Original Articles

Forecasting COVID-19 Chile' second outbreak by a generalized SIR model with constant time delays and a fitted positivity rate

Patricio Cumsille[a,d,*], Óscar Rojas-Díaz[b,c], Pablo Moisset de Espanés[d], Paula Verdugo-Hernández[e]

[a] *Department of Basic Sciences, Faculty of Sciences, University of Bío-Bío Campus Fernando May, Av. Andrés Bello 720, Casilla 447, Chillán, Chile*
[b] *Neurovision AI, San Eugenio 3494, Santiago, Chile*
[c] *Department of Informatics Engineering, Faculty of Engineering, University of Santiago of Chile, Santiago, Chile*
[d] *Centre for Biotechnology and Bioengineering (CeBiB) University of Chile, Beauchef 851, Santiago, Chile*
[e] *Escuela de Pedagogía en Ciencias Naturales y Exactas, Facultad de Ciencias de la Educación, Universidad de Talca, Chile*

Abstract

The COVID-19 disease has forced countries to make a considerable collaborative effort between scientists and governments to provide indicators to suitable follow-up the pandemic's consequences. Mathematical modeling plays a crucial role in quantifying indicators describing diverse aspects of the pandemic. Consequently, this work aims to develop a clear, efficient, and reproducible methodology for parameter optimization, whose implementation is illustrated using data from three representative regions from Chile and a suitable generalized SIR model together with a fitted positivity rate. Our results reproduce the general trend of the infected's curve, distinguishing the reported and real cases. Finally, our methodology is robust, and it allows us to forecast a second outbreak of COVID-19 and the infection fatality rate of COVID-19 qualitatively according to the reported dead cases.

© 2021 International Association for Mathematics and Computers in Simulation (IMACS). Published by Elsevier B.V. All rights reserved.

Keywords: Mathematical modeling; Parameter estimation method; Predictive modeling; Computational model; Epidemiological modeling

1. Introduction

From the early stage of the pandemic, the Chilean government has strengthened the health system by increasing the number of intensive care unit beds and mechanical ventilation devices for critically ill patients. Currently, the government is enhancing tools for tracing capture on cases and contacts of confirmed COVID-19 patients. A collaborative effort among the main universities, the ministries of health and science, technology, knowledge, and innovation produced a web platform for data analysis. The site provides key indicators to represent the state and

---

evolution of the pandemic. These indicators are meant to provide timely and clear information to the authorities, communication media, citizenship, and the scientific community to understand the pandemic evolution and improve decision-making.

Mathematical modeling plays a crucial role in quantifying indicators describing diverse aspects of the pandemic. Many research groups in Chile have developed mathematical models, and providing a comprehensive review would ask for much unnecessary effort. Therefore, we selected some that have produced periodic reports on COVID-19, which we comment on next and partly use ideas. It is worth noting that, in general, these works do not provide details on the methodology, which would be desirable to reproduce results scientifically.

The Center for Mathematical Modeling (CMM) group provided mainly simulations, not predictions. Their results aimed to estimate the maximum capacity required for intensive care unit beds and simulate different opening scenarios (for teaching and economic activities). They developed complex modeling based on a generalization of the SEIR compartmental model, including variables for hospitalized, separating those requiring critical services, and even age structures, and containing several parameters making it difficult to deal with in the sense of forecasting; see the website. [7].

The COVID-19 University of Concepción (UdeC) group provided simulations and predictions. Their results aim to predict the curve of real infected for different datasets encompassing the Ñuble, Bío-Bío regions, and the entire country and simulate different scenarios with or without confinements. They developed a variant of the SEIR compartmental model, adding a fraction to account for the proportion of real infected observed by the RT-PCR tests and carrying out parameter estimation to calibrate each dataset. In our work, we used some of their ideas and compared some of our results of calibration for the Ñuble Region; see the website [6].

We want to note that the research group of the previous paragraph and other manuscripts [8] support their ideas in the fact that there are unreported cases or that only a fraction of the real infected is observed. Indeed, data is measured with great uncertainty, as we explain in Section 2, and consequently has been reviewed and changed retrospectively. In this sense, we aim to obtain a suitable model that fits the data and forecast the general trend of the real infected' curve, despite the uncertainty in data. We believe that the way is not constructing a complex model. In that case, the parameter estimation required to make the model calibration becomes very difficult or even impossible to do in real-time or does not necessarily provide suitable results. Concerning this, we cite the work [14] that used a complex generalization of the SEIR model proposed in [21], and applied it Chilean data until April 14, 2020, obtained that the peak of infected would occur between April and May, an unsuitable prediction that may be explained because data were near to its zero equilibrium, as we will discuss in Section 4.

Parameter estimation involves solving an inverse problem: given a model and measurements of some state or output variables, the parameters that characterize the system, i.e., those producing a good fit of the model with the data, need to be identified [1,11]. This problem is difficult since no unique analytical or numerical solution is usually available [1,11]. Even if a unique solution was available, suitable initial guesses are required by optimization solvers to compute suitable parameter estimates [1,11]. From the previous discussion, the main goal of the present study was to describe a hybrid approach [10], i.e., a methodology that holistically mixes mathematical modeling and experimental design, which is required, as shown in the literature, for better understanding the studied system by fitting parameters of a given model with a specific scenario and for obtaining models with predictive capability. In particular, we aim to develop a clear, efficient, and reproducible methodology for optimization parameters, whose implementation is illustrated by using data from three representative regions from Chile and a suitable generalized SIR model. The methodology relies on a numeric procedure called the Trust-Region-Reflective optimization algorithm. Simultaneously, we computed the model's predictive power using goodness-of-fit criteria and assessed parameter uncertainty through sensitivities that yield the standard errors.

2. A generalized SIR model with constant time delays and fitted positivity rate

The classical SIR model's main assumptions are that the mean infection and recovery rates are positive constants over time, which is not necessarily realistic for a given disease. In COVID-19, when a susceptible has just been infected does not instantaneously show symptoms. There should be a time-delay period due to the SARS-CoV-2 incubation. Similarly, the infected subjects are not recovered or dead instantaneously, but this would occur within a time-delay period. Indeed, these time-delay periods for both phenomena, incubation, and removed (by recovery or dead), are well documented and reported, e.g., in [23].

In this work, we consider the following generalization of the classical SIR model (adapted from the model from Book by Hairer et al. [15] p. 295; [19]), which we will call generalized SIR model with constant time delays:

$$\frac{dS}{dt}(t) = -\frac{\beta(t)}{N}S(t)I(t-\tau_1), \tag{1a}$$

$$\frac{dI}{dt}(t) = \frac{\beta(t)}{N}S(t)I(t-\tau_1) - \gamma I(t-\tau_2), \tag{1b}$$

$$\frac{dR}{dt}(t) = \gamma I(t-\tau_2). \tag{1c}$$

The parameters of the model (1) are $\beta(t)$ that corresponds to the average number of contacts per person per time, multiplied by the probability of disease transmission in exposure between a susceptible and an infectious subject (see first report at the web site [7]), or more simply, $\beta(t)$ is the mean rate of disease transmission; γ is the mean removal rate, which in the classical SIR model represents the average time of infection duration; τ_1 is the mean incubation time of disease; τ_2 is the median time from onset to clinical recovery or death caused by disease, or the duration time of disease until recovery or death.

Parameter $\beta(t)$ depends essentially on two factors: the disease characteristics and the contact rate within the population [12]. Since we cannot modify the disease characteristics, authorities can foster or even impose incentives to each individual to reduce his/her contact rate with other individuals in the population, causing a decrease in $\beta(t)$. Consequently, we consider $\beta(t)$ variable in time to consider mitigation strategies dynamics of COVID-19 such as scenarios confinement and opening at different periods.

The variables of model (1) are:

1. S, the number of susceptible individuals;
2. I, the number of infectious individuals;
3. R, the number of removed (and immune), or deceased individuals.

The initial conditions have to satisfy $S(t_0) + I(t_0) + R(t_0) = N$, where N is the size of the population under study for a closed system and taking into account that $(S + I + R)'(t) = 0$ for all $t > 0$, and where t_0 is a day chosen suitably after the pandemic began, and which varies for every studied population (in Chile, the first infected subject was detected on March 02, 2020). We remark that the model (1) considers the symptomatic and presymptomatic subjects since the term $I(t-\tau_1)$ at the right-hand side of Eq. (1a) represents the number of infected cases that do not manifest symptoms yet since the infection is under incubation.

The solution of model (1) will be defined in $[t_0, T[$, where T is defined as the first time at which the infected number becomes zero, i.e., $I(T) = 0$. From Eq. (1b), we have that for every $t \in [t_0, T[$, $I'(t) > 0$ if and only if

$$\frac{\beta(t)}{N}S(t)I(t-\tau_1) - \gamma I(t-\tau_2) > 0,$$

or equivalently

$$R_e(t) := R_e^{SIR}(t)R_I(t) > 1, \quad \text{where} \quad R_e^{SIR}(t) := \frac{\beta(t)S(t)}{\gamma N} \quad \text{and} \quad R_I(t) := \frac{I(t-\tau_1)}{I(t-\tau_2)}. \tag{2}$$

In Eq. (2), $R_e(t)$ is the effective reproductive number of the generalized SIR model with constant time delays (1), $R_e^{SIR}(t)$ is the effective reproductive number in the classical SIR model, and $R_I(t)$ is the ratio of the latent over the removed infected.

It is worth to note that the generalized SIR model with constant time delays can generate a complex dynamics since, as we will see in Section 4, by contrast to the classical SIR model, it can simulate more than one local maximum, providing thus a way to explain a COVID-19 second-outbreak, as already observed in some European countries.

We assume that infectious subjects are detected only when RT-PCR tests are applied to them, and their results are positive. Therefore the observation of this variable is equivalent to the number of positive RT-PCR tests. In this sense, the number of asymptomatic subjects in the population is underestimated, given that they are difficult to detect since they do not manifest any symptoms. Therefore this variable is difficult to observe with accuracy, so we did not consider it in our modeling directly.

Remark. The generalized SIR model with constant time delays would produce better prediction results than the classical SEIR model since no observation of the exposed population is available. This is because, similarly to the asymptomatic population, the exposed are difficult to observe. Indeed, the Chilean government database [16,17], apart from the symptomatic cases, counts the asymptomatic cases, and there is no way to know how many become symptomatic ones.

By contrast, the symptomatic cases are detected as symptoms manifest themselves, and the RT-PCR test confirms the infection. Therefore, this variable is relatively well-observed, provided that the RT-PCR tests are reliable and enough are available. Taking this into account, we assume that the daily number of infected with symptoms reported by the ministry of health, denoted by $I_r(t)$, is underestimated since it depends on the availability and proper application of RT-PCR tests; see the presentation on COVID-19 in Chile in [2]. Consequently, we assume that $I_r(t)$ is a fraction of the actual number of infected cases $I(t)$, i.e.,

$$I_r(t) = f(t)I(t). \tag{3}$$

In Eq. (3), $f(t)$ is the ratio between positive RT-PCR tests number and the real infected cases for day t, which accounts for the real positivity rate. It is worth noting that $f(t)$ is related to the *positivity rate of detected cases*, denoted by $f_r(t)$, reported by the Ministry of Health, and defined as the proportion of tests that result positive to the total number of applied tests for day t, denoted by $T(t)$. Let us denote by $R_{IT}(t) = \frac{I(t)}{T(t)}$ the ratio of actual infected cases to the total number of applied tests, and by $f^A(t) = \frac{I_r^A(t)}{I(t)}$ the proportion of asymptomatic infected subjects to the actual infected ones, where $I_r^A(t)$ stands for the asymptomatic infected reported for day t. Since the number of positive detected cases corresponds to the total number of confirmed infected reported with and without symptoms, $I_r(t) + I_r^A(t)$, then it easily follows that

$$f_r(t) = \frac{I_r(t) + I_r^A(t)}{T(t)} = \left[f(t) + f^A(t) \right] R_{IT}(t), \tag{4a}$$

$$f(t) = \frac{f_r(t)}{R_{IT}(t)} - f^A(t). \tag{4b}$$

From (4b), one has that if $R_{IT}(t) \geq 1$ then $f(t) < f_r(t)$. In addition, assuming that $R_{IT}(t) \leq c$ for some constant $0 < c < 1$ satisfying that $\frac{1-c}{c} \geq M$ where $M > 0$ is the maximum of $\frac{I_r^A(t)}{I_r(t)}$, which can be determined from data, from (4a) it can be proven that $f_r(t) \leq f(t)$.

We modeled $f(t)$ in the same way as Cabrera-Vives et al.; see the first report in [6]. That is to say, $f(t)$ is an inverted Sigmoid-type function such that if $I(t)$ is small enough, which occurred during the beginning of the outbreak, then an important fraction of the real infected cases are detected ($I_r(t) \approx I(t)$). On the contrary, when $I(t)$ is large enough, which occurred just before the quarantines were imposed, then only a small fraction $0 < a < 1$ of the real infected cases are detected ($I_r(t) \approx aI(t)$). Precisely, $f(t)$ is defined as

$$f(t) = 1 + \frac{a-1}{1 + e^{-k(I(t)-I_m)}} \tag{5}$$

whose parameters are a, k and I_m, where a and k represent the minimum and the decay rate of $f(t)$, respectively. On the other hand, the measure of how large/small is $I(t)$ is given by a threshold I_m such that $I(t) \ll I_m$ implies $f(t) \approx 1$, and $I(t) \gg I_m$ implies $f(t) \approx a$.

3. Material and methods

This section describes the data and the methodology used for estimating the optimal parameters that produce the model' fit to the data.

The data corresponds to the reported Chilean government's daily official cases at the regional level in the current year, which we used to predict the COVID-19 spreading. We chose data from some representative regions of the north, south, and central zones. Specifically, we worked with the Antofagasta, the Metropolitan, and the Ñuble regions, where the Metropolitan Region is the most populous and which capital is Santiago of Chile, the country's capital. The data we used is available from the web site [17]; also, the countrywide level data is at [16]. The data in [17] contains not only the number of confirmed cases but also the size of targeted populations (N) and the periods of quarantines per commune.

3.1. Scenarios for the fittings and forecastings

We fitted the model (1) to each dataset by taking into account different scenarios that encompass, at least, three phases of the pandemic spreading: an early stage characterized by low dissemination and levels of daily new infected cases relatively low, followed by a fast-propagation under quarantine, a measure taken by the Ministry of Health as a reaction to the first phase in a selective way by sectors that may encompass several communes of the region at different periods, characterized by a fast increasing of daily new infected cases and finally, a spreading slowdown stage where the quarantine is relaxed. In a general way, to account for the phases described, we modeled the mean rate of COVID-19 transmission $\beta(t)$ by a piecewise constant function as

$$\beta(t) = \sum_{i=1}^{n} \beta_i \mathbf{1}_{[t_{i-1},\, t_i[}(t) \qquad \text{for every } t \in [t_0,\, T[\, . \tag{6}$$

In Eq. (6), β_i denotes the mean rate of transmission for every scenario $i = 1, \ldots, n$, $n \geq 3$ is the number of scenarios, $\mathbf{1}_{[t_{i-1},\, t_i[}(t)$ stands for the indicator (or characteristic) function of the time interval $[t_{i-1},\, t_i[$ that corresponds to the $i-$th scenario, t_0 is the initial time, and T is the maximal time of existence of model solutions, defined in Section 2. For the three datasets considered, the initial condition was imposed at a day t_0 chosen with respect to day 1 that corresponds to March 03, 2020, when it began to measure the infected in the three targeted regions.

From the definition (6), we denote the model parameters as the vector
$\theta = (\beta_1,\, \beta_2,\, \cdots,\, \beta_n,\, \gamma,\, \tau_1,\, \tau_2,\, a,\, k,\, I_m) \in \mathbb{R}^{n+6}$, where $n = 3$ is the minimum of scenarios to calibrate the model (1) to the data. In our computations, for the Antofagasta Region dataset, we split the second phase into two to capture the real behavior of COVID-19 expansion more accurately due to two differentiated periods of quarantine in that region, taking $n = 4$ in that case.

Finally, we simulated a possible change of scenario, which was evaluated by computing the relative errors predicting data after the final calibration time. To do that, we solved the model (1) by taking a mean rate of transmission $\beta(t)$, as in (6), adding one more scenario to the n computed from the calibration, and letting vary β_{n+1} to verify the value that forecasts the best possible the data after the calibration. To be precise, we solved the model (1) by redefining $\beta(t)$ as

$$\beta(t) = \sum_{i=1}^{n} \hat{\beta}_i \mathbf{1}_{[t_{i-1},\, t_i[}(t) + \beta_{n+1} \mathbf{1}_{[t_n,\, t_{n+1}[}(t) \qquad \text{for every } t \in [t_0,\, T[\, , \tag{7}$$

where $(\hat{\beta}_1,\, \hat{\beta}_2,\, \cdots,\, \hat{\beta}_n) \in \mathbb{R}^n$ denotes the parameters estimated for the fitting of every dataset (computed together with $(\hat{\gamma},\, \hat{\tau}_1,\, \hat{\tau}_2,\, \hat{a},\, \hat{k},\, \hat{I}_m)$), the time t_n represents a new change of scenario, t_{n+1} extends until the last predicted datum, and β_{n+1} describes the value of the mean rate of transmission after the time t_n ($\beta_{n+1} = \hat{\beta}_n$ means no change of scenario).

3.2. Identifiability analysis of model' parameters

In this section, we explain in detail the methodology to conduct the identifiability analysis of model' parameters, i.e., finding the parameters of the generalized SIR model with constant time delays (1) that best fit to the every dataset.

3.2.1. Direct and inverse problem

Given a parameters vector $\theta \in \mathbb{R}^{n+6}$, the *direct problem* consists of finding the (unique) solution $(S, I, R)(t, \theta)$ of model (1) with initial condition imposed at a suitable initial time, after the pandemic began. The solution to the direct problem is required to solve the inverse problem. The present study aimed to forecast and simulate different scenarios of COVID-19 expansion in Chile, for which we developed a clear, efficient, and reproducible methodology to solve the parameter identification problem. This procedure is referred to as the *inverse problem*, i.e., given data that provides observations of the variable $I(t)$, namely $I_r(t)$, for some time points t to identify the parameters vector $\theta \in \mathbb{R}^{n+6}$ such that the mathematical model (1) fits the data in the sense of the least-squares. More precisely, to

solve the inverse problem of parameter estimation, we have to find the vector $\theta \in \mathbb{R}^{n+6}$ that minimizes the sum of squares

$$SS(\theta) = \frac{1}{N^2} \sum_{j=1}^{M} \left[(I_r)_j - f(j, \theta_2) I(j, \theta_1) \right]^2 . \tag{8}$$

The objective function defined in (8) corresponds to the sum of squares of the absolute errors $\frac{(I_r)_j - f(j,\theta_2)I(j,\theta_1)}{N}$ relative to the size of the targeted population N, i.e., of the differences between the daily number of infected reported with symptoms, $(I_r)_j$, and the theoretical daily symptomatic infected $f(j, \theta_2) I(j, \theta_1)$, which is a fraction of the real symptomatic infected $I(j, \theta_1)$, corresponding to the model solution evaluated at (j, θ_1) at days $j = 1, \ldots, M$, for a given value of the parameters vector θ, and where M is the size of each dataset. The variable $I(t, \theta_1)$ depends on t and $\theta_1 = (\beta_1, \beta_2, \cdots, \beta_n, \gamma, \tau_1, \tau_2) \in \mathbb{R}^{n+3}$, while the fraction $f(t, \theta_2)$ depends on t and $\theta_2 = (a, k, I_m) \in \mathbb{R}^3$, with $\theta = (\theta_1, \theta_2) \in \mathbb{R}^{n+6}$.

The minimum of the sum of squares $SS(\theta)$ is designated as $\hat{\theta}$ for every dataset. The vector $\hat{\theta}$ is called the *nonlinear least-squares estimator* denoted as *nonlinear LSE* hereafter. To minimize $SS(\theta)$, we applied the *Trust-Region Interior Reflective* (TIR) method implemented in Matlab© as the subroutine *lsqnonlin*, specially adapted for solving nonlinear least-squares minimization problems.

Below, we describe the entire methodology extensively. In Section 3.2.2, we provide details on the implementation of the lsqnonlin solver, specifically on the procedure to assess the quality of the optimal solution $\hat{\theta}$, and on the stopping criterion associated with the algorithm, both concerning the convergence. Also, we computed the fit performance by some goodness-of-fit criteria and represented graphically the traditional sensitivity functions associated with the nonlinear LSE to assess parameter uncertainty, as explained in Section 3.2.3.

3.2.2. Parameter estimation

Convergence of TIR method, as described in [9], is theoretically achieved under general conditions. However, in practice, the convergence depends strongly on the initial parameter estimations, which has to be relatively close to the optimal solution. We were able to efficiently minimize the objective function by a trial and error process, by executing the codes hundreds of times, to guess the suitable initial parameters vector that is different for every dataset [1,11]. It is worth noting that the initial parameters that mostly influence the fitting results are $(\beta_1, \beta_2, \cdots, \beta_n)$ that describe the mean rate of COVID-19 transmission, according to the scenarios of expansion. As usual, there is a tradeoff among how well the model outputs fit the empirical data, the number of parameters (increasing n), and the risk of overfitting. Moreover, there is the human factor involved in selecting the n time intervals associated with each parameter β, although changes in lockdown policies determine reasonable interval boundaries. We will estimate parameters using hand-picked intervals to address this issue, justifying our choices in Section 4. In Section 4.4, we will explore our results' robustness using equal length intervals and varying values of n.

To evaluate the objective function $SS(\theta)$, we numerically solved the model (1) at days $j = 1, \ldots, M$ for different parameters vectors θ depending on every dataset by applying a Runge–Kutta type formulae [22]. We invoked the subroutine *dde23* implemented in Matlab© , designed for solving delay differential equations (DDE) systems with constant time delays. In addition, we had to reconstruct the function of history for the model (1) by interpolating the daily reported data of infected with symptoms, recovered and dead considered for every dataset. The interpolation used is a shape-preserving piecewise cubic as devised by the *interp1* Matlab subroutine with the option *pchip*.

As for the options chosen for the lsqnonlin solver, we first choose the TIR method as the optimization algorithm, which is the default. Secondly, the maximum number of iterations was set at 1,000, while the maximum number of function evaluations was set at 20,000. Finally, the function tolerance was set at $1.0e - 17$, whereas the norm of step tolerance, denoted by *tol*, was set variable between $1.0e - 14$ and $1.0e - 08$, depending on the targeted dataset. The norm of the step $\delta_\theta^{(p)} > 0$ measures the change between two successive iterates $\theta^{(p)}$ and $\theta^{(p+1)}$, which is defined as:

$$\delta_\theta^{(p)} = \| \theta^{(p+1)} - \theta^{(p)} \|. \tag{9}$$

On the other hand,

$$\delta_{SS}^{(p)} = \frac{|SS(\theta^{(p)}) - SS(\theta^{(p+1)})|}{1 + |SS(\theta^{(p)})|} , \tag{10}$$

is the relative change in the sum of squares. The stopping criterion for our algorithm is defined as:

$$\delta_{SS}^{(p)} < 1.0e - 17 \quad \text{or} \quad \delta_{\theta}^{(p)} < tol. \tag{11}$$

When the previous condition is met, the algorithm stops at iteration $p + 1$ and returns an approximation of the nonlinear LSE $\hat{\theta} = \theta^{(p+1)}$, as well as an approximation of the Jacobian matrix χ evaluated at $(j, \hat{\theta})$ for every dataset; see Section 3.2.3 for details. To check the convergence of the algorithm, we also provided the first-order optimality measure, which measures how close the approximation is to the actual minimum of the sum of squares under the criterion of first-order partial derivatives. The first-order optimality measure is defined as the infinity norm of the gradient of the objective function evaluated at the nonlinear LSE $\hat{\theta}$, i.e., the maximum absolute value of the partial derivatives of the objective function with respect to the θ variables:

$$\|\nabla SS(\hat{\theta})\|_\infty = \max_{\ell=1,\dots,n+6} \left| \frac{\partial SS}{\partial \theta_\ell}(\hat{\theta}) \right| . \tag{12}$$

Other metrics that quantify convergence are the norm of the step $\delta_\theta^{(p)}$, the relative change in the sum of squares $\delta_{SS}^{(p)}$ (see Eqs. (9)–(10)), and the number of iterations $p + 1$. Optionally, lsqnonlin displays the metrics of convergence at the end of its execution.

We computed the model parameters $\theta = (\beta_1, \beta_2, \cdots, \beta_n, \gamma, \tau_1, \tau_2, a, k, I_m) \in \mathbb{R}^{n+6}$ within the following bounds: $1 \le \tau_1 \le 14$, $7 \le \tau_2 \le 56$ days (the average for the time of incubation and removal, by recovery or death, is 5 and 14 days respectively; see [23]), $0 \le a \le 1$, $\min\{I_r(t)\} \le I_m \le \max\{I_r(t)\}$ in the time interval considered, and all the rest of parameters are positive (i.e., between 0 and $+\infty$).

3.2.3. Goodness-of-fit criteria and parameter uncertainty

We used statistical methods (from the context of nonlinear least-squares regression) to quantify the fit performance of the model to the data. More precisely, once estimating the nonlinear LSE $\hat{\theta}$, one may compute several goodness-of-fit criteria, which evaluate how well the model (1) fits each dataset. We calculated $\hat{\sigma}^2$ and $\hat{\sigma}$, the unbiased variance of the residuals, and the root mean square error (RMSE) defined respectively in [3,13] as

$$(\hat{\sigma})^2 = \frac{1}{M-n-6} \sum_{j=1}^{M} \left[(I_r)_j - f(j,\theta_2)I(j,\theta_1)\right]^2 = N^2 \frac{SS(\hat{\theta})}{M-n-6}, \qquad \hat{\sigma} = N\sqrt{\frac{SS(\hat{\theta})}{M-n-6}} . \tag{13}$$

If the residuals $\frac{[(I_r)_j - f(j,\theta_2)I(j,\theta_1)]}{N}$ were normally distributed or if the datasets sizes M were sufficiently large, then the estimated covariance matrix, and the sensitivities associated with the nonlinear LSE $\hat{\theta}$ would be expressed as

$$Cov = \hat{\sigma}^2 \left[\chi^t \chi\right]^{-1}, \quad \text{where,} \tag{14}$$

$$\chi_{j,l} = \frac{\partial(fI)}{\partial \theta_l}(j, \hat{\theta}) \quad \text{for every } j = 1, \dots, M; \text{ and every } l = 1, \dots, n+6 . \tag{15}$$

$$S_{\hat{\theta}_l}(t) = \chi_{t,l} \quad \text{for every } t; \text{ and each } l = 1, \dots, n+6 \text{ fixed.} \tag{16}$$

The matrix χ is called the *traditional sensitivity function*, and it quantifies the variation of the state variable $f(t,\theta_2)I(t,\theta_1)$ with respect to changes in the parameters vector components $\theta = (\theta_1, \theta_2) \in \mathbb{R}^{n+6}$. Sensitivity analysis can provide information about the relevance of data measurements to identify parameters; it then yields the basis for new tools to design inverse problem studies; see [4] for details. By using the covariance matrix given in (14), we can compute the standard errors, $se_{\hat{\theta}}$, and the normalized standard errors, $nse_{\hat{\theta}}$, associated to the nonlinear LSE $\hat{\theta}$ from which we could quantify the accuracy of the parameter estimate. Both quantities are defined by

$$se_{\hat{\theta}_l} = \sqrt{(Cov)_{l,l}} , \qquad nse_{\hat{\theta}_l} = 100 \cdot \frac{se_{\hat{\theta}_l}}{\hat{\theta}_l} , \quad \text{for every } l = 1, \dots, n+6. \tag{17}$$

The quantities defined by (14)–(17) are asymptotic, i.e., they are valid only if each dataset size M is sufficiently large because the nonlinear least-squares estimators are asymptotically normally distributed [3,13]. We computed the $nse_{\hat{\theta}}$ and the sensitivities of parameters $\hat{\theta}_l$, $S_{\hat{\theta}_l}(t)$, with a $nse_{\hat{\theta}_l}$ larger than 100%; see (16). Also, we plotted $S_{\hat{\theta}_l}(t)$ to justify why some parameters were estimated relatively reliably even if their $nse_{\hat{\theta}_l}$ were larger than 100%.

4. Results and discussion

In this section, we show and discuss the numerical results obtained according to the generalized SIR model with constant time delays (1), for each dataset corresponding to the selected regions of Chile for performing the calibration and prediction. We start by describing how we chose the different scenarios, followed by the fitting and prediction by using a figure that depicts the actual and calibrated infected, $I(t)$ and $f(t)I(t)$, respectively, computed according to our model. Next, we discuss parameter estimation's reliability by depicting the traditional sensitivity functions for those parameters having a *nse* larger than 100% if it applies; see (16)–(17). A table follows that shows quantitative measures to verify the convergence of the optimization algorithm. Concretely, the table shows the number of iterations made by the optimization solver to meet the stopping criterion, the sum of squares' optimal value, the RMSE, the first-order optimality measure, the norm of step, and the estimated parameters' values. Finally, we show numerical results to simulate an eventual change of scenario, which effect we measured by computing the relative errors for predicted data after the final time of calibration and discussing the number of infected' curve's long-time behavior.

We should warn that the actual values would depend on people behavior and of the measures taken by sanitary authorities of the state. Therefore, the results of the article should consider as a general trend on curves' behavior at simulation' conditions.

4.1. Results for the Antofagasta Region

Next, we show and discuss fitting and forecasting results for the Antofagasta Region, whose capital is Antofagasta, located 1093 km to the North of Santiago of Chile straight line. The region has $N = 691,854$ inhabitants, and it is representative of the country by its important ore activity. We fitted data starting from day 36 up to day 201, corresponding to April 7 and September 19, respectively (dataset size $M = 165$). We chose April 7 as the initial time because the number of active accumulated cases was 69.

The region had four of nine communes under confinement (Antofagasta, Mejillones, Calama, and Tocopilla) in the following periods. Antofagasta and Mejillones from May 5 to May 29, and from June 23 to September 28; Calama from June 9 to September 21, and Tocopilla from June 23 to August 9. On average and taking into account the population density, we defined four scenarios for Antofagasta Region as follows: $[t_0, t_1[= [36, 64[$ for the early expansion free of quarantine (between April 7 and May 5); $[t_1, t_2[= [64, 84[$ for the phase of fast propagation under the first quarantine (between May 5 and May 29); $[t_2, t_3[= [84, 113[$ for a phase of quarantine' relaxation with relatively fast propagation (between May 29 and June 23); and $[t_3, t_4[= [113, 201[$ for a phase of slowdown expansion under the second quarantine (between June 23 and September 19).

Fig. 1 depicts the model fitted to the Antofagasta Region dataset, plotting them each four-time points.
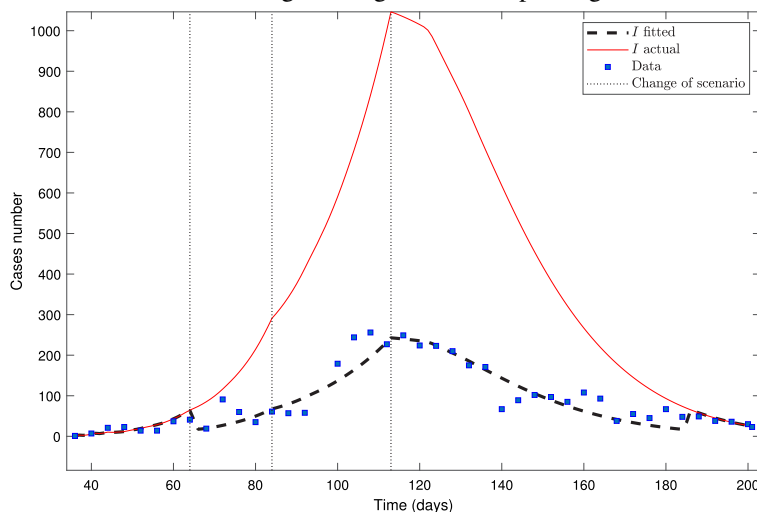

Fig. 1. Calibration for the Antofagasta Region dataset from April 7 up to September 19.

Fig. 1 shows that the calibrated and actual curves of infected are very different from each other, except in scenario 1 and near the end of scenario 4 of the pandemic. We expected this result since a few infected people would be almost all detected, so the calibrated and actual curves would coincide. On the other hand, we observe that the calibrated curve of infected fits quite well the data (infected reported $I_r(t)$). Indeed, since the dataset is large enough and encompasses the four stages of the pandemic, we estimated the parameters reliably. Only parameter $\hat{k}$, corresponding to the decay rate of fraction $f(t)$ (see (5)), has a *nse* larger than 100%; see Table 1. Its sensitivity function, $S_{\hat{k}}(t)$, varies in the interval $[-1.2e{-}05, 7.8e{-}06]$ for all t within the time interval of calibration; therefore, its estimate is unreliable. However, all the rest of the parameters have a *nse* less than 100%, so fitting is relatively reliable. Indeed, from Table 1, one can observe that the overall fitting is suitable since the error is quite low (sum of squares and RMSE).

Table 1

Fit performance for the Antofagasta Region dataset from April 7 up to September 19.

Quantity	Symbol	Value
Iterations	$p+1$	37
Sum of squares	$SS(\hat{\theta})$	$3.1149e-07$
RMSE	$\hat{\sigma}$	31.0148
First-order optimality	$\|\nabla SS(\hat{\theta})\|_\infty$	$2.71e-06$
Norm of step	$\delta_\theta^{(p)}$	$6.6371e-08$
Mean transmission rate	$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)$	$(3.9591e-01, 3.0843e-01, 2.3631e-01, 1.6749e-01)$
Normalized standard errors	$nse_{\hat{\beta}_l}, l=1,2,3,4$	(4.50%, 12.08%, 4.75%, 1.53%)
Mean removal rate	$\hat{\gamma}$	$1.8455e-01$
Normalized standard errors	$nse_{\hat{\gamma}}$	$\ll 0.01\%$
Mean time delays	$(\hat{\tau}_1, \hat{\tau}_2)$	(8.4047, 10.0222)
Normalized standard errors	$nse_{\hat{\tau}_l}, l=1,2$	(10.05%, 5.79%)
Parameters of fraction $f(t)$	$(\hat{a}, \hat{k}, \hat{I}_m)$	(23.20%, 1.0243, 68.8445)
Normalized standard errors	$nse_{\hat{a}, \hat{k}, \hat{I}_m}$	(37.02%, 512.03%, 15.99%)

The trend for the mean transmission rates is realistic. Indeed, $\hat{\beta}_1$ is the largest for scenario 1 (free of quarantine), followed by $\hat{\beta}_2$ for scenario 2 (with quarantine), followed by $\hat{\beta}_3$ for scenario 3 (free of quarantine), and finally $\hat{\beta}_4$ is the least of the four for scenario 4 (with quarantine). The same trend is observed for the fittings of the other datasets.

As explained in Section 3.1, we will show different predictions assuming that $\beta(t)$ may vary after September 19. Concretely, we run simulations by considering a mean transmission rate $\beta(t)$ defined by (7) taking $n=4$. The predicted data (not fitted) encompasses September 20 to October 05, $t_5 = 217$ that corresponds to this last date. Table 2 shows the range (minimum and maximum) and the average of the relative errors of the model' forecasting for different values of β_5, and where $t_4 = 210$ to try of capturing that none commune of the Antofagasta Region remained confined since September 28 (day 210).

Table 2

Predictions of the infected curve under different scenarios for the Antofagasta Region.

Value	Minimum	Maximum	Mean
$\beta_5 = 2\hat{\beta}_4$	$1.2416e-01$	1.3163	$4.3568e-01$
$\beta_5 = 1.5\hat{\beta}_4$	$1.4015e-01$	1.3163	$4.7957e-01$
$\beta_5 = 3\hat{\beta}_4$	$1.7351e-02$	1.3163	$4.9284e-01$
$\beta_5 = 1.25\hat{\beta}_4$	$1.3951e-01$	1.3163	$5.2653e-01$
$\beta_5 = 1.125\hat{\beta}_4$	$1.2278e-01$	1.3163	$5.4590e-01$
$\beta_5 = \hat{\beta}_4$	$1.4071e-01$	1.3163	$5.7411e-01$
$\beta_5 = 0.75\hat{\beta}_4$	$1.4134e-01$	1.3163	$6.2148e-01$
$\beta_5 = 0.5\hat{\beta}_4$	$1.4198e-01$	1.3163	$6.6890e-01$
$\beta_5 = 0.4\hat{\beta}_4$	$1.4227e-01$	1.3163	$6.8756e-01$
$\beta_5 = 0.3\hat{\beta}_4$	$1.5134e-01$	1.3163	$7.0866e-01$
$\beta_5 = 0.25\hat{\beta}_4$	$1.5275e-01$	1.3163	$7.1859e-01$
$\beta_5 = 0.125\hat{\beta}_4$	$1.4228e-01$	1.3827	$7.3947e-01$

From Table 2, we observe that $\beta_5 = 2\hat{\beta}_4$ yields the least relative error mean, which implies that after the quarantine finished on September 28, the mean rate of transmission doubled, producing a negative change of scenario. Our model forecasts that, if these conditions were kept, a COVID-19 second-outbreak would occur in

the Antofagasta Region. The peak of this new outbreak would occur around January 24, 2021, with an estimate of 102,154 real infected and 23,703 daily reported infected.

4.2. Results for the Metropolitan Region

Next, we show and discuss the results of fitting and forecasting for the Metropolitan Region (RM), whose capital is Santiago of Chile, also the capital of Chile. We fitted data starting from day 15 until day 203 corresponding to March 17 and September 21, respectively (dataset size M=189). We chose March 17 as the starting day because the number of newly infected with symptoms was 29, with 152 active accumulated cases.

The region has $N = 8,125,072$ inhabitants (around 41.76% of the total inhabitants of Chile) and 52 communes. From these, 47 communes have been under selective quarantine at different periods (few of which are yet), the most of which on average between May 7 and June 30. Therefore, we defined the three scenarios for MR as follows: $[t_0, t_1[= [15, 66[$ for the first stage that encompasses the period from March 17 to May 7 (early expansion free of quarantine on average); $[t_1, t_2[= [66, 120[$ for the second phase that ranges from May 7 to June 30 (fast propagation under quarantine); and $[t_2, t_3[= [120, 203[$ for the third step that goes from June 30 up to September 21 (slowdown expansion with relaxation of quarantine on average).

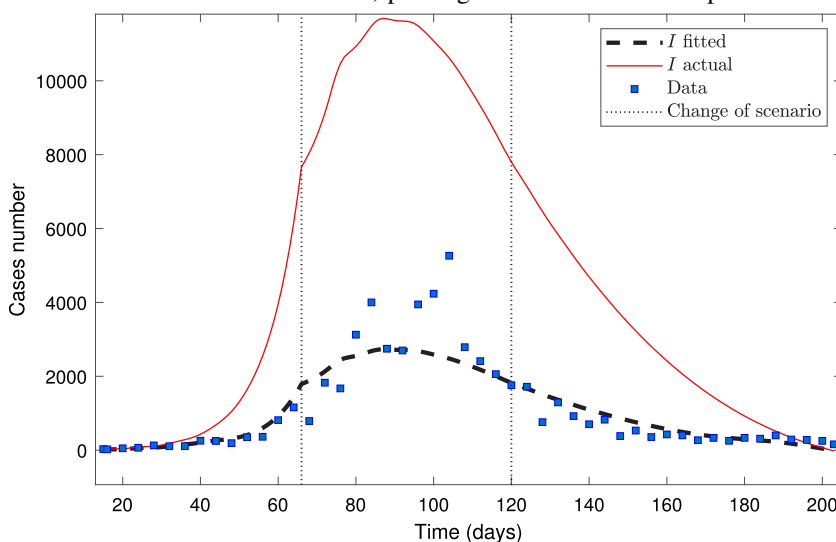Fig. 2 depicts the model fitted to the MR dataset, plotting them each four-time points.

Fig. 2. Calibration for MR dataset from March 17 up to September 21.

Fig. 2 shows that curves of calibrated and actual infected are very different from each other. On the other hand, we observe a good calibration of the model to the data since the computed curve of infected, $f(t)I(t)$, fits quite well the daily reported infected, $I_r(t)$. Indeed, since the dataset is large enough and encompasses the three stages of the pandemic, we estimated the parameters reliably, which is reinforced by the fact that the maximum nse is 89.39%, meaning accurate calibration results. To further appreciate the suitable fitting of the model, Table 3 presents its performance.

As explained in Section 3.1, we will show different predictions assuming that $\beta(t)$ may vary after September 21. Concretely, we run simulations by considering a mean rate of transmission $\beta(t)$ defined by (7) for $n = 3$. The predicted data (not fitted) encompasses September 22 to 28, $t_4 = 210$ that corresponds to this last date. Table 4 shows the range (minimum and maximum) and the average of the relative errors of the model' forecasting for different values of β_4 and for $t_3 = 182$ to try of capturing that almost none commune of the MR remained confined since August 31 (day 182).

From Table 4, we observe that $\beta_4 = 1.125\hat{\beta}_3$ yields the least relative error (minimum and mean), which reinforces the assumption that the quarantine is valid for the MR between May 7 and June 30, on average, and that there was a little change of scenario from August 31 until September 28. Our model forecasts that, if these conditions were kept, a COVID-19 second-outbreak would occur in the MR. The peak of this new outbreak would occur around May 24, 2021, with an estimate of 70,937 real infected and 16,568 daily reported infected.

Table 3
Fit performance for the MR dataset from March 17 up to September 21.

Quantity	Symbol	Value
Iterations	$p + 1$	34
Sum of squares	$SS(\hat{\theta})$	$9.0760e - 07$
RMSE	$\hat{\sigma}$	581.8176
First-order optimality	$\|\nabla SS(\hat{\theta})\|_\infty$	$9.07e - 06$
Norm of step	$\delta_\theta^{(p)}$	$1.5147e - 08$
Mean transmission rate	$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$	$(3.7426e - 01, 1.6679e - 01, 1.7155e - 01)$
Normalized standard errors	$nse_{\hat{\beta}_l}, l = 1, 2, 3$	$(6.14\%, 9.73\%, 8.69\%)$
Mean removal rate	$\hat{\gamma}$	$1.7429e - 01$
Normalized standard errors	$nse_{\hat{\gamma}}$	9.54%
Mean time delays	$(\hat{\tau}_1, \hat{\tau}_2)$	$(7.7609, 11.5809)$
Normalized standard errors	$nse_{\hat{\tau}_l}, l = 1, 2$	$(2.76\%, 3.60\%)$
Parameters of fraction $f(t)$	$(\hat{a}, \hat{k}, \hat{I}_m)$	$(23.36\%, 2.5772e - 03, 123.8884)$
Normalized standard errors	$nse_{\hat{a}, \hat{k}, \hat{I}_m}$	$(44.54\%, 89.39\%, \ll 0.01\%)$

Table 4
Predictions of the infected curve under different scenarios for the MR.

Value	Minimum	Maximum	Mean
$\beta_4 = 1.125\hat{\beta}_3$	$1.0763e - 01$	$3.8723e - 01$	$2.2169e - 01$
$\beta_4 = 1.25\beta_3$	$1.4605e - 02$	$7.3981e - 01$	$3.0414e - 01$
$\beta_4 = \hat{\beta}_3$	1.1228	1.5040	1.3079
$\beta_4 = 0.125\hat{\beta}_3$	$4.1640e - 01$	6.4217	2.6687
$\beta_4 = 0.75\beta_3$	3.3321	5.0633	4.0505
$\beta_4 = 1.5\beta_3$	1.3802	2.5700	1.9916
$\beta_4 = 0.25\hat{\beta}_3$	1.5586	7.1982	3.7048
$\beta_4 = 0.4\hat{\beta}_3$	3.1184	7.4159	4.5583
$\beta_4 = 0.5\hat{\beta}_3$	3.6698	7.1802	4.8149
$\beta_4 = 2\hat{\beta}_3$	6.0875	$1.0406e + 01$	7.7380
$\beta_4 = 3\hat{\beta}_3$	$2.2805e + 01$	$4.3345e + 01$	$2.9752e + 01$

4.3. Results for the Ñuble Region

Next, we show and discuss fitting and forecasting results for the Ñuble Region, located 403.8 km to the south of Santiago of Chile, and whose capital city is Chillán (population $N = 511, 551$). On September 21, Chillán was the fourth city with more active infected per 100,000 inhabitants (407 cases). The three scenarios considered here are more easily identifiable than for the MR and the Antofagasta Region since the quarantines in the Ñuble Region were imposed mostly in Chillán at a single date within the time interval of calibration.

We split the results into two parts; first, we did the fitting for the data from March 21 to April 27 and next from March 21 to September 02. Therefore, the scenarios are the following: $[t_0, t_1[= [19, 28[$ for the first stage that encompasses the period from March 21 up to March 30 (early expansion); $[t_1, t_2[= [28, 52[$ for the second phase that ranges from March 30 up to April 23 (fast propagation under quarantine); and $[t_2, t_3[$ where $t_2 = 52$ and $t_3 = 57$ or $t_3 = 184$ for the third step that goes from April 23 to April 27 and to September 2 (slowdown expansion, free of quarantine), corresponding to the first and second results, respectively.

4.3.1. Results for fitting until April 27

To compare with the only quantitative results we know of for the fitting of data from the Ñuble Region (see the first report by Cabrera-Vives et al. [6]), we did the fitting starting from day 19 (March 21) until day 57 (April 27) (dataset size $M = 39$), when the accumulated infected at the Ñuble Region were 58. It is worth noting that, because of the small number of data points, the third scenario does not represent a spreading's slowdown. We performed this fitting only to compare our results with that of Cabrera-Vives and collaborators' first report.
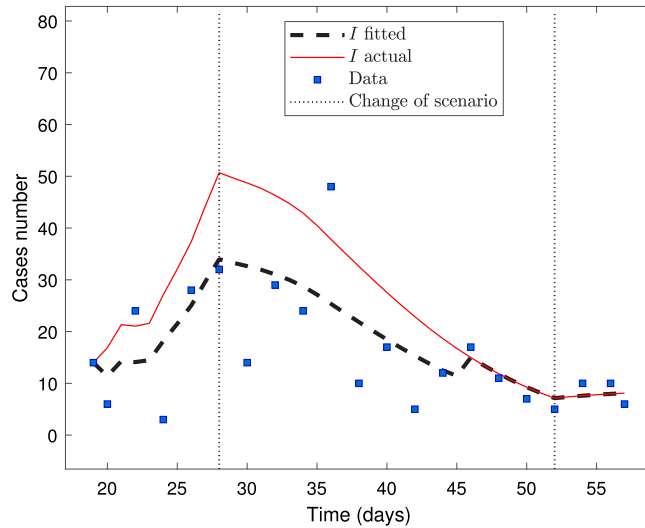
Fig. 3. Calibration for the Ñuble Region dataset from March 21 up to April 27.

Fig. 3 depicts the model fitted to the dataset for the Ñuble Region, plotting the data each two-time points.

From Fig. 3, we observe that the calibrated curve of infected does not fit suitably the data, despite that the optimization solver converged (the last change in the relative sum of squares is less than the value of the function tolerance). Indeed, the calibrated curve does not represent the trend of not fitted data from April 28 to May 02 (not reported here). Following the findings in [4], the explanation is that the data used to make the fitting is close to the zero equilibrium (the early stage of the pandemic), suggesting limited ability to determine some parameters reliably. This is related to the traditional sensitivity functions, depicted in Fig. 4, for the parameters estimated with a *nse* larger than 100%; see Table 5.
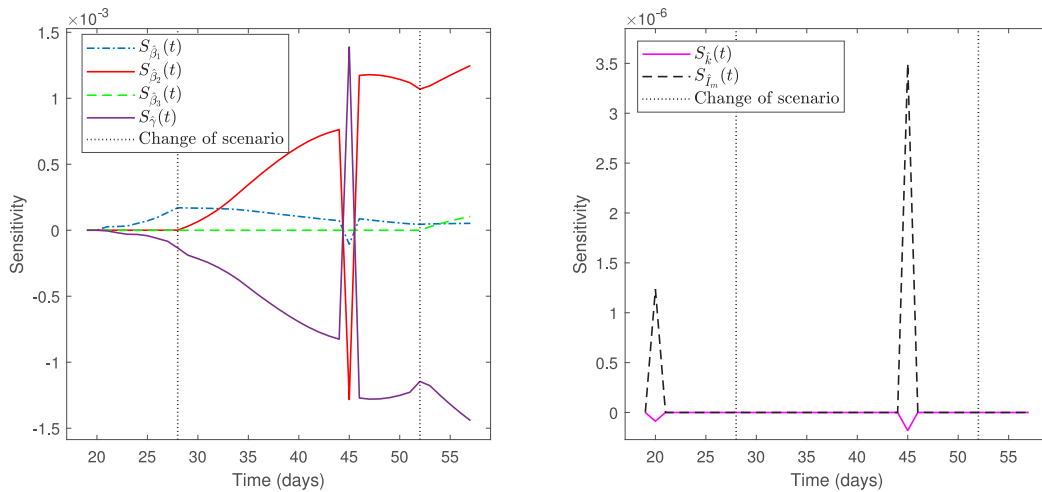

Fig. 4. Sensitivity functions for parameters estimated with a nse larger than 100%.

From Fig. 4(right), we observe that parameters $\hat{k}$ and $\hat{I}_m$ (see Eq. (5)) have the smallest sensitivities (very near to zero for every t), which implies that their estimations are not reliable. By contrast, Fig. 4(left) shows that parameter $\hat{\gamma}$ has the largest sensitivity, which means that its estimate is relatively better than the other parameters having a *nse* larger than 100%. In the case of parameters $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$, $S_{\hat{\beta}_1}(t)$ and $S_{\hat{\beta}_3}(t)$ are small in the regions of scenarios 1 and 3, where $\hat{\beta}_1$ and $\hat{\beta}_3$ represent the mean rates of transmission, respectively; by contrast, $S_{\hat{\beta}_2}(t)$ is large enough in the region of scenario 2, where $\hat{\beta}_2$ describes the mean rate of transmission. Consequently, the estimation of $\hat{\beta}_2$

is relatively better than those of $\hat{\beta}_1$ and $\hat{\beta}_3$. We will see that fitting the data in the three phases of the pandemic provides better results.

Table 5 shows the quantitative results of the fitting.

Table 5
Fit performance for the Ñuble Region dataset from March 21 up to April 27.

Quantity	Symbol	Value
Iterations	$p + 1$	41
Sum of squares	$SS(\hat{\theta})$	$2.0949e - 08$
RMSE	$\hat{\sigma}$	13.5179
First-order optimality	$\|\nabla SS(\hat{\theta})\|_\infty$	$2.19e - 08$
Norm of step	$\delta_\theta^{(p)}$	$1.0606e - 10$
Mean transmission rate	$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$	$(3.7849e - 01, 2.2784e - 02, 1.1015e - 01)$
Normalized standard errors	$nse_{\hat{\beta}_l}, l = 1, 2, 3$	(148.27%, 1163.89%, 363.29%)
Mean removal rate	$\hat{\gamma}$	$7.5470e - 02$
Normalized standard errors	$nse_{\hat{\gamma}}$	347.77%
Mean time delays	$(\hat{\tau}_1, \hat{\tau}_2)$	(5.4927, 7.1330)
Normalized standard errors	$nse_{\hat{\tau}_l}, l = 1, 2$	(44.74%, 25.31%)
Parameters of fraction $f(t)$	$(\hat{a}, \hat{k}, \hat{I}_m)$	(67.02%, 7.7127, 16.3431)
Normalized standard errors	$nse_{\hat{a}, \hat{k}, \hat{I}_m}$	(71.45%, 16, 164.15%, 409.00%)

The results are comparable with that from the first report by Cabrera-Vives et al. [6] (a similar RMSE), although we did the fitting by using the real date at which the authorities imposed the quarantine. Our model also forecasts that, if the same conditions of the third scenario were kept, the outbreak's peak would occur around February 20, 2021, that qualitatively coincides with the prediction obtained by Cabrera-Vives et al. [6]; see the bottom of Figure 1 from the first report. However, according to the previous discussion, the fitting obtained from data near an equilibrium point produces wrong prediction results [4,5].

4.3.2. Results until September 2

Now, we show the results of fitting and predictions for the data from Ñuble Region, widening the time window from March 12 to September 02 (dataset size $M = 175$), considering the same constraints as before (quarantine from March 30 to April 23), but now $t_3 = 184$ instead of 57 (day 184 corresponds to September 2). Fig. 5 depicts the model fitted to the dataset for the Ñuble Region, plotting the data at each four-time points
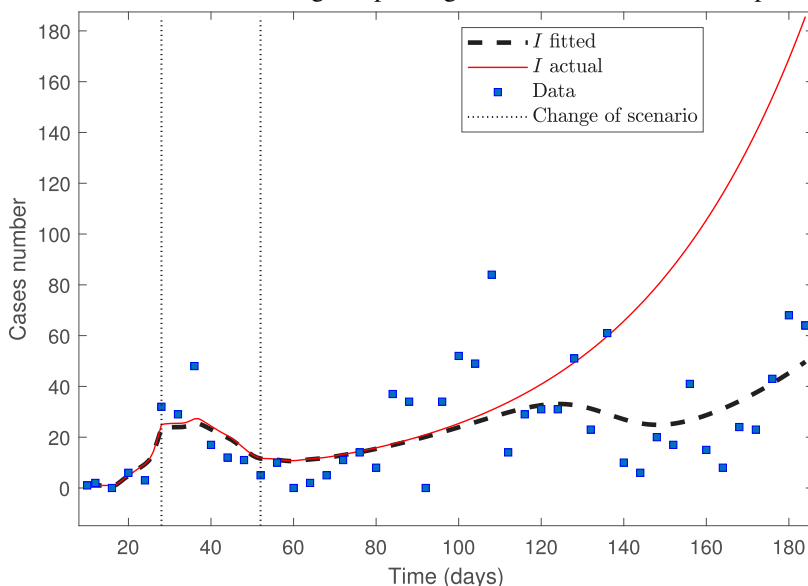

Fig. 5. Calibration for the Ñuble Region dataset from March 21 up to September 2.

From Fig. 5, we observe that the calibrated curve of infected fits quite well the data. In this case, since the dataset is large enough and encompasses data from the three stages of the pandemic, this time, we estimated the

parameters reliably, except maybe for $\hat{k}$ that is the only with a *nse* larger than 100%; see Table 6. Fig. 6 depicts the sensitivity function of $\hat{k}$.
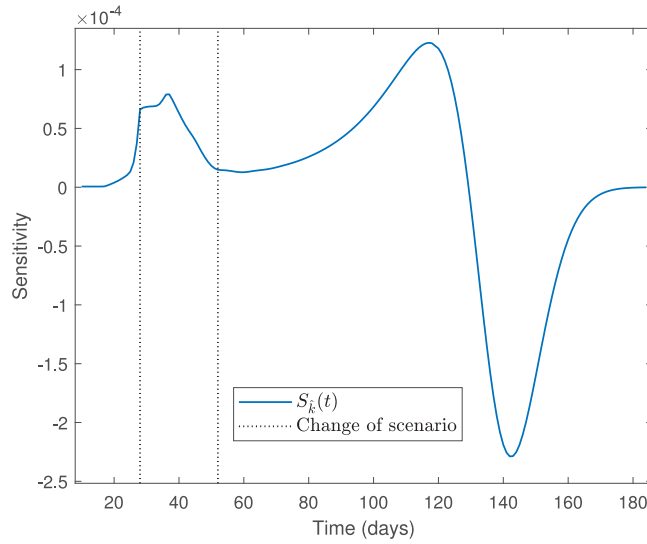

Fig. 6. Sensitivity function for parameter $\hat{k}$.

From Fig. 6, we observe that $S_{\hat{k}}(t)$ is relatively large in the region corresponding to scenario 3, and therefore estimate of $\hat{k}$ is relatively reliable. The result is better than the last fitting (data until April 27); see Fig. 4 and the discussion below it.

Table 6 shows the corresponding quantitative results of the fitting.

Table 6

Fit performance for the Ñuble Region dataset from 12 March to 02 September.

Quantity	Symbol	Value
Iterations	$p + 1$	68
Sum of squares	$SS(\hat{\theta})$	$1.3061e - 07$
RMSE	$\hat{\sigma}$	14.4360
First-order optimality	$\|\nabla SS(\hat{\theta})\|_\infty$	$3.65e - 07$
Norm of step	$\delta_\theta^{(p)}$	$1.1492e - 08$
Mean transmission rate	$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$	$(1.3248, 1.3487e - 01, 1.9592e - 01)$
Normalized standard errors	$nse_{\hat{\beta}_l}, l = 1, 2, 3$	$(19.65\%, 73.85\%, 39.71\%)$
Mean removal rate	$\hat{\gamma}$	$1.7578e - 01$
Normalized standard errors	$nse_{\hat{\gamma}}$	47.84%
Mean time delays	$(\hat{\tau}_1, \hat{\tau}_2)$	$(7.4305, 9.4810)$
Normalized standard errors	$nse_{\hat{\tau}_l}, l = 1, 2$	$(13.78\%, 13.73\%)$
Parameters of fraction $f(t)$	$(\hat{a}, \hat{k}, \hat{I}_m)$	$(26.78\%, 9.5539e - 02, 50.9513)$
Normalized standard errors	$nse_{\hat{a}, \hat{k}, \hat{I}_m}$	$(90.91\%, 163.47\%, 26.74\%)$

Consistently with the results obtained fitting the dataset until April 27, our model forecasts that, if the same conditions of the third scenario were kept, the outbreak's peak would occur around March 02, 2021. However, we would expect a change of scenario since September 2 because authorities re-imposed confinement in Chillán due to the high increase in daily new cases.

As explained in Section 3.1, we will show different predictions assuming that $\beta(t)$ may vary after September 2 due to the confinement (helpful or counterproductive measure depending on people behavior). Concretely, we run simulations by defining $\beta(t)$ as in (7) (with $n = 3$) to predict the data (not fitted) between September 3 and September 22 (data number = 20), and putting $t_3 = 184$ and $t_4 = 204$ that correspond to September 2 and 22, respectively. Table 7 shows the range (minimum and maximum) and the average of the relative errors of the model' forecasting for different values of β_4.

From Table 7, we observe that $\beta_4 = 0.75\hat{\beta}_3$ yields the least relative error maximum and mean. The interpretation is that quarantine effectively reduced the transmission rate, and therefore there was a positive change of scenario

Table 7
Predictions of the infected curve under different scenarios for the Ñuble Region.

Value	Minimum	Maximum	Mean
$\beta_4 = 0.75\hat{\beta}_3$	$1.8775e-02$	1.0677	$3.0481e-01$
$\beta_4 = 0.5\hat{\beta}_3$	$2.6869e-02$	2.9773	$7.2569e-01$
$\beta_4 = \hat{\beta}_3$	$5.2774e-02$	2.8704	$7.9289e-01$
$\beta_4 = 1.125\hat{\beta}_3$	$5.7721e-03$	3.8428	1.1393
$\beta_4 = 0.4\hat{\beta}_3$	$6.0453e-02$	4.8728	1.2234
$\beta_4 = 1.25\hat{\beta}_3$	$2.1683e-02$	5.0685	1.5241
$\beta_4 = 0.25\hat{\beta}_3$	$2.6635e-02$	7.1286	2.0765
$\beta_4 = 1.5\hat{\beta}_3$	$1.0001e-01$	8.2770	2.3404
$\beta_4 = 0.125\hat{\beta}_3$	$1.0034e-02$	8.4642	2.7584
$\beta_4 = 3\hat{\beta}_3$	$2.2141e-01$	$3.9640e+01$	9.1426

since September 2. Our model forecasts that, if these conditions were kept, a COVID-19 second-outbreak would not occur in the Ñuble Region.

Finally, Fig. 7 depicts the calibrated and real curves of infected for the Ñuble Region until September 25, by taking $\beta_4 = 0.75\hat{\beta}_3$ as before. The third peak corresponds to September 2, the final calibration date.
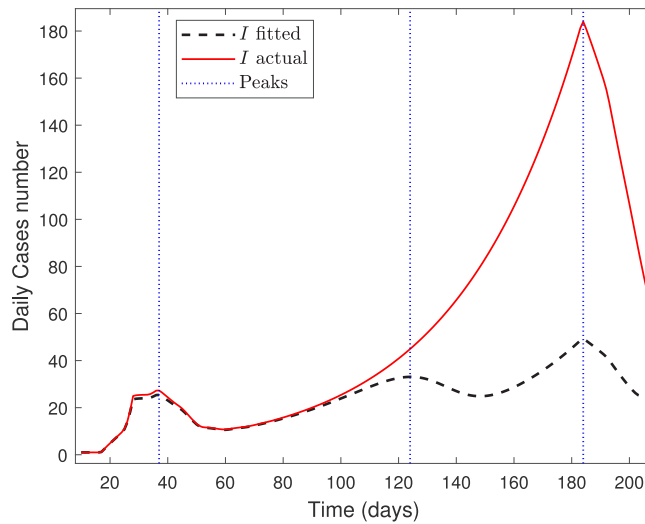

Fig. 7. Forecasted curve of infected for Ñuble Region from March 21 to September 25.

The result of Fig. 7 coincides with a report by the Chilean Society of Intensive Medicine, published on September 25, which pointed out that Chillán underwent the third peak of the COVID-19 outbreak by this date; see [20].

4.4. On the robustness of the optimization algorithm and the estimation of the infection fatality rate

To assess the robustness of the β parameter estimations we repeated the fitting processed we described in Section 4.2 after modifying some assumptions. Instead of assuming that $\beta(t)$ is piecewise constant and defining the three scenarios based on actual policy changes, here we considered $\beta(t)$ to be piecewise linear with $n-1$ equal length time intervals. More formally we replaced Eq. (6) with

$$\beta(t) = \sum_{i=1}^{n} \beta_i \mathbf{T}_i(t) \qquad (18)$$

where $\mathbf{T}_i$ is the tent function with value of one at t_i and tapers off to zero at t_{i-1} and t_{i+1}. Thus, we determine n values for the $\hat{\beta}_i$'s. We tried 5 values for n and we summarize the results in Table 8.

The results in Table 8 are, up to a point, compatible with the values shown in Table 3. For example, we can see that for values of $i \approx n/2$ the $\hat{\beta}_i$'s are in the $[0.13, 0.19]$ range consistently for $\hat{\beta}_2$ from Table 3. Moreover, for $n=4$

Table 8
Estimators for $\beta(t)$ using $n - 1$ segments piecewise linear functions.

n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
8	$7.5e - 01$	$8.2e - 02$	$2.1e - 01$	$1.4e - 01$	$1.8e - 01$	$2.1e - 01$	$1.2e - 01$	$4.2e - 01$
7	$1.3e - 01$	$1.6e - 01$	$1.7e - 01$	$1.6e - 01$	$9.2e - 02$	$5.8e - 02$	$2.4e - 01$	
6	$1.7e - 01$	$3.3e - 01$	$1.3e - 01$	$1.7e - 01$	$2.2e - 01$	$8.7e - 02$		
5	$5.3e - 01$	$1.9e - 01$	$1.4e - 01$	$1.9e - 01$	$9.0e - 02$			
4	$1.7e - 01$	$1.5e - 01$	$1.3e - 01$	$3.9e - 01$				

and $n = 5$ the same trend follows as in Table 3, i.e., the values for $\hat{\beta}_i$ are relatively large for both small and large i's (first and last scenario). We already stated in Section 4 that this was a consequence of the different scenarios considered for each dataset. The same trend is not seen if $n = 7$. This is natural, as having many time intervals, and therefore many $\hat{\beta}_i$'s parameters to adjust will tend to cause overfitting. Indeed, we observe across rows of Table 8 a relatively large variability (range $[0.13, 0.75]$) among the $\hat{\beta}_1$'s. The same can be said about the values for $\hat{\beta}_n$ (range $[0.09, 0.42]$). One consequence of this observation is that predictions are likely to be very inaccurate if we simply assume that $\hat{\beta}_n$ will hold during an extrapolation beyond the last data point after trying the fitting with a single n. Nonetheless, we can use the ranges the $\hat{\beta}_n$'s lie on for a scenario analysis as we did in Section 4 (Tables 2, 4 and 7). Another consequence is the difficulty to estimate precisely the effectiveness of the quarantine procedures. This follows from the variability within each row of Table 8. In principle, quarantines should lower the values for β. However, because of the inherent inaccuracy of the estimation during the initial disease propagation, and the enforcement of quarantines as soon as the cases multiply (and thus the estimation becomes reliable), computing the temporal changes in β accurately is more than challenging. Yet, for small values of n (say 4 or 5), the qualitative behavior of the β_i's we expect is observed.

Furthermore, it is of practical interest to estimate the lethality of a disease. The *infection fatality rate* (IFR), the probability of dying for an infected person, is computed as the ratio of dead subjects to the number of infected people for the targeted population [18]. However, according to our model, there is a time delay between that an infected person recovers or dies (τ_2), and therefore to compute the IFR at a given day t, we divided $D(t)$ (the number of death at t) into the number of infected at time $t - \tau_2$. To exemplify the model's usefulness, we will show IFR estimations computed for each dataset by combining empirical data and model outputs. To do that, we will use τ_2 from Tables 1, 3 and 6, $D(t)$ the confirmed number of deaths for each region reported in [17], and the actual number of infected $I(t)$ computed from the outputs of model (1). As usual, we divided the process into the scenarios, as described in Section 4, plus the pandemic's entire period used in the calibration. Table 9 shows the IFR medians in percentage.

Table 9
IFR median estimation for each dataset.

Region	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Entire
Antofagasta	0	0	$6.8758e - 01$	$8.8174e - 01$	$7.9727e - 01$
Metropolitan	1.1396	$4.4294e - 01$	1.0444	Does not apply	$9.6547e - 01$
Ñuble	0	0	0	Does not apply	0

We computed the median since it is more appropriate for skewed distribution, such as the confirmed number of deaths, $D(t)$, which are raw data. We did not plot the IFR since it presents many oscillations and a few outlier values because of uncertainty in data and the numerical error of calibration results. The values of Table 9 are consistent with the recent results in [18], where the author claims that IFR $\in [0\%, 1.54\%]$, which implies that the curve of real infected provided by the model is accurate since it yields a realistic value for the IFR's median in the case of MR. By contrast, for the Ñuble and Antofagasta Regions, the confirmed death cases are not large enough to yield an accurate IFR's estimation (for neither mean nor median).

5. Conclusions

We described and successfully implemented a clear, efficient, and reproducible parameter estimation methodology to a generalized SIR model with constant time delays that can reproduce complex dynamics for COVID-19. We illustrated our methodology and modeling by carrying out parameter estimation for three datasets corresponding to

Chile's three representative regions, although this can be applied to any country. From a methodological viewpoint, we assessed the reliability of estimated parameters and showed that when the data are located in the transition from the zero to the non-zero equilibria (encompassing all the stages of pandemic spread), the parameters are reliably estimated. Also, we verified our optimization methodology's robustness by considering an arbitrary number of n scenarios.

The numerical results allow us to forecast the general trend of the infected' curve, the calibrated and the real, and provide some predictions that allow us to prognosticate a possible COVID-19 second-outbreak. This prognosis is valid only if the conditions that allowed arrive at the estimated mean transmission rate are kept. In the absence of an effective vaccine or drug, these conditions are essentially translated into self-care. Finally, we showed that our model is precise enough to reproduce the infection fatality rate fairly accurately, according to the reported dead cases.

There are two lines of research that we would like to tackle. First, we expect to apply our model and methodology to other regions/communes in Chile. Because of the heterogeneity among their socio-economic conditions, the pandemic has affected them differently. On the other hand, we intend to calibrate our model to the reported dead to estimate more accurately and quantitatively the infection fatality rate of COVID-19, including structures of age and the dynamics of interactions among individuals of diverse age groups.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

References

[1] G. Badillo, P. Cumsille, L. Segura-Ponce, G. Pataro, G. Ferrari, An efficient optimization methodology of respiration rate parameters coupled with transport properties in mass balances to describe modified atmosphere packaging systems, Inverse Probl. Sci. Eng. 28 (10) (2020) 1361–1383, http://dx.doi.org/10.1080/17415977.2020.1717488.

[2] R. Baeza-Yates, Work group of data sciences on Covid-19 in Chile, 2020, https://www.youtube.com/watch?v=T26vGwcbxH4&feature=share&fbclid=.

[3] H.T. Banks, M. Davidian, J.R. Samuels, K.L. Sutton, An inverse problem statistical methodology summary, in: G. Chowell, J.M. Hyman, L.M.A. Bettencourt, C. Castillo-Chavez (Eds.), Mathematical and Statistical Estimation Approaches in Epidemiology, Springer Netherlands, Dordrecht, 2009, pp. 249–302, http://dx.doi.org/10.1007/978-90-481-2313-1_11.

[4] H.T. Banks, S. Dediu, S.L. Ernstberger, Sensitivity functions and their uses in inverse problems, J. Inverse Ill-Posed Probl. 15 (7) (2008) 683–708, http://dx.doi.org/10.1515/jiip.2007.038, https://doi.org/10.1515/jiip.2007.038.

[5] H.T. Banks, S.L. Ernstberger, S.L. Grove, Standard errors and confidence intervals in inverse problems: sensitivity and associated pitfalls, J. Inverse Ill-Posed Probl. 15 (1) (2007) 1–18, http://dx.doi.org/10.1515/JIIP.2007.001, https://doi.org/10.1515/JIIP.2007.001.

[6] G. Cabrera-Vives, C. Donoso-Oliva, M. Martínez, R. Molina, A. Sánchez, Informe proyecciones COVID-19 UdeC, 2020, https://github.com/guille-c/Covid-19/tree/master/Informes.

[7] A. Cancino, P. Gajardo, R. Lecaros, C. Muñoz, H. Ramírez, J. Ortega, COVID-19 en Chile, 2020, http://www.cmm.uchile.cl/?p=37663.

[8] M. Candia Reyes, G. Vergara-Hermosilla, Estimación de casos no reportados de infectados de COVID-19 en Chile, el Maule y la Araucanía durante marzo de 2020, Working Paper or Preprint, 2020, URL https://hal.archives-ouvertes.fr/hal-02560526.

[9] T.F. Coleman, Y. Li, An Interior Trust Region approach for nonlinear minimization subject to bounds, SIAM J. Optim. 6 (2) (1996) 418–445, http://dx.doi.org/10.1137/0806023.

[10] P. Cumsille, A. Coronel, C. Conca, C. Quiñinao, C. Escudero, Proposal of a hybrid approach for tumor progression and tumor-induced angiogenesis, Theoret. Biol. Med. Model. 12 (1) (2015) 13, http://dx.doi.org/10.1186/s12976-015-0009-y.

[11] P. Cumsille, M. Godoy, Z.P. Gerdtzen, C. Conca, Parameter estimation and mathematical modeling for the quantitative description of therapy failure due to drug resistance in gastrointestinal stromal tumor metastasis to the liver, PLoS One 14 (5) (2019) 1–27, http://dx.doi.org/10.1371/journal.pone.0217332.

[12] R. Elie, E. Hubert, G. Turinici, Contact rate epidemic control of COVID-19: an equilibrium view, Math. Model. Nat. Phenom. 15 (2020) 35, http://dx.doi.org/10.1051/mmnp/2020022.

[13] W.H. Greene, Econometric Analysis, seventh ed., Pearson Education, 2012.

[14] C. Guerrero-Nancuante, R. Manríquez P., An epidemiological forecast of COVID-19 in Chile based on the generalized SEIR model and the concept of recovered, Medwave 20 (04) (2020) e7898, http://dx.doi.org/10.5867/medwave.2020.04.7898.

[15] E. Hairer, S.P. Nørsett, G. Wanner, Solving Ordinary Differential Equations I: Nonstiff Problems, Springer Series in Computational Mathematics, vol. 8, Springer-Verlag Berlin Heidelberg, 1993, URL https://www.springer.com/gp/book/9783540566700.

[16] M. de Ciencia Tecnología Conocimiento e Innovación, Datos-COVID19, 2020, https://github.com/MinCiencia/Datos-COVID19/blob/master/output/producto5/TotalesNacionales.csv.

[17] M. de Ciencia Tecnología Conocimiento e Innovación, Datos-COVID19, 2020, https://github.com/MinCiencia/Datos-COVID19/blob/master/output/producto3/TotalesPorRegion.csv.

[18] J. Ioannidis, Infection fatality rate of COVID-19 inferred from seroprevalence data, Bull. World Health Organ. (2020) arXiv:https://www.who.int/bulletin/online_first/BLT.20.265892.pdf?ua=1. URL https://www.who.int/bulletin/online_first/en/.

[19] M. Maleewong, Time delay epidemic model for COVID-19, MedRxiv (2020) 2020.05.23.20111500. http://dx.doi.org/10.1101/2020.05.23.20111500. URL http://medrxiv.org/content/early/2020/05/26/2020.05.23.20111500.abstract.

[20] S. Núñez, A. Meleán, They warn that Chillán is going through the third peak of infections by Covid-19 (Advierten que Chillán atraviesa el tercer peak de contagios por Covid-19), The discussion" Newspaper (Diario la Discusión). URL: http://www.ladiscusion.cl/advierten-que-chillan-atraviesa-el-tercer-peak-de-contagios-por-covid-19/.

[21] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of COVID-19 in China by dynamical modeling, MedRxiv (2020) http://dx.doi.org/10.1101/2020.02.16.20023465, arXiv:https://www.medrxiv.org/content/early/2020/02/18/2020.02.16.20023465.full.pdf. URL https://www.medrxiv.org/content/early/2020/02/18/2020.02.16.20023465.

[22] L.F. Shampine, S. Thompson, Solving DDEs in MATLAB, Appl. Numer. Math. 37 (2001) 441–458, http://dx.doi.org/10.1016/s0168-9274(00)00055-6, URL http://gen.lib.rus.ec/scimag/index.php?s=10.1016/s0168-9274(00)00055-6.

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