The Dengue Cases Using Grey Model

Preecha Khrueasom, Chalermchai Puripat, Shawiss Puripat and Vadhana Jayathavaj

October 24, 2022

The Dengue cases using Grey Model

Preecha Khrueasom[1] Chalermchai Puripat[2*] Shawiss Puripat[3]

Vadhana Jayathavaj[4]

[1] Dhonburi Rajabhat University, Bangkok 10600, Thailand
preecha.k@dru.ac.th
[2] Kasem Bundit University, Bangkok 10250, Thailand
chalermchai.pur@kbu.ac.th
[3] Bangpoo Industial Estate Office, Samut Prakan 10280, Thailand
pu.shawiss_st@tni.ac.th
[4]Pathumthani University, Pathumthani 12000, Thailand
vadhana.j@ptu.ac.th

Abstract. The forecasting of Dengue cases was only relied on ARIMA (Auto Regressive Integrated Moving Average), the Grey Model has been acceptable worldwide. The GM(1,1) and GM(1,1) expanded with periodic correction (GM(1,1)EP)model were tested with annual Dengue cases, the determined Grey models showed good prediction results for the normal year 2018, but was not good for the peak year 2019. The roll forward with the Grey models GM(1,1)EP for monthly data were not achieved good accuracy results due to the data may lost their originality. The sophisticated model will be explored to achieve the accuracy of prediction.
Keywords: Dengue cases; Forecasting; Grey model; Time series

1. Introduction

Dengue virus infection caused by mosquitoes and developed Dengue Fever (DF), Dengue Hemorrhagic Fever (DHF), and Dengue shock syndrome (DSS) (WHO SEARO, 2020). Bureau of Vector Borne Disease (VB), Department of Disease Control (DDC), Ministry of public Health (MOPH) used ARIMA model (Auto Regressive Integrated Moving Average) to produce the Dengue Fever Forecast in 2018, 2019 and 2020 with Mean Average Percentage Error (MAPE) of ARIMA(1,0,2)(1,1,2), ARIMA model (1,1,0) and ARIMA model (1,1,2) (1,1,2) at 53.62%, 31.51%, and 17.09%, respectively. The 2019 forecast was difference from the statistics in 2019 as shown in Table 1.

Table 1. Monthly Dengue cases actual and forecast in 2019.

Month	Actual	Forecast	% Difference
1	5,292	4,744	-10.36%
2	4,900	4,205	-14.18%
3	5,356	4,368	-18.45%
4	4,901	4,614	-5.86%
5	8,305	7,469	-10.07%
6	18,560	12,446	-32.94%
7	22,394	13,414	-40.10%

8	18,187	12,103	-33.45%
9	14,325	9,142	-36.18%
10	12,747	7,870	-38.26%
11	10,317	7,577	-26.56%
12	4,622	6,339	37.15%
Total	129,906	94,291	-27.42%

In 1982, Deng established the grey system theory (Deng, 1982, 1989), and the grey forecasting model has been successfully used in finance, physical control, engineering and economics. The advantages of the grey forecasting model include: (a) it can be used in circumstances with relatively little data; as low as four observations were reported to estimate the outcome of an unknown system; and (b) it can use a first-order differential equation to characterize a system (Liu & Lin, 2010). However, grey models have not been explored in the seasonal time series forecasting, a ratio-to-moving-average method was purposed to remove the seasonality in a seasonal time series before modeling a grey model, and the GM(1,1) grey model with depersonalized data, outperformed other models (Tseng, Yu, & Tzeng, 2001). To explore the opportunity introducing the Grey time series forecasting that would be appropriated in both annual and monthly seasonal of Dengue cases, the Grey model was trial with the past data from MOPH

2 Materials and Methods

2.1 Dengue cases

Number of cases and deaths of DF, DHF and DSS classified by month and province from 2003 to 2019 (17 years), and 2020 (January and February) from National Disease Surveillance (Report 506) (Center of Epidemiological Information, 2020).
The total of the Dengue cases was combined from DF, DHF, and DSS cases.

2.2 Grey system theory for time series forecasting

2.2.1 The Grey Model First Order One Variable - GM(1,1) model

Grey Model First Order One Variable - GM(1,1) is a basic model with its computational efficiency among GM(1,1) (a grey model, where is the order of difference equation and is the number of variables). GM(1,1) is most widely used in various fields, i.e. agriculture, ecology, medicine, environment, etc. and also in the time series forecasting model. The Accumulation Generating Operation (AGO) applies to the primitive data in order to smooth the randomness, the differential equation is solved and the Inverse Accumulated Generating Operation (IAGO) is applied to find the predicted values of original data (Deng, 1989). Consider that denotes the number of deaths from cerebrovascular diseases of non-negative sequence and is the sample size of the data. After applying AGO to using Eq. (3), the monotonic increasing sequence is obtained. is the mean sequence that is generated from using Eq. (5). The least square estimate sequence of the grey difference equation of GM(1,1) is defined in Eq. (6) The whitening equation is shown in Eq. (7). is a sequence of parameters that can be found in Eq. (8). According to Eq. (7), the solution of at time is in Eq. (11), and by IAGO, the original sequence can be expressed in Eq. (12) (Liu and Lin, 2010). And they are the forecast values of the individual values and the accumulated values, respectively.

$$X^{(0)} = \left(x^{(0)}(1),\ldots,x^{(0)}(n)\right) \tag{1}$$

$$X^{(1)} = \left(x^{(1)}(1),\ldots,x^{(1)}(n)\right) \tag{2}$$

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i),\ k = 1,2,\ldots,n \tag{3}$$

$$Z^{(1)} = (z^{(1)}, z^{(2)}, \ldots, z^{(n)}) \tag{4}$$

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k+1),\quad k = 2,3,\ldots,n \tag{5}$$

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{6}$$

The first order ordinary differential equation of $X^{(1)}$ as:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = b \tag{7}$$

a and *b* are called the developing coefficient and grey input, respectively.

$$\begin{bmatrix} a \\ b \end{bmatrix} = \left(B^T B\right)^{-1} B^T Y \tag{8}$$

Where

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \ldots & \ldots \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{9}$$

$$Y = \left(x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)\right)^T \tag{10}$$

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a} \tag{11}$$

$$\hat{x}^{(0)}(k+1) = \left(1 - e^{a}\right)\left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak}, k = 1,2,\ldots,n \tag{12}$$

2.2.2 Expanded forms of GM(1,1) model

Expanded forms of GM(1,1) model (GM(1,1)E) provide better simulation accuracies than the difference model by transforming GM(1,1) into Equations (13)and(14) (Liu and Lin, 2010, pp. 109-116).

$$x^{(0)}(k) = \beta - \alpha x^{(1)}(k-1) \tag{13}$$

Where

$$\beta = \frac{b}{1+0.5a} \text{ and } \alpha = \frac{a}{1+0.5a}$$

$$\hat{x}^{(0)}(k) = \beta - \alpha x^{(0)}(1)e^{-a(k-2)} \tag{14}$$

The monthly Dengue cases were removed the seasonality by a ratio-to-moving-average method before modeling a grey model.

2.2.3 GM(1,1) expanded with periodic correction model

Improved forecasting precision using error modification of Grey models has been shown as an error correction model (Lin et al., 2013), modification of GM(1,1) model using Fourier series of error residuals (Kayacan, Ulutus, & Kaynak, 2010), and residual modification of Grey Verhulst model on times series error correction (Guo, Song, & Ye, 2005). The error residuals in Eq. (15) can be expressed in Fourier series as Eq. (16).

$$\varepsilon^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k) \tag{15}$$

$$\varepsilon^{(0)}(k) \cong \frac{1}{2}a_0 + \sum_{i=1}^{z}\left[a_i \cos\left(\frac{2\pi i}{T}k\right) + b_i \sin\left(\frac{2\pi i}{T}k\right)\right], \quad k = 2,3,\ldots,n. \tag{16}$$

$$T = n-1 \text{ and } z = \frac{(n-1)}{2} - 1.$$

T will be an integer number and z will be selected as an integer number (Guo, Song, & Ye, 2005). Eq. (16) can be rewritten as Eq. (17) where P and C matrixes can be defined as Eqs. (18)-(20). Fourier series correction can be obtained as Eq. (21).

$$\varepsilon^{(0)} \cong PC \tag{17}$$

$$P = \begin{bmatrix} 1/2 & \cos\left(2\frac{2\pi}{T}\right) & \sin\left(2\frac{2\pi}{T}\right) & \cos\left(2\frac{2\pi 2}{T}\right) & \sin\left(2\frac{2\pi 2}{T}\right) & \ldots & \cos\left(2\frac{2\pi z}{T}\right) & \sin\left(2\frac{2\pi z}{T}\right) \\ 1/2 & \cos\left(3\frac{2\pi}{T}\right) & \sin\left(3\frac{2\pi}{T}\right) & \cos\left(3\frac{2\pi 2}{T}\right) & \sin\left(3\frac{2\pi 2}{T}\right) & \ldots & \cos\left(3\frac{2\pi z}{T}\right) & \sin\left(3\frac{2\pi z}{T}\right) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1/2 & \cos\left(n\frac{2\pi}{T}\right) & \sin\left(n\frac{2\pi}{T}\right) & \cos\left(n\frac{2\pi 2}{T}\right) & \sin\left(n\frac{2\pi 2}{T}\right) & \ldots & \cos\left(n\frac{2\pi z}{T}\right) & \sin\left(n\frac{2\pi z}{T}\right) \end{bmatrix} \tag{18}$$

$$C = \begin{bmatrix} a_0 & a_1 & b_1 & a_2 & b_2 & \ldots & a_n & b_n \end{bmatrix}^T \tag{19}$$

$$C \cong \left(P^T P\right)^{-1} P^T \varepsilon^{(0)} \tag{20}$$

$$\hat{x}_r^{(0)}(k) = \hat{x}^{(0)}(k) - \hat{\varepsilon}^{(0)}(k) \tag{21}$$

2.2.4 Model accuracy evaluation

The accuracy evaluation terms that are used to examine the accuracy of the models in this study are as follows.

The mean absolute percentage error (MAPE) is the average of the absolute value of relative percentage errors.

The root mean square error (RMSE) is the root of the average sum squares of the error.

The closer the correlation coefficient (CC) is closed to 1, the better the prediction.

The closer the coefficient of efficiency (CE) closed is to 1, the more the prediction matches the actual situation.

MAPE, RMSE, CC, and CE are shown in Equations. (22)-(25), respectively (Lin, et al., 2013).

$$MAPE = \left(\frac{1}{n-1} \sum_{k=1}^{n} \left| \frac{\hat{x}(k) - x^{(0)}(k)}{x^{(0)}(k)} \right| \right) \times 100\% \quad (22)$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n} \left(\hat{x}(k) - x^{(0)}(k) \right)^2}{n-1}} \quad (23)$$

$$CC = \frac{\sum_{k=2}^{n} \left(x^{(0)}(k) - \bar{x}(k) \right) \left(\hat{x}(k) - \bar{\hat{x}}(k) \right)}{\sqrt{\sum_{k=2}^{n} \left(x^{(0)}(k) - \bar{x}(k) \right)^2 \sum_{k=2}^{n} \left(\hat{x}(k) - \bar{\hat{x}}(k) \right)^2}} \quad (24)$$

$$CE = 1 - \frac{\sum_{k=2}^{n} \left(x^{(0)}(k) - \hat{x}(k) \right)^2}{\sum_{k=2}^{n} \left(x^{(0)}(k) - \bar{x}(k) \right)^2} \quad (25)$$

The difference of the forecast value from the real value is the ultimate accuracy.

2.3 Forecasting

The GM(1,1) model and GM(1,1) expanded with periodic correction model or GM(1,1)EP were applied to the data in both annually and monthly basis.

2.3.1 Annually basis

The Dengue cases of the year 2003 to 2017 were used to forecast the Dengue cases of the year 2018. The model accuracy and the different from actual in 2018 were reported.

The Dengue cases of the year 2003 to 2018 were used to forecast the Dengue cases of the year 2019. The model accuracy and the different from actual in 2019 were reported.

2.3.2 Monthly basis

Both real values and deseasonalized values of Dengue cases brought to Grey models.

The monthly Dengue cases were deseasonalized by a ratio-to-moving-average method before modeling a grey model.

The monthly Dengue cases 10 years from the year 2009 to 2018 of both actual data and deseasonalized data were brought to forecast the monthly Dengue cases in

2019 by the roll forward forecasting. The roll forward forecasting bring the new forecast to be the actual value for the period k+1 to forecast the period k+2 until k+12 while the eliminating the 1st period, the 2nd period, until the 11st period, respectively. The roll forward forecasting used only 120 months to forecast the 121st month.

The forecasted values using deseasonalized data were seasonalized before comparing with the actual in 2019.

The model accuracy and the different from actual monthly cases in 2019 were reported.

3 Results

3.1 Seasonal Index

The ratio-to-moving-average method was applied 120 months from 2009 to 2018 as shown in Table 2.

Table 2: Seasonal Index using Dengue cases during 2009-2018.

Month	2009	…	2018	Seasonal Index
1	2,614	…	2,244	0.49
2	2,057	…	1,996	0.39
3	2,324	…	2,606	0.44
4	2,947	…	3,104	0.46
5	6,234	…	7,291	0.87
6	8,569	…	13,612	1.51
7	7,184	…	14,125	1.81
8	7,302	…	12,420	1.78
9	5,016	…	9,081	1.32
10	4,640	…	7,592	1.02
11	4,724	…	7,256	0.98
12	3,040	…	5,885	0.58
Total	56,651	…	87,212	0.49

To find the trend line from the deseasonalized 120 months of 2009 to 2018, the regression equation was not significant (Significance F of 0.814 and the p-value of the intercept and the time period were 0.000 and 0.814).

3.2 Annual Forecast for 2018 and 2019

The GM(1,1) and GM(1,1)EP using 15 years Dengue cases from 2003 to 2017 to forecast the Dengue cases in 2018 as shown in Figure 1, and Table 3.

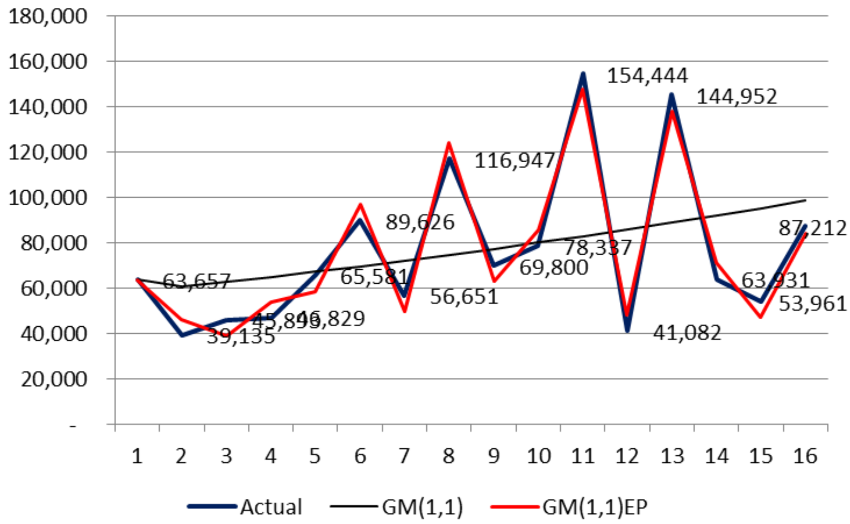

Figure 1: Forecast Dengue cases for the year 2018 (period 16 in X axis)

Table 3: Annual Dengue cases actual and forecast using GM(1,1), and GM(1,1) EP for 2018.

Year	A.D.	Actual	Forecast GM(1,1)	GM(1,1)EP
1	2003	63,657	63,657	63,657
2	2004	39,135	60,469	46,171
3	2005	45,893	62,615	38,857
4	2006	46,829	64,836	53,865
5	2007	65,581	67,137	58,545
6	2008	89,626	69,519	96,662
7	2009	56,651	71,986	49,615
8	2010	116,947	74,540	123,983
9	2011	69,800	77,185	62,764
10	2012	78,337	79,923	85,373
11	2013	154,444	82,759	147,408
12	2014	41,082	85,696	48,118
13	2015	144,952	88,736	137,916
14	2016	63,931	91,885	70,967
15	2017	53,961	95,145	46,925
16	**2018**	**87,212**	**98,521**	**84,040**
MAPE			38.86%	10.07%
2018 Forecast different from Actual			12.97%	-3.64%

The GM(11,1) and GM(1,1)EP using 16 years Dengue cases from 2003 to 2018 to forecast the Dengue cases in 2019 as shown in Figure 2, and Table 4.

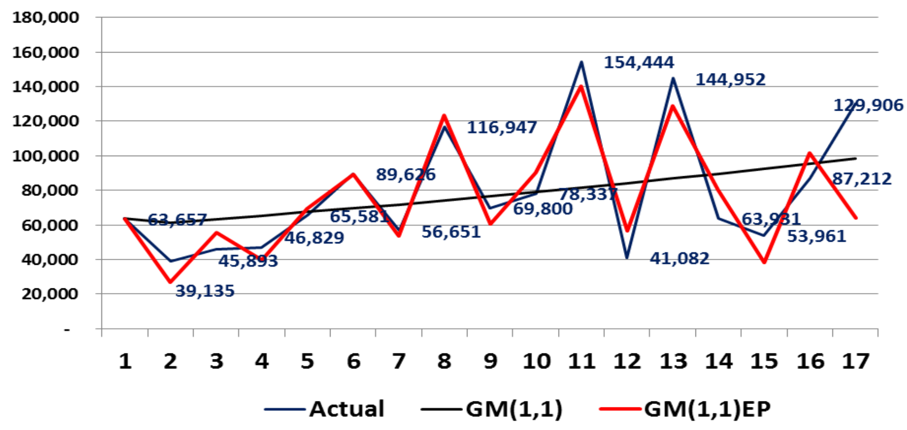

Figure 2: Forecast Dengue cases for the year 2019 (period 17 in X axis)

Table 4: Annual Dengue cases actual and forecast using GM(1,1), and GM(1,1) EP for 2019.

Year	A.D.	Actual	Forecast GM(1,1)	GM(1,1)EP
1	2003	63,657	63,657	63,657
2	2004	39,135	61,382	26,798
3	2005	45,893	63,351	55,739
4	2006	46,829	65,383	39,903
5	2007	65,581	67,480	69,283
6	2008	89,626	69,644	89,309
7	2009	56,651	71,878	53,569
8	2010	116,947	74,184	123,293
9	2011	69,800	76,563	60,467
10	2012	78,337	79,019	90,249
11	2013	154,444	81,554	140,473
12	2014	41,082	84,170	56,501
13	2015	144,952	86,869	128,759
14	2016	63,931	89,656	80,190
15	2017	53,961	92,532	38,346
16	2018	87,212	95,500	101,500
17	2019	**129,906**	**98,563**	**63,829**
MAPE			36.47%	16.11%
2019 Forecast different from Actual			-24.13%	-50.87%

3.3 Monthly forecast for 2019

The roll forward forecast using GM(1,1)EP using 10 years monthly Dengue cases from 2009 to 2018 (120 months), the monthly data during 2017 to 2019 were shown in Figure 3.

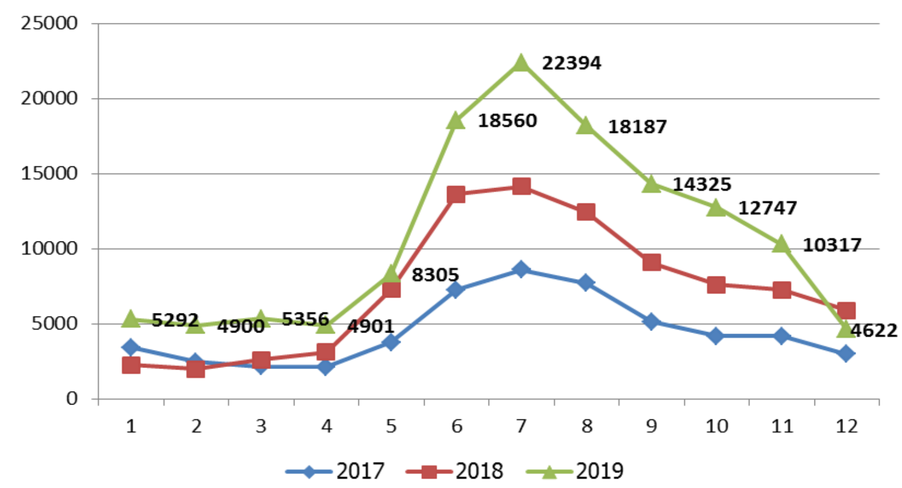

Figure 3: Monthly Dengue cases during 2017 to 2019

To forecast the monthly Dengue cases for 2019 using the roll forward forecast using GM(1,1)EP with 10 years monthly Dengue cases from 2009 to 2018 (120 months) as shown in Table 6.

To forecast the monthly Dengue cases for 2019 using the roll forward forecast using GM(1,1)EP with 10 years deseasonalized monthly Dengue cases from 2009 to 2018 (120 months) as shown in Table 7.

Table 5: Monthly Dengue cases actual and forecast using the recursive GM(1,1) Periodic Correction for the year 2019.

	2019		
Month	**Actual**	**Forecast**	**Difference**
1	5,292	1,850	-65%
2	4,900	1,632	-67%
3	5,356	1,782	-67%
4	4,901	4,597	-6%
5	8,305	6,799	-18%
6	18,560	5,408	-71%
7	22,394	5,502	-75%
8	18,187	3,071	-83%
9	14,325	2,492	-83%
10	12,747	2,151	-83%
11	10,317	232	-98%
12	4,622	232	-95%
Total	129,906	35,748	-72%
MAPE		3.49%	

Table 6: Monthly Dengue cases deseasonalized actual and forecast using the re cursive GM(1,1) Periodic Correction for the year 2019.

Month	Actual	Deseasonal-ized Forecast	Seasonalized	Difference
1	5,292	4,949	2,437	-54%
2	4,900	4,669	1,817	-63%
3	5,356	5,619	2,450	-54%
4	4,901	6,176	2,847	-42%
5	8,305	4,710	4,078	-51%
6	18,560	2,670	4,025	-78%
7	22,394	2,552	4,618	-79%
8	18,187	1,757	3,126	-83%
9	14,325	2,180	2,870	-80%
10	12,747	1,960	2,009	-84%
11	10,317	2,156	2,119	-79%
12	4,622	3,816	2,215	-52%
Total	129,906		34,611	-73%
MAPE		1.11%		

4 Discussion

The ratio-to-moving-average method was applied to 120 months from 2009 to 2018, the high seasonal indexes were in rainy season and lower in winter and summer, because rainy season is a favorable environment condition for mosquitoes to breed.

The GM(1,1) and GM(1,1)EP using 15 years Dengue cases from 2003 to 2017 to forecast the Dengue cases in 2018 had MAPE at 38.86% and 10.07%, respectively. The difference from actual were 12.97% and -3.64%, respectively.

The GM(1,1) and GM(1,1)EP using 16 years Dengue cases from 2003 to 2018 to forecast the Dengue cases in 2019 had MAPE at 36.47% and 16.11%, respectively. The difference from actual in 2019 were -24.13% and -50.87%, respectively.

To forecast the monthly Dengue cases for 2019 using the roll forward forecast using GM(1,1)EP with 10 years monthly Dengue cases and deseasonalized monthly Dengue cases from 2009 to 2018 (120 months) were not shown the appropriate solution, even though they had very low MAPE but monthly difference were very high. These situations may come from the data loss their originality towards the roll forward forecasting. To forecast data with seasonality with deseasonalized before building GM(1,1) grey model were not worked with Dengue cases (Tseng, Yu & Tzeng, 2001; WANG, et al., 2005; Xia & Wong, 2014).

5 Conclusions

The case of annual forecast in 2018 was very acceptable, but for the annual forecast in 2019 the Grey model cannot cover the cyclical effect. The monthly forecasts with both actual and deseasonalized data are still in doubts. The sophisticated model will be explored to achieve the accuracy of prediction.

Acknowledgments: I would like to thank Dhonburi Rajabhat University for supporting this research and Professor Adisak Pongpullponsak as us great teacher.

References

1. Author, F.: Article title. Journal 2(5), 99–110 (2016).
2. Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
3. Bureau of Vector Borne Disease, Department of Disease Control, Ministry of public Health. (2018, 2019, 2020). Dengue Fever Forecast 2019 (in Thai). pp. 3, 5, 5-6.
4. Center of Epidemiological Information, Bureau of Epidemiology, Ministry of Public Health. (2020).
5. National Disease Surveillance (Report 506) DF, DHF and DSS. Total 1. (Number of cases and deaths by month and province) [46][47][48][49][50][51] [52][53][54][55][56][57][58] [59][60] [61][62]. Retrieved from http://www.boe.moph.go.th/boedb/surdata/disease.php?dcontent =old &ds=262766.
6. Deng, J.L. (1982). Control problems of Grey Systems, *System and Control Letters*, 5,288-94.
7. Deng, J.L. (1989). Introduction to grey system theory, *The Journal of Grey System*, 1 (1), 1–24.
8. Guo, Z., Song, X., & Ye, J. (2005).
9. A Verhulst model on time series error corrected for port throughput forecasting. *Journal of the Eastern Asia Society for Transportation Studies*, *6*, 881-891. DOI: http://doi.org/10.11175/easts.6.881
10. Kayacan, E., Ulutas, B., & Kaynak O. (2010). Grey system theory-based models in time series prediction. *Expert Systems with Applications, 37* (2010), 1784-1789. DOI: 10.1016/j.eswa.2009.07.064
11. Lin, Y. H., Chiu, C. C., Lin, Y. J., & Lee, P. C. (2013). Rainfall prediction using innovative grey model with the dynamic index. Journal of Marine Science and Technology, 21(1), 63-75. DOI: 10.6119/JMST-011-1116-1.
12. Liu, S. & Lin, Y. (2010). *Grey systems theory and application*. Verlag: Springer.
13. Tseng, F.M., Yu, H. C., & Tzeng, G. H. (2001). Applied Hybrid Grey Model to Forecast Seasonal Time Series. *Technological Forecasting and Social Change*, 67, 291–302.
14. WANG Qi-jie, LIAO Xin-hao, ZHOU Yong-hong, ZOU Zheng-rong, ZHU Jian-jun, & PENG Yue. (2005, Oct). Hybrid grey model to forecast monitoring series with seasonality. *J. CENT. SOUTH UNIV. TECHNOL*, 12(5), 623-627.
15. WHO SEARO. (2020). Comprehensive Guidelines for Prevention and Control of Dengue and Dengue Haemorrhagic Fever Revised and Expanded Edition. (SEARO Technical Publication Series No. 60). India.
16. Xia, M., & Wong, W.K. (2014). A seasonal discrete grey forecasting model for fashion retailing. *Knowledge-Based Systems*, 57 (2014), 119–126.