



Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Disease

journal homepage: www.elsevier.com/locate/apjtd

Document heading

doi:10.1016/S2222-1808(13)60084-5

© 2013 by the Asian Pacific Journal of Tropical Disease. All rights reserved.

Generating temporal model using climate variables for the prediction of dengue cases in Subang Jaya, Malaysia

Nazri Che Dom^{1,2*}, A Abu Hassan¹, Z Abd Latif³, Rodziah Ismail²¹School of Biological Sciences, University of Science Malaysia, Pulau Pinang, Malaysia²Faculty of Health Sciences, University of Teknologi MARA, Selangor, Malaysia³Faculty of Architecture, Planning and Surveying, University of Teknologi MARA, Selangor, Malaysia

PEER REVIEW

Peer reviewer

Dr Zulhabri Ismail

Centre of Studies for Construction Management & Law, Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA, Malaysia.

Tel: 03-55444191, 03-55444353

E-mail: zulhabri@ieec.org

Comments

This is a good study in which the authors have developed a forecasting model for the incidence of dengue cases in Subang Jaya using time series analysis. The result of this study are interesting and revealed that the ARIMA model developed, closely described the trends of dengue incidence and confirmed the existence of dengue fever cases in Subang Jaya for the year 2005 to 2010.

Details on Page 360

ABSTRACT

Objective: To develop a forecasting model for the incidence of dengue cases in Subang Jaya using time series analysis.

Methods: The model was performed using the Autoregressive Integrated Moving Average (ARIMA) based on data collected from 2005 to 2010. The fitted model was then used to predict dengue incidence for the year 2010 by extrapolating dengue patterns using three different approaches (*i.e.* 52, 13 and 4 weeks ahead). Finally cross correlation between dengue incidence and climate variable was computed over a range of lags in order to identify significant variables to be included as external regressor.

Results: The result of this study revealed that the ARIMA (2,0,0) (0,0,1)₅₂ model developed, closely described the trends of dengue incidence and confirmed the existence of dengue fever cases in Subang Jaya for the year 2005 to 2010. The prediction per period of 4 weeks ahead for ARIMA (2,0,0) (0,0,1)₅₂ was found to be best fit and consistent with the observed dengue incidence based on the training data from 2005 to 2010 (Root Mean Square Error=0.61). The predictive power of ARIMA (2,0,0) (0,0,1)₅₂ is enhanced by the inclusion of climate variables as external regressor to forecast the dengue cases for the year 2010.

Conclusions: The ARIMA model with weekly variation is a useful tool for disease control and prevention program as it is able to effectively predict the number of dengue cases in Malaysia.

KEYWORDS

Dengue, Forecasting, ARIMA model, Climate, Malaysia

1. Introduction

Recently, modelling of dengue cases has become an interest among many epidemiologists in Malaysia. Much has been discussed regarding the effective and predictive models which are able to accurately predict the occurrence of dengue cases [1,2]. In the absence of neither effective vaccines nor a specific anti-viral treatment, a suitable

predictive model is needed to support the primary anti dengue control which aims to destroy the breeding containers thus interrupting the transmission of dengue virus by infected adult mosquitoes.

In an attempt to forecast future events, most of the models developed by researchers, used dengue database obtained from the Ministry of Health to formulate the predictive model as opposed to using vector population data. To date,

*Corresponding author: Nazri Che Dom, School of Biological Sciences, Universiti Sains Malaysia, Penang.

Tel: +6019 6143465

E-mail: nazried@salam.uitm.edu.my

Foundation Project: Supported by Universiti Sains Malaysia (USM)–LRGS grant 304/PBiology/650575/U112), Universiti Teknologi MARA (UiTM) and Ministry of Higher Education (MOHE) Malaysia.

Article history:

Received 22 Jul 2013

Received in revised form 29 Jul, 2nd revised form 5 Aug, 3rd revised form 18 Aug 2013

Accepted 15 Sep 2013

Available online 28 Oct 2013

many models experience high degree of complexity as they often require a very large data base comprising several different variables. Collecting such data is always costly and eventually this may result in the scarcity of suitable data sets to be used in formulating the model. Autoregressive integrated moving average (ARIMA) model in time series analysis is an example of statistical modelling approach, which require less data than the deterministic simulation methodology. There have been researches indicating that the time series modelling is more appropriate than the simple trend fitting approach, despite the fact that, this trend suffers model specification error[3]. In response to this, Box and Jenkins had effectively put together the relevant information in a comprehensive manner for the user to understand and apply this time series model[4]. In addition, it is also possible for a non-expert to easily adopt the model methodology for prediction of dengue cases.

ARIMA models are particularly useful in modelling the temporal dependence structure of a time series as they explicitly assume temporal dependence between observations[5]. Through the modelling of the temporal structure, particularly for seasonal disease, prediction made with ARIMA models have been shown to be more accurate than those obtained by other statistical methods[3,6,7]. Prediction can be made for a certain number of periods ahead, with this number being given by the highest order of the parameters of the model[5].

The ARIMA models have been used successfully in epidemiology to monitor and predict infectious disease, such as malaria and hepatitis A incidence[7], influenza and pneumonia deaths[6], as well as other infectious disease incidence[8–11]. Apart from that, it can also to be applied in the management of the health facilities and in isolation of syndromatic surveillance[12,13].

Promprou *et al.* used the univariate time series analysis method, ARIMA (1,0,1) to model and forecast the monthly cases of dengue haemorrhagic fever cases in southern Thailand[14]. The results of their study showed that the regressive forecast curves of dengue cases were consistent with the pattern of actual values. Luz *et al.* used seasonal ARIMA to model dengue incidence in Rio de Janeiro, Brazil[15]. They found that the number of dengue cases in a month can be estimated by the number of dengue cases occurring one, two and twelve months ahead. Their findings indicated that ARIMA models were useful tools for monitoring dengue incidence in Rio de Janeiro and could be applied for surveillance and predicting trends in dengue incidence.

It has been proposed that climate variables can increase the predictive power of dengue models[16]. Increased temperature has been associated with dengue in tropical country[17–21]. A group of corresponding researchers has also mentioned that rainfall has a positive correlation with dengue incidence[17,20,22]. In the past two decades, several studies have documented the relationship between weather variables and dengue incidence. However, the use of weather

variables as a predictor in influencing the occurrence of DF cases has yet to be explored and established.

Therefore, this work attempts to use ARIMA models to monitor and predict dengue incidence in Subang Jaya based on dengue database from 2005 to 2009 by using the Box–Jenkins approach. The fitted model was then used to predict dengue incidence for the year 2010. Finally, the impact of climate variables (rainfall, temperature and relative humidity) was incorporated with the prediction of dengue incidence and outbreak by selecting the best fitting model. Hence, this study will hopefully enhance the efficiency of dengue surveillance program, and thus help in controlling dengue outbreak.

2. Materials and methods

This study covers the dengue incidence and meteorological data in Subang Jaya between the periods of 2005 until 2009. During the study period, stable dengue control programs were implemented each year in Subang Jaya. This study was primarily conducted using the statistical package XLstat for windows to develop ARIMA models. The ARIMA model was analysed with the Box–Jenkins approach which was appropriate for a longer forecasting period. Specifically, it consists of three types of parameters in the model namely; the autoregressive parameters (p), the number of differencing phases (d), and moving average parameters (q). This basis of the Box–jenkins approach to modelling time series is summarized in Figure 1 and consists of three phases; identification, estimation / testing and application.

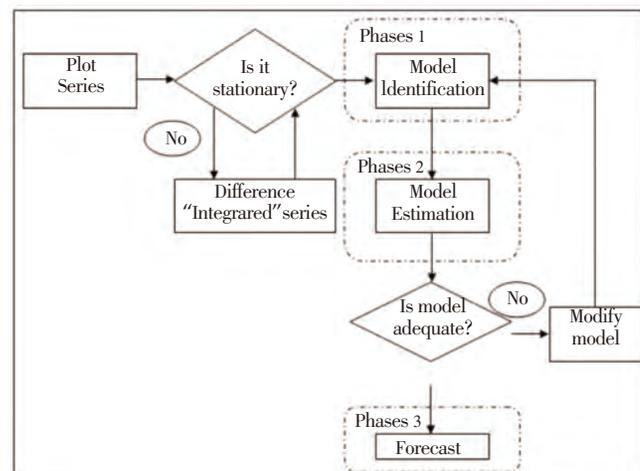


Figure 1. The Box Jenkins model building process.

2.1. Model identification

The first step is to determine whether the time series is in a stationary or non-stationary condition by using the mean range plot. If non-stationary, it has to be transformed into a stationary time series by applying a suitable degree of differencing to the dataset. The input series for ARIMA needs to be stationary, with a constant mean, variance and autocorrelation through time. Basically, the number of time

series need to be differenced to achieve stationary condition is reflected in the (d) parameter.

In order to determine the necessary level of differencing, the plot of the data and autocorrelogram were examined. Significant changes in each level (strong upward or downward changes) usually require first order of non-seasonal (lag=1) differencing whereas strong changes of slope require second order of non seasonal differencing. In normal practice, seasonal pattern of the data does not required respective seasonal differencing. If the estimated autocorrelation coefficient decline slowly at longer lags, first order differencing is usually needed. At this stage the order of non seasonal (p, d, q) and seasonal (P, D, Q), autoregressive parameters (p and P) and moving average parameters (q and Q) are needed to be ascertained to yield an effective model.

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} - \delta_1 a_{t-1} - \delta_2 a_{t-2} - \dots - \delta_q a_{t-q} + a_t$$

Consequently, with the series in stationary condition by transformation or differencing, an ARIMA model was developed. A tentative model was made to express each observation as a linear function of the previous value of the series (autoregressive parameter) and of the past random shock (moving average parameters). The general form of the tentative model was given below:

Where:

Y_t denotes the number of DF cases at time t , Y_{t-1} denotes the number of DF cases at time $t-1$, Y_{t-p} denotes the number of DF cases at time $t-p$, $\phi_1, \phi_2, \dots, \phi_p$ is autoregressive parameters (of order p), $\delta_1, \delta_2, \dots, \delta_p$ is moving average parameters (of order q), a_t denotes a time series of random shock or white noise process at time t , a_{t-1} denotes the white noise process at time $t-1$, δ is a constant term.

The random shock a_t is a value that is assumed to be randomly selected from a normal distribution that has a mean of 0 and variance that is constant at every time period. The random shock $a_t, a_{(t-1)}, \dots, a_{(t-q)}$ are assumed to be statistically independent.

In order to analyse the time series of dengue incidence in Subang Jaya between the year 2005 to 2009, observation (s) is defined as the number of epidemiological weeks in one year ($s=52$). Then, the temporal structure of seasonal and non seasonal autoregressive parameters (p, P), moving average parameters (q, Q) parameters were determine by assessing the analysis of autocorrelation function (ACF) and partial autocorrelation function (PACF). The Akaike Information Criterion (AIC) was used to select the best fit model with fewer parameters in order to proceed to the next process. Then, validation of the final model was based on the residual analysis (The residual value must be equivalent to white noise) using Ljung–Box test.

2.2. Model estimation and testing

In the identification phase, the specific number and types of ARIMA parameters to be used were estimated. The major

tools used in these phase are plot of the series, correlogram of autocorrelation, and partial autocorrelation. The decision is not straightforward and in less typical cases requires not only experience but also a good deal of experimentation with alternative models (as well as the technical parameters of ARIMA). However, a majority of empirical time series patterns can be sufficiently approximated using one of the 5 basic models that can be identified based on the shape of the autocorrelogram and partial autocorrelogram.

The general recommendation concerning the selection of parameters to be estimated (based on ACF and PACF) also apply to seasonal models. The main difference is that in seasonal series, ACF and PACF will show sizable coefficient at multiples intervals of the seasonal lag (in addition of their overall pattern reflecting the non-seasonal components of the series).

Although the selected model may appear to be the best among those models considered, it is also necessary to run diagnostic checking in order to verify the adequacy of the model. A good model should not only provide sufficiently accurate forecast, it should also be parsimonious as well as produce statistically independent residuals that contain only white noise without systematic components (the correlogram of residuals should not reveal any serial dependencies). A good indicator of the model is by plot the residuals and inspects for any systematic trends and examines the autocorrelogram of residuals (there should be no serial dependency between residuals).

For a good forecasting model, the residual left over after fitting the model should be closed to white noise. Therefore, if the ACF and PACF of the residuals were obtained, there should be no significance for both autocorrelation and partial autocorrelation.

Subsequently, a portmanteau test can also be applied to the residuals as an additional test of best fit. Insignificant value of the test suggests the model is adequate and there is no need to consider further refinement of the model. On the contrary, significant values of the test suggest the model is inadequate and thus other ARIMA model need to be considered in the process.

2.3. Application phase

Another straightforward and common measure of the reliability of the model is the accuracy of its forecast generated based on partial data so that the forecast can be compared with known (original) observations. Some of the data at the end of the series omitted before the models are estimated. Then the models are compared on the basis of how accurate they forecast which have been withheld rather than how well they forecast the same data which has been used for prediction modelling.

In this study, the fitted ARIMA model adopted from dengue incidence data in Subang Jaya from 2005 to 2009 was used to calculate the predicted values and their 95% prediction intervals. Three different approaches were designed to estimate

the predicted values for the year 2010. First approach was applied to predict 2010 data with a one-year (52 weeks) lag. The second and third method was iterative approach. They were applied to predict dengue incidence with 3 months (13 weeks) and one month (4 weeks) lag, respectively. The selection of this approach was reflected by the on-going nature of dengue surveillance where the health authorities in charge of disease monitoring continuously work with new information as it arrives. The basic idea was to superimpose reference lines, called control limit on a time series plot.

The final test for an ARIMA model was its ability to forecast. The observed data for this period were included in the database in order to update the model and to compare the prediction values of dengue occurrence throughout the year 2010. To validate, the predictive value was assessed by calculating the root mean square error (RMSE) and the statistical difference of the error. This test is to evaluate the median of the distribution and the difference of the error among 3 approaches is statistically different from zero at the 5% significance level.

As a tropical country, Malaysia experience slight variation of temperature, abundant rainfall and high humidity throughout the year. This condition has significantly impact on population size, maturation period, feeding characteristics and survival rates of *Aedes* mosquitoes. In order to improve the predictive model, climate variables were used as a predictors. To facilitate selection of these explanatory variables, cross correlation coefficient was computed between climatic variables and dengue incidence with effect of different time lags, ranging from 1 to 16 weeks for each weather predictor. Cross correlation coefficient of each weather variable and dengue cases as well as literature reports were also examined to estimate maximum lag term.

As a criterion for assessing the predictive ability of the model to forecast the number of DF cases, the model adequacy was assess by checking whether (i) the model assumption were satisfied, (ii) the error were normally distributed and (iii) all residual ACF were equal to zero by using Q-statistics Box–Ljung test.

3. Results

A total of 4898 dengue fever cases have been recorded by Vector Control Unit, MPSJ from 2005 to 2009. The monthly numbers of dengue fever cases notified are shown in Table 1. Note that dengue cases in Subang Jaya increased yearly in the study period and reach the highest record in 2008. During the study period, recorded dengue incidence increased with greater magnitude and intensity from 805 cases in year 2005 to 1191 cases in 2008. The incidence then decreased to 1033 cases in 2009. The highest dengue cases in the study period were reported in February of year 2007 with 208 cases being recorded.

Table 1

Number of recorded cases of dengue between 2005 and 2009 in Subang Jaya.

Month	Year (Number of dengue fever cases)				
	2005	2006	2007	2008	2009
January	122	23	72	93	101
February	101	86	208	127	120
March	44	90	100	95	68
April	42	125	70	122	52
May	125	125	103	151	110
June	44	89	53	94	74
July	55	106	58	144	67
August	37	84	39	92	69
September	78	60	62	77	74
October	61	40	48	77	71
November	46	60	44	56	92
December	50	77	47	63	135
Total	805	965	904	1191	1033

Since data for climate predictors were collected continuously year round, trend and seasonality pattern in collected data need to be identified in order to control for unmeasured cofounders which might influence some parts of the seasonal and long term time trends variation in dengue cases. Daily weather data were aggregated to weekly average which comprised a total of 260 weeks period (52 weeks in a year for 5 years). This analysis is independent of any spatial location or week of transmission of dengue fever cases. Spatial and temporal location is not a factor in this analysis.

The plot of weekly mean temperature and weekly cumulative precipitation for the study period showed increasing trends in both climatic variables, however the plot of weekly mean precipitation didn't reveal consistently distinctive seasonal pattern compared to weekly mean temperature which show consistent pattern during 2005–2009 (Figure 2).

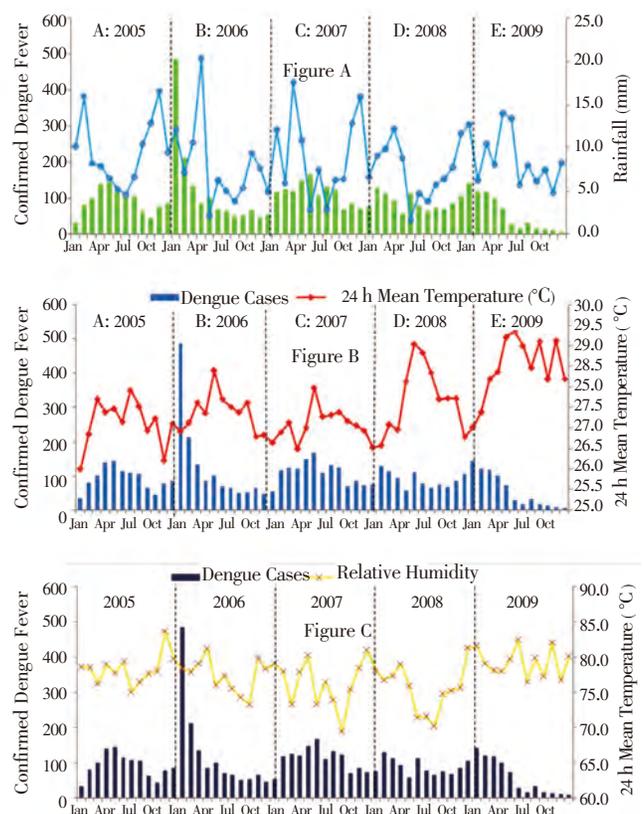


Figure 2. A: monthly rainfall and total dengue cases, B: temperature and total dengue cases, C: relative humidity and the total dengue cases in year 2005–2009.

In the first step in analyzing the time series of dengue incidence, a natural logarithm transformation was performed to stabilize the variance of the series. In order to stabilize the variance, seasonal and regular differencing was applied. Figure 3 showed the transformation of the series with the lowest dispersion whereby it indicated that dengue incidence has no significant trends.

The temporal dependence structure was then determined by assessing the analysis of autocorrelation and Partial auto correlation. The plot of ACF and PACF (Figure 4A and Figure

4B) described the temporal dependence structure indicating that the seasonal (P,D,Q) and non seasonal (p,d,q) parameters are needed in the model development. After differencing, a significant cut offs at one week lag and another at lag 52 weeks were observed on the plot of ACF (Figure 4C). These two cut offs were less marked on the plot PACF (Figure 4D) and evolved more gradually over time, compared to the plot ACF. The analysis from the correlograms of the series suggests that p value should be equal to 1 or 2 and q value equal to 0 or 1 of moving average parameters.

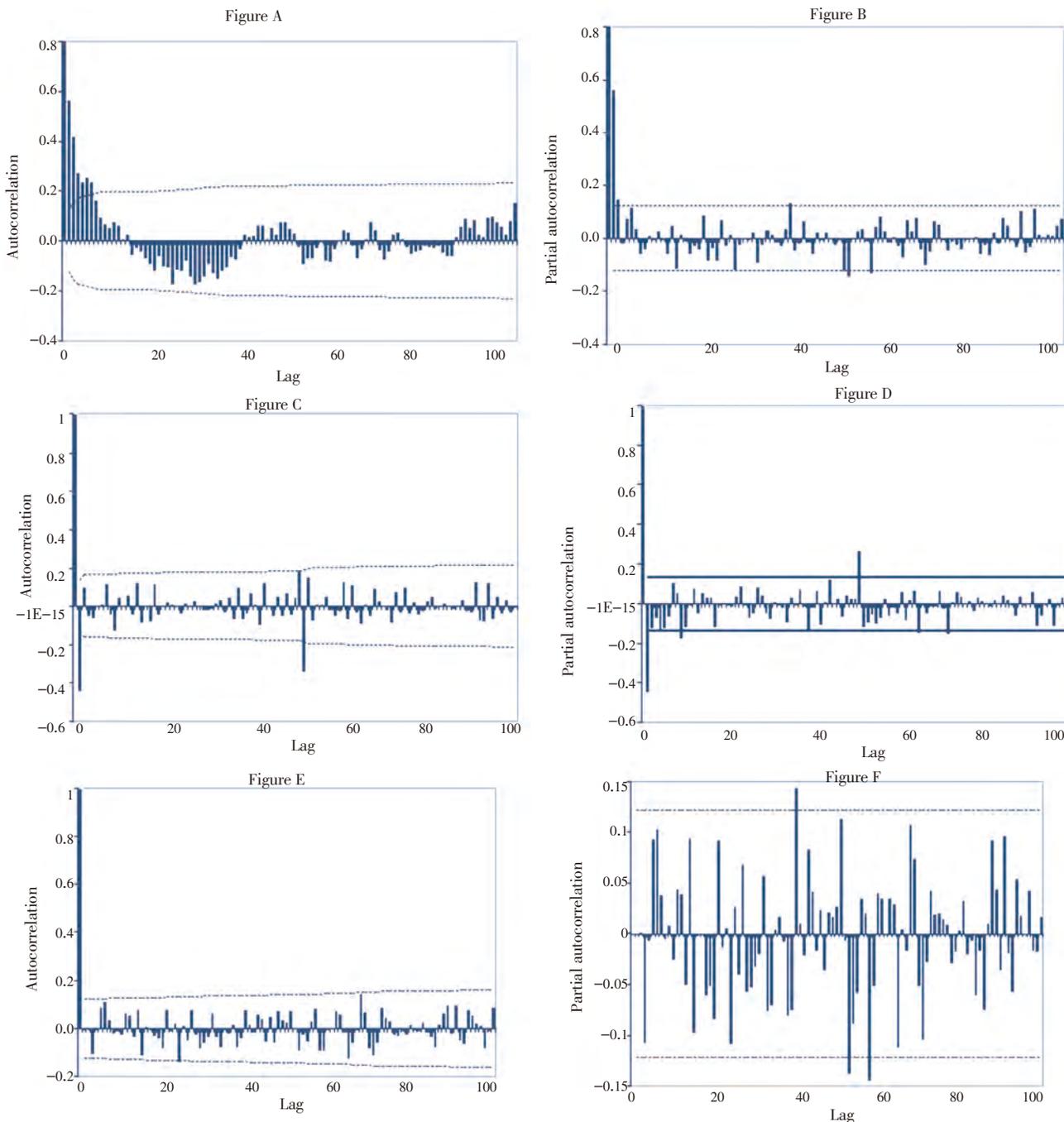


Figure 4. A: ACF plot of original dengue incidence, B: PACF plot of original dengue incidence, C: ACF plot of integrated dengue incidence, D: PACF plot of integrated dengue incidence, E: ACF of residual after applying ARIMA (2,0,0) (0,0,1)₅₂ model, F: PACF of residual after applying ARIMA (2,0,0) (0,0,1)₅₂ model. The X axis gives the number of lags in weeks and the Y axis, the value of the correlation coefficient comprised between -1 and 1. Dotted lines indicate 95% confidence interval.

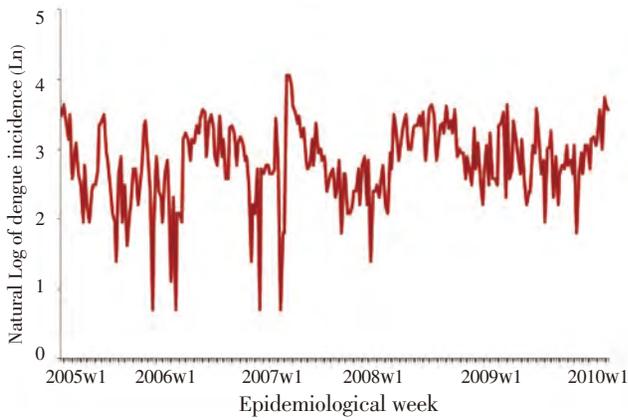


Figure 3. Natural logarithm of DF incidence in Subang Jaya for the period 2005 to 2009.

Both ACF and PACF were utilized to explore a set of models based on the training data from 2005 to 2009. Table 2 showed AIC values and Mean Absolute Percent Error (MAPE) for the ARIMA models corresponding to difference choice of p, d and q . Among these models, ARIMA (2,0,0) (0,0,1)₅₂ had both lowest AIC (360.093) and MAPE (16.858) values and appeared to be the best ARIMA model to fit the occurrence of dengue incidence.

Table 2
AIC values for different ARIMA (p, d, q) (0,0,1)₅₂ model and MAPE values.

Models	MAPE*	AIC*
(2,0,0) (0,0,1) ₅₂	16.858	360.093
(2,0,1) (0,0,1) ₅₂	16.552	367.534
(2,1,1) (0,0,1) ₅₂	17.171	369.534
(1,0,0) (0,0,1) ₅₂	18.445	408.503
(1,0,1) (0,0,1) ₅₂	18.487	367.959
(1,1,1) (0,0,1) ₅₂	17.152	369.441

* Lower values of AIC and MAPE are preferable.

Therefore, the following ARIMA (2,0,0) (0,0,1)₅₂ model was found to be the best to fit the occurrence of dengue incidence based on the training data from 2005 to 2009. The analysis of residual on ACF and PACF plots (Figure 4E and Figure 4F) assessed the absence of persistent temporal correlation. For the model, the plot of the residual autocorrelation function died out after one lag and the residual autocorrelations fell within 95% confidence limit. This suggested that autocorrelation function of residuals at different lag times in the ARIMA (2,0,0) (0,0,1)₅₂ model did not differ from zero. The observed and predicted dengue fever cases in Subang Jaya from 2005 to 2009 matched reasonably well.

Furthermore, the selected ARIMA model fitted observed data from 2005 to 2009 was used to forecast the dengue incidence in 2010 by extrapolating the pattern several weeks ahead. In this model three approaches were designed to extrapolate the pattern of dengue incidence in 2010 by using timely interval of one year (52 weeks) lag, 3 months (13 weeks) and one month (4 weeks) lag. Predicted 2010 values were shown in Table 3, where they are compared with the observed number of dengue cases. Table 3 presented out

of sample predicted values obtained from the ARIMA (2,0,0) (0,0,1)₅₂ model by extrapolating the pattern several weeks ahead. The model was able to describe the pattern in weekly dengue cases and it produced good fit of predicted cases when plotted against observed data.

Table 3
Observed number of dengue cases in 2010 and corresponding out of sample predicted values obtained from ARIMA models.

Epidemiological week	Observed values, 2010	Out-of-sample predicted values for 2010		
		4-weeks ahead	13 weeks ahead	52 weeks ahead
W1	35	35	34	38
W2	33	32	32	36
W3	22	30	30	33
W4	24	25	25	27
W5	24	21	21	24
W6	39	22	22	20
W7	36	28	27	25
W8	43	33	32	30
W9	36	34	34	31
W10	17	34	34	31
W11	36	23	23	21
W12	22	23	23	21
W13	17	26	25	23
W14	29	18	19	23
W15	11	21	21	25
W16	20	17	17	21
W17	10	15	15	18
W18	7	14	14	17
W19	9	9	9	11
W20	8	9	9	10
W21	12	9	9	11
W22	8	10	10	12
W23	3	10	11	12
W24	6	6	6	7
W25	2	5	5	6
W26	9	4	4	4
W27	11	5	5	4
W28	16	10	11	9
W29	4	13	13	12
W30	4	9	9	8
W31	6	5	5	4
W32	2	6	6	5
W33	1	4	4	4
W34	3	2	2	2
W35	1	2	2	3
W36	4	2	2	3
W37	4	3	3	3
W38	3	5	5	5
W39	2	4	4	5
W40	1	3	3	4
W41	5	2	3	4
W42	3	3	4	5
W43	2	5	5	5
W44	3	3	5	4
W45	1	3	5	4
W46	1	2	4	4
W47	2	1	4	3
W48	2	2	4	2
W49	3	3	3	3
W50	4	3	4	3
W51	6	3	5	4
W52	5	4	6	5
Total	617	619	634	656

The graph in Figure 5 compares the number of dengue cases observed in 2010 with the values obtained by all pattern of weeks. The 4 weeks step approach showed the smallest difference between observed and predicted values (RMSE=0.61) when compared to the 52 weeks step approach (RMSE=0.83) and to the 13 weeks step approach (RMSE=0.76). The observed and estimated values agreed very closely in the time series. However, the difference between residuals predicted 13 weeks ahead and those predicted 4 weeks ahead was not statistically significant. The prediction for the following months was the best compromise for helping the health authorities to take measures to mitigate transmission, morbidity and mortality.

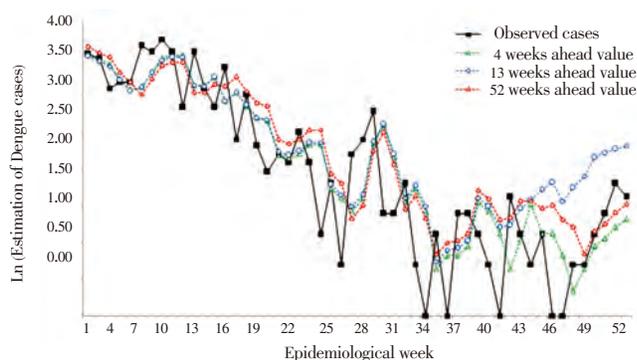


Figure 5. Natural logarithm of dengue incidence in Subang Jaya for 2010. Solid line (filled squared): observed values during the period, Dashed line: ARIMA model (2,0,0) (0,0,1)_{s2} model.

Subsequently, this study was assessed to improve the predictive power of the model. Therefore, a climatic variable was incorporated as an external regressor in the univariate time series ARIMA model. Table 4 described the characteristics of the ARIMA model for climatic variables. A regular differencing was applied for all climatic variables except for relative humidity and rainy day variables. On the contrary, a seasonal differencing was applied for all climatic variables and the residual was kept for the multivariate analysis.

Climatic variables identified as the most interconnected to dengue incidence were accounted one by one, due to their strong interconnection. In order to facilitate the selection of climate variables as external regressor, cross correlations between residuals of dengue incidence and those climatic variables over a range of 16 weeks was analysed. Figure 6 showed the cross correlation functions between dengue fever cases and climatic variables after applying ARIMA model. The following explanatory variables were the most potential external regressor based on the correlation coefficient of the residual of the time series: Lag-10 maximum temperature (Pearson correlation: $r=0.129$, $P=0.017$), Lag-5 minimum temperature (Pearson correlation: $r=0.122$, $P=0.075$), Lag-10 average temperature (Pearson correlation: $r=0.169$, $P=0.014$), Lag-11 relative humidity (Pearson correlation: $r=0.163$, $P=0.012$). Meanwhile, precipitation and rainy day was not correlated with dengue incidence directly over a range of

Table 4

Characteristics of ARIMA models for climatic variables: coefficient, standard errors of residuals, AIC, P -value after Ljung-Box test of residuals.

Climatic variables	ARIMA (p,d,q) (P,D,Q) _s	AR1	MA1	SMA	Sd (residuals)	AIC	P -value (Ljung-box test)*
Precipitation	(1,1,1) (0,1,1) _{s2}	0.140*	-1.000*	-1.000*	1.339	782.502	0.22
Relative humidity	(0,0,1) (0,1,1) _{s2}	-	0.462*	-0.694*	4.208	1204.913	0.52
Minimum temperature	(1,1,1) (0,1,1) _{s2}	0.160*	-0.936*	-1.000*	0.998	565.991	0.45
Maximum temperature	(1,1,1) (0,1,1) _{s2}	0.283*	-0.931*	-0.843*	0.976	521.573	0.57
Average temperature	(1,1,1) (1,1,1) _{s2}	0.307*	-0.929*	-0.973*	0.928	477.077	0.61
Rainy day	(0,0,1) (0,1,1) _{s2}	-	0.204*	-0.981*	1.743	879.488	0.62

AR: autoregressive, MA: moving average, SMA: Seasonal moving average, *: $P<0.05$ significant. * The residual value must be equivalent to white noise.

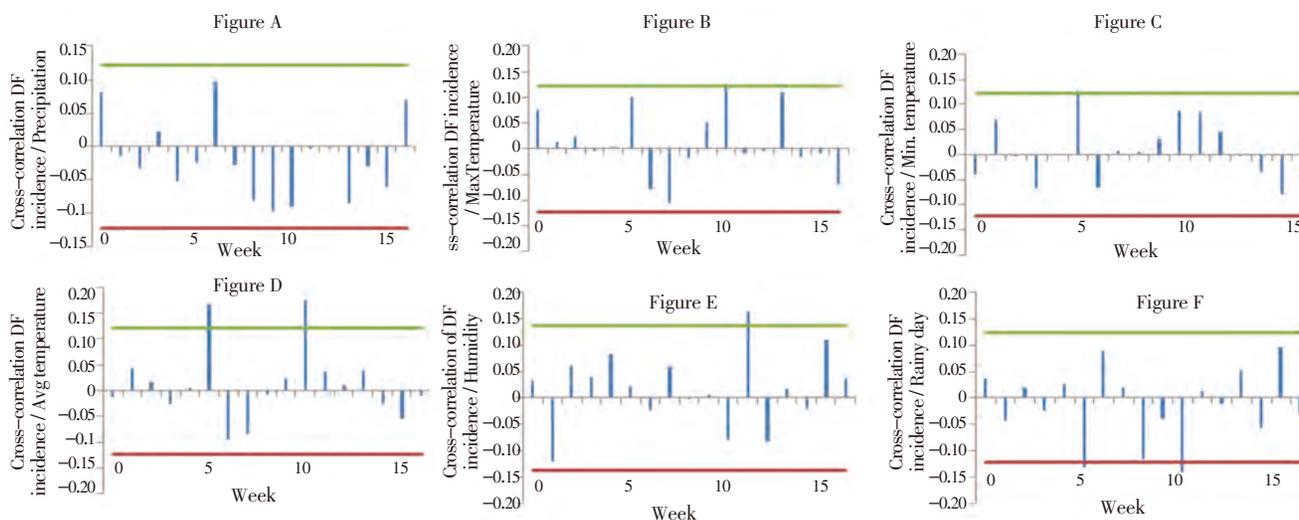


Figure 6. Cross correlation functions between dengue fever (DF) cases and meteorological variables after applying ARIMA model. The x axis gives the number of lags in weeks. Dotted lines indicate 95% confidence interval. Only positive lags are taken into account. A: Accumulative precipitation, B) Maximum temperature, C) Minimum temperature, D) Mean Temperature, E) Relative Humidity, F) Rainy day.

Table 5

Characteristics of univariate and multivariate model using climate variables which have most correlated to dengue incidence: Coefficients, standard error and *P*-value of parameters, AIC, RMSE for prediction.

ARIMA model	Coefficient	Standard error	<i>t</i> -statistics	<i>P</i> -value	AIC	RMSE
None					360.093	0.76
Average Temperature_Lag10	0.208	0.035	1.128	0.219	362.812	0.75
Minimum Temperature_Lag5	0.248	0.041	1.654	0.023	359.190	0.71
Relative humidity_Lag11	0.218	0.054	2.343	0.025	362.462	0.75

Last column gives the predictive power as measured by the RMSE of the four ARIMA models using 4 weeks (1 month) ahead approach. Parameter estimated by maximum likelihood.

16 weeks-lags ($r < 0.1$, $P = \text{non significant}$). Thus, the result of this study indicated that the climatic variables varied accordingly to a range of lag time.

The identification of climate variables that significantly correlated with dengue incidence were then tested with ARIMA (2,0,0) (0,0,1)₅₂ model. This is carried out by incorporating these independent variables, either the Lag-5 minimum temperature or the Lag-11 average temperature. These models have improved predictive power as measured by the RMSE as shown in Table 5. The prediction for 2010 on a monthly basis were improved after the introduction of either minimal temperature at Lag-5 weeks (RMSE=0.71) or the average temperature at Lag-11 weeks (RMSE=0.75) (Figure 7).

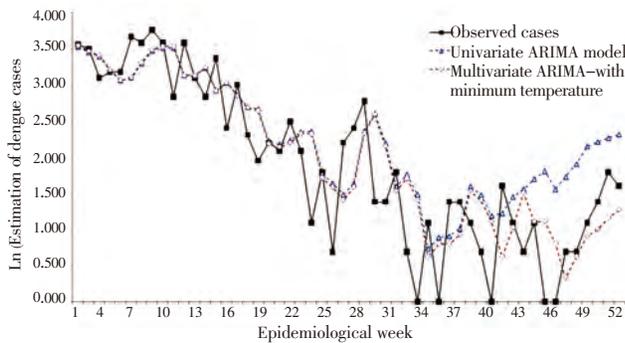


Figure 7. Natural logarithm of dengue incidence in Subang Jaya for 2010. Solid line (filled square): observed values during the period; Blue (triangle): Univariate ARIMA (2,0,0) (0,0,1)₅₂ models 13 weeks ahead values; Red (Circle): multivariate ARIMA (2,0,0) (0,0,1)₅₂ models 13 weeks ahead values with minimum temperature lag-5.

4. Discussion

An early epidemic prediction tool is critical for evaluating the risk of an outbreak. Early identification allows early interventions and prevention of an epidemic instead of an epidemic management. There is a need to develop an early warning system that should be able to identify and quantify the risk of dengue in a population. The present system of prediction of dengue outbreak is based on the use of various entomological indices without incorporating epidemiological component. The ARIMA model is a useful tool for interpreting and applying surveillance data. It has a great potential to be used as a decision support tool in order to improve contingency planning and prevention by public health interventions in the field.

The result of this study revealed that the ARIMA model

developed, closely described the trend of dengue incidence and confirmed the existence of dengue fever cases in Subang Jaya. The analysis of auto-regression and moving average parameters of the developed model demonstrated the prediction for the year 2010. The developed model can be estimated by the number of dengue fever cases using the extrapolation of several week patterns ahead. In this study, the prediction for the year 2010 per period of 4 weeks ahead was found agreed with the observed dengue incidence.

Moreover, this study subsequently assessed alternative ARIMA model incorporating climate variables as external regressor. Cross correlation between dengue incidence and climate variables was compute over a range of lags in order to identify significant variables to be included as external regressor. Although the observed correlation are week (below 0.2) but they are statistically significant which means these variables have a powerful strength to improve the reliability of developed predictive model. It was found that temperature variables is a key component to introduce in the model as external regressor either minimal or average temperature in order to improve the prediction capability of the model.

Climate variables are key components in dengue transmission cycle and affect dengue incidence in multiple ways. For instance rainfall plays an indefinite role on dengue incidence while temperature and relative humidity affect this transmission in several ways. Rainfall has the potential to effect either to increase the transmission of vector borne disease by promoting the proliferation of breeding places for mosquitoes or by eliminating breeding sites via heavy rainfall that destroy existing breeding sites. Hence, interrupting the development of eggs or larvae out of the pool. In this study rainfall parameters failed to be included as the best variable to fit in predictive model. Concurrently, temperature indirectly influences the development and digestion of mosquitoes by reducing the duration of the gonotrophic cycle and female size. An increased temperature accelerates viral dissemination within the mosquito, reducing the extrinsic incubation period in the water. Besides that, higher temperature may increase the ratio of the standing crop of pupae to the number of adult females.

In recent years, the ability to predict local and regional climate in terms of accuracy and lead times has rapidly been improved due to advances in technology. This has allowed a better understanding of the interaction between climate and the temporal distribution of dengue fever as well as stimulating research interest on epidemic prediction modelling. As systematic mosquito data were not available

in the study area, this study explored the climate variables to develop dengue forecasting model based on the delay effect of climate variables on dengue incidence.

Several studies have documented scientific evidence on the impact of climate condition on the life cycle dynamic of both vectors and virus. From mosquito hatching to human case appearance, several successive phase occurred resulting cumulative lags observed in the study. These phases include larval and pupal development (10–21 d), gonotrophic cycle (3 to 7 d per cycle), extrinsic incubation in mosquitoes (7 to 15 d) and incubation in human (1–12 d)[23–25]. The lag between weather data and DF incidence data will differ depending on the respective lag between the biological cycle and clinical symptoms. Thus, the lag is expected to be shorter for minimum temperature that is usually associated with adult mosquito mortality and expected to be longer for high relative humidity which is related to adult survival and hatching. On the contrary, the mean temperature is involved in all biological cycles of mosquitoes that take more time to influence the dengue incidence.

The results from this model are consistent with all of this assumption that is similar with other studies dealing with the effect of climate on dengue outbreak. Many studies have also dealt with the effect of climate on dengue outbreak. In Thailand, the dengue incidence was positively correlated with the average temperature at lag 3–4 months[25]. In Taiwan, there was significant positive correlation with the maximum temperature at lag 1–4 months, the minimum temperature at lag 1–3 months and the relative humidity at lag 1–3 months[26]. In Brazil, positive association were found between the minimum and maximum temperature and dengue transmission at lag–0[15], and in the city of Guangzhou in China, the minimum temperature and relative humidity were positively correlated with dengue incidence at lag–1 month.

It should be acknowledged that dengue transmission is a very complicated problem. The risk of transmission varies in space and time and the dynamics of the disease is dependent on seasonal changes in weather and immunity. Dengue transmission is particularly sensitive to rainfall, temperature, and humidity, which is associated with the monsoon season. Along with weather variables, other environmental and host factors such as community intervention measures and human behaviour also influence mosquito populations and the degree of contact between human being and vector.

The development of this model hopefully can be used to monitor and predict dengue incidence in Subang Jaya. This is in line with an urgent need to improve approaches for monitoring and predicting dengue incidence in order to reduce spreading of DF cases globally. Hence there is a potential of the ARIMA model to be used in the estimation and prediction of dengue cases, thus supporting the existing intervention program. Accurate predictions for even a few months ahead provide an invaluable opportunity to mount a vector control intervention. This study proves the ability of ARIMA model to be used as a simple, precise and low cost functional dengue early warning system and thus help to develop an efficient dengue control program.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

The authors sincerely thank Dr. Roslan Mohamed Hussin, Director of Health Department MPSJ for providing ground data on DF cases for this research work. The contribution of research funding from Universiti Sains Malaysia (USM) – (LRGS grant 304/PBiology/650575/U112), Universiti Teknologi MARA (UiTM) and Ministry of Higher Education (MOHE) Malaysia are also duly acknowledged.

Comments

Background

Modeling of dengue cases has become an interest among many epidemiologists in Malaysia. ARIMA models are particularly useful in modelling the temporal dependence structure of a time series as they explicitly assume temporal dependence between observations. Through the modelling of the temporal structure, particularly for seasonal disease, prediction made with ARIMA models have been shown to be more accurate than those obtained by other statistical methods (Box *et al.*, 2011).

Research frontiers

This study proves the ability of ARIMA model to be used as a simple, precise and low cost functional dengue early warning system and thus help to develop an efficient dengue control program.

Related reports

The results from this model are consistent and similar with other studies dealing with the effect of climate on dengue outbreak. Many studies have also dealt with the effect of climate on dengue outbreak, *e.g.*, Watts *et al.* (1987); Focks *et al.* (2005); Luz *et al.* (2008).

Innovations & breakthroughs

The ARIMA model was found to be the best to fit the occurrence of dengue incidence based on the training data from 2005 to 2009. The analysis of residual on ACF and PACF plots assessed the absence of persistent temporal correlation. For the model, the plot of the residual autocorrelation function died out after one lag and the residual autocorrelations fell within 95% confidence limit. This suggested that autocorrelation function of residuals at different lag times in the ARIMA model did not differ from zero. The observed and predicted dengue fever cases in Subang Jaya from 2005 to 2009 matched reasonably well.

Applications

In this study, three different approaches were designed

to estimate the predicted values for the year 2010. First approach was applied to predict 2010 data with a one year (52 weeks) lag. The second and third method was iterative approach. They were applied to predict dengue incidence with 3 months (13 weeks) and one month (4 weeks) lag, respectively. The selection of this approach was reflected by the on-going nature of dengue surveillance where the health authorities in charge of disease monitoring continuously work with new information as it arrives. The basic idea was to superimpose reference lines, called control limit on a time series plot.

Peer review

This is a good study in which the authors have developed a forecasting model for the incidence of dengue cases in Subang Jaya using time series analysis. The result of this study are interesting and revealed that the ARIMA model developed, closely described the trends of dengue incidence and confirmed the existence of DF cases in Subang Jaya for the year 2005 to 2010. The ARIMA model with weekly variation is a useful tool for disease control and prevention program as it is able to effectively predict the number of dengue cases in Malaysia.

References

- [1] Lu L, Lin H, Tian L, Yang W, Sun J, Liu Q. Time series analysis of dengue fever and weather in Guangzhou, China. *BMC Public Health* 2009; **9**(1): 395.
- [2] Wongkoon S, Jaroensutasinee M, Jaroensutasinee K. Development of temporal modeling for prediction of dengue infection in Northeastern Thailand. *Asian Pac J Trop Med* 2012; **5**(3): 249–252.
- [3] Farmer RD, Emami J. Models for forecasting hospital bed requirements in the acute sector. *J Epidemiol Commun Health* 1990; **44**(4): 307–312.
- [4] Box GE, Jenkins GM, Reinsel GC. *Time series analysis: forecasting and control*. New York: Wiley; 2011.
- [5] Helfenstein U. The use of transfer function models, intervention analysis and related time series methods in epidemiology. *Int J Epidemiol* 1991; **20**(3): 808–815.
- [6] Choi K, Thacker SB. An evaluation of influenza mortality surveillance, 1962–1979 I. Time series forecasts of expected pneumonia and influenza deaths. *Am J Epidemiol* 1981; **113**(3): 215–226.
- [7] Nobre FF, Monteiro ABS, Telles PR, Williamson GD. Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology. *Stat Med* 2001; **20**(20): 3051–3069.
- [8] Allard R. Use of time-series analysis in infectious disease surveillance. *Bull World Health Organ* 1998; **76**(4): 327–333.
- [9] Helfenstein U. Box-jenkins modelling of some viral infectious diseases. *Stat Med* 1986; **5**(1): 37–47.
- [10] Tong S, Hu W. Climate variation and incidence of Ross river virus in Cairns, Australia: a time-series analysis. *Environ Health Perspect* 2001; **109**(12): 1271–1273.
- [11] Trottier H, Philippe P, Roy R. Stochastic modeling of empirical time series of childhood infectious diseases data before and after mass vaccination. *Emerg Themes Epidemiol* 2006; **3**: 9.
- [12] Earnest A, Chen MI, Ng D, Sin LY. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. *BMC Health Serv Res* 2005; **5**: 36.
- [13] Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak* 2003; **3**: 2.
- [14] Promprou S, Jaroensutasinee M, Jaroensutasinee K. Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA Models. *Dengue Bull* 2006; **30**: 99–106.
- [15] Luz PM, Mendes BV, Codeço CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *Am J Trop Med Hyg* 2008; **79**(6): 933–939.
- [16] Hurtado-Díaz M, Riojas-Rodríguez H, Rothenberg SJ, Gomez-Dantés H, Cifuentes E. Short communication: impact of climate variability on the incidence of dengue in Mexico. *Trop Med Int Health* 2007; **12**(11): 1327–1337.
- [17] Corwin AL, Larasati RP, Bangs MJ, Wuryadi S, Arjoso S, Sukri N, et al. Epidemic dengue transmission in southern Sumatra, Indonesia. *Trans Royal Soc Trop Med Hyg* 2001; **95**(3): 257–265.
- [18] Nakhapakorn K, Tripathi N. An information value based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence. *Int J Health Geogr* 2005; **4**: 13.
- [19] Bangs MJ, Larasati RP, Corwin AL, Wuryadi S. Climatic factors associated with epidemic dengue in Palembang, Indonesia: implications of short-term meteorological events on virus transmission. *Southeast Asian J Trop Med Public Health* 2006; **37**(6): 1103–1116.
- [20] Focks DA, Barrera R. Dengue transmission dynamics: assessment and implications for control. Report of the scientific working group meeting on dengue[R]. Geneva: WHO, 2006: p. 92–109.
- [21] Arcari P, Tapper N, Pfueller S. Regional variability in relationships between climate and dengue/DHF in Indonesia. *Singapore J Trop Geogr* 2007; **28**(3): 251–272.
- [22] Chadee DD, Shrivnauth B, Rawlins SC, Chen AA. Climate, mosquito indices and the epidemiology of dengue fever in Trinidad (2002–2004). *Ann Trop Med Parasitol* 2007; **101**(1): 69–77.
- [23] Yang HM, Macoris MLG, Galvani KC, Andrighetti MTM, Wanderley DMV. Assessing the effects of temperature on the population of *Aedes aegypti*, the vector of dengue. *Epidemiol Infect* 2009; **137**(8): 1188–1202.
- [24] Watts DS, Burke BA, Harrison RE, Whitmore A, Nisalak. Effect of temperature on the vector efficiency of *Aedes aegypti* for dengue 2 virus. *Am J Trop Med Hyg* 1987; **36**(1): 143–152.
- [25] Focks D, Alexander N, Villegas E. Multicountry study of *Aedes aegypti* pupal productivity survey methodology: findings and recommendations. Geneva: WHO, 2006.
- [26] Wu PC, Lay JG, Guo HR, Lin CY, Lung SC, Su HJ. Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *Sci Total Environ* 2009; **407**(7): 2224–2233.