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

Fran Sérgio Lobato  
Federal University of Uberlandia: Universidade Federal de Uberlandia

Gustavo B Libotte (gilbott@lncc.br)  
National Laboratory for Scientific Computing  https://orcid.org/0000-0002-4583-6026

Gustavo M. Platt  
FURG: Universidade Federal do Rio Grande

Research Article

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

Posted Date: May 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-172625/v1

License: ☇️  📧 This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License

Version of Record: A version of this preprint was published at Nonlinear Dynamics on July 7th, 2021. See the published version at https://doi.org/10.1007/s11071-021-06680-0.
Mathematical modelling of the second wave of COVID-19 infections using deterministic and stochastic SIDR models

Fran Sérgio Lobato · Gustavo Barbosa Libotte · Gustavo Mendes Platt

Received: date / Accepted: date

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. The objective of this work is to propose a methodology for estimating parameters of compartmental models by means of a dynamic data segmentation approach. 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. 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.

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. Political, economic and social measures have been adopted distinctively by the countries, in an attempt to mitigate the effects of the pandemic [45]. 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 [18, 27, 31]. Other factors, such as the identification of reinfections and the seasonal immunity can also be associated with such an increase [14, 23, 30, 44].

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. Researches show that the virus associated to COVID-19 has been mutating [89]. 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 [13]. According to Ghanbari [17], 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 [2, 5, 10, 29]. To date, few countries have started vaccination or are in the final stages of preparing the vaccination plan, such as the United Kingdom [20] and Russia [7]. Libotte et al. [28] 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. [35] 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.

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. [6] proposed a robust approach based on epidemic renormalization group framework to simulate the second wave for different countries. These authors conducted statistical analyses on different levels of human interaction to find the second wave, that they indicate to occur between July 2020 and January 2021. Nori Junior et al. [36] proposed a two-wave statistical model based on the superposition of normal distributions. The future scenario can be predicted based on parameters estimated for the first wave. Castro [8] 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 [17] proposed a SID-based model with two sub-populations (which are classified by means of immunity) of infected individuals, in order to simulate the second wave in Iran. This author indicates that the second wave may be more severe than the first. Pedro et al. [37] proposed a SEIR model that considers a time-varying parameter to evaluate the impact 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 [9] studied a stochastic SEQR model to analyze the epidemic in India. In turn, He et al. [19] 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. [40] incorporated a time-delay in the stochastic SIRC model. Faranda and Alberti [15] proposed a stochastic SEIR model to analyze the impact of the relaxation of isolation measures in France and Italy. Tang et al. [42] proposed a stochastic SEIHRQ model to investigate the imported cases on local epidemic in the Shaanxi-China province. This model consider the effects of intermittent inflow with a Poisson distribution on the likelihood of multiple outbreaks.

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 piece-wise 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 Section 2 considering the deterministic approach and a stochastic version of the model. The formulation of the proposed inverse problem is detailed in Section 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 Section 5. Finally, conclusions are presented in Section 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 SIDR Model

In order to represent the dynamic behavior of COVID-19 epidemic, the SIDR model is considered. This model considers the interaction among the individuals of the population, which are classified as Susceptible, Infected, Dead and Recovered individuals. We employed 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

\[
\begin{align*}
\frac{dS}{dt} &= -\beta SI \\
\frac{dI}{dt} &= \beta SI - (\alpha + \gamma) I \\
\frac{dD}{dt} &= \gamma I \\
\frac{dR}{dt} &= \alpha I .
\end{align*}
\]

Parameters $\beta$, $\alpha$ and $\gamma$ represent, respectively, the transmission rate, death rate, and recovery rate. The initial conditions are: $D(0) = 0$, $R(0) = 0$, $I(0) = \frac{I(0)}{2}$ and $S(0) = 1 - D(0) - R(0) - I(0)$, since we have the constraint $S(t) + I(t) + D(t) + R(t) = 1$. $\mathbf{I}(0)$ represents the first element in the vector $\mathbf{I}$ of reported infected cases (i.e., the number of infected individuals at $t = 0$), as we will discuss in Section 3. $\varphi$ is a quantity related to the population size and the fraction of reported cases. We will use a new variable to represent $I(0)$,

\[
\delta = \frac{\mathbf{I}(0)}{\varphi} ,
\]

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 ($R_0$) 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 [3]. When $R_0 > 1$, the disease has a spreading potential, with a tendency to increase the number of new cases.

The basic reproduction number for the SIDR model is represented by [3]

\[
R_0 = \frac{\beta}{\alpha + \gamma} .
\]

It can be a large value ($R_0 \gg 1$) when $\alpha + \gamma$ is small compared to $\beta$. Thus, in the early stage of COVID-19 spread in different countries, we can observe high values for $R_0$ considering different numerical approaches [6,21].

\[1\]

[24][25][32][33][38] As mentioned by Marimuthu et al. [32], 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, in real-cases, a population is not entirely susceptible to the disease, as consequence of immunity, for example. In practice, the effective (or time-dependent) reproduction number ($R_t$) can provide more realistic results. $R_t$ represents the expected number of secondary cases arising from a primary case infected at time $t$ [43].

To evaluate this parameter, various methodologies can be used, such as, Exponential Growth and Time Dependent Methods [32]. In this work, we adopt $R_t$, defined as [12]

\[
R_t = R_0 S(t) ,
\]

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

2.2 Stochastic SIDR 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 [15] 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 modeled by a specific probability distribution. In this case, for each control parameter (in our case, $\beta$, $\gamma$ and $\alpha$), a new differential equation with stochastic contribution is formulated. Thus the following stochastic differential equations are employed [15]:

\[
\frac{d\Phi_{un}}{dt} = -\Phi_{un} + \Phi_{un0} + \eta\phi \xi(t) \]

where $\Phi_{un}$ is the control parameter with uncertainty ($\Phi_{un} \in \{\beta_{un}, \gamma_{un}, \alpha_{un}\}$), $\Phi_{un}(0) = \Phi_{un0}$ is the initial condition related to the corresponding parameters, $\eta\phi$ is the standard deviation for the control parameter with uncertainty, and $\xi(t)$ is a random number represented by a specific distribution type. In this case, the stochas-
tive SIDR model is given by

\[
\begin{align*}
\frac{dS}{dt} &= -\beta_{un}SI \\
\frac{dI}{dt} &= \beta_{un}SI - (\alpha_{un} + \gamma_{un})I \\
\frac{dD}{dt} &= \gamma_{un}I \\
\frac{dR}{dt} &= \alpha_{un}I \\
\frac{d\beta_{un}}{dt} &= -\beta_{un} + \beta_{un0} + \eta_{\beta}\xi(t) \\
\frac{d\gamma_{un}}{dt} &= -\gamma_{un} + \gamma_{un0} + \eta_{\gamma}\xi(t) \\
\frac{d\alpha_{un}}{dt} &= -\alpha_{un} + \alpha_{un0} + \eta_{\alpha}\xi(t)
\end{align*}
\]

with appropriated initial conditions (discussed in Section 4.2.2).

3 Formulation of the Inverse Problem

In the SIDR model described by Eqs. (1), the parameters $\beta$, $\gamma$, $\alpha$ and $\delta$ must be calculated in order to allow the numerical solution of the system of ordinary differential equations. It is important to mention that since $\delta$ 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 $\beta$, $\gamma$ and $\alpha$ are taken to be constant during the integration of the differential system, for the sake of simplicity. However, according to Castro [8], both the transmission rate ($\beta$) and death rate ($\gamma$) can be chosen as time-dependent parameters to better capture the effects of policies defined to mitigate the pandemic. Following Cheynet [11], we define $\beta$ as:

\[
\beta(t) = \beta_0 \exp \left( - (\beta_1 (t - \beta_2))^2 \right)
\]

where $\beta_0$ (day$^{-1}$), $\beta_1$ (day$^{-1}$) and $\beta_2$ (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 $\beta_0$, $\beta_1$, $\beta_2$, $\gamma$ and $\alpha$ must be calculated. In addition, for the first wave, the quantity related to the initial value for the infected individuals ($\delta$) must be estimated, as well as the time instant $\tau$ where each set of parameters are active, i.e., the interface between the end of the first wave and the start of the second wave. In summary, 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, $\theta$ (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 Eqs. (1) for a discrete time set and a particular approximation of the vector $\theta$ produces vectors of responses for each compartment of the model, i.e., $S$, $D$, $I$ and $R$. We adopt reported data related to infected and dead individuals, which are respectively denoted by $I$ and $D$, in the parameter estimation procedure. Then, we are interested in an approximation of the vector $\theta$ that minimizes the error between the model responses $I(\theta)$ and $D(\theta)$, and the reported data, $I$ and $D$, respectively. In other words, we aim to minimize the quantity

\[
\mathcal{F}(\theta) = \frac{||I(\theta) - I||^2}{(\max (I))^2} + \frac{||D(\theta) - D||^2}{(\max (D))^2},
\]

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

Furthermore, in practice, the basic reproduction number $R_0$ and the effective reproduction number $R_t$ cannot assume arbitrarily large values. Considering this situation, we introduce an inequality constraint to assure that the solution of the inverse problem will represent the adherence to the observed values for $R_t$ in the pandemic, as follows

\[
\frac{\beta(t)}{\gamma + \alpha} S(t) \leq R_{\text{max}},
\]

which holds $\forall t \in 0 \leq t \leq t_f$, where $S(t)$ is the simulated normalized susceptible population at time $t$ and $R_{\text{max}}$ 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 $\beta$. Obviously, the use of $R_0$ or even $R_t$ as a parameter to analyze the spread of the disease must be carefully conducted [1]. The main idea here is to avoid unrealistic values for $R_t$.

Mathematically, the inverse problem is formulated as [33]:

\[
\theta^* = \arg \min_{\theta} \mathcal{F}(\theta)
\]

subject to Eq. (9), where $\theta^*$ is the optimal value of $\theta$. 
In this case, the normalized SIDR 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 ($F$). It is important to mention that the constrained optimization problem is transformed into an equivalent unconstrained problem by using the Static Penalty Function Method [16].

4 Data, Numerical Algorithms and Error Metrics

4.1 Data

The data employed for parameter estimation are the daily number of infected individuals ($I$) and the cumulative number of deaths ($D$) 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 [17].

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 [11] (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 $G_{\text{max}} = 1000$, and strategy rand/1/bin (see Storn and Price [11] 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, $R_{\text{max}}$ is considered to be equal to 10, based on recent studies regarding $R_t$ in Germany [23]. The penalization parameter (required to use the Static Penalty Function method) is equal to $10^5$. The deterministic SIDR model is numerically solved using the fourth-order Runge-Kutta method [11] 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 [20]. In this work, we use the Milstein method [23]. 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 $\beta$, $\gamma$, and $\alpha$.

In this context, to define a type of distribution to represent the stochastic contribution. As suggested by Faranda and Alberti [15], $\gamma_{un}$ and $\alpha_{un}$ is normally distributed, and $\beta_{un}$ follows a lognormal distribution, defined as

$$
\gamma_{un} \sim \mathcal{N}\left(\bar{\gamma}^f, \eta_\gamma\right)
$$

$$
\alpha_{un} \sim \mathcal{N}\left(\bar{\alpha}^f, \eta_\alpha\right)
$$

$$
\beta_{un} \sim \mathcal{LN}\left(\bar{\beta}^f - 0.5\kappa^2, \eta_\beta\right)
$$

where $\kappa$ is a proportionality constant, and $\eta_\gamma$, $\eta_\alpha$, and $\eta_\beta$ 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, $\eta_\gamma = \kappa$, $\eta_\alpha = \kappa\bar{\alpha}^f$, and $\eta_\beta = \kappa\bar{\beta}^f$. 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 ($\Theta$), used to determine the goodness-of-fit of the model. Mathematically, this metric is given...
by \( [10] \)

\[
\Theta(Z) = 1 - \frac{\|Z - Z\|^2}{\|Z - IZ\|^2}
\]  

(12)

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

The second metric represents the root-mean-squared error (\( \Gamma \)), calculated as \( [10] \)

\[
\Gamma(Z) = \frac{\|Z - Z\|}{N^{1/2}}
\]  

(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 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 in the best estimator for \( \theta \) (the best results in Table 1), is equal to \( F = 0.1324 \). This corroborates the affinity between the profiles resulting from the integration of \( I \) and \( D \), 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, \( \Gamma \) 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 essentially constant, i.e., the number of new infections between the end of the first wave and the beginning of the second wave remains very low. On the other hand, the number of recovered individuals increases during the epidemic period, as expected. Such behaviors demonstrate that the system presents a behavior that is consistent with the dynamics of an epidemic.

Regarding \( \beta \), 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. 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 $\beta_1 \to 0$ and, therefore, $\beta(t) \approx \beta_0$ (see Eq. 7). A possible criticism to this approach is related to the discontinuity in the profile of $\beta$ 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 $\beta$ are expected. Moreover, with this assumption we avoid discontinuities in the profiles for the individuals in the compartments.

The effective reproduction number $R_t$, with $R_0$ being calculated using the values shown in Table 1, is presented in Fig. 2. In general, for each wave, $R_t$ decreases significantly compared to the initial value. This 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 controlled ($R_t \to 0$). In addition, Fig. 2 shows the behavior of $R_t$ obtained by Petrova et al. [38] using the SID model, and by Khalilie et al. [25] 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 $R_t$ approximately equal to 1.5 for the first wave and 6.50 for the second wave. According to the Robert Koch Institute [22], 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.

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
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.

| Parameter                  | First wave     | Second wave    |
|----------------------------|----------------|---------------|
| $\beta_0^t$ (day$^{-1}$)   | 0.47815 ± 0.02454 | 0.78693 ± 0.02434 |
| $\beta_1^t$ (day$^{-1}$)   | 0.02830 ± 0.00112  | 0.00558 ± 0.00009  |
| $\beta_2^t$ (day)          | 0.38865 ± 0.06656  | 0.17129 ± 0.02332  |
| $\gamma^t$ (day$^{-1}$)    | 0.00502 ± 0.00012   | 0.00071 ± 0.00001  |
| $\alpha^t$ (day$^{-1}$)    | 0.05472 ± 0.00155   | 0.02349 ± 0.00001  |
| $\delta$                  | $1.09 \times 10^{-6}$ ± $1.64 \times 10^{-9}$ | $\tau$ (day) 187.08035 ± 11.74343 |

With the analyzed data, as illustrated in Fig. 4, qualitatively, the results are similar to those obtained for the data from Germany (see Fig. 3), but differ in terms of the magnitude of the data and results. The error metrics point out that the fit between the response 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 adequately.

In turn, Fig. 4 presents the behavior of $R_t$ for each wave, considering the parameters shown in Table 2. Again, our results are in good agreement with those obtained by Petrova et al. 38 and Khalilae et al. 29 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 $R_t$, 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 15 (the basic reproduction number $R_0$ obtained by these authors is, approximately, equal to 4, using a stochastic SEIR model). Due to stochastic nature of this model, the value of this parameter can fluctuate up to 6. In turn, Cacciapaglia et al. 6 reported that the maximum value that $R_t$ can reach is close to 10 for the Italian case.

5.2 Stochastic Analysis

In order to analyze the influence of uncertainties, this section considers the simulation of the stochastic SIDR 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, $t > \tau$. The first wave is simulated with the same parameters obtained in the previous sections (Tables 1 and 2). In practice, for the second wave, uncertainties are incorporated into the parameters $\beta$, $\gamma$ and $\alpha$. The initial conditions for the stochastic SIDR model are those obtained for the first wave using the deterministic approach. In this sense, for Germany, $\beta_{un}(\tau) = 0.09491$ day$^{-1}$, $\gamma_{un}(\tau) = 0.00058$ day$^{-1}$, and $\alpha_{un}(\tau) = 0.00854$ day$^{-1}$; for Italy, $\beta_{un}(\tau) = 0.14485$ day$^{-1}$, $\gamma_{un}(\tau) = 0.00071$ day$^{-1}$, and $\alpha_{un}(\tau) = 0.02349$ day$^{-1}$. In addition, since $\beta$ 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 $\kappa$, with $\kappa \in \{0.0, 0.1, 0.2, 0.3, 0.4\}$, as defined in Eq. (11). The fluctuation of each curve in relation to deterministic solution (where $\kappa = 0$ is the reference profile) is shown in Fig. 5a. For both Germany and Italy, the increase in the value $\kappa$ 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 peak. We may also note that a high value for $\kappa$ 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 SIDR model.

In addition, for both analyzed cases, $R_t$ decreases over time (see Fig. 5b). It should be clear that these profiles represent only the second wave and, therefore, it is expected a high value in the beginning of the wave (which will produce the second peak, as discussed previously). Furthermore, the higher is the value of $\kappa$, 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 $\kappa$ value—we obtain a higher value for $R_t$ (for the same time instant $t$).

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.

### 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 SIDR model is conducted. For this purpose, the parameters of the compartmental SIDR 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.

### Declarations

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. Gustavo Libotte is sup-
ported 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.

Conflicts 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.

Availability of data and material
The data used in this work are available at [https://www.worldometers.info/coronavirus/](https://www.worldometers.info/coronavirus/).

Code availability
The codes used to reproduce the results presented in this work are available upon request to the corresponding author.

References
1. 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). DOI 10.1038/d41586-020-02009-w
Modeling the second wave of COVID-19

2. 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. Frontiers in Immunology 11, 1880 (2020). DOI 10.3389/fimmu.2020.01880

3. Anastassopoulos, 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)

4. Atkinson, K.A.: An Introduction to Numerical Analysis. New York: John Wiley & Sons (1989)

5. Baviskar, T., Raut, D., Bhatt, L.K.: Deciphering vaccines for COVID-19: where do we stand today? Immunopharmacology and Immunotoxicology pp. 1–14 (2020). DOI 10.1080/08923973.2020.1837867

6. Cacciaipaglia, G., Cot, C., Sannino, F.: Second wave COVID-19 pandemics in Europe: a temporal playbook. Scientific Reports 10, 15514–15522 (2020). DOI https://doi.org/10.1038/s41598-020-72611-5

7. Callaway, E.: Russia announces positive COVID-vaccine results from controversial trial. Nature (2020). DOI 10.1038/d41586-020-03209-0. URL https://doi.org/10.1038/d41586-020-03209-0

8. Castro, F.: Modelling of the Second (and Subsequent) Waves of the Coronavirus Epidemic. Spain and Germany as Case Studies (2020). DOI https://doi.org/10.1101/2020.06.12.20129429. MedRxiv preprint

9. Chau, A.L., Singh, R.K.B.: Stochastic approach to study control strategies of COVID-19 pandemic in India. Epidemiology and Infection 148, 1–9 (2020). DOI https://doi.org/10.1017/S0950268820001946

10. Cheng, C.L., Shalabh, Garg, G.: Coefficient of determination for multiple measurement error models. Journal of Multivariate Analysis 126, 137 – 152 (2014). DOI https://doi.org/10.1016/j.jmva.2014.01.006. URL http://www.sciencedirect.com/science/article/pii/S0047259X14000141

11. Cheynet, E.: Generalized SEIR Epidemic Model (Fitting and Computation) (2020). DOI 10.5281/ZENODO.3911854. Zenodo

12. Delamater, P.L., Street, E.J., Leslie, T.F., Yang, Y.T., Jacobsen, H.K.: Complexity of the basic reproduction number. Emerging Infectious Diseases 25, 1–4 (2019)

13. van Dorp, L., Richard, D., Tan, C.C., Shaw, L.P., Bakker, M., Klein, M., Loens, K., Jubbink, M.F., Matser, A., Kissella, C.M., Rueda, P., Leven, M., Goossens, H., Prins, M., Sastre, P., Deijs, M., van der Hoek, L.: Seasonal coronavirus protective immunity is short-lasting. Nature Medicine (2020). DOI 10.1038/s41467-020-19818-2

14. Edridge, A.W.D., Kaczorowska, J., Hoste, A.C.R., Bakker, M., Klein, M., Loens, K., Jubbink, M.F., Matser, A., Kissella, C.M., Rueda, P., Leven, M., Goossens, H., Prins, M., Sastre, P., Deijs, M., van der Hoek, L.: Seasonal coronavirus protective immunity is short-lasting. Nature Medicine (2020). DOI 10.1038/s41591-020-0965-6

15. Faranda, D., Alberti, T.: Modelling the second wave of COVID-19 infections in France and Italy via a stochastic SEIR model. Chaos: An Interdisciplinary Journal of Nonlinear Science 30, 111101–111132 (2020). DOI https://hal.archives-ouvertes.fr/hal-02668318v3

16. Funk, C.D., Lafferrière, C., Ardakani, A.: A snapshot of the global race for vaccines targeting SARS-CoV-2 and the COVID-19 pandemic. Frontiers in Pharmacology 11, 937 (2020). DOI 10.3389/fphar.2020.00937

17. Ghanbari, B.: On forecasting the spread of the COVID-19 in Iran: The second wave. Chaos, Solitons and Fractals 140, 110176–110184 (2020)

18. 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 pp. 1525–1534 (2020). DOI 10.1016/S0140-6736(20)32007-9

19. He, S., Tang, S., Rong, L.: A discrete stochastic model of the COVID-19 outbreak: Forecast and control. Mathematical Biosciences and Engineering 17, 2792–2804 (2020). DOI 10.3934/mbe.2020153

20. Higham, D.J.: An algorithmic introduction to numerical simulation of stochastic differential equations. SIAM Review 43(3), 525–546 (2001). DOI http://www.jstor.org/stable/2649798

21. 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). DOI https://doi.org/10.1371/journal.pone.0236464

22. Institute, R.K.: Coronavirus disease 2019 (COVID-19)—daily situation report of the robert koch institute. URL https://corona.rki.de Accessed December 29, 2020

23. Iwasaki, A.: What reinfections mean for COVID-19. The Lancet Infectious Diseases (2020). DOI 10.1016/S1473-3099(20)30783-0

24. 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. Journal of Medical Internet Research 22, 1–11 (2020). DOI https://doi.org/10.2196/19007

25. Khaitie, 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. medRxiv (2020). DOI 10.1101/2020.04.04.20053637

26. 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). DOI 10.1038/d41586-020-03441-8

27. 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). DOI 10.1016/S0140-6736(20)30746-7

28. 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. Computer Methods and Programs in Biomedicine 196, 105664–105677 (2020)

29. Liu, Y., Wang, K., Massoud, T.F., Paulmurugan, R.: SARS-CoV-2 vaccine development: An overview and perspectives. ACS Pharmacology & Translational Science 3(5), 844–858 (2020). DOI 10.1021/acspts.0c00109

30. 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. Nature Medicine 26(8), 1200–1204 (2020). DOI 10.1038/s41591-020-0965-6
31. López, L., Rodó, X.: The end of social confinement and COVID-19 re-emergence risk. Nature Human Behaviour 4(7), 746–755 (2020). DOI 10.1038/s41562-020-0908-8
32. 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. Clinical Epidemiology and Global Health 9, 57–61 (2021). DOI https://doi.org/10.1016/j.cegh.2020.06.012
33. Milstein, G.N.: Approximate integration of stochastic differential equations. Theory of Probability & Its Applications 19, 557–600 (1975). DOI 10.1137/1119062
34. Moura Neto, F.D., Silva Neto, A.J.: An Introduction to Inverse Problems with Applications. Springer Berlin Heidelberg, Berlin, Heidelberg (2013). DOI 10.1007/978-3-642-32557-1
35. Mukandavire, Z., Nyabadza, F., Mahunguza, 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). DOI 10.1371/journal.pone.0236003
36. 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). DOI https://doi.org/10.1101/2020.05.08.20095133. MedRxiv preprint
37. Pedro, S.A., Ndjomatchoua, F.T., Jentsch, P., Tchuenche, J.M., Anand, M., Bauch, C.T.: Conditions for a second wave of COVID-19 due to interactions between disease dynamics and social processes. Frontiers in Physics 2020, 1–9 (2020)
38. Petrova, T., DmitriSoshnikov, Grunin, A.: Estimation of time-dependent reproduction number for global COVID-19 outbreak. Preprints (2020). DOI 10.20944/preprints202006.0289.v1
39. Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muurato, A.E., Zou, J., Fontes-Garfias, C.R., Mchandani, 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). DOI 10.1038/s41586-020-2895-3
40. Rihan, F.A., Alsakaji, H.J., Rajivgauthi, C.: Stochastic SIRC epidemic model with time-delay for COVID-19. Advances in Difference Equations 2020:502, 1–20 (2020). DOI https://doi.org/10.1186/s13662-020-02964-8
41. Storn, R., Price, K.: Differential Evolution—a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 11(4), 341–359 (1997). DOI 10.1023/A:100820291328
42. 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). DOI https://doi.org/10.1101/s41586-020-2895-3
43. Thompson, R.N., Stockwin, J.E., R. D. van Gaalen, J.A.P., Kamvar, Z.N., Denmarsh, P.A., Dahlquist, 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). DOI https://doi.org/10.1016/j.epide.2019.100356
44. 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). DOI 10.1016/S1473-3099(20)30764-7
45. Tisdell, C.A.: Economic, social and political issues raised by the covid-19 pandemic. Economic Analysis and Policy 68, 17–28 (2020). DOI https://doi.org/10.1016/j.eap.2020.08.002
46. Vanderplaats, G.N.: Numerical Optimization Techniques for Engineering Design with Applications. McGraw-Hill, New York (1984)
47. Worldometer: Coronavirus Live Tracker. URL https://www.worldometers.info/coronavirus/#countries. Accessed December 29, 2020
48. 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. Frontiers of Medicine 14, 613–622 (2020). DOI https://doi.org/10.1007/s11684-020-0787-4