Estimation and monitoring of COVID-19's trasmissibility from publicly available data

Antonio Silveira*, **Antonio Pereira**

Department of Electrical and Biomedical Engineering, Federal University of Pará, 66075-110, Belém, PA, Brazil

Correspondence*:
Corresponding Author
asilveira@ufpa.br

ABSTRACT

The COVID-19 pandemic began in the city of Wuhan, China, at the end of 2019 and quickly spread worldwide. The disease is caused by contact with the SARS-CoV-2 virus, which probably jumped from an animal host to humans. SARS-CoV-2 infects various tissues in the body, notably the lungs, and patients usually die from respiratory complications. Mathematical models of the disease have been instrumental to guide the implementation of mitigation strategies aimed at slowing the spread of the disease. One of the key parameters of mathematical models is the basic reproduction ratio R_0, which measures the degree of infectivity of affected individuals. The goal of mitigation is to reduce R_0 as close or below 1 as possible, as it means that new infections are in decline. In the present work, we use the recursive least-squares algorithm to establish the stochastic variability of a time-varying $R_0(t)$ from eight different countries: Argentina, Belgium, Brazil, Germany, Italy, New Zealand, Spain, and the United States of America. The proposed system can be implemented as an online tracking application providing information about the dynamics of the pandemic to health officials and the public at large.

Keywords: COVID-19, Epidemic Spreading, Pattern Recognition, Mathematical modelling, Transmission dynamics, Disease Prediction.

1 INTRODUCTION

On the 11th of March 2020, the World Health Organization (WHO) declared the 2019 coronavirus disease (COVID-19) a global pandemic (1). COVID-19 is caused by the SARS-CoV-2 coronavirus and was first reported in Wuhan, China, in December 2019 (2). Since then, COVID-19 has spread globally with a total of 4,525,497 laboratory-confirmed cases and 307, 395 deaths (3). The median incubation period of COVID-19 is 5.1 days and nearly all infected persons who have symptoms will do so within 12 days of infection (4). However, an unprecedented characteristic of COVID-19 is its asymptomatic transmission (5), which contributes to its elevated transmissibility (6). So far, there is no specific treatment for the disease and many research teams are currently working on a vaccine that, optimistically, will only be available in 2021. Meanwhile, hospital structures around the globe (e.g. intensive care units' (ICU) beds, ventilators, etc.) are becoming overwhelmed with new patients and the increasing caseload will prove most catastrophic for poor countries, which lack the adequate health-care capacity to deal with the unparalleled demand posed by COVID-19 (7; 8).

So far, efforts to contain the spread of the disease have focused on the adoption of non-pharmaceutical interventions (NPI) based on population behavioral change and social distancing, such as banning large gatherings, enforcing the use of masks, washing hands, or imposing severe lockdowns (9; 10). Due to significant uncertainties regarding the transmissibility of SARS-CoV-2 as well as other political, social, and economic considerations, it is necessary to delineate effective social distancing policies which are able to alleviate COVID-19's burden on healthcare structure and borrow time for the development of a

vaccine or drug candidates while also simultaneously reducing the socioeconomic strain of living in a locked-down, confined society (11). The effective monitoring of the epidemic's dynamics plays a crucial role in the ongoing containment effort, but will also continue to do so for some time when social distancing control measures are relaxed and the first wave of the pandemic is followed some months later by second or third waves of infection that may be more severe than the first (12).

Mathematical models, by providing a quantitative framework for hypothesis evaluation and the estimation of changes in transmission of infectious diseases over time and space, can indicate whether containment measures are having a measurable effect while guiding the design of alternative interventions (13). Mathematical models vary in many aspects, including complexity in terms of the number of variables and parameters used, spatial and temporal resolution (e.g., discrete vs. continuous-time), and design (e.g., deterministic or stochastic) (14). Mechanistic models of the susceptible-infected-recovered SIR type (15; 16) are the standard framework for a wide array of infectious diseases, including COVID-19 (see, for example, (17; 18)). However, parameter estimates for a given model are subject to two major sources of uncertainty: the noise intrinsic to the data and the underlying assumptions used for ascertaining parameter estimates (14).

The basic reproduction number (19), R_0, or the average number of new infections caused by an infectious individual (20), is widely used to characterize pandemics and other infectious outbreaks. For instance, the average R_0at the start of the SARS pandemic in 2003 was estimated to be around 2.75 and was later reduced to <1 due to intervention strategies, including isolation and quarantine activities (21). There are many challenges for estimating R_0, such as the fact that disease can be spread by asymptomatic individuals and the scarce availability of testing supplies (17). R_0 is rarely measured directly and modeled R_0 values are thus dependent on model structures and assumptions (22). Several methods have been proposed to track trends in R_0 during the course of an epidemic (23; 24; 25; 26; 27). The access to reliable estimates of R_0 could provide useful information about the efficacy of containment measures and allow their effective management in order to keep hospitalization rates within a desired approximate range (28).

The least-squares algorithm (LSA) is one of the most popular predictive methods in machine-learning and has been used in many scientific and engineering applications (29). Its offshoot, the recursive least-squares algorithm (RLSA), has been used for real-time estimation applications in diverse areas such as signal and data processing, communications, and control systems (30; 31; 32). The LSA has also been used in epidemiology for calibrating mathematical models' parameters based on time series data while also generating disease forecasts in the near or long terms (33; 14; 34; 35).

In the present study, we model the transmission dynamic of COVID-19 in eight countries (Table 1) using LSA-based techniques. Our goal is to contribute to understanding the spread of SARS-CoV-2 and compare R_0 uncertainty arising from noise in the time series data gathered from public online sources. We used machine learning algorithms to optimally estimate a time-varying $R_0(t)$, which allows the monitoring of the ongoing pandemic in almost real-time. We compared the daily reports on R_0 to those estimated in the present work in order to assess if the available official cases data are reliable to support the analysis of SARS-CoV-2 spread and the control measures by health officials. Besides, LSA-based techniques were used to reveal clues regarding the dynamics of $R_0(t)$ and its stochasticity in terms of the linear power of the estimation error and provide information on how such stochastic behavior is correlated to the outcome of the ongoing pandemics.

Table 1. Sociodemographic data for the eight countries targeted in this study and associated COVID-19 data. COVID-19 data accessed on 22/05/20 on `https://coronavirus.jhu.edu/map.html`. Intervention data according to (36). GDP per capita in US$. Interventions are categorised into Alert Levels 1 to 4 according to the New Zealand framework (L1 = Level 1, Prepare; L2 = Level 2, Reduce; L3 = Level 3, Restrict; L4 =Level 4, Eliminate) (36).

Country	Population	GDP	HDI	Covid-19		Date of 100th case	Day control started (in days since 100th case)			
				Cases	Deaths		L1	L2	L3	L4
Argentina	44,938,712	9,887	0.830	9,918	416	03/19/20	prior	prior	prior	prior
Belgium	11,589,623	46,724	0.919	56,511	9,212	03/06/20	4	6	11	-
Brazil	210,147,125	8,955	0.761	312,000	20,112	03/13/20	6	-	-	-
Germany	83,149,300	46,653	0.939	179,000	8,320	03/01/20	10	12	13	21
Italy	60,317,116	32,947	0.883	228,000	32,486	02/23/20	7	10	17	27
New Zealand	4,984,340	41,616	0.921	1,154	21	03/19/20	prior	2	4	6
Spain	47,007,367	30,774	0.893	235,000	28,628	03/2/20	8	13	13	13
USA	328,239,523	67,426	0.920	1,610,000	92,513	03/8/20	4	12	12	12

2 METHODS

2.1 Data sources

The COVID-19 data used in this report is publicly available from The Center for Systems Science and Engineering of the Johns Hopkins University (JSU CCSE) (37), which maintains a Repository on Github (38).

2.2 Procedures

The transmission dynamics of the COVID-19 outbreak is usually described by a compartmental model (SEIR), where (S) susceptible – (E) exposed – (I) infected – (R) removed (39; 40). In a closed population of P_n individuals, the transitions between the compartments (cf. Fig. 1) are described through the following set of differential equations:

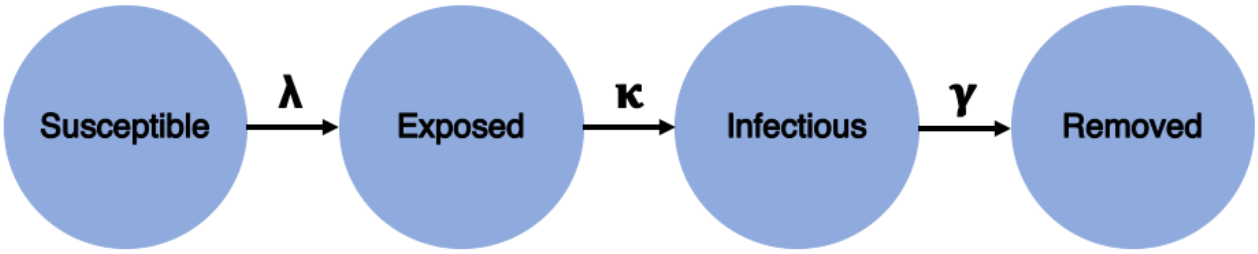

Figure 1. SEIR Model Structure

$$\frac{d}{dt}s(t) = -\bar{\lambda}s(t)i(t) \tag{1}$$

$$\frac{d}{dt}e(t) = \bar{\lambda}s(t)i(t) - \kappa e(t) \tag{2}$$

$$\frac{d}{dt}i(t) = \kappa e(t) - \gamma i(t) \tag{3}$$

$$\frac{d}{dt}r(t) = \gamma i(t) \tag{4}$$

where $\bar{\lambda} = \lambda/P_n$, λ is the infection rate, γ is the remove rate, and κ is the incubation rate. From these parameters, it is possible to calculate the basic reproduction ratio, $R_0 = \lambda/\gamma$. Thus, R_0 is not solely dependent on the infection rate, but also on the frequency of removals due to death or recoveries.

Assuming that the incubation period of the disease is instantaneous and the duration of infectivity is the same as the length of the disease, we can consider both groups E and I as contagious and $E(t) := I(t)$. Also, according to the Akaike information criterion (AIC), the simpler SIR model performs much better than an SEIR model in representing the information contained in the confirmed-case data available for COVID-19 (41). The basic SIR model is described by the set of Kermack–McKendrick equations (42):

$$\frac{d}{dt}s(t) = -\frac{\lambda}{P_n}s(t)i(t) \tag{5}$$

$$\frac{d}{dt}i(t) = \frac{\lambda}{P_n}s(t)i(t) - \gamma i(t) \tag{6}$$

$$\frac{d}{dt}r(t) = \gamma i(t) \tag{7}$$

2.3 Discrete-time SIR system parametric estimation in real-time

Traditionally, mathematical epidemiology models have been approached with a continuous-time perspective, due in part to the fact that these are more tractable mathematically (43). However, in order to use machine learning techniques there is a need for a discrete-time equivalent realization to cope with daily-sampled data (44). Due to the slow dynamics of the pandemics, a first order continuous to discrete Euler approximation can be applied to the Kermack–McKendrick equations.

For a general $f(t)$ function, a Backward discrete-time derivative approximation is:

$$\frac{d}{dt}f(t) := \frac{f(k+1) - f(k)}{T_s} = \Delta f(k+1)/T_s \tag{8}$$

where $T_s = 1$ is the sampling interval in days and $\Delta = 1 - q^{-1}$ is the discrete difference operator, defined in q^{-1}, the backward shift operator domain. The discrete-time approximations of eqs. (5) to (7) are given by the following difference equations, respectively:

$$s(k) = s(k-1) - \lambda(k-1)\frac{s(k-1)i(k-1)}{P_n} \tag{9}$$

$$i(k) = i(k-1) + \lambda(k-1)\frac{s(k-1)i(k-1)}{P_n} - \gamma(k-1)i(k-1) \tag{10}$$

$$r(k) = r(k-1) + \gamma(k-1)i(k-1) \tag{11}$$

The discrete-time SIR system described above considers time-varying parameters in order to continuously adapt the model as new data becomes available. Using the time-series of infections and removals (due to death or recovery), EQs. (9) to (11) can be used to estimate the model parameters.

Since Eq. (11) has an exclusive dependence with $\gamma(k)$, this poses a direct estimation problem which can be stated as "*for N registered samples, minimize the following quadratic cost function:*"

$$J_r = \frac{1}{2}\sum_{0}^{N} e_r^2(k) = \frac{1}{2}\sum_{k=0}^{N} [r(k) - \hat{r}(k)]^2 \tag{12}$$

Eq. (12) is based on the estimation error of $r(k)$. By applying the recursive least-squares method to minimize J_r, it is possible to optimally estimate $\hat{\gamma}(k)$ using the following equation:

$$\hat{r}(k) = r(k-1) + \hat{\gamma}(k-1)i(k-1) \tag{13}$$

We assume that the estimation error is Gaussian, $e_r(k) \sim (0, \sigma_{e_r}^2)$, with zero mean and variance $\sigma_{e_r}^2$, such that $r(k) = \hat{r}(k) + e_r(k)$. The estimator gain, the parametric estimation and error variance minimization are solved recursively, as follows:

$$L_r(k) = \frac{p_r(k-1)i(k-1)}{f_{\mathrm{f}} + i^2(k-1)p_r(k-1)} \tag{14}$$

$$\hat{\gamma}(k) = \hat{\gamma}(k-1) + L_r(k)\,[r(k) - \hat{r}(k)] \tag{15}$$

$$p_r(k) = [1 - L_r(k)i(k-1)]p_r(k-1) \tag{16}$$

where f_{f} is the forgetting factor of the recursive least-squares estimator and the closer it is to one the lesser the estimator forgets. Similarly, the error covariance matrix can be reset periodically to prioritize more recent data. Specifically in this work, $p_r(k)$ is a scalar vector, since only a single parameter is being estimated. Choices to initialize the covariance matrix may vary, depending on prior available covariance information or other positive definite matrix. The higher its magnitude the higher the estimator gain in the transitory dynamical stage of the estimation procedure.

In regard to Eq. (11) and the estimation of $\hat{\gamma}(k)$, since the infection numbers increase before any removal report is available during the first stages of the pandemic, during this period $\hat{\gamma}(k)$ tends to zero and eventually makes the time-varying estimated reproduction number tend to infinity:

$$\hat{R}_0(k) = \frac{\hat{\lambda}(k)}{\hat{\gamma}(k) \to 0} = \infty \tag{17}$$

Since Eq. (10) depends on both SIR parameters, due to its stronger dependence of $\hat{\gamma}(k)$ and Eq. (11), its estimated value is substituted into Eq. (10) and another recursive least-squares problem is constructed to

estimate $\hat{\lambda}(k)$, using

$$\hat{i}(k) = i(k-1) + \hat{\lambda}(k-1)\frac{\hat{s}(k-1)i(k-1)}{P_n} - \hat{\gamma}(k)i(k-1) \tag{18}$$

The solution is akin to the minimizing the estimation error $e_i(k) = i(k) - \hat{i}(k)$, using the following equations for estimator gain, parametric estimation update, and error covariance minimization, respectively:

$$\hat{i}(k) = i(k-1) + \hat{\lambda}(k-1)\frac{\hat{s}(k-1)i(k-1)}{P_n} - \hat{\gamma}(k)i(k-1) \tag{19}$$

$$\hat{\lambda}(k) = \hat{\lambda}(k-1) + L_i(k)\left[i(k) - \hat{i}(k)\right] \tag{20}$$

$$p_i(k) = \left\{1 - L_i(k)\left[\frac{\hat{s}(k-1)i(k-1)}{P_n}\right]\right\} p_i(k-1) \tag{21}$$

Time-series data for Eq. (9) is not available and the evolution of the Susceptible compartment in time is in fact estimated based on the known initial condition (i.e., the population P_n) and on the estimated $\hat{\lambda}(k)$. Thus, it is always estimated and fed back to (19), such that the correct form to represent it, within this time-varying SIR model, is by rewriting Eq. (9) based on the estimated Susceptible:

$$\hat{s}(k) = \hat{s}(k-1) - \hat{\lambda}(k)\frac{\hat{s}(k-1)i(k-1)}{P_n} \tag{22}$$

The estimation of the time-varying reproduction number, based on the derived discrete-time SIR model, is given by:

$$\hat{R}_0(k) = \frac{\hat{\lambda}(k)}{\hat{\gamma}(k)} \tag{23}$$

We also adopted two modifications to the nominal $\hat{R}_0(k)$ equation: a moving 4-day average to compensate for the randomness of daily updates on incidence data, as seen in the German Daily Situation Report of the Robert Koch Institute on COVID-19 (45), and the proportion of susceptible individuals in the population, known as the effective reproduction number (14):

$$\bar{R}_0(k) = \frac{\hat{s}(k)}{P_n}\left[\frac{\hat{R}_0(k) + \hat{R}_0(k-1) + \hat{R}_0(k-2) + \hat{R}_0(k-3)}{4}\right] \tag{24}$$

With this formulation, it is possible to analyze the transmission ratio of the pandemics on a daily basis, as with a sensor. Besides, estimations of the transmission ratio produce a dynamic representation from the perspective of the time-series of $\hat{R}_0(t)$, which allows the modeling of its dynamics and its randomness in order to assess stochastic properties correlated to the time-varying reproduction number, which might reflect how health authorities have been handling the challenges posed by the pandemics in each country considered in this work.

2.4 Modeling of R_0

Henceforth we assume that we are able to estimate the reproduction number on a daily basis and it is thus possible to consider it as another output of the proposed pandemic model. Thus, relying on the time-series of $\hat{R}_0(k)$ and knowing it is correlated with the number of infected and removed individuals, we deploy machine learning techniques to identify a dynamic system that fits the data.

Differently from the real-time monitoring/sensing procedure adopted to estimate $\hat{R}_0(k)$, now we are interested in obtaining a general model that can describe the dynamics of a system, $\hat{R}_0(q^{-1})$, for a certain period of interest. In order to model such a system, we use the non-recursive least-squares estimation technique (46) and propose a black-box polynomial model:

$$\hat{R}_0(q^{-1}) = \frac{\hat{B}(q^{-1})I(q^{-1}) + \hat{C}(q^{-1})R(q^{-1}) + e_{R_0}(q^{-1})}{\hat{A}(q^{-1})} \tag{25}$$

$$\begin{aligned} \hat{A}(q^{-1}) &= 1 + \hat{a}_1 q^{-1} + \hat{a}_2 q^{-2} \\ \hat{B}(q^{-1}) &= \hat{b}_0 q^{-1} + \hat{b}_1 q^{-2} \\ \hat{C}(q^{-1}) &= \hat{c}_0 q^{-1} + \hat{c}_1 q^{-2} \end{aligned} \tag{26}$$

where $e_{R_0}(q^{-1})$ is the Gaussian process based on the estimation error $e_{R_0}(k)$ of estimated polynomials shown in (26).

This second-order autoregressive with exogenous inputs (ARX) based model structure is assumed considering the fundamental simplicity of the SIR model, in which the Infected and Removed systems together form a second-order system.

The non-recursive least-squares estimator is a batch processing technique, used to optimally estimate the set of parameters that minimizes a quadratic performance index as the one shown in (12), but using a vector-matrix form of error, $\mathbf{e}_{R_0} = \hat{\mathbf{R}}_\mathbf{0} - \mathbf{\Phi}\hat{\theta}$. This vector-matrix system is defined as:

$$\mathbf{e}_{R_0} = \begin{bmatrix} e(0) & \cdots & e(N) \end{bmatrix}^T \tag{27}$$

$$\hat{\mathbf{R}}_\mathbf{0} = \begin{bmatrix} \hat{R}_0(0) & \cdots & \hat{R}_0(N) \end{bmatrix}^T \tag{28}$$

$$\hat{\theta}^\mathbf{T} = \begin{bmatrix} \hat{a}_1 & \hat{a}_2 & \hat{b}_0 & \hat{b}_1 & \hat{c}_0 & \hat{c}_1 \end{bmatrix} \tag{29}$$

$$\mathbf{\Phi}^\mathbf{T} = \begin{bmatrix} \phi^T(0) & \cdots & \phi^T(N) \end{bmatrix} \tag{30}$$

The above equations represent, respectively, the vector of errors, the vector of observed outputs, the estimated parameters vector and the matrix of regressors. The latter is based on the vectors of regressors formed up to N registered samples, with such vectors defined as:

$$\phi^T(k) = \begin{bmatrix} \hat{r}_0(k-1) & \hat{r}_0(k-2) & i(k-1) & i(k-2) & r(k-1) & r(k-2) \end{bmatrix} \tag{31}$$

The solution to obtain the estimated parameters is straightforward and given by

$$\hat{\theta} = \left(\mathbf{\Phi^T \Phi}\right)^{-1} \mathbf{\Phi^T \hat{R}_0} \tag{32}$$

By assuming an ARX linear model it is possible to evaluate the pandemics from the perspective of linear stochastic systems theory, assessing how the reproduction number decays linearly in time in different countries and how the random nature of events associated with the pandemics affects the model's uncertainties, i.e., $e_{R_0}(k) \sim (0, \sigma^2_{e_{R_0}})$. This calculated uncertainty can give us some clues regarding the effectiveness of pandemic control measures.

We propose that by analyzing the linear power of the $\hat{R}_0(q^{-1})$ estimation error, i.e. $\sigma^2_{e_{R_0}}$, we can use this stochastic property to generate *stochastic ratio* curves based on the estimated linearized model of Eq. (25). The higher the variability associated with the stochastic ratio the higher we expect the variance or linear power of the error to be and, consequently, the more uncertain is the official COVID-19 data reported by health authorities.

3 RESULTS

We used publicly available data to validate the algorithms and the estimated time-varying SIR model parameters. This section is organized in the following way: we first present the estimated results based on the number of infectious and removals available from different countries. Then, we present time-series of daily estimates on $R_0(k)$ based on the RLSA-estimated $\hat{\lambda}(k)$ and $\hat{\gamma}(k)$. These results are followed by the presentation of linearized estimated outputs generated by the ARX-based $\hat{R}(q^{-1})$ models (shown in (25)) using non-recursive LSA batch processing of 30-days of $R_0(k)$ estimates. The modeling residuals were assumed to be Gaussian (zero mean). Estimated error variances were used to produce 200 discrete Gaussian sequences as surrogates to additive white noises, depicted by $e_{R_0}(k)$ in (25). These 200 additive noise sequences were used to generate, for every analyzed country, 200 stochastic reproduction ratios, shown together with the linearized ratio and the real-time estimated ratio.

3.1 Discrete-time SIR model estimation results

The estimators were implemented with a forgetting factor of 0.98 and the covariance matrices were reset to ones every seven days to prioritize more recent data (47). Also, we adopted a moving 4-day average for $\hat{R}_0(k)$, shown in (24), to compensate for daily random effects (45). We only used data from 22 March 2020 onwards, when all eight countries already had more than 100 infections reported.

Figures 2 and 3 show the dynamics of both Infected and Removed cases using the SIR model.

Fig. 4 presents the $\hat{R}_0(k)$ of the eight different countries during a period of two weeks. Despite the large variability of both Argentina and the United States, there are other countries that are already reducing the number of new infections and where the frequency of removals has increased, such as in Germany, Italy, and New Zealand (see Table 2).

The trajectory of the curves shown in Fig. 4 can also be associated with some extraordinary events that occurred during the same period. For example, after 3 May 2020, when some U.S. states had relaxed social distancing guidelines, it is possible to observe a corresponding phasic increase in the basic reproduction ratio of the U.S., which was also reported in the Washington Post on 9 May 2020 (48).

One interesting trajectory shown in Fig. 4 regards Argentina. For most of the time, the estimated reproduction ratio of Argentina was one of the highest and comparable only to the U.S., despite its low number of infected individuals (cf. Fig. 2). This apparent contradiction is related to the low number

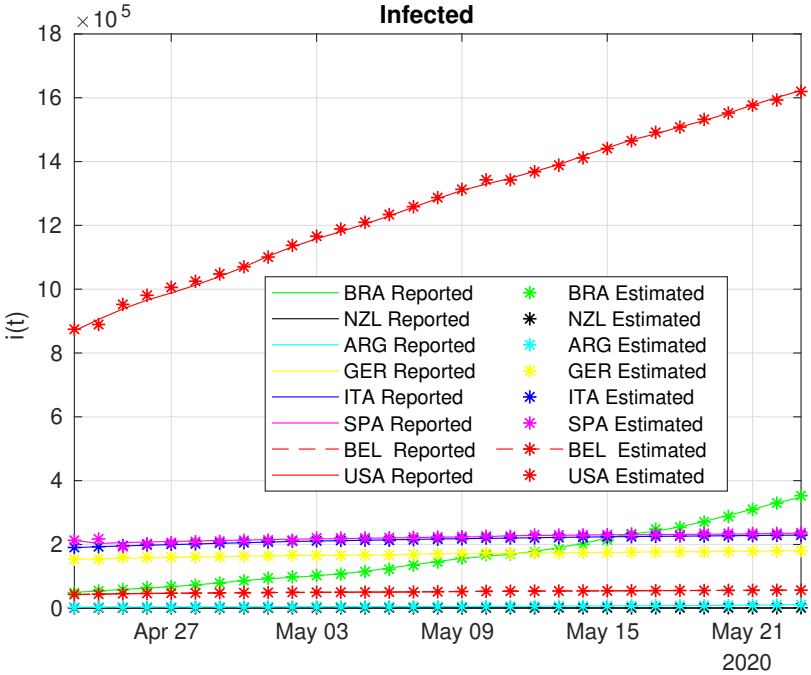

Figure 2. Estimation of Infected individuals using the RLSA technique.

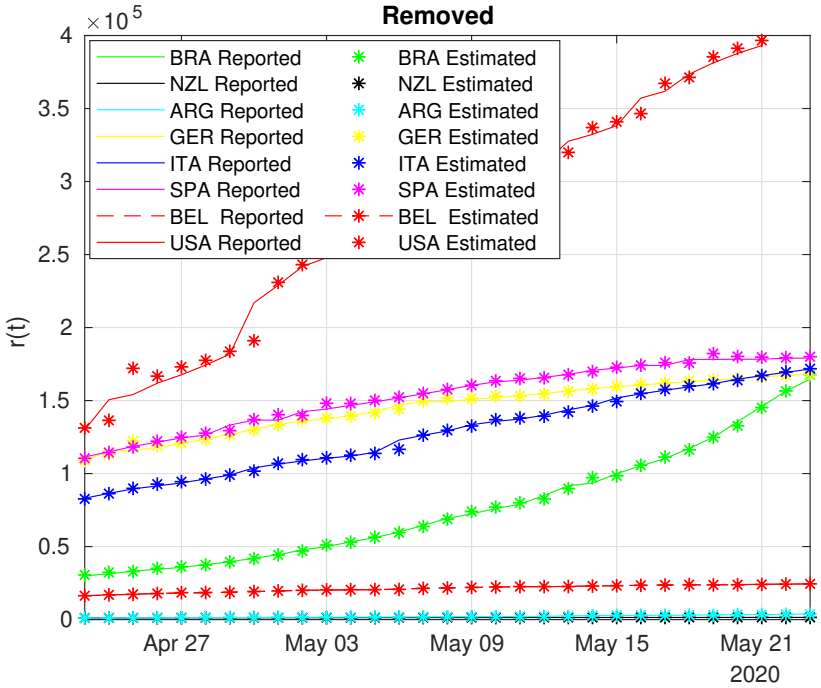

Figure 3. Estimation of Removed individuals using the RLSA technique.

of removals at the beginning of the pandemic that tends to raise the $\hat{R}_0(k)$ (see Eq. 17). Recoveries in Argentina, based on data of May 16 (cf. Table 2), are approximately 30.5% of its total infected, close to

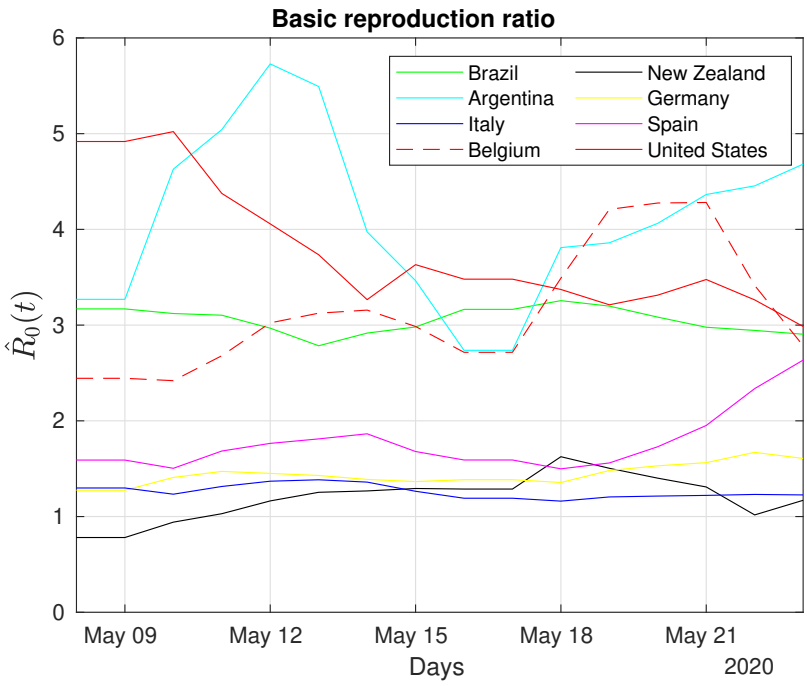

Figure 4. Two weeks of daily monitoring the pandemics: dynamics of estimated reproduction number according to available official data and SIR model.

Table 2. Recoveries (May 21, 2020).

Country	Recoveries
Argentina	30.5%
Belgium	26.6%
Brazil	40.5%
Germany	88.2%
Italy	59.0%
New Zealand	96.6%
Spain	70.3%
United States	23.5%

Belgium with 26.6% and whose transmission ratio is below Argentina's, reinforcing the notion that the transmission ratio is not a static parameter and is best approached by dynamical systems theory.

The results displayed in Figure 4 seem at odds with a recent report that claimed that Brazil's R_0 had recently dropped to 1.4 (49). However, on the same day a Brazilian newspaper quoted this report, the country had a record number of new infections and deaths (50). This reinforces the notion that machine learning methods might give better and faster clues regarding the severity of the pandemic in terms of $\hat{R}_0(k)$.

3.2 Assessing the pandemics through the R_0 dynamics

Figures 5 to 8 show $\hat{R}_0(k)$ for a 30-day period for the eight investigated countries. These figures show linearized R_0 together with the respective real-time estimated data that originated this second stage estimation. The variance of the estimation error is then used to generate Gaussian sequences which are superimposed on the linearized R_0 estimate, the stochastic ratio which synthetically reproduces the

stochasticity and uncertainties of R_0 estimates. Both Germany and Italy (Fig. 5) are through a period of consistently decaying $R_0(t)$. Despite the hard way COVID-19 hit Italy before, its stochastic ratio currently is among the lowest.

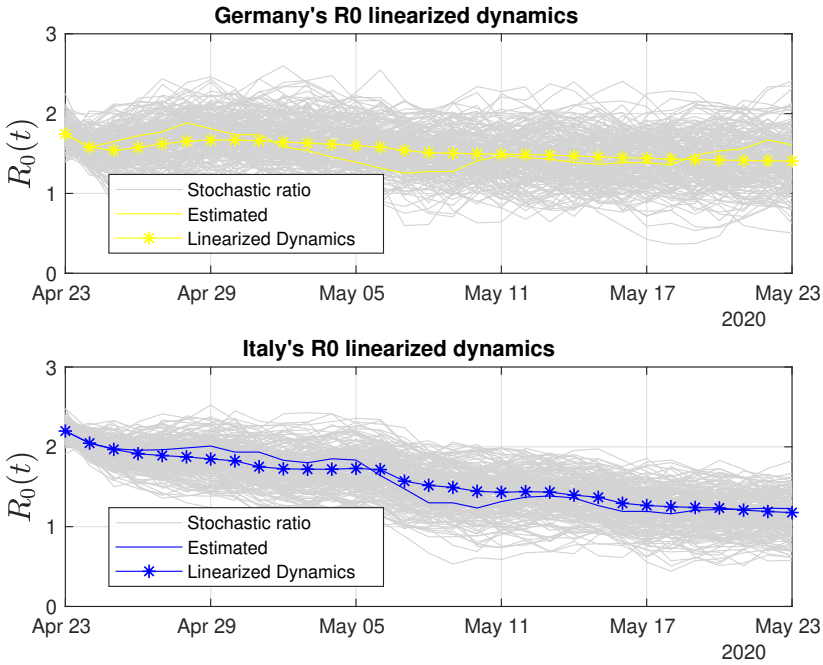

Figure 5. R0 linearized dynamics based on Germany's and Italy's after 30-days of real-time monitoring. Both countries have low variability in the stochastic ratio curve.

Brazil and the United States (Fig. 6) are the two most populous countries of our sample and the most hard-hit by COVID-19 among them. Both countries also have been struggling with their uneven response to the pandemics (51). This outcome is captured by our R_0 sensors, with the Brazilian stochastic ratio being in decrease, as Recovered data becomes more available (see Fig. 6).

In Fig. 7 we present linearized R_0 estimates for Spain and Belgium. Spain's estimates have suffered the influence of annotation errors (by subtracting infected individuals in 24 Apr. 2020) which eventually provoked oscillations in the real-time ratio estimate, making it zero-cross on April 27th. Such an error was washed out by the linearized estimate and has certainly contributed to its increased degree of uncertainty (see Table 3).

Belgium's degree of uncertainty was the worst among all the countries, at least during the period we analyzed. One possible clue to understand why Belgium's stochastic ratio became so variable is to consider the impossibility to linearize its dynamics and the associated increase in error. However, our estimation procedure considers the error as Gaussian, and its mean value in fact approaches zero. Thus, the linearized estimate based on 30-days of data has a high probability to be close to the values shown.

Belgium has also shown an increased R_0 during the analyzed period. The low number of recoveries (see Table 2) influenced the frequency of removals and *the basic reproduction number, which is the ratio between the frequency of infections and the frequency of removals* (see eq. (23)). Even though a recent report (52) proposed that Belgium's R_0 at the beginning of May 2020 was 0.8, according to our adaptive

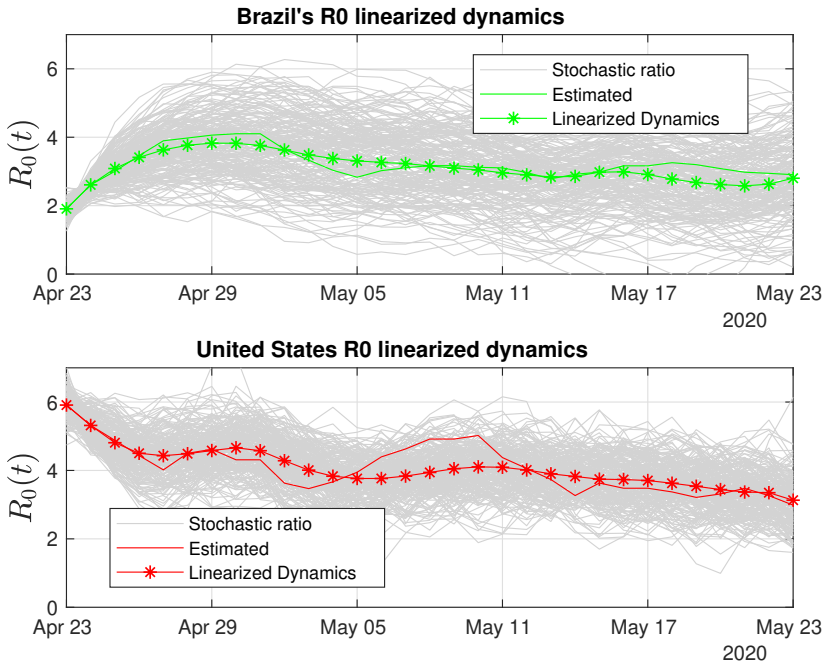

Figure 6. Brazil and the United States of America showing increased variability in the stochastic ratio curve, which copes with data uncertainty and difficulties to face and control COVID-19 spread.

Table 3.
Pandemic's uncertainty (up to 23 May 2020).

Country	Power	Uncertainty
Argentina	0.3099	30.9%
Belgium	0.2426	24.2%
Brazil	0.0546	5.4%
Germany	0.0220	2.2%
Italy	0.0113	1.1%
New Zealand	0.0349	3.4%
Spain	0.0521	5.2%
United States	0.1835	18.3%

SIR-based R_0 estimations, the current situation in Belgium is uncertain with a probable R_0 close to 3.0, which is the mean value of 30-days linearized dynamics (see Table 4).

Figure 8 shows the trends for Argentina and New Zealand. The R_0 estimates for the two countries, which have the lowest number of infections of the group, suggests that several recent R_0 reports are considering a ratio solely based on the number of infections and population, without estimating and taking into account the frequency of removals, which makes R_0 increase.

We compiled a table of the uncertainty of the estimated stochastic ratios by country (see Table 3), where the percentage of uncertainty is proportional to the linear power of the Gaussian estimation error. Table 4 shows the estimated R_0 estimates by country and demonstrate the usefulness of our approach to provide a real-time picture of the pandemic that can be used to support decision-making.

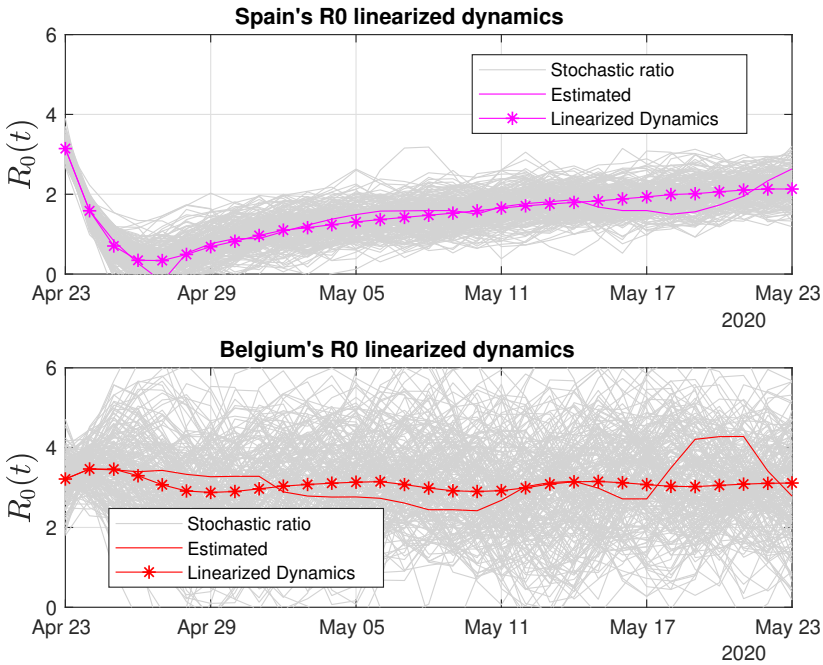

Figure 7. Spain's and Belgium's R0 linearized dynamics. Both countries are exhibiting high variability of the stochastic ratio, which warns of possible problems with the official data or difficulties to stabilize the pandemics.

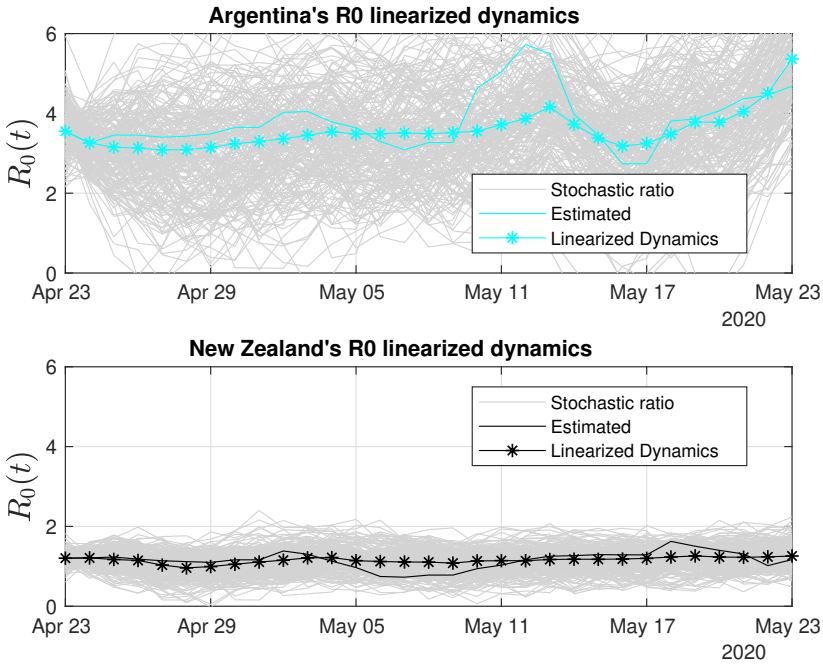

Figure 8. Argentina's and New Zealand's linearized R0 dynamics. New Zealand's has been showing a stable control of the pandemics, which copes with the low variability stochastic ratio curve.

4 DISCUSSION

Almost two months after the WHO declared COVID-19 a pandemic much is still unknown about the SARS-CoV-2 virus while it spreads around the world. Research efforts for the development of a vaccine are

Table 4.
R0 estimates (on/up to 23 May 2020).

Country	Real-time	Linearized	30-days Mean
Argentina	4.6	5.3	3.5
Belgium	2.7	3.1	3.0
Brazil	2.9	2.8	3.0
Germany	1.6	1.4	1.5
Italy	1.2	1.1	1.5
New Zealand	1.1	1.2	1.1
Spain	2.6	2.1	1.4
United States	2.9	3.1	4.0

being accelerated but still is a distant target (53). At this moment, however, the most effective interventions are NPIs aimed at reducing SARS-CoV-2 transmission rates in order to decrease mortality by increasing the fraction of severe cases having access to scarce medical resources, such as mechanical ventilation (54).

Mathematical modeling is a valuable instrument to gauge epidemics' dynamics evaluate the effects of interventions aimed to control its spread. A crucial parameter is R_0, the basic reproduction number which is closely followed by health officials and the public alike and translates into the infectiousness of a disease. As the world transitions to a gradual release from social distancing measures, many questions still remain about the SARS-CoV-2 virus and there is all but the inevitability of secondary waves of infection ahead. Thus, the continuing use of mathematical models to track the disease will remain a necessity. However, the utility of models depends on the quality of the data they are fed. And there are many doubts regarding the data on COVID-19 cases publicly available. For instance, due to the lack of widespread testing and sub notifications on the cause of deaths due to the disease, it is almost impossible to have a definitive picture of transitions between compartments of the most used model to track the disease. Thus, in this work, we provide some answers to that degree of uncertainty regarding the results presented by those epidemiological models.

During this pandemic, we are interacting with it in an unprecedented manner, trying to control its spread in real-time, as in feedback control systems. Therefore, we are interfering in COVID-19's dynamics, affecting its behavior and parameters, and even discussing the development of real-time monitoring techniques to work against it using automation and control systems technologies. However, this closed-loop control system needs humans in the loop and some data annotation mistakes may occur, such as: seasonal dynamic changes related to workers shifts, weekend reduced reports and even possible political interference in the data log may compromise the stability of estimators due to these disturbances.

Our comparative approach shows that the strict measures adopted by some countries managed to stabilize the epidemics, such as Germany, Italy, Spain, and New Zealand. In others, such as the U.S. and Brazil, the delay in adopting such measures and lack of coordination proved decisive to keep R_0 values high and with a high degree of uncertainty. The real-time estimation of model parameters such as R_0 allows the gain of essential insight into the underlying epidemic process and provides robustness in the face of imperfect data. This strategy can be used to provide useful information and be implemented as an online tracking application providing information about the dynamics of the pandemic to health officials and the public at large.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

A.S. and A.P. conceived and wrote the manuscript.

DATA AVAILABILITY STATEMENT

The datasets [GENERATED/ANALYZED] for this study can be found in the [NAME OF REPOSITORY] [LINK].

REFERENCES

1. WHO announces COVID-19 outbreak a pandemic. Available at: `http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic`

2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, et al. A Novel Coronavirus from Patients with Pneumonia in China 2019. *New England Journal of Medicine* (2020) **382**:727–733. doi:10.1056/nejmoa2001017

3. Global situation report No. 118. Available at: `https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200517-covid-19-sitrep-118.pdf?sfvrsn=21c0dafe_6`

4. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine* (2020) **172**:577. doi:10.7326/m20-0504

5. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, Taylor J, Spicer K, Bardossy AC, Oakley LP, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *New England Journal of Medicine* (2020) doi:10.1056/nejmoa2008457

6. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic Transmission the Achilles' Heel of Current Strategies to Control Covid-19. *New England Journal of Medicine* (2020) doi:10.1056/nejme2009758

7. El-Sadr WM, Justman J. Africa in the Path of Covid-19. *New England Journal of Medicine* (2020) doi:10.1056/nejmp2008193

8. Kirby T. South America prepares for the impact of COVID-19. *The Lancet Respiratory Medicine* (2020) doi:10.1016/s2213-2600(20)30218-6

9. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *The Lancet* (2020) **395**:931–934. doi:10.1016/s0140-6736(20)30567-5

10. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nature Human Behaviour* (2020) doi:10.1038/s41562-020-0887-9

11. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A. The outbreak of COVID-19 coronavirus and its impact on global mental health. *International Journal of Social Psychiatry* (2020)002076402091521. doi:10.1177/0020764020915212

12. Xu S, Li Y. Beware of the second wave of COVID-19. *The Lancet* (2020) **395**:1321–1322. doi:10.1016/s0140-6736(20)30845-x

13. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. *Nature* (2006) **442**:448–452. doi:10.1038/nature04795

14. Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty identifiability, and forecasts. *Infectious Disease Modelling* (2017) **2**:379–398. doi:10.1016/j.idm.2017.08.001

15. MKendrick AG. Applications of Mathematics to Medical Problems. *Proceedings of the Edinburgh Mathematical Society* (1925) **44**:98–130. doi:10.1017/s0013091500034428

16. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A Containing Papers of a Mathematical and Physical Character* (1927) **115**:700–721. doi:10.1098/rspa.1927.0118

17. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* (2020) **368**:489–493. doi:10.1126/science.abb3221

18. Weitz JS, Beckett SJ, Coenen AR, Demory D, Dominguez-Mirazo M, Dushoff J, Leung C-Y, Li G, Magalie A, Park SW, et al. Intervention Serology and Interaction Substitution: Modeling the Role of Shield Immunity in Reducing COVID-19 Epidemic Spread. (2020) doi:10.1101/2020.04.01.20049767

19. Viceconte G, Petrosillo N. COVID-19 R0: Magic number or conundrum?. *Infectious Disease Reports* (2020) **12**: doi:10.4081/idr.2020.8516

20. Ridenhour B, Kowalik JM, Shay DK. Unraveling R0: Considerations for Public Health Applications. *American Journal of Public Health* (2014) **104**:e32–e41. doi:10.2105/ajph.2013.301704

21. Riley S. Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. *Science* (2003) **300**:1961–1966. doi:10.1126/science.1086478

22. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the Basic Reproduction Number (R0). *Emerging Infectious Diseases* (2019) **25**:1–4. doi:10.3201/eid2501.171901

23. Obadia T, Haneef R, Boëlle P-Y. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Medical Informatics and Decision Making* (2012) **12**: doi:10.1186/1472-6947-12-147

24. Bettencourt LMA, Ribeiro RM. Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases. *PLoS ONE* (2008) **3**:e2185. doi:10.1371/journal.pone.0002185

25. Wallinga J. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *American Journal of Epidemiology* (2004) **160**:509–516. doi:10.1093/aje/kwh255

26. Cauchemez S, Boëlle P-Y, Donnelly CA, Ferguson NM, Thomas G, Leung GM, Hedley AJ, Anderson RM, Valleron A-J. Real-time Estimates in Early Detection of SARS. *Emerging Infectious Diseases* (2012) **12**:110–113. doi:10.3201/eid1201.050593

27. Cauchemez S, Boëlle PY, Thomas G, Valleron AJ. Estimating in real time the efficacy of measures to control emerging communicable diseases.. *Am J Epidemiol* (2006) **164**:591–7.

28. Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, Sachs JD, Wang Z, Meyers LA, Singer BH, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proceedings of the National Academy of Sciences* (2020) **117**:9122–9126. doi:10.1073/pnas.2004064117

29. Nievergelt Y. A tutorial history of least squares with applications to astronomy and geodesy. *Journal of Computational and Applied Mathematics* (2000) **121**:37–72. doi:10.1016/s0377-0427(00)00343-5

30. Park G, Choi SB. An Integrated Observer for Real-Time Estimation of Vehicle Center of Gravity Height. *IEEE Transactions on Intelligent Transportation Systems* (2020)1–12. doi:10.1109/tits.2020.2988508

31. Puttagunta V, Kalpakis K. Adaptive Methods for Activity Monitoring of Streaming Data.. in *ICMLA, vol. 2*

32. Jaros R, Martinek R, Kahankova R, Koziorek J. Novel hybrid extraction systems for fetal heart rate variability monitoring based on non-invasive fetal electrocardiogram. *IEEE Access* (2019) **7**:131758–131784.

33. Smirnova A, Chowell G. A primer on stable parameter estimation and forecasting in epidemiology by a problem-oriented regularized least squares algorithm. *Infectious Disease Modelling* (2017) **2**:268–275. doi:10.1016/j.idm.2017.05.004

34. Banks HT, Hu S, Thompson WC. *Modeling and inverse problems in the presence of uncertainty*. CRC Press (2014).

35. Shen J. A Recursive Bifurcation Model for Predicting the Peak of COVID-19 Virus Spread in United States and Germany. (2020) doi:10.1101/2020.04.09.20059329

36. Binny RN, Hendy SC, James A, Lustig A, Plank MJ, Steyn N. Effect of Alert Level 4 on effective reproduction number: review of international COVID-19 cases. (2020) doi:10.1101/2020.04.30.20086934

37. Novel Coronavirus (COVID-19) Cases Data - Humanitarian Data Exchange. Available at: `https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases`

38. CSSEGISandData/COVID-19. Available at: `https://github.com/CSSEGISandData/COVID-19`

39. Anderson RM, May RM. *Infectious diseases of humans: dynamics and control*. Oxford university press (1992).

40. Brauer F, Castillo-Chavez C. "Epidemic Models," in *Texts in Applied Mathematics* (Springer New York), 345–409. doi:10.1007/978-1-4614-1686-99

41. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic?. *Infectious Disease Modelling* (2020) **5**:271–281. doi:10.1016/j.idm.2020.03.001

42. Anderson RM, May RM. Population biology of infectious diseases: Part I. *Nature* (1979) **280**:361–367. doi:10.1038/280361a0

43. Brauer F, Feng Z, Castillo-Chavez C. Discrete epidemic models. *Mathematical Biosciences & Engineering* **7**:1.

44. Gómez S, Arenas A, Borge-Holthoefer J, Meloni S, Moreno Y. Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *EPL (Europhysics Letters)* (2010) **89**:38009. doi:10.1209/0295-5075/89/38009

45. RKI - Coronavirus SARS-CoV-2 - Aktueller Lage-/Situationsbericht des RKI zu COVID-19. Available at: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Gesamt.html

46. Ljung L. *System identification (2nd ed.): theory for the user*. Prentice Hall PTRUpper Saddle River, NJUnited States (1999).

47. Kevin Systrom. Available at: http://systrom.com/blog/author/ksys1983/

48. Coronavirus flares as states and countries ease social distancing guidelines. Available at: https://www.washingtonpost.com/national-security/coronavirus-flares-as-states-and-countries-ease-social-distancing-guidelines 2020/05/09/cccb3c0c-9219-11ea-9e23-6914ee410a5f_story.html

49. Filho RL, Lichtenthaler DG. A dynamic model for Covid-19 in Brazil.. (2020) doi:10.1101/2020.05.10.20097550

50. Brazil Registers 888 New Deaths from Coronavirus; Total Number of Deaths Exceeds 18 thousand. Available at: https://www1.folha.uol.com.br/internacional/en/scienceandhealth/2020/05/brazil-registers-888-new-deaths-from-coronavirus-total-number-of-deaths-exce shtml

51. Brazil, Once a Leader, Struggles to Contain Virus Amid Political Turmoil. (2020) Available at: https://www.nytimes.com/2020/05/16/world/americas/virus-brazil-deaths.html

52. Basic reproduction number of novel coronavirus in Belgium falls to 0.6. Available at: https://www.vrt.be/vrtnws/en/2020/05/04/basic-reproduction-number-of-novel-coronavirus-in-belgium-falls/

53. Graham BS. Rapid COVID-19 vaccine development. *Science* (2020)eabb8923. doi:10.1126/science.abb8923

54. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, Zhang C, Boyle C, Smith M, Phillips JP. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *New England Journal of Medicine* (2020) doi:10.1056/nejmsb2005114