A PREDICTIVE MODEL FOR DENGUE CASES IN SOUTHERN LAM DONG PROVINCE, VIETNAM, 2007-2017

Nguyen, A.D.T¹, Kositanont, U.¹*, Hinjoy, S.², Iamsirithaworn, S ²

¹ Faculty of Public Health, Thammasat University
² Department of Disease Control, Ministry of Public Health

*Corresponding author: Uraiwan Kositanont, Faculty of Public Health, Thammasat University, Klong Nueng, Klong Luang District, Pathum Thani 12120, Thailand.
Email: uriwan.kos@mahidol.ac.th

https://doi.org/10.32827/ijphcs.7.2.82

ABSTRACT

Background: Dengue is arthropod-borne disease and one of major causes of morbidity and fatality in the Western Pacific Region. Climate is considered one of the main factors for dengue transmission. Objective of the study is to determine correlations between climatic factors and dengue, using reported cases in Southern Lam Dong Province, Vietnam, 2007-2017 and to validate the predictive model for number of dengue cases in 2017 using data from 2007 to 2016.

Materials and Methods: This was a retrospective quantitative study. Spearman’s Rank test was used to examine the correlation between each climatic factor and dengue reported cases. Seasonal Autoregressive Integrated Moving Average (SARIMA) models using the training data set from 2007 to 2016, correlative factors of dengue cases, the Bayes Information Criterion (BIC) and improved Box-Jenkins models, were applied to predict dengue cases during 2017. There is a wide range of potential confounders for annual dengue epidemics such as mosquito ecology, population density, population immunity and dengue cycle. Amongst these factors, population density was forced into the predictive model. Data analysis was done using Excel and SPSS version 16.

Result: There were significant correlations between dengue cases and climatic factors, consisting of minimum temperature ($r = 0.384$, $p < 0.01$) and relative humidity ($r = 0.372$, $p < 0.01$). The SARIMA $(1,2,1) \times (1,1,1)_{12}$ model at lag one month was the best fitted model for predicting dengue cases.

Conclusion: Predicted cases from time series model would be imperative for controlling and preventing the occurrence of dengue epidemics in the community. This study used secondary data, so it was difficult to control the occurrence of missing data point. Climatic and non-climate factors should be considered in predictive models for dengue epidemiology in the future.

Keywords: climatic factors, dengue cases, correlation, predictive model, Lam Dong Province
1.0 Introduction

Dengue is a vector-borne disease and it is one of primary public health problems in Vietnam and other tropical and subtropical countries (WHO, 2009). This is the most speedily transmitting arthropod-borne viral disease in the world. The prevalence of dengue fever has increased steadily 30 times between 1960-2010 and they also continue to spread into other regions (WHO, 2009). The first cases of dengue was recorded during 1779 and 1780 in Cairo, Philadelphia and Batavia (Murray et al., 2013). From April to June of 1998, many countries in Southeast Asia, Latin America and the Western Pacific reported strangely number of dengue and DHF cases and the prolonged drought due to El Nino phenomenon was believed to be a contributing factor for the epidemic (Tipayamongkholgul et al., 2009; WHO, 1998). Today, there are nearly 100 million cases infected by dengue viruses with 500,000 obvious cases annually in more than 100 countries in Americas, Africa, Eastern Mediterranean, especially Southeast Asia and the Western Pacific are the areas where affected most seriously (WHO, 2018). Between 2001 and 2008, the highest numbers of dengue morbidity and fatality was found in four nations in Western Pacific Region, including Vietnam (WHO, 2009).

In Vietnam, the percentage of dengue haemorrhage in adults constituted for 44% by which changing in population structure with low birth rate from 1999 to 2014, (Cummings et al., 2009; Egger & Coleman, 2007; Simmons & Farrar, 2009). In Central Highlands Provinces, most of dengue cases were in adults (more than 15 years old). Furthermore, four serotypes of dengue virus (DENV) have been found in this region (Tuan et al., 2017). According to Vietnam Ministry of Health, 43,000 cases are recorded in 53 Provinces, with 28 deaths in Vietnam (Vo et al., 2017). Dengue is a year-round disease in Vietnam, but usually increases from June to November. In 2013, 66,000 dengue cases and 42 deaths have been reported. Over 85% of dengue cases and 90% of deaths occur in the Southern Provinces of Vietnam (WHO, 2018).

There is a wide range of risk factors for dengue epidemic such as weather change, social-economic variables of community, susceptible groups and ambient environment. However, climate is believed as one of main factors of dynamics of dengue transmission (Kuno, 1995; Morin et al., 2013). Precisely, temperature could have some impacts on the capacity of vector survival, life length of arthropods, or human interaction. Temperature, rainfall regime and relative humidity are often used in statistical analysis to evaluate the association between dengue and weather factors and built the predict model for dengue illness in some areas around the world (Gharbi et al., 2011; Guzmán & Kouri, 2003; H. V. Pham et al., 2011; Winwanitkit, 2005). In rainy season, egg, larvae as well as pupae of Aedes mosquitoes are flushed away by the rain in the short term, breeding habitat for mosquito is expanded for a long term. Whereas, there were an association between dry season and warmer temperatures and the growth of dengue virus and its vectors. Several former studies assessed the association between dengue occurrence and climate factors in Southeast Asian countries. (H. V. Pham et al., 2011; Thai et al., 2010; Toai et al., 2016; Vu et al., 2014).

Association between climate and dengue was found in a study in Can Tho city, Vietnam but only focused on the hospitalised cases (P. T. Nguyen et al., 2016). Furthermore, the influence of climatic variables on dengue cases with controlled confounding factors has not been
studied in Vietnam. Being aware of prevention is better than treatment, the study aims to determine the correlation between dengue cases and climatic factors in Southern Lam Dong Province, Vietnam. From that, dengue prediction model was created which is useful for preparing against dengue timely based on local climate data.

2.0 Materials and Methods

2.1 Study area

Southern Lam Dong Province (latitude 11°38’31” N and longitudes 108°26’0” E) which is consisted of five districts and one city. The area is in the tropical climate area with monsoon. The altitude varies from 400 metres to 1000 metres above sea level. There are two seasons in a year; dry season (December to April next year) and rainy season (May to November). The annual average temperature varies from 16.6 degree Celsius to 27.4 degree Celsius, the precipitation ranges from 2,500mm to 3,000mm and the humidity is normally above 80%. The population were 600,408 (in 2014) (T. C. Nguyen, 2015). These areas composed of variety of ethnics such as Kinh, K’ho, Tay, Nung. The industry in the Southern Lam Dong includes tea, coffee production, textile manufacture and mining. This area was chosen to study as similar characteristics in climatic condition and high number of dengue cases which represented the whole area.

Figure 1: Administrative maps of Lam Dong Province, Vietnam.
2.2 Data collection

This was a retrospective quantitative study. Daily dengue cases and annually population data were obtained from Lam Dong Preventive Medicine Centre. Daily climatic data were obtained from the National Centre for Hydro-meteorological Forecasting. Identified data from Jan 2007 to Dec 2017 were collected in the study.

Population data were obtained from Preventive Medicine Centres and local Health Centres in Southern Lam Dong Province. Only annually population data are available, monthly dengue incidence (DI) rates were calculated by formula:

\[ DI = \frac{\text{monthly dengue cases}}{\text{annually average population}} \times 100,000 \]

2.3 Data analysis

After collecting, daily dengue cases and daily climatic data were calculated into monthly data. Descriptive analysis was used to describe distribution of dengue cases. Spearman’s Rank test was used to examine the correlation between each climatic factor and reported dengue cases.

Seasonal Autoregressive Integrated Moving Average (SARIMA) models were used to predict number of dengue cases in 2017. There is a wide range of potential confounders for annual dengue epidemics such as mosquito ecology, population density, population immunity and dengue cycle. Amongst these factors, population density was chosen to support the model.

The Autoregressive Integrated Moving Average (ARIMA) model was enhanced from AR, MA and ARMA models, where ARMA is a combination of AR and MA. The ARIMA models were applied with the Box – Jenkins method, using previous observations and lag timing value to predict values in the future.

In the Auto-regressive AR model, the present value of the time series \( x_t \) associated to its former value \( (x_{t-1}, x_{t-2}, \ldots) \) and the current residual \( \varepsilon_t \). The model can be shown as:

\[ x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \] (1)

In the Moving Average MA model, the current value of the time series \( x_t \) associated to its present and former residual series \( \varepsilon_t, \varepsilon_{t-1}, \ldots \). The model can be shown as:

\[ x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{q-1} \] (2)
The ARMA (Autoregressive Moving Average) model is a combination of AR and MA models, where the present value of the time series $x_t$ associated to its former value $(x_{t-1}, x_{t-2}...)$ and present as well as former residual series $\epsilon_t, \epsilon_{t-1},...$. The model could be shown as:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q-1}$$  \hspace{1cm} (3)

Differencing process combine with ARMA model was necessary for dealing with non-stationary time series and forming the ARIMA model.

The model SARIMA (p,d,q) x (P,D,Q)s, an extension of ARIMA model, composing of non-seasonal part and seasonal part with Auto-Regressive AR (p), (P); Integrated I (d), (D) and Moving Average MA (q), (Q), respectively. “s” is length of the seasonal period; s=7 if a daily data time series in weekly cycle or s=12 if a daily data time series in monthly cycle. In this study, s=12 was used for analysing.

The SARIMA models were used with the Box – Jenkins method. The first step of the approach was to examine whether the observations was stationary or non-stationary. If non-stationary, an appropriate degree of differencing (d) and (D) will be applied to convert the time series into stationary. Then Auto-correlation function (ACF) and Partial Auto-correlation function (PACF) was tested to determine applicable values of p or P and q or Q. Two types of set in the series was established which involve training data (from January 2007 to December 2016) and validation data (in 2017) to perform data effectively.

Different formulations of the AR and MA terms were modelled. The final model was selected based on data analysis, previous researches and judging three measures: the root of mean square error (RMSE), the mean absolute percentage error (MAPE) and Normalised Bayesian Information Criterion (BIC) (Etebong, 2014). Models with lower of BIC, RMSE as well as higher MAPE and R-Square value were used for predicted values in the training as well as validation dataset. The Ljung – Box statistic was used to exam the appropriateness and sufficiency of the model (Ljung & Box, 1978).

The final selected model was used to predict the number of monthly cases until the December 2017. One-step forecast was applied to find the predicted value for training dataset while dynamic forecasting was used in validation dataset and the 2017 projection, starting on January 2007 to the December 2016. Based on observed values and predicted values of the preceding month, dynamic forecasts were compared with one-step forecast, respectively. Consequently, there were vulnerable to collected errors over time. All analyses were performed using Excel software, SPSS software version 16.0 and the level of significance was set at 1%.
3.0 Result

3.1 Distribution of dengue cases and dengue incidence

![Graph showing distribution of dengue cases by year in Southern Lam Dong Province, Vietnam, 2007-2017](image)

Figure 2: Distribution of dengue cases by year in Southern Lam Dong Province, Vietnam, 2007-2017

3038 dengue cases were reported during the study period. The highest (n=260) and lowest (n=8) number of cases were recorded in 2016 and 2011, respectively. The extraordinary peak was observed in 2016, followed by a sharp decline in 2017. Within the 11-year study period, peaks in the number of cases were observed in the months of June and August and then there was declined steadily by December of the same year.

3.2 Correlation of climatic factors and dengue reported cases

Table 1: Correlation of climatic factors and dengue reported cases

<table>
<thead>
<tr>
<th>Climatic factors</th>
<th>Median</th>
<th>Mean±S.D</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum temperature (°C)</td>
<td>21.60</td>
<td>21.05±1.53</td>
<td>.384** .000</td>
</tr>
<tr>
<td>Maximum temperature (°C)</td>
<td>29.83</td>
<td>30.12±1.43</td>
<td>-.118 .177</td>
</tr>
<tr>
<td>Average temperature (°C)</td>
<td>24.48</td>
<td>24.46±1.06</td>
<td>.160 .067</td>
</tr>
<tr>
<td>Average rainfall (mm)</td>
<td>12.52</td>
<td>12.28±5.81</td>
<td>0.153 0.082</td>
</tr>
<tr>
<td>Average relative humidity (%)</td>
<td>84.83</td>
<td>83.65±5.35</td>
<td>.372** .000</td>
</tr>
<tr>
<td>Extreme wind velocity (m/s)</td>
<td>2.18</td>
<td>2.24±0.64</td>
<td>-.210 .016</td>
</tr>
</tbody>
</table>

As shown in table 1, daily temperature ranged from 21.60°C to 29.83°C, precipitation was 12.28±5.81mm and the daily relative humidity was 83.65±5.53% during the period 2007-2017. Besides, there was a significant correlation of climatic factors and monthly dengue reported cases, including monthly minimum temperature and average relative humidity (p<0.000).
3.3 Predictive model for dengue reported cases

The life cycle of *Aedes* mosquito last from 4 days to 1 month, depending on the climatic statement. Based on immune system and virus types, incubation and infection of dengue virus in human blood last 10-12 days and 4-13 days, respectively (CDC, 2018; Rodenhuis-Zybert et al., 2010). Therefore, total time for life cycle of *Aedes* mosquito and dengue virus lasts from 18 days to 2 months. In this study, lag 1 and 2 months were chosen to build the predictive model using monthly climatic data.

**Table 2**: Model statistic using correletively climatic variables

<table>
<thead>
<tr>
<th>ARIMA (p,d,q) x (P,D,Q)</th>
<th>Model Fit statistics</th>
<th>Ljung-Box Q (18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p d q P D Q R-squared</td>
<td>RMSE</td>
<td>MAPE</td>
</tr>
<tr>
<td>Lag 1 month</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 1 1 1 1</td>
<td>.776</td>
<td>21.513</td>
</tr>
<tr>
<td>Lag 2 months</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 1 1 1 1</td>
<td>.767</td>
<td>21.920</td>
</tr>
</tbody>
</table>

Table 2 shows model fit statistics and significant value in Ljung-box test for each SARIMA model. The best model was SARIMA (1,2,1) x (1,1,1)\textsubscript{12} at lag 1 month, which fitted with the conditions. There was no significant difference in the value of the Ljung-Box Statistic (p=0.386), with a value of 14.882 for 14 df, hence failing to reject the null hypothesis of white noise which means the model has sufficiently matched the correlation in the time series. Moreover, the model was better as it has the low value of RMSE and BIC and the high value of the R-square.

**Figure 3**: Predicting monthly total dengue cases in 2017 by training data from 2007 to 2016 in Southern Lam Dong Province, Vietnam
Model SARIMA using correlative variables at lag 1 month was established. The results showed that SARIMA (1,2,1) x (1,1,1)_{12} using monthly relative humidity at lag 1 month was the most fitted model with low value RMSE of 21.338, BIC value of 6.429 and high value R-Square of 0.775. Estimation by Ljung-Box test with Q=14.364 for 14 d.f and p-value of 0.423 shown no autocorrelation between residuals at different lag times.

**Figure** Error! No text of specified style in document.: Predicted cases for dengue reported cases in Southern Lam Dong Province, 2017

Figure 4 shows the distribution of dengue cases in 2017, both for real cases and predicted cases. The real cases were lower than the predicted cases from February to September. The correlation between the real cases and predicted cases was relatively strong (r=0.680, p =0.015).

**4.0 Discussion**

The purpose of this retrospective quantitative study was to build a predictive model for dengue case using climatic variables in Southern Lam Dong Province, Vietnam. Via data of dengue reported cases from 2007 to 2017, there were not obvious in the annual epidemic cycle for dengue cases in Southern Lam Dong Province. In this period, there were the highest number of cases in 2016, accounting for nearly a half of total cases within 11 years.

While the dengue incidence rates in Southern area were relatively constant in most years in the study period. In 2016, however, the results increased extraordinarily. It was supposed that “El Nino” phenomenon was a factor for this change. “El Nino” was an alarm problem on global scale in the period 2015-2016 which increased in temperature, decreased in precipitation; drought and other disasters from “El Nino” causes plenty of serious damages such as in economic, agriculture and human health (FAO, 2016a). Vietnam was also impacted by El Nino in 2016 (FAO, 2016b). Extended drought conditions could promote Aedes mosquito breeding, because of increasing in water storage containers around home (Jansen & Beebe, 2010). The change in immunity population could be another factors for increasing in the dengue incidence rate (Schmidt et al., 2011).
Rainfall is an advantage factors for creating Aedes breeding sites. However, extended rainfall has a negative effect on breeding grounds and wash the larva and eggs away (Lim et al., 2013). Also, the combination of high temperature and rainfall contribute to the significant growth of mosquitoes (T. K. L. Pham et al., 2015). This was appropriate to this study when most of dengue cases occurred in the rainy season, from May to November. The highest number of cases were observed from June to August. This was similar to previous studies which conducted in the Southeast Asia such as Vietnam, Thailand, Philippines (Campbell et al., 2013; Tran et al., 2018; Undurraga et al., 2017).

The correlation between climate factors and dengue cases were found in monthly minimum temperature and relative humidity. In Can Tho City, it was demonstrated that there were a positive correlation between dengue hospitalization rate and humidity with a lag of one month (P. T. Nguyen et al., 2016). In Ho Chi Minh City, there is a positive association between relative humidity and a negative association between temperature and dengue incidence but not for rainfall. In Ha Noi capital, rainfall and temperature were positively correlated with dengue incidence (Vu et al., 2014).

Lagged effects of climatic factors on dengue has been found in previous studies (Choi et al., 2016; Ehelepola et al., 2015; Wang et al., 2012). The differences in correlation of climatic factors and dengue cases at each lag time between this study and others could be some discrepancies in regions, geography, climate and the quality of health care systems between study areas and other areas (Promprou et al., 2006). However, it was supposed that lag 1 month is suitable with Dengue life cycles, one for Aedes mosquito which lasts 4 days to three weeks and one for dengue virus in the human body which lasts about 5 days (WHO, 1997).

For predicting dengue cases, the models were trained using correlative variables at different lag months and population density as independent variables. This study identified that SARIMA (1,2,1) x (1,1,1)_{12} at lag 1 month is superior to other models for data series in predicting monthly dengue cases having a low BIC value and high R-square. The results of the model indicated that the predictive values were higher than the real values.

The values in SARIMA model have been different among regions. In Thailand, they showed different SARIMA models for dengue incidence in provinces in Northern Thailand (Silawan et al., 2008). A study in Can Tho, Vietnam showed that SARIMA model including AR(1) and SMA(1,12), with support of relative humidity at lag of 1 month was the best model for dengue hospitalisation (Toai et al., 2016). It could be explained that SARIMA could be built based on data series. As these data are distinguished among areas, the values in the model are not the same. However, by identifying factors for dengue and calculating, the predictive model could be established to predict the trend of dengue in near future.

A prediction for dengue cases from January 2017 to December 2017 was verified. Even the higher monthly predicted cases than the monthly real cases, the different numbers were 0-50 cases. As the time series model based on training data from 2007 to 2016; the number of dengue cases was extremely high in 2016 could affect the predicted cases in 2017 (Figure 4.20). Beside the climatic factor, population density and dengue cycle, immune population are factors for the extraordinary number of dengue in 2016 (Piedrahita et al., 2018; Thai et al., 2005). To improve the accuracy of the prediction model, other factors such as population immunity should force into the model. Monthly predicted cases were nearly close to monthly
real cases and high correlation (r=0.680), the model could be applied to predict distribution of dengue cases (Figure 4.21). The distribution of dengue cases in the predictive model would be imperative for controlling and preventing the occurrence of dengue epidemics in the community.

5.0 Conclusion and recommendation

From this study, SARIMA models using minimum temperature and relative humidity was established. SARIMA (1,2,1) x (1,1,1)_12 model is the excellent prediction model for dengue cases in Southern Lam Dong province, Vietnam. Further study should be suggested to combine climatic and non-climatic factors such as entomological, virological and anthropological factors in the predictive model for dengue cases.

Acknowledgement

We would like to express my gratitude to all of my lecturers, office staff of the faculty of Public Health, Thammasat University, which support me to conduct this study. We thank Lam Dong Province Preventive Medicine Centre, the local Health Centres and National Hydro-Meteorological station for allowing me to collect data, provide full support and information during the process of collection.

Declaration

Author(s) declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

Authors contribution

Author 1: An Dao Thien Nguyen  
Author 2: Uraiwan Kositanont  
Author 3: Soawapak Hinjoy  
Author 4: Sopon Iamsirithaworn
References


Nguyen, A.D.T, Kositanont, U., Hinjoy, S., Iamsirithaworn, S. https://doi.org/10.32827/ijphcs.7.2.82


Nguyen, A.D.T, Kositanont, U., Hinjoy, S., Iamsirithaworn, S. https://doi.org/10.32827/ijphcs.7.2.82


