# Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) considering its particular characteristics. The case of China

**Preprint** · March 2020

DOI: 10.13140/RG.2.2.21543.29604

CITATIONS

0

**4 authors**, including:

Benjamin Ivorra
Complutense University of Madrid
**121** PUBLICATIONS **737** CITATIONS

SEE PROFILE

Miriam Ruiz Ferrández
Universidad de Almería
**16** PUBLICATIONS **24** CITATIONS

SEE PROFILE

María Vela
Complutense University of Madrid
**17** PUBLICATIONS **100** CITATIONS

SEE PROFILE

# Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) considering its particular characteristics. The case of China

23 March 2020

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

MOMAT Research Group (`https://www.ucm.es/momat`)
Interdisciplinary Mathematics Institute
Complutense University of Madrid
emails: [a]ivorra@ucm.es; [b]mrferrandez@ual.es; [c]maria.vela@ucm.es; [d]angel@mat.ucm.es

**Abstract**

In this paper we develop a mathematical model for the spread of the coronavirus disease 2019 (COVID-19). We use a compartmental model (but not a SIR, SEIR or other general purpose model) and take into account the known special characteristics of this disease, as the existence of infectious undetected cases. 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 about 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. The model is also able to estimate the needs of beds in hospitals for intensive care units. Finally, we also study the behavior of the outputs returned by our model when considering incomplete data (by truncating them at some dates before and after the peak of daily reported cases). By comparing those results with real observation we can estimate the error produced by the model when identifying the parameters at early stages of the epidemic.

**Keywords:** Mathematical model, COVID-19, coronavirus, SARS-CoV-2, Pandemic, simulation

**Background:** Coronavirus disease 2019 (COVID-19) is an infectious disease emerging in China in December 2019 that rapidly spread around that country and subsequently many others. On 9 February 2020 our group developed a first tentative mathematical model of the outbreak (see [12]), based on the Be-CoDiS model (see [11]) and the background of our group in the mathematical modeling of epidemics (see `https://www.ucm.es/momat/epidemics`). We calibrated the model with the available data reported by authorities by 8 February 2020 and reported a forecast. It fitted the data well for 3 weeks (see [13]), even with a sudden unexpected increase of official reported cases on 17 February 2020 due to a change in the World Health Organization guidelines to count cases (see Figure 3). On 21 February 2020 COVID-19 started spreading locally around Italy (see [25]). This country became the epicenter of COVID-19, which spread worldwide. WHO declared COVID-19 to be a pandemic on 11 March 2020 [24]. To the best of our knowledge this fact was not forecasted for any scientific paper or report in early February 2020. The main reason is the above-mentioned bad quality of the official reported data, due in part to the fact that COVID-19 is a disease caused by a new virus, called SARS-CoV-2, which has generated a completely exceptional new worldwide emergency situation.

**Objective:** The main goal of this paper is to develop a mathematical model well adapted to COVID-19 (not just using a SIR, SEIR or other general purpose epidemic models), able to estimate, considering different scenarios, the

number of cases, deaths and needs of beds in hospitals for intensive care, in territories where COVID-19 is (or may be) a very serious health problem. We study here the case of China (including the Chinese Mainland, Macao, Hong-Kong and Taiwan, as done by the World Health Organization in its reports about COVID-19; see [23]), 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. Finally, we also study the behavior of the outputs returned by our model when considering incomplete data (by truncating them at some dates before and after the peak of daily reported cases). By comparing those results with real observation we can estimate the error produced by the model when identifying the parameters at early stages of the epidemic.

**Limitations:**

1. Bad quality of official reported data, mainly due to: a) changes in guidelines to count cases in China (as mentioned above); b) Possible relevant number of undocumented infections; c) uncertainty about some of the characteristics of SARS-CoV-2.

2. 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).

3. 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 omitted.

4. The current model is only suitable for countries or territories with a relevant number of people infected by COVID-19, where local spread is very important. The between-country spread has not been modeled in this work. We will do that in another work, based on the BeCoDis model (see [11] and [12]), but the movement of people between countries or territories is not well documented.

5. This study has been done under the pressure of the urgency of the current situation due to COVID-19. Part of the material developed here can be probably improved.

**Summary of the results:** From the model and simulations considered in this work, we found the following novelties and results:

- The value of the basic reproduction number $R_0$, for COVID-19 in China, is 3.3701. 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.

- The model estimates that, including undetected cases, around 195700 persons could have been infected in China.

- Undetected cases could represent around 60% of the total number of cases.

- The undetected cases (i.e. asymptomatic cases), may have caused around 23% of the total infections.

- Model fits quite well the date and amount of persons of the peak of hospitalized people.

- We propose the use of a filtered version of the data reported by the WHO, in order to smoothly distribute during the previous dates, the sudden increase of 17414 cases reported on 17 February 2020.

- We developed a compartmental model well adapted to the characteristics of COVID-19, taking into account undetected cases and a method to estimate unknown parameters. The corresponding simulations returned outputs fitting quite well the data reported by the WHO.

- Focusing on the parameter estimation procedure, results show that estimating the epidemic at early stages (before the peak) could generate poorly estimated results.

# 1 Introduction

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 [34]). 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 [10, 26]).

This is a new virus and a completely new situation [36]. On 30 January 2020, WHO declared it to be a Public Health Emergency of International Concern [30]. 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 [24]. To date, there are 332,930 and 14510 official reported cases and deaths, respectively, and there is no vaccine specifically designed for this virus, with proven effectiveness.

The model developed here is based on the Be-CoDiS model (see [11]), designed to be able to study the spread of human diseases worldwide. It was initially used for the 2014-2016 ebola outbreak (see [11]) and has also been used for the 2018-2020 ebola outbreak in the Democratic Republic of the Congo (see [5, 6]), 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`.

Here, we have adapted the 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 spread is very important. The between-country spread will be included in the model in another future work, based on the BeCoDis model (see [11] and [12]).

We apply here the model to the case of China (including the Chinese Mainland, Macao, Hong-Kong and Taiwan, as done by the World Health Organization in its reports about COVID-19; see [23]), 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.

# 2 Mathematical formulation of the model

## 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:

- 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 one of the Infectious states.
- Infectious that will be detected (denoted by $I_{\mathrm{d}}$): The person can infect other people, starts developing clinical signs and will be detected and reported by authorities (when arriving to the compartments $H_R$ or $H_D$). After this period, people in this compartment are taken in charge by sanitary authorities and we classify them as Hospitalized.
- Infectious that will not be detected (denoted by $I_{\mathrm{u}}$): The person can infect other people and may start developing clinical signs but will not be detected and reported by authorities. After this period, people in this compartment pass to the Recovered state.
- 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 state.

- 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 state.
- Dead (denoted by $D$): The person has not survived the disease.
- Recovered (denoted by $R$): The person has survived the disease, is no longer infectious and has developed a natural immunity to the disease pathogen.

The authorities may apply various control measures in order to control the COVID-19 spread (see [3]):

- Isolation: Infected people are isolated from contact with other people. Only sanitary professionals are in contact with them. However, contamination of those professionals also occur (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 persons 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 of sanitary resources: The number of operational beds and sanitary personal available to detect and treat affected persons is increased, producing a decrease in the infectious period.
- Increase the number of tests in order to increase the percentage of detected infected people.

## 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 [29]). Furthermore, according to [36], 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 [23]).

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 states $S$, $E$, $I_{\mathrm{d}}$, $I_{\mathrm{u}}$, $H_R$, $H_D$, $R$ 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 $t$ all people in all the considered territories are in the susceptible state, 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 model is presented in Figure 1.

The choice of using a deterministic model instead of a stochastic one is done as a first approach, since such kind of models presents 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 ([4]), 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, as soon as possible.

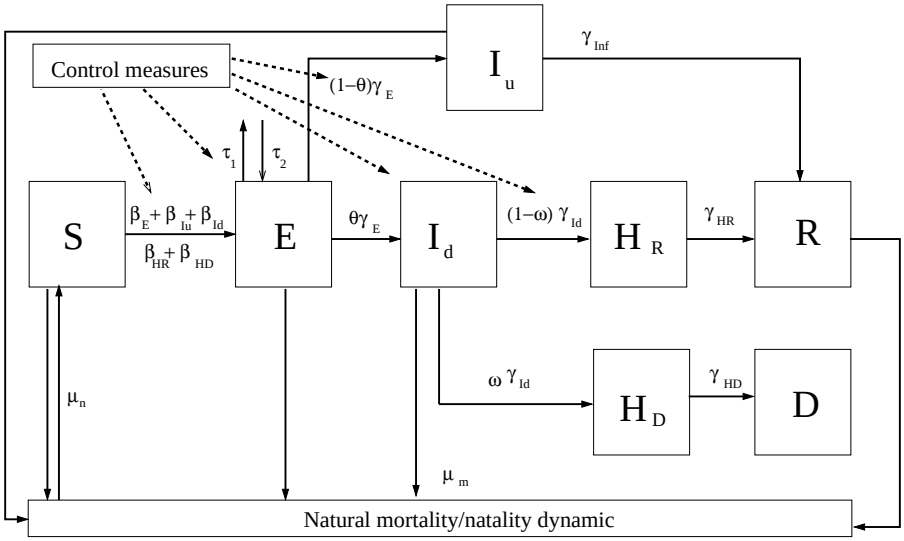


Figure 1: Diagram summarizing the model for COVID-19 given by system (1).

## 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, [2]).

We assume that people in a territory are characterized to be in one of those states, described in Section 2.1: $S$, $E$, $I_{\rm d}$, $I_{\rm u}$, $H_R$, $H_D$, $R$ or Dead $D$. 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 persons. 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:

$$
\begin{aligned}
\frac{\mathrm{d}S^{(i)}}{\mathrm{d}t}(t) &= -\frac{S^{(i)}(t)}{N^{(i)}}\left(m_E^{(i)}(t)\beta_E^{(i)}E^{(i)}(t) + m_{I_\mathrm{u}}^{(i)}(t)\beta_{I_\mathrm{u}}^{(i)}I_\mathrm{u}^{(i)}(t) + m_{I_\mathrm{d}}^{(i)}(t)\beta_{I_\mathrm{d}}^{(i)}I_\mathrm{d}^{(i)}(t)\right) \\
&\quad -\frac{S^{(i)}(t)}{N^{(i)}}\left(m_{H_R}^{(i)}(t)\beta_{H_R}^{(i)}H_R^{(i)}(t) + m_{H_D}^{(i)}(t)\beta_{H_D}^{(i)}H_D^{(i)}(t)\right) \\
&\quad -\mu_m^{(i)}S^{(i)}(t) + \mu_n^{(i)}\left(S^{(i)}(t) + E^{(i)}(t) + I_\mathrm{u}^{(i)}(t) + I_\mathrm{d}^{(i)}(t) + H_R^{(i)}(t) + R^{(i)}(t)\right), \\
\frac{\mathrm{d}E^{(i)}}{\mathrm{d}t}(t) &= \frac{S^{(i)}(t)}{N^{(i)}}\left(m_E^{(i)}(t)\beta_E^{(i)}E^{(i)}(t) + m_{I_\mathrm{u}}^{(i)}(t)\beta_{I_\mathrm{u}}^{(i)}I_\mathrm{u}^{(i)}(t) + m_{I_\mathrm{d}}^{(i)}(t)\beta_{I_\mathrm{d}}^{(i)}I_\mathrm{d}^{(i)}(t)\right) \\
&\quad +\frac{S^{(i)}(t)}{N^{(i)}}\left(m_{H_R}^{(i)}(t)\beta_{H_R}^{(i)}H_R^{(i)}(t) + m_{H_D}^{(i)}(t)\beta_{H_D}^{(i)}H_D^{(i)}(t)\right) \\
&\quad -\mu_m^{(i)}E^{(i)}(t) - \gamma_E E^{(i)}(t) + \tau_1^{(i)}(t) - \tau_2^{(i)}(t), \\
\frac{\mathrm{d}I_\mathrm{d}^{(i)}}{\mathrm{d}t}(t) &= \theta^{(i)}(t)\gamma_E E^{(i)}(t) - (\mu_m^{(i)} + \gamma_{I_\mathrm{d}}^{(i)}(t))I_\mathrm{d}^{(i)}(t), \\
\frac{\mathrm{d}I_\mathrm{u}^{(i)}}{\mathrm{d}t}(t) &= (1-\theta^{(i)}(t))\gamma_E E^{(i)}(t) - (\mu_m^{(i)} + \gamma_{\mathrm{Inf}})I_\mathrm{u}^{(i)}(t), \\
\frac{\mathrm{d}H_R^{(i)}}{\mathrm{d}t}(t) &= (1-\omega^{(i)}(t))\gamma_{I_d}^{(i)}(t)I_{I_d}^{(i)}(t) - \gamma_{H_R}H_R^{(i)}(t), \\
\frac{\mathrm{d}H_D^{(i)}}{\mathrm{d}t}(t) &= \omega^{(i)}(t)\gamma_{I_\mathrm{d}}^{(i)}(t)I_{I_d}^{(i)}(t) - \gamma_{H_D}^{(i)}(t)H_D^{(i)}(t), \\
\frac{\mathrm{d}R^{(i)}}{\mathrm{d}t}(t) &= \gamma_{H_R}^{(i)}(t)H_R^{(i)}(t) - \mu_m^{(i)}R^{(i)}(t), \\
\frac{\mathrm{d}D^{(i)}}{\mathrm{d}t}(t) &= \gamma_{H_D}^{(i)}(t)H_D^{(i)}(t),
\end{aligned}
\tag{1}
$$

where:

- $i \in \{1, \ldots, N_C\}$, with $N_\mathrm{C} \in \mathbb{N}$ being the number of countries/territories/areas considered.
- $N^{(i)}$ is the number of people in territory $i$ before the start of the pandemic.
- $\theta(t)$ is the percentage of infected people that is detected and documented by the authorities.
- $\mu_n^{(i)} \in [0,1]$ is the natality rate ($\mathrm{day}^{-1}$) in territory $i$: the number of births per day and per capita.
- $\mu_m^{(i)} \in [0,1]$ is the mortality rate ($\mathrm{day}^{-1}$) 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).
- $\omega^{(i)}(t) \in [0,1]$ is the disease fatality percentage in territory $i$ at time $t$: the percentage of persons who do not survive the disease.

- $\beta_E^{(i)}, \beta_{I_u}^{(i)}, \beta_{I_d}^{(i)}, \beta_{H_R}^{(i)}, \beta_{H_D}^{(i)} \in \mathbb{R}^+$ are the disease contact rates (day$^{-1}$) of a person in the corresponding states, in territory $i$.

- $\gamma_E$, $\gamma_{\text{Inf}}$ are the transition rates (day$^{-1}$) from state $E$ to states $I_{\text{d}}$ or $I_{\text{u}}$ and from state $I_{\text{d}}$ or $I_{\text{u}}$ to state $R$, respectively. They are the same for all the territories.

- We assume that the infectious time from the beginning of state $I_{\text{d}}$ or $I_{\text{u}}$ to the beginning of state $R$ is $\frac{1}{\gamma_{\text{Inf}}}$, we have that

$$\frac{1}{\gamma^{(i)}(t)} + \frac{1}{\gamma_{H_R}^{(i)}(t)} = \frac{1}{\gamma_{\text{Inf}}}, \text{ which implies that } \gamma_{H_R}^{(i)}(t) = \frac{\gamma_{\text{Inf}}\gamma_{I_{\text{d}}}^{(i)}(t)}{\gamma_{I_{\text{d}}}^{(i)}(t) - \gamma_{\text{Inf}}}.$$

  We could have also assumed the case of $\gamma_{H_R}^{(i)}$ not dependent on $\gamma_{\text{Inf}}$ and $\gamma_{I_{\text{d}}}^{(i)}$. In this case we would have needed another compartment $R_{\text{u}}$ for recovered people coming from state $I_{\text{u}}$, in order to be able to estimate the output $\text{Hos}^{(i)}(t)$ described in Section 2.4.

- $\gamma_{I_d}^{(i)}(t)$, $\gamma_{H_D}^{(i)}(t) \in (0, +\infty)$ denote the transition rates (day$^{-1}$) from states $I_{\text{d}}$, $H_D$ to state $H_R$, $D$, respectively.

- $m_E^{(i)}(t)$, $m_{I_{\text{u}}}^{(i)}(t)$, $m_{I_{\text{d}}}^{(i)}(t)$, $m_{H_R}^{(i)}(t)$, $m_{H_D}^{(i)}(t) \in [0, 1]$ (%) are functions representing the efficiency of the control measures applied to the corresponding states, 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 [11]).

System (1) is completed with initial data $S^{(i)}(t_0)$, $E^{(i)}(t_0)$, $I_{\text{d}}^{(i)}(t_0)$, $I_{\text{u}}^{(i)}(t_0)$, $H_R^{(i)}(t_0)$, $H_D^{(i)}(t_0)$, $R^{(i)}(t_0)$ and $D^{(i)}(t_0)$ given in $[0, \infty)$, for $i$=1,.., $N_C$.

We point out that the 8th equation of system (1) is not coupled with the other equations. Thus, we can solve the first seven equations of that system and the solution of the last one can be computed as follows:

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

We could have also considered 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 Figure 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.

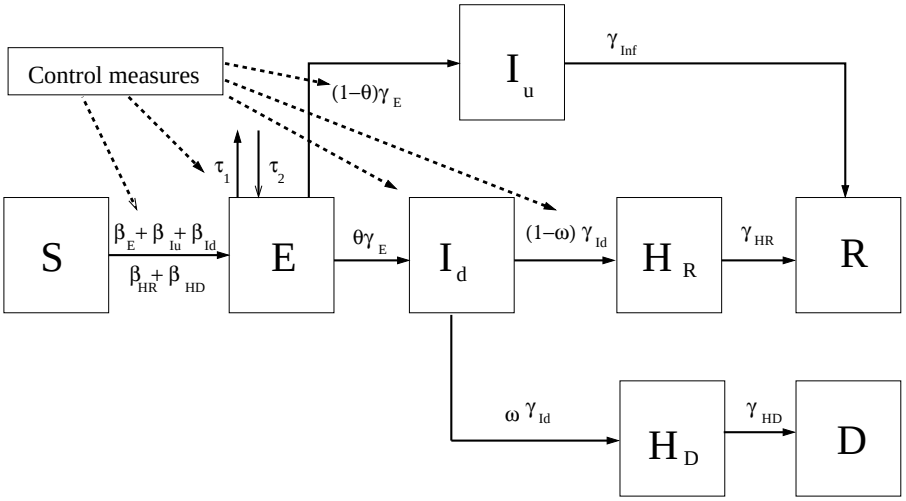


Figure 2: Diagram summarizing the simplified version of the model for COVID-19 given by system (3).

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t}(t) &= -\frac{S(t)}{N}\Big(m_E(t)\beta_E E(t) + m_{I_{\mathrm{u}}}(t)\beta_{I_{\mathrm{u}}} I_{\mathrm{u}}(t) + m_{I_{\mathrm{d}}}(t)\beta_{I_{\mathrm{d}}} I_{\mathrm{d}}(t)\Big) \\
&\quad -\frac{S(t)}{N}\Big(m_{H_R}(t)\beta_{H_R} H_R(t) + m_{H_D}(t)\beta_{H_D} H_D(t)\Big), \\
\frac{\mathrm{d}E}{\mathrm{d}t}(t) &= \frac{S(t)}{N}\Big(m_E(t)\beta_E E(t) + m_{I_{\mathrm{u}}}(t)\beta_{I_{\mathrm{u}}} I_{\mathrm{u}}t) + m_{I_{\mathrm{d}}}(t)\beta_{I_{\mathrm{d}}} I_{\mathrm{d}}(t)\Big) \\
&\quad +\frac{S(t)}{N}\Big(m_{H_R}(t)\beta_{H_R} H_R(t) + m_{H_D}(t)\beta_{H_D} H_D(t)\Big) - \gamma_E E(t) + \tau_1(t) - \tau_2(t), \\
\frac{\mathrm{d}I_{\mathrm{d}}}{\mathrm{d}t}(t) &= \theta(t)\gamma_E E(t) - \gamma_{I_{\mathrm{d}}}(t) I_{\mathrm{d}}(t), \\
\frac{\mathrm{d}I_{\mathrm{u}}}{\mathrm{d}t}(t) &= (1-\theta(t))\gamma_E E(t) - \gamma_{\mathrm{Inf}} I_{\mathrm{u}}(t), \\
\frac{\mathrm{d}H_R}{\mathrm{d}t}(t) &= (1-\omega(t))\gamma_{I_d}(t) I_{\mathrm{d}}(t) - \gamma_{H_R} H_R(t), \\
\frac{\mathrm{d}H_D}{\mathrm{d}t}(t) &= \omega(t)\gamma_{I_{\mathrm{d}}}(t) I_{\mathrm{d}}(t) - \gamma_{H_D}(t) H_D(t), \\
\frac{\mathrm{d}R}{\mathrm{d}t}(t) &= \gamma_{H_R}(t) H_R(t), \\
\frac{\mathrm{d}D}{\mathrm{d}t}(t) &= \gamma_{H_D}(t) H_D(t),
\end{aligned}
\tag{3}
$$

We remark that, with this simplification the 7th and 8th equation 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 two equations can be

computed by using (2) and

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

For the numerical simulations presented in section 4, the first six equations of System (3) were numerically solved with the classic Runge–Kutta method of 4 stages and order 4 (RK4), with 4 h 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

  $$c_{\mathrm{m}}(t) = H_R(t) + H_D(t) + R(t) + D(t)$$

  and can be also computed as follows:

  $$c_{\mathrm{m}}(t) = c_{\mathrm{m}}(t_0) + \int_{t_0}^{t} \frac{\mathrm{d}\Big(H_R + H_D + R + D\Big)}{\mathrm{d}t}(s)\mathrm{d}s = c_{\mathrm{m}}(t_0) + \int_{t_0}^{t} \gamma_{I_{\mathrm{d}}}(s) I_{\mathrm{d}}(s)\mathrm{d}s.$$

- $d_{\mathrm{m}}(t)$: The model cumulative number of deaths (due to COVID-19), at day $t$ (in territory $i$), which is given by $D(t)$.

- $\mathrm{R}_0$ and $\mathrm{R}_{\mathrm{e}}$: 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, $\mathrm{R}_{\mathrm{e}}(0) = \mathrm{R}_0$. Typically, the spread of the disease slows down when $\mathrm{R}_{\mathrm{e}}(i,t) < 1$.

  We apply the Next Generation Matrix method (see [32]) to system (3) and obtain that:

  $$\mathrm{R}_0 = \frac{\Big(\left(\beta_{H_R}\left(1-\omega\right)\theta\gamma_E + \beta_E\gamma_{\mathrm{Inf}} + \beta_{I_u}\gamma_E\left(1-\theta\right)\right)\gamma_{I_d} + \left(\beta_{I_d} - \beta_{H_R}\left(1-\omega\right)\right)\theta\gamma_E\gamma_{\mathrm{Inf}}\Big)\gamma_{H_D} + \beta_{H_D}\omega\theta\gamma_E\gamma_{I_d}\gamma_{\mathrm{Inf}}}{\gamma_E\gamma_{I_d}\gamma_{H_D}\gamma_{\mathrm{Inf}}},$$

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

  $$\mathrm{R}_e(t) = \frac{U_e(t)}{\gamma_E\gamma_{I_d}\gamma_{H_D}\gamma_{\mathrm{Inf}}}\frac{S(t)}{N},$$

  where

  $$\begin{aligned} U_e(t) \;=\; & \Big(\left(m_{H_R}\beta_{H_R}\left(1-\omega\right)\theta\gamma_E + m_E\beta_E\gamma_{\mathrm{Inf}} + m_{I_u}\beta_{I_u}\gamma_E\left(1-\theta\right)\right)\gamma_{I_d} \\ & + \left(m_{I_d}\beta_{I_d} - m_{H_R}\beta_{H_R}\left(1-\omega\right)\right)\theta\gamma_E\gamma_{\mathrm{Inf}}\Big)\gamma_{H_D} + m_{H_D}\beta_{H_D}\omega\theta\gamma_E\gamma_{I_d}\gamma_{\mathrm{Inf}}, \end{aligned}$$

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

- The number of people detected by the authorities as recovered from COVID-19 can be estimated from the model as follows:

$$\mathrm{Rec}(t) = \frac{\theta(1-\omega)}{\Big(1-\theta\Big) + \Big(\theta(1-\omega)\Big)} R(t),$$

- Hos($t$): The number of people in hospital considering only severe and critical cases is estimated as follows:

$$\mathrm{Hos}(t) = H_D(t) + p(t)\left(H_R(t) + (\mathrm{Rec}(t) - \mathrm{Rec}(t - C_o))\right),$$

  where $p(t)$ is the ratio, at time $t$, of people in state $H_R$ that are hospitalized and $C_{\mathrm{o}}$ 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 persons 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.

- $\Gamma_E(t), \Gamma_{I_u}(t), \Gamma_{I_d}(t)$, and $\Gamma_H(t)$: which correspond to the amount of people infected during the time interval $[t_0, T]$, by contact with persons in state $E$, $I_u$, $I_s$ and $H = H_R + H_D$, respectively. They are given by:

$$\begin{aligned}
\Gamma_E(t) &= \int_{t_0}^{t} m_E(s)\beta_E E(s)\frac{S(s)}{N}\mathrm{d}s,\\
\Gamma_{I_u}(t) &= \int_{t_0}^{t} m_{I_d}(s)\beta_{I_u} I_u(s)\frac{S(s)}{N}\mathrm{d}s,\\
\Gamma_{I_d}(t) &= \int_{t_0}^{t} m_{I_d}(s)\beta_{I_d} I_d(s)\frac{S(s)}{N}\mathrm{d}s,\\
\Gamma_H(t) &= \int_{t_0}^{t} \Big(m_{H_R}(s)\beta_{H_R} H_R(s) + m_{H_D}(s)\beta_{H_D} H_D(s)\Big)\frac{S(s)}{N}\mathrm{d}s.
\end{aligned}$$

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])

# 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 [1]).

There are some studies about the relevance of humidity and temperature for the spread of COVID-19. In [18] 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 ([35]) 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 [23]) we have obtained the reported cases per day (i.e. persons who enter each day in states $H_R$ or $H_D$) and the reported COVID-19 deaths per day. For each day $t$ they are denoted by $c_{\mathrm{r}}(t)$ and $d_{\mathrm{r}}(t)$, 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 [36]). 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.

Table 1: Summary of some parameters calibrated for COVID-19 in China, which can be found in [9, 17, 27]. A brief description (**Description**) and the range of the considered values (**Value**) are reported.

| **Notation** | **Value** | **Description** | **Reference** |
|---|---|---|---|
| $\gamma_E$ | 0.1818 | Transition rate of a person in state $E$ (day$^{-1}$) | [27] |
| $\gamma_{I_d}(t)$ | [0.1493,1.4286] | Transition rate of a person in state $I$(day$^{-1}$) at time $t$ | [17] |
| $\gamma_{H_R}(t)$ | [0.0752,0.1370] | Transition rate of a person in state $H$ to state $R$ (day$^{-1}$) at time $t$ | [27] |
| $\gamma_{H_D}(t)$ | [0.0493,0.0699] | Transition rate of a person in state $H$ to state $D$ (day$^{-1}$) at time $t$ | [27] |
| $\gamma_{\mathrm{Inf}}$ | 0.0714 | Transition rate of a person in state $I_u$ to state $R$ (day$^{-1}$) at time $t$ | [27] |
| $C_{\mathrm{o}}$ | 14 | The period of convalescence (day) | [9] |

## 3.3 Control measures

Focusing on the application of the control measures, we multiply the disease contact rates (i.e. $\beta_E$, $\beta_I$, $\beta_{H_R}$ and $\beta_{H_D}$) by decreasing functions simulating the reduction of these rates as the control measures efficiency is increased. Here, we have considered the functions (see [15]):

$$m_E(t) = m_I(t) = m_H(t) = m_D(t) = \begin{cases} (m_0 - m_1)\exp\Big(-\kappa_1(t-\lambda_0)\Big) + m_1 & t \in [t_0, \lambda_1) \\ (m_1 - m_2)\exp\Big(-\kappa_2(t-\lambda_1)\Big) + m_2 & t \in [\lambda_1, \lambda_2) \\ \dots \\ (m_{q-1} - m_q)\exp\Big(-\kappa_q(t-\lambda_{q-1})\Big) + m_q & t \in [\lambda_{q-1}, \infty), \end{cases} \tag{4}$$

where, for every $j \in \{0, ..., q\}$, $m_j \in [0, 1]$ 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 $\lambda_j$ 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 Fatality rate

The fatality rate $\omega(t)$ depends on the territory $i$ and time $t$. As osbserved in other epidemics (see, e.g. [11]), 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:

$$\omega(t) = m_{I_\mathrm{d}}(t)\overline{\omega} + (1 - m_{I_\mathrm{d}}(t))\underline{\omega}, \tag{5}$$

where $\overline{\omega} \in [0, 1]$ is the fatality rate observed when no control measures are applied (i.e. $m_{I_\mathrm{d}}(t) \equiv 1$); and $\underline{\omega} \in [0, 1]$ is the fatality rate observed when control measures are fully applied (i.e. $m_{I_\mathrm{d}}(t) \equiv 0$). 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 $d_{I_\mathrm{d}}$, $d_{I_\mathrm{u}} = \frac{1}{\gamma_{\mathrm{Inf}}}$, $d_{H_R}$ and $d_{H_D}$ the mean duration in days of a person from state $I_\mathrm{d}$ to $H_R$ or $H_D$, from state $I_\mathrm{u}$ to $R$, from state $H_R$ to $R$ and from state $H_D$ to $D$, respectively, without the application of control measures. Additionally, we consider that:

- According to [16], the transition rate from $E$ to $I_d$ or $I_u$ depends only on the disease and, therefore, is considered constant.
- According to [16, 17], it has been observed that the value of $\gamma_{I_\mathrm{d}}(t)$ can be increased due to the application of control measures (i.e. people with symptoms are detected earlier). As a consequence, the values of $\gamma_{H_R}(t)$ and $\gamma_{H_D}(t)$ are decreased (i.e. people with symptoms stay under observation more time).
- According to [16], $d_{H_R}$ is lower than $d_{H_D}$. Thus, we consider $d_{H_D} = d_{H_R} + \delta_R$, where $\delta_R \geq 0$.
- As highlighted previously, we consider that $d_{I_\mathrm{u}} = d_{I_\mathrm{d}}(t) + d_{H_R}(t)$.

To simulate those effects, following [11], we use the following functions:

$$\gamma_{I_\mathrm{d}}(t) = \frac{1}{d_{I_\mathrm{d}} - g(t)}\ (\mathrm{day}^{-1}), \quad \gamma_{H_R}(t) = \frac{1}{d_{I_\mathrm{u}} - d_{I_\mathrm{d}} + g(t)}\ (\mathrm{day}^{-1}), \quad \gamma_{H_D}(t) = \frac{1}{d_{H_R} + \delta_R}\ (\mathrm{day}^{-1}),$$

where $g(t) = d_g(1 - m_{I_\mathrm{d}}(t))$ represents the decrease of the duration $d_{I_\mathrm{d}}$ due to the application of control measures, at time $t$; and $d_g$ is the maximum number of days that $d_{I_\mathrm{d}}$ can be decreased due to the control measures.

### 3.6 Disease contact rates

Here, following the idea proposed in [11, 16], we assume that there exist a relationship between the contact rates $\beta_{I_\mathrm{d}}$, $\beta_{I_\mathrm{u}}$, $\beta_E$, $\beta_{H_R}$ and $\beta_{H_D}$.

More precisely, it was reported that people in states $E$, $I_\mathrm{u}$, $H_R$ and $H_D$ are less infective than people in state $I_\mathrm{d}$ (due to their lower virus load or isolation measures, see [16]). Thus, we consider that $\beta_E = C_E\beta_{I_\mathrm{d}}$, $\beta_{I_\mathrm{u}} = C_u\beta_{I_\mathrm{d}}$ and $\beta_{H_R}(t) = \beta_{H_D}(t) = C_H(t)\beta_{I_\mathrm{d}}(t)$, where $C_E$, $C_\mathrm{u}$, $C_H(t) \in [0, 1]$. 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 parameter with a multiobjective technique

First, we determine the set of unknown model parameters to be identified by the approach presented here. We denote this set by $\Omega_p$. Then, we follow the methodology proposed in [6].

More precisely, we define a multi-objective optimization problem. This problem is based on the minimization of several objectives functions which compute the difference between some of the model outputs and reported data (such as, the number of cases or deaths) for a particular set of values for $\Omega_p$.

To solve this multi-objective optimization problem, the algorithm called *Weighting Achievement Scalarizing Function Genetic Algorithm* (WASF-GA) is applied [7, 8, 31]. From a general point of view, WASF-GA is a population-based evolutionary algorithm that explores the feasible region looking for efficient solutions using an iterative procedure. As a result, WASF-GA obtains a set of solutions covering a region of the Pareto Optimal front. Finally, we must specify a specific criteria to choose a particular point.

Here, we apply the WASF-GA configurations and parameters considered in [6], Section 4.1.2.

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

### 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 [33]).

Following [16, 17], we set $d_E = 5.5$ days (i.e., $\gamma_E = 1/5.5$ days$^{-1}$), $d_{I_{\rm d}} = 6.7$days, $d_g = 5.7$ days, $d_{I_{\rm u}}$=14 days (i.e., $\gamma_{\rm inf} = 1/14$ days$^{-1}$).

Additionally, according to [28], on 20 February 2020, 2055 healthcare workers were infected in China. At this date, 74651 cases were reported in China. From this data, we assume that the infection due to contact with persons in state $H_R$ or $H_D$ should represent around $100\frac{2055}{74651}\% = 2.75\%$ of the number of cases. Thus, we compute $\beta_H(t) = C_H(t)\beta_{I_d}$ such that

$$\frac{C_H\beta_{I_d}((1-\omega)\frac{1}{\gamma_{H_R}} + \omega\frac{1}{\gamma_{H_D}})}{C_H\beta_{I_d}((1-\omega)\frac{1}{\gamma_{H_R}} + \omega\frac{1}{\gamma_{H_D}}) + \beta_{I_d}\frac{1}{\gamma_I} + \beta_E\frac{1}{\gamma_E}} = 0.0275.$$

In the previous expression, coefficients depend on territory $i$ and time $t$. This implies that,

$$C_H(t) = \frac{0.0275(\beta_{I_d}\frac{1}{\gamma_I(t)} + \beta_E\frac{1}{\gamma_E})}{(1-0.0275)\beta_{I_d}((1-\omega(t))\frac{1}{\gamma_{H_R}(t)} + \omega(t)\frac{1}{\gamma_{H_D}(t)})}.$$

Following data reported by the WHO, we have aggregated data from the Chinese Mainland, Macao, Hong-Kong and Taiwan. Additionally, we assume that the movement of people in and out of China is negligible in orden to estimate the number of cases and deaths in China. Therefore, for the sake of simplicity, we assume in this section that $\tau_1 = \tau_2 \equiv 0$. Nevertheless, once the number of new cases is under small, one can include them in the model.

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 [20]). This supposed a sudden increase of around 15.000 cases in one day ([14]).

In addition, up to 16 February 2020, the WHO decided to continue reporting only cases confirmed with laboratory tests in its situation reports ([19]). On 17 February 2020, the WHO changed its protocol to count cases and decided (see [21]) 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 ([20]).

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 [22]).

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 figures.

In order to smoothly distribute the sudden increase of 17410 cases in the number of cases reported before the 17 February 2020, we distributed those 17410 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.

We report in Figure 3, the evolution of $c_{\mathrm{r}}$ and $c_{\mathrm{ar}}$ in China from 12 January 2020 up to 21 March 2020.

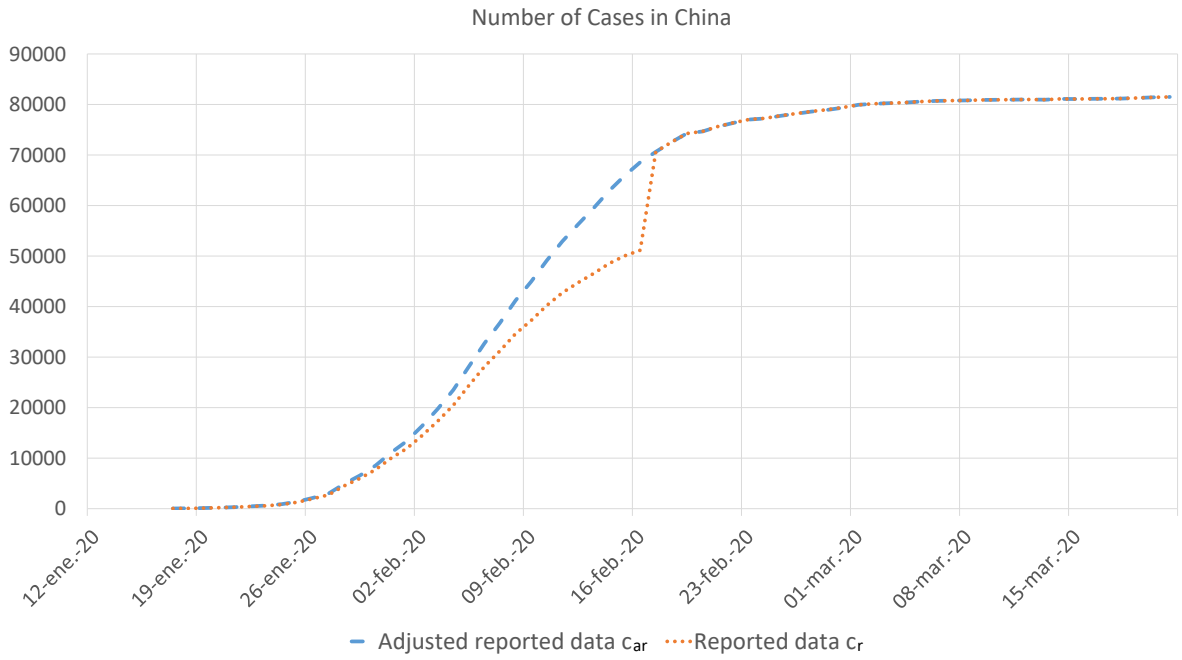


Figure 3: Evolution of the number of cases in China: Reported ($c_{\mathrm{r}}$) and adjusted reported data ($c_{\mathrm{ar}}$).

.

According to [16], 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 [16]). Therefore, we consider here that figures 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

$$\theta(t) = \begin{cases} 0.86 & \text{if 1 December 2019} \leq t \leq \text{24 January 2020} \\ \text{linear continuous} & \text{if 24 January 2020} \leq t \leq \text{8 February 2020} \\ 0.35 & \text{if 8 February 2020} \leq t \end{cases}$$

Thus, even if the reported deaths $d_{\mathrm{r}}(t)$ can be considered a good estimation of the real figures, neither the reported cases $c_{\mathrm{r}}(t)$ nor the adjusted reported cases $c_{\mathrm{r}}(t)$ described above would be a good estimation of the real number of infected cases $c_{\mathrm{real}}(t)$, which can be computed as

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

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 ([23]). Therefore, we considered the implementation of control measures after this date, in order to represent the real situation of the measures imposed:

$$m_E(t) = m_{I_u}(t) = m_{I_d}(t) = m_{H_R}(t) = m_{H_D}(t) = \begin{cases} 1, & \text{if } t \in [1 \text{ December } 2019, \lambda_1) \\ \exp\Big( -\kappa_1(t-\lambda_1)\Big), & \text{if } t \in [\lambda_1, 21 \text{ March } 2020] \end{cases}$$

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: $\beta_{I_{\mathrm{d}}}$, $C_E$, $C_u$, $\delta_R$, $\delta_\omega$, $\underline{\omega}$ and $\kappa_1$. All those parameters are bounded according to values found in the literature (see [16]). 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 .

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** |
|---|---|---|
| $\beta_{I_{\mathrm{d}}}$ | [0,0.5] | disease contact rate (day$^{-1}$) of a person in states $I_{\mathrm{d}}$. |
| $C_E$ | [0,1] | used to estimate $\beta_E = C_E\beta_{I_{\mathrm{d}}}$ |
| $C_u$ | [0,1] | used to estimate $\beta_{I_{\mathrm{u}}} = C_u\beta_{I_{\mathrm{d}}}$ |
| $\delta_R$ | [7,14] | used to estimate $d_{H_D} = d_{H_R} + \delta_R$ |
| $\delta_\omega$ | [0,0.2] | used to estimate $\overline{\omega} = \underline{\omega} + \delta_\omega$ |
| $\underline{\omega}$ | [0.005,0.05] | minimum fatality rate during the epidemic |
| $\kappa_1$ | [0,1] | efficiency of the control measures |

## 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 $c_{\mathrm{ar}}(t)$ and $d_{\mathrm{r}}(t)$, from the index case up to different values of the final date $t_{\mathrm{f}}$. We set $t_{\mathrm{f}}$ to the following particular values:

- $t_{\mathrm{f}}$=21 March 2020: this date corresponds to that of the last data available before running the numerical experiments. We denote this experiment by $\mathbf{EXP}_{21\mathrm{M}}$.
- $t_{\mathrm{f}}$=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 $\mathbf{EXP}_{25\mathrm{F}}$.
- $t_{\mathrm{f}}$=8 February 2020: this date is considered as the inflection point for which the daily number of reported cases start to decrease, we denote this experiment by $\mathbf{EXP}_{08\mathrm{F}}$.

- $t_{\mathrm{f}}$=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 $\mathbf{EXP}_{29\mathrm{J}}$.

The objective is to show how the model and the parameter estimation methodology deal with incomplete data.

In Table 3, we report 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 $\beta_{I_{\mathrm{d}}}$, $C_E$, $C_u$ and $\kappa_1$ show a progressive evolution to the value obtained with $\mathbf{EXP}_{21\mathrm{M}}$ 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 $\mathbf{EXP}_{21M}$ and $\mathbf{EXP}_{29\mathrm{J}}$. 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 estimated results.

Additionally, parameters $\delta_\omega$ and $\underline{\omega}$ change drastically when comparing the outputs of $\mathbf{EXP}_{21M}$ with those of $\mathbf{EXP}_{25\mathrm{F}}$, $\mathbf{EXP}_{08\mathrm{F}}$ and $\mathbf{EXP}_{29\mathrm{J}}$. This is consistent with real observation reported by the WHO, where the number of reported death drastically decreased during last month (see Figure 4). Finally, focusing on $\delta_R$ we observe that the value generated with $\mathbf{EXP}_{21M}$ clearly differs from the one obtained with $\mathbf{EXP}_{25\mathrm{F}}$, $\mathbf{EXP}_{08\mathrm{F}}$ and $\mathbf{EXP}_{29\mathrm{J}}$. We note that this particular parameter was already reported in the literature (see, e.g. [16]) as being difficult to be estimated (considering a range of one month). This radical change of value could be due to the clear decrease in the number of daily reported cases observed since the end of February. 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.

Table 3: Results obtained from the parameter identification step for experiments $\mathbf{EXP}_{21\mathrm{M}}$, $\mathbf{EXP}_{25\mathrm{F}}$, $\mathbf{EXP}_{08\mathrm{F}}$ and $\mathbf{EXP}_{29\mathrm{J}}$.

| **Notation** | $\mathbf{EXP}_{21\mathrm{M}}$ | $\mathbf{EXP}_{25\mathrm{F}}$ | $\mathbf{EXP}_{8\mathrm{F}}$ | $\mathbf{EXP}_{29\mathrm{J}}$ |
|---|---|---|---|---|
| $\beta_{I_{\mathrm{d}}}$ | 0.3373 | 0.2771 | 0.4526 | 0.5000 |
| $C_E$ | 0.8450 | 0.9426 | 0.2665 | 0.1543 |
| $C_u$ | 0.3622 | 0.6336 | 0.9255 | 0.9988 |
| $\delta_R$ | 7.0004 | 13.9 | 13.2 | 13.7 |
| $\delta_\omega$ | 0.1815 | 0.1070 | 0.1476 | 0.1128 |
| $\underline{\omega}$ | 0.0050 | 0.0458 | 0.0330 | 0.0500 |
| $\kappa_1$ | 0.1200 | 0.1272 | 0.1647 | 0.2354 |

In Figure 4, we depict the evolution of the number of cases and deaths in China from 1 December 2019 to 21 March 2020 from observed data and results obtained at the end of experiments $\mathbf{EXP}_{21\mathrm{M}}$, $\mathbf{EXP}_{25\mathrm{F}}$, $\mathbf{EXP}_{25\mathrm{F}}$ and $\mathbf{EXP}_{29\mathrm{J}}$. We observe that $\mathbf{EXP}_{29\mathrm{J}}$ forecasts poorly and underestimates the evolution of the epidemic up to the 21 March 2020. However, this forecast captures reasonably the deceleration of the daily number of cases observed at the end of February. This indicates that, as expected, forecasting at an early stage of the epidemic (before the inflection point in the curve of the reported cases) is a quite difficult task. However, in all cases, it seems that our approach underestimates the number of cases: $-40\%$ of final cases for $\mathbf{EXP}_{29\mathrm{J}}$, $-11\%$ for $\mathbf{EXP}_{25\mathrm{F}}$ and less than $-1\%$ for $\mathbf{EXP}_{25\mathrm{F}}$ and $\mathbf{EXP}_{21\mathrm{M}}$. Thus, for forecast of COVID-19 epidemics performed at an early stage of the epidemic, results could vary up to 40%.

Focusing on $\mathbf{EXP}_{25\mathrm{F}}$ and $\mathbf{EXP}_{08\mathrm{F}}$, both forecasts could be considered as acceptable, with a total number of reported cases at a reasonable distance of the real values. However, focusing on the number of deaths, it seems that they overestimate the dynamic of the reported deaths near 21 March 2020. This is not a surprising result as, as said previously, the number of reported daily deaths drastically decreased at the end of February.

We now focus in results returned at the end of Experiment $\mathbf{EXP}_{21\mathrm{M}}$ to analyze the main key factor that may explain the dynamics of the COVID-19 epidemic in China up to 21 March 2020.

In Figure 5, we report the evolution of the reported cases and $I_u(t)$, $I_d(t)$ and $I_u(t)+I_d(t)$ estimated by our model. We observe that the final number of undetected cases (around 114.600 cases) estimated by our model represents around 59% of the total number of cases (around 195.700 cases). Furthermore, it seems that control measures have not contained their growing as fast as for the detected cases. This results is interesting as it seems to indicate that,

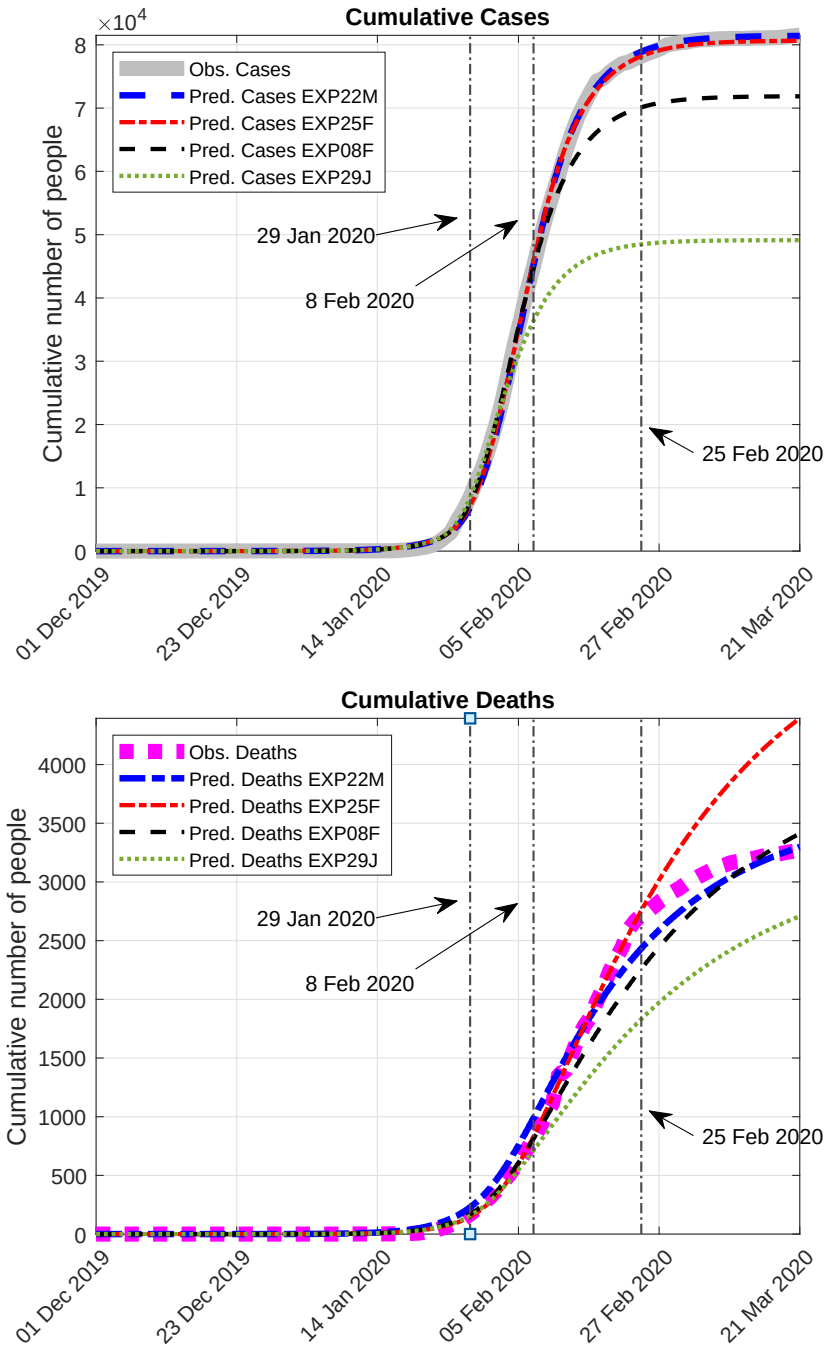


Figure 4: Evolution of the number of **(TOP)** cases and **(BOTTOM)** deaths in China from 1 December 2019 to 21 March 2020: Adjusted observed data, and estimation obtained with $\mathbf{EXP}_{21\mathrm{M}}$, $\mathbf{EXP}_{25\mathrm{F}}$, $\mathbf{EXP}_{08\mathrm{F}}$ and $\mathbf{EXP}_{29\mathrm{J}}$

.

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.

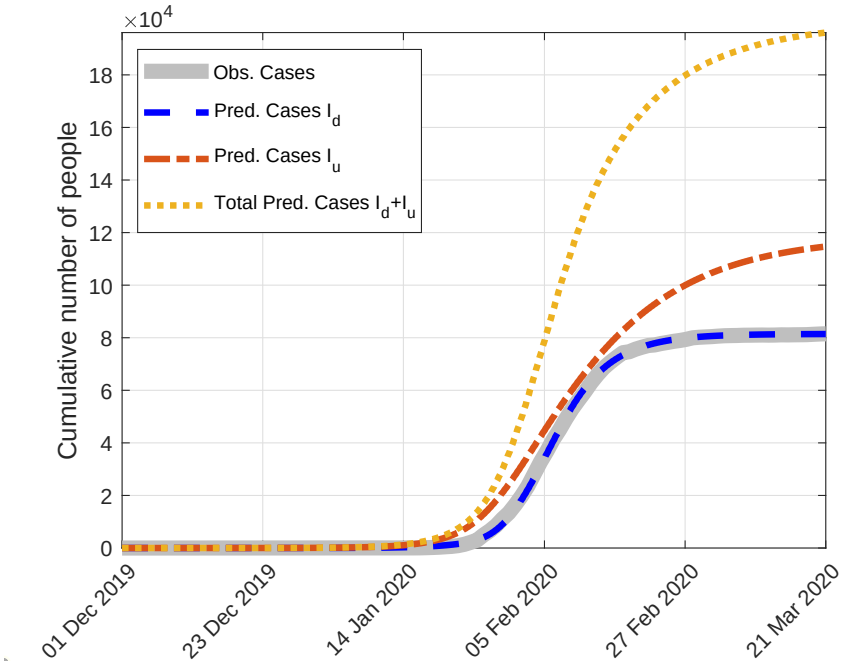


Figure 5: Evolution of the number of cases in China from 1 December 2019 to 21 March 2020: Adjusted reported data and $I_u(t)$, $I_d(t)$ and $I_u(t) + I_d(t)$ obtained with $\mathbf{EXP}_{21\mathrm{M}}$.

In Figure 6, we report the evolution of Hospitalized and Recovered people estimated by our model. The model estimates that the peak of number of persons hospitalized at the same time has been reached around 16 February 2020, with around 54.485 hospitalized patients. According to data reported in

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

the peak was reached on 17 February 2020 with around 58000 hospitalized persons at the same time. The model estimates quite reasonably the number of hospitalized people. However, the obtained results underestimate the real observations by around 6%. Focusing on Recovered persons, our model return a final number of 47.884 persons, which clearly underestimate the real observation (around 72.000 recovered people on 21 March 2020). Further investigation on the recovery rate (and convalescence period) should be done to better estimate this particular output.

In Figure 7, we report the evolution of the effective reproduction number $\mathrm{R_e}(t)$. 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 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 $\mathrm{R}_0$ is 3.37, which is bigger than other values reported in the literature (between 2.5 and 3.0, see [16]). This could be explained by the fact that we have taken into account the undetected persons.

Additionally, persons in state $E$ seem to be one of the main sources of infection and represent around 67% of the infections. This could be explained by the fact that, due to the application of control measures, people in state $I_\mathrm{d}$ are quickly hospitalized. Persons in state $I_\mathrm{d}$ caused around 23% of infections. Those results seem to indicate that persons 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 0.3% of the infections (it should be around 3%). A more precise approach should be proposed to estimate better the value of $C_H$.

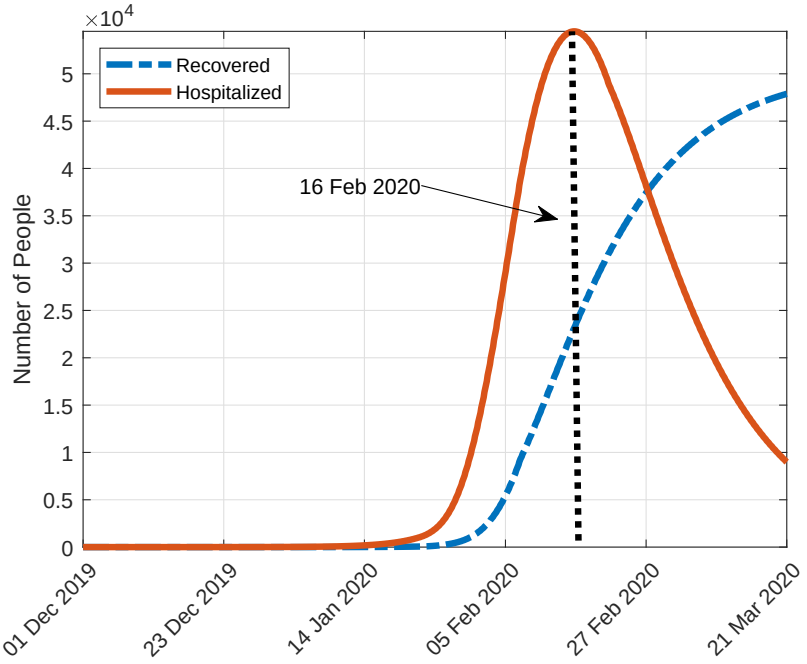


Figure 6: Evolution of the number of Hospitalized and Recovered people in China from 1 December 2019 to 21 March 2020 obtained with $\mathbf{EXP}_{21\mathrm{M}}$.

Table 4: Results obtained from the parameter identification step for experiments $\mathbf{EXP}_{21\mathrm{M}}$, $\mathbf{EXP}_{25\mathrm{F}}$, $\mathbf{EXP}_{08\mathrm{F}}$ and $\mathbf{EXP}_{29\mathrm{J}}$.

| **Notation** | **Value** |
|---|---|
| $R_0$ | 3.3701 |
| $\beta_{I_E}$ | 0.2850 |
| $\beta_{I_\mathrm{d}}$ | 0.3373 |
| $\beta_{I_\mathrm{u}}$ | 0.1222 |
| $\beta_{H_R}$ | [0.0038, 0.0126] |
| $\beta_{H_D}$ | [0.0038, 0.0126] |
| $\omega(t)$ | [0.0052, 0.1865] |
| $\Gamma_E(t_\mathrm{f})$ | 134710 |
| $\Gamma_{I_\mathrm{d}}(t_\mathrm{f})$ | 18471 |
| $\Gamma_{I_\mathrm{u}}(t_\mathrm{f})$ | 46353 |
| $\Gamma_H(t_\mathrm{f})$ | 520 |

**Acknowledgments:** This research was partially supported by the Spanish Government under project MTM2015-64865-P; and the research group MOMAT (Ref. 910480) supported by the Complutense University of Madrid.

# References

[1] World Bank. World Bank Open Data. `https://data.worldbank.org/`, March 2020.

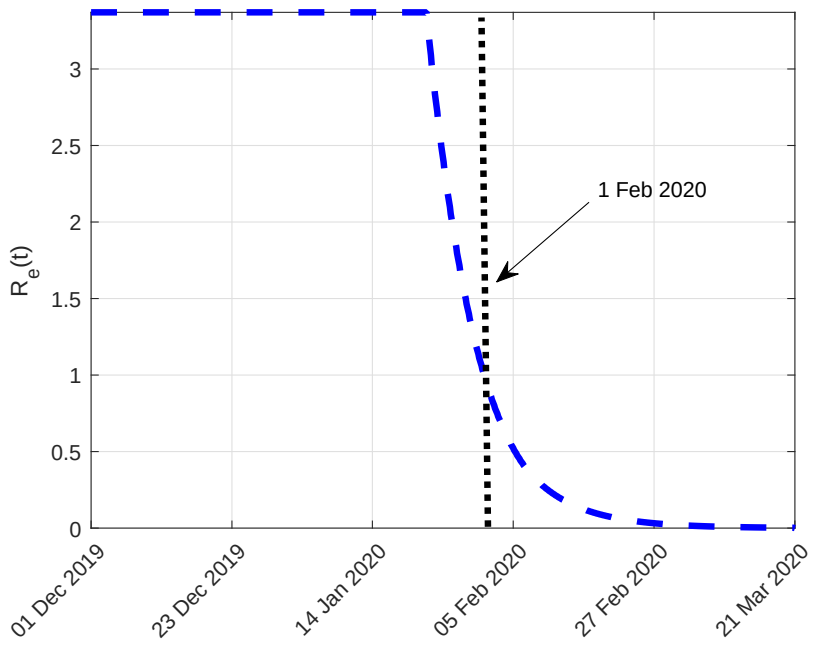


Figure 7: Evolution of the effective reproduction number $\mathrm{R_e}(t)$ in China from 1 December 2019 to 21 March 2020 obtained with $\mathbf{EXP}_{21\mathrm{M}}$.


[2] F Brauer and C Castillo-Chávez. *Mathematical Models in Population Biology and Epidemiology*. Texts in Applied Mathematics, Springer, 2001.

[3] S Chen, J Yang, W Yang, and T Barnighausen. COVID-19 control in China during mass population movements at New Year. *The Lancet*, 395(10226):764–766, 2020.

[4] O Diekmann, H Heesterbeek, and T Britton. *Understanding Infectious Disease Dynamics*. Princeton Series in Theoretical and Computational Biology, Princeton University Press, 2013.

[5] M R Ferrández, B Ivorra, P M Ortigosa, A M Ramos, and J L Redondo. Application of the Be-CoDis model to the 2018-19 Ebola Virus Disease outbreak in the Democratic Republic of Congo. *ResearchGate Preprint*, 23 July 2019:1–17, 2019.

[6] M R Ferrández, B Ivorra, J L Redondo, A M Ramos, and P M Ortigosa. A multi-objective approach to estimate parameters of compartmental epidemiological models. Application to Ebola Virus Disease epidemics. *researchgate.net*, pages 1–49, 2020.

[7] M. R. Ferrández, S. Puertas-Martín, J. L. Redondo, B. Ivorra, A. M. Ramos, and P. M. Ortigosa. High-performance computing for the optimization of high-pressure thermal treatments in food industry. *The Journal of Supercomputing*, 75:1187–1202, 2018.

[8] M. R. Ferrández, J. L. Redondo, B. Ivorra, A. M. Ramos, and P. M. Ortigosa. Preference-based multi-objectivization applied to decision support for high-pressure thermal processes in food treatment. *Applied Soft Computing*, 79:326–340, 2019.

[9] European Centre for Disease Prevention and 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, March 2020.

[10] A E Gorbalenya, S C Baker, R S Baric, R J de Groot, C Drosten, A A Gulyaeva, and et al. Severe acute respiratory syndrome-related coronavirus: the species and its viruses - a statement of the Coronavirus Study Group. *bioRxiv*, pages 1–20, 2020.

[11] B Ivorra, A M Ramos, and D. Ngom. 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. *Bulletin of Mathematical Biology*, 77(9):1668–1704, 2015.

[12] B Ivorra and AM Ramos. Application of the Be-CoDis mathematical model to forecast the international spread of the 2019 Wuhan coronavirus outbreak. *ResearchGate Preprint*, 9 February 2020:1–13, 2020.

[13] B Ivorra and AM Ramos. Validation of the forecasts for the international spread of the coronavirus disease 2019 (covid-19) done with the Be-CoDis mathematical model. *ResearchGate Preprint*, 28 February 2020:1–7, 2020.

[14] Johns Hopkins University (JHU). Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE). https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6, March 2020.

[15] P Lekone and B Finkenstädt. Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics*, 62(4):1170–1177, 2006.

[16] R Li, S Pei, B Chen, Y Song, T Zhang, W Yang, and J Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 2020.

[17] T Liu, J Hu, M Kang, L Lin, H Zhong, J Xiao, and et al. Transmission dynamics of 2019 novel coronavirus (2019-ncov). *bioRxiv*, 2020.

[18] W Luo, M Majumder, and Liu D. The role of absolute humidity on transmission rates of the COVID-19 outbreak. *MedRxiv*, 2020.

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

[20] World Health Organization. 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, February 2020.

[21] World Health Organization. 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, February 2020.

[22] World Health Organization. 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, February 2020.

[23] World Health Organization. Coronavirus disease (COVID-2019) Situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/, March 2020.

[24] World Health Organization. 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, March 2020.

[25] World Health Organization. Joint WHO and ECDC mission in Italy to support COVID-19 control and prevention efforts. http://www.euro.who.int/en/countries/italy/news/news/2020/2/joint-who-and-ecdc-mission-in-italy-to-support-covid-19-control-and-prevention-efforts, February 2020.

[26] World Health Organization. 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, February 2020.

[27] World Health Organization. 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, February 2020.

[28] World Health Organization. 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, February 2020.

[29] World Health Organization. Rolling updates in coronavirus disease COVID-19. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen, January 2020.

[30] World Health Organization. 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)., January 2020.

[31] A. B. Ruiz, R. Saborido, and M. Luque. A preference-based evolutionary algorithm for multiobjective optimization: the weighting achievement scalarizing function genetic algorithm. *Journal of Global Optimization*, 62(1):101–129, 2015.

[32] P Van den Driessche and J Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1-2):29–48, 2002.

[33] R Verity, L C Okell, I Dorigatti, P Winskill, C Whittaker, N Imai, G Cuomo-Dannenburg, H Thompson, P Walker, and et al. Estimates of the severity of COVID-19 disease. *medRxiv*, 2020.

[34] C Wang, P W Horby, FG Hayden, and G F Gao. A novel coronavirus outbreak of global health concern. *Lancet*, 395:470–473, 2020.

[35] M Wang, M D Aili Jiang, and L Gong. Temperature Significantly Change COVID-19 transmission in 429 cities. *MedRxiv*, 2020.

[36] Y Yang, Q B Lu, M J Liu, Y X Wang, A R Zhang, and et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in china. *Medrxiv*, 2020.