# Dynamic Modelling of Reported Covid-19 Cases and Deaths with Continuously Varying Case Fatality and Transmission Rate Functions

Mingdong Lyu1, Andrew Moore2, and Randolph Hall2

1 Mobility, Behaviour, and Advanced Powertrains Department, National Renewable Energy Laboratory, Denver, 80401, United States.

2 Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, 90089, United States. Correspondence should be addressed to Mingdong Lyu; mingdonl@usc.edu

**Keywords:** Infectious disease modeling; COVID-19; Time-varying parameters; SEIRD

## Abstract

This paper develops and applies an enhanced SEIRD (Susceptible-Exposed-Infectious-Recovered-Death) model with time-varying case fatality and transmission rates for the COVID-19 pandemic. Our aim is to accurately characterize time-variations in transmission and fatality rates relative to reported cases and deaths with a function that utilizes a small set of parameters. The time-varying functions, when integrated into the SEIRD model, efficiently characterize dynamic changes in fatality and transmission rates, which result from public health interventions, changes in medical care, changing human behaviour, and potential changes in the virus itself.

## Introduction

COVID-19 has challenged the world to react to a new contagious virus in the absence of effective medical treatment and vaccines. Over the course of the two years of the pandemic from the outbreak in December 2019, when the first cases were confirmed in Wuhan, China, until March 24th 2022, 213 countries and territories reported nearly 490 million confirmed cases and a death toll exceeding 6 million persons [1]. Waiting for effective clinical care and vaccination, countries reacted to the pandemic by controlling travel, implementing large-scale quarantine, restricting gatherings, requiring hygiene measures and screening for possible cases.

Shaw and Kennedy's examination of reproduction numbers (called the R value) [2], illustrates how rates of disease transmission can change over time as a consequence of changes in human behavior that alter rates of contact between infected and susceptible people, and alter probability of infection upon exposure. The authors also illustrate how rates of transmission (and hence reproduction numbers) depend on contact behavior within regions or localities. In each example,

variation in human behavior over time and space affects disease transmission and rates of new infections.

Compared to the global SARS epidemic in 2002 and MERS in 2012, COVID-19 has a relatively long incubation period, originally estimated to have a mean time of 5 days [3], [4]. COVID-19 was also found to be transmissible while individuals are asymptomatic. Meanwhile, disease severity is widely variable, depending on age, comorbidities, baseline health and access to care. Even those with mild or no symptoms, often young adults, may transmit the disease to others. These factors, combined with limited testing and inconsistent adherence to public health measures, made the virus impossible to contain. Policymakers are also facing a dilemma, balancing the goal of maintaining economic activity against saving lives through strict measures, learning about the effectiveness of interventions and the nature of the disease.

The paper aims to improve the understanding of how COVID-19 is spread by developing a variation of the Susceptible-Exposed-Infected-Recovered-Death (SEIRD) model. Our novel innovation is representing the transmission rate and case fatality rate as continuously varying Sigmoid functions of time. The functions are optimally fit to historical data on confirmed cases and deaths. The functions, when integrated into the SEIRD model, reflect the various factors that can affect the spread of COVID-19, such as the implementation of public health interventions, changes in medical care, changing human behavior, and potential changes in the virus itself, which can affect transmission rates and death rates. We applied our model to all 50 American states to derive insights into how the disease has spread in different localities, which is influenced by population health, disease exposure, localized public health interventions and messaging, in addition to other place specific factors.

## Prior Research

Prior research on COVID-19 has estimated disease-specific parameters, such as the basic reproduction number and latent period [5]–[9], demonstrating why the disease is highly transmissible. Mathematical models have also been used to analyze transmission scenarios for communicable disease and inform policy makers of possible futures and the effects of interventions. For example, according to a statistical guideline model published in 2015 [10], the state of New York reacted to the urgent shortage of ventilators by requesting more ventilators from the federal government and implementing new interventions, such as closures of schools and restaurants [11].

Another use of disease transmission models has been to predict and plan for future demands on the healthcare system, such as demands for hospital beds (ICU in particular) and needs for health care resources, such as ventilators. Toward that goal, [12] provides a statistical model of death data to predict future fatalities, assuming that social distancing measures are maintained. From

the projected fatality data, they estimated hospital utilization with an individual-level microsimulation model based on the historical statistics of age-specific ICU admission. [13] simulates the COVID-19 outbreak, parameterized with the US population demographics, with a compartmental model under different scenarios of self-isolation, projecting hospital utilization and recognizing the mitigation effect of self-isolation on hospital capacity.

Due to the limits of testing methods, the long incubation period, and cases with mild or no symptoms and delayed reporting, there is potentially a huge (and unknown) number of unreported cases, which could affect the future evolution of the epidemic. Some researchers, therefore, have used the SIR (symptomatic-infectious-recovered) model and SEIRD to estimate the number of undetected cases [14]–[17]. Some approaches also incorporate transportation information (such as human migration data and community mobility data) to analyze the impact of travel on disease transmission and thus the effect of travel restriction [18]–[21]. However, studies using typical SEIRD or SIR typically assume the transmission rate and death rate to be constant over time.

With changes in human behavior, clinical treatment and intervention policies, the transmission and fatality rate vary over time. Therefore, SEIRD models with constant parameters cannot accurately depict the spread of disease. Some researchers have considered time dependency of transmission parameters. One approach is to multiply "modulate factors" and transmission parameters, where the modulate factors vary with respect to the intervention policy. For example, Ray et al. extend the SIR model by introducing a time-varying transmission rate modifier $\pi(t)$ to modify the basic transmission rate $\beta$, where $\pi(t)$ is determined by the in-home isolation rate and in-hospital isolation rate together [22]. Another approach is to represent transmission rate $\beta$ by a piecewise function, with discrete values $\beta_i, with\ i = 1, \dots, n$ representing corresponding time periods $\Delta t_i$. Piccolomini et al. compare two piecewise time-dependent infection rate functions and fit the infection rate function, incubation period, and death rate for each uniformly divided time interval [23]. In another paper, Santamaŕia and Hortal utilize segmented regressions to create a piecewise time-dependent model for reproduction numbers within 16 Spain regions [24]. Jonas et al. used a Bayesian Markov Chain Monte Carlo method to infer the $\beta_i$ for each time period and then identify potential changing points in the spread of COVID-19.

Some researchers have used a precise functional form for the time variation of transmission parameters. Godio et al. modified the recovery rate to a sinusoidal function with six parameters and adjusted the transmission rate according to mobility trends [25]. Cotta et al proposed an exponentially decreasing form of transmission rate: $\beta(t) = \beta_0 exp^{-\gamma(t-t_x)}$, where $\beta_0$ is the transmission rate without any intervention, and $t_x$ is the time when the interventions are implemented [26].Cotta further simulated different intervention measures through five different scenarios and pointed out that improving sanitary habits, with more intensive testing for isolation, is essential to contain the disease. Li et al. introduces a novel epidemiological model,

DELPHI, which captures the impact of under-detection and government intervention on the spread of COVID-19 [27]. The model, applied across 167 geographical areas, has successfully predicted large-scale epidemics and has been used to analyze the effectiveness of various government interventions. The authors find that mass gathering restrictions and school closings were among the most effective measures in reducing the rate of infection during the early stages of the pandemic. One problem with the four approaches is the introduction of many parameters to depict time dependency, thus risking overfitting and increasing the computational cost when analyzing multiple regions at the same time. Further, the exponentially decreasing form of transmission rates may be too restrictive, as it demands that rates of change in transmission decline continuously. For these reasons, we will focus on a function that is neither overly general (subject to over-fitting) nor as restrictive as the exponential form.

In our research we investigate the use of a concise formulation through which continuously time varying transmission and case fatality rates are modeled with a small number of parameters that fit reported data. Like [9], [28]–[30], we utilize a type of logistic function (i.e., a Sigmoid function), but not simply to model reported cases or model reported deaths over time, but to instead model both reproduction rate and case fatality rate within an integrated SEIRD model (a version of which we initially developed in 2020 [31]). Our innovation is to improve the classical SEIRD model through an approach that adapts to the dynamic pattern of transmission under different epidemic scenarios. Thus, we provide insights into transmissibility of the disease while modeling historical data on confirmed cases and confirmed deaths in 50 states of the US.

## The Proposed Time Varying Model

We draw from the SEIRD compartmental model, which divides the population into five groups: susceptible(S), exposed(E), infected(I), recovered(R) and dead(D). SEIRD utilizes differential equations to model the evolution of the number of people in these states over time. Susceptible individuals can catch the virus through contact with infected people and transition into the exposed state. Exposed people are in a latent state until they progress to the infectious state, occurring at a rate inversely proportional to the incubation period (thus, exposed is defined as a state in which people are not yet infectious). Infected people eventually progress into either the dead state, if they succumb to the disease, or into the recovered state, with different rates. Those who have recovered are assumed to be no longer susceptible to contracting the disease in the SEIRD model.

We introduce death rate $\alpha(t)$ as a time varying function, representing the proportion of infectious individuals who eventually die from the disease, by date t. Those who eventually die transfer from the infected to the dead state at a rate of $\rho$, representing the inverse of the time from becoming infectious until time of death. In our model, $\rho$ is assumed to be constant over

time. Those who eventually recover do so at the rate $\gamma$, representing the inverse of the time from becoming infectious until recovery. We will also later derive the effective reproduction number$Rep(t)$, representing the average number of persons who are exposed to the disease by each infectious person, as a function of time. Taking these factors into account, the system of equations of the proposed SEIRD model is given by Equation (1):

$$\frac{dS(t)}{dt} = -\beta(t) \cdot I(t) \cdot \frac{S(t)}{N}$$

$$\frac{dE(t)}{dt} = \beta(t) \cdot I(t) \cdot \frac{S(t)}{N} - \sigma \cdot E(t)$$

$$\frac{dI(t)}{dt} = \sigma \cdot E(t) - \big(1 - \alpha(t)\big) \cdot \gamma I(t) - \alpha(t) \cdot \rho \cdot I(t) \quad (1)$$

$$\frac{dR(t)}{dt} = \big(1 - \alpha(t)\big) \cdot \gamma \cdot I(t)$$

$$\frac{dD(t)}{dt} = \alpha(t) \cdot \rho \cdot I(t)$$

where:

$S(t)$ = number of people in susceptible state at time t

$E(t)$ = number of people in exposed, but uninfected at time t

$I(t)$ = number of people in infectious state at time t

$D(t)$ = number of people who have died at time t

$R(t)$ = number of people who have recovered at time t

N = total number of people

$\beta(t)$ = transmission rate at time t

$\sigma$ = transformation rate from exposed to infectious, which is the reciprocal of the incubation period

$\alpha(t)$ = likelihood of eventual death of a person who is infected at time t

$\gamma$ = transformation rate from infectious to recovered, which is the reciprocal of the recovery time

$\rho$ = transformation rate from infectious to dead

Changes in intervention policy, global events and medical care affect $\alpha(t)$ and $\beta(t)$. While, in theory, these functions may change erratically as a consequence of discrete events, such as new public health measures, we hypothesize that such discrete events do not suddenly alter either function. Therefore, we seek to understand whether a simple continuous model, with a minimal set of parameters, might accurately represent historical data. For illustration, at the enactment of

a new intervention policy, the public may not react suddenly, and neither do the transmission parameters. The public may become used to the policy after a period of adaptation, and eventually the effective reproduction number will stabilize. In addition, the public responds to both government policies and communication about the disease. Communication comes from many, sometimes conflicting, sources. How the public at large absorbs and responds to such often confusing messages may be gradual.

A natural function to describe this pattern of change is the Sigmoid function. Equation (2) is the general form of the Sigmoid function, where $k$ determines the slope of the function and $a$ determines the $x$ value at the middle point (i.e., point of time when y=.5).

$$S(x) = \frac{1}{1 + e^{k(x-a)}} \tag{2}$$

Thus, we define the function for transmission rate and death rate Equation (3) and (4).

$$\beta(t) = \beta_{end} + \frac{\beta_{start} - \beta_{end}}{1 + e^{m\cdot(x-a)}} \tag{3}$$

$$\alpha(t) = \alpha_{end} + \frac{\alpha_{start} - \alpha_{end}}{1 + e^{n\cdot(t-b)}} \tag{4}$$

where,

$\beta_{start}$ is the starting reproduction number

$\beta_{end}$ is the ending reproduction number

$\alpha_{start}$ is the starting death rate, ranging from 0 to 1

$\alpha_{end}$ is the ending death rate, ranging from 0 to 1

$m, n, a, b$ are the shape parameters

Note that the Sigmoid function does not generalize to instances where rates both decline and increase over time. Such situations demand a multi-phase model, as discussed later. Nevertheless, as we will show, the sigmoid function produces low error rates in predicting cases and deaths in the early months of the pandemic in the United States.

## Parameter Estimation and Model Fitting

Parameters in Eqs. 1 were estimated with the objective of minimizing the weighted summation of squared error between cumulative predicted and measured confirmed cases and the summation of squared error between cumulative predicted and cumulative confirmed deaths. Our analysis is based on the period from the day of first reported case in each state until 07/28/2020, across all 50 American states. For each state of the United States, we chose a start date of 4 days prior to the date of the first confirmed case. Four days was chosen based on the information from CDC [4].

Two methods were used for different sets of parameters, as described below. To estimate the shape parameters $m, n, a, b$ and the starting/ending parameters $\beta_{start}, \beta_{end}, \alpha_{start}, \alpha_{end}$ , we fit Eqs. 1 to the cumulative confirmed case numbers and the cumulative confirmed death numbers with the nonlinear least square method. Other parameters were derived from prior research.

**Parameters Derived from Prior Research**

As mentioned in the other studies [32]–[34], the median incubation period was 4 days. Among 305 hospitalized patients and 10,647 recorded deaths, the median time of hospitalization was 8.5 days and the median interval from illness onset to death was 10 days (IQR =6 - 15 days). We assume the median hospitalization time is the median time for infectious people to stop being contagious. Hence, we set these parameters as the inverse of these time values: $\sigma = 1/4, \gamma = 1/8.5, \rho = 1/10$.

**Parameters Derived from Optimization**

The remaining parameters are derived for each American state by optimizing the fit of the model to historical case and death data, where the objective is to minimize a weighted sum of daily squared error over the analysis period. We utilized a search algorithm that required initialization and a constrained search space, as explained below.

We define the model function M(t; $[\beta_{start}, \beta_{end}, m, a, \alpha_{start}, \alpha_{end}, n, b]$): $t \to R^2$, where M(t; $[\beta_{start}, \beta_{end}, m, a, \alpha_{start}, \alpha_{end}, n, b]$) = $[\hat{I}(t) + \hat{R}(t) + \hat{D}(t), \hat{D}(t)]$ and the reported case number and death number at time t is [Cases(t), Deaths(t)]. Because it is unlikely for transmission and death rates to change drastically in a single day, we set upper bounds for $m \; and \; n$ at 0.33 (meaning that rates do not suddenly change in less than three days) and initialize the search at 0.25. We permit the turning point of the sigmoid function to occur on any day in the timeline; we set $a, b \in [0,125]$, where 125 is the length of the period from March 1st to July 28th, in days (as of March 1 few states had reported cases). Prior research suggests that the initial effective reproduction number is around 3 [7], equivalent to a transmission rate of 0.75, which we use for initialization. Because transmission rates vary significantly among locations due to local conditions (such as crowding), we bound $\beta_{start} \in [0.5,7.5]$ and $\beta_{end} \in [0,2.5]$, thus permitting a wide range of results.

To summarize, the parameters set $P = [\beta_{start}, \beta_{end}, m, a, \alpha_{start}, \alpha_{end}, n, b]$ is initialized as [0.75,0.5,0.25,10,0.4,0.1,0.25,10]. Then the parameter optimization problem is formulated in Equations (5):

$$min_P \| M(\mathrm{t}; P) - [Cases(t), Deaths(t)] \|_2^2 \quad (5)$$

$$s.t. \quad 0.5 \le \beta_{start} \le 7.5$$

$$0.1 \le \beta_{end} \le 2.5$$

$$0 \le \alpha_{\mathrm{start}} \le 1$$

$$0 \leq \alpha_{end} \leq 1$$
$$0.01 \leq m \leq 0.33$$
$$0.01 \leq n \leq 0.33$$
$$0 \leq a \leq 125$$
$$0 \leq b \leq 125$$

The number of reported deaths is smaller than the number of reported cases in all locations. Thus, treating errors in death estimation and case estimation the same will lead to underfitting of the death data, in preference to minimizing the errors in case data. Therefore, considering the accuracy of the reported death data and the fitting accuracy, we optimized a weighted sum of squared death and case data, multiplying w by deaths during the fitting process. The adjusted objective function is shown as Equation (6):

$$\min_p \left\| \left( \hat{I}(t) + \hat{R}(t) + \hat{D}(t) - Cases(t) \right)^2 + w * \left( \hat{D}(t) - Deaths(t) \right)^2 \right\|_2 \quad (6)$$

The parameters are estimated by solving the nonlinear constrained least-squares problem in Equation (5), utilizing the Levenberg–Marquardt algorithm (LMA). The LMA algorithm adaptively varies the parameter updates between the gradient descent update and the Gauss-Newton update and accelerates to a local minimum [31]. The LMA is implemented to our model fitting by the *lmfit* package in Python. In our analysis we utilized w = 20 to yield similar error percentages for deaths and cases.

**Data Limitations**

We recognize that reported cases and deaths are not the same as actual infections and actual deaths, which are unknowable. Daily confirmed cases are influenced by widely varying testing rates and policies, which change over time. At the beginning of the epidemic, the limited test kits were restricted to those who suffer from severe symptoms and those who are in a higher risk of exposure. Death data was likely to be more accurate but can suffer from reporting errors, due to how deaths are attributed to COVID-19 (or not), the timing of filing reports and the general accuracy of reporting. For these reasons, our model is fit to reported data.

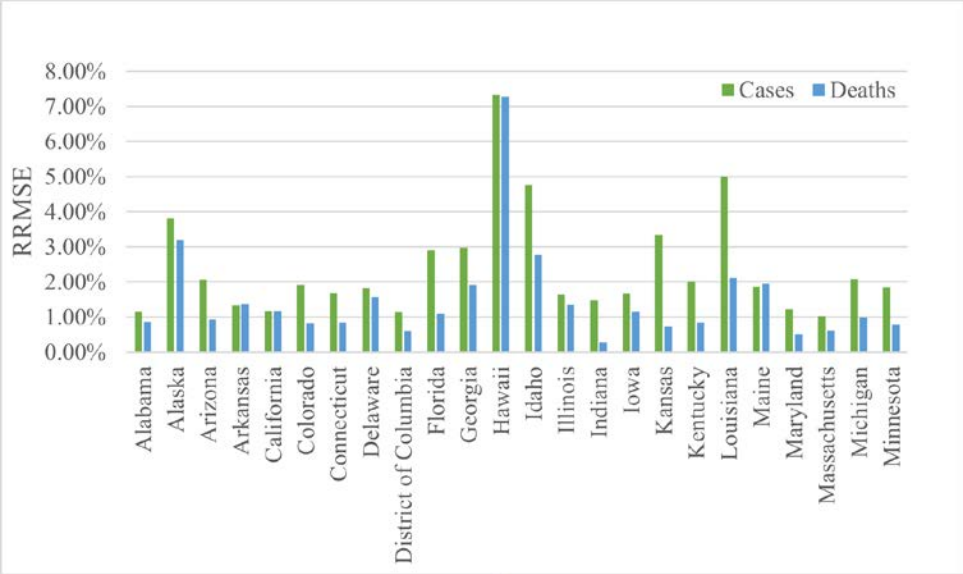
Reporting has also shown a consistent day-of-week variation across many locations, with weekend data differing from weekday data. This variation is more likely the consequence of different patterns of healthcare staffing, and differences in how patients present for testing by day of the week, rather than differences in disease transmission by day of the week. To smooth out these effects, we model the moving 7-day average data instead of the daily reported data.

# Results

## Model Accuracy

The first case of COVID-19 in the United States was reported on January 20, 2020 [35]. As of July 31, 2020, a total of 4,665,469 cases and 155,863 deaths had been reported across the states and territories of America [36].We fit the model with the dataset of 7-day moving average cases and deaths for the 50 states, provided by the COVID-19 tracking project led by *The Atlantic* (derived from the Centers for Disease Control), for the period from the date of the first reported cases to July 31st. The fitting accuracy across all states is presented in Figure 1, measured by the relative root mean square error (RRMSE) (explained in Supplementary Materials S1).

The fitting accuracy of the reported cases ranges from 0.54% to 7.34% and of the reported deaths ranges from 0.29% to 7.28%. The average and median RRMSEs for deaths are 1.61% and 1.33%; for cases, the average and median values are 2.30% and 1.88%. RRMSE fell below 5% by both measures for all states except Hawaii, Idaho, Louisiana, Montana and Wyoming.



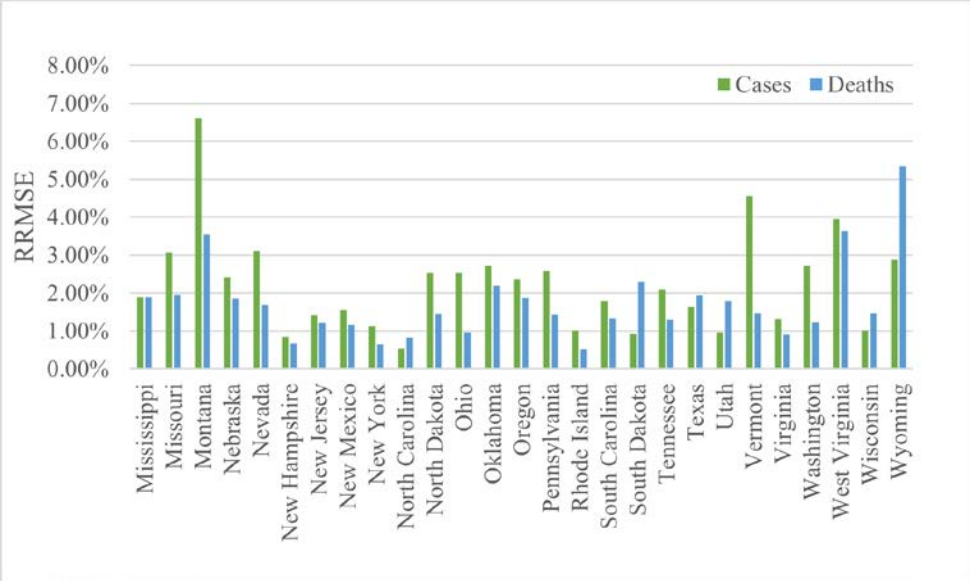
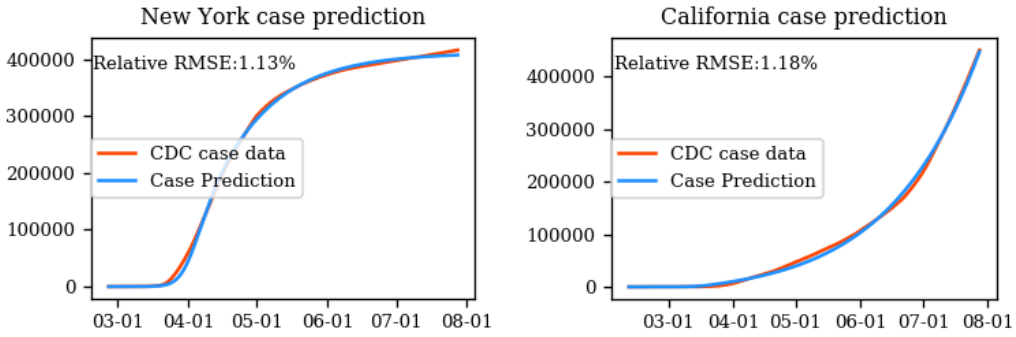



Figure 1: Relative root mean squared error (RRMSE) for case and death data across all states

Figures 2 and 3 show the specific fitting results for cases and deaths by day for the two states with the largest number of cases (New York and California) as well as two other states for which the fit is less accurate (Florida and Hawaii). For New York and California, the fitting results closely coincide with CDC data. Examining Florida and Hawaii, the CDC data follows a pattern of two phases, which is not as well captured by our model. For Hawaii, the curve flattened for a period and then rose. As discussed later, our basic model characterizes the transmission dynamic for a period with one phase (i.e. the curve should become flat at most once) but can be modified as a multi-phase model.



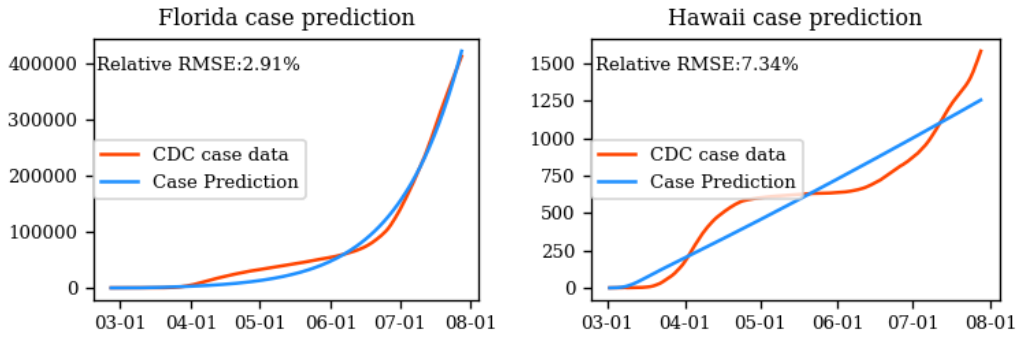



Figure 2: Fitting results for case data: New York, California, Florida and Hawaii

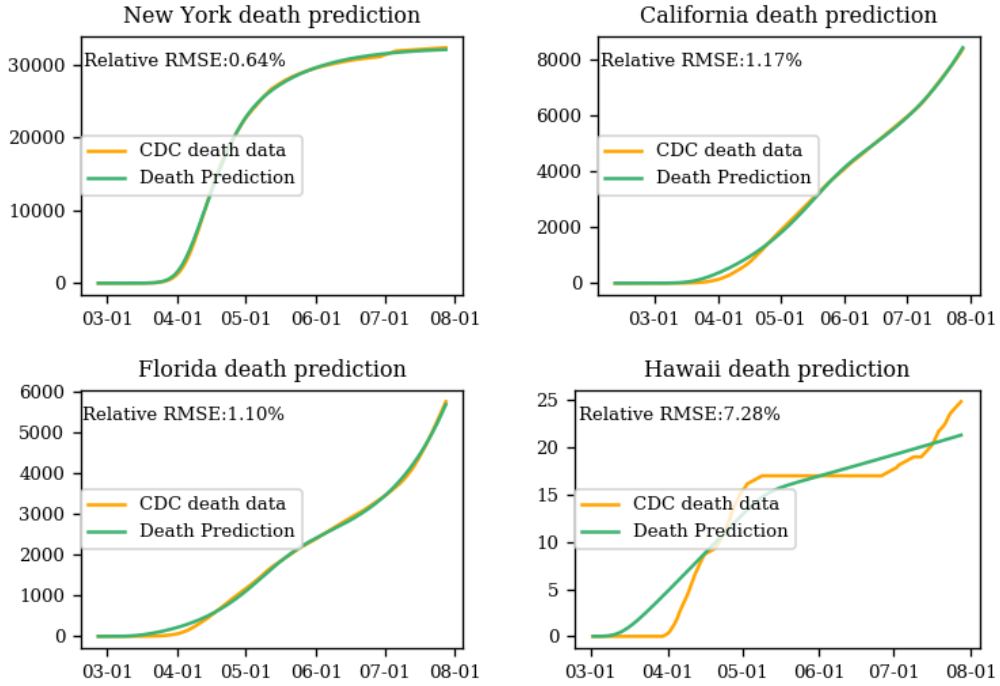



Figure 3: Fitting results for death data: New York, California, Florida and Hawaii

## Effective Reproduction Number Calculation and Trends

Effective reproduction number at any time t, which we define as $Rep(t)$, is the average number of people in a population who are infected per infectious case, where everyone is susceptible to the disease. $Rep(t)$ measures the transmission potential of infectious diseases [37]. When $Rep(t) > 1$, the rate of new cases will increase over time, until the population loses susceptibility to the disease. When $Rep(t) < 1$, the rate of new cases will decline over time.

$Rep(t)$ can be estimated with the next-generation matrix method (explained in Supplementary Material S2) [38], [39].

At the beginning of the epidemic, $Rep(t)$ reflects the natural transmissibility of COVID-19, i.e. the basic reproduction number $R_0$ in the absence of intervention. With the evolution of the epidemic, $Rep(t)$ changes dynamically, as do the transmission rate $\beta(t)$ and death rate $\alpha(t)$,

which are influenced by both the intervention policy and population immunity. Figures 4 and 5 show the fitted $Rep(t)$ at the start of the epidemic across all states (defined by first reported case) and fitted $Rep(t)$ on July 31st. We see that $Rep(t)$ ranges from 1.27 to 16.49, with a median value of 2.87. We found that by July 31st, the reproduction number had fallen below 1 in all 50 states, with a median value of 0.37. It should be kept in mind that this optimal fit is a reflection of the reported data on cases and deaths. Increasingly aggressive testing may cause rates of reported cases to grow faster than the rate of growth for actual infections.

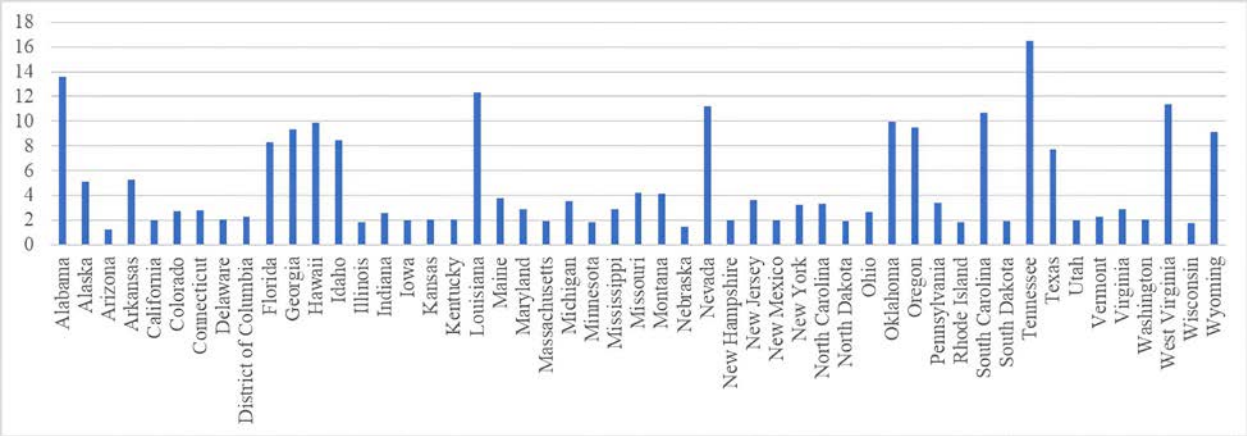



Figure 4: Fitted $Rep(t)$ on the date of first reported case across all 50 states

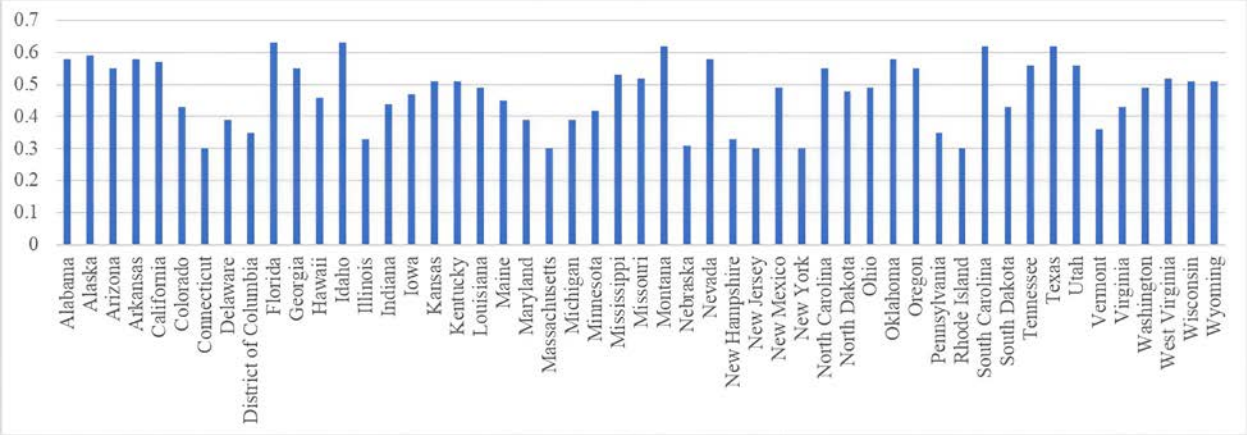



Figure 5: Fitted $Rep(t)$ on July 28th across all 50 states

For illustration, Figure 6 shows our estimated history of $Rep(t)$ for New York, California, Florida and Hawaii. Time 0 in these graphs is the day of the first reported case, which varies from state to state. In these cases, the effective reproduction number both stabilized and became smaller than 1 with time, with the change occurring over a period of 10 to 30 days.

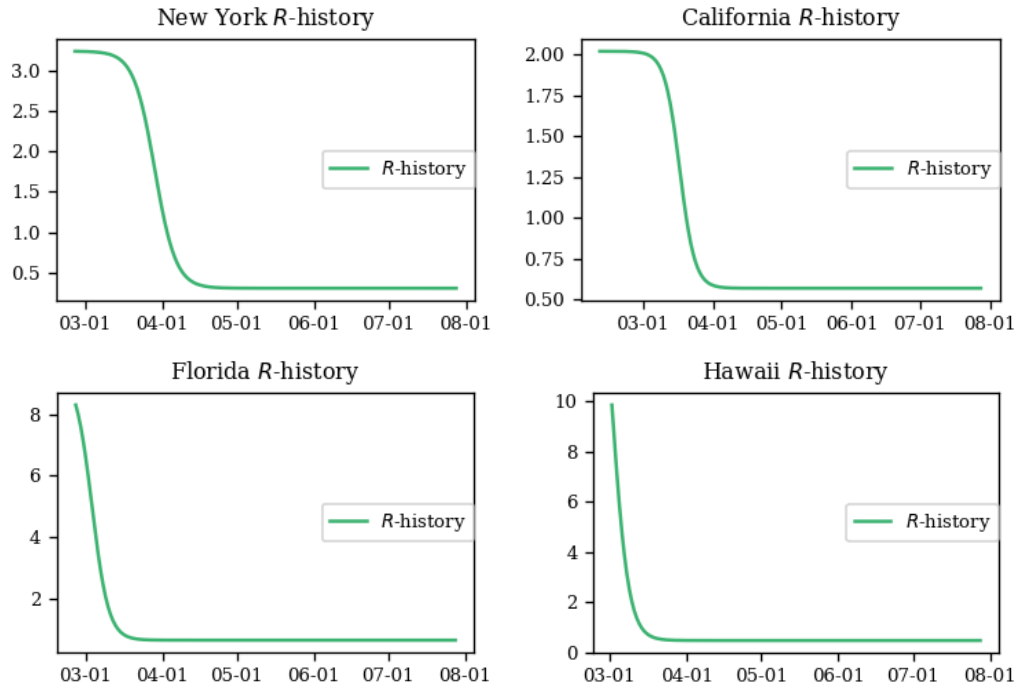
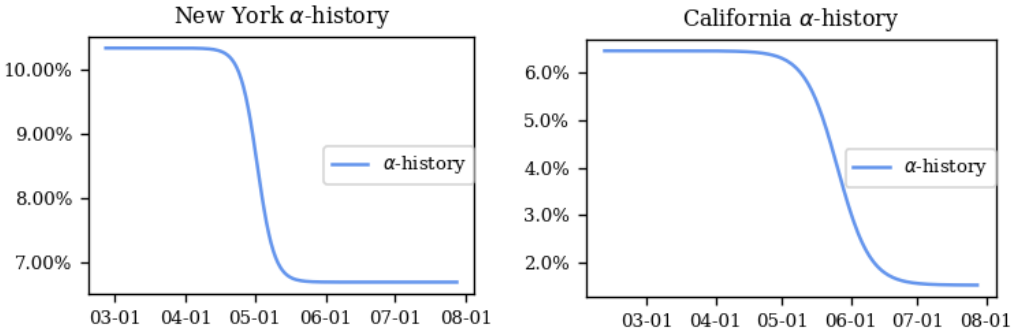



Figure 6: Estimated effective reproduction number (R) by date: New York, California, Florida and Hawaii

As noted, in the early stages of an epidemic, the reproduction number may seem particularly large not only because the disease spreads rapidly but also because the rate of testing is increasing.

**Death Rate Trends**

Death rate is another measure that shows the change in virus outcomes over time, reflecting the health system's ability to deal with the flood of infected people. Figure 7 provides examples. From the historical plot, we see the hardest-hit states, like New York and Florida, experienced a much higher death rate in the early stage than the average 3% death rate in the United States. The relatively high death rate could be caused by the lack of effective medical treatment and hospital overload. It could also reflect limited testing of patients, whereby only the sickest patients were recorded as cases. With improvement of medical treatment, and increased testing, the death rate per confirmed case for most states decreased to a much smaller value.



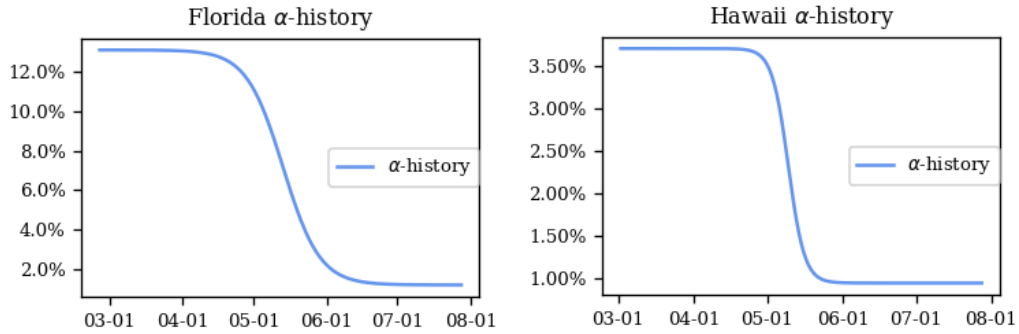


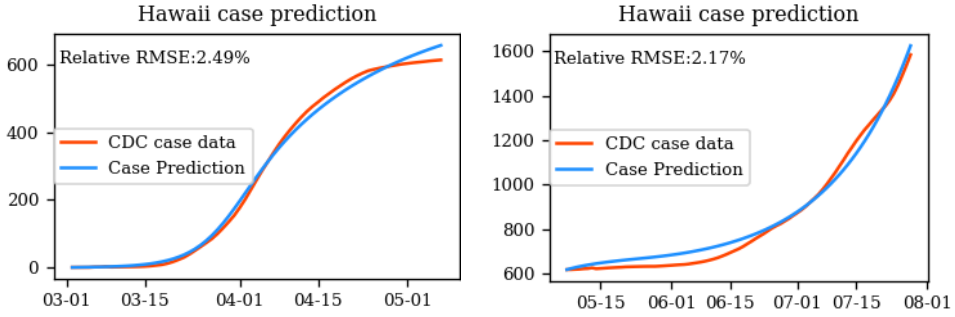

Figure 7: Predicted death rate ($\alpha$) by date: New York, California, Florida and Hawaii

## Multi-Phase Model

Our model fits reported cases and deaths within 2% error in most states. However, because the model is premised on the assumption that transmission rates do not at first go down, and then later go up, it needs to be modified for states that exhibit multiple waves of the disease within the study period. Data for Hawaii – for which the model has the poorest fit – indicate this pattern.

For such locations, we propose an alternate multi-phase model. The Hawaii Department of Health announced the first positive case on Oahu, Hawaii, on March 6th, 2020, and then immediately enacted a stay-at-home order on March 25th. From April 19th to May 7th, the case curve flattened. The state announced on May 7th that Hawaii would embark on the first phase of reopening. The data reflect a second wave of coronavirus commencing on or about May 7.

We divide the Hawaii timeline into two periods, the first from March 6th until May 7th, and the second from May 7th to July 28th. We fit the first stage with the initialization of one exposed people at the start. To initialize the second phase, we use the predicted number of exposed, infectious and recovered people from the first phase, combined with the reported deaths as of May 7th. With this modification, the RRMSE for cases declines below 2.5% and the RRMSE for deaths declines below 2.7%. The fitting results in Figure 8 show that our two-phase model captures the transmission pattern more precisely than the single-phase model.



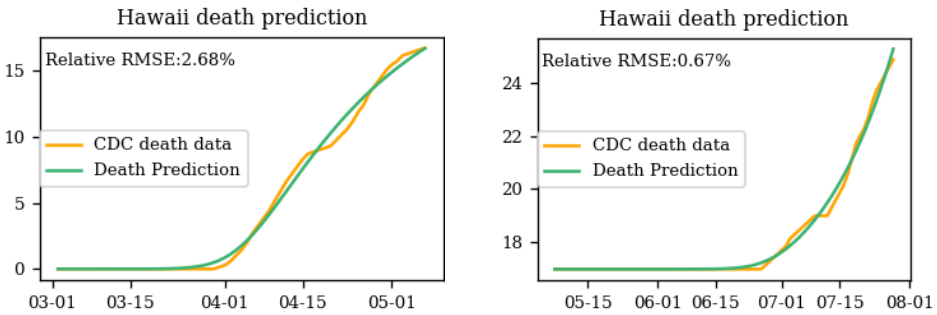
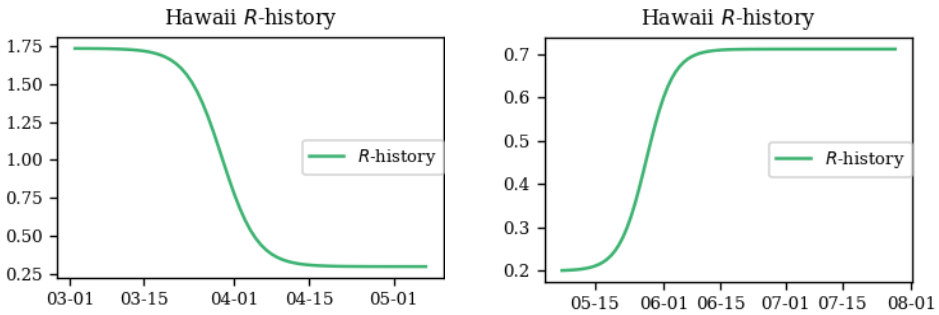



Figure 8: Two phase fitting results for Hawaii (left graphs phase 1, right graphs phase 2)

The histories for estimated (two-phase) effective reproduction number and death rate are shown in Figure 9. The first phase showed a decline in the reproduction number after the initial announcement of the stay-at-home order. However, with the reopening, the reproduction number increased, explaining increases in case rates. Death rates, by contrast, exhibit a peculiar behaviour, increasing over time in each phase, with a discontinuity when transitioning from the first phase to the second. Beyond exhibiting two phases, Hawaii has a small number of deaths, with no deaths occurring in the transition period between phases. We surmise that the function, while representing the data well, is peculiar because of the unusual pattern in deaths within Hawaii.



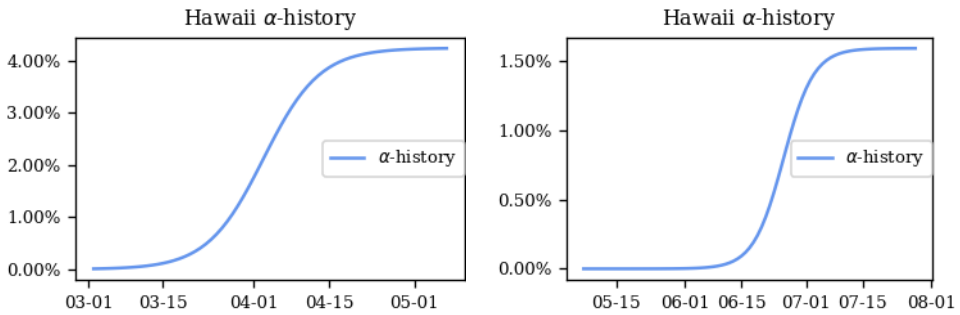



Figure 9: Estimated effective reproduction number and death rate by date for two phase model (left graphs phase 1, right graphs phase 2)

To summarize, our time-varying model produced comparatively small error rates for most states when fit to historical data. Over the time period studied, the reproduction rate of disease declined in most states, which was characterized well by the Sigmoid function for transmission rate. The multi-phase model improves the fit to historical data in states that demonstrated both a decline and increase in transmission during the study period, as demonstrated for Hawaii. As has been apparent during the pandemic, both increases and decreases are possible, as public health rules and human behaviour change over time.

## Conclusions

We have developed an extension of the SEIRD model that represents dynamic changes in death and transmission rates over time using a continuous Sigmoid function, under the hypothesis that these rates change continuously, rather than immediately upon implementation of public health policies or treatments.

We showed that the model fit historical data for the United States well for most states for the early months of the pandemic, with a median RRMSE of 1.33% for deaths and 1.88% for cases among the 50 states. We found that the reproduction number varied between 1.27 and 16.49 at the start of the pandemic among the 50 states, with a median of 2.87, meaning that case rates were growing throughout the country. By July 28, the reproduction rate fell below 1 in all 50 states, with a median of 0.37, meaning case rates were dropping throughout the country. Our time-varying model tracked the underlying changes in transmission rates, as well as death rates, that occurred during that six-month period.

Those states with poorer fits experienced multiple waves of the disease. Using Hawaii as an example, we showed that a multi-phase extension of the model provides a more accurate fit, where the transition from one phase to the next is defined by changes in public health policy, disease variants or behaviour that affect rates of transmission or case death rates. Using just two phases, the RRMSE for cases dropped to 2.5% and the RRMSE for deaths dropped to 2.7%

An advantage of our model is the small number of parameters needed to depict dynamic changes in transmission and death rates. Thus, the model provides an efficient method for quantifying differences among regions, and over time, in the spread and outcomes of the disease. By examining historical trends, the model can be applied to analyze how variations in simple parameters can lead to fewer or more cases and deaths.

In the future, we intend to investigate ranges of uncertainty in parameters, and also apply the model in the optimization of vaccine distribution. We will also develop multi-region extensions of the model, which represent spread of disease from one region to another, or perhaps within sub-regional groups. In all of these examples, representation of transmission and death rates with a small number of parameters can be the foundation for more complex analyses.

Our research models case and death data as they are reported. We recognize that the true number of cases may differ from reported values, as might the number of deaths. The variations from state to state reflect, in part, the actual spread and outcomes of disease, the extent to which cases are detected and reported, and how deaths have been classified. Throughout the COVID-19 pandemic, data accuracy has challenged all efforts to model the spread of the disease.

## Data Availability

The data that support the findings of this study are openly available at https://covid19datasource.usc.edu/.

## Conflicts of Interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest or non- financial interest in the subject matter or materials discussed in this manuscript.

## Funding Statement

Research was supported by the University of Southern California through the Zumberge Innovation Fund and the Center for Undergraduate Research in Viterbi Engineering (CURVE).

## Acknowledgments

This manuscript was submitted as a pre-print in the link "https://www.medrxiv.org/content/10.1101/2020.09.25.20201905v1.full.pdf " [40].

# Supplementary Materials

## S1. Relative Root Mean Square Error (RRMSE)

## S2. Next-generation Method

## S3. Basic SEIR Model with Constant Transmission Parameters

# References


[1] Y. Fang, Y. Nie, and M. Penny, "Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis," *J. Med. Virol.*, vol. 92, no. 6, pp. 645–659, 2020, doi: 10.1002/jmv.25750.

[2] C. L. Shaw and D. A. Kennedy, "What the reproductive number R0 can and cannot tell us about COVID-19 dynamics," *Theor. Popul. Biol.*, vol. 137, pp. 2–9, 2021.

[3] B. L. Tesini, "Coronaviruses and Acute Respiratory Syndromes (COVID-19, MERS, and SARS) - Infectious Diseases," *Merck Manuals Professional Edition*. Merck Manuals, Jul-2020.

[4] "Clinical presentation," *Centers for Disease Control and Prevention*. Centers for Disease Control and Prevention, Oct-2022.

[5] J. A. Backer, D. Klinkenberg, and J. Wallinga, "Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020," *Euro Surveill.*, vol. 25, no. 5, pp. 1–6, 2020, doi: 10.2807/1560-7917.ES.2020.25.5.2000062.

[6] S. Zhao *et al.*, "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *Int. J. Infect. Dis.*, vol. 92, pp. 214–217, 2020, doi: 10.1016/j.ijid.2020.01.050.

[7] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of COVID-19 is higher compared to SARS coronavirus," *J. Travel Med.*, vol. 27, no. 2, pp. 1–4, 2020, doi: 10.1093/jtm/taaa021.

[8] G. G. Katul, A. Mrad, S. Bonetti, G. Manoli, and A. J. Parolari, "Global convergence of COVID-19 basic reproduction number and estimation from early-time SIR dynamics," *PLoS One*, vol. 15, no. 9, p. e0239800, 2020.

[9] Y. Alimohamadi, M. Taghdir, and M. Sepandi, "Estimate of the basic reproduction number for COVID-19: a systematic review and meta-analysis," *J. Prev. Med. Public Heal.*, vol. 53, no. 3, p. 151, 2020.

[10] N. Y. S. D. of H. New York State Task Force on Life & the Law, "VENTILATOR ALLOCATION GUIDELINES," 2015.

[11] B. M. Rosenthal and J. Goldstein, "N.Y. May Need 18,000 Ventilators Very Soon. It Is Far Short of That." The New York Times, 2020.

[12] IHME COVID-19 health service utilization forecasting team, "Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months | Institute for Health Metrics and Evaluation," *MedRxiv*, 2020, doi: 10.1101/2020.03.27.20043752.

[13] S. M. Moghadas *et al.*, "Projecting hospital utilization during the COVID-19 outbreaks in the United States," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 16, pp. 9122–9126, 2020, doi: 10.1073/pnas.2004064117.

[14] S. Zhao *et al.*, "Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak," *J. Clin. Med.*, vol. 9, no. 2, p. 388, 2020, doi: 10.3390/jcm9020388.

[15] H. Nishiura *et al.*, "Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)," *Int. J. Infect. Dis.*, vol. 94, pp. 154–155, 2020, doi: 10.1016/j.ijid.2020.03.020.

[16] M. G. Pedersen and M. Meneghini, "Quantifying undetected COVID-19 cases and effects of containment measures in Italy," *Res. Prepr. (online 21 March 2020) DOI*, vol. 10, no. 3, 2020.

[17] A. Mahajan, R. Solanki, and N. Sivadas, "Estimation of undetected symptomatic and asymptomatic cases of

COVID-19 infection and prediction of its spread in the USA," *J. Med. Virol.*, vol. 93, no. 5, pp. 3202–3210, 2021.

[18] A. Radulescu and K. Cavanagh, "Management strategies in a SEIR model of COVID 19 community spread," pp. 1–13, 2020.

[19] G. Luo, M. L. McHenry, and J. J. Letterio, "Estimating the prevalence and risk of COVID-19 among international travelers and evacuees of Wuhan through modeling and case reports," *PLoS One*, vol. 15, no. 6, pp. 1–13, 2020, doi: 10.1371/journal.pone.0234955.

[20] S. Zhao *et al.*, "The association between domestic train transportation and novel coronavirus (2019-nCoV) outbreak in China from 2019 to 2020: A data-driven correlational report," *Travel Med. Infect. Dis.*, vol. 33, no. January, pp. 2019–2021, 2020, doi: 10.1016/j.tmaid.2020.101568.

[21] J.-T. Wei *et al.*, "Impacts of transportation and meteorological factors on the transmission of COVID-19," *Int. J. Hyg. Environ. Health*, vol. 230, p. 113610, 2020.

[22] D. Ray *et al.*, "Predictions, role of interventions and effects of a historic national lockdown in India's response to the COVID-19 pandemic: data science call to arms," *Harvard data Sci. Rev.*, vol. 2020, no. Suppl 1, 2020.

[23] E. Loli Piccolomini and F. Zama, "Monitoring Italian COVID-19 spread by a forced SEIRD model," *PLoS One*, vol. 15, no. 8, p. e0237417, 2020, doi: 10.1371/journal.pone.0237417.

[24] L. Santamar\'\ia and J. Hortal, "COVID-19 effective reproduction number dropped during Spain's nationwide dropdown, then spiked at lower-incidence regions," *Sci. Total Environ.*, vol. 751, p. 142257, 2021.

[25] 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, 2020, doi: 10.3390/ijerph17103535.

[26] R. M. Cotta, C. P. Naveira-Cotta, and P. Magal, "Mathematical parameters of the COVID-19 epidemic in Brazil and evaluation of the impact of different public health measures," *Biology (Basel).*, vol. 9, no. 8, p. 220, 2020.

[27] M. L. Li, H. T. Bouardi, O. S. Lami, T. A. Trikalinos, N. Trichakis, and D. Bertsimas, "Forecasting COVID-19 and analyzing the effect of government interventions," *Oper. Res.*, vol. 71, no. 1, pp. 184–201, 2023.

[28] D. Faranda and T. Alberti, "Modeling the second wave of COVID-19 infections in France and Italy via a stochastic SEIR model," *Chaos An Interdiscip. J. Nonlinear Sci.*, vol. 30, no. 11, p. 111101, 2020.

[29] T. Alberti and D. Faranda, "On the uncertainty of real-time predictions of epidemic growths: A COVID-19 case study for China and Italy," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 90, p. 105372, 2020.

[30] A. Bessadok-Jemai and A. A. Al-Rabiah, "Predictive approach of COVID-19 propagation via multiple-terms sigmoidal transition model," *Infect. Dis. Model.*, vol. 7, no. 3, pp. 387–399, 2022.

[31] H. P. Gavin, "The Levenburg-Marqurdt Algorithm For Nonlinear Least Squares Curve-Fitting Problems," *Duke Univ.*, pp. 1–19, 2019.

[32] W. Guan *et al.*, "Clinical characteristics of coronavirus disease 2019 in China," *N. Engl. J. Med.*, vol. 382, no. 18, pp. 1708–1720, 2020, doi: 10.1056/NEJMoa2002032.

[33] J. A. W. Gold *et al.*, "Characteristics and Clinical Outcomes of Adult Patients Hospitalized with COVID-19 — Georgia, March 2020," *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 69, no. 18, pp. 545–550, 2020, doi: 10.15585/mmwr.mm6918e1.

[34] F. May *et al.*, "Characteristics of Persons Who Died with COVID-19 — United States ," *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 69, no. 28, pp. 923–929, 2020.

[35] M. L. Holshue *et al.*, "First case of 2019 novel coronavirus in the United States," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 929–936, 2020, doi: 10.1056/NEJMoa2001191.

[36] A. Madrigal and R. Meyer, "The COVID Tracking Project," *The COVID Tracking Project*, 2020. [Online]. Available: https://covidtracking.com/.

[37] G. Chowell, J. M. Hyman, L. M. A. Bettencourt, and C. Castillo-Chavez, "The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends," *Math. Stat. Estim. Approaches Epidemiol.*, pp. 1–363, 2009, doi: 10.1007/978-90-481-2313-1.

[38] P. Van Den Driessche, “Reproduction numbers of infectious disease models,” *Infect. Dis. Model.*, vol. 2, no. 3, pp. 288–303, 2017, doi: 10.1016/j.idm.2017.06.002.

[39] P. den Driessche and J. Watmough, “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission,” *Math. Biosci.*, vol. 180, no. 1–2, pp. 29–48, 2002.

[40] M. Lyu and R. Hall, “Dynamic Modeling of Reported COVID-19 Cases and Deaths with Continuously Varying Case Fatality and Transmission Rate Functions,” *medRxiv*, 2020.