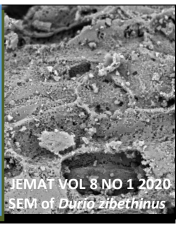
HIBISCUS PUBLISHER

JOURNAL OF ENVIRONMENTAL MICROBIOLOGY AND TOXICOLOGY

Website: https://journal.hibiscuspublisher.com/index.php/JEMAT



Predictive Mathematical Modelling of the Total Number of COVID-19 Cases for the Kingdom of Saudi Arabia

Abubakar Aisami[1] and Mohd Yunus Abd Shukor[2]*

[1]Department of Biochemistry, Gombe State University, P.M.B. 127, Tudun Wada, Gombe, Gombe State, Nigeria

[2]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

HISTORY

Received: 17th July 2020
Received in revised form: 25th July 2020
Accepted: 31st July 2020

KEYWORDS

SARS-CoV-2
COVID-19
Total number of cases
growth curve
MMF model

ABSTRACT

Throughout this article, we discuss different development models such as Von Bertalanffy, Baranyi-Roberts, Morgan-Mercer-Flodin (MMF), modified Richards, modified Gompertz, modified Logistics and Huang throughout fitting and evaluating the COVID-19 disease pattern in the context of the cumulative number of SARS-CoV-2 infection cases in the Kingdom of Saudi Arabia as of July 15th, 2020. The MMF model with the biggest adjusted R^2 and the lowest RMSE values established to be the best. The Accuracy, as well as the Bias Factors, are found to have values close to unity (1.0). The parameters generated from the MMF model include the maximum growth rate (log) of 0.03 (95% CI from 0.030 to 0.036), curve constant (δ) that affects the inflection point of 1.100 (95% CI from 1.029 to 1.171) and the highest possible number of cases (y_{max}) of 2,485,187 (95% CI from 1,468,970 to 4,204,419). The MMF model projected that COVID-19 will end in about 567 days (95% CI of 483 to 714) days from 15th of July 2020 centred on the lower bound of the 95% CI from the calculated maximum number of total cases (y_{max}). The MMF assumed that the total number of cases for the Kingdom of Saudi Arabia on 15 August and 15 September 2020 would be 384,258 (95 per cent CI from 368,567 to 400,618) and 508,412 (95 per cent CI from 482,797 to 535,387) respectively. The predictive ability of the model used in this study is an effective instrument for epidemiologists to track and assess the severity of COVID-19 in the Kingdom of Saudi Arabia over the coming months. Nevertheless, like any other model, due to the unpredictability of the COVID-19 situation locally and globally, these values must be taken with caution.

INTRODUCTION

INTRODUCTION

An evolving novel coronavirus first identified in 2019 in Wuhan, China triggered a massive spread of the disease known as coronavirus (COVID-19) and shortly after it was pronounced a pandemic [1]. On 2 March, a Saudi citizen from Iran via Bahrain screened COVID-19 positively and was subsequently identified and confirmed as the first case in the Kingdom of Saudi Arabia by the Ministry of Health [2].

The growth curve of viruses and microorganisms on a substrate surface such as nutrients and other organisms like humans usually preceded a sigmoidal pattern, beginning with the lag phase just after t = 0, preceded by the logarithmic phase and afterwards the organism enters the stationary period and eventually moves to the death or declining growth phase. There is numerous sigmoidal features curve, there are various sigmoidal functions such as Von Bertalanffy, Baranyi-Roberts, modified Richards, modified Gompertz and modified Logistics [3]) including Morgan-Mercer-Flodin (MMF) [4] that are used describe the organism growth. The useful parameters of the growth curve include the maximum defined growth rate (µm), the lag time and asymptotic values.

COVID-19 pandemic analysis can be conducted using mathematical models including theoretical, quantitative, and simulation. Models such as the modified Gompertz, von Bertalanffy and logistics have been used with the strong predictive ability to model the COVID-19 pandemic [5]. The main goal of this work is to evaluate several accessible models such as Logistic [3,6], Gompertz [3,7], Richards [3,8], Morgan-Mercer-Flodin (MMF) [4], Baranyi-Roberts [9], Von Bertalanffy [10,11], Buchanan three-phase [12] and more recently Huang model [13] in fitting and examining the epidemic pattern of the COVID-19 in the framework of the total infection case of SARS-CoV-2 in the Kingdom of Saudi Arabia as of 15th of July 2020.

MATERIALS AND METHODS

Data for the total number of infected cases from the Kingdom of Saudi Arabia as of 15th of July 2020 was acquired from Worldommeter [14]. Data were first converted to logarithmic values and the time after first infected were utilized for time zero.

Statistical analysis

The statistically 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 [15].

The RMSE was calculated according to Eq. (1), where Pd_i is 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.

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}\left(Pd_i-Ob_i\right)^2}{n-p}} \qquad \text{(Eqn. 1)}$$

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;

$$Adjusted\left(R^2\right)=1-\frac{RMS}{s_Y^2} \qquad \text{(Eqn. 2)}$$

$$Adjusted\left(R^2\right)=1-\frac{\left(1-R^2\right)\left(n-1\right)}{\left(n-p-1\right)} \qquad \text{(Eqn. 3)}$$

The Akaike information criterion (AIC) [16] was calculated as follows;

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

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 [17].

Accuracy Factor (AF) and Bias Factor (BF) as suggested by Ross [4] were calculated as follows;

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

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

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	$y=\frac{A}{\left\{1+\exp\left[\frac{4\mu_m}{A}(\lambda-t)+2\right]\right\}}$
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	$y=y_{max}-\frac{(y_{max}-\beta)}{1+(\mu_m t)^{\delta}}$
Baranyi-Roberts	4	$y=A+\mu_m x+\frac{1}{\mu_m}\ln\left(e^{-\mu_m x}+e^{-h_0}-e^{-}\right.$ $-\ln\left(1+\frac{e^{\mu_m x+\frac{1}{\mu_m}\ln\left(e^{-\mu_m x}+e^{-h_0}-e^{-\mu_m x-h_0}\right)}-1}{e^{\left(y_{\max}-A\right)}}\right)$
Von Bertalanffy	3	$y=K\left[1-\left[1-\left(\frac{A}{K}\right)^3\right]\exp^{-\left(\mu_m x/3K^{\frac{1}{3}}\right)}\right]^3$
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–λ), IF $\lambda \leq X \geq X_{MAX}$ Y = Y_{MAX}, IF $X \geq X_{MAX}$

Note:
A= maximum no of cases lower asymptote;
y_{max}= maximum no of cases upper asymptote;
μ_m= maximum specific growth rate;
v= affects near which asymptote maximum no of cases occurs.
λ=lag time
e = exponent (2.718281828)
t = time after first case is reported
α,β,δ 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=μ_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

All of the curves tested show visually acceptable fitting except the Buchanan-3-phase model (**Figs 1** to **6**). The best performance was the MMF model 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 were the closest to 1.0. The poorest performance was the modified logistics model (**Table 2**). The coefficients for the MMF model are shown in **Table 3**.

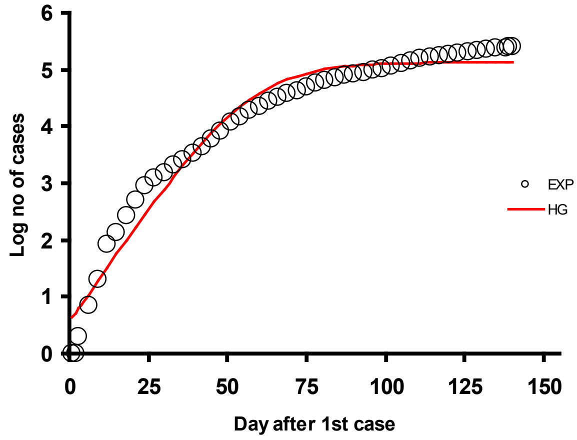


Fig. 1. Total no of SARS-CoV-2 cases in the Kingdom of Saudi Arabia as of 15th of July 2020 as modelled using the Huang model.

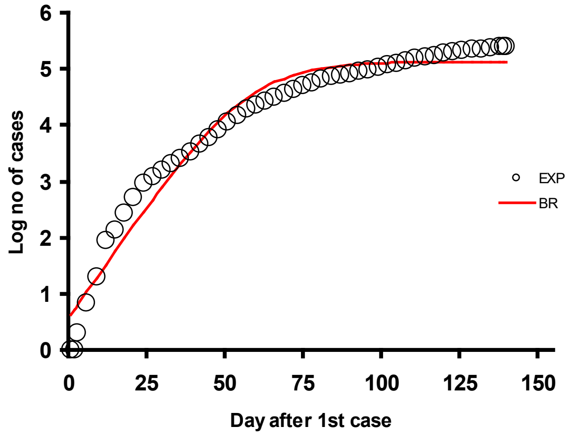


Fig. 2. Total no of SARS-CoV-2 cases in the Kingdom of Saudi Arabia as of 15th of July 2020 as modelled using the Baranyi-Roberts model.

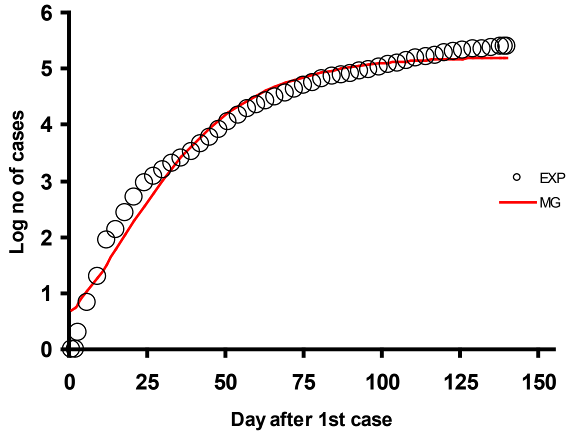


Fig. 3. Total no of SARS-CoV-2 cases in the Kingdom of Saudi Arabia as of 15th of July 2020 as modelled using the modified Gompertz model.

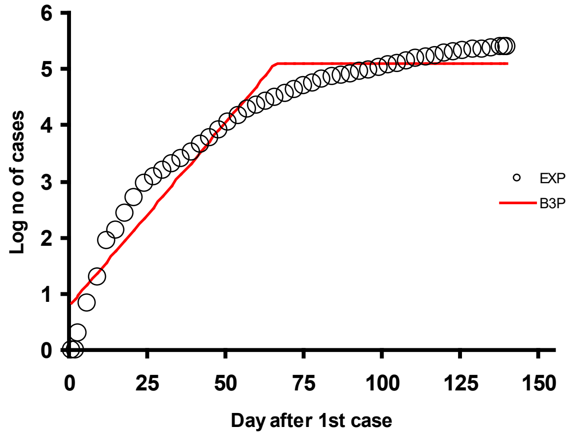


Fig. 4. Total no of SARS-CoV-2 cases in the Kingdom of Saudi Arabia as of 15th of July 2020 as modelled using the Buchanan-3-phase model.

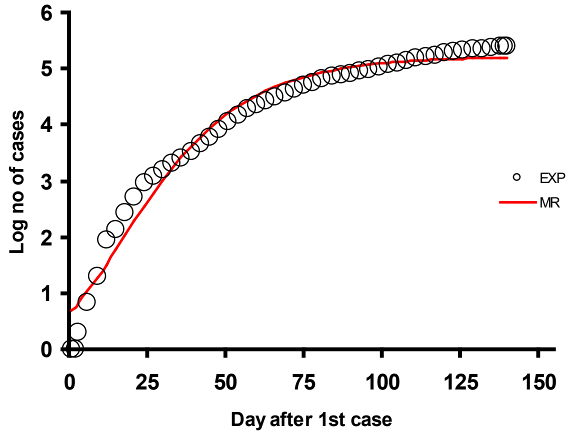


Fig. 5. Total no of SARS-CoV-2 cases in the Kingdom of Saudi Arabia as of 15th of July 2020 as modelled using the modified Richard model.

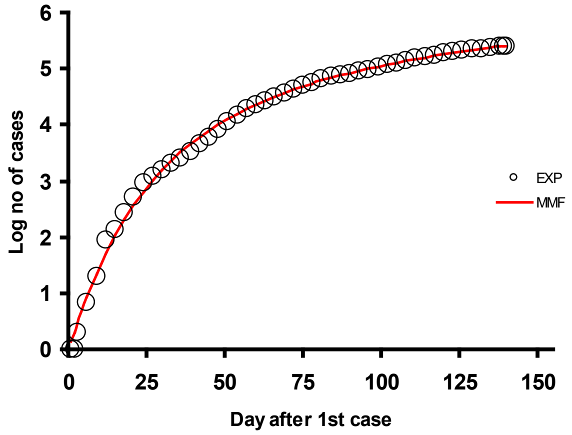


Fig. 6. Total no of SARS-CoV-2 cases in the Kingdom of Saudi Arabia as of 15th of July 2020 as modelled using the MMF model.

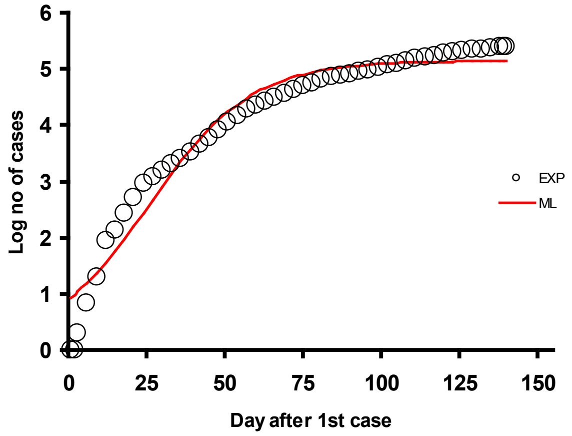


Fig. 7. Total no of SARS-CoV-2 cases in the Kingdom of Saudi Arabia as of 15th of July 2020 as modelled using the modified logistics model.

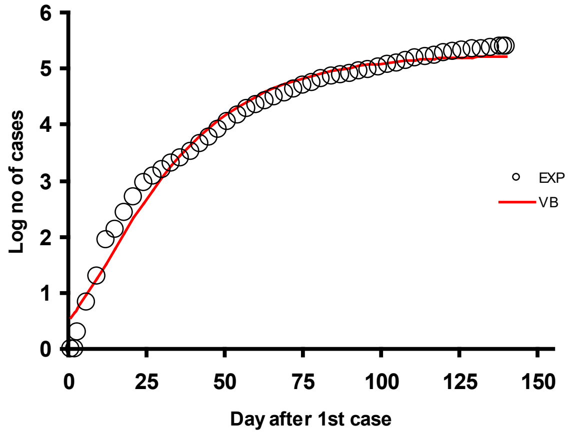


Fig. 8. Total no of SARS-CoV-2 cases in the Kingdom of Saudi Arabia as of 15th of July 2020 as modelled using the von Bertalanffy model.

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

Model	*p*	RMSE	R^2	adR^2	AF	BF	AICc
Huang	4	0.284	0.965	0.962	1.032	1.00	-110.75
Baranyi-Roberts	4	0.284	0.965	0.962	1.032	1.00	-110.75
modified Gompertz	3	0.252	0.971	0.969	1.045	1.00	-126.22
Buchanan-3-phase	3	0.371	0.939	0.935	1.045	1.00	-87.43
modified Richards	4	0.254	0.971	0.969	1.024	1.00	-121.74
MMF	3	0.087	0.997	0.997	1.006	1.00	-232.72
modified Logistics	3	0.321	0.951	0.948	1.030	1.00	-101.73
von Bertalanffy	3	0.219	0.979	0.977	1.021	1.00	-140.12

Note: p is no of parameter

Table 3. Coefficients as modelled using the MMF model.

Parameters	Value	95% Confidence interval
μ_m	0.03	0.030 to 0.036
δ	1.100	1.029 to 1.171
y_{max}	2,485,187	1,468,970 to 4,204,419

Table 4. Predictions of COVID-19 pandemic for The Kingdom of Saudi Arabia based on the MMF model.

Prediction	Mean	95% Confidence interval
Maximum number of total cases by the end of COVID-19	2,485,187	1,468,970 to 4,204,419
Maximum number of total cases by 15th of August 2020	384,258	368,567 to 400,618
Maximum number of total cases by 15th of September 2020	508,412	482,797 to 535,387

The parameters obtained from the MMF model include maximum growth rate (log) of 0.03 (95% CI from 0.030 to 0.036), curve constant (δ) that affects the inflection point of 1.100 (95% CI from 1.029 to 1.171) and the maximal total number of cases (ymax) of 2,485,187 (95% CI from 1,468,970 to 4,204,419). The MMF model predicted that COVID-19 will end about 567 days (95% CI of 483 to 714) days from 15th of July 2020 based on the lower bound of the 95% CI from the calculated maximum number of total cases (ymax) 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 The Kingdom of Saudi Arabia on the coming 15th of August and 15th of September 2020 will be 384,258 (95% CI of 368,567 to 400,618) and 508,412 (95% CI of 482,797 to 535,387), 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 model was initially established to define a wide variability of nutrient-response relations in higher organisms [4]. As of current, the model has found usefulness in numerous modelling exercises concerning animals such as rabbit, sheep, horse, microorganisms [18–22], the yield of oil palm [23], ethanol [24] and even in finance [25]. Whether the predicted data is correct or not will be contingent on a case by case basis and include the effectiveness of lockdown, mutation of the virus that increases the infectivity rate of the virus to name a few. Certainly, the models will be revisited every few months to remodel the data so a better prediction can be obtained. Moreover, after extensive Movement Control Order (MCO) steps, these data were modelled and little increase to no cases occurs at the end of the modelling era. except if the MCO is removed as expected in the coming weeks after the collected data, unless cases keep increasing then the findings of this modeling exercise will no longer be accurate and another modeling exercise will be carried out using data at the end of the modeling evaluation as the benchmark can be conducted.

CONCLUSION

In conclusion, the MMF model was the best in modelling the total number of SARS-CoV-2 cases in the Kingdom of Saudi Arabia 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 (μ_m), the curve constants (δ) and the maximal total number of cases (Y_{max}). The model allows for the prediction of a total number of cases and this prediction will vary according to the various number of factors. Despite this, 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 the Kingdom of Saudi Arabia in months to come

REFERENCES

1. World Health Organization. WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19; 2020 [Internet]. 2020 Mar. Available from: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020
2. MOH News. M.O.H. Reports First Case of Coronavirus Infection; 2020. 2020 Mar 2 [cited 2020 Apr 5]; Available from: https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2020-03-02-002.aspx
3. Zwietering M, Jongenburger I, Rombouts F, Van’t Riet K. Modeling of the bacterial growth curve. Appl Environ Microbiol. 1990; 56(6):1875–1881.
4. Morgan PH, Mercer LP, Flodin NW. General Model for nutritional responses of higher organisms. Proc Natl Acad Sci. 1975; 72(11):4327–4331.
5. Jia L, Li K, Jiang Y, Guo X, others. Prediction and analysis of Coronavirus Disease 2019. ArXiv Prepr ArXiv200305447 [Internet]. 2020; Available from: http://arxiv.org/abs/2003.05447
6. RICKER W. 11 growth rates and models. In: fish physiology. 1979. p. 677.
7. Gompertz B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. Philos Trans R Soc Lond. 1825 ;(115):513–583.
8. Richards F. A flexible growth function for empirical use. J Exp Bot. 1959; 10(2):290–301.
9. Baranyi J, Roberts TA. Mathematics of predictive food microbiology. Int J Food Microbiol. 1995; 26(2):199–218.
10. Babák L, Šupinová P, Burdychová R, others. Growth models of Thermus aquaticus and Thermus scotoductus. Acta Univ Agric Silvic Mendel Brun. 2013; 60(5):19–26.
11. López S, Prieto M, Dijkstra J, Dhanoa MS, France J. Statistical evaluation of mathematical models for microbial growth. Int J Food Microbiol. 2004; 96(3):289–300.
12. Buchanan RL. Predictive food microbiology. Trends Food Sci Technol. 1993; 4(1):6–11.
13. Huang L. Optimization of a new mathematical model for bacterial growth. Food Control. 2013; 32(1):283–288.
14. Worldometers. COVID-19 Coronavirus Pandemic. 2020 Jul 15; Available from: https://www.worldometers.info/coronavirus/#countries
15. Halmi, MI., Shukor M., Johari W.L.W WL. Shukor M. Mathematical Modeling of the Growth Kinetics of *Bacillus* sp. on Tannery Effluent Containing Chromate. J Environ Bioremediation Toxicol. 2014; 2(1):6–10.
16. Akaike H. Factor analysis and AIC. Psychometrika. 1987; 52:371–386.
17. Motulsky HJ, Ransnas LA. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. Fed Am Soc Exp Biol. 1987; 1(5):365–374.
18. Santos SA, Oliveira MR de, Sereno JR, others. Uso de modelos não-lineares para o ajuste de curvas de crescimento de cavalos Pantaneiros. Pesqui Agropecuária Bras. 1999; 34(7):1133–1138.
19. Topal M, Bolukbasi Ş. Comparison of nonlinear growth curve models in broiler chickens. J Appl Anim Res. 2008; 34(2):149–152.
20. Tariq MM, Iqbal F, Eyduran E, Bajwa MA, Huma ZE, Waheed A. Comparison of non-linear functions to describe the growth in Mengali sheep breed of Balochistan. Pak J Zool. 2013; 45(3):661–665.
21. 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):100–106.
22. Kemper CM. Growth and development of the brush-tailed rabbit-rat (Conilurus penicillatus), a threatened tree-rat from northern Australia. Aust Mammal. 2020;
23. Khamis A. Nonlinear growth models for modeling oil palm yield growth. J Math Stat. 2005; 1(3):225–233.
24. Germec M, Turhan I. Ethanol production from acid-pretreated and detoxified tea processing waste and its modeling. Fuel. 2018; 231:101–109.
25. Wijeratne A, Karunaratne J. Morgan-Mercer-Flodin model for long term trend analysis of currency exchange rates of some selected countries. Int J Bus Excell. 2014; 7(1):76–87.