Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Dynamics of COVID-19 using inverse problem for coefficient identification in SIR epidemic models

Tchavdar T. Marinov\textsuperscript{a,}\textsuperscript{*}, Rossitza S. Marinova\textsuperscript{b}

\textsuperscript{a} Department of Natural Sciences, Southern University at New Orleans, 6801 Press Drive, New Orleans, LA 70126, U.S.A.
\textsuperscript{b} Department of Mathematical and Physical Sciences, Concordia University of Edmonton, 7128 Ada Boulevard, Edmonton, AB, T5B 4E4, CANADA

\section*{A R T I C L E  I N F O}

Article history:
Received 22 June 2020
Accepted 10 July 2020
Available online 15 July 2020

\section*{A B S T R A C T}

This work deals with the inverse problem in epidemiology based on a SIR model with time-dependent infectivity and recovery rates, allowing for a better prediction of the long term evolution of a pandemic. The method is used for investigating the COVID-19 spread by first solving an inverse problem for estimating the infectivity and recovery rates from real data. Then, the estimated rates are used to compute the evolution of the disease. The time-dependent parameters are estimated for the World and several countries (The United States of America, Canada, Italy, France, Germany, Sweden, Russia, Brazil, Bulgaria, Japan, South Korea, New Zealand) and used for investigating the COVID-19 spread in these countries.

\section*{1. Introduction}

The COVID-19 coronavirus appeared in late 2019 and quickly spread across many countries. According to [14] and [37], by June 21, 2020, there were more than 9 million confirmed cases of infected people, with more than 470,000 reported deaths globally. Governments closed the so-called non-essential businesses and services for weeks in order to slow down the growth of infections – especially among vulnerable populations – and thus, save lives. Numerous mathematical models are developed to forecast the future of the COVID-19 epidemic spread worldwide. They are used to assist governments in making decisions to cope with the virus and its consequences. As of June 2020, there are no vaccines for this highly contagious disease; nor efficient antiviral drugs.

Mathematical modeling and forecasting the spread of epidemic diseases has a long history, see [3], [5], [10], and [29]. The earliest published paper on mathematical modeling of spread of disease was carried out in 1766 by Daniel Bernoulli. Trained as a physician, Bernoulli created a mathematical model to defend the practice of inoculating against smallpox [7], [15]. Infectious diseases include measles, malaria, varicella, HIV, Ebola, and SARS. A systematic review of the risk of death associated with Middle East respiratory syndrome (MERS) as well as risk factors for associated complications is given in [25]. An analysis and forecast of COVID-19 spreading in China, Italy and France was given in [11]. There is no vaccination for some of the infectious diseases, only preventive practices – see [16].

Kermack and McKendrick [17] introduced their model in 1921. Their theory is a mathematical hypothesis proposed to explain the rapid rise and fall in the number of infected people with a contagious illness in a closed population over time. It is the origin of SIR (Susceptibles–Infected–Recovered) type models. This formalism is the basis of all current modeling of the dynamics and evolution of infectious diseases, see [4], [26], [33], [34], and [35]. The models are useful in understanding the basic principles of the system, even in choosing a proper policy of infectious disease control. COVID-19 attracted attention in recent publications [12], [19], [27], [31], [32], and [39].

The dynamical systems for epemics are highly nonlinear – they include diverse characteristics, from population specifics to the immune system of an individual. Every new epidemics must be studied and mathematical models constructed to answer important questions on handling the disease. The parameters in the SIR model, the rate of transmission $a$ and the rate of removal $b$, depend on the evolution of the epidemic disease over time, see [1].

Complex and realistic mathematical models are used to assist policy decision making. Recent relevant works include [13]. This work aims to create a method that can accurately identify the time dependent parameters of the SIR system using real data and then use the computed parameter values to predict the spread of the epidemics. If some conditions change, then the projection based
on the “old” data will no longer be accurate and will require adjustment.

The parameter estimation is referred to as the inverse modeling problem [6]. It means adjusting the parameters of a mathematical model to reproduce measured data [23]. Model deficiencies are usually due to inaccurate parameters and must always be addressed. The inverse problem is crucial for calibrating the model and for controlling the model parameters. Approaches involving inverse problems can be successfully applied to a variety of important processes, including the spread of infectious diseases, allowing epidemiologists and public health specialists to make predictions on epidemics [5].

The present work extends the method proposed in [22] for finding optimum values for the infectivity and recovery rates. In [22], these values are assumed to be constant over the whole interval because we applied the method to a short outbreak of influenza instead of considering the long term evolution of a pandemic as with the case of COVID-19. Here, we assume the infectivity and recovery rates are functions of time, namely \( a = a(t) \) and \( b = b(t) \).

2. Problem formulation

2.1. The direct problem

The SIR model is undoubtedly the most famous mathematical model for the spread of an infectious disease. A constant population of size \( N \) is divided into three classes: susceptible \( S \), infectives \( I \) and removed \( R \). Removed individuals \( R \) are no longer susceptible nor infectious for whatever reason; for example, they have recovered from the disease and are now immune, or they have been vaccinated, or perhaps they have died from the disease. Since the isolated infectious people are still capable of infecting individuals from class \( S \) (see for example the big number of infected medical personnel), we count these people as members of the class \( I \) [9].

The diagram for the SIR model is given on Fig. 1.

![Diagram of SIR epidemic model](image)

There exist other models considering additional events and features, such as birth and death rates (non-constant population), vaccine effect, reinfection, latent period, and so on. In this work, we use the SIR model, because the considered time period is relatively short.

We assume that the population under study is well mixed so that every person has equal probability of coming into contact with every other person. This major assumption does not hold in many situations [26], e.g. sexually transmitted diseases and COVID-19 social distancing.

Let \( a \Delta t \) be the probability that a random infectious individual infects a random susceptible individual during the time period \( \Delta t \). Then with \( S(t) \) susceptible and \( I(t) \) infectious population, the expected number of newly infectious individuals in the total population \( N \) during time \( \Delta t \) is \( aS(t)\Delta t \), i.e., the rate of change of the class \( S(t) \) is \( -aSI \) where \( a > 0 \).

We also assume that infectious individuals leave the \( I(t) \) class with rate \( b \) and they move directly into the \( R(t) \) class. There are reports for recovered from COVID-19 individuals who are re-infected; hence, we added a dashed arrow from the class \( R \) to the class \( I \).

Since the number of these cases at the present moment is very limited, we cannot estimate the rate \( c \) and we do not include this option into the system.

\[
\frac{dS}{dt} = -aSI \tag{1}
\]

\[
\frac{dI}{dt} = aSI - bl. \tag{2}
\]

\[
\frac{dR}{dt} = bl. \tag{3}
\]

Since the differential equations (1) and (2) do not depend on \( R \), it is convenient to split the system into two parts – equations (1) and (2) as one system, and equation (3) by itself. If the coefficients \( a \) and \( b \) are given, we can derive initial conditions from the given data and solve the problem numerically by any of the well known methods for initial value problems (IVPs). We refer the reader to [2] for more information on methods for solving IVPs numerically.

The main question is how to find the coefficients \( a \) and \( b \) for a given infection, e.g. COVID-19. The problem for the estimation of the constants \( a \) and \( b \) is an inverse problem solved in [22]. A similar approach for identifying coefficients in an Euler-Bernoulli equation from over-posed data is used in [21] and [24].

2.2. The inverse problem

We assume that (from the available data) the values of \( S \) and \( I \) are known at two time moments, initial time moment \( T_I \) and final time moment \( T_F \),

\[
S(T_I) = S_I, \quad I(T_I) = I_I, \tag{4}
\]

and

\[
S(T_F) = S_F, \quad I(T_F) = I_F. \tag{5}
\]

If the coefficients \( a \) and \( b \) are known, the solution of equations (1) and (2) can be determined using the initial conditions (4). In general, the terminal conditions (5) are not satisfied exactly. Therefore, if the coefficients \( a \) and \( b \) are given, the problem is overdetermined.

Let us assume that the coefficients \( a \) and \( b \) are constant and unknown. In this case, the general solution of the system (1), (2) depends on four constants – two constants from the integration and the unknown coefficients \( a \) and \( b \). The number of conditions/equations in (4) and (5) is also four; hence, the problem for identifying the coefficients along with the functions \( S \) and \( I \) is well-posed, provided that an exact solution exists (or can be obtained). This type of problems belongs to the class of the so-called inverse problems.

Still, the problem for obtaining \( (S, I) \) and \( (a, b) \) from equations (1) and (2) under the conditions (4) and (5) could also be ill-posed. For arbitrary values of \( S_I, S_F, I_I, I_F \), there may be no solution \( (S, I, \ a, \ b) \) satisfying the equations (1), (2) and all of the conditions in (4), (5). For this reason, we assume that the problem is posed correctly after Tikhonov [36], i.e., it is known a priori that a solution of the problem exists. In other words, the boundary data have “physical meaning” and, therefore, a solution exists.

We are now ready to construct an algorithm for approximating the solution \( (S, I) \), \( (a, b) \) of the inverse problem (1), (2), (4), (5). For estimating the coefficients \( a \) and \( b \) of the SIR model, we use an approach that is very similar to the variational method used in [22]. This method is a generalization of the Least Squares Method. More details about the method for transforming the inverse problem into a correct direct problem can be found in [8], [20], [21], [24].

2.3. Coefficients \( a \) and \( b \) depend on time

The original SIR model assumes that the coefficients \( a \) and \( b \) are constants. In real live situations, e.g. the COVID-19 pandemic,
the coefficients vary in time, i.e., \( a = a(t) \) and \( b = b(t) \). The coefficients are affected by governments’ restrictions, societal experiences, treatment, among others. A similar idea was proposed in [28].

Let \( D \) be a data set of function values \( S(t_k), l(t_k) \) at some time moments \( t_1, t_2, \ldots. \) We assume there exists an algorithm for solving the inverse problem (1), (2), given the conditions (4), (5) and the data set \( D \). Here, we consider \( a \) and \( b \) to be piece-wise constant (step) functions of time

\[
a(t) = a_k \text{ and } b(t) = b_k \text{ for } t_{k-1} < t < t_k.
\]

The constants \( a_k \) and \( b_k \) can be estimated by solving the inverse problem (1), (2), (4), (5) using \( S_0 = S(t_{k-1}) \), \( S_F = S(t_k) \), \( I_0 = l(t_{k-1}) \), and \( T_F = t_k \), for \( k = 2, 3, \ldots \)

2.4. The reproduction rate and the herd immunity

Epidemics grow if the derivative \( dI/dt > 0 \) and decrease if \( dI/dt < 0 \). An important parameter, which characterizes infectious diseases, is the reproduction rate (also called reproduction number or ratio) of the infection

\[
\rho(t) = \frac{a(t)S(t)}{b(t)}.
\]

From equation (2) and (7), it follows that

\[
dI/dt = aSI - bl = \rho b - bl = (\rho - 1)bl.
\]

Consequently, the epidemic grows if \( \rho(t) > 1 \) and decreases if \( \rho(t) < 1 \). The reproduction rate \( \rho(t) \) represents the expected number of secondary infections produced by a single primary infected person. The reproduction rate in Hubei Province (China) using case-report data from January 11 to February 6, 2020 was estimated in [18] as 4.1192 (with 95% confidence interval 4.0473-4.1912). The reproduction rate \( \rho(t) \) also relates to the fraction of the population that gets sick.

Another important characteristics of an infectious disease is the so-called herd immunity – the minimum fraction of the population, \( f \), that is required to have immunity in order to prevent an epidemic [30]. The stability analysis of the SIR model shows that \( f = 1 - 1/\rho \). The society can acquire herd immunity in two ways. The “natural” way is to let the epidemic spread until the required fraction of the population, \( f \), obtains immunity. The most common “non-natural” way is vaccination. Since vaccines are not presently available for COVID-19, some countries, e.g. Sweden, tried the “natural” way. As of this moment, their approach is questionable, according to the results for their pandemic spread, see Section 4.7.

Many countries have been practicing social distancing and other measures to slow down the transmission rate \( a \) of the epidemic.

3. Solving the inverse problem for the time dependent infectivity and recovery rates

A method for solving the inverse problem in the case of constant coefficients \( a \) and \( b \) is given in [22], along with a discussion of some theoretical aspects. Here, we present the numerical algorithm for estimating the coefficients if they depend on time, i.e., \( a = a(t) \) and \( b = b(t) \). The inverse problem for the time dependent infectivity and recovery rates follows the idea of the method used in [24]. The original problem is replaced by a minimization problem. For the correct embedded problem, a difference scheme and a numerical algorithm can be constructed.

3.1. Minimization problem using data values at two time moments

First, we start with the sub-problem in case of using data values at two time moments. Since we are concerned with the numerical solution of the system (1–3), we seek approximations of

\[
T_F \quad t_0 \quad t_1 \quad t_2 \quad t_3 \quad \ldots \quad t_{n-3} \quad t_{n-2} \quad t_{n-1} \quad t_n.
\]

the functions \( S(t), l(t), R(t) \) at the discrete set \( \{t_0, t_1, \ldots, t_n\} \) of points in the interval \( [T_0, T_F] \), where \( T_0 \) is the initial time moment, \( T_F \) is the final time moment, and \( n \) is an integer greater than 1. The mesh of equidistant points is shown in Fig. 2.

We define the step size as \( \tau = \frac{T_F - T_0}{n} \). The nodes are the equidistant points \( t_k = T_0 + k\tau, \) \( k = 0, 1, \ldots, n \). Now we are ready to discretize equations (1) and (2)

\[
S_k = S_{k-1} - \tau a\tilde{S}_k\tilde{I}_k,
\]

\[
l_k = I_{k-1} + \tau (a\tilde{S}_k\tilde{I}_k - b_k),
\]

where \( S_k = (\tilde{S}_{k-1} + \tilde{S}_k)/2, \) \( I_k = (\tilde{I}_{k-1} + \tilde{I}_k)/2 \) for \( k = 1, 2, \ldots, n \).

This discretization secures the second order of approximation \( O(\tau^2) \). Since the problem (8)-(9) is non-linear, we use an iterative procedure, assuming that \( S_k \) and \( I_k \) are given from the previous iteration. Consider the function

\[
\Phi(a, b, S_1, S_2, \ldots, S_n, I_1, I_2, \ldots, I_n) = \frac{1}{2} \sum_{k=1}^{n-1} \left((S_k - S_{k-1})^2 + (I_k - I_{k-1} - \tau (a\tilde{S}_k\tilde{I}_k - b_k))^2\right).
\]

for \( k = 1, 2, \ldots, n-1 \). Adding the initial and terminal conditions (4) and (5), we obtain a well-posed linear system of \( 2(n+1) \) equations for the unknown sets of values \( \{S_0, S_1, S_2, \ldots, S_n\} \) and \( \{I_0, I_1, I_2, \ldots, I_n\} \).

3.1.1. Equations for \( S \) and \( I \)

The necessary conditions for minimization of the function \( \Phi \) with respect to its arguments \( S_k \) and \( I_k \) are

\[
\frac{\partial \Phi}{\partial S_k} = 0, \quad \frac{\partial \Phi}{\partial I_k} = 0.
\]

Conditions (11) yield the following difference equations

\[
S_{k-1} - 2S_k + S_{k+1} = \tau a(S_k\tilde{I}_k - S_{k+1}\tilde{I}_{k+1}),
\]

\[
l_{k-1} - 2I_k + I_{k+1} = \tau (-a\tilde{S}_k\tilde{I}_k + a\tilde{S}_{k+1}\tilde{I}_{k+1} + b_k - b_{k+1}).
\]

for \( k = 1, 2, \ldots, n-1 \). Adding the initial and terminal conditions (4) and (5), we obtain a well-posed linear system of \( 2(n+1) \) equations for the unknown sets of values \( \{S_0, S_1, S_2, \ldots, S_n\} \) and \( \{I_0, I_1, I_2, \ldots, I_n\} \).

3.1.2. Equations for \( a \) and \( b \)

We rewrite the function \( \Phi \) in the form

\[
\Phi = \alpha_0 \alpha + \alpha_10{\alpha} + \alpha_0{b} + \alpha_{20}{a^2} + \alpha_{11}ab + \alpha_{02}b^2,
\]

where

\[
\alpha_0 = \sum_{k=1}^{n-1} (l_k - I_{k-1})^2 + (S_k - S_{k-1})^2.
\]
The necessary conditions for minimization of the function $\Phi$ with respect to $a$ and $b$ are as follow

$$ \frac{\partial \Phi}{\partial a} = \alpha_{10} + 2\alpha_{20}a + \alpha_{11}b = 0, \quad (21) $$

$$ \frac{\partial \Phi}{\partial b} = \alpha_{01} + \alpha_{11}a + 2\alpha_{02}b = 0. \quad (22) $$

The solution of the system $(21), (22)$ is

$$ a = -\frac{2\alpha_{02}\alpha_{10} - \alpha_{01}\alpha_{11}}{-\alpha_{11}^2 + 4\alpha_{02}\alpha_{20}}, \quad (23) $$

$$ b = -\frac{\alpha_{10}\alpha_{11} - 2\alpha_{01}\alpha_{20}}{-\alpha_{11}^2 + 4\alpha_{02}\alpha_{20}}. \quad (24) $$

3.2. Minimization problem using the entire dataset

Let us assume that the number of infectious individuals at some time moments $\nu_1, \nu_2, \ldots, \nu_m$ is known and given by

$$ I(\nu_l) = \sigma_l, \quad (25) $$

for $l = 1, 2, \ldots, m$, while the values of the coefficients $a$ and $b$ are unknown.

Suppose that for every $1 \leq l \leq m$, there exists index $k_l$, such that $\nu_l = t_{k_l}$, i.e., the set of time moments $\{\nu_1, \nu_2, \ldots, \nu_m\}$ is a sub-
set of the set of mesh nodes \( \{t_0, t_1, t_2, \ldots, t_n\} \). To simplify the calculations, let us introduce the notations \( \chi_k \) such as: if there exists \( k \in \{0, 1, 2, \ldots, n\} \) such that \( t_k = t_l \), \( 1 \leq l \leq m \), then \( \chi_k = \sigma_l \) and \( \mu_k > 0 \), otherwise \( \chi_k = 0 \) and \( \mu_k = 0 \), where \( \mu_k \) is the weight of equation (25) in the LSM.

Similarly to Subsection 3.1, consider the function

\[
\Phi(a, b, S_1, \ldots, S_{n-1}, I_1, \ldots, I_{n-1}) = \sum_{k=1}^{n-1} \left[ \epsilon_k^2 + \delta_k^2 + \mu_k (I_k - \sigma_l)^2 \right] 
\]

\[
= \sum_{k=1}^{n-1} \left( S_k - S_{k-1} + \tau a \tilde{S}_k \tilde{I}_k \right)^2 
+ \left( I_k - I_{k-1} - \tau (a \tilde{I}_k \tilde{I}_k - b \tilde{I}_k) \right)^2 + \mu_k (I_k - \chi_k)^2 \right]. \tag{26}
\]

where \( \epsilon_k \) and \( \delta_k \) are the residuals of the equations (8) and (9), respectively.

### 3.2.1. Equations for \( S \) and \( I \)

The necessary conditions for minimizations of the function \( \Phi \) with respect to \( S_k \) and \( \tilde{I}_k \) are given by equations (11). For the case under consideration, we obtain

\[
S_{k-1} - 2S_k + S_{k+1} = \tau a (\tilde{S}_k \tilde{I}_k - \tilde{S}_{k+1} \tilde{I}_{k+1}). \tag{27}
\]

\[
I_{k-1} - (2 + \mu_k)I_k + I_{k+1} = \tau (-a \tilde{I}_k \tilde{I}_k + a \tilde{S}_k \tilde{I}_k + b \tilde{I}_k \tilde{I}_k) - \mu_k \chi_k, \tag{28}
\]

for \( k = 1, 2, \ldots, n-1 \). Adding the initial and terminal conditions (4), (5), we obtain a well-posed linear system with \( (n+1) \) equations for the unknown set of values \( (S_0, S_1, S_2, \ldots, S_n) \) and \( (I_0, I_1, I_2, \ldots, I_n) \).

#### 3.2.2. Equations for \( a \) and \( b \)

The equations for \( a \) and \( b \) in this case are the same as the equations (23) and (24), derived in Subsection 3.1.2.

### 3.3. Algorithm for solving the inverse problem

We use the following approach for solving the system (12), (13), (4), (5):

i) With given initial values for \( \tilde{S}, \tilde{I}, a, b \), the system (12), (13), (4), (5) is solved for the functions \( S \) and \( I \).

ii) The deviation of the values of \( S \) and \( I \) from \( \hat{S} \) and \( \hat{I} \) are computed as:

\[
\left\| \frac{S - \hat{S}}{S} \right\| < \varepsilon_0, \quad \left\| \frac{I - \hat{I}}{I} \right\| < \varepsilon_0. \tag{29}
\]

If they are smaller than a given tolerance \( \varepsilon_0 \), then the algorithm proceeds to iii); otherwise, \( \hat{S} \) and \( \hat{I} \) are replaced by \( S \) and \( I \), respectively, and the calculations return to i);

iii) The coefficients \( a \) and \( b \) are computed from (23) and (24). If the difference between the new and old values of \( a \) and \( b \)

\[
|a - \hat{a}| \leq \varepsilon_0, \quad |b - \hat{b}| \leq \varepsilon_0, \tag{30}
\]

is less than \( \varepsilon_0 \) then the calculations terminate; otherwise, the iterations continue, go back to i).

We use \( \varepsilon_0 = 5 \cdot 10^{-10} \) in all calculations presented here.
3.4. Parameters $a$, $b$, and $\rho$ as functions of time

The data set (25) contains the number of infected individuals for every day for a period of $m$ days. We divide the set into subsets of fixed length of $P$ days ($\text{say } P = 7, \text{ or } 10, \text{ or } 14$ days), as shown in Fig. 3. Next, we apply the algorithm presented in subsection 3.3 for every sub-interval $[k−P+1, k]$, for $k = P, P + 1, \ldots, m$. Consequently, we obtain $a_k$ and $b_k$ as functions of time. For every sub-interval we calculate the value of the reproduction rate

$$\rho_k = \frac{a_k S_k}{b_k}. \quad (31)$$

4. Data mining and projections

Models can help us understand what would be the worst case scenario and plan proper actions to achieve the best possible outcome under the given constraints. They cannot accurately predict what will happen.

The forecasts of the epidemic spread are based on the available COVID-19 data as of June 21, 2020, at [37] and [38]. Our experiments indicate that the optimal value of $\varepsilon_0$ is $10^{-8}$. The parameters $a, b,$ and $\rho$ as functions of time are computed with $\tau = 1/40$ and $\mu = \tau$. They are estimated with calculations performed over 7, 10, and/or 14 day intervals. The dynamic of the COVID-19 disease is projected based on the values of the parameters taken at the last day of the period used for the transmission are recovery rates estimation.

This section presents our results for the COVID-19 pandemic. In particular, the following characteristics are important for understanding the infectious disease dynamics:

- Estimated up-to-date transmission rate $a(t)$;
- Estimated up-to-date removal rate $b(t)$;
- Estimated up-to-date reproduction rate $\rho(t)$;
- Projected COVID-19 infectious cases $I(t)$ using the up-to-date estimated rates, taken at the last time period, for which data is available.

The functions $I(t)$, $S(t)$, and $R(t)$ satisfy the direct problem, if the coefficients are known. After estimating $a$ and $b$, we find the numerical solution of the direct problem (1)–(3) with proper initial conditions in order to project the future behavior of the epidemic. In all calculations, we use second order Runge–Kutta method, a type of predictor-corrector method. The first step is:

$$S_k^p = S_{k−1} − \tau a_{k−1} S_{k−1}, \quad (32)$$

$$I_k^p = I_{k−1} + \tau a_{k−1} S_{k−1} − \tau b_{k−1}, \quad (33)$$

$$R_k^p = R_{k−1} + \tau b_{k−1}. \quad (34)$$

Next, the corrector step improves the prediction obtained in the first step:

$$S_k^c = S_k^p − \tau a_{k} I_{k−1}^p S_{k−1}, \quad (35)$$

$$I_k^c = I_k^p + \tau a_{k−1}^p S_{k−1}^p − \tau b_{k} I_{k−1}^p, \quad (36)$$

$$R_k^c = R_k^p + \tau b_{k−1}^p. \quad (37)$$

Finally,

$$S_k = 0.5(S_k^p + S_k^c), \quad I_k = 0.5(I_k^p + I_k^c), \quad R_k = 0.5(R_k^p + R_k^c). \quad (38)$$
for $k = 1, 2, \ldots$. We use $\tau = 0.01$ for solving the direct problem numerically by the Runge–Kutta method.

It must be noted that these forecasts are valid as far as the rates remain as estimated; however, since COVID-19 is a highly contagious epidemic, for which very little is known, the rates of transmission, removal, and reproduction may suddenly change due to measures and/or events affecting the spread. Over the last several months, people learned how to protect themselves from being infected; consequently, this helped keep the transmission and reproduction rates low.

4.1. The world

Fig. 4 (a)–(c) presents the obtained values of the parameters $a$, $b$, and $\rho$ as functions of time, calculated over 7, 10, and 14 day intervals. The values of $a$, $b$, and $\rho$ for the last day are given in Table 1. Forecasts, based on the last 7, 10, and 14 day interval estimations of $a$ and $b$ are given in Fig. 4(d). The reproduction rate $\rho$ is still greater than one although decreasing.

| days  | $a$      | $b$       | $\rho$    |
|-------|----------|-----------|-----------|
| 7     | 5.2162527e-12 | 0.0289885 | 1.3993350 |
| 10    | 5.2443698e-12 | 0.0295486 | 1.3802095 |
| 14    | 5.0981280e-12 | 0.0284969 | 1.3912376 |

4.2. The United States of America (USA)

Fig. 5 (a)–(c) presents the obtained values of the parameters $a$, $b$, and $\rho$ as functions of time, calculated over 7, 10, and 14 day intervals. The values of $a$, $b$, and $\rho$ for the last day are given in Table 2. The projections, based on the last 7, 10, and 14 day interval estimations of $a$ and $b$ are given in Fig. 5(d). The transmission rate $a$ and the reproduction rate $\rho$ have increased lately.

4.3. Canada

Fig. 6 (a)–(c) presents the obtained values of the parameters $a$, $b$, and $\rho$ as functions of time, calculated over 7, 10, and 14 day intervals. The values of $a$, $b$, and $\rho$ for the last day are given in Table 3. The projections, based on the last 7, 10 and 14 day interval estimations of $a$ and $b$ are given in Fig. 6(d). Canada started its re-opening some services and businesses in several provinces around May 14, 2020. However, during the month of June, the transmission rate $a$ and the reproduction rate $\rho$ steadily decrease.
Fig. 8. Estimated values of $a$, $b$, $\rho$ used for projections of $I$ for France.

Fig. 9. Estimated values of $a$, $b$, $\rho$ used for projections of $I$ for Germany.
Fig. 10. Estimated values of $a$, $b$, $\rho$ used for projections of $I$ for Sweden.

Fig. 11. Estimated values of $a$, $b$, $\rho$ used for projections of $I$ for Russia.
Fig. 12. Estimated values of $a$, $b$, $\rho$ used for projections of $I$ for Brazil.

Fig. 13. Estimated values of $a$, $b$, $\rho$ used for projections of $I$ for Bulgaria.
4.4. Italy

Fig. 7 (a)–(c) presents the obtained values of the parameters $a$, $b$, and $\rho$ as functions of time, calculated over 7, 10, and 14 day intervals. The values of $a$, $b$, and $\rho$ for the last day are given in Table 4. The projections, based on the last 7, 10, and 14 day interval estimations of $a$ and $b$ are given in Fig. 7(d). The data shows that the epidemic is slowing down significantly and it may disappear soon.

4.5. France

Fig. 8 (a)–(c) presents the obtained values of the parameters $a$, $b$, and $\rho$ as functions of time, calculated over 7, 10, and 14 day intervals. The values of $a$, $b$, and $\rho$ for the last day are given in Table 5. Three projections, based on the last 7, 10 and 14 day interval estimations of $a$ and $b$ are given in Fig. 8(d). The reproduction rate has been less than one since the middle of May, but, the end of May it is again greater than one and the epidemic spread is increasing.

4.6. Germany

Fig. 9 (a)–(c) presents the obtained values of the parameters $a$, $b$, and $\rho$ as functions of time, calculated over 7, 10, and 14 day intervals. The values of $a$, $b$, and $\rho$ for the last day are given in Table 6. The projected values, based on the last 7, 10, and 14 day interval estimations of $a$ and $b$ are given in Fig. 9(d). The reprodu-
...tion rate has been less than one since the middle of April, leading to decrease of the epidemic spread. For the last week the reproduction rate tends to increase.

4.7. Sweden

Fig. 10 (a)–(c) presents the obtained values of the parameters \( a, b, \) and \( \rho \) as functions of time, calculated over 7, 10, and 14 day intervals. The values of \( a, b, \) and \( \rho \) for the last day are given in Table 7. Three forecasts, based on the last 7, 10 and 14 day interval estimations of \( a \) and \( b \) are given in Fig. 10(d). The authors could not find detailed data for Sweden after May 30.

4.8. Russia

Fig. 11 (a)–(c) presents the obtained values of the parameters \( a, b, \) and \( \rho \) as functions of time, calculated over 7, 10, and 14 day intervals. The values of \( a, b, \) and \( \rho \) for the last day are given in Table 8. Three projections, based on the last 7, 10 and 14 day interval estimations of \( a \) and \( b \) are given in Fig. 11(d). Russia flattened the curve of the reproduction rate \( \rho \), with values being around one lately.

4.9. Brazil

Fig. 12 (a)–(c) presents the obtained values of the parameters \( a, b, \) and \( \rho \) as functions of time, calculated over 7, 10, and 14 day intervals. The values of \( a, b, \) and \( \rho \) for the last day are given in Table 9. Three projections, based on the last 7, 10, and 14 day interval estimations of \( a \) and \( b \) are given in Fig. 12(d). The transmission rate \( a \) decreases; however, the reproduction rate \( \rho \) is still greater than one.

4.10. Bulgaria

Fig. 13 (a)–(c) presents the obtained values of the parameters \( a, b, \) and \( \rho \) as functions of time, calculated over 7, 10, and 14 day intervals. The values of \( a, b, \) and \( \rho \) for the last day are given in Table 10. Three projections, based on the last 7, 10, and 14 day interval estimations of \( a \) and \( b \) are given in Fig. 13(d). Bulgaria flattened the curve of the reproduction rate \( \rho \), with values being around one lately.

| Table 7 | The values of \( a, b, \) and \( \rho \) for the last day under consideration (Sweden). |
|---------|-----------------------------------------------|
| days    | \( a \)     | \( b \)     | \( \rho \)     |
| 7       | 1.9300015e-09| 0.0021713  | 8.9453266     |
| 10      | 2.2363626e-09| 0.0024685  | 9.1161725     |
| 14      | 2.1259462e-09| 0.0020484  | 10.4443928    |

| Table 8 | The values of \( a, b, \) and \( \rho \) for the last day under consideration (Russia). |
|---------|-----------------------------------------------|
| days    | \( a \)     | \( b \)     | \( \rho \)     |
| 7       | 2.3654099e-10| 0.0315732  | 1.0891498     |
| 10      | 2.4217924e-10| 0.0339433  | 1.0372133     |
| 14      | 2.4741041e-10| 0.0334696  | 1.0746127     |

| Table 9 | The values of \( a, b, \) and \( \rho \) for the last day under consideration (Brazil). |
|---------|-----------------------------------------------|
| days    | \( a \)     | \( b \)     | \( \rho \)     |
| 7       | 2.9974405e-10| 0.0462894  | 1.3700484     |
| 10      | 3.2268384e-10| 0.0525649  | 1.2988252     |
| 14      | 3.2543792e-10| 0.0405913  | 1.4778407     |
Table 10
The values of $a$, $b$, and $\rho$ for the last day under consideration (Bulgaria).

| days | $a$       | $b$       | $\rho$       |
|------|-----------|-----------|--------------|
| 7    | 8.1879008e-09 | 0.0285207 | 1.9937457    |
| 10   | 9.3700495e-09 | 0.0290925 | 2.2367532    |
| 14   | 9.244636e-09  | 0.0349212 | 1.8384449    |

Fig. 16. Estimated values of $a$, $b$, and $\rho$ for New Zealand.

intervals. The values of $a$, $b$, and $\rho$ for the last day are given in Table 10. The projections, based on the last 7, 10, and 14 day interval estimations of $a$ and $b$ are given in Fig. 13(d). The reproduction rate $\rho$ was less than one for a while, however, is started to increase about two weeks ago.

4.11. Japan

Fig. 14 (a)–(c) presents the obtained values of the parameters $a$, $b$, and $\rho$ as functions of time, calculated over 7, 10, and 14 day intervals. The values of $a$, $b$, and $\rho$ for the last day are given in Table 11. The projections, based on the last 7, 10, and 14 day interval estimations of $a$ and $b$ are given in Fig. 14(d). The reproduction rate became less than one early May, causing the COVID-19 spread to decrease.

4.12. South Korea

Fig. 15 (a)–(c) presents the obtained values of the parameters $a$, $b$, and $\rho$ as functions of time, calculated over 7, 10, and 14 day intervals. The values of $a$, $b$, and $\rho$ for the last day are given in Table 12. The projections, based on the last 7, 10, and 14 day interval estimations of $a$ and $b$ are given in Fig. 15(d). The reproduction rate became greater than one during the last three weeks.

4.13. New Zealand

Fig. 16 (a)–(c) presents the obtained values of the parameters $a$, $b$, and $\rho$ as functions of time, calculated over 7, 10, and 14 day intervals. New Zealand has only one active case since May 29. Therefore, we present at Fig. 16(d) the reported data and the numerical results obtained with our model. Results show that COVID-19 is disappearing in New Zealand.

Table 11
The values of $a$, $b$, and $\rho$ for the last day under consideration (Japan).

| days | $a$       | $b$       | $\rho$       |
|------|-----------|-----------|--------------|
| 7    | 5.4674305e-10 | 0.0843019 | 0.8201525    |
| 10   | 4.784543e-10  | 0.0902789 | 0.6701985    |
| 14   | 4.0639329e-10 | 0.0763858 | 0.6727947    |
5. Comparisons of reproduction rates $\rho$

There are similarities between the epidemic spread for some countries. Fig. 17 presents the obtained values of the reproduction rate $\rho$ as function of time, calculated over 7 day intervals, for four European countries: France; Germany; Italy; and Bulgaria. The obtained values of the reproduction rate $\rho$ as function of time, also calculated over 7 day intervals, for the world, the United States of America, Russia, and Brazil are shown in Fig. 18.

6. Conclusion

The inverse problem for estimating the time-dependent transmission and removal rates in the SIR epidemic model is derived and solved. The minimization problem uses the entire dataset with data available on June 21, 2020 for estimating the non-constant rates. The obtained numerical results demonstrate that the transmission and removal rates and the unknown functions are accurately estimated. The numerically computed rates are used for forecasting the COVID-19 pandemic for the world and a number of countries. The results of this research give insight of the pandemic in parts of the world and could help in determining policy. The SIR model is a good choice for the short period of time of this epidemic; however, it possesses known limitations in case of a long term infectious disease. In future, we plan to use other models. Depending on future developments of the disease, we may consider models addressing non-constant population, latency, reinfection, and vaccine.

Declaration of Competing Interest

No conflict of interests.

References

[1] Acero L, Morao J-A, Santonja F-J, Villanueva RJ. A deterministic model for highly contagious diseases: the case of varicella. Phys A 2016;450:278–86.
[2] Ackleh AS, Karfourt RB, Allen EJ. Classical and modern numerical analysis: theory, methods and practice. Chapman & Hall/CRC Numerical Analysis and Scientific Computing. CRC Press; 2009.
[3] Anderson RM, May RM. Infectious diseases of humans. Oxford UK: Oxford University Press; 1991.
[4] Azam S, Macías-Díaz JE, Ahmed N, Khan I, Iqbal MS, Rafiq M, et al. Numerical modeling and theoretical analysis of a nonlinear advection-reaction epidemic system. Comput Methods Programs Biomed 2020;193:105429.
[5] Bauch C, Earn D. Interepidemic intervals in forced and unforced SEIR models. In: Ruan S, Wolkowicz GS, Wu J, editors. Dynamical systems and their applications in biology. New York: American Mathematical Society; 2003. p. 33–43.
[6] Bellomo N, Preziosi L. Modelling mathematical methods and scientific computation. CRC Press Inc; 1995.
[7] Bernoulli D, Blower S. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. Rev Med Virol 2004;14:275–88.
[8] Christov CI, Marinov TT. Identification of heat-conduction coefficient via method of variational imbedding. Math Comput Model 1996;27(3):109–16.
[9] Coroescu A, Huanca F, Sepúlveda M. A note on the existence and stability of an inverse problem for a SIS model. Comput Model Appl 2019;77(12):3186–94.
[10] Diekmann O, Heesterbeek JAJ. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. New York: John Wiley & Sons, Incorporatted; 2000.
[11] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in china. Italy France ChaosSoliton Fractal 2020;134:109761. doi: 10.1016/j.chaos.2020.109761.
[12] Ferguson N.M., et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand at spiral. 2020. 10.25561/77482.
[13] Jin X, Jia J. Qualitative study of a stochastic SIRS epidemic model with information intervention. Phys A 2020;547:123865.
[14] Coronavirus resource center. Johns Hopkins University & Medicine. https://coronavirus.jhu.edu/.
[15] Hethcote HW. The mathematics of infectious diseases. SIAM 2000;42:599–653.
[16] Arifel KM, Kabir KK, Tanino J. Analysis of SIR epidemic model with information spreading of awareness. Chaos Soliton Fractals 2019;119:118–25.
[17] Kermack WC, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond Ser A 1927;115:700–21.
[18] Kim S, Choi S, Ko Y, Kim M, Jung E. Risk estimation of the SARS-cov-2 acute respiratory disease outbreak outside china. Theor Biol Med Model 2020;17(9), doi:10.1186/s12976-020-00127-6.
[19] Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study Lancet Infect Dis.; 2020. Online First. doi:10.1016/S1473-3099(20)30144-4.
[20] Marinov TT, Christov CI, Marinova RS. Novel numerical approach to solitary-wave solutions identification of booussinesq and korteweg–de vries equations. Int J Bifurcat Chaos 2005;15(2):557–65.
[21] Marinov TT, Marinova RS. Coefficient identification in euler–bernuoli equation from over–posed data. J Comput Appl Math 2010;235(2):450–9.
[22] Marinov TT, Marinova RS, Omojola J, Jackson M. Inverse problem for coefficient identification in SIR epidemic models. Comput Math Appl 2014;67:2218–27.
[23] Marinov TT, Marinova RS, Christov CI. Coefficient identification in elliptic partial differential equation. In: Likor JW, Margenov S, editors. Lecture Notes in Computer Science 3743. Springer; 2006. p. 372–9.
[24] Marinov TT, Vatsala A. Inverse problem for coefficient identification in euler–bernuoli equation. Comput Math Appl 2008;56(2):400–10.
[25] Matsuyma R, Nishiura H, Kutsuna S, Hayakawa K, Ohmagari N. Clinical determinants of the severity of middle east respiratory syndrome (MERS): a systematic review and meta-analysis. BMC Public Health 2016;16(1203). doi: 10.1186/s12889-016-3881-4.
[26] Murray JD. Mathematical biology i an introduction. Rev. ed. of: Mathematical biology. 2nd ed. c. Murray JD, editors. Berlin; 2006.
[27] Nishiura H, Linton NM, Akhevatshyan AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 2020;93:284–6.
[28] Reis RF, de Melo Quintela B, de Oliveira Campos J, Comes JM, Rocha BM, Lobosco M, dos Santos RW. Characterization of the COVID-19 pandemic and the impact of uncertainties. Mitigad Strateg Underreport Cases South Korea Italy Brazil Chaos Solitons Fractals 2020;136:109888. doi: 10.1016/j.chaos.2020.109888.
[29] Rhodes JA, Allman ES. Mathematical models in Biology: An Introduction. New York: Cambridge UP; 2003. Pp. 280–301.
[30] Rodphong P, Auewarakul P. Viral evolution and transmission effectiveness. World J Virol 2012;1(5):131–4.
[31] Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, Yan P, Chowell G. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th. 2020. Infect Dis Model 2020;5:256–63.

[32] Sheeren MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. J Adv Res 2020;24:91–8.

[33] Smith D, Moore L. The SIR model for spread of disease. Convergence; 2004.

[34] Takeuchi Y, Iwasa Y, Sato K. Mathematics for life science and medicine. Springer-Verlag Berlin Heidelberg; 2007.

[35] Tarantola A. Inverse problem theory and methods for model parameter estimation. SIAM; 2005.

[36] Tikhonov AN, Arsenin V. Methods for solving incorrect problems. Nauka, Moscow; 1974.

[37] Novel Coronavirus-(ncov) situation reports - world health organization (WHO); 2019. https://www.who.int/worldometerCOVID-19.

[38] COVID-19 coronavirus pandemic. https://www.worldometers.info/coronavirus/.

[39] Li Y, Wang B, Peng R, Zhou C, Zhan Y, Liu Z, Jiang X, Zhao B. Mathematical modeling and epidemic prediction of COVID-19 and its significance to epidemic prevention and control measures. Ann Infect Dis Epidemiol 2020;5(1):1052.