BACHELOR’S THESIS / BIOMEDICAL ENGINEERING 2020

# A Mathematical Model to Estimate the Dynamics of the Covid-19 Pandemic Using the Ongoing Public Domain Data

Clàudia Saiz Vila

upf.
Universitat
Pompeu Fabra
*Barcelona*

Bachelor thesis on Biomedical Engineering

Universitat Pompeu Fabra

# A Mathematical Model to Estimate the Dynamics of the Covid-19 Pandemic Using the Ongoing Public Domain Data

Clàudia Saiz Vila

Supervisor: Prof. Dr. Ralph G. Andrzejak

Department of Information and Communication Technologies

September 2020

Universitat
Pompeu Fabra
*Barcelona*

# Acknowledgments

I would like to acknowledge everyone who played a role in this thesis and in my academic accomplishments. Especially, I would like to thank:

**Ralph Andrzejak**, for giving me the chance to carry out the thesis in his team when I most needed it, and for his valuable advice and feedback.

**Ann Blucher** and **Alex Aparicio**, for their help and willingness.

**Mònica Ballester**, for her daily support, even in the distance.

**Aaron Vilalta, Alberto Tiana, Irene Sainz, Mar Pascual, Marc Vargas, Núria Moya** and **Oriol Cuxart** for these wonderful years of friendship, which I hope will last for many more years. Thank you for being my family and making Barcelona my home. Especially to **Núria Moya**, for sharing these days doing the thesis and for being part of me.

**My family**, for all the support, trust, patience, help and love. Especially to **my father**, thank you for all your time and dedication, I owe you most of my achievements.

# Abstract

At the end of December 2019, it was found that a new coronavirus (SARS-CoV-2) was causing pneumonia-like illness in the city of Wuhan, China. This virus started to spread very rapidly causing a global large-scale infection. The Covid-19 pandemic has produced and it is still generating a brutal impact on society, forcing the lockdown of many countries as well as the collapse of their healthcare system, leading to a considerable growth in the number of deaths. During the outbreak, most of the information and dynamics of the virus was unknown and unpredictable. Therefore, the proposed study aims to create a stochastic mathematical model based on probabilities to estimate the dynamics of the outbreak of the Covid-19 pandemic using the available public domain data. By estimating the probabilities of getting the infection and subsequently recovering or dying from it, the epidemic curves of the cumulative sum of detected infected cases, recoveries and deaths were simulated for Germany, Italy and South Korea from 22nd January to 30th June 2020. Furthermore, using the outputs provided by the proposed model, a more accurate case fatality ratio was calculated and different lockdown scenarios such as its anticipation or delay were discussed. Results have been analyzed with respect to the political and healthcare strategies that each country has followed during the pandemic.

# Keywords

Covid-19; SARS-CoV-2; pandemic; mathematical model; probability; dynamics; epidemic curves; case fatality ratio; lockdown;

# Index

# 1. INTRODUCTION

## 1.1. Motivation

### 1.1.1. Covid-19

At the end of December 2019, the World Health Organization (WHO) communicated that an unknown new virus was causing pneumonia-like illness in the city of Wuhan, China. It was found out that a coronavirus (CoV) was causing this disease. CoV is a type of virus that can cause from a common cold to more severe diseases like Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV) [1]. This novel CoV was named SARS-CoV-2, and the disease that it causes, Covid-19 [2].

The symptoms detected so far are highly variable since the virus affects people in very different ways. The most common symptoms are fever, dry cough and fatigue, although some people can be asymptomatic [3]. The WHO states that the majority of the infected people are developing mild to moderate symptoms, which enables a recovery without hospitalization. Around 1 out of 5 people who contract Covid-19 develop a severe illness and experience breathing difficulties. Elderly people and people suffering from high blood pressure, heart or lung affections, diabetes or cancer are more likely to develop severe conditions [3]. Furthermore, it has been demonstrated that the virus is highly contagious between human beings. The principal way of transmission occurs when an infected person expels small droplets when coughing, sneezing or speaking and another one inhales these droplets. However, if these droplets fall on objects or surfaces and someone else touches them, they might also get the infection. That is the reason why hygiene, face masks and an adequate safety distance between people have been and continue to be of vital importance for reducing the transmission of the virus. Several diagnostic techniques have been urgently developed both to detect the virus RNA and to identify the antibodies created by the immune system when interacting with it. The administered treatment is mainly symptomatic, since unfortunately, there is no current vaccine for SARS-CoV-2. Further research is crucial to adequately treat Covid-19.

Therefore, the aforementioned characteristics of Covid-19 have empowered the virus to spread rapidly worldwide, reaching all continents and affecting more than 17 million of the world population, having caused more than 675000 deaths by the end of July [4]. In fact, on 30th January 2020 the WHO declared Covid-19 a public health emergency of international concern [5].

### 1.1.2. Mathematical models in epidemiology

The Covid-19 pandemic has caused a brutal impact on society, forcing the closure of borders and a strict lockdown of many countries as well as the collapse of the healthcare system and severe damage to the world economy. This situation has led to a state of uncertainty, which is still present to date. Hence, mathematical models are needed in order to measure the impact of Covid-19, as well as to predict the future dynamics of the current pandemic. Mathematical models have been used in the past for planning the public health response to different infections [6] as well as to inform public health policy.

Most of the mathematical models found in literature try to describe the dynamics of the evolution of Covid-19 and its infection rate using fraction derivates like the Susceptible, Exposed, Infected and Recovered (SEIR) model [7] or more complex models like the SEIHRD model [8]. Other models use stochastic techniques to simulate the outbreak and to predict the future of the pandemic [9]. Moreover, additional mathematical models try to simulate the effect of control strategies such as the lockdown [10].

### 1.1.3. Mortality rate

Two distinct rates can be used to measure the risk of death caused by any virus: the infection fatality rate (IFR) and the case fatality rate (CFR). The first one measures the proportion of deaths among all infected people. On the other hand, CFR computes the proportion of individuals with fatal outcomes among the confirmed infected cases. At an early stage of the pandemic, all mortality ratios were calculated based on the total number of detected cases, given rise to a CFR between 0.1% and 25% depending on which country was studied [11].

However, computing these ratios is a complex procedure since different biases have to be considered. First of all, not enough people have been tested of Covid-19. During the course of the pandemic, each country has decided how to manage the diagnosis depending on the intensity of the outbreak. As a result, some countries have tested a large part of the population, such as South Korea, while others have chosen to only test the severe cases, Italy being one of them. Insufficient testing can lead to an overestimation of the CFR, since the number of confirmed cases is much lower than the real amount of existing cases, as asymptomatic individuals are not considered. Since we are faced with an ongoing pandemic, there is a delay between contracting the disease and death; people can be infected for a long time before becoming sick enough to be at risk of death. Regarding Covid-19, the time between the beginning of the infection and the consequent death has been estimated to be around 2 to 3 weeks or more [12]. In fact, people might die from the long-term effects caused by SARS-CoV-2 even if there is no longer an active infection. Therefore, by the time the CFR is calculated some individuals that might die in the future are not considered, leading to a lower CFR which does not truly reflect the situation.

## 1.2. Objectives

The aim of this project is to create a stochastic mathematical model to describe the ongoing dynamics of the Covid-19 pandemic and to individually estimate the probability of recovery or death for every infected patient. Initially, we develop a mathematical model based on the different probabilities that a patient has in terms of getting the infection or not, and of recovering or dying at a later stage if infected. We make use of the public domain data reported by several countries to calculate a more accurate CFR by reducing the bias caused because of the delay between illness and death. The available data and the computed CFR have been used to perform a parameter estimation to design a mathematical model similar to the real dynamics of the pandemic. The second part of the study aims to use the mathematical model to study the impact of the lockdown.

The study, which was started at the beginning of the outbreak, was initially conducted for eleven different countries: Andorra, China, France, Germany, Italy, Netherlands, Singapore, South Korea, Spain, United Kingdom and United States of America. However, during the course of the pandemic, some of them were discarded due to missing information or irregularities in the reporting system. Netherlands, Spain and United Kingdom were discarded for their lack of reported recovered cases. Furthermore, Spain, China and Andorra changed the reporting system during the outbreak resulting in data with many variations. Regarding USA, one of the major problems encountered when reporting the cases was that not all the states were following the established guidelines [13]. The number of recovery and death cases reported by France was lower in comparison to the total reported infected cases, leading to a greater number of active cases over time. Finally, Singapore was discarded due the low number of reported deaths. Thus, Germany, Italy and South Korea were the countries selected in order to apply the mathematical model developed in this study, using their available data.

# 2. METHODS

## 2.1.Data source

The data used for this project was obtained from the Johns Hopkins University Coronavirus Resource Center (CSSE) [14]. This database consists of 3 different Comma-Separated Values (CSV) format files; one for the confirmed cases, and two others for cases resulting in death or recovery. Each of them is provided with the following information: province or state, country or region, latitude, longitude, date and the cumulative sum of the cases. All this data was analyzed and represented graphically as epidemic curves by using MATLAB software.

Some irregularities were found in the available data. Thus, the Worldometer source [4], which provided a database similar to CSSE, was used to compare both datasets. Both databases showed similarities between them for each studied country. Therefore, if a significant strange behavior on the epidemic curves in the CSSE database was found, such as a suspicious high increase in cases reported in a single day, the value of total cases of the CSSE dataset was modified.

## 2.2. Mathematical formulation of the model

The proposed model was constructed based on stochastic procedures for studying the dynamics of the novel virus SARS-CoV-2 using MATLAB software.

The model is based on a fixed number of individuals ($n$) constructing a population for a given country ($c$). The time unit was set to one day ($d$) since the reported data has been provided on daily basis. The total number of individuals at any day ($d$), are symbolized as $N(d,c)$ and can be divided into four exclusive groups (Eq. 2.1): the individuals susceptible to be infected $S(d,c)$, the Covid-19 active infected people $I(d,c)$, the people who have recovered from the infection $R(d,c)$, and the people who have unfortunately died from Covid-19 $M(d,c)$, where:

$$N(d,c) = S(d,c) + I(d,c) + R(d,c) + M(d,c) \tag{2.1}$$

Initially, at $d_0 = 0$, all of the population is susceptible to get the infection, meaning that $N(d_0,c) = S(d_0,c)$. Then, during the interval $[d_{0+1}, d_{max}]$ any individual can go through the different states with a certain probability $p$. Figure 1 shows a decision tree that represents all the states and probabilities.

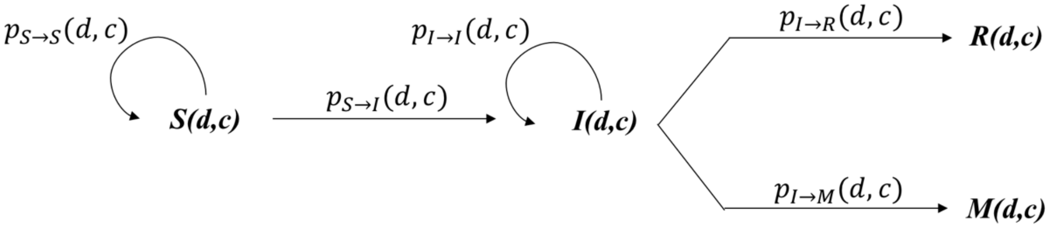

***Figure 1: Probability decision tree used to define the model.*** *$p_{S\to S}(d,c)$ is the probability of going from being susceptible to remaining susceptible; $p_{S\to I}(d,c)$ is the probability of going from susceptible to being infected; $p_{I\to I}(d,c)$ is the probability of going from being infected to remaining infected; $p_{I\to R}(d,c)$ is the probability of recovering from the infection; and $p_{I\to M}(d,c)$ is the probability of dying from the infection.*

Each day, the mathematical model examines at which state does every individual belongs to. Thus, if an individual is susceptible, she/he can stay susceptible or get the infection. If an individual is infected, she/he can stay infected, recover or, unfortunately, die.

Additionally, we assumed that $p_{S\to S}(d,c)$ and $p_{S\to I}(d,c)$ varied over time depending on the number of active cases $I(d,c)$. Thus, these probabilities can be described as follows:

$$\begin{aligned} p_{S\to S}(d,c) &= p_{0(S\to S)} - \alpha \cdot I(d-1,c) \\ p_{S\to I}(d,c) &= \ p_{0(S\to I)} + \alpha \cdot I(d-1,c) \end{aligned} \tag{2.2}$$

where $p_{0(S\to S)}$ is the initial probability of staying susceptible but not getting the infection, $p_{0(S\to I)}$ is the initial probability of getting the infection and $\alpha$ is the extent to which the number of active cases of the day before affects these probabilities.

Moreover, the time interval between when an individual gets infected and recovers or dies, needs also to be considered in the model. This variable is denoted as $(T)$ and affects the probabilities of Eq. 2.3:

$$
\begin{aligned}
p_{I\to I}(d,c) &= p_{0(I\to I)} - \beta \cdot T \\
p_{I\to R}(d,c) &= p_{0(I\to R)} + {}^{\beta}/_{2} \cdot T \\
p_{I\to M}(d,c) &= p_{0(I\to M)} + {}^{\beta}/_{2} \cdot T
\end{aligned}
\tag{2.3}
$$

where $p_{0(I\to I)}$ is the initial fixed probability of staying infected, $p_{0(I\to R)}$ is the initial fixed probability of recovering, $p_{0(I\to M)}$ is the initial fixed probability of dying, and $\beta$ is the magnitude in which the number of active infected days $T$ affects these probabilities. An individual that suffers from an active infection will finally recover or die. Thus, the parameter $\beta$ affects equally the probabilities $p_{I\to R}(d,c)$ and $p_{I\to M}(d,c)$.

To simplify the model, some assumptions were made. Firstly, we assumed that an individual can only get infected once so when recovered, the infected person becomes immune to the virus. However, how Covid-19 affects the immune system is still under study so further information is required. Secondly, we assumed that an infected person can infect others during all the course of the disease without taking into account any infectious period, such as incubation time gaps. Finally, we assumed that the first infected individual of the population is externally infected with a probability $p_{0(S\to I)}$.

Furthermore, since all values are probabilities, the following constrains were applied:

$$
\begin{gathered}
0 \leq p(d,c) \leq 1 \\
p_{S\to S}(d,c) + p_{S\to I}(d,c) = 1 \\
p_{I\to I}(d,c) + p_{I\to R}(d,c) + p_{I\to M}(d,c) = 1
\end{gathered}
\tag{2.4}
$$

Therefore, the model outputs are an estimation of $I(d,c)$, $R(d,c)$ and $M(d,c)$. Moreover, the model also returns an estimate of the days that pass between when an individual gets the infection and when subsequently recovers or dies from it.

The variables used for the mathematical model were found in literature and are shown in Table 1. The mathematical model was applied for three different countries: Germany, Italy and South Korea using data from the 22nd January 2020 to the 30th June 2020. A 0.5% of the total population of each country was established as the total individuals (*n*).

***Table 1: Variables used for the mathematical model.*** *The total number of individuals is considered as a 0.5% of the total population of each country.*

| Notation | Description | Value | Source |
|---|---|---|---|
| $d_{max}$ | Total days | 161 | Data source (see Section 2.1) |
| ***N(d,Germany)*** | Total individuals for Germany | 418500 | [15] |
| ***N(d,Italy)*** | Total individuals for Italy | 302000 | [15] |
| ***N(d,South Korea)*** | Total individuals for South Korea | 256000 | [15] |

## 2.3. CFR calculations

The CFR measures the severity of an infection, computed among all reported infected cases. During an ongoing pandemic, overestimation or underestimation of the CFR may occur due to delays of case resolution. Three different methods to calculate the CFR were studied and used to adjust the mathematical model.

First, the CFR can be calculated as follows:

$$L(d,c) = \frac{M(d,c)}{I(d,c)} \cdot 100 \tag{2.5}$$

However, results obtained with Eq. 2.5 provide an underestimation of the CFR since all active cases are considered. Since we are dealing with an ongoing pandemic, there is a delay from illness to death or recovery, so people can be infected for a long time before becoming sick enough to be at a risk of death.

One simple solution to mitigate this bias is to restrict the analysis to resolved cases:

$$U(d,c) = \frac{M(d,c)}{R(d,c) + M(d,c)} \cdot 100 \tag{2.6}$$

However, results obtained with Eq. 2.6 are overestimating the CFR, since sick people that die tend to be reported faster than infected people that will further recover. The existing bias both in Eq. 2.5 and 2.6 is caused by the no consideration of the active cases that will recover or die in the future.

Finally, we used the mathematical model outputs to mitigate this bias, by adding when active cases are going to die. Thus, the CFR can be calculated as:

$$CFR(d,c) = \frac{M(d,c) + E(d,c)}{I(d,c)} \cdot 100 \tag{2.7}$$

where $E(d,c)$ represents the active cases that will end up dying from the infection. Since the mathematical model estimates when an individual will recover or die, $E(d,c)$ was computed by examining the active infected cases $I(d,c)$ of a certain day that will further die. Therefore, this method to calculate the CFR uses information from the future to update the information of the present.

# 3. RESULTS

## 3.1. Model parameters estimation

Some of the parameters used in the model have been estimated based on published literature. However, despite the effort to use robust parameters, most of them have been manually estimated by running a large amount of simulations. In each simulation, a graphical comparison between epidemic curves was done in order to decide which were the parameters that required a better adjustment. The parameters estimated for Germany, Italy and South Korea are shown in Tables 2, 3 and 4.

***Table 2: Estimates of the parameters and initial values of the model for Germany.***

| Notation | Description | Value | Source |
|---|---|---|---|
| $\boldsymbol{p_{0(S\to S)}}$ | initial probability of staying susceptible | $1-4e^{-7}$ | Estimated |
| $\boldsymbol{p_{0(S\to I)}}$ | initial probability of becoming infected | $4e^{-7}$ | Estimated |
| $\boldsymbol{\alpha}$ | Magnitude in which active cases affect the probability $p_{0(S\to S)}$ and $p_{0(S\to I)}$ | $5.5e^{-7}$ (From day 1 to 70) | Estimated using [16] |
| | | $1.3e^{-7}$ (From day 71 to 147) | Estimated using [16] |
| | | $1.6e^{-7}$ (From day 148 to 161) | Estimated using [16] |
| $\boldsymbol{p_{0(I\to I)}}$ | initial fixed probability of staying infected | 0.9487 | Estimated using CFR |
| $\boldsymbol{p_{0(I\to R)}}$ | initial fixed probability of recovering | 0.0489 | Estimated using CFR |
| $\boldsymbol{p_{0(I\to M)}}$ | initial fixed probability of dying | 0.0024 | Estimated using CFR |
| $\boldsymbol{\beta}$ | Magnitude in which the days a patient has been infected affect the probabilities $p_{0(I\to I)}$, $p_{0(I\to R)}$ and $p_{0(I\to M)}$ | 0.000026 | Estimated using [20] and [21] |

***Table 3: Estimates of the parameters and initial values of the model for Italy.***

| Notation | Description | Value | Source |
|---|---|---|---|
| $\boldsymbol{p_{0(S\to S)}}$ | initial probability of staying susceptible | $1-8.5e^{-7}$ | Estimated |
| $\boldsymbol{p_{0(S\to I)}}$ | initial probability of becoming infected | $8.5e^{-7}$ | Estimated |
| $\boldsymbol{\alpha}$ | Magnitude in which active cases affect the probability $p_{0(S\to S)}$ and $p_{0(S\to I)}$ | $8.47e^{-7}$ (From day 1 to 56) | Estimated using [17] |
| | | $3.6e^{-7}$ (From day 57 to 161) | Estimated using [17] |
| $\boldsymbol{p_{0(I\to I)}}$ | initial fixed probability of staying infected | 0.9487 | Estimated using CFR |
| $\boldsymbol{p_{0(I\to R)}}$ | initial fixed probability of recovering | 0.0435 | Estimated using CFR |
| $\boldsymbol{p_{0(I\to M)}}$ | initial fixed probability of dying | 0.0078 | Estimated using CFR |
| $\boldsymbol{\beta}$ | Magnitude in which the days a patient has been infected affect the probabilities $p_{0(I\to I)}$, $p_{0(I\to R)}$ and $p_{0(I\to M)}$ | 0.000026 | Estimated using [20] and [21] |

***Table 4: Estimates of the parameters and initial values of the model for South Korea.***

| Notation | Description | Value | Source |
|---|---|---|---|
| $\boldsymbol{p_{0(S\to S)}}$ | initial probability of staying susceptible | $1-1.1e^{-6}$ | Estimated |
| $\boldsymbol{p_{0(S\to I)}}$ | initial probability of becoming infected | $1.1e^{-6}$ | Estimated |
| $\boldsymbol{\alpha}$ | Magnitude in which active cases affect the probability $p_{0(S\to S)}$ and $p_{0(S\to I)}$ | $9.0e^{-7}$ (From day 1 to 70) | Estimated using [19] |
| | | $1.0e^{-7}$ (From day 46 to 126) | Estimated using [19] |
| | | $2.0e^{-7}$ (From day 127 to 161) | Estimated using [19] |
| $\boldsymbol{p_{0(I\to I)}}$ | initial fixed probability of staying infected | 0.9487 | Estimated using CFR |
| $\boldsymbol{p_{0(I\to R)}}$ | initial fixed probability of recovering | 0.05035 | Estimated using CFR |
| $\boldsymbol{p_{0(I\to M)}}$ | initial fixed probability of dying | 0.00095 | Estimated using CFR |
| $\boldsymbol{\beta}$ | Magnitude in which the days a patient has been infected affect the probabilities $p_{0(I\to I)}$, $p_{0(I\to R)}$ and $p_{0(I\to M)}$ | 0.000026 | Estimated using [20] and [21] |

The initial probabilities of staying susceptible $p_{0(S\to S)}$ and getting infected $p_{0(S\to I)}$ were estimated based on when the pandemic started. These probabilities mark the first cases of infected people, that took place when the virus started spreading. Therefore, very low parameters were selected in order to simulate the first detected cases. The probability that corresponds to South Korea is the highest, since it was the country where the virus arrived before, followed by Italy and then Germany. These parameters are only useful to initialize the outbreak, then $p_{S\to S}(d,c)$ and $p_{S\to I}(d,c)$ are governed by the other part of the equation (see Eq. 2.2).

Furthermore, in order to reproduce the control measures that each government has been dictating to limit the spread of the disease, different $\alpha$ values were adjusted for diverse periods of time, representing social distancing over time. Control strategies do not have an immediate impact on the dynamics of the pandemic. Therefore, the delay between the application of these strategies and their effect has been estimated, so $\alpha$ values decreased during this time interval.

Regarding Germany, schools and nurseries closed on 13th March. Two days later, the borders with five neighboring countries closed. Finally, on 22nd March, a lockdown and national curfew were established by the government [16]. Nevertheless, we estimated the effect of the lockdown 8 days later, on 1st April. On the other hand, the effect of the un-lockdown in Germany has been estimated on 17th June since the number of new daily infections started to increase again.

In the case of Italy, the government initially established some restrictions in 11 municipalities in late February. On 8th March, restrictions were also applied for the whole region of Lombardy and 14 other provinces. One day later, a national lockdown was imposed, restricting the movement of the population and closing all non-essential businesses [17]. However, we estimated the effect of the lockdown 8 days later, on 17th March. The progressive un-lockdown carried out in Italy has had no effect since the number of new infections did not increase, at least until 30th June.

In contrast to Germany and Italy, South Korea decided to start fighting the virus by doing a massive testing of the population. As the outbreak increased, approximately 600 test centers were established in order to test patients outside the health care system. They did from 15,000 to 20,000 tests per day [18]. This early detection was followed by an extensive tracking and tracing of the South Korean population; infected patients were isolated and places visited by many people, such as schools and churches, were closed. Nonetheless, a national lockdown, as imposed in the other countries, was never established [19]. Therefore, we estimated that the day these effective control strategies caused a positive effect in the reduction of Covid-19 cases was on 3rd March. Despite the efforts of the government and the population to fight the virus, the number of new daily infected people began to rise again on 27th May.

The initial probabilities of staying infected $p_{0(I \to I)}$, recovering $p_{0(I \to R)}$ or dying $p_{0(I \to M)}$ were set constant over time. The probability of dying in Italy was set higher than in Germany and South Korea since more people have died from Covid-19 in this country.

Moreover, these probabilities were adjusted using the CFR as input information of every country. The CFR was calculated in three different ways: *L(d,c)*, *U(d,c)* and *CFR(d,c)* (see Section 2.3). The results of CFR for Germany, Italy and South Korea are shown in Figure 2, 3 and 4. The results obtained for *L(d,c)* and *U(d,c)* follow the same dynamics in all countries, where *L(d,c)* increases and *U(d,c)* decreases over time. At the end of the outbreak of the pandemic, both curves converge since all active infected patients end up either recovering or dying from the disease. The initial probabilities were adjusted until the CFR calculated form the model also converged with the other two curves.

It is fundamental to emphasize that the CFR calculated with the outputs of the model (magenta curves in Figures 2, 3 and 4) begins to increase before the other two curves and finally remains constant over time. The reason for this might be because the active infected cases that will die in the future are considered on those given days. However, these CFR values slightly diminish at the end of June because as the data stops being recorded, there are active cases with an active infection that have not been determined whether they will recover or die.

Finally, the parameter $\beta$ was adjusted so that the time interval between getting the infection to recovering was an average of 2 weeks [20] and the time gap between getting the infection and dying from it was between 2 and 3 weeks (approximately 17 days) [21].

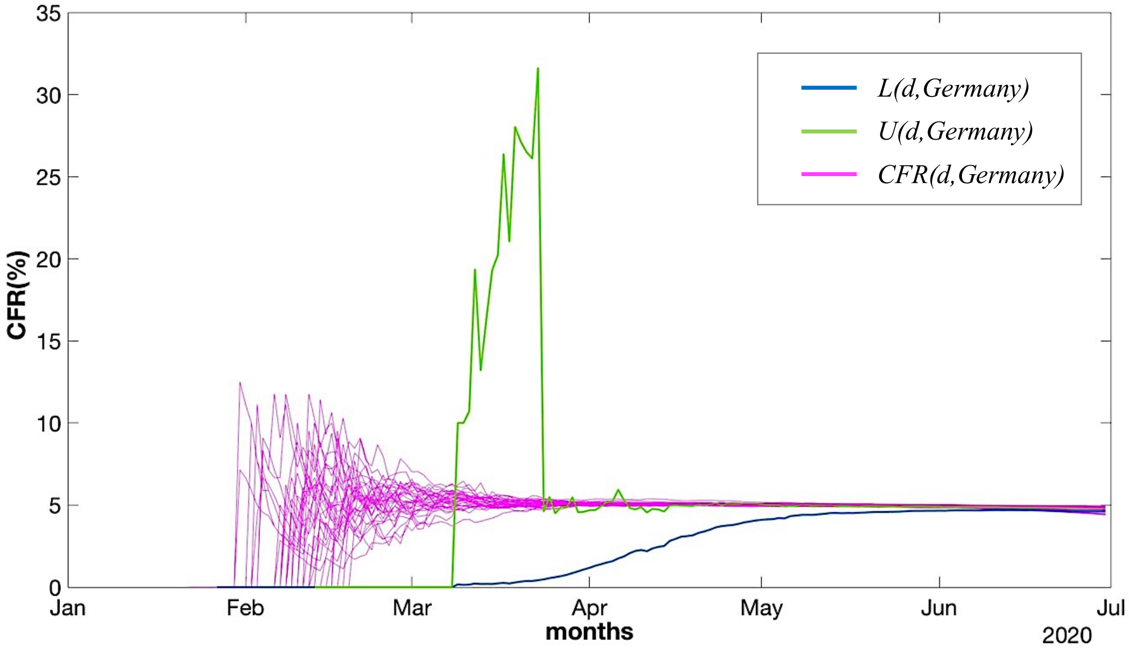

***Figure 2: Different estimations of Germany's CFR.*** *The CFR estimated with 30 simulations of the model shows values from 4.410 % to 4.893% on 30th June, while L(d,c) reaches values of 4.6% and U(d,c) reaches values of 4.805%.*

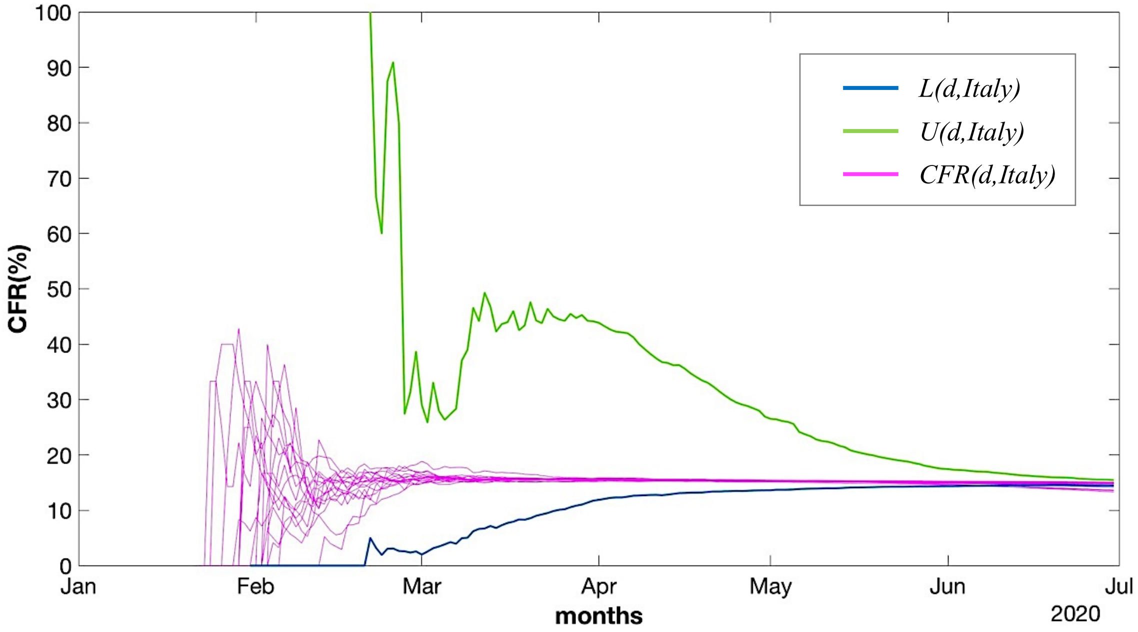

***Figure 3: Different estimations of Italy's CFR.*** *The CFR estimated with 30 simulations of the model shows values from 14.140% to 15.100% on 30*$^{th}$ *June, while L(d,c) reaches values of 14.45% and U(d,c) reaches values of 15.45%.*

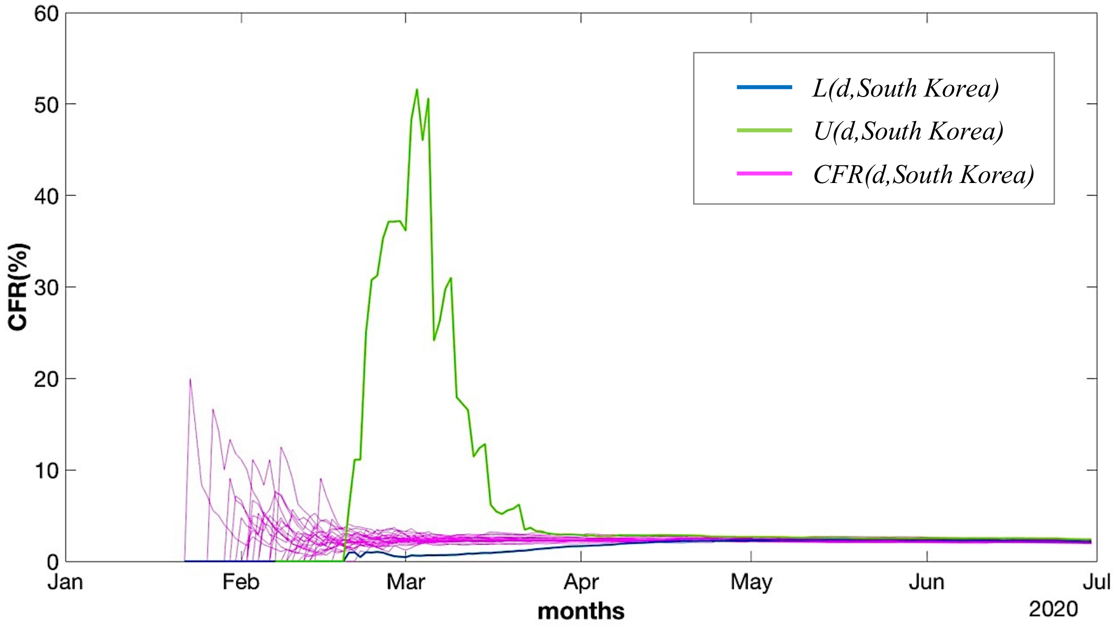

***Figure 4: Different estimations of South Korea's CFR.*** *The CFR estimated with 30 simulations of the model shows values from 1.986% to 2.448% on 30*$^{th}$ *June, while L(d,c) reaches values of 2.195% and U(d,c) reaches values of 2.371%.*

### 3.2. Numerical Simulations

In this section, results of numerical experiments (Figures 5 to 13) are shown to illustrate the efficiency and robustness of the developed mathematical model, as well as to compare the results for the three different countries, Germany, Italy and South Korea. A total of 30 simulations were carried out for each country by using the parameters explained in Section 3.1.

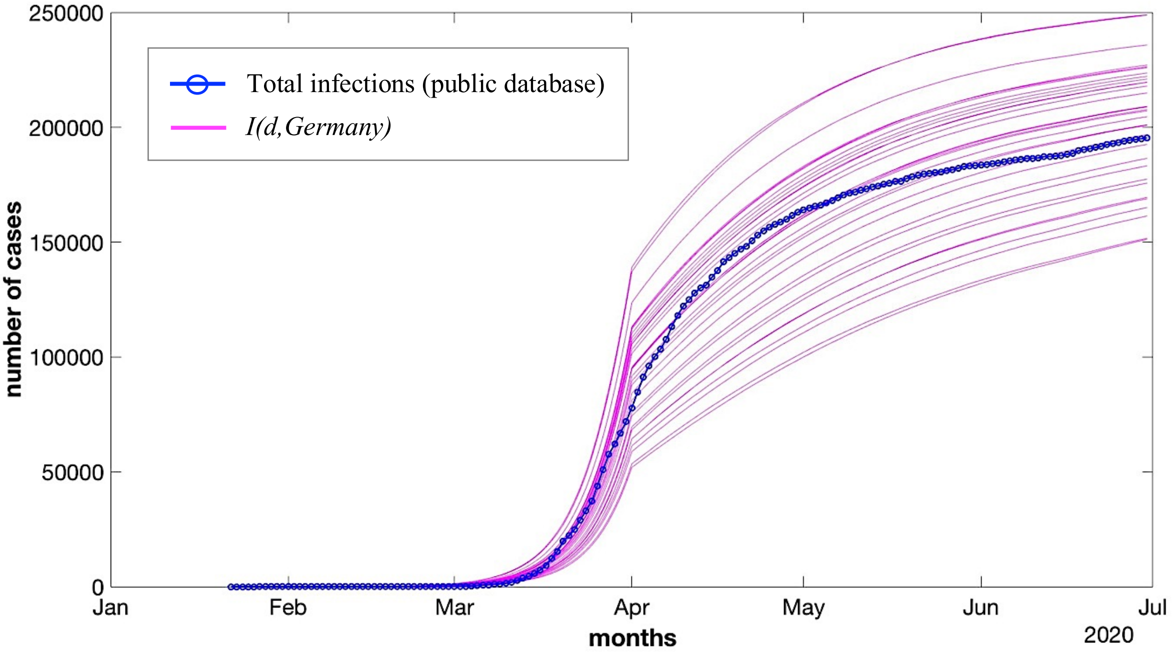

***Figure 5: Comparison between the curve of total confirmed cases obtained from the data and the ones from simulations obtained using the model for Germany.*** *Magenta curves represent the 30 simulations obtained from the model.*

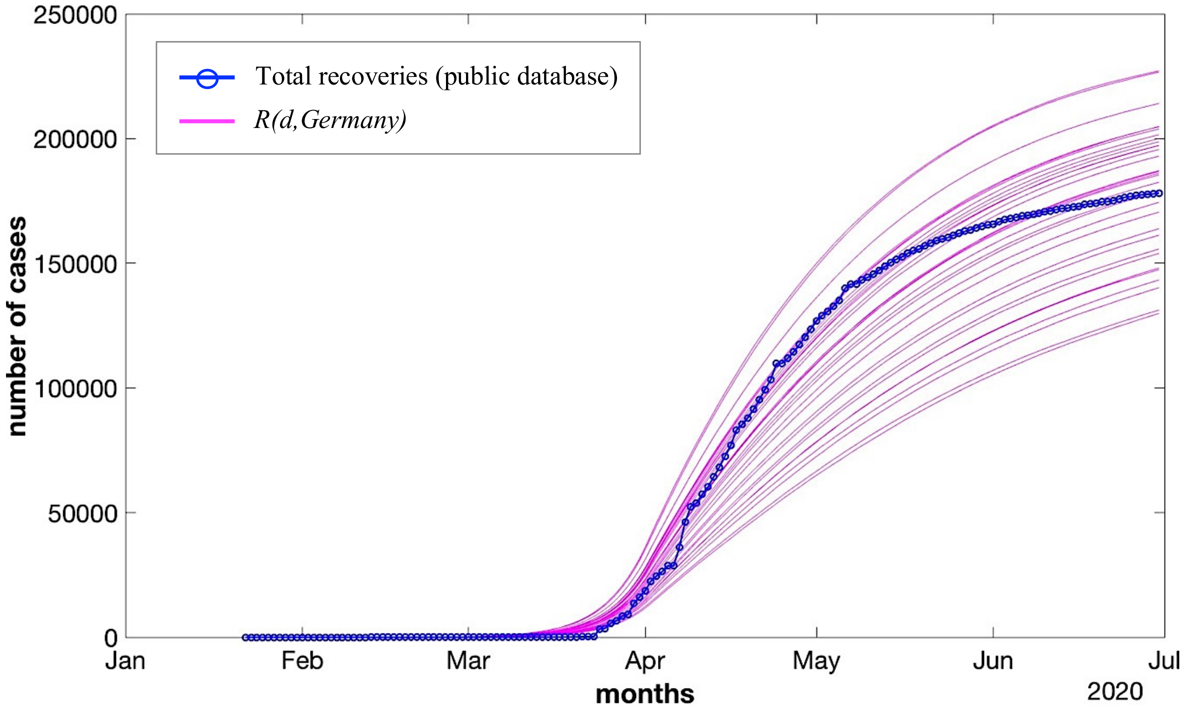

***Figure 6: Comparison between the curve of total recovered cases obtained from the data and the ones from the simulations using the model for Germany.*** *Magenta curves represent the 30 simulations obtained from the model.*

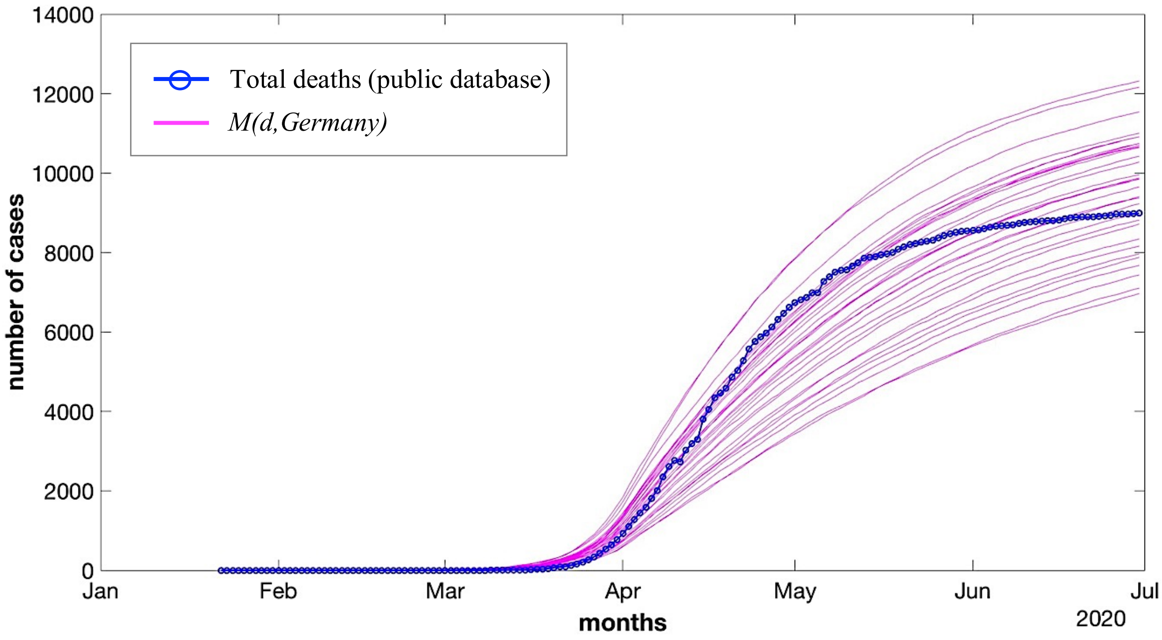

***Figure 7: Comparison between the curve of total death cases obtained from the data and the ones from the simulations using the model for Germany.*** *Magenta curves represent the 30 simulations obtained from the model.*

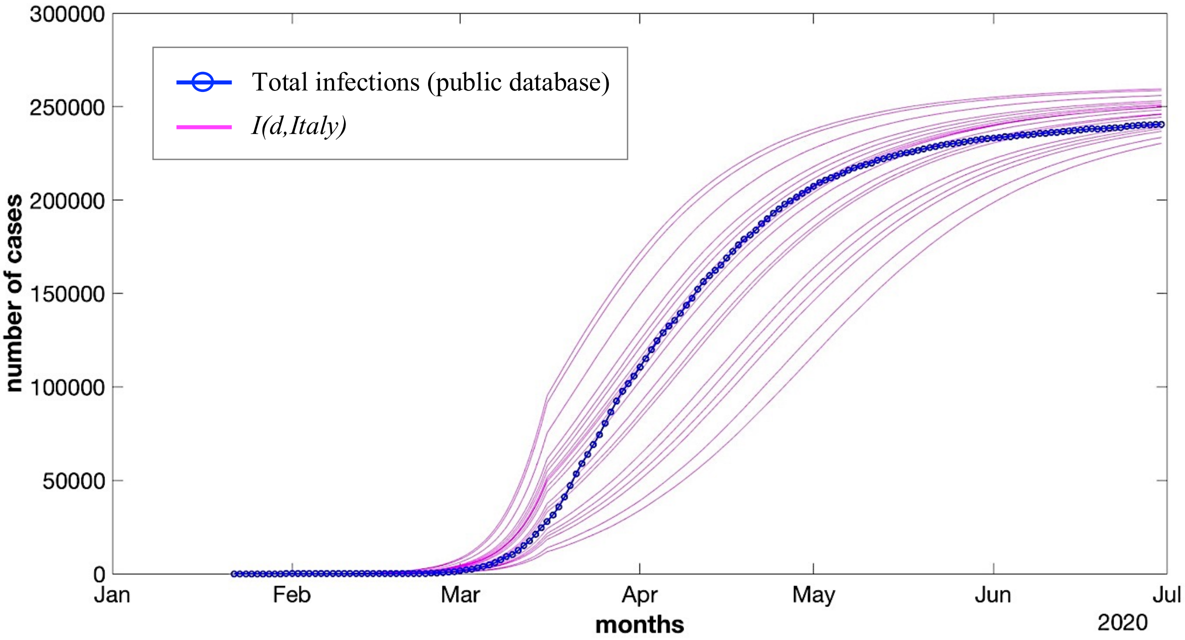

***Figure 8: Comparison between the curve of total confirmed cases obtained from the data and the ones from the simulations using the model for Italy.** Magenta curves represent the 30 simulations obtained from the model.*

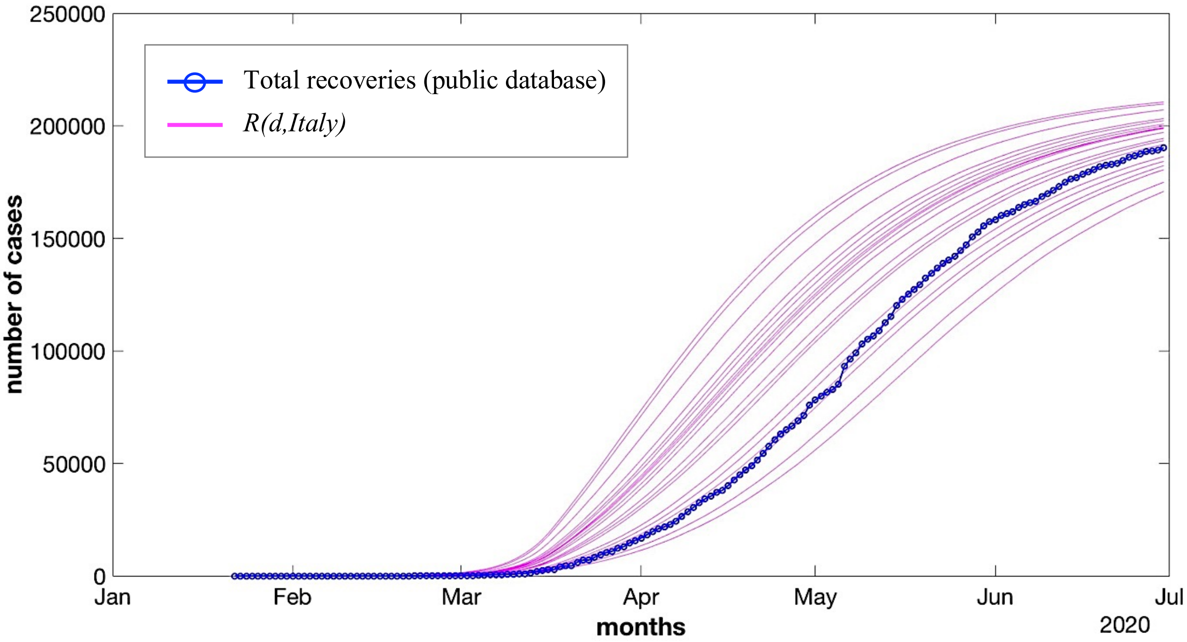

***Figure 9: Comparison between the curve of total recovered cases obtained from the data and the ones from the simulations using the model for Italy.** Magenta curves represent the 30 simulations obtained from the model.*

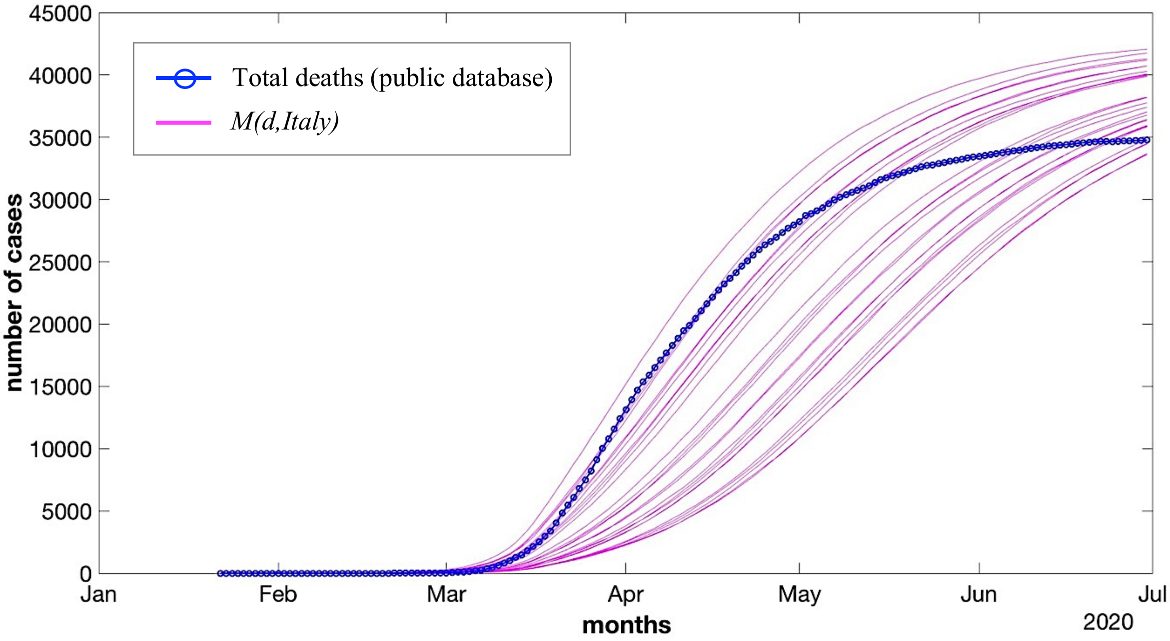

***Figure 10: Comparison between the curve of total deaths obtained from the data and the ones from simulations using the model for Italy.*** *Magenta curves represent the 30 simulations obtained from the model.*

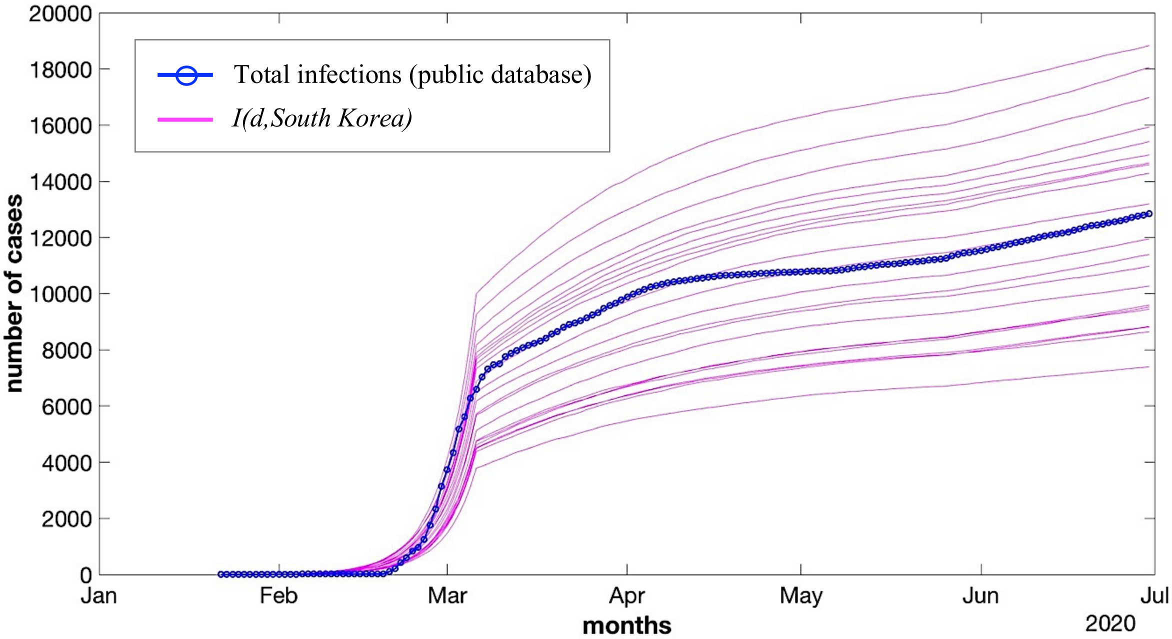

***Figure 11: Comparison between the curve of total confirmed cases obtained from the data and the ones from the simulations using the model for South Korea.*** *Magenta curves represent the 30 simulations obtained from the model.*

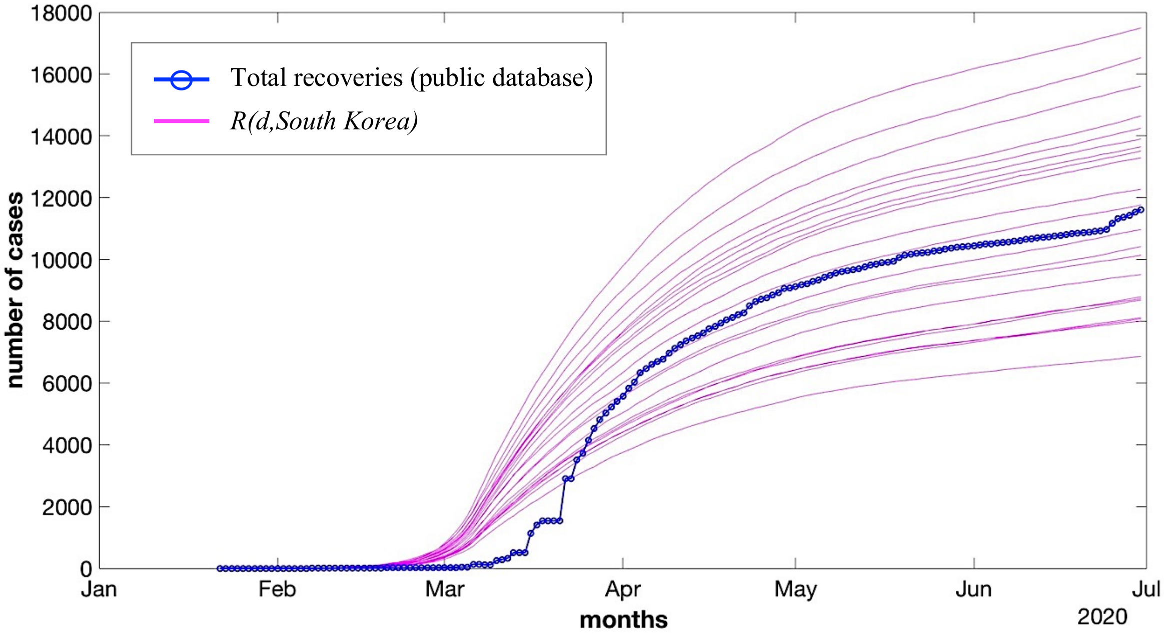

***Figure 12: Comparison between the curve of total recovered cases obtained from the data and the ones from the simulations using the model for South Korea.*** *Magenta curves represent the 30 simulations obtained from the model.*

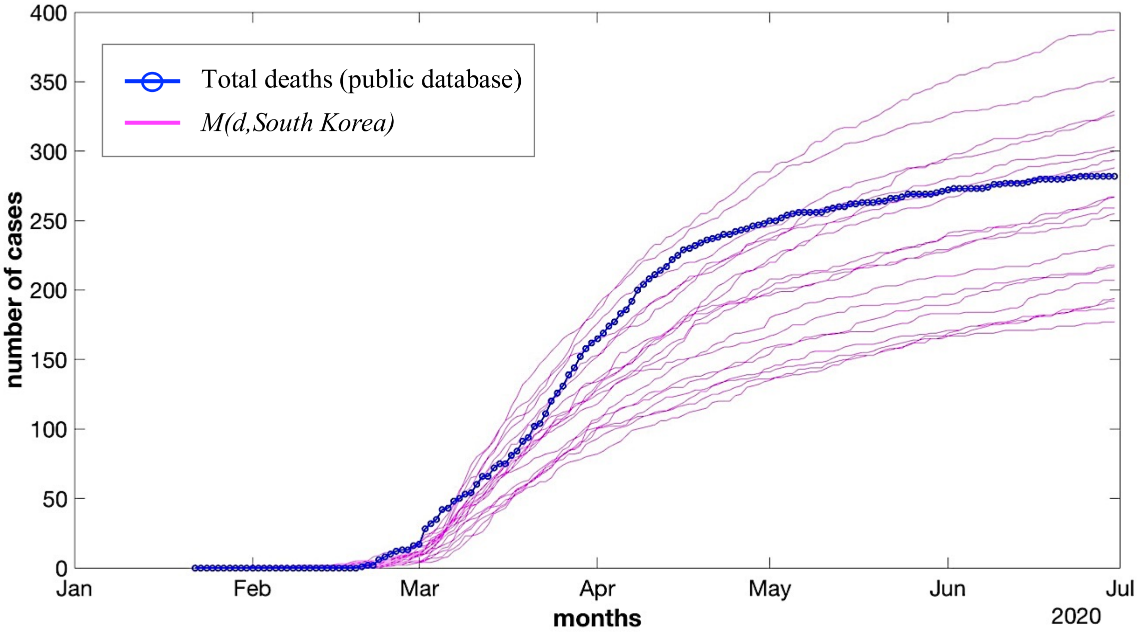

***Figure 13: Comparison between the curve of total deaths obtained from the data and the ones from the simulations using the model for South Korea.*** *Magenta curves represent the 30 simulations obtained from the model.*

Results show that almost all reported data is included in between the stochastic simulation as seen graphically in Figures 5-13. However, the cumulative death cases curve for the different countries, especially for Germany and Italy, using the public database, is quite different from the one obtained with the model. Regarding the robustness of the model, it has been seen that the simulations obtained for South Korea and Italy are more robust than the ones for Germany, which led to more dissimilar simulations. In fact, the difference between the lowest and highest simulation is approximatively of 90,000 cases in Germany, 20,000 in Italy and 10,000 in South Korea. Despite the satisfactory robustness of the model of South Korea, the dynamics shown in Figures 11-13 may seem widespread since the total number of cases is significantly lower than the ones reported by Germany and Italy.

## 3.3. Effect of changing model parameters

### 3.3.1. Change on initial probabilities

In this section, the effect of changing the initial probabilities $p_{0(S\to S)}$ and $p_{0(S\to I)}$ is analyzed to determine how robust is the model when a perturbation is introduced in one of the estimated parameters. Since $p_{0(S\to S)} + p_{0(S\to I)} = 1$, we decided to decrease $p_{0(S\to S)}$ and increase $p_{0(S\to I)}$, and the other way round. Therefore, the modification of the probabilities caused an advance or delay of the beginning of the outbreak. Ten simulations were carried out for each parameter modification, shown in Figures 14, 15 and 16.

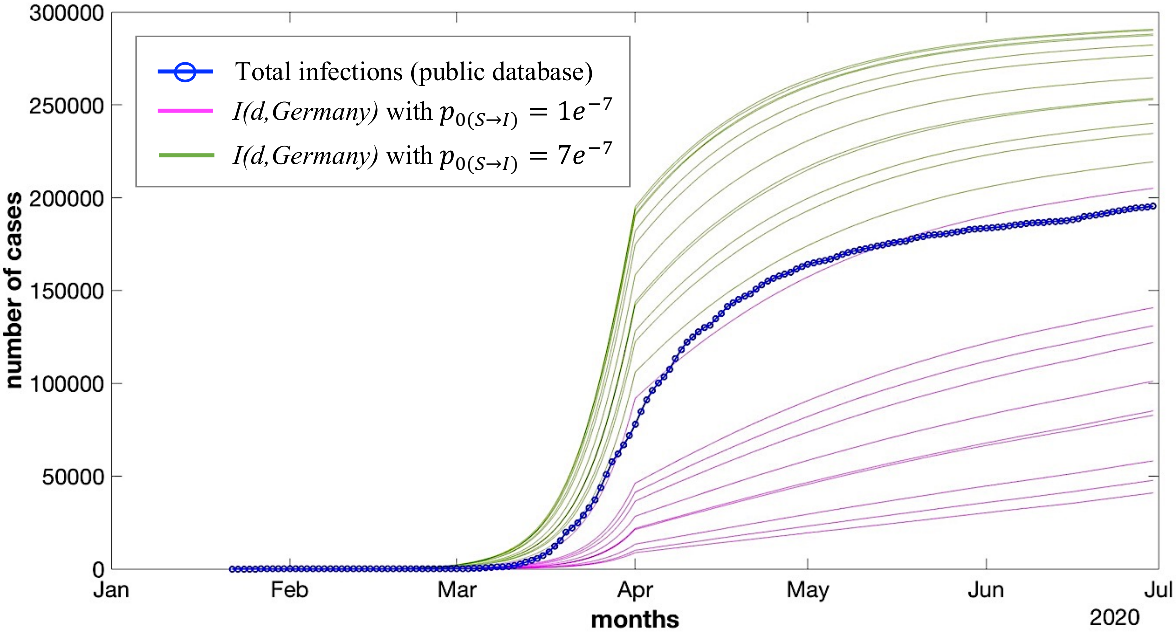

***Figure 14: Effect of changing $p_{0(S\to S)}$ and $p_{0(S\to I)}$ in Germany.*** *The estimated probability $p_{0(S\to I)}$ to approximate the beginning of the outbreak was set to 4e$^{-7}$. Results show 10 simulations obtained from increasing and decreasing this probability by 3e$^{-7}$.*

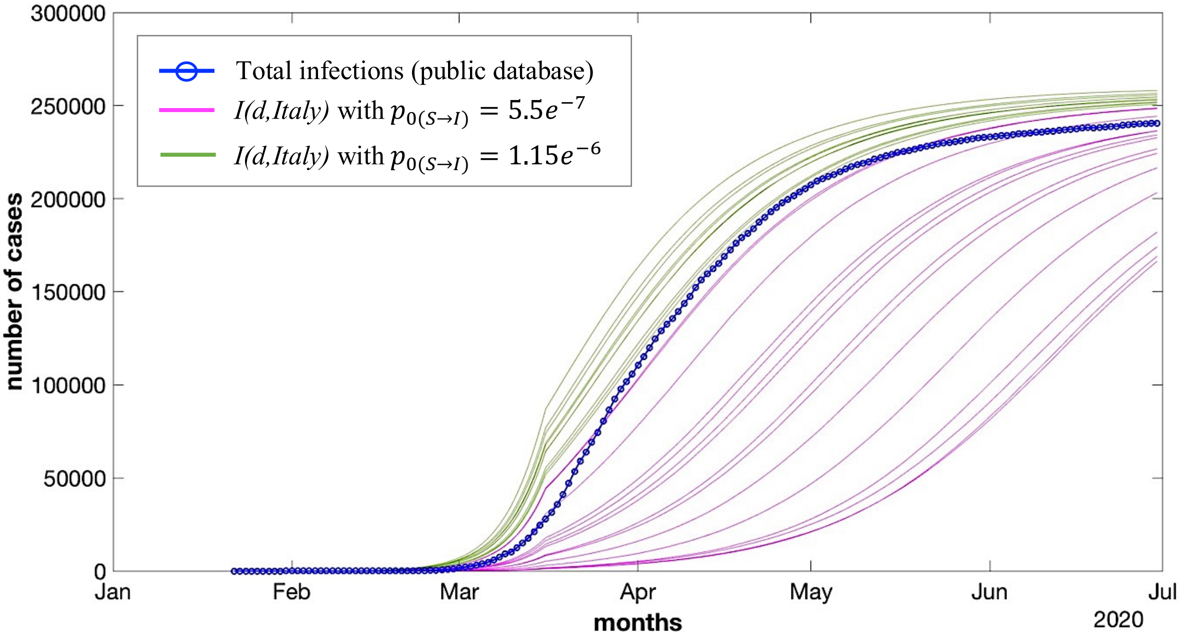

***Figure 15: Effect of changing $\mathbf{p_{0(S\to S)}}$ and $\mathbf{p_{0(S\to I)}}$ in Italy.*** *The estimated probability $p_{0(S\to I)}$ to approximate the beginning of the outbreak was set to $8.5e^{-7}$. Results show 10 simulations obtained from increasing and decreasing this probability by $3e^{-7}$.*

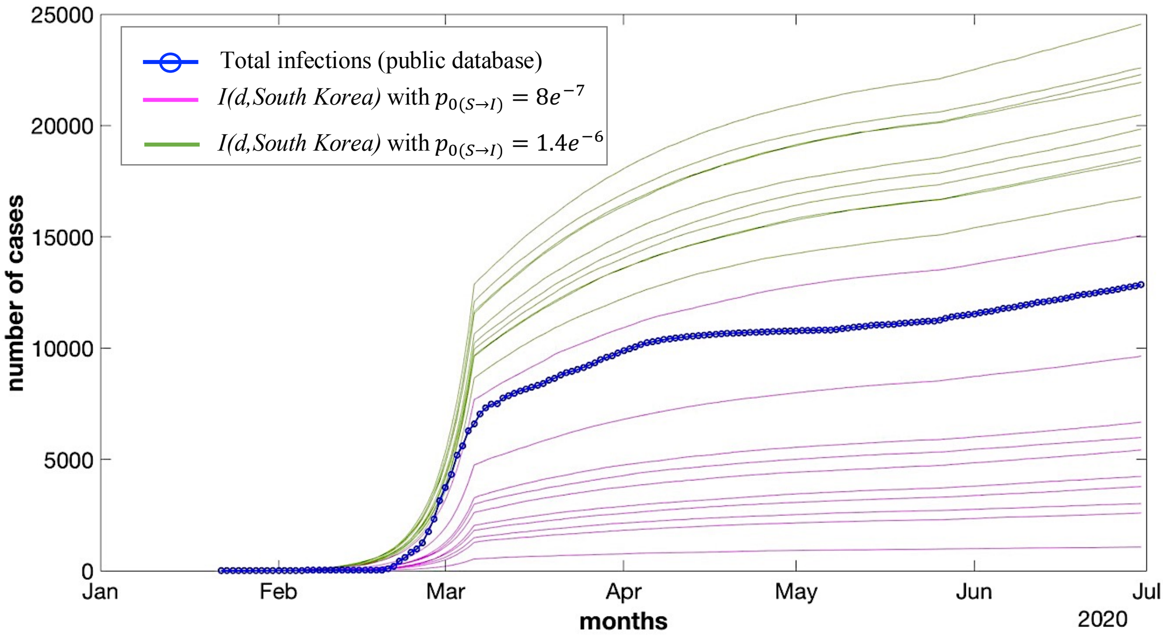

***Figure 16: Effect of changing $\mathbf{p_{0(S\to S)}}$ and $\mathbf{p_{0(S\to I)}}$ in South Korea.*** *The estimated probability $p_{0(S\to I)}$ to approximate the beginning of the outbreak was set to $1.1e^{-6}$. Results show 10 simulations obtained from increasing and decreasing this probability by $3e^{-7}$.*

Results show that by increasing the initial probability of getting the infection $p_{0(S\rightarrow S)}$ the outbreak starts before it really has. On the other hand, by reducing this probability, the outbreak starts afterwards. Furthermore, if $p_{0(S\rightarrow S)}$ increases, the model shows greater robustness, meaning that the range between the two more extreme simulation curves is reduced. On the other hand, if this probability decreases, the different simulations differ more between them. The aforementioned perturbations explain the difference in the robustness of the model between Germany, and Italy or South Korea. Since the Covid-19 outbreak started later in Germany, the probability of someone getting the infection is lower and therefore, the system is less robust.

Furthermore, the CFR was obtained for these different simulations to determine how changing probabilities $p_{0(S\rightarrow S)}$ and $p_{0(S\rightarrow I)}$ affects this ratio. Results are shown in Figures 17, 18 and 19. Decreasing $p_{0(S\rightarrow S)}$ makes the outbreak start later in time. Therefore, individuals begin to die later than before. Thus, on 30th June there are many active infected cases, which make the CFR decrease, as seen at the end of the CFR simulations obtained with the model, especially in the case of Italy, where the proportion of deaths is higher than in the other countries.

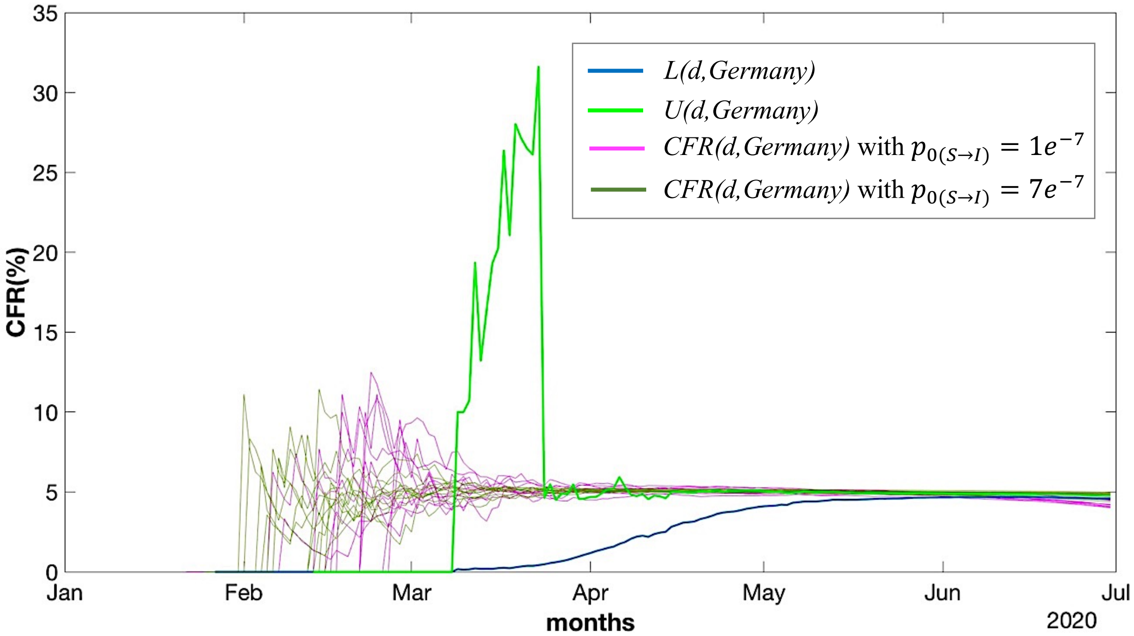

***Figure 17: Effect of changing $\mathbf{p}_{0(S\rightarrow S)}$ and $\mathbf{p}_{0(S\rightarrow I)}$ in Germany's CFR.*** *Results show 10 CFR simulations obtained from increasing $p_{0(S\rightarrow I)}$ by $3e^{-7}$ and 10 CFR simulations obtained from decreasing $p_{0(S\rightarrow I)}$ by $3e^{-7}$.*

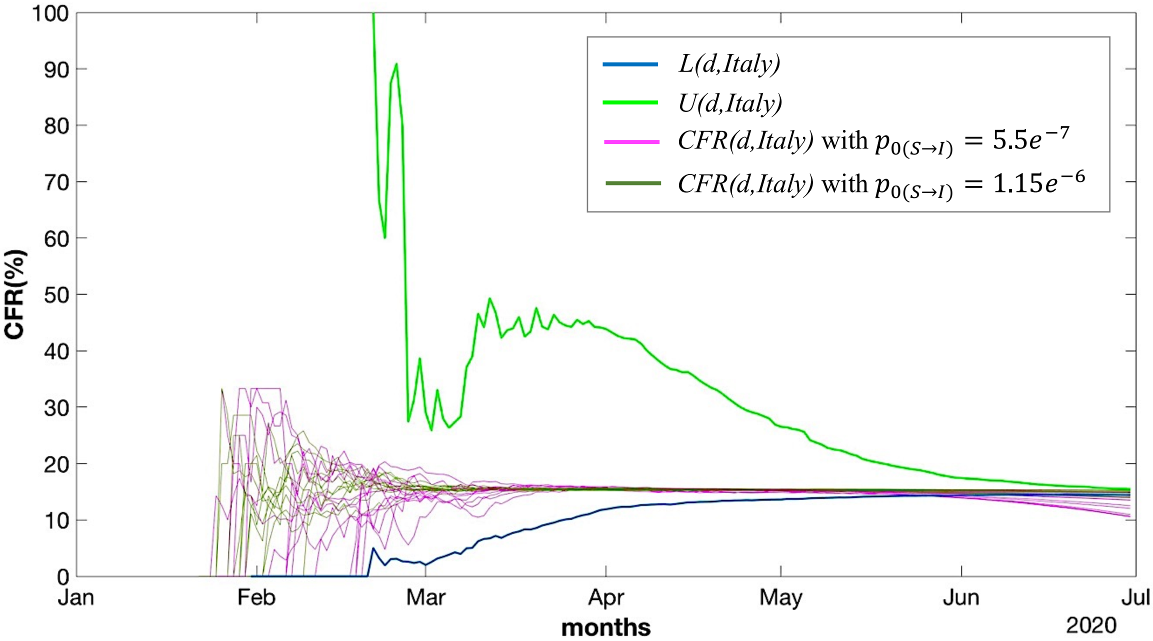

***Figure 18: Effect of changing $p_{0(S\to S)}$ and $p_{0(S\to I)}$ in Italy's CFR.*** *Results show 10 CFR simulations obtained from increasing $p_{0(S\to I)}$ by $3e^{-7}$ and 10 CFR simulations obtained from decreasing $p_{0(S\to I)}$ by $3e^{-7}$.*

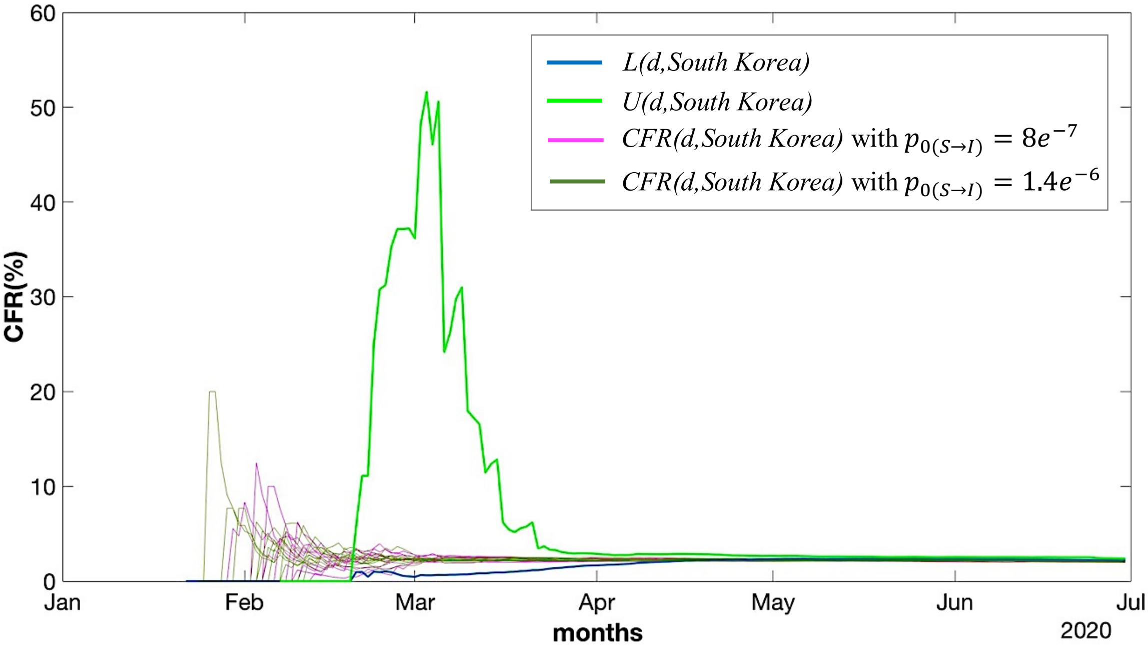

***Figure 19: Effect of changing $p_{0(S\to S)}$ and $p_{0(S\to I)}$ in South Korea's CFR****. Results show 10 CFR simulations obtained from increasing $p_{0(S\to I)}$ by $3e^{-7}$ and 10 CFR simulations obtained from decreasing $p_{0(S\to I)}$ by $3e^{-7}$.*

### 3.3.2. Different lockdown scenarios

Another useful application of the model is to assess the potential impact of the control strategies. Therefore, this section presents the effect that the modification of the start date of the lockdown has on the total number of infected cases.

We assumed that the probabilities and parameters of the mathematical model remain the same, that the epidemic curves follow the same dynamics as represented in previous sections, and that the lockdown also has an effect 8 days after commencing.

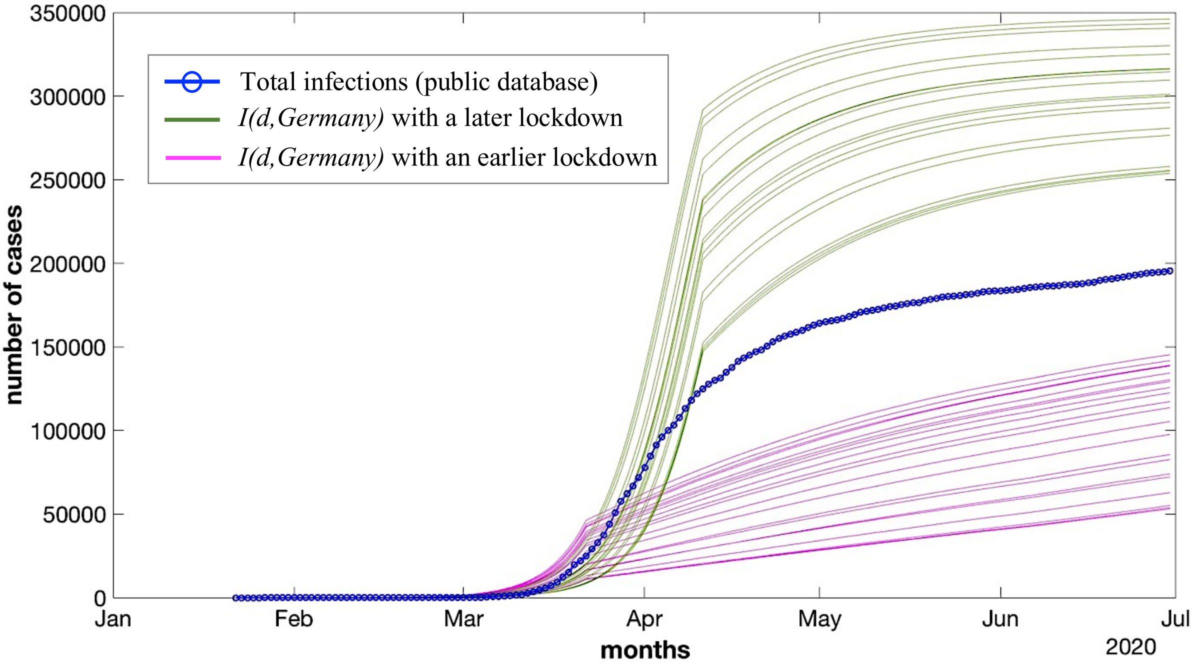

***Figure 20: Comparison of the cumulative infected cases with different lockdown scenarios using the Germany's model.*** *Green curves are 20 simulations made by assuming the lockdown would have been 10 days after and magenta curves represent 20 simulations made by assuming a 10 days earlier lockdown.*

If Germany had confined the country 10 days earlier, the number of predicted infected cases would have decreased around 25% and 72%. On the other hand, if it had taken 10 more days to confine the country, results show that cases would have increased from 29% to 77%.

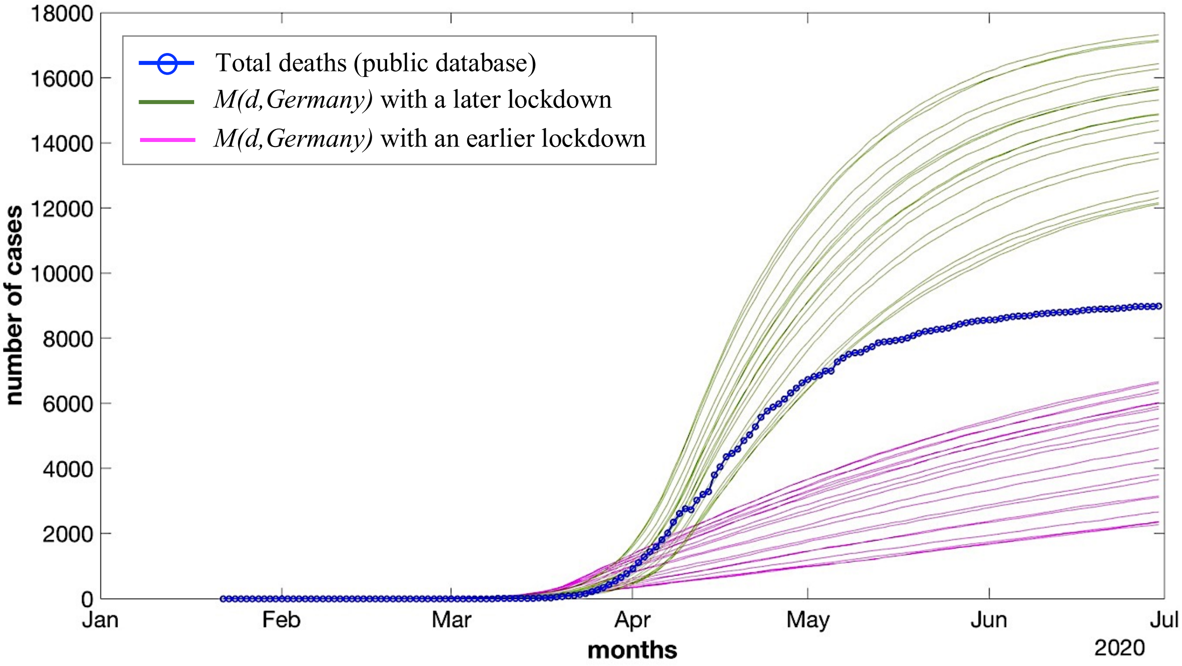

***Figure 21: Comparison of the cumulative death cases with different lockdown scenarios using Germany's model.*** *Green curves are 20 simulations made by assuming the lockdown would have been 10 days after and magenta curves represent 20 simulations made by assuming a 10 days earlier lockdown.*

Consequently, the number of predicted deaths would have decreased from 25% to 74% if the lockdown had started 10 days earlier, and would have increased from approximatively 34% to 85% if the lockdown had started later, as seen in Figure 21. These variations in the number of predicted deaths and total cases are due to the randomness of the model.

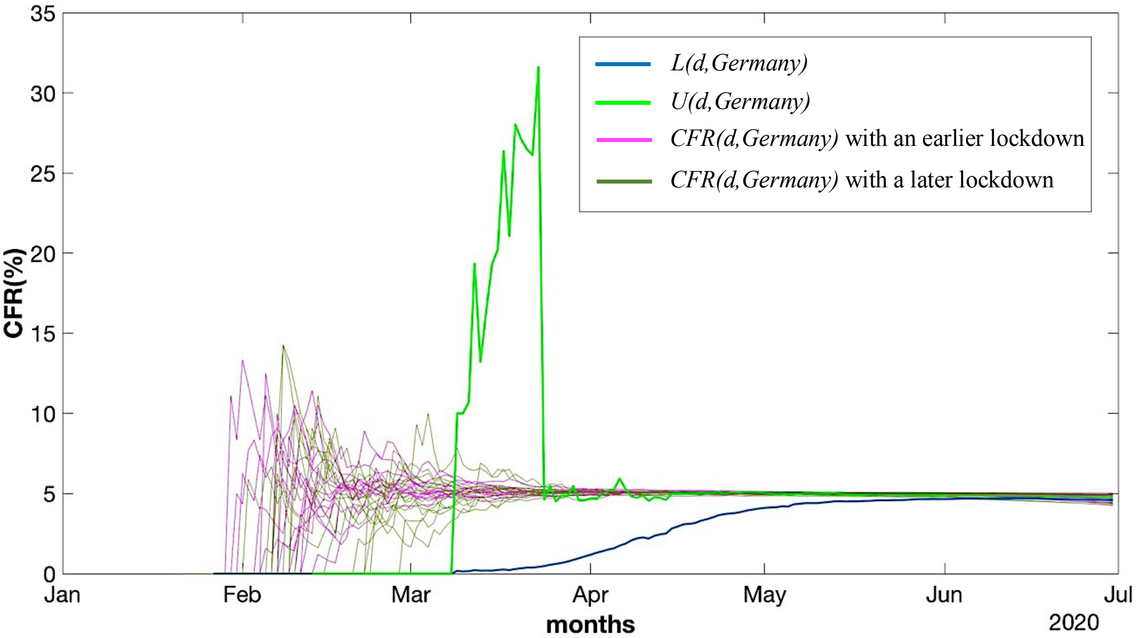

***Figure 22: Comparison of the CFR with different lockdown scenarios.***

Figure 22 shows that the CFR obtained considering the different lockdown scenarios does not vary since the probability of dying remains the same. Since we fixed the probabilities of dying or staying infected, the proportion of deaths over total infected cases remains the same despite variations in the lockdown starting date. Thus, when computing the CFR, since the total of infected cases and the number of death cases increase or decrease in the same proportion, the ratio does not change.

Results were also obtained for Italy since the country also established a national lockdown.

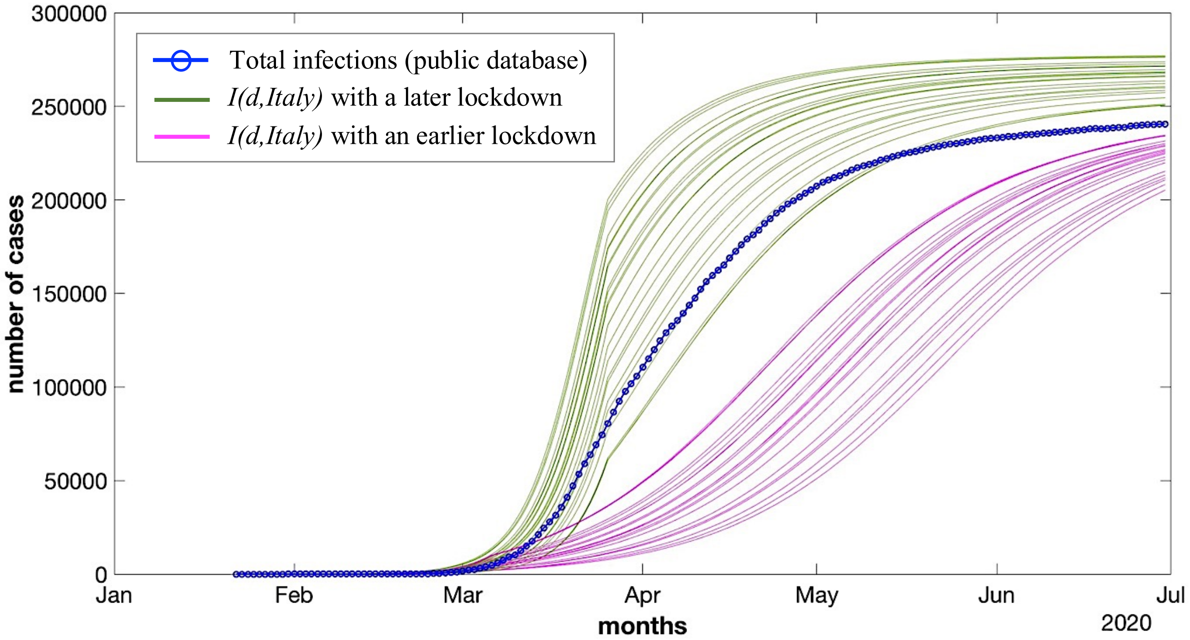

***Figure 23: Comparison of the cumulative infected cases with different lockdown scenarios, using Italy's model.*** *Green curves are 20 simulations made by assuming the lockdown would have been 10 days after and magenta curves represent 20 simulations made by assuming a 10 days earlier lockdown.*

If Italy had confined the country 10 days earlier, the number of predicted infected cases would have decreased approximatively from 2% to 15%. On the other hand, if the lockdown had started 10 days later, the model predicts that cases would have increased from 4% to 16%.

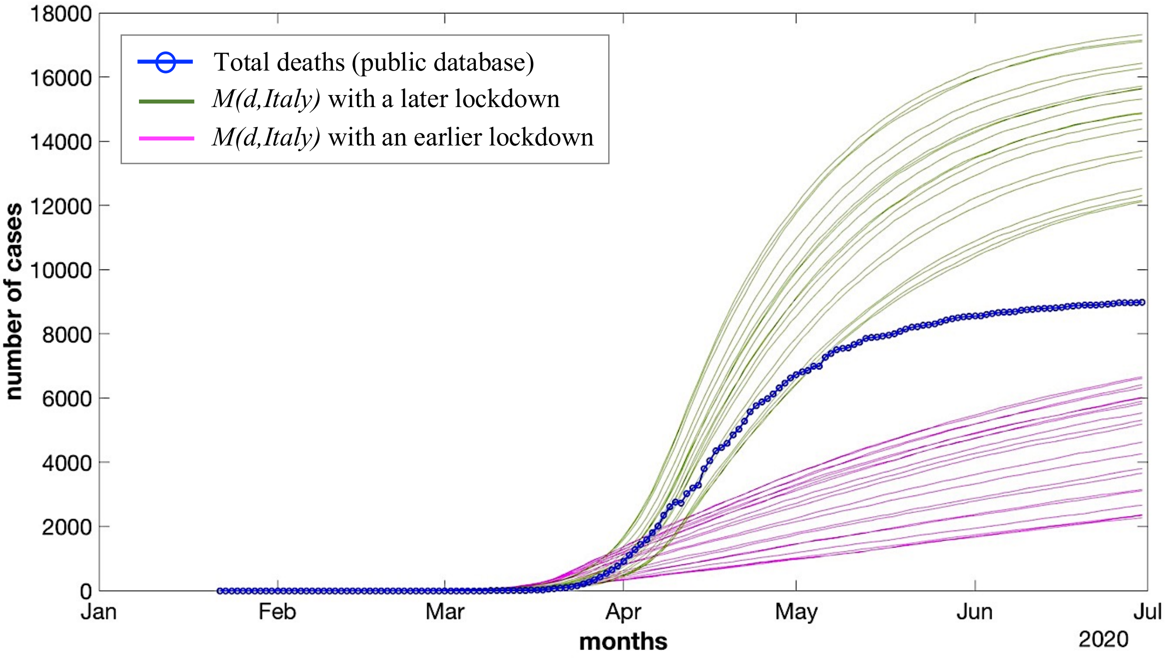

***Figure 24: Comparison of the cumulative death cases with different lockdown scenarios, using Italy's model.*** *Green curves are 20 simulations made by assuming the lockdown would have been 10 days after and magenta curves represent 20 simulations made by assuming a 10 days earlier lockdown.*

Consequently, the number of predicted deaths would have decreased from 3% to 20% provided a 10 days earlier lockdown, and would have increased from 6% to 20% if the lockdown had started 10 days later, leading to a similar CFR in both scenarios.

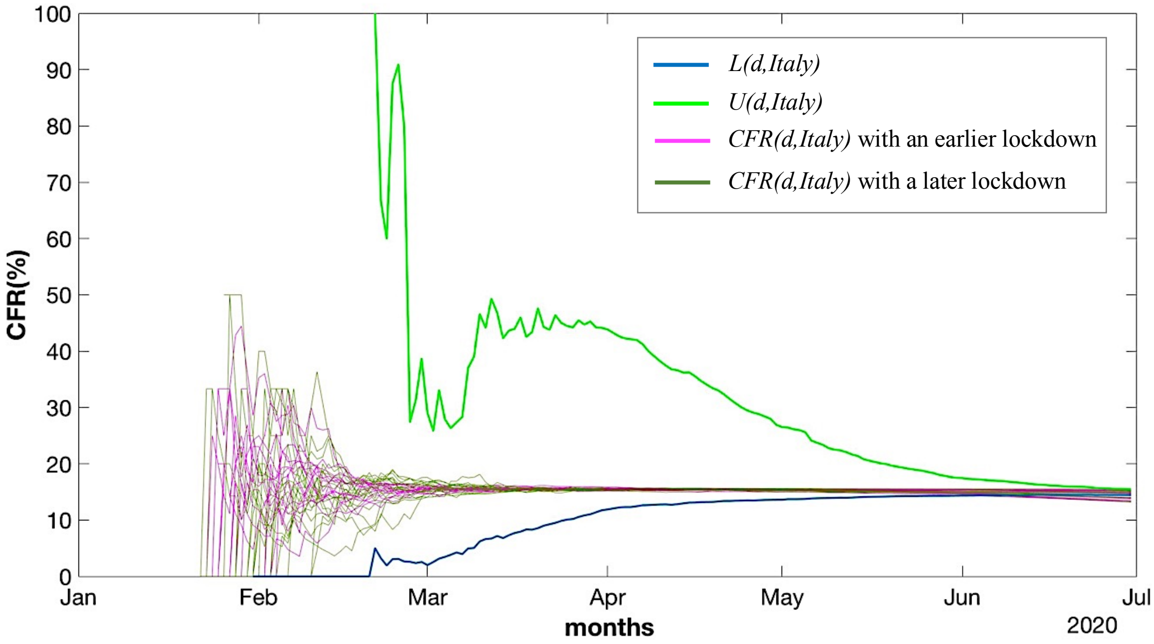

***Figure 25: Comparison of the CFR with different lockdown scenarios.***

# 4. DISCUSSION

## 4.1. Mathematical model

In this project, a mathematical model to reproduce the dynamics of the Covid-19 pandemic has been designed. A new model was developed following the probability that an individual has of being infected from SARS-CoV-2, as well as the probabilities that this individual has to recover or die from the infection. Furthermore, this model was used to calculate the CFR of Covid-19 by considering the active infected cases that will eventually die, as well as to simulate different scenarios related to the lockdown start date.

The outputs obtained from the simulations approximately fit the data reported by Johns Hopkins Coronavirus Resource Center, and the magnitude and date of the number of infected cases was correctly estimated. In this study, we represented social distancing as a sudden and abrupt change but a gradual decline of the probability of getting infected could have been considered. Further studies could be related to changing the effect of the control strategies to see if results would improve.

The estimation of the dynamics of the number of death cases has not shown such good results. This can be explained due to the large variability of dying from Covid-19. We assumed a fixed probability of dying over time. However, published studies suggest that this probability not only varies across countries but also across individuals. In fact, the data obtained so far indicates that elderly people and individuals with pre-existing health conditions such as cardiovascular diseases, diabetes, chronic respiratory diseases, hypertension or cancer have a higher risk of dying from Covid-19 [22]. Furthermore, we assumed that the time interval between getting the infection and dying from it was around to 2 to 3 weeks. However, recent data suggests that patients suffering from Covid-19 are dying much later after getting rid of the virus, due to the pathological consequences that SARS-CoV-2 has caused in their organism [23]. There are other considerations that may affect the probability of dying over time such as the hospital saturation that many countries suffered during the pandemic, as well as the lack of knowledge about the best treatment to administer, since this SARS-CoV-2 is a novel and unknown virus. Thus, all the aforementioned reasons make the representation of a good approximation of the dynamics of dying from Covid-19 challenging.

Additionally, the current model has demonstrated better robustness for Italy and South Korea than for Germany. By looking at the results obtained when changing the initial probabilities of getting the infection, we can conclude that the lack of robustness for Germany's model is due to the fact that very small probabilities are used in order to simulate the beginning of the outbreak, which was later than in Italy and South Korea. Robustness results could be improved by considering a smaller population or by predefining the day in which the outbreak started in every country.

## 4.2. Case fatality rate

From the model and simulations considered in this work, the CFR of different countries could have been calculated by reducing the time delay that exists between getting the infection to dying from it. The CFR observed for Germany was lower in comparison to Italy. Germany did massive testing over most of the population [24]. Therefore, the number of confirmed cases increased, mainly because asymptomatic individuals were taken into consideration, which made the CFR smaller. However, other factors such as culture, randomness or having a competent healthcare system can also decrease the CFR. In particular, Germany has an efficient healthcare system, with a hospital capacity greater than many other countries. In fact, Germany has 8.3 hospital beds for every 1,000 people [25] and 33 intensive care units (ICU) for every 100,000 individuals [26]. Another key element that can also influence this low rate of Germany is that most cases have occurred among young people and it has been shown that the virus is more lethal in older people [27]. In fact, around 67% of the infected cases in Germany have occurred in people between 15 and 59 years old.

Regarding Italy, results show that the CFR computed on 30th June had really high values. This could be due to several reasons such as a healthcare system that was not prepared for the pandemic, which led to the collapse of the system. In fact, before the outbreak, Italy had 3.4 hospital beds for every 1,000 people [25] and just 8 ICU for every 100,000 people [26]. Another explanation for the high CFR values is the lack of massive testing that Italy carried out at the beginning of the outbreak, which led to considering only the most sever infected patients. Moreover, more infections occurred in older people than in Germany, which led to a higher number of deaths. In fact, Italy is the second country in the world with the largest percentage of older adults since 22.8% of the population is older than 65 years old. Furthermore, it is known that the 85% of those who died due to Covid-19 in Italy, were over 70 years old [28].

Focusing on South Korea, the CFR was estimated to be the smallest compared with the ones of Italy and Germany. South Korea performed an early and indiscriminate testing as a first move to fight the infection. In fact, it was the first country to start doing massive tests over the population [18], more than Germany. This led to a reduction of the number of new infections in the subsequent days after the massive testing, so the curve started to flatten. Moreover, since the more tests are done, the more infected people are detected, there was an increase in the number of reported infected cases, making the CFR lower. In addition, the government of South Korea decided to develop tools to extensively track and trace their population. This strategy also proved to be very useful to reduce the transmission of Covid-19, leading to fewer new daily infected cases and to an early detection of the infected in order to quickly treat these patients and thus, reduce the probability of dying. Furthermore, South Korea had from the beginning a clear idea on how to treat the patients and an efficient healthcare system. In fact, South Korea has 11.5 hospital beds for every 1,000 people [25] and 10 ICU for every 100,000 people [26].

## 4.3. Lockdown scenarios

Finally, a study of different lockdown scenarios was also carried out. Despite the differences obtained in both total cases and death cases in the different countries due to the small robustness of the models, especially for Germany, results demonstrated that an earlier lockdown would have decreased the number of infected cases and consequently, the number of deaths. If we take the average between the percentage ranges obtained for the infected cases provided the lockdown had started 10 days earlier, Germany would have decreased around 48% of reported cases and deaths. Regarding Italy, around 8% fewer cases would have been reported if the lockdown had been declared 10 days earlier. This difference can be explained as Italy confined the country when only 9,172 infected cases had been reported, whereas Germany announced the lockdown when there were 24,870 reported cases. Therefore, we can conclude that these governments announced the lockdown not based on the total number of reported cases but on the increment in daily cases, since in Italy cases increased much faster and earlier than in Germany leading to a prior lockdown. The same but opposite dynamics have been observed if the lockdown had started 10 days later.

On the other hand, the strategy of South Korea needs to be considered more carefully as this country was able to reduce the number of infected cases without harming the economy of the country since no lockdown was imposed. Nevertheless, in order to be able to follow South Korea, a good heath system and predisposition of the population is required.

### 4.4. Limitations

The main limitations of the current study should be discussed. First, the public domain data used did not have good quality due to the huge controversy on how countries are reporting Covid-19 cases. Moreover, the number of total infected cases is probably higher than reported since there are many undiagnosed cases, mainly young asymptomatic people. Although some mathematical models existing in literature have approximated the real number of total infected cases by considering the undiagnosed individuals [29], this has not been included in the current work. Therefore, an important limitation of the study is that it is only based on the diagnosed and reported cases. Another constraint is that parameters were manually adjusted and the results of the mathematical model were graphically compared. The use of a more automatic methodology to estimate the parameters and algorithms to calculate the accuracy of the mathematical model would provide more reliable results than those obtained in the current study.

Future studies can add complexity to the model by incorporating more variables such as the age population ranges with distinct probabilities of getting infected, recovering or dying, or the unreported infected cases, in order to obtain more accurate curves of the dynamics of the Covid-19.

# 5. LIST OF FIGURES

# 6. LIST OF TABLES

# 7. BIBLIOGRAPHY

[1] World Health Organization, Novel Coronavirus situation. Report -1 (January 2020). URL https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4. Web; visited 15-03-2020.

[2] World Health Organization, Naming the coronavirus disease and the virus that causes it (2020). URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it. Web; visited 15-03-2020.

[3] World Health Organization, Q&A on coronaviruses COVID-19 (2020). URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses. Web; visited 31-07-2020.

[4] Worldometer, Covid-19 coronavirus pandemic (2020).
URL https://www.worldometers.info/coronavirus/. Web; visited 25-08-2020.

[5] BBC, Coronavirus declared global health emergency by WHO (January 2020). URL https://www.bbc.com/news/world-51318246. Web; visited 15-03-2020.

[6] Ferguson F. *et al* (April 2006). Strategies for mitigating an influenza pandemic, *Nature*, 442(7101):448-51.

[7] Hou C. *et al* (April 2020). The effectiveness of quarantine of Wuhan City against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis, *Journal of Medical Virology*, Volume 92, Issue 7.

[8] Ivorra B. *et al* (2020). Mathematical modelling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China, *Elsevier Public Health Emergency Collection*, PMC7190554.

[9] He S. And Tang S (January 2020). A discrete stochastic model of the COVID-19 outbreak: Forecast and control, *Mathematical Biosciences and Engineering*, 17(4):2792-2804.

[10] Kyrychko Y. et al (July 2020). Mathematical modelling of dynamics and containment of COVID-19 in Ukraine, *COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv*.

[11] World Health Organization, Estimating mortality from Covid-19 (2020). URL https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19 .Web; visited 15-08-2020.

[12] Science Media Center, Expert reaction to WHO Director-General's comments that 3.4% of reported COVID-19 cases have died globally (March 2020). URL https://www.sciencemediacentre.org/expert-reaction-to-who-director-generals-comments-that-3-4-of-reported-covid-19-cases-have-died-globally/. Web; visited 20-03-2020.

[13] CNN, More than half of states may be undercounting coronavirus cases by not following CDC guidelines (June 2020).
URL https://edition.cnn.com/2020/06/09/health/us-coronavirus-tuesday/index.html. Web; visited 31-07-2020.

[14] GitHub, Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE (2020). URL https://github.com/CSSEGISandData/COVID-19. Web; visited 30-06-2020.

[15] Worldometer, Countries in the world by population (2020). URL https://www.worldometers.info/world-population/population-by-country/. Web; visited 20-08-2020.

[16] Wikipedia, COVID-19 pandemic in Germany (2020).
URL https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Germany. Web; visited 15-07-2020.

[17] Wikipedia, COVID-19 pandemic in Italy (2020).
URL https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Italy. Web; visited 15-07-2020.

[18] Our World in Data, Emerging COVID-19 success story: South Korea learned the lessons of MERS (June 2020). URL https://ourworldindata.org/covid-exemplar-south-korea. Web; visited 15-07-2020.

[19] Wikipedia, COVID-19 pandemic in South Korea (2020).
URL https://en.wikipedia.org/wiki/COVID-19_pandemic_in_South_Korea. Web; visited 30-07-2020.

[20] Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (February 2020). URL https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf. Web; visited 30-04-2020.

[21] Verity R. et al (June 2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis, *Lancet Infect Dis*., 20(6):669-677.

[22] Our World in Data, Mortality Risk of Covid-19 (2020).
URL https://ourworldindata.org/mortality-risk-covid. Web; visited 20-07-2020.

[23] Ministerio de Sanidad, Enfermedad por coronaviurs, COVID-19 (April 2020). URLhttps://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/20200417_ITCoronavirus.pdf. Web; visited 30-08-2020.

[24] EuroNews, Germany increases its COVID-19 tests to 500,000 per week (March 2020). URL https://www.euronews.com/2020/03/27/germany-increases-its-covid-19-tests-to-500-000-per-week. Web; visited 15-07-2020.

[25] IndexMundi (January 2020).
URL https://www.indexmundi.com/g/r.aspx?v=2227&l=es. Web; visited 17-07-2020.

[26] EOM, La capacidad de las unidades de cuidados intesivos (2020). URL https://elordenmundial.com/mapas/cuantas-camas-uci-tiene-cada-pais/ Web; visited 17-07-2020.

[27] Robert Koch Institute, Coronavirus Disease 2019 Daily Situation Report (April 2020).
URL https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-04-29-en.pdf?__blob=publicationFile Web; visited 17-07-2020.

[28] Statista Research Department, Coronavirus (COVID-19) deaths in Italy (August 2020). URL https://www.statista.com/statistics/1105061/coronavirus-deaths-by-region-in-italy/ Web; visited 17-07-2020.

[29] Ruiz M. *et al* (July 2020). A simple but complex enough θ-SIR type model to be used with COVID-19 real data. Application to the case of Italy, *Epidemiological modelling of human diseases*, preprint.