SEHS
science, engineering and health studies

Research Article

Science, Engineering and Health Studies
https://li01.tci-thaijo.org/index.php/sehs
ISSN (Online): 2630-0087

Forecasting modeling of the number of cumulative COVID-19 cases with deaths and recoveries removal in Thailand

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

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

***Corresponding author**:
Rapin Sunthornwat
rapin@pit.ac.th

Received: *16 September 2020*
Revised: *28 November 2020*
Accepted: *29 November 2020*
Published: *2 June 2021*

Citation:
Areepong, Y., and Sunthornwat, R. (2021). Forecasting modeling of the number of cumulative COVID-19 cases with deaths and recoveries removing in Thailand. Science, Engineering and Health Studies, 15, 21020004.

ABSTRACT

Coronavirus disease 2019 (COVID-19) could become one of the problems for the healthcare system in Thailand, thus forecasting the number of cumulative cases could be helpful to mitigate this. The present research was conducted to forecast the number of cumulative COVID-19 cases in Thailand and analyze any trends in the data. The predictive modeling was based on logistic models for cumulative COVID-19 cases with and without removing the deaths and recoveries data. Confidence intervals for and validation of the forecasting models were also given. The results showed that the logistic model performed better than the logistic model after deaths and recoveries removal but analysis of the behavior of the number of cumulative COVID-19 cases in Thailand by using the latter model was still pertinent. This research provided a predictive modeling tool to help the authorities developing policies for controlling the spread of COVID-19 in Thailand.

Keywords: coronavirus disease 2019; biostatistics; logistic with removing model; confidence interval; stability analysis

1. INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a new infectious disease of the respiratory system that emerged at the end of 2019. The virus that causes this disease, SARS-CoV-2, could have crossed the species barrier by human interaction with bats, musk, badgers, echidna, among others. This outbreak has now spread to all countries in the world, including Thailand. Attempts by the Thai government to control the spread of COVID-19 has caused many problems such as economic regression, health, and unemployment. Actions for controlling the spread of COVID-19, such as social distancing, the closure of academic institutions, etc., have been enforced to flatten the curve of COVID-19 cases.

Forecasting the number of cumulative COVID-19 cases could be a major factor for developing the policies to control the spread of COVID-19. For example, acquiring adequate personal protective equipment based on the number of COVID-19 cases is necessary to adequately equip healthcare workers who are at the front line in tackling the disease.

Several researchers have developed models to forecast the number of COVID-19 cases, and classical forecasting, machine learning, and mechanistic models of the cumulative confirmed, deaths, and recovered cases of COVID-19 have been suggested. However, predicting the continuous increase in confirmed COVID-19 cases is still uncertain. Research has been conducted to help authorities plan and make decisions on the best course of action in the COVID-19 pandemic (Petropoulos and Makridakis, 2020; Shinde et al., 2020; Liu et al., 2020). A simple iteration method has been applied to estimate the growth in the number of COVID-19 cases while accounting for the expected recoveries and deaths to make the trend of the COVID-19 outbreak more comprehensible. The results showed that when the daily growth rate is less than 5%, then the number of cases in the COVID-19 outbreak will plateau (Perc et al., 2020). Research has been conducted on forecasting models for the spread of COVID-19 based on three regional-scale models: exponential growth, a self-exciting branching process, and compartmental models for susceptible-infected-resistant (SIR) and susceptible-infectious-recovered-dead (SIRD) populations. The findings revealed that these forecasting models can be applied in simulations to determine whether social distancing or other measures will slow down the spread of COVID-19 (Bertozzia et al., 2020; Anastassopoulou et al., 2020). Research into simulating the spread of COVID-19 in Hubei was conducted to forecast the average number of reinfections, incubation period, and the average number of recoveries per day (Li et al., 2020), and was extended to analyze the situations in Korea, Italy, and Iran. Forecasting confirmed cases, deaths, and recoveries from COVID-19 in China during the early stage was conducted by applying the Fractional Grey model (Liu et al., 2020). Forecasting the cumulative number of COVID-19 deaths based on the Boltzmann function was estimated in Hubei and Wuhan city (Gao et al., 2020). Time series models such as auto-regressive integrated moving average (ARIMA), Holt-Winters additive, trigonometric seasonal formulation, the Box-Cox transformation as well as machine learning have been adopted for forecasting the percentage of COVID-19 active cases per capita (Papastefanopoulos et al., 2020).

The aim of the present research is to forecast the number of cumulative COVID-19 cases in Thailand based on logistic models before and after removing the deaths and recoveries.

2. MATERIALS AND METHODS

The mathematical and statistical backgrounds for this research (Edwards and Penney, 2008) were presented. The assumptions for this research are based on a closed system or equivalence by using the lockdown policy, and strong state quarantine policy. The cumulative COVID-19 cases in Thailand presented as an S-shaped (sigmoidal) curve. Thus, forecasting models for the number of cumulative COVID-19 cases are based on two models presenting S-shaped curves as follows.

2.1 Data collection

The daily number of cumulative COVID-19 cases in Thailand comprised the data for this research and were collected from Worldometer (2020). The in-sample data were from January 13, 2020, for the first case to August 7, 2020, while the out-of-sample data for measuring the accuracy and validation of the forecasting models were from August 8-24, 2020.

2.2 Forecasting model 1: the logistic model for all COVID-19 cases in Thailand

Let $x(t)$ be the number of cumulative COVID-19 cases at any time (t). P denotes the maximum cases or carrying capacity. The logistic model for forecasting model 1 can be defined by applying a logistic differential equation with initial cases x_0 as follows:

$$\frac{dx}{dt} = kx(P-x) \quad (1),$$

where k is the growth rate. The solution of logistic differential equation is called logistic model, which can bederived by separable method as follows:

$$\int \frac{dx}{x(P-x)} = \int kdt$$

$$\int \left[\frac{1/P}{x} - \frac{1/P}{x-P}\right] dx = \int kdt$$

$$ln\left[\frac{|x|^{1/P}}{|x-P|^{1/P}}\right] = kt + A; A \text{ is an arbitrary constant.}$$

Taking exponential function both side above equation, the result becomes $\frac{x}{x-P} = w\exp(kPt)\,; w = \pm\exp(AP)$.

$$x = \frac{P}{1-\frac{1}{w}\exp(-kPt)} \quad (2)$$

with initial condition of the cases $t = 0$ and $x = x_0$, then $\frac{1}{w} = \frac{x_0-P}{x_0}$. Thus, the logistic model is

$$x(t) = \frac{P}{1-\left(\frac{x_0-P}{x_0}\right)\exp(-kPt)} \quad (3)$$

where P and k are unknown parameters of logistic model.

2.3 Forecasting model 2: the logistic model for COVID-19 cases after removing deaths and recoveries

In forecasting model 2, the recoveries and deaths are removed from the forecasting model because these cases cannot further transmit the SARS-CoV2 virus. Thus, the number of cumulative cases will be decreased to flatten the S-shaped curve of the forecasting model. The logistic model after removing deaths and recoveries can be defined with initial cases x_0 as follows:

$$\frac{dx}{dt} = kx(P-x) - R \quad (4),$$

where R is removing parameter as the recoveries and deaths, k is the growth rate, P is the carrying capacity.

$$\frac{dx}{dt} = kxP - kx^2 - R$$

$$\frac{dx}{dt} = k(N-x)(x-T) \quad (5),$$

where $T = \frac{kP-\sqrt{(kP)^2-4Rk}}{2k}$ for $(kP)^2 - 4Rk \geq 0$ and $N = \frac{kP+\sqrt{(kP)^2-4Rk}}{2k}$ for $(kP)^2 - 4Rk \geq 0$.

Using separable method, Eq. 5 can be carried out as follows:

$$\frac{dx}{(x-N)(x-T)} = -kdt$$
$$\int \left[\frac{1/(N-T)}{x-N} - \frac{1/(N-T)}{x-T}\right] dx = -\int kdt$$
$$ln\left(\left|\frac{x-N}{x-T}\right|^{\frac{1}{N-T}}\right) = kt + A; A \text{ is an arbitrary constant.}$$

Taking exponential function to above equation, the result becomes

$$\frac{x-N}{x-T} = wexp(-k(N-T)t); w = \pm exp(A(N\text{-}T))$$

$$x(t) = \frac{N-Twexp(-k(N-T)t)}{1-wexp(-k(N-T)t)} \quad (6)$$

with initial cases $t = 0$ and $x = x_0$, then $w = \frac{x_0-N}{x_0-T}$.
Thus, the logistic model with removing the deaths and recoveries can be written as follows:

$$x(t) = \frac{N-T\left(\frac{x_0-N}{x_0-T}\right)\exp(-k(N-T)t)}{1-\left(\frac{x_0-N}{x_0-T}\right)\exp(-k(N-T)t)} \quad (7)$$

where N, T, and k are unknown parameters of logistic model with removing the deaths and recoveries.

2.4 Analysis of the forecasting models

This is based on a stability analysis of the critical points. Differential equation $\frac{dx}{dt} = f(x)$ or the forecasting model is referred to as an autonomous differential equation of the first degree. The definitions and asymptotical behavior of the forecasting models are given by the following equations.

Definition 1: For some c such that $f(x = c) = 0$, then c is called a critical point or an equilibrium point.

Definition 2: Equilibrium point $x = c$ is said to be stable if x_0 sufficiently close to c, then $x(t)$ remains close to c for all $t > 0$. In other words, the critical point $x = c$ is stable if, for each $\varepsilon > 0$, there exists $\delta > 0$ such that

$|x_0 - c| < \delta$ implies that $|x(t) - c| < \varepsilon$
for all $t > 0$.

For logistic model, equilibrium points can be carried out $\frac{dx}{dt} = kx(P-x) = 0$. Obviously, there are two equilibrium points: $x = 0$ and $x = P$. Asymptotic behavior of logistic model, when t approaches ∞, then $x(t)$ approaches P in case of $x_0 > P$ or $0 < x_0 < P$. Namely, $x = P$ is stable but $x = 0$ is unstable. For logistic model with removing the deaths and recoveries, equilibrium points can be carried out $\frac{dx}{dt} = k(N-x)(x-T) = 0$. Obviously, there are two equilibrium points: $x = N$ called new limiting case and $x = T$ called threshold case for $N > T$. Asymptotic behavior of logistic model with removing the deaths and recoveries, if $x_0 > T$ then $x(t)$ approaches N. By contrast, if $x_0 < T$ then $x(t)$ approaches 0. Namely, $x = N$ is stable but $x = T$ is unstable.

To apply the forecasting model for estimating the number of cumulative COVID-19, inflection point for the maximum spreading of COVID-19 is analyzed. The minimum number of cumulative COVID-19 cases can be removed and derived in order to reduce the number of cumulative COVID-19 cases to zero. In addition, to break off the number of cumulative COVID-19 cases with size at M for reducing infectious people and controlling of COVID-19 spreading, the maximum remained (initial) size of cumulative COVID-19 cases in the system will be permitted and analyzed this situation.

The inflection of logistic model can be derived as

$$\frac{d^2x}{dt^2} = \frac{d}{dt}\left[\frac{dx}{dt}\right]$$
$$= \frac{d}{dt}[kx(P-x)]$$
$$= x\frac{d}{dt}[kP-kx] + (kP-kx)\frac{dx}{dt}$$
$$= 2k^2x\left[\frac{P^2}{2} - \frac{3P}{2}x + x^2\right]$$
$$= 2k^2x(x-P)\left(x-\frac{P}{2}\right)$$

The inflection point occurs at the second derivative to be zero.

$$2k^2x(x-P)\left(x-\frac{P}{2}\right) = 0$$
$$x = 0, x = P, \text{ and } x = \frac{P}{2}$$

By the testing of inflection point based on calculus, $x = 0$ and $x = P$ are not satisfied with inflection point testing but $x - \frac{P}{2}$ is only satisfied with inflection point testing. Thus, $x_l - \frac{P}{2}$ is inflection point of logistic model which returns the maximum spreading point of COVID-19. Also, the time estimated by logistic model for the maximum spreading of COVID-19 corresponds to

$$t_l = \frac{1}{kP}\ln\left[\left(\frac{x_l-P}{x_0-P}\right)\frac{x_0}{x_l}\right] \quad (8)$$

The inflection of logistic model with removing the deaths and recoveries can be derived as follows:

$$\frac{d^2x}{dt^2} = \frac{d}{dt}\left[\frac{dx}{dt}\right]$$
$$= \frac{d}{dt}[kx(P-x) - R]$$
$$= (kP - 2xk)\frac{dx}{dt}$$
$$= (kP-2xk)[kx(P-x) - R]$$
$$= -kPR + (k^2P^2 + 2kR)x - (k^2P + 2k^2P)x^2 + 2k^2x^3$$

The inflection point occurs at the second derivative to be zero.

$$-kPR + (k^2P^2 + 2kR)x - (k^2P + 2k^2P)x^2 + 2k^2x^3 = 0 \quad (9)$$

The roots of Eq. 9 are carried out with analytical method, the numerical method is applied for the inflection point x_r of logistic model with removing (Kharab and Guenther, 2011). The time estimated by logistic model with removing the deaths and recoveries for the maximum spreading of COVID-19 corresponds to

$$t_r = \frac{1}{k(N-T)}\ln\left[\left(\frac{x_r-N}{x_r-T}\right)\left(\frac{x_0-T}{x_0-N}\right)\right] \quad (10)$$

2.5 Parameter estimation for the proposed forecasting models

Let t_b be the nearest time of predicting time $t_{predict}$. Also, $x(t_b)$ denoted by x_b is the nearest value before predicting value $x(t_{predict})$.

For model 1 (logistic model), Eq. 3 becomes

$$x_b = \frac{P}{1 - \left(\frac{x_0 - P}{x_0}\right) exp(-kPt_b)}$$

$$(x_0 - P)exp(-kPt_b) = \frac{x_0}{x_b}(x_0 - P)$$

$$exp(-kPt_b) = \frac{x_0}{x_b}\left(\frac{x_b - P}{x_0 - P}\right)$$

By taking natural logarithm, the result becomes

$$-kP = \frac{1}{t_b}\ln\left[\frac{x_0}{x_b}\left(\frac{x_b-P}{x_0-P}\right)\right] \quad (11)$$

Thus, the forecasting model for logistic model with predicting time $t_{predict}$ is

$$x(t_{predict}) = \frac{P}{1-\left(\frac{x_0-P}{x_0}\right)exp\left(\frac{t_{predict}}{t_b}\ln\left[\frac{x_0}{x_b}\left(\frac{x_b-P}{x_0-P}\right)\right]\right)} \quad (12)$$

For model 2 (logistic model with removing the deaths and recoveries), Eq. 7 becomes

$$x_b = \frac{N - T\left(\frac{x_0 - N}{x_0 - T}\right) exp(-k(N-T)t_b)}{1 - \left(\frac{x_0 - N}{x_0 - T}\right) exp(-k(N-T)t_b)}$$

$$x_b - x_b\left(\frac{x_0 - N}{x_0 - T}\right) exp(-k(N-T)t_b) =$$

$$N - T\left(\frac{x_0 - N}{x_0 - T}\right) exp(-k(N-T)t_b)$$

$$exp(-k(N-T)t_b) = \left(\frac{x_b - N}{x_b - T}\right)\left(\frac{x_0 - N}{x_0 - T}\right)$$

By taking natural logarithm, the result becomes

$$-k(N-T) = \frac{1}{t_b}\ln\left[\left(\frac{x_b-N}{x_b-T}\right)\left(\frac{x_0-N}{x_0-T}\right)\right] \quad (13)$$

Thus, the forecasting model for logistic model with removing the deaths and recoveries with predicting time $t_{predict}$ is as follows:

$$x(t_{predict}) = \frac{N-T\left(\frac{x_0-N}{x_0-T}\right)exp\left(\frac{t_{predict}}{t_b}\ln\left[\left(\frac{x_b-N}{x_b-T}\right)\left(\frac{x_0-N}{x_0-T}\right)\right]\right)}{1-\left(\frac{x_0-N}{x_0-T}\right)exp\left(\frac{t_{predict}}{t_b}\ln\left[\left(\frac{x_b-N}{x_b-T}\right)\left(\frac{x_0-N}{x_0-T}\right)\right]\right)} \quad (14)$$

Let N_0 the maximum remained (initial) size of cumulative COVID-19 cases in the system will be allowed when new limiting case of logistic model with removing the deaths and recoveries is determined as the number of cumulative COVID-19 cases with size at *M*.

$$\frac{kP + \sqrt{(kP)^2 - 4R_{min}k}}{2k} = M$$

$$R_{min} = \frac{k}{4}\left[P^2 - \left(\frac{2M}{P} - P\right)^2\right]$$

Therefore, N_0should be less than T and then the number of cumulative COVID-19 cases will approach zero. That is,

$$N_0 < \frac{kP-\sqrt{(kP)^2-4R_{min}k}}{2k}.$$

2.6 Accuracy and validation of the proposed forecasting models

Accuracy and validation of forecasting are the method for scoring to measure the precision of the model. Root mean square percentage error (RMSPE), coefficient of determination (R^2), and confidence interval (CI) were measured to compare between proposed two forecasting models.

RMSPE is defined as error measurement.

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_{actual}(t_i)-x_{predict}(t_i)}{x_{actual}(t_i)}\right)^2},$$

where $x_{actual}(t_i)$ is the actual number of cumulative COVID-19 cases at time t_i, $x_{predict}(t_i)$ is the predicted number of cumulative COVID-19 cases at time t_i.

Coefficient of determination is defined as the ratio between explained variance in the dependent variable, which can be predicted by the independent variable and unexplained variance. If R^2 approaches 1, then the independent variable can highly explain the dependent variable, i.e.,

$$R^2 = \frac{Explained\ variance}{Total\ variance}; 0 \leq R^2 \leq 1$$

Confidence interval of $100(1-\alpha)\%$ in $x_{predict}(t)$of predicting the predicted number of cumulative COVID-19 cases at time *t*. There are two levels of confidence interval as follow:

$$\text{Upper confidence interval (UCI)} = x_{predict}(t) + t_{\frac{\alpha}{2},df}(SE)$$

$$\text{Lower confidence interval (LCI)} = x_{predict}(t) - t_{\frac{\alpha}{2},df}(SE),$$

where *SE* is the standard error of *x(t)* and *df* is the degree of freedom of *t* statistic.

3. RESULTS AND DISCUSSION

The parameters for this research were *P* = 4,000 COVID-19 cases and *k* = 0.03. The actual number of cumulative COVID-19 cases as of August 8, 2020, in Thailand was 3,348 cases.

3.1 Validation of the forecasting models

The number of cumulative COVID-19 cases in Thailand was estimated and compared between logistic models with and without removing the deaths and recoveries, as shown in Figure 1. The parameter settings were *P* = 4,000 COVID-19 cases and *k* = 0.03. It can be seen that the logistic model for the complete data fit the observations better than the logistic model after removing the deaths and recoveries. The daily number of cases ranged from 20 to 60 until March 15, 2020, after which it was greater than 100 (*t* = 63). Afterward, the number of cases rose sharply to 3,000 on May 8, 2020 (*t* = 117), and then became steady until August 31, 2020.

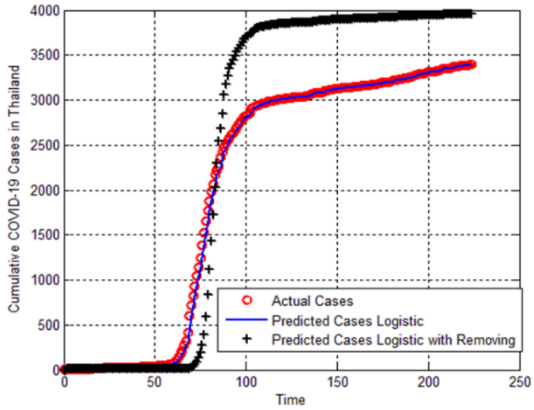

Figure 1. Forecasting model for cumulative COVID-19 cases in Thailand

As reported in Table 1, the predicted number of cumulative COVID-19 cases in Thailand from August 8 (t = 209) to August 24 (t = 224), 2020, and the R^2 and $RMSPE$ were used to measure the model accuracy and validate the models. It was found that the logistic model with complete data returned R^2 = 0.99 and RMSPE = 0.001521, whereas the logistic model after removing the deaths and recoveries returned R^2 = 0.97 and RMSPE = 0.172782.

Table 1. Predicted cumulative COVID-19 cases for out of sample in Thailand

t	$x_{actual}(t)$	**Logistic model**		**Logistic with removing model**	
		$x_{predictl}(t)$	**LCI-UCI**	$x_{predictl}(t)$	**LCL-UCI**
209	3351	3348.01	3343.07-3352.96	3954.43	2700.93-5207.93
210	3351	3351.01	3346.07-3355.96	3954.88	2701.37-5208.38
211	3351	3351.00	3346.06-3355.94	3955.17	2701.67-5208.67
212	3356	3351.00	3346.06-3355.94	3955.17	2701.67-5208.67
213	3359	3356.02	3351.08-3360.97	3955.41	2701.91-5208.92
214	3376	3359.01	3354.07-3363.96	3956.04	2702.54-5209.54
215	3376	3376.08	3371.14-3381.02	3957.13	2703.63-5210.64
216	3377	3376.00	3371.06-3380.94	3958.67	2705.16-5212.17
217	3378	3377.00	3372.06-3381.95	3958.71	2705.21-5212.21
218	3381	3378.00	3373.06-3382.95	3958.84	2705.34-5212.34
219	3382	3381.01	3376.07-3385.96	3959.06	2705.56-5212.56
220	3389	3382.00	3377.06-3386.95	3959.36	2705.86-5212.86
221	3390	3389.03	3384.09-3393.97	3959.75	2706.24-5213.25
222	3390	3390.00	3385.06-3394.95	3960.37	2706.87-5213.88
223	3395	3390.00	3385.06-3394.94	3960.46	2706.95-5213.96
224	3397	3395.02	3390.08-3399.97	3960.66	2707.16-5214.16
R^2		0.99		0.97	
$RMSPE$		0.001521		0.172782	

Note: t = time, LCI = lower confidence interval, and UCI = upper confidence interval

3.2 Equilibrium and stability analysis

The equilibrium and stability analysis were conducted to predict trends in the cumulative COVID-19 cases. Namely, the behavior of the number of cumulative COVID-19 cases will converge or diverge to the equilibrium points as shown in Figures 2-4.

The equilibrium point for the logistic model with complete data was only 4,000 cases and maximum $\frac{dx}{dt}$ occurred at x_l = 2,000. As shown in Figure 2, the equilibrium points for the logistic model after removing deaths and recoveries were T = 19 and N = 3,981, and maximum $\frac{dx}{dt}$ occurred at $x_r \approx$ 2,008. This means that the fastest spread of COVID-19 in Thailand occurred at $x_r \approx$ 2,008 and the peak time for the spread of COVID-19 has already passed.

Figure 3 demonstrates the behavior of the logistic model solutions when asymptotically approaching equilibrium point P = 4,000. As this point is stable, each solution will approach it. Thus, the number of cumulative COVID-19 cases in Thailand will approach 4,000 cases according to the logistic models when applied over a long period.

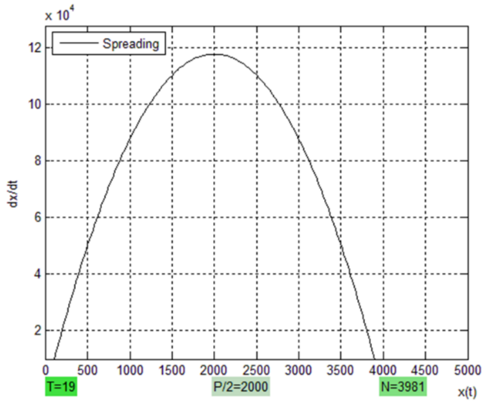

Figure 2. Equilibrium points of cumulative COVID-19 cases on logistic with removing

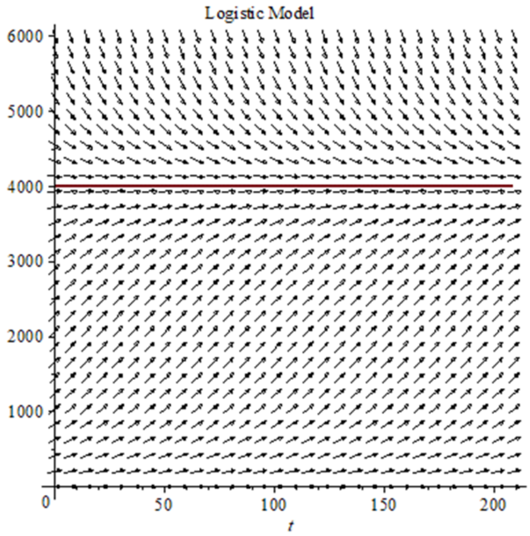

Figure 3. Stability of cumulative COVID-19 cases on logistic

Figure 4 illustrates that the behavior of the solution for the logistic model after removing deaths and recoveries approached equilibrium point *N* = 3,981, which is stable, and diverged at equilibrium point *T* = 19, which is unstable. Thus, the number of cumulative COVID-19 cases in Thailand will approach *N* = 3,981 cases, which is the new equilibrium point differing from that for the logistic model with complete data at 4,000 cases and will diverge at *T* = 19 cases, which is the threshold point. This implied that if the remaining number of cumulative COVID-19 cases in Thailand is between 19 and 3,981, then the cumulative cases will approach the equilibrium point of 3,981. On the contrary, if the remaining number of cumulative cases in Thailand is less than 19 (the threshold point), then the number of cases will approach zero.

The methods in the literature have been used to predict the number of total cases of COVID-19 in the early stage when the spread was exponentially increasing or straight with a slope from approximately January to May, 2020. The existing forecasting models for the number of total COVID-19 cases in the early stage are based on linear modeling. The performance of the Fractional Grey model used to forecast the number of total cases of COVID-19 in China during the early stage had a mean error of 4.66 (Liu et al., 2020). The ARIMA model adopted to forecast the number of total COVID-19 cases in the US, Italy, and Russia had a mean error of 0.004862. The models can be applied to fit the real number of total COVID-19 cases in the early stage over a short period. However, when analyzing over a long period, deaths and recoveries should be removed from the model to represent the real situation, as we did in the current research. The performance of the logistic model without and with the deaths and recoveries data in terms of mean error was 0.001521 and 0.172782, respectively. In a comparison of the existing models and proposed models, the Fractional Grey model had the biggest error and the proposed logistic model with complete data had the smallest mean error. Over a long period, the government's policies such as physical distancing, face mask-wearing, hand washing with alcohol gel, state quarantine when traveling from outside of the country, etc. for controlling the COVID-19 outbreak have been applied to flatten the exponential curve or straight line of the total COVID-19 cases to an S-curve, which is similar to the curve generated by the logistic model after removing the deaths and recoveries. Therefore, this logistic model is more suitable for the real situation than the existing linear models because the number of COVID-19 cases is not likely to increase exponentially or in a straight line.

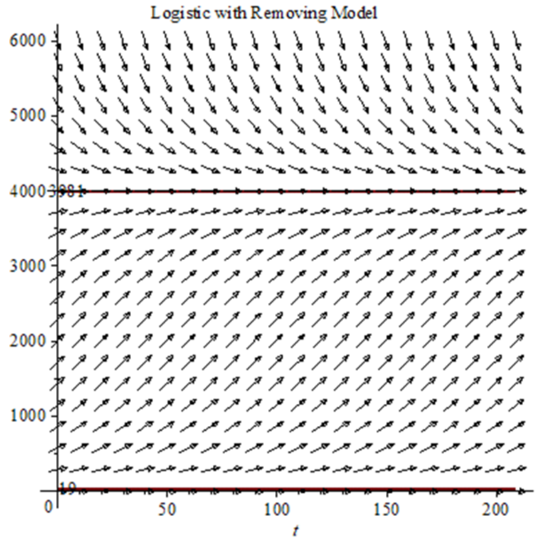

Figure 4. Stability analysis of cumulative COVID-19 cases on logistic with removing the deaths and recoveries

4. CONCLUSION

Forecasting models for predicting the number of cumulative COVID-19 cases in Thailand were proposed in this research. These are based on logistic model with and without deaths and recoveries, which were compared via a stability analysis. The results showed that the logistic model with complete data performed better than the logistic model after removing deaths and recoveries based on a lower *RMSPE* and a higher R^2. It is possible that the logistic model after removing deaths and recoveries has many parameters estimation and estimating with it is complex, whereas the logistic model with complete data has parameters that are easy to estimate and the estimating process with the model is not complex. For the first wave, the maximum spread of COVID-19 in Thailand (around 2,000 cases) was correctly estimated by both forecasting models. This means that Thailand has passed the peak point of the COVID-19 outbreak. Moreover, if the number of cumulative COVID-19 cases in Thailand remains between 19 and 3,981 cases, the COVID-19 outbreak is still ongoing and the number of cumulative COVID-19 cases will approach the carrying capacity of 4,000 cases. On the other hand, the number of cumulative COVID-19 cases in Thailand will converge to zero if the remained number of cumulative COVID-19 is less than 19 cases.

Future research should be focused on approximating the maximum number of tourists that can enter Thailand while controlling the number of COVID-19 cases to avoid further outbreaks. The application of both forecasting models to predict and analyze the COVID-19 spread in other countries would also be interesting to research.

REFERENCES

Anastassopoulou, C., Russo, L., Tsakris, A., and Siettos, C. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE,* 15(3), 1-21.

Bertozzia, A. L., Franco, E., Mohler, G., Short, M. B., and Sledge, D. (2020). The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America,* 117(29), 16732-16738.

Edwards, C. H., and Penney, D. E. (2008). *Elementary Differential Equations with Boundary Value Problems,* 6th, New Jersey: Pearson Education, Inc., pp. 480-553.

Gao, Y., Zhang, Z., Yao, W., Ying, Q., Long, C., and Fu, X. (2020). Forecasting the cumulative number of COVID-19 deaths in China: a Boltzmann function-based modeling study. *Infection Control and Hospital Epidemiology,* 41(7), 841-843.

Kharab, A., and Guenther, R. B. (2011). *An Introduction to Numerical methods: A MATLAB Approach*, 3rd, USA: CRC Press., pp. 39-82.

Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., Wang, D., Chen, G., Zhang, J., Peng, H., and Shao, Y. (2020). Propagation analysis and prediction of the COVID-19. *Infectious Disease Modelling,* 5, 282-292.

Liu, D., Clemente, L., Poirier, C., Ding, X., Chinazzi, M., Davis, J., Vespignani, A., and Santillana, M. (2020). Real-time forecasting of the COVID-19 outbreak in Chinese provinces: machine learning approach using novel digital data and estimates from mechanistic models. *Journal of Medical Internet Research,* 22(8), e20285.

Liu, L., Chen, Y., and Wu, L. (2020). Forecasting confirmed cases, deaths, and recoveries from COVID-19 in China during the early stage. *Mathematical Problems in Engineering,* 2020, 1-4.

Papastefanopoulos, V., Linardatos, P., and Kotsiantis, S. (2020). COVID-19: a comparison of time series methods to forecast percentage of active cases per population. *Applied Science*, 10(11), 3880.

Perc, M., Miksić, N., Slavinec, M., and Stozer, A. (2020). Forecasting COVID-19. *Frontiers in Physics,* 8, 127.

Petropoulos, F., and Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS ONE,* 15(3), e0231236.

Shinde, G. R., Kalamkar, A. B., Mahalle, P. N., Dey, N., Chaki, J., and Hassanien, A. E. (2020). Forecasting models for Coronavirus Disease (COVID-19): a survey of the state-of-the-art. *Springer Nature Computer Science,* 1, 1-15.

Worldometer. (2020). Reported cases and deaths by country, territory, or conveyance [Online URL: https://www.worldometers.info/coronavirus/country] accessed on August 31, 2020.