Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 1 (2015), pp. 267-273

http://www.ripublication.com

An Approach Model To Forecast The Number Of DHF Cases In Chanthaburi Province, Thailand

J. Mekparyup[1] and K. Saithanu[*2]

[1,2]Department of Mathematics, Faculty of Science, Burapha University
169 Muang, Chonburi, Thailand
[1]jatupat@buu.ac.th, [2]ksaithan@buu.ac.th

Abstract

The presented research was to propose an approach model to forecast the dengue hemorrhagic fever (DHF) cases in Chanthaburi province, Thailand. The seasonal ARIMA model was applied by defining parameters p=1, d=0, q=2, P=1, D=0 and Q=2 with S=12 and no constant or represented with ARIMA$(1,0,2)(1,0,2)_{12}$ which was generated by time series data with root mean square error (RMSE) of 26.68. For considering accuracy, the seasonal ARIMA was determined by the correlation coefficient (R) between the observed and forecasted DHF cases which was 0.964 with p-value=0.000.

Mathematics Subject Classification: 62-07

Keywords: Dengue Hemorrhagic Fever, ARIMA model

INTRODUCTION

Dengue virus is usually passed by Aedes aegypti and albopictus mosquitoes which are contagious. Dengue virus infection in humans causes dengue fever which shows many symptoms in patients with infection. The symptoms are Asymptomatic, undifferential viral syndrome, dengue fever and Dengue Hemorrhagic Fever (DHF) [1]. Dengue virus is a single stranded RNA positive-strand virus of the family Flaviviridae which are 4 serotypes; DEN1, DEN2, DEN3 and DEN4 [2][3]. The first outbreak of dengue fever occurs in the year 1779 in Asia, Africa and North America and then in Philippines in the year 1949. For Thailand, the first patient of DHF was found in the year 1949 and the first outbreak occurred in Bangkok in the year 1958 [4]. The Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health of Thailand, estimated that the number of DHF cases could increase to 100,000 – 200,000 in the year 2013 and the number of cases who died of DHF probably high as

100 – 200 without strictness outbreak prevention. Chanthaburi province is in the east of Thailand facing the epidemic of DHF [5] so the objective of this study is to fit an appropriate model to forecast the number of DHF cases preventing the epidemic of DHF in Chanthaburi province.

MATERIALS AND METHODS

The number of DHF cases in Chanthaburi province was collected from the Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health of Thailand, since January 2007 to December 2013.

1. THE STATIONARY TIME SERIES

Firstly, stationary time series was checked from sample with Autocorrelation function (SACF) and Partial autocorrelation function (SPACF) as Equation 1 and Equation 2 consequently before fitting the seasonal ARIMA model .

$$r_k = \frac{\sum_{t=1}^{n-1}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2} \tag{1}$$

$$\phi_{kk} = \begin{cases} r_1;\ k = 1 \\ \dfrac{r_k - \sum_{j=1}^{k-1} \phi_{(k-1)j} r_{k-1}}{1 - \sum_{j=1}^{k-1} \phi_{(k-1)j} r_j};\ k = 2,\ 3,\ \ldots \end{cases} \tag{2}$$

where Z_t be a variable at time t, k be time lag, $\bar{Z} = \sum_{t=1}^{n} Z_t / n$, $\phi_{kj} = \phi_{(k-1)j} - \phi_{kk}\phi_{(k-1)(k-j)}$; $j = 1,\ 2,\ \ldots,\ k-1$. The graph of SACF and SPACF reduce rapidly to zero or the graph is zero all the time then it could be concluded that time series is stationary. Time series set should be at least 50 proposed by Box and Jenkins [6].

2. THE SEASONAL ARIMA MODEL

Secondly, parameters p, d, q, P, D and Q of multiplicative seasonal autoregressive integrated moving average model of order $(p,d,q)(P,D,Q)_S$ or ARIMA$(p,d,q)(P,D,Q)_S$ is estimated as following Equation 3 [6].

$$\begin{aligned}&(1-\phi_1 B-\ldots-\phi_p B^p)(1-\Phi_1 B^s-\ldots-\Phi_P B^{P_s})(1-B)^d(1-L)^D Z_t\\&=(1-\theta_1 B-\ldots-\theta_q B^q)(1-\Theta_1 B^s-\ldots-\Theta_Q B^{Q_s})a_t\end{aligned} \quad (3)$$

where a_t be the random error at time t, ϕ_i and Φ_I be regression coefficient at time i and I; i=1, 2, …, p; I=1, 2, …, P, θ_j and Θ_J be moving average coefficient at time j and J; j=1, 2, …, q; J=1, 2, …, Q.

3. CHECKING OF THE FITTED MODEL

3.1 THE RESIDUALS

The residuals of the SACF graph was plotted to check the selected seasonal ARIMA model. The model is fitted if the forecast error is within $(1-\alpha)100\%$ confidence interval

3.2 THE LJUNG-BOX Q STATISTIC

To validate the parameter estimation in Equation 3, the Ljung-Box Q statistic was calculated as of Equation 4 [7].

$$Q=\{(n-d)[(n-d)+2]\}\sum_{j=1}^{k}\frac{r_j^2}{[(n-d)-j]} \quad (4)$$

where k be the distance of time lag, n be the number of observations of the time series, d be the order of the time series variances and r_j be autocorrelation at time lag j. The Q statistic is chi-square distribution with degrees of freedom k–p–q when n is large.

4. MEASUREMENT OF THE FITTED MODEL

Lastly, the fitted seasonal ARIMA model was validated by the root mean square error (RMSE) as of Equation 5, time series plot between the observed (OBS) and the forecasted (FOR) data of DHF cases and scatter plot of the correlation coefficient (R) as of Equation 6.

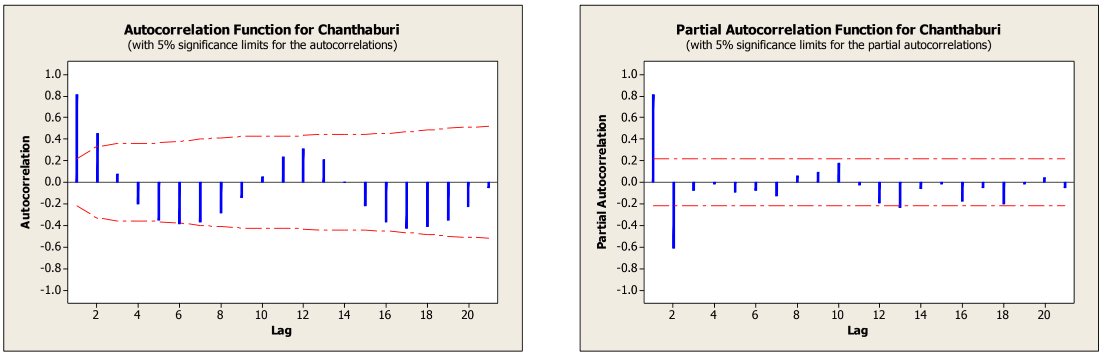
$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(OBS_i-FOR_i)^2} \quad (5)$$

$$R=\frac{\sum_{i=1}^{n}(OBS_i-\overline{OBS})(FOR_i-\overline{FOR})}{\sqrt{\sum_{i=1}^{n}(OBS_i-\overline{OBS})^2\sum_{i=1}^{n}(FOR_i-\overline{FOR})^2}} \quad (6)$$

where OBS_i be the i^{th} observed value, $\overline{OBS}$ be the average of OBS, FOR_i be the i^{th} forecasted value and $\overline{FOR}$ be the average of FOR.

RESULTS AND DISCUSSION

The SACF and SPACF graph were plotted to check the stationary condition of DHF cases in Chanthaburi province.



(a) (b)

Figure 1: Checking the stationary condition (a) SACF graph, (b) SPACF graph

Figure 1 and 2 illustrated that the time series was nonstationary then the parameters of the seasonal ARIMA model were estimated by p=1, d=0, q=2, P=1, D=0 and Q=2 with S=12 and no constant or $ARIMA(1,0,2)(1,0,2)_{12}$ shown in Equation 7.

$$Z_t = \phi_1 Z_{t-1} + \Phi_1 Z_{t-12} - \phi_1 \Phi_1 Z_{t-13} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \Theta_1 a_{t-12}$$

$$+\theta_1 \Theta_1 a_{t-13} + \theta_2 \Theta_1 a_{t-14} - \Theta_2 a_{t-24} + \theta_1 \Theta_2 a_{t-25} + \theta_2 \Theta_2 a_{t-26} \qquad (7)$$

As of Table 1, all parameters estimation of the fitted model was verified following; ϕ_1(T=4.71, p-value=0.000), Φ_1 (T=30.69, p-value=0.000), θ_1(T = –2.24, p-value=0.028), θ_2(T = –2.47, p-value=0.016), Θ_1(T=11.57, p-value=0.000) and Θ_2(T = –6.77, p-value=0.000). Then the fitted seasonal ARIMA model for forecasting the number of DHF cases in Chanthaburi province was written as of Equation 8.

Table 1: The estimator values and testing of the parameters of ARIMA(1,0,2)(1,0,2)$_{12}$

Parameters	Estimation	Error	T statistic	p-value
ϕ_1(AR 1)	0.6102	0.1296	4.71	0.000
Φ_1 (SAR 12)	1.0010	0.0326	30.69	0.000
θ_1(MA 1)	–0.3175	0.1415	–2.24	0.028
θ_2(MA 2)	–0.3298	0.1337	–2.47	0.016
Θ_1(SMA 12)	1.3250	0.1145	11.57	0.000
Θ_2(SMA 24)	–0.7959	0.1176	–6.77	0.000

$$Z_t = 0.6102Z_{t-1} + 1.001Z_{t-12} - 0.6108Z_{t-13} + a_t + 0.3175a_{t-1} + 0.3298a_{t-2}$$
$$-1.325a_{t-12} - 0.4207a_{t-13} - 0.437a_{t-14} + 0.7959a_{t-24} + 0.2527a_{t-25}$$
$$+0.2625a_{t-26} \quad (8)$$

Then the residuals of the SACF graph was plotted to check the fitted seasonal ARIMA model shown in Figure 2. It was concluded that the forecast error was within 95% confidence interval.

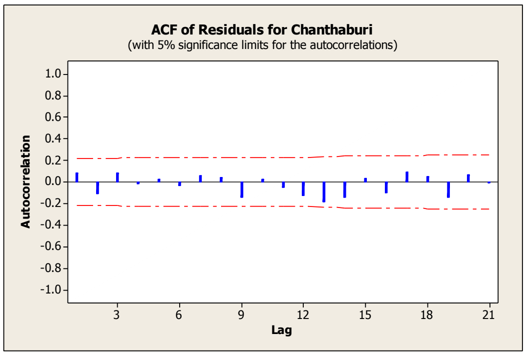


Figure 2: The residuals of SACF graph of ARIMA(1,0,2)(1,0,2)$_{12}$

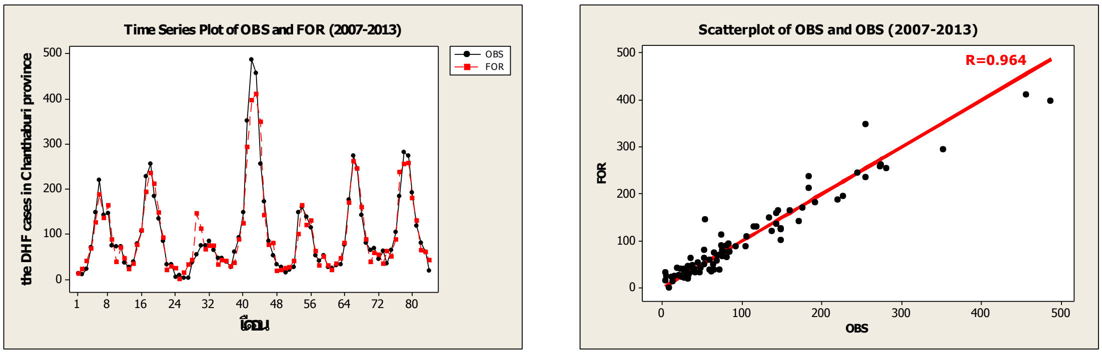
Next, The Ljung-Box Q statistic was determined to validate the forecasted DHF cases in the next succeeding 4 years displayed in Table 2.

Table 2: The Q statistic of ARIMA(1,0,2)$(1,0,2)_{12}$

Lag	12	24	36	48
Chi-square statistic	6.8	19.8	35.1	43.4
Degree of freedom	6	18	30	42
p-value	0.341	0.345	0.238	0.412

The ARIMA(1,0,2)$(1,0,2)_{12}$, was fitted to forecast the DHF cases in Chanthaburi province with Q_{12}=6.8 (p-value=0.341), Q_{24}=19.8 (p-value=0.345), Q_{36}=35.1 (p-value=0.238) and Q_{48}=43.4 (p-value=0.412).

Finally, the accuracy of the ARIMA(1,0,2)$(1,0,2)_{12}$ model was tested by computing the RMSE which was 26.68, comparing time series plot between the observed and forecasted DHF cases as Figure 3a and creating scatter plot between the observed and forecasted DHF cases as Figure 3b contained with 0.964 for the correlation coefficient with p-value=0.000. The ARIMA(1,0,2)$(1,0,2)_{12}$ model was then concluded to appropriately fit to forecast the number of DHF cases in Chonthaburi province because the forecasted values were very close to the observed values as displaying in both of time series and scatter plot. Moreover, the value of correlation coefficient was very high with 0.964. The forecasted DHF cases in Chanthaburi province from 2014 to 2015 also demonstrated in Table 3.



(a) **(b)**

Figure 3: Comparison between the OBS and the FOR; (a) Time series plot, (b) Scatter plot

Table 3: The forecasted DHF cases in Chanthaburi province using ARIMA(1,0,2)(1,0,2)$_{12}$

Month / Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2014	3.61	17.2	22.17	56.49	212.14	289.49	275.04	151.17	86.75	54.83	55.67	42.65
2015	50.44	39.24	47.14	81	192.58	289.23	279.63	173.23	97.99	74.5	68.55	36.04

ACKNOWLEDGEMENT

The authors wish to thank to the Bureau of Epidemiology, National Trustworthy and Competent Authority in Epidemiological Surveillance and Investigation, Thailand for kindly support data.

REFERENCES

[1] Chastel, C., 2012, "Eventual role of asymptomatic cases of dengue for the introduction and spread of dengue viruses in non-endemic regions," Frontiers in physiology, 3, 1-4.

[2] Weaver, S. C., & Vasilakis, N., 2009, "Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease," Infection, Genetics and Evolution, 9(4), 523-540.

[3] Mohammed, H. P., Ramos, M. M., Rivera, A., Johansson, M., Muñoz – Jordan, J. L., Sun, W., & Tomashek, K. M., 2010, "Travel – Associated Dengue Infections in the United States, 1996 to 2005," Journal of travel medicine, 17(1), 8-14.

[4] The Bureau of Epidemiology, 2013, "Dengue fever," Retrived July 30, 2013, from the Bureau of Epidemiology Web site: http://www.boe.moph.go.th/fact/Dengue_ Haemorrhagic_Fever.htm

[5] Pattamak, C., 2014, "Prediction of the number of the cases of dengue hemorrhagic fever in Eastern region of Thailand, 2013-2017," Journal of Health Science, 23, 804-812.

[6] Box, G. E. P., & Jenkins, G. M., 1976, "Time series analysis forecasting and control (2nd ed.)," San Francisco: Holden-day.

[7] Ljung, G. M., & Box, G. E., 1979, "The likelihood function of stationary autoregressive-moving average models," Biometrika, 66(2), 265-270.