doi:10.20944/preprints202007.0186.v1

Modeling Epidemics as First-order Systems – COVID-19 Example

Abdallah Abusam

Water Research Center, Kuwait Institute for Scientific Research

P. O. Box 24885, Sadat 13109, Kuwait

aabusam@kisr.edu.kw

Abstract

The semi-logarithmic plot of the cumulative number of cases of epidemics resembles the response of a first-order systems for a step load. This similarity was utilized to develop a first order model that can be used for extracting information about the dynamics of infectious disease epidemics. The developed model was validated using COVID-19 data of China. It was also heuristically fitted to other 13 countries. Obtained results indicated that the model can reliably forecasts the number of infected person, epidemic growth speed towards steady-state condition (process time constant, T), and time to reach steady-state condition (4T). The developed model will help public health authorities in developing more effective control strategies of epidemics.

Keywords: Infectious Diseases, Epidemics, Dynamical Modeling, COVID-19, First-order Systems.

1.0 Introduction

Mathematical models are useful tools that help provide insights into infectious disease dynamics (Caccavo 2020). They can be utilized to get vital information about dynamics of epidemics such as epidemics growth rate, time to reach the peak and the number of infected persons. Such information is of great importance for public health and other concerned authorities. It is useful to develop appropriate control measures and be prepared to deal with consequences of epidemics (Rabbani et al 2020). Due to increase in epidemics occurrence, use of mathematical models has increased tremendously in the last decades. In fact, mathematical modeling has become the main tool for developing effective control strategies for infectious disease epidemics (Anderson and May, 1991).

Large numbers of mathematical models have been developed to describe the growth patterns of infectious disease epidemics (Lin et al. 2020). Epidemiological models can be divide broadly into statistical and dynamical models. The commonly used dynamical models are either logistic growth models or compartmental models. The first logistic model was proposed by Verhulst in 1834 and modified by Richards in 1959 (Batista, 2020). On the other hand, the first compartmental SIR (susceptible, infected and recovered model) model was developed by the Scottish Kermack and McKendrick in 1927 (Sanglier et al. 2020). Since then, many new versions of both the logistic growth and compartmental models have been developed (Batista 2020; Liu 2020; Ma 2020; Roda

2020). Despite that, however, the original versions of these models are still in use. As a matter of fact, they have been widely used for modeling COVID-19.

During epidemics there is always a need for a fast and effective response (Sanglier et al. 2020). Effective response, however, requires reliable mathematical models that not only fit the profiles of epidemics, but also reliably explain their dynamics. Unfortunately, results of most of the existing models are unreliable because their parameters have no physical meanings and they cannot be accurately estimated from data. Roda et al. 2020 reported that forecast reliabilities of epidemiological models are greatly affected by the identifiability problem of their parameters. Abusam et al. (2020) have shown that parameters of Verhulst and Richards models cannot be identified from data and their predictions are unreliable. According to Louchet 2020, many countries could not reliably predict the progress of the on-going COVID-19 pandemic.

For developing effective control strategies of epidemics, there is always a need for a reliable model that can accurately capture the dynamics of the epidemics (Chowell et al. 2016). In this paper, principles of systems engineering were used to develop a new dynamical model that extracts accurate and physically meaningful information about the dynamics of infectious disease epidemics.

2.0 Materials and Methods

2.1 Data

Data used in this paper was obtain from Worldometers website (Worldometers, 2020). Except for China, the data is the daily records of the cumulative confirmed cases, from the 15 February to 30 June 2020. Notice that the data of China used here is for the first wave of COVID-19 epidemic which occurred between 22 January to 2 March 2020.

2.1 Model development

The semi- logarithmic plot of cumulative cases of an epidemic grows asymptotically towards a steady state condition (Fig. 1). Data plotted in Fig. 1 is for COVID-19 in China (Worledimeters, 2020) and SARs, which occurred in Singapore in 2003 (Ang, 2007). The two curves shown in Fig. 1 resemble the response of first order systems to step loads. Accordingly, it has been assumed that the logarithm of cumulative cases of an epidemic can be modelled as a first order system subjected to a sudden step load (i.e. increase in number of infected people). Therefore, important information about the dynamics of epidemics (e.g. time constant and time to reach steady-state condition) can accurately be extracted from the data. Notice that conversion of data from arithmetic scale to logarithm scale will just condense it but will not change it. Further, any information obtained in logarithmic scale can easily be converted back to arithmetic scale through calculation of the anti-logarithm. Furthermore, all information about time is and will remain in arithmetic scale.

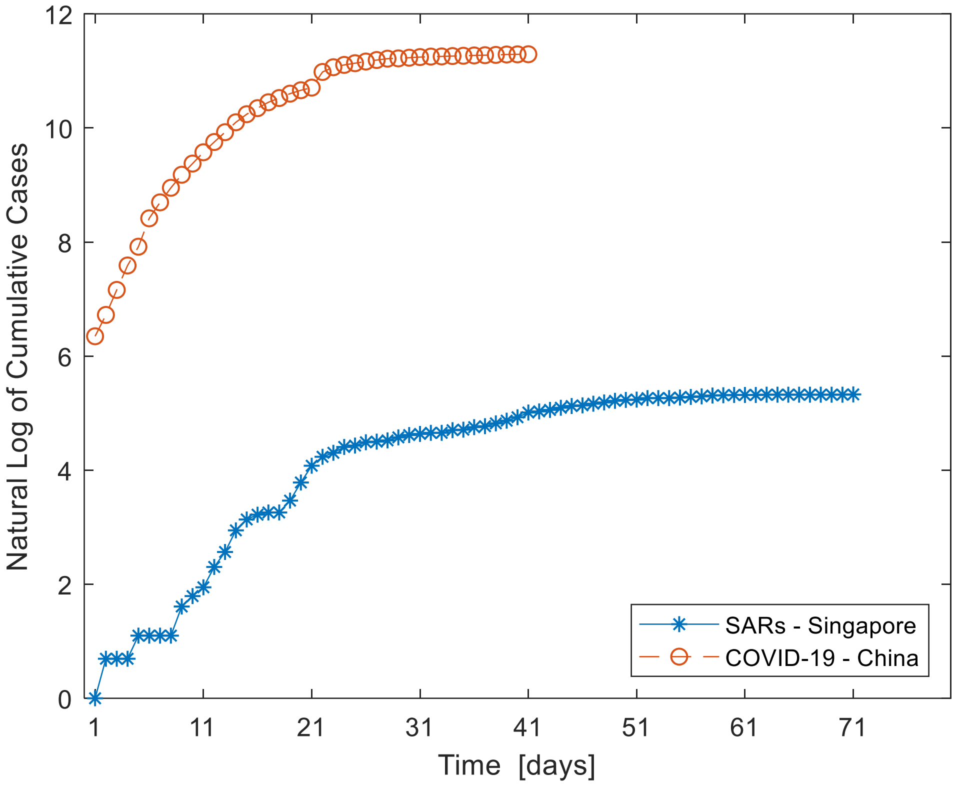

Fig. 1. Natural Logarithms of Cumulative Cases of SARs of 2003 in Singapore and COVID-19 in China.

First order systems are modeled by a first order differential equations. Denoting for the logarithm of the cumulative cases by C’ (pronounced C prime), the response of the system to a step input load u (a sudden change in number of infections) can thus be described mathematically as follows:

$$T \cdot \frac{dC'(t)}{dt} + C'(t) = k \cdot u(t) \text{ ....(Eqn. 1)}$$

Where

C’: Logarithm of the cumulative number of infected cases.

T: Process time constant (day).

u: Step input load (sudden change in number of infected persons).

k: process gain (increase in number of cumulative infected persons).

u: process input load (sudden increase in number of infected persons).

Solution of Eqn. 1 is (Bequette, 1998):

$$C'(t) = k \cdot du \cdot \left[1 - e^{\frac{-t}{T}}\right] \text{ ....(Eqn. 2)}$$

Where

du: Change of input load w.r.t. initial steady state conditions.

Notice that it is not necessary that the initial steady-state be equal to zero (no infected person). Initial steady-state greater than zero means that certain number of infection was sustained before the epidemic appeared.

Assuming the product of k and du equals K, Eqn. 2 can be re-written as:

$$C'(t) = K \cdot \left[1 - e^{\frac{-t}{T}}\right] \text{ ....(Eqn. 3)}$$

After rearranging and taking the natural logarithm, Eqn. 3 can also be re-written as a straight line equation:

$$ln[K - C'(t)] = -\frac{t}{T} + ln[K] \text{ ....(4)}$$

Hence for first-order system, the plot of ln[K-C’] is linear (Eqn. 4). To test that, ln[K-C’] of data for Kuwait and Oman was plotted and presented in Fig. 2. From this figure, it is apparent that the plots are not perfectly straight lines, but they can reasonably be approximated by straight lines, except for the first day(s) of the epidemic. Since modeling is a simplification of reality, it is acceptable to assume the plots of ln[K-C’] is a perfect straight lines, when the first day(s) is excluded. That means the initial steady-state condition before the epidemic, C’(0), should be considered an unknown parameter to be estimated from data. This assumption seems to be reasonable since the first record(s) of data is often not reflecting the true number of persons initially infected.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2020 doi:10.20944/preprints202007.0186.v1

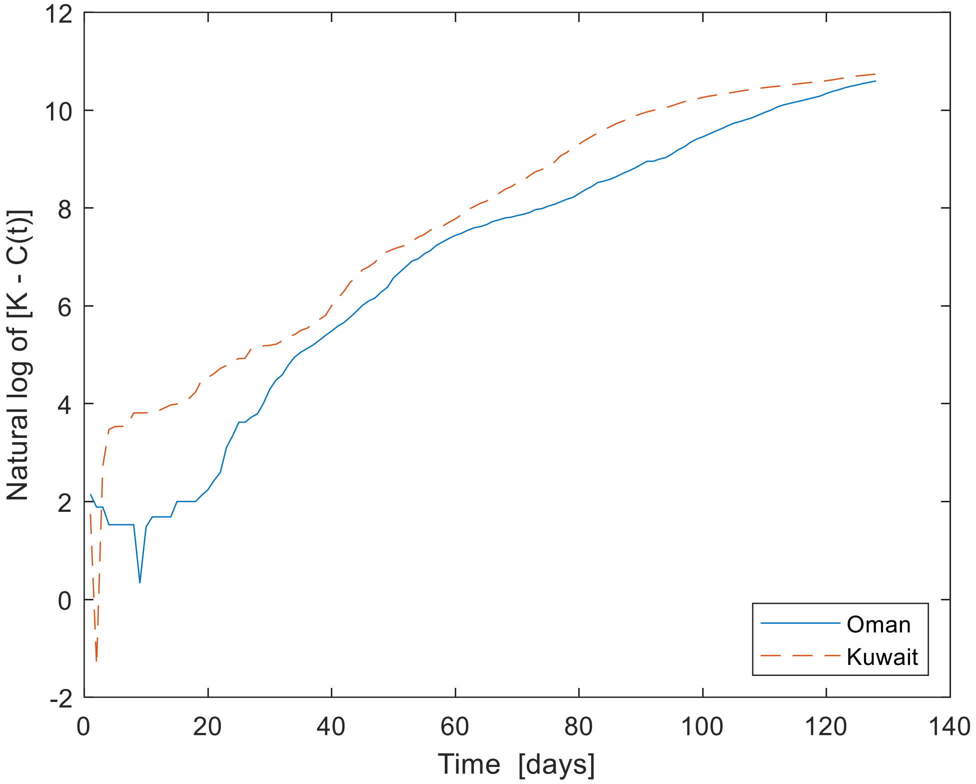

Fig. 2. Shape of the Natural log of K – C(t) for Oman and Kuwait

With unknown initial steady-state condition, the model will be:

$$C'(t) = C'(0) + K \cdot \left[1 - e^{\frac{-t}{T}}\right] \text{ ....(Eqn. 5)}$$

The developed model (Eqn. 5) has only three unknown parameters (C’(0), K and T) to be estimated from data. The parameter C’(0) is the number of infected persons just before the epidemic become apparent. The parameter T is the speed at which the epidemic grow (time constant of the epidemic, T) to reach the steady-state condition. From the value of T, time will be taken to reach steady-state condition can also be estimated. Since first order systems achieve more than 98% of their ultimate gain in time equals 4T, therefore they are considered reached steady-state at 4T (Bequette, 1998). Thus, accurate estimation of these physically meaningful parameters will provide public health decision makers with very important information about the dynamics of epidemics.

2.3 Model Calibration and Validation

Data of the first wave of COVID-19 epidemic in China was mainly chosen to test the model (Eqn. 5) because it had already reached its end and thus all actual information about it is available.

For calibration, the model was run with data of the first 20 days only. The least square technique, through MATALB function *lsqcurvefit,* was used to fit model's prediction to the data. Calibration of the model had resulted in estimation of the values of the three unknown parameters: C'(0), T and K.

To validate the model, it was then run with whole 41-days data, *using the same* parameters' values obtained during the calibration step.

Dynamics of epidemics usually differ from a community to another and thus from a country to another. To demonstrate the ability of the model in identifying the different dynamics of the same epidemics in different countries, the model was also heuristically fitted to COVID-19 data of the following 13 countries: Bahrain, France, Kuwait, Italy, Netherlands, Oman, Qatar, Singapore, Spain, Saudi Arabia, UAE, UK, and USA.

3.0 Results and Discussion

Values obtained from model calibration for the three parameters C'(0), T and K, were 5.7296 ± 0.0027, 9.6455 ± 0.0149 and 5.6893 ± 0.0022, respectively. Notice that values of C'(0) and K are in logarithmic scale, whereas that of T is in linear scale. It should also be noted that the standard errors, which were estimated using the Jacobian matrix (Abusam et al 2000 and 2001), are much smaller than parameters' values. These results clearly indicate that the three parameters can be identified accurately and reliably from the data.

The anti-log of the value estimated for C'(0) (5.7296) is 308. This means that there were initially 308 persons infected before the step increase in number of infections which resulted in the appearance of the epidemic. Given that the anti-log of the prediction for the first day is 520 (571 from data). The step increase in number of infection is thus 212 persons (520 – 308). Therefore, the sudden increase in number of infected persons by 212 had caused the epidemic to appear.

From model calibration, the estimated process time constant (T) is 9.65 days. This means that the epidemic will reach steady-state condition in about 39 days (4 x 9.65 days). Therefore, that is a highly reasonable estimate since the epidemic in China was considered to be over in 41 days. The epidemic in China finished so fast because very strict control measures had been implemented (Lin et al 2020). In short, results of model calibration seemed to be very realistic and accurate.

Fig. 3. presents the validation results of the model. This figure clearly shows that the model fits well the data. That is also confirmed from the small value of root mean square errors (RMSE), which was 0.1608 in logarithmic scale.

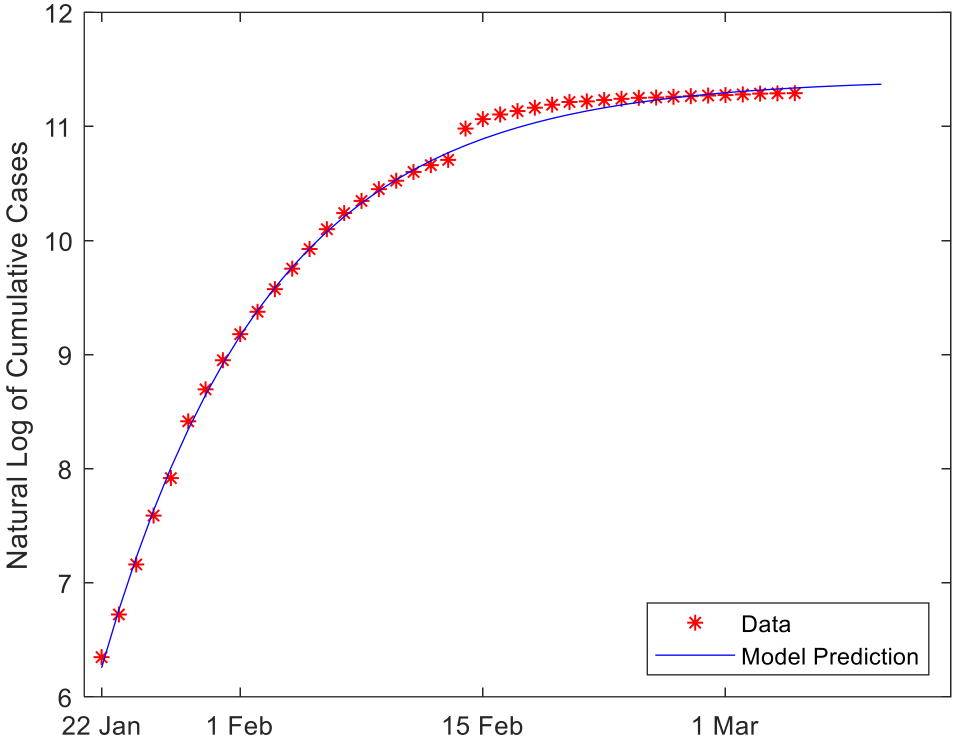

Fig. 3. Results of Model Calibration and Validation with data of China

Further, the model predicts the cumulative number of infected persons of day 41 to be 83885 persons (anti-log of 11.3372), which is only 4.66% more than the reported value (80151). Such good and accurate forecast indicates that the developed model can be used reliably to forecast the dynamics of COVID-19 epidemic.

Results of the heuristic fitting to the data of 13 countries other than China are given in Table 1. The small values of RMSE, which ranges from 0.0881 to 0.4504, indicate that the model fits very well the dynamics of COVID-19 at the different countries. This can also be seen from Fig. 4 which presents the model fit to the data of Italy as an example. The highest value of RMSE (0.4504) is obtained for Singapore. As can be from Fig. 5, the fit of model with data of Singapore is not as good as that, for example, for Italy (Fig. 4). The relatively large value estimated for the time constant of Singapore (93 days) indicates the epidemic there still has long time (day 376) to go before approaching its end. In spite of the relatively poor fit, however, all the model parameters were also accurately estimated from the data of Singapore as indicated by small standard errors (Table 1).

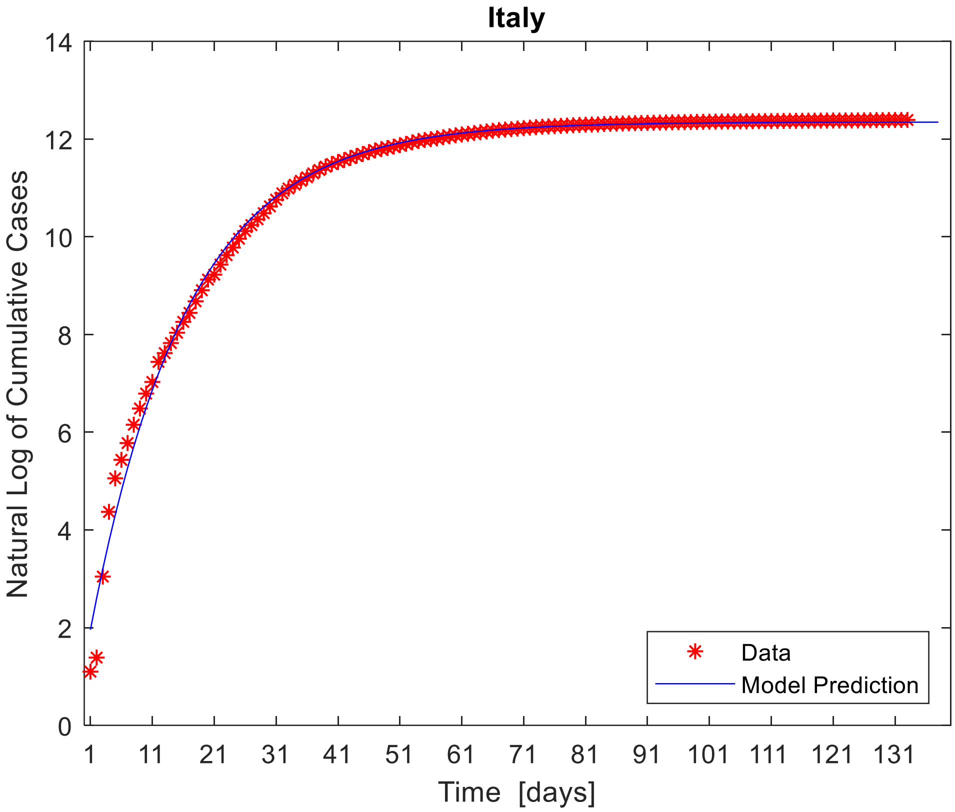

Fig. 4. Results of Heuristically fitting the Model to COVID-19 Data of Italy

Table 1. Estimated Values of the Parameters When the Model Fitted Heuristically to COVID-19 data of 13 Countries

Country	**Ln(C’0)**	**T (days)**	**Ln(K)**	**Tss (days)**	**Ln(RMSE)**
Bahrain	2.9222 ± 0.0199	74.7071 ± 0.9632	8.8036 ± 0.0437	299	0.2441
France	1.4287 ± 0.0060	18.0408 ± 0.0171	10.5958 ± 0.0057	72	0.1061
Kuwait	2.5815 ± 0.0246	108.8433 ± 2.1438	12.4205 ± 0.1270	435	0.2794
Italy	1.2715 ± 0.0205	15.6460 ± 0.0462	11.0756 ± 0.0199	63	0.1899
Netherlands	0.6278 ± 0.0043	16.4657 ± 0.0116	10.1666 ± 0.0042	66	0.0881
Oman	0.8628 ± 0.0108	85.4577 ± 0.5237	12.5832 ± 0.0332	342	0.1826
Qatar	1.2787 ± 0.0718	41.9013 ± 0.8362	10.7472 ± 0.0684	168	0.4314
Singapore	3.1592 ± 0.0634	93.8916 ± 4.0394	10.6020 ± 0.2013	376	0.4504
Spain	0.1000 ± 0.0176	18.1273 ± 0425	12.5877 ± 0.0169	73	0.1822
Saudi Arabia	0.3714 ± 0.0275	34.6036 ± 0.2035	11.8775 ± 0.0250	138	0.2590
UAE	0.7539 ± 0.0590	64.4939 ± 1.3758	11.9359 ± 0.0835	258	0.4196
UK	0.4853 ± 0.0379	27.8111 ± 0.1747	12.3680 ± 0.0350	111	0.2944
USA	1.9445 ± 0.1083	29.1825 ± 0.5137	13.0000 ± 0.0994	117	0.5022

Tss: Time to steady-state condition

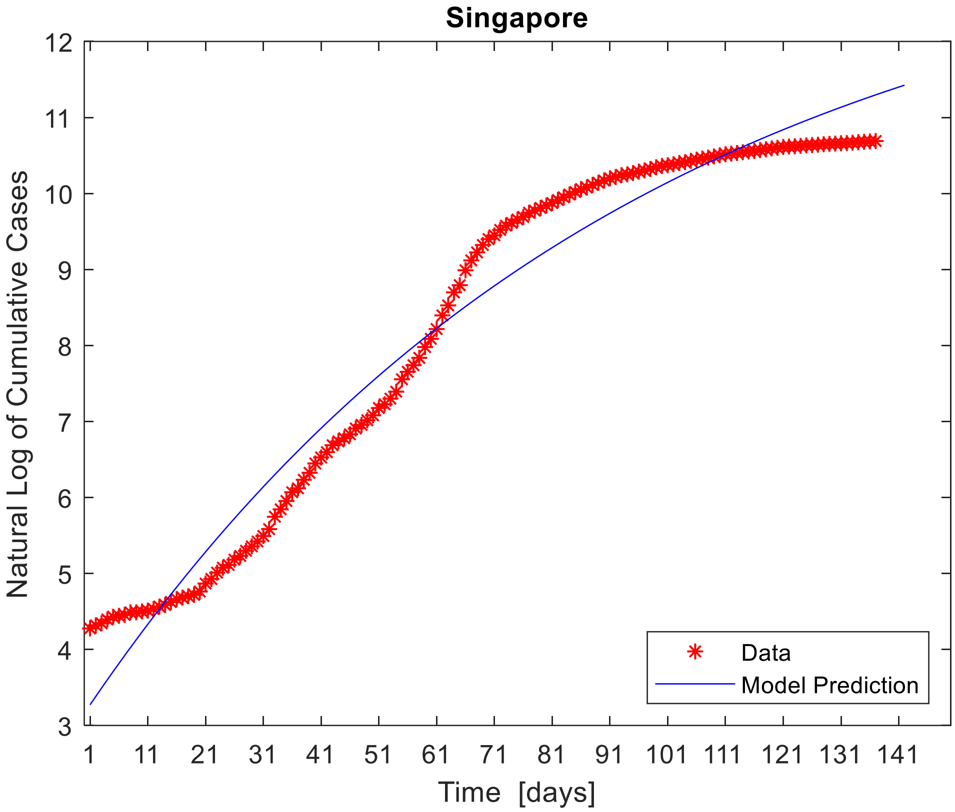

Fig. 5. Model Heuristically fitted to COVID-19 Data of Singapore

From heuristic fitting to different countries, as expected, almost completely different values of the model's parameters were obtained for the different countries (Table 1). This demonstrate clearly that the model is able to predict the difference in dynamics of the epidemic under different conditions. Table 1 shows that the values estimated for the initial steady-state conditions (C'(0)), process gain (K) and the process time constant (T) are unique for each country. However, it can be noticed that the initial steady-state conditions (C'(0)) for the European countries (France, Italy, Netherlands and Spain) is relatively smaller (15 -20 days) than that for the other countries, e.g. for Kuwait (109 days). This means that the model predicted that the epidemic in the four mentioned European countries will reach steady-state conditions in only 60 to 80 days, while it that will take 435 days in Kuwait. This seems to be true. The four European countries have already started to ease the control measures and to return normal life, while Kuwait is still implementing some strict regulations.

It becomes clear from above that the developed model can be used reliably for extracting the dynamics of COVID-19 epidemic. However, the same model can also be used to model other epidemics. As shown for SARs in Fig. 1, the semi-logarithmic plots of the epidemiological curves look like a response of a first-order system to a step load. Therefore, the same model can also be used to modeled other epidemics.

4.0 Conclusions

A simple first-order system model, with physically meaningful parameters, has been developed for infectious disease epidemics and illustrated using the data of COVID-19 pandemic.

Model calibration has indicated that all the three unknown parameters of the model (state initial condition, process gain and process time constant) can be accurately identified from data.

The developed model has also proved to be reliable at revealing the difference in the dynamics of COVID-19 pandemic in other 13 countries.

Acknowledgement

COVID-19 data was obtained from Worldometer website.

References

1. Abusam, A., Abusam, R. and Al-Anzi, B. (2020). Adequacy of Logistic Models for Describing the Dynamics of COVID-19 Pandemic, Submitted to Infectious Disease Modelling.
2. Abusam, A., K. J. Keesman & Van Straten G. (2001). Oxygen transfer rate estimation in oxidation ditches from clean water measurements, Water Research 35(8):2058-2064.
3. Abusam, A., K.J. Keesman, Van Straten G., Spanjers H. and Meinema K. (2000). Parameter estimation procedure for complex non-linear systems: calibration of ASM No.1 for N-removal in a full-scale oxidation ditch, Wat. Sci. Tech. 43(7):357-365.
4. Anderson, RM., May, R.M. (1991). Infectious diseases of humans. Oxford University Press; Oxford, UK.
5. Anderson, R.M., May, RM. Infectious diseases of humans. Oxford University Press; Oxford: 1991.
6. Ang, K. -C (2007). A simple stochastic model for an epidemic – numerical experiments with matlab, The Electronic Journal of Mathematics and Technology, Volume 1, Number 2
7. Bequette, B.W. (1998). Process dynamics – modeling, analysis and simulation, Prentice Hall, Upper Saddle River, New Jersey, USA.
8. Chowell, G. Sattenspiel, L., Bansal, S. and Viboud, C. (2016). Mathematical models to characterize early epidemic growth – A review, Physics of Life Reviews 18:16-97. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5348083/

9. Sanglier, G., Cesteros, S., Lopez, E.J. and Gonzalez, R.A. (2020). Mathematical model and data anlysis to determine the number of confirmed infection due to COVID-19 in Spain, Modern Applied Science 14:60-74.
10. Lin, X., Hewings, G., Zheng, S. and Li, X. (2020). Modeling the situation of COVID-19 and effects of different control strategies in China with dynamic differential equation and parameter estimation. https://www.researchgate.net/publication/340879118_Modeling_the_situation_of_COVID-19_and_effects_of_different_containment_strategies_in_China_with_dynamic_differential_equations_and_parameters_estimation_Preprint
11. Louchet, F. (2020). A brief theory of epidemic kinetics. https://www.preprints.org/manuscript/202004.0265/v1
12. Rabbani, H.S., Osei-Bonsu, K., Osei-Bonsu, P.K., Seers, T.D. (2020). Modeling COVID-19 using the fundamentals of fluid dynamics. https://www.medrxiv.org/content/10.1101/2020.06.24.20139071v1
13. Caccavo, D. (2020). Chinese and Italian COVID-19 outbreaks can be correctly described by modified SIRD model, Available at: https://www.researchgate.net/publication/340078430_Chinese_and_Italian_COVID-19_outbreaks_can_be_correctly_described_by_a_modified_SIRD_model.
14. Batista, M. (2020). Estimation of the final size of the COVID-19 Epidemic, Available at: https://www.researchgate.net/publication/339355847_Estimation_of_the_final_size_of_the_COVID-19_epidemic.
15. Liu, X., Hewings, G., Wang, S., Qin, M., Xiang, X., Zheng, S. & Li, X. (2020). Modelling the situation of COVID-19 and effects of different containment strategies in China with dynamic differential equations and parameters estimation. Available at: https://www.researchgate.net/publication/340879118_Modeling_the_situation_of_COVID-19_and_effects_of_different_containment_strategies_in_China_with_dynamic_differential_equations_and_parameters_estimation_Preprint
16. Ma, J. (2020). Estimating epidemic exponential growth rate and basic reproduction number, Infectious Disease Modelling 5:129-141. Available at: https://www.researchgate.net/publication/338454727_Estimating_Epidemic_Exponential_Growth_Rate_And_Basic_Reproduction_Number
17. Roda, W. C. Varughese, M. B. Han, D. & Li, M. Y. (2020). Why it is difficult to predict accurately the COVID-19 epidemic? Infectious Disease Modelling 5: 271-281. Available at: https://www.researchgate.net/publication/339850601_Why_Is_It_Difficult_to_Accurately_Predict_the_COVID-19_Epidemic
18. Worldometer (2020). COVID-19 coronavirus outbreak. World meters, 2020. Available at: https://www.worldometers.info/coronavirus/#countries.