

Real-Time Prediction of the COVID-19 Epidemic in Thailand using Simple Model-Free Method and Time Series Regression Model

Rati WONGSATHAN

*Department of Electrical Engineering, Faculty of Engineering and Technology,
North-Chiang Mai University, Chiang Mai 50230, Thailand*

(Corresponding author's e-mail: rati@northcm.ac.th, rati1003@gmail.com)

Received: 28 April 2020, Revised: 22 March 2021, Accepted: 5 April 2021

Abstract

The novel coronavirus 2019 (COVID-19) pandemic was declared a global health crisis. The real-time accurate and predictive model of the number of infected cases could help inform the government of providing medical assistance and public health decision-making. This work is to model the ongoing COVID-19 spread in Thailand during the 1st and 2nd phases of the pandemic using the simple but powerful method based on the model-free and time series regression models. By employing the curve fitting, the model-free method using the logistic function, hyperbolic tangent function, and Gaussian function was applied to predict the number of newly infected patients and accumulate the total number of cases, including peak and viral cessation (ending) date. Alternatively, with a significant time-lag of historical data input, the regression model predicts those parameters from 1-day-ahead to 1-month-ahead. To obtain optimal prediction models, the parameters of the model-free method are fine-tuned through the genetic algorithm, whereas the generalized least squares update the parameters of the regression model. Assuming the future trend continues to follow the past pattern, the expected total number of patients is approximately 2,689 - 3,000 cases. The estimated viral cessation dates are May 2, 2020 (using Gaussian function), May 4, 2020 (using a hyperbolic function), and June 5, 2020 (using a logistic function), whereas the peak time occurred on April 5, 2020. Moreover, the model-free method performs well for long-term prediction, whereas the regression model is suitable for short-term prediction. Furthermore, the performances of the regression models yield a highly accurate forecast with lower RMSE and higher R^2 up to 1-week-ahead.

Keywords: COVID-19, Predictive model, Model-free method, Time series regression, Genetic algorithm

Introduction

A novel coronavirus 2019, known as COVID-19, which was originated in Wuhan, China, in December 2019, has spread rapidly, resulting in around 1.6 million infected people and over 95,000 deaths worldwide (as reported on April 11, 2020). Now, it has been declared a pandemic, which has adversely affected human health. Scientists are still creating an effective vaccine against this disease. The COVID-19 entered Thailand on January 12, 2020, with its first confirmed case (South China Morning Post, January 13, 2020). The situation is getting worse with the increasing infected rates, which has become serious due to a super-spreader event, the boxing match in a stadium in Bangkok (the capital city) on March 6, 2020. On April 7, 2020, the cases rose after some workers in Bangkok returned to their provinces after the government declared a curfew.

Since the severe acute respiratory syndrome (SARS) (2003) and the Middle East respiratory syndrome (MERS) (2012), including COVID-19, are also coronaviruses, the infection distribution and the subsequent model for the dynamic spread could be similar. In various research [1-3], the transmission

models of these epidemics have been proposed to estimate the basic reproduction number. These models represent the expected number of patients directly generated by 1 case. Nonetheless, this parameter is only coarsely quantitative to assess the epidemic. In addition, Kermack and McKendrick have proposed the well-known susceptible-infectious-recovered (SIR) model [4] based on medical information. Recently, the susceptible-exposed-infectious-recovered (SEIR) model has been proposed to predict the number of COVID-19-patients [5]. However, these mathematical-based models, dealing with the ordinary differential equations-driven method and many idealistic assumptions, are not realistic and give inconsistent predictions. To address the problems, the predictive models, ranging from linear and simple to nonlinear and complex models, succeed in applying to predict the spread of SARS and MERS, such as [1-9], are developed for COVID-19 [10-12].

To avoid the difficulties while getting in time for predicting the ongoing spread of the COVID-19, this study applied two approaches: a model-free method and a multiple linear regression (MLR) model (1 kind of machine-learning) due to their simplicity but acceptable accuracy. The model-free method is based on three explicit functions likely to be the results in epidemiology, namely the Gaussian function, hyperbolic-tangent function, and logistic function. Whereas the MLR model is based on the assumption that the future spread of infections depends on the number of infected people, some past data are well-correlated and can be used to predict future data through statistical regression. In literature, for example, the Gaussian function has been used to model the spreading prediction epidemics [13]. Moreover, in [14-16], the dynamic accumulations of the SARS patients have been fitted using the widely used Richard function among the growth-curve types. However, having many parameters, Richard's function is a disadvantage. Recently, the logistic function has been employed to forecast short-term daily accumulated COVID-19-infection cases [16]. Besides, the MLR model is used to predict the newly confirmed COVID-19-patients [17]. From a review of the previous works, there is still no research for the prediction model of the COVID-19 epidemic in Thailand. During the outbreak, lacking a reliable prediction model leads to the inability to plan for short/long-term and control the infectious transmission. Apart from the right and practical measures, theoretical and quantitative analysis based on the predictive model is needed to predict the epidemic trend. The overall concept of this work is illustrated in **Figure 1**. Both the model-free method and time series MLR model online perform predictions by projecting current and future trends using live reported data. For the model-free method, the Gaussian function describes the characteristics of the newly infected patients and estimates the cessation date. While the logistic and hyperbolic-tangent functions capture the epidemic growth rate and estimate the cessation date, including the total patients. Due to the numerical difficulties in solving parameters and selecting proper initial parameters used in previous research, this work is fine-tuned through the genetic algorithm (GA)-an evolutionary search algorithm (DE Goldberg, 1989). For the regression models that treat some past values as the predictors, they respond to predict the trend of the newly infected patients and accumulations for a 1-day, 1-week, and 1-month ahead. Their parameters are updated by the generalized least square (GLS) method.

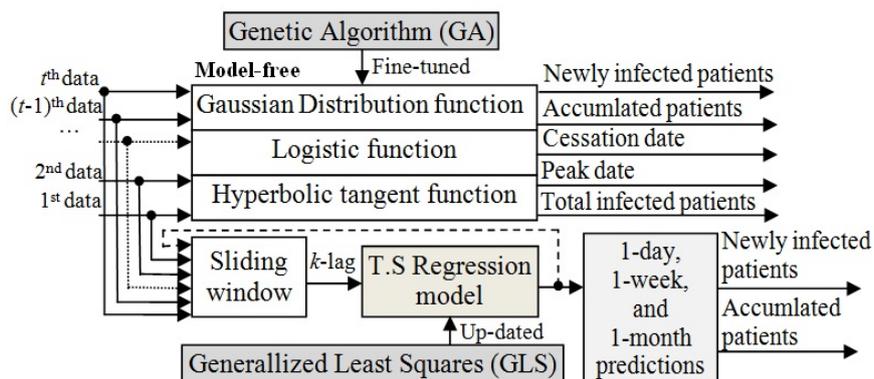


Figure 1 The overall framework for modeling COVID-19 prediction.

Materials and methods

To make the COVID-19 prediction model happen in a timely manner, the newly infected patients and accumulated patient’s data are made openly available online, publicly by the department of disease control of Thailand through the official website <https://covid19.th-stat.com/th/api>. The data during January 1, 2020 to April 17, 2020, were used in training the proposed predictive models. The trends are depicted in **Figure 2** by onset of infection. The characteristics of the newly infected patients followed the bell shape (**Figure 2(a)**). It is seen that after detecting the 1st confirmed patient on January 12, 2020, the viral spread appeared very slow for a while. When the 1st super-spreader occurred, the infected rate sharply reached the highest peak on March 22, 2020, and gradually declined after the curfew on April 3, 2020. It can be suggested that this measure can highly restrict the spread of COVID-19. The spreading infection may explode unless the measures are not taken. Besides, the epidemic growth rate follows an S-shaped curve (**Figure 2(b)**) where the 1st phase of the epidemic, characterized by the slow slope, continues to the 2nd phase of the epidemic, with a sharply increasing slope, on January 16, 2020.

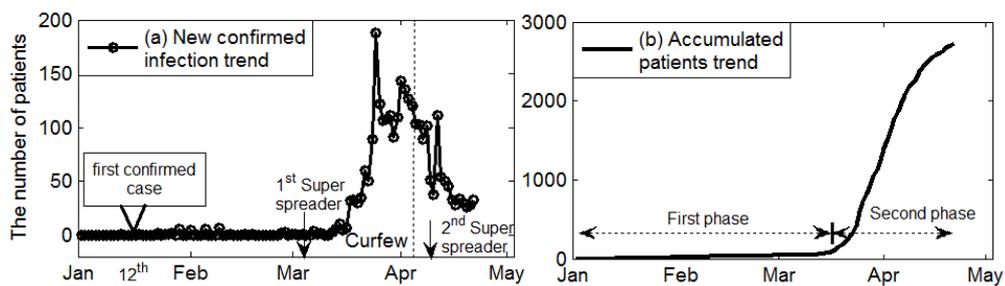


Figure 2 The trend of the number of new COVID-19 patients, and accumulations of Thailand.

By assuming the entire pattern of trends can be mapped out, the model-free method and time series regression model used in this study detail as follows. For the model-free method, considering the trend of the newly infected patients (**Figure 2(a)**), it naturally increases at the initial and decreases after reaching the peak with some fluctuations. Here, the probability of newly infected patients is assumed to follow the Gaussian function, which is expressed as,

$$f_G(t, l_1, l_2, l_3) = l_1 \exp\left(-\left(\frac{t-l_2}{2l_3}\right)^2\right), \tag{1}$$

where t denotes the time as the dependent variable, l_1 , l_2 , and l_3 are referred to the maximum number of infected patients, the peak date, and the dispersion close to or away from the peak date. The ending day of viral spread can be determined at the tail of the function reaching the nearest zero.

For epidemiology, the S-shape-like functions including logistic and hyperbolic-tangent functions, were mainly used to describe the growth of the infected patients (**Figure 2(b)**). The general mathematical expression of the logistic function is expressed as,

$$f_L(t, j_1, j_2, j_3) = \frac{j_1}{1 + e^{-(t-j_2)/j_3}}, \tag{2}$$

where j_1 , j_2 , and j_3 are referred to the total number of patients at the ending period, the infected rate and the time at the maximum number of patients, respectively. Also, the ending day can be determined at the

day at which the number of patients is equal to j_1 by rounding to the nearest integer. Whereas, the general mathematical expression of the hyperbolic-tangent function is expressed as,

$$f_T(t, k_1, k_2, k_3, k_4) = \frac{k_1}{2} \left(\frac{1 - e^{-(t-k_2)/k_3}}{1 + e^{-(t-k_2)/k_3}} \right) + k_4, \tag{3}$$

where k_1, k_2, k_3 and k_4 are referred to the total number of infected patients at the ending period, the infected rate, inflection, and bias, respectively.

In this work, the GA is employed to fine-tune the parameters of those functions of the free-model method for achieving accurate prediction models. It is based on evolutionary searching for the fittest survival and worked with the coding of the parameters combining the mechanic of natural selection and genetic processes, such as crossover, and mutation. The process of GA (**Figure 3**) comprises 4 steps detailed as follows.

In step 1, the M -chromosome, composed of the genes represented the solution of parameters of functions (l 's or j 's or k 's), including some of the GA's parameters (the probability of crossover P_c and mutation P_m) for comprehensive optimization, are randomly generated and encoded in a binary string of N_{bit} -bit a gene, providing $2^{N_{bit}}$ -point, which is normalized into the pre-defined searching space.

In step 2, the individual i^{th} -chromosome at n^{th} -iteration is assessed by the fitness function regarding the error $E_{i,n}$ between the confirmed and prediction values, $f_{i,n} = 1 / (\sum_{j=1}^N E_{i,n}^j + 1)$, where N is the total number of updated data. The higher the fitness value, the higher the scoring chromosome.

In step 3, to maintain some fittest chromosomes by the elitism strategy, they are saved by $M_{elit}\%$ which can pass to the next generation unconditionally, whereas the rest are selected by assigning the bias of selection to those with higher fitness values through the roulette wheel method.

Lastly, in step 4, the selected 2 chromosomes, according to P_c , are subsequently crossed at random positions to change the information from each other in the crossover stage. For getting out of the local minima, the newly produced chromosomes are mutated at the random position with P_m . So, the former chromosomes are replaced by the generated offspring. The process of GA is then repeated step 1- step 4, as long as without improving the convergence rate or meeting the maximum generation.

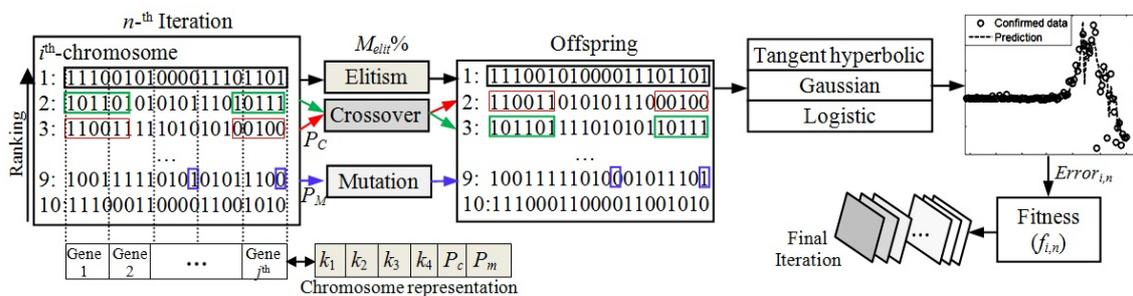


Figure 3 The optimization of parameters of the model-free method through genetic algorithm.

For the time-series MLR (D. A. Freedman, 2009), a statistical technique is used to identify the relationship of the independent variable (predictor) set to the dependent variable. Here, due to the limited independent data, e.g., the transportation of people of each province during the returning hometown period, population interaction activity, medical data of patients, weathers, that impact on the dynamic change of the number of the predicted results, so some past data with a certain number of delays or lagged

values (lag), $y(t-1)$, $y(t-2)$, ..., $y(t-lag)$, and their interaction terms (the combination of product terms) are used for predicting a future value $y(t)$. The MLR model can be expressed as,

$$y(t) = \alpha + \sum_{i=1}^{lag} \beta_i y(t-i) + \sum_{i=1}^{lag-1} \sum_{j=i+1}^{lag} \delta_{i,j} y(t-i)(t-j) + \sum_{i=1}^{lag-2} \sum_{j=i+1}^{lag-1} \sum_{k=j+1}^{lag} \gamma_{i,j,k} y(t-i)y(t-j)y(t-k) + \dots \quad (5)$$

Selecting the appropriate lagged value is important. To make Eq. (5) more accurate, the parameters of the MLR are updated through the GLS to minimize the sum of squared residual.

To prepare the data in training the prediction models, the dates are changed into the number to set as the independent variable. Let January 12, 2020, be the 1st day in prediction. Some past data required initially to fit the functions of the free-model method, whereas at least 30, 37, and 60 past data required to train : test (75 % : 25 %) the regression model for 1-day, 1-week, and 1-month ahead, respectively. Therefore, the free-model method can be used at the beginning of the viral spread, and the MLR model is subsequently applied to predict in the middle of February, March, and April. To assess the prediction models, 2 error metrics: the root mean square error (RMSE) and the R-squared or coefficient of determination (R^2), are used, where the lesser the RMSE the more accurate the prediction, and the higher of R^2 the more completely explained the variations of the observation by the predictor set.

Results and discussion

For the model-free method, in the GA optimization stage, the pre-setting values are as follows, $M = 10$, $N_{bit} = 10$, $N = [10, 11, \dots]$, $M_{elit} = 10\%$, double point crossover, and multiple point mutation. The searching spaces of the parameters for the Gaussian function, l_1 , l_2 , and l_3 are $[1, 300]$, $[10, 200]$, and $[5, 50]$, respectively, those of the logistic function, j_1 , j_2 , and j_3 are $[14, 2000]$, $[10, 100]$, and $[1, 50]$, respectively, and those of the hyperbolic tangent, k_1 , k_2 , k_3 , and k_4 are $[1, 3000]$, $[5, 500]$, $[0.001, 1]$, and $[5, 1500]$, respectively. The GA's parameters, P_c and P_m , are $[0.65, 0.95]$, and $[0.05, 0.35]$, respectively. After meeting the stopping criterion of maximum generation set as 10,000, the best chromosomes (solutions) represented parameters of the model-free method with the optimized P_c and P_m , including the dynamic functions according to the number of training data from January 12, 2020, to January 31, 2020, to February 29, 2020, to March 31, 2020, and to April 17, 2020, for the logistic, hyperbolic tangent, and Gaussian distribution functions are shown in **Figures 4(a) - 4(d)**, **Figures 5(a) - 5(d)**, and **Figures 6(a) - 6(d)**, respectively.

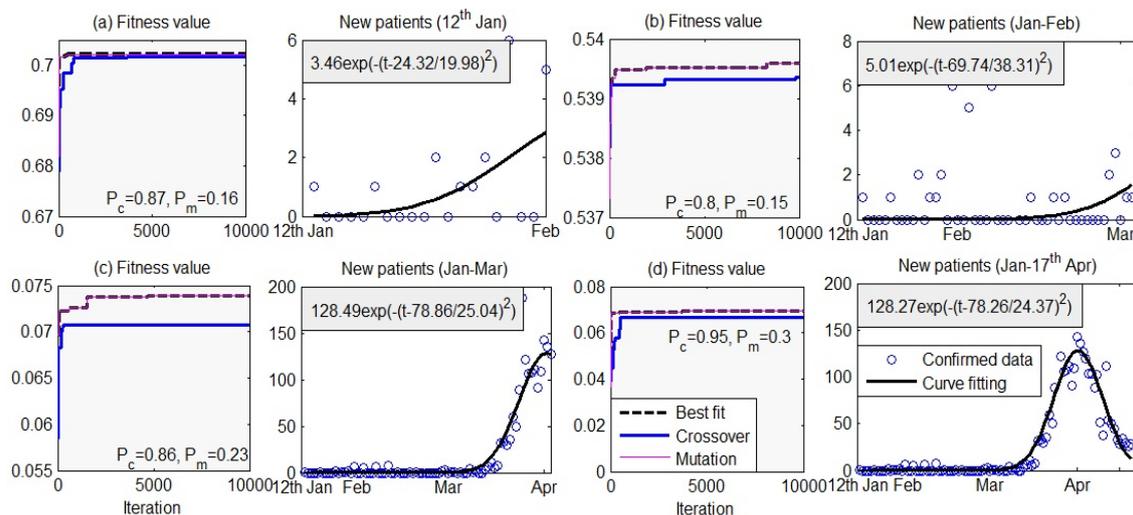


Figure 4 The convergence of the best chromosomes, and resulting fitted-curves of the Gaussian function.

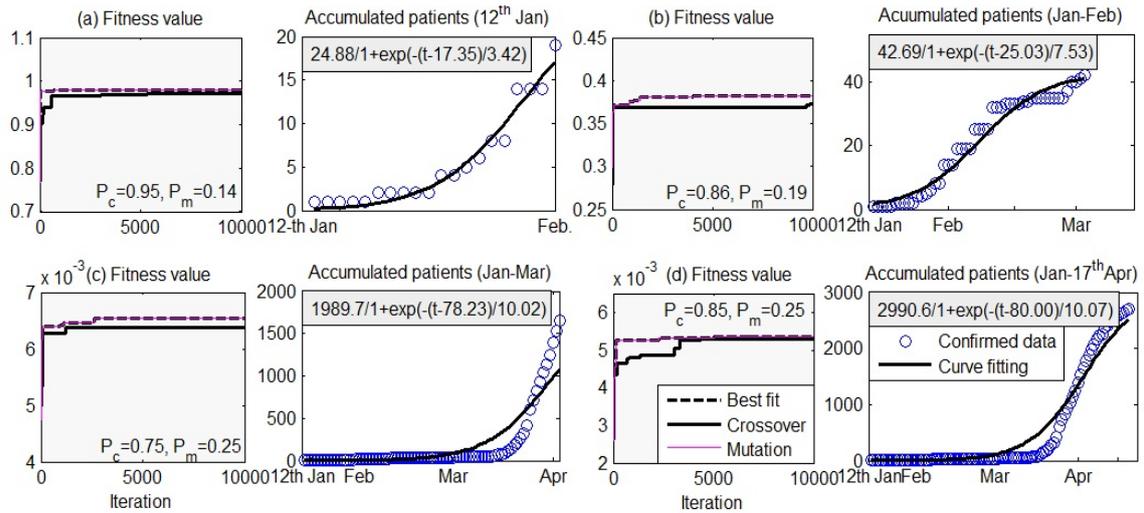


Figure 5 The convergence of the best chromosomes, and resulting fitted-curves of the logistic function.

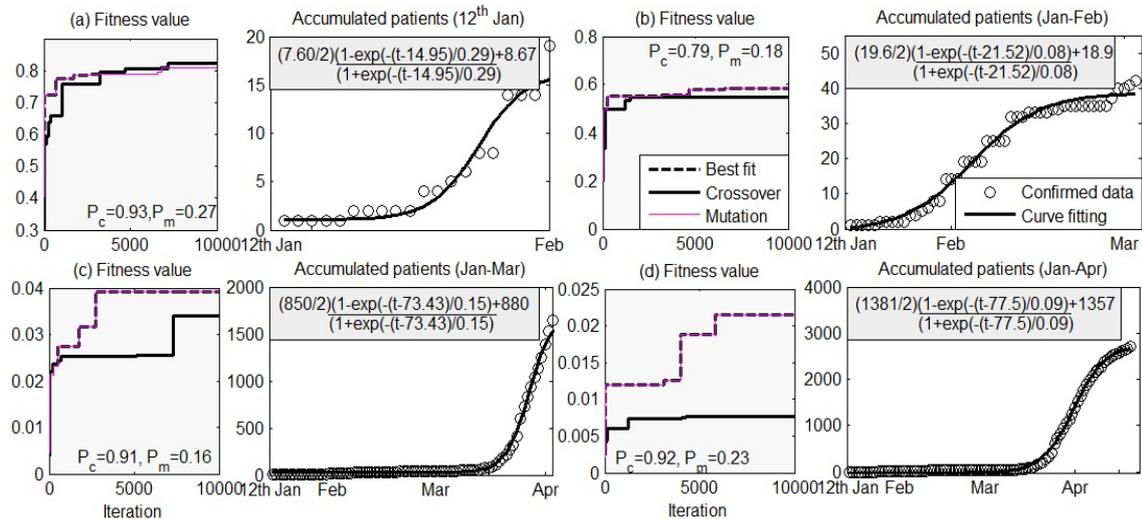


Figure 6 The convergence of the best chromosomes, and resulting fitted-curves of the hyperbolic tangent function.

In the test, the prediction results of the free-model method (i.e., the newly infected patient by the Gaussian function, and the accumulated patients by the hyperbolic-tangent and logistic functions) are shown in Figures 7 and 8, respectively. The results of parameters of the predicted functions (i.e., k_1 , k_2 , k_3 , and k_4), including the RMSE, R^2 , inflection/turning points, the ending date of viral spread, and the total patients, are presented in Table 1. After assessing the goodness-of-fit of the used functions in model-free method, they have a high R^2 , whereas the Gaussian function gives lower RMSE than the rest.

For the prediction on infected patients using the GDF of the 4 periods (Figure 7), the results showed that the cessation (ending) day of viral spread is at around May 2 - 5, 2020. Whereas, for the prediction on accumulated patients using hyperbolic tangent and logistic functions of the 4 periods (Figure 8), the results showed that the cessation day varies with the past data is not consistent. However,

the cessation days using the data from January 12 to April 17, 2020, are May 4, 2020, and June 5, 2020, predicted by the hyperbolic-tangent and logistic functions, respectively. Moreover, the results indicate that total patients are about 2,750 and 3,000 cases. Furthermore, the turning point (peak time) of the Gaussian function and the inflection point of the hyperbolic tangent and logistic functions are located on April 2 - 5, 2020 that corresponds to using curfew and social distance measures.

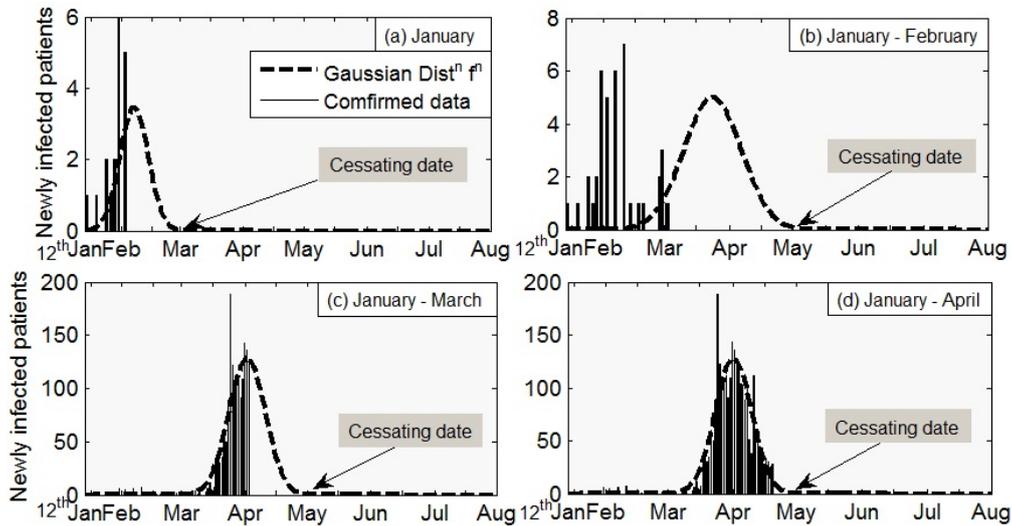


Figure 7 Prediction results of the confirmed infected patients using Gaussian distribution function.

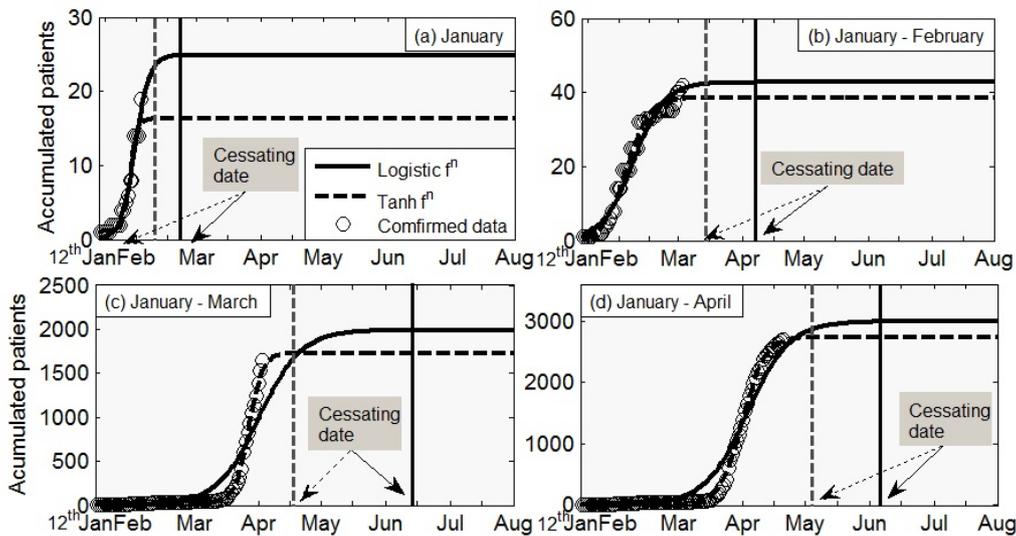


Figure 8 Prediction results of the accumulated patients using logistic and hyperbolic tangent functions.

Table 1 The optimized parameters of the model-free method with their performances and forecasts.

Function	Training period	Parameter (95 % C.I.)*				RMSE	R ²	Predictions		
								Cessation day	Inflection point	Total patients
		(SD.)	(SD.)	(SD.)	(SD.)					
		<i>l</i> ₁	<i>l</i> ₂	<i>l</i> ₃	-					
Gaussian function	12Jan. - 31Jan.	3.46 (0.32)	24.32 (1.81)	19.98 (1.03)	-	1.42	0.90	3 May	6 Feb.	-
	12Jan. - 29Feb.	5.01 (0.14)	69.74 (4.31)	38.31 (3.32)	-	1.85	0.79	5 May	22 Mar.	-
	12Jan. - 31Mar.	128.49 (5.3)	78.86 (3.33)	25.04 (2.03)	-	13.53	0.92	2 May	3 Apr.	-
	12Jan. - 17Apr.	128.27 (4.9)	78.26 (2.83)	24.37 (1.99)	-	14.36	0.88	2 May	5 Apr.	-
	average					7.79	0.87	-	-	-
		<i>j</i> ₁	<i>j</i> ₂	<i>j</i> ₃	-					
Logistic function	12Jan. - 31Jan.	24.88 (2.5)	17.35 (1.5)	3.42 (0.22)	-	1.01	0.91	22 Feb.	1 Feb.	25
	12Jan. - 29Feb.	42.69 (2.8)	25.03 (2.0)	7.53 (0.9)	-	2.61	0.79	24 Mar.	8 Feb.	43
	12Jan. - 31Mar.	1989.7 (13.3)	78.23 (3.41)	10.02 (0.87)	-	153.0	0.74	5 Jun.	4 Apr.	2,000
	12Jan. - 17Apr.	2990.6 (23.3)	80.00 (3.14)	10.07 (1.1)	-	186.9	0.83	5 Jun.	5 Apr.	3,000
	average					85.88	0.82	-	-	-
		<i>k</i> ₁	<i>k</i> ₂	<i>k</i> ₃	<i>k</i> ₄					
Hyperbolic tangent function	12Jan. - 31Jan.	7.60 (0.87)	14.95 (1.23)	0.29 (0.03)	8.67 (0.88)	0.08	0.95	10 Feb.	28 Jan.	16
	12Jan. - 29 Feb.	19.66 (1.87)	21.52 (1.98)	0.08 (0.00)	18.96 (1.74)	1.71	0.94	13 Mar.	5 Feb.	38
	12Jan. - 31Mar.	850.06 (15.3)	73.43 (3.35)	0.15 (0.00)	880.1 (12.7)	25.54	0.97	20 Apr.	27 Mar.	1,750
	12Jan. - 17Apr.	1381.3 (53.9)	77.54 (4.6)	0.09 (0.00)	1357.7 (60.3)	46.55	0.96	4 May	2 Apr.	2,750
	average					18.47	0.95	-	-	-

* C.I. and S.D. stands for confident interval and standard deviation of the error

For the MLR, time lag is chosen to be 4 (*lag*) by considering the autocorrelation plots (**Figure 9**) of which the correlation coefficients are more than 0.8. Therefore, the prediction model can be represented through 16 parameters (*k*₀, *k*₁, *k*₂, ..., and *k*₁₅), which not shown here, by

$$\begin{aligned}
 y(t + Lead) = & k_0 + k_1y(t-1) + k_2y(t-2) + k_3y(t-3) + k_4y(t-4) + k_5y(t-1) \times y(t-2) \\
 & + k_6y(t-1) \times y(t-3) + \dots + k_{10}y(t-3) \times y(t-4) + k_{11}y(t-1) \times y(t-2) \times y(t-3), \quad (4) \\
 & + \dots + k_{14}y(t-2) \times y(t-3) \times y(t-4) + k_{15}y(t-1) \times y(t-2) \times y(t-3) \times y(t-4)
 \end{aligned}$$

where *y*(*t* + *Lead*) represents the predicted variables for 1-day-ahead (*Lead* = 1), 1-week-ahead (*Lead* = 7), and 1-month-ahead (*Lead* = 30).

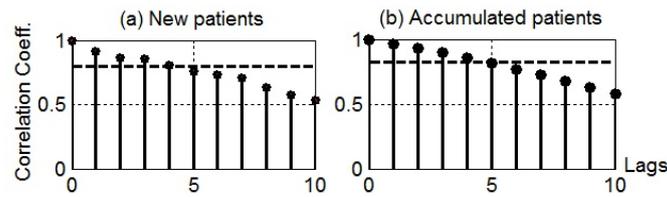


Figure 9 The autocorrelation plots.

For the prediction using the time-series MLR, the results are shown in Figure 10. It is seen that the MLR models can predict the number of variables well, especially for 1-day and 1-week ahead. By using the RMSE (Table 2) for evaluation, the predictions 1-day-ahead and 1-week-ahead are more accurate than those for 1-month ahead with regards to the RMSE. Whereas by using the R^2 for assessment, the variation of prediction variable for 1-day-ahead and 1-week-ahead can be explained well by using 4-time lags of past value as the predictors.

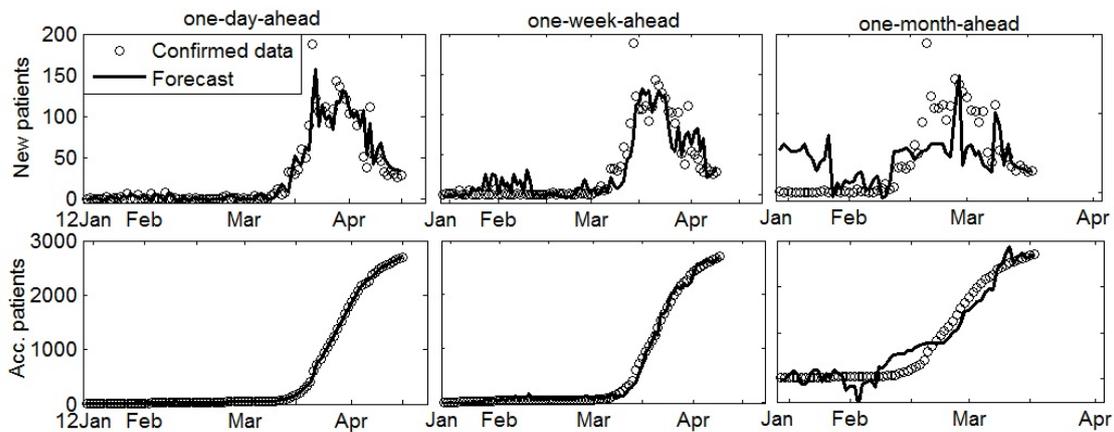


Figure 10 Prediction results by using the time series MLR.

Table 2 The prediction performances of the MLR.

Variable	Period	Performances						Prediction (cases)
		1-day-ahead		1-week-ahead		1-month-ahead		
		R^2	RMSE	R^2	RMSE	R^2	RMSE	
Newly infected case	12Jan. - 31Jan.	0.91	0.46	0.53	1.46	-	-	-
	12Jan. - 29Feb.	0.53	1.14	0.52	1.28	0.89	0.28	-
	12Jan. - 31Mar.	0.97	7.00	0.88	15.19	0.44	38.53	-
	12Jan. - 17Apr.	0.89	13.85	0.78	19.83	0.30	40.70	-
	average	0.82	5.61	0.67	9.44	0.54	26.50	-
Cumulatively infected case	12Jan. - 31Jan.	0.99	0.35	0.99	0.23	-	-	-
	12Jan. - 29Feb.	0.99	1.07	0.96	1.92	1.0	0	2,689
	12Jan. - 31Mar.	1.0	5.43	1.0	21.64	0.88	151.59	2,689
	12Jan. - 17Apr.	1.0	11.98	0.99	60.86	0.93	247.06	2,695
	average	0.995	4.70	0.98	21.16	0.93	132.88	-

Conclusions

In this study, the model-free method and the time-series MLR model are implemented online to predict the ongoing infected cases of COVID-19 in Thailand during the 1st and second phases of the epidemic. The prediction results from January 12, 2020, to April 17, 2020, using the model-free method based on the Gaussian, hyperbolic-tangent, and logistic functions matched the number trend of the daily confirmed patients and its cumulative cases. Also, the prediction results using the MLR model are accurate up to a 1-week-ahead of the forecast. Though, at the moment of proposing the predictive models (April 17, 2020), while the COVID-19 is not terminated, the other results of cessation day and total infected patients are still not proven. In the future, the predictive model for the spread of COVID-19 can be further improved by incorporating the other time series, such as the transportation of people of each province during the outbreak, population interaction activity, medical data of patients, and weathers, into the MLR model.

Acknowledgements

The author directly acknowledges gratefully the Research Institute of North-Chiang Mai University.

References

- [1] X Tan, E Feng and G Xu. SARS epidemic modeling and the study on its parameter control system. *J. Eng. Math.* 2003; **20**, 39-40.
- [2] Y Bai and Z Jin. Prediction of SARS epidemic by BP neural networks with online prediction strategy. *Chaos Solitons Fractals* 2005; **26**, 559-69.
- [3] A Earnest, MI Chen, D Ng and LY Sin. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. *BMC Health Serv. Res.* 2005; **5**, 36.
- [4] OB Da'ar and AE Ahmed. Underlying trend, seasonality, prediction, forecasting and the contribution of risk factors: An analysis of globally reported cases of Middle East Respiratory Syndrome Coronavirus. *Epidemiol. Infect.* 2018; **146**, 1343-49.
- [5] K Nah, S Otsuki, G Chowell and H Nishiura. Predicting the international spread of Middle East respiratory syndrome (MERS). *BMC Infect. Dis.* 2016; **16**, 356.
- [6] WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemics - I. *Bull Math. Biol.* 1991; **53**, 33-55.
- [7] W Yang, A Karspeck and J Shaman. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLoS Comput. Biol.* 2004; **10**, e1003583.
- [8] TK Yamana, S Kandula and J Shaman. Individual versus superensemble forecasts of seasonal influenza outbreak in the United States. *PLoS Comput. Biol.* 2017; **13**, e1005801.
- [9] EL Ray and NG Reich. Prediction of infectious disease epidemics via weighted density ensembles. *PLoS Comput. Biol.* 2018; **14**, e1005910.
- [10] TM Chen, J Rui, QP Wang, ZY Zhao, JA Cui and L Yin. A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infect. Dis. Poverty* 2020; **9**, 24.
- [11] Q Li, X Guan, P Wu, X Wang, L Zhou, Y Tong, R Ren, KSM Leung, EHY Lau, JY Wong, X Xing, N Xiang, Y Wu, C Li, Q Chen, D Li, T Liu, J Zhao, M Liu, W Tu, C Chen, L Jin, R Yang, Q Wang, S Zhou, R Wang, H Liu, Y Luo, Y Liu, G Shao, H Li, Z Tao, Y Yang, Z Deng, B Liu, Z Ma, Y Zhang, G Shi, TTY Lam, JT Wu, GF Gao, BJ Cowling, B Yang, GM Leung and Z Feng. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Eng. J. Med.* 2020; **382**, 1199-207.
- [12] CY Yang and J Wang. A mathematical model for the novel coronavirus epidemic in Wuhan, China. *Math. Biosci. Eng.* 2020; **17**, 2708-24.
- [13] J Gu, Z Gao and W Li. Modeling of epidemic spreading with white Gaussian noise. *Chinese Sci. Bull.* 2011; **56**, 3683-88.

- [14] G Zhou and G Yan. Severe acute respiratory syndrome epidemic in Asia. *Emerg. Infect. Dis.* 2003; **9**, 1608-10.
- [15] YH Hsieh, JY Lee and HL Chang. SARS epidemiology modeling. *Emerg. Infect. Dis.* 2004; **10**, 1165-7.
- [16] K Roosa, Y Lee, R Luo, A Kirpich, R Rothenberg, JM Hyman, P Yang and G Chowell. Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13-23, 2020. *J. Clin. Med.* 2020; **9**, 596.
- [17] B Pirouz, SS Haghshenas, SS Haghshenas and P Piro. Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. *Sustainability* 2020; **12**, 2427.