ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-7605-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICASSP39728.2021.9414762

AN IMPROVED DATA DRIVEN DYNAMIC SIRD MODEL FOR PREDICTIVE MONITORING OF COVID-19

Pushpendra Singh[*†], *Amit Singhal* [δ], *Binish Fatimah* [△], *Anubha Gupta*[‡]

[†]Deptt. of ECE, National Institute of Technology Hamirpur, Hamirpur, India.
[δ]Deptt. of ECE, Bennett University, Greater Noida, India.
[△]Deptt. of ECE, CMR Institute of Technology, Bengaluru, India.
[‡]SBILab, Deptt. of ECE, IIIT-Delhi, India.

ABSTRACT

COVID-19 pandemic spreaded across the world in early 2020. It forced many countries to impose lockdown to prevent surge in the number of infected cases. There has been a huge impact on social and economic activities worldwide. In this work, we carry out the functional modeling of COVID-19 infection trends using two models: the Gaussian mixture model (GMM) and the composite logistic growth model (CLGM). Unlike the traditional SIRD models that use numerical data fitting, we utilize the best data-fitted curves employing GMM and/or CLGM to construct the Susceptible-Infected-Recovered-Dead (SIRD) pandemic model. Further, we derive the explicit expressions of time-varying parameters of the SIRD model unlike most works that consider static parameters without any closed form solution. The proposed parameterized dynamic SIRD model is generically applicable to any pandemic, can capture the day-to-day dynamics of the pandemic and can assist the governing bodies in devising efficient action plans to deal with the prevailing pandemic.

Index Terms— COVID-19 modeling, Gaussian mixture model, Composite logistic growth function, Time-varying reproduction number, dynamic SIRD model

1. INTRODUCTION

The world is dealing with a pandemic situation caused by the rapid spread of novel coronavirus disease (COVID-19). It started towards the end of year 2019 and has resulted in more than 1.05 million deaths worldwide [1] and loss of livelihoods for many more. COVID-19 is highly contagious and hence, strict social distancing norms need to be followed to arrest the spread of this disease [2]. In order to ensure the safety of the population, a lot of restrictions have been imposed by the governing authorities. The virus primarily enters the body through mouth and nose [3]. Thus, people have been advised to wear face masks, whenever they step out of their homes.

Biologists across the globe are trying to develop a vaccine against this disease, while some researchers are analyzing this challenging situation by providing mathematical models to capture and predict the trend of infections. These models would assist the authorities in devising appropriate strategies to cope up with the pandemic more effectively. Some parts of the world are experiencing a bigger second wave of fresh cases [1], after it was thought that the pandemic had already passed. Hence, a careful examination of the current situation is paramount in this need of the hour.

Susceptible-Infected-Removed (SIR) [4, 5] is the most popular and widely-used mathematical model to assess the daily variations in the number of people getting infected. It has been applied in [6] for early prediction of COVID-19 outbreak in mainland China. Some variations of the SIR model have also been proposed in the literature [7, 8]. For example, a Susceptible-Infected-Recovered-Dead (SIRD) model is obtained by considering 'recovered' and 'dead' cases separately instead of a single category of 'removed' cases. This model is studied in [7] to analyze the trend of infections in China, Italy and France. Susceptible-Exposed-Infectious-Removed (SEIR) model is explored in [8]. Amongst other efforts to analyze the pandemic, authors in [9] suggest the use of long short-term memory (LSTM) model for predicting the number of cases in the near future. Prediction is performed using a Gaussian mixture model (GMM) in [10] while an auto-regressive integrated moving average (ARIMA) model is employed in [11]. A recent study [12] identifies situational information from social media platforms to gather relevant information, enabling a quick response by the local authorities.

The original SIRD model has the following limitations: (1) A closed form solution of the SIRD model is not available in the literature. (2) It assumes all parameters to be constant, while these parameters change with time in the real scenario. (3) The parameters are first estimated from the given data using the optimization algorithms. Then, the SIRD model is solved by numerical methods. Since many independent datasets are used, the problem translates to a multi-objective function minimization and is computationally complex. (4) Generally, the SIRD model uses a single wave to fit the data,

*Corresponding author (spushp@nith.ac.in; spushp@gmail.com). This work was supported by NIT Hamirpur, (HP) India.

which are actually multi-modal. Hence, it is not able to predict and track the pandemic accurately. The modified SIRD model proposed in this work solves all these issues and uses the actual data for infected, recovered and dead cases. Thus, it provides the best results for continuous monitoring and prediction of the pandemic. Since, India is the second most populous country of the world and has witnessed a large number of cases, results on COVID-19 data modeling are presented for India. But the method is generic and can be applied to any other country.

The main contributions of this study are as follows: (1) We have connected the SIRD model with Gaussian mixture model (GMM) and/or the composite logistic growth model (CLGM). (2) Each of the SIRD waves obtained from the real-data for COVID-19 are decomposed into multiple waves to accommodate the GMM and/or CLGM, thereby, capturing the multi-modal trend for data. (3) Instead of estimating parameters by considering them as static, we have derived the explicit expressions for the SIRD parameters considering them as functions of time. This looks natural because the death rate, growth rate, spread, mean infection time, etc. of a pandemic change over time depending on the local efforts, citizens' response, treatment availability, and administrative decisions on managing the pandemic. Thus, these time-varying parameters reflect the dynamic trends and predictions are made for the future. (4) The proposed framework can be used with other growth functions such as beta, Kumaraswamy, Irwin–Hall, Fréchet, or Weibull to build any variant of SIRD model.

The paper is organized as follows. The proposed SIRD model with GMM, CLGM, and dynamic parameter modeling is presented in Section-2. Simulation results are provided in Section-3. Conclusions are presented in Section-4.

2. PROPOSED DATA-DRIVEN SIRD MODEL

In this section, first we discuss the classical Susceptible-Infected-Recovered-Dead (SIRD) model. The classical SIRD epidemic model [4,7] is defined by the following initial-value problem:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{-\beta IS}{N}, \qquad S(0) = S_0 \geq 0, \tag{1a}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta SI}{N} - \gamma I - \mu I, \qquad I(0) = I_0 \geq 0, \tag{1b}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I, \qquad R(0) = R_0 \geq 0, \tag{1c}$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \mu I, \qquad D(0) = D_0 \geq 0, \tag{1d}$$

where $S(t)$, $I(t)$, $R(t)$, and $D(t)$ are the numbers of susceptible, infected, recovered, and dead cases, respectively, β is the contact/infection rate (i.e., the average number of contacts per person per unit time), γ is the recovery rate ($1/\gamma$ represents the average infectious period), and μ is the death rate. The total population size, assumed to be a constant, is obtained as

$$S(t) + I(t) + R(t) + D(t) = N. \tag{2}$$

It is evident from (1) and (2) that $\frac{\mathrm{d}S}{\mathrm{d}t} + \frac{\mathrm{d}I}{\mathrm{d}t} + \frac{\mathrm{d}R}{\mathrm{d}t} + \frac{\mathrm{d}D}{\mathrm{d}t} = 0$.

For an outbreak of the pandemic, $\frac{\mathrm{d}I}{\mathrm{d}t} > 0$, i.e., the change in the infected cases is positive implying a positive growth in infected cases. Thus, from (1b), one can obtain the reproduction number (RN), $\mathfrak{R}_0 = \frac{\beta S}{N(\gamma+\mu)} > 1$. It is an important parameter and refers to the number of fresh infections caused by an infected person. The value of $\mathfrak{R}_0 > 1$ indicates that the infection is growing with time, $\mathfrak{R}_0 = 1$ indicates flattening of the infection, while $\mathfrak{R}_0 < 1$ shows that the infection is decreasing and a pandemic outbreak will eventually disappear. We propose to decompose the data into multiple waves and obtain

$$I = \sum_{i=1}^{M_1} I_i, \quad R = \sum_{i=1}^{M_2} R_i, \quad D = \sum_{i=1}^{M_3} D_i, \tag{3a}$$

$$S = N - \sum_{i=1}^{M_1} I_i - \sum_{i=1}^{M_2} R_i - \sum_{i=1}^{M_3} D_i, \tag{3b}$$

where the total cumulative cases $C(t) = I(t) + R(t) + D(t)$ and $N = C(T)$ with T denoting the time-duration of the pandemic, i.e., $t \in [0, T]$.

The data for cumulative, infected, recovered and dead cases due to COVID-19 pandemic are available online for most of the countries. The waves for $C(t)$, $R(t)$, and $D(t)$ exhibit a monotonically increasing trend. Hence, they are fitted using integrals of GMM, while $I(t)$ can be fitted directly with real-data using GMM as

$$C(t) = \sum_{i=1}^{M_1} C_i(t) = \sum_{i=1}^{M_1} \int_0^t a_i \exp\left(-\left(\frac{\tau - b_i}{c_i}\right)^2\right) \mathrm{d}\tau, \tag{4a}$$

$$I(t) = \sum_{i=1}^{M_2} I_i(t) = \sum_{i=1}^{M_2} a_i \exp\left(-\left(\frac{t - b_i}{c_i}\right)^2\right), \tag{4b}$$

$$R(t) = \sum_{i=1}^{M_3} R_i(t) = \sum_{i=1}^{M_3} \int_0^t a_i \exp\left(-\left(\frac{\tau - b_i}{c_i}\right)^2\right) \mathrm{d}\tau, \tag{4c}$$

$$D(t) = \sum_{i=1}^{M_4} D_i(t) = \sum_{i=1}^{M_4} \int_0^t a_i \exp\left(-\left(\frac{\tau - b_i}{c_i}\right)^2\right) \mathrm{d}\tau, \tag{4d}$$

and the daily cases are fitted and obtained as

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = \sum_{i=1}^{M_1} a_i \exp\left(-\left(\frac{t - b_i}{c_i}\right)^2\right). \tag{5}$$

The other derivatives can be computed from (4) in a similar manner. Likewise, $C(t)$, $I(t)$, $R(t)$, and $D(t)$ can be fitted using the CLGM as

$$C(t) = \sum_{i=1}^{M_1} C_i(t) = \sum_{i=1}^{M_1} \frac{K_i}{1 + A_i \exp(-r_i(t-\tau_i))}, \tag{6a}$$

$$I(t) = \sum_{i=1}^{M_2} I_i(t) = \sum_{i=1}^{M_2} \frac{K_i A_i r_i \exp(-r_i(t-\tau_i))}{(1 + A_i \exp(-r_i(t-\tau_i)))^2}, \tag{6b}$$

$$R(t) = \sum_{i=1}^{M_3} R_i(t) = \sum_{i=1}^{M_3} \frac{K_i}{1 + A_i \exp(-r_i(t-\tau_i))}, \tag{6c}$$

$$D(t) = \sum_{i=1}^{M_4} D_i(t) = \sum_{i=1}^{M_4} \frac{K_i}{1 + A_i \exp(-r_i(t-\tau_i))}, \tag{6d}$$

where the daily cases are given by

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = \sum_{i=1}^{M_1} \frac{K_i A_i r_i \exp(-r_i(t-\tau_i))}{(1 + A_i \exp(-r_i(t-\tau_i)))^2}. \tag{7}$$

The number of Gaussian (or LGM) waves and the parameters a_i, b_i, c_i (or K_i, A_i, r_i, τ_i) for each wave are estimated by the minimization of the objective function (with $\tau_1 = 0$) given by the sum of squares for residuals of values [13, 14]. The minimization process uses the simplex search method in order to estimate the optimal values of the unknown parameters. Finally, the time-varying parameters of the SIRD model are computed from (1) as

$$\beta = -\frac{\mathrm{d}S}{\mathrm{d}t} \times \frac{N}{IS}, \qquad \gamma = \frac{\mathrm{d}R}{\mathrm{d}t} \times \frac{1}{I}, \tag{8a}$$

$$\mu = \frac{\mathrm{d}D}{\mathrm{d}t} \times \frac{1}{I}, \quad \text{and} \quad \mathfrak{R}_0 = \frac{\beta S}{N(\gamma + \mu)}. \tag{8b}$$

3. RESULTS AND DISCUSSION

In this section, we present results obtained for COVID-19 progression in India using the proposed model. The data is obtained from the Worldometer [1] and WHO daily situation report [2]. The data from March 01, 2020 to October 01, 2020 is used for fitting the model and thereafter, the future values are predicted using the obtained model. Considering the actual data till October 16, 2020, the predicted values for the last 15 days are compared. The values for the root mean square error (RMSE) are provided for both the models in Table 1. Since the RMSE with both the growth function models are comparable for India, SIRD curves and parameter values are provided only for the GMM for the sake of brevity, although results can be obtained for CLGM in a similar manner.

Fig. 1 shows the trend of cumulative, susceptible, infected, recovered, and death cases estimated using GMM. The actual values are also shown for all of these. Since the data for the susceptible cases is not available, it has been estimated using the proposed model.

From Fig. 1, the estimated values are found to agree with the actual data within small margins of error. The error between the estimated and actual cases can be attributed to

Table 1: The RMSE values obtained by GMM and CLGM for fitting and prediction of SIRD waves

SIRD waves	GMM		CLGM	
	Fitting	Prediction	Fitting	Prediction
$C(t)$	4.25e+03	1.31e+04	4.03e+03	1.31e+04
$I(t)$	1.68e+04	5.44e+04	2.54e+04	5.55e+04
$R(t)$	3.59e+03	8.79e+03	3.56e+03	8.74e+03
$D(t)$	126.2	192.8	132.1	198.1

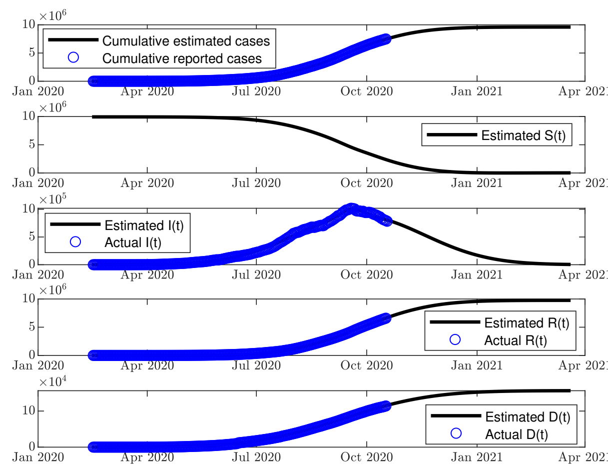

Fig. 1: Cumulative cases and SIRD waves (susceptible, infected, recovered and death cases) for COVID-19 data of India fitted from March 01, 2020 to October 01, 2020 by GMM and predicted thereafter.

the daily variations in testing and other effects due to policy changes. The optimum parameter values of GMM obtained for fitting the data are given in Table 2. These parameters are estimated within the 95% confidence interval. It is observed that the cumulative, recovered, and death cases are modelled as a single Gaussian wave, while the infected cases are modelled by two Gaussian waves.

Table 2: Parameters obtained by fitting GMM for cumulative, infected, recovered, and death cases of COVID-19 data of India from March 01, 2020 to October 01, 2020

SIRD waves	a_i	b_i	c_i
Cumulative $C(t)$	8.476e+04	212.2	63.88
Infected $I(t)$	8.813e+05	224.2	76.53
	1.297e+05	217.7	13.07
Recovered $R(t)$	8.328e+04	224.7	66.1
Dead $D(t)$	1067	209.6	82.74

The wave-forms for susceptible, infected, recovered, and death cases estimated using GMM are used to calculate the parameters of the SIRD model from (8). Unlike the traditional SIRD model, wherein the infection rate, recovery

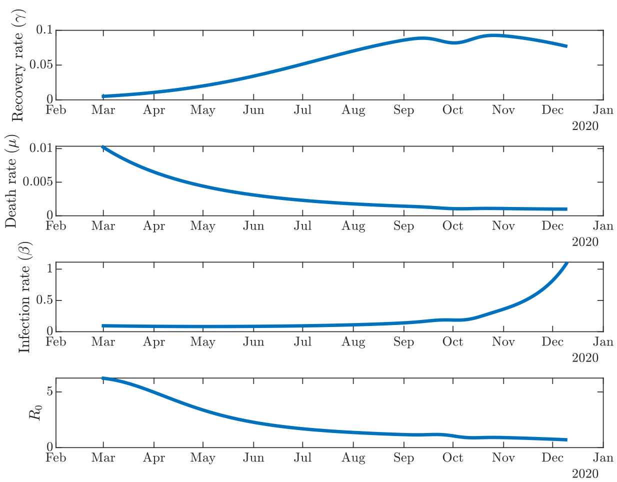

Fig. 2: Time-varying SIRD parameters estimated for COVID-19 data of India using the modeling of (8) based on the SIRD model with GMM: infection rate (β), recovery rate (γ), death rate (μ), and reproduction number $\mathfrak{R}_0$.

rate, death rate and reproduction number are generally taken to be constant parameters, the proposed modeling approach presents closed-form explicit expressions of these parameters. The estimated parameter waves for India are shown in Fig. 2. This approach provides realistic modeling of the parameters because with the actual situation of the pandemic and policy decisions, these parameter values will change with time. Further, these time-varying parameters can be tracked by the health sector and the administrative authorities for managing the resources and implementing policies. The recovery rate γ increased consistently to a value of 0.088 till September 12, 2020 and has shown slight variations since then. On the other hand, death rate μ has monotonically decreased to a current value of 0.001. On March 12, 2020, $\mathfrak{R}_0$ was 5.93, and has been gradually decreasing since then owing to the lock-downs imposed by the government. Its estimated value on October 17, 2020 is 0.89. In line with this observation of $\mathfrak{R}_0 < 1$, we also note from the $I(t)$ curve that the infected cases have started declining.

Fig. 3 (top subfigure) shows the cumulative cases $C(t)$, Fig. 3 (middle subfigure) shows fresh cases reported on a daily basis, and daily cases' growth rate is depicted in Fig. 3 (bottom subfigure). The SIR model has been adapted from the MATLAB files provided by Milan Batista [15]. The predicted values for the total cases and cases per day converge to the actual values provided till October 17, 2020. The growth factor of daily cases, defined by $\frac{\mathrm{d}C/\mathrm{d}t}{C} \times 100\%$, was 13% on March 3, 2020, started declining from the end of April, 2020, declined below 5% on May 22, 2020, and has now reduced to less than 1% in the month of October. This shows that the growth factor is on a constant decline. The reproduction number, β and γ (by definition, $\gamma_{\text{SIR}} = \gamma_{\text{SIRD}} + \mu_{\text{SIRD}}$) values estimated using SIR model are 1.93, 0.073, and 0.038, respectively. However, these constant values are not able to capture the variations as a function of time. Also, the traditional SIR and SIRD models generate RMSE of 28635 and 30475, respectively, for the cumulative cases, whereas the proposed method results in a lower RMSE of 4250 with GMM, and RMSE of 4030 with CLGM as shown in Table 1.

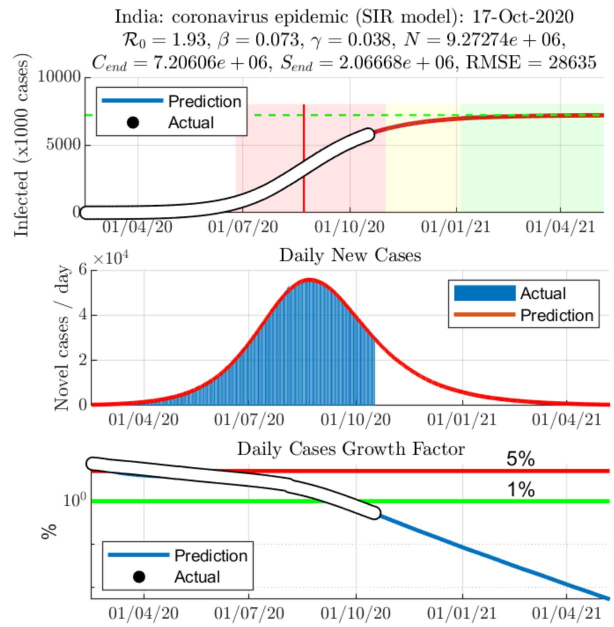

Fig. 3: The SIR model fitting for the COVID-19 cases of India (actual data: March 1, 2020 to October 17, 2020). Top subfigure: modeling of the total (cumulative) number of people infected $C(t)$; middle subfigure: modeling of new cases reported on daily basis; and bottom subfigure: predicted versus actual value of growth factor of daily cases computed as $\frac{\mathrm{d}C/\mathrm{d}t}{C} \times 100\%$.

4. CONCLUSION

This paper has presented a modified SIRD model that fits the pandemic data using well-defined functions of GMM and/or CLGM, unlike the traditional SIRD model that is based on pure data fitting. Subsequently, these functions are integrated into the SIRD model to derive explicit expressions of SIRD parameters that are found to be time-varying. In actual pandemic tracking, these time-varying parameters, say the reproduction number, may help to see the implications of pandemic management policy decisions or the pandemic spread in the real time. As per the data available up to October 17, 2020, the proposed model estimates the final epidemic size in India as 9596498. This optimistic value is supported by the fact that $\mathfrak{R}_0 < 1$ by the second week of October, 2020. However, this estimation will be valid only if the new infected patients follow the quarantine conditions, and the social distancing norms are maintained. The estimation might also change when the unrestricted international travel will resume. The performance of the proposed model is far superior than its counterparts existing in the literature.

5. REFERENCES

[1] Worldometers, https://www.worldometers.info/coronavirus/country/india/, Retrieved Oct. 17, 2020.

[2] World Health Organization, "Coronavirus disease (COVID-2019) situation reports," https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/, Retrieved Oct. 17, 2020.

[3] L. Wang, Y. Wang, D. Ye, Q. Liu, "Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence," Int J Antimicrob Agents, vol. 55, pp. 105948, 2020.

[4] W.O. Kermack, A.G. McKendrick, "Contributions to the mathematical theory of epidemics," Proc Roy Soc London Ser A, vol. 115, pp. 700–721, 1927.

[5] H.W. Hethcote, "The mathematics of infectious diseases," SIAM Review, vol. 42, no. 4, pp. 599–653, 2020.

[6] L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, D. Liu, "Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model," IEEE Acc, vol. 8, pp. 51761–51769, 2020.

[7] D. Fanelli, F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," Chaos Sol Frac, vol. 134, pp. 109761, 2020.

[8] M. Peirlinck, K. Linka, F.S. Costabal, E. Kuhl, "Outbreak dynamics of COVID-19 in China and the United States," Biomech Mod Mechanobiol, pp. 1–15, 2020.

[9] A. Tomar, N. Gupta, "Prediction for the spread of COVID-19 in India and effectiveness of preventive measures," Sci Tot Environ, vol. 728, pp. 138762, 2020.

[10] A. Singhal, P. Singh, B. Lall, S.D. Joshi, "Modeling and prediction of COVID-19 pandemic using Gaussian mixture model," Chaos Sol Frac, vol. 138, pp. 110023, 2020.

[11] N. Chintalapudi, G. Battineni, F. Amenta, "COVID-19 disease outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach," J Microbiol Immun Inf, vol. 53, no. 3, pp. 396–403, 2020.

[12] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang, T.-L. Gao, W. Duan, K. K. fai Tsoi, F.-Y. Wang, "Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on Weibo," IEEE Trans Comput Soc Syst, vol. 7, no. 2, pp. 556–562, 2020.

[13] M. Batista, "Estimation of a state of corona 19 epidemic in august 2020 by multistage logistic model: A case of EU, USA, and World," medRxiv, pp. 1–9, 2020; https://doi.org/10.1101/2020.08.31.20185165.

[14] J.C. Lagarias, J.A. Reeds, M.H. Wright, P.E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," SIAM J Optimization, vol. 9, no. 1, pp. 112–147, 1998.

[15] M. Batista, "fitVirusCOVID19, MATLAB Central File Exchange," Retrieved Oct. 20, 2020, https://www.mathworks.com/matlabcentral/fileexchange/74658-fitviruscovid19