Research paper

Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China

B. Ivorra [a], M.R. Ferrández [b], M. Vela-Pérez [a], A.M. Ramos [a,*]

[a] *MOMAT Research Group, Interdisciplinary Mathematics Institute, Complutense University of Madrid, Spain*
[b] *Computer Science Department, University of Almería, Spain*

A B S T R A C T

In this paper we develop a mathematical model for the spread of the coronavirus disease 2019 (COVID-19). It is a new θ-SEIHRD model (not a SIR, SEIR or other general purpose model), which takes into account the known special characteristics of this disease, as the existence of infectious undetected cases and the different sanitary and infectiousness conditions of hospitalized people. In particular, it includes a novel approach that considers the fraction θ of detected cases over the real total infected cases, which allows to study the importance of this ratio on the impact of COVID-19. The model is also able to estimate the needs of beds in hospitals. It is complex enough to capture the most important effects, but also simple enough to allow an affordable identification of its parameters, using the data that authorities report on this pandemic.

We study the particular case of China (including Chinese Mainland, Macao, Hong-Kong and Taiwan, as done by the World Health Organization in its reports on COVID-19), the country spreading the disease, and use its reported data to identify the model parameters, which can be of interest for estimating the spread of COVID-19 in other countries. We show a good agreement between the reported data and the estimations given by our model. We also study the behavior of the outputs returned by our model when considering incomplete reported data (by truncating them at some dates before and after the peak of daily reported cases). By comparing those results, we can estimate the error produced by the model when identifying the parameters at early stages of the pandemic. Finally, taking into account the advantages of the novelties introduced by our model, we study different scenarios to show how different values of the percentage of detected cases would have changed the global magnitude of COVID-19 in China, which can be of interest for policy makers.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Modeling and simulation are important decision tools that can be useful to control human and animal diseases [1–4]. However, since each disease exhibits its own particular biological characteristics, the models need to be adapted to each specific case in order to be able to tackle real situations [5,6].

Coronavirus disease 2019 (COVID-19) is an infectious disease emerging in China in December 2019 that has rapidly spread around China and many other countries (see [7]). On 11 February 2020, the World Health Organization (WHO) renamed the epidemic disease caused by 2019-nCoV as strain severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (see [8,9]).

This is a new virus and a completely new situation [10]. On 30 January 2020, WHO declared it to be a Public Health Emergency of International Concern [11]. As of 11 March 2020, the disease was confirmed in more than 118,000 cases reported globally in 114 countries, more than 90 percent of cases are in just four countries (two of those China and the Republic of Korea - have significantly declining epidemics) and WHO declared it to be a pandemic, the first one caused by a coronavirus [12]. On 1 April 2020 there are 872,481 and 43,275 official reported cases and deaths, respectively, and there is no vaccine specifically designed for this virus, with proven effectiveness.

There are some mathematical models in the literature that try to describe the dynamics of the evolution of COVID-19. Three phenomenological models are presented in [13], which were validated with outbreaks of other diseases different from COVID-19, trying to generate and assess short-term forecasts of the cumulative number of reported cases. Other works (see e.g. [14]), propose SEIR type models with little variations and some of them incorporate stochastic components. COVID-19 is a disease caused by a new virus, which is generating a worldwide emergency situation and needs a model taking into account its known specific characteristics. In particular, it would be convenient to develop a model which incorporates the following:

- the effect of undetected infected people (see [15]), being able to show the dependence of the impact of COVID-19 on the percentage of detected cases over the real total infected cases. Here, we denote by θ the fraction of detected cases over the real total infected cases. We note that there exist in the literature various works that propose estimations of θ in several countries (including China) for the COVID-19 pandemic (see, e.g., [15, 16]).
- the effect of different sanitary and infectiousness conditions of hospitalized people (differentiating those with mild and severe conditions that will recover from those who will finally die),
- the estimation of the needs of beds in hospitals (which is one of the major problems for policy makers addressing COVID-19).

The main goal of this paper is to develop a mathematical model well adapted to COVID-19 (not a SIR, SEIR or other general purpose epidemic models), taking into account the special characteristics mentioned previously. The model should be able to estimate, considering different scenarios, the number of cases, deaths and needs of beds in hospitals, in territories where COVID-19 is (or may be) a very serious health problem. It needs to be complex enough to capture the most important effects, but also simple enough to allow an affordable identification of its parameters, using the data that authorities report on this pandemic.

The θ-SEIHRD model developed here is based on the Be-CoDiS model (see [17]), designed to be able to study the spread of human diseases worldwide. That model was initially used for the 2014–2016 ebola outbreak (see [17]) and has also been used for the 2018–2020 ebola outbreak in the Democratic Republic of the Congo (see [18,19]), in both cases with very successful forecasts. Other works of our group related with the mathematical modeling of epidemics can be seen on the website https://www.ucm.es/momat/epidemics. In this paper, we have adapted that model to the specific case of COVID-19. Only within-country disease spread is considered in this paper for territories with a relevant number of people infected by COVID-19, where local transmission is the major cause of the disease spread. We point out that, when available, the parameters of the model are taken from the literature (see, e.g. [15,20]). Undocumented parameters or specific parameters to our approach are identified by using a multi-objective optimization approach, developed in a previous work (see [19]). We point out that we use a novel approach that relates the disease fatality rate with the percentage of detected cases over the real total infected cases, which allows to study the importance of this percentage on the impact of COVID-19.

Next, we validate our approach by studying the specific case of China (including the Chinese Mainland, Macao, Hong-Kong and Taiwan, as done by the World Health Organization in its reports on COVID-19; see [21]), which is the country with more data available to date. A good identification of the model parameters for this study case may be of interest for the application of the model to other territories. With that aim, we discuss and compare the results returned by our model with data reported by authorities on COVID-19 (see, e.g. [22–24]). In particular, we describe the evolution of the effective reproduction number R_e and compare the values with those found in the literature (see e.g. [25]).

We also study the behavior of the outputs returned by our model when considering incomplete data (truncated at some dates before and after the peak of daily reported cases). By comparing those results with real observation data we can estimate the error produced by the model when identifying the parameters at early stages of the epidemic.

Finally, we study different scenarios to show how different values of the percentage of detected cases would have changed the impact of COVID-19 in China, which can be of interest for policy makers.

2. Mathematical formulation of the model

In this section, we describe the θ-SEIHRD model. First of all, we present the epidemiological characteristics of COVID-19. Then, we introduce a general and a detailed description of our approach. Finally, we detail some of the model outputs that will be used for the numerical experiments performed later.

2.1. Epidemiological characteristics of COVID-19

According to the known characteristics of the COVID-19 pandemic, we assume that each person is in one of the following compartments (see [10]):

- Susceptible (denoted by S): The person is not infected by the disease pathogen.
- Exposed (denoted by E): The person is in the incubation period after being infected by the disease pathogen, and has no visible clinical signs. The individual could infect other people but with a lower probability than people in the infectious compartments. After the incubation period, the person passes to the Infectious compartment I.
- Infectious (denoted by I): After the incubation period, it is the first compartment of the infectious period, where nobody is expected to be detected yet. The person has finished the incubation period, may infect other people and starts developing clinical signs. After this period, people in this compartment can be, either taken in charge by sanitary authorities (and we classify them as hospitalized), or not detected by authorities and continue as infectious (but in other compartment, I_u).
- Infectious but undetected (denoted by I_u): After being in the compartment I, the person can still infect other people, have clinical signs but is not be detected and reported by authorities. We assume that only people with low or medium symptoms can reach this compartment, not the people who will die. After this period, people in this compartment pass to the Recovered compartment R_u.
- Hospitalized or in quarantine at home (but detected and reported by the authorities) that will recover (denoted by H_R): The person is in hospital (or in quarantine at home) and can still infect other people. At the end of this state, a person passes to the Recovered compartment R_d.
- Hospitalized that will die (denoted by H_D): The person is hospitalized and can still infect other people. At the end of this state, a person passes to the Dead compartment.
- Dead by COVID-19 (denoted by D): The person has not survived the disease.
- Recovered after being previously detected as infectious (denoted by R_d): The person was previously detected as infectious, survived the disease, is no longer infectious and has developed a natural immunity to the virus. When a person enters this compartment he/she remains in hospital for a convalescence period of C_o days (average).
- Recovered after being previously infectious but undetected (denoted by R_u): The person was not previously detected as infectious, survived the disease, is no longer infectious and has developed a natural immunity to the virus.

The authorities may apply various control measures in order to control the COVID-19 spread (see [26]):

- Isolation: Infected people are isolated from contact with other people. Only sanitary professionals are in contact with them. However, contamination of those professionals also occurs (see [27]). Isolated patients receive an adequate medical treatment that reduces the COVID-19 fatality rate.
- Quarantine: Movement of people in the area of origin of an infected person is restricted and controlled (e.g. quick sanitary check-points at the airports) to avoid that possible infected people spread the disease.
- Tracing: The objective of tracing is to identify potential infectious contacts which may have infected a person or spread COVID-19 to other people. Increase the number of tests in order to increase the percentage of detected infected people.
- Increase of sanitary resources: The number of operational beds and sanitary personal available to detect and treat affected people is increased, producing a decrease in the infectious period for the compartment I.

2.2. General description

The model is used to evaluate the spread of a human disease within some territories during a fixed time interval.

At the beginning of the simulation, the model parameters are set by the user (for instance, the values considered for COVID-19 are described in Section 3). We consider as time $t = 0$ the 1 December 2019 (7 days before the date that appears in the literature as the most probable date for the index case in China). Here we took into account that, according to the World Health Organisation's website, the first confirmed COVID-19 case in China was on December 8 (see [28]). Furthermore, according to [10], the earliest symptom onset of confirmed patients can be traced back to 7 December 2019. We set then $t = 0$ that day at 10AM CET (this is a technical adjustment, since each day the WHO provides data as reported by national authorities by that hour; see [21]).

We can start our simulation at any initial time $t_0 \geq 0$, considering that only susceptible people live in the territories that are free of the disease, whereas the number of people in compartments S, E, I, I_u, H_R, H_D, R_d, R_u, and D of the infected territories are set to their corresponding values. Then, during the time interval $[t_0, t_0 + T_{max}]$, with $T_{max} \in \mathbb{N}$ being the maximum number of simulation days, the within-country daily spread procedures (described in Section 2.3) are applied. If at the end of a simulation day t there is less than 1 person in each compartment E, I, I_u, H_R and H_D in all the territories considered, the simulation is stopped. Else, the simulation is stopped when $t = t_0 + T_{max}$. Furthermore, the control measures are also implemented and they can be activated or deactivated, when starting the model, in order to quantify their effectiveness to reduce the magnitude and duration of a COVID-19 epidemic.

A diagram summarizing the main structure of our full model is presented in Fig. 1. After describing the full model, we will present a simplified version (the one we will use for our simulations), with an also simpler diagram (see Fig. 2).
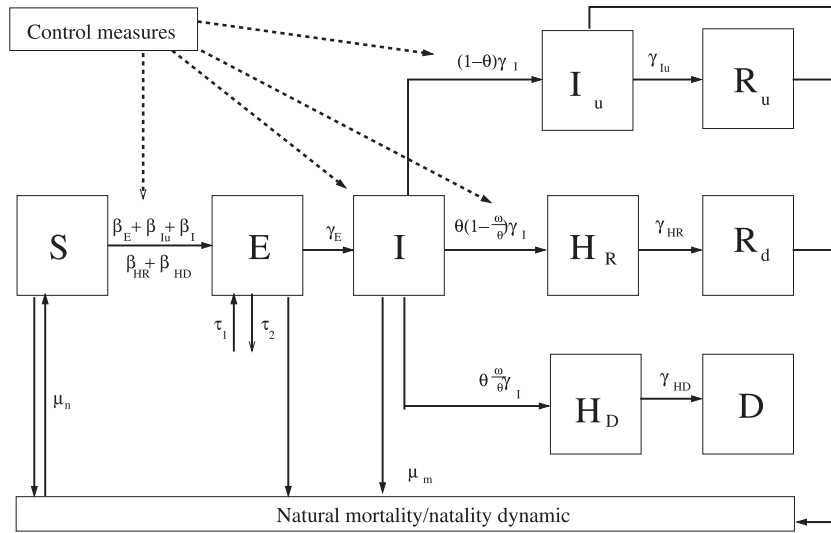
Fig. 1. Diagram summarizing the model for COVID-19 given by system (1).
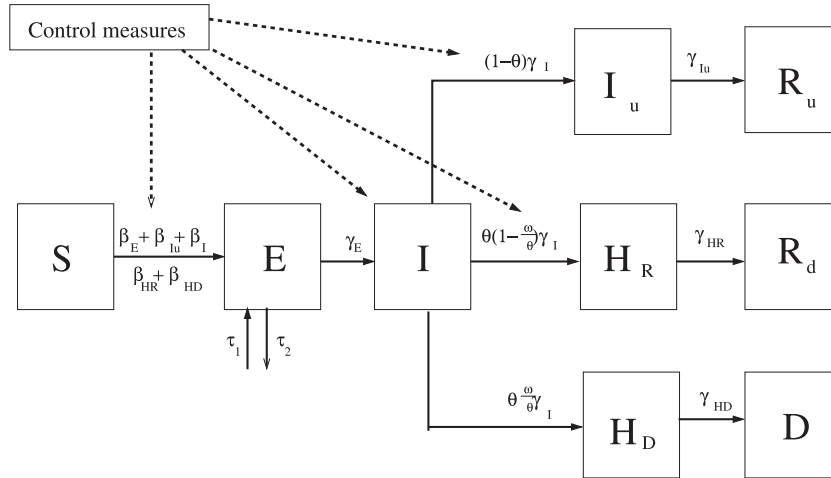

Fig. 2. Diagram summarizing the simplified version of the model for COVID-19 given by system (3).

The choice of using a deterministic model instead of a stochastic one is done as a first approach, since such kind of models present some advantages, such as: a low computational complexity allowing a better calibration of the model parameters or the possibility of using the theory of ordinary differential equations for suitably analyzing and interpreting the model. Furthermore, according to [29], deterministic models should be the first tool to be used when modelling a new problem with few data. The authors of that work also note that the stochastic models are not suitable when it is difficult or impossible to determine the distribution probability, are difficult to analyze and require more data for the calibration of the model.

As said above, in this work we will only consider the within-country disease spread of territories where, starting from suitable values of t_0, the COVID-19 pandemic is already spreading by its own, with a relatively negligible dependence of the international movement of people. The between-country disease spread will be considered in another work.

2.3. The mathematical model

The dynamic disease spread within a particular contaminated territory i is modeled by using a deterministic compartmental model (see, for instance, [5]).

We assume that people in a territory are characterized to be in one of those compartments, described in Section 2.1: *S, E, I, I_u, H_R, H_D, R_d, R_u* or *D*. For the sake of simplicity we assume that, at each time, the population inside a territory is homogeneously distributed (this can be improved by dividing some territories into a set of smaller regions with similar

characteristics). Thus, the spatial distribution of the epidemic inside a territory is omitted. We also assume that new births are susceptible people. We do not consider here movement of people between territories.

Under those assumptions, the evolution of the compartments mentioned above is modeled by the following system of ordinary differential equations (which is simplified below):

$$
\begin{aligned}
\frac{dS^{(i)}}{dt}(t) = &-\frac{S^{(i)}(t)}{N^{(i)}}\left(m_E^{(i)}(t)\beta_E^{(i)}E^{(i)}(t) + m_I^{(i)}(t)\beta_I^{(i)}I^{(i)}(t) + m_{I_u}^{(i)}(t)\beta_{I_u}^{(i)}(\theta^{(i)}(t))I_u^{(i)}(t)\right) \\
&-\frac{S^{(i)}(t)}{N^{(i)}}\left(m_{H_R}^{(i)}(t)\beta_{H_R}^{(i)}(t)H_R^{(i)}(t) + m_{H_D}^{(i)}(t)\beta_{H_D}^{(i)}(t)H_D^{(i)}(t)\right) \\
&-\mu_m^{(i)}S^{(i)}(t) + \mu_n^{(i)}\left(S^{(i)}(t) + E^{(i)}(t) + I^{(i)}(t) + I_u^{(i)}(t) + R_d^{(i)}(t) + R_u^{(i)}(t)\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{dE^{(i)}}{dt}(t) = &\frac{S^{(i)}(t)}{N^{(i)}}\left(m_E^{(i)}(t)\beta_E^{(i)}E^{(i)}(t) + m_I^{(i)}(t)\beta_I^{(i)}I^{(i)}(t) + m_{I_u}^{(i)}(t)\beta_{I_u}^{(i)}(\theta^{(i)}(t))I_u^{(i)}(t)\right) \\
&+\frac{S^{(i)}(t)}{N^{(i)}}\left(m_{H_R}^{(i)}(t)\beta_{H_R}^{(i)}(t)H_R^{(i)}(t) + m_{H_D}^{(i)}(t)\beta_{H_D}^{(i)}(t)H_D^{(i)}(t)\right) \\
&-\mu_m^{(i)}E^{(i)}(t) - \gamma_E E^{(i)}(t) + \tau_1^{(i)}(t) - \tau_2^{(i)}(t),
\end{aligned}
$$

$$
\frac{dI^{(i)}}{dt}(t) = \gamma_E E^{(i)}(t) - (\mu_m^{(i)} + \gamma_I^{(i)}(t))I^{(i)}(t),
$$

$$
\frac{dI_u^{(i)}}{dt}(t) = (1 - \theta^{(i)}(t))\gamma_I^{(i)}(t)I^{(i)}(t) - (\mu_m^{(i)} + \gamma_{I_u}^{(i)}(t))I_u^{(i)}(t),
$$

$$
\frac{dH_R^{(i)}}{dt}(t) = \theta^{(i)}(t)\left(1 - \frac{\omega^{(i)}(t)}{\theta^{(i)}(t)}\right)\gamma_I^{(i)}(t)I^{(i)}(t) - \gamma_{H_R}^{(i)}(t)H_R^{(i)}(t),
$$

$$
\frac{dH_D^{(i)}}{dt}(t) = \omega^{(i)}(t)\gamma_I^{(i)}(t)I^{(i)}(t) - \gamma_{H_D}^{(i)}(t)H_D^{(i)}(t),
$$

$$
\frac{dR_d^{(i)}}{dt}(t) = \gamma_{H_R}^{(i)}(t)H_R^{(i)}(t) - \mu_m^{(i)}R_d^{(i)}(t),
$$

$$
\frac{dR_u^{(i)}}{dt}(t) = \gamma_{I_u}^{(i)}(t)I_u^{(i)}(t) - \mu_m^{(i)}R_u^{(i)}(t),
$$

$$
\frac{dD^{(i)}}{dt}(t) = \gamma_{H_D}^{(i)}(t)H_D^{(i)}(t), \tag{1}
$$

where:

- $i \in \{1, \ldots, N_C\}$, with $N_C \in \mathbb{N}$ being the number of countries/territories/areas considered.
- $N^{(i)}$ is the number of people in territory i before the start of the pandemic.
- $\mu_n^{(i)} \in [0, 1]$ is the natality rate (day^{-1}) in territory i: the number of births per day and per capita.
- $\mu_m^{(i)} \in [0, 1]$ is the mortality rate (day^{-1}) in territory i: the number of deaths per day and per capita (or, equivalently, the inverse of the mean life expectancy (day) of a person).
- $\omega^{(i)}(t) \in [\underline{\omega}^{(i)}, \overline{\omega}^{(i)}] \subset [0, 1]$ is the case fatality rate in territory i at time t: the proportion of deaths compared to the total number of infectious people (detected or not detected). Here, $\underline{\omega}^{(i)}$ and $\overline{\omega}^{(i)}$ are the minimum and maximum case fatality rates for territory i, respectively.
- $\theta^{(i)}(t) \in [\overline{\omega}^{(i)}, 1]$ is the fraction of infected people that are detected and documented by the authorities in territory i at time t. We are assuming, for the sake of simplicity, that all deaths due to COVID-19 are detected and reported, so that $\theta^{(i)}(t) \geq \overline{\omega}^{(i)}$. Estimations of θ for the COVID-19 pandemic in several countries (including China) can be found in the literature (see, e.g., [15, 16]).
- $\beta_E^{(i)}, \beta_I^{(i)}, \beta_{I_u}^{(i)}, \beta_{H_R}^{(i)}, \beta_{H_D}^{(i)} \in \mathbb{R}^+$ are the disease contact rates (day^{-1}) of a person in the corresponding compartments, in territory i (without taking into account the control measures).
- $\beta_{I_u}^{(i)}(\theta) \in \mathbb{R}^+$ is the disease contact rate (day^{-1}) of a person in the compartment I_u in territory i (without taking into account the control measures), where the fraction of infected people that are detected are θ.
- $\gamma_E \in (0, +\infty)$ is the transition rate (day^{-1}) from compartment E to compartment I. It is the same for all the territories.
- $\gamma_I^{(i)}(t) \in (0, +\infty)$ is the transition rate (day^{-1}) from compartment I to compartments I_u, H_R or H_D in territory i at time t.
- $\gamma_{I_u}^{(i)}(t), \gamma_{H_R}^{(i)}(t)$ and $\gamma_{H_D}^{(i)}(t) \in (0, +\infty)$ denote the transition rates (day^{-1}) from compartments I_u, H_R and H_D to compartments R_u, R_d and D, respectively, in territory i at time t.
- $m_E^{(i)}(t), m_I^{(i)}(t), m_{I_u}^{(i)}(t), m_{H_R}^{(i)}(t), m_{H_D}^{(i)}(t) \in [0, 1]$ (%) are functions representing the efficiency of the control measures applied to the corresponding compartments, in territory i at time t.

- $\tau_1^{(i)}(t)$ is the people infected that arrives to territory i from other territories per day. $\tau_2^{(i)}(t)$ is the people infected that leaves territory i per day. Both can be modeled following the between-country spread part of the Be-CoDiS model, see [17]).

System (1) is completed with initial data $S^{(i)}(t_0)$, $E^{(i)}(t_0)$, $I^{(i)}(t_0)$, $I_u^{(i)}(t_0)$, $H_R^{(i)}(t_0)$, $H_D^{(i)}(t_0)$, $R_d^{(i)}(t_0)$, $R_u^{(i)}(t_0)$ and $D^{(i)}(t_0)$ given in $[0, \infty)$.

We point out that the 9th equation of system (1) is not coupled with the other equations. Thus, we can solve the first eight equations of that system and the solution of the last one can be computed as follows:

$$D^{(i)}(t) = D^{(i)}(t_0) + \int_{t_0}^{t} \gamma_{H_D}^{(i)}(s) H_D^{(i)}(s)\,\mathrm{d}s. \tag{2}$$

From a modeling point of view, the term $\frac{\omega^{(i)}(t)}{\theta^{(i)}(t)}$ in the 5th equation of system (1) corresponds to the apparent fatality rate of the disease (obtained by considering only detected cases) in territory i at time t, and $\omega^{(i)}(t)$ is the real fatality rate of the COVID-19 (taking into account detected and undetected cases).

We could have also considered sub-compartments organized by ages, etc., but that would complexify the model and would hinder the identification of parameters. For instance, we could use in each compartment a different natality rate for the people infected (according to their age distribution). Actually, since the natality and mortality (not from COVID-19) do not seem to be important factors for COVID-19 (at least for relatively short periods of time) one may consider the simplified model with $\mu_m = \mu_n = 0$. We will consider this case in the rest of this paper. The corresponding new diagram summarizing the main structure of the simplified model can be seen in Fig. 2 and the resulting system equations, after removing the index i denoting different territories (for the sake of simplicity), is given by the system (3) below.

$$\frac{\mathrm{d}S}{\mathrm{d}t}(t) = -\frac{S(t)}{N}\left(m_E(t)\beta_E E(t) + m_I(t)\beta_I I(t) + m_{I_u}(t)\beta_{I_u}(\theta(t))I_u(t)\right)$$
$$-\frac{S(t)}{N}\left(m_{H_R}(t)\beta_{H_R}(t)H_R(t) + m_{H_D}(t)\beta_{H_D}(t)H_D(t)\right),$$

$$\frac{\mathrm{d}E}{\mathrm{d}t}(t) = \frac{S(t)}{N}\left(m_E(t)\beta_E E(t) + m_I(t)\beta_I I(t) + m_{I_u}(t)\beta_{I_u}(\theta(t))I_u(t)\right)$$
$$+\frac{S(t)}{N}\left(m_{H_R}(t)\beta_{H_R}(t)H_R(t) + m_{H_D}(t)\beta_{H_D}(t)H_D(t)\right) - \gamma_E E(t) + \tau_1(t) - \tau_2(t),$$

$$\frac{\mathrm{d}I}{\mathrm{d}t}(t) = \gamma_E E(t) - \gamma_I(t)I(t),$$

$$\frac{\mathrm{d}I_u}{\mathrm{d}t}(t) = (1-\theta(t))\gamma_I(t)I(t) - \gamma_{I_u}(t)I_u(t),$$

$$\frac{\mathrm{d}H_R}{\mathrm{d}t}(t) = \theta(t)\left(1 - \frac{\omega(t)}{\theta(t)}\right)\gamma_I(t)I(t) - \gamma_{H_R}(t)H_R(t),$$

$$\frac{\mathrm{d}H_D}{\mathrm{d}t}(t) = \omega(t)\gamma_I(t)I(t) - \gamma_{H_D}(t)H_D(t),$$

$$\frac{\mathrm{d}R_d}{\mathrm{d}t}(t) = \gamma_{H_R}(t)H_R(t),$$

$$\frac{\mathrm{d}R_u}{\mathrm{d}t}(t) = \gamma_{I_u}(t)I_u(t),$$

$$\frac{\mathrm{d}D}{\mathrm{d}t}(t) = \gamma_{H_D}(t)H_D(t). \tag{3}$$

We remark that, with this simplification the 7th, 8th and 9th equations of system (3) are not coupled with the other equations. Thus, we can solve the first six equations of that system and the solution of the last three equations can be computed by using (2) and

$$R_d(t) = R_d(t_0) + \int_{t_0}^{t} \gamma_{H_R}(s)H_R(s)\,\mathrm{d}s,$$

$$R_u(t) = R_u(t_0) + \int_{t_0}^{t} \gamma_{I_u}(s)I_u(s)\,\mathrm{d}s.$$

For the numerical simulations presented in Section 4, the first six equations of system (3) were numerically solved with the classic 4th order Runge–Kutta method with four stages (RK4) and with four hours as time step (which was tested to see that it is suitable to get stable results).

2.4. Outputs of the model

Here, we present the outputs of the mathematical model (for territory i; we continue avoiding the use of that index in the notation, for the sake of simplicity), used to analyze the results of the simulations performed in Section 4:

- $c_{\mathrm{m}}(t)$: The model cumulative number of COVID-19 cases (in country i) at day t, which is given by

$$c_{\mathrm{m}}(t) = H_R(t) + H_D(t) + R_{\mathrm{d}}(t) + D(t),$$

and can be also computed as follows:

$$c_{\mathrm{m}}(t) = c_{\mathrm{m}}(t_0) + \int_{t_0}^{t} \frac{\mathrm{d}\left(H_R + H_D + R_{\mathrm{d}} + D\right)}{\mathrm{d}t}(s)\mathrm{d}s = c_{\mathrm{m}}(t_0) + \int_{t_0}^{t} \theta(s)\gamma_I(s)I(s)\mathrm{d}s.$$

- $d_{\mathrm{m}}(t)$: The model cumulative number of deaths (due to COVID-19), at day t (in territory i), which is given by $D(t)$.
- R_0 and R_e: The basic reproduction number and the effective reproduction number of COVID-19 (for territory i).
  The basic reproduction number is defined as the number of cases one infected person generates on average over the course of its infectious period, in an otherwise uninfected population and without special control measures. It depends on the considered population (therefore, it may be different for different territories), but does not change during the spread of the disease.
  The effective reproduction number is defined as the number of cases one infected person generates on average over the course of its infectious period. Part of the population can be already infected and/or special control measures may have been implemented. It depends on the considered population and changes during the spread of the disease. Furthermore, $\mathrm{R}_e(0) = \mathrm{R}_0$. Typically, the spread of the disease slows down when $\mathrm{R}_e(i, t) < 1$.
  We apply the Next Generation Matrix method (see [30]) to system (3) and obtain that:

$$\mathrm{R}_0 = \frac{\left(\left(\left(\beta_{I_{\mathrm{u}}}(1-\theta)\gamma_{H_R} + \beta_{H_R}\gamma_{I_{\mathrm{u}}}(\theta-\omega)\right)\gamma_I + \beta_I\gamma_{H_R}\gamma_{I_{\mathrm{u}}}\right)\gamma_E + \beta_E\gamma_I\gamma_{H_R}\gamma_{I_{\mathrm{u}}}\right)\gamma_{H_D} + \beta_{H_D}\omega\gamma_E\gamma_I\gamma_{H_R}\gamma_{I_{\mathrm{u}}}}{\gamma_E\gamma_I\gamma_{H_D}\gamma_{H_R}\gamma_{I_{\mathrm{u}}}},$$

where, for the sake of simplicity of notation, all previous coefficients correspond to their particular values at time $t = 0$. Furthermore,

$$\mathrm{R}_e(t) = \frac{U_e(t)}{\gamma_E\gamma_I\gamma_{H_D}\gamma_{H_R}\gamma_{I_{\mathrm{u}}}} \frac{S(t)}{N},$$

where

$$U_e(t) = \left(\left(\left(m_{I_{\mathrm{u}}}\beta_{I_{\mathrm{u}}}(1-\theta)\gamma_{H_R} + m_{H_R}\beta_{H_R}\gamma_{I_{\mathrm{u}}}(\theta-\omega)\right)\gamma_I + m_I\beta_I\gamma_{H_R}\gamma_{I_{\mathrm{u}}}\right)\gamma_E + m_E\beta_E\gamma_I\gamma_{H_R}\gamma_{I_{\mathrm{u}}}\right)\gamma_{H_D}$$
$$+ m_{H_D}\beta_{H_D}\omega\gamma_E\gamma_I\gamma_{H_R}\gamma_{I_{\mathrm{u}}},$$

and, again in order to simplify the notation, all previous coefficients correspond to their particular values at time t.

- $\mathrm{Hos}(t)$: The number of people in hospital is estimated as follows:

$$\mathrm{Hos}(t) = H_D(t) + p(t)(H_R(t) + (R_{\mathrm{d}}(t) - R_{\mathrm{d}}(t - C_{\mathrm{o}}))),$$

where $p(t)$ is the fraction, at time t, of people in compartment H_R that are hospitalized and C_{o} days is the period of convalescence (i.e. the time a person is still hospitalized after recovering from COVID-19). This function can help to estimate and plan the number of clinical beds needed to treat all the COVID-19 cases.

- MHos: The maximum number of hospitalized people at the same time (in territory i) during the time interval $[t_0, T]$. It is computed as:

$$\mathrm{MHos} = \max_{t \in [t_0, T]} \mathrm{Hos}(t).$$

This number can help to estimate and plan the number of clinical beds needed to treat all the COVID-19 cases.

- $\Gamma_E(t)$, $\Gamma_{I_u}(t)$ and $\Gamma_H(t)$: The number of people infected during the time interval $[t_0, T]$, by contact with people in compartment E, I_u and $H = H_R + H_D$, respectively. They are given by:

$$\Gamma_E(t) = \int_{t_0}^{t} m_E(s)\beta_E E(s)\frac{S(s)}{N}\mathrm{d}s,$$

$$\Gamma_{I_u}(t) = \int_{t_0}^{t} m_{I_u}(s)\beta_{I_u} Iu(s)\frac{S(s)}{N}\mathrm{d}s,$$

$$\Gamma_H(t) = \int_{t_0}^{t} \left(m_{H_R}(s)\beta_{H_R}H_R(s) + m_{H_D}(s)\beta_{H_D}H_D(s)\right)\frac{S(s)}{N}\mathrm{d}s.$$

We point out that $c_{\mathrm{m}}(t)$ and $d_{\mathrm{m}}(t)$ can be compared with the corresponding values reported by WHO (see [27])

Table 1

Summary of some parameters calibrated for COVID-19 in China, which can be found in [20,27,34]. A brief description (**Description**) and the range of the considered values (**Value**) are reported.

Notation	Value	Description	Reference
γ_E	0.1818	Transition rate of a person in compartment E (day^{-1})	[27]
$\gamma_I(t)$	[0.1493,1.4286]	Transition rate of a person in compartment I(day^{-1}) at time t	[20]
$\gamma_{H_R}(t)$	[0.0752,0.1370]	Transition rate of a person in compartment H_R to compartment $R_{\rm d}$ (day^{-1}) at time t	[27]
$\gamma_{I_u}(t)$	[0.0752,0.1370]	Transition rate of a person in compartment I_u to compartment $R_{\rm u}$ (day^{-1}) at time t	[27]
$C_{\rm o}$	14	The period of convalescence (day)	[34]

3. Model parameter estimation for COVID-19

Some of the parameters used in the simulations presented in Section 4 have been found in the literature. However, despite the effort to use the maximum amount of robust parameters as possible, and due to lack of information of the behavior of the SARS-CoV-2, some of them have been estimated using empirical assumptions or techniques for parameter identification. This part should be improved as soon as new information is available.

We now detail each kind of parameter by its category.

3.1. Territory indicators obtained from databases

We have built a database where we have recorded the following data, that we use to estimate the parameters of the model and to compare with the outputs of the numerical simulations presented in Section 4

We obtain the countries' population (N) for year 2018 from the World Data Bank, downloaded on 5 March 2020 (see [31]).

There are some studies about the relevance of humidity and temperature for the spread of COVID-19. In [32], it is shown that the observed patterns of COVID-19 are not completely consistent with the hypothesis that high absolute humidity may limit the survival and transmission of this virus. Furthermore, in [33] it has been found that the lower is the temperature, the greater is the survival period of the SARS-CoV-2 outside the host. Since there is no clear scientific evidence of the effect of the humidity and the temperature on the SARS-CoV-2, we have not included these two factors in our model (this would need to be revised in case of appearing new findings regarding this topic).

Furthermore, from the WHO Coronavirus disease (COVID-2019) situation reports (see [21]), we have obtained the reported cases per day (i.e. people who enter each day in compartments H_R or H_D) and the reported COVID-19 deaths per day. For each day t, they are denoted by $c_{\rm r}(t)$ and $d_{\rm r}(t)$, respectively. We have those data starting from 21 January 2020. Data from 12 to 21 January 2020 were obtained from the following website:

https://bnonews.com/index.php/2020/02/the-latest-coronavirus-cases/

We assume that the index case started his incubation period in China on 1 December 2019, as the earliest symptom onset of confirmed patients can be traced back to 7 December 2019 (see [10]). In order to complete the missing information for the days without data between 8 December 2019 and 10 January 2020, we use a cubic hermite interpolation method considering the fact that before 8 December 2019 no cases were reported.

3.2. Parameters found in the literature

Data given in Table 1 were calibrated for the cases in China, the first big source of COVID-19. However, due to the spread of this disease, new studies should be performed to analyze the behavior of COVID-19 in other sanitary, population and climatic conditions. Currently, very few studies accepted for the scientific community are available.

3.3. Control measures

Focusing on the application of the control measures, we multiply the disease contact rates (i.e. β_E, β_I, β_{I_u}, β_{H_R} and β_{H_D}) by decreasing functions simulating the reduction of these rates as the control measures efficiency is increased. Here, we have considered the functions (see [35]):

$$m_E(t) = m_{I_u}(t) = m_I(t) = m_{H_R}(t) = m_{H_D}(t) = \begin{cases} (m_0 - m_1)\exp\left(-\kappa_1(t-\lambda_0)\right) + m_1 & t \in [t_0, \lambda_1], \\ (m(\lambda_1) - m_2)\exp\left(-\kappa_2(t-\lambda_1)\right) + m_2 & t \in (\lambda_1, \lambda_2], \\ \dots \\ (m(\lambda_{q-1}) - m_q)\exp\left(-\kappa_q(t-\lambda_{q-1})\right) + m_q & t \in (\lambda_{q-1}, \infty), \end{cases} \quad (4)$$

where, for every $j \in \{0, \ldots, q\}$, $m_j \in [0, 1]$ measures the intensity of the control measures (greater value implies lower value of disease contact rates), $\kappa_j \in [0, +\infty)$ (day^{-1}) simulates the efficiency of the control measures (greater value implies lower value of disease contact rates) and $\lambda_j \in [t_0, \infty]$ denotes the first day of application of each control strategy ($\lambda_0 \in [0, \infty]$ is the first day of application of a control strategy that was being used before t_0, if any). Here, $q \in \mathbb{N}$ is the number of changes of control strategy.

The values of λ_j are typically taken from the literature (using the dates when the territories implement special control measures). Some of the values of m_j can be also sometimes known. The rest of the parameters need to be calibrated as explained in Section 3.7.

3.4. Case fatality rate

The case fatality rate $\omega(t)$ depends on the territory i and time t. We recall that ω is the proportion of deaths compared to the total number of infectious people (i.e. including detected and undetected people). As observed in other epidemics (see, e.g. [17]), it can be affected by the application of control measures (such as: earlier detection, better sanitary conditions, etc.). Thus, we propose to consider the following definition:

$$\omega(t) = m_I(t)\overline{\omega} + (1 - m_I(t))\underline{\omega}, \tag{5}$$

where $\overline{\omega} \in [0, 1]$ is the case fatality rate when no control measures are applied (i.e. $m_I(t) \equiv 1$); and $\underline{\omega} \in [0, 1]$ is the case fatality rate when the implemented control measures are fully applied (i.e. $m_I(t) \equiv m_q$ for $t \in [\lambda_{i-q}, \lambda_q)$). Constants $\overline{\omega}$ and $\underline{\omega}$ are estimated with the calibration process mentioned in Section 3.7.

We note that $\overline{\omega} \geq \underline{\omega}$, thus, we consider $\overline{\omega} = \underline{\omega} + \delta_\omega$, with $\delta_\omega \in [0, 1 - \underline{\omega}]$.

3.5. Transition rates

We denote by d_I, d_{I_u}, d_{H_R} and d_{H_D} the (average) duration in days of a person in compartment I, I_u, H_R and H_D, respectively, without the application of control measures. Additionally, we consider that:

- According to [15], the transition rate from E to I depends only on the disease and, therefore, is considered constant.
- According to [15,20], the value of $\gamma_I(t)$ can be increased due to the application of control measures (i.e. people with symptoms are detected earlier). As a consequence, the values of $\gamma_{I_u}(t)$, $\gamma_{H_R}(t)$ and $\gamma_{H_D}(t)$ can be decreased (e.g. people with symptoms stay under observation during more time).
- According to [15], d_{H_R} is lower than d_{H_D}. Thus, we consider $d_{H_D} = d_{H_R} + \delta_R$, where $\delta_R > 0$.
- We assume, for the sake of simplicity, that the infectious period for undetected people is the same than that of the hospitalized people that survive the disease. Thus, $d_{I_u} = d_{H_R}$. Furthermore, we assume that the addition of the times a person is in the compartments I and I_u is constant. Therefore, following [17], we use the following functions:

$$\gamma_I(t) = \frac{1}{d_I - g(t)} \ (\text{day}^{-1}), \quad \gamma_{I_u}(t) = \gamma_{H_R}(t) = \frac{1}{d_{I_u} + g(t)} \ (\text{day}^{-1}), \quad \gamma_{H_D}(t) = \frac{1}{d_{I_u} + g(t) + \delta_R} \ (\text{day}^{-1}),$$

where $g(t) = d_g(1 - m_I(t))$ represents the decrease of the duration of d_I due to the application of control measures, at time t; and d_g is the maximum number of days that d_I can be decreased due to the control measures.

3.6. Disease contact rates

Regarding β_{I_u}, we may consider

$$\beta_{I_u}(\theta) = \begin{cases} \overline{\beta}_I & \text{, if } \theta = \overline{\omega}, \\ \text{nonincreasing} & \text{, if } \theta \in (\overline{\omega}, 1), \\ \underline{\beta}_I & \text{, if } \theta = 1, \end{cases}$$

where $\underline{\beta}_I$ and $\overline{\beta}_I$ are suitable lower and upper bounds, respectively. For the sake of simplicity, here we take $\overline{\beta}_I = \beta_I$.

Furthermore, following the idea proposed in [15,17], we assume that there exists a relationship between the contact rates β_I, β_E, β_{H_R}, β_{H_D} and $\underline{\beta}_I$.

More precisely, we assume that people in compartments E, H_R, H_D and I_u are less infectious than people in compartment I (due to their lower virus load or isolation measures, see [15]). Thus, we consider that $\beta_E = C_E\beta_I$, $\beta_{H_R}(t) = \beta_{H_D}(t) = C_H(t)\beta_I$ and $\underline{\beta}_I = C_u\beta_I$, where C_E, $C_H(t)$ and $C_u \in [0, 1]$. In Section 4.1, we show how to compute the function $C_H(t)$ for the case of China (something similar can be done for other territories).

3.7. Identification of some parameters with a multi-objective technique

First, we determine the set of unknown model parameters to be identified by the approach presented here. We denote this set by Ω. The specific parameters considered to be included in Ω will depend on the available information about the

disease in the considered territory. For the case of China considered in this work, we list the parameters in Ω at the end of Section 4.1.

Once Ω is defined, we follow the methodology proposed in [19]. More precisely, we define a multi-objective optimization problem based on the minimization of the relative errors between the cumulative number of cases and deaths reported by the authorities (i.e. c_{r} and d_{r}) and the cumulative number of cases and deaths estimated by the model when considering a particular set of values for Ω (denoted by c_{m}^{Ω} and d_{m}^{Ω}, respectively). To do so, we consider the following multi-objective optimization problem:

$$\begin{cases} \min \ \left\{ f_c(\Omega), f_d(\Omega) \right\} \\ \text{subject to } \Omega \in S \subset \mathbb{R}^n, \end{cases} \tag{6}$$

where S is the feasible region (i.e. upper and lower bounds of each parameter), n is the number of parameters in Ω, and f_c and f_d are the objective functions defined as:

$$f_c(\Omega) = \frac{\left\| c_{\mathrm{r}} - c_{\mathrm{m}}^{\Omega} \right\|_{L^2([t_0, t_0 + T_{\max}])}}{\left\| c_{\mathrm{r}} \right\|_{L^2([t_0, t_0 + T_{\max}])}} \quad \text{and} \quad f_d(\Omega) = \frac{\left\| d_{\mathrm{r}} - d_{\mathrm{m}}^{\Omega} \right\|_{L^2([t_0, t_0 + T_{\max}])}}{\left\| d_{\mathrm{r}} \right\|_{L^2([t_0, t_0 + T_{\max}])}}, \tag{7}$$

where $\|g\|_{L^2(t_0, t_0 + T_{\max})} = \left(\int_{t_0}^{t_0 + T_{\max}} g(t)^2 \mathrm{d}t \right)^{1/2}$ denotes the L^2 norm of function g in $[t_0, t_0 + T_{\max}]$. We note that the choice of this norm reduces the risk of over-fitting to the reported data [36].

Next, to solve this multi-objective optimization problem, the algorithm called Weighting Achievement Scalarizing Function Genetic Algorithm (WASF-GA) is applied [37–39]. We choose this particular multi-objective algorithm as it has been applied with success in a previous work to identify some parameters of an epidemiological model simulating the evolution of Ebola Virus Disease outbreaks in Democratic Republic of the Congo [38].

From a general point of view, WASF-GA is an evolutionary algorithm that manages a population of possible solutions (called individuals), trying to improve them iteratively towards the optimal set (called Pareto optimal set). To do that, it uses two procedures at each iteration: reproduction and replacement. Firstly, the reproduction method applies some classical genetic operators (here: selection, mutation, and crossover) to the population of the previous iteration to obtain a new offspring population. Secondly, the replacement procedure decides which solutions from the previous iteration and the offspring populations should compose the population for the next iteration. The main feature of WASF-GA is that this decision is based on preferences, which are defined as reference values for the objectives. Indeed, it classifies the individuals using an achievement scalarizing function (see [39]) characterized by the L_{∞} distance between the objective function values and the reference values with different weights. The considered sample of weight vectors is built following [39] to ensure they are well distributed. As a result, WASF-GA returns a set of solutions covering a region of the Pareto optimal front.

In this work, we set the number of iterations to 15,000, the population size to 50 individuals, the number of weight vectors to 50, and the reference values for the objectives to zero, corresponding to the ideal situation of a total matching between the model outputs and the reported data. The remaining input parameters of WASF-GA are those reported in [39].

Since WASF-GA is a stochastic algorithm, it may return different Pareto front approximations for different executions of the same instance. Then, to guarantee that results do not depend on the stochasticity of the algorithm, we perform 60 runs of each experiment. Among all the 60 sets of solutions, we extract the most balanced solution, meaning the one whose objective values are the closest to the reference values (in our case, zero) when considering the Euclidean distance.

4. Numerical simulations and calibration of the model for the case of China

We now propose some numerical experiments to illustrates the efficiency of our approach by considering the particular case of the COVID-19 in China. First, we detail the considered parameters related to this scenario. Then, we present and analyze the obtained results.

4.1. Considered data

We performed several numerical simulations for the case of China. For all the cases below we have used $N = 1.400.812.636$ people and $p(t) \equiv 1$, as it was decided to hospitalize all cases to reduce onward transmission (see [40]).

Following [15,20], we set $d_E = 5.5$ days (i.e. $\gamma_E = 1/5.5$ days^{-1}), $d_I = 6.7$ days, $d_g = 6$ days, $d_{I_{\mathrm{u}}} = 14 - d_I = 7.3$ days (i.e. $\gamma_{I_{\mathrm{u}}} = 1/7.3$ days^{-1}).

For $\beta_{I_{\mathrm{u}}}$, we use the linear relationship

$$\beta_{I_{\mathrm{u}}}(\theta) = \underline{\beta_I} + \frac{\beta_I - \underline{\beta_I}}{1 - \omega(t)}(1 - \theta).$$

Additionally, according to [41], on 20 February 2020, 2055 healthcare workers were infected in China. At this date, 74,651 cases were reported in China. From this data, we assume that the infection due to contact with people in compartment H_R or H_D should represent around $100 \frac{2055}{74651}\% = 2.75\%$ of the number of cases. Thus, we compute $\beta_{H_R}(t) = \beta_{H_D}(t) = C_H(t)\beta_I$ such

that

$$\frac{C_H \beta_I \theta \left(\left(1 - \frac{\omega}{\theta}\right) \frac{1}{\gamma_{H_R}} + \frac{\omega}{\theta} \frac{1}{\gamma_{H_D}} \right)}{C_H \beta_I \theta \left(\left(1 - \frac{\omega}{\theta}\right) \frac{1}{\gamma_{H_R}} + \frac{\omega}{\theta} \frac{1}{\gamma_{H_D}} \right) + \frac{\beta_I}{\gamma_I} + \frac{\beta_E}{\gamma_E} + (1 - \theta) \frac{\beta_{I_u}}{\gamma_{I_u}}} = 0.0275.$$

This implies that,

$$C_H = \frac{0.0275 \left(\frac{\beta_I}{\gamma_I} + \frac{\beta_E}{\gamma_E} + (1 - \theta) \frac{\beta_{I_u}}{\gamma_{I_u}} \right)}{(1 - 0.0275) \beta_I \theta \left(\left(1 - \frac{\omega}{\theta}\right) \frac{1}{\gamma_{H_R}} + \frac{\omega}{\theta} \frac{1}{\gamma_{H_D}} \right)}.$$

In the previous expression, the coefficients depend on territory i and time t.

Following data reported by the WHO, we have aggregated data from the Chinese Mainland, Macao, Hong-Kong and Tai-wan. Additionally, we assume that the movement of people in and out of China is negligible in order to estimate the number of cases and deaths in China. Therefore, for the sake of simplicity, we assume in this section that $\tau_1 = \tau_2 \equiv 0$.

Reported data for China have some inconveniences for modeling purposes. On 13 February 2020, the Hubei (the Chinese province with the largest number of cases and deaths) authorities decided to report as infected not only people who was positive in laboratory tests, but also clinically diagnosed cases in that province (see [22]). This supposed a sudden increase of around 15,000 cases in one day (see [42]).

In addition, up to 16 February 2020, the WHO decided to continue reporting only cases confirmed with laboratory tests in its situation reports (see [43]). On 17 February 2020, the WHO changed its protocol to count cases and decided (see [23]) to include in its reports the data provided by the Hubei province. This supposed a sudden increase of around 20,000 cases in one day (see [22]).

On 21 February 2020 China (and WHO) informed that they have "revised their guidance on case classification for COVID-19, removing the classification of 'clinically diagnosed' previously used for Hubei province, and retaining only 'suspected' and 'confirmed' for all areas, the latter requiring laboratory confirmation. Some previously reported 'clinically diagnosed' cases are thus expected to be discarded over the coming days as laboratory testing is conducted and some are found to be COVID-19-negative" (see [24]).

Thus, different criteria have been used in the official reports, with even different data in Chinese and WHO reports from 13 February 2020 up to 16 February 2020. This hinders the calibration of our mathematical model and makes very difficult an accurate forecast. Therefore we have filtered the data to get more reliable values.

In order to smoothly distribute the sudden increase of 17,410 cases in the number of cases reported before the 17 February 2020, we distributed those 17,410 cases to previous dates according to the following formula:

$$c_{\mathrm{ar}}(t) = c_{\mathrm{r}}(t) + 17410 \frac{\sum_{\tau=\bar{t}}^{t} c(\tau)}{\sum_{\tau=\bar{t}}^{t_a} c_{\mathrm{r}}(\tau)},$$

where $c_{\mathrm{ar}}(i, t)$ denotes the adjusted number of reported cases in country i at time t, $\bar{t} = 12$ January 2020 is the date of the first available data and $t_{\mathrm{a}} = 16$ February 2020. This filtered function is the one used for the identification of parameters mentioned in Section 3.7.

This filtering procedure highlights the fact that reported data could be inconsistent. In this case, since our model repro-duces the biology of the pandemic, if reported data are incorrect (with the exception of small noise) and are not consistent with the biology of the pandemic, the model would not be able to fit the data in a reasonable way. Thus, the use of this model could asses the fact that some data are reported incorrectly.

We show in Fig. 3, the evolution of c_{r} and c_{ar} in China from 12 January 2020 up to 28 March 2020.

According to [15], 86% of all infections were undocumented prior to 23 January 2020, when travel restrictions where applied in Wuhan. Furthermore, after Wuhan's closure measures (later extended to much of China), the dynamics of the epidemic changes. Between 24 January 2020 and 8 February 2020, it is estimated that the percentage of invisible infected people drops to 35% (see [15]). Therefore, we consider here that the numbers of deaths that we have in our database, reported by authorities, are very close to the real ones, but the real total cases are underestimated. Actually, according to the percentages given above, we take

$$\theta(t) = \begin{cases} 1 - 0.86 = 0.14, & \text{if 1 December 2019} \le t \le 24 \text{ January 2020}, \\ \text{linear continuous}, & \text{if 24 January 2020} \le t \le 8 \text{ February 2020}, \\ 1 - 0.35 = 0.65, & \text{if 8 February 2020} \le t. \end{cases} \quad (8)$$
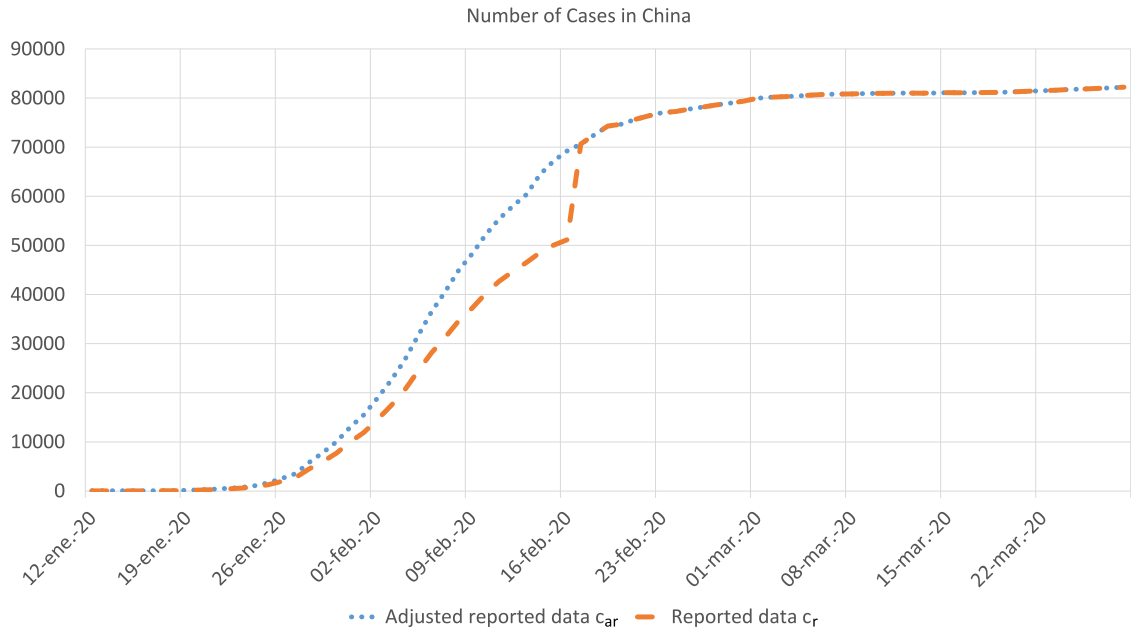
Fig. 3. Evolution of the number of cases in China: Reported (c_r) and adjusted reported data (c_{ar}).

Table 2
Summary of the parameters to be identified with the multiobjective technique presented in
Section 3.7, for the COVID-19 in China. A brief description (**Description**) and the range of the
boundary values (**Range**) considered during the optimization process are reported.

Notation	Range	Description
β_I	[0,0.75]	disease contact rate (day^{-1}) of a person in compartment I.
C_E	[0,0.4]	used to estimate $\beta_E = C_E \beta_I$
C_u	[0.4,0.9]	used to estimate $\underline{\beta}_I = C_u \beta_I$
δ_R	[7,21]	used to estimate $d_{H_D} = d_{H_R} + \delta_R$
δ_ω	[0,0.2]	used to estimate $\overline{\omega} = \underline{\omega} + \delta_\omega$
$\underline{\omega}$	[0.005,0.022]	minimum case fatality rate during the epidemic
κ_1	[0,0.2]	efficiency of the control measures

Thus, even if the reported deaths $d_r(t)$ can be considered a good estimation of the real numbers, neither the reported cases $c_r(t)$ nor the adjusted reported cases $c_{ar}(t)$ described above would be a good estimation of the real number of infected cases $c_{real}(t)$, which can be computed as

$$c_{real}(t) = \frac{c_{ar}(t)}{1 - \theta(t)}.$$

In order to effectively prevent further exportation of infected individuals to the rest of the country, the Chinese government decided to impose a cordon sanitaire (i.e. all public transportation and all outbound trains and flights were halted) around Wuhan and neighboring municipalities on 23 January 2020 ([21]). Therefore, we considered the implementation of control measures after this date, in order to represent the real situation of the measures imposed:

$$m_E(t) = m_{I_u}(t) = m_I(t) = m_{H_R}(t) = m_{H_D}(t) = \begin{cases} 1, & \text{if } t \in [1 \text{ December } 2019, \lambda_1), \\ \exp\left(-\kappa_1(t - \lambda_1)\right), & \text{if } t \in [\lambda_1, 29 \text{ March } 2020], \end{cases}$$

where $\lambda_1 = 23$ January 2020 and the rest of parameters need to be identified.

Using the multiobjective optimization process mentioned in Section 3.7, we determine the following parameters: β_I, C_E, C_u, δ_R, δ_ω, $\underline{\omega}$ and κ_1. All those parameters are bounded according to values found in the literature (see [15]). In Table 2, we recall the meaning of those parameters and their range of values considered during the optimization process for the particular case of COVID-19 in China.

From a numerical point of view, as explained above, we use the classic 4th order Runge–Kutta method with four stages (RK4) and with four hours as time step in order to approximate the solution of system (3). The numerical model and the algorithm WASF-GA are implemented in Java. Experiments are performed in a computer with an I7-4790 3.6 Ghz CPU and

Table 3
Results obtained from the parameter identification step for experiments EXP_{29M}, EXP_{25F}, EXP_{08F} and EXP_{29J}.

Notation	EXP_{29M}	EXP_{25F}	EXP_{8F}	EXP_{29J}
β_I	0.2834	0.3653	0.3718	0.4012
C_E	0.3806	0.0800	0.0828	1.4306×10^{-5}
C_u	0.4001	0.4053	0.4271	0.8559
δ_R	7.0000	9.7528	7.2773	12.6250
δ_ω	0.0195	0.0155	1.3265×10^{-5}	4.6360×10^{-6}
$\underline{\omega}$	0.0161	0.0220	0.0220	0.0220
κ_1	0.1090	0.1007	0.1415	0.2000

32Gb of Ram. One execution of the model takes around 0.01 s and the full parameter identification process takes around 660 s.

4.2. Results

We present and discuss the results returned by our model, considering the parameters that were fixed for the case of China and the parameters to be identified (presented in Section 4.1).

In particular, to show the robustness of the proposed approach, we have run the experiment by considering data $c_{ar}(t)$ and $d_r(t)$, from the index case up to different values of the final date t_f. We set t_f to the following particular values:

- t_f=29 March 2020: this date corresponds to that of the last data available before running the numerical experiments. We denote this experiment by **EXP_{29M}**.
- t_f=25 February 2020: this date is a representative date for which the epidemic in China was decreasing (i.e. the daily number of reported cases was decreasing during several days). We denote this experiment by **EXP_{25F}**.
- t_f=8 February 2020: this date is considered as the inflection point for which the daily number of reported cases starts to decrease, we denote this experiment by **EXP_{08F}**.
- t_f=29 January 2020: this date is a representative date for which the epidemic in China was growing (i.e. the daily number of reported cases was increasing during several days), we denote this experiment by **EXP_{29J}**.

The objective is to show how the model and the parameter estimation methodology deal with incomplete data. This study also illustrates the robustness of our approach to the over-fitting of the model parameters with respect to the reported data.

In Table 3, we show the values of the optimized parameters obtained at the end of each experiment. From those results, on the one hand we observe that the values of β_I, C_E, C_u, $\underline{\omega}$, δ_ω and κ_1 show a progressive evolution to the value obtained with **EXP_{29M}**, when increasing the amount of available data for fitting the model to observed data. There is a noticeable difference between the values returned at the end of **EXP_{29M}** and **EXP_{29J}**. This indicates that a reasonable amount of data should be available to estimate well some of the model parameters. Estimating those parameters at an early stage of the epidemic could produce poorly consistent results.

Additionally, for parameters C_u, $\underline{\omega}$, δ_ω and κ_1, we observe that the results returned by **EXP** $_{29M}$ and **EXP** $_{25F}$ are quite similar.

Additionally, focusing on δ_R, we observe that the value generated with **EXP_{29M}** and **EXP_{08F}** clearly differs from the one obtained with **EXP_{25F}** and **EXP_{29J}**. We note that this particular parameter was already reported in the literature (see, e.g. [15]) as being difficult to be estimated (considering a range of one month). Again, this shows that the estimation of some key parameter could change dramatically during an epidemic and re-calibration should be performed periodically to capture the recent behavior of the epidemic.

In Fig. 4, we show the evolution of the number of cases and deaths in China from 1 December 2019 to 29 March 2020 for the reported data (i.e. $c_{ar}(t)$ and $d(t)$) and for the results obtained at the end of experiments **EXP_{29M}, EXP_{25F}, EXP_{25F}** and **EXP_{29J}**.

Focusing on the number of cases, we observe that results obtained with **EXP_{29M}, EXP_{25F}** reproduce quite accurately the evolution of the number of cases. In particular, we found a final number of cases around 80,980 and 83,800, considering **EXP_{29M}** and **EXP_{25F}**, respectively (corresponding to an error of less than 1.8% with respect to the reported values). The final number of cases obtained with **EXP_{08F}** is around 65,950, which correspond to an underestimation of -20% with respect to real reported data. Finally, we observe that **EXP_{29J}** forecasts poorly and clearly underestimates the evolution of the epidemic up to the 29 March 2020, with only 52,740 cases (−36% with respect to the reported value). However, all forecasts capture reasonably the deceleration of the daily number of cases observed at the end of February. This indicates that forecasting at an early stage of the epidemic (before the inflection point in the curve of the reported cases) is quite a difficult task.

Focusing on the number of deaths, it seems that all results do not capture efficiently the drastic decrease occurring at the end of February. However, the result returned by **EXP_{29M}** shows a slope similar to that of the reported deaths. In all cases, the obtained estimations are in a reasonable range compared to the reported final value (with an error of $\pm$ 24%).
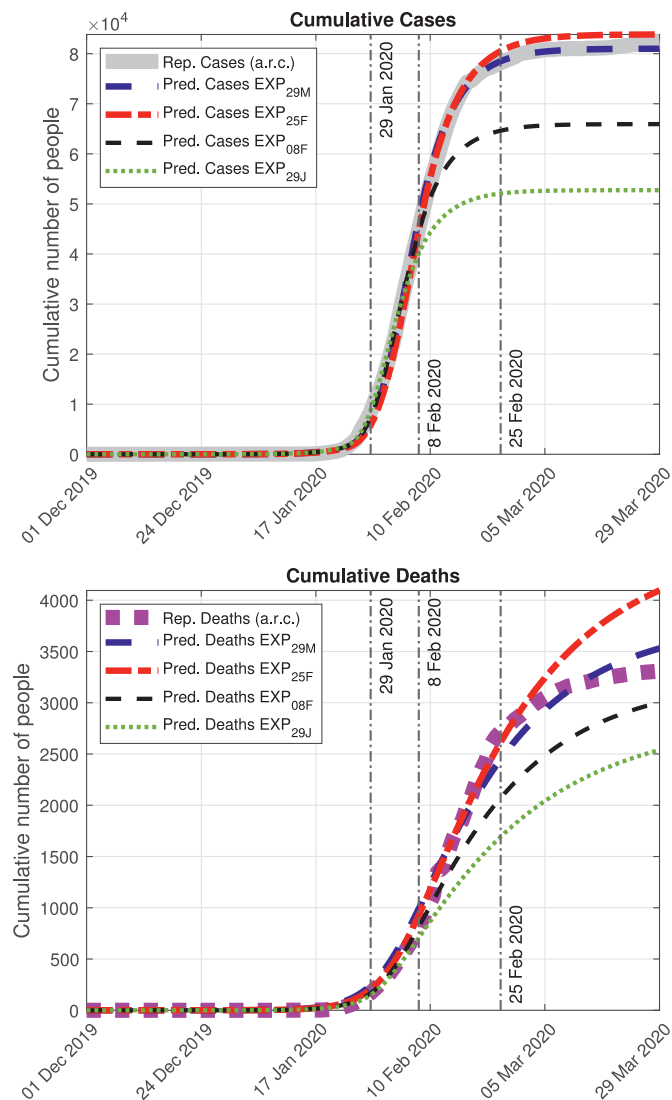
Fig. 4. Evolution of the number of **(TOP)** cases and **(BOTTOM)** deaths in China from 1 December 2019 to 29 March 2020: Adjusted reported cases (a.r.c) and deaths, and estimation obtained with **EXP$_{29M}$**, **EXP$_{25F}$**, **EXP$_{08F}$** and **EXP$_{29J}$**.

We now focus on the results returned at the end of Experiment **EXP$_{29M}$** to analyze the main key factor that may explain the dynamics of the COVID-19 epidemic in China up to 29 March 2020.

In Fig. 5, we show the evolution of the adjusted reported cases and the detected, undetected and total (i.e. detected plus undetected) cases estimated by our model. We observe that the final number of undetected cases (around 87,100 cases) estimated by our model represents around 52% of the total number of cases (around 168,100 cases). Furthermore, it seems that control measures have not contained their growing as fast as for the detected cases. These results are interesting as they indicate that, despite the relative control of the epidemic in China, there may still exist an undetected source of infected people that could cause an increase of the epidemic in a near future, if the control measures are significantly relaxed.

In Fig. 6, we show the evolution of Hospitalized and Recovered people estimated by our model and reported data. The model estimates that the peak of the number of people hospitalized at the same time has been reached around 17 February 2020, with 64,704 hospitalized patients. According to data reported in

https://www.worldometers.info/coronavirus/country/china/

the peak was reached on 17 February 2020 with around 58,000 hospitalized people at the same time. The model estimates quite reasonably the number of hospitalized people. However, the obtained results slightly overestimate the real observations by around 11.6%. Focusing on Recovered people, our model returns a final number of 75,100 people, which is very close to the real observation (around 75,130 recovered people on 29 March 2020). This shows that the proposed methodology is a reasonable decision tool to estimate the number of beds in hospitals during an epidemic or pandemic.
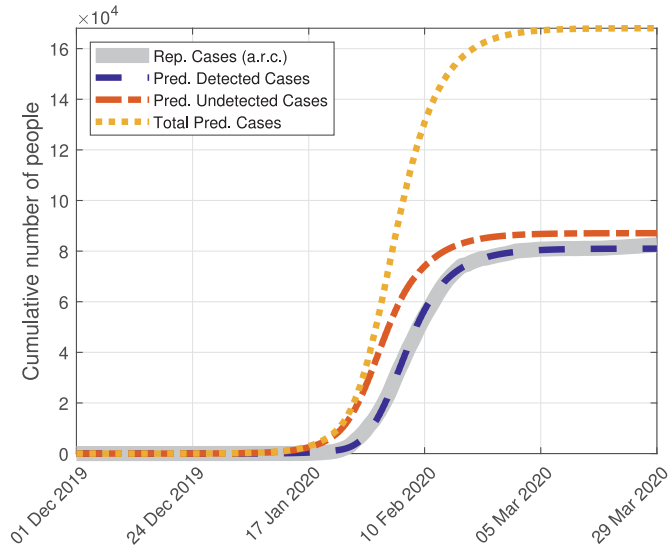
Fig. 5. Evolution of the number of cases in China from 1 December 2019 to 29 March 2020: Adjusted reported cases (a.r.c.), as well as detected, undetected and total cases obtained with **EXP$_{29M}$**.

Additionally, in Fig. 7, we show the evolution of Hospitalized people in China considering the adjusted reported cases (i.e. $c_{ar}(t)$ minus the number of reported deaths and recovered people at time t) and the results returned at the end of experiments **EXP$_{29M}$**, **EXP$_{25F}$**, **EXP$_{08F}$** and **EXP$_{29J}$**. We observe that in all cases, the estimated date for the peak of people in Hospital is close to the real observed date (17 February 2020). This seems to indicate that our approach is able to detect early a reasonable approximated date for this peak. Focusing on the magnitude of this peak, **EXP$_{29M}$**, **EXP$_{25F}$** and **EXP$_{08F}$** give acceptable estimations (with error of $\pm$ 13.8%). However, as expected, the magnitude of the peak returned with **EXP$_{29J}$** underestimates the reported data (with an error of -24%).

In Fig. 8, we present the evolution of the daily number of new people in the exposed compartment, new cases and new deaths in China from 1 December 2019 to 29 March 2020 considering the results returned at the end of experiments **EXP$_{29M}$**. We also show the adjusted daily number of new reported cases (a.r.c.) and the daily number of new reported deaths. These outputs are of interest to estimate the behavior of the epidemic curve (see [44]), and they can be used to study the efficiency of control measures to flatten this curve.

Here, we observe that considering the daily number of new cases and new deaths, the predicted values fit reasonably well the reported ones (taking into account that there were several errors and approximations in the reported number of deaths, see, e.g. [45]). Focusing on the number of new daily people in the exposed compartment E, we observe that the peak was reached on 27 January 2020, suggesting that the isolation measures taken in China on 23 January 2020 were efficient to quickly reduce the magnitude of the epidemic. The peak for the new daily cases and deaths (which is the visible part for the society) is delayed due to the transition rates between each compartment of the disease. Therefore, despite the effects of the control measures are immediate, this effect is visible for the society after a non negligible delay (in this case around two weeks).

In Fig. 9, we show the evolution of the effective reproduction number $R_e(t)$. We observe that this value decreases since the application of the control measures and it remains below 1 after 1 February 2020. These results are consistent with the fact that a decrease in the daily number of reported cases was observed at the beginning of February 2020.

In Table 4, we present some important parameters and threshold values related to this epidemic. In particular, we observe that in our case the basic reproduction ratio R_0 is 4.2732, which is bigger than other values reported in the literature (between 2.5 and 3.0, see [15]). This could be explained by the fact that we have taken into account the undetected people.

Additionally, we observe that people in compartment E represent around 26% of the infections. People in compartment I_u caused around 37% of infections. Those results suggest that people with no or few symptoms were the major source of infection. However, it seems that our model underestimates the infection due to hospitalized people with only 1% of the infections (it should be around 3%). A more precise approach should be proposed to estimate better the value of C_H.

Finally, we analyze the impact of the detection parameter $\theta(t)$ on the global magnitude of the epidemic. With that aim, considering the model parameters obtained at the end of **EXP$_{29M}$**, we replace $\theta(t)$ by a constant function with values 0.14, 0.25, 0.5, 0.75 and 1 and run the model up to 29 March 2020. In Fig. 10 we show the evolution of the total number of cases (detected plus undetected) returned by each scenario. We observe that increasing the detection of affected people has a major impact on the global magnitude of the epidemic. The model estimates a reduction on the total amount of cases from 206,500, when detecting only 14% of cases, to less than 2,000 total cases, when detecting 100% of cases (which is an
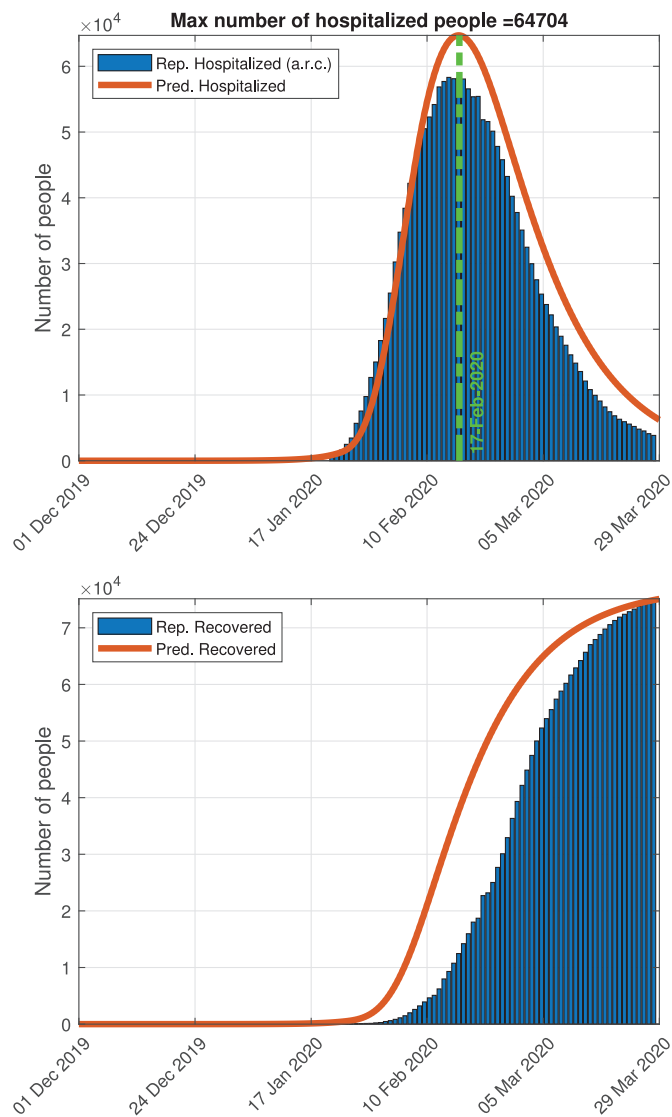
Fig. 6. Evolution of the number of Hospitalized **(Top)** and Recovered **(Bottom)** people in China from 1 December 2019 to 29 March 2020 considering reported data and estimation obtained with **EXP$_{29M}$**.

Table 4
Results obtained *with experiment* **EXP$_{29M}$**.

Notation	Value
R_0	4.2732
β_E	0.1079
β_I	0.2834
β_{I_u}	[0.1739, 0.2650]
β_{H_R}	[0.0052, 0.0924]
β_{H_D}	[0.0052, 0.0924]
$\omega(t)$	[0.0161, 0.0356]
$\Gamma_E(t_f)$	43,560
$\Gamma_I(t_f)$	61,309
$\Gamma_{I_u}(t_f)$	61,575
$\Gamma_H(t_f)$	1,640
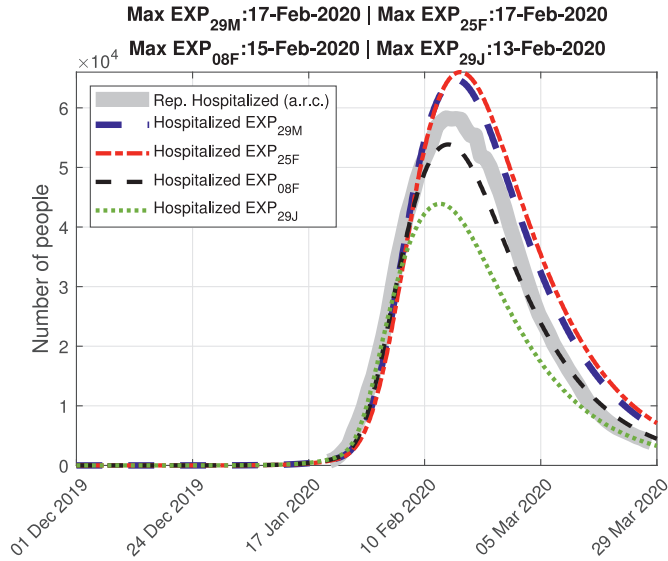
Fig. 7. Evolution of the number of Hospitalized people in China from 1 December 2019 to 29 March 2020, considering the adjusted reported cases (a.r.c.) and the results returned at the end of experiments EXP_{29M}, EXP_{25F}, EXP_{08F} and EXP_{29J}.
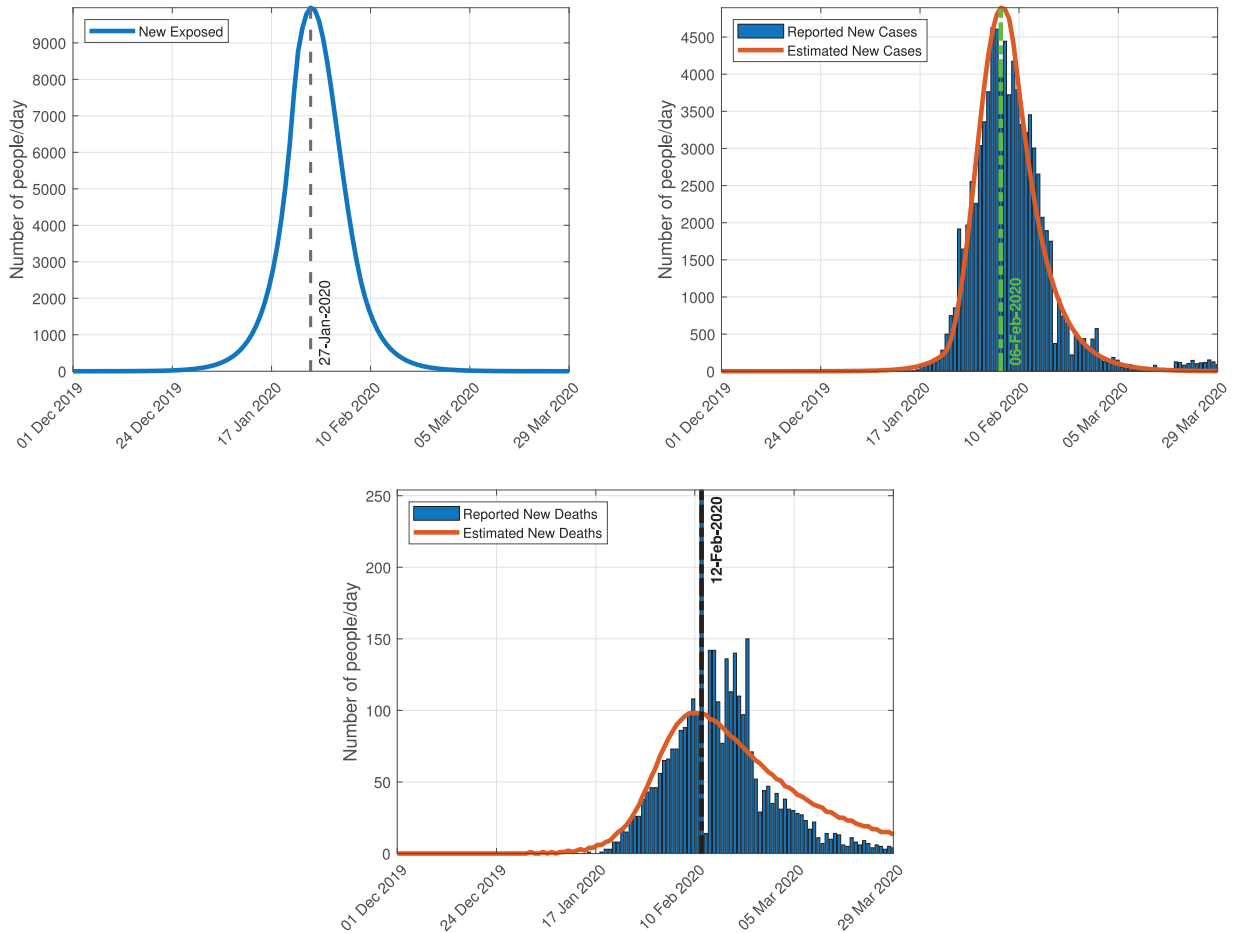
Fig. 8. Evolution of the daily number of new people in the exposed compartment (**Top-Left**), new cases (**Top-Right**) and new deaths (**Bottom**) in China from 1 December 2019 to 29 March 2020 considering the results returned at the end of experiment EXP_{29M}. We also report the adjusted daily number of new reported cases and the daily number of new reported deaths.
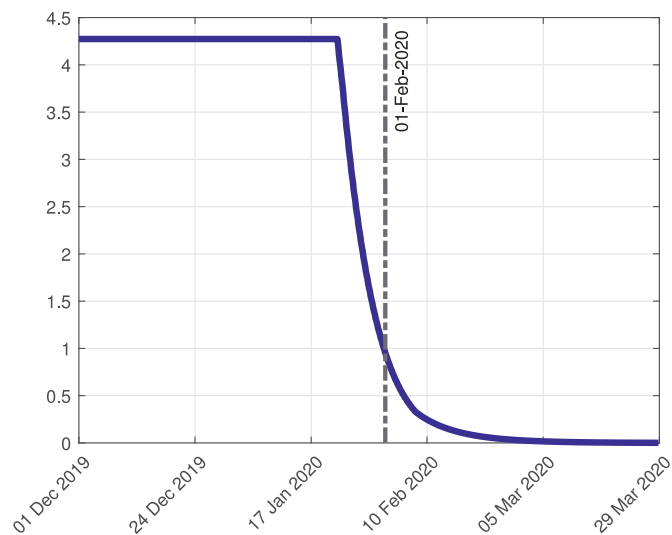
Fig. 9. Evolution of the effective reproduction number $R_e(t)$ in China from 1 December 2019 to 29 March 2020 obtained with **EXP$_{29M}$**.
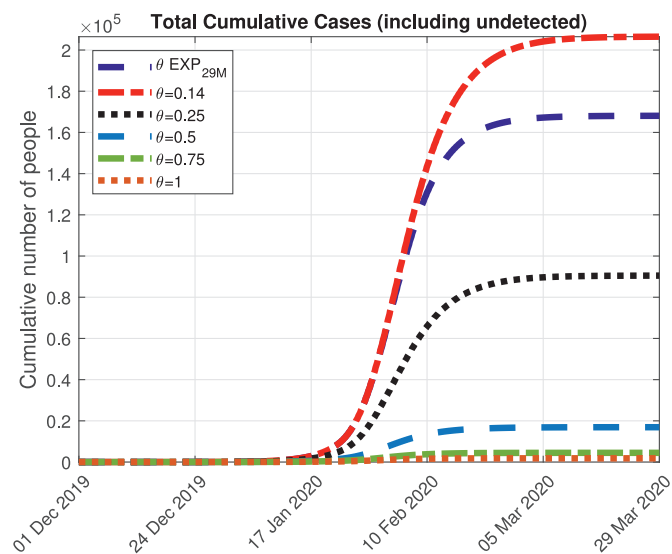

Fig. 10. Evolution of the total number of cases (detected plus undetected) in China from 1 December 2019 to 29 March 2020, considering parameters from **EXP$_{29M}$** and the function θ given in (8), and also scenarios in which $\theta(t)$ is given the constant values of 0.14, 0.25, 0.5, 0.75 and 1.

ideal case). Those results suggest that a detection campaign should be applied to reduce the impact of current COVID-19 epidemics.

5. Conclusions and discussion

In this work, we developed a new θ-SEIHRD model adapted to the characteristics of COVID-19. This model considers some important characteristics of the COVID-19, such as the undetected infectious people and the effect of the control measures. In particular, it includes a novel approach that considers the fraction θ of detected cases over the real total infected cases, which allows to study the importance of this ratio on the impact of COVID-19. We also introduced a method to estimate some of the unknown parameters of this model. Additionally, we proposed the use of a filtered version of the data reported by the WHO, in order to smoothly distribute along the previous dates the sudden increase of 17,414 cases reported on 17 February 2020.

From a general point of view, the outputs returned by the simulation fit quite well the data reported by the WHO. In particular, they estimated reasonably the date and magnitude of the peaks corresponding to the number of new cases, new deaths and amount of hospitalized people. This indicates that the proposed methodology can be used as an useful decision tool for policy makers. However, considering an estimation performed at an early stage of the epidemic could produce poor results.

Focusing on the study of the peaks of some relevant curves of the epidemic (e.g. new exposed people, new cases, new deaths and hospitalized people), results show that the control measures in China were efficient to quickly reduce the magnitude of the epidemic. However, due to the transition rates between the disease compartments, the visible impact for the society has a delay of two weeks after the beginning of those control measures.

From the model and simulations considered in this work, we estimate that the value of the basic reproduction number R_0, for COVID-19 in China, is 4.2732. We note that this value is higher than the one usually reported in the literature (between 2.5 and 3 see, e.g. [15]). This could be due to the fact that our model also estimates the amount of undetected cases, which increases the initial growth of the epidemic. Additionally, the effective reproduction number R_e decreased, mainly due to the application of control measures, and reached values lower than 1 after 1 February 2020.

Considering the total cases, the model estimates that, including undetected cases, around $168,100$ people could have been infected in China. Those undetected cases could represent around 52% of the total number of cases. Furthermore, they may have caused around 37% of the total infections. These results are consistent with other conclusions found in the literature (see [46]). Other studies found that asymptomatic or mild cases represent about $40\% - 50\%$ of all infections (see [47]) whereas in [48] it is estimated that the percentage of symptomatic COVID-19 cases reported is 33%.

Finally, we also proposed a study of the impact of the percentage of detection of cases and obtained that the magnitude of the epidemic can be drastically reduced when increasing this percentage. This results could be used as a recommendation to countries currently actively affected by COVID-19.

However, we also note some limitations in the proposed methodology.

We assume that the population inside the considered countries or territories is homogeneously distributed (this could be improved by dividing some of them into a set of smaller territories with similar characteristics). Thus, the spatial distribution of the epidemic inside a territory is not taken into account. Additionally, the current model is only suitable for countries or territories with a relevant number of people infected by COVID-19, where local transmission is the major cause of the disease spread. The between-country spread has not been modeled in this work. Moreover, since there is no clear scientific evidence of the effect of the humidity and the temperature on SARS-CoV-2, we have not included these two factors in our model (this would need to be revised in case of new findings regarding this topic).

In this work we also point out the fact that we worked with official reported data, whose quality has been affected by: i) changes in guidelines to count cases in China (as mentioned above); ii) uncertainty about the number of undocumented infections; and iii) uncertainty about some of the characteristics of SARS-CoV-2.

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.

CRediT authorship contribution statement

B. Ivorra: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **M.R. Ferrández:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **M. Vela-Pérez:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **A.M. Ramos:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing.

Acknowledgments

References

[1] Anderson M. Population biology of infectious diseases: part 1. Nature 1979;280:361–7.

[2] Ivorra B, Martínez-López B, Sánchez-Vizcaíno JM, Ramos AM. Mathematical formulation and validation of the Be-FAST model for classical swine fever virus spread between and within farms. Ann Oper Res 2014;219(1):25–47.

[3] Martínez-López B, Ivorra B, Ramos AM, Sánchez-Vizcaíno JM. A novel spatial and stochastic model to evaluate the within- and between-farm transmission of classical swine fever virus. i. general concepts and description of the model. Vet Microbiol 2011;147(3–4):300–9.

[4] Thieme HR. Mathematics in population biology. Mathematical Biology Series. Princeton University Press; 2003.

[5] Brauer F, Castillo-Chávez C. Mathematical Models in Population Biology and Epidemiology. Texts in Applied Mathematics, Springer; 2001. ISBN 978-1-4757-3516-1. https://www.springer.com/gp/book/9781441931825.

[6] Yan D, Cao H. The global dynamics for an age-structured tuberculosis transmission model with the exponential progression rate. Applied Mathematical Modelling 2019;75:769–86. doi:10.1016/j.apm.2019.07.003. http://www.sciencedirect.com/science/article/pii/S0307904X19304044.

[7] Wang C, Horby PW, Hayden F, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020;395:470–3. doi:10.1016/S0140-6736(20)30185-9.

[8] Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. Severe acute respiratory syndrome-related coronavirus: the species and its viruses - a statement of the coronavirus study group. bioRxiv 2020:1–20. doi:10.1101/2020.02.07.937862.

[9] Organization WH. Naming the coronavirus disease (COVID-19) and the virus that causes it. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it; 2020a.

[10] Wang L, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (covid-19) implicate special control measures. J Med Virol 2020. doi:10.1002/jmv.25748.

[11] Organization W.H.. Statement on the second meeting of the international health regulations (2005). emergency committee regarding the outbreak of novel coronavirus (2019-ncov). https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov); 2020b.

[12] Organization W.H.. Director-general's opening remarks at the media briefing on COVID-19 - 11 march 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020; 2020c.

[13] Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, et al. Real-time forecasts of the COVID-19 epidemic in china from february 5th to february 24th, 2020. Infect Dis Modell 2020;5:256–63. doi:10.1016/j.idm.2020.02.002.

[14] Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020. doi:10.1016/S1473-3099(20)30144-4.

[15] Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-cov2). Science 2020. doi:10.1126/science.abb3221.

[16] Russell T.W., Hellewell J., Abbott S., Golding N., Gibbs H., Jarvis C.I., et al. Using a delay-adjusted case fatality ratio to estimate under-reporting. Report from the Centre for Mathematical Modelling of Infectious Diseases. https://cmmid.github.io/topics/covid19/severity/global_cfr_estimates.html; 2020a.

[17] Ivorra B, Ramos AM, Ngom D. Be-CoDiS: A mathematical model to predict the risk of human diseases spread between countries. validation and application to the 2014 ebola virus disease epidemic. Bull Math Biol 2015;77(9):1668–704. doi:10.1007/s11538-015-0100-x.

[18] Ferrández MR, Ivorra B, Ortigosa PM, Ramos AM, Redondo JL. Application of the Be-CoDis model to the 2018-19 ebola virus disease outbreak in the democratic republic of congo. ResearchGate Preprint 2019;23 July 2019:1–17. doi:10.13140/RG.2.2.13267.63521/2.

[19] Ferrández MR, Ivorra B, Redondo JL, Ramos AM, Ortigosa PM. A multi-objective approach to estimate parameters of compartmental epidemiological models. application to ebola virus disease epidemics.. researchgatenet 2020:1–49. doi:10.13140/RG.2.2.25778.56006.

[20] Liu T, Hu J, Kang M, Lin L, Zhong H, Xiao J, et al. Transmission dynamics of 2019 novel coronavirus (2019-ncov). bioRxiv 2020. doi:10.1101/2020.01.25.919787.

[21] Organization W.H.. Coronavirus disease (COVID-2019) Situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/; 2020d.

[22] Organization W.H.. Coronavirus disease 2019 (COVID-19) situation report – 24. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200213-sitrep-24-covid-19.pdf?sfvrsn=9a7406a4_4; 2020e.

[23] Organization W.H.. Coronavirus disease 2019 (COVID-19) situation report – 28. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200217-sitrep-28-covid-19.pdf?sfvrsn=a19cf2ad_2; 2020f.

[24] Organization W.H.. Coronavirus disease 2019 (COVID-19) situation report – 31. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200220-sitrep-31-covid-19.pdf?sfvrsn=dfd11d24_2; 2020g.

[25] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 2020;27(2). doi:10.1093/jtm/taaa021.

[26] Chen S, Yang J, Yang W, Barnighausen T. COVID-19 control in China during mass population movements at New Year. Lancet 2020;395(10226):764–6. doi:10.1016/S0140-6736(20)30421-9.

[27] Organization W.H.. Report of the WHO-China Joint Mission on Coronavirus Disease 2019. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf; 2020h.

[28] Organization W.H.. Rolling updates in coronavirus disease COVID-19. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen; 2020i.

[29] Diekmann O, Heesterbeek H, Britton T. Understanding Infectious Disease Dynamics. Princeton Series in Theoretical and Computational Biology, Princeton University Press; 2013. ISBN 978-1-4008-4562-0. http://www.jstor.org/stable/j.cttq9530.

[30] Van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 2002;180(1–2):29–48. doi:10.1101/2020.02.22.20025791.

[31] Bank W.. World Bank Open Data. https://data.worldbank.org/; 2020.

[32] Luo W, Majumder M, D L. The role of absolute humidity on transmission rates of the COVID-19 outbreak.. MedRxiv 2020. doi:10.1101/2020.02.12.20022467.

[33] Wang M, Aili Jiang MD, Gong L. Temperature significantly change COVID-19 transmission in 429 cities. MedRxiv 2020. doi:10.1101/2020.02.22.20025791.

[34] for Disease Prevention E.C., Control. Discharge criteria for confirmed COVID-19 cases – when is it safe to discharge COVID-19 cases from the hospital or end home isolation?https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-Discharge-criteria.pdf; 2020.

[35] Lekone P, Finkenstädt B. Statistical inference in a stochastic epidemic seir model with control intervention: ebola as a case study.. Biometrics 2006;62(4):1170–7. doi:10.1111/j.1541-0420.2006.00609.x.

[36] Rhys HI. Machine Learning with R, the tidyverse, and mlr. Manning Publications; 2020. ISBN 9781617296574. https://books.google.es/books?id=jRzYDwAAQBAJ.

[37] Ferrández MR, Puertas-Martín S, Redondo JL, Ivorra B, Ramos AM, Ortigosa PM. High-performance computing for the optimization of high-pressure thermal treatments in food industry. J Supercomput 2018;75:1187–202. doi:10.1007/s11227-018-2351-4.

[38] Ferrández MR, Redondo JL, Ivorra B, Ramos AM, Ortigosa PM. Preference-based multi-objectivization applied to decision support for high-pressure thermal processes in food treatment. Appl Soft Comput 2019;79:326–40. doi:10.1016/j.asoc.2019.03.050.

[39] Ruiz AB, Saborido R, Luque M. A preference-based evolutionary algorithm for multiobjective optimization: the weighting achievement scalarizing function genetic algorithm. J Global Optim 2015;62(1):101–29.

[40] Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of COVID-19 disease. MedRxiv 2020. doi:10.1101/2020.03.09.20033357.

[41] Organization W.H.. 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; 2020j.

[42] (JHU) J.H.U.. Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE). https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6; 2020.

[43] news B.. Tracking coronavirus: Map, data and timeline. https://bnonews.com/index.php/2020/01/the-latest-coronavirus-cases/; 2020.

[44] for Disease Control C., Prevention. Using an epi curve to determine mode of spread. https://www.cdc.gov/training/QuickLearns/epimode/; 2020.

[45] Organization W.H.. Coronavirus disease 2019 (COVID-19) situation report – 25. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200214-sitrep-25-covid-19.pdf?sfvrsn=61dda7d_2; 2020k.

[46] Qiu J. Covert coronavirus infections could be seeding new outbreaks. Nature 2020. doi:10.1038/d41586-020-00822-x.

[47] Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond princess cruise ship, yokohama, japan 2020. Euro Surveill 2020;25(10). doi:10.2807/1560-7917.ES.2020.25.10.2000180.

[48] Russell TW, Hellewell J, Abbott S, Jarvis CI, van Zandvoort K, et al. Using a delay-adjusted case fatality ratio to estimate under-reporting. Centre for Mathematical Modeling ofInfectious Diseases Repository 2020. https://cmmid.github.io/topics/covid19/severity/global_cfr_estimates.html.