Extracting transmission and recovery parameters for an adaptive global system dynamics model of the COVID-19 pandemic

Craig Stuart Carlson, David Milton Rubin, Vilma Heikkilä, Michiel Postema

Extracting transmission and recovery parameters for an adaptive global system dynamics model of the COVID-19 pandemic

Craig S. Carlson[1,*]
David M. Rubin[1]

[1]School of Electrical and Information Engineering
University of the Witwatersrand, Johannesburg
1 Jan Smuts Laan
Braamfontein 2050
South Africa

Vilma Heikkilä[2]
Michiel Postema[1,2]

[2]BioMediTech
Faculty of Medicine and Health Technology
Tampere University
Korkeakoulunkatu 3
33720 Tampere
Finland

***Abstract*—Accurately modelling the susceptibility, infection, and recovery of populations with regards to the COVID-19 pandemic is highly relevant for the implementation of countermeasures by governing bodies. In the past year, several thousands of articles on COVID-19 modelling were published. The spread of the pandemic has frequently been modelled using the Susceptible-Infected-Recovered (SIR) epidemic model owing to the low level of complexity. In recognition of its simplicity, we developed an SIR model to represent the spread of disease on a global scale, irrespective of mutation and countermeasures. The SIR parameters were reverse-engineered from aggregated global data. This model is the first to retrospectively deduce the initial incidence. The average transmission and recovery parameters were computed to be $0.33\,\text{week}^{-1}$ and $0.23\,\text{week}^{-1}$, respectively. These values lie well within the range of reported values on COVID-19 determined from geographically different regions. The model was simulated in the Ventana® simulation environment Vensim® for a 65-weeks duration and an adjusted initial infection incidence, which was presumed three times the reported initial infection incidence. The simulated data visually aligns with the real incidence data. We attribute the discrepancy between the presumed initial value and the reported value to lack of testing facilities on the starting date of 1 March 2020. Our parameter extraction suggests a novel methodology to quantify undertesting retrospectively in epidemics.**

***Index Terms*—system dynamics; coronavirus model; SIR model; parameter extraction.**

I. Introduction

The coronavirus disease (COVID-19) pandemic threw the world into a state of turmoil from late 2019, with 160 074 267 confirmed cases reported on 13 May 2021 [1]. Due to the highly contagious nature and rapid spread of COVID-19 [2], a need for tools to forecast the spread of the virus has emerged. Being able to evaluate the number of infections could aid governments to determine necessary precautions to slow down the spread of the disease and assess the required capacity of healthcare facilities to treat the disease.

Epidemic characteristics have been described by logistic growth models [3], [4], natural growth models [5], auto regressive integrated moving average models [6], Susceptible-Infected-Recovered (SIR) compartment models [4], [6], [7], and derivations thereof [8]–[11]. The simplest modelling method is the SIR model as it described by two parameters. Such a simplistic model is not to be used to give explanations about the virus itself. Key modelling characteristics include the rate of transmission, the rate of recovery and the basic reproduction number. For COVID-19, the basic reproduction number lies on the range 1.00–3.15 [12], [13].

Whilst any number of parameters can be added, the challenge in designing a mathematical model is estimating the actual parameter values to achieve accurate results. This requires the fitting of the model to actual data to extract parameter values. As model parameters are fitted for a geographic region of interest and compared to a geographic data subset there is great variation of parameters in a model. These region-specific parameters may not describe the spread of a disease on a global scale.

In this study an SIR model was implemented to represent the spread of COVID-19 on a global scale, with a simple set of parameters considered. The set of transmission and recovery parameters was reverse-engineered from aggregated global data. Special attention was paid to the deduction of the initial incidence. This is highly relevant, as undertesting in epidemics leads to false parameter estimations and consequently to false predictions. Any effects of mutations and countermeasures were not included in the model.

This work has been based on research supported in part by the National Research Foundation of South Africa, Grant Number 127102, and by the Academy of Finland, Grant Number 340026.

*Correspondence to: Craig.Carlson@wits.ac.za

II. Theory

The SIR epidemic model is based on earlier epidemics research that propose the division of the population into compartments [14]–[19]. This model simplifies the dynamics of the disease to three compartments — those at risk for infection; those infectious and infected with the disease; and those no longer infected due to recovery. These compartments are named Susceptible; Infected; and Recovered, respectively.

Each compartment, called a stock, represents an accumulation of people at time t. There are two flow rates associated with the stocks in the basic SIR model, namely the infection rate and the recovery rate. The infection rate is the rate of susceptible people becoming infected, and the recovery rate is the rate of infected people recovering from the disease.

The model is governed by a set of three ordinary differential equations which describe how the three stocks change over time [20]. The rate of change of the susceptible stock is

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta \frac{I}{N} S, \tag{1}$$

where I is the infected population size, N is the total population size, S is the susceptible population size and β is the transmission parameter of disease.

Moreover, the rate of change of the infected stock is

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta \frac{I}{N} S - \gamma I, \tag{2}$$

where γ is the recovery parameter.

The rate of change of the recovered stock is

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I, \tag{3}$$

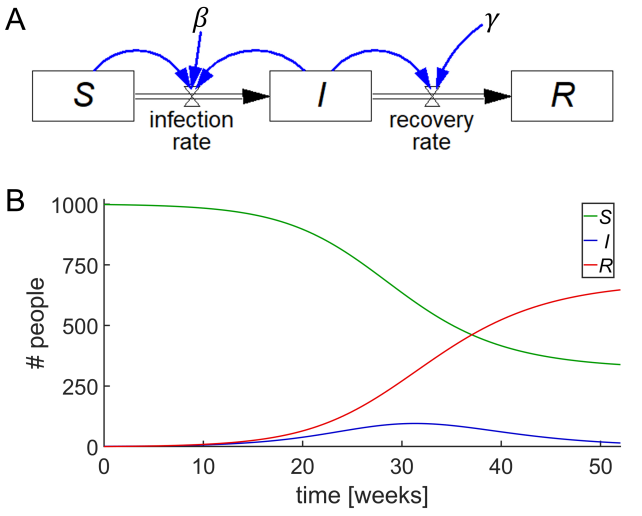

Fig. 1. A graphical representation of the basic SIR model (A) and associated graphs of the solutions of the ordinary differential equations (B), demonstrating the behaviour of a viral epidemic with starting conditions $S(0) = 999$ people, $I(0) = 1$ person, $R(0) = 0$ people, and model parameters $\beta = 0.5$ week^{-1}, $\gamma = 0.3$ week^{-1}.

where R is the recovered population.

Figure 1 shows the graphical representation of the basic SIR model described by equations (1)–(3). The graph shows the number of people in each compartment on a given day.

This basic model can be extended by including other compartments such as death, quarantine, and hospitalisation. Adding parameters to a model inherently increases the complexity of the solution. The influence of lockdowns, closing of borders, and rolling out a vaccination programme on the susceptibility might be included in a model [21], whilst a mutation of the virus might change the transmission parameter [22].

For convenience, the potential severity of an infectious disease is expressed by the basic reproduction number, R_0 [23]. If the susceptible population is equal to the total population, the basic reproduction number is approximated by

$$R_0 \approx \frac{\beta}{\gamma}. \tag{4}$$

If $R_0 > 1$, the epidemic is sustained.

III. Methods

The basic SIR model used in this study is shown in Figure 1A. The model was implemented in the Ventana® simulation environment Vensim® (Ventana Systems, Inc., Harvard, MA, USA). Solutions were computed for a 65-week duration. The starting conditions of the simulations were those of the well-documented situation on 1 March 2020 [24].

The transmission and recovery parameters were reverse-engineered from aggregated global data of infection and recovery numbers [24], assuming a constant world population size $N = 7.8 \times 10^9$ people, and taking $S = N - I - R$. These were substituted into equations (1)–(3). The input set of SIR parameters, normalised by N, is shown in Figure 2.

The reverse-engineered vectors $\beta(t)$ and $\gamma(t)$ were subjected to five-degree polynomial fitting using the fit.m routine of MATLAB® (The MathWorks, Inc., Natick, MA, USA), yielding $\beta(t) = \sum_{i=0}^{5} \beta_i t^i$ and $\gamma(t) = \sum_{i=0}^{5} \gamma_i t^i$. Both polynomial representations were subsequently used in an adaptive global system dynamics model for long-term prediction of the COVID-19 pandemic spread.

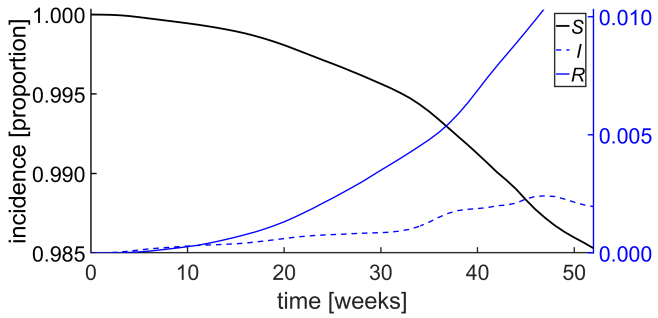

Fig. 2. Input SIR values used to compute β and γ.

TABLE I
POLYNOMIAL COEFFICIENTS FOR β AND γ CURVE FITTING.

Polynomial degree i	β_i	γ_i
0	0.7993	0.2507
1	-0.0899	-0.03592
2	0.005827	0.00471
3	-0.0001697	-0.0002116
4	2.306×10^{-6}	3.983×10^{-6}
5	-1.246×10^{-8}	-2.705×10^{-8}

For comparison, parameters were extracted from literature. To find relevant articles for the first dataset, we used the PubMed® database (National Library of Medicine, Bethesda, MD, USA) using the query (COVID-19) ∩ (SIR model). Only articles published between 1 January 2020 and 28 February 2021 were included. Of the initial selection of 60 SIR-like models, only 21 stated values for the transmission and recovery parameters. A full analysis of the query outcome has been presented by Heikkilä [25]. The values for the transmission parameter in these 21 articles were averaged, and so were the values for the recovery parameter.

IV. Results and Discussion

Figure 3 demonstrates graphical representations of the transmission parameter $\beta(t)$, the recovery parameter $\gamma(t)$, and the basic reproduction number $R_0(t)$. While their average values are $\bar{\beta} = 0.33\,\text{week}^{-1}$ and $\bar{\gamma} = 0.23\,\text{week}^{-1}$, it is evident that there is a high degree of variability over the course of the epidemic. This suggests the value of incorporating time-varying transmission and recovery parameters into the SIR model.

The values extracted from articles containing regional data extend over the ranges $0.13 \leq \beta \leq 0.70\,\text{week}^{-1}$ and $0.04 \leq \gamma \leq 0.29\,\text{week}^{-1}$. Thus, the average global values lie within these ranges.

The polynomial coefficients of the transmission and recovery parameters are shown in Table I. Clearly, the coefficients become negligible beyond third degree.

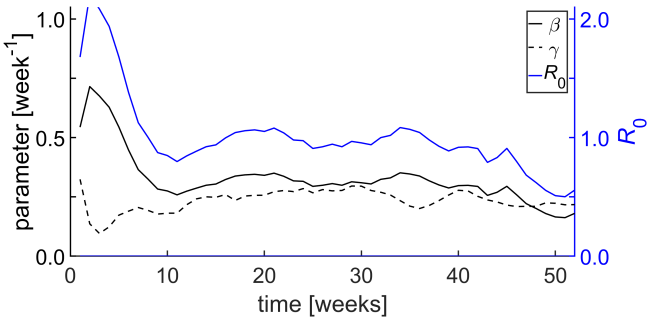

Fig. 3. Transmission parameter β (—), recovery parameter γ (- -), and basic reproduction number R_0 (—), all as a function of time.

The forward modelling using the transmission and recovery coefficients from Table I is shown in Figure 4, for an interval of 65 weeks, and an adjusted initial infection incidence $I(0)$, which was presumed three times the reported initial infection incidence. The simulated data visually aligns with the real incidence data. We noted that the simulated output became computationally unstable beyond 15 weeks after the reported data. Although a lower initial incidence yields a graph of similar shape (not shown), the peak value is only a third. We attribute this discrepancy with the reported value to lack of testing facilities on the starting date of 1 March 2020. Our parameter extraction suggests a novel methodology to quantify undertesting retrospectively in epidemics.

V. Conclusion

The average transmission and recovery parameters were computed to be $0.33\,\text{week}^{-1}$ and $0.23\,\text{week}^{-1}$, respectively. These values lie well within the range of reported values on COVID-19 determined from geographically different regions.

We demonstrate that the extraction of transmission and recovery parameters from real SIR data is feasible for an adaptive global system dynamics model of the COVID-19 pandemic. This study suggests a novel methodology for forward prediction and for the quantification of undertesting early in the epidemic.

This model presented in this paper is the first to retrospectively deduce the initial incidence.

References

[1] World Health Organization, "Coronavirus disease (COVID-19) outbreak," *World Health Organization*, 2021. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[2] S. Sanche, Y. Lin, C. Xu, E. Romero-Severson, N. Hengartner and R. Ke, "High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2," *Emerg. Infect. Dis.*, vol. 26, no. 7, pp. 1470–1477, 2020. [DOI: 10.3201/eid2607.200282]

[3] Y. Zou, et al., "Outbreak analysis with a logistic growth model shows COVID-19 suppression dynamics in China," *PLoS One*, vol. 15, no. 6, e0235247, 2020. [DOI: 10.1371/journal.pone.0235247]

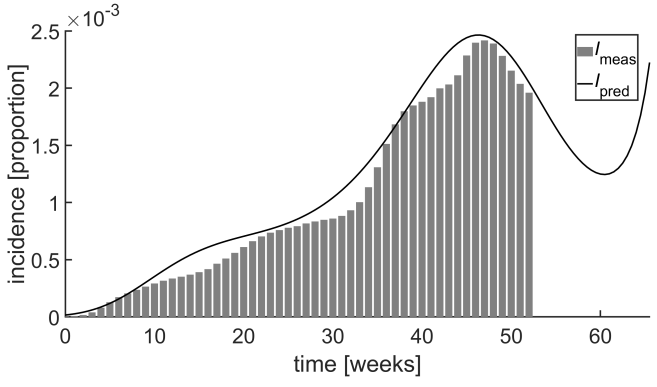

Fig. 4. Modelled incidence I_{pred} using reverse-engineered transmission and recovery coefficients (—) and real incidence data I_{meas} (bars), normalised by the total population (assumed constant).

[4] B. Malavika, S. Marimuthu, M. Joy, A. Nadaraj, E.S. Asirvatham and L. Jeyaseelan, "Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models," *Clin. Epidemiol. Glob. Health*, vol. 9, no. 1, pp. 26–33, 2021. [DOI: 10.1016/j.cegh.2020.06.006]

[5] N.E. Huang and F. Qiao, "A data driven time-dependent transmission rate for tracking an epidemic: a case study of 2019-nCoV," *Sci. Bull.*, vol. 65, no. 6, pp. 425–427, 2020. [DOI: 10.1016/j.scib.2020.02.005]

[6] K.A. Abuhasel, M. Khadr and M.M. Alquirash, "Analyzing and forecasting COVID-19 pandemic in the Kingdom of Saudi Arabia using ARIMA and SIR models," *Comput. Intell.*, pp.1–14, 2020. [DOI: 10.1111/coin.12407]

[7] U. Nguemdjo, F. Meno, A. Dongfack and B. Ventelou, "Simulating the progression of the COVID-19 disease in Cameroon using SIR models," *PLoS One*, vol. 15, no. 8, e0237832, 2020. [DOI: 10.1371/journal.pone.0237832]

[8] A. Alsayed, H. Sadir, R. Kamil and H. Sari, "Prediction of epidemic peak and infected cases for COVID-19 disease in Malaysia," *Int. J. Environ. Res. Public Health*, vol. 17, no. 11, 4076, 2020. [DOI: 10.3390/ijerph17114076]

[9] A.H.A. Mehra, M. Shafieirad, Z. Abbasi and I. Zamani, "Parameter estimation and prediction of COVID-19 epidemic turning point and ending time of a case study on SIR/SQAIR epidemic models," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–13, 2020. [DOI: 10.1155/2020/1465923]

[10] D.M. Rubin, et al. "Facilitating understanding, modeling and simulation of infectious disease epidemics in the age of COVID-19," *Front. Public Health*, vol. 9, 593417, 2021. [DOI: 10.3389/fpubh.2021.593417]

[11] S. Zhao and H. Chen, "Modeling the epidemic dynamics and control of COVID-19 outbreak in China," *Quant. Biol.*, vol. 8, pp. 11–19, 2020. [DOI: 10.1007/s40484-020-0199-0]

[12] M. Al-Raeei, "The basic reproduction number of the new coronavirus pandemic with mortality for India, the Syrian Arab Republic, the United States, Yemen, China, France, Nigeria and Russia with different rate of cases," *Clin. Epidemiol. Glob. Health*, vol. 9, pp. 147–149, 2021. [DOI: 10.1016/j.cegh.2020.08.005]

[13] W. He, G.Y. Yi and Y. Zhu, "Estimation of the basic reproduction number, average incubation time, asymptomatic infection rate, and case fatality rate for COVID19: meta-analysis and sensitivity analysis," *J. Med. Virol.*, vol. 92, pp. 2543–2550, 2020. [DOI: 10.1002/jmv.26041]

[14] R. Ross, "An application of the theory of probabilities to the study of a priori pathometry. – Part I," *Proc. R. Soc. A.*, vol. 92, no. 638, pp. 204–230, 1916. [DOI: 10.1098/rspa.1916.0007]

[15] H.P. Hudson and R. Ross, "An application of the theory of probabilities to the study of a priori pathometry. — Part II," *Proc. R. Soc. A.*, vol. 93, no. 650, pp. 212–225, 1917. [DOI: 10.1098/rspa.1917.0014]

[16] W.O. Kermack and A.G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc. R. Soc. A.*, vol. 115, no. 772, pp. 700–721, 1927. [DOI: 10.1098/rspa.1927.0118]

[17] W.O. Kermack and A.G. McKendrick, "Contributions to the mathematical theory of epidemics. II. — The problem of endemicity," *Proc. R. Soc. A.*, vol. 138, no. 834, pp. 55–83, 1932. [DOI: 10.1098/rspa.1932.0171]

[18] W.O. Kermack and A.G. McKendrick, "Contributions to the mathematical theory of epidemics. III. — Further studies of the problem of endemicity," *Proc. R. Soc. A.*, vol. 141, no. 843, pp. 86–94, 1933. [DOI: 10.1098/rspa.1933.0106]

[19] A.L. Bertozzi, E. Franco, G. Mohler, M.B. Short and D. Sledge, "The challenges of modeling and forecasting the spread of COVID-19," *Proc. Natl Acad. Sci.*, vol. 117, no. 29, pp. 16732–16737, 2020. [DOI: 10.1073/pnas.2006520117]

[20] Z. Ma and J. Li, *Dynamical Modeling And Analysis Of Epidemics.* World Scientific, Singapore, 2009. [DOI: 10.1142/6799]

[21] S. SeyedAlinaghi, et al., "Reinfection risk of novel coronavirus (COVID-19): a systematic review of current evidence," *World J. Virol.*, vol. 9, no. 5, pp. 79–90, 2020. [DOI: 10.5501/wjv.v9.i5.79]

[22] J.A. Plante, B.M. Mitchell, K.S. Plante, K. Debbink, S.C. Weaver and V.D. Menachery, "The variant gambit: COVID-19's next move," *Cell Host & Microbe*, vol. 29, no. 4, pp. 508–515, 2021. [DOI: 10.1016/j.chom.2021.02.020]

[23] R. Breban, R. Vardavas and S. Blower, "Theory versus data: how to calculate R_0?" *PLoS One*, vol. 2, no. 3, e282, 2007. [DOI: 10.1371/journal.pone.0000282]

[24] Worldometer, "COVID-19 coronavirus pandemic," *Worldometer*, 2021. [Available online: https://www.worldometers.info/coronavirus/]

[25] V. Heikkilä, *A quantitative analysis of susceptible-infected-removed models of the coronavirus.* B.Sc. Thesis. Tampere University, Finland, 2021.