

Developing time-based model for the prediction of breeding activities of dengue vectors using early life cycle variables and epidemiological information in Northern Malaysia

Nur Aida, H.^{1*}, Abu Hassan, A.², Anita, T.³, Nurita, A.T.⁴, Dieng, H.⁵, Suhaila, A.H.², Rajasaygar, S.², Nur Hafiza, M.² and Farida, A.⁶

¹School of Food Science and Technology, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

²School of Biological Sciences, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia

³School of Distance Education, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia

⁴International School of Kuantan, 25200 Bandar Indera Mahkota, Kuantan, Pahang, Malaysia

⁵Institute of Biodiversity and Environmental Conservation, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

⁶Faculty of Veterinary Medicine, Syiah Kuala University, Banda Aceh, Indonesia

*Corresponding author e-mail: aida.hashim@umt.edu.my; nuraida203@yahoo.com

Received

; received in revised form

; accepted

Abstract: Autoregressive integrated moving average (ARIMA) was applied to make real-time predictions on the *Aedes* egg populations in three selected dengue hotspots of Penang, Malaysia. The weekly ovitrap collection was carried out to determine the abundance of *Aedes* eggs in field population in some selected areas. The ARIMA models were able to estimate actual egg abundance using two criteria. The first criteria is to determine the reliability of statistics and the second is to measure the accuracy of forecasting ability of the model equation. The parsimonious model with a lowest order of AR or MA and RMSE value of the forecast for each data set was considered the best. ARIMA (1,0,0), ARIMA (2,0,0) and ARIMA (0,1,1) models were judged to be the best fit for the suburban, urban squatter and urban area data sets respectively. The models were able to forecast the number of eggs within a range of one to eleven weeks. The developed models were able to estimate the egg abundance adequately to permit their use in *Aedes* control programme in Penang Island. Thus, it can be a useful tool for health officials to improve the management of mosquito control and alert the public to reduce the possibility of dengue outbreaks.

INTRODUCTION

In developing countries, vector borne disease such as dengue fever (DF) not only become a major threat to the communities' health, but also cause a huge economic liability to the health services (Shepard & Suaya, 2009; Guzman & Isturiz, 2010; Kongsin *et al.*, 2010). Its geographical spread is increasing; only nine countries have been recorded with dengue epidemics before 1970. Currently, more than 100 countries all over the world have

experienced dengue endemic. Regions such as the Southeast Asia, Western Pacific and Americas are affected tremendously (WHO, 2013). Its occurrence has increased dramatically, with up to 100 million cases of DF occurring annually world-wide which resulted in 250,000 to 500,000 cases of dengue haemorrhagic fever (DHF) each year in recent decades (Shu & Huang, 2004; WHO, 2004). Out of this number, an estimated 50 million dengue infections occur each year in tropical and subtropical regions alone and approximately 24,000

results in death (Rigau-Perez & Gubler, 1999; Barclay, 2008; Suaya *et al.*, 2009; Guzman & Isturiz, 2010).

In Malaysia, the number of reported dengue cases has been escalating year to year for the last 20 years. The cases reported has seen 18 fold increase from a total of 6543 cases in 1995 to 120,836 cases in 2015. Numerous national anti-mosquito and cleanliness campaigns have not succeeded in reducing the number of reported dengue cases. This condition has raised the concerns of the public and officials in the Ministry of Health (MOH, 2009; MOH, 2015). The causes for the increase of the dengue epidemic are more likely associated with the breakdown of vector control, and the presence of multiple dengue virus serotypes co-circulating (KKM, 2002; KKM, 2005). Rapid population growth, unplanned city development, inadequate public health infrastructures, increased international movement and ineffective vector control programme as well as climate change have been said to contribute to the spread of the disease (Knowlton *et al.*, 2009). Currently, the world was alerted with new emerging mosquito-borne disease, Zika. This disease shared similar vector with dengue, which is *Aedes aegypti*. WHO (2016) reported Asian countries such as Singapore, Philippines, Malaysia, Vietnam and Thailand continue to report new cases. However, it is uncertain whether the recently increase in the number of reported Zika cases is due to an actual increase in the incidence or whether this is the result of enhanced surveillance and awareness.

The presence and abundance of *Ae. aegypti* and *Aedes albopictus* are vital to the transmission of dengue. In the absence of vaccination and effective drugs, vector control has been considered as a vital tool in the prevention and control of DF, DHF and chikungunya virus infection (WHO, 2011), largely by community involvement through insecticide application, source reduction and public awareness campaign (Corbel *et al.*, 2004; Farias *et al.*, 2009). With these approaches, comes the need for rapid and accurate estimations of the

existing field population size and dynamics.

The recent advances in computer technology have generated a considerable interest in population modelling, with the aim of getting better insights into the complexities of population dynamics, mosquito control and disease transmission. Mathematical modelling becomes an interesting tool for creating forecast and control strategies of the vectors (Focks *et al.*, 1995; Rohani *et al.*, 2011). The current indices used for *Aedes* surveillance are not sensitive and accurate enough for forecasting vector densities and dengue outbreaks (Focks *et al.*, 2000; Focks, 2003; Nazri *et al.*, 2012a). Therefore, mathematical models were introduced as tools that are more accurate and useful in creating an early warning mechanism for the prediction and forecasting of dengue outbreaks. Dengue forecasting models have gained much interest in the past decades. Review of literature on the subject shows that a relationship exists between vector population and dengue incidence (Mahadev *et al.*, 2004; Tewari *et al.*, 2004; Gubler, 2005; Chadee, 2009; Luz *et al.*, 2009). Previous studies also tend to use historical data of dengue incidence instead of vector population data to predict dengue occurrence (Nor Azura & Naomie, 2008; Hii *et al.*, 2009; Rohani *et al.*, 2011; Nazri *et al.*, 2012b; Wongkoon *et al.*, 2012). However, previous studies did not indicate the severity of vector occurrence and infestation. Adult mosquito abundance can be predicted using generalized linear models (GLM). Interaction between lagged temperature, relative humidity and one week lagged mosquito abundance is significantly influenced the current mosquito abundance (Simões *et al.*, 2013). In addition, mosquito population dynamics are not the same in different geographical areas. The occurrence of dengue is probably area specific due to the climate influence (Scott & Morrison, 2003). However, in dengue hotspot areas where space spraying and other vector control programmes are regularly conducted, climate will have low predictive power for vector abundance. Therefore, a more

reliable model that is independent of environmental parameters is needed as a predictor to forecast vector population abundance.

In the present study, Autoregressive Integrated Moving Average (ARIMA) models were used to generate the temporal dependence structure of a time series for early prediction of *Aedes* egg population abundance. The main characteristic of time-series modeling is that it only models the relationship between the observed number of eggs at time t (y_t) from the past observations (y_1, y_2, \dots, y_{t-1}), without using any other variables. The main objective of this study is to develop univariate time-series models to estimate and forecast the fluctuation of *Aedes* egg abundance in selected dengue hotspots using vector population data alone. This forecasting offers the potential for improved contingency planning of public health intervention.

MATERIALS AND METHODS

Data source and study sites

In this study, the data for *Aedes* eggs were collected using ovitraps from February 2008 to March 2010. The ovitrap survey was carried out in three locations in Penang Island where dengue transmission is common: Permatang Damar Laut (PDL) (05.27544° N, 100.27104° E) a suburban coastal area; Sungai Nibong Kecil (SGN) (05.33001° N, 100.28722° E) an urban squatter area; Sungai Dua (SG2) (05.35076° N, 100.30251° E) an urban area. The ovitrap survey was conducted outdoors in order to examine the outdoor breeding potential of *Aedes* mosquitoes in artificial containers.

a) Permatang Damar Laut (PDL)

PDL is a suburban area located close to the sea and approximately 1km from Penang International Airport in the southwest district of the island. This area was listed as a dengue hotspot in 2005 (KKM, 2005) and continuously recorded dengue cases throughout the study period. The houses in the area are built systematically, however

some of these houses are surrounded by bushes that are rarely cleared and there are many garbage dump sites made by villagers.

b) Sungai Nibong Kecil (SGN)

SGN is a squatter residential area of a township in the Southwestern coastal area of Penang Island. This area is the most important dengue hotspot in Penang (Maimusa *et al.*, 2017). The area has poor vegetation cover characterised by the presence of shrubs along the drainage system which separate the area from a construction site. The houses in the area, often wooden or brick houses with zinc roofing sheets are built close to one another. There are also several abandoned and dilapidated houses in the area. The drainage system is very poor and often clogged with waste, thus the area is regularly flooded during rainy seasons. Discarded artificial containers and plastic bags can be found littered in the surrounding area of houses. All homes have a tap water supply system, but in some, householders store water in cement tanks for bathing and washing purposes.

c) Sungai Dua (SG2)

SG2 is an urban residential area consisting of ten blocks of flats, each separated by small roads and parking lots. A small river flows alongside the buildings and the area is surrounded by shrubs, weeds and banana plants. The drainage and sanitation system are generally adequate but garbage in the form of artificial containers are often dumped into the drainage system that separates the flats and shop lot area. There are welding and auto repair shops nearby with an accumulation of discarded tires and unwanted vehicles parts.

Trapping, Collecting and Identification of *Aedes* mosquitoes

Ovitraps were used to determine abundance of *Aedes* mosquito populations in this study. The ovitraps were made of milk tin cans sized 15 cm high and 7.5 cm in diameter. All the cans were coated on the outer and inner surfaces with black glossy

paint prior to use. The oviposition substrate used in each can was a hardboard paddle, measuring 3 cm in width, 12.5 cm in length and 0.3 cm thick, which was placed in the can lengthwise with the upper (dry) portion resting against the upper rim of the can and the bottom half of the paddle submerged in the water.

Ovitrap were placed outdoors at selected houses. Thirty ovitraps were placed on the ground at each sampling area with a distance of at least 10 m between each ovitraps. The ovitraps were half-filled with seasoned tap water and a paddle was placed in each ovitrap as an oviposition substrate, which was replaced every six days.

Sample size was calculated using negative binomial model (Elliot, 1973). A reasonable estimate of the field population density as one that performs with a standard error less than 40% (Elliot, 1973). Based on preliminary sampling, this sample size (30 ovitraps) collected more than 75% of the total egg population from each site. This was sufficient for egg population representation.

Servicing of ovitraps was done every six days throughout the 26-months (108 weeks) study period from February 2008 to March 2010. Servicing of ovitraps involved the collection and replacement of paddles, cleaning the cans and replenishing the water in each can. The content of each ovitrap (water and larvae) was poured into individual zip-lock plastic bags (26 cm x 18 cm), sealed and labelled. The paddles were also placed in individual plastic bags and brought back to the laboratory for egg and larvae counting. Each paddle was examined under a dissecting microscope to count and record all eggs present. The number of eggs (from each paddle) and larvae (in the water from each ovitrap) were then pooled together and counted as eggs. Mosquito species identification was carried out by hatching the eggs and rearing the larvae until the emergence of adults. Upon emergence, the adults were identified to species level using keys by

Rueda (2004) and the number of species per ovitrap was recorded. The egg data were summed up and averaged to create mean weekly egg per ovitrap according for each study site.

Non seasonal Box-Jenkins (1970) components

The ARIMA is used to determine the patterns and estimate values of time series data (Box & Jenkins, 1970). The parameters of the ARIMA model were assessed using the Statistical Package for Social Sciences (SPSS) package version 21.0. There are four steps involved in ARIMA process; identification, estimation, diagnostic checking and forecasting. Each of these four steps is explained by the egg abundance. In the identification process, the ACF and PACF generated from the data series are used as a guide to choose one or more ARIMA appropriate models. Comparisons are made between the estimated ACF and PACF from the observed data with various theoretical ACF and PACF. A tentative model which has theoretical ACF and PACF functions that most resemble the estimated ACF and PACF is chosen. In the estimation process which involves estimating the general fitness of the generated models, a model is considered fitting if it satisfies certain statistical criterion. This is done by firstly estimating the parameters (coefficients) and secondly generating the estimated values for the observations. The estimation process will search for estimated coefficients that minimize the differences between the actual and the forecasted data. The estimated coefficients are also subjected to certain conditions known as stationarity and inevitability conditions.

ARIMA procedure: The autoregressive integrated moving averages (ARIMA) model is built from three components. They are Autoregressive (AR), time series after differenced (d) and Moving Average (MA) components. Various combinations of AR and MA will produce different models.

The AR procedure of order (p) is:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \dots + \varepsilon_t$$

The MA procedure of order (q) is:

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \dots + \varepsilon_t$$

ARIMA model equation (p, d, q) is:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \dots + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \dots + \varepsilon_t$$

Where ε_t 's are independent, normally distributed, zero mean and constant variance σ^2 for $t = 1, 2, \dots, n$.

Model adequacy analysis: The values of Root Mean Square Error (RMSE), Maximum Absolute Percentage Error (MaxAPE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Maximum Absolute Error (MaxAE), R^2 , Stationary R^2 and Normalized BIC are commonly used to assess the time series model adequacy. Lowest value for each reliability statistics indicate the model efficiency to estimate and forecast of the egg abundance.

The estimation and diagnostic checking procedures may be performed repeatedly, going back and forth, each time revising and improving the model until an estimated model superior to other models which satisfies all assumptions is found. The best model is obtained with diagnostic of Bayesian's Information Criteria (BIC). Lower value of BIC was preferable.

The Q statistic was used in order to determine the residual adequacy. The formula is:

$$Q = \frac{n(n+2)\sum r_k^2}{(n-k)}$$

r_k = The residual autocorrelation at lag k

n = The number of residuals

The value obtained from Q statistic will be compared with Chi square distribution critical value. The model is considered poor if the p-value of Q statistic is low ($p < \alpha$).

Once the model's fitness has been confirmed, it is then ready to be used to generate the forecasts for future values. The forecast values may be in terms of single-valued items or in terms of confidence intervals. The confidence interval estimates provide the probabilistic measures of certainty and uncertainty associated with the forecast values.

Performance or suitability of the model for making forecasts can be investigated by comparing forecast values and real observed values. The criterion normally used to differentiate between a poor forecast model and a good forecast model is called the error "measure". Typically, for a time series of length N, the ARIMA model is usually estimated using $\{Y_1, Y_2, \dots, Y_{N-m}\}$. The forecasting ability of the model is then tested on the remaining m observations. These m observations are sometime referred to as holdout observations. The holdout is a historical series point that is not used in the computation of the model parameters, thus removing its effect on the computation of forecasts. By forcing the model to predict values that actually known (observed values), the model forecasting ability can be evaluated. Thus, m is chosen such that the within-sample observations consist of 95% of the original sample series. This method was illustrated in the present study by holding out the data from week 98 through week 108 in each data series (PDL, SGN and SG2). The data prior to week 98 are used to build the model which is then used to forecast the mean number of eggs in week 98 to week 108.

All tests were conducted at the 5% level of significance, and data analysis was performed using the Time Series analysis (ARIMA) in Statistical Packages for Social Science version 21.0 (SPSS 21.0).

RESULTS

Permatang Damar Laut (PDL)

The autocorrelation function (ACF) and partial autocorrelation function (PACF) correlogram of the PDL mean number of

eggs were generated for a lag of 26 and are shown in Figure 1A and 1B respectively. The ACF of PDL tails off rather sharply; indicating that the mean of the series is stationary. As shown in Figure 1B, the PACF of the series seems to cut off at lag number one and PACF at other lags are generally small and insignificant. No obvious seasonal pattern is observed on both ACF and PACF. Thus, the observations suggest that the PDL series follows ARIMA (1, 0, 0) model.

The PDL series was then fitted with the ARIMA (1, 0, 0) process along with the constant and the result is shown in Table 1a. Since the coefficient of AR (1) is significant at the 5% level and the stationary condition is satisfied, analysis is further carried out by checking the adequacy of the model and generating plots of the ACF and PACF residuals (Figure 2).

Generally, all the ACF and PACF of the residuals fall within Bartlett's intervals (Figure 2). The plots also show that the individual residuals are white noise and group residuals was not significant ($\chi^2=23.384$, $df = 17$, $P>0.05$) which is indicated by the Box-Ljung Q statistic. This suggests that the individual and group residuals are not autocorrelated. In addition, this model has minimum normalized BIC (Table 1b). Thus, the ARIMA (1,0,0) model fitted to PDL series is adequate and able to fully capture the autocorrelation among the observations.

Thus, the general form of ARIMA (1,0,0) model is as follows:

$$Y_t = c + \phi_1 Y_{(t-1)} + \varepsilon_t \quad (1)$$

Where c is the constant, ϕ_1 was the parameter or the model coefficient and ε_t was a time-series of random shocks or white noise process at time t . The coefficients were estimated as an autoregressive $\phi_1 = 0.702$ ($t_1=9.548$, $p<0.05$) and the autoregressive coefficients were very close to their limit of stationarity, $-1<\phi_1<1$. The final model is as follows:

$$Y_t = 29.190 + 0.702Y_{(t-1)} + \varepsilon_t \quad (2)$$

This model revealed that the mean number of eggs per ovitrap at time t were approximately 70.2% of the mean number of eggs per ovitrap at time $t-1$ plus a white noise process. The observed and fitted mean number of eggs from week 1 (February 2008) to week 97 (January 2010) matched reasonably well (Figure 5A). The estimated data (fitted values minus observed values) were normally distributed (Shapiro-Wilk, $p > 0.05$). No significant difference was found between the actual number of eggs and the fitted number of eggs by the ARIMA (1,0,0) model (independent t-test $t = -0.11$, $df = 96$, $p = 0.991$).

The model in equation (2) was used to forecast the mean number of eggs at time t (Y_t) for 11 sequential future weeks (i.e. week 98-week 108) based on the last obtainable data point Y_{97} (i.e. mean number of eggs in week 97) as the forecasting origin. The ARIMA (1,0,0) model is capable to forecast the number of eggs per ovitrap from 1 to 4 weeks before sampling is carried out. The forecasted values minus observed values were normally distributed (Shapiro-Wilk $p > 0.05$). The actual number of eggs and the forecasted number of eggs was not significantly different (Independent t-test $t = -2.442$, $df = 3$, $p = 0.092$). The mean number of eggs after the fourth week of forecast increased and reached a maximum forecasted level in week 108 (March 2010) (Figure 5A). The forecasted values after week 100 began to level out because the autoregressive (AR) process was no longer influencing the forecast.

Sungai Nibong Kecil (SGN) and Sungai Dua (SG2)

The Figures of ACF and PACF are shown in Figure 1C and D, and 1E and F for SGN and SG2 respectively. In the SGN series, the tentative ARIMA models are described and the chosen model has a minimum normalized BIC (Bayesian Information Criterion). Similarly, the same criteria were applied for the SG2 series but with the differentiation process.

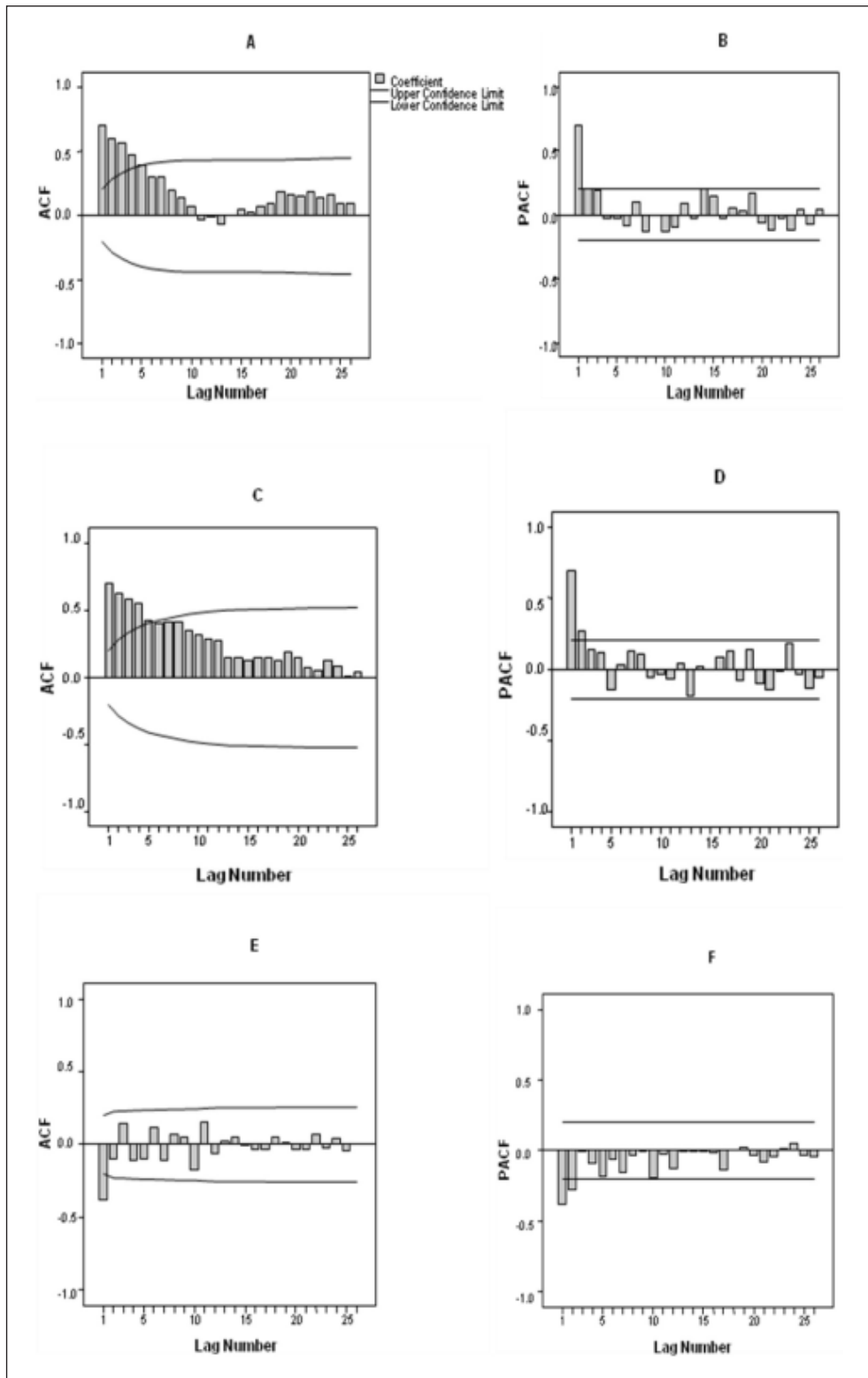


Figure 1. ACF and PACF of the mean number of eggs per ovitrap in different study sites. The letters [A, B] refer to ACF and PACF in PDL accordingly; [C, D] and as [E, F] indicate the same parameters in SGN and SG2 respectively.

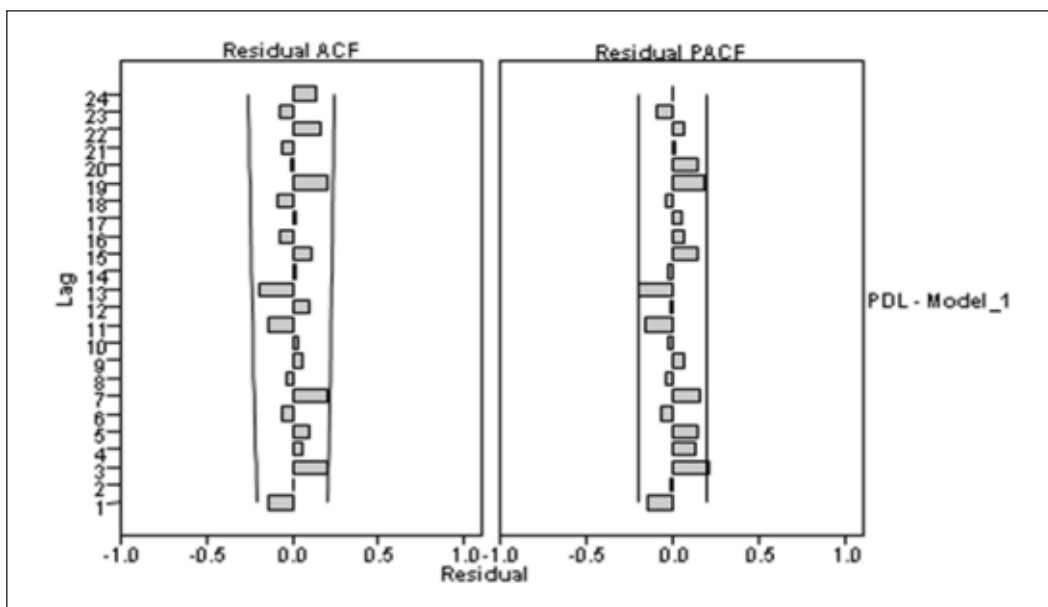


Figure 2. ACF and PACF of the residuals of the ARIMA (1,0,0) model.

Table 1a. Estimated model parameters of ARIMA (1,0,0) with constant and model statistics

Variable	Coefficient	S.E.	t-statistic	p-sig
C	29.190	4.276	6.826	0.000
AR(1)	0.702	0.074	9.548	0.000
Stationary r^2	0.495			
R^2	0.495			
RMSE	12.822			
MAPE	50.950			
MaxAPE	363.427			
MAE	9.666			
MaxAE	44.981			
BIC	5.197			

Table 1b. BIC values of ARIMA (p,d,q) for PDL series

ARIMA (p,d,q)	BIC values
(1,0,0)	5.197
(2,0,0)	5.206
(3,0,0)	5.221

The parameters of the models together with normalized BIC values are presented in Tables 2a and b and 3a and b for SGN

and SG2 respectively. For the SGN series, the model with the best fit is ARIMA (2,0,0). While ARIMA (0,1,1) was the most suitable model for SG2 series as this model had the lowest BIC value.

The ACF and PACF of the residuals for SGN are shown in Figure 3. All lags of ACF and PACF residuals fall within the Bartlett's intervals. Thus the individual residuals are white noise and the Box-Ljung Q test was not significant ($\chi^2 = 19.241$, $df = 16$, $p > 0.05$) which suggest that the residuals are not

Table 2a. Estimated model parameters of ARIMA (2,0,0) with constant and model statistics for SGN

Variable	Coefficient	S.E.	t-statistic	p-sig
C	38.150	7.134	5.347	0.000
AR(1)	0.502	0.099	5.051	0.000
AR(2)	0.273	0.099	2.746	0.007
Stationary r^2	0.529			
R^2	0.529			
RMSE	16.551			
MAPE	48.915			
MaxAPE	309.453			
MAE	12.204			
MaxAE	49.503			
BIC	5.754			

Table 2b. BIC values of ARIMA (p,d,q) for SGN series

ARIMA (p,d,q)	BIC values
(1,0,0)	5.754
(2,0,0)	5.790
(3,0,0)	5.792

Table 3a. Estimated model parameters of ARIMA (0,1,1) with constant and model statistics for SG2

Variable	Coefficient	S.E.	t-statistic	p-sig
MA(1)	0.596	0.079	7.518	0.000
Difference	1			
Stationary r^2	0.214			
R^2	0.529			
RMSE	17.670			
MAPE	46.492			
MaxAPE	309.453			
MAE	409.932			
MaxAE	12.223			
BIC	79.247			

Table 3b. BIC values of ARIMA (p,d,q) for SG2 series

ARIMA (p,d,q)	BIC values
(0,1,1)	5.787
(0,1,2)	5.827
(1,1,1)	5.850
(2,1,0)	5.881

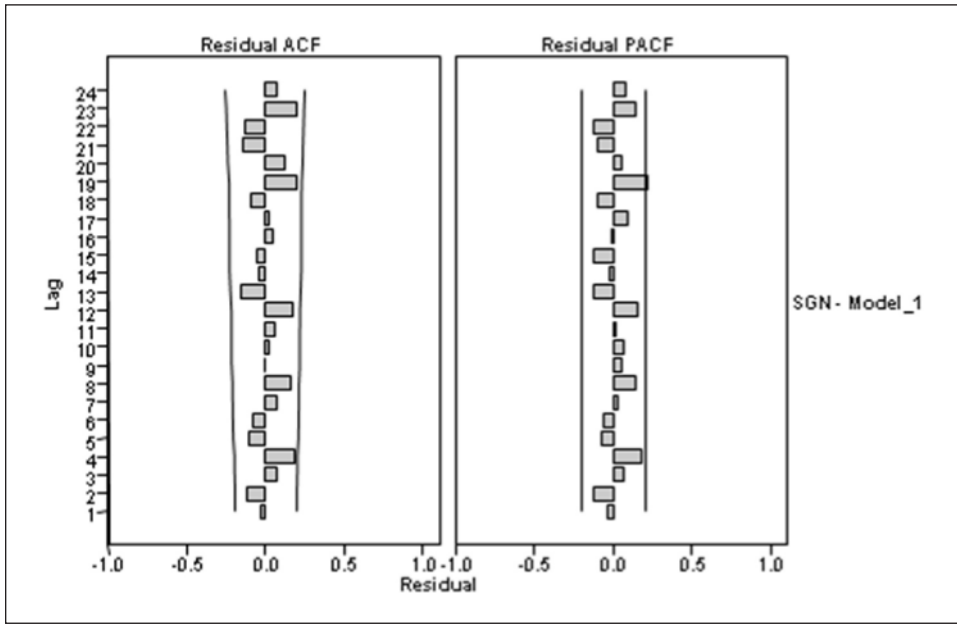


Figure 3. ACF and PACF of the residuals of the ARIMA (2,0,0) model for SGN.

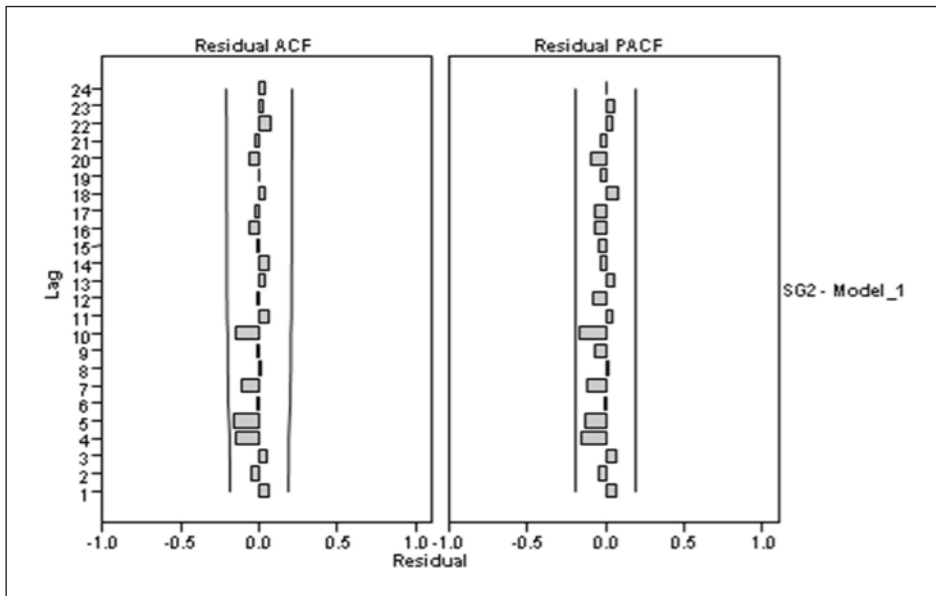


Figure 4. ACF and PACF of the residuals of the ARIMA (0,1,1) model for SG2.

autocorrelated. The general form of ARIMA (2,0,0) model is as follows:

$$Y_t = c + \phi_1 Y_{(t-1)} + \phi_2 Y_{(t-2)} + \varepsilon_t \quad (1)$$

Where c is the constant term, ϕ_1 and ϕ_2 were the parameters or the model coefficients

and ε_t was a time-series of random shocks or white noise process at time t . The coefficients were estimated as an autoregressive $\phi_1 = 0.502$ ($t_1 = 5.051$, $p < 0.05$) and autoregressive $\phi_2 = 0.273$ ($t_2 = 2.746$, $P < 0.05$). The autoregressive coefficients values follow their limit of stationarity

conditions, $\phi_1 + \phi_2 < 1$, $-1 < \phi_2 < 1$, $\phi_1 - \phi_2 < 1$. The final model of ARIMA (2,0,0) for SGN is as follows:

$$Y_t = 38.150 + 0.502Y_{(t-1)} + 0.273Y_{(t-2)} + \varepsilon_t \quad (2)$$

This model revealed that the mean number of eggs per ovitrap at time t were approximately 77.75% of the mean number of eggs per ovitrap at time t_{-1} plus a white noise process. The model demonstrates that the current mean number of eggs in SGN is influenced by the mean number of eggs in the previous 2 weeks. The observed and fitted mean number of eggs from week 1 (February 2008) to week 97 (January 2010) matched reasonably well (Figure 5B).

The model, equation (2), was used to forecast mean number of eggs at time t (Y_t) for 11 sequential future weeks (i.e. week 98-week 108) based on the last obtainable data point Y_{97} (i.e. mean number of eggs in week 97) as the forecasting origin. The ARIMA (2,0,0) model is able to forecast the number of eggs per ovitrap from 1 to 11 weeks before sampling is commenced. The observed number of eggs and forecasted number of eggs did not differ significantly (Wilcoxon Signed Ranks Test $z = -0.711$, $p = 0.477$). Figure 5B shows the ability of the model to forecast future egg numbers in SGN up to an 11 week interval.

Table 4a shows the results of estimation of coefficients for ARIMA (0,1,1) model without a constant for SG2 series. The model coefficient of MA(1) is significant and the value is less than 1 which fulfils the stationary condition of MA(1) process.

All lags of ACF and PACF residuals fall within Bartlett's intervals, indicating that the individual residuals are white noise which means there was no autocorrelation. The Box-Ljung Q test is insignificant ($\chi^2 = 12.438$, $df = 17$, $p > 0.05$) suggesting that the residuals were also not autocorrelated as a group. The results show that ARIMA (0,1,1) is the most adequate model and satisfies the rule of parsimony for SG2 data series.

The general form of the ARIMA (0,1,1) model is as follows:

$$Y_t = Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} \quad (1)$$

Where θ_1 is the moving average parameter or the model coefficient, ε_{t-1} is the error term or white noise process at time t-1. The coefficients were estimated as moving average $\theta_1 = 0.596$ ($t_1 = 7.518$, $p < 0.05$), coefficient values were stationary ($-1 < \theta_1 < 1$) and ε_t was distributed with a mean 0 and variance 1.

The final model is as follow:

$$Y_t = Y_{t-1} + \varepsilon_t - 0.596\varepsilon_{(t-1)} \quad (2)$$

The current value mean number of eggs per ovitrap at time t (Y_t) is therefore, a weighted average of past values plus an innovation. This model demonstrated that the current mean number of eggs in SG2 was influenced by the number of eggs in the previous week. Figure 5C shows that the predictions from ARIMA (0,1,1) performed reasonably well for both training and validation data from week 1 (February 2008) to week 97 (January 2010). The data for this model were normally distributed (Shapiro-Wilk $p > 0.05$) and no significant difference was found between the actual and fitted number of eggs (independent t-test $t = -0.565$, $p = 0.574$).

This model is capable of forecasting the number of eggs per ovitrap from 1 to 2 weeks before sampling. However, after the short interval of one to two weeks, the forecast values in the model essentially 'levels out' (Figure 5C). This characteristic occurred when the moving average (MA) ceased to affect the forecast, which means the model is no longer able to forecast egg abundance. Therefore, MA is more suitable for short term forecasting of one to two weeks.

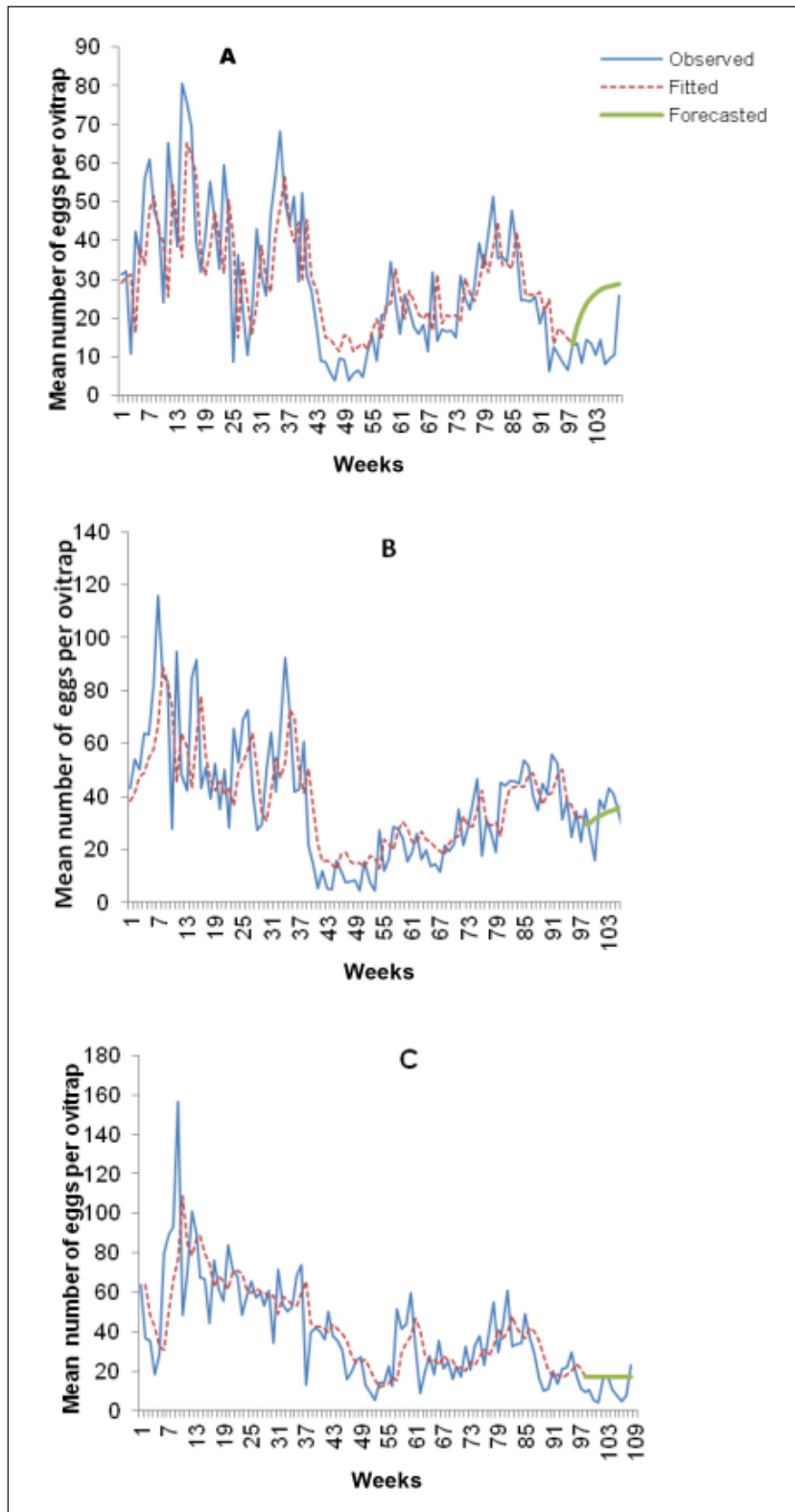


Figure 5. Observed, fitted and forecast data of *Aedes* eggs for PDL (A), SGN (B) and SG2(C) series.

DISCUSSION

The present study clearly showed that a population model based on the population count can be developed using time series analysis which is ARIMA. Based on the results, the estimated data plots are corresponding to the actual data.

The model ARIMA (1,0,0) successfully forecasted the egg abundance pattern in PDL from 1 to 4 weeks before sampling was conducted, while ARIMA (2,0,0) was able to forecast the egg abundance in SGN from 1 to 11 weeks before commencement of sampling. For autoregressive models ARIMA (1,0,0) and ARIMA (2,0,0), each value in the model (autoregression) is a weighted average of recent values of the series. Since these values are weighted averages of previous values, the effect of a given disturbance in an autoregressive process decrease as time passes. The model ARIMA (0,1,1) was able to forecast the mean number of eggs in SG2 one week prior to sampling. This model is a moving average process where each value in the series is an average of the most recent residuals. The residual affects the system for a finite number of periods (the order of the moving-average) and then briskly stops to affect it. Therefore in practical terms, autoregressive (AR) processes are more useful for modelling longer-term effects as demonstrated in PDL and SGN series, while moving average (MA) processes are more useful for modelling short-term fluctuations as demonstrated in SG2 series.

The vector density forecast as demonstrated in the present study could alert health officials to intensify mosquito control efforts and to inform the public about the increased risk of acquiring dengue. Conventional vector control is usually conducted after dengue outbreaks are reported (Rohani *et al.*, 2011). Therefore, once the dengue virus is introduced into a human population through the vectors, it is often too late to kill the infected mosquitoes. Thus fogging after or during the outbreak has little impact on the spread of the disease and by the time an

outbreak is reported, 7-10 days would have passed before control is conducted. Therefore, by using forecasting models such as ARIMA, control measure can be conducted before the outbreak occurs.

To use the ARIMA models for the mosquito control programme, three vital information is required. The first information is the number of dengue cases occurring in the area. The information on dengue cases is to determine the onset level before the egg population number reaches the risk of dengue transmission. The second information is the cost of chemical control (i.e. fogging and ULV). This information can be used to estimate the cost of control interventions. The third information is a constant record of *Aedes* egg populations to be incorporated in the models to forecast when the *Aedes* population will essentially reach the transmission risk which will help create more cost-efficient control programmes.

The Malaysian Ministry of Health (MOH) has documented the number of dengue cases and chemical interventions conducted in the areas. Thus, the MOH only needs to maintain a constant collection of ovitrap data. The continuous egg data collection will permit the variation in the *Aedes* population to be quickly integrated with the developed model. To ensure the success of vector management and to reduce dengue transmission risk in the area, it is crucial for the vector surveillance information to be recorded constantly although it might be tedious to collect data continuously. Using the ARIMA models to forecast *Aedes* populations requires the MOH operator or personnel to take a small sample size, for example 10-30 ovitraps per week, depending on the size of the area. With sufficient data, it may prove possible to apply ARIMA models for biweekly or monthly ovitrap data depending on the type of forecast that is required, whether biweekly or monthly forecasts. In dengue hot spot areas, a more frequent data collection schedule might be more meaningful in terms of producing more accurate predictions.

The abundance of mosquito vector strongly related to disease transmission. Therefore by recognizing the biological and environmental factors that contributed to vector abundance will assist the Ministry of Health to estimate the suitable period for the most economical and effective vector control programme. ARIMA models provide useful tools for administrators and the Ministry of Health in planning the budget required in mosquito control programmes and the exact time to start interventions. This model can also be used to optimize dengue prevention by providing estimates on vector population fluctuations. Accurate predictions for even a few weeks ahead as demonstrated in the present study provides an invaluable advantage as it provides an opportunity for fast mobilization of preventive vector control measures or prepared for hospital demand. Previous researchers (Regis *et al.*, 2008) used egg averages to recognize areas with high vector abundance. Arboviral surveillance activities were increased corresponding with the increase of vector population in order to prevent dengue and other vector borne diseases. Information on the type of virus circulating, vector control programme and vector abundance forecast in the particular area can be an early warning system components of dengue transmission risk (Rohani *et al.*, 2011).

The findings of the present study demonstrate that vector egg data from constant surveillance together with simple time series model is adequate to produce prediction on the peaks of vector abundance. Authorities such as Ministry of Health or Municipal council can adopt these approaches to obtain optimum vector control and thus decrease the risk vector borne disease among the community.

Urbanization and population growth in Penang Island has fostered the dengue fever transmission especially in residential areas where inappropriate waste management at the household level has created an abundance of artificial containers in the surrounding premises which will eventually become mosquito breeding sites. Information from epidemiological data in

Malaysia shows that dengue virus infection is predominantly in urban areas, where 61.8% of the total population lives, in which rapid industrial and economic development has created many man-made opportunities for *Aedes* mosquito breeding (Teng, 2001). In southern Malaysia, 76% of dengue fever cases are reported within residential areas followed by 8% from squatter areas and 5% from rural areas (Seng *et al.*, 2005). This information has reaffirmed the association of dengue transmission with common breeding grounds associated with humans especially in developing and domestic environments. Surveys on human behaviour in regards to key activities which could encourage dengue transmission such as gardening and waste management might reveal important information regarding the widespread occurrence of dengue fever in the domestic environment. Entomologic surveys (i.e. larval and pupal surveys) that might assist in determining possible breeding site attributes of the dengue vectors, coupled with the time series model proposed, will produce a very effective and reliable early warning tool to combat dengue outbreaks.

Acknowledgments. We would like to thank the Dean of the School of Biological Sciences of Universiti Sains Malaysia (USM) for the facilities provided. We also thank Miss Rosmawati Hussein and Mr. Kalimuthu Supramaniam for their technical assistance. This work was supported by the National Science Fellowship Scheme from the Ministry of Science, Technology & Innovation; and the Research University Postgraduate Research Grant Scheme (USM-RU-PRGS) [grant number 1001/PBIOLOGY/841011].

REFERENCES

- Barclay, E. (2008). Is climate change affecting dengue in the Americas? *Lancet* **371**: 973-974.
- Box, G.E.P. & Jenkins, G.M. (1970). *Time series analysis: forecasting and control* San Francisco, Holden-Day.

- Chadee, D.D. (2009). Dengue cases and *Aedes aegypti* indices in Trinidad, West Indies. *Acta Tropica* **112**: 174-80.
- Corbel, V., Duchon, M., Zaim, M. & Hougand, J.M. (2004). Dinotefuran: a potential neonicotinoid insecticide against resistant mosquitoes. *Journal of Medical Entomology* **41**: 712-717.
- Elliot, J.M. (1973). *Some methods for the statistical analysis of samples of benthic invertebrates*. The Ferry House, Ambleside, Cumbria: Freshwater Biological Association Scientific Publication.
- Farias, D.F., Cavalheiro, M.G., Viana, S.M., De Lima, G.P.G., Da Rocha-Bezerra, L.C.B., Ricardo, N.M.P.S. & Carvalho, A.F.U. (2009). Insecticidal action of sodium anacardate from Brazilian cashew nut shell liquid against *Aedes aegypti*. *Journal of the American Mosquito Control Association* **25**: 386-389.
- Focks, D.A., Daniels, E., Haile, D.G. & Keesling, J.E. (1995). A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *The American Journal of Tropical Medicine and Hygiene* **53**: 489-506.
- Focks, D.A., Brenner, R.J., Hayes, J. & Daniels, E. (2000). Transmission thresholds for dengue in terms of *Aedes aegypti* pupae per person with discussion of their utility in source reduction efforts. *American Journal of Tropical Medicine and Hygiene* **62**: 11-18.
- Focks, D.A. (2003). A review of entomological sampling methods and indicators for dengue vectors TDR/IDE/Den/03.1. Geneva, Switzerland: World Health Organization. 38p.
- Gubler, D.J. (2005). The emergence of epidemic dengue fever and dengue hemorrhagic fever in the Americas: a case of failed public health policy. *Revista Panamericana de Salud Pública* **17**: 221-4.
- Guzman, A. & Isturiz, R.E. (2010). Update on the global spread of dengue. *International Journal of Antimicrobial Agents* **1**: S40-2.
- Hii, Y.L., Rocklov, J., Ng, N., Tang, C.S., Pang, F.Y. & Sauerborn, R. (2009). Climate variability and increase in intensity and magnitude of dengue incidence in Singapore. *Global Health Action* [Internet] 2. doi: 10.3402/gha.v2i0.2036. [Accessed January 11, 2010]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20052380>
- KKM (Kementerian Kesihatan Malaysia). 2002. Laporan Tahunan 2002. Cawangan Penyakit Bawaan Vektor, Kuala Lumpur, Malaysia.
- KKM (Kementerian Kesihatan Malaysia). 2005. Laporan Tahunan 2005. Cawangan Penyakit Bawaan Vektor, Pulau Pinang, Malaysia.
- Knowlton, K., Solomon, G. & Rotkin-Ellman, M. (2009). Natural Resources Defence Council (NRDC) Fever Pitch: Mosquito-Borne Dengue Fever Threat Spreading in the Americas. *Natural Resources Defence Council*: 1-16.
- Kongsin, S., Jiamton, S., Suaya, J.A., Vasanawathana, S., Sirisuvan, P. & Shepard, D.S. (2010). Cost of dengue in Thailand. *Dengue Bulletin* **34**: 77-88.
- Luz, P.M., Mendes, B.V.M., Codeço, C.T., Struchiner, C.J. & Galvani, A.P. (2008). Time Series Analysis of Dengue Incidence in Rio de Janeiro, Brazil. *American Journal of Tropical Medicine and Hygiene* **79**: 933-939.
- Mahadev, P.V.M., Fulmali, P.V. & Mishra, A.C. (2004). A preliminary study of multilevel geographic distribution and prevalence of *Aedes aegypti* (Diptera: Culicidae) in the state of Goa, India. *Indian Journal of Medical Research* **120**: 173-82.
- Maimusa, A.H., Abu Hassan, A., Abu Kassim, N.F., Hamdan, A., Dieng, H. & Rahim, J. 2017. Contribution of public places in proliferation of dengue vectors in Penang Island, Malaysia. *Asian Pacific Journal of Tropical Biomedicine* **7**(3): 183-187.

- MOH (Ministry of Health). 2009. Annual Report 2009. Kuala Lumpur, Malaysia: Vector-borne Disease Branch, Ministry of Health, Malaysia.
- MOH (Ministry of Health). 2015. Press Release 2015. Kuala Lumpur, Malaysia: Vector-borne Disease Branch, Ministry of Health, Malaysia. [Accessed on April 2, 2016]. Available from: http://www.infosihat.gov.my/infosihat/isusemasa/denggi_2016/pdf/KPK_UNTUK_DENGGI_MINGGU_01_2016.pdf
- Nazri, C.D., Biswajeet, P., Abu Hassan, A., Zulkiflee, A.L. & Rodziah, I. (2012a). Coupling of remote sensing data and environmental related parameters for Dengue transmission risk assessment in Subang Jaya, Malaysia, *Geocarto International* [Internet] DOI:10.1080/10106049.2012.696726. [accessed May 24 2012]. Available from: <http://www.tandfonline.com/loi/tgei20>
- Nazri, C.D., Zulkiflee, A.L., Abu Hassan, A., Rodziah, I. & Biswajeet, P. (2012b). Manifestation of GIS Tools for Spatial Pattern Distribution Analysis of Dengue Fever Epidemic in the City of Subang Jaya *Malaysia Environment Asia* **5**: 82- 92.
- Nor Azura, H. & Naomie, S. (2008). A Comparative study for back propagation neural network and nonlinear models for predicting dengue outbreak. *Jurnal Teknologi Maklumat* **20**: 97-114.
- Regis, L., Montero, A.M., Mello-Santos, M.A.V., Silveira, J.J.C., Furtado, A.F., Acroli, R.V., Santos, J.M., Nakazawa, M., Carvalho, M.S. & Sauza, W.V. (2008). Developing new approaches for detecting and preventing *Aedes aegypti* population outbreaks: Basis for surveillance, alert and control system. *Memorias Institute Oswaldo Cruz, Riode Jeneiro* **103**: 50-59.
- Rigau-Perez, J.G. & Gubler, D.J. (1999). Is there an inapparent dengue explosion? Reply. *Lancet* **353**(9158): 1101-1101.
- Rohani, A., Suzilah, I., Malinda, M., Anuar, I., Mohd Mazlan, I., Salmah Maszaitun, M., Topek, O., Tanrang, Y., Ooi, S.C., Rozilawati, H. & Lee, H.L. (2011). *Aedes* larval population dynamics and risk for dengue epidemics in Malaysia. *Tropical Biomedicine* **28**: 237-248.
- Rueda, L.M. (2004). Pictorial keys for the identification of mosquitoes (Diptera: Culicidae) associated with Dengue virus transmission. *Zootaxa* **589**: 1-60.
- Scott, T.W. & Morrison, A.C. (2003). *Aedes aegypti* density and the risk of dengue virus transmission. In: Takken W, Scott TW (eds). 2003. Ecological aspects for application of genetically modified mosquitoes. FRONTIS, Dordrecht, The Netherlands, pp 187-206.
- Seng, S.B., Chong, A.K. & Moore, A. (2005). Geostatistical Modelling, Analysis and Mapping of Epidemiology of Dengue Fever in Johor State, Malaysia Presented at SIRC 2005. In: The 17th Annual Colloquium of the Spatial Information Research Centre. 2005 November 24-25; Dunedin. New Zealand: University of Otago. p 15.
- Shepard, D.S. & Suaya, J.A. (2009). Economic evaluation of dengue prevention. In W. Ungar (Ed.), Economic Evaluation in Child Health. Oxford: Oxford University Press.
- Shu, P.Y. & Huang, J.H. (2004). Current advances in dengue diagnosis. *Clinical and Diagnostic Laboratory Immunology* **11**: 642-650.
- Simões, T.C., Codeço, C.T., Nobre, A.A. & Eiras, Á.E. (2013). Modeling the non-stationary climate dependent temporal dynamics of *Aedes aegypti*. *PloS One* **8** (8): e64773. doi: 10.1371/journal.pone.0064773.

- Suaya, J.A., Shepard, D.S., Siqueira, J.B., Martelli, C.T., Lum, L.C., Tan, L.H., Kongsin, S., Jiamton, S., Garrido, F., Montoya, R., Armien, B., Huy, R., Castillo, L., Caram, M., Sah, B.K., Sughayyar, R., Tyo, K.R. & Halstead, S.B. (2009). Cost of dengue cases in eight countries in the Americas and Asia: a prospective study. *The American Journal of Tropical Medicine and Hygiene* **80**(5): 846-55.
- Teng, A.K. & Singh, S. (2001). Epidemiology and new initiatives in the prevention and control of dengue in Malaysia. *Dengue Bulletin* **25**: 7-14.
- Tewari, S.C., Thenmozhi, V., Katholi, C.R., Manavalan, R., Munirathinam, A. & Gajanana, A. (2004). Dengue vector prevalence and virus infection in a rural area in south India. *Tropical Medicine & International Health* **9**: 499-507.
- WHO (World Health Organization). (2004). Using Climate to Predict Infectious Disease Outbreaks: A Review. Geneva: World Health Organization. p 55.
- WHO (World Health Organization). (2011). Comprehensive guidelines for prevention and control of dengue and dengue haemorrhagic fever. [Accessed on April 7, 2015]. Available from: http://apps.searo.who.int/pds_docs/B4751.pdf
- WHO (World Health Organization). (2013). Dengue and severe dengue: Fact sheet N°117. [Accessed December 5 2013]. Available from: <http://www.who.int/mediacentre/factsheets/fs117/en/>
- WHO (World Health Organization) (2016). Zika situation report. [Accessed on September 20, 2016]. Available from: <http://www.who.int/emergencies/zika-virus/situation-report/22-september-2016/en/>
- Wongkoon, S., Jaroensutasinee, M. & Jaroensutasinee, K. (2012). Development of temporal modelling for prediction of dengue infection in Northeastern Thailand. *Asian Pasific Journal of Tropical Medicine*: 249-252.