Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 1 (2015), pp. 287-293

http://www.ripublication.com

A Seasonal ARIMA Model For Forecasting The Dengue Hemorrhagic Fever Patients In Rayong, 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 proposed research objectified to determine an appropriate model, the seasonal ARIMA model, and forecast the dengue hemorrhagic fever cases in Rayong, Thailand. The estimation of 7 parameters in the seasonal ARIMA model was defined by p=1, d=0, q=2, P=0, D=0, Q=3 and S=12 with no constant or $ARIMA(1,0,2)(0,0,3)_{12}$. Then the performance of model was measured with 35.68 of root mean square error (RMSE) and 0.942 of the correlation coefficient (R) between the observed and forecasted cases (p-value=0.000).

Mathematics Subject Classification: 62-07

Keywords: Dengue Hemorrhagic Fever, Seasonal ARIMA model

1. Introduction

Epidemic of dengue fever is more than 100 countries, mostly in tropical and sub-tropical regions [1][2]. The outbreak of dengue fever caused by many factors, for example, 1) Social change from rural to urban led to a larger urban societies becoming overcrowded, lack of hygiene management and consequently increased mosquito breeding [3], 2) Transportation is more convenient and comfortable causing the spread of dengue [4], 3) The growth of tourism conducts many tourists travel across countries where the outbreak of dengue fever explodes, especially, the countries in Southeast Asia and the Pacific Rim [5] and it was reported that tourists from Europe infected dengue virus after visiting to the Asia region [6], 4) The global warming has a part to the spread of infected mosquitoes with dengue virus in several areas in the world [7][8][9]. In Thailand, the Bureau of Vector - Borne Diseases and the Bureau of Epidemiology, Department of Diseases, Ministry of Public Health is responsible for monitoring dengue fever. Currently, there are 99,452 patients of dengue hemorrhagic fever, a rate of 154.75 per thousand populations, and 94 patients died, a mortality rate of 0.15 per thousand populations. The number of dengue hemorrhagic fever cases in the past 5 years, 2009 – 2013, displayed the dengue hemorrhagic fever was significantly increasing trend [10]. Rayong is one of the eastern provinces in Thailand encountering the epidemic of dengue fever [11]. The number of dengue hemorrhagic fever cases probably rise to 100,000 – 120,000 cases without controlling the outbreak. Then the objective of proposed research aimed to fitting the model to forecast the number of dengue hemorrhagic fever cases in the year 2014 – 2015 in Rayong and present this results to the Bureau of the Vector - Borne Diseases and the Bureau of Epidemiology of Thailand for more efficient planning, controlling and preventing the outbreak of the disease.

2. Materials And Methods

The number of dengue hemorrhagic fever patients in Rayong was collected from the Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health of Thailand, since January 2007 to December 2013.

3. Building The Seasonal Arima Model

3.1 The Stationary Time Series

The Sample Autocorrelation function (SACF) and the Sample Partial autocorrelation function (SPACF) were plotted to check the stationary time series as of Equation 1 and Equation 2 respectively.

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

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

where Z_t be a variable at time t, k be time lag, $\bar{Z} = \sum_{t=1}^{n} Z_t / n$, $\hat{\phi}_{kj} = \hat{\phi}_{(k-1)j} - \hat{\phi}_{kk} \hat{\phi}_{(k-1)(k-j)}$; $j = 1, 2, \ldots, k-1$. Notice that time series set should be more than 50 [12].

3.2 Parameters Estimation

Seven parameters of the ARIMA(p,d,q)$(P,D,Q)_S$, seasonal ARIMA model, were estimated as of Equation 3 [12].

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

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.

4. Checking The Fitted Seasonal Arima Model

4.1 The Residuals

Verifying the seasonal ARIMA model, the SACF graph of residuals was plotted to check. In case of the residuals was within $(1-\alpha)100\%$ confidence interval then the seasonal ARIMA model is fitted.

4.2 The Ljung-Box Q Statistic

Validating the estimators of the seasonal ARIMA model, the Ljung-Box Q statistic was determined as of Equation 4 [13].

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

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. When n is large, the Q statistic was chi-square distribution with degrees of freedom k–p–q.

5. Measurement Of The Fitted Seasonal Arima Model

The root mean square error (RMSE) was calculated to validate the fitted seasonal ARIMA model as of Equation 5. Also, two graphs, were plotted to measure the fitted seasonal ARIMA model. One was the time series plot between the observed (OBS) and the forecasted cases (FOR) of dengue hemorrhagic fever and the other was the scatter plot of the correlation coefficient (R) displayed in Equation 6.

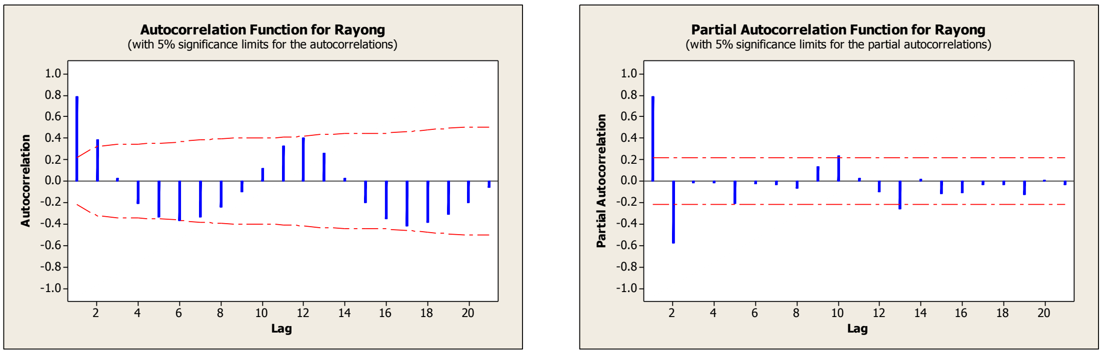
$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(OBS_i-FOR_i)^2} \quad (5)$$

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

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.

6. Results And Discussion

Examining the stationary time series, the SACF and SPACF graph were plotted as of Figure 1. Obviously, the time series was not stationary so the parameters of the seasonal ARIMA model should be estimated by p=1, d=0, q=2, P=0, D=0 and Q=3 with S=12 and no constant or ARIMA$(1,0,2)(0,0,3)_{12}$ which was illustrated in Equation 7.



(a) (b)

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

$$Z_t = \phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13} + \theta_2 \Theta_1 a_{t-14} - \Theta_2 a_{t-24} + \theta_1 \Theta_2 a_{t-25} + \theta_2 \Theta_2 a_{t-26} + \theta_1 \Theta_3 a_{t-27} - \Theta_3 a_{t-36} + \theta_2 \Theta_3 a_{t-38} \quad (7)$$

Table 1: The estimator values and testing of the parameters of ARIMA$(1,0,2)(0,0,3)_{12}$

Parameters	Estimates	Error	T statistic	p-value
ϕ_1 (AR 1)	0.7163	0.0920	7.79	0.000
θ_1 (MA 1)	–0.6676	0.1258	–5.31	0.000
θ_2 (MA 2)	–0.4083	0.1218	–3.35	0.001
Θ_1 (SMA 12)	–0.4308	0.1165	–3.70	0.000
Θ_2 (SMA 24)	–0.6883	0.1237	–5.57	0.000
Θ_3 (SMA 36)	–0.4645	0.1248	–3.72	0.000

All parameter estimates of ARIMA$(1,0,2)(0,0,3)_{12}$ were tested following ϕ_1 (T=7.79, p-value=0.000), θ_1 (T = –5.31, p-value=0.000), θ_2 (T = –3.35, p-value=0.001), Θ_1 (T= –3.70, p-value=0.000), Θ_2 (T = –5.57, p-value=0.000) and Θ_3

(T = –3.72, p-value=0.000) as of Table 1. The conclusion was that ARIMA(1,0,2)(0,0,3)$_{12}$ was appropriate to forecast the number of dengue hemorrhagic fever patients in Rayong as showing in Equation 8.

$$\begin{aligned} Z_t &= 0.7163Z_{t-1} + a_t + 0.6676a_{t-1} + 0.4083a_{t-2} + 0.4308a_{t-12} \\ &+0.2876a_{t-13} + 0.1759a_{t-14} + 0.6883a_{t-24} + 0.4595a_{t-25} \\ &+0.281a_{t-26} + 0.3101a_{t-27} + 0.4645a_{t-36} + 0.1897a_{t-38} \end{aligned} \tag{8}$$

After testing all estimates of the model, then the SACF graph of residuals was plotted to validate ARIMA(1,0,2)(0,0,3)$_{12}$ shown in Figure 2 which the residuals was within 95% confidence interval.

The forecasted dengue hemorrhagic fever patients were verified in the next succeeding 4 years. The Ljung-Box Q statistic values showed in Table 2 with Q_{12}=6.6 (p-value=0.361), Q_{24}=11.9 (p-value=0.853), Q_{36}=20.9 (p-value=0.890) and Q_{48}=31.6 (p-value=0.879). The results exhibited that ARIMA(1,0,2)(0,0,3)$_{12}$ was fitted to forecast the dengue hemorrhagic fever cases in Rayong.

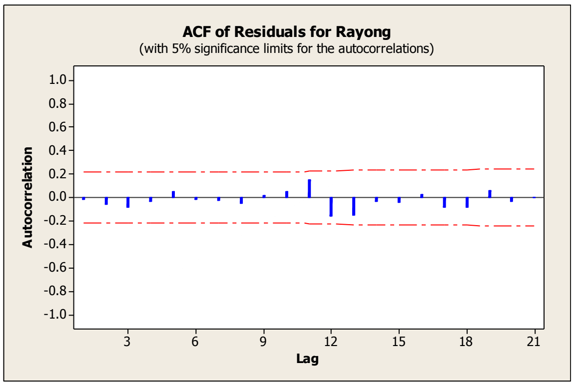
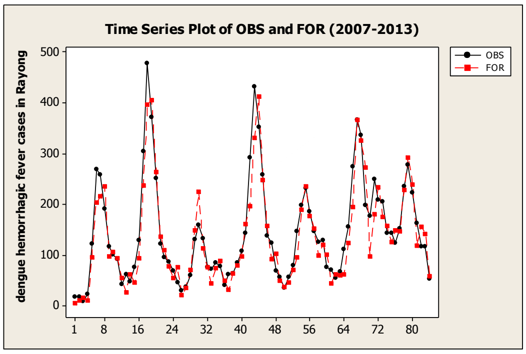


Figure 2: The SACF graph of the residuals of ARIMA(1,0,2)(0,0,3)$_{12}$

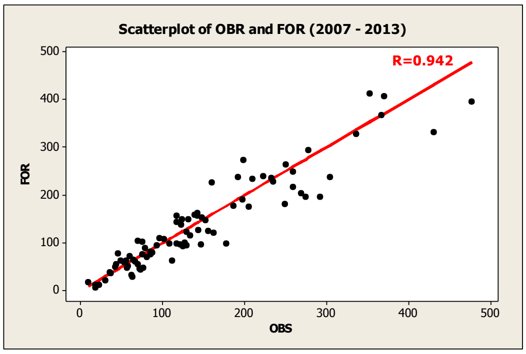
Table 2: The Q statistic of ARIMA(1,0,2)(0,0,3)$_{12}$

Lag	12	24	36	48
Chi-square statistic	6.6	11.9	20.9	31.6
Degree of freedom	6	18	30	42
p-value	0.361	0.853	0.890	0.879

The RMSE of 35.68 was computed to measure accuracy of ARIMA(1,0,2)(0,0,3)$_{12}$. Time series plot between the OBS and the FOR depicted that the forecasted cases were very close to the observed cases shown in Figure 3a and scatter plot between OBS and FOR was generated with the correlation coefficient of 0.942 (p-value=0.000) displayed in Figure 3b. Then the dengue hemorrhagic fever cases in Rayong were forcasted from 2014 to 2015 and 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 cases in Rayong by ARIMA(1,0,2)$(0,0,3)_{12}$

Year \ Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2014	49.55	26.95	42.93	65.61	120.83	191.27	190.3	160.38	79.87	86.24	120.92	102.94
2015	101.78	69.39	69.82	61.09	73.09	109.18	103.79	77.83	42.24	33.78	42.54	29.01

7. 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.

8. References

[1] Rajapakse, S., Rodrigo, C., & Rajapakse, A., 2012, "Treatment of dengue fever. Infection and drug resistance," 5, 103-112.

[2] Brady, O. J., Gething, P. W., Bhatt, S., Messina, J. P., Brownstein, J. S., Hoen, A. G., Moyes, C. L., Farlow, A. W., Scott, T. W., & Hay, S. I., 2012, "Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS neglected tropical diseases," 6(8), e1760.

[3] Arunachalam, N., Tana, S., Espino, F., Kittayapong, P., Abeyewickrem, W., Wai, K. T., Tyagi, B. K., Kroeger, A., Sommerfeld, J., & Petzold, M, 2010, "Eco-bio-social determinants of dengue vector breeding: a multicountry study in urban and periurban Asia," Bulletin of the world Health Organization, 88(3), 173-184.

[4] Mahabir, R. S., Severson, D. W., & Chadee, D. D., 2012, "Impact of road networks on the distribution of dengue fever cases in Trinidad, West Indies," Acta tropica, 123(3), 178-183.

[5] Reilly, M. W., & Kuehne, E. A., 2013, “Dengue Fever in a 16-Year-Old Girl After Travel in Asia,” Journal of Pediatric Health Care, 27(4), 308-311.

[6] Pongsumpun, P., Patanarapelert, K., Sriprom, M., Varamit, S., & Tang, I. M., 2004, “Infection risk to travelers going to dengue fever endemic regions,” Southeast Asian journal of tropical medicine and public health, 35(1), 155-159.

[7] Khasnis, A. A., & Nettleman, M. D., 2005, “Global warming and infectious disease,” Archives of medical research, 36(6), 689-696.

[8] Barclay, E., 2008, “Is climate change affecting dengue in the Americas?,” The Lancet, 371(9617), 973-974.

[9] Beebe, N. W., Cooper, R. D., Mottram, P., & Sweeney, A. W., 2009, “Australia's dengue risk driven by human adaptation to climate change,” PLoS neglected tropical diseases, 3(5), e429.

[10] The Bureau of Epidemiology, 2014, “Dengue fever,” Retrived December 30, 2014, from the Bureau of Epidemiology Web site: http://www.boe.moph.go.th/fact/Dengue_ Haemorrhagic_Fever.htm

[11] 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.

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

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