Modified Verhulst Logistic Growth Model Applied to COVID-19 Data in Indonesia as One Example of Model Refinement in Teaching Mathematical Modeling

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

The Verhulst logistic function is among the most popular function to describe a growth phenomenon. Initially the theory is applied in studying the growth of living organism populations. However now it finds the applications in any growth phenomenon, including social, education, and engineering. There is a huge number of applications of the logistic function in various field. One of the strength of the model is its capability in estimating the carrying capacity or the maximum level of the growth. This upper bound is very important to obtain and has many practical implication. However, in some circumstances the function may fail to estimate this upper bound, especially when the growth is still at the beginning phase. In pedagogical context of mathematical modeling this failure is regarded as a good example in explaining the modeling process, in which when a model fails to comply with the reality, one should proceed to refine the model following the full cycle of modeling process. In this paper we present a modified growth model of the Verhulst logistic function, since when it is applied to the COVID-19 pandemic data in Indonesia, it cannot estimate the carrying capacity satisfactory. The modification has improved the estimation performance in terms of the root of the mean square error measure (RMSE).

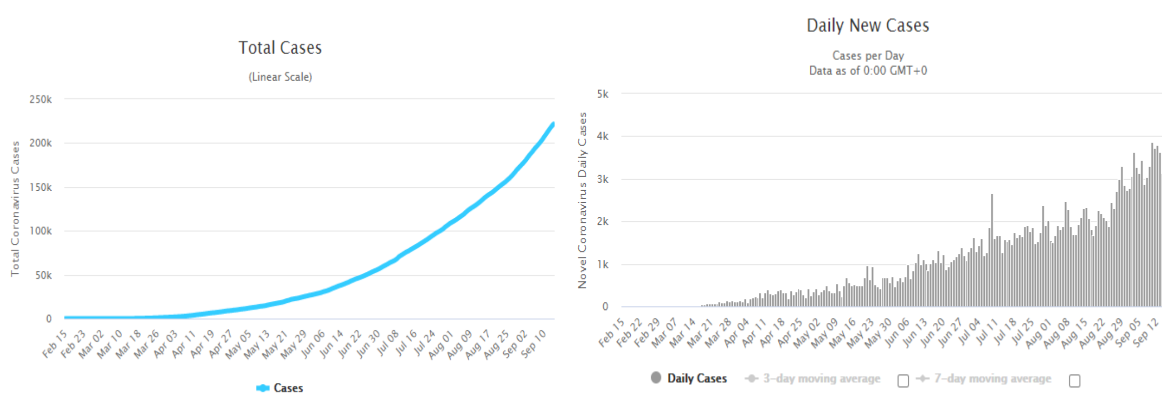
Keywords
COVID-19 Pandemic, Empirical Model, Indonesia, Verhulst Logistic Equation, Pedagogical Aspect.

1. Introduction
Teaching mathematical modeling is not only transfering the knowlegde of modeling concepts to students, but also should provide real experiences and full involvement in developing a reasonably good and realistic model. The system that is being modeled and the problem that is being addressed should also seen as a familiar system and problem. The appearence of COVID-19 as a pandemic can be seen as a good example. In this paper we will present a mathematical model that can be used as a good example in teaching how to refine a model to produce a better model according to Meyer (1984) criteria. Acccording to Meyer, a model can be called a good model if it has reasonably good in terms of accuracy, descriptive realism, precision, robustness, generality, and fruitfulness. The fulfillment to those criteria might be partially (it is not necessarilly to fulfill all the seven criteria). To give a feeling to the problem in COVID-19 pandemic, in the following we give a brief review of the disease transmission background.

COVID-19 appeared only about a year 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. 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). See also the following website address: https://www.hawaiinewsnow.com/video/2020/09/03/experts-states-confusing-covid-reporting-makes-it-difficult-understand-spread-covid-community/. This is among the reasons why so far most infected countries still struggling combating the disease. 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).

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.

Since it first appearence in Wuhan, the propagation in many countries still in increasing phase, including in Indonesia. Figure 1.a shows that so far 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 (Figure 1. b). This trend also holds for the cumulative confirmed cases worldwide (Anonymous 2020).

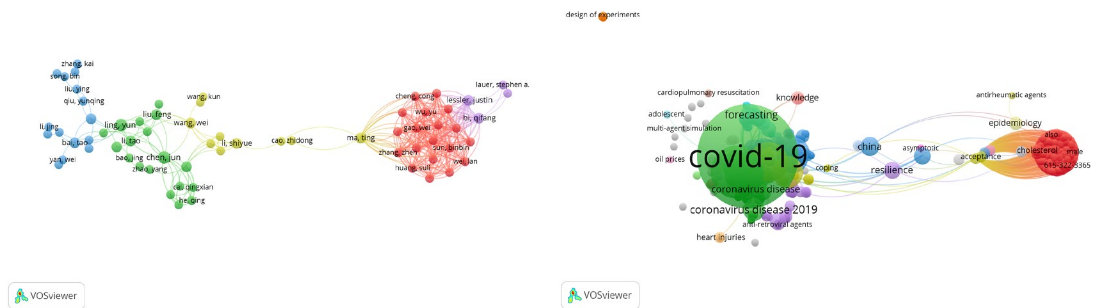


(a) (b)

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/

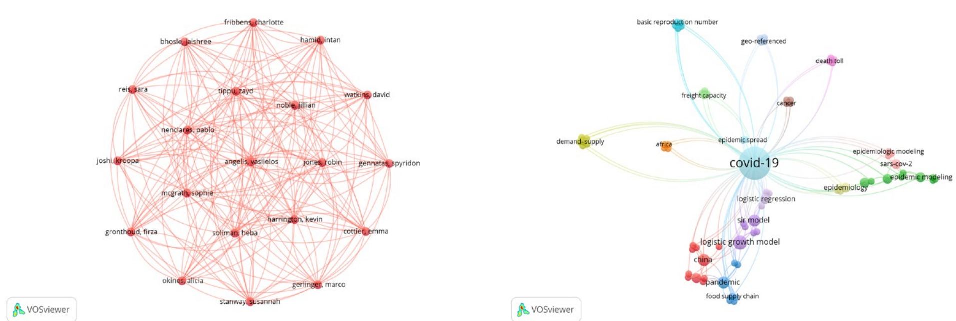
So far there are a lot of mathematical models have been developed to study the COVID-19 data, using different mathematical methods and approaches. Some using mechanical white box models and others using empirical black box models. Both have equally strengths and weaknesses, depending on the target of 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. Some implication regarding control action and intervention are also suggested in their works.

We undertook a brief bibliogrhaphic analyis using Publish or Perish software application (https://harzing.com/resources/publish-or-perish) by searching publication for the keywords "COVID-19" and "logistic" from Google Scholar database and Scopus data base, and found more than 1000 papers. But the keyword "logistic" might also refer to other concept, such as those in supply chain and economy. We select only the related ones and the results from VOSViewer visualization software application (https://www.vosviewer.com/) we obtain keywords relation (to other keywords) and authors relation in Figure 2 (Google Scholar data base) and Figure 3 (Scopus data base). The result is shown in Figure 2.a (Google Scholar data base) indicates that most collaborating authors bearing chinesse name. It does not mean that the institutions are in China. However result in Figure 3.a (Scopus data base) show that (when look at the authors' names) the collaborating authors are more diverse, coming from different countries. Figures 2.b and 3.b show other related keywords that appeared in the publication of the papers bearing those two keywords, "COVID-19" and "logistic". The figures show that the logistic growth function (or the modified- and generalized- logistic growth) have been applied in many context of COVID-19.



(a) (b)

Figure 2. VOSviewer representation of collaborating authors who published papers bearing the keywords "COVID-19" and "logistic" (a) and other keyword that appear in the papers (b). The metadata of publication are from Google Scholar data base, retrieved on 14 September 2020.



• (a) (b)

Figure 3. VOSviewer representation of collaborating authors who published papers bearing the keywords "COVID-19" and "logistic" (a) and other keyword that appear in the papers (b). The metadata of publication are from Scopus data base, retrieved on 14 September 2020.

The Verhulst logistic function is among the most popular equation to describe a growth phenomenon. Besides paper publication on logistic growth model, there are also software application that readily used by those who do not want to involve in too many mathematical analysis, such as in Wolframcloud (2020). There is a huge number of applications of the logistic equation in various field. One of the strength of the model is its capability in estimating the carrying capacity or the maximum level of the growth. This upper bound is very important to obtain and has many practical implication. However, in some circumstances the model may fail to estimate this upper bound, especially when the growth is at the beginning phase. We will show the case when logistic model is unable to give a satisfactory performance. We will also present a modified growth model of the Verhulst logistic equation, since when it is apply to the COVID-19 data pandemic in Indonesia, it cannot estimate the carrying capacity satisfactory. The modification has improve the estimation in terms of the root of the mean square error measure. The following section will review briefly the Verhulst logistic equation.

2. Method

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 classical Verhulst logistic model to fit the time series of the total confirmed cases. 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*. 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.

2.1 The Morgan-Mercer-Flodin Growth 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

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

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

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

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.

2.2 The Curve Fiting Approach

There are many modifications available for the classical Verhulst logistic function in literatures. In general the modification process is done through a new formulation such as described in Meyer's flowchat in Appendix 1. To be specific, here we give a flexibility for the equation to reach the inflection point by modifying equation (2) to the folowing form

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

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

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

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

3. 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). The results are presented in Tables 1 and 2, and Figures 4 and 5.

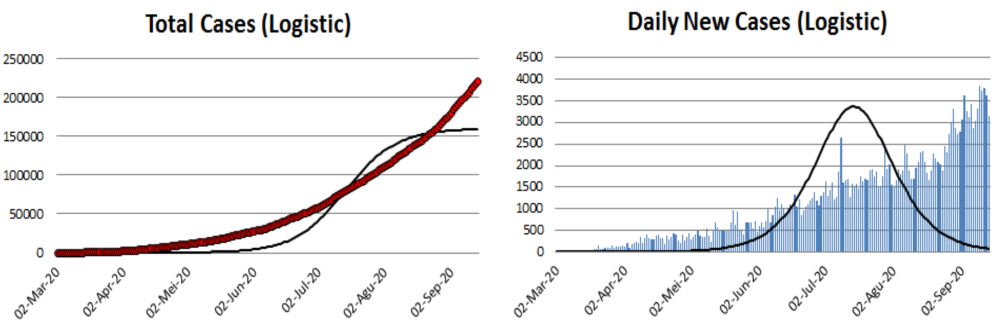
Table 1 shows that the modified Verhulst / modified logistic model is better in terms of the RMSE. In the modified logistic model, K is no more representing the carrying capacity. The carrying capacity is given by K^{α} (Table 2). The growth of the modified logistic model is slower than the original logistic model (0.02761 < 0.0840880). However it can reach a higher maximum growth (353,785 < 160,068). See also Figures 4.a and 5.a for comparison. In regards to daily new cases, the original logistic model does not fit to the actual data of COVID-19 in Indonesia (Figure 4.b). Meanwhile, the modified logistic model fits to most of the actual data of COVID-19 in Indonesia, except the later data (Figure 5.b). This suggests that we still need a further modification to the existing modified model. This is currently under investigation.

Table 1. Resulting values of parameters

Model	**Equation**	*K*	*r*	**α**	**RMSE**
Verhulst Logistic	(1)	160,068	0.0840880	-	18,221
Modified Verhulst Logistic	(3)	301.195	0.02761	2.23844	2,761

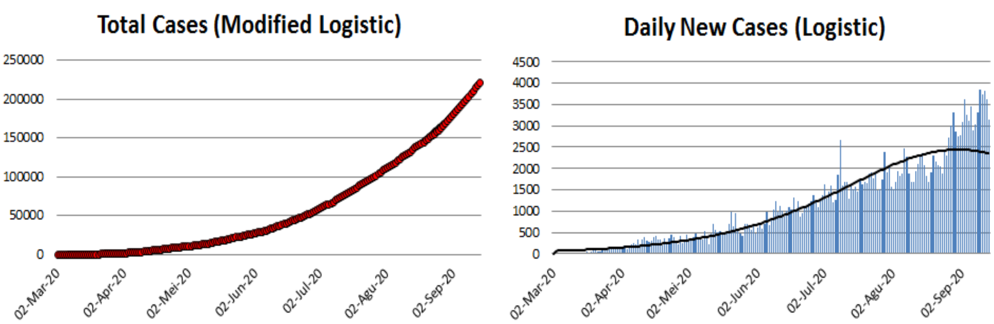
Table 2. Resulting prediction

Model	**Max Total Confirmed Cases**	**Inflection Time**	**Total Conf. Cases at Inflection**	**Peak New Cases**	**Peak Time**	**Total to Dec 2020**
Verhulst Logistic	K =160,068	134,267	80,034	3,364	Day 135	160,067
Modified Verhulst Logistic	K^{α} = 353,785	181,401	176,893	2,442	Day 182	342,189



(a) (b)

Figure 4. Total Coronavirus Cases (a) and Daily New Cases (b) in Indonesia pandemic data fitted by the logistic function. 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 show that the logistic function fails to estimate the carrying capacity *K*. The model shows that at this date the disease should have already reached the carrying capacity *K*, which is in fact untrue (a). The inaccuracy becomes apparent when we plot the daily new cases which clearly depart form the observed data (b).



(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. Figure 5.a shows that the modified logistic equation is able to estimate the carrying capacity K^{α} and produces curve that satisfactorily fits the data (compared to that in Figure 4.a). However if we look at the result for the daily new cases, the model shows that at this date the disease should also have already reached the carrying capacity *K*, which is in fact untrue (b).

4. 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: 1827 / UN6.3.1 / LT / 2020.

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

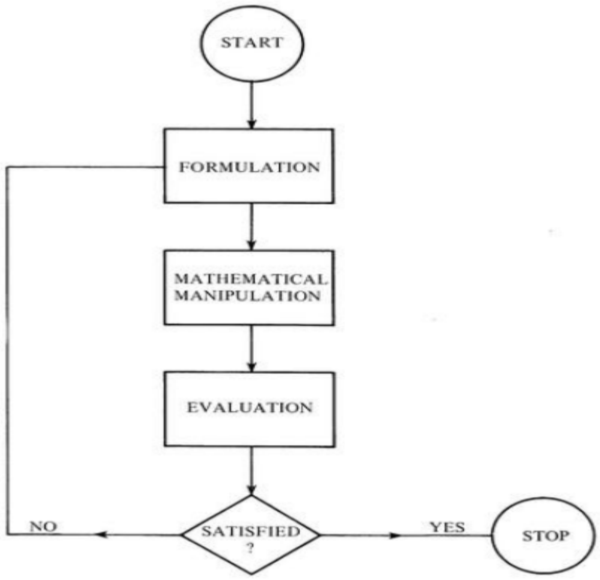
World Health Organization, Coronavirus disease (COVID-19) Pandemic. https://www.who.int/ emergencies/diseases/ novel-coronavirus-2019, Online; accessed 26 August 2020.

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.

Appendix

1. Flow Chart of Modeling Cycle According to Meyer (1984)



2. Inflection point of the modified logistic model

Let $X(t)=\dfrac{K^{\alpha}}{1+\dfrac{(K-X_0)e^{-rt}}{X_0}}$ be the modified logistic model, then we have the first derivative of the growth model is given by $X'(t)=\dfrac{K^{\alpha}Are^{-rt}}{\left(1+Ae^{-rt}\right)^2}$ with $A=\dfrac{K-X_0}{X_0}$.

Hence, the second derivative is given by $X''(t)=\dfrac{2K^{\alpha}A^2r^2\left(e^{-rt}\right)^2}{\left(1+Ae^{-rt}\right)^3}-\dfrac{K^{\alpha}Ar^2e^{-rt}}{\left(1+Ae^{-rt}\right)^2}$.

Solving the equation $X''(t)=0$ for t yields $t_i=\dfrac{\ln(A)}{r}$ and by substituting this critical value to the growth function will end up to $X(t_i)=\dfrac{K^{\alpha}}{2}$. The proof for the original logistic model is straight forward.

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.