Current Applied Science and Technology Vol. 18 No. 2 May-August 2018

DOI: 10.14456/cast.2018.1

Dengue Forecasting Model using SARIMA Model to Predict the Incidence of Dengue in Thailand

Sadiporn Phuthomdee, Pichitpong Soontornpipit, Chukiat Viwatwongkasem and Jutatip Sillabutra*

Department of Biostatistics, Faculty of Public Health, Mahidol University, Bangkok, Thailand

Received: 15 September 2017, Revised: 25 December 2017, Accepted: 4 July 2018

Abstract

Dengue is one of the major public health problems in the tropical countries of the world. SARIMA model is a popular method used for forecasting dengue incidence. The aim of this study was to determine optimal model for forecasting the dengue incidence. SARIMA model with Box-Jenkins approach was conducted to forecast dengue incidence using the previous data from 2006 to 2015. Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC) and Root Mean Square Error (RMSE) were used to determine their accuracy. The results showed that SARIMA (6, 0, 3) $(0, 1, 1)_{52}$ were the best model that fitted with the actual data. It had the smallest AIC and BIC (3827.60 and 3873.30, respectively) and RMSE (0.8420).

Keywords: dengue incidence, SARIMA, forecasting model

1. Introduction

Dengue fever is a serious public health problem in the word. Billion people are now at risk with this disease. WHO estimated that 500,000 people requiring hospitalization each year and about 2.5 % of those died [1]. Moreover, there is now no available licensed dengue vaccine. Dengue vaccine candidates are now under phases of development [2]. To control the spread of disease, disease surveillance and mosquito vector eradication are implemented [1]. However, the dengue incidences have dramatically increased around the world in recent decades [1]. In the South-East Asia region, dengue is increasing with an exponential form every three to five years [3]. In 2015, the morbidity and mortality rates of Thailand are higher than those of five previous year. The incidence rate of Thailand was 219.46 per 100,000 populations (142,925 cases). There were 328.28 per 100,000 population in central, 201.10 per 100,000 population in the Northern, 166.21 per 100,000 population in the North-eastern and 127.52 per 100,000 population in the Southern, respectively. The death rate was 0.22 per 100,000 populations (141 cases). Female and male were equally incidence (1.01:1). The groups of aged 15 - 24 years old were the high incidence rate of dengue infection [4, 5].

*Corresponding author: Tel.: +66-2354-8530 Fax: +66-2354-8534

E-mail: jutatip.sil@mahidol.ac.th

DOI: 10.14456/cast.2018.1

The mathematical or statistical models were famously used to forecast and describe the phenomena of diseases. Forecasting is important as it is benefit for management, health risk assessment, and planning the equipment for control and prevention. [6, 7]. Moreover, the forecasting is taken into accounted as the Early Warning system [8]. The results of forecasting were not only for understanding size of problem, susceptibility host, period of disease, area, and pattern of disease, but also disease control [6].

Time Series Analysis was basically used to predict the incidence case in many diseases such as influenza [9, 10], malaria [11-13], and dengue [14-19]. The model has different degrees of complexity based upon many factors: previous data of dengue cases, climatic variables such as rainfall, evaporation, temperature, humidity, water budget, breeding sites, population density, virus serotypes and microclimates [19-22], Many studies revealed that Seasonal Autoregressive Integrated Moving Average (SARIMA) was used as the model for predicting dengue incidence [23].

SARIMA is reported as the great model for dengue forecasting due to dengue is none stationary problem. The seasonal was taken into account in model. In 2014, Monika S Sitepu *et al.* showed that SARIMA $(1,0,1)(0,1,1)_{12}$ was the best model for predicting dengue in Jakarta [16]. Bhatnagar *et al.* [6] revealed that forecasting model for dengue incidence in Rajasthan was SARIMA $(0,0,1)(0,1,1)_{12}$. The previous studies in Thailand revealed that the forecasting models of dengue incidence in Thailand were SARIMA $(2,0,1)(0,2,0)_{12}$ [24] and SARIMA $(2,0,1)(0,1,1)_{12}$ [25]. According to the previous studies, the most of forecasting models were proposed based on 12 months period of time interval. However, some studies were proposed 52 weeks as time interval in the models [15, 26]. In Thailand, only SARIMA models based on 12 months period of time interval were conducted. There is no study using 52 weeks of time interval in SARIMA model. Therefore, this study aims to determine optimal model for forecasting the dengue incidence in Thailand using SARIMA model with 52 weeks of time interval period. AIC and BIC were used to select the best model, and performance of model was measured by RMSE.

2. Materials and Methods

The SARIMA with Box-Jenkins approach was used to predict dengue incidence. The historical data from 2006 to 2015 were used in analysis. The performance of model was measured by RMSE. The presence of these two components determine the choice of SARIMA $(p,d,q)(P,D,Q)_s$ model equation:

$$Y_{t+k} = \frac{\psi_q(\alpha)\Psi_Q(\alpha)^s \varepsilon_t \Gamma_P(\alpha)^s}{\Gamma_P(\alpha)^s \gamma_p(\alpha)(1-\alpha)^d(1-\alpha^{sD})} \qquad (1)$$

where $\square_{\square+\square}$ is evidence weekly dengue incidence, $\square_{\square}(\square)$ is coefficient of moving average (MA) at q order, $\Psi_{\square}(\square)$ is coefficient of seasonal moving average (SMA) at Q order and S seasonal period, $\square_{\square}(\square)$ is coefficient of autoregressive (AR) at p order, $\Gamma_{\square}(\square)^{\square}$ is coefficient of seasonal autoregressive (SAR) at P order and S seasonal period, d is order of different week period, D is order of seasonal different period, $\square_{\square}$ is white noise time or residual of time**,** t is weekly period time and k is weekly ledge time

According to previous studies, SARIMA $(p,d,q)(P,D,Q)_s$ were set as SARIMA $(p,d,q)(0,1,1)_s$ [12, 13]. In this study, the 52-week (s=52) was used in model. Therefore, SARIMA $(p,d,q)(0,1,1)_{52}$ was selected as the basic structure of candidate model. Logarithmic transformation was used to adjust data to meet the criteria of equally mean and variance in each period. Based on mean-rang plot analysis, Autoregressive function and Partial Autocorrelation function were used to identify the order of Autoregressive part, order of Moving Average part and difference. The order of *d* was randomly from 0, 1, and 2. The order of *p* was randomly from 0 to 6 and the order of *q* was randomly from 0 to 3 (Figure 2). The first criteria for selection model was smallest AIC, BIC and RMSE and highest of Likelihood Ratio Test (LRT).

Finally, the evidence data in 2016 was used to compare with the estimation of model. The performance of model was determined by RMSE. The statistical software R [27] was used for all analysis.

3. Results and Discussion

The evidence dengue incidence in Thailand from 2006 - 2015 showed in Figure 1. It showed that there were two outbreaks in early 2010 to mid-2010, and early 2013 to late 2013. Large of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were showed in Figure 2A and 2B, respectively. The sharp decrease in ACF value from lag 1 to lag 50 and the sharp increase in PACF value from after lag 50 indicated that there was evidence of long-term trend. Therefore, the first different term was taken into account. When considering the first different term as shown in Figure 2B, it showed that sharp of ACF of first difference of weekly dengue incidence were decreased after lag 1.

SARIMA $(6,0,3)(0,1,1)_{52}$ was selected as the best model with the smallest AIC (3827.60). BIC and RMSE were 3873.30 and 0.8420, respectively. The estimated variation of dengue incidence ($\hat{\sigma}^2$) from this models was 171.30. The model parameters were significance (*p*-value <0.0001) with 1stAR (-1.1042, S.E.=0.4364), 2nd AR (-0.2041, S.E.=0.3523), 3rd AR (-0.4611, S.E.=0.3804), 4th AR (-0.2251, S.E.=0.2725), 5th AR (-0.2474, S.E.=0.1115), 6th AR (-0.1029, S.E.=0.0723), 1st MA (1.1185, S.E.=0.4410), 2nd MA (0.4741, S.E.=0.1228), 3rd MA (0.1891, S.E.=0.4432) and seasonal MA (-0.8699, S.E.=0.0464). The best fit model was model SARIMA $(6,0,3)(0,1,1)_{52}$. The comparison between actual dengue fever data in 2016 and forecast data form SARIMA $(6,0,3)(0,1,1)_{52}$ was presented in Figure 3.

In this study, SARIMA $(6,0,3)(0,1,1)_{52}$ was the most optimal predictive model which showed that the smallest AIC, BIC and RMSE. Our results were difference on AR, MA and SMA, when comparing with previous studies. SARIMA $(0,0,1)(0,1,1)_{12}$ was offered with no seasonal differences presented in Rajasthan [6]. Moreover, SARIMA $(1,0,1)(0,1,1)_{12}$ was the best model for predicting DHF cases in Jakarta [16]. According to time series plots analysis, AR and MA were determined and depended on timely interval of lags. AR and MA of this study were designed based on the dengue incidence of epidemiological weeks in one year (s= 52 weeks), but the previous studies were designed based on the dengue incidence of epidemiological months in one year (s= 12). Moreover, the model may depend on geographical area. SARIMA$(0,1,1)(0,1,1)_{52}$, SARIMA $(0,0,1)(0,1,1)_{12}$, SARIMA $(1,0,1)(0,1,1)_{12}$ were the models conducted in island [6, 15, 16]. The different SARIMA models in different areas of Thailand were also found. The SARIMA models conducted

DOI: 10.14456/cast.2018.1

in northern part and north-eastern part were SARIMA(2,0,1)(0,2,0)$_{12}$ and SARIMA(2,0,1)(0,1,1)$_{12}$, respectively [24, 25].

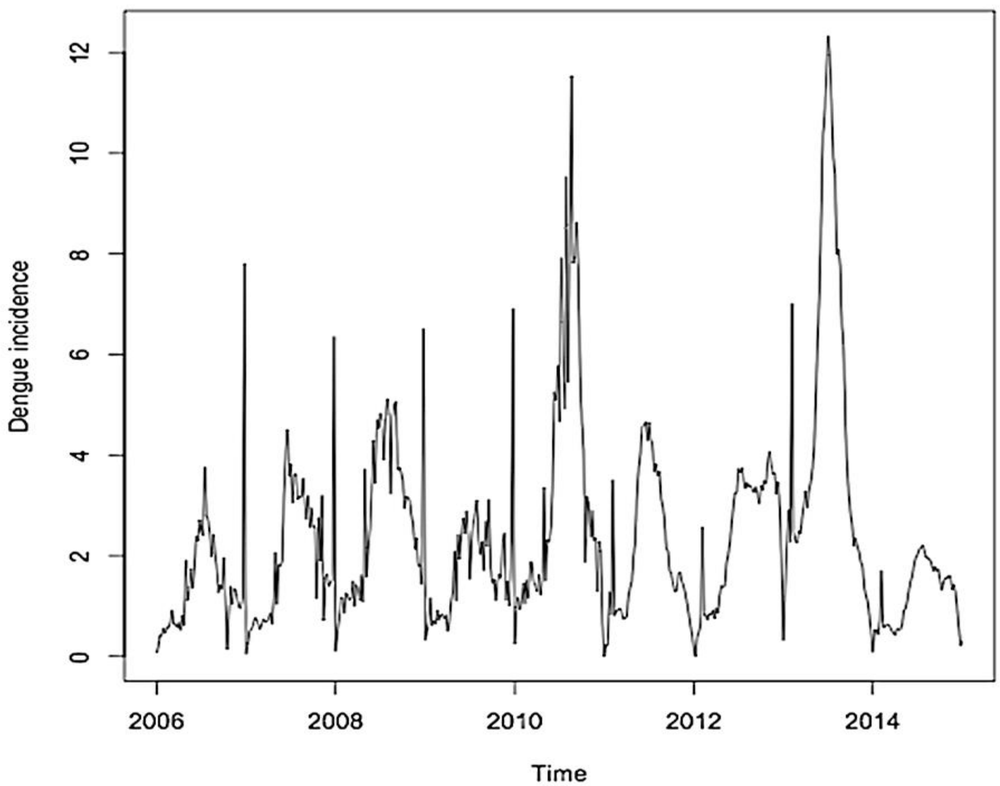

Figure 1. Weekly dengue incidence (per 100,000 populations) in Thailand from January 2006 to December 2015

DOI: 10.14456/cast.2018.1

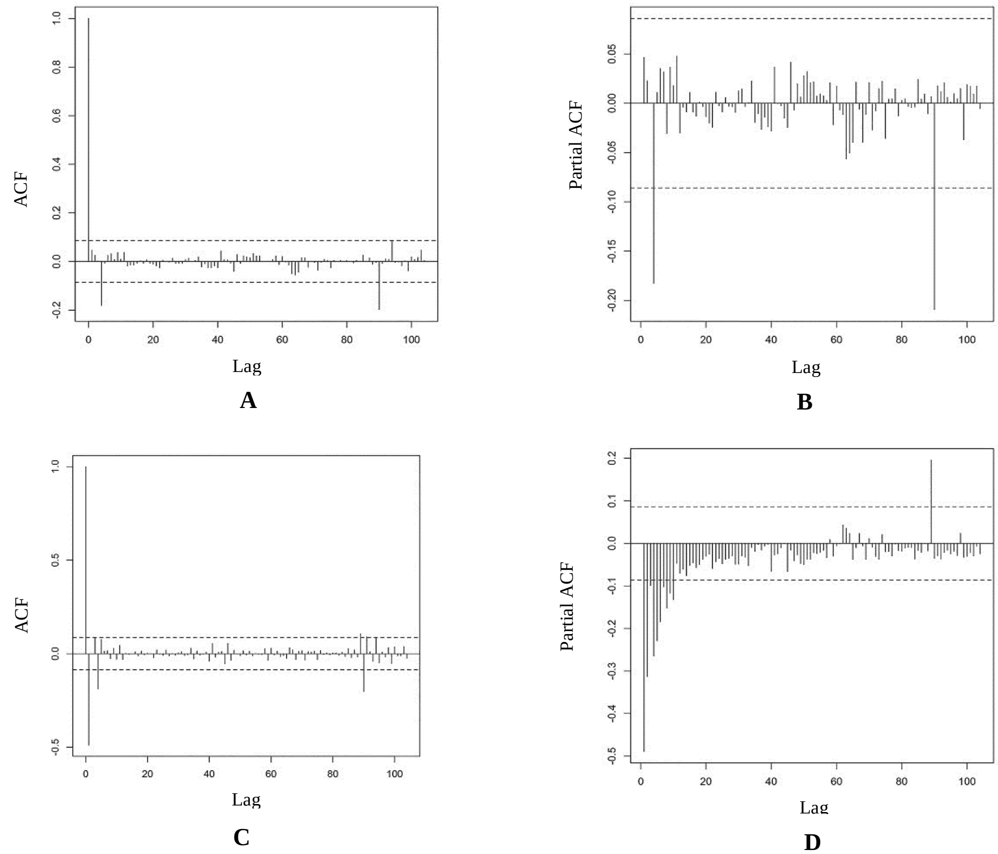

Figure 2. A and B) ACF and PACF of logarithm dengue incidence (per 100,000 populations) in Thailand from January 2006 to December 2015 C and D) ACF and PACF after differencing of logarithm dengue incidence (per 100,000 populations) in Thailand from January 2006 to December 2015

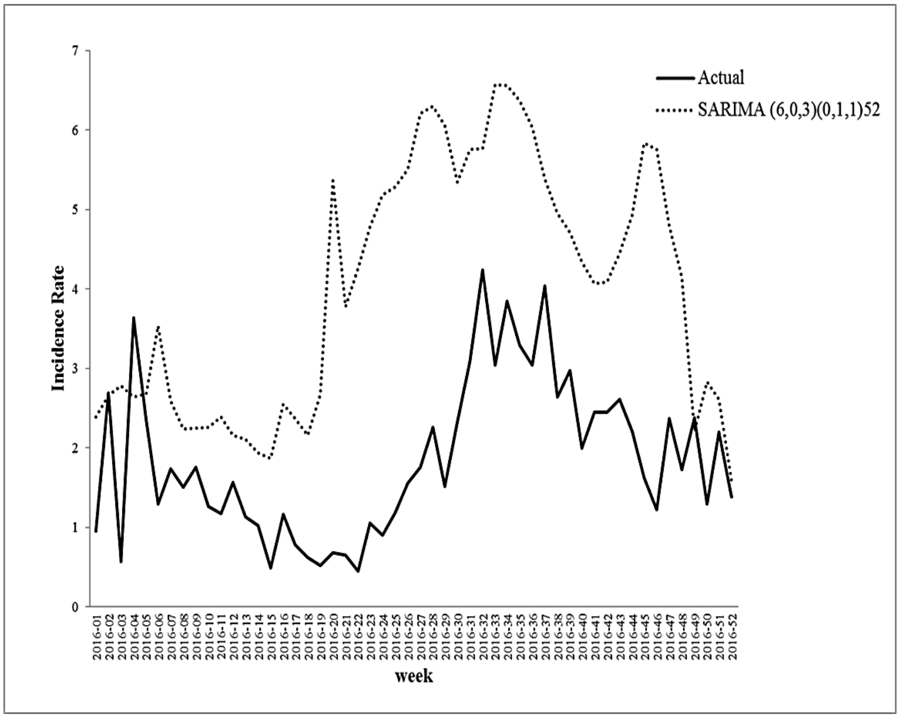

Figure 3. Comparative of dengue incidence rate in 2016 between actual data and forecast data from SARIMA (6,0,3)(0,1,1)$_{52}$.

4. Conclusions

Seasonal Autoregressive Moving Average (SARIMA) model was used to forecast dengue incidence. This model was the most popular model used for prediction in many communicable diseases including influenza, malaria, and dengue. Moreover, it was benefit to epidemiological surveillance and policy makers to manage and prevent the outbreak.

In this study, the forecasting model of dengue incidence was developed based on the secondary data of historical data from 2006 to 2015. The seasonal period with 52 weeks in a year was taken into accounted. SARIMA (6,0,3)(0,1,1)$_{52}$ was presented as the best fitted model with the actual data. However, dengue is also associated with the multiple risk factors such as climate, migration, the characteristic of virus and etc. Therefore, the further study should have taken the risk factors in the model to describe the pattern of dengue.

5. Acknowledgements

We are greatly thankful to Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health, Bureau of Vector - Borne Disease, Department of Disease Control, Ministry of Public Health, Department of Provincial Administration, Ministry of Interior, and the Thai Meteorological

Department, Ministry of Digital Economy and Society for their support the secondary data to developed model in this study.

References

[1] World Health Organization, 2009. *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control*. [online] Available at: www.who.int/tdr/publications/documents/dengue-diagnosis.pdf

[2] Ghosh, A. and Dar, L., 2015. Dengue vaccines: Challenges, development, current status and prospects. *Indian Journal of Medical Microbiology*, 33, 3-15.

[3] WHO Regional Office for South-East Asia., 2011. *Comprehensive Guidelines for Prevention and Control of Dengue and Dengue Hemorrhagic Fever*.

[4] Bureau of Epidemiology, Ministry of Public Health, 2016. *Surveillance Report (R 506) DHF Total*. [online] Available at: http://www.boe.moph.go.th/boedb/surdata/506wk/ y60/d262766_4860.pdf

[5] Bureau of epidemiology, Ministry of Public Health, 2015. *Surveillance Report (R 506) DHF Total*. [online] Available at: http://www.boe.moph.go.th/boedb/surdata/506wk/y59/d262766_5259.pdf

[6] Bhatnagar, S., Lal, V., Gupta, S.D. and Gupta, O.P., 2012. Forecasting incidence in Rajasthan, using time Series analysis. *Indian Journal Public Health*, 56(4), 281-285.

[7] Suthachana, S., Kaewnokkao, W., Suangtho, P., Sayumpurujinun, S. and Kongyu, S., 2012. Forecasting the situation of influenza in Thailand in 2012. In: Bureau of Epidemiology, Department of Diseases Control, Ministry of Public Health, *Workshop Conference in Analyzed and Forecasting*. Bangkok, Thailand, 12-14 September 2012. Bangkok: Ministry of Public Health.

[8] Drake, J.M., 2005. Fundamental limits to the precision of early warning systems for epidemic of infectious disease. *PLoS Medicine*, 2(5), e144. http://doi.org/10.1371/journal.pmed.0020144.

[9] Soebiyanyo, R.P., Adimi, F. and Kaing, R.K., 2010. Modeling and predicting seasonal influenza transmission in warm region using climatological parameters, *PLoS ONE*, 5(3), e9450. http://doi.org/10.1371/journal.pone.

[10] Parmsan, C., 2007. Influence of species, latitudes and methodologies on estimates of phonological response to global warming. *Global Chang Biology*, 13, 1860-1872.

[11] Loha, E. and Lindijorn, B., 2010. Model variation in prediction incidence of *Plasmodium falciparum* malaria using 1998-2007 morbidity and meteorological data from South Ethiopia. *Malaria Journal* 9, 166-173.

[12] Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N.J. and Kaewkungwal, J., 2010. Development of temporal modeling for forecasting and prediction of malaria infections using time series and ARIMAX analysis: A case study in endemic district of Bhutan. *Malaria Journal*, 9, 251-265.

[13] Tian, L., Bi, Y., Ho, S.C., Liu, W., Liang, S. and Goggins, W.B., 2008. One-year delayed effect of fog on malaria transmission: A time series analysis in the rain forest area of Mengla County, south-west China. *Malaria Journal*, 7, 110-123.

[14] Dom, N.C., Hassan, A.A., Ltif, Z.A. and Ismail R., 2012. Generating temporal model using climate variable for prediction of dengue case in subbing jaya, Malaysia. *Asian Pacific Journal of Tropical Disease*, 2, 352-361.

[15] Gharbi, M., Quenel, P., Gustave, J., Cassadou, S., Ruche, G.L., Girdary, L. and Marrama L., 2010. Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting

model using climate variables as predictors, *BMC Infectious Diseases* 11, 166. http://doi.org/10.1186/1471-2334-11-166.

[16] Sitepu, M. S., Kaewkungwal, J., Lupleriop, N., Soonthornworasiri, N., Silawan, T., Poungsombat, S. and Lawpoolsri, S., 2013. Temporal patterns and a disease forecasting model for dengue hemorrhagic fever in Jakarta based on 10 years of surveillance data. *Southeast Asian Journal Tropical Medicine and Public Health*, 44(2), 206-217.

[17] Promprou, S., Jaroensutasinee, M. and Jaroensutasinee, K., 2006. Forecasting dengue hemorrhagic fever cases in Southern Thailand using ARIMA models. *Dengue Bulletin*, 30, 99-106.

[18] Institute of Research, Knowledge Management and Standards for Disease Control, Ministry of Public Health, 2017. A Predictive Model for Dengue Hemorrhagic Fever Epidemics in 8 Provinces, Northern Thailand. [online] Available at: http://irem2.ddc.moph.go.th/research/4803.

[19] Schmidt, W.P., Suzuki, M., Dinh Thiem, V., White, R.G., Tsuzuki, A., Yoshida, L.M., Yanai, H., Haque, U., Tho, L.H., Duc Anh, D. and Ariyoshi, K., 2011. Population density, water supply, and the risk of dengue fever in Vietnam: Cohort study and spatial analysis. *PLoS Medicine*, 8(8), e1001082. http://doi.org/10.1371/ journal.pmed.1001082.

[20] Thammapalo, S., Nagao, Y., Sakamoto, W., Saengtharatip, S., Tsujitani, M., Nakamura, Y., Coleman, P.G. and Davies, C., 2008. Relationship between transmission intensity and incidence of dengue hemorrhagic fever in Thailand. *PLoS Neglected Tropical Diseases*, 2(7), e263. http://doi.org/10.1371/journal.pntd.0000263.

[21] Mammen, M.P., Pimgate, C., Koenraadt, C.J.M., Rothman, A.L., Aldstadt, J., Nisalak, A., Jarman, R.G., Jones, J.W., Srikiatkhachorn, A., Ypil-Butac, C.A., Getis, A., Thammapalo, S., Morrison, A.C., Libraty, D.H., Green, S. and Scott, T.W., 2008. Spatial and temporal clustering of dengue virus transmission in Thai villages. *PLoS Medicine*, 5(11), e205. http://doi.org/10.1371/journal.pmed.0050205.

[22] Schreiber, K.V., 2001. An investigation of relationships between climate and dengue using a water budgeting technique. *International Journal of Biometeorology*, 45(2), 81-89.

[23] Halide, H. and Ridd, P., 2008. A Predictive model hemorrhagic fever epidemics. *International Journal of Environmental Health Research*, 18(4), 253-265.

[24] Wongkoon, S., Pollar, M., Jaroensutasinee, M. and Jaroensutasinee, K., 2007. Predicting DHF incidence in Northern Thailand using time series analysis technique. *International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, 1(8), 484-488.

[25] Silawan, T., Singhasivanon, P., Kaewkungwal, J., Nimmanitya, S. and Suwonkerd, W., 2008. Temporal patterns and forecast of dengue infection in Northeastern Thailand. *Southeast Asian Journal of Tropical Medicine and Public Health*, 39(1), 90-98.

[26] Dom, N.C., Hassan, A.A., Latif, Z.A. and Ismail, R., 2012. Generating temporal model using climate variables for the prediction of dengue case in Sunbang Jaya, Malaysia. *Asian Pacific Journal of Tropical Disease*, 3(5), 352-361.

[27] The R Foundation, 2017. *The R Project for Statistical Computing*. [online] Available at: http:// www.R-project.org/.