# Predictive Mathematical Modelling of the Total Number of COVID-19 Cases for Brazil

Garba Uba[1], H. M., Yakasai[2], Abdussamad Abubakar[3] and Mohd Yunus Abd Shukor[4]*

[1]Department of Science Laboratory Technology, College of Science and Technology,
Jigawa State Polytechnic, Dutse, PMB 7040, Nigeria.
[2]Department of Biochemistry, College of Basic Medical science, Bayero University,
Kano, PMB 3001- Nigeria.
[3]Department of Microbiology, Faculty of Science, Bauchi State University Gadau,
P. M. B. 67 Itas Gadau, Bauchi State.
[4]Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia.

*Corresponding author:

Prof. Dr. Mohd Yunus Shukor,
Department of Biochemistry,
Faculty of Biotechnology and Biomolecular Sciences,
Universiti Putra Malaysia,
43400 UPM Serdang, Selangor,
Malaysia.
Email: mohdyunus@upm.edu.my

**ABSTRACT**

In this paper, we present different growth models such as Von Bertalanffy, Baranyi-Roberts, Morgan-Mercer-Flodin (MMF), modified Richards, modified Gompertz, modified Logistics and Huang in fitting and analyzing the epidemic trend of COVID-19 in the form of total number of infection cases of SARS-CoV-2 in Brazil as of 15th of July 2020. The MMF model was found to be the best model with the highest adjusted $R^2$ value with the lowest RMSE value. The Accuracy and Bias Factors values were close to unity (1.0). The parameters obtained from the MMF model include maximum growth rate (log) of 0.026    (95% CI from 0.024 to 0.028), curve constant ($\delta$) that affects the inflection point of 1.094 (95% CI from 1.024 to 1.165) and maximal total number of cases ($y_{max}$) of 66,527,316 (95% CI from 35,156,044 to 143,548,943). The MMF predicted that the total number of cases for Brazil on the coming 15th of August and 15th of September 2020 will be 2,993,850 (95% CI of 3,407,196 to 2,630,649) and 4,676,829 (95% CI of 5,553,936 to 3,938,240), respectively. The predictive ability of the model utilized in this study is a powerful tool for epidemiologist to monitor and assess the severity of COVID-19 in Brazil in months to come. However, as with any other model, these values need to be taken with caution due to the unpredictability of the COVID-19 situation locally and globally.

## INTRODUCTION

A novel corona virus, 2019-nCoV, was identified as the cause of a respiratory disease outbreak that originated in Wuhan, China, and spread to several other countries around the world [1]. As the global death toll of COVID-19 continues to rise, there is an increasing awareness that SARS-COV-2 mortality is inequitably distributed among vulnerable populations. Such vulnerable groups include the elderly, densely populated, low socio-economic status, refugees and minorities. All groups are vulnerable. The rates of these classes are higher, placing them at a high risk of infection and for serious disease consequences [2,3]. At first, modelling studies centered on the Chinese epidemic, and in particular on the epidemic dynamics in Wuhan City and in Hubei Province[4]. Much effort has been made at this early stage to evaluate surveillance data from China in order to obtain parameter estimates such as basic reproduction number ($R_0$), case fatality rate and incubation period [5]. For the first attempts at Susceptible-Exposed-Infectious-Recovered (SEIR) style dynamic models, parameter estimates were 'borrowed' from what was known about other corona viruses (SARS-CoV and MERS-CoV) and/or obtained by fitting the models to the monitoring data [6].

The growth curve of viruses and microorganisms on substrates such as nutrients or other organisms, including

humans, typically followed a sigmoidal pattern, beginning with the lag section just after t = 0, preceded by the logarithmic segment and then entering the stationary period and eventually heading into the period of death or decreasing growth. There are various sigmoidal functions to define the growth curve of organisms, such as Von Bertalanffy, Baranyi-Roberts, modified Richards, modified Gompertz and modified Logistics [7] including Morgan-Mercer-Flodin (MMF) [8]. The growth curve valuable parameters include the maximum specific growth rate ($\mu_m$), the lag period and the asymptotic values.

COVID-19 pandemic analyzes can be performed using statistical models, including theoretical, quantitative and simulation. For the analysis of the COVID-19 pandemic [9], strong predictive ability was employed models, such as updated Gompertz and Bertalanffy and logistics. The purpose of the paper is to test many available models like: "Logistic" (Ricker 1979[10]; [10], "Baranyi-Roberts" [11], "Von Bertalanffy"[12] ; [13] and "Margan" [14], "Baranyi Robert" [11]; [15] [11]" Morgan et al. and three-phase (Buchan 1993) recently Huang model [16] in fitting and analyzing the epidemic trend of COVID-19 in the form of total infection case of SARS-CoV-2 in Brazil as of 15th of July 2020.

## MATERIALS AND METHODS

Data were collected from Worldommeter (Worldometer 2020) for the total number of cases infected from Brazil as of 15 July 2020. Data were first converted to logarithmic values, and the time after first infected was utilized for time zero.

### Statistical analysis

Statistical significant **difference** between the models was calculated through various methods including the adjusted coefficient of determination ($R^2$), accuracy factor (AF), bias factor (BF), Root-Mean-Square Error (RMSE) and corrected AICc (Akaike Information Criterion) as before [17].
The RMSE was calculated according to Eq. (1), where $Pd_i$ are the values predicted by the model and $Ob_i$ are the experimental data, $n$ is the number of experimental data, and $p$ is the number of parameters of the assessed model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Pd_i - Ob_i)^2}{n-p}}$$
(eqn.1)

The adjusted $R^2$ is used to calculate the quality of nonlinear models according to the formula where RMS is Residual Mean Square and $S_y^2$ is the total variance of the y-variable ad calculated as follows;

$$Adjusted\ (R^2) = 1 - \frac{RMS}{s_Y^2}$$
(eqn. 2)

$$Adjusted\ (R^2) = 1 - \frac{(1-R^2)(n-1)}{(n-p-1)}$$
(eqn.3)

The Akaike information criterion (AIC) [18] was calculated as follows;

$$AICc = 2p + n\ln\left(\frac{RSS}{n}\right) + 2(p+1) + \frac{2(p+1)(p+2)}{n-p-2}$$
(eqn. 4)

Where $n$ is the number of data points and $p$ is the number of parameters of the model. The model with the smallest AICc value is highly likely correct [19].

Accuracy Factor (AF) and Bias Factor (BF) as suggested by Ross [20] were calculated as follows;

$$\text{Bias factor} = 10^{\left(\frac{\sum_{i=1}^{n}\log(Pd_i/Ob_i)}{n}\right)}$$
(eqn. 5)

$$\text{Accuracy factor} = 10^{\left(\frac{\sum_{i=1}^{n}\log|(Pd_i/Ob_i)|}{n}\right)}$$
(eqn. 6)

## FITTING OF THE DATA

Fitting of the bacterial growth curve using various growth models (**Table 1**) was carried out using GraphPad Prism (v 8.0 trial version).

**Table 1**. Models used in this study.

| Model | p | Equation |
|---|---|---|
| Modified Logistic | 3 | $y = \dfrac{A}{\left\{1+\exp\left[\frac{4\mu_m}{A}(\lambda-t)+2\right]\right\}}$ |
| Modified Gompertz | 3 | $y = A\exp\left\{-\exp\left[\frac{\mu_m e}{A}(\lambda-t)+1\right]\right\}$ |
| Modified Richards | 4 | $y = A\left\{1+v\exp(1+v)\exp\left[\frac{\mu_m}{A}(1+v)\left(1+\frac{1}{v}\right)(\lambda-t)\right]\right\}$ |
| Morgan-Mercer-Flodin (MMF) | 4 | $y = y_{max} - \dfrac{(y_{max}-\beta)}{1+(\mu_m t)^\delta}$ |
| Baranyi-Roberts | 4 | $y = A + \mu_m x + \frac{1}{\mu_m}\ln\left(e^{-\mu_m x}+e^{-h_0}-e^{-\mu_m x - h_0}\right) - \ln\left(1+\frac{e^{\mu_m x + \frac{1}{\mu_m}\ln\left(e^{-\mu_m x}+e^{-h_0}-e^{-\mu_m x-h_0}\right)}-1}{e^{(y_{max}-A)}}\right)$ |
| Von Bertalanffy | 3 | $y = K\left[1-\left[1-\left(\frac{A}{K}\right)^{\frac{1}{3}}\right]\exp^{-\left(\mu_m x/3K^{\frac{1}{3}}\right)}\right]^3$ |
| Huang | 4 | $y = A + y_{max} - \ln\left(e^A + \left(e^{Y_{max}} - e^A\right)e^{-\mu_m B(x)}\right)$ $B(x) = x + \frac{1}{\alpha}\ln\frac{1+e^{-\alpha(x-\lambda)}}{1+e^{\alpha\lambda}}$ |
| Buchanan Three-phase linear model | 3 | $Y = A$, IF $X < $ LAG $Y = A + K(X-\lambda)$, IF $\lambda \le X \ge X_{MAX}$ $Y = Y_{MAX}$, IF $X \ge X_{MAX}$ |

Note:
A= maximum no of cases lower asymptote;
$y_{max}$= maximum no of cases upper asymptote;
$\mu_m$= maximum specific growth rate;
v= affects near which asymptote maximum no of cases occurs.
$\lambda$=lag time
e = exponent (2.718281828)
t = time after first case is reported
$\alpha,\beta,\delta$ and k = curve fitting parameters
$h_0$ = a dimensionless parameter quantifying the initial physiological state of the reduction process. The lag time (h$^{-1}$) or (d$^{-1}$) can be calculated as $h_0 = \mu_m$
When data at time zero is 0 (Day after 1st case log 1=0 for COVID-19) the MMF is reduced to a 3-parameter model

## RESULTS AND DISCUSSION

The coefficient of these models are displayed in (**Fig 1 to 8**), their values determine the shape of the curves. All the curves tested show visually acceptable fitting with the exception of the Buchanan-3-phase model (**Figs 3**). The best performance was the MMF model (**fig 2**) with the lowest value for RMSE, AICc and the highest value for adjusted $R^2$. The AF and BF values were also excellent for the model with their values which were the closest to 1.0. The worst performance among them was seen in **fig 3** Buchnan-3- phase model (**Table 2**). This model proved to be unfit for data due to high residual variance and high nonlinearity values[21].The coefficients for the MMF model are shown in **Table 3**, a good agreement between experimental data and predicted values was obtained.

The fitness order to actual data according to $R^2$ values as well as MSE of all models explained the Total Number of COVID-19 Cases for Brazil in the following order; MMF> von Bertalanffy > modified Gompertz > modified Richards > Baranyi-Roberts > Huang > modified Logistics. Thus, MMF was determined to be the best fitted models and Buchanan-3-phase the worst[22].
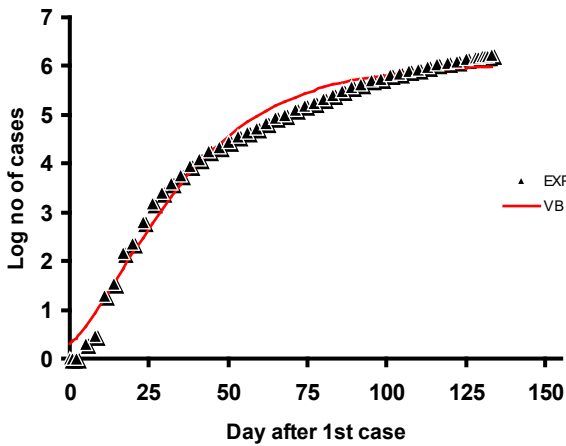


**Fig. 3**. Total no of SARS-CoV-2 cases in Brazil as of 15th of July 2020 as modelled using the Buchanan-3-phase model.



**Fig. 1**. Total number of SARS-CoV-2 cases in Brazil as of 15th of July 2020 as modelled using the von Bertalanffy model.



**Fig.4**. Total number of SARS-CoV-2 cases in Brazil as of 15th of July 2020 as modelled using the modified Gompertz model.
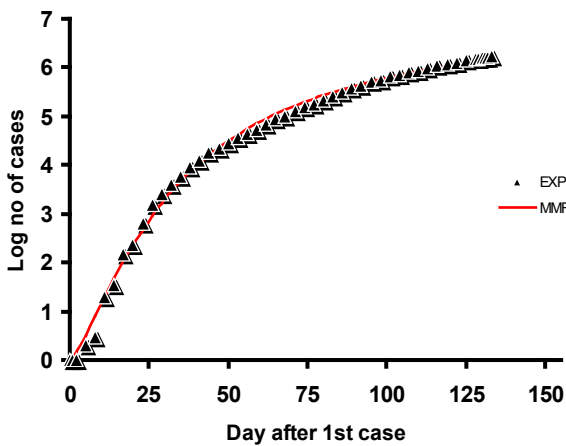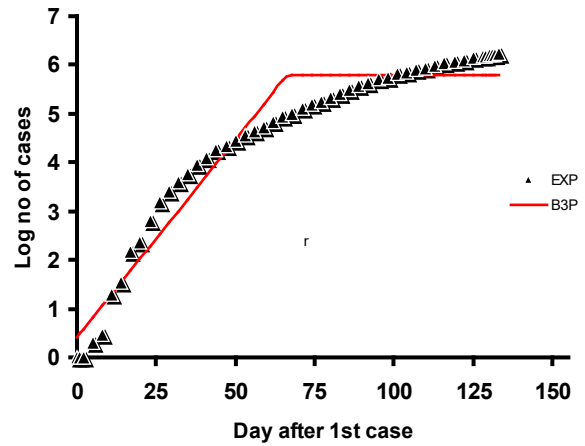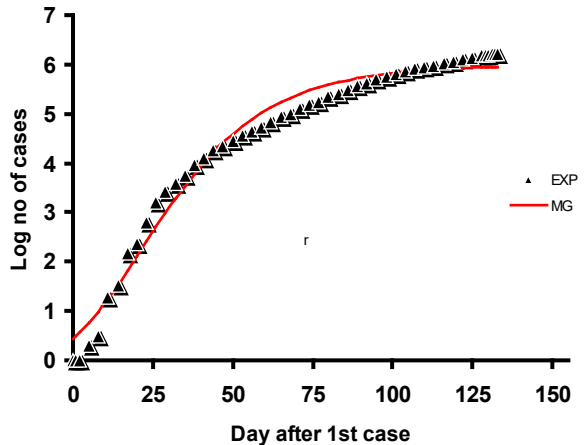


**Fig. 2**. Total number of SARS-CoV-2 cases in Brazil as of 15th of July 2020 as modelled using the MMF model.
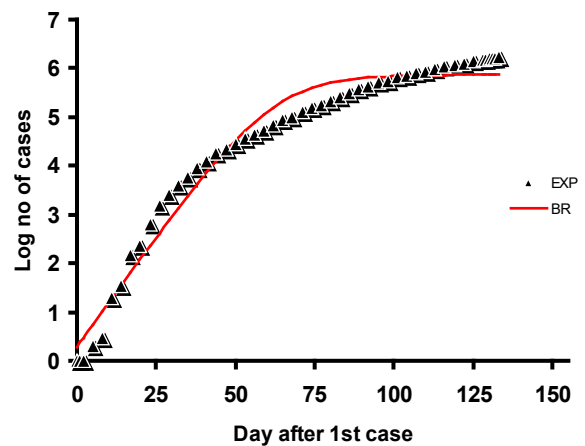


**Fig. 5**. Total number of SARS-CoV-2 cases in Brazil as of 15th of July 2020 as modelled using the Baranyi-Roberts model.
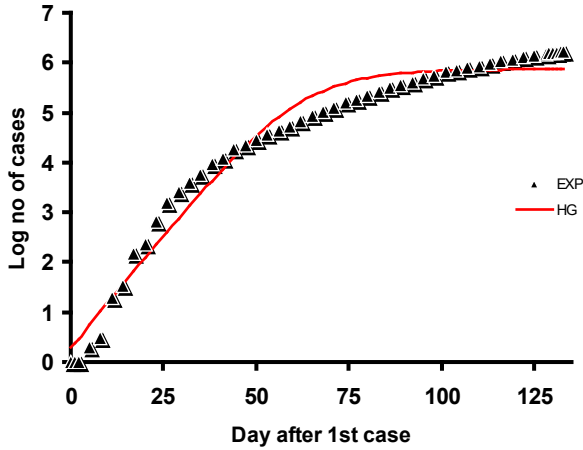
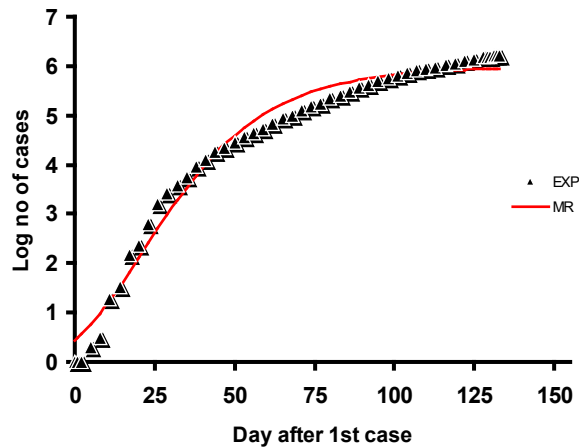**Fig. 6**. Total number of SARS-CoV-2 cases in Brazil as of 15th of July 2020 as modelled using the Huang model.



**Fig. 7**. Total number of SARS-CoV-2 cases in Brazil as of 15th of July 2020 as modelled using the modified Richard model.
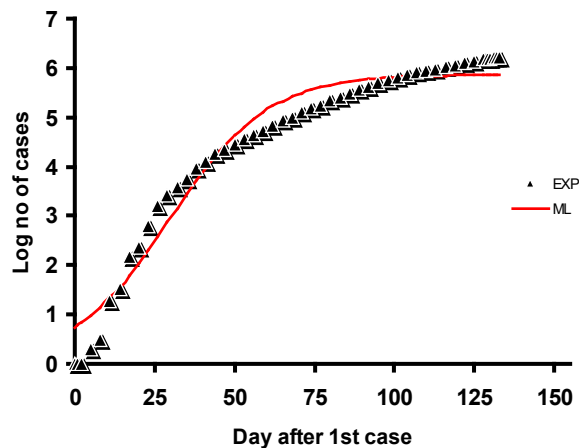


**Fig. 8**. Total number of SARS-CoV-2 cases in Brazil as of 15th of July 2020 as modelled using the modified logistics model.

**Table 2.** Statistical tests for the various models utilized in modelling the total no of SARS-CoV-2 cases in Brazil as of 15th of July 2020.

| Model | p | RMSE | $R^2$ | $adR^2$ | AF | BF | AICc |
|---|---|---|---|---|---|---|---|
| Huang | 4 | 0.330 | 0.970 | 0.968 | 1.046 | 0.99 | -95.73 |
| Baranyi-Roberts | 4 | 0.330 | 0.971 | 0.968 | 1.046 | 0.99 | -95.74 |
| modified Gompertz | 3 | 0.279 | 0.978 | 0.976 | 1.062 | 1.00 | -115.73 |
| Buchanan-3-phase | 3 | 0.433 | 0.948 | 0.945 | 1.062 | 0.99 | -72.01 |
| modified Richards | 4 | 0.282 | 0.978 | 0.976 | 1.034 | 1.00 | -111.25 |
| MMF | 3 | 0.116 | 0.996 | 0.996 | 1.013 | 1.00 | -203.74 |
| modified Logistics | 3 | 0.382 | 0.956 | 0.953 | 1.045 | 0.99 | -84.45 |
| von Bertalanffy | 3 | 0.232 | 0.985 | 0.984 | 1.029 | 1.00 | -134.16 |

Note: p is no of parameter

**Table 3**. Coefficients as modelled using the MMF model.

| Parameters | Value | 95% Confidence interval | |
|---|---|---|---|
| $\mu_m$ | 0.026 | 0.024 | to 0.028 |
| $\delta$ | 1.094 | 1.024 | to 1.165 |
| $y_{max}$ | 66,527,316 | 35,156,044 | to 143,548,943 |

**Table 4**. Predictions of COVID-19 pandemic for Brazil based on the MMF model.

| Prediction | Mean | 95% Confidence interval | |
|---|---|---|---|
| Maximum number of total cases by the end of COVID-19 | 66,527,316 | 35,156,044 to | 143,548,943 |
| Maximum number of total cases by 15th of August 2020 | 2,993,850 | 3,407,196 to | 2,630,649 |
| Maximum number of total cases by 15th of September 2020 | 4,676,829 | 5,553,936 to | 3,938,240 |

The parameters obtained from the MMF model include maximum growth rate (log) of 0.026 (95% CI from 0.024 to 0.028), curve constant ($\delta$) that affects the inflection point of 1.094 (95% CI from 1.024 to 1.165) and maximal total number of cases ($y_{max}$) of 66,527,316 (95% CI from 35,156,044 to 143,548,943). The MMF model predicted that COVID-19 will end about 720 days (95% CI of 498 to 942) days from 15th of July 2020 based on the lower bound of the 95% CI from the calculated maximum number of total cases ($y_{max}$) while the mean and upper 95% CI bound values failed to be predicted by the software for their number of days. The MMF predicted that the total number of cases for Brazil on the coming 15th of August and 15th of September 2020 will be 2,993,850 (95% CI of 3,407,196 to 2,630,649) and 4,676,829 (95% CI of 5,553,936 to 3,938,240), respectively. This prediction has to be taken with caution since the model failed to predict the number of days for the mean and upper 95% CI values and the number of days for COVID-19 to end may be much larger.

The MMF was initially developed as a model to describe nutrient-response relationships in a variety of organisms especially higher ones [8]. It has been utilized to model a variety of growth rates of animals including sheep, rabbit, horse and even microorganisms [23–27], palm oil's yield [28], ethanol [29] and even in economics [30]. If the predicted data is right, this will depend on case by case and includes lockdown effectiveness. Viral mutations that increase the infectivity rate of the virus will change the course of the model outcome. Probably, the models should be revisited every few months to restructure the data and have a stronger forecasting ability.

## CONCLUSION

In conclusion, the MMF model was the best model and first in modelling the number of total infected case of Covid-19 based on statistical tests such as corrected AICc (Akaike Information Criterion), bias factor (BF), adjusted coefficient of determination ($R^2$) and root-mean-square error (RMSE). Parameters obtained from the fitting exercise were maximum growth rate ($\mu_m$), the curve constants ($\delta$) and maximal total number of cases ($Y_{max}$). The model predicts the total number of cases in Brazil and varies according to different numbers of factors. Notwithstanding this, an important method for epidemiologists in this study to track and determine the extent of COVID-19 in Brazil in the upcoming months is the predictive ability of a sample.

## REFERENCE

1. Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. Mil Med Res [Internet]. 2020 Dec [cited 2020 Jul 23];7(1):11. Available from: https://mmrjournal.biomedcentral.com/articles/10.1186/s40779-020-00240-0

2. Shadmi E, Chen Y, Dourado I, Faran-Perach I, Furler J, Hangoma P, et al. Health equity and COVID-19: global perspectives. Int J Equity Health [Internet]. 2020 Jun 26 [cited 2020 Jul 23];19(1):104. Available from: https://doi.org/10.1186/s12939-020-01218-z

3. Dorn A van, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. The Lancet [Internet]. 2020 Apr [cited 2020 Jul 23];395(10232):1243–4. Available from: https://linkinghub.elsevier.com/retrieve/pii/S014067362030893X

4. Canada PHA of. Modelling scenarios of the epidemic of COVID-19 in Canada, CCDR 46(6) [Internet]. aem. 2020 [cited 2020 Jul 23]. Available from: https://www.canada.ca/en/public-health/services/reports-publications/canada-communicable-disease-report-ccdr/monthly-issue/2020-46/issue-6-june-4-2020/predictive-modelling-covid-19-canada.html

5. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med [Internet]. 2020 Apr [cited 2020 Jul 23];26(4):450–2. Available from: http://www.nature.com/articles/s41591-020-0820-9

6. Tuite AR, Fisman DN, Greer AL. Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. Can Med Assoc J [Internet]. 2020 May 11 [cited 2020 Jul 23];192(19):E497–505. Available from: http://www.cmaj.ca/lookup/doi/10.1503/cmaj.200476

7. Zwietering MH, Jongenburger I, Rombouts FM, Van't Riet K. Modeling of the bacterial growth curve. Appl Environ Microbiol. 1990;56(6):1875–81.

8. Morgan PH, Mercer LP, Flodin NW. General model for nutritional responses of higher organisms. Proc Natl Acad Sci. 1975 Nov 1;72(11):4327–31.

9. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature [Internet]. 2020 Jun [cited 2020 Jul 24];582(7812):389–94.Available from: http://www.nature.com/articles/s41586-020-2284-y

10. Zwietering MH, Rombouts FM, van 't Riet K. Modeling of the Bacterial Growth Curve. Appl Environ Microbiol [Internet]. 1990 [cited 2020 Jul 15];56(6):1875–81. Available from: https://AEM.asm.org/content/56/6/1875

11. Baranyi J, Roberts TA. Mathematics of predictive food microbiology. Int J Food Microbiol [Internet]. 1995 Jul [cited 2020 Jul 25];26(2):199–218. Available from: https://linkinghub.elsevier.com/retrieve/pii/016816059400121L

12. López S, Prieto M, Dijkstra J, Dhanoa MS, France J. Statistical evaluation of mathematical models for microbial growth. Int J Food Microbiol [Internet]. 2004 Nov [cited 2020 Jul 25];96(3):289–300. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168160504002041

13. Babák L, Šupinová P, Burdychová R. Growth models of Thermus aquaticus and Thermus scotoductus. Acta Univ Agric Silvic Mendel Brun [Internet]. 2012 [cited 2020 Jul 25];60(5):19–26. Available from: https://acta.mendelu.cz/60/5/0019/

14. Buchanan A. The re-emergence of correctional intervention. Crim Behav Ment Health [Internet]. 1993 Nov [cited 2020 Jul 25];3(4):527–8. Available from: http://doi.wiley.com/10.1002/cbm.1993.3.4.527

15. Loibel S, Andrade MG, Val JBR do, Freitas AR de. Richards growth model and viability indicators for populations subject to interventions. An Acad Bras Ciênc [Internet]. 2010 Dec [cited 2020 Jul 25];82(4):1107–26. Available from: http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0001-37652010000400028&lng=en&nrm=iso&tlng=en

16. Huang L. Optimization of a new mathematical model for bacterial growth. Food Control. 2013;32(1):283–8.

17. Halmi MIE, Shukor MS, Johari WLW, Shukor MY. Modeling the growth curves of Acinetobacter sp. strain DRY12 grown on diesel. J Environ Bioremediation Toxicol. 2014;2(1):33–7.

18. Akaike H. Factor analysis and AIC. Psychometrika. 1987;52(3):317–32.

19. Motulsky HJ, Ransnas LA. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. FASEB J Off Publ Fed Am Soc Exp Biol. 1987;1(5):365–74.

20. Ross T, McMeekin TA. Predictive microbiology. Int J Food Microbiol. 1994;23(3–4):241–64.

21. Nesetrilova H (Ceska ZU. Comparison of several growth models for cattle. Czech J Anim Sci - UZPI Czech Repub [Internet]. 2001 [cited 2020 Jul 24]; Available from: https://agris.fao.org/agrissearch/search.do?recordID=CZ2001001300

22. Topal M, Bolukbasi ŞC. Comparison of Nonlinear Growth Curve Models in Broiler Chickens. J Appl Anim Res [Internet]. 2008 Dec [cited 2020 Jul 24];34(2):149–52. Available from: http://www.tandfonline.com/doi/abs/10.1080/09712119.2008.9706960

23. Santos SA, Souza G da S e, Oliveira MR de, Sereno JR. Uso de modelos não-lineares para o ajuste de curvas de crescimento de cavalos pantaneiros. Pesqui Agropecuária Bras. 1999 Jul;34(7):1133–8.

24. Topal M, Bolukbasi ŞC. Comparison of nonlinear growth curve models in broiler chickens. J Appl Anim Res. 2008 Dec 1;34(2):149–52.

25. Tariq M, Iqbal F, Eyduran E, Bajwa M, Huma Z, Waheed A. Comparison of non-linear functions to describe the growth in Mengali sheep breed of Balochistan. Pak J Zool. 2013 Jun 1;45:661–5.

26. Augustine A, Imelda J, Paulraj R, David NS. Growth kinetic profiles of Aspergillus niger S14 a mangrove isolate and Aspergillus oryzae NCIM 1212 in solid state fermentation. Indian J Fish. 2015;62(3):100–6.

27. Kemper CM. Growth and development of the brush-tailed rabbit-rat (Conilurus penicillatus), a threatened tree-rat from northern Australia. Aust Mammal [Internet]. 2020 Jun 5 [cited 2020 Jul 20]; Available from: https://www.publish.csiro.au/am/AM19027

28. Khamis A, Ismail Z, Haron K, Mohammed AT. Nonlinear Growth Models for Modeling Oil Palm Yield Growth. J Math Stat. 2005 Sep 30;1(3):225–33.

29. Germec M, Turhan I. Ethanol production from acid-pretreated and detoxified tea processing waste and its modeling. Fuel. 2018 Nov 1;231:101–9.

30. Wijeratne AW, Karunaratne JA. Morgan-Mercer-Flodin model for long term trend analysis of currency exchange rates of some selected countries. Int J Bus Excell. 2013 Dec 2;7(1):76–87.