

## Predictive Models for Cumulative Confirmed COVID-19 Cases by Day in Southeast Asia

Yupaporn Areepong<sup>1</sup> and Rapin Sunthornwat<sup>2,\*</sup>

<sup>1</sup>Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand

<sup>2</sup>Industrial Technology Program, Pathumwan Institute of Technology, Bangkok, 10330, Thailand

\*Corresponding Author: Rapin Sunthornwat. Email: rapin@pit.ac.th

Received: 25 June 2020; Accepted: 07 September 2020

**Abstract:** Coronavirus disease 2019 outbreak has spread as a pandemic since the end of year 2019. This situation has been causing a lot of problems of human beings such as economic problems, health problems. The forecasting of the number of infectious people is required by the authorities of all countries including Southeast Asian countries to make a decision and control the outbreak. This research is to investigate the suitable forecasting model for the number of infectious people in Southeast Asian countries. A comparison of forecasting models between logistic growth curve which is symmetric and Gompertz growth curve which is asymmetric based on the maximum of Coefficient of Determination and the minimum of Root Mean Squared Percentage Error is also proposed. The estimation of parameters of the forecasting models is evaluated by the least square method. In addition, spreading of the outbreak is estimated by the derivative of the number of cumulative cases. The findings show that Gompertz growth curve is a suitable forecasting model for Indonesia, Philippines, and Malaysia and logistic growth curve suits the other countries in South Asia.

**Keywords:** Coronavirus disease 2019; Gompertz function; least square estimation; logistic function

### 1 Introduction

Coronavirus disease 2019 is an emerging infectious disease which is shortly named as COVID-19. The human's lung which is easily attacked by Coronavirus is the main part of human body because there are a lot of receptors which match the spike proteins of Coronavirus. Then, Coronavirus will damage the human lungs. This is the main point of severe acute respiratory syndrome in the human respiratory system. The patient zero which gets the Coronavirus 2019 disease of the world was found at the end of year 2019 in Wuhan, China. Consequently, the disease has spread to all areas and it has caused the health problems and economic problems for all counties in the world [1,2]. Southeast Asia is one of the areas in the world have been affected by the outbreak of COVID-19. Southeast Asia consists of 11 countries, Brunei, Cambodia, East Timor, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam. It is located in the tropical zone of the world and the disease spreads in the summer season



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(around March to May, 2020) of Southeast Asia. The situation (May 30, 2020) of COVID-19 transmission in Southeast Asia is likely declining. However, one of the causes of COVID-19 spreading in Southeast Asia is evacuation of migrant workers who are vulnerable populations to their hometown when each country in Southeast Asia acts suddenly lockdown policy. They cannot abide by the International Health Regulations. For example, the social and physical distancing cannot practice to this group due to a lot of living in migrant workers' dormitories. Then, the migrant workers have become disease carriers and transmit the disease to their countries. However, the authorities of each country have detected this group to the state quarantine process in order to control the spreading of disease [3–5].

To control the outbreak of disease, forecasting of the number of infectious people is very important to make a decision and develop policy of the authorities. A lot of research about estimating growth curve of population and predictive models has been studied. Forecasting of the future and distant changes of new processes and products by using logistic curve was studied [6]. Bacterial growth experiments and population dynamic equations based on logistic growth curve was investigated [7]. Fitting data was used by logistic growth curve to develop the study of empirical description of plant growth [8]. Researchers analyzed and fitted Gompertz growth curve to study the bacterial growth, and tumor growth. Also, the parameters of logistic and Gompertz growth curve are estimated by the least square method. The growth curve based on Gompertz growth curve was analyzed by using mathematical and statistical analysis [9]. Gompertz growth curve was used for estimation of the bacterial growth [10,11]. Gompertz growth curve was initially fitted by data of tumor growth [12]. Estimated parameters of logistic growth curve was approximated by using the least square method [13]. Data and estimated parameters were fitted by logistic growth curve based on the least square method [14]. The least square method was adopted to estimate parameters of Gompertz growth curve [15]. Remarkably, the number of infectious cases of COVID-19 looks like S-curve. Also, logistic and Gompertz growth curve shape S-curve with different parameters to represent in the cases. Furthermore, a lot of research has been studied about spreading of COVID-19. The spreading of COVID-19 of 11 European countries was studied during starting time of outbreak until 4 May, 2020—after a big outbreak at China. The effects of non-medical measures such closure of academic institutes and lockdown policies were investigated. The proposed models which can be backward calculated from the death cases to estimate the spreading with allowing for the lag time between infectious and death time. The data were internationally merged for the time-varying reproduction number. The estimation was reported that the non-medical measures derived the reproductive number to be less than one with confidence interval at 99%. Briefly, the 11 European countries will have 12–15 millions of infected individuals at 14 May, 2020. The non-medical measures especially lockdowns effects to decrease the COVID-19 spreading [16]. After the spreading of COVID-19 at Wuhan, China, the outbreak threatened the health systems of all the world. The conceptual model for dynamical system of COVID-19 outbreak was studied at Wuhan, China with Chinese authority such as establishing special hospitals and limitation in immigration of people across borders. The individual behaviors were are restricted and studied to correspond to the authorities' policies such as extension of holiday, measures of quarantine, etc. The models were estimated by with two main components based on the 1918 influenza pandemic and the trend of pandemic with reporting ratio [17]. The outbreak of COVID-19 has transmitted to all countries in the world. The number of infectious cases in the world is increasing. The events, festivals, and tournaments have been canceled or postponed as well as the academic institutions were closed more than 14 days, incubation period of COVID-19. This outbreak affects the economics and health system in the world. The mathematical model (dynamical system) based on differential equation, SIR compartmental model,

of COVID-19 spreading was modelled to describe the behavior of COVID-19 outbreak. The parameters in the model were estimated to analyze the COVID-19 outbreak situation in China for developing the efficient measures to control the spreading [18]. The outbreak model of COVID-19 and analysis of effects from lockdown policy in India were conducted. The assumptions of model are based on lockdown policy, quarantine, and infectious cases. The trend and rate of COVID-19 spreading of 18 states in India were forecasted. The results showed that the peak time and ending time of spreading in India will be July, 2020 and March, 2021, respectively. Besides, the number of infectious cases in India was approximately estimated over 19 lakhs [19]. The dynamical model with time of COVID-19 outbreak in China, Italy, and France was analyzed and proposed. The simple susceptible-infected-recovered-deaths model as dynamical model for this research indicated that the kinetic parameter did not affect to the analysis of outbreak each country whereas the infection and death rates appear to be more variable [20]. The reproductive number of COVID-19 was estimated and presented on the difference and parameter selection which reflect the dynamical transmission with the reproductive number. The risk of transmission on COVID-19 outbreak was also estimated. The method for detecting infectious cases and risk assessment on COVID-19 outbreak improved rapidly by sensor technology to shorten detection time and led to quickly diagnose the cases. Moreover, the diagnosis of cases through dynamical system based on the reproductive number was proposed. Then, the suitable measures, self-quarantine and isolation, were intervened to control the spreading and to evaluate risk of COVID-19 transmission [21].

However, forecasting models for estimating the growth curve of COVID-19 infectious people are a novel topic. The number of COVID-19 infectious people is interesting to compare the results from two forecasting models, logistic and Gompertz growth curves. Therefore, the main purpose of this research is to construct the forecasting models for prediction of the cumulative number of COVID-19 cases by day in Southeast Asia. Logistic which is symmetric growth S-curve with respect to its inflection point and Gompertz which asymmetric growth S-curve are selected to be the fitting functions of the cumulative number of cases which look like S-curve with the least square method for the parameters estimation. Furthermore, the comparison among the models to be suitable model is based on optimization of Coefficient of Determination and Root Mean Squared Percentage Error. Next section relates to mathematical and statistical background as the materials and methods for this research. Then, the results and discussion, and conclusion are also proposed for this research.

## **2 Materials and Methods**

This section presents data collection, the forecasting models—logistic growth curve model and Gompertz growth curve model—for predicting the number of daily cumulative confirmed COVID-19 cases of each country in Southeast Asia. Besides, the algorithm for forecasting models of COVID-19 cases is provided. The two growth curve models are derived from the differential equation as the followings.

### **2.1 Data Collection**

Data or observations as the secondary data for the cumulative daily confirmed COVID-19 cases are gathered from the website Worldometer [22]. Worldometer is operated by oversea developers, researchers, and volunteers to provide statistics collected around the world. This is a website which also proposed the data about COVID-19 on real time daily. The cumulative daily confirmed COVID-19 cases as data for this research in each country on Southeast Asia are collected from February 15, 2020 ( $t=0$ ) to May 20, 2020 ( $t=95$ ). The collected data are perfect and no missing values.

## 2.2 Forecasting Models and Its Analytics

Let  $l(t)$  be the number of individuals based on logistic growth curve of daily cumulative confirmed COVID-19 cases of each country in Southeast Asia at time  $t \geq 0$ . The logistic growth curve is a solution of differential equation which was initiated by Verhulst [23],

$$\frac{d}{dt}l(t) = rl(t) \left[ 1 - \frac{l(t)}{C} \right]; \quad l(t=0) = l_0 \quad (1)$$

where  $l_0 > 0$  is an initial condition for initial population size of cumulative COVID-19 cases,  $r > 0$  is a growth rate parameter, and  $C$  is a carrying capacity parameter. The solution of Eq. (1) can be carried out by partial fraction and separable method [24] as

$$\begin{aligned} \frac{dl(t)}{l(t) \left( 1 - \frac{l(t)}{C} \right)} &= r dt \\ \int \left( \frac{1}{C-l(t)} + \frac{1}{l(t)} \right) dl(t) &= \int r dt \\ -\ln|C-l(t)| + \ln|l(t)| &= rt + A \\ \ln \left| \frac{l(t)}{C-l(t)} \right| &= rt + A \end{aligned}$$

Taking exponential function into both sides, the result is

$$\frac{l(t)}{C-l(t)} = B \exp(rt); \quad B = \pm \exp(A) \text{ where } A, B \text{ are arbitrary constant.}$$

With initial condition  $l(t=0) = l_0$ , the result is obtained as

$$l(t) = \frac{C}{1 + K \exp(-rt)} \text{ where } K = \frac{C-l_0}{l_0}. \quad (2)$$

Let  $g(t)$  be the number of individuals based on Gompertz growth curve of daily cumulative confirmed COVID-19 cases in Southeast Asia with growth curve  $D$  which is intrinsic growth rate and represents growth rate per capita at time  $t \geq 0$ . Physically, this rate is the parameter to control Gompertz growth curve to be S-curve which represents the COVID-19 curve of the number of infectious cases. The Gompertz growth curve is also a solution of differential equation [25] based on an exponential growth differential equation

$$\frac{d}{dt}g(t) = \gamma(t)g(t); \quad g(t=0) = g_0 \quad (3)$$

where  $\frac{d}{dt}\gamma(t) = -D\gamma(t); \quad \gamma(t=0) = \gamma_0$ .

$$\frac{d}{dt}\gamma(t) = -D\gamma(t)$$

$$\frac{d\gamma(t)}{\gamma(t)} = -Ddt$$

$$\ln |\gamma(t)| = -Dt + A$$

$$\gamma(t) = B \exp(-Dt); \quad B = \pm \exp(A)$$

With initial condition  $\gamma(t=0) = \gamma_0$ , the result is

$$\gamma(t) = \gamma_0 \exp(-Dt) \tag{4}$$

By substitution of Eq. (4) into Eq. (3), the result is obtained as

$$\frac{d}{dt}g(t) = \gamma_0 \exp(-Dt) g(t)$$

$$\frac{dg(t)}{g(t)} = \gamma_0 \exp(-Dt) dt$$

$$\ln |g(t)| = -\frac{\gamma_0}{D} \exp(-Dt) + A \text{ where } A \text{ is arbitrary constant.}$$

With initial condition  $g(t=0) = g_0$ , the result is

$$\ln \left| \frac{g(t)}{g_0} \right| = \frac{\gamma_0}{D} [1 - \exp(-Dt)]$$

By taking exponential function into both sides, the result is obtained as

$$g(t) = g_0 \exp\left(\frac{\gamma_0}{D} [1 - \exp(-Dt)]\right) \tag{5}$$

Next, analysis of logistic growth curve model and Gompertz growth curve model is an essential understanding for application the models to pandemic of COVID-19. There are three aspects for analysis of forecasting models—asymptotical behavior, the maximum growth rate behavior, and symmetrical behavior.

Asymptotical behavior of forecasting models will occur when time tends to infinity. That is, the cumulative number of COVID-19 infectious people is asymptotically estimated by the steady state which is state of system for long time in the future.

For logistic growth curve model,

$$l(t) = \lim_{t \rightarrow \infty} \left( \frac{C}{1 + K \exp(-rt)} \right) = C.$$

For Gompertz growth curve model,

$$g(t) = \lim_{t \rightarrow \infty} \left[ g_0 \exp\left(\frac{\gamma_0}{D} [1 - \exp(-Dt)]\right) \right] = g_0 \exp\left(\frac{\gamma_0}{D}\right).$$

The maximum spreading behavior of the cumulative number of COVID-19 infectious people is the peak point of growth rate of COVID-19 cases.

For logistic growth curve model, to find the maximum growth rate of logistic growth curve with peak time, the saddle point is computed as

$$\frac{d^2}{dt^2}l(t) = 0$$

$$-C(1 + K \exp(-rt))^{-2} (Kr^2 \exp(-rt)) + 2C(1 + K \exp(-rt))^{-3} (Kr \exp(-rt))^{-2} = 0$$

$$t = \frac{\ln(K)}{r}$$

$$l(t) = \frac{C}{1 + K \exp(-\ln(K))} = \frac{C}{2}$$

For Gompertz growth curve model, to find the maximum growth rate of Gompertz growth curve with peak time, the saddle point is computed as

$$\frac{d^2}{dt^2} g(t) = 0$$

$$-g_0 \lambda D^2 \exp(\lambda - Dt - \lambda \exp(-Dt)) + g_0 \lambda^2 D^2 \exp(\lambda - 2Dt - \lambda \exp(-Dt)) = 0; \quad \lambda = \frac{\gamma_0}{D}$$

$$t = \frac{\ln(\lambda)}{D}; \quad \lambda = \frac{\gamma_0}{D}$$

$$g(t) = g_0 \exp(\lambda - \lambda \exp(-\ln(\lambda))) = g_0 \exp(\lambda - 1)$$

Symmetrical behavior of forecasting models is the behavior of the fitted data which are the cumulative number of COVID-19 infectious people whether it is symmetric data or not.

For logistic growth curve model, it is symmetric about its saddle point.

For Gompertz growth curve model, it is asymmetric.

Generally, the definition of the first derivative of function  $f(t)$  which represents logistic or Gompertz growth curve at time of a point  $t_*$  is defined as

$$\frac{d}{dt} f(t_*) = \lim_{h \rightarrow 0} \frac{f(t_* + h) - f(t_*)}{h}$$

Also, it can be estimated by  $\frac{f(t_* + h) - f(t_* - h)}{2h}$  and the second derivative is  $\frac{d}{dt} \left[ \frac{d}{dt} f(t_*) \right]$ .

### 2.3 Parameters Estimation

Let  $f(t; a, b, c)$  denotes a function of logistic or Gompertz growth curve with three parameters  $a, b, c$  for any time  $t \geq 0$ . The parameters are estimated by fitting actual total COVID-19 cases  $x_t$  for any time  $t \geq 0$  with function logistic or Gompertz growth curve. Here, the least square method is applied to estimate the parameters based on the minimum of the square of deviation function  $d(a, b, c)$  as

$$\text{Min}_{a,b,c} \left\{ d(a, b, c) = \sum_{t=0}^n (x_t - f(t; a, b, c))^2 \right\}$$

where  $a$  represents  $l_0$  or  $g_0$  for logistic or Gompertz growth curve, respectively,  $b$  represents  $C$  or  $\gamma_0$  for logistic or Gompertz growth curve, respectively,  $c$  represents  $r$  or  $D$  for logistic or Gompertz growth curve, respectively.

To minimize the square of deviation function, taking partial derivatives on the square of deviation function is carried out and equals them to zero. This yields the linear equation system to solve for the parameters  $a, b, c$  as

$$\frac{\partial}{\partial a}d(a, b, c) = 0$$

$$\frac{\partial}{\partial b}d(a, b, c) = 0$$

$$\frac{\partial}{\partial c}d(a, b, c) = 0$$

### 2.4 Accuracy of Forecasting Models

The accuracy of forecasting model is checking the validation of forecasting model with a comparison between forecasted value and actual value via the Root Mean Squared Percentage Error, appropriation of forecasting model via the Coefficient of Determination, and the trust of forecasting model via confidence interval. Let  $f(t)$  be the actual value and  $\hat{f}(t)$  be the forecasted value.

Root Mean Squared Percentage Error (RMSPE)

$$RMSPE = \sqrt{\frac{1}{n} \sum_{t=0}^n \left( \frac{f(t) - \hat{f}(t)}{f(t)} \right)^2}$$

Coefficient of Determination ( $R^2$ )

$$R^2 = \frac{\text{Explained Variance}}{\text{Total Variance}}$$

$$= \frac{\sum_{t=1}^n (\hat{f}(t) - \bar{f}(t))^2}{\sum_{t=1}^n (f(t) - \bar{f}(t))^2}$$

Confidence Interval (CI)  $100(1 - \alpha)\%$  of  $f(t)$

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

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

where  $SE$  is the standard error of  $f(t)$  and  $df$  is the degree of freedom.

### 2.5 Algorithm for Forecasting Model of COVID-19 Cases

Let  $\theta$  be a set of parameters for logistic or Gompertz growth curve.

---

**Algorithm 1:** Parameter estimation and Comparison of the predictive model

---

*Step 1: Setting actual value*

Population  $\leftarrow$  the actual number of cumulative COVID-19 cases

Time  $\leftarrow$  0 to length(Population) – 1 with step 1

*Step 2: Setting initial value for predictive models*

$r, D \leftarrow$  Rate of change of present Time with respect to previous time

$\theta_l \leftarrow$  initial setting  $l_0, C, r$  for logistic growth curve

$\theta_g \leftarrow$  initial setting  $g_0, \gamma_0, D$  for Gompertz growth curve

*Step 3: Applying the least square method for parameter estimation*

Fitted value  $\leftarrow$  Leastsquare(growth curve,  $\theta$ , Population, Time)

*Step 4: Calculating error to compare predictive models between actual value and fitted value to be a suitable model based on the minimum of root mean squared percentage error and based on the maximum of coefficient of determination*

*Step 5: Forecasting the future value and its confident interval*

*Step 6: Comparing the suitable predictive model for spreading of COVID-19 by using the derivative*

---

### 3 Results and Discussion

This section illustrates the results which consist of estimated parameters of forecasting models and validation of forecasting models. Also, interpretation of results is discussed and analyzed the number of total COVID-19 cases and of its spreading in Southeast Asia.

#### 3.1 Estimated Parameters and Forecasting Models

This section provides the estimated Parameters of forecasting models, logistic growth curve and Gompertz growth curve, for countries in Southeast Asia.

**Table 1:** Estimated parameters and suitable model

Country	Logistic growth curve					Gompertz growth curve				
	$l_0$	$C$	$r$	$R^2$	$RMSPE$	$g_0$	$\gamma_0$	$D$	$R^2$	$RMSPE$
Brunei	0.076	138.318	0.228	0.997*	1.235*	4.081e-05	1.361	0.090	0.982	2.158
Cambodia	0.003	120.509	0.302	0.996*	0.425*	4.151e-06	1.391	0.080	0.949	0.894
East Timor	6.352e-08	23.937	0.325	0.989*	0.464*	5.765e-08	1.033	0.051	0.905	2.043
Indonesia	62.889	23159.869	0.076	0.997	26.843	0.394	0.328	0.028	0.999*	4.522*
Laos	0.001	19.160	0.215	0.992*	0.083*	6.630	0.909	0.061	0.953	0.190
Malaysia	35.352	6680.308	0.106	0.995	1.726	0.003	0.875	0.059	0.998*	0.414*
Myanmar	0.008	182.374	0.160	0.994*	0.148*	0.001	0.557	0.038	0.972	0.608
Philippines	62.911	13836.474	0.079	0.992	25.945	0.069	0.479	0.039	0.997*	2.851*
Singapore	4.548	30766.634	0.119	0.997*	0.446*	0.119	0.381	0.029	0.988	0.689
Thailand	1.988	2961.349	0.165	0.998*	0.448*	0.015	0.703	0.057	0.972	0.724
Vietnam	1.438	287.001	0.137	0.989*	0.303*	0.002	0.878	0.072	0.985	0.403

Note: \* Appropriate value.



Tab. 1 shows the estimated parameters of forecasting models, logistic growth curve and Gompertz growth curve. The selected model is based on the maximum of Coefficient of Determination which informs how the independent variable (time) can be explain the dependent variable (the number of total COVID-19 cases) and the minimum of Root Mean Squared Percentage Error. The results show the suitable forecasting model for each country in Southeast Asia as follows:

The suitable forecasting model for Indonesia, Philippines, and Malaysia is Gompertz growth curve which is asymmetric.

The suitable forecasting model for Singapore, Thailand, Vietnam, Cambodia, Brunei, Myanmar, East Timor, and Laos is logistic growth curve which is symmetric.

Fig. 1 shows the estimate vs. actual total number of COVID-19 cases in Singapore, Indonesia, Philippines, and Malaysia based on selected model.

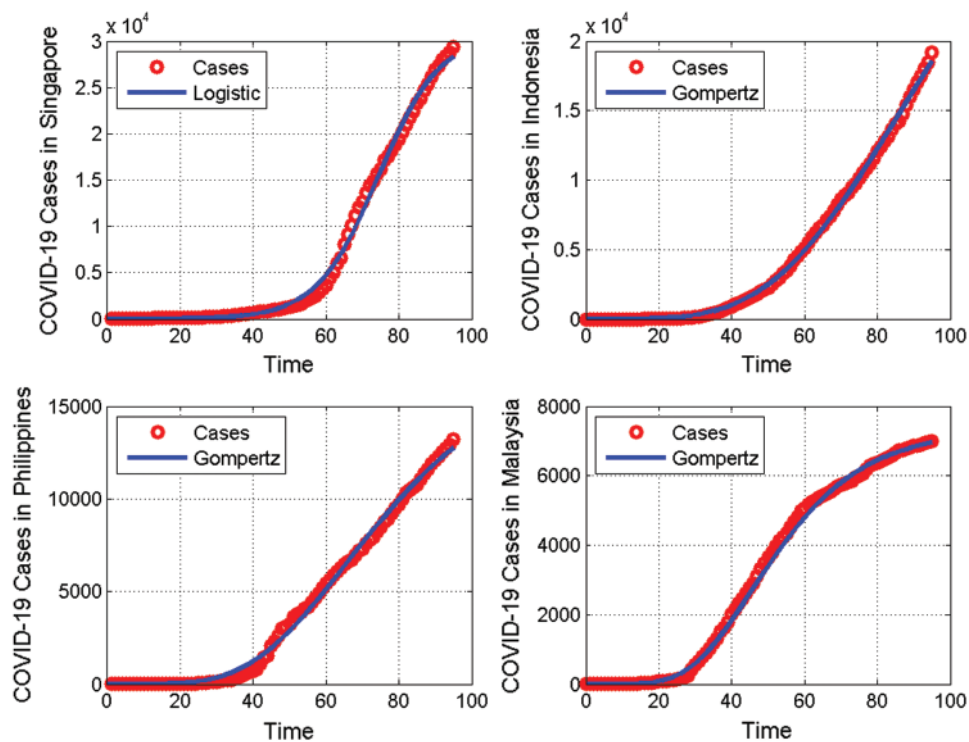
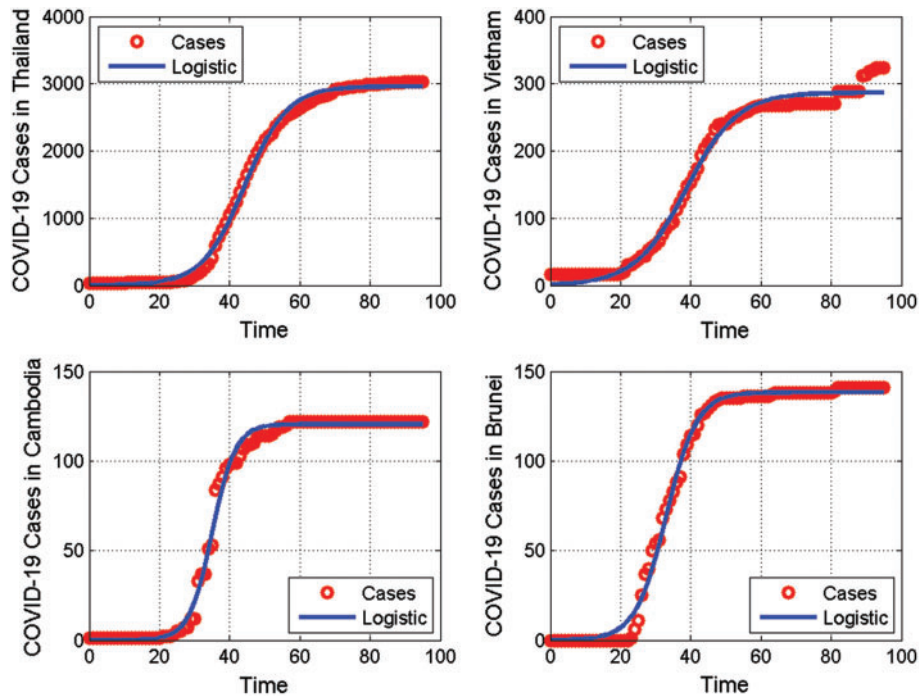


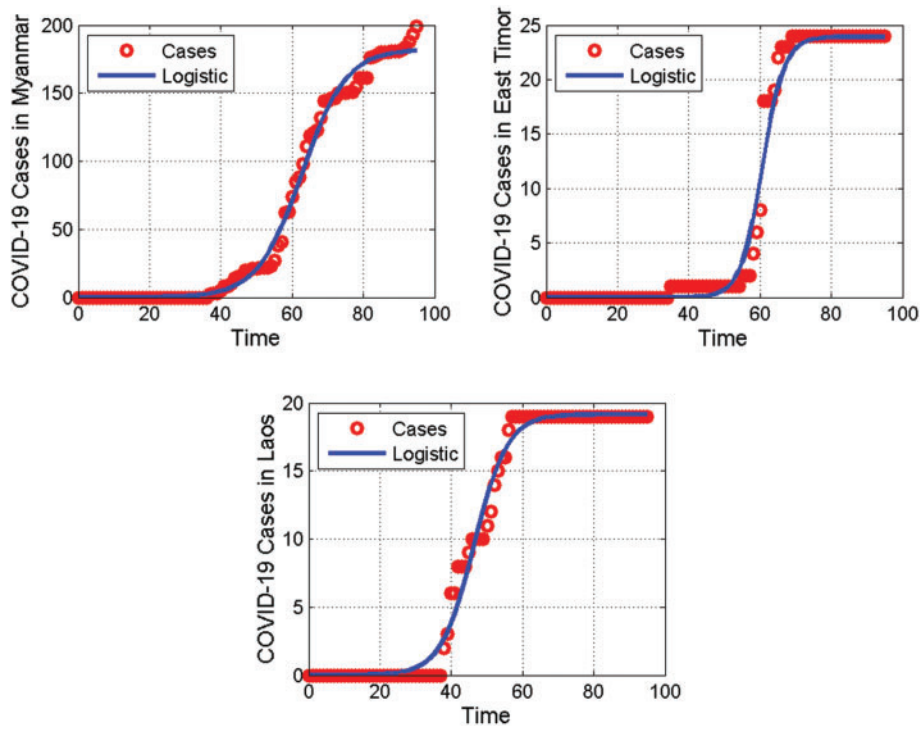
Figure 1: Estimate COVID-19 total cases in Singapore, Indonesia, Philippines, and Malaysia

Fig. 2 shows the estimate vs. actual total number of COVID-19 cases in Thailand, Vietnam, Cambodia, and Brunei based on selected model.

Fig. 3 shows the estimate vs. actual total number of COVID-19 cases in Myanmar, East Timor, and Laos based on selected model.



**Figure 2:** Estimate COVID-19 total cases in Thailand, Vietnam, Cambodia, and Brunei



**Figure 3:** Estimate COVID-19 total cases in Myanmar, East Timor, and Laos

### 3.2 Spreading of COVID-19 and Forecasting the Number of Total Cases in Southeast Asia

The results of behavior and estimating spread of COVID-19 cases in Southeast Asian Countries are provided in this section.

Tab. 2 shows the asymptotical infectious size, peak time of spreading which is calculated at saddle point, and the maximum growth rate which is calculated by the derivative at peak time of COVID-19 spreading in Southeast Asia. The approximate total infectious population size during outbreak is asymptotical infectious size. The time for the fastest spreading (peak time) of the COVID-19 in Southeast Asia is occurred at the saddle point. Also, the time in which the spread of the disease starts to slow down is after the peak time. Indonesia has just passed the peak time for approximately 15 days, approximately 25 days for Singapore, and approximately 35 days for Philippines, with respect to starting February 15, 2020 to ending May 25, 2020. Peak time is important because the spread after peak time will slowly decrease. The first two countries in Southeast Asia of the maximum growth rates of spreading are Singapore and Indonesia respectively. The duration of approaching peak time of Indonesia is longer than the other countries in Southeast Asia. The maximum of asymptotical infectious size is Singapore, about 0.526%. The time for the fastest spreading of the COVID-19 in Singapore is about 74.366 days after the beginning outbreak ( $t = 0$ ). At the peak of pandemic, approximately 912.194 new people get the disease per day in Singapore. In the same way, the time for the fastest spreading of the COVID-19 in Indonesia is about 86.173 days after the beginning outbreak ( $t = 0$ ). At the peak of pandemic, approximately 421.487 new people get the disease per day in Indonesia.

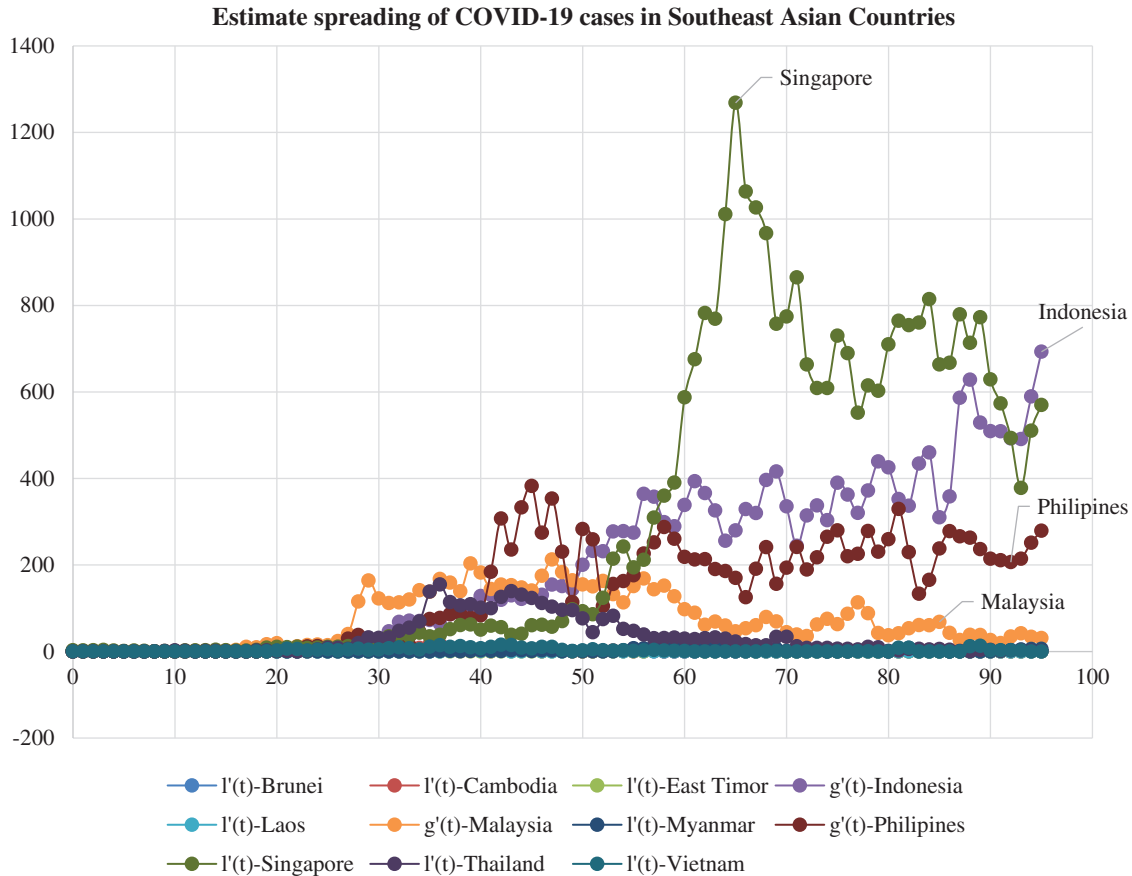
**Table 2:** Asymptotically infectious size and peak time of COVID-19 spreading in Southeast Asian countries

Country	Population	Selected model	Asymptotical infectious size (%)	Peak time ( <sup>th</sup> day)	Maximum growth rate (cases/day)
Brunei	437,479	Logistic	138.318 (0.032)	32.968	7.872
Cambodia	16,718,965	Logistic	120.509 (0.001)	35.044	9.095
East Timor	1,318,445	Logistic	23.937 (0.002)	60.688	1.947
Indonesia	273,523,615	Gompertz	40369.752 (0.015)	86.173	421.487
Laos	7,275,560	Logistic	19.160 (0.001)	46.465	1.029
Malaysia	32,365,999	Gompertz	7336.969 (0.023)	45.558	159.608
Myanmar	54,409,800	Logistic	182.374 (0.001)	35.044	7.315
Philippines	109,581,078	Gompertz	17562.819 (0.016)	65.373	249.143
Singapore	5,850,342	Logistic	30766.634 (0.526)	74.366	912.194
Thailand	69,799,978	Logistic	2961.349 (0.004)	44.317	122.047
Vietnam	97,338,579	Logistic	287.001 (0.001)	38.719	9.805

Note: Population as of May 25, 2020 ( $t = 100$ ).

Fig. 4 show the spreading of COVID-19 in Southeast Asia via the derivative of the number of daily total cases. Namely, the significance of derivative is represented by the speed of spreading of COVID-19. The spreads in Malaysia, Thailand, Vietnam, Cambodia, Brunei, Myanmar, East Timor, and Laos have decreased because the spreading in these countries exceeds the peak time

and then the spreading will slowly decrease after peak time. Furthermore, some countries of them hardly spread such as Laos, East Timor.



**Figure 4:** Estimate the derivative of COVID-19 cases in Southeast Asian countries

On the other hand, the spreading in Philippines fluctuates though it has exceeded peak time. After peak time, the spreading in Philippines is continuously sideways and the suitable forecasting model for Philippines is Gompertz growth curve which is asymmetric.

The outbreak in Singapore is very severe in Southeast Asia due to a lot of the number of total cases but it has passed the peak time and the spreading has gradually declined after peak time. It is likely that the number of total cases in Singapore will continuously decrease because the suitable forecasting model for Singapore is symmetric growth curve which is logistic growth curve model.

On the other hand, Indonesia is one of the countries that the spreading of COVID-19 is severe because it is continually increasing. However, the suitable forecasting model for Indonesia, Gompertz growth curve which is asymmetric, shows that the spreading in Indonesia has just passed the peak time for approximately 15 days. It delays in passing the peak time with comparing to the other countries in Southeast Asia. It is possible that the spreading in Indonesia is still increasing because the suitable forecasting model is asymmetric.

Moreover, the forecasted number of total COVID-19 cases and confidence interval with 95% confidence of total COVID-19 cases in Southeast Asian are demonstrated in [Tab. 3](#).

**Table 3:** Forecasted infectious cases by selected models in Southeast Asia with  $\alpha = 0.05$

Countries	Time			
	$t = 96$ May 21, 2020	$t = 97$ May 22, 2020	$t = 98$ May 23, 2020	$t = 99$ May 24, 2020
Brunei	141/138.317 131.136– 145.498	141/138.317 131.136– 145.498	141/138.3175 131.136– 145.498	141/138.318 131.136– 145.498
Cambodia	123/120.509 113.393– 127.624	123/120.509 113.393– 127.624	124/120.509 113.393– 127.624	124/120.5089 113.393– 127.624
East Timor	24/23.937 21.599– 26.274	24/23.937 21.599– 26.274	24/23.937 21.599– 26.274	24/23.937 21.599– 26.274
Indonesia	20162/18943.627 18552.095– 19335.159	20796/19348.965 18957.433– 19740.496	21745/19751.271 19359.739– 20142.803	22271/20150.336 19758.803– 20541.867
Laos	19/19.165 17.585– 20.7443	19/19.165 17.585– 20.744	19/19.165 17.585– 20.744	19/19.165 17.585– 20.744
Malaysia	7059/6974.608 6712.536– 7236.679	7137/6994.922 6732.851– 7256.993	7185/7014.123 6752.051– 7276.194	7245/7032.269 6770.198– 7294.341
Myanmar	199/181.501 170.296– 192.705	199/181.629 170.425– 192.834	201/181.739 170.535– 192.944	201/181.833 170.628– 193.037
Philippines	13434/12920.558 12401.611– 13439.506	13597/13071.460 12552.512– 13590.408	13777/13218.318 12699.369– 13737.266	14035/13361.177 12842.228– 13880.125
Singapore	29812/28570.642 26312.718– 30828.566	30426/28800.533 26542.608– 31058.456	31068/29007.838 26749.914– 31265.761	31616/29194.478 26936.554– 31452.402
Thailand	3037/2960.759 2834.881– 3086.636	3037/2960.849 2834.971– 3086.726	3040/2960.925 2835.046– 3086.802	3040/2960.989 2835.111– 3086.867
Vietnam	324/286.887 261.628– 312.145	324/286.901 261.642– 312.159	324/286.91 261.655– 312.172	325/286.925 261.666– 312.183

Note: Actual/estimate and LCI-UCI.

To sum up, three countries should carefully monitor the situation of COVID-19 spreading are Indonesia, Singapore, and Philippines.

#### 4 Conclusion

This research studies the forecasting models to estimate the number of cumulative COVID-19 cases in Southeast Asian countries. The forecasting models—logistic growth curve which symmetric and Gompertz growth curve which is asymmetric—are adopted to estimate the COVID-19 cases in Southeast Asian countries. The suitable forecasting model, logistic or Gompertz growth curve, is selected by the maximum of Coefficient of Determination and the minimum of Root Mean Squared Percentage Error. The spreading based on estimate derivatives of COVID-19 cases is compared among the countries in Southeast Asia. Furthermore, the forecasted number of cumulative COVID-19 cases and its confidence interval are also demonstrated. The findings showed that the suitable forecasting model for Indonesia, Philippines, and Malaysia is Gompertz growth curve which is asymmetric. These countries are in the severe spreading group in Southeast Asia. On the other hand, the number of cumulative infectious people of the other countries is estimated by logistic growth curve, which is a suitable forecasting model and is symmetric. The countries in case of logistic growth curve model except Singapore are the slight spreading group in Southeast Asia. The results imply that the forecasting model which is symmetric such a logistic growth curve model can help the authorities' decision to control the spreading COVID-19 easier than case of asymmetric such a Gompertz growth curve. That is, logistic growth curve will start to decline for the half distance or symmetry property of logistic growth curve slow after peak time. When it comes spreading in Singapore, the number of cumulative infectious people in Singapore is estimated by logistic growth curve which is symmetric. After peak time, the number of cases in Singapore will slowly decline by the symmetry property of logistic growth curve. Under comparison of the derivative amongst all countries in Southeast Asia, Singapore, Indonesia, Philippines, and Malaysia are the countries in which COVID-19 spreading is quite severe. Singapore has passed the peak time for approximately 25 days. Philippines has passed the peak time for approximately 35 days. Malaysia has passed the peak time for approximately 55 days. However, Indonesia has just passed the peak time for approximately 15 days. Also, the duration of approaching peak time of Indonesia is longer than the other countries in Southeast Asia. It implies that Singapore, Indonesia, Philippines, and Malaysia should carefully monitor COVID-19 spreading situation. Furthermore, the forecasted infectious cases in the future of all countries in Southeast Asia for  $t=96$  to  $t=99$  are in 95% of confidence interval.

However, the external factors can increase the number of cases in each country because some cases infected from outside country and can be detected by the authorities at the airport. The forecasting models do not include such the cases and it based on the past information to forecast the future. The future research should be focused on adding the exogenous variable or external variable to the models and establish the better forecasting models. Also, the active infected cases should be focused for estimation because these cases directly relate to the medical planning and controlling the COVID-19 outbreak such as the plan of the number of hospital bed and ventilator, the need for personal protective equipment, etc.

**Acknowledgement:** The authors have to thank Pathumwan Institute of Technology and King Mongkut's University of Technology North Bangkok for support and encouragement during the writing of this research.

**Funding Statement:** The research was funding by King Mongkut's University of Technology North Bangkok Contract No. KMUTNB-61-GOV-03-23.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

1. Harapan, H., Naoya, I., Amanda, Y., Wira, W., Syntat, K. et al. (2020). Coronavirus disease 2019 (COVID-19): A literature review. *Journal of Infection and Public Health*, *13*(5), 667–673. DOI 10.1016/j.jiph.2020.03.019.
2. Li, H., Liu, S. M., Yu, X. H., Tang, S. L., Tang, C. K. (2020). Coronavirus disease 2019 (COVID-19): Current status and future perspectives. *International Journal of Antimicrobial Agents*, *55*(5), 1–8.
3. Camille, B. (2020). The Disproportionate Effect of COVID-19 on migrant workers in ASEAN. The Diplomat <https://thediplomat.com/2020/05/the-disproportionate-effect-of-covid-19-on-migrant-workers-in-asean>.
4. Deshwal, V. K. (2020). COVID-19: A comparative study of Asian, European, American continent. *International Journal of Scientific Research and Engineering Development*, *3*(2), 436–440.
5. Djalante, R., Lassa, J., Nurhidayah, L., Minh, H., Mahendradhata, Y. et al. (2020). The ASEAN's responses to COVID-19: A policy sciences analysis. *SSRN Electronic Journal*. DOI 10.2139/ssrn.3595012.
6. Kucharavy, D., Roland, D. G. (2015). Application of logistic growth curve. *Procedia Engineering*, *131*(2015), 280–290. DOI 10.1016/j.proeng.2015.12.390.
7. Rosalind, J. A., Bartłomiej, W. (2019). Bacterial growth: A statistical physicist's guide. *Reports on Progress in Physics*, *82*(1), 1–55.
8. Nelder, J. A. (1961). The fitting of a generalization of the logistic curve. *Biometrics*, *17*(1), 89–110. DOI 10.2307/2527498.
9. Kathleen, M. C. T., Even, T. (2017). The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the unified-Richards family. *PLoS One*, *12*(6), 1–17.
10. Teleken, J., Robazza, W., Gomes, G. (2011). Mathematical modeling of microbial growth in milk. *Food Science and Technology (Campinas)*, *31*(4), 891–896. DOI 10.1590/S0101-20612011000400010.
11. Zwietering, M. H., Jongenburger, I., Rombouts, F. M., Van't Rie, K. (1990). Modeling of the bacterial growth curve. *Applied and Environment Microbiology*, *56*(6), 1975–1881. DOI 10.1128/AEM.56.6.1875-1881.1990.
12. Laird, A. K. (1964). Dynamics of tumor growth. *British Journal of Cancer*, *13*(3), 490–502. DOI 10.1038/bjc.1964.55.
13. Oliver, F. R. (1964). Methods of estimating the logistic growth function. *Journal of the Royal Statistical Society*, *13*(1), 57–66.
14. Ayiomamitis, A. (1986). Logistic curve fitting and parameter estimation using nonlinear noniterative least-squares regression analysis. *Computers and Biomedical Research*, *19*(2), 142–150. DOI 10.1016/0010-4809(86)90012-1.
15. Jukić, D., Kralik, G., Scitovski, R. (2004). Least-squares fitting Gompertz curve. *Journal of Computational and Applied Mathematics*, *169*(2), 359–375. DOI 10.1016/j.cam.2003.12.030.
16. Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A. et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. DOI 10.1038/s41586-020-2405-7.
17. Lin, Q., Zhao, S., Gao, D., Lou, Y., Yang, S. et al. (2020). A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International Journal of Infectious Diseases*, *93*, 211–216. DOI 10.1016/j.ijid.2020.02.058.

18. Khrapov, P., Loginova, A. (2020). Mathematical modelling of the dynamics of the corona virus COVID-19 epidemic development in China. *International Journal of Open Information Technologies*, 8(4), 13–16.
19. Sahoo, B. K., Sapra, B. K. (2020). A data driven epidemic model to analyse the lockdown effect and predict the course of COVID-19 progress in India. *Chaos, Solitons & Fractals*, 139, 110034. DOI 10.1016/j.chaos.2020.110034.
20. Fanelli, D., Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, 134, 109761. DOI 10.1016/j.chaos.2020.109761.
21. Tang, B., Bragazzi, N. L., Li, Q., Tang, S., Xiao, Y. et al. (2020). An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). *Infectious Disease Modelling*, 5, 248–255. DOI 10.1016/j.idm.2020.02.001.
22. Worldometer 2020. <https://www.worldometers.info/coronavirus/country>.
23. Verhulst, P. F. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondence Mathematique et Physique*, 1838(10), 113–121.
24. Martha, L. A., James, P. B. (2018). *Introductory differential equations*, 5th edition. United Kingdom: Academic Press.
25. Pearl, R., Reed, L. J. (1923). On the mathematical theory of population growth. *Metron*, 3, 6–19.