

Received January 8, 2021, accepted January 9, 2021, date of publication January 18, 2021, date of current version January 25, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3051929

A Generalized Mechanistic Model for Assessing and Forecasting the Spread of the COVID-19 Pandemic

HAMDI FRIJI¹, RABY HAMADI¹, HAKIM GHAZZAI¹, (Senior Member, IEEE),
HICHEM BESBES², AND YEHIA MASSOUD¹, (Fellow, IEEE)

¹School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030, USA

²Higher School of Communication of Tunis, University of Carthage, Tunis 2083, Tunisia

Corresponding author: Hakim Ghazzai (hghazzai@stevens.edu)

ABSTRACT Since early 2020, the world has been afflicted with an unprecedented global pandemic. The SARS-CoV-19 (COVID-19) has levied massive economic and public health costs across many countries. Due to its virulence, the pathogen is rapidly propagating throughout the world in such a way that makes it incredibly challenging for officials to contain its spread. Therefore, there is a pressing need for national and local authorities to have tools that aid in their ability to assess and extrapolate the future trends of the spread of COVID-19, so they may make rational and informed decisions that minimize public harm. Mechanistic models are prominent mathematical tools that are used to characterize epidemics. In this paper, we propose a generalized mechanistic model with eight states characterizing the COVID-19 pandemic evolution from a susceptible state to discharged states while passing by quarantined and hospitalized states. The parameters of the model are determined by solving a fitting optimization problem with three observed inputs: the number of infected, deceased, and reported cases. The model's objective function is weighted over the training days so as to guide the fitting algorithm towards the latest pandemic period and lead to more accurate trend predictions for a stronger forecast. We solve the fitting problem with the Levenberg-Marquardt algorithm; we compare the performance of the model generated from this algorithm to the one of another state-of-the-art fitting algorithm as well as to the one of another compartmental model widely used in literature. We test the model on the COVID-19 data from four highly afflicted countries. The fitting algorithm has been validated graphically and through numerical metrics, and results show significantly accurate results for most of the countries. Once the model's parameters are estimated, forecasting results are derived and uncertainty regions of the expected scenarios are provided.

INDEX TERMS COVID-19, fitting optimization, mathematical modeling, mechanistic model, nonlinear differential equations, pandemic.

I. INTRODUCTION

Coronaviruses are a large family of viruses that trigger an infection in the contaminated's nose, sinuses, or upper throat, and even fatal acute respiratory distress syndrome (ARDS) [1], [2]. In addition to humans, the coronaviruses may also infect animals including camels, cattle, cats, and bats. Even though coronaviruses originated in animals can rarely infect humans, the Middle East Respiratory Syndrome CoronaVirus (MERS-CoV) and the Severe Acute Respiratory

Syndrome CoronaVirus (SARS-CoV) have broken the rule and shifted to humans. The same happened with the new coming virus, named SARS-CoV-2 by International Committee on Taxonomy of Viruses that causes a disease named Coronavirus Disease 2019 (COVID-19) by the World Health Organization (WHO). Generally, the human-to-human transmission occurs when a contaminated person's coughs, sneezes, or even talks. Indeed, released droplets may stay in the air and can infect people within a 6 feet distance [3]. The virus can also be transmitted through aerosol or surface transmission as researchers have shown that the virus may live in the air up to three hours and on steel for two or three days [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott¹.

Originally starting from Wuhan, Hubei a province of China, the first cases of COVID-19 were announced in the late of 2019. The outbreak quickly spread within China due to the massive transportation and large population mobility before the Chinese Spring Festival, and subsequently, it spread all over the world. The WHO declared the outbreak to be a Public Health Emergency of International Concern on January 30th, 2020 and recognized it as a pandemic on March 11th, 2020. Generally, coronaviruses are not dangerous, however, the new COVID-19 comes with a considerable mortality rate. Currently, more than 12 millions cases of COVID-19 are confirmed resulting on more than 550 thousand deaths in 216 different countries, areas, and territories as indicated by WHO through the COVID-19 situation dashboard website. These numbers are expected to dramatically increase. Most of these cases are located in the United States where more than 3 million confirmed cases and more than 110 thousand confirmed deaths have been reported representing around 25% and 24% of the global results, respectively [5].

In addition to the high fatality rates, the rapid spread of COVID-19 has engendered tremendous social and economic implications all over the world [6]–[10]. For example, R. Y. Kim identified the pandemic as an accelerator of the structural change in consumption and the digital transformation in the marketplace. He also explored how the pandemic accelerated the growth of e-commerce [11]. The global economy has also been impacted since the usage of human distancing, as an easiest solution to limit the spread of the disease. Thus, non-essential services have been shut-down, and hence, hundreds and even thousands of people lost their jobs. Furthermore, since most of the world nations have closed their borders, the international trade has collapsed. Additionally, the impact of the COVID-19 pandemic has reached the telecommunication industry. Accordingly, Internet Providers (IPs) have recently reported a massive and huge traffic due to the lock-down and stay-home orders applied by several states in the USA and other regions and countries [12]. Over just few months, COVID-19 has changed the world and specifically our habits and daily lives and the situation does not seem to return to its normal soon in many countries. Therefore, it is very important to understand the behavior and expected trend of this disease and predict its spread. Forecasting the future evolution of this pandemic will also help assess its future consequences on different social and economic sectors.

In literature, mathematical modelling has been widely used to model pandemics since they provide a quantitative framework that scientists can assess to build hypotheses on the potential underlying mechanisms explaining patterns in the real observed data at both spatial and temporal scales [13]–[17] and especially to model the COVID-19 outbreak [18]–[23]. Mathematical models may have high different complexity levels depending on the number of variables and parameters used to characterize the dynamic states of the system, their spatial and temporal resolutions, e.g., dis-

crete vs. continuous time, and their design, e.g., deterministic or stochastic [24]. The two main mathematical models used to characterize pandemics are phenomenological and mechanistic models:

- **Phenomenological models** are used to describe the relationship between the patterns of the data without a specific basis on the physical laws or mechanisms using an empirical approach, i.e., phenomenological models serve to describe and explain the reasons of the observed interactions between the patterns as they are. Two useful phenomenological models to characterize pandemic growth patterns exist: the Generalized Growth Model (GGM) and the Generalized Richards Model (GRM) [24]–[28]. Although they are commonly used, these models have limitations and downsides because of their subjectivity. Thus, establishing the reliability and validity of these approaches can be challenging, which makes these models very subjective and requires elaborate efforts to output accurate results.

- **Mechanistic models** are used to explain patterns in the observed data while considering key physical laws or mechanisms involved in the dynamics of the investigated problem. The Susceptible-Infected-Recovered (SIR) and Susceptible-Exposed-Infectious-Recovered (SEIR) models are the least complex mechanistic models [29], and they are frequently used to characterize pandemics [24], [30]–[33]. Tuen Wai Ng *et al.* studied the SARS pandemic and compared the SIR and SEIR models' performance in characterising it. Then, they proposed a new SEIRP model by adding a protected state and evaluated its performance on multiple use-cases of SARS outbreak [34]. Another variant of mechanistic models are used in [35] to study the Ebola pandemic, where Tae Sug Do *et al.* presented a model called SLIRD to better understand the spread of the disease and tested their model on the outbreak in Nigeria. Moreover, the authors of [36] presented a calibration of a SEIR–SEI pandemic model to describe the dynamics of the 2016 Zika virus outbreak in Brazil. The authors of [19], investigated the lifting of control measurement, in Wuhan, China, especially social distancing and the effect of returning to work using an SEIR model.

Recently, the authors of [37] presented a generalized SEIR model to characterize the COVID-19 outbreak in different hot spots in China and applied inverse inference to the starting date of the outbreak. Similarly, the study in [38] examined the baseline SIR model along with the SEIR model to characterize the spread in Canada. This study modelled the social distancing through the isolation of a subset of the susceptible population. Another recent study, presented in [39], proposed a comprehensive solution to estimate the mortality due to SARS-Cov-2 infection. An SEIR mathematical model describing infection transmission and death was developed to estimate the case-fatality, the symptomatic case fatality, and the overall infection fatality ratios. The mechanistic model has been applied to different impacted regions of the world with the objective of comparing the different indicators and assess the mortality dynamics per age group. Moreover, in [21]

presented a SIDARTHE model, an extension of the SIR model, which distinguishes between detected and undetected cases and between the different Severity of Illness (SoI) and applied it to characterize the spread in Italy. In [40], the authors presented a basic SEIR model to model the COVID-19 spread in India and used it to simulate and evaluate different scenarios of the spread based on the transmission rate and finally, they proposed potential future evolution of the spread and possible mitigation. Along with mathematical modelling, in [41], the authors employed a machine learning algorithm namely, the decision tree algorithm, to study the indicators responsible for the evolution of COVID-19 pandemic using multi-sourced data. The study identified the population density as one of the most important determinants of infection spread. Moreover, the authors of [42] addressed the future forecasting of COVID-19 spread by identifying the threatening factors using supervised machine learning models. They focused on forecasting the number of newly infected cases, the number of deaths, and the number of recoveries in the next ten days after the training period.

Although the previously mentioned studies, at some points, succeeded to model the pandemic, they only focus on specific countries or regions and their models are not generalized to fit different regions simultaneously. The fitting is also usually done using one target data set containing the number of confirmed cases. Moreover, the presented models are usually based on simplistic models, e.g., SIR and SEIR and do not provide an advanced overview about the different scenarios that a contaminated person may face.

In this paper, we propose to investigate the evolution of COVID-19 in different countries of the world using an extended mechanistic model. The developed model is composed of eight states: protected, susceptible, exposed, infectious, recovered, hospitalized, quarantined, and deaths. These states aim to characterize the possible behaviour of the crowd and their effects on the pandemic spread. The evolution of the pandemic is then modeled as an Ordinary Differential Equation (ODE) system interconnecting the different states. In order to forecast the evolution of COVID-19, we proceed with an identification process that aims to estimate the values of the coefficients of the model. To this end, a fitting optimization problem is developed having as inputs three target data sets, i.e., observations: confirmed infected cases, death cases, and recovered cases. The objective function to be optimized is modeled as a weighted non-linear least squares function and is solved by the Levenberg-Marquardt (LM) algorithm. The weights are added to guide the focus of the fitter towards a particular period of the trend. The model is fitted and tested on multiple countries and show accurate performances. The impact of the weights on the fitting performance has been also evaluated. The performances of the fitting algorithms are also compared to the ones of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) [43] algorithm for various metrics. The estimated parameters are then used to forecast the future evolution of the pandemic given the current observation. Then, uncertainty regions are provided

to visualize the expected upper and lower limits of the future evolution of the number of cases and deaths. To the best of our knowledge, we are the first to characterize a pandemic with a developed model presenting a loop between the states while taking into consideration all possible movements between them. Moreover, we are the first to employ a weighted fitting process that endorses the forecasting process by focusing on the last trends of the spread. We also provide a comparison of the estimated, best, and worse cases with regards to the infection rate of the disease to estimate how well the countries operated against the pandemic and finally, to validate our forecast results, we compare the performance of our proposed model to those of a Random Walk (RW) approach with drift [44] and show that our proposed model outperforms the RW in most of the investigated scenarios. Finally, we have validate the efficiency of our proposed model with the state-of-the-art SEIRD model composed of five compartments, namely, susceptible, exposed, infected, recovered and deceased cases.

The rest of the paper is organized as follows. In Section II, we present the developed model and the different parameters that characterize the COVID-19 spread. Then, we present the proposed weighted fitting technique in Section III. The Section IV is devoted to evaluate the performance of the fitting algorithm as well as expected evolution of the COVID-19 spread in the upcoming period using the proposed fitting technique. The paper is concluded in Section VI.

II. PROPOSED MECHANISTIC MODEL

In this section, we present a generalized model to characterize the COVID-19 pandemic composed with eight different states or phenotypes. These states are supposed to characterize all the life cycle of an infected case before the infection, during the infection, and after being discharged, i.e., either deceased or recovered as shown in Fig. 1. Each state in our model is supposed to characterize, at a given unit of time t , in our case day, a specific population's behaviour as follows:

- $S(t)$: A state representing the total number of susceptible cases that can be infected with the COVID-19 pandemic.
- $P(t)$: A state representing the total number of protected cases that are taking the required precautions, i.e., quarantine, social distancing, etc. This state represents the category of people that are very unlikely to be infected with the COVID-19.
- $E(t)$: A state representing the total population exposed to the pandemic and are infected but not yet reported. The population within this state are supposed to be infected but not yet be infectious, i.e., in a latent period. An exposed case can be, or not, reported.
- $I(t)$: A state representing the number of confirmed infected cases with infectious capabilities that are not respecting the stay-home rules, if exist, and propagating the disease.
- $Q(t)$: A state representing the number of confirmed cases that are quarantined and are not able to infect other people. In this paper, we suppose that not all the confirmed cases are quarantined.

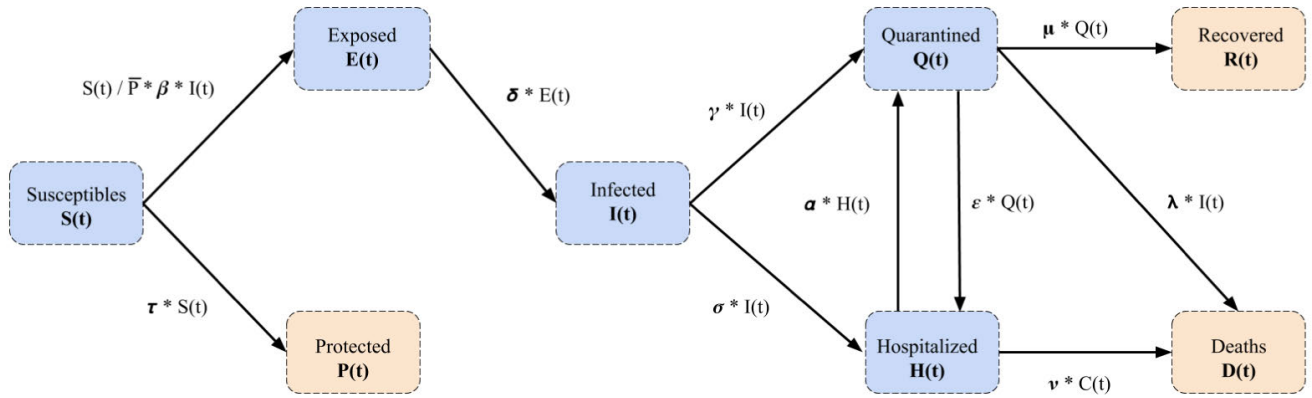


FIGURE 1. The proposed mechanistic model architecture to characterize the COVID-19 pandemic spread, where the orange boxes are the final states and other boxes are intermediate states.

- $H(t)$: A state representing the number of confirmed cases that are hospitalized. This state includes the cases who require an Intensive Care Unit (ICU) bed to be treated. In this paper, we suppose that the hospitalized cases are quarantined and hence, they do not transmit the disease to other people.
- $D(t)$: A state representing the total number of deaths due to the COVID-19 pandemic.
- $R(t)$: A state representing the total number of recovered cases from the COVID-19 pandemic. We assume that the recovered cases are supposed to gain an immunity against the disease and cannot be infected again [45].

Given the definitions of the states presented earlier, we define that the total population, denoted by \bar{P} , as $\bar{P} = S(t) + P(t) + E(t) + I(t) + Q(t) + H(t) + R(t) + D(t)$ for each day t . We can also define the total number of confirmed infected cases, denoted by I^{rep} , and the total number of infected but not reported cases, denoted by I^{Nrep} , for each day t , as follows:

$$I^{rep}(t) = I(t) + H(t) + Q(t), \quad \forall t, \quad (1a)$$

$$I^{Nrep}(t) = E(t), \quad \forall t. \quad (1b)$$

To characterize the possible interactions between the previously presented states, we represent each possible transition from one compartment to another with an arrow as shown in Fig. 1. Each arrow is labeled with a rate describing the amount of time required for the transition to take place multiplied by the population of the group of individuals that the transition applies to [46]. The rates between the different compartments are defined as follows:

- δ^{-1} : The length of the incubation period, i.e., the average period from exposure to symptoms onset.
- σ^{-1} : The expected amount of time during which the health of an infected person will get worse and require hospitalization.
- γ^{-1} : The average quarantine period.
- α^{-1} : The average period for hospitalized cases to be moved from ICU bed to quarantine.

- ϵ^{-1} : The expected number of days for a quarantined case to get worse and require hospitalization.
- λ : The infected population’s fatality rate.
- ν : The hospitalized population’s fatality rate.
- χ : The quarantined population’s fatality rate.
- μ : The quarantined population’s recovery rate.

We also employ two other parameters to characterize the disease and the possible protection methods that healthy people may employ to protect themselves:

- τ : A protection rate to quantify the possible protection measurements, i.e., social distancing, mask usage, quarantine respect, etc.
- β : An infection rate reflecting the expected number of people that an infected person infects per day.

In our model, we suppose that an infected case can be healed only after passing by the quarantine state because the COVID-19 usually has symptoms, even light ones, that will oblige the contaminated persons, sooner or later, to stay home. In other terms, our assumption is to suppose that the COVID-19 symptoms with different degrees (light to severe) will appear on all the infected cases and will force them to stay home, i.e., quarantined for at least one day before being totally healed from the diseases. Since our model is based on a constant infection rate β , considering that an infected person will be placed on quarantine before its final recovery will allow our model to learn that an infected case in state I will not affect people daily, i.e., this assumption will make our model closer to reality by allowing it to automatically learn that an infected person may or may not infect people daily and will help, when needed, determine accurate basic reproduction number values regardless the constant parameters. Other studies are considering similar scenario where recovery is only possible through quarantine can be found in [37], [47].

Moreover, we assume that an infected case may require intensive care as soon as its contamination is confirmed and hence, can immediately hospitalized in state H . The infected cases may also choose to be quarantined in state Q and

hence, will not be able to spread the disease or they may also keep contaminating other people and spreading the disease in state I . We also assume that hospitalized cases are quarantined by default so they do not spread the virus but they may pass away while in intensive care, otherwise they may be moved to the quarantine state Q before their final discharge, i.e., either to recovered state R or the death state D . Moreover, in the model, we suppose that an infected case may move in loop between the quarantined and hospitalized states multiple times. Also, we assume that infected cases may die while being on quarantine or hospitalized at any instant of time otherwise they will recover. Finally, our analysis exclusively encloses the COVID-19 cases and hence, we only consider the death cases that are only due to the COVID-19.

The proposed model, shown in Fig. 1, works as a closed system and is designed to track the local spread inside multiple countries to help officials take convenient protective policies against this pandemic such as border opening dates and stay-home orders. It can be translated into an ODE system that resumes the variation speed between the different states. The ODE system is expressed as follows:

$$\frac{\partial S(t)}{\partial t} = -\beta \cdot \frac{S(t)}{\bar{P}} \cdot I(t) - \tau \cdot S(t), \tag{2a}$$

$$\frac{\partial P(t)}{\partial t} = \tau \cdot S(t), \tag{2b}$$

$$\frac{\partial E(t)}{\partial t} = \beta \cdot \frac{S(t)}{\bar{P}} \cdot I(t) - \delta \cdot E(t), \tag{2c}$$

$$\frac{\partial I(t)}{\partial t} = \delta \cdot E - \gamma \cdot I - \lambda \cdot I - \sigma \cdot I, \tag{2d}$$

$$\frac{\partial Q(t)}{\partial t} = \gamma \cdot I(t) + \alpha \cdot H(t) - \epsilon \cdot Q(t) - \chi \cdot Q(t) - \mu \cdot Q(t), \tag{2e}$$

$$\frac{\partial H(t)}{\partial t} = \sigma \cdot I(t) + \epsilon \cdot Q(t) - \alpha \cdot H(t) - \nu \cdot H(t), \tag{2f}$$

$$\frac{\partial R(t)}{\partial t} = \mu \cdot Q(t), \tag{2g}$$

$$\frac{\partial D(t)}{\partial t} = \lambda \cdot I(t) + \chi \cdot Q(t) + \nu \cdot H(t). \tag{2h}$$

An initialization of the system is needed in order to solve the ODE system (2). The initial conditions will be denoted as follows by $S_0, P_0, E_0, I_0, Q_0, H_0, D_0$, and R_0 for $S(t), P(t), E(t), I(t), Q(t), H(t), D(t)$, and $R(t)$, respectively. This model will be used to characterize the spread of the COVID-19 pandemic. To this end, curve fitting will be applied using three target data sets to estimate the parameters of the model. Once estimated, the ODE will be solved by integrating each equation in (2) over the desired period to determine the population of each state with respect to the initial conditions provided.

Lastly, our proposed model has many novel contributions compared to previous studies on COVID-19 spread. In Table 1, we introduce a high-level comparison between our model and those used in these studies.

TABLE 1. High-level comparison of the proposed model and previous studies on COVID-19 spread.

Feature	Proposed work	[43]	[48]	[49]
Number of states	8	4	4	6
Protected state	✓	✗	✗	✓
Quarantine state	✓	✗	✗	✓
Hospitalized state	✓	✗	✗	✗
Weighted model	✓	✗	✗	✗
Last trends favored	✓	✗	✗	✗

III. CURVE FITTING MODEL

The goal of this section is to present the curve fitting model and discuss the algorithm employed to solve the corresponding optimization problem, i.e., the Levenberg-Marquard method. Finally, we discuss the employed techniques to validate the efficiency of our model.

The optimization phase is based on three target real-world data set (real data or observations), collected from official data sources. Each data set represents a vector of length N where N is the number of training days, i.e., period during which real data is observed and collected. The vectors are defined as follows:

- Deaths (D_r): the elements of this vector indicate the total number of confirmed deaths due to COVID-19 at each day officially reported.
- Recovered (R_r): the elements of this vector indicate the total number of recovered cases at each day officially reported.
- Infected (I_r): each element of I_r contains the total number of confirmed infected cases officially reported. This accounts for previous infected cases plus the new reported cases that day.

In this paper, we investigate a multiple data set optimization problem, in which we aim to fit the curve of the model to the actual data by minimizing an objective function that takes into consideration the errors between every real and estimated data at each training day. In the sequel, we denote by n the n^{th} training day where $n = 1, \dots, N$. Hence and for notation purposes only, we use this discrete notation n instead of the continuous one t used in (2). Finally, we propose to employ a weighted Non-Linear Least Squares (NLLS) method where the objective function that calculates the Mean Squared Error (MSE) between the predicted and the real data values given the estimated values of the fitting variables is expressed as follows:

$$\hat{\theta} = \arg \min_{\theta} \sum_{n=1}^N w_n ((\hat{I}_r(n, \theta) - I_r(n))^2 + (\hat{R}(n, \theta) - R_r(n))^2 + (\hat{D}(n, \theta) - D_r(n))^2), \tag{3}$$

where $I_r(n), R_r(n)$, and $D_r(n)$ are the n^{th} element of the real data of the infected, recovered, and deaths vectors, respectively. The estimated vectors $\hat{I}_r(n, \theta) = \hat{I}(n, \theta) + \hat{Q}(n, \theta) + \hat{H}(n, \theta) + \hat{R}(n, \theta) + \hat{D}(n, \theta)$, $\hat{R}(n, \theta)$, and $\hat{D}(n, \theta)$ are the

predicted data of the total number of $I_r(n)$, $R_r(n)$, and $D_r(n)$ given a value of a predict vector θ , respectively. The optimized vector $\hat{\theta}$ includes the list of all target parameters as follows:

$$\hat{\theta} = [\hat{\beta} \ \hat{\delta} \ \hat{\rho} \ \hat{\mu} \ \hat{\lambda} \ \hat{\chi} \ \hat{\alpha} \ \hat{\sigma} \ \hat{\epsilon} \ \hat{\nu} \ \hat{\tau}]^T, \quad (4)$$

and finally, w_n are the weights of the data where $w_n \in [0, 1]$ and $\sum_{n=1}^N w_n = 1$. Notice that the $\hat{I}_r(n, \theta)$ is expressed as the sum of all the states that involve infected, recovered, and deceased persons as per the definition of $I_r(n)$. The idea of employing weights is to endorse the fitting process and force the model to prioritize the last trends of the spread for each country and prevent the initial periods from miss-leading the model, and hence, ensure that the forecast follows the same latest trends to provide accurate forecast results that are on the verge of reality. While uniform weights maintain the same importance level for all the observed results for each day, the piecewise constant, linear, and exponential weights will increase the importance of the latest observed results and neglect the earliest ones uniformly, linearly, and exponentially, respectively. A convenient choice of the weights, that does not completely neglect the first observations, will not necessarily lead to an overfitting since the predictions for a given day t are highly correlated with the data of the last recent period. In other words, the number of active infected cases reported in the recent few past days are those who are going to spread the disease and contaminate a part of the susceptible population S . This correlation relationship between day t and the previous days is perfectly modeled using the weights that will give a higher importance with respect to time so that our model will learn two major properties to help the forecast: i) the smooth evolution of the weights will help optimize the parameters of the model and so learn the relationship between the states and ii) the last trends will help achieve accurate forecasting results.

A. MODEL WEIGHTS

The objective function of the fitting problem is weighted according to the training days. Hence, the values of the weights will allow the possibility to orient the focus of the fitting model towards some specific days more than the others. This will allow, for example, to relax the focus on the fitting model on the first periods of the training where the number of infected cases is very low. Indeed, the ultimate objective is to predict the evolution of the COVID-19 spread right after the training phase. In this paper, we test the following three configurations of weights:

- Uniform function: this is the typical model, which is investigated in most of the previous studies. In this model, a uniform weight is assigned to each training day, i.e., $w_n = \frac{1}{N}, \forall n$. Hence, all the days are treated similarly by the fitting model.

- Piecewise constant function: in this case, some consecutive days are given additional priority more than the others, e.g., $w_1 = \dots = w_{n^0} = p_1$ and $w_{n^0+1} = \dots = w_N = p_2$ where $n^0 p_1 + (N - n^0) p_2 = 1, p_1 < p_2$, and $n^0 \in [1, N]$.

In this example, the fitting model will focus less on the n^0 first days than the remaining days. Hence, less fitting errors will be tolerated for the first days.

- Non-decreasing function: in this model, the first day will be assigned the least weight value while the last day will be assigned the highest weight value. The weights are assigned to the remaining days following a non-decreasing function, e.g., linear or exponential function. Hence, the fitting model will progressively give more importance to last days. This is a very important choice since the first days of the pandemic are usually characterised with a low number of deaths, infected cases, and recovered persons. Moreover, as we aim to forecast the future evolution of the pandemic, it is worthwhile to have accurate fitting at the end of the fitting period so as to obtain exact trends at the last portion of each data.

B. OPTIMIZATION PROBLEM

The ODE system presented in (2) could be expressed in the following non-linear matrix form:

$$\dot{X}(t) = A \times X(t) + B \times (S(t) \cdot I(t)), \quad \forall t, \quad (5)$$

where

$$X(t) = [S(t) P(t) E(t) I(t) Q(t) H(t) R(t) D(t)]^T, \quad (6a)$$

$$\dot{X}(t) = [\dot{S}(t) \dot{P}(t) \dot{E}(t) \dot{I}(t) \dot{Q}(t) \dot{H}(t) \dot{R}(t) \dot{D}(t)]^T, \quad (6b)$$

$$B = \frac{\beta}{P} [-1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0]^T, \quad (6c)$$

Note that $\dot{X}(t)$ is the derivative with respect to time of $X(t)$. Finally, (7) as shown at the bottom of the next page.

Hence, the Non-Linear Least squares optimization problem for the proposed mechanistic model for COVID-19 characterization:

$$(P) \quad \hat{\theta} = \arg \min_{\theta} \sum_{n=1}^N w_n ((\hat{I}_r(n, \theta) - I_r(n))^2 + (\hat{R}(n, \theta) - R_r(n))^2 + (\hat{D}(n, \theta) - D_r(n))^2),$$

subject to:

$$\text{ODE model: } \dot{X}(n, \theta) = A \times X(n, \theta) + B \times (S(n, \theta) \cdot I(n, \theta)), \quad \forall n,$$

$$\text{Initialization: } P(0) = P_0, S(0) = S_0, E(0) = E_0, I(0) = I_0, Q(0) = Q_0, H(0) = H_0, D(0) = D_0, \text{ and } R(0) = R_0. \quad (8)$$

C. FITTING ALGORITHM AND EVALUATION METRIC

Due to the non-convexity of the problem, the NLLS optimization problem given in (8) cannot be analytically and optimally solved.

In this case, there are two major approaches. It is possible to exploit meta-heuristic algorithms to solve the NLLS problem [50]–[52] or employ numerical optimization algorithms that attempt to reach local minima with gradient-based techniques. The latter method is exploited in this paper. The main representative of this class of algorithms is the Newton algorithm. A downside of this algorithm is that it is considerably time-consuming due to the need for line searches and the

computation of the Hessian matrix. To solve NLLS estimation problems, it is often better to exploit the quadratic structure of the cost function, as it is done by the Gauss-Newton (GN) [53] and Levenberg-Marquard (LM) [54], [55] algorithms, which are, by far, the most popular NLLS optimization algorithms. In this study, we exploit the fitting algorithm to estimate the parameters presented in (2) that minimize the squared error between the real data $I_r(n)$, $R_r(n)$, and $D_r(n)$ and their corresponding predicted values.

1) NLLS OPTIMIZATION ALGORITHM:
LEVENBERG-MARQUARDT (LM)

The LM algorithm [54] is an iterative technique to solve NLLS problems. This method is a combination of the Gradient Descent algorithm, that is efficient for early iterations, but performs slowly when it gets close to the best-fit values, and the GN algorithm, that is inefficient in early iterations, but performs perfectly when close to the best-fit values. The LM method uses steepest descent in early iterations and then gradually switches to the GN approach which means that in many cases it finds a solution even if it starts very far from the final minimum value and hence, it guarantees convergence. In fact, LM method can be seen as a generalization of the GN algorithm as its normal equations to estimate the fitting parameters are derived from those of the GN method with an additional damping factor that is adjusted every iteration.

LM is slightly more computationally demanding than GN, but it converges for all initialization combinations that are far away from the solution, where GN often fails [56]. LM shows global convergence properties and is, therefore, the preferred method of choice in common NLLS problems and consequently for our problem [56]. Finally, to ensure an efficient fitting, especially that the optimization target function is non-convex, we consider multiple starting point optimization (i.e., multiple starting points of θ) to select the best fitting solution among the tested ones.

2) FITTING ALGORITHM VALIDATION

The model validation is the most important step in the model building process. Our model validation process is composed of two levels:

- Graphical validation: In this level, we validate the model by comparing the original data and the fitted data in the same graph. This method is important to visualize the overall variation of the model through time.

- Numerical validation: In this level, we calculate numerical metrics to evaluate the error between the estimation and the real data and to verify the wellness of the fit. The employed metrics are classified into three types [57]:

- Scale-dependent metrics: These metrics provide insights about the difference between the predicted and measured values. The most commonly used are: Mean Absolute Error (MAE) and Normalized Root Mean Squared Error (NRMSE) normalized by the difference between maximum and minimum actual data. Note that our observed data, in practice, is not a constant data. Their expressions are given as follows:

$$AE_n = |y_n - \hat{y}_n|, \quad \forall n, \tag{9a}$$

$$MAE = \frac{1}{N} \sum_{n=1}^N AE_n, \tag{9b}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2}, \tag{9c}$$

$$NRMSE = \frac{RMSE}{\max(y_n) - \min(y_n)}, \tag{9d}$$

- Percentage-error metrics: These metrics provide insights about the percentage of the difference between the approximated and observed values. The most commonly used is the Symmetric Absolute Percent Error (SAPE). Its expression and its mean are given as follows:

$$SAPE_n = \frac{|y_n - \hat{y}_n|}{\left(\frac{y_n + \hat{y}_n}{2}\right)}, \quad \forall n, \tag{10}$$

$$SMAPE = \frac{1}{N} \sum_{n=1}^N SAPE_n. \tag{11}$$

- Relative-error metrics: These metrics provide insights about the error with regards to its real observed values. The most commonly used are the Mean Relative Absolute Error (MRAE) and the coefficient of determination R^2 , which are expressed as follows:

$$MRAE = \frac{1}{N} \sum_{n=1}^N \frac{|y_n - \hat{y}_n|}{|y_n - \bar{y}|}, \tag{12a}$$

$$R^2 = 1 - \frac{\sum_{n=1}^N (y_n - \hat{y}_n)^2}{\sum_{n=1}^N (y_n - \bar{y})^2}, \tag{12b}$$

$$A = \begin{pmatrix} -\tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta & -(\gamma + \lambda + \sigma) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma & -(\mu + \epsilon + \chi) & \alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma & \epsilon & -(\alpha + \nu) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & \chi & \mu & 0 & 0 & 0 \end{pmatrix} \tag{7}$$

where \bar{y} is the mean value of the observed data.

Similarly to the percentage-error metrics, the relative-error metrics takes into account the percentage change between the estimated and the observed values of the data. However, it provides insights on the quality of the training in general by considering the order of amplitude of the data that may significantly vary between different parts of the training period. The values of these metrics are independent of the population of the studied countries. A perfect fitting should achieve an MRAE and R^2 close to 0 and 1, respectively.

Our model is tested and validated using those two levels of validation in the following section.

IV. RESULTS & DISCUSSION

In this section, we present the simulation results, investigate the proposed model performance, and validate the use of the proposed fitting technique. To this end, we start by scrutinizing the COVID-19 spread in Russia to highlight the effectiveness of the training of the LM algorithm and the use of different weights in the fitting objective function. Afterwards, we present the fitting results for all the other investigated countries, namely Brazil, Italy, and USA using the proposed approach and the best weight combination. We also compare the performances of the employed LM algorithm to those of another fitting approach, i.e., BFGS [43]. The proposed mechanistic model is investigated using a real world data set containing the number of reported infected, deceased, and recovered cases for the different countries. The Novel Coronavirus (COVID-19) Cases Data data set used in this paper is obtained from The Humanitarian Data Exchange website.¹ The initialization values of each country are given in Table 2. In this study, the training period starts on January 22nd, 2020 and ends on June 4th, 2020 unless otherwise stated. For some countries, we suppose that $E_0 = 1$ even though no confirmed cases are reported, namely for the case of Russia and Brazil, to take into consideration the delay of an exposed case to be reported, i.e., the time from being exposed to the virus until the appearance of the symptoms. Finally, the estimated parameters obtained from the fitting are used to forecast the future pandemic spread for each country by integrating the ODE equations over the desired period. The selected countries are chosen due to their high number of confirmed cases reported to WHO. They are classified among the highest infected countries and witnessing varying trends,

¹https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases

TABLE 2. Initialization of the ODE model states in millions for each investigated country.

Country	S_0	P_0	E_0	I_0	Q_0	H_0	R_0	D_0
Russia	145	0	10^{-6}	10^{-6}	0	0	0	0
Italy	60	0	10^{-6}	10^{-6}	0	0	0	0
Brazil	212	0	10^{-6}	10^{-6}	0	0	0	0
USA (Phase I)	331	0	10^{-6}	10^{-6}	0	0	0	0
USA (Phase II)	212	118	1	2.06	0.58	0.14	0.55	0.11

which might be challenging for our model. The forecast of the spread for each country is over a period $\Delta t = 365$ days, starting from January 22nd, 2020. This work is implemented using python 3.7. As for the ODE integrator, we use the SciPy library where we employ the ‘odeint’ function that solves systems of first-order ordinary differential equations using ‘lsoda’ implementation of the FORTRAN library ‘odepack’. For the fitting, we employ the python library ‘Lmfit’ that solves the formulated optimization problem (P) using a predefined algorithm (e.g., LM and BFGS algorithms). Taking advantage of the ‘odeint’ function, we use a multi-start optimization to run the ‘Lmfit’ and then, randomly select initialization values achieved the best fitting performance (i.e., minimizing the objective function of the optimization problem (P)). Note that, we selected the same initialization values for both fitting algorithms (LM and BFGS).

A. TRAINING PHASE AND WEIGHTS IMPACT: THE CASE OF RUSSIA

During this phase, we employ the LM algorithm to solve the fitting optimization problem and estimate the ODE system parameters. To this end, we investigate the effect of weights employment in the fitting process. In this paper, three forms of weights are considered: piecewise constant, linear, and exponential weights. The expressions of the employed normalized weight functions are given, respectively, as follows:

$$\forall n = 1, \dots, N, \quad w_n^{pw} = \begin{cases} 25 \cdot 10^{-4} & \text{if } n < 70 \\ 125 \cdot 10^{-4} & \text{if } n \geq 70 \end{cases} \quad (13a)$$

$$w_n^{lnr} = A_l n + B_l, \text{ where } A_l = 8.75 \cdot 10^{-5}, B_l = 1.5 \cdot 10^{-3}, \quad (13b)$$

$$w_n^{exp} = A_e \exp(B_e n) + C_e, \text{ where } A_e = 3 \cdot 10^{-5}, B_e = 4 \cdot 10^{-2}, C_e = 6 \cdot 10^{-3}. \quad (13c)$$

Note that the weight functions are obtained after normalization over the number of training days ($N = 135$ days) and that the sum of their values is equal to 1. Fig. 2 plots the curves corresponding to the considered normalized weight functions.

In Fig. 3, we represent the variation of the MSE error with respect to the training iterations for the case of Russia employing the three different weight functions and the LM algorithm. It is shown that, in all cases, the LM reaches stagnation at different minimum values, which confirms the convergence of the model. However, these values are very close to each other for all weight functions. Note that the plotted MSE in this figure considers the different data sets all together, i.e., the infected, deceased, and recovered data. Indeed, it plots the achieved weighted objective function expressed in (3) during the training process.

Graphically, we visualize the trained fitting of the different observed states separately using the LM algorithm for the different functions of the weights in Fig. 4. By observation,

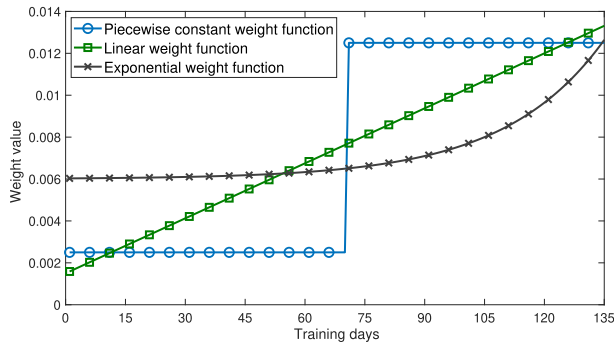


FIGURE 2. Plots of the investigated normalized weight functions.

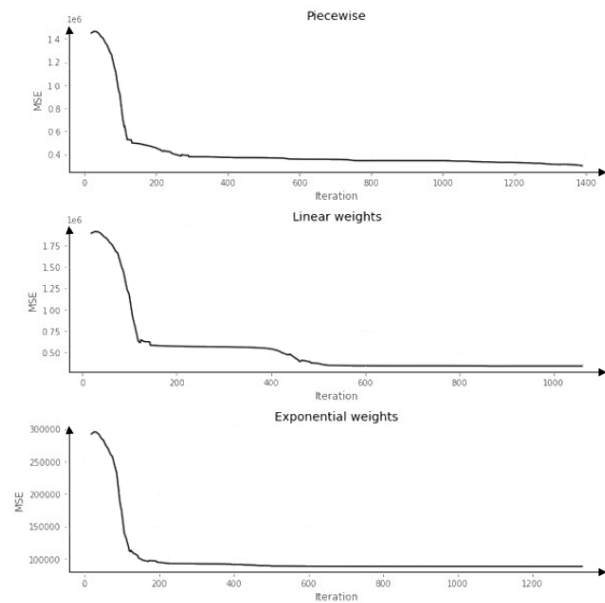


FIGURE 3. Weighted MSE evolution during the training phase for each weight function.

these graphs confirm that the exponential and linear weights outperform the piecewise constant function of the weights. Indeed, although the fitted data do not exactly fit the real data at the beginning of the pandemic, it successfully follows it at the latest training days, which makes the models ready to forecast with respect to these trends.

The fitting results for each weight function using the LM algorithm are numerically corroborated in Table 3 where we provide the achieved average values for the different validation metrics. The table shows that the use of exponential weights allowed to obtain less error than the other

TABLE 3. Russia: Numerical validation of the LM fitting algorithm using the different weight functions.

Weight Function	MAE	NRMSE	SMAPE	MRAE	R^2
Piecewise	89476.19	0.78	3.13	1.12	0.862
Linear	68215.24	0.43	2.98	0.86	0.903
Exponential	44215.39	0.21	1.22	0.58	0.947

investigated weights in terms of all the evaluation metrics: state-dependent, percentage-error, and relative-error metrics. For instance, the MAE with the exponential weight provides is around 44215 which is 51% less than the MAE obtained with the piecewise constant weight function. In terms of R^2 , that reflects the goodness of fit, we can notice that the exponential function of the weights provide the highest value 0.94, which is very close to 1, compared to the R^2 achieved by the piecewise constant weight function (i.e., $R^2 = 0.86$). This proves the efficiency of the fitting of the exponential weights. Indeed, as mentioned earlier, the non-decreasing functions of the weights allow the fitting algorithm to focus more on the latest training days in progressive way and due to the nature of the pandemic which starts with very small number of cases, the fitting errors are extremely reduced for relatively high number of infected, deceased, and recovered cases.

Finally, in Fig. 5, we plot the daily values for each training day of the APE, RAE, and AE metrics using the exponential function of the weights. The figure clearly illustrates the focus of the fitting on the last days of the training. Indeed, the SMAPE is significantly reduced for the different states during that period. The RAE is again confirming the efficiency of the fitting where the relative-error is less than 2% for the whole training period except some particular spots as the fitting is not exactly perfect. Finally, the AE highlights the absolute error for each day and state. Although it may reach the order of thousands, this error remains small compared to the real total number of infected and recovered cases.

Throughout the rest of this section, we proceed with the exponential function of the weights to perform the fitting process.

B. FITTING PERFORMANCE AND VALIDATION

In this section, we present the results of the proposed weighted fitting technique using the LM algorithm for different countries. Afterwards, we compare the efficiency of the proposed fitting model to another fitting algorithm, i.e., the BFGS.

1) THE CASES OF RUSSIA, ITALY, AND BRAZIL

We propose to provide the fitting results for several countries to prove that our model is capable of providing accurate fitting in many situations. Indeed, we chose Italy as one of the first affected countries witnessing one of the highest death rates during the first four months of the pandemic, and which is showing now stable and much safer situation. Russia is also a very interesting scenario to study due to its unique spread profile (i.e., Russia witnessed a tremendous spike in COVID-19 virus infections after a period characterized with a low infection rate). Finally, we investigate the case of Brazil, which is currently witnessing its spike.

The results presented in Fig. 4, Fig. 6, and Fig. 7 show that the model can adapt the last trend of the data for all previously mentioned countries with some minor errors. Indeed, due to the existence of three different data sets having different orders of magnitude and trends, the fitting algorithm may not

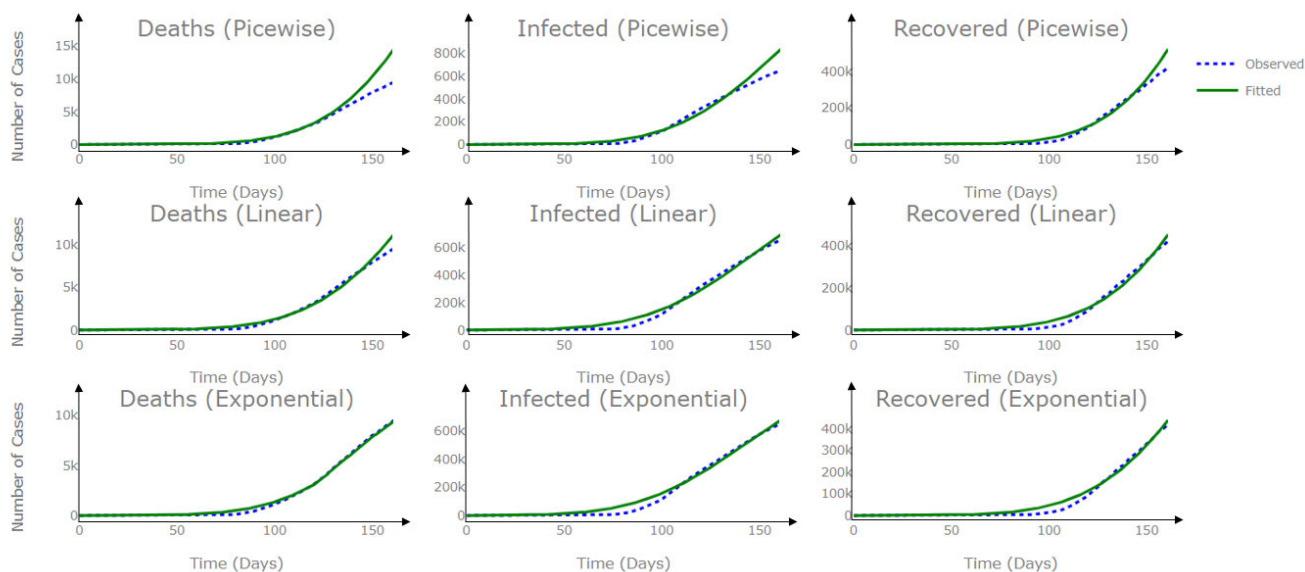


FIGURE 4. Russia: Graphical validation of the LM algorithm for different weight functions.

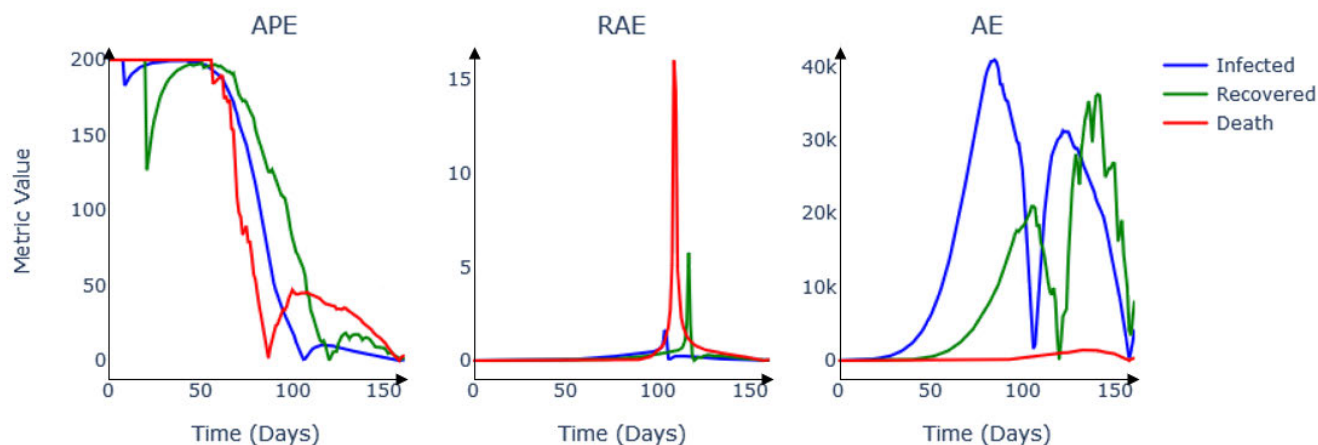


FIGURE 5. Russia: Illustration of the values of the APE, RAE, and AE metrics versus the training days using the exponential weight function.

fit exactly all the curves but globally the fitting is effective as confirmed by the achieved metrics presented in Table 4. Italy is showing the weakest results among the three investigated countries. Although it achieves similar R^2 as Russia, the fitting of Italy reaches around twice the MRAE value obtained with Russia. It is also confirmed graphically in Fig. 6 for the infected and due its small population. Notice also the

TABLE 4. Numerical validation of the LM fitting algorithm using the exponential weight function for Russia, Italy, and Brazil.

Country	MAE	NRMSE	SMAPE	MRAE	R^2
Russia	44215.39	0.21	1.22	0.58	0.947
Brazil	62419.17	0.277	1.12	0.34	0.995
Italy	9769.24	1.81	1.3	0.94	0.951

MAE of Italy is low compared to Russia and Brazil because of the low number of infected cases reported as well as its small population. The fitting of Brazil is the most accurate as confirmed graphically and with the achieved evaluation metrics ($R^2 = 0.99$ and $MRAE = 0.34$). Minor errors are noticed mainly during the first period of the pandemic as tolerated by the utilized weight function.

Furthermore, in this study, we compare the performance of the proposed fitting algorithm to the one of another fitting algorithm employed in [43] where the BFGS algorithm is used to resolve the fitting problem. In Fig. 8, we plot an example of the fitting for the case of Russia using the exponential weight function and BFGS algorithm. It is shown that the BFGS partially fails in fitting all the three data sets, especially the latest days of the infected and recovered cases. In Table 5, we provide a performance comparison between the

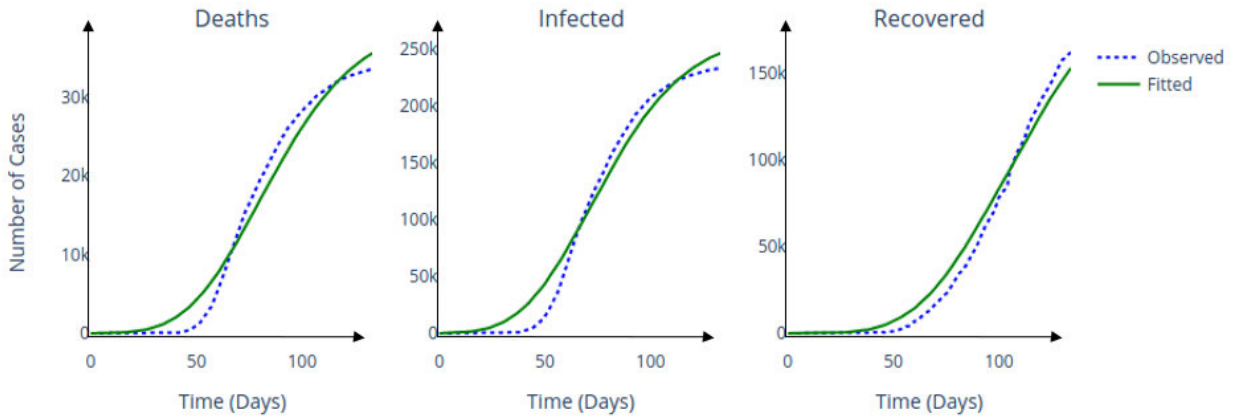


FIGURE 6. Italy: Graphical validation of the LM algorithm using the exponential weight function.

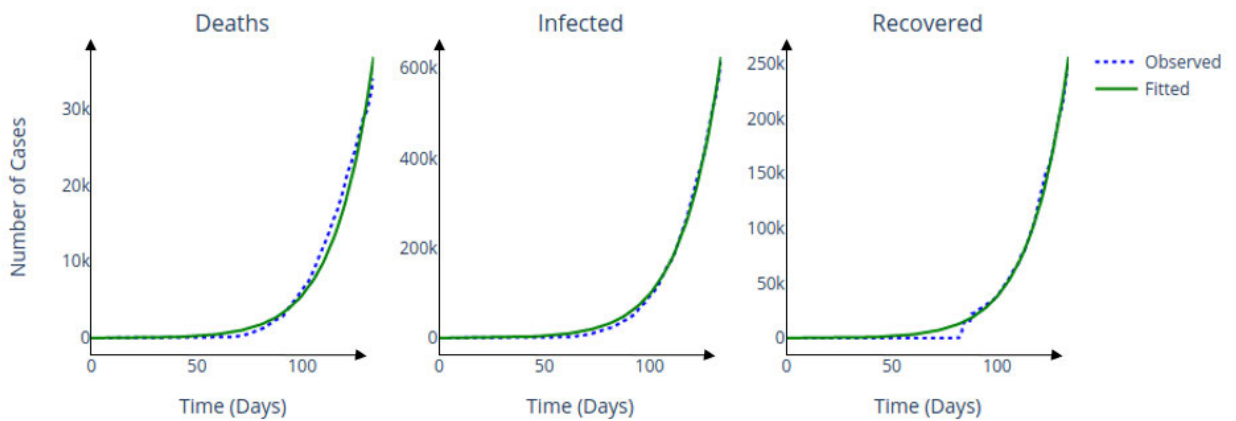


FIGURE 7. Brazil: Graphical validation of the LM algorithm using the exponential weight function.

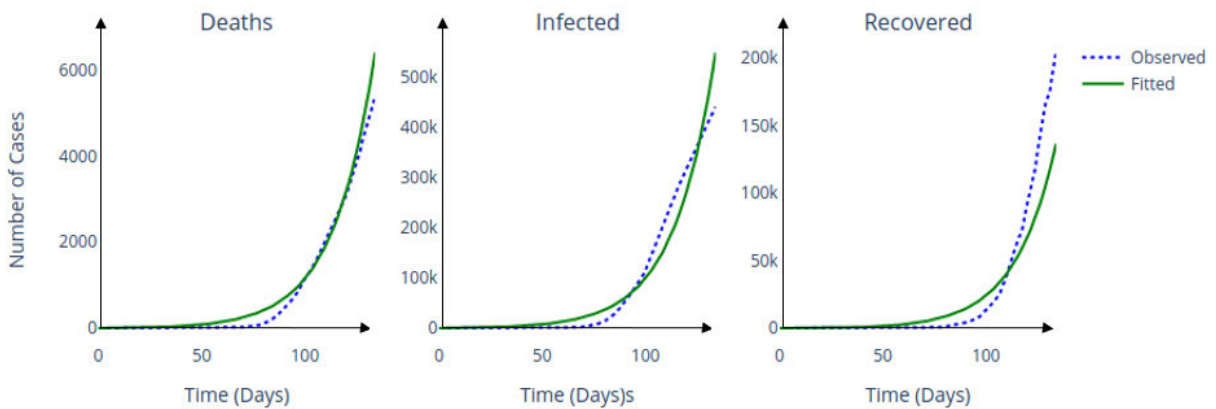


FIGURE 8. Russia: Fitting results of the BFGS algorithm using the exponential weight function.

LM and BFGS algorithms for all the considered metrics and the investigated countries. We notice that LM considerably outperforms the BFGS especially in Italy and then, Russia. However, both algorithms achieve similar results for the case of Brazil with a certain advantage for the BFGS in terms of MAE. In general, the LM algorithm provides much more accurate results than the BFGS.

2) THE CASE OF USA

In this study, we consider the USA as a special case. Indeed, unlike other countries which are controlled by a single authority, the USA is composed of fifty different states. Each one of them is governed independently of the other and in a different manner. For example, some states are imposing the lockdown, e.g., NY and NJ, and the stay home order at the beginning of

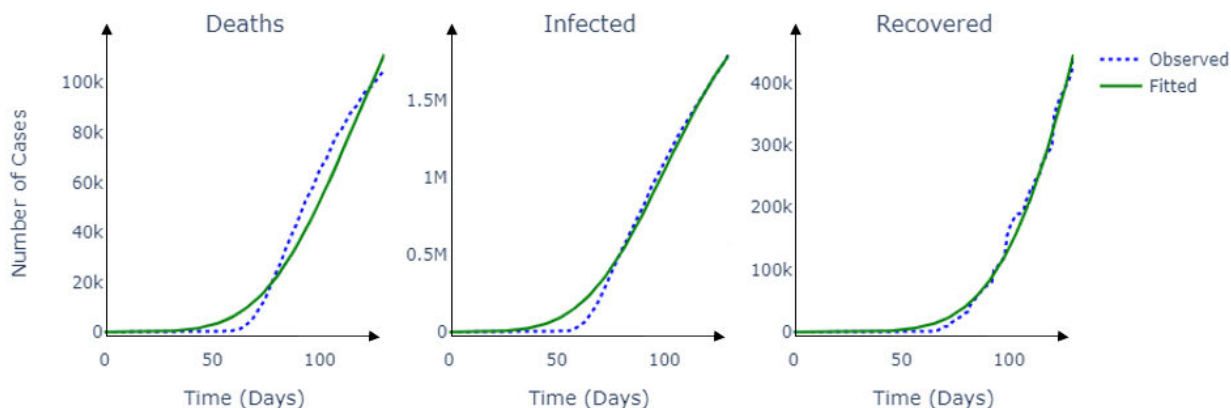


FIGURE 9. USA (Phase I): Graphical validation of the LM algorithm using the exponential weight function for the first period of the pandemic (January 22nd, 2020 till June 13th, 2020).

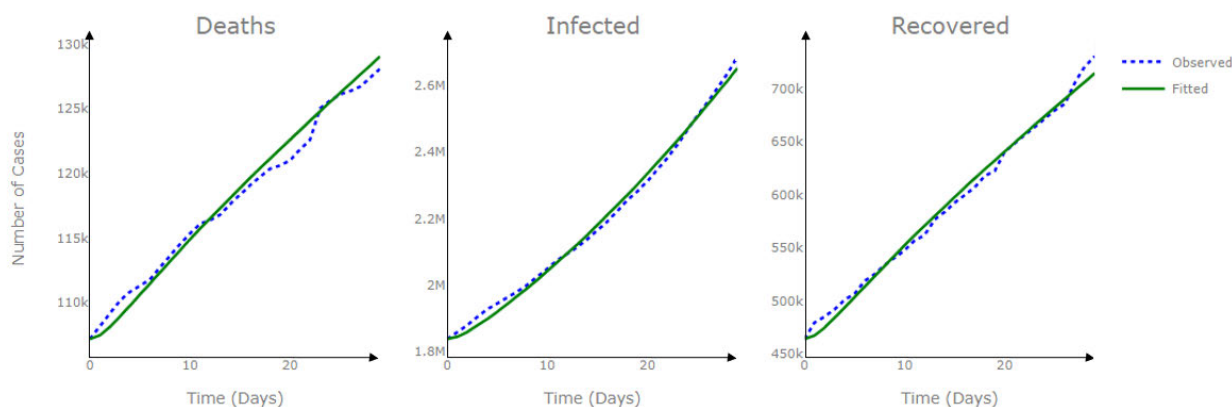


FIGURE 10. USA (Phase II): Graphical validation of the LM algorithm using the exponential weight function for the first period of the pandemic (June 13th, 2020 till July 7th, 2020).

TABLE 5. Comparison between the LM and BFGS fitting algorithms using the exponential weight function for Russia, Italy, and Brazil.

Country	Algorithm	MAE	NRMSE	SMAPE	MRAE	R^2
Russia	BFGS	488155.78	1.24	1.5	1.72	0.857
	LM	44215.39	0.21	1.22	0.58	0.947
Brazil	BFGS	45801.68	0.122	1.04	0.284	0.994
	LM	62419.17	0.277	1.12	0.34	0.995
Italy	BFGS	122355.74	2.45	2.7	1.24	0.116
	LM	9769.24	1.81	1.3	0.94	0.951

the pandemic (early March) while others do not or impose it tardily, e.g., FL. Moreover, the Black lives matter protests on different US states were unpredictable and their impact on the spread can not be measured and/or characterized and hence, this can have a huge impact on the spread of the COVID-19. Hence, treating the US as a single entity is not appropriate and may lead to inaccurate results due to the heterogeneity of the states. Therefore, one possible option is to investigate each

state separately as a single homogeneous entity. However, due to the non-availability of the data in our database that we used with other countries, we do not include these results in this study. In fact, we propose to deal with the USA case by dividing the fitting period into two periods or phases so that the model will be aware that two waves exist, one dominated by the northeastern states, NY, NJ, and CT, during the first four months of the pandemic and the other is dominated by southern states such as TX, FL, and CA starting from June 13th. In this scenario, we can obtain relatively accurate fitting results as shown in Fig. 9 and 10 with $R^2 = 0.992$ and $R^2 = 0.993$, respectively, as given in Table 6. In the latter table, we also provide the fitting results for the whole period (Phase I and Phase II combined) where two COVID-19 waves are registered. As explained earlier, the fitting was not successful with a poor $R^2 = 0.36$. Nevertheless, along the two periods, the model successfully fits the real data and shows acceptable results despite the previously mentioned characteristics of the studied country. In other words, for each period, different parameters of the mechanistic model are

TABLE 6. USA: Numerical validation of the LM fitting algorithm using the exponential weight function.

Country	MAE	NRMSE	SMAPE	MRAE	R ²
USA (Phase I)	84589.82	0.425	1.09	0.59	0.992
USA (Phase II)	22141.24	0.312	0.94	0.46	0.993
USA (Phase I & II)	103257.24	2.877	1.79	1.83	0.363

estimated unlike the other studied countries in the previous sections. Hence, for the USA, we only prove that our model is able to fit the reported cases representing the spread for these two periods. Otherwise, combining the two periods will require the adoption of fitful model parameters that vary with time, e.g., $\beta(n)$ instead of constant β . This study is more elaborate and will be the focus of our future extension of this work.

The estimated parameters of the model of each country are provided in Table 7. These values are used to forecast the future evolution of the pandemic in the next section.

TABLE 7. Estimated model parameters for each investigated country.

Parameters	Russia	Italy	Brazil	USA (Phase I)	USA (Phase II)
β	1.279	1.248	1.039	1.436	1.482
δ	0.041	0.050	0.017	0.043	0.0513
ρ	0.002	0.014	$1.5 \cdot 10^{-7}$	$1.7 \cdot 10^{-5}$	$1.8 \cdot 10^{-6}$
μ	$1.5 \cdot 10^{-4}$	0.015	$2.1 \cdot 10^{-4}$	0.004	0.001
λ	0.011	0.149	0.099	0.048	0.015
χ	0.002	$4 \cdot 10^{-12}$	$2.1 \cdot 10^{-6}$	0.001	0.003
α	0.349	0.077	0.835	0.347	1.968
σ	0.135	0.135	0.144	0.126	0.125
ϵ	1.237	0.568	$7.05 \cdot 10^{-5}$	1.527	0.909
ν	0.007	$5 \cdot 10^{-7}$	0.076	$7 \cdot 10^{-4}$	$2.5 \cdot 10^{-6}$
τ	0.016	0.023	0.002	0.020	0.025

C. EXPECTED SPREAD AND FORECASTING VALIDATION

In this section, we extrapolate the COVID-19 spread in the studied countries using our fitted models to forecast its evolution during the next period. We also discuss in details the evolution of each state. In addition, we provide the upper and lower bounds of the expected spread representing the worst and best cases scenarios for these countries based on our model. We particularly select to investigate the forecasting results of Russia and Italy as they are already reach their peaks unlike Brazil. The case of USA is also omitted due to the inaccurate fitting results that we discussed the reasons earlier. Finally, we compare the forecasting performance of our model to the one of the RW model.

The forecasting per state, for Russia and Italy, are shown in Figs. 11 and 12, respectively. In these figures, the total sick represents the sum of the population of the states E , I , Q , and H , that are already contaminated with the virus. As for the confirmed cases, it stands for all the reported cases represented by the states I , H , and Q . The fitting model based on the current spread parameters expects that the cumulative number of infected cases will reach about

seven million cases and 250 thousands in Russia and Italy, respectively. The forecast results for Italy show that the infected cases are being reported as soon as the exposed cases start to appear as it shown by evolution of the pink (exposed cases) and light green curves (confirmed cases) of Fig. 12. Moreover, the big gap between the confirmed cases and the exposed cases (confirmed cases is always higher than exposed cases) reflects that the country is efficiently identifying the infected cases, which proves the effectiveness of the prevention approach employed by this country to avoid higher propagation of the virus. However, for Russia, Fig. 11 shows that the number of exposed cases is more important than the number of confirmed cases which explains the huge number of cumulative cases that are expected to be reported by the end of the forecast period. Nevertheless, these results may change in practice. The local authorities, in the case of Russia, may impose stricter policies which may limit the propagation of the spread. Recall that our forecast are based on the real data observed in each country and hence, the forecast will extrapolate the evolution of the model states based on the spread data. Hence, if a country is not adopting an effective strategy at the beginning of the pandemic, the estimated results might be relatively critical as it was expected by our model for Russia.

Also, in Figs. 11 and 12, we observe that the majority of confirmed cases are at the quarantine or hospitalized state, i.e., either being quarantined at home (red curve) or being quarantined at hospitals (black curve), which proves the effectiveness of the stay home orders taken in these countries and their policies to limit the spread. However, we can notice that, in Italy, the quarantining process starts around 45 days since the beginning of the training period, which is around March 10th the day when the stay home order was announced). In Russia, the lockdown starts around 75 days (end of March) and the number of cases is expected to be very high if no actions are taken to stop the spread. By the start of 2021 we expect that few infected cases will be reported and the spread will be significantly reduced for both countries.

Since our fitting model is trained during the first four months of the pandemic spread, the forecast results may not be accurate in practice. Indeed, several policies can be further ordered and social distancing will be much better applied and hence, these actions may slow down the spread. In other scenarios, people will not conveniently apply the COVID-19 prevention guidelines anymore, which may speed up the propagation of the virus. Therefore, in Figs. 13 and 14, we provide the uncertainty regions that may characterize the COVID-19 spread in terms of number of active cases and number of deaths. The objective is to use the parameters estimated by our model and vary the infection rate, β , that reflects the daily number of contaminated person that an infected person can infect. By varying this parameter, we can evaluate the best and worst case scenarios that may occur for these countries. In Figs. 13 and 14, we also provide the observed results during the seventeen days right after the fitting period (fit/forecast split in the figure). The objective is to show that the fitting and the resulting forecast are

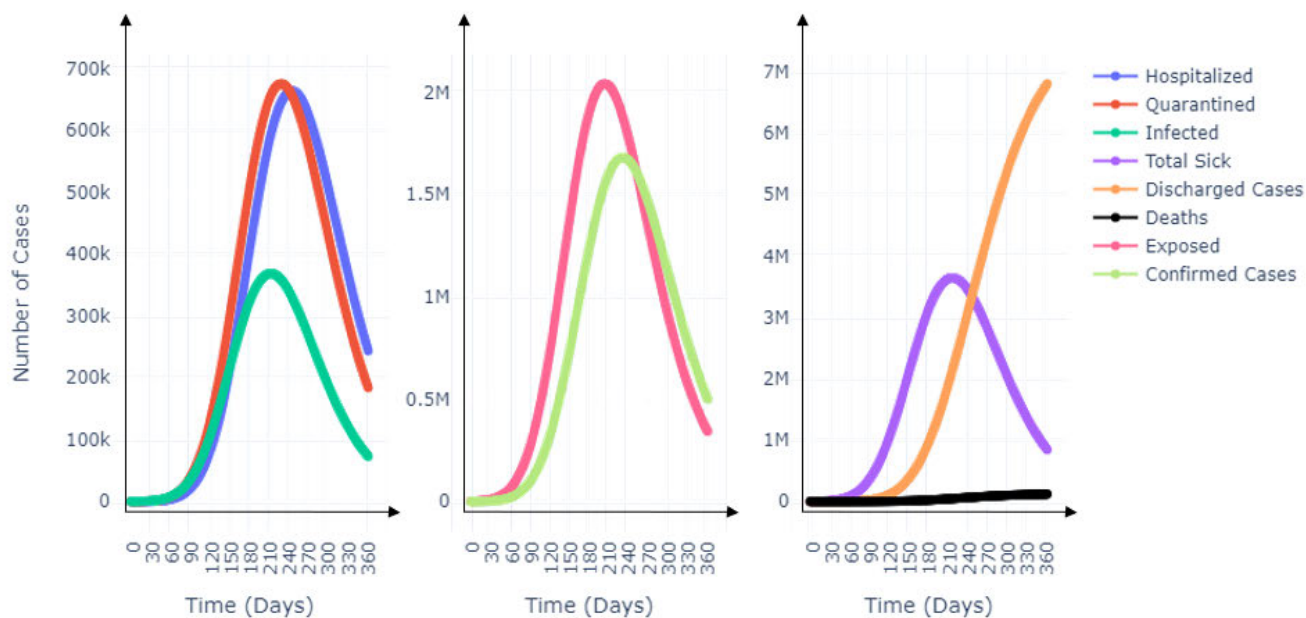


FIGURE 11. Russia: COVID-19 spread forecasting results and expected evolution of the different states based on the estimated parameters.

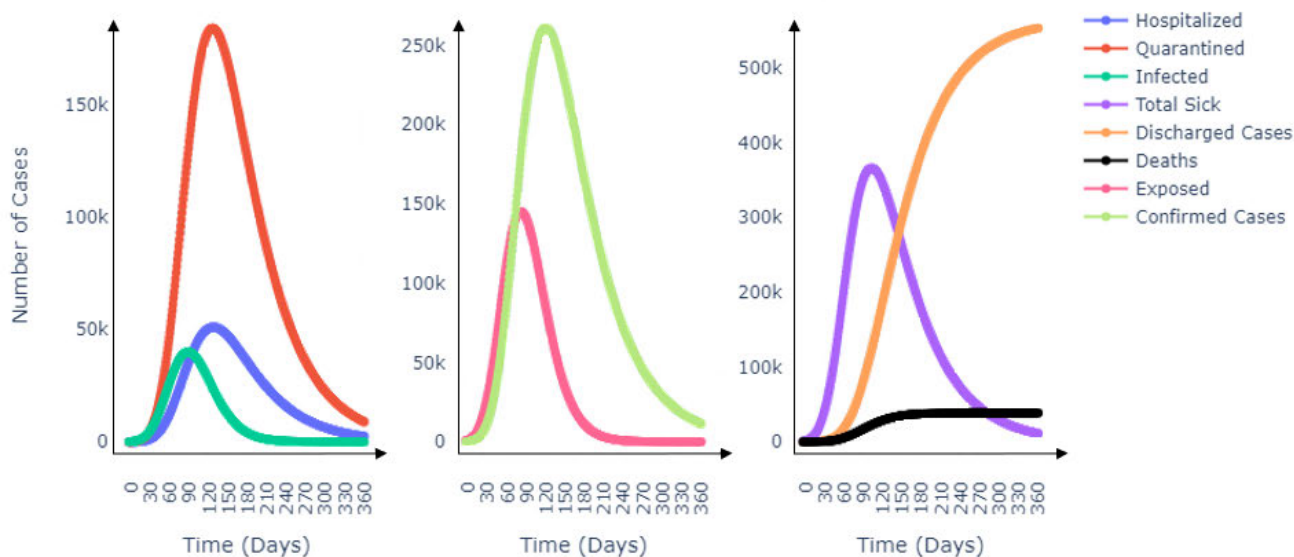


FIGURE 12. Italy: COVID-19 spread forecasting results and expected evolution of the different states based on the estimated parameters.

accurate and have the same trends as it is observed in reality in terms of reported infected and deceased cases. Regarding the expected evolution, we notice that for Italy the number of active cases will significantly diminish and the number of deceased persons will reach its maximum and slightly increases. However, for Russia, if no actions are taken by the authorities and officials, our model expects that the number of infected cases (new cases plus active cases) will keep increasing until reaching a peak in August and then, it will

start decreasing significantly. Moreover, the number of deaths is expected to continue growing and slowing down starting from September.

In the sequel, we provide two comparisons of our proposed model with a reference model, namely the RW approach with drift algorithm. The comparison consists of forecasting the data for the next 31 days right after the fitting period. We compare the results graphically and numerically to evaluate the efficiency of the proposed model. This comparison targets

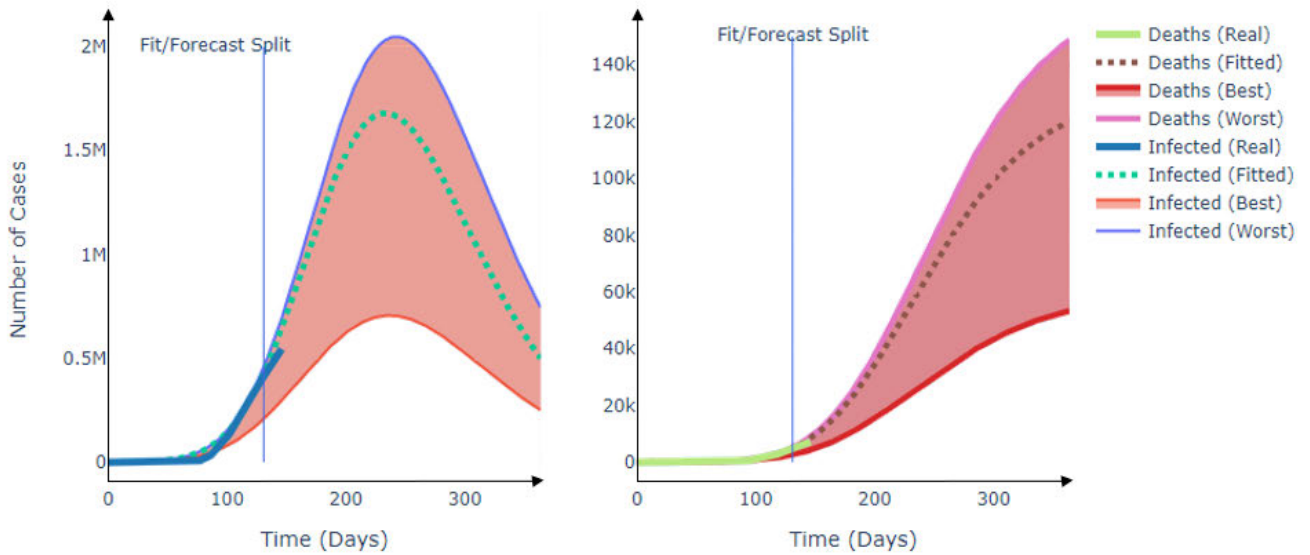


FIGURE 13. Russia: Best and worst possible spread scenarios by varying the infection rate β .

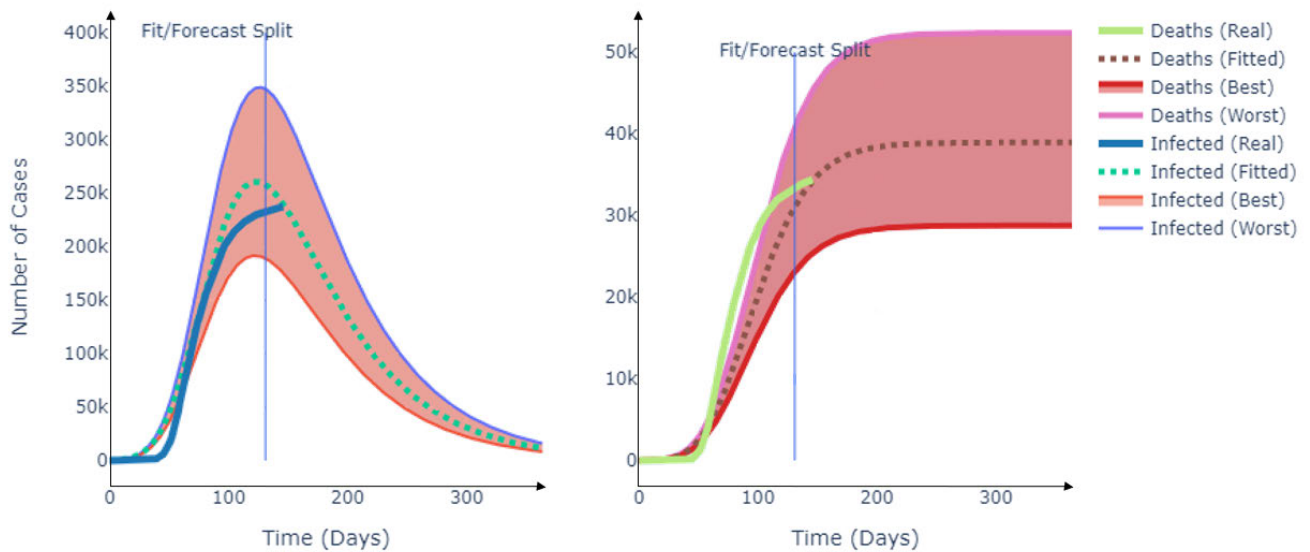


FIGURE 14. Italy: Best and worst possible spread scenarios by varying the infection rate β .

only the infected and deaths states due to the lack of data for the other states, which prevents us to build their respective RW models.

Initially, the RW model is built using the observed data during the fitting period. Then, it is used to forecast the evolution of the number of active and deceased cases. We have generated two random forecasting attempts that we denoted by RW model (Test I) and RW model (Test II). In addition to that, we have provided the forecast based on the estimated mean (RW mean) and the forecasts based on a 75% prediction interval, named RW upper bound and RW lower bound where there is 75% chance that the predicted observations will fall within their ranges.

Figs. 15 and 16 plot the forecast number of deceased and infected cases for Russia and Italy, respectively using our proposed model (with exponential weights) and the RW model. In each figure, six curves are presented with different colors: the observed data during the whole period (red), our proposed model during the fitting and forecasting periods (black), the RW mean (yellow), the RW model (Test I) and (Test II) (black and brown, respectively), and finally, the 75% prediction interval identified by the two green limits. The curves related to the RW model are plotted during the fitting period. The split between the fitting and forecasting periods are indicated by the purple vertical line. For fairness reasons, we train the models (proposed and RW) over the same fitting

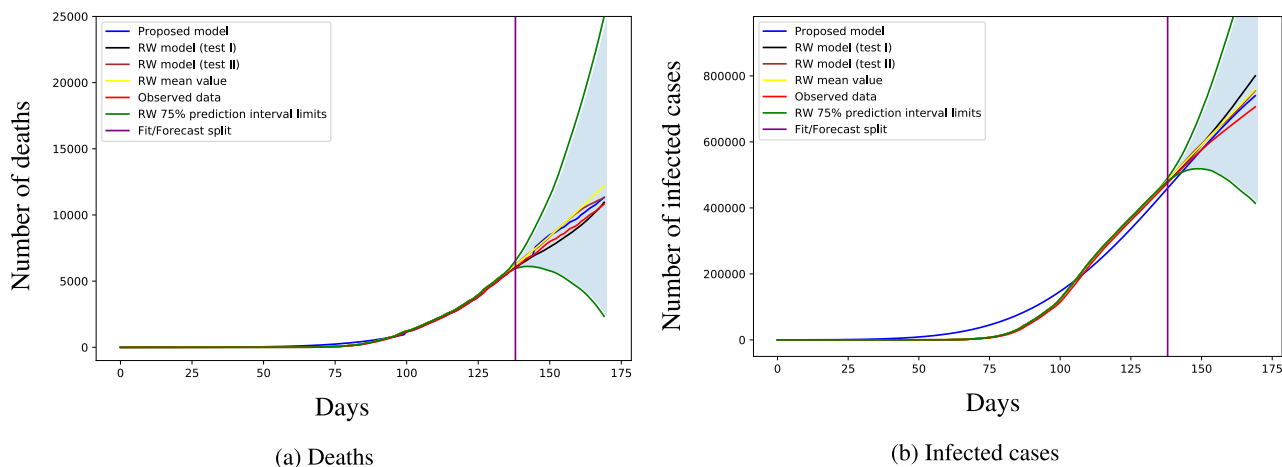


FIGURE 15. Russia: Forecast of the infected and deceased cases. Comparisons between the proposed and the RW models.

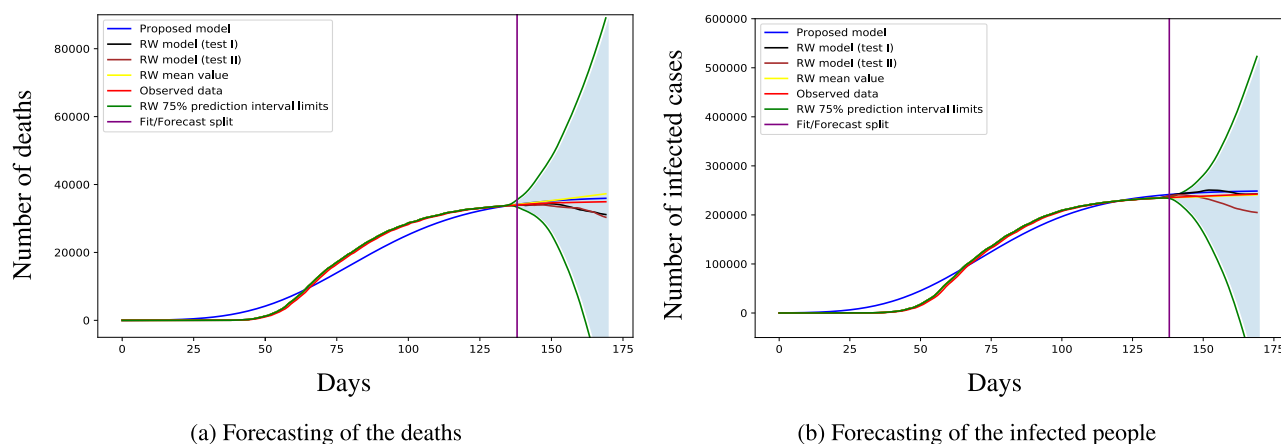


FIGURE 16. Italy: Forecast of the infected and deceased cases. Comparisons between the proposed and the RW models.

period (135 days) and investigating their prediction performance for a forecasting period of 31 days.

The numerical results related to Figs. 15 and 16 are given in Tables 8 and 9, respectively, where the NRMSE and SMAPE metrics for each forecasting attempt are provided. The results show that the RW model has a wide range of possible manifestations due to the huge variation of the observed data during the fitting phase. A random attempt to forecast the observed data may not necessary provide accurate results

as shown in the Tables. For instance, in terms of NRMSE for the case of Russia, our model, that provides a deterministic forecast, achieves NRMSE values equal to 0.99 and 0.066 for the infected and deceased, while the RW attempts are quite far from these results as indicated in Table 8. Through these tables, we also investigate the forecasting results using two different weight functions: the uniform (Uni.), no weights and the exponential (Exp.) weights. It is shown that NRMSE and SMAPE values using the exponential weights are slightly

TABLE 8. Russia: Numerical validation of the forecast results.

Investigated states	Infected		Deaths	
	NRMSE	SMAPE	NRMSE	SMAPE
RW model (Test I)	0.42	0.18	0.134	0.044
RW model (Test II)	0.27	0.122	0.411	0.143
Proposed model (Exp.)	0.099	0.052	0.066	0.0304
Proposed model (Uni.)	0.19	0.075	0.084	0.051
RW mean	0.226	0.121	0.175	0.068
RW upper bound	2.188	0.680	0.744	0.350
RW lower bound	1.935	1.312	1.716	0.447

TABLE 9. Italy: Numerical validation of the forecast results.

Investigated states	Infected		Deaths	
	NRMSE	SMAPE	NRMSE	SMAPE
RW model (Test I)	19.05	0.36	3.515	0.145
RW model (Test II)	1.566	0.045	6.28	0.363
Proposed model (Exp.)	0.80	0.024	0.457	0.022
Proposed model (Uni.)	0.91	0.038	0.62	0.035
RW mean	2.47	0.063	0.121	0.006
RW upper bound	66.75	0.742	38.68	1.421
RW lower bound	84.197	1.545	36.279	0.714

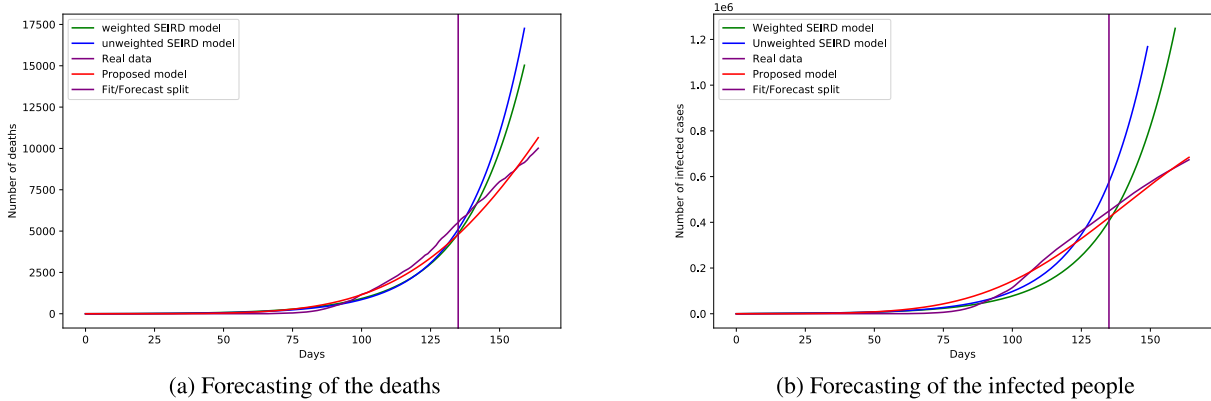


FIGURE 17. Russia: Forecast of the infected and deceased cases. Comparisons between the proposed and two versions of an SEIRD model.

lower than the ones obtained with the uniform weight functions. This result is due to the more accurate fitting obtained with the exponential weights that guarantees a certain correlation between the forecast observations and the fitted ones.

However, if we consider the mean value, which is very representative of the RW model, we can notice that our proposed model outperforms in most of the cases (e.g., infected cases for Italy and infected and deceased cases for Russia). In many cases, the RW model fails in accurately predicting the observed data due to its wide prediction interval even if the RW mean is showing accurate results. On the contrary, the deterministic prediction of our model has efficiently predicted the future pandemic spread.

The achieved metrics given in Tables 8 and 9 clarify and corroborate the graphical results and proves that our approach effectively outperforms the RW model in most of the cases.

D. MODEL COMPARISON WITH THE SEIRD MODEL

In this section, we compare our proposed model to a simple model with five states only (Susceptible, Exposed, Infected, Recovered, Deaths) as described in [58] and defined in this paper as the SEIRD model. The graphical results shown in Fig. 17 present the forecasting attempt of the observed data using our proposed model in addition to weighted and unweighted versions of the SEIRD model. At the training phase, we can notice that the three models perform, approximately, with the same level of efficiency. On the contrary, at the forecasting phase, we can easily notice that our model outperforms the two versions of the SEIRD model. In fact, our proposed model follows the trend of the observed data during the forecasting period. Additionally, from these figures, we can deduce the effect of the weights on the forecast performance. In fact, the forecast of the number of infected cases proves that the weighting technique helps in improving the forecast of the observed data but it will reduce the efficiency of the fitting. Indeed, if the model is trained to perfectly follow the training data, then the model will over-fit the training data and it will perform badly during the forecast. Table 10 contains the NRMSE and SMAPE metrics, defined

TABLE 10. Russia: Numerical validation of the forecast results.

Investigated States Metrics	Deaths		Infected	
	NRMSE	SMAPE	NRMSE	SMAPE
Weighted SEIRD (Exp.)	0.115	1.0825	0.401	1.68
Unweighted SEIRD	0.161	1.0824	0.400	1.64
Proposed model	0.032	1.0419	0.036	1.05

in Section III-C2, of the proposed model as well as the weighted and unweighted versions of the SEIRD model. The numerical results in Table 10 confirm the graphical results and prove that the outperformance of the proposed model compared to the two investigated versions of the SEIRD model. For instance, in terms of NRMSE, the proposed model achieves 0.032 while the SEIRD model cannot reach below 0.11 for the forecast deaths. Similar observations are noticed for the SMAPE metric and the forecast infected cases as well.

V. LIMITATION OF THE STUDY

In this study, we have proposed a mechanistic model for assessing and forecasting the spread of the COVID-19 pandemic. The pandemic of interest is very novel and even though it has already spread all over the world, scientists still do not have enough knowledge to model it properly, especially due to the lack of data and evidence about its contamination process. In this study, we have made some assumptions to conceive our model, which may impact its performance. First, we have considered a closed system, which does not distinguish between local and non-local cases. Hence, the spread is assumed to be originated from a first local case or a first imported case. At the time of this study, there is no available data that provide accurate and precise statistics about non-local cases and whether they are responsible for the virus spread or not. Therefore, the model only considers the local spread of the virus and discards any cases imported from outside of the country of interest. The study also assumes long-term immune memory. In other words, cases that are recovered are assumed to be immunized against

the virus and hence, can not be contaminated again. Indeed, the available data set does not provide detailed information about these scenarios and there is no strong evidence for short-term immune memory.

Another future improvement of this study could be the concrete consideration of asymptomatic cases [59]. In this study, we have assumed that every case is symptomatic at least for a short period of time, e.g., 1 day, and hence, each infected case must pass by a quarantine state even for a short period. Statistically, around 30% of patients who tested positive were asymptomatic and may spread the virus. In this study, we did not distinguish between these the asymptomatic and symptomatic cases due to the absence of related data for each of the investigated countries. Therefore, we have supposed that the COVID-19 symptoms with different degrees (light to severe) will appear on all the infected cases and will force them to stay home, i.e., quarantined for at least one day before being totally healed from the diseases.

This study is based on an ODE system with non-variant parameters and especially a fixed infection rate β . By considering that an infected person will be placed on quarantine before its final recovery will allow our model to learn that an infected case in state I will not affect people daily unlike typical SIR and SEIR models. This assumption will make our model closer to reality by allowing it to automatically learn that an infected person may or may not infect people daily and will help, when needed, determine accurate basic reproduction number values regardless of the constant parameters. Other works are considering a similar scenario where recovery is only possible through quarantine can be found in [37]. However, a more elaborate and accurate model would consider a time-varying infection rate. In that case, more precise results could be obtained.

Finally, our model assumes that the investigated countries have centralized governmental systems. Hence, the human decisions, such as border openings, lock-down, indoor services, and mask mandates, and their impacts on the disease spread have uniform consequences. As a result, the prediction of the future evolution of the pandemic within federal countries, such as the USA, can not be accurately performed with our model due to the fact that every state has its own regulations and policies and hence, it is more convenient to investigate each state separately and devise a generalized model for the whole country.

All these assumptions have made our forecasting results valid for a short-period of time, e.g., 30 days as shown by our forecasting results. Our study does not claim long-term prediction of the disease spread unless similar human factors and management practices are maintained.

VI. CONCLUSION

In this study, we developed a mechanistic model composed of eight states to characterize the COVID-19 spread in different countries. An ODE system is formulated to mathematically model the interactions between the different states. A curve fitting using real world observed data sets is developed to

determine accurate fitting solutions and estimate the parameters of the ODE model. We use the LM algorithm to solve and estimate the ODE system parameters. We have shown that the employed algorithms outperform a reference model, i.e., BFGS. The fitting technique is represented along with the model that aims to endorse its forecasting results by focusing on the last trends of the data, and hence, giving accurate results that are on the verge of these trends. Multiple countries with different spread trends are investigated with the proposed work. We reveal that our model is able to fit the real world data of different countries showing different spread trends, namely, Russia, Brazil, Italy, and USA. Despite of the challenges and the various human factors, i.e., social distancing, quarantine, wearing masks that are not directly characterised by our model and that may affect the spread and its evolution. It is shown that accurate forecasting results are obtained, e.g., Italy and Russia.

The proposed model is not only one of the rarest extended mechanistic models that include all the possible eight states that an individual may go through during the illness cycle but also it is the only model that employs a weighted fitting technique to follow the latest trends in addition that the proposed model is trained on three target data sets. However, this work can be endorsed by the employment of variable parameters that change over the time to better fit the variations of the data and hence, the challenges related to the case of USA can be addressed. Moreover, once more detailed data becomes available, it will be very interesting and challenging to extend our model and investigate more complicated and elaborate scenarios, e.g., considering the re-contamination of healed cases and incoming non-local cases.

REFERENCES

- [1] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. Zhu, Y. Tai, C. Bai, T. Gao, J. Song, P. Xia, J. Dong, J. Zhao, and F.-S. Wang, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *Lancet Respiratory Med.*, vol. 8, no. 4, pp. 420–422, Apr. 2020.
- [2] A. Nasir, K. Shaukat, I. A. Hameed, S. Luo, T. M. Alam, and F. Iqbal, "A bibliometric analysis of corona pandemic in social sciences: A review of influential aspects and conceptual structure," *IEEE Access*, vol. 8, pp. 133377–133402, Jul. 2020.
- [3] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, and X. Xing, "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia," *New England J. Med.*, vol. 382, no. 13, pp. 1199–1207, Mar. 2020.
- [4] N. Van Doremalen, T. Bushmaker, D. H. Morris, M. G. Holbrook, A. Gamble, B. N. Williamson, A. Tamin, J. L. Harcourt, N. J. Thornburg, S. I. Gerber, and J. O. Lloyd-Smith, "Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1," *New England J. Med.*, vol. 382, no. 16, pp. 1564–1567, 2020.
- [5] P. Staszekiewicz, I. Chomiak-Orsa, and I. Staszekiewicz, "Dynamics of the COVID-19 contagion and mortality: Country factors, social media, and market response evidence from a global panel analysis," *IEEE Access*, vol. 8, pp. 106009–106022, Jun. 2020.
- [6] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study," *Lancet*, vol. 395, no. 10225, pp. 689–697, Feb. 2020.
- [7] F. Song, N. Shi, F. Shan, Z. Zhang, J. Shen, H. Lu, Y. Ling, Y. Jiang, and Y. Shi, "Emerging 2019 novel coronavirus (2019-nCoV) pneumonia," *Radiology*, vol. 295, no. 1, pp. 210–217, Apr. 2020.

- [8] H. Shi, X. Han, N. Jiang, Y. Cao, O. Alwalid, J. Gu, Y. Fan, and C. Zheng, "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: A descriptive study," *Lancet Infectious Diseases*, vol. 20, no. 4, pp. 425–434, Apr. 2020.
- [9] "Coronavirus disease 2019 (COVID-19): Situation report of the 1st July 163," World Health Org., Geneva, Switzerland, Tech. Rep. 163, 2020. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200701-covid-19-sitrep-163.pdf?sfvrsn=c202f05b_2
- [10] V. Chamola, V. Hassija, V. Gupta, and M. Guizani, "A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact," *IEEE Access*, vol. 8, pp. 90225–90265, May 2020.
- [11] R. Y. Kim, "The impact of COVID-19 on consumers: Preparing for digital sales," *IEEE Eng. Manag. Rev.*, vol. 48, no. 3, pp. 212–218, Sep. 2020.
- [12] Z. R. Alashhab, M. Anbar, M. M. Singh, Y.-B. Leau, Z. A. Al-Sai, and S. A. Alhayja'a, "Impact of coronavirus pandemic crisis on technologies and cloud computing applications," *J. Electron. Sci. Technol.*, 2020, Art. no. 100059, doi: 10.1016/j.jnlest.2020.100059.
- [13] G. Sertsov, N. Wilson, M. Baker, P. Nelson, and M. G. Roberts, "Key transmission parameters of an institutional outbreak during the 1918 influenza pandemic estimated by mathematical modelling," *Theor. Biol. Med. Model.*, vol. 3, no. 1, p. 38, Nov. 2006.
- [14] A. Flahault, E. Vergu, L. Coudeville, and R. F. Grais, "Strategies for containing a global influenza pandemic," *Vaccine*, vol. 24, nos. 44–46, pp. 6751–6755, Nov. 2006.
- [15] A. K. Lugnér, M. van Boven, R. de Vries, M. J. Postma, and J. Wallinga, "Cost effectiveness of vaccination against pandemic influenza in European countries: Mathematical modelling analysis," *BMJ*, vol. 345, p. 4445, Jul. 2012.
- [16] Pandemic Influenza Outbreak Research Modelling Team (Pan-InfORM), "Modelling an influenza pandemic: A guide for the perplexed," *Can. Med. Assoc. J.*, vol. 181, nos. 3–4, pp. 171–173, Jul. 2009.
- [17] D. K. Menon, B. L. Taylor, S. A. Ridley, and U. on behalf of the Intensive Care Society, "Modelling the impact of an influenza pandemic on critical care services in England," *Anaesthesia*, vol. 60, no. 10, pp. 952–954, Oct. 2005.
- [18] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, and R. M. Eggo, "Early dynamics of transmission and control of COVID-19: A mathematical modelling study," *Lancet Infectious Diseases*, vol. 20, pp. 553–558, May 2020.
- [19] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, and N. Davies, "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study," *Lancet Public Health*, vol. 5, no. 5, pp. e261–e270, May 2020.
- [20] Q. Lin, S. Zhao, D. Gao, Y. Lou, S. Yang, S. S. Musa, M. H. Wang, Y. Cai, W. Wang, L. Yang, and D. He, "A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action," *Int. J. Infectious Diseases*, vol. 93, pp. 211–216, Apr. 2020.
- [21] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. D. Filippo, A. D. Matteo, and M. Colaneri, "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy," *Nature Med.*, vol. 26, no. 6, pp. 855–860, Jun. 2020.
- [22] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, "Data-based analysis, modelling and forecasting of the COVID-19 outbreak," *PLoS ONE*, vol. 15, no. 3, pp. 1–21, Mar. 2020.
- [23] S. L. Chang, N. Harding, C. Zachreson, O. M. Cliff, and M. Prokopenko, "Modelling transmission and control of the COVID-19 pandemic in Australia," 2020, *arXiv:2003.10218*. [Online]. Available: <http://arxiv.org/abs/2003.10218>
- [24] G. Chowell, "Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts," *Infectious Disease Model.*, vol. 2, no. 3, pp. 379–398, Aug. 2017.
- [25] G. Chowell, L. Sattenspiel, S. Bansal, and C. Viboud, "Mathematical models to characterize early epidemic growth: A review," *Phys. Life Rev.*, vol. 18, pp. 66–97, Sep. 2016.
- [26] B. Pell, Y. Kuang, C. Viboud, and G. Chowell, "Using phenomenological models for forecasting the 2015 Ebola challenge," *Epidemics*, vol. 22, pp. 62–70, Mar. 2018.
- [27] K. Wu, D. Darcet, Q. Wang, and D. Sornette, "Generalized logistic growth modeling of the COVID-19 outbreak: Comparing the dynamics in the 29 provinces in China and in the rest of the world," *Nonlinear Dyn.*, vol. 101, no. 3, pp. 1561–1581, Aug. 2020, doi: 10.1007/s11071-020-05862-6.
- [28] C. Viboud, L. Simonsen, and G. Chowell, "A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks," *Epidemics*, vol. 15, pp. 27–37, Jun. 2016.
- [29] F. Yang, H. Liu, H. Qi, and X. Liu, "SEIR evolutionary simulation model of the infectious disease emergency," in *Proc. Int. Conf. Ind. Informat.-Comput. Technol., Intell. Technol., Ind. Inf. Integr. (ICIICII)*, Wuhan, China, Dec. 2016, pp. 315–318.
- [30] B. Xu, J. Cai, D. He, G. Chowell, and B. Xu, "Mechanistic modelling of multiple waves in an influenza epidemic or pandemic," *J. Theor. Biol.*, vol. 486, Feb. 2020, Art. no. 110070.
- [31] J. Lessler and D. A. T. Cummings, "Mechanistic models of infectious disease and their impact on public health," *Amer. J. Epidemiol.*, vol. 183, no. 5, pp. 415–422, Mar. 2016.
- [32] W. Chen, "A mathematical model of Ebola virus based on SIR model," in *Proc. Int. Conf. Informat.-Comput. Technol., Intell. Technol., Ind. Inf. Integr.*, Wuhan, China, Dec. 2015, pp. 213–216.
- [33] Y. Maki and H. Hirose, "Infectious disease spread analysis using stochastic differential equations for SIR model," in *Proc. 4th Int. Conf. Intell. Syst., Model. Simul.*, Bangkok, Thailand, Jan. 2013, pp. 152–156.
- [34] T. W. Ng, G. Turinici, and A. Danchin, "A double epidemic model for the SARS propagation," *BMC Infectious Diseases*, vol. 3, no. 1, p. 19, Sep. 2003.
- [35] T. S. Do and Y. S. Lee, "Modeling the spread of Ebola," *Osong Public Health Res. Perspect.*, vol. 7, no. 1, pp. 43–48, Feb. 2016.
- [36] E. Dantas, M. Tosin, and A. Cunha, "Calibration of a SEIR–SEI epidemic model to describe the Zika virus outbreak in Brazil," *Appl. Math. Comput.*, vol. 338, pp. 249–259, Dec. 2018.
- [37] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of COVID-19 in China by dynamical modeling," 2020, *arXiv:2002.06563*. [Online]. Available: <http://arxiv.org/abs/2002.06563>
- [38] J. Yawney and S. A. Gadsden, "A study of the COVID-19 impacts on the Canadian population," *IEEE Access*, vol. 8, pp. 128240–128249, Jul. 2020.
- [39] A. Hauser, M. J. Counotte, C. C. Margossian, G. Konstantinoudis, N. Low, C. L. Althaus, and J. Riou, "Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe," *PLOS Med.*, vol. 17, pp. 1–17, Jul. 2020.
- [40] R. Gopal, V. K. Chandrasekar, and M. Lakshmanan, "Dynamical modelling and analysis of COVID-19 in India," 2020, *arXiv:2005.08255*. [Online]. Available: <http://arxiv.org/abs/2005.08255>
- [41] Y. Zhang, Y. Li, B. Yang, X. Zheng, and M. Chen, "Risk assessment of COVID-19 based on multisource data from a geographical viewpoint," *IEEE Access*, vol. 8, pp. 125702–125713, Jun. 2020.
- [42] F. Rustam, A. A. Reshi, A. Mehmood, S. Ullah, B.-W. On, W. Aslam, and G. S. Choi, "COVID-19 future forecasting using supervised machine learning models," *IEEE Access*, vol. 8, pp. 101489–101499, May 2020.
- [43] A. Lojić Kapetanović and D. Poljak, "Modeling the epidemic outbreak and dynamics of COVID-19 in Croatia," 2020, *arXiv:2005.01434*. [Online]. Available: <http://arxiv.org/abs/2005.01434>
- [44] K. Linka, M. Peirlinck, and E. Kuhl, "The reproduction number of COVID-19 and its correlation with public health interventions," *MedRxiv*, Jul. 2020. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7359536/>, doi: 10.1101/2020.05.01.20088047.
- [45] R. D. Kirkcaldy, B. A. King, and J. T. Brooks, "COVID-19 and postinfection immunity: Limited evidence, many remaining questions," *JAMA*, vol. 323, no. 22, pp. 2245–2246, Jun. 2020.
- [46] M. Small and D. Cavanagh, "Modelling strong control measures for epidemic propagation with Networks—A COVID-19 case study," *IEEE Access*, vol. 8, pp. 109719–109731, Jun. 2020.
- [47] C. Groendyke and A. Combs, "Modifying the network-based stochastic SEIR model to account for quarantine," Aug. 2020, *arXiv:2008.01202*. [Online]. Available: <https://arxiv.org/abs/2008.01202>
- [48] R. Gupta, G. Pandey, P. Chaudhary, and S. K. Pal, "SEIR and regression model based COVID-19 outbreak predictions in India," Apr. 2020, *arXiv:2004.00958*. [Online]. Available: <https://arxiv.org/abs/2004.00958>
- [49] S. X. D. Tan and L. Chen, "Real-time differential epidemic analysis and prediction for COVID-19 pandemic," Apr. 2020, *arXiv:2004.06888*. [Online]. Available: <https://arxiv.org/abs/2004.06888>

- [50] M. J. Mahmoodabadi, "Epidemic model analyzed via particle swarm optimization based homotopy perturbation method," *Informat. Med. Unlocked*, vol. 18, 2020, Art. no. 100293. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352914819303569>
- [51] A. Godio, F. Pace, and A. Vergnano, "SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence," *Int. J. Environ. Res. Public Health*, vol. 17, no. 10, p. 3535, May 2020.
- [52] D. Akman, O. Akman, and E. Schaefer, "Parameter estimation in ordinary differential equations modeling via particle swarm optimization," *J. Appl. Math.*, vol. 2018, pp. 1–9, Sep. 2018.
- [53] M. Kommenda, M. Affenzeller, G. Kronberger, and S. M. Winkler, "Non-linear least squares optimization of constants in symbolic regression," in *Computer Aided Systems Theory—EUROCAST 2013*, R. Moreno-Diaz, F. Pichler, and A. Quesada-Arencibia, Eds. Berlin, Germany: Springer, 2013, pp. 420–427.
- [54] G. Ramos-Llorden, G. Vegas-Sanchez-Ferrero, M. Bjork, F. Vanhevel, P. M. Parizel, R. S. J. Estepar, A. J. D. Dekker, and J. Sijbers, "NOVIFAST: A fast algorithm for accurate and precise VFA MRI T_1 mapping," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2414–2427, Nov. 2018.
- [55] M. Lourakis, "A brief description of the Levenberg-Marquardt algorithm implemented by Levmar," *Found. Res. Technol.*, vol. 4, pp. 1–6, Jan. 2005.
- [56] J. Chen and W. Li, "Convergence of Gauss-Newton's method and uniqueness of the solution," *Appl. Math. Comput.*, vol. 170, no. 1, pp. 686–705, Nov. 2005.
- [57] A. Botchkarev, "A new typology design of performance metrics to measure errors in machine learning regression algorithms," *Interdiscipl. J. Inf. Knowl. Manage.*, vol. 14, pp. 45–76, Jan. 2019.
- [58] E. L. Piccolomini and F. Zama, "Monitoring Italian COVID-19 spread by a forced SEIRD model," *PLoS ONE*, vol. 15, no. 8, pp. 1–17, Aug. 2020.
- [59] K. Kabir and J. Tanimoto, "Evolutionary game theory modeling to represent the behavioral dynamics of economic shutdowns and shield immunity in the COVID-19 pandemic," *Roy. Soc. Open Sci.*, vol. 7, pp. 1–8, Sep. 2020.



HAMDI FRIJI received the National Diploma of Engineering degree (Hons.) in information and communication technology from the Higher School of Communication of Tunis (SUP'COM), University of Carthage, Tunisia, in 2020. He worked as a Research Assistant with the Stevens Institute of Technology, Hoboken, NJ, USA, in 2020. His research interests include intersection of artificial intelligence, mathematical modeling, optimization, the Internet of Things, and graph theory.



RABY HAMADI received the Diplome National d'Ingenieur degree (Hons.) in information and communication technology engineering from the Ecole Supérieure des Communications de Tunis (SUP'COM), University of Carthage, Tunisia, in 2020. He was a Research Assistant with the Stevens Institute of Technology, Hoboken, NJ, USA, in 2020. His general research interests include intersection of artificial intelligence, finance, crowdsourcing, optimization, graph theory, mathematical modeling, and the Internet of Things.



HAKIM GHAZZAI (Senior Member, IEEE) received the Diplome d'Ingenieur degree (Hons.) in telecommunication engineering and the master's degree in high-rate transmission systems from the Ecole Supérieure des Communications de Tunis (SUP'COM), Tunis, Tunisia, in 2010 and 2011, respectively, and the Ph.D. degree in electrical engineering from KAUST, Saudi Arabia, in 2015. He was a Visiting Researcher with Karlstad University, Sweden, and a Research Scientist with the Qatar Mobility Innovations Center (QMIC), Doha, Qatar, from 2015 to 2018. He is currently a Research Scientist with the Stevens Institute of Technology, Hoboken, NJ, USA. He has authored over 120 papers in peer-reviewed journals and conferences. His general research interests include intersection of wireless networks, UAVs, the Internet of Things, intelligent transportation systems, and optimization. He is on the Editorial Board of the *IEEE COMMUNICATIONS LETTERS*, the *IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY*, and *Frontiers in Communications and Networks*.



HICHEM BESBES received the Diploma degree in electrical engineering, the Diplome d'Etudes Approfondies (DEA) (master's diploma) degree in systems theory, and the Ph.D. degree in electrical engineering from ENIT, University El Manar, Tunisia, in 1991, 1991, and 1999, respectively, and the Habilitation à Diriger des Recherches (HDR) degree in telecommunications from the Ecole Supérieure des Communications de Tunis (SUP'COM), in June 2005. He was the former Head of the Department of Applied Mathematics and Signal Processing, Higher School of Engineering in Communication in Tunisia (Sup'Com). He has more than 25 years of experience in higher education and ICT. He was a Postdoctoral Fellow of Concordia University, from 1999 to 2000, and a Visiting Scholar with Colorado State University, in 2011. He is currently a Professor in wireless communication with the University of Carthage, the former President, the Chairman, and a member of the Board of the National Telecommunication Regulatory Authority of Tunisia (INTT), for more than five years.



YEHIA MASSOUD (Fellow, IEEE) received the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA. He is currently the Dean of the School of Systems and Enterprises, Stevens University of Science and Technology, Hoboken, NJ, USA. He has authored over 300 papers in peer-reviewed journals and conferences. He has held several academic and industrial positions, including a member of the technical staff with the Advanced Technology Group, Synopsys, Inc., CA, USA, a tenured faculty with the Departments of Electrical and Computer Engineering and Computer Science, Rice University, Houston, TX, USA, the W. R. Bunn Head of the Department of Electrical and Computer Engineering, The University of Alabama, Birmingham, AL, USA, and the Head of the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA, USA. He has served as a Distinguished Lecturer by the IEEE Circuits and Systems Society and as an Elected Member of the IEEE Nanotechnology Council. He was selected as one of ten MIT Alumni Featured by MIT's Electrical Engineering and Computer Science Department, in 2012. He was a recipient of the Rising Star of Texas Medal, in 2007, the National Science Foundation CAREER Award, in 2005, the DAC Fellowship, in 2005, the Synopsys Special Recognition Engineering Award, in 2000, several best paper award nominations, and two best paper awards from the IEEE International Symposium on Quality Electronic Design, in 2007, and the IEEE International Conference on Nanotechnology, in 2011. He has served as an Editor of *Mixed-Signal Letters—The Americas* and also as an Associate Editor for the *IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS AND IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS*.

...