Modeling the Epidemic Outbreak and Dynamics of COVID-19 in Croatia

Ante Lojić Kapetanović*, Dragan Poljak†

Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture
University of Split
Split, Croatia
*alojic00@fesb.hr, †dpoljak@fesb.hr

Abstract— **The paper deals with a modeling of the ongoing epidemic caused by Coronavirus disease 2019 (COVID-19) on the closed territory of Republic of Croatia. Using official public information on the number of confirmed infected, recovered and deceased individuals, the modified SEIR compartmental model is developed to describe the underlying dynamics of the epidemic. Fitted modified SEIR model provides the prediction of the disease progression in near future, considering strict control interventions by means of social distancing and quarantine for infected and at-risk individuals introduced at the beginning of COVID-19 spread on February, 25th by Croatian Ministry of Health. Assuming the accuracy of provided data and satisfactory representativeness of the model used, basic reproduction number is derived. Obtained results portray potential positive developments and justify the stringent precautionary measures introduced by the Ministry of Health.**

Index Terms—**SARS-CoV-2, COVID-19, epidemic modeling, compartmental epidemiology models**

I. INTRODUCTION

Coronavirus 2019-2020 epidemic of coronavirus disase 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), began in Wuhan, China, in late December 2019 [1]. Although Europe had time to prepare for the case of epidemic outbreak, the lack of relevant medical data and the novelty of COVID-19 disease have made it the new epicenter of the current pandemic with more reported cases and deaths than the rest of the world combined, excluding China, at the time of writing this paper.

The paper aims to model the disease dynamics by determining epidemic parameters in order to estimate the number of infected individuals over time on the closed territory of Croatia and to predict the pace of COVID-19 disease spreading by using the basic reproduction number R_0. Since control measures in Croatia, such as tracing close contacts, quarantining infected cases, promoting social distancing and self-protection using protective equipment, i.e. medical masks and gloves in public area, started with the first confirmed case on February, 25th, R_0 explicitly indicates the effectiveness of the current control measures and facilitate the adoption and implementation of potential additional measures. Thus, the ultimate goal of this paper is to accurately determine R_0 as the expected number of secondary infections caused by a single infectious symptomatic or asymptomatic individual in a fully susceptible population [2].

Even though Coronavirus-based diseases are well documented and modeled with a satisfactory level of accuracy in computational epidemiology literature [3], there are a few novel properties of COVID-19 disease that need to be considered. Firstly, long incubation period will cause time delay between real dynamics and the actual status. Furthermore, both symptomatic and asymptomatic individuals are capable of being infectious carriers of SARS-CoV-2. Finally, the disease transmission is achieved via respiratory droplets and is extremely difficult to prevent due the well resilient pathogen hardly affected by external atmospheric conditions [4].

In order to determine epidemic parameters, stochastic models are well adopted and preferred in the current research [5], but during the initial stage of the epidemic, the data are sparse and the epidemic dynamics are better described using deterministic modeling. Generalizations of Susceptible-Exposed-Infected-Recovered (SEIR) model are used for China [6], [7] and heavily affected parts of Europe [8], [9]. However, due to the change of diagnostics and the lack of implementation of control measures in the initial stage of the outbreak, early published models were forced to add many free parameters and fit the data in multiple phases. Such complicated models are prone to over-fitting.

This paper introduces the modified version of SEIR model with a single additional parameter to fit and implicitly included additional compartment for quarantined and self-isolated individuals; the paper is organized as follows: Section II outlines the mathematical modeling of COVID-19 and methods for predicting the spread of the mentioned disease; Section III is dedicated to the computational results and discussion; lastly, conclusion is presented in Section IV. Scraping of official data, the modified SEIR model fitting details and optimization method are given in Appendix A, while additional details on epidemic parameters are given in Appendix B.

II. TOOLS FOR MATHEMATICAL MODELING OF COVID-19

A. The Modified SEIR Compartmental Model

SEIR model, based on the early work of Kermack and McKendrick [10]–[12], is compartmental, population-based epidemiological model used for an appropriate mathematical representation of underlying dynamics of an infectious disease. Such model, readily generalized and computationally tractable
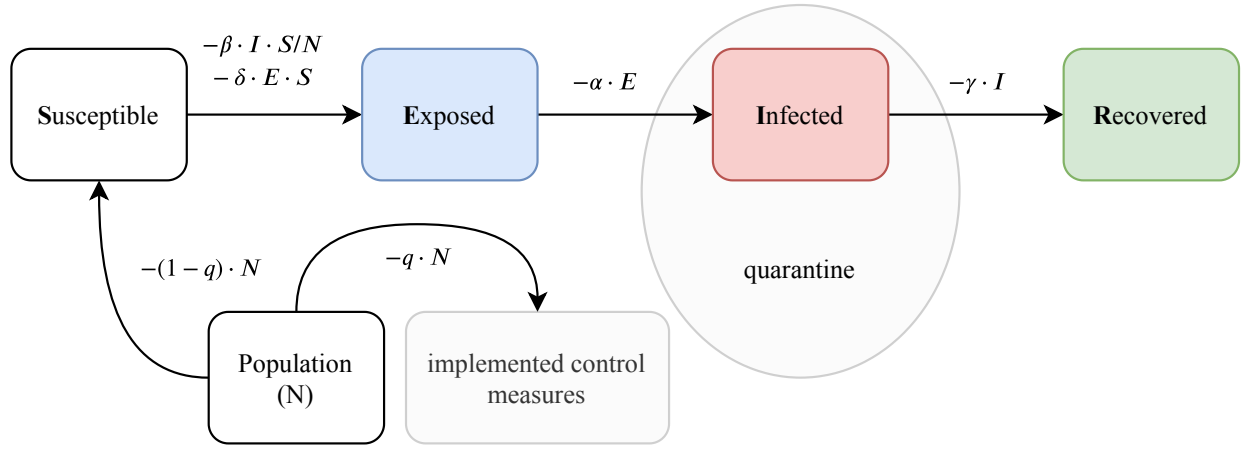
Fig. 1. The modified SEIR compartmental model of COVID-19 disease.

is a perfect candidate to simulate a propagation of a disease without much of a prior knowledge on a disease. The entire population, which is assumed to be constant in time (no vital dynamics by means of birth and death rate are taken into an account) is divided into 4 compartments. The rate of change over time for each compartment is expressed in terms of set of 4 coupled ordinary differential equations (ODEs). Note that the nature of differential equations dealing with time dependant quantities is to assume continuous variables, which is not the case for both the number of people and observed time steps. In order to obtain the discrete time equivalent, the changes of individuals in each compartment are modeled via the set of difference equations over an arbitrary discrete-time step t (typically a day). It is worth emphasizing that the discretization of variables is usually carried out as the prequel to stochastic modeling [13]. Nevertheless, continuous compartmental models provide more detailed epidemic dynamics [6], thus the modified SEIR model is defined in terms of the following set of ODEs:

$$\frac{dS}{dt} = -\beta \cdot \frac{I}{N} \cdot S - \delta \cdot E \cdot S \tag{1}$$

$$\frac{dE}{dt} = \beta \cdot \frac{I}{N} \cdot S - \alpha \cdot E + \delta \cdot E \cdot S \tag{2}$$

$$\frac{dI}{dt} = \alpha \cdot E - \gamma \cdot I \tag{3}$$

$$\frac{dR}{dt} = \gamma \cdot I \tag{4}$$

where,

- S is the susceptibles compartment;
- E is the exposed compartment;
- I is the infected compartment and
- R is the recovered (or deceased) compartment.

Simplified visual representation of the modified SEIR model of COVID-19 disease based on the set of equations (1)-(4) is shown in Fig. 1. It is assumed that only a small portion of the total population N is susceptible to the infection due to

timely and effective measures implemented from the first day of the epidemic.

Free parameters are in order: the transition rate β, which stands for the average number of contacts per unit time multiplied by the probability of transmission per each contact, the direct transition rate δ between exposed and susceptible compartment; the reciprocal value of the incubation period α and the rate of recovery or mortality γ. The mentioned expansion with respect to standard SEIR model lies in the expression $\delta \cdot E \cdot S$, which allows exposed individuals, considered to be in the incubation phase, to transmit the infection onto susceptible individuals [14].

B. R_0 - The Key Epidemic Parameter

R_0 (read 'R naught') is the expected number of secondary infections in a sufficiently large population without prior immunity to a disease. The assumption is well aligned with the current COVID-19 disease outbreak, since there is no maternal immunity nor there is a functional vaccine yet. R_0 number depends on both biological characteristics of the pathogen and the host, and on social circumstances, as follows:

$$R_0 = \frac{\beta}{\gamma + \delta} \tag{5}$$

The importance of R_0 lies in the fact that it represents somewhat of a threshold value from which one could determine the further progress of the disease. Namely, if $R_0 > 1$, the pathogen is expected to invade a population - the epidemic expansion is likely to occur. On the other hand, if $R_0 < 1$, the ongoing disease will vanish over time. The goal of preventive measures is to keep R_0 below the mentioned threshold. In other words, R_0 directly depends on the total infectious period,

$$\tau = \frac{1}{\gamma + \delta} \tag{6}$$

which is considered to be constant biological property and on the probability of disease transmission per contact per unit time β. The transition rate β can be implicitly lowered by
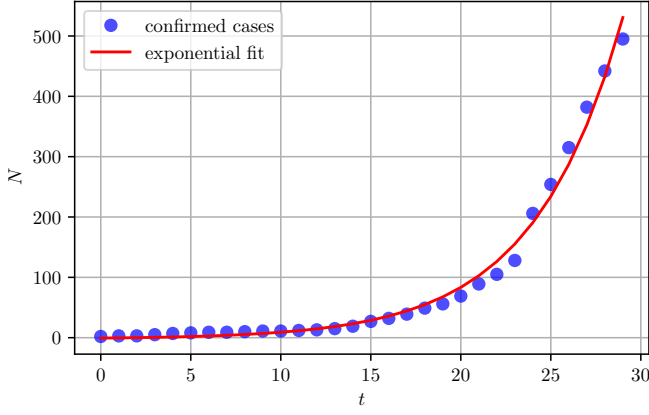
Fig. 2. Exponential function fitted to the number of confirmed infected cases, N, during the first $t = 30$ days of COVID-19 disease outbreak in Croatia. Even though, control measures were implemented from the first day of the disease, because of the incubation period of 3-15 days, the effect of those measures was shown with a time lag.

lowering the susceptible population through social distancing and by lowering the transition rate e.g. enhanced hygiene, wearing protective clothing.

In this paper, R_0 is determined via expression given in (5), with parameters obtained through fitting the actual data, time series of confirmed infections and recovered individuals from February, 25th to April, 11th, to the modified SEIR model. The data are obtained from the official publicly available sources [15].

III. RESULTS AND DISCUSSION

An early stage of the epidemic spread of an infectious disease, shown in Fig. 2, is often characterized by exponential growth, with the rapidity of growth dependent on the force of infection, formally defined as the per capita rate at which susceptible individuals contract infection and mathematically formulated with the following expression:

$$F = \beta \cdot I \tag{7}$$

After an early stage, exponential growth is physically unfeasible because of the finite size of the population. Exponential function overestimates the total number of infected individuals, shown in Fig. 3a, where actual growth soon starts to asymptotically approach the value of the maximum infections and is more accurately represented via the logistic function. However, since the current status of the growth is unknown, logistic function fitted to the portion of the data provides grossly underestimated predictions, depicted in Fig. 3b.

Instead of observing confirmed cases over time, a more appropriate capture of the current trend of the disease spread is achieved by observing the change itself. Fig. 4 charts the number of new cases averaged over previous 7 days. Obviously, COVID-19 disease spread in Croatia is in a downward trend. If the influence of time is instead taken implicitly and the growth rate is plotted against the current number of infections in logarithmic scale, the resulting comparison between the actual
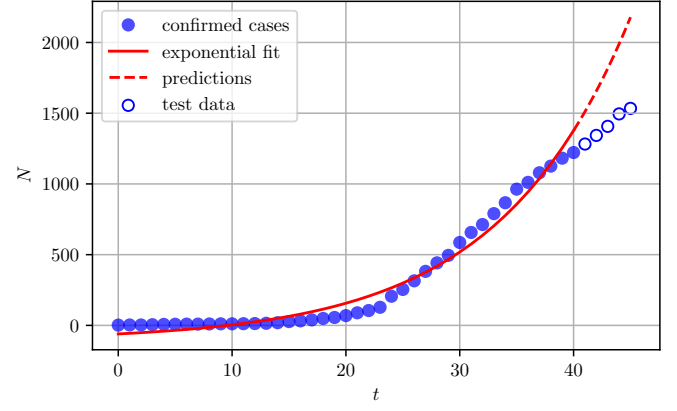
(a) Exponential function fitted to the confirmed number of infected individuals N over time t. The dashed line represents prediction for the upcoming 5 days.
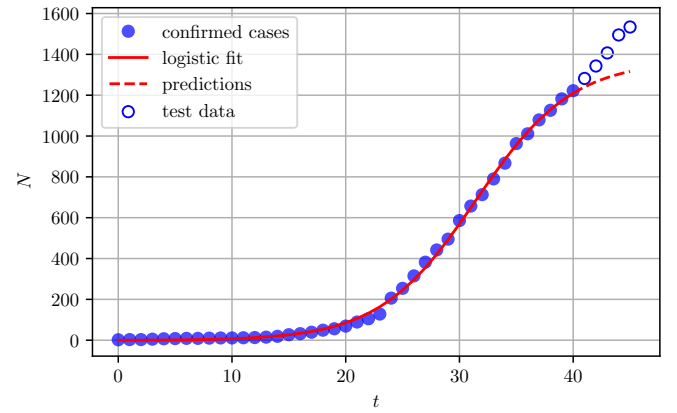
(b) Logistic function fitted to the confirmed number of infected individuals N over time t. The dashed line represents prediction for the upcoming 5 days.

Fig. 3. Exponential and logistic functions fitted to the data divided into training and testing set.

growth of COVID-19 disease and the exponential growth will be easier to compare, see Fig. 5. Once the trend begins to steeply decline, the epidemic is considered to be terminated, i.e. the reproductive number drops below 1.

The modified SEIR model, introduced in the section II-A, is fitted to the number of confirmed, recovered and death cases over time. Using the estimated parameters, described in detail in Appendix B, of the fitted model, reproduction number is calculated for different phases of the epidemic. The data are fitted using the iterative Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm. Within this framework, for each iteration of optimization, the set of ODEs (1)-(4) is solved by the 4th order Runge-Kutta method. The optimization procedure and the solution of the set of ODEs are presented in more detail in section A. Fig. 6 depicts fitted modified SEIR model using different sizes of the train data set. Fig. 6a shows the model fitted using 88% of the data set, while the rest of the data are used in order to assess goodness of fit using the out-of-sample data, Fig. 6b shows the SEIR model fitted using 96% of the data set, while Fig.
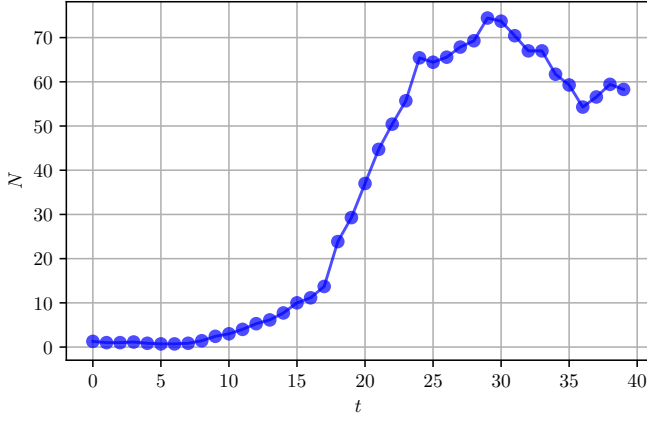
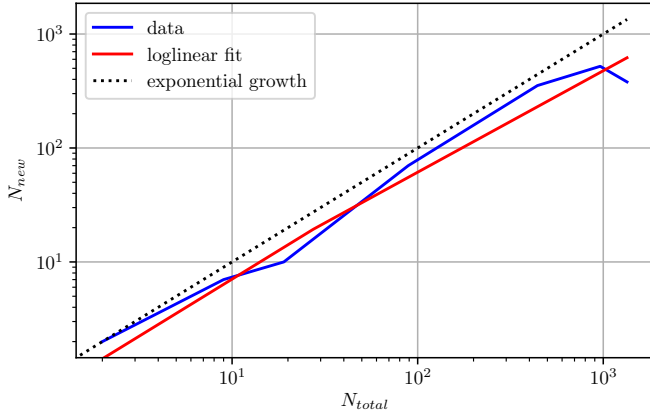Fig. 4. Time series of new confirmed cases N averaged over 7-day period.

Fig. 5. Logarithmic number of new cases, N_{new}, vs. logarithmic number of confirmed cases, N_{total}, averaged over 7-day period (blue line) compared to the theoretical exponential function (black dashed line) in log-scale.

6c deals with the complete data set. Fitting the data provides optimal values for epidemiological parameters α, β, γ and δ. R_0 is then calculated for different phases of the epidemic and the resulting values are 1.78, 1.63 and 1.36 for 88% of the data, 96% of the data and the complete data set, respectively. These results imply the effectiveness of the control measures implemented to combat the epidemic as R_0 decreases with each increase of the data set. In case there is no change in control measures, one could infer that the positive downward trend will continue up until the end of May when the number of confirmed infected cases will reach its maximum. The maximum point is also the inflection point indicating the moment at which $R_0 < 1$ and after which, with the retention of the control measures, there are no new infections.
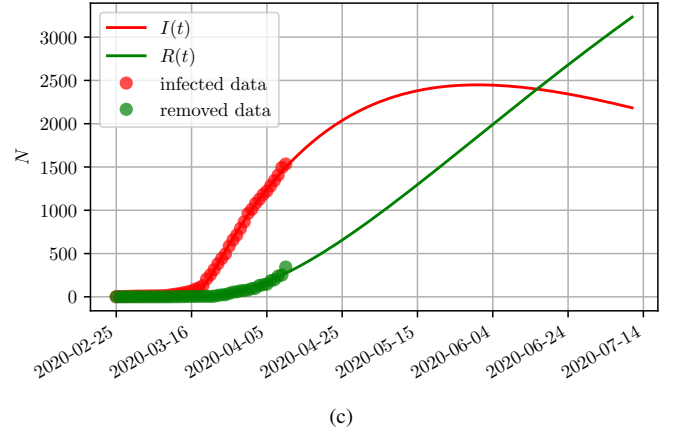
IV. CONCLUSION

In this paper, the slightly modified SEIR compartmental epidemiological model, regarded as an opener to the subject, is proposed in order to analyze and predict COVID-19 epidemic in Croatia. Transmission parameters of the model are optimized using the available data on the number of confirmed,
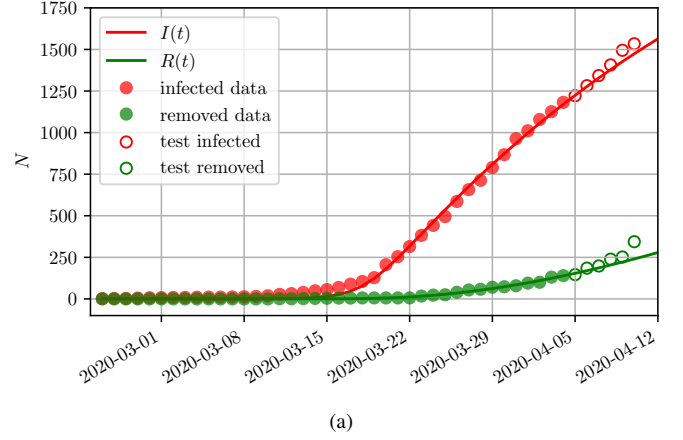
(a)
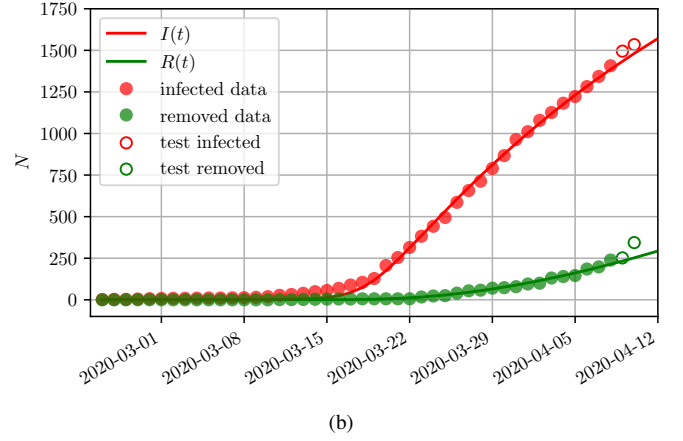
(b)

(c)

Fig. 6. The modified SEIR model fitted to different sizes of the training data set. The top two figures compare the actual data with the predicted solution of the infected and the recovered (or deceases) differential equations defined in (3) and (4) respectively. The bottom figure shows the prediction of the near future.

recovered and deceased individuals, respectively. The model assumes a sufficient percentage of a population $q \cdot N$ being voluntarily self-isolated, shown in Fig. 1. The percentages of self-isolated people, together with initial number of exposed, infected and recovered individuals are taken as initial values in order to solve the set of ODEs.

The parameters are used to calculate R_0 for different stages of the spread of COVID-19 in Croatia. The values of R_0 improve as the data approach closer to April 11th. At this point one considers the government control measures, implemented to slow down the disease spread, to be effective and should be continued until R_0 number decreases and remains sufficiently long period of time below the threshold value of 1.

Based on fitted SEIR model, the climax of the infection is expected to occur around the end of May. The inflection point should follow soon after that date, which ultimately means that the epidemic is in a decline period.

ACKNOWLEDGMENT

APPENDIX A
OUTLINE OF MATHEMATICAL MODELING

The number of confirmed, recovered and deceased people is collected from available data for the period of February, 25th to April, 11th. The modified SEIR model consists of 4 free parameters β, γ, α and δ to be optimally determined by minimizing the loss function $J(\beta, \gamma, \alpha, \delta)$. The loss function is formulated as the weighted summation of l_2 norm between the actual number of confirmed, recovered and deceased people and their respective approximations. The loss function is written using the following expression:

$$J(\beta, \gamma, \alpha, \delta) = ||I(t) - \hat{I}(t)||_2 + ||R(t) - \hat{R}(t)||_2 \quad \text{(A.1)}$$

The approximations are obtained by solving the set of deterministic ODEs and are marked with $\hat{I}(t)$ for approximated number of infections over time and $\hat{R}(t)$ for approximated total sum of recovered and deceased individuals over time.

The minimization is performed using L-BFGS method where for each set of parameter values, the 4th order Runge-Kutta method is employed to solve the set of ODEs with known initial values. For each of the differential equations given in (1)-(4), initial values are determined using available data on the number of susceptible individuals S_0, the number of exposed individuals E_0, the number of infected individuals I_0 and the number of recovered and deceased individuals R_0 at the first day of the epidemic. Both, the minimization and the initial value problem for the set of ODEs are solved using SciPy, open-source scientific software built on the NumPy extension of Python programming language.

APPENDIX B
LIST OF PARAMETERS

- β - transition or infectious rate; controls the rate of spread which represents the probability of transmitting disease between a susceptible and an infected individual per contact per unit time;
- γ - recovery or mortality rate; determined by the average duration of infection which eventually lead to recovery

or death, where individual is no longer classified as susceptible;
- α - incubation rate; reciprocal value of the incubation period. In the case of COVID-19 disease, an exposed individual is infectious in incubation asymptomatic phase;
- δ - direct transition rate between susceptible and exposed individual. As mentioned, an exposed individual is capable of shedding the infection even in the asymptomatic phase;
- q - quarantine or self-isolation rate; the percentage of the total population insusceptible to COVID-19 disease.

REFERENCES

[1] World Health Organization. (2020) Coronavirus disease (COVID-19) outbreak. [Online]. Available: https://www.who.int/health-topics/coronavirus

[2] M. Kretzschmar, "Measurement and modeling: Infectious disease modeling," in *Reference Module in Biomedical Sciences*. Elsevier, 2016.

[3] X.-N. Han, S. J. de Vlas, L.-Q. Fang, D. Feng, W.-C. Cao, and J. D. F. Habbema, "Mathematical modelling of SARS and other infectious diseases in China: a review," *Tropical Medicine & International Health*, vol. 14, pp. 92–100, Nov. 2009.

[4] W. jie Guan, Z. yi Ni, Y. Hu, W. hua Liang, C. quan Ou, J. xing He, L. Liu, H. Shan, C. liang Lei, D. S. Hui, B. Du, L. juan Li, G. Zeng, K.-Y. Yuen, R. chong Chen, C. li Tang, T. Wang, P. yan Chen, J. Xiang, S. yue Li, J. lin Wang, Z. jing Liang, Y. xiang Peng, L. Wei, Y. Liu, Y. hua Hu, P. Peng, J. ming Wang, J. yang Liu, Z. Chen, G. Li, Z. jian Zheng, S. qin Qiu, J. Luo, C. jiang Ye, S. yong Zhu, and N. shan Zhong, "Clinical characteristics of 2019 novel coronavirus infection in China," Feb. 2020.

[5] M. Marathe and A. K. Vullikanti, "Computational epidemiology," *Communications of the ACM*, vol. 56, pp. 88–96, 07 2013.

[6] P. Liangrong, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of covid-19 in china by dynamical modeling," 02 2020.

[7] S. Zhao and H. Chen, "Modeling the epidemic dynamics and control of COVID-19 outbreak in china," *Quantitative Biology*, vol. 8, no. 1, pp. 11–19, Mar. 2020.

[8] D. Cereda, M. Tirani, F. Rovida, V. Demicheli, M. Ajelli, P. Poletti, F. Trentini, G. Guzzetta, V. Marziano, A. Barone, M. Magoni, S. Deandrea, G. Diurno, M. Lombardo, M. Faccini, A. Pan, R. Bruno, E. Pariani, G. Grasselli, A. Piatti, M. Gramegna, F. Baldanti, A. Melegaro, and S. Merler, "The early phase of the covid-19 outbreak in lombardy, italy," 2020.

[9] L. R. Lopez and X. Rodo, "A modified SEIR model to predict the COVID-19 outbreak in spain: simulating control scenarios and multiscale epidemics," Mar. 2020.

[10] W. O. Kermack and A. G. McKendrick, "Contributions to the mathematical theory of epidemics—i," *Bulletin of Mathematical Biology*, vol. 53, no. 1-2, pp. 33–55, Mar. 1991.

[11] ——, "Contributions to the mathematical theory of epidemics—II. the problem of endemicity," *Bulletin of Mathematical Biology*, vol. 53, no. 1-2, pp. 57–87, Mar. 1991.

[12] ——, "Contributions to the mathematical theory of epidemics—III. further studies of the problem of endemicity," *Bulletin of Mathematical Biology*, vol. 53, no. 1-2, pp. 89–118, Mar. 1991.

[13] B. Cantó, C. Coll, and E. Sánchez, "Analysis of a class of discrete sir epidemic model," 08 2015.

[14] W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, and V. J. Lee, "Presymptomatic Transmission of SARS-CoV-2 — Singapore," *Morbidity and Mortality Weekly Report*, vol. 69, no. 14, pp. 411–415, Apr. 2020.

[15] Croatian institue of public health. (2020) Official government website for accurate and verified infromation on Coronavirus. [Online]. Available: https://www.koronavirus.hr/en