model that can be varied to obtain the best fit to the data, namely t_1, t_2 and N'; instead of t_1 and t_2, we elected to use instead as independent parameters t_1 and $\Delta t = t_2 - t_1$. The best fit here is defined in the Root-Mean-Square (RMS) sense.

We were able to obtain reasonable fits for 50 countries from the entire database [21]. The resulting values for the parameters t_1, t_2, are shown in Fig. 2 ($\Delta t = t_2 - t_1$ is the distance between each pair of values). The averages and standard deviations for this set are $\langle t_1 \rangle = 17.81 \pm 6.58$, $\langle \Delta t \rangle = 29.20 \pm 9.16$, giving $\langle t_2 \rangle = \langle t_1 \rangle + \langle \Delta t \rangle = 47.19$. The values of the parameters involved span a wide range. For other countries in the database, the data are either too noisy or have not reached the point where the FSIR model can provide a good fit: specifically, the model needs to include data up to the maximum of the curve, otherwise it does not give meaningful values for the fitting parameters.

Instead of including all 50 countries in the following discussion, we have chosen to focus on 10 countries that span the whole range of parameter values

and could hopefully provide some insight to the behavior of the pandemic. The choice of the 10 countries also aims to represent parts of the world more heavily or less heavily impacted by the disease, as well as more typical cases. Here we defined the impact as the total number N_T of infected individuals during the first wave of the pandemic, as predicted by the FSIR model; this number is scaled by the population of the country, N_P, in Fig. 3. In particular, we have included 3 countries in which the impact was small, China, Greece and Australia for which $(N_T/N_P) < 500$ infected per million, 3 countries in which the impact was moderate, Denmark, Germany and France for which $1,000 < (N_T/N_P) < 2,000$ infected per million, and four countries where the impact was large, Switzerland, Italy, USA and Spain for which $(N_T/N_P) > 3,000$ infected per million. The average value for t_1 for this set of 10 countries is $\langle t_1 \rangle = 20.04$ and for Δt it is $\langle \Delta t \rangle = 27.31$.

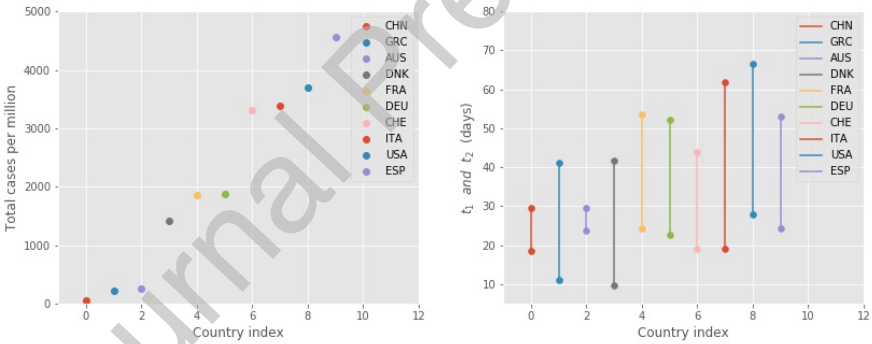


Figure 3: **Left**: the estimated total cases (N_T) scaled by the population of each country. **Right**: t_1 and t_2 values (dots) in 10 countries, as obtained in the FSIR model by fitting the raw data reported in [21], including data up to April 28, 2020 (see also Fig. 4 for specific examples).

In Fig. 4 we give some examples of the actual fits for the "outlier" countries (China, Greece, USA and Spain). To have a measure of the fit that is comparable between different countries, we defined the "quality of fit" as:

$$Q_{\text{fit}} = \frac{1}{N'}\text{RMSE} \tag{9}$$

Index	Country	(Symbol)	t_0 (days)	t_1 (days)	Δt (days)	N'	$Q_{\rm fit}$ (%)	$N_{\rm T}$	D (days)
0	China	(CHN)	17	18.5	11.1	7,343	16.19	81,100	−59
1	Greece	(GRC)	65	11.0	30.1	79	36.91	2,400	−4
2	Australia	(AUS)	61	23.8	5.7	1,143	6.61	6,500	−14
3	Denmark	(DNK)	75	9.63	32.0	258	21.30	8,200	7
4	France	(FRA)	57	24.4	29.2	4,272	20.97	124,800	1
5	Germany	(DEU)	57	22.6	29.7	5,246	17.22	156,000	0
6	Switzerland	(CHE)	59	19.1	24.7	1,142	14.62	28,100	−6
7	Italy	(ITA)	53	19.0	42.8	4,774	15.13	204,000	3
8	United States	(USA)	59	27.9	38.7	31,314	12.58	1,210,300	14
9	Spain	(ESP)	56	24.3	28.7	7,417	14.42	212,800	0

Table 1: The values of the various parameters that enter in the FSIR model of Eq. (3), for the 10 countries considered (see text for details). The countries have been indexed according to their N_T values, scaled by the population of each country. The next-to-last column includes the values for the *expected* total number of cases N_T when the number of infections has dropped to near zero, and is an *extrapolated* value (rounded to the nearest 100). The last column includes the number of days D (counting from April 28) until the value of N_T has reached 99% of its final value; a negative number (for China, Australia, Switzerland, and Greece) indicates that this date has already passed.

which is expressed as a percentage (multiplied by a factor of 100). The resulting values of the parameters, including our choices of t_0, are given in Table 1.

The values of the parameters obtained reveal interesting behavior.

- t_0: The value of this parameter is similar for all countries, except for China with $t_0 = 17$. This simply reflects the fact that the pandemic originated in China and then spread through the rest of the world. The rest of the countries have starting dates of the exponential increase within one week from the earliest, Italy (with $t_0 = 53$) to the latest, such as Greece and Denmark (with $t_0 = 65$, and 75, respectively). The time lag between most countries and China is approximately 6 weeks.

- t_1: This value indicates the position of the mid-point of the sigmoid representing the behavior of the susceptible population, S. The shorter it is, the sooner the country experiences the exponential increase in the infected cases, thus urgently necessitating the introduction of health interventions and measures to limit the spread of the disease. The three countries with

the shortest t_1 values are Greece, Denmark and China; unsurprisingly, these countries also have of the lowest number of cases per million, as shown in Fig. 3.

- Δt: This value indicates the lag between the sigmoid that describes the recovered population ($\tilde{R}$) and the sigmoid of the susceptible population ($\tilde{S}$). As such, it can be interpreted as the effective rate of removal (γ in the SIR model). In Table 1 we present the values of Δt for each country. The average of Δt is close to 27.5 days ($\sim$ 4 weeks), a value consistent with a recently reported estimated median time of approximately 2 weeks from onset to clinical recovery for mild cases, and 3–6 weeks for patients with severe or critical disease ([22], [23], [24]). Australia and China show an unusual low value, Δt= 5.7, and 11.1 days, respectively. The value of this parameter has a significant effect on the total expected number of cases (see below).

- N': the value of this parameter is representative of the number of daily cases near the peak of the $\tilde{I}$ curve. It is close to reported values for this quantity for all the countries. Interestingly, if we assume that the total number of susceptible individuals is close to the population of each country, which in all cases is in the range of $N \sim 10^7 - 10^9$, then the ratio $n' = N'/N \to 0$, as we assumed for the FSIR model earlier.

In Table 1 we also include the values for the quality of the fit, which range from 6.6 (AUS) and 12.6 (USA) to 36.8 (Greece), representing a measure of the relative noise in the data; the noise is largest for Greece because the numbers are rather small. We have also considered fitting the FSIR model to seven-day running averages of the reported cases, and this in general makes almost no difference to the value of the parameters or the quality of the fit (see Fig. 4 for examples).

Using the analytic expression for $\tilde{I}(t)$ we can extrapolate to long times and try to obtain an estimate for the total value of cases over a long period, when the

number of daily cases of infection have essentially dropped to negligible levels (this corresponds to $\tilde{I}(t) \approx 0$). We call this asymptotic value N_T and report it in Table 1.

The average (over all countries in the set) of Δt is 27.31 ± 11.35 (median of Δt is 29.48), the average of $\overline{\gamma}$ is 0.058 ± 0.042, and the average of $1/\Delta t$ is 0.052 ± 0.047, (all numbers reported to 3 significant digits). It should be noted that the average of Δt over the set of the 50 countries mentioned earlier is 29.20 ± 9.16.

The FSIR-estimated average of $\Delta t = 27.31 \pm 11.35$ yields an average reproduction number of $\overline{R}_0 \approx 6.83 \pm 2.84$ (7.37, if we consider the median value of Δt). Initial estimates of the early dynamics in Wuhan, China, suggested a value for R_0 in the range 2.2–2.7. For China, the FSIR model estimates $\overline{R}_0 = 2.77$. However, the FSIR estimates for the rest of the countries in the set, suggest much higher values of $\overline{R}_0$. By conducting an elaborate analysis of datasets and data sources, estimating distributions of epidemiological parameters, and integrating uncertainties in parameters values, Sanche *et al.*[10] reported a median $R_0 = 5.7$ (95% CI 3.8 – 8.9) for China. The FSIR estimated values of $\overline{R}_0$ for the countries we study fall almost entirely within this range.

The estimated $\overline{R}_0$ values correspond to the model's fitted parameters for actual data. Since $\beta = \alpha = 0.25$, a fixed value as explained above, the adjustable fitted parameter Δt is the one essentially yielding the value of $\overline{R}_0$ through the relation $\overline{R}_0 = \beta/\overline{\gamma} \approx \alpha \Delta t$ (with $\Delta t \approx \overline{\gamma}$). Large values of R_0 have been reported in the literature ([10], [25]). However, fitting real country data tends to produce higher values of Δt because a seemingly single epidemic wave in a country can be the aggregate result of the superposition of smaller or larger sub-epidemics [26].

Fig. 5 depicts the FSIR-obtained values of $\overline{\gamma}$, plotted in conjuction with the $1/\Delta t$ values, for the 10 countries considered. For each country, the values are very close in magnitude, as was explained in a previous section. It should be noted that $\overline{\gamma}$ is calculated as the time average of the coefficient $\gamma(t)$, as presented in Eq. (5), over the time period starting from t_0 until the expected total number

of infected people, N_T, has reached 99% of its final value. Fig. 5 also depicts the $\overline{R}_0$ values for each country (calculated as $\overline{R}_0 \approx \alpha \Delta t$, where $\alpha \approx 0.25$ and the value of each country's Δt is presented in Table 1).

As measured by the estimate of the basic reproduction number $\overline{R}_0$, Italy is the country most adversely affected by the disease ($\overline{R}_0 \approx 10.69$), followed by the USA ($\overline{R}_0 \approx 9.78$), Denmark ($\overline{R}_0 \approx 8.0$), Greece ($\overline{R}_0 \approx 7.53$), Germany ($\overline{R}_0 \approx 7.44$), France ($\overline{R}_0 \approx 7.30$), and Spain ($\overline{R}_0 \approx 7.17$). Greece, although it suffered relatively small number of cases, has a large value for the basic reproduction number ($\overline{R}_0 \approx 7.53$). USA, Spain, and Italy suffer the highest numbers of expected total cases (N_T), whereas in the case of Greece the expected total number of cases is one of the lowest in the set, presumably due to the fast implementation of measures imposed by the government and followed by the citizens. It should be noted that Greece has one of the smaller t_1 values. Australia, with $\overline{R}_0 \approx 1.42$, has the lowest value of $\overline{R}_0$.

4. Assessment and conclusions

The model presented here offers several advantages, but also has certain limitations. As a model of the simple SIR type, it restricts the compartmentalization of the population in only 3 classes (S, I, R). It is a simple deterministic model, which does not take into consideration age, gender, spatial position or any other factors. It assumes homogeneous mixing, that is, individuals make contact at random, the transmission and recovery rates are positive and the same for all individuals, there is no vaccine available, the total population size is constant and large, and any recovered person obtains permanent immunity.

Its main strength is the simplicity and the insight it offers to how the pandemic affects large populations such as countries, states, or cities. The model relies on only three parameters all of which are obtained by directly fitting the reported data of daily populations of infected individuals. The eect of interventions and measures can be inferred from the values of the adjustable parameters t_1, Δt, and N', so there is no need to impose specic time-dependent conditions on the model parameters themselves (it should be noted that the time it takes

for infectious individuals to recover is independent of non-pharmaceutical interventions). As such, it is particular useful in obtaining adjustable parameters at the beginning of the an epidemic, but not after different non-pharmaceutical interventions are applied, or if there is a second peak or multiple waves (sub-epidemics) in the number of infectious individuals at later times.

An additional advantage of the model is that the values of the parameters derived from the data for different countries can be used to estimate the quantity $\overline{R}_0$, a crucial measure of how each country has been affected. This is again in reasonable agreement with the picture described qualitatively in news reports about each country, and offers a quantitative assessment of the severity of the situation. For instance, as Fig. 5 shows, low values of the parameters $\overline{\gamma}$ and $1/\Delta t$, which are directly obtained from the model, are indicative of high $\overline{R}_0$ values, as in the cases of Italy, USA and Spain, all of which have suffered relatively large effects. In contrast, high values of these parameters are indicative of relatively mild effects, as is the case for China and Australia. Greece seems to be an exception, having average values for these parameters (similar to several other countries), yet suffered relatively mild effects as judged by the total number of infected and the fatalities. In this respect, our model can also offer useful insight: it is not just the value of $\overline{R}_0$ that matters, but also how soon measures are adopted, as described by the parameter t_1, and how strictly they are enforced, as described by the parameter Δt. Thus, the model can provide some guidance on the relative merits of different approaches.

Finally, the projections of the total amount of infected individuals for the first wave, as predicted by extending the time evolution of the model to the future, are quite sensible. In this sense, the model offers useful estimates of when the total number of infections of the first wave will be reached in each country.

Nevertheless, the model has important limitations.

First, it contains no information on, and therefore can make no predictions about, the case mortality rate. The number of fatalities (case fatality) is roughly proportional to the total number of infected cases, although the constant of

proportionality varies in each country, ranging from a high of about 0.15 for Belgium, 0.14 for France, 0.13 for Italy, to a low of about 0.01 for Australia, and about 0.05 for Greece, China, and Denmark [27]. Fig. 6 presents the case fatality ratio and the deaths (COVID-related deaths) per 100K of the population, for each country. The case fatality ratio represents the mortality per absolute number of cases, that is, the total confirmed cases within a country. Greece, Denmark, and China have low values of case fatality ratios and deaths per 100K of the population, and so have Germany and the USA. Australia has the lowest ratio. In the other end, France, Italy, and Spain have the highest ratios. Additional information is ndeed to explain these results, possibly having to do with the capabilities of the health care system in each conutry, which is not related simply to the dynamics of the disease spread in the population; such factors are not contained in the mathematical model considered here.

Second, the extrapolation to future cases of infection is only a *lower limit*. This point has been discussed in an elegant mathematical analysis of the data by Fokas *et al.* [28]. We emphasize here that this shortcoming is *not* traced to the analytical FSIR model, but rather to the underlying differential equations assumed to describe the phenomenon (the original SIR model). The reason is that any set of linear differential equations that describe exponential decay and growth, as those of Eq.'s (1) do, will necessarily have exponential (logistic) behavior of the tails; this is captured by our analytical model, but apparently does not represent the actual data for the long-term behavior of the total number of infected individuals. This is evident in Fig. 4: in the countries that have long passed the peak of the reported cases, the tail does not asymptote to a constant value, as the sigmoid (logistic) model predicts, but the number actually keeps growing at a slow rate. To capture this behavior, non-linear terms are needed in the underlying differential equations (see [28]).

We suggest that, despite its limitations, the simple analytical model presented here can be useful for a quantitative evaluation of different efforts to contain the pandemic.

Acknowledgement: This work grew out of a lecture and a lab exercise on numerical solutions of ordinary differential equations, presented on April 10 2020 in the course "APMTH-10: Computing for Science and Engineering", which the three authors teach. We gratefully acknowledge support from the Computing in Engineering Education (CEE) group of the Active Learning Labs in developing and offering this course.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions: E.K. conceived and developed the proposed model, and supervised the study. E.A. designed the code. G.N., E.A., and E.K. performed the fitting experiments and analyzed the results. E.K. wrote the initial draft of the manuscript. All authors critically revised, improved, and reviewed the manuscript in various ways, and gave final approval for publication.

[1] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics", Proc. Roy. Soc. A, **115**, 772 (1927)

[2] M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, "Dynamical patterns of epidemic outbreaks in complex heterogeneous networks", J Theor Biol. 2005;235(2):275–288. doi:10.1016/j.jtbi.2005.01.011

[3] M. J. Ferrari, S. Bansal, L. A. Meyers, O. N. Bjørnstad, "Network frailty and the geometry of herd immunity", Proc Biol Sci. 2006;273(1602):2743–2748. doi:10.1098/rspb.2006.3636

[4] E. Volz, "SIR dynamics in random networks with heterogeneous connectivity", J Math Biol. 2008 Mar;56(3):293-310. doi: 10.1007/s00285-007-0116-4

[5] C. Groendyke, D. Welch, and D. R. Hunter, "Bayesian Inference for Contact Networks Given Epidemic Data", Scandinavian Journal of Statistics, Vol. 38, No. 3 (September 2011), pp. 600-616

[6] E. Tagliazucchi, P. Balenzuela, M. Travizano, G.B. Mindlin, P.D. Mininni. "Lessons from being challenged by COVID-19". Chaos Solitons Fractals. 2020;137:109923. doi:10.1016/j.chaos.2020.109923

[7] Q.H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, A. Vespignani. "Measurability of the epidemic reproduction number in data-driven contact networks". Proceedings of the National Academy of Sciences 2018: 115 (50) 12680-12685. DOI: 10.1073/pnas.1811115115

[8] M. Zhang, A. Verbraeck, R. Meng, B. Chen, and X. Qiu, (2016) "Modeling Spatial Contacts for Epidemic Prediction in a Large-Scale Artificial City" Journal of Artificial Societies and Social Simulation 19 (4) 3 (downloaded from http://jasss.soc.surrey.ac.uk/19/4/3.html on 4/9/2020). doi: 10.18564/jasss.3148-616

[9] J. Jumper, K. Tunyasuvunakool, P. Kohli, D. Hassabis, and the AlphaFold Team, "Computational predictions of protein structures associated with COVID-19", Version 2, DeepMind website, 8

April 2020, https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19

[10] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, R. Ke, "High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2" . Emerg Infect Dis. 2020 Jul (accessed on April 15, 2020), https://doi.org/10.3201/eid2607.200282. DOI: 10.3201/eid2607.200282

[11] Q. Li , X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, et al., "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia", N Engl J Med. 2020;382:1199–207

[12] N. Imai, I. Dorigatti, A. Cori, S. Riley, N. M. Ferguson, "Estimating the potential total number of novel coronavirus cases in Wuhan City, China" [cited 2020 Feb 2]. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/2019-nCoV-outbreak-report-17-01-2020

[13] C. Rothe, M. Schunk, P. Sothmann, G. Bretzel, G. Froeschl, C. Wallrauch, et al., "Transmission of 2019-nCoV infection from an asymptomatic contact in Germany", N Engl J Med. 2020;382:970–1

[14] L. Wynants, B. Van Calster Ben, M. Bonten, G. Collins, T. Debray, M. De Vos et al., "Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal" BMJ 2020; 369 :m1328

[15] G. Giordano, F. Blanchini, R. Bruno, et al. "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy". Nat Med 2020: 26, 855–860. https://doi.org/10.1038/s41591-020-0883-7

[16] G. Barmparis, G. Tsironis, "Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach", Chaos, Solitons & Fractals, 2020, 109842, ISSN 0960-0779, https://doi.org/10.1016/j.chaos.2020.109842.

[17] E. Franco, "A feedback SIR (fSIR) model highlights advantages and limitations of infection-based social distancing", arXiv:2004.13216 [q-bio.PE], downloaded on April 29, 2020.

[18] P. Nadler, S. Wang, R. Arcucci, X. Yang, and Y. Guo "An Epidemiological Modelling Approach for Covid19 via Data Assimilation", arXiv:2004.12130 [stat.AP], downloaded on April 29, 2020.

[19] 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

[20] J. Dehning, J. Zierenberg, F.P. Spitzner, M. Wibral, J. Pinheiro Neto, M. Wilczek, V. Priesemann. "Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions". Science 2020. DOI: 10.1126/science.abb9789.

[21] https://opendata.ecdc.europa.eu/covid19/casedistribution/csv

[22] Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) 16-24 February 2020 https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

[23] F. Zhou , T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, et al. "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study", Lancet. 2020;395:1054–62

[24] R. Wölfel, V.M. Corman, W. Guggemos, et al. "Virological assessment of hospitalized patients with COVID-2019". Nature 581, 465–469 (2020). https://doi.org/10.1038/s41586-020-2196-x

[25] Y. Liu, A. Gayle, A. Wilder-Smith, J. Rocklöv. "The reproductive number of COVID-19 is higher compared to SARS coronavirus". Journal of Travel Medicine, Volume 27, Issue 2, March 2020, taaa021, https://doi.org/10.1093/jtm/taaa021.

[26] G. Chowell, A. Tariq, J.M. Hyman. "A novel sub-epidemic modeling framework for short-term forecasting epidemic waves". BMC Med 2019: 17(164). https://doi.org/10.1186/s12916-0191406-6

[27] https://coronavirus.jhu.edu/data/mortality

[28] A. Fokas, N. Dikaios, and G. Kastis. "Predictive mathematical models for the number of individuals infected with COVID-19". medRxiv doi: https://doi.org/10.1101/2020.05.02.20088591

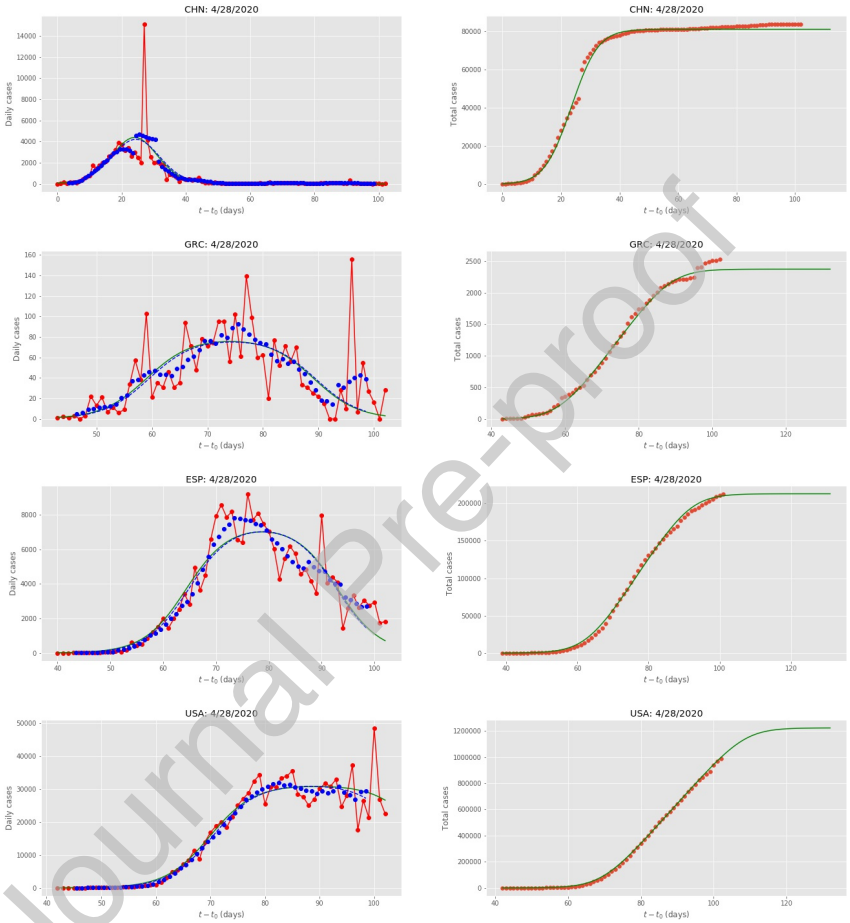


Figure 4: Data for China (CHN), Greece (GRC), USA and Spain (ESP) and the United States of America (USA). Red dots are the daily data reported in Ref. [21], the green lines are the fits by the FSIR model. The blue dots are seven-day running averages and the blue dashed lines are the fits by the FSIR model; in all cases the green and blue-dashed lines are essentially indistinguishable, except for USA near the end of the examined period.

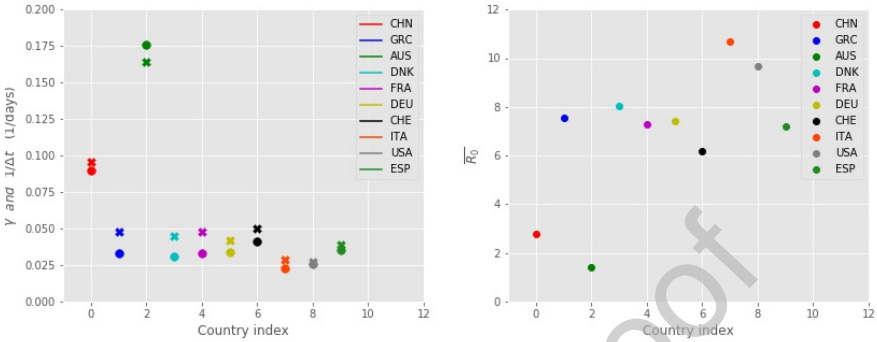


Figure 5: **Left**: Estimates of $\overline{\gamma}$ (filled x symbols) and $1/\Delta t$ (filled circles) for the 10 countries considered in Table 1; these values are useful in estimating the reproduction number $\overline{R}_0$. **Right**: Estimated values for the reproduction number $\overline{R}_0$ for the 10 countries considered, as obtained from the FSIR model. The larger the value $\overline{R}_0$, the more adversely affected by the disease the country is.

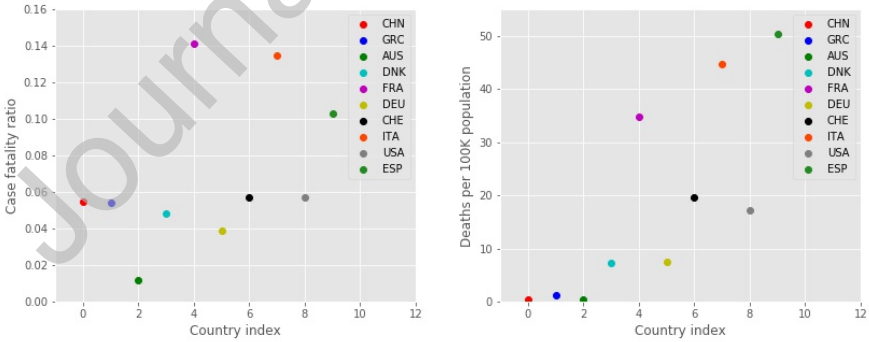


Figure 6: **Left**: Case fatality ratios as reported in [27] for 10 countries. **Right**: Deaths per 100K inhabitants for each country; these values closely follow the trends of the expected total number of infections scaled by the population, see Fig. 3.

MANUSCRIPT ENTITLED:

The first 100 days: modeling the evolution of the COVID-19 pandemic

Re-Submitted for publication to CHAOS, SOLITONS and FRACTALS Journal, on June 24, 2020.

Authors:

Efthimios Kaxiras [1,2] (first author), **Georgios Neofotistos** [1,4], **Eleni Angelaki** [3,4]

Affiliation:

[1] Institute for Applied Computational Science,
Harvard J.A. Paulson School of Engineering and Applied Sciences,
Harvard University, Cambridge, MA, USA

[2] Physics Department, Harvard University, Cambridge, MA, USA

[3] Harvard J.A. Paulson School of Engineering and Applied Sciences,
Harvard University, Cambridge, MA, USA

[4] Physics Department, University of Crete, Heraklion, Greece

Declaration of interests

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: