Engineering and Applied Science Research 2021;48(4):432-445 Research Article

EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Predictive models for the number of cumulative cases for spreading coronavirus disease 2019 in the world

Rapin Sunthornwat*[1)] and Yupaporn Areepong[2)]

[1)]Industrial Technology Program, Faculty of Science and Technology, Pathumwan Institute of Technology, Bangkok 10330, Thailand
[2)]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

Received 27 July 2020
Revised 14 December 2020
Accepted 21 December 2020

Abstract

The coronavirus disease outbreak in 2019 (COVID-19) has caused major economic and healthcare problems worldwide. At this time, the worldwide outbreak has passed its peak, while the greatest number of cases has been in the USA, Brazil, and India. Measures and policies for controlling the outbreak have been developed by authorities to protect the population of each country, and forecasting the number of infectious people is an important factor for developing them. This research was conducted to identify a suitable forecasting model for estimating the cumulative daily number of infectious people worldwide. Sample countries with severe outbreaks were selected from each continent. Herein, forecasting models based on logistic, Richards, and Gompertz growth curves are derived and their suitability for forecasting the COVID-19 rates in each sample country and worldwide are analyzed. Moreover, estimating the growth curve parameters is based on the least-squares method. The results show that the Gompert growth curve is the most suitable for estimating the cumulative number of infectious people worldwide.

Keywords: Coronavirus disease 2019, Logistic growth curve, Richards growth curve, Gompertz growth curve, Least square estimation

1. Introduction

The coronavirus disease outbreak in 2019 (COVID-19) is caused by a new virus, Severe Acute Respiratory Syndrome–related Coronavirus-2 (SARS-CoV-2), which belongs to the same family of viruses as the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and SARS-CoV. It was first discovered in Wuhan, Hubei province, China, in December 2019, after which it rapidly spread worldwide. The most likely transmission route is from contact with a COVID-19 carrier via the aerosol effect from coughing and sneezing.

Most governments have recommended that people abide by the regulations from the World Health Organization for controlling the spread of COVID-19. The recommended fundamental ways for protection against COVID-19 infection are washing your hands with water or cleaning them with alcohol gel and wearing a mask when going outside of your home. Moreover, the authorities have combined advice on the preventive measures with providing knowledge about COVID-19 to the people. One of the factors for developing the measures is based on estimating the number of infectious people and depends on the current situation of the spread of the virus which is described by modeling the growth curve of confirmed infections in the population.

Thus far, many researchers have applied growth curves to analyze biological problems and study populations. A new efficient process design to help decision-making based on forecasting is by using S-growth curves. The number of daily COVID-19 cases has been explored by applying log-transformation (the inverse of an exponential function) [1], while the confirmed cases of COVID-19 have been explained by using sub-exponential growth [2]. These growth curves are unbounded since they are based on exponential functions, but exponential growth curves do not correspond realistically to the population size and have a carrying capacity of the maximum value for the environment. On the other hand, the logistic, Richards, and Gompertz growth curves are based on the carrying capacity of the environment. A non-linear model for estimating the number of COVID-19-infected people in Brazil based on the Chapman-Richard model has been evaluated [3]. Thus, the logistic growth curve has been applied to a forecasting system [4]. A growth curve for tumors has been estimated by applying the Gompertz model based on the whole population [5]. Analyzing population growth curves for birds and mammals has also been conducted, with logistic growth curves being fitted to population datasets based on Akaike's information criterion [6]. The fitted function for the yield of many kinds of rice has been modeled by applying the Richards growth curve (an asymmetric function), such as indices for the growth and seed vigor of rice plants [7].

Least-squares estimation of the parameters of logistic growth curves has been investigated for fitting durable customer goods data in China [8]. It has also been used to investigate weight-age data of male and female California turkeys by using linear modeling, and the logistic and Richards growth curves [9]; the results show that the logistic model was the most reliable.

*Corresponding author. Tel.: +66 2104 9099
Email address: rapin@pit.ac.th
doi: 10.14456/easr.2021.46

Forecasting models have an important role in aiding decision-making by the authorities for developing and supporting plans in the future. Especially, a forecasting model of the total number of COVID-19 cases is necessary to help develop measures for controlling the disease outbreak by measures such as lockdown during the outbreak. This research was conducted to investigate predictive models for analyzing the COVID-19 outbreak situation worldwide. The logistic, Richards, and Gompertz growth curves with least-squares estimation for their parameters are the fitting models applied to estimate the daily cumulative COVID-19 cases.

2. Materials and methods

Here, we cover the differential equations and statistical method as the background for this research.

2.1 Derivation of growth curves from differential equations

The logistic, Gompertz, and Richards growth curves for estimating the number of COVID-19 cases worldwide are based on solutions for differential equations.

Let L(t), R(t) and G(t) be logistic differential equation, Richards differential equation, and Gompertz differential equation, respectively. These differential equations were used for describing the growth of the number of population such as the number of living cells in population over time or bacterial growth, the number of cells in the tumor over time or cancer growth, and the number of the world population or world population growth.

Logistic differential equation [10] is written as

$$\frac{d}{dt}L(t) = lL(t)\left[1-\frac{L(t)}{C}\right]; L(t=0) = L_0 \quad (1)$$

where $L(t)$ is the number of cumulative COVID-19 infectious population by logistic growth curve at time *t*,
C is the carrying capacity or the maximum COVID-19 infectious population size as $t \to \infty$,
L_0 is the initial condition,
l is the intrinsic growth rate by logistic growth curve or growth rate per capita.

The solution of logistic differential equation is a logistic growth curve and it can be carried out as the followings.

$$\frac{d}{dt}L(t) = lL(t) - \frac{lL^2(t)}{C}; L(t=0) = L_0 \quad (2)$$

Based on Bernoulli differential equation [11], the transformation of variable is applied as $z = L^{-1}$. Then, Equation (2) becomes

$$\frac{dz}{dt} = -L^2\frac{dz}{dt}, -L^2\frac{dz}{dt} = lL - \frac{lL^2}{C}, \frac{dz}{dt} + lz = \frac{l}{C} \quad (3)$$

Equation (3) is linear differential equation. It can be solved by integrating factor $\exp(\int l dt) = \exp(lt)$

Multiplying integrating factor through both sides of Equation (3), the result is $\frac{d[z\exp(lt)]}{dt} = \frac{l}{C}\exp(lt)$, $z = \frac{1}{C} + A\exp(-lt)$; A is an arbitrary constant. $L^{-1} = \frac{1+AC\exp(-lt)}{C}$.

With initial condition $L(0) = L_0$, the result becomes

$$L(t) = \frac{C}{1+A\exp(-lt)}; A = \frac{C-L_0}{L_0} \quad (4)$$

Richards differential equation [12] is written as

$$\frac{d}{dt}R(t) = rR(t)\left[1-\left(\frac{R(t)}{C}\right)^{\beta}\right]; R(t=0) = R_0 \quad (5)$$

where $R(t)$ is the cumulative number of COVID-19 infectious population by Richards growth curve at time t ,
C is the carrying capacity or the maximum COVID-19 infectious population size as $t \to \infty$,
R_0 is the initial condition,
r is the intrinsic growth rate by Richards growth curve or growth rate per capita.
β is the shape parameter of Richards growth curve

The solution of Richards differential equation is Richards growth curve and it can be carried out as the followings.

$$\frac{dR(t)}{dt} - rR(t) = -\frac{r}{C^{\beta}}R^{\beta+1}(t); R(t=0) = R_0 \quad (6)$$

Equation (6) is form Bernoulli's differential equation, the solution can be carried out by transformation of Equation (6) to the linear differential equation as $v = R^{1-(\beta+1)} = R^{-\beta}$; $\frac{dv}{dt} = -\beta R^{-\beta-1}\frac{dR}{dt}$

Equation (6) becomes, $\frac{dv}{dt} + r\beta R^{-\beta} = \frac{r\beta}{C^{\beta}}$

$$\frac{dv}{dt} + r\beta v = \frac{r\beta}{C^{\beta}} \quad (7)$$

Multiplying both sides of Equation (7) by integrating factor $\exp(\int r\beta dt) = exp(r\beta t)$

$$\frac{d[v exp(r\beta)]}{dt} = \frac{r\beta}{C^{\beta}} \exp(r\beta t) \quad (8)$$

Taking integral both sides of Equation (8), the result becomes $\int d[v \exp(r\beta)] = \int \frac{r\beta}{C^{\beta}} exp(r\beta t)\, dt$; $v = \frac{1}{C^{\beta}} + A\, exp(-r\beta t)$; A is an arbitrary constant.

$$R^{-\beta}(t) = \frac{1}{C^{\beta}} + A exp(-r\beta t);\ R^{\beta}(t) = \frac{C^{\beta}}{1 + AC^{\beta}\exp(-r\beta t)}$$

Initial condition $R(0)=R_0$, the constant A is $A=\frac{C^{\beta}-R_0^{\beta}}{(CR_0)^{\beta}}$

The solution of Richards differential equation is

$$R(t)=\left[\frac{C^{\beta}}{1+Aexp(-r\beta t)}\right]^{\frac{1}{\beta}};\ A = \frac{C^{\beta}-R_0^{\beta}}{R_0^{\beta}} \quad (9)$$

Gompertz differential equation [13] is written as

$$\frac{d}{dt} G(t) = \mu_0\, exp(-gt)\, G(t);\ G(t=0) = G_0 \quad (10)$$

where $G(t)$ is the cumulative number of COVID-19 infectious population by Gompertz growth curve at time t,
μ_0 is an arbitrary constant from exponential growth,
G_0 is the initial condition,
g is the intrinsic growth rate by Gompertz growth curve or growth rate per capita.

The Gompertz growth curve as the solution of Gompertz differential equation can be carried out as the followings.

$$\int \frac{dG(t)}{G(t)} = \int \mu_0 \exp(-gt)\, dt;\ ln|G(t)| = -\mu_0 \frac{exp(-gt)}{g} + A;$$ A is an arbitrary constant.

With initial condition $G(0)=G_0$, the result becomes

$$ln|G_0| = \frac{-\mu_0}{g} + A$$

$$A = ln|G_0| + \frac{\mu_0}{g}$$

The solution of Gompertz differential equation is

$$ln\left|\frac{G}{G_0}\right| = \frac{\mu_0}{g}[1 - exp(-gt)] \quad (11)$$

Taking the exponential both sides of Equation

$$G(t)=G_0\exp(\frac{\mu_0}{g}\ [1 - \exp(-gt)]) \quad (12)$$

2.2 Analysis of the predictive models

The predictive models for estimating the cumulative daily number of COVID-19 infectious people for this research are based on the logistic, Richards, and Gompertz growth curves. The mathematical analysis of the growth curves is very important for predicting the maximum growth rate of COVID-19 infectious people.

The logistic growth curve is symmetric with respect to its inflection point, whereas the Gompertz growth is asymmetric. The Richards growth curve is symmetric with respect to its inflection point when $\beta = 1$ (under which condition it becomes a logistic growth curve) and asymmetric when $\beta \neq 1$.

The asymptotic behavior of growth curves is occurred as large time, $t \rightarrow \infty$. Logistic growth curve approaches the carrying capacity as large time.

$$L = \lim_{t\rightarrow\infty} \frac{C}{1+A\, exp(-lt)} = C$$

Richards growth curve also approaches the carrying capacity as large time.

$$R = \lim_{t\rightarrow\infty}\left[\frac{C^{\beta}}{1+\frac{C^{\beta}-R_0^{\beta}}{R_0^{\beta}}exp(-r\beta t)}\right]^{\frac{1}{\beta}} = C$$

Gompertz growth curve approaches $G_0\, exp\left(\frac{\mu_0}{g}\right)$ as large time.

$$G(t)= \lim_{t\to\infty} G_0 \exp\left(\frac{\mu_0}{g}[1-\exp(-gt)]\right) = G_0 exp\left(\frac{\mu_0}{g}\right)$$

The maximum growth rate of the number of COVID-19 infectious people is occurred at inflection point which the point for the second derivative equals to zero.

First of all, the definition of first derivative for any function $f(t)$ is given as

$$\frac{df}{dt} = \lim_{h\to o}\frac{f(t+h)-f(t)}{h} \approx \frac{f(t+h)-f(t-h)}{2h}$$

and the second derivative is given as $\frac{d^2f}{dt^2}$

The maximum growth rate is occurred at the inflection point or the peak time which carried out by the second derivative of growth curves.

For logistic growth curve,

$$\frac{d}{dt}L(t) = lL(t)\left[1-\frac{L(t)}{C}\right]; L(t=0) = L_0$$

$$\frac{d}{dt}L(t) = lL(t) - \frac{lL^2(t)}{C}; L(t=0) = L_0$$

$$\frac{d^2L(t)}{dt^2} = \frac{d}{dt}\left[lL(t) - \frac{lL^2(t)}{C}\right]$$

$$= l\frac{dL(t)}{dt} - \frac{2l}{C}L(t)\frac{dL(t)}{dt}$$

$$= l\frac{dL(t)}{dt}\left[1-\frac{2L(t)}{C}\right]$$

and set the second derivative equal to zero, $\frac{d^2L(t)}{dt^2} = 0$

$$l\frac{dL(t)}{dt}\left[1-\frac{2L(t)}{C}\right]=0$$

The maximum growth rate occurs of Logistic growth curve at $L(t)=\frac{C}{2}$ when $\frac{dL(t)}{dt} = 0$

For Richards growth curve,

$$\frac{dR(t)}{dt} - rR(t) = -\frac{r}{C^\beta}R^{\beta+1}(t); R(t=0) = R_0$$

$$\frac{dR(t)}{dt} = rR(t) - \frac{rR^{\beta+1}(t)}{C^\beta}; R(t=0) = R_0$$

$$\frac{d^2R(t)}{dt^2} = \frac{d}{dt}\left[rR(t) - \frac{rR^{\beta+1}(t)}{C^\beta}\right]$$

$$= r\frac{dR(t)}{dt} - \frac{r(\beta+1)R^\beta(t)}{C^\beta}\frac{dR(t)}{dt}$$

$$= r\frac{dR(t)}{dt}\left[1-\frac{(\beta+1)R^\beta(t)}{C^\beta}\right]$$

and set the second derivative equal to zero, $\frac{d^2R(t)}{dt^2} = 0$.

$$r\frac{dR(t)}{dt}\left[1-\frac{(\beta+1)R^\beta(t)}{C^\beta}\right]=0$$

The maximum growth rate occurs of Richards growth curve at $R(t)=\frac{C}{(\beta+1)^{\frac{1}{\beta}}}$ when $\frac{dR(t)}{dt} = 0$

For Gompertz growth curve,

$$\frac{d}{dt}G(t) = \mu_0 \exp(-gt)\, G(t); G(t=0) = G_0$$

$$\frac{dG(t)}{dt} = G_0\frac{\mu_0}{g}g\, exp\left(\frac{\mu_0}{g}[1-\exp(-gt)]\right)\exp(-gt)$$

$$\frac{d^2G(t)}{dt^2} = G_0\frac{\mu_0}{g}g\, exp\left(\frac{\mu_0}{g} - gt - \frac{\mu_0}{g}exp\left(-\frac{\mu_0}{g}t\right)\right)\left[\frac{\mu_0}{g} - gt - \frac{\mu_0}{g}\exp(-gt)\right]$$

$$= -G_0\frac{\mu_0}{g}g^2 exp\left(\frac{\mu_0}{g} - gt - \frac{\mu_0}{g}\exp(-gt)\right) + G_0\left(\frac{\mu_0}{g}\right)^2 g^2 exp\left(\frac{\mu_0}{g} - 2gt - \frac{\mu_0}{g}\exp(-gt)\right)$$

and set the second derivative equal to zero, $\frac{d^2G(t)}{dt^2} = 0$

$$t = \frac{\ln(\frac{\mu_0}{g})}{g}$$

Thus, the maximum growth rate occurs of Gompertz growth curve at $G(t) = G_0 exp\left(\frac{\mu_0}{g} - 1\right)$

2.3 Computation of the parameter estimates and the veracity of the predictive model

The unknown constants in the growth curves for estimation of total COVID-19 cases are called parameters. For logistic growth, there are three parameters for estimation, namely *C, l, L0* for Richards growth curve, there are four parameters for estimation, namely *C, r, R0*, β Similarly, there are three parameters of Gompertz growth curve for estimation, namely $\mu_{0,}\ g, G_0$. Let θ be a set of parameters of the growth curves. Define the sum square of error between the actual value and fitted value of the growth curves for estimation of total COVID-19 cases as

The unknown constants in the growth curves for estimating the total COVID-19 cases are treated as parameters. The logistic growth curve has three parameters to estimate *(C, l* and *L0)*, the Richards growth curve has four (*C, r, R0* and β), and the Gompertz growth curve has three ($\mu_{0,}\ g$ and G_0). Let θ be the set of parameters for the growth curves. The sum-of-squares of the error (SSE) between the actual and fitted values of the growth curves for estimating the total COVID-19 cases can be defined as

$$\text{SSE}=\sum_{i=1}^{n}\left(f(t_i)-\hat{f}(t_i)\right)^2$$

Where $f(t_i)$ is actual value of the number of cumulative COVID-19 cases at time t_i,
$\hat{f}(t_i)$ is actual value of the number of cumulative COVID-19 cases at time t_i,
n is the number of the number of cumulative COVID-19 cases.

The least-squares method for parameter estimation is applied to determine the minimum SSE as follows:

$$\min_{\theta}(SSE)=\min_{\theta}\left(\sum_{i=1}^{n}\left(f(t_i;\theta)-\hat{f}(t_i;\theta)\right)^2\right)$$

The extreme value by using the optimization method in multivariate calculus is applied to solve this problem. The partial derivative of the SSE with respect to each parameter to determine the critical points is derived as

$$\frac{\partial}{\theta}(SSE)=\frac{\partial}{\theta}\left(\sum_{i=1}^{n}\left(f(t_i;\theta)-\hat{f}(t_i;\theta)\right)^2\right)=0$$

The Levenberge-Marquardt Algorithm [14, 15] for solving the roots of this non-linear equation to determine the critical points of the SSE function. The solutions of this equation are critical points θ^* for testing the minimum value by using the second partial derivative with *m* parameters. Let *H* denote the Hessian matrix of the second partial derivative,

$$H=\begin{bmatrix}\frac{\partial^2 SSE}{\partial\theta_1^2} & \frac{\partial^2 SSE}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 SSE}{\partial\theta_1\partial\theta_m}\\ \frac{\partial^2 SSE}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 SSE}{\partial\theta_2^2} & \cdots & \frac{\partial^2 SSE}{\partial\theta_2\partial m}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial^2 SSE}{\partial\theta_m\partial\theta_1} & \frac{\partial^2 SSE}{\partial\theta_m\partial\theta_2} & \cdots & \frac{\partial^2 SSE}{\partial\theta_m^2}\end{bmatrix}$$

Let D_k be the determinant of the Hessian matrix in the variables $\theta_1,\theta_2,\theta_3,\dots,\theta_k$ for $k=1,2,3,\dots,m$. If $D_k(\theta^*)>0$ and $(-1)^k D_k(\theta^*)>0$ then *SSE* has a minimum at critical point.

The predictive models were validated by measuring their accuracy, precision, and trustworthiness to determine their veracity. The accuracy of the predictive model is based on the minimum error, which is the difference between the actual and fitted values. The precision of the predictive model is based on the maximum of the coefficient of determination, i.e. by how much the independent variable (time) can explain the dependent variable (the number of cumulative COVID-19 cases). Moreover, the trust of the predictive model is based on the confidence interval for the predicted values. Let *f(t)* be the actual value and $\hat{f}t$ be the fitted value.

The error is the value of the squared 2-norm of the estimate vector *X* and actual vector *Y* of cumulative COVID-19 cases. That is, $Error=\|Y-X\|^2$

Coefficient of Determination (R^2) $R^2=\frac{Explined\ Variance}{Total\ Variance}$

$$R^2=\frac{\sum_{i=1}^{n}(f(t_i)-\overline{\hat{f}(t_i)})^2}{\sum_{i=1}^{n}(\hat{f}(t_i)-\overline{\hat{f}(t_i)})^2}$$

Confidence Interval (CI) 100(1-α)% of *f(t)*

Upper Confidence Interval (UCI) = $\hat{f}(t)+t_{\frac{\alpha}{2},df}(SE)$

Lower Confidence Interval (LCI) = $\hat{f}(t)-t_{\frac{\alpha}{2},df}(SE)$

where SE is the standard error of *f(t)*,
df is the degree of freedom,
$1\text{-}\alpha$ is confidence level.

2.4 Data collection

The secondary data for this research are the number of daily cumulative COVID-19 cases reported on the website World meter [10] run by a group of scientific researchers who gather data and information for statistical evaluation worldwide. The number of cumulative COVID-19 cases recorded and reported daily from February 15, 2020 (*t=0*), to June 19, 2020 (*t=125*), a total of 126 days divided into 122 days for the within-sample and 4 days for the out-sample, were used in the study. The sampling technique was classified as the most severely affected country by continent. The data are summarized in Table 1.

Table 1 Sampling plan for data collection

Continent	Countries
Africa	South Africa, Egypt, Nigeria, Ghana
Asia	India, Iran, Turkey, Pakistan
Europe	Russia, UK, Spain, Italy
North America	USA, Mexico, Canada, Dominican Republic
Australia	Australia, New Zealand, French Polynesia, New Caledonia
South America	Brazil, Peru, Chile, Colombia

3. Results

Here, we report the output for this research, namely the estimated parameters and forecasting models concerning the spread of COVID-19 for each country and worldwide.

3.1 Estimated parameters and forecasting models

The parameters for the logistic, Richards, and Gompertz growth curve models are reported in Table 2. The most suitable forecasting models for the daily cumulative COVID-19 cases for Africa, Asia, Europe, North America, Australia, and South America continent were selected based on the minimum squared 2-norm and the maximum coefficient of determination values, as shown in Figures 1 to 6, respectively. Moreover, Figure 7 shows the most suitable forecasting model of the daily cumulative COVID-19 cases worldwide.

Table 2 Estimated parameters and suitable model

Countries	Logistic growth curve $L_0/l/C$ $R^2_{Logis}/Error_{Logis}$	Richards growth curve $R_0/r/C/\beta$ $R^2_{Richards}/Error_{Richards}$	Gompertz growth curve $L_0/\mu_0/g$ $R^2_{Gomp}/Error_{Gomp}$
South Africa	82.9264/0.0567/1000378.9285 0.9997*/1.0139e+07*	18.6676/ 0.0739/147066.5511/1.1082 0.9951/3.1811e+08	1.0024/0.1828/0.0162 0.9503/2.7821e+10
Egypt	95.1858/0.0539/163096.7018 0.9512/1.1863e+07	4.3449/ 0.0845/ 92577.9489/ 1.1406 0.9993*/3.3071e+04*	38.1241/ 0.0803/ 0.0054 0.9632/1.1348e+07
Nigeria	13.2337/0.0688/22297.0021 0.9865/9.6259e+06	0.1594/0.1041/33316.0521/0.7678 0.9643/1.2399e+08	0.1019/0.2491/0.0186 0.9978*/7.6956e+06*
Ghana	10.9309/ 0.0731/ 12805.2107 0.9783/1.2704e+07	0.0201/ 0.1205/ 23928.0542/ 0.5831 0.9651/1.0620e+08	0.0042/ 0.3981/ 0.0257 0.9971*/9.6393e+06*
India	304.8635/0.0641/654325.7546 0.9217/4.4792e+008	197.37091/0.0684/379751.7446/1.2711 0.9531/2.9731e+11	7.4312/0.1781/0.0132 0.9615*/8.4167e+007*
Iran	6803.0241/0.0428/204517.9913 0.9794*/ 9.3585e+09*	197.3709/ 0.0685/ 379751.7446/ 1.2711 0.9657/2.9731e+11	10.0002/ 0.1782/ 0.0282 0.9564/1.0436e+12
Turkey	257.5967/0.0995/166068.4316 0.9944*/3.6300e+09*	437.2818/0.0632/359661.5833/1.5288 0.97861/4.1805e+11	10.0002/0.1788/0.0182 0.9621/7.8181e+11
Pakistan	422.7768/0.0492/1095243.4580 0.9979*/3.7735e+008*	21.0881/0.0801/288956.1378/1.1688 0.9801/1.0290e+010	10.0002/0.1767/0.0199 0.9735/1.1477e+011
Russia	277.8570/ 0.0822/ 559568.7415 0.9978*/9.0662e+09*	288.6071/0.0729/1074421.3369/1.5029 0.9764/9.6665e+11	10.0003/ 0.1827/0.0201 0.9031/5.0509e+12
UK	997.1581/0.0808/291095.8566 0.9974/4.7665e+09*	0.9053/0.1216/593715.5694/0.2813 0.9673/9.8859e+10	9.0002/0.1726/0.01938 0.8961/2.8831e+12
Spain	1684.3488/0.0944/284296.1137 0.9951*/9.6756e+09*	114.4965/0.0831/582377.8026/1.2787 0.8794/1.9823e+12	2.0010/0.1798/0.0120 0.8689/2.8904e+12
Italy	2690.5934/0.0911/230607.7172 0.9962*/4.7064e+009*	170.5765/0.0780/474579.8350/1.3655 0.8711/3.0093e+012	3.0004/0.1707/0.0166 0.8963/1.5386e+012
USA	15/0.0812/4115676 0.7781/1.1770e+014	7.1606/0.1117/4492267.8165/0.3023 0.9608*/2.8601e+012*	15.0001/0.1701/0.0389 0.8961/1.2933e+014
Mexico	141.2758/0.0653/231580.3867 0.9993*/1.6280e+008*	39.3668/0.0763/293673.7942/1.2555 0.8912/1.5898e+010	10.0002/0.1782/0.0234 0.7913/2.3305e+011
Canada	400.0502/0.0752/98877.6065 0.9951/4.8434e+008	12.0101/0.0899/198303.4810/0.4938 0.8973/1.3101e+010	0.0606/0.5449/0.0376 0.9977*/4.5830e+008*
Dominican Republic	132.7059/0.0559/27203.3057 0.8971/2.6256e+07	48.1301/0.0607/46541.9967/1.3782 0.7895/1.4428e+09	3.3146/0.2076/0.0218 0.9989*/6.9086e+06*
Australia	1.6378/0.1996/ 6969.1371 0.9959* /4.6228e+06*	207.4113/ 0.0422/ 14670.2237/0.5586 0.8961/2.9161e+08	0.0014/ 1.0049/0.0647 0.9751/4.5902e+07
New Zealand	0.0769/ 0.2132/ 1491.0701 0.9990* /5.5732e+04*	62.5261/ 0.0406/ 3007.9987/ 1.7104 0.8921/5.0640e+07	0.0021/ 0.7381/ 0.0543 0.6252/3.6505e+06
French Polynesia	0.0739/ 0.1564/ 59.3560 0.9952* /411.0837*	0.0166/ 0.0906/ 119.9991/ 0.2711 0.8721/1.9969e+04	0.0012/ 0.6531/ 0.0599 0.8836/1.9233e+03
New Caledonia	2.4662/ 0.4177/ 18.4963 0.9876* /104.4317*	34.2205/ 0.0018/ 41.9992/ 10.2092 0.692/2.8332e+03	3.4778/ 1.1717/ 0.0753 0.9741 /526.8915
Brazil	417.0724/0.0712/1481495.9642 0.9998/1.6953e+09*	1.1063/0.1191/1946286.4472/0.6221 0.8971/6.0974e+10	10.0003/0.1846/0.0286 0.7618/1.0336e+13
Peru	1200.2618/0.0704/299809.5313 0.9129/7.8941e+008	55.8899/0.0784/465983.5125/1.0879 0.8201/4.4420e+010	0.0802/0.3569/0.0223 0.9991*/5.7994e+008*
Chile	128.5940/0.0673/397059.1560 0.9993*/2.6878e+008*	32.4794/0.0801/420011.8291/1.1151 0.9782/1.1053e+010	10/0.1605/0.0173 0.9631/3.4767e+011
Colombia	109.5881/0.0550/136227.2525 0.9412/1.1462e+007	20.9606/ 0.0723/106125.9104/ 1.2606 0.8629/1.4362e+009	26.1497/ 0.0977/ 0.0079 0.9997*/7.7138e+006*
The World	138559.7559/0.0446/10990215.0535 0.9861/6.7886e+012	66793.6234/0.0459/16252332.6872/0.75059 0.9217/1.1567e+013	18365.6829/0.1317/0.0197 0.9976*/1.8539e+012*

Note: * Optimal value

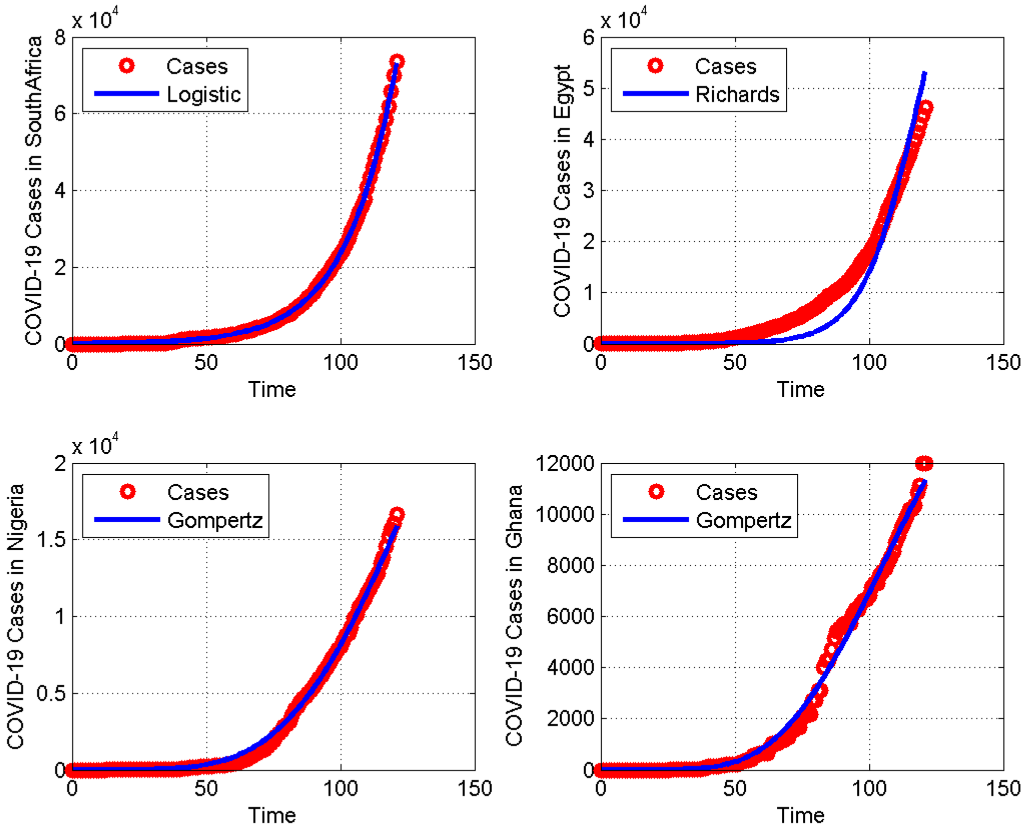

Figure 1 Estimate the cumulative number of COVID-19 cases in Africa

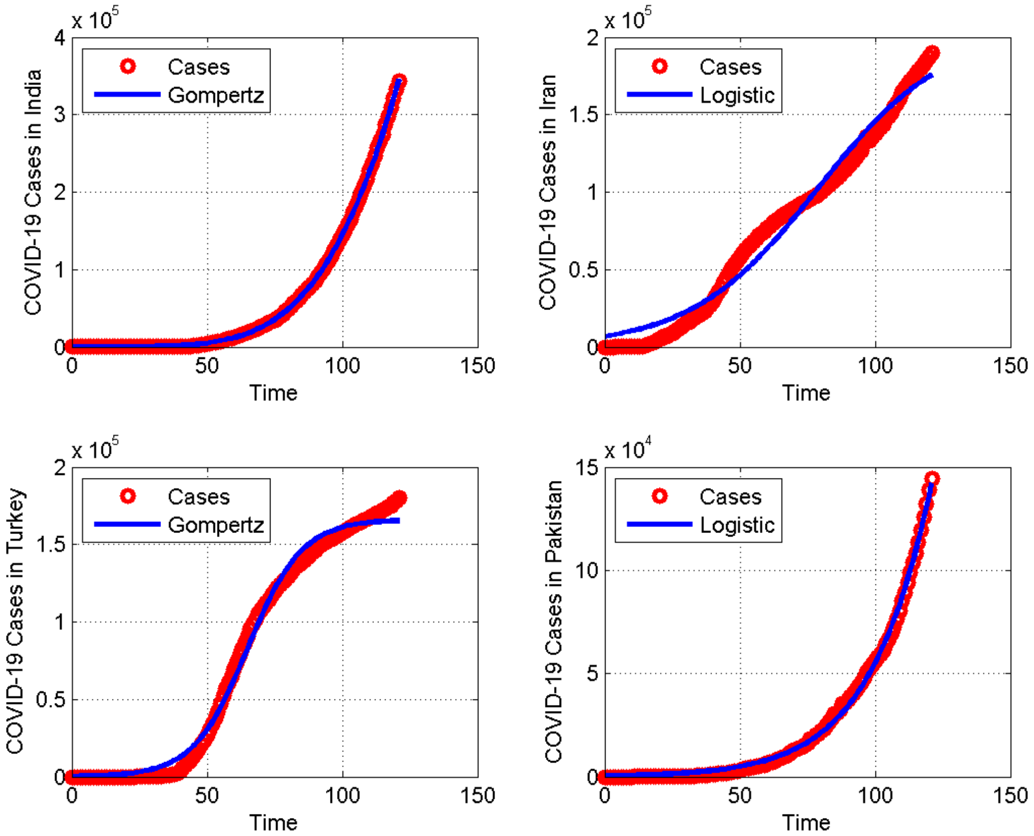

Figure 2 Estimate the cumulative number of COVID-19 cases in Asia

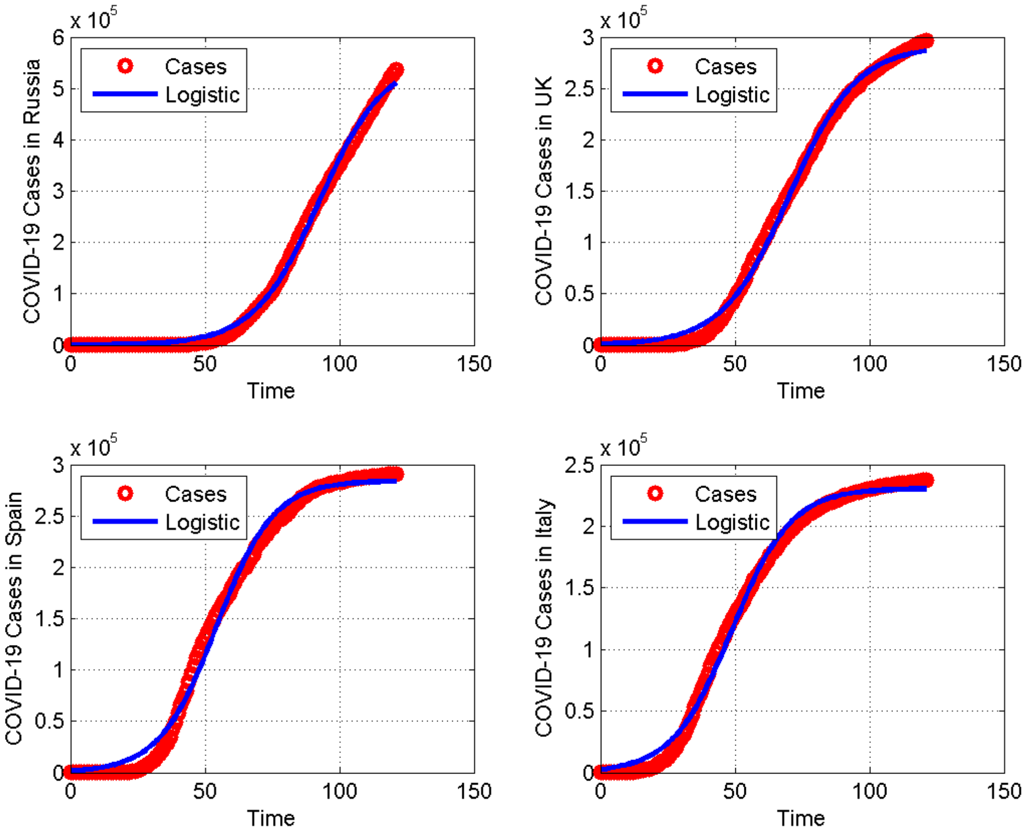

Figure 3 Estimate the cumulative number of COVID-19 cases in Europe

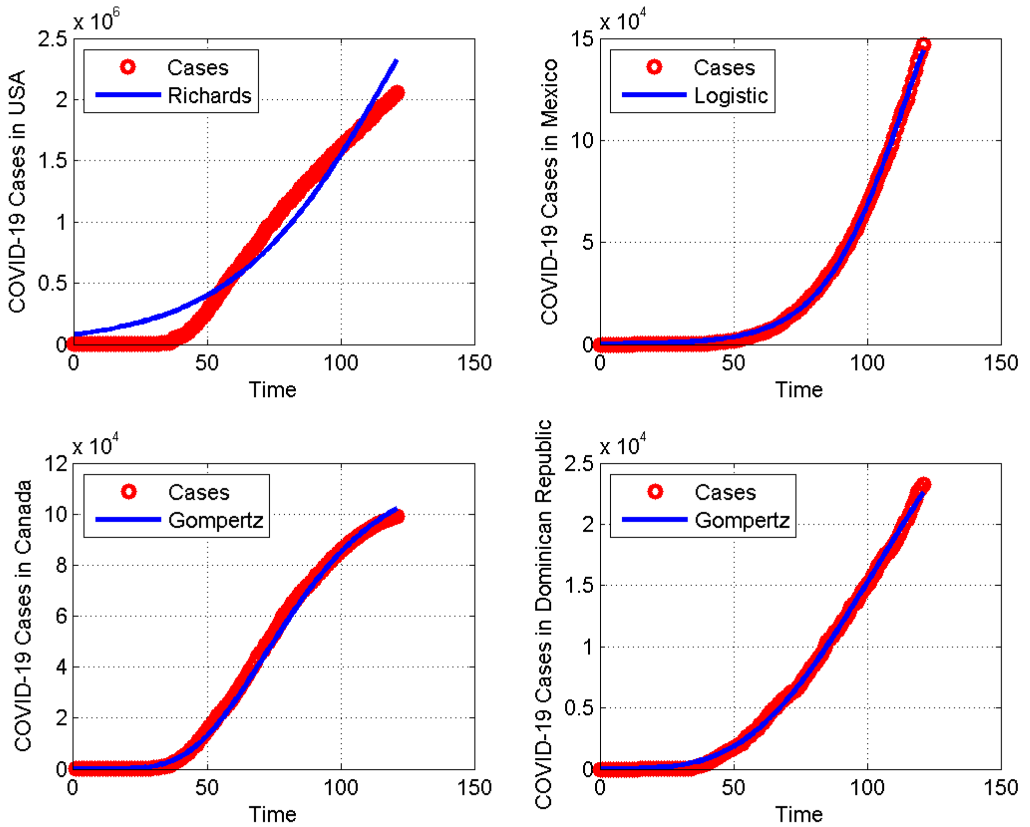

Figure 4 Estimate the cumulative number of COVID-19 cases in North America

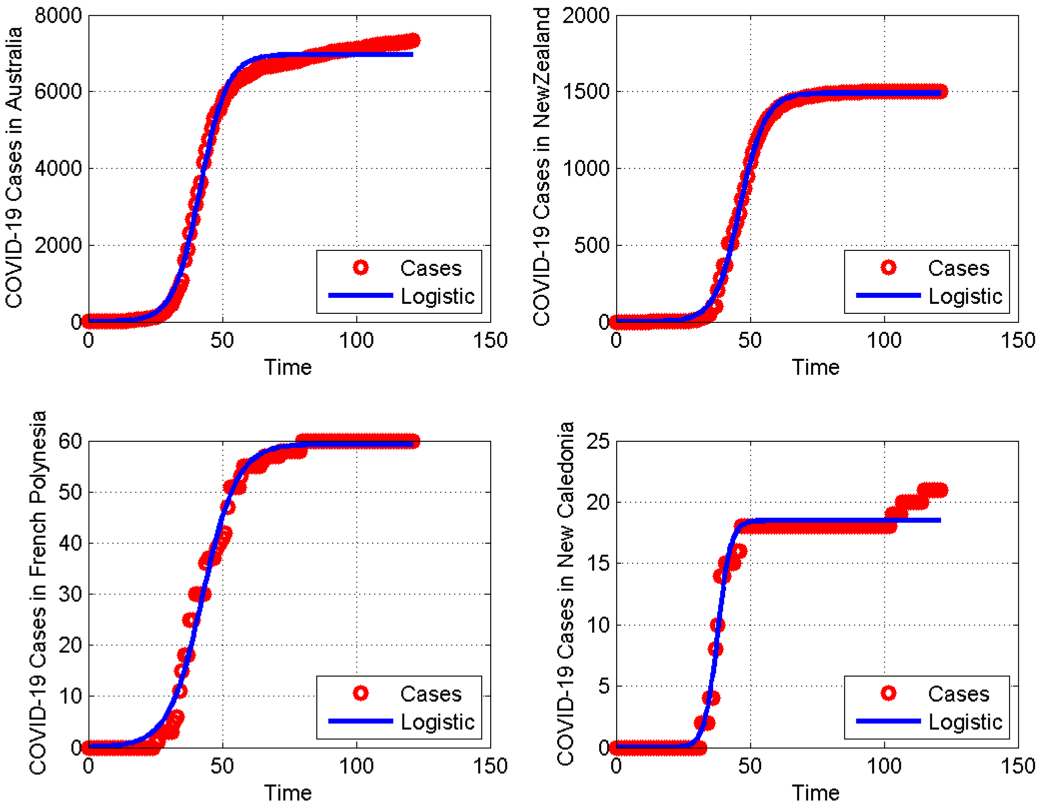

Figure 5 Estimate the cumulative number of COVID-19 cases in Australia

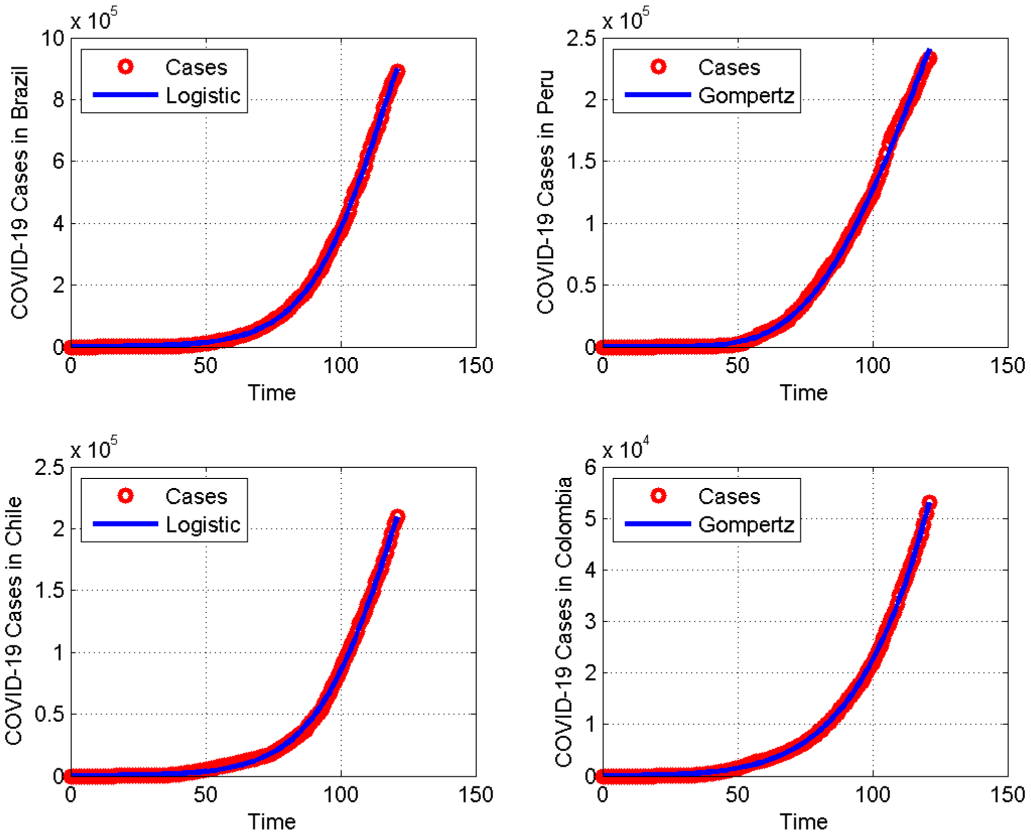

Figure 6 Estimate the cumulative number of COVID-19 cases in South America

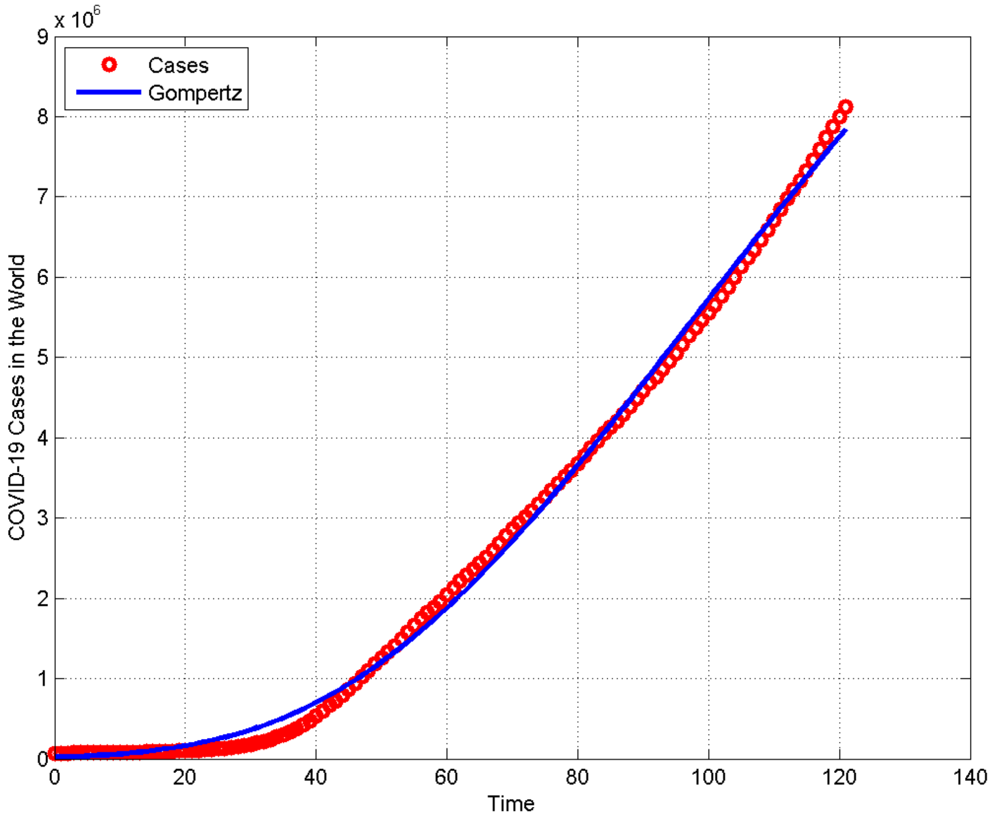

Figure 7 Estimate the cumulative number of COVID-19 cases in the World

3.2 The spread of COVID-19 and forecasting the number of cumulative cases

The suitable model, infectious population size, peak time, and the maximum growth rate are reported in Table 3 and the forecasting of the infectious population with confidence intervals is listed in Table 4. Figures 8 and 9 show the spread of COVID-19 via the derivative of the number of daily cases for each country and worldwide, respectively.

Table 3 Asymptotically infectious size and peak time of COVID-19 spreading

Countries	Population	Selected model	Asymptotical infectious size	Peak time (th day)	Maximum growth rate (cases/ day)
South Africa	59,283,506	Logistic	1000378.93 (1.69%)	165.79	14176.55
Egypt	102,263,012	Richards	92577.9489 (0.09%)	153.11	2139.72
Nigeria	205,934,216	Gompertz	64107.15 (0.03%)	138.89	440.05
Ghana	31,048,247	Gompertz	22605.0976 (0.07%)	106.62	213.71
India	1,379,567,062	Gompertz	5225298.19 (0.38%)	196.55	25427.37
Iran	83,956,634	Logistic	204517.99 (0.24%)	78.75	2187.46
Turkey	84,309,377	Logistic	166068.43 (0.20%)	64.99	4131.34
Pakistan	220,732,413	Logistic	1095243.46 (0.50%)	159.67	13477.57
Russia	145,932,745	Logistic	559568.74 (0.38%)	92.52	11502.64
UK	67,875,356	Logistic	291095.86 (0.43%)	70.22	5879.22
Spain	46,754,281	Logistic	284,296.11 (0.61%)	54.24	6,712.77
Italy	60,464,181	Logistic	230607.72 (0.38%)	48.74	5250.73
USA	330,944,050	Richards	4492267.82 (1.36%)	97.34	48620.82
Mexico	128,888,616	Logistic	231580.39 (0.18%)	113.36	3780.10
Canada	37,731,670	Gompertz	118999.75 (0.32%)	71.10	1646.24
Dominican Republic	10,844,393	Gompertz	44410.16 (0.41%)	103.07	356.89
Australia	25,490,075	Logistic	6969.14 (0.03%)	41.86	347.79
New Zealand	4,821,005	Logistic	1491.07 (0.03%)	46.32	79.46
French Polynesia	280,859	Logistic	59.36 (0.02%)	42.77	2.32
New Caledonia	285,410	Logistic	18.49 (0.01%)	37.89	1.93
Brazil	212,512,836	Logistic	1481495.96 (0.7%)	114.87	26359.75
Peru	32,956,067	Gompertz	703654.31 (2.14%)	124.15	5779.47
Chile	19,111,018	Logistic	397059.16 (2.08%)	119.39	6680.21
Colombia	50,865,177	Gompertz	6214863.65 (12.22%)	318.81	18043.30
The World	7,792,648,720	Gompertz	14448083.09 (0.2%)	96.08	104953.09

Note: Population as of Saturday, June 20, 2020 (t=126)

Table 4 Forecasted infectious cases by suitable models with $\alpha = 0.05$

Countries	Time			
	t **=122** **June 16, 2020**	***t*** **=123** **June 17, 2020**	***t*** **=124** **June 18, 2020**	***t*** **=125** **June 19, 2020**
South Africa	76334/ 77136.24 74217.38-80055.12	80412/81269.52 78350.65-84188.39	83890/85603.73 82684.86-88522.60	87715/90146.43 87227.56-93065.30
Egypt	47856/55322.24 43144.40-67500.07	49219/57447.26 45269.42-69625.09	50437/59523.49 47345.65-71701.32	52211/61543.30 49365.47-73721.13
Nigeria	17148/16282.36 14750.10-17814.62	17735/16700.15 15167.89-18232.41	18480/17120.63 15588.37-18652.89	19147/17543.64 16011.38-19075.90
Ghana	12193/11526.57 9811.70-13241.45	12590/11725.24 10010.36-13440.11	12929/11922.15 10207.28-13637.02	13203/12117.25 10402.38-13832.12
India	354161/357905.69 356240.51-359570.88	367264/370739.43 369074.24-372404.62	381091/383855.62 382190.43-385520.81	395812/397254.31 395589.12-398919.50
Iran	192439/176730.76 150126.53-203334.99	195051/177742.15 151137.92-204346.38	197647/178722.09 152117.86-205326.32	200262/179671.19 153066.96-206275.42
Turkey	181298/165499.48 143407.09-187591.87	182727/165553.21 143460.81-187645.59	184031/165601.87 143509.48-187694.26	185245/165645.95 143553.56-187738.34
Pakistan	148921/147562.11 141218.18-153906.04	154760/153928.01 147584.07-160271.93	160118/160540.45 154196.52-166884.38	165062/167407.79 161063.86-173751.72
Russia	545458/514047.54 461676.59-566418.49	553301/517369.79 464998.85-569740.74	561091/ 520468.03 468097.08-572838.97	569063/523354.71 470983.76-575725.65
UK	298136/286722.61 274317.97-299127.24	299251/287057.31 274652.67-299461.94	300469/287366.72 274962.09-299771.36	301815/287652.72 275248.09-300057.35
Spain	291408/283824.48 266150.98-301497.98	291763/283866.92 266193.42-301540.42	292348/283905.55 266232.05-301579.05	292655/283940.70 266267.21-301614.20
Italy	237500/ 230316.19 217990.10- 242642.29	237828/ 230341.54 218015.45- 242667.63	237760/ 230364.69 218038.60- 242690.78	238011/ 230385.82 218059.73- 242711.91
USA	2079592/2361047.77 2057185.56- 2664909.99	2098106/2398837.84 2094975.62- 2702700.05	2126027/2436541.25 2132679.03- 2740403.46	2149166/2474136.89 2170274.67- 2777999.10
Mexico	150264/147614.84 145322.37-149907.32	154863/151077.06 148784.58-153369.53	159793/154471.08 152178.61-156763.55	165455/157791.90 155499.43-160084.37
Canada	99467/102685.47 98839.03-106531.90	99853/103245.79 99399.36-107092.23	100220/103788.33 99941.89-107634.76	100629/104313.54 100467.10-108159.9
Dominican Republic	23686/22922.57 22450.31-23394.82	24105/23252.49 22780.24-23724.75	24645/23579.88 23107.63-24052.14	25068/23904.66 23432.40-24376.91
Australia	7347/6969.14 4998.18-8940.09	7370/6969.14 4998.18-8940.09	7391/6969.14 4998.18-8940.09	7409/6969.14 4998.18-8940.09
New Zealand	1506/1491.07 1274.66-1707.48	1506/1491.07 1274.66-1707.48	1507/1491.07 1274.66-1707.48	1507/1491.07 1274.66-1707.48
French Polynesia	60/59.3558 55.71-62.99	60/59.3558 55.71-62.99	60/59.3558 55.71-62.99	60/59.3559 55.71-62.99
New Caledonia	21/18.4963 16.66-20.33	21/18.4963 16.66-20.33	21/18.4963 16.66-20.33	21/18.4963 16.66-20.33
Brazil	928834/924882.21 917484.29-932280.13	960309/949386.05 941988.13-956783.97	983359/973403.78 966005.85-980801.69	1038568/996890.78 989492.86-1004288.69
Peru	237156/246444.96 242118.08-250771.84	240908/252220.34 247893.47-256547.22	244388/257999.09 253672.21-262325.9	247925/263778.26 259451.38-268105.14
Chile	215871/215892.25 212946.60-218837.91	220628/222499.43 219553.78-225445.08	225103/229053.16 226107.51-231998.81	231393/235539.45 232593.80-238485.11
Colombia	54931/55052.92 54553.90-55551.94	57046/57136.79 56637.77-57635.81	60217/59282.22 58783.20-59781.24	63276/61490.39 60991.37-61989.41
The World	8270313/7933525.82 7688883.63-8178168.01	8416901/8027061.19 7782419.01-8271703.38	8558504/8119838.29 7875196.11-8364480.48	8740991/8211842.38 7967200.19-8456484.56

Note: Actual/ Estimate and LCI- UCI

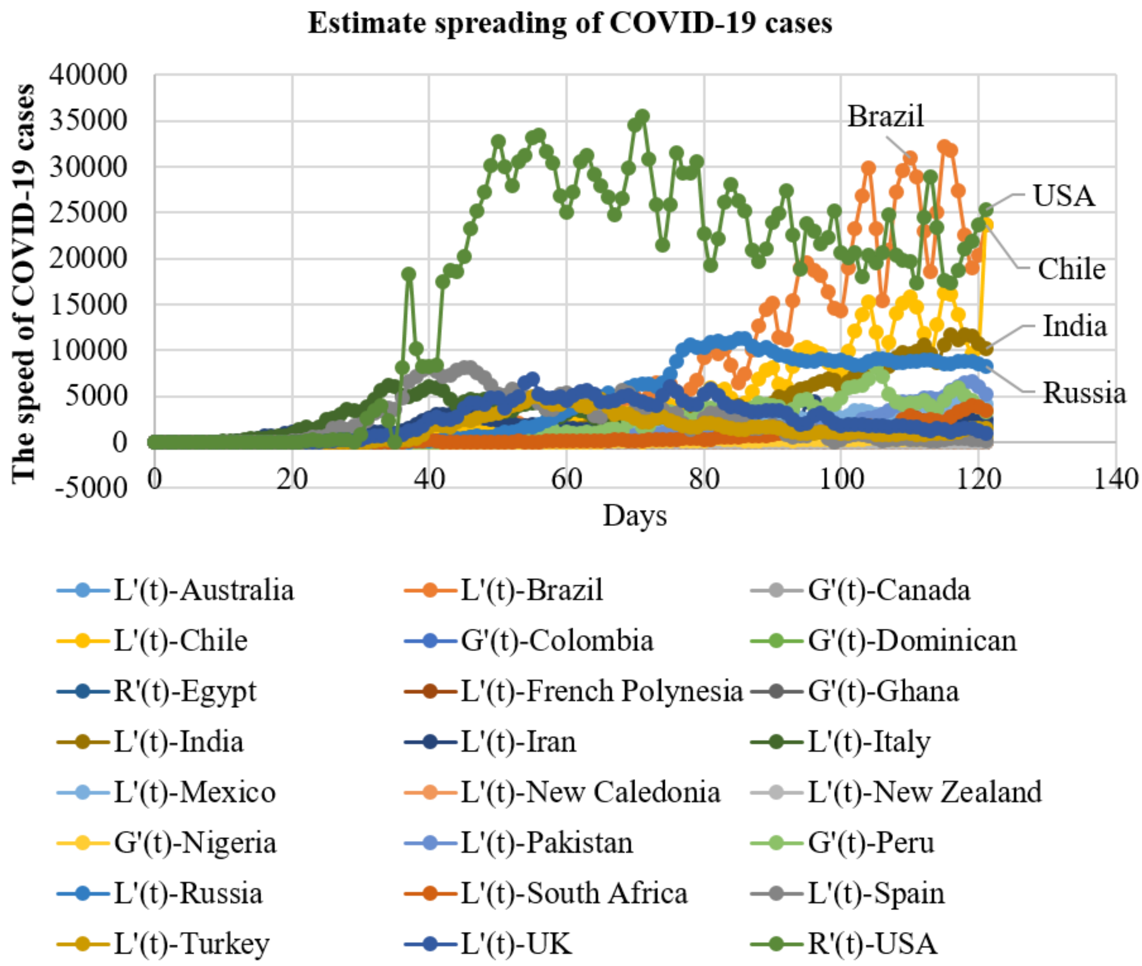

Figure 8 Spreading of COVID-19 cases estimated by the derivative

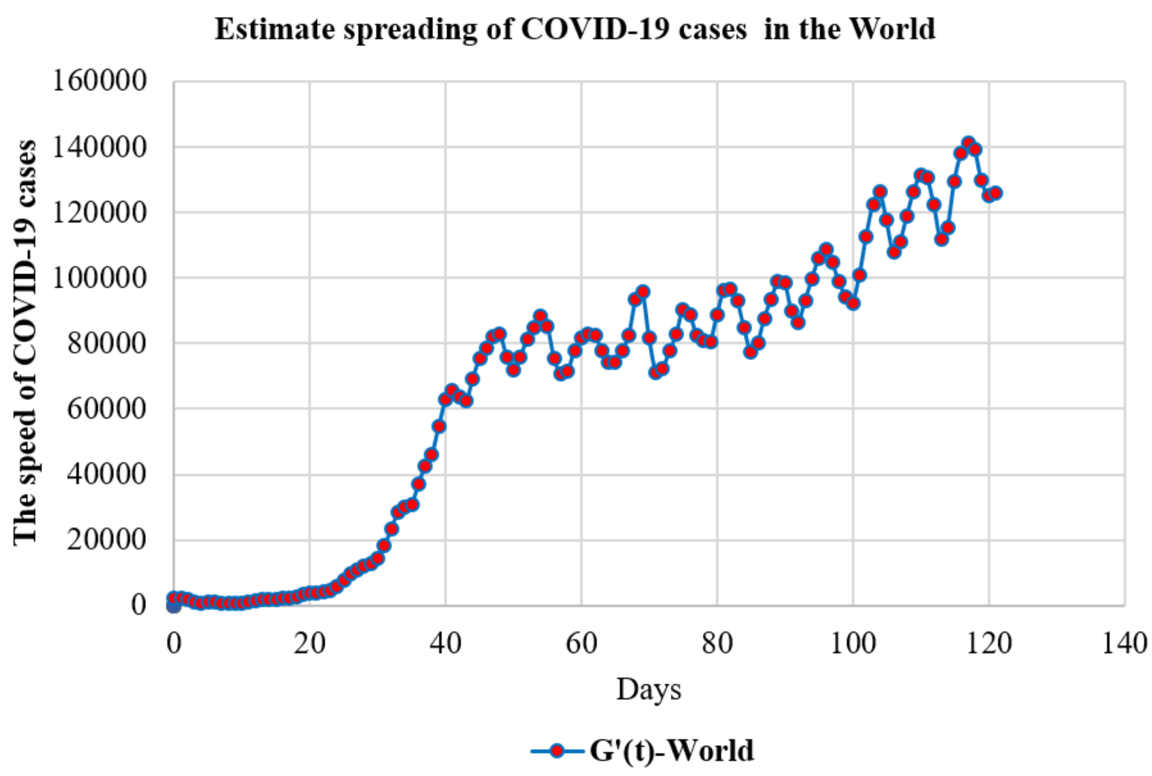

Figure 9 Spreading of COVID-19 cases estimated by the derivative in the World

4. Discussion

This research was conducted to identify the most suitable forecasting model based on the minimum squared 2-norm and maximum coefficient of determination for estimating the cumulative daily COVID-19 cases worldwide and by the most severely affected country in each continent. This was achieved by analyzing the spread of COVID-19 based on the derivative for representing the speed of the COVID-19 outbreak by country and worldwide.

In Africa, the total number of COVID-19 cases by country was suitably fitted by using the Gompertz growth curve, which is asymmetric. South Africa attained the maximum growth rate of COVID-19 cases at 14176.55 per day. Moreover, it was confirmed that South Africa, Egypt, and Nigeria have not yet reached the peak time of infection whereas Ghana has exceeded it (analyzed at June 20, 2020).

In Asia, the total COVID-19 cases in most countries were suitably fitted by applying the logistic growth curve, which is symmetric. India obtained the maximum growth rate of COVID-19 cases at 25427.37. Furthermore, India and Pakistan have not yet reached the peak time of infection whereas Turkey and Iran have already exceeded it (for June 20, 2020).

In Europe, the total COVID-19 cases of most countries were suitably fitted by using the logistic growth curve. Russia showed the maximum growth rate of COVID-19 cases at 11502.64 per day. Moreover, all of the countries tested have already exceeded the peak time of infection (for June 20, 2020).

In North America, the total COVID-19 cases of most countries were suitably fitted by applying the Gompertz growth curve, which is asymmetric, whereas the total COVID-19 cases for the USA is a whole were fitted well by using the Richards growth curve. The USA attained the maximum growth rate of COVID-19 cases at 48620.82 per day. Moreover, all of the countries have already exceeded the peak time of infection (for June 20, 2020).

In Australia, the total COVID-19 cases for most countries were suitably fitted by applying the logistic growth curve, which is symmetric. Australia obtained the maximum growth rate of COVID-19 cases at 347.79 per day. Furthermore, all of the countries have already exceeded the peak time of infection (for June 20, 2020).

In South America, the total COVID-19 cases were suitably fitted by applying both the logistic and Gompertz growth curves, which are symmetric and asymmetric, respectively. Brazil obtained the maximum growth rate of COVID-19 cases at 26359.75 per day. Moreover, Brazil, Peru, and Chile have exceeded the peak time of infection whereas Colombia has not yet reached it because of quite a long peak time at 318.81 days (for June 20, 2020).

Worldwide, the most suitable forecasting model for estimating the total COVID-19 cases was the Gompertz growth curve model, which is asymmetric. The maximum growth rate was 104953.09 COVID-19 cases per day. Notably, only the total COVID-19 cases in the USA and Egypt could be fitted by applying the Richards growth curve. Moreover, the worldwide forecasting model shows that the spread of COVID-19 has exceeded the peak time of infection at 96.08 days (for June 20, 2020). It was forecasted that the COVID-19 disease will affect approximately 0.2% of the world's population. The countries which will have an infectious population of more than 1% are South Africa (1.69%), the USA (1.36%), Peru (2.14%), Chile (2.08%), and Colombia (12.22%). The others are less than 1%. The highest maximum growth rate is in the USA (48620.82 cases per day), followed by Brazil (26359.75 cases per day), India (25427.37 cases per day), Colombia (18043.30 cases per day), and South Africa (14176.55 cases per day). To illustrate the spread of COVID-19, evaluation by applying the derivative of total COVID-19 cases in Figures 8 and 9 shows that the USA attained the maximum, followed by Brazil, Chile, India, and Russia. Meanwhile, the spread of COVID-19 worldwide has continuously increased but has exceeded the peak time of infection. The predicted values and confidence intervals are reported in Table 4. For example, the predicted cumulative COVID-19 cases in South Africa on June 16, 2020, is 77136.24 with 95% confidence intervals (74217.38–80055.12).

The forecasting models of COVID-19 data from existing research are based on exponential growth [1] and sub-exponential growth [2] of the COVID-19 by using the exponential growth curve. In reality, the growth curve of the COVID-19 outbreak is not exponential because of the flattening the curve policies, such as social distancing, hand washing, mask-wearing, etc., implemented by most governments. This has resulted in the COVID-19 cases more closely following the S-curve proposed in this research, which is more like the logistic, Richards, and Gompertz growth curves. Moreover, we evaluated the maximum growth rate and the peak time of COVID-19 infections at the inflection point of the growth curves, and thus could more accurately forecast the number of cumulative COVID-19 cases. Notably, the maximum growth rate and the peak time of COVID-19 infections are not used in the generalized growth curve model [3] for the average number of COVID -19 infected individuals. In addition, we estimated the spread of the COVID-19 outbreak based on an estimated derivative of the total COVID-19 cases. The approach used in this research could be an important weapon for fighting the spread of COVID-19 by identifying when to action policies for flattening the growth curve of the COVID outbreak.

The proposed forecasting models include uncontrollable external factors. For example, people living outside of their country of origin can return home and any positive cases should be detected while they are under state quarantine. However, although this scenario is included in the proposed forecasting model, these individuals should not be included because it can cause an unreal second wave of infection to manifest. Another limitation of this research is that the parameters of the models will change over time, and so they are only accurate in the short term. and will need to be recalculated in the long term. In future research, we will focus on forecasting models for the real second wave of COVID-19 infections.

5. Conclusions

A suitable forecasting model for analyzing the number of daily cumulative COVID-19 cases worldwide is the Gompertz growth curve, which is asymmetric. The USA is in first place for the spread of COVID-19, as shown graphically based on the derivative of the spread fitted by applying the Richards growth curve, which is asymmetric. Brazil is in second place, as most suitably fitted with the logistic growth curve, while India is third, as most suitably fitted with the Gompertz growth curve. However, the logistic model, which uses a symmetric growth curve, most often occurred in regions where the spread of COVID-19 has been well-controlled, such as Australia and Asia. Of particular note is that the spread of COVID-19 in Colombia should be carefully monitored because its peak time of infection is much longer than the other countries. Besides, the spread of COVID-19 in Chile has rapidly increased and its growth curve resembles those of the USA and Brazil. Worldwide, the spread of COVID-19 curve using the Gompert model showed a continuous increase but the peak time of infection had already passed. The results show that COVID-19 will infect approximately 0.2% of the world's population for short time period of the prediction.

6. References

[1] Li Y, Liang M, Yin X, Liu X, Hao M, Hu Z, et al. COVID-19 epidemic outside China: 34 founders and exponential growth. J Investig Med. 2020:1-4.
[2] Maier BF, Brockmann D. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in mainland china. arXiv:2002.07572. 2020:1-9.
[3] Amaral MTR, Katiane SC, Marinho GA, Carlos RP. Generalized growth curve model for COVID-19 in Brazilian states. Revista Brasileira de Biometria. 2020;38:125-46.
[4] Kucharavy D, Roland DG. Application of logistic growth curve. Procedia Eng. 2015;131:280-90.
[5] Vaghi C, Rodallec A, Fanciullino R, Ciccolini J, Mochel JP, Mastri M, et al. Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. PLoS Comput Biol. 2020;16:1-24.
[6] Eberhardt LL, Breiwick JM. Models for population growth curves. Int Sch Res Not. 2012;2012:1-7.
[7] Yoshitaka H. Calculation of population parameters using Richards function and application of indices of growth and seed vigor to rice plants. Plant Prod Sci. 1999:129-35.
[8] Li-Qun J. Analysis of a modified logistic model for describing the growth of durable customer goods in China. Math Comput Appl. 2013;18:30-7.
[9] Ersoy IE, Mendes M, Keskin S. Estimation of parameters of linear and nonlinear growth curve models at early growth stage in California Turkeys. Eur Poultry Sci. 2007;71:175-80.
[10] Verhulst PF. Notice sur la loi que la population suit dans son accroissement. Math Phys. 1838;1838:113-21.
[11] Martha LA, James PB. Introductory Differential Equations. 5th ed. New York: Academic Press; 2018.
[12] Tsoularis A, Wallace J. Analysis of logistic growth models. Math Biosci. 2002;179:21-55.
[13] Worldometer [Internet]. 2020 [cited 2020 May 26]. Available from: https://www.worldometers.info/coronavirus/#countries.
[14] Levenberg K. A method for the solution of certain non-linear problems in least squares. Q J Mech Appl Math. 1994;2:164-8.
[15] Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. SIAM J Appl Math. 1963;11:431-41.