How Good or Bad is the Prediction of Black-Box COVID-19 Models After One Year Has Lapped?

Hennie Husniah
Department of Industrial Engineering
Langlangbuana University
Bandung, 40116, INDONESIA
hennie.husniah @gmail.com

Asep K. Supriatna
Department of Mathematics
Padjadjaran University
Jatinangor, 45363, INDONESIA
aksupriatna@gmail.com

Abstract

In this paper we present a report on the performance of two black-box mathematical models of COVID-19 prediction. The models are the growth equation models: modified Verhulst Logistic Model and the Morgan-Mercer-Flodin (MMF) Model. Both models predict the cumulative numbers of infected people using the COVID-19 Data of Indonesia from begining day of the infection to end of September 2020. The COVID-19 pandemic data in Indonesia from the Worldometer website (Worldometer 2020) will be used as the underlying data for the curve fitting to those two models. We use the time series data of the total confirmed cases and the daily new cases to fit to the model models. We used the data starting on 2 March, the official first day of the reported pandemic cases in Indonesia, up to 30 November 2020. We used the models to fit the time series of the total confirmed cases by minimizing the minimum root of the mean square error (RMSE) to obtained the parameterized models. After a year has been lapped, we compare the prediction from those parameterized models to the actual number of infected people for the same country, from the beginning of epidemic to 30 August 2021. The results show that for certain condition (different assumptions of the two models) the models could predict the number in a relatively good accuracy, at least qualitatively could portray the shape of the cumulative number of infected people curve.

Keywords
COVID-19 Pandemic, Empirical Model, Indonesia, Verhulst Logistic Equation, Morgan-Mercer-Flodin Equation.

1. Introduction
COVID-19 appeared as a pandemic is only about two years ago officially. However the impact is very devastating to almost all aspect of human life. It is a very dangerous new disease in terms of health impact, economy, and other human aspects. As a new disease it is already has been declared as a Pandemic by the World Health Organization (WHO) on March 11, 2020. This disease is caused by the Corona SARS-2 virus and is thought to have originated in Wuhan, China. At that time, more than 118,000 cases were recorded in 114 countries with 4,291 people losing their lives (WHO 2020). To see the widespread of the disease, we note that as of April 11, 2020, the figure has increased dramatically with a total number of more than 1.5 million cases of infection and more than one hundred fatalities and at the end of July 2020, there were 16,839,692 recorded positive cases of infection with 661,379 deaths, and even currently those number have reached to nearly 27 million and one million, respectivelly (Worldometer 2020). COVID-19 so far is still regarded as a disease that difficult to understand, let alone to control, due to several reasons such as lack of data and confusing report (Caudill 2020). This is among the reasons why so far most infected countries still struggling combating the disease.

1.1 Objectives
In this paper we compare the prediction from both models after the COVID-19 pandemic make a continous progression up to this date, specifically we will show that the models predict differently for the long-term prediction of the pandemic.

2. Literature Review
Since the announcement of the pandemic, almost every country has made very intensive effort to combat the disease, albeit with a wide variety of responses. Efforts are generally directed to handling cases of infection, prevention of transmission and development of early detection methods for monitoring transmission of the disease. Some examples of the responses are documented in Djalante et al. (2020), Kaguyo et al. (2020). Various collaborative research efforts are made to develop better strategies for controlling the spread of disease based on a scientific basis including the use of mathematical modeling to understand COVID-19 data.

So far there are a lot of mathematical models have been developed to study the COVID-19 data, using different mathematical methods and approaches. Some using mechanical white box models and others using empirical black box models. Both have equally strengths and weaknesses, depending on the target of the modeling purposes. Among the empirical models are (Shen 2020, Zou et al. 2020, Aviv-Sharon and Aharoni 2020, Wang et al. 2020, Wu et al. 2020, Ghosh et al. 2020). All the above mentioned authors have utilized the logistic growth function (or the modified logistic of the classical Verhulst model) in analyzing the COVID-19 data for various countries. The results are quite good and fit to the real data which have sigmoid curves. Besides Verhulst model, there are also the Morgan-Mercer-Flodin (MMF) model.

Husniah and Supriatna (2020) use a modified Verhulst logistic equation to model the cumulative number of COVID-19 infected people in Indonesia. They also use the MMF equation to model the same data (Husniah and Supriatna and Husniah, 2020). Both models give a relatively the same degree of accuracy in predicting the cumulative number of infected people and able to capture roughly the portrait of the pandemic shown in Figure 1. In this paper we compare the prediction from both models after the COVID-19 pandemic make a continous progression up to this date, specifically we will show that the models predict differently for the long-term prediction of the pandemic. Here the Verhulst logistic model and the MMF model which parameterized by the Indonesia COVID-19 data (from the beginning of the pandemic to 30 September 2020) are able to predict Figure 2, but with completely different assumption, which will be shown later on in this paper. In the following section we will give a brief methodological parameterization as described in Husniah and Supriatna (2020) and Supriatna and Husniah (2020).

3. Methods
We obtained the pandemic data in Indonesia from the Worldometer website (Worldometer 2020). The data which are available from the website include total confirmed cases, daily new cases, daily active cases, daily death, etc. However, we only use the time series data of the total confirmed cases to fit with the logistic model. We used the data starting on 2 February, the official first day of the reported pandemic cases in Indonesia, up to 14 September 2020. We used the modified Verhulst logistic model and the MMF model to fit the time series of the total confirmed cases. For the Verhulst model, the parameters refer to the asymptotic value (carrying capacity or the maximum number of total confirmed cases, K) and the logistic growth rate or steepness of the curve (r). In applying the logistic equation to the pandemic data we denoted that $X(t)$ is the cumulative of confirmed case at time t. For the MMF model, L_∞ is the carrying capacity or the maximum value of the curve and β is the initial value of the growth. In the case of MMF model we give two different methods as follows. We used the MMF model to fit the time series of the total confirmed cases by minimizing the minimum root of the mean square error (RMSE) of the total confirmed cases (Method 1) and by minimizing the RMSE of the daily new cases (Method 2). The calculation is done using Solver in the Microsoft Excel application by choosing the GRG Nonlinear (Generalized Reduced Gradient) for the oftimization to find the minimum root of the mean square error as the measure.

3.1 The Verhulst Logistic Function
The Verhulst logistic equation is among the most popular equation to describe a growth phenomenon. For biologists, techniques or models to understand the dynamics of growing and shrinking populations of living organisms is vital. Among the early researchers is Thomas Malthus who first pointed out that populations grew exponentially (Malthus 1798). This is true in some situation, e.g. in a relatively unlimited resources. However, in fact most populations grow up approaching an upperbound. A model of population growth that considers this upperbound, which later is called

carrying capacity, is the logistic model of population growth which has been formulated by Pierre François Verhulst in 1838. It is a sigmoid curve that describes "the growth of a population as exponential followed by a decrease in growth, and bound by a carrying capacity due to environmental pressures" (Renshaw 1991). The logistic function has the form

$$X(t) = \frac{K}{1 + \frac{(K - X_0)e^{-r(t-t_0)}}{X_0}} . \tag{1}$$

Where K is the carrying capacity or the maximum value of the curve, r is the the logistic growth rate or steepness of the curve, and X_0 is the X value of the sigmoid's midpoint. Appendix 1 shows that the logistic growth function has an inflection point at (t_i, x_i) with

$$t_i = \frac{\ln\left(\frac{K - X_0}{X_0}\right)}{r} , \quad X_i = \frac{K}{2} . \tag{2}$$

It is easy to prove that the function is initially increases with the curve concave up until the time t_i, then afterwards concave down approaching the carrying capacity K forming a sigmoid curve.

3.2 The Modified Verhulst Logistic Function

There are many modifications available for the classical Verhulst logistic function in literatures. In general the modification process is done through a new formulation. To be specific, here we give a flexibility for the equation to reach the inflection point by modifying equation (2) to the folowing form

$$X(t) = \frac{K^\alpha}{1 + \frac{(K - X_0)e^{-rt}}{X_0}} . \tag{3}$$

This form gives an inflection point slightly different to that in (2), (see also Appendix 2), i.e.

$$t_i = \frac{\ln\left(\frac{K - X_0}{X_0}\right)}{r} , \quad X_i = \frac{K^\alpha}{2} . \tag{4}$$

Time to inflection point is the same but the height of the curve at inflection time is different.

3.3 The Morgan-Mercer-Flodin Growth Function

The MMF equation is among available equations to describe a growth phenomenon, although it is not as popular as the Verhulst logistic growth equation. For biologists, techniques or models to understand the dynamics of growing and shrinking populations of living organisms is vital. Among the early researchers is Thomas Malthus who first pointed out that populations grew exponentially (Malthus 1798). This is true in some situation, e.g. in a relatively unlimited resources. However, in fact most populations grow approaching to an upperbound. A model of population growth that considers this upperbound, which later is called carrying capacity, is the logistic model of population growth which has been formulated by Pierre François Verhulst in 1838. It is a sigmoid curve that describes "the growth of a population as exponential, followed by a decrease in growth, and bound by a carrying capacity due to environmental pressures" (Renshaw 1991). Despite its popularity, the logistic equation sometimes perform unsatisfactorily (Supriatna and Husniah, 2020). There are a lot of modification of the logistic or new sigmoid equation available in the literature. One of theim is the one developed by Morgan et al. (1975) which better known as the Morgan-Mercer-Flodin (MMF) growth equation.

The MMF function has the form (Tariq et al., 2013)

$$X(t) = L_\infty - \frac{L_\infty - \beta}{1+(kt)^\delta} \tag{5}$$

where L_∞ is the carrying capacity or the maximum value of the curve and β is the initial value of the growth. Appendix 1 shows that the MMF growth function has an inflection point at (t_i,x_i) with

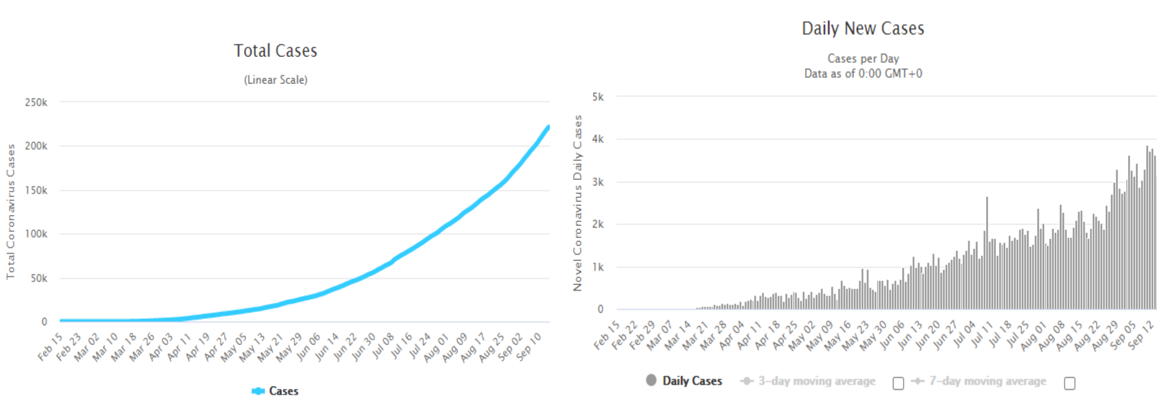
$$t_i = \frac{1}{k}\left(\frac{\delta-1}{\delta+1}\right)^{1/\delta} \tag{6}$$

$$X(t_i) = \frac{L_\infty\left(\frac{\delta-1}{\delta+1}\right)\delta k}{\left(\frac{\delta-1}{\delta+1}\right)^{1/\delta}\left(1+\frac{\delta-1}{\delta+1}\right)} - \frac{\left(L_\infty\left(\frac{\delta-1}{\delta+1}\right)+\beta\right)\left(\left(\frac{\delta-1}{\delta+1}\right)\delta k\right)}{\left(1+\left(\frac{\delta-1}{\delta+1}\right)\right)^2\left(\frac{\delta-1}{\delta+1}\right)^{1/\delta}} \tag{7}$$

It is easy to prove that the function is initially increases with the curve concave up until the time t_i, then afterwards concave down approaching the carrying capacity L_∞ forming a sigmoid curve.

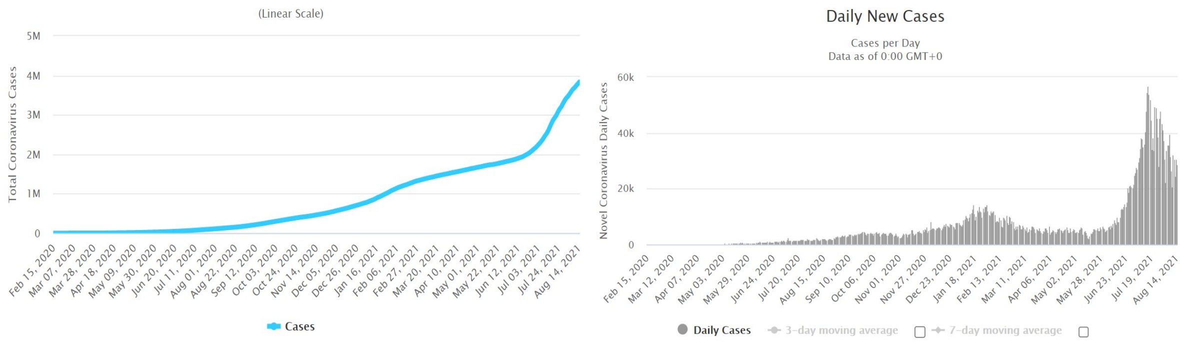
4. Data Collection

Since it first appearence in Wuhan, the propagation in many countries still in increasing phase, including in Indonesia. Figure 1.a shows that as September 2020 there is no indication that it would decelerate the propagation velocity. Theoretically all disease if it is untreated will experience a sigmoid curve growth for the cumulative confirmed cases, since eventually most people will be contracted by the disease. Consequently, this will make the number of susceptible declines and makes the difficulty for the disease to find the target. This will also make the number of the daily new cases declines. This theoretical situation has not yet appeared in the case of Indonesia as September 2020 (Figure 1. b). This trend also holds for the cumulative confirmed cases worldwide (Anonymous 2020). As August 2021 the portrait of the pandemic in In donesia is prortrayed in Figure 2.



(a) (b)

Figure 1. Total Coronavirus Cases (a) and Daily New Cases (b) in Indonesia from the beginning of the disease (2 March 2020) to 14 September 2020. The figures show that the disease still in the exponential growth phase. Source: https://www.worldometers.info/coronavirus/country/indonesia/



(a) (b)

Figure 2. Total Coronavirus Cases (a) and Daily New Cases (b) in Indonesia pandemic Indonesia from the beginning of the disease (2 March 2020) to date as August 2021.

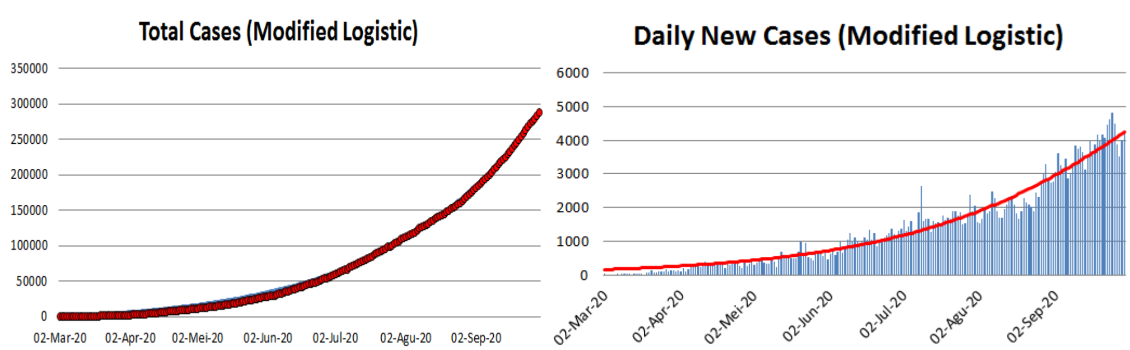
5. Results and Discussion

We obtained pandemic data in Indonesia from the Worldometer website (Worldometer 2020). The data which are available from the website include total confirmed cases that we fitted to the classical Verhulst logistic function (1), and the modified logistic function (3), and the MMF function (5). The resulting parameters are presented in Table 1.

Table 1. Resulting values of parameters

Model	Equation	Parameters				RMSE
Verhulst Logistic	(1)	$K = 160{,}068$	$r = 0.0840880$	-	-	18,221
Modified Verhulst Logistic	(3)	$K = 301.195$	$r = 0.02761$	$\alpha = 2.23844$	-	2,761
MMF (Method 1)	(5)	$L = 1{,}714{,}829{,}160.66$	$k = 0.000187372$	$\beta = 0.000000001$	$\delta = 2.71584844575$	1584
MMF (Method 2)	(5)	$L = 2{,}148{,}297{,}910.89$	$k = 0.00027652$	$\beta = 0$	$\delta = 3.15329536631$	274

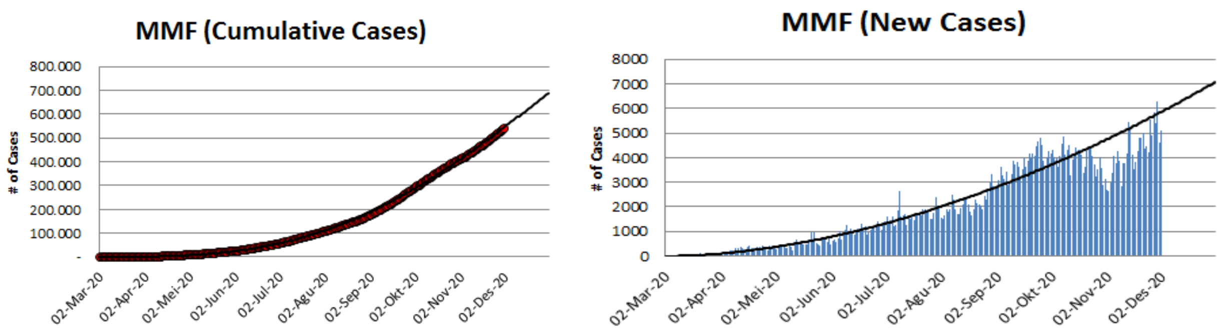
All the equations discussed above with parameter values in Table 1 are able to model the total number of infection during the period of February 2020 to September 2020 (Figure 1) as described in Husniah and Supriatna (2020) and Supriatna and Husniah (2020). The real data in Figure 1 can be satisfactorilly approximated by the modified Verhulst logistic equation as seen in Figure 3 and by the MMF equation as seen in Figure 4.



(a) (b)

Figure 3

Total Coronavirus Cases (a) and Daily New Cases (b) in Indonesia pandemic data fitted by the modified logistic equation. The data used to parameterize the equation are taken from the beginning of the disease to (2 March 2020) to 14 September 2020. The figures fits in agreement with the cumulative and new cases in Figure 1.



(a) (b)

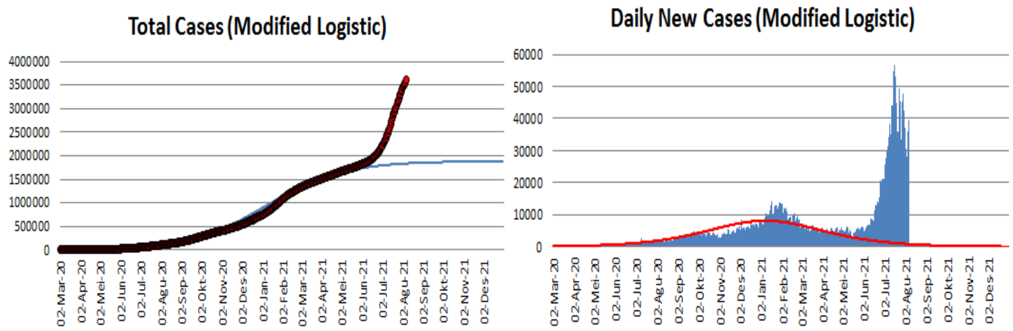
Figure 4

Total Coronavirus Cases (a) and Daily New Cases (b) in Indonesia pandemic data fitted by the MMF growth function (Method 1). The data used to parameterize the equation are taken from the beginning of the disease (2 March 2020) to 30 November 2020. The figures fits in agreement with the cumulative and new cases in Figure 1.

5.1 Discussion on the Modified Verhulst Logistic Model

The following discussion will emphasize on the total number of infection during the period of February 2020 to August 2021 (Figure 2), almost a year after the parameterization of the growth equations were carried out. Figure 2 shows that in the early 2021 (January/February 2021) the number of new cases begins to decrease. This figure is satisfactorily captured by the modified logistic model (red curve in Figure 5.b). However the modified logistic model fails to capture the surge of new cases in the middle of 2021 (June/July 2021). This failure can be explained partly by the assumption of the presence of new COVID-19 variants in Indonesia.

As July 2021 it was detected that there has been a sharp increase in the number of Covid-19 new cases as reported in mass-media, especially in Jakarta (Sofa, 2021). The report pointed out that some new variants such as Alpha (B.1.17), Beta (B.1.351), and Delta (B.1.617.2) are also found. The Indonesian Health Research and Development Agency shows that on 20 June 2021, there were 33 cases of Alpha variant, 4 cases of Beta variant, and 57 cases of Delta variant in Jakarta. The biggest contribution to the sharp increase is the Delta variant which has faster virus transmissions and a higher risk of being hospitalized (Sofa, 2021). A warning just before the surge of new cases happens has been made by the Governor of Jakarta. He warned that although at that time the number of new cases has begin to decline (in Jakarta), there could be another outbreak for two reasons, (i) there already found new variant in India (Delta variant) which is far more harmful and devastating, that might be enter / imported to Indonesia, (ii) there is a long holiday in Indonesia, that might be increase people mobility, and might change people behaviour towards the disease, i.e. less aware and less cautios to the disease (https://www.youtube.com/watch?v=HCs0OOoTS9A). Considering this situation, we may assume that Indonesia at the time has experienced two waves of COVID-19. The first wave is succesfully described by modified Verhulst logistic model and presumably will end if there is no new COVID-19 variant appear as suggested by red curve in Figure 5.b., while the surge of new cases might be caused by different variant and should have different equation. If in fact the surge is caused by the same virus, then in this case the modified Verhulst logistic equation fails to describe the COVID-19 transmission.

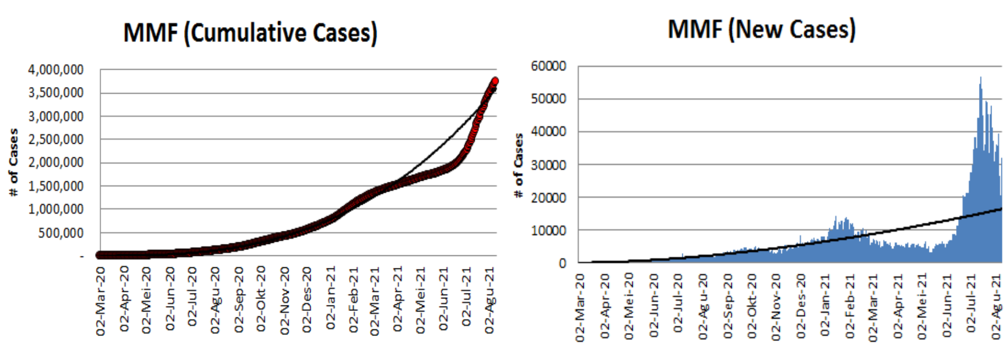


(a) (b)

Figure 5

Total Coronavirus Cases (a) and Daily New Cases (b) in Indonesia pandemic data fitted by the modified logistic equation. The data used to parameterize the equation are taken from the beginning of the disease to (2 March 2020) to 14 September 2020. In (a), the thin blue curve is the predicted cases and the red circles with black boundary is the actual cases (appeared as a thick black curve), while in (b) the red curve is the predicted cases and the blue bars is the actual cases.

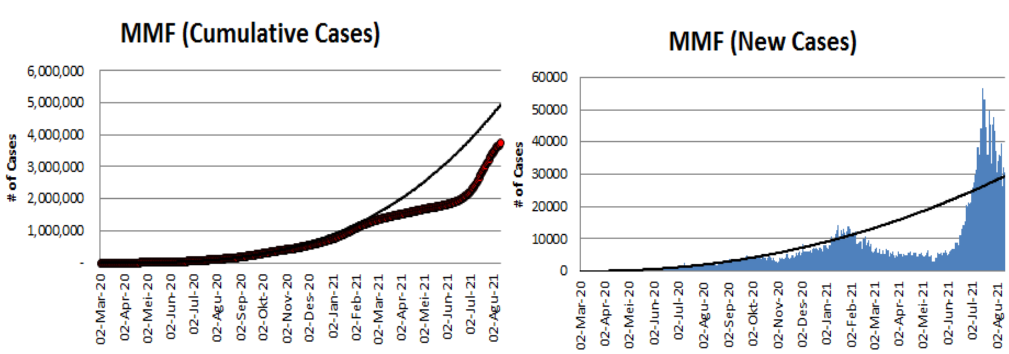
5.2 Discussion on the MMF Model



(a) (b)

Figure 6

Total Coronavirus Cases (a) and Daily New Cases (b) in Indonesia pandemic data fitted by the MMF equation. The data used to parameterize the equation are taken from the beginning of the disease to (2 March 2020) to 30 September 2020 using Method 1. In (a), the thin black curve is the predicted cases and the red circles with black boundary is the actual cases (appeared as a thick black curve), while in (b) the black curve is the predicted cases and the blue bars is the actual cases.



(a) (b)

Figure 6

Total Coronavirus Cases (a) and Daily New Cases (b) in Indonesia pandemic data fitted by the MMF equation. The data used to parameterize the equation are taken from the beginning of the disease to (2 March 2020) to 30 September 2020 using Method 2. In (a), the thin black curve is the predicted cases and the red circles with black boundary is the actual cases (appeared as a thick black curve), while in (b) the black curve is the predicted cases and the blue bars is the actual cases.

6. Conclusion

We modeled the pandemic data of Indonesia (the total/cumulative confirmed cases data) from the Worldometer website (Worldometer 2020) using two growth functions: the original logistic function and the modified logistic function. The results show that the logistic function fails to estimate the carrying capacity K of the total confirmed cases data. The model shows that, in terms of total confirmed cases pandemic data of Indonesia, at this date the disease should have already reached the carrying capacity K, which is in fact untrue. The inaccuracy becomes apparent when we plot the daily new cases which clearly depart form the observed data. Meanwhile, the results also show that the modified logistic equation is able to estimate the carrying capacity K^{α} of the total confirmed cases and produces curve that satisfactorily fits the data. However if we look at the result for the daily new cases, the model also shows that at this date the disease should have already reached the carrying capacity K^{α}, which is also untrue. This concludes that while the modified logistic model is able to mimic the total confirmed cases data, it still fails to fit the daily new cases data. Further refinement of the model still need to be done. This is currently under investigation. There is some reasons why the logistic growth model does not fit for COVID-19 data. One of them is when it is applied too early in which the disease still in the exponential growth phase, such as in the case of Indonesia COVID-19 data (https://www.researchgate.net/post/Logistic_Growth_Model_Is_it_suitable_for_COVID-19). The process of the logistic model refinement above bearing a good pedagogical example in teaching mathematical modeling.

Acknowledgements

This research was funded by the Indonesian Government through the scheme "PENELITIAN HIBAH RISET DAN TEKNOLOGI / BADAN RISET DAN INOVASI NASIONAL" with contract number: 1207 / UN6.3.1 / PT.00 / 2021. We thank the anonymous referees who have read the earlier version of the manuscript and raised some cronstructive comments that helped to improve the content and the presentation of the paper.

References

Anonymous, available at https://ourworldindata.org/grapher/total-covid-cases-region?time=2020-01-05..latest, downloaded 14 September 2020.

Aviv-Sharon, E. and Aharoni, A. Generalized logistic growth modeling of the COVID-19 pandemic in Asia, *Infectious Disease Modelling*, vol. 5, pp. 502-509, 2020.

Caudill, L., Lack of data makes predicting COVID-19's spread difficult but models are still vital, *The Conversation*, April, 2020. https://theconversation.com/lack-of-data-makes-predicting-covid-19s-spread-difficult-but-models-are-still-vital-135797.

Djalante, R. et al., Review and analysis of current responses to COVID-19 in Indonesia: period of January to March 2020, *Progress in Disaster Science*, vol. 6, 2020.

Ghosh, P., Ghosh, R., and Chakraborty, B., COVID-19 in India: statewise analysis and prediction, *JMIR Public Health Surveill* , vol. 6, no. 3, e20341, 2020.

Kuguyo, O., Kengne, A.P., and Dandara, C. Singapore COVID-19 Pandemic Response as a Successful Model Framework for Low-Resource Health Care Settings in Africa? *A Journal of Integrative Biology*, vol. 24, no. 8, pp. 470-478, 2020, http://doi.org/10.1089/omi.2020.0077.

Malthus, T. R.. *An Essay on the Principle of Population*. Oxfordshire, England: Oxford World's Classics. p. 13. ISBN 978-1450535540.

Meyer, W. J. Concepts of Mathematical Modeling, Dover Publication Inc, New York, 1984. ISBN 978-0486435152.

Renshaw, E., *Modelling Biological Populations in Space and Time*, Cambridge University Press, 1991, https://doi.org/10.1017/CBO9780511624094.

Shen, C. Y., Logistic growth modelling of COVID-19 proliferation in China and its international implications, , *International Journal of Infectious Diseases*, vol. 96, pp. 582-586, 2020.

Verhulst, P. F., *Notice sur la loi que la population poursuit dans son accroissement*. *Correspondance Mathématique et Physique. 10: 113–121*.

Wang, P., Zheng, X., Li, J., and Zhu, B., Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics, *Chaos, Solitons and Fractals*, vol. 139, pp. 1-7. 2020.

Wu, K., Darcet, D., Wang, Q., and Sornette, D., Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world, Nonlinear Dynamics, https://doi.org/10.1007/s11071-020-05862-6.

Wolfram, https://www.wolframcloud.com/obj/covid-19/Published/Logistic-Growth-Model-for-COVID-19.nb

World Health Organization, Coronavirus disease (COVID-19) Pandemic. https://www.who.int/ emergencies/diseases/ novel-coronavirus-2019, Online; accessed 26 August 2020.

Worldometer, https://www.worldometers.info/coronavirus/.

Zou, Y., Pan, S., Zhao, P., Han, L., Wang, X, Hemerik, L., Knops, J., and van der Werf, W., Outbreak analysis with a logistic growth model shows COVID-19 suppression dynamics in China, *PLoS ONE* 15(6): e0235247, 2020, https://doi.org/10.1371/journal.

Sofa, A. (2021). Things You Need To Know About Covid-19 Delta Variant. https://corona.jakarta.go.id/en/artikel/covid-19-varian-delta-dan-hal-hal-yang-harus-kamu-perhatikan (26/09/2021).

Biographies

Hennie Husniah is a Lecturer at Department of Industrial Engineering, Langlangbuana University. She received her PhD from Department of Industrial, Engineering, Institut Teknologi Bandung, Indonesia. Her research interests include product warranties, analysis of warranty data, maintenance and reliability models. Her works have been published in the following journals: International Journal of Collaborative Enterprise, Quality Technology and Quantitative Management, Computer and Industrial Engineering, Journal of the Indonesian Mathematical Society, International Journal of Applied Mathematics and Statistics,Journal of Information, Jurnal Teknik Industri, Jurnal Teknologi, Computer & Indusrtrial Engineering Journal and many international proceedings published by IEEE, AIP, IOP, etc.

Asep K. Supriatna is a full Professor in the Department of Mathematics, Padjadjaran University, Indonesia. Currently he holds a position as the head of the department. His primary research interest are in Mathematical Bioeconomics Modeling, Mathematical Epidemiology, and Industrial Mathematics. His educational background includes the following: Post Doctoral (Applied Mathematics, ITB & University of Twente, The Netherlands, 2006-2008); Doctoral (Applied Mathematics, University of Adelaide, Australia, 1998); Master (Mathematics, ITB, 1992); Bachelor (Mathematics, UNPAD, 1988). Numerous papers have been published in several scientific journals, including Bulletin of Mathematical Biology, Mathematical Biosciences, Communication in Biomathematical Sciences, Computational and Mathematical Methods in Medicine, etc**.**