Mathematical modelling of the second wave of COVID-19 infections using deterministic and stochastic SIRD models

Fran Sérgio Lobato · Gustavo Barbosa Libotte · Gustavo Mendes Platt

Received: 30 December 2020 / Accepted: 28 June 2021 / Published online: 7 July 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract Recently, various countries from across the globe have been facing the second wave of COVID-19 infections. In order to understand the dynamics of the spread of the disease, much effort has been made in terms of mathematical modeling. In this scenario, compartmental models are widely used to simulate epidemics under various conditions. In general, there are uncertainties associated with the reported data, which must be considered when estimating the parameters of the model. In this work, we propose an effective methodology for estimating parameters of compartmental models in multiple wave scenarios by means of a dynamic data segmentation approach. This robust technique allows the description of the dynamics of the disease without arbitrary choices for the end of the first wave and the start of the second. Furthermore, we adopt a time-dependent function to describe the probability of transmission by contact for each wave. We also assess

the uncertainties of the parameters and their influence on the simulations using a stochastic strategy. In order to obtain realistic results in terms of the basic reproduction number, a constraint is incorporated into the problem. We adopt data from Germany and Italy, two of the first countries to experience the second wave of infections. Using the proposed methodology, the end of the first wave (and also the start of the second wave) occurred on 166 and 187 days from the beginning of the epidemic, for Germany and Italy, respectively. The estimated effective reproduction number for the first wave is close to that obtained by other approaches, for both countries. The results demonstrate that the proposed methodology is able to find good estimates for all parameters. In relation to uncertainties, we show that slight variations in the design variables can give rise to significant changes in the value of the effective reproduction number. The results provide information on the characteristics of the epidemic for each country, as well as elements for decision-making in the economic and governmental spheres.

F. S. Lobato
Chemical Engineering Faculty, Federal University of
Uberlândia, Uberlândia, Brazil
e-mail: fslobato@ufu.br

G. B. Libotte (✉)
National Laboratory for Scientific Computing, Petrópolis,
Brazil
e-mail: glibotte@lncc.br

G. M. Platt
Graduate Program in Agroindustrial Systems and
Processes, School of Chemistry and Food, Federal
University of Rio Grande, Santo Antônio da Patrulha,
Brazil
e-mail: gmplatt@furg.br

Keywords Mathematical modeling of COVID-19 · Second wave · Inverse problem · Stochastic model · Uncertainty quantification

1 Introduction

Since the end of 2019, the world has been experiencing the consequences of the increase in cases of COVID-19.

In political, economic, and social context, various measures have been adopted distinctively by the countries, in an attempt to mitigate the effects of the pandemic [45, 54]. After a significant increase in the number of infections and deaths, the proposed measures had an effect to a certain extent, causing the number of new infections to be reduced in some locations. However, the relaxation of such measures, in view of the supposed control of the spread of the disease, caused the number of cases to rise again [20, 29, 34]. Other factors, such as the identification of reinfections and the seasonal immunity, can also be associated with such an increase [15, 25, 33, 53]. In the sense of understanding the burden caused by the spread of diseases, numerous studies in different fields of science have been conducted, among which we may mention especially the development and improvement of mathematical models [42], environmental studies [3, 19, 30, 48, 49] and geopolitical aspects [46].

Some relevant works that analyze the early stages of the spread of the disease have helped to understand important aspects of the pandemic. In this context, Perc et al. [42] proposed a flexible model to predict the spread of COVID-19 based on minimal and maximal growth rates (for the number of infected individuals) in a certain period for the USA, Slovenia, Iran, and Germany. Hâncean et al. [21] studied the initial spread of COVID-19 in Romania, presenting an analysis of imported cases (mainly Romanian citizens returning from Italy) using human-to-human transmission networks. Ribeiro et al. [46] analyzed the relationship between the size of cities and the spread of COVID-19 in Brazil, using power-law models relating the size of the population and the number of cases/deaths. The results indicate that small cities are more affected in the early stages of the pandemic; on the other hand, large cities are more favorable environments for the spread of SARS-CoV-2 in the long term.

The analysis of the effects of environmental factors and the use of spatial data infrastructure, such as the Geographic Information System (GIS), have also been explored in research on the spread of COVID-19. Roher et al. [48] considered the influence of weather scenarios—for instance, dust of particulate matter—in the spread of COVID-19 in some locations in Europe, such as Paris, London, and the Canary Islands. The influence of aerosols on the infection fatality rate of COVID-19 in Germany, Italy, and Spain was studied by Li et al. [30], focusing on the aerosol optical depth.

In the same vein, Gupta et al. [19] analyzed the effect of weather on the spread of the disease in India with a detailed statistical approach. Using a support vector machine approach, Gupta et al. [19] established an association between daily transmission and weather parameters. Al-Kindi et al. [3] employed GIS techniques to investigate the spatiotemporal pattern of the COVID-19 pandemic in the territory of Oman. Stofel et al. [49] analyzed the effect of the pandemic on carbon dioxide emissions during 2020, presenting an interesting comparison between solar radiation management and an antipyretic medicine (referring to global warming as a fever). The authors warn that, until the end of the pandemic, governments must do even more to reduce greenhouse gas emissions.

The so-called second wave of a pandemic is characterized by an expressive increase in the number of cases (and consequently also in the number of deaths caused by the disease), after a significant drop in the number of new infections during the first wave. Research shows that the virus associated to COVID-19 has been mutating [44]. However, the occurrence of new cases, as well as a possible increase in the severity of the disease or in the transmissibility, is not yet associated with such biological alteration [14]. According to Ghanbari [18], the conditions for the occurrence of a second wave of infections are due to the increase in the number of cases within specific groups, such as people in the same age-group, in specific places, or even the persistence of the spread of the first wave.

Vaccines have been widely studied since the onset of the pandemic [4, 7, 17, 32]. Several countries have started vaccination, but so far, only a few have achieved high vaccination coverage, as the UK [28] and Israel [35]. Libotte et al. [31] proposed a strategy to determine an optimal control strategy for vaccine administration in COVID-19 pandemic considering a hypothetical scenario in relation to the Chinese population. Mukandavire et al. [39] estimated the basic reproduction number and critical vaccination coverage to control the disease for different hypothetical vaccine efficacy scenarios in South Africa. However, some countries face a gradual increase in the number of cases of the disease, while vaccination is still a long way off. Priesemann et al. [45] claimed for a pan-European effort to mitigate the effects of further waves of COVID-19, aiming at (i) reducing the number of cases, (ii) keeping the number of cases at a low level, and (iii) develop a

longer-term common vision to guide the management of the pandemic.

In view of the threat of a new wave of infections, several efforts have been concentrated on trying to predict their behavior and thus provide mechanisms for decisions aimed at mitigating the disease. In this context, Cacciapaglia et al. [8] proposed a robust approach based on epidemic renormalization group framework to simulate the second wave for different countries. The authors conducted statistical analyses on different levels of human interaction to obtain an estimate of the occurrence of the second wave in Europe, which would most likely take place between July 2020 and January 2021. Nori Junior et al. [40] proposed a two-wave statistical model based on the superposition of normal distributions, where future scenarios can be predicted based on parameters estimated for the first wave. Castro [9] proposed a SEIR-based model with time-dependent transmission and mortality rates to represent the second wave considering data from Spain and Germany. The proposed model was employed to simulate the post-confinement epidemic under several scenarios. Ghanbari [18] proposed a SID-based model with two sub-populations (which are classified by immunological criteria) of infected individuals, in order to simulate the second wave in Iran. Pedro et al. [41] proposed a SEIR model that considers a time-varying parameter to evaluate the influence of schools and workplaces closure on the transmission of SARS-CoV-2.

Usually, such models and related studies do not consider the uncertainties associated with simulations and data. In this context, stochastic models have been widely used as a way to incorporate such uncertainties into computational simulations. Chanu and Singh [10] studied a stochastic SEQIR model to analyze the epidemic in India. In turn, He et al. [22] proposed a discrete-time stochastic SEIHR model considering the governmental measures to control the spread of the disease, as well as the clinical condition of individuals. Rihan et al. [47] incorporated a time delay in the stochastic SIRC model. Faranda and Alberti [16] proposed a stochastic SEIR model to analyze the impact of the relaxation of isolation measures in France and Italy. Tang et al. [51] proposed a stochastic SEIHRQ model to investigate the imported cases on local epidemic in the Shaanxi-China province. This model takes into account the effects of intermittent inflow with a Poisson distribution on the likelihood of multiple outbreaks.

Typically, compartmental models have constant parameters, which may not be sufficiently descriptive, affecting the goodness-of-fit statistics and, consequently, the long-term simulations. In addition, such models may not be able to describe the behavior of the second wave of an epidemic, without taking advantage of arbitrary choices. The major contributions of this work are to provide a generalized methodology that is capable of identifying the interface between two epidemiological curves, so that compartmental models with time-dependent parameters can fit to more complex epidemiological data accurately. Our goal is to determine the parameters of the SIDR model that best describe the behavior of the second wave in Germany and Italy. We adopt a piecewise time-dependent transmission rate for each wave. The determination of the parameters of the model considers a constraint related to the effective reproduction number, in order to provide more realistic results. In addition, we propose to analyze the uncertainties of the simulations using a stochastic SIDR model, taking into account the incorporation of additional differential equations to represent the variation of some parameters. This work is organized as follows. The description of the SIDR model is presented in Sect. 2, considering the deterministic approach and a stochastic version of the model. The formulation of the proposed inverse problem is presented in Sect. 3. Section 4 describes the data employed, the numerical methods used to solve the deterministic and stochastic problems and the definition of some error metrics—used to evaluate the performance of the proposed approach. The results obtained by solving the inverse problem are presented and discussed in Sect. 5. Finally, conclusions are drawn in Sect. 6.

2 Epidemiological models

In this section we describe the deterministic and stochastic versions of the SIDR compartmental model used in the mathematical modelling, as well as the definition of the initial conditions and some considerations regarding the basic reproduction number and the effective reproduction number. For the stochastic SIDR model, we present the stochastic differential equations that describe the uncertainties in the control parameters.

2.1 Deterministic SIRD model

In order to simulate the dynamic behavior of COVID-19 epidemic, we adopt the SIRD model. This model considers the interaction among individuals in the population, which are classified as susceptible, infected, dead and recovered individuals. We employ a normalized version of the model, in the sense that the quantities S , I , D and R are divided by the population size N . The variations of S , I , D and R over time in this model are represented by

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - (\alpha + \gamma) I \\ \frac{dD}{dt} &= \gamma I \\ \frac{dR}{dt} &= \alpha I.\end{aligned}\quad (1)$$

Parameters β , α and γ represent, respectively, the transmission rate, death rate, and recovery rate. The initial conditions are $D(0) = 0$, $R(0) = 0$, $I(0) = \mathcal{I}(0)/\varphi$, and $S(0) = 1 - D(0) - R(0) - I(0)$, since the constraint given by $S(t) + I(t) + D(t) + R(t) = 1$ holds for all t . The term $\mathcal{I}(0)$ represents the first element in the vector $\mathcal{I}$ of cases reported (i.e., the number of infected individuals at $t = 0$), as we will discuss in Section 3. In turn, φ is a quantity related to the population size and the fraction of cases reported. We use a new variable to represent $I(0)$,

$$\delta = \frac{\mathcal{I}(0)}{\varphi}, \quad (2)$$

since this parameter will be an argument of the optimization procedure in the inverse problem.

In order to predict the potential spreading of an infectious disease in a population, the basic reproduction number ($\mathcal{R}_0$) may be considered as an epidemiological metric. This parameter represents the average number of secondary cases resulting from a single case, taking into account a population totally susceptible to the disease [5]. When $\mathcal{R}_0 > 1$, the disease has a spreading potential, with a tendency to increase the number of new cases.

The basic reproduction number for the SIRD model is given by [5]

$$\mathcal{R}_0 = \frac{\beta}{\alpha + \gamma}. \quad (3)$$

It can be a large value ($\mathcal{R}_0 \gg 1$) when $\alpha + \gamma$ is small compared to β . Thus, in the early stage of COVID-19 spread in different countries, we can observe high values for $\mathcal{R}_0$ considering different numerical approaches [8, 24, 26, 27, 36, 43, 57]. As mentioned by Marimuthu et al. [36], this parameter encompasses the rate of contact for the individuals in the population, the probability of transmission by contact (not every contact is capable to transmit the disease) and the estimated period in which the population is affected by the disease. On the other hand, a population may not be entirely susceptible to the disease, as consequence of immunity, for instance. In practice, the effective (or time-dependent) reproduction number ($\mathcal{R}_t$) can provide more realistic results. $\mathcal{R}_t$ represents the expected number of secondary cases arising from a primary case of infection at time t [52].

To evaluate this parameter, various methodologies can be used, such as exponential growth and time-dependent methods [36]. In this work, we adopt $\mathcal{R}_t$ defined as [13]

$$\mathcal{R}_t = \mathcal{R}_0 S(t), \quad (4)$$

where $S(t)$ is the susceptible population at a given time t .

2.2 Stochastic SIRD model

In the context of modeling the dynamics of epidemics, there are several factors that incorporate uncertainty in the parameter estimation process. Considering such uncertainties means obtaining results that are consistent with the real scenario. In order to evaluate the influence of uncertainties in such epidemiological models, the control parameters can be stochastically perturbed [16] and, for this purpose, the original deterministic model is rewritten as a stochastic model, where the parameters that characterize the dynamics of the epidemic are defined by specific probability distributions. In this case, for each control parameter (in our case, β , γ and α), a new differential equation with stochastic contribution is formulated. Thus the following stochastic differential equations are employed [16]:

$$\frac{d\Phi_{un}}{dt} = -\Phi_{un} + \Phi_{un0} + \eta_{\Phi} \xi(t) \quad (5)$$

where Φ_{un} is the stochastic control parameter, which in our case corresponds to $\Phi_{un} \in \{\beta_{un}, \gamma_{un}, \alpha_{un}\}$. In turn, $\Phi_{un}(0) = \Phi_{un0}$ is the initial condition related to

the corresponding parameter, η_ϕ is the standard deviation for the stochastic control parameter, and $\xi(t)$ is a random number represented by a specific distribution type. In this case, the stochastic SIRD model is given by

$$\begin{aligned}\frac{dS}{dt} &= -\beta_{\text{un}}SI \\ \frac{dI}{dt} &= \beta_{\text{un}}SI - (\alpha_{\text{un}} + \gamma_{\text{un}})I \\ \frac{dD}{dt} &= \gamma_{\text{un}}I \\ \frac{dR}{dt} &= \alpha_{\text{un}}I \\ \frac{d\beta_{\text{un}}}{dt} &= -\beta_{\text{un}} + \beta_{\text{un}0} + \eta_\beta \xi(t) \\ \frac{d\gamma_{\text{un}}}{dt} &= -\gamma_{\text{un}} + \gamma_{\text{un}0} + \eta_\gamma \xi(t) \\ \frac{d\alpha_{\text{un}}}{dt} &= -\alpha_{\text{un}} + \alpha_{\text{un}0} + \eta_\alpha \xi(t)\end{aligned}\quad (6)$$

with appropriated initial conditions (discussed in Sect. 4.2.2).

3 Formulation of the inverse problem

In the SIRD model described by Eq. (1), the parameters β , γ , α and δ must be calculated in order to allow the numerical solution of the system of ordinary differential equations. It is important to mention that since δ is a quantity related to the effective number of infected individuals (see Eq. (2)), it is employed to define the initial conditions for susceptible, dead, and recovered individuals, i.e., $S(0) = 1 - I(0) = 1 - \delta$, $D(0) = 0$ and $R(0) = 0$, respectively. For this purpose, it is necessary to formulate and solve an inverse problem through the minimization of the difference between calculated values and reported data.

Commonly, parameters β , γ and α are taken to be constant during the integration of the differential system, for the sake of simplicity. However, according to Castro [9], both the transmission rate (β) and death rate (γ) can be chosen as time-dependent parameters to better capture the effects of policies defined to mitigate the pandemic. Following Cheynet [12], we define β as:

$$\beta(t) = \beta_0 \exp\left(-(\beta_1(t - \beta_2))^2\right), \quad (7)$$

where β_0 (day^{-1}), β_1 (day^{-1}) and β_2 (day) are parameters that define the transmission rate of a particular epidemic.

The other parameters of the model are considered to be constant. Thus, for each wave, the parameters β_0 , β_1 , β_2 , γ and α must be calculated. In addition, for the first wave, the quantity related to the initial value for the infected individuals (δ) must be estimated, as well as the time instant τ where each set of parameters are effective, i.e., the interface between the end of the first wave and the start of the second wave. The first wave is defined within $0 \leq t \leq \tau$, and the second wave holds for $\tau < t \leq t_f$, where t_f is the final time. It is important to mention that the model continuity is guaranteed at $t = \tau$.

We will define a vector of parameters, θ (to be estimated), as

$$\theta^T = (\beta_0^f, \beta_1^f, \beta_2^f, \tau, \gamma^f, \alpha^f, \delta, \beta_0^s, \beta_1^s, \beta_2^s, \gamma^s, \alpha^s)$$

where the superscripts f and s refer, respectively, to the first and the second waves.

The numerical solution of the system given by Eq. (1) for a discrete time set and a particular approximation of the vector θ produces vectors of responses for each compartment of the model, i.e., $\mathbf{S}$, $\mathbf{D}$, $\mathbf{I}$ and $\mathbf{R}$. We adopt reported data related to infected and dead individuals, which are respectively denoted by $\mathcal{I}$ and $\mathcal{D}$, in the parameter estimation procedure. Then, we are interested in an approximation of the vector θ that minimizes the error between the model responses $\mathbf{I}(\theta)$ and $\mathbf{D}(\theta)$, and the reported data, $\mathcal{I}$ and $\mathcal{D}$, respectively. In other words, we aim to minimize the quantity

$$\mathcal{F}(\theta) = \frac{\|\mathbf{I}(\theta) - \mathcal{I}\|^2}{(\max(\mathcal{I}))^2} + \frac{\|\mathbf{D}(\theta) - \mathcal{D}\|^2}{(\max(\mathcal{D}))^2}, \quad (8)$$

where $\max(\mathcal{I})$ and $\max(\mathcal{D})$ are the highest values in the vectors $\mathcal{I}$ and $\mathcal{D}$ (just for scaling purposes).

Furthermore, in practice, the basic reproduction number $\mathcal{R}_0$ and the effective reproduction number $\mathcal{R}_t$ cannot assume arbitrarily large values. Considering this situation, we introduce an inequality constraint to ensure that the solution of the inverse problem will represent the adherence to the observed values for $\mathcal{R}_t$ in the pandemic, as follows

$$\frac{\beta(t)}{\gamma + \alpha} S(t) \leq \mathcal{R}_{\max}, \quad (9)$$

which holds $\forall t \in 0 \leq t \leq t_f$. $S(t)$ is the simulated normalized susceptible population at time t and $\mathcal{R}_{\max}$

is the maximum value that the effective reproduction number can take (which is further discussed). We also consider that $\beta = \beta(t)$, according to Eq. (7), indicating the time dependency for β . Obviously, the use of $\mathcal{R}_0$ or even $\mathcal{R}_t$ as a parameter to analyze the spread of the disease must be carefully conducted [2]. The main idea here is to avoid unrealistic values for $\mathcal{R}_t$.

Mathematically, the inverse problem is formulated as [38]:

$$\theta^* = \arg \min_{\theta} \mathcal{F}(\theta) \quad (10)$$

subject to Eq. (9), where θ^* is the optimal value of θ .

In this case, the normalized SIRD model must be simulated considering the parameters calculated by the optimizer, in order to obtain the number of infected and dead individuals estimated by the model and, consequently, the value of the objective function ($\mathcal{F}$). It is important to mention that the constrained optimization problem is transformed into an equivalent unconstrained problem by using the static penalty function method [55].

4 Data, numerical algorithms and error metrics

4.1 Data

The data employed for parameter estimation are the daily number of infected individuals ($\mathcal{I}$) and the cumulative number of deaths ($\mathcal{D}$) caused by COVID-19 in Germany and Italy. We use 297 records for each time series, both ranging from February 02, 2020 to November 25, 2020 [56].

4.2 Numerical algorithms

Specific algorithms for deterministic and stochastic models are presented below.

4.2.1 Deterministic approach

Each inverse problem is solved by using the differential evolution algorithm [50] (DE) in the optimization step, considering the following parameters: population size $NP = 50$, crossover probability $CR = 0.8$, amplification parameter $F = 0.8$, maximum number of generations $G_{\max} = 1000$, and strategy rand/1/bin (see Storn and Price [50] for a detailed description of the

algorithm). The search space is given by $0 \leq \beta_0^{\{f,s\}} \leq 1$, $0 \leq \beta_1^{\{f,s\}} \leq 1$, $0 \leq \beta_2^{\{f,s\}} \leq 1$, $0 \leq \gamma^{\{f,s\}} \leq 1$, $0 \leq \alpha^{\{f,s\}} \leq 1$, $0 \leq \delta \leq 1$, and $80 \leq \tau \leq 200$. The evolutionary process is interrupted if a given maximum number of generations is reached.

To evaluate the inequality constraint, $\mathcal{R}_{\max}$ is considered to be equal to 10, based on recent studies regarding $\mathcal{R}_t$ in Germany [27]. The penalization parameter (required to use the static penalty function method) is equal to 10^5 . The deterministic SIRD model is numerically solved using the fourth-order Runge–Kutta method [6] considering 1000 equally spaced points. In addition, for each wave, the model is simulated for t_f equal to 297 days.

4.2.2 Stochastic approach

To solve a stochastic differential equation, several numerical methods with different characteristics can be found in the literature [23]. In this work, we use the Milstein method [37]. We do not estimate the parameters of the stochastic model, but only use it for simulation purposes, and therefore, the optimization procedure is not employed in this step. In order to assess the influence of uncertainties, we consider the results regarding the first wave, which are previously obtained using the deterministic approach. In this case, we investigate the sensitivity of the calculated parameters only considering the second wave (when $t > \tau$), whereas the first wave is simulated considering the parameters estimated by means of the deterministic approach. In practice, this means that, for the second wave, the uncertainties are associated only to β , γ , and α .

In this context, to simulate the stochastic SIRD model given by Eq. (6) for the second wave, the initial conditions for S , I , D and R are taken as the simulated values at $t = \tau$ considering the estimated parameters for the first wave (this strategy ensures the continuity of the waves). In addition, the initial conditions for β_{un} , γ_{un} and α_{un} are defined as the average values for the first wave (denoted as $\bar{\beta}^f$, $\bar{\gamma}^f$, and $\bar{\alpha}^f$, respectively).

For each parameter, it is required to define a type of distribution to represent the stochastic contribution. As suggested by Faranda and Alberti [16], γ_{un} and α_{un} are normally distributed, and β_{un} follows a lognormal distribution, defined as

Table 1 Results obtained for the proposed inverse problem using reported data from Germany. The values shown correspond to the best result and the standard deviation calculated from the computed values

First wave		Second wave	
β_0^f (day ⁻¹)	0.31349 ± 0.00244	β_0^s (day ⁻¹)	0.09517 ± 0.00425
β_1^f (day ⁻¹)	0.02190 ± 0.00122	β_1^s (day ⁻¹)	0.00013 ± 0.00075
β_2^f (day)	0.10140 ± 0.00989	β_2^s (day)	0.00033 ± 0.00003
γ^f (day ⁻¹)	0.00290 ± 0.00011	γ^s (day ⁻¹)	0.00058 ± 0.00001
α^f (day ⁻¹)	0.04556 ± 0.00352	α^s (day ⁻¹)	0.00854 ± 0.00043
δ	$1.91 \times 10^{-5} \pm 2.97 \times 10^{-8}$	τ (day)	166.18015 ± 5.96786

$$\begin{aligned}
 \gamma_{\text{un}} &\sim \mathcal{N}(\bar{\gamma}^f, \eta_\gamma) \\
 \alpha_{\text{un}} &\sim \mathcal{N}(\bar{\alpha}^f, \eta_\alpha) \\
 \beta_{\text{un}} &\sim \mathcal{LN}(\bar{\beta}^f - 0.5\kappa^2, \eta_\beta),
 \end{aligned} \tag{11}$$

where κ is a proportionality constant, and η_γ, η_α , and η_β are the corresponding standard deviations, which are essentially defined as a proportion of the average value of the parameter obtained in the first wave, that is, $\eta_\beta = \kappa$, $\eta_\gamma = \kappa\bar{\gamma}^f$ and $\eta_\alpha = \kappa\bar{\alpha}^f$. We also consider $\Delta t = 1$ in the Milstein method to simulate the process, with final time equal to 600 days and 100 independent runs to obtain the average values.

4.3 Error metrics

In order to compare the quality of solution, two metrics are considered in this work. The first is the coefficient of determination (Θ), used to determine the goodness-of-fit of the model. Mathematically, this metric is given by [11]

$$\Theta(Z) = 1 - \frac{\|\mathbf{Z} - \mathbf{Z}\|^2}{\|\mathbf{Z} - \mathbf{1}\bar{Z}\|^2} \tag{12}$$

where $\bar{Z}$ is the mean value of $\mathbf{Z}$ and $Z = I$ or D , $\mathbf{Z} = \mathbf{I}$ or $\mathbf{D}$, $\mathbf{Z} = \mathbf{I}$ or $\mathbf{D}$. The vector $\mathbf{1}$ represents a vector of ones with same length of $\mathbf{Z}$. In practice, the quantity Θ gives the percentage variation in Z compared to the set of variables considered.

The second metric represents the root-mean-squared error (Γ), which is calculated as [11]

$$\Gamma(Z) = \frac{\|\mathbf{Z} - \mathbf{Z}\|}{N^{1/2}}. \tag{13}$$

Similarly to Eq. (12), $Z = I$ or $Z = D$. This metric quantifies the differences between the reported data and the responses of the model.

5 Results

In this section, we present the results for the proposed inverse problem considering the deterministic approach, as well as the analyzes regarding the stochastic approach, for both Germany and Italy.

5.1 Deterministic Inverse Problem

The first problem addressed is the solution of the deterministic inverse problem for the two waves. The results are shown below.

5.1.1 Germany

Initially, we are interested in verifying the model's ability to capture information about both waves using the proposed approach. To that end, we run DE for 20 times, in such a way that each new result obtained by solving the deterministic inverse problem is independent of the others, that is, the initial population is different in each run, and the randomness of the genetic operators guarantees the diversity of the results. This strategy ensures that favorable results are not obtained by chance. Table 1 presents the best result and the corresponding standard deviation for each estimated parameter. In view of the value of the standard deviation corresponding to each parameter, it is clear that the proposed methodology is robust in the sense of estimating the model parameters for the analyzed data set. The

Fig. 1 **a** Simulation of the SIRD model with parameters obtained from the solution of the inverse problem given by Eq. (10), using epidemiological data from Germany in the period between February 2 and November 25, 2020. The readings for compartments S ,

I and R are on the left and for compartment D are on the right; **b** Transmission rate of the disease in the same period, where the vertical blue dashed line represents the transition between the first and second waves

visual inspection of Fig. 1a, which shows the simulation of the model compared to the reported data, demonstrates that the set of parameters is in good agreement with the dynamics of the disease. The value obtained for the objective function, evaluated at the best estimator for θ (the best results in Table 1), is equal to $\mathcal{F} = 0.1324$. This corroborates the similarity between the profiles resulting from the integration of $\mathbf{I}$ and $\mathbf{D}$, with the analyzed data. The error metrics indicate that, considering the reported data on the number of infected individuals, $\Theta(I) = 0.99425$. For data on the number of dead individuals, $\Theta(D) = 0.99287$. In relation to the root-mean-squared error, Γ is equal to 546.89075 and 334.28045, respectively for infected and dead individuals.

Figure 1a shows the drop over time in the population of susceptible individuals, for each wave. In addition, between the two waves, it is observed a phase where this group of individuals remains nearly constant, i.e., the number of new infections between the end of the first wave and the start of the second wave remains very low. On the other hand, the number of recovered individuals increases during the epidemic period, as expected. Such results demonstrate that the system presents a behavior that is consistent with the dynamics of an epidemic.

Regarding β , whose behavior is shown in Fig. 1b, initially it assumes the maximum value within the analyzed range, since at the beginning of the epidemic, the number of immune individuals is very small and, therefore, new infections are more likely to occur. As the time goes by and the gradual mitigation of the disease, this value decreases, tending to zero, showing the proximity of the end of the first wave. During the second wave, this parameter is nearly constant since $\beta_1^s \rightarrow 0$ and, therefore, $\beta(t) \approx \beta_0^s$ (see Eq. 7). A possible criticism to this approach is related to the discontinuity in the profile of β at $t = \tau$. It must be borne in mind that this behavior is a consequence of the choice of a “double-wave” model, where two profiles of β are expected. Moreover, with this assumption we avoid discontinuities in the profiles for the individuals in the compartments.

The effective reproduction number $\mathcal{R}_t$, with $\mathcal{R}_0$ being calculated using the values shown in Table 1, is presented in Fig. 2. In general, for each wave, $\mathcal{R}_t$ decreases significantly compared to the initial value. The high value at the initial point represents the start of the epidemic period. Thus, as different measures are taken to mitigate the epidemic, the natural tendency of this parameter is to reduce until the epidemic is contained ($\mathcal{R}_t \rightarrow 0$). In addition, Fig. 2 shows the behav-

ior of $\mathcal{R}_t$ obtained by Petrova et al. [43] using the SID model, and by Khailaie et al. [27] using the SECIR model, both for the first wave. The results are quite similar, mainly in terms of the behavior of the curves. On average, we obtained $\mathcal{R}_t$ approximately equal to 1.5 for the first wave and 6.50 for the second wave. According to the Robert Koch Institute [1], preliminary information indicates that, for the second wave, this value is close to 3. Such difference may be due to type of approach used to determine each parameter. The value calculated here is obtained considering a particular set of reported data.

5.1.2 Italy

For the data from Italy, we adopt the same methodology described for the German case: we run the optimizer for 20 times and obtain a set of independent results, of which the best result and the standard deviation corresponding to each parameter are shown in Table 2. Again, it is clear that the simulated profiles, considering the best parameters obtained, have good agreement with the analyzed data, as illustrated in Fig. 3. Qualitatively, the results are similar to those obtained for the data from Germany (see Fig. 1), but differ in terms of the magnitude of the data and results. The error metrics point out that the fit between the responses of the model and the data is also satisfactory, with $\Theta(I) = 0.93987$ and $\Gamma(I) = 2678.75$ for data related to the number of infected individuals, and $\Theta(I) = 0.92678$ and $\Gamma(I) = 3555.69$ for dead individuals. Even with a greater deviation than in the results for the German case, the fitting still reflects the behavior observed in the data set fairly well.

In turn, Fig. 4 presents the behavior of $\mathcal{R}_t$ for each wave, considering the parameters shown in Table 2.

Fig. 2 Estimate of the effective reproduction number of the disease in Germany, in the period between February 2 and November 25, 2020. The values are calculated using Eq. (4) and considering the transmission rate shown in Fig. 1b. The results are compared with the estimates obtained with SECIR [27] and SID [43]

Again, our results are in good agreement with those obtained by Petrova et al. [43] and Khailaie et al. [27] using the SID and SECIR models, respectively. This shows that the proposed methodology is capable of obtaining good estimates for the model parameters and, consequently, for $\mathcal{R}_t$, as illustrated in Fig. 4. In this case, the average values regarding the effective reproduction number are, approximately, 1.31 and 4.7 for the first and second waves, respectively. This result is in line with that obtained by Faranda and Alberti [16] (the basic reproduction number $\mathcal{R}_0$ obtained by these authors is, approximately, equal to 4, using a stochastic SEIR model). Due to the stochastic nature of this model, the value of this parameter can fluctuate up to 6. In turn, Cacciapaglia et al. [8] reported that the maximum value that $\mathcal{R}_t$ can reach is close to 10 for the Italian case.

Table 2 Results obtained for the proposed inverse problem using reported data from Italy. The values shown correspond to the best result and the standard deviation calculated from the computed values

First wave		Second wave	
β_0^f (day ⁻¹)	0.47815 ± 0.02454	β_0^s (day ⁻¹)	0.78693 ± 0.02434
β_1^f (day ⁻¹)	0.02830 ± 0.00112	β_1^s (day ⁻¹)	0.00558 ± 0.00009
β_2^f (day)	0.38865 ± 0.06656	β_2^s (day)	0.17129 ± 0.02332
γ^f (day ⁻¹)	0.00502 ± 0.00012	γ^s (day ⁻¹)	0.00071 ± 0.00001
α^f (day ⁻¹)	0.05472 ± 0.00155	α^s (day ⁻¹)	0.02349 ± 0.00001
δ	$1.09 \times 10^{-6} \pm 1.64 \times 10^{-9}$	τ (day)	187.08035 ± 11.74343

Fig. 3 **a** Simulation of the SIRD model with parameters obtained from the solution of the inverse problem given by Eq. (10), using epidemiological data from Italy in the period between February 2 and November 25, 2020. The readings for compartments S , I

and R are on the left and for compartment D are on the right; **b** Transmission rate of the disease in the same period, where the vertical blue dashed line represents the transition between the first and second waves

Fig. 4 Estimate of the effective reproduction number of the disease in Italy, in the period between February 2 and November 25, 2020. The values are calculated using Eq. (4) and considering the transmission rate shown in Fig. 3b. The results are compared with the estimates obtained with SECIR [27] and SID [43]

5.2 Stochastic analysis

In order to analyze the influence of uncertainties, this section considers the simulation of the stochastic SIRD model. For this purpose, at first we consider the results obtained for the first wave, in both cases (Germany and Italy). Our focus is to analyze the dynamics of the population and the behavior of the parameters of the

model for the second wave, when $t > \tau$. The first wave is simulated with the same parameters obtained in the previous sections (Tables 1 and 2 for Germany and Italy, respectively). In practice, for the second wave, uncertainties are incorporated into the parameters β , γ , and α . The initial conditions for the stochastic SIRD model are those obtained for the first wave using the deterministic approach. In this sense, for Germany, we have $\beta_{\text{un}}(\tau) = 0.09491 \text{ day}^{-1}$, $\gamma_{\text{un}}(\tau) = 0.00058 \text{ day}^{-1}$, and $\alpha_{\text{un}}(\tau) = 0.00854 \text{ day}^{-1}$; for Italy, $\beta_{\text{un}}(\tau) = 0.14485 \text{ day}^{-1}$, $\gamma_{\text{un}}(\tau) = 0.00071 \text{ day}^{-1}$, and $\alpha_{\text{un}}(\tau) = 0.02349 \text{ day}^{-1}$. In addition, since β is a time-dependent function, its average value is taken in this analysis.

Figure 5 presents the average profiles and the effective reproduction number (only for the second wave) considering the influence of uncertainties, with different values for κ , where $\kappa \in \{0.0, 0.1, 0.2, 0.3, 0.4\}$, as defined in Eq. (11). The fluctuation of each curve in relation to the deterministic solution (where $\kappa = 0$ is the reference profile) is shown in Fig. 5a. For both Germany and Italy, the increase in the value of κ implies in more distancing with respect to the reference profile, i.e., the peak localization referring to the second wave gets further from the reference value. This variation also affects the maximum value observed for each

Fig. 5 **a** Stochastic simulations of the SIRD model for the COVID-19 epidemic in Germany and Italy. The variation of the mean of the distribution corresponding to the transmission rate in the second wave is considered, according to Eq. (11), by assign-

ing different values for κ ; **b** Variation of the effective reproduction number over time for each particular value of κ represented by the area in light blue. The curve shown in dark blue represents the average values for each time instant

peak. We may also note that a high value for κ at a given time represents an increase in the number of susceptible individuals (for both countries), and thus requires a proportional decrease in the populations of infected, recovered and dead individuals, according to the hypotheses of the SIRD model.

In addition, for both cases analyzed, $\mathcal{R}_t$ decreases over time (see Fig. 5b). Such profiles represent only the second wave and, therefore, it is expected a high value in the start of the wave (which will produce the second peak, as discussed previously). Furthermore, the

higher is the value of κ , the lower is the initial effective reproduction number for the second wave, which is consonant to the displacement of the second peak to the right in the number of infected individuals, i.e., with a more quantity of susceptible individuals—as a consequence of a higher value of κ —we obtain a higher value for $\mathcal{R}_t$ (for the same time instant t).

Figure 6 provides another perspective on the variation of the average profiles and the values of $\mathcal{R}_t$ in Germany and Italy considering uncertainties, as shown in Fig. 5. Initially, note that for $\kappa \in [0.1, 0.4]$, the mean

Fig. 6 Variation of the mean of the β_{un} distribution in relation to the variation of κ . The gray squares show the four values of κ used in the stochastic simulations (see Fig. 5) and, for each of

these values, 2000 samples of β_{un} are drawn, where the dashed horizontal lines represent the corresponding mean of the distribution ($\beta^f - 0.5\kappa^2$) for the particular value of κ

value of β_{un} is strictly decreasing, given the particular choice of β^f . Such monotonicity provides an indication of the behavior of the average profiles at a given time $t > \tau$: the number of infected and dead individuals tends to decrease, whereas the number of susceptible and recovered individuals increases, preserving the total size of the population in both countries, as κ increases. Clearly, the shift of the average profiles is reflected in the value of the effective reproduction number, as can be seen by comparing the corresponding frames in Fig. 5.

The behavior of the profiles for different values of κ in Fig. 5a is in agreement with the mean of the samples in Fig. 6. Even with the values of κ being equally spaced, the profiles in Fig. 5a do not shift linearly as κ increases. Note that $\eta_\beta = \kappa$ in Eq. (11) and, therefore, the mean and standard deviation of the samples are coupled and cannot be analyzed individually. In addition, β_{un} follows a lognormal distribution, which is not symmetrical around the mean value. Therefore, the behavior of the profiles is a consequence of the adopted distributions. Also note that values greater than $\kappa = 0.4$ can produce negative transmission rates (see the samples in frame 4 of Fig. 6).

6 Conclusions

In this work we proposed a framework to simulate the second wave of COVID-19 pandemic. We analyze data from Germany and Italy, which have well-defined behavior regarding the second wave. In addition, a sensitivity analysis considering a stochastic SIRD model is conducted. For this purpose, the parameters of the SIRD compartmental model, piece-wise defined (for each wave), are obtained using DE. The inverse problem considers a piece-wise time-dependent transmission rate for each wave and an inequality constraint is adopted, in order to avoid unrealistic values for the effective reproduction number. In general, the parameters obtained for the first wave, especially regarding the effective reproduction number, are in agreement with those reported by other authors.

In relation to uncertainty analysis, we can conclude that small perturbations in the design variables vector can result in significant variations in the value of the effective reproduction number. In this case, this fluctuation can help to understand different scenarios of the pandemic, as well as to define policies to mitigate the disease. Finally, the formulation of the inverse problem with an inequality constraint, associated with a piece-wise time-dependent transmission rate for each wave

and the sensitivity analysis represent the main contributions of this work.

Funding This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. Gustavo Libotte is supported by a postdoctoral fellowship from the Institutional Training Program (PCI) of the Brazilian National Council for Scientific and Technological Development (CNPq), grant number 303185/2020-1.

Availability of data and material The data used in this work are available at <https://www.worldometers.info/coronavirus/>.

Declarations

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest. The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Code availability The source code used to generate the results is publicly available in the repository: https://github.com/gustavolibotte/Modeling_second_wave_COVID-19.

References

1. Coronavirus disease 2019 (COVID-19)—daily situation report of the Robert Koch Institute. <https://corona.rki.de>. Accessed December 29, 2020
2. Adam, D.: A guide to R—what the reproduction number can and can't tell us about managing COVID-19. *Nature* **583**, 346–348 (2020). <https://doi.org/10.1038/d41586-020-02009-w>
3. Al-Kindi, K.M., Alkharusi, A., Alshukaili, D., Nasiri, N.A., Awadhi, T.A., Charabi, Y., El Kenawy, A.M.: Spatiotemporal assessment of COVID-19 spread over oman using GIS techniques. *Earth Syst. Environ.* **4**, 797–811 (2020). <https://doi.org/10.1007/s41748-020-00194-2>
4. Alturki, S.O., Alturki, S.O., Connors, J., Cusimano, G., Kutzler, M.A., Izmirly, A.M., Haddad, E.K.: The 2020 pandemic: Current SARS-CoV-2 vaccine development. *Front. Immunol.* **11**, 1880 (2020). <https://doi.org/10.3389/fimmu.2020.01880>
5. Anastassopoulou, C., Russo, L., Tsakris, A., Siettos, C.: Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE* **3**(15), 1–21 (2020)
6. Atkinson, K.A.: *An Introduction to Numerical Analysis*. Wiley, New York (1989)
7. Baviskar, T., Raut, D., Bhatt, L.K.: Deciphering vaccines for COVID-19: where do we stand today? *Immunopharmacol. Immunotoxicol.* (2020). <https://doi.org/10.1080/08923973.2020.1837867>
8. Cacciapaglia, G., Cot, C., Sannino, F.: Second wave COVID-19 pandemics in Europe: a temporal playbook. *Sci. Rep.* **10**, 15514–15522 (2020). <https://doi.org/10.1038/s41598-020-72611-5>
9. Castro, F.: Modelling of the Second (and Subsequent) Waves of the Coronavirus Epidemic. Spain and Germany as Case Studies. (2020) <https://doi.org/10.1101/2020.06.12.20129429>. MedRxiv preprint
10. Chanu, A.L., Singh, R.K.B.: Stochastic approach to study control strategies of COVID-19 pandemic in India. *Epidemiol. Inf.* **148**, 1–9 (2020). <https://doi.org/10.1017/S0950268820001946>
11. Cheng, C.L., Shalabh, G.G.: Coefficient of determination for multiple measurement error models. *J. Multivar. Anal.* **126**, 137–152 (2014). <https://doi.org/10.1016/j.jmva.2014.01.006>
12. Cheynet, E.: Generalized SEIR epidemic model (Fitting and Computation) (2020). <https://doi.org/10.5281/ZENODO.3911854>. Zenodo
13. Delamater, P.L., Street, E.J., Leslie, T.F., Yang, Y.T., Jacobsen, K.H.: Complexity of the basic reproduction number. *Emerg. Inf. Diseases* **25**, 1–4 (2019)
14. van Dorp, L., Richard, D., Tan, C.C., Shaw, L.P., Acman, M., Balloux, F.: No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun* (2020). <https://doi.org/10.1038/s41467-020-19818-2>
15. Edridge, A.W.D., Kaczorowska, J., Hoste, A.C.R., Bakker, M., Klein, M., Loens, K., Jebbink, M.F., Matser, A., Kinsella, C.M., Rueda, P., Ieven, M., Goossens, H., Prins, M., Sastre, P., Deijs, M., van der Hoek, L.: Seasonal coronavirus protective immunity is short-lasting. *Nat. Med.* (2020). <https://doi.org/10.1038/s41591-020-1083-1>
16. Faranda, D., Alberti, T.: Modelling the second wave of COVID-19 infections in France and Italy via a stochastic SEIR model. *Chaos Interdiscip. J. Nonlinear Sci.* **30**, 111101–111132 (2020). <https://doi.org/10.1063/5.0015943>
17. Funk, C.D., Laferrière, C., Ardakani, A.: A snapshot of the global race for vaccines targeting SARS-CoV-2 and the COVID-19 pandemic. *Front. Pharmacol.* **11**, 937 (2020). <https://doi.org/10.3389/fphar.2020.00937>
18. Ghanbari, B.: On forecasting the spread of the COVID-19 in Iran: The second wave. *Chaos Solitons Fractals* **140**, 110176–110184 (2020)
19. Gupta, A., Pradhan, B., Maulud, K.N.A.: Estimating the impact of daily weather on the temporal pattern of COVID-19 outbreak in India. *Earth Syst. Environ.* **4**, 523–534 (2020). <https://doi.org/10.1007/s41748-020-00179-1>
20. Han, E., Tan, M.M.J., Turk, E., Sridhar, D., Leung, G.M., Shibuya, K., Asgari, N., Oh, J., García-Basteiro, A.L., Hanefeld, J., Cook, A.R., Hsu, L.Y., Teo, Y.Y., Heymann, D., Clark, H., McKee, M., Legido-Quigley, H.: Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. *The Lancet* (2020). [https://doi.org/10.1016/S0140-6736\(20\)32007-9](https://doi.org/10.1016/S0140-6736(20)32007-9)
21. Hancean, M.G., Perc, M., Lerner, J.: Early spread of COVID-19 in Romania: imported cases from Italy and human-to-human transmission networks. *R. Soc. Open Sci.* **7**, 1–8 (2020). <https://doi.org/10.1098/rsos.200780>
22. He, S., Tang, S., Rong, L.: A discrete stochastic model of the COVID-19 outbreak: forecast and control. *Math. Biosci. Eng.* **17**, 2792–2804 (2020). <https://doi.org/10.3934/mbe.2020153>
23. Higham, D.J.: An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* **43**(3), 525–546 (2001)

24. Hong, H.G., Li, Y.: Estimation of time-varying reproduction numbers underlying epidemiological processes: a new statistical tool for the COVID-19 pandemic. *PLoS ONE* **e0236464**, 1–15 (2020). <https://doi.org/10.1371/journal.pone.0236464>
25. Iwasaki, A.: What reinfections mean for COVID-19. *Lancet Inf. Diseases* (2020). [https://doi.org/10.1016/S1473-3099\(20\)30783-0](https://doi.org/10.1016/S1473-3099(20)30783-0)
26. Jung, S.Y., Jo, H., Son, H., Hwang, H.J.: Real-world implications of a rapidly responsive COVID-19 spread model with time-dependent parameters via deep learning: Model development and validation. *J. Med. Internet Res.* **22**, 1–11 (2020). <https://doi.org/10.2196/19907>
27. Khailaie, S., Mitra, T., Bandyopadhyay, A., Schips, M., Mascheroni, P., Vanella, P., Lange, B., Binder, S., Meyer-Hermann, M.: Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for political measures. *BMC Med.* **19**(32) (2021). <https://doi.org/10.1186/s12916-020-01884-4>
28. Ledford, H., Cyranoski, D., Noorden, R.V.: The UK has approved a COVID vaccine—here's what scientists now want to know. *Nature* **588**(7837), 205–206 (2020). <https://doi.org/10.1038/d41586-020-03441-8>
29. Leung, K., Wu, J.T., Liu, D., Leung, G.M.: First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *The Lancet* **395**(10233), 1382–1393 (2020). [https://doi.org/10.1016/S0140-6736\(20\)30746-7](https://doi.org/10.1016/S0140-6736(20)30746-7)
30. Li, W., Thomas, R., El-Askary, H., Piechota, T., Struppa, D., Ghaffar, K.A.A.: Investigating the significance of aerosols in determining the coronavirus fatality rate among three european countries. *Earth Syst. Environ.* **4**, 513–522 (2020). <https://doi.org/10.1007/s41748-020-00176-4>
31. Libotte, G.B., Lobato, F.S., Platt, G.M., Neto, A.J.S.: Determination of an optimal control strategy for vaccine administration in COVID-19 pandemic treatment. *Comput. Methods Programs Biomed.* **196**, 105664–105677 (2020)
32. Liu, Y., Wang, K., Massoud, T.F., Paulmurugan, R.: SARS-CoV-2 vaccine development: An overview and perspectives. *ACS Pharmacol. Transl. Sci.* **3**(5), 844–858 (2020). <https://doi.org/10.1021/acspstsci.0c00109>
33. Long, Q.X., Tang, X.J., Shi, Q.L., Li, Q., Deng, H.J., Yuan, J., Hu, J.L., Xu, W., Zhang, Y., Lv, F.J., Su, K., Zhang, F., Gong, J., Wu, B., Liu, X.M., Li, J.J., Qiu, J.F., Chen, J., Huang, A.L.: Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**(8), 1200–1204 (2020). <https://doi.org/10.1038/s41591-020-0965-6>
34. López, L., Rodó, X.: The end of social confinement and COVID-19 re-emergence risk. *Nat. Human Behav.* **4**(7), 746–755 (2020). <https://doi.org/10.1038/s41562-020-0908-8>
35. Mallapaty, S.: Vaccines are curbing COVID: Data from israel show drop in infections. *Nature* **590**(7845), 197–197 (2021). <https://doi.org/10.1038/d41586-021-00316-4>
36. Marimuthu, S., Joy, M., Malavika, B., Nadaraj, A., Asirvatham, E.S., Jeyaseelan, L.: Modelling of reproduction number for COVID-19 in India and high incidence states. *Clin. Epidemiol. Global Health* **9**, 57–61 (2021). <https://doi.org/10.1016/j.cegh.2020.06.012>
37. Milstein, G.N.: Approximate integration of stochastic differential equations. *Theory Probab. Appl.* **19**, 557–600 (1975). <https://doi.org/10.1137/1119062>
38. Moura Neto, F.D., Silva Neto, A.J.: An Introduction to Inverse Problems with Applications. Springer, Berlin (2013). <https://doi.org/10.1007/978-3-642-32557-1>
39. Mukandavire, Z., Nyabadzha, F., Malunguza, N.J., Cuadros, D.F., Shiri, T., Musuka, G.: Quantifying early COVID-19 outbreak transmission in south Africa and exploring vaccine efficacy scenarios. *PLOS ONE* **15**(7), e0236003 (2020). <https://doi.org/10.1371/journal.pone.0236003>
40. Nori Junior, A.D., da Silva, B.A., Pizzol, F.D., Porto, L.M.: A two-wave Epidemiological Model of COVID-19 Outbreaks using MS-Excel (2020). <https://doi.org/10.1101/2020.05.08.20095133>. MedRxiv preprint
41. Pedro, S.A., Ndjomatchoua, F.T., Jentsch, P., Tchuente, J.M., Anand, M., Bauch, C.T.: Conditions for a second wave of COVID-19 due to interactions between disease dynamics and social processes. *Front. Phys.* **2020**, 1–9 (2020)
42. Perc, M., Miksic, N.G., Slavinec, M., Stozer, A.: Forecasting COVID-19. *Front. Phys.* **8**, 1–5 (2020). <https://doi.org/10.3389/fphy.2020.00127>
43. Petrova, T., DmitriSoshnikov, Grunin, A.: Estimation of time-dependent reproduction number for global COVID-19 outbreak. *Preprints* (2020). <https://doi.org/10.20944/preprints202006.0289.v1>
44. Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., Mirchandani, D., Scharton, D., Bilello, J.P., Ku, Z., An, Z., Kalveram, B., Freiberg, A.N., Menachery, V.D., Xie, X., Plante, K.S., Weaver, S.C., Shi, P.Y.: Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* (2020). <https://doi.org/10.1038/s41586-020-2895-3>
45. Priesemann, V., Brinkmann, M.M., Ciesek, S., Cuschieri, S., Czypionka, T., Giordano, G., Gurdasani, D., Hanson, C., Hens, N., Iftekhar, E., Kelly-Irving, M., Klimek, P., Kretzschmar, M., Peichl, A., Perc, M., Sannino, F., Schernhammer, E., Schmidt, A., Staines, A., Szczurek, E.: Calling for pan-European commitment for rapid and sustained reduction in SARS-CoV-2 infections. *Lancet* **397**, 92–93 (2021). [https://doi.org/10.1016/S0140-6736\(20\)32625-8](https://doi.org/10.1016/S0140-6736(20)32625-8)
46. Ribeiro, H.V., Sunahara, A.S., Sutton, J., Perc, M., Hanley, Q.S.: City size and the spreading of COVID-19 in Brazil. *PLoS ONE* **15**, 1–12 (2020). <https://doi.org/10.1371/journal.pone.0239699>
47. Rihan, F.A., Alsakaji, H.J., Rajivganthi, C.: Stochastic SIRC epidemic model with time-delay for COVID-19. *Adv. Differ. Equa.* **2020**(502), 1–20 (2020). <https://doi.org/10.1186/s13662-020-02964-8>
48. Rohrer, M., Flahault, A., Stoffel, M.: Peaks of fine particulate matter may modulate the spreading and virulence of COVID-19. *Earth Syst. Environ.* **4**, 789–796 (2020). <https://doi.org/10.1007/s41748-020-00184-4>
49. Stoffel, M., Stephenson, D.B., Haywood, J.M.: Antipyretic medication for a feverish planet. *Earth Syst. Environ.* **4**, 75–762 (2020). <https://doi.org/10.1007/s41748-020-00182-6>
50. Storn, R., Price, K.: Differential Evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* **11**(4), 341–359 (1997). <https://doi.org/10.1023/A:1008202821328>

51. Tang, S., Tang, B., Bragazzi, N.L., Xia, F., Li, T., He, S., Ren, P., Wang, X., Peng, Z., Xiao, Y., Wu, J.: Stochastic Discrete Epidemic Modeling of COVID-19 Transmission in the Province of Shaanxi Incorporating Public Health Intervention and Case Importation (2020). <https://doi.org/10.1101/2020.02.25.20027615>. MedRxiv preprint
52. Thompson, R.N., Stockwin, J.E., R. D. van Gaalen, J.A.P., Kamvar, Z.N., Demarsh, P.A., Dahlgvist, E., Li, S., Miguel, E., Jombart, T., Lessler, J., Cauchemez, S., Cori, A.: Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* **29**, 100356–100362 (2019). <https://doi.org/10.1016/j.epidem.2019.100356>
53. Tillett, R.L., Sevinsky, J.R., Hartley, P.D., Kerwin, H., Crawford, N., Gorzalski, A., Laverdure, C., Verma, S.C., Rossetto, C.C., Jackson, D., Farrell, M.J., Van Hooser, S., Pandori, M.: Genomic evidence for reinfection with SARS-CoV-2: a case study. *The Lancet Infectious Diseases* (2020). [https://doi.org/10.1016/S1473-3099\(20\)30764-7](https://doi.org/10.1016/S1473-3099(20)30764-7)
54. Tisdell, C.A.: Economic, social and political issues raised by the covid-19 pandemic. *Econ. Anal. Policy* **68**, 17–28 (2020). <https://doi.org/10.1016/j.eap.2020.08.002>
55. Vanderplaats, G.N.: Numerical Optimization Techniques for Engineering Design with Applications. McGraw-Hill, New York (1984)
56. Worldometer: Coronavirus Live Tracker. <https://www.worldometers.info/coronavirus/#countries>. Accessed December 29, 2020
57. Xu, C., Dong, Y., Yu, X., Wang, H., Tsamlag, L., Zhang, S., Chang, R., Wang, Z., Yu, Y., Long, R., Wang, Y., Xu, G., Shen, T., Wang, S., Zhang, X., Wang, H., Cai, Y.: Estimation of reproduction numbers of COVID-19 in typical countries and epidemic trends under different prevention and control scenarios. *Front. Med.* **14**, 613–622 (2020). <https://doi.org/10.1007/s11684-020-0787-4>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.