

# What Can We Estimate from Fatality and Infectious Case Data? A case Study of Covid-19 Pandemic

Semra Ahmetolan<sup>1</sup>, Ayse Humeyra Bilge<sup>2</sup>, Ali Demirci<sup>1</sup>, Ayse Peker-Dobie<sup>1,2\*</sup>, Onder Ergonul<sup>3</sup>

<sup>1</sup>Department of Mathematics, Faculty of Science and Letters, Istanbul Technical University, 34469, Istanbul, Turkey

<sup>2</sup> Department of Industrial Engineering, Faculty of Engineering and Natural Sciences, Kadir Has University, 34083, Istanbul, Turkey

<sup>3</sup>Infectious Diseases and Clinical Microbiology Department, School of Medicine, Koc University, Istanbul, Turkey

## \* Correspondence:

Ali Demirci  
demircial@itu.edu.tr

**Keywords:** COVID-19, Coronavirus, SIR Model, Parameter Estimation

## Abstract

Daily case reports and daily fatalities for China, South Korea, France, Germany, Italy, Spain, Iran, Turkey, the United Kingdom and the United States over the period January 22, 2020 - April 20, 2020 are analysed using the Susceptible-Infected-Removed (SIR) model. For each country, the Susceptible-Infected-Removed (SIR) models fitting *cumulative infective case data* within 5% error are analysed. It is shown that the quantity that can be the most robustly estimated from the normalized data, is the timing of the maximum and timings of the inflection points of the proportion of infected individuals.

## 1 INTRODUCTION

The Coronavirus disease 2019 (Covid-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a highly contagious disease affecting huge numbers of people all over the world. The earliest case was identified in China in December 2019. After the first diagnosis, the disease spread very quickly to other countries, in spite of efforts to slow and stop the spread of COVID-19, such as self-isolation, quarantine, social distancing, contact tracing, and travel limitations. As a result of its rapid spread and very high infection rates, the World Health Organization (WHO) declared Covid-19 a pandemic in March 2020 [1].

Even though the pandemic has passed its early stage and there are 90% fewer cases in China as a consequence of successful containment measures, the disease is rapidly expanding in Europe, America, Asia, Middle East and Africa. Despite the application of travel restrictions by many

countries, there have been no substantial delays in the arrival of the pandemic in non-affected areas. This fact was also observed for the H1N1 epidemic in 2009 [2]. Increased globalization makes infectious diseases everyone's problem, as experienced in this pandemic: the massively increased demand for intensive care treatment of Covid-19 patients has caused severe disruption of health care systems throughout the world.

In most countries the infection is in its early phase in terms of the duration of the infection. A great deal of effort has been invested in the estimation of epidemic parameters of Covid-19 in the early stage for China and some other countries [3], [4], [5], [6], [7], [8], [9], [10]. In [3] the authors analysed the temporal dynamics of the disease in China, Italy and France in the period between 22nd of January and 15<sup>th</sup> of March 2020. In [4], the potential for sustained human-to-human transmission to occur in locations outside Wuhan is assessed based on the estimations of how transmission in Wuhan varied between December, 2019, and February, 2020. The difficulty related to the accurate predictions of the pandemic is discussed in [5]. In [6], the authors used phenomenological models valid during previous outbreaks to generate and assess short-term forecasts of the cumulative number of confirmed reported cases in Hubei province and for the overall trajectory in China [7]. Epidemic analysis of the disease in Italy is presented in [8] by means of dynamical modelling [9]. Forecasting Covid-19 is investigated in [10]. In addition, the change in the epidemic behaviour of various countries can be traced by the use of data driven systems [11].

One of the common features of these works is the existence of variations in these parameter estimations. In the present work, we show that the number of reported cases provides the most accurate representation of the number of removed individuals and that the quantity that can be most robustly estimated from normalized data, is the timing of the maximum and timings of the inflection points of the proportion of infected individuals. These values correspond to the peak of the epidemic and to the highest rates of increase and the highest rates of decrease in the number of infected individuals. The stability of the estimations is discussed by comparing predictions based on data with long time spans.

## **2 DATA AND METHODS**

Publicly accessible data that have been released by the state offices of each country are used for the analysis. The data set of each country is collected according to published official reports and available at the website <http://www.worldometers.info/coronavirus/> (last access: 27 April 2020). Updated data are also available at the website <http://epikhas.khas.edu.tr/>. The last data in this work was collected on the 18th of April, 2020. Data covers the period January 22-April 18, 2020 and in the following, “Day 1” corresponds to January 22, 2020. The Susceptible-Infected-Removed (SIR) model [12] is used for analysis. It should be noted that Chinese state-provided data up to April 16, 2020 is used in the analysis. This data was provided prior to the revisions made after this date.

## 2.1 SIR and SEIR models

The Susceptible-Infected-Removed (SIR) model is a system of ordinary differential equations modelling the spread of epidemics in a closed population, under the assumption of permanent immunity and homogeneous mixing [12]. These equations are

$$S' = -\beta S I, \quad I' = \beta S I - \eta I, \quad R' = \eta I \quad (1)$$

If the disease has an incubation period, then the Susceptible-Exposed-Infected-Removed (SEIR) model governing its spread is

$$S' = -\beta S I, \quad E' = \beta S I - \varepsilon E, \quad I' = \varepsilon E - \eta I, \quad R' = \eta I. \quad (2)$$

where  $\beta$  is the contact rate,  $T = 1/\eta$  is the mean infectious period and  $1/\varepsilon$  is the incubation period.

Since the Covid-19 infection has an incubation period, the right model to use is the SEIR system. But, in previous work [14] it was shown that the parameters of the SEIR model cannot be determined from the time evolution of the normalized curve of removed individuals. Thus the SEIR model should not be used in the absence of additional information that might be obtained by clinical studies. Nevertheless, the SIR model can be used with some modifications.

### 2.1 Relation between the basic reproduction number and the total number of removed individuals

The ratio  $\beta/\eta$ , called the Basic Reproduction Number and denoted as  $R_0$ , is the key parameter in both the SIR and SEIR models. This number is related to the growth rate of the number of infected individuals in a fully susceptible population and determines the final value of  $R$ ,  $R_f$ , that is the proportion of individuals that will be affected by the disease. This proportion includes individuals who gain immunity without showing symptoms, those who are treated, as well as disease-related fatalities. The reciprocal of the parameter  $\eta$ ,  $T=1/\eta$  is considered as a representative of the mean infectious period.

In [14], it was shown that the normalized  $R(t)$  for the SEIR model with parameters  $(\beta, \varepsilon, \eta)$  and the normalized  $R_I(t)$  for the SIR model with parameters  $(\beta_I, \eta_I)$  have the same final value ( $R_f$ ), the same  $R_m$  value at the peak of the number of infected individuals, and the same slope ( $R'_m$ ) at that point (after a time shift of these curves), provided that  $1/\eta_I = 1/\varepsilon + 1/\eta$ . It follows that if one has to work with the normalized data, the SIR model can be used to model diseases with an incubation period, provided that the sum of the incubation and the infection periods of the SEIR model is used as an effective infectious period for the SIR model, unless there are reasonable estimates for  $R_0$  and/or for the duration of incubation and/or infection periods.

It was also observed in [14] that the normalized curves of removed individuals are practically indistinguishable for moderate values of  $R_0$ , such as in ordinary flu. As the Basic Reproduction Number for Covid-19 is high, one cannot expect that the shifted normalized curves for the SIR and SEIR models coincide, but as clinical information on Covid-19 parameters is yet unclear, the SIR system is adopted as the basic model in the present work.

The relation between  $R_0$  and  $R_f$  is determined as follows. Note that  $R(t)$  is monotone increasing, and hence it can be used as an independent variable, instead of  $t$ . The derivative of  $S$  with respect to  $R$  is given by

$$dS/dR = S'/R' = -\beta/\eta S = -R_0 S. \quad (3)$$

Assuming initial conditions  $S \rightarrow 1$  and  $R \rightarrow 0$  as  $t$  approaches negative infinity, one can obtain the following by integrating (3)

$$S = e^{-R_0 R}. \quad (4)$$

As  $t$  approaches positive infinity,  $I \rightarrow 0$ , and  $S(t)$  and  $R(t)$  approach their final values  $S_f$  and  $R_f$ , respectively. Then  $S+I+R=1$  yields

$$R_f + e^{-R_0 R_f} = 1. \quad (5)$$

Consequently,  $R_0$  is derived as

$$R_0 = -R_f \ln(1 - R_f). \quad (6)$$

Therefore, the values  $S_m=S(t_m)$ ,  $I_m=I(t_m)$  and  $R_m=R(t_m)$ , as obtained in [14] are

$$S_m = 1/R_0, \quad R_m = -R_f \ln(R_0)/\ln(1 - R_f), \quad I_m = 1 - S_m - R_m. \quad (7)$$

Here,  $t_m$  refers to the time at which the number of infectious cases reaches its maximum.

The values  $S_m$ ,  $R_m$  and  $I_m$  are crucial in determining the proportion of individuals that need to be vaccinated in order to reduce the proportion of susceptible individuals below the threshold value  $S_m=1/R_0$ .

The graph of  $R_f$  versus  $R_0$  is shown on Figure 1, together with the ranges of  $R_0$  for well-known diseases. It can be seen that for  $R_0>2.5$ ,  $R_f$  is greater than 90%. The figure also shows that the increase in  $R_f$  with respect to  $R_0$  is very slow for  $R_0>3$ . It is generally accepted that the  $R_0$  for Covid-19 is greater than 3 despite all containment measures [13], [15], [16]. Thus, unless vaccination is applied, one would expect that at least 95% of the population would be affected by the disease. In addition, the knowledge of its precise value would have little effect on the planning of healthcare measures. It should also be kept in mind that containment measures provide a temporary control of the spread of the epidemic, just to the point of reducing the burden of the epidemic to a manageable size.

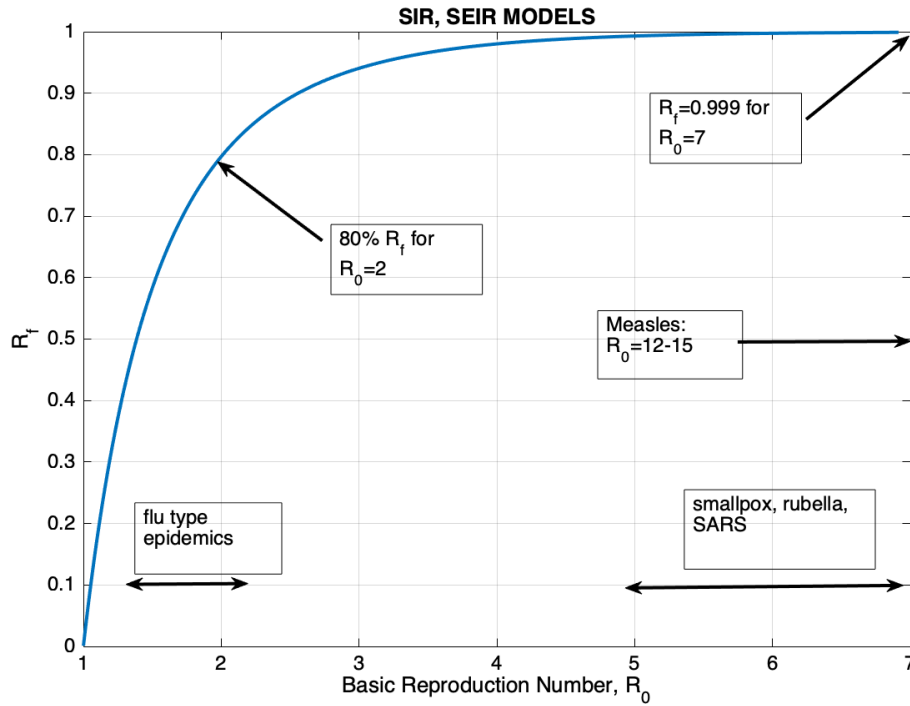


Figure 1. The graph  $R_f$ , the final proportion of individuals that would be affected by the disease, versus the Basic Reproduction Number  $R_0$ .

According to the Centers for Disease Control and Prevention (CDC), it is still unknown when viral shedding begins or how long it lasts for, and nor is the period of COVID-19's infectiousness known. Like infections with MERS-CoV and SARS-CoV, SARS-CoV-2 RNA may be detectable in the upper or lower respiratory tract for weeks after illness onset, though the presence of viral RNA is no guarantee of the presence of the infectious virus. It has been reported that the virus was found without any symptoms being shown (asymptomatic infections) or before symptoms developed (pre-symptomatic infections) with SARS-Cov-2, though the role they may play in transmission remains unknown. According to prior studies, the incubation period of SARS-CoV-2, like other coronaviruses, may last for 2-14 days [17].

To illustrate an example for an SIR model,  $R_0$ ,  $T$  and  $R(t_0)$  are chosen as 3, 10 and  $10^{-3}$ , respectively and the related graphs are given on Figure 2.

From Figure 2, it can be seen that for parameter values  $R_0=3$ ,  $T=10$  day, the duration of the epidemic is about 100 days. The peak of the epidemic occurs approximately at day 35. Note that the derivative of  $I(t)$  vanishes at time  $t_m$  when  $S(t_m)=1/R_0$ . In this example,  $S(t_m)=0.3333$ ,  $I(t_m)=0.3005$  and  $R(t_m)=0.3662$ . The final values of  $S(t)$  and  $R(t)$  are  $S_f=0.0595$  and  $R_f=0.9404$  at the end of the epidemic.

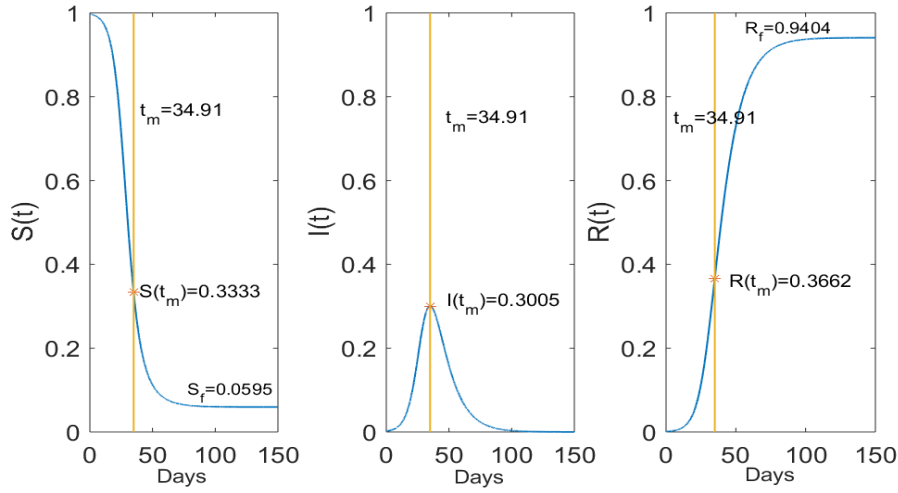


Figure 2: The time evolution of  $S(t)$ ,  $I(t)$  and  $R(t)$  for  $R_0=3$ ,  $T=10$  days and  $R(0)=10^{-3}$ .

## 2.2 Representative data for the proportion of removed individuals

It is in general accepted that the number of fatalities represents the number of removed individuals and the number of confirmed cases represents the number of infected individuals. The proportionality constants are unknown, but as long as they are constant, one can work with the normalized case reports and normalized fatalities and look for the determination of the epidemic parameters from the shape of these normalized curves. In Section 4, it will be shown that for the Covid-19 data, total cases would be a better representative of the number of removed individuals.

## 3 OVERVIEW OF DATA

According to the SIR and the SEIR models, given by the equations (1) and (2), the rate of change of the number of removed individuals is proportional to the number of infectious cases. In terms of observations, this corresponds to the fact that the ratio of, for example, daily fatalities to daily infectious cases should be constant. In the literature on the analysis of historical epidemics, fatality reports are usually the only available data, hence models are necessarily based on the assumption that cumulative fatalities represent cumulative number of removed individuals. For the Covid-19 pandemic, as daily fatality and infectious case reports are available, further evaluation of the representation of  $R(t)$  in terms of fatality data is presented. Daily infections and total fatalities are displayed on Figure 3, for all countries. From these graphs, it is difficult to see whether the relation  $R' = \eta I$  is satisfied or not. For this, first, daily fatalities as representatives of the derivatives of  $R(t)$  will be compared with daily infectious cases as representatives of  $I(t)$  on Figures 4-6. Based on these comparisons, it will be concluded that total infectious cases would be better representative of the number of Removed individuals,  $R(t)$ .

### 3.1 Time evolution of daily infections and total fatalities

Normalized daily infectious cases and total fatalities are shown on Figure 3.

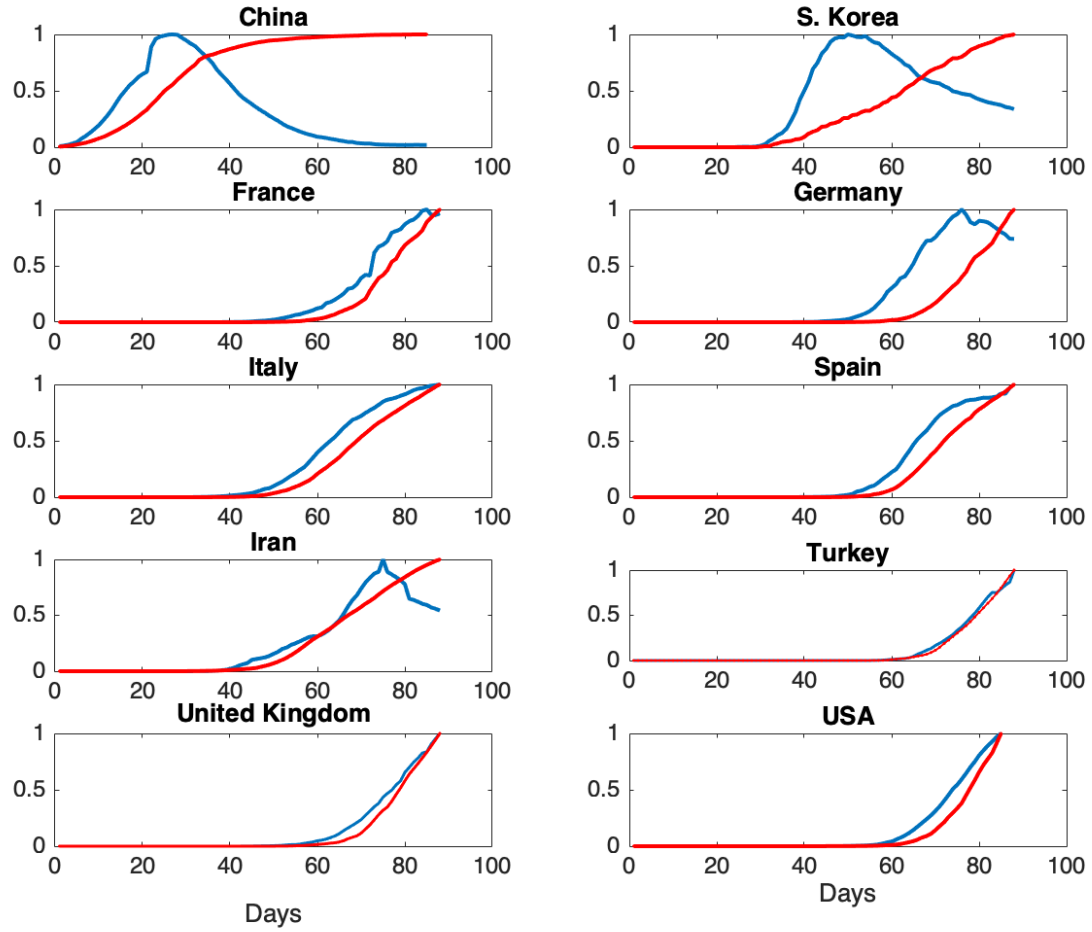


Figure 3. Normalized daily case reports (Blue) and normalized total fatalities (Red) for each country. The horizontal axis represents days starting from January 22.

From Figure 3, it can be seen that the epidemic cycle has been completed in China over the course of about 70 days. The jump in total fatalities is due to a change in the reporting scheme. As our analysis is based on total infectious cases, this change has no effect on the models. For South Korea, the epidemic is in a state of slow decrease at the end of about 60 days, but the rate of infections is still high. This qualitative behaviour is an indication of the fact that  $R_0$  for South Korea is expected to be much higher than the one for China. For France, Germany and Iran, the epidemic is in the decline phase. For the rest of the countries, further analysis is needed in order to assess epidemic phase.

### 3.2 Time evolution of daily infections and daily fatalities

Daily infectious cases and daily fatalities are given on Figures 4-6. Daily fatalities are represented by the derivative of total fatalities, computed with 3 different methods.

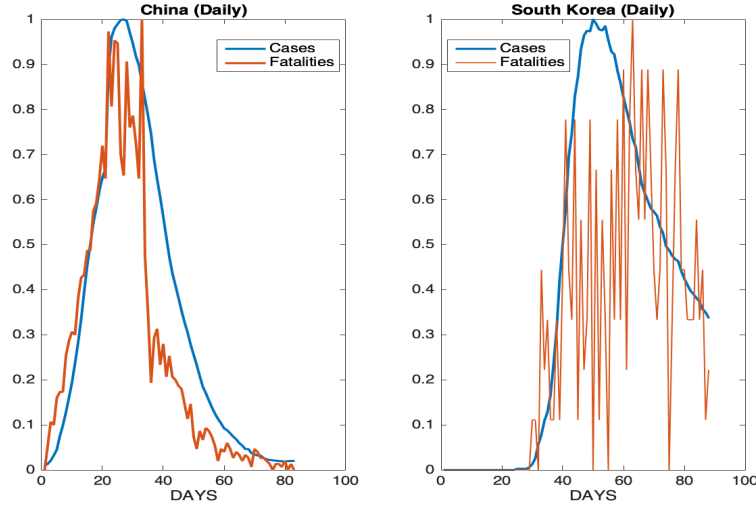


Figure 4. Daily cases (Blue) and daily fatalities (Red) for China and South Korea. The horizontal axis represents days starting from January 22.

Note that in China, fatalities occur earlier than hospitalizations in the initial phase. This early phase is followed by a “stationary” period, over which the number of hospitalizations is around their peak and the number of daily fatalities oscillates around a mean. This intermediate period ends with a sharp decrease in fatalities (the reasons should be investigated). During the third phase, the decrease in the fatalities is faster than the decrease in the hospitalizations. Thus the data for China has 3 phases while the data for South Korea looks much like the stationary phase of the data for China. It should be also pointed out that the number of fatalities in South Korea is very low as compared with China, hence it would be expected that the number of infections is a better representation of the number of infections.

Daily infectious cases and daily fatalities for France, Germany, Italy and Spain, and for Iran, Turkey, the United Kingdom and the United States are shown on Figures 5 and 6, respectively.

From these graphs, one can see that in Germany, the number of daily infectious cases leads the number of daily fatalities, but for other countries either they coincide or the situation is reversed. The underlying reasons for this behaviour should be analysed in more detail, using country specific information on strategies for testing infectious cases and treatment procedures applied in the course of the epidemic.

#### 4 ESTIMATION OF THE SIR MODEL PARAMETERS

As noted above, the knowledge of  $R_0$  determines the total proportion of individuals that would be affected,  $R_f$ . Furthermore, the peak of  $I(t)$  occurs at the time  $t_m$ , at which the proportion of susceptible individual falls to the value  $1/R_0$ . This information is useful for the determination of the proportion of people that have to be vaccinated in order to drag the proportion of susceptible individuals below this threshold. The Basic Reproduction Number is “defined” as the number of new infections per unit time in a fully susceptible population. Thus, it is a quantity that might be



measured by direct on-site observations. On the other hand, the knowledge of  $R_0$  by itself does not give any information on the timing of the progress of the epidemic.

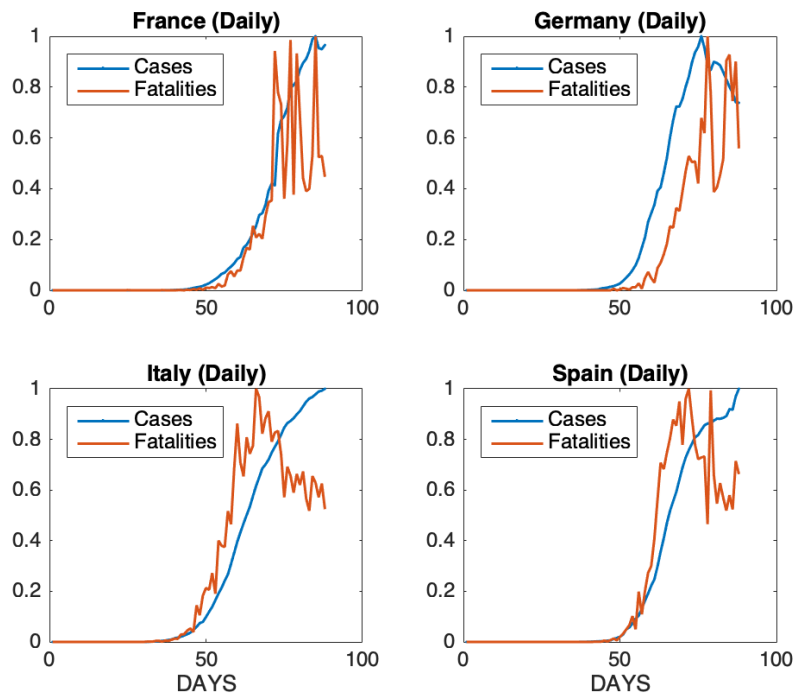


Figure 5. Daily cases (Blue) and daily fatalities (Red) for France, Germany, Italy and Spain. The horizontal axis represents days starting from January 22.

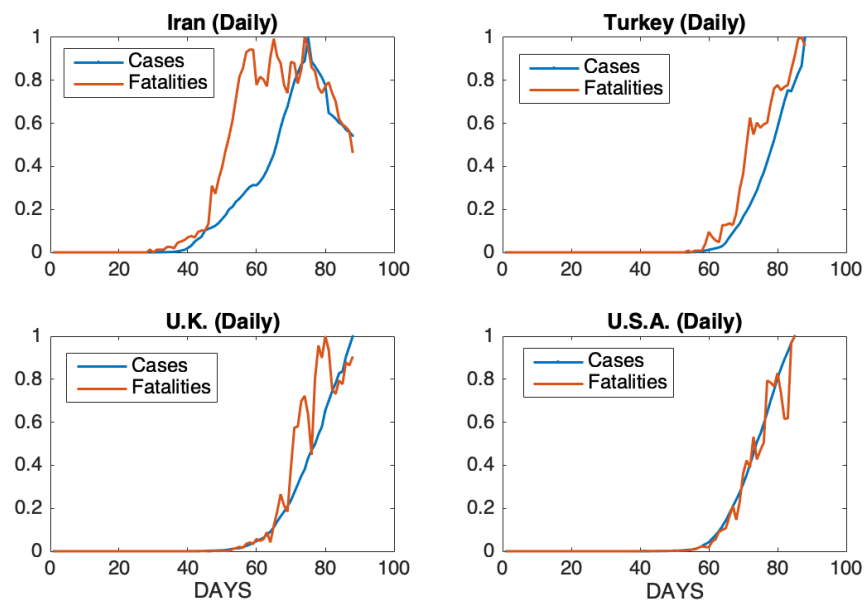


Figure 6. Daily cases (Blue) and daily fatalities (Red) for Iran, Turkey, United Kingdom and United States. The horizontal axis represents days starting from January 22

In the present work, the determination of the following parameters is discussed:

- 1) The Basic Reproduction Number  $R_0$ ,
- 2) The mean duration of the infectious period  $T$ ,
- 3) The time  $t_m$  (days) at which the number of infectious cases reaches its maximum, i.e, the first derivative of  $I(t)$  is zero,
- 4) The time  $t_a$  (days) at which the rate of increase in the number of infectious cases reaches its maximum, i.e, the time at which the second derivative of  $I(t)$  is zero and the first derivative is positive,
- 5) The time  $t_b$  (days) at which the rate of decrease in the number of infectious cases reaches its maximum, i.e, the time at which the second derivative of  $I(t)$  is zero and the first derivative is negative.

It will be seen that  $R_0$  and  $T$  can be estimated only for China where the spread of the epidemic is over. For other countries,  $R_0$  and  $T$  cannot be estimated from the normalized data, but the timings of the key events,  $t_m$ ,  $t_a$  and  $t_b$  can be determined quite reliably.

#### 4.1 Methods for estimating the parameters $R_0$ , $T$ , $t_m$ , $t_a$ , and $t_b$

These parameters are determined by a “brute force” approach: The models are run for a broad range of parameters. Then the difference between data and the model is compared by using various norms. Finally, the models that match data within 5% are selected. If the scatter plot of the errors versus the parameter to be estimated has a sharp minimum, it is concluded that the corresponding parameter can be determined from the shape of the normalized data.

The parameter ranges for the SIR model are

$$1.5 < R_0 < 10, \quad 2 < T < 30, \quad (8)$$

and the initial values are chosen as

$$R_{ini} = 10^{-k}, \quad S_{ini} = e^{-(\beta/\eta)R_{ini}}, \quad I_{ini} = 1 - S_{ini} - R_{ini}, \quad (9)$$

where  $1 < k < 10$ . For South Korea, these parameter ranges are extended appropriately.

#### 4.2 Selection of representative Data for $R(t)$

In the SIR model, since  $R' = \eta I$ ; that is, the rate of change in the number of removed individuals is proportional to the number of infected individuals, it is expected that the cumulative cases are proportional to cumulative fatalities. Thus, the SIR model predicts the simultaneity of the daily fatalities and daily infections. The verification of this fact requires the availability of data both for infections and for fatalities. To the best of our knowledge, historical data studied in the literature includes fatalities only, and the data for the 2009 H1N1 epidemic collected at certain major hospitals [18] is unique in the sense of reflecting information on both infections and fatalities. The peculiarity of this data is a shift of about 8 days between total infections and total fatalities, the peak of infections occurring 8 days prior to the peak of fatalities. This time shift was explained by a multi-stage SIR model [19].

Cumulative cases and cumulative fatalities for Covid-19 do not show such a clear time shift. On the contrary, in China and Korea, fatalities increase faster than infections. In Germany, there is a slight lead for infections, while for other countries the two curves more or less coincide. The lead of fatalities over infections that is observed in China and in Korea is an unexpected fact, which is possibly due to the irregularities in the statistics, in medical treatment practices, etc. We should also note that the progression of the Covid-19 epidemic is unique in the sense that new treatment methods are applied during the initial phase in China and these methods have been applied in other countries.

For China, several programs were run, first by fitting the predicted  $R(t)$  to the total fatality data, then to the cumulative infectious case data. In the first case, about 700 models fitting cumulative fatalities within 5% error and about 3000 models that fit cumulative infections within 5% error are found. Furthermore, in the latter case, the minima for the quantities that were aimed to be determined were much sharper. For South Korea, as it will be explained later, the model matching was not successful. For other countries, as the difference between total infections and total fatalities was negligible, total infections are used as a representative of  $R(t)$  of the SIR model.

Our main result is that it is not possible to determine the Basic Reproduction Number and the mean duration of the infectious period from the shape of the *normalized* data (unless there are reasonable estimates for either of these parameters). In order to make a stable determination of the parameters  $R_0$  and  $T$  by using the early stage data, a certain period of time has to pass. This period is approximately 70 days for 2009 A(H1N1) epidemic [20]. However, this period for Covid-19 is still uncertain. This is possibly the reason why the parameters for countries other than China and South Korea can not be established. On the other hand, the timings of the peak of the infectious cases, the peak of the rate of increase and the rate of decrease of the infectious cases can be determined more precisely from the shape of the normalized data.

### 4.3 Simulations for SIR models with $R_0/T = \text{constant}$

The ‘best’ estimations of the parameters  $R_0$  and  $T$  lie on a curve that is nearly linear when a SIR model is used to fit the data of an epidemic. This fact has been observed in previous work [20], in the study of the H1N1 epidemic and it was explained by the fact that the duration of the epidemic pulse (appropriately defined in terms of a fraction of the peak of infections) was nearly invariant for values of  $R_0$  and  $T$ , with  $R_0/T$  constant.

In order to visualize this situation, the solutions of this system of differential equations of the SIR model (1) for parameter ranges  $1.5 < C < 10$ ,  $R_0 = 2 \cdot C$ ,  $T = 10 \cdot C$ , and for  $R_0/T = 1/5$  (constant) are obtained. The graphs of normalized solutions (after an appropriate time shift) are given in Figure 7.

## 5 RESULTS FOR EACH COUNTRY

The scatter plots of the mean infectious period  $T$  versus  $R_0$ , and the scatter plots of the modelling error versus the parameters are presented in Figures 9-18 where  $I_t$  and  $I_{tt}$  represent the values of the first and the second derivatives of  $I(t)$  at the last day of the data April 18, 2020, respectively. The error stands for the relative error between the normalized  $R(t)$  of the model and normalized total infectious cases, in the  $L_2$  norm.

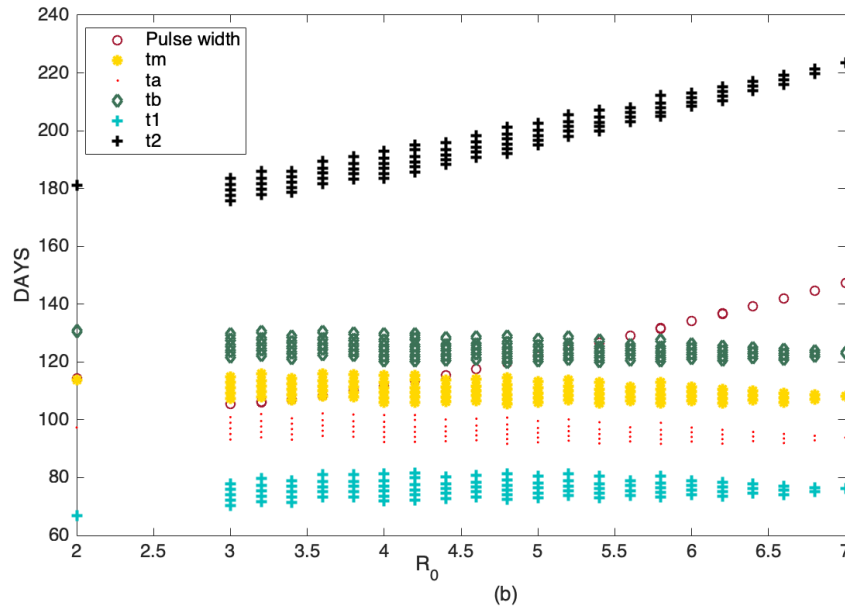
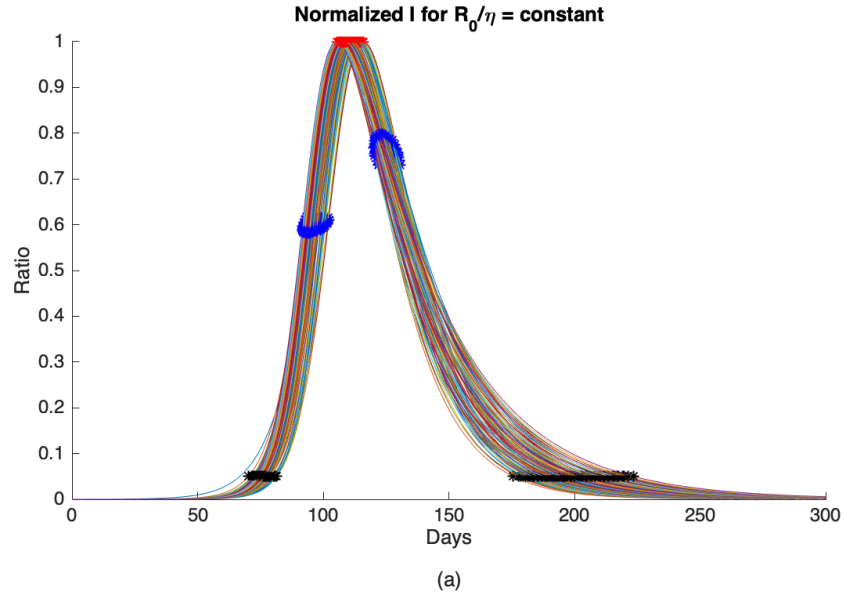


Figure 7 (a) Normalized values of  $I(t)$  for  $1.5 < C < 10$ ,  $R_0 = 2 \cdot C$ ,  $T = 10 \cdot C$ , and for  $R_0/T = 1/5$ , together with the inflection points ( $t_a$ ,  $t_b$ ), the peak point ( $t_m$ ) and the timing of the initial ( $t_1$ ) and final ( $t_2$ ) points when the 5% of maximum value  $I(t)$  epidemic, (b) Dependency of  $t_1$ ,  $t_a$ ,  $t_m$ ,  $t_b$ ,  $t_2$  on  $R_0$ .

### 5.1 *Scatter plot of the mean duration of the infectious period versus the basic reproduction number*

In Figures 9-18, the first graph, in the upper left of the panel is the scatter plot of the mean duration of the infectious period,  $T$ , with respect to the basic reproduction number  $R_0$ , for models that fit

data within 5% error in the norm described above. For all countries, the “best” parameters lie on a curve, instead of being agglomerated around a mean. This indicates that although the SIR model fitting normalized data is unique, the parameters  $R_0$  and  $T$  cannot be determined precisely from normalized data. The colors blue, red and yellow in Figures 9-18 represent the results according to whether the last day of the analysis,  $t_f$ , is 78, 83 and 88, respectively.

### **5.1.1 Scatter plot of the modelling error versus the basic reproduction number and versus the mean duration of the infectious period**

In Figures 9-18, the second (first row, right panel) and the third (second row, left panel) graphs display the scatter plot of the modelling error with respect to  $R_0$  and  $T$  respectively. For China, there are well defined minima in the modelling errors at nearly  $R_0=3$  and  $T=9$ . For South Korea, the minima of the error in  $R_0$  seems to be located beyond  $R_0=8$ , and the minimal error in  $T$  corresponds to  $T=25$  approximately. These parameter values are not in the ranges reported in the literature. Furthermore, these values are not stable under the variation of the last day of the analysis. It is therefore concluded that the data for South Korea shows completely different characteristic that might be explained by the fact that approximately 27.4 percent of the confirmed coronavirus patients in South Korea were in their 20s. For all of the remaining countries, the ranges of  $R_0$  and  $T$  corresponding minimal modelling errors are too large to attempt any reasonable estimation for these parameters.

### **5.1.2 Timing of the peak of the maximum for $I(t)$**

The fourth (second row, right panel) graph in Figures 9-18 shows the scatter plot of the modelling error versus  $t_m$ , the timing of the peak of the number of infections. For all of the countries analysed, this parameter can be estimated quite sharply. In order to study the reliability of this estimation, the model matching process is repeated for  $t_f=78, 83$  and  $88$ .

### **5.1.3 Timing of the inflection points of $I(t)$**

The ratio of infected individuals  $I(t)$  has two inflection points. The first inflection point ( $t_a$ ) is located at the left of the maximum ( $t_m$ ) whereas the second one ( $t_b$ ) is located at the right of  $t_m$ .  $t_a$  and  $t_b$  correspond to the highest rate of increase and decrease in  $I(t)$ , respectively. In Figures 9-19, the right and left panels of the third row display scatter plot of the error in these quantities. Their variation with respect to  $t_f$  is also investigated.

### **5.1.4 Final values of the first and second derivatives of $I(t)$**

The values of the first and second derivatives at  $t_f$  are shown on the fourth row, left and right panels, respectively. If the first derivative is positive (negative), the  $I(t)$  is in the rising (falling) phase, while if the second derivative is positive (negative) the curve is concave up (down).

The epidemic phases which are categorized by the sign of the first and the second derivatives of  $I(t)$  are given in Figure 8.

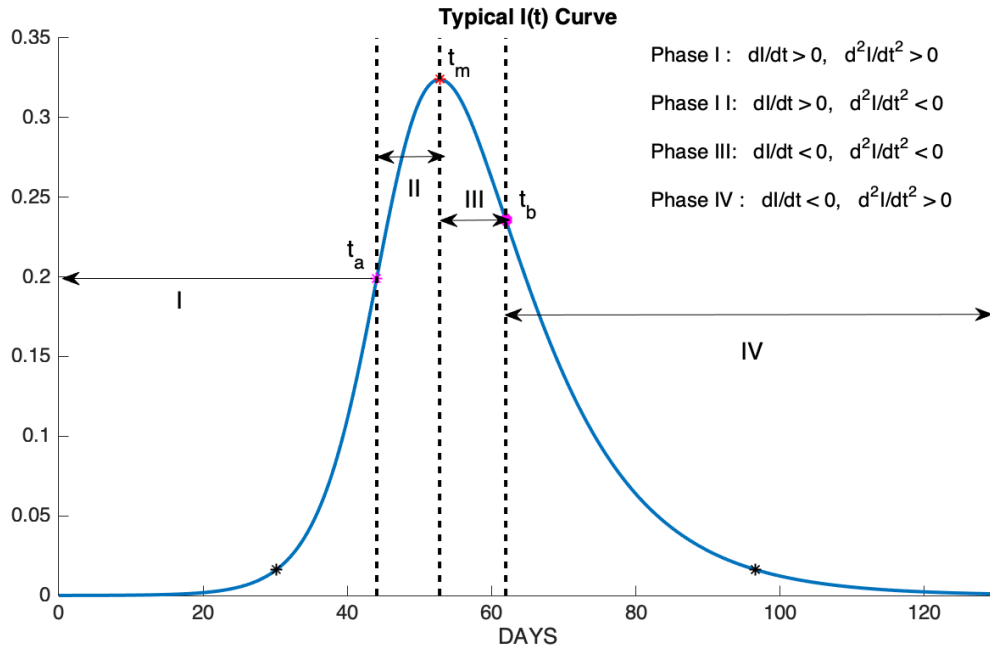


Figure 8. Phase I: slow increase, Phase II: fast increase, Phase III: fast decrease, Phase IV: slow decrease.

Estimation of parameters for each country and for  $t_f=88$  is summarized in Table 1.

Table 1 Timing of the phases of the epidemic.

	China ( $t_f=85$ )	S. Korea	France	Germany	Italy	Spain	Iran	Turkey	U. Kingdom	U. States
$R_0$	3	8	-	-	-	-	-	-	-	-
$T$	9	25	-	-	-	-	-	-	-	-
$t_m$	26	50	86	76	81	81	75	88-92	87-92	90-92
$t_a$	18	41	74	65	65	67	63	78	77	75
$t_b$	35	59	95-104	88-90	97-100	93-95	86-88	96-104	100-108	97-102

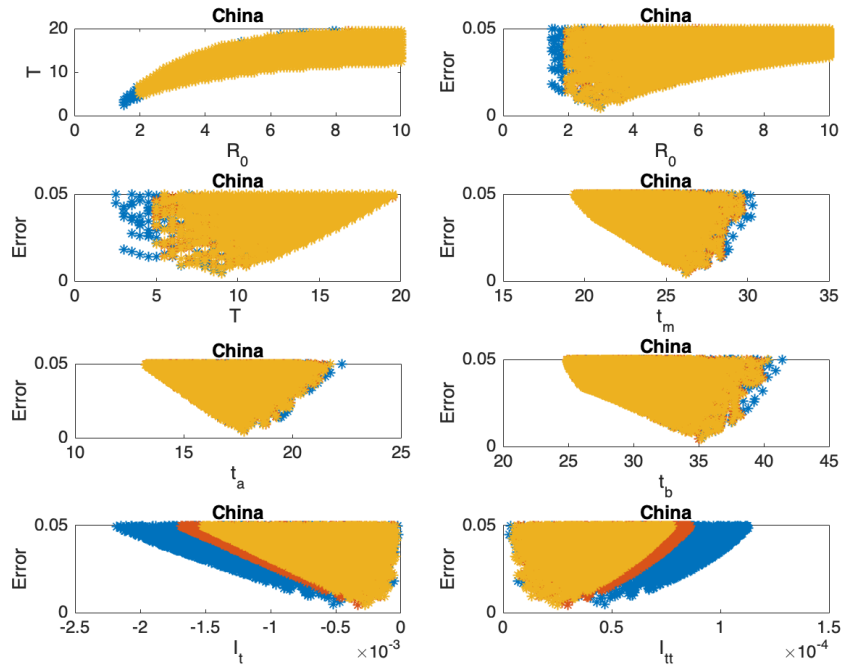


Figure 9. China: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is in Phase IV.

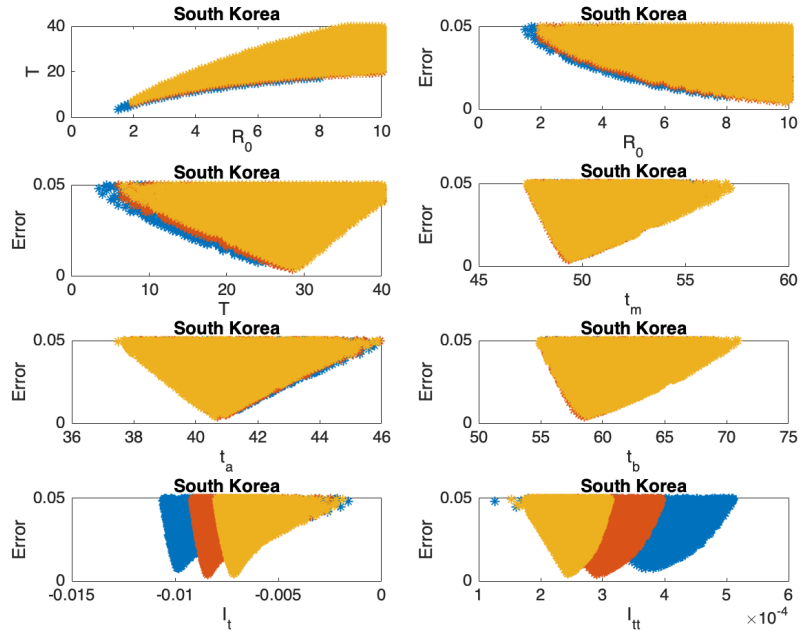


Figure 10. South Korea: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is in Phase IV. The values for  $R_0$  and  $T$  don't seem to fall in reasonable ranges and the data for South Korea should be studied more closely.

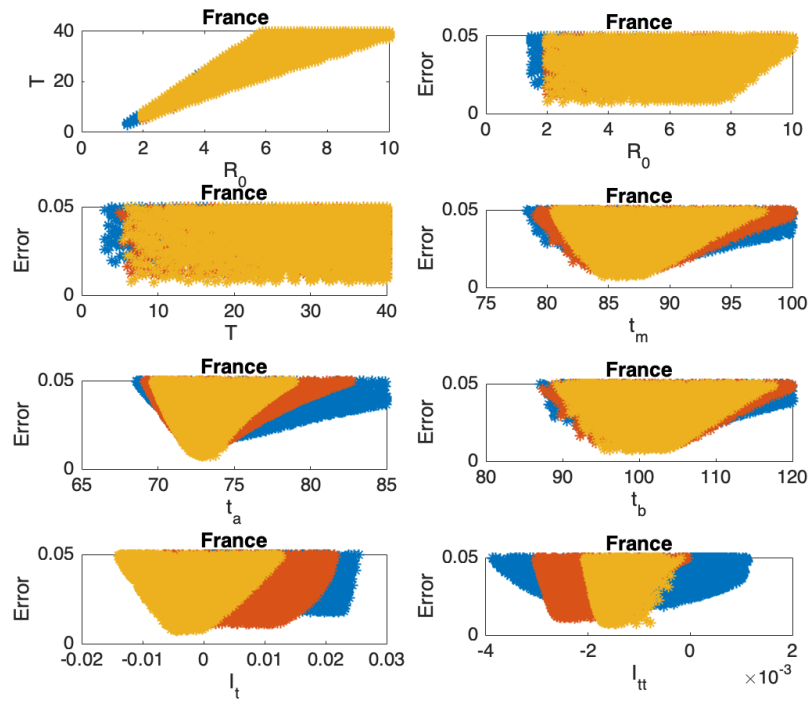


Figure 11. France: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is at the beginning of Phase III.

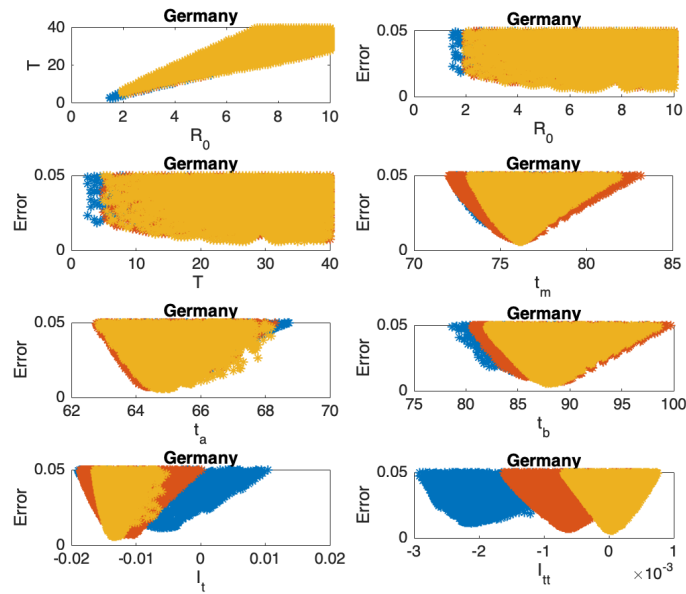


Figure 12. Germany: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is at the beginning of Phase IV.



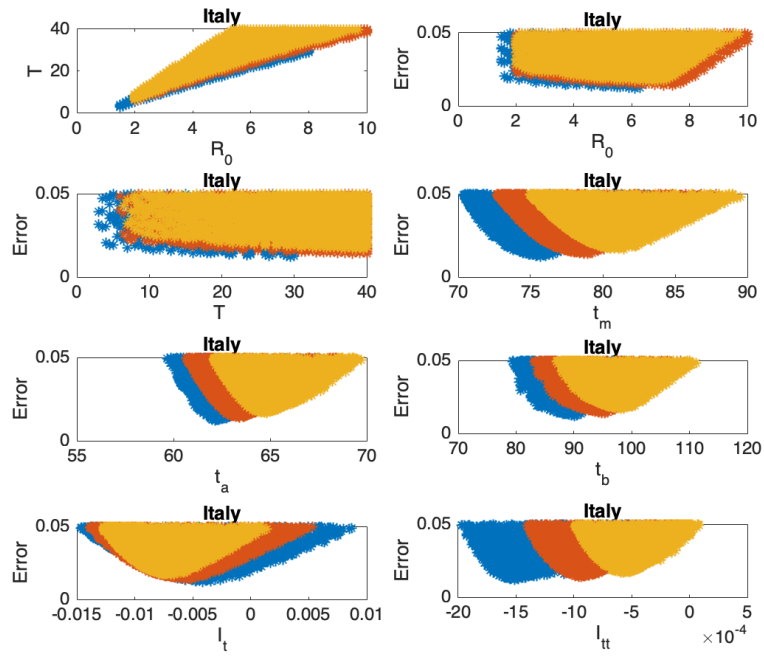


Figure 13. Italy: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is in Phase III.

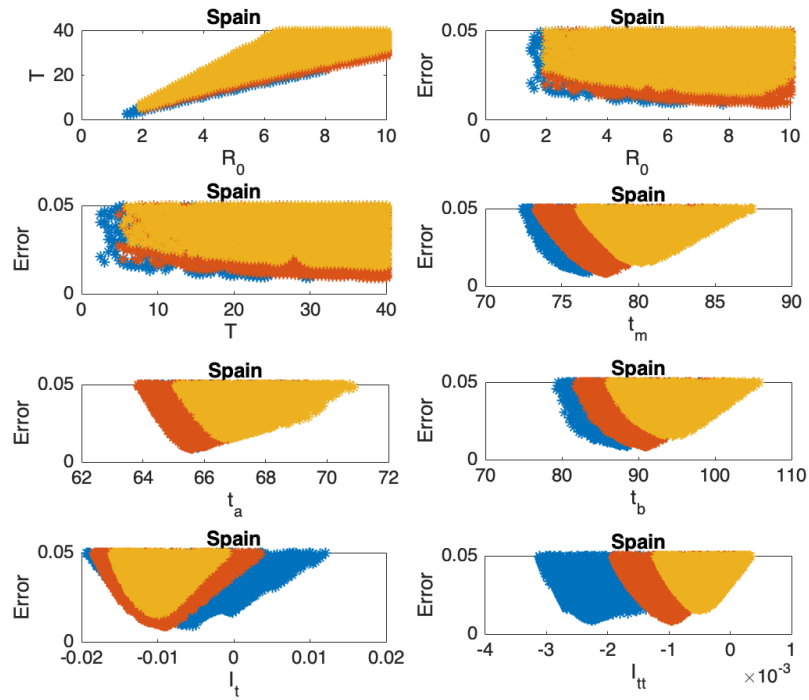


Figure 14. Spain: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is in Phase III.

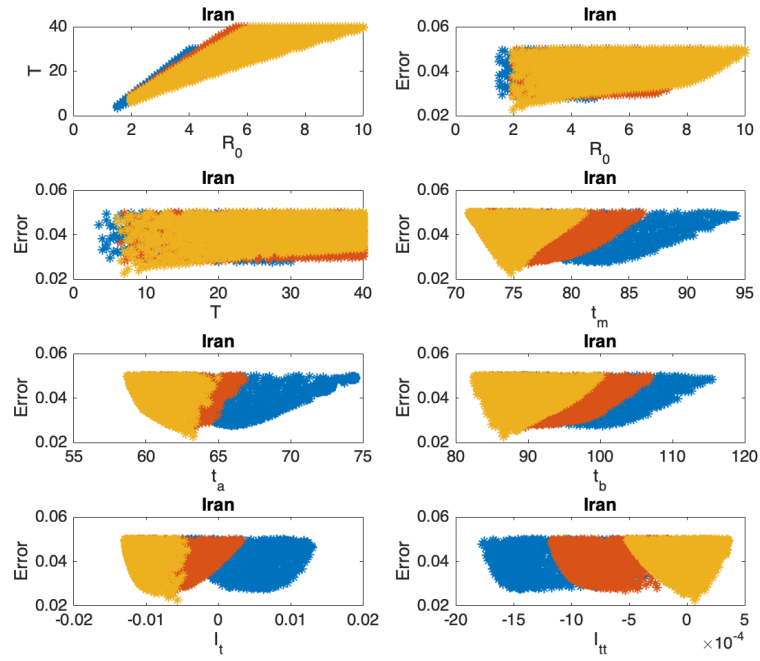


Figure 15. Iran: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is at the beginning of Phase IV.

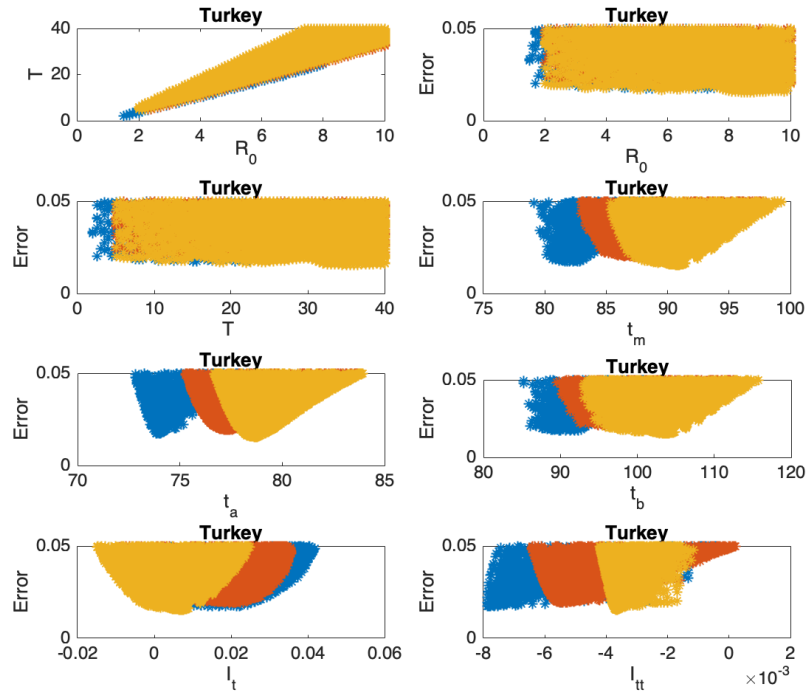


Figure 16. Turkey: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is at the end of Phase II.

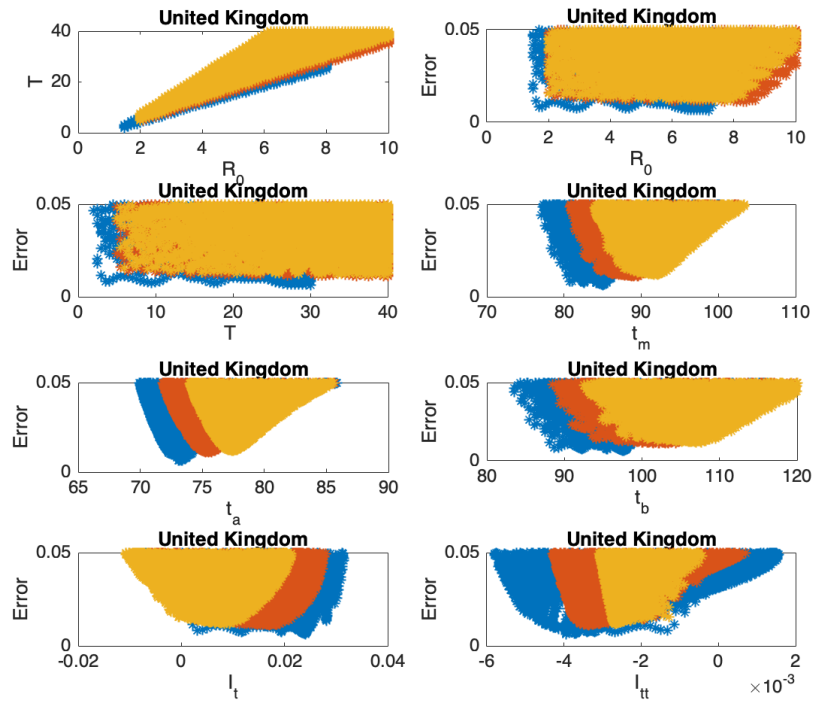


Figure 17. United Kingdom: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is in Phase II.

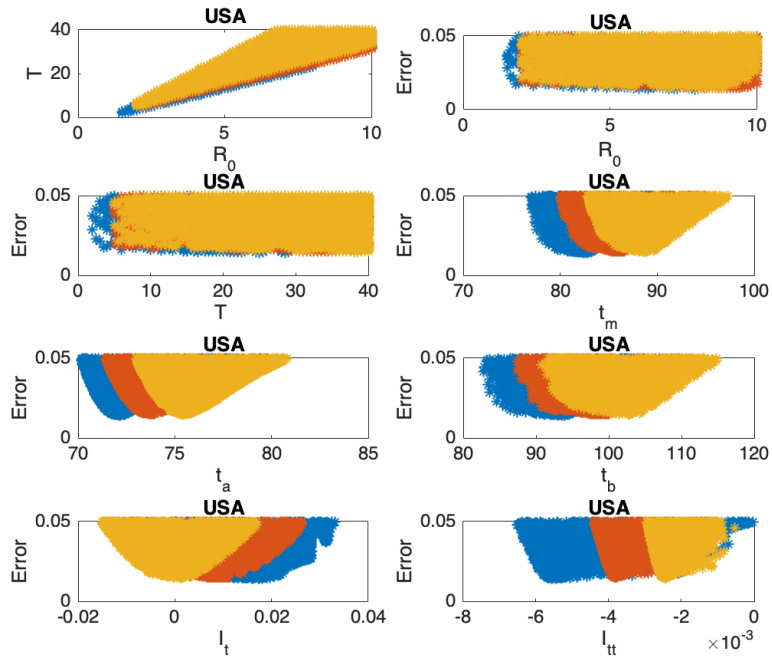


Figure 18. United States: The 7<sup>th</sup> and 8<sup>th</sup> graphs indicate that the epidemic is at the end of Phase II.

## 6 MODELLING VERSUS FORECAST

In Section 5, it can be seen that although  $R_0$  and  $T$  cannot be determined, it was possible to estimate  $t_m$ ,  $t_a$  and  $t_b$  quite sharply from data. In this section, the reliability of these estimates is discussed by comparing predictions based on data with different time spans.

The best SIR models fitting data for 78, 83 and 88 days are obtained, and data and graphs of 10 best models for each time span are plotted in Figures 19-23. For China and South Korea, for which the epidemic cycle is more or less complete, estimations based on time spans varying by 5 days give the same result as can be observed in Figure 19.

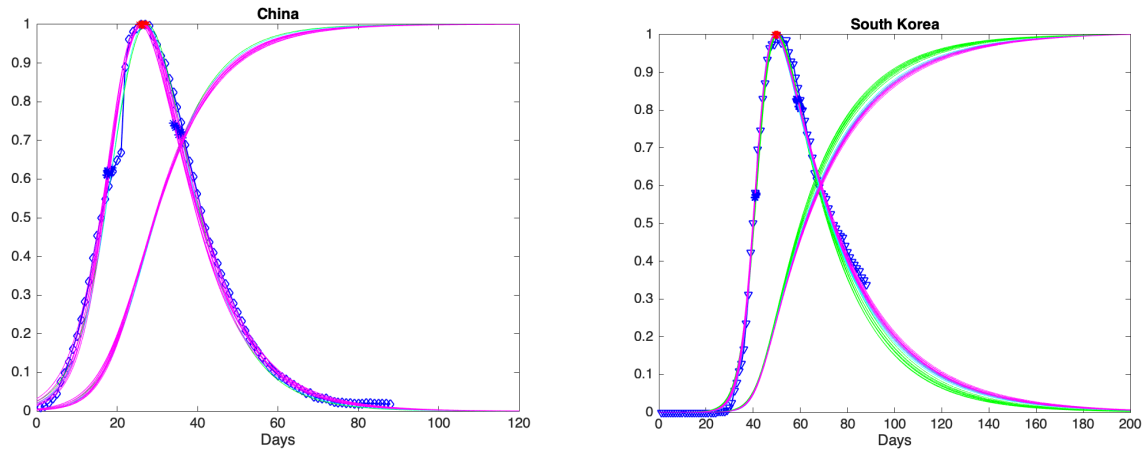


Figure 19. China and South Korea: Graphs of  $I(t)$  and  $R(t)$  for the best 10 SIR models for each time span (diamond: real data).

On the other hand, for those countries that are as yet before or around the peak of the epidemic, the situation may be different, as can be observed in Figure 20-23.

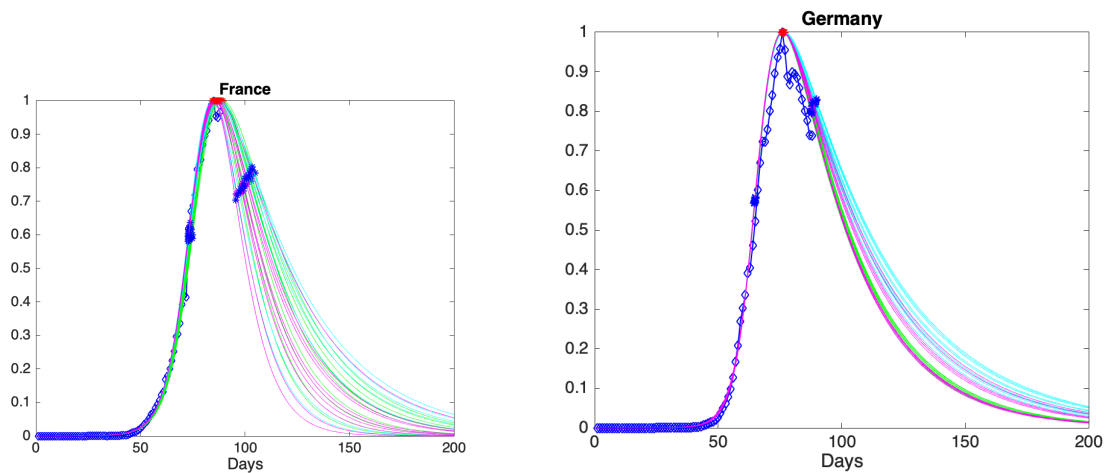


Figure 20. France and Germany: Graphs of  $I(t)$  for the best 10 SIR models for each time span (diamond: real data).

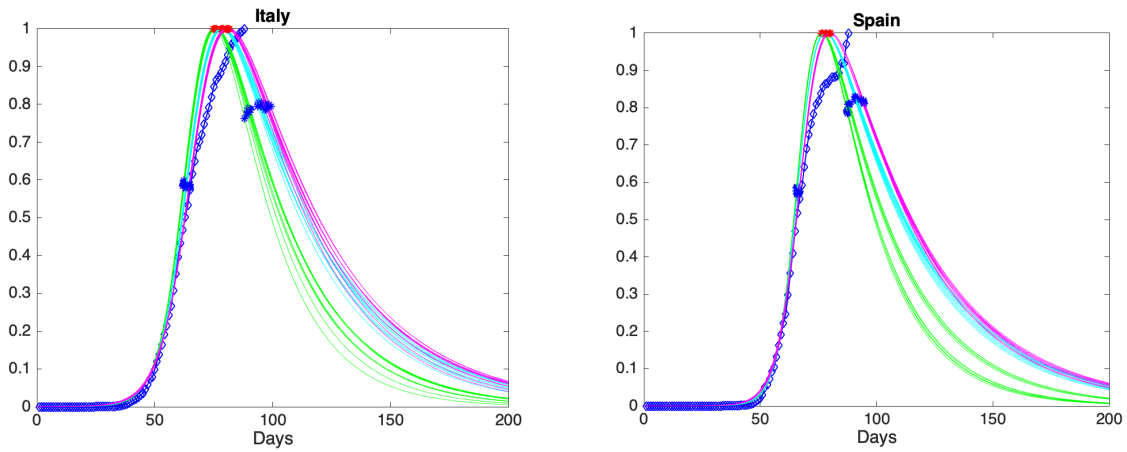


Figure 21. Italy and Spain: Graphs of  $I(t)$  for the best 10 SIR models for each time span (diamond: real data).

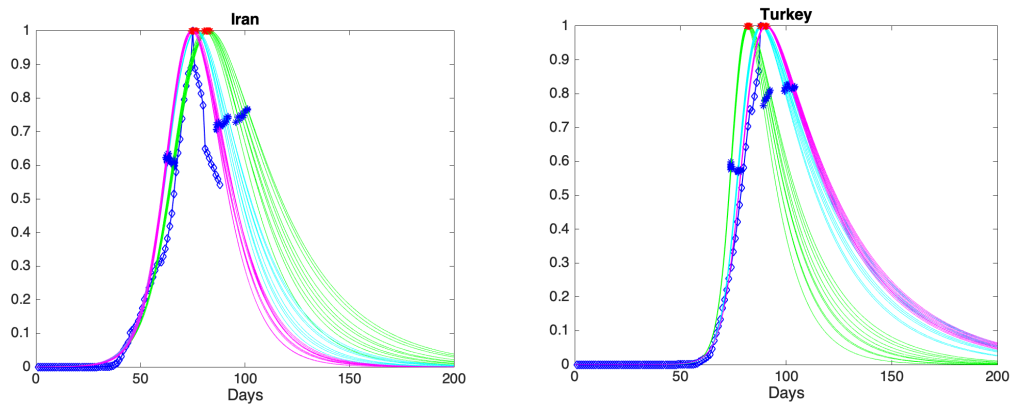


Figure 22. Iran and Turkey: Graphs of  $I(t)$  for the best 10 SIR models for each time span (diamond: real data).

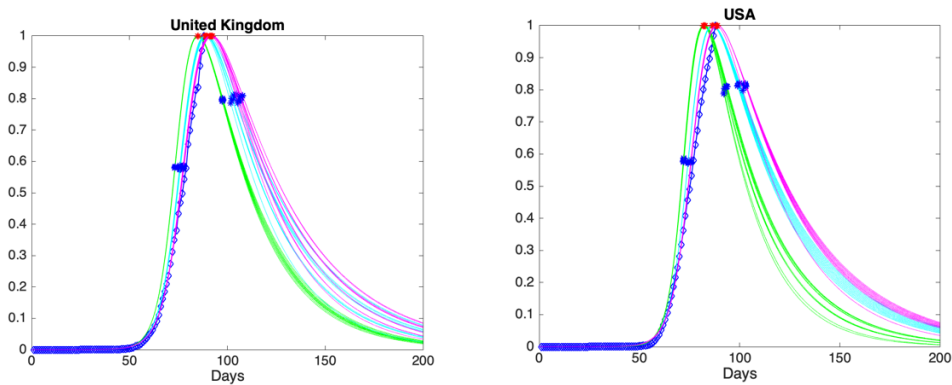


Figure 23 United Kingdom and United States of America: Graphs of  $I(t)$  for the best 10 SIR models for each time span (diamond: real data).

## 7 CONCLUSION

The epidemic parameters of Covid-19 for ten selected countries are estimated by using the data released by the state offices. These parameters include the basic reproduction number, mean duration of infectious period, the time at which the number of infectious cases reaches its maximum, the time at which the rate of increase in the number of infectious cases reaches its maximum, the time at which the rate of decrease in the number of infectious cases reaches its maximum. For each country, the best Susceptible-Infected-Removed (SIR) models fitting *cumulative case data* are obtained. A wide variety of intervals with different scales of the parameters, basic reproduction number  $R_0$  and infectious period  $T$ , are observed. More specifically, the basic reproduction number and mean duration of infectious period are estimated only for China since the spread of the disease there is over. These parameters are found to be 3 and 5, respectively. The fact that the median incubation and infection periods are approximately 5 days, supports the observations for  $R_0$  and  $T$ . However, the basic reproduction number and infectious period for other countries cannot be predicted from the normalized data but the timing of key events can be estimated quite reliably. To summarize, we show that the quantity that can be the most robustly estimated from the normalized data, is the timing of the highest rate of increase in the number of infections, i.e, the inflection point of the number of infected individuals. However, it should be pointed out that the analysis performed by the SIR model for South Korea provides dissimilar results which can be explained by the unique age distribution nature of the confirmed cases.

## 8 REFERENCES

1. World Health Organization. (2020). Coronavirus disease 2019 (COVID-19): situation report, 52. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> [Accessed April 27, 2020].
2. Tizzoni M., Bajardi P., Poletto C., Ramasco JJ., Balcan D., Gonçalves B., Niccola P., Colizza V., Vespignani A. (2012). Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Medicine* 10(1): 165. doi:10.1186/1741-7015-10-165.
3. Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*. 134, 109761. doi:10.1016/j.chaos.2020.109761.
4. Kucharski, Adam J., et al. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*. doi: 10.1016/S1473-3099(20)30144-4.
5. Roda WC., Varughese MB., Han D., Li, MY. (2020). Why is it difficult to accurately predict the COVID-19 epidemic?. *Infectious Disease Modelling*. 5, 271-281. Doi: 10.1016/j.idm.2020.03.001.
6. Roosa K., Lee Y., Luo R., Kirpich A., Rothenberg R., Hyman JM., Yan P., Chowell G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, 5, 256-263. doi:10.1016/j.idm.2020.02.002.
7. Li L., Yang Z., Dang Z., Meng C., Huang J., Meng H., Wang D., Chen G., Zhang J., Peng H., Shao, Y. (2020). Propagation analysis and prediction of the COVID-19. *Infectious Disease Modelling*, 5, 282-292. Doi: 10.1016/j.idm.2020.03.002.

8. Mangoni L., Pistilli M. (2020). Epidemic analysis of Covid-19 in Italy by dynamical modelling. Available at SSRN: <https://ssrn.com/abstract=3567770>.
9. Jia W., Han K., Song Y., Cao W., Wang S., Yang S., Wang J., Kou F., Tai P., Li J., Liu M., He Y. (2020). Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *Frontiers in Medicine*. Accepted Paper. Doi: 10.3389/fmed.2020.00169.
10. Perc M., Miksić NG., Slavinec M., Stožer, A. (2020). Forecasting COVID-19. *Frontiers in Physics*, 8, 127. doi: 10.3389/fphys.2020.00127.
11. Batista, M. (2020). Estimation of the final size of the COVID-19 epidemic. *medRxiv*. doi: 10.1101/2020.02.16.20023606.
12. Hethcote HW. (1976). Qualitative analyses of communicable disease models. *Mathematical Biosciences*, 28(3-4), 335-356. doi: 10.1016/0025-5564(76)90132-2.
13. Shen M., Peng Z., Xiao Y., Zhang L. (2020). Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. *bioRxiv*. doi: 10.1101/2020.01.23.916726.
14. Bilge AH., Samanlioglu F., Ergonul, O. (2015). On the uniqueness of epidemic models fitting a normalized curve of removed individuals. *Journal of mathematical biology*, 71(4), 767-794. doi: 10.1007/s00285-014-0838-z.
15. Tang B., Wang X., Li Q., Bragazzi NL., Tang S., Xiao Y. Wu J. (2020). Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *Journal of Clinical Medicine*, 9(2), 462. doi: 10.3390/jcm9020462.
16. Read, JM., Bridgen JR., Cummings DAT., Ho A., Jewell CP. (2020). Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *MedRxiv*. doi: 10.1101/2020.01.23.200118549.
17. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html> [Accessed April 27, 2020]
18. Samanlioglu F., Bilge AH., Ergonul O. (2012). A Susceptible-Exposed-Infected-Removed (SEIR) model for the 2009-2010 A/H1N1 epidemic in Istanbul. *arXiv preprint arXiv:1205.2497*.
19. Dobie AP., Demirci A., Bilge AH., Ahmetolan S. (2019). On the time shift phenomena in epidemic models. *arXiv preprint arXiv:1909.11317*.
20. Bilge AH., Samanlioglu F. (2018). Determination of epidemic parameters from early phase fatality data: A case study of the 2009 A (H1N1) pandemic in Europe. *International Journal of Biomathematics*, 11(02), 1850021. doi: 10.1142/S1793524518500213.