Acta Manilana 60 (2012), pp. 19–26
Printed in the Philippines
ISSN: 0065–1370

Forecasting dengue incidence in the National Capital Region, Philippines: using time series analysis with climate variables as predictors

Arjelou C. dela Cruz, Jonathan A. Lubrica, Bill Vincent D.C. Punzalan, & Mary C. Martin*

Department of Mathematics and Physics, College of Science
University of Santo Tomas, España Boulevard, 1015 Manila, Philippines

The Philippines, like many of the developing countries, is among the most vulnerable to the impacts of climate change. As changes in rainfall patterns exacerbate the incidence of vector-borne diseases such as dengue, the country is in need of efficient tools to provide for better dengue prevention and control programs. The National Capital Region (NCR) consistently belongs to the regions with the highest number of dengue cases per year. This study provides a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model that could predict dengue outbreaks based on the past dengue incidences and climate variables in NCR. Monthly dengue incidence data from January 2005 to December 2010 were fitted into a SARIMA model (RMSE = 0.3365). All the climatic variables showed positive correlation with the number of dengue cases. The predictive power of the SARIMA model was enhanced by the inclusion of climatic variables as external regressors to forecast the year 2011.

Keywords: dengue incidence, climate variables, predictors

Introduction

Dengue fever is a disease caused by dengue viruses carried and transmitted by female, day-biting mosquitoes. It causes severe flu-like illness and may be potentially fatal with the complications of a dengue hemorrhagic fever (DHF). DHF was first recognized in the Philippines during the epidemic in 1953 [1]. The primary vector responsible for dengue fever transmission is the *Aedes aegypti*. It is particularly susceptible to climate variability and climatic change. The World Health Organization considers dengue to be the most important vector-borne viral disease, potentially affecting 2.5 billion people in tropical and sub-tropical countries throughout the world [2].

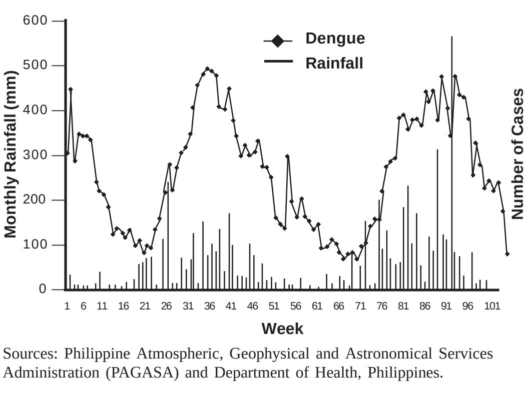
J. Rocklöv and A. Wilder-Smith [3] cited weather variability, seasonal and long-term climatic variations have been shown to be associated with dengue activity. Meteorological variables and its consequences on vector population will be of assistance to target control measures and policies and contribute to the development of climate-based control and surveillance measures [4]. In the Philippines, there is a clear correlation between the incidence of dengue fever and rainfall [5] as shown in Fig. 1.

**To whom correspondence should be addressed*
E-mail: mcmartin@mnl.ust.edu.ph

Dengue cases in NCR usually are on the rise in the months of July to October. As of September 2011, there were 21,088 reported dengue cases compared to 16,237 cases reported the same month in 2010 (Data from Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA).

Environments and climatic variables such as temperature, humidity, season and precipitation significantly influence mosquito development. Temperature affects the development of the mosquitoes, as well as dengue viral development [6]. A study in Trinidad and Tobago indicated that it takes six months for environmental conditions measured by temperature alone to affect dengue incidence. It further revealed that higher temperatures may hasten the incubation period resulting in greater transmission [7].

The efficiency of vector control campaigns can be enhanced if an early warning of a dengue outbreak can be carried out. Several mathematical models using statistical tools had been developed to measure relationships between climatic variables and dengue. Recent studies in Bangladesh [8], French West Indies [9], China [10], and Brazil [11] employed the use of time-series analyses in developing autoregressive integrated moving average (ARIMA) models to predict dengue incidence. There are no similar models developed yet in the Philippines.



Sources: Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) and Department of Health, Philippines.

Figure 1. Monthly rainfall and number of dengue fever cases in the Philippines, 2008–2009.

The objectives of this study are to develop a model that could predict dengue outbreaks in the National Capital Region (NCR) based on the past dengue incidences and climate variables such as relative humidity, rainfall, minimum, mean and maximum temperature, and to quantify the strength of the relationship between dengue incidence and the identified climate variables.The model is intended to give a prediction of future dengue outbreaks, which will aid in providing an appropriate, efficient, and timely implementation of dengue prevention activities.

Methods

Settings. The National Capital Region is the smallest administrative region in the Philippines but the most populous at 11,855.975 (2010 census) and having a density of 18,641 per square kilometers [12]. Its area is approximately 636 square kilometers. Climate in Metro Manila is classified into dry season (November to May) and the rainy season (June to October). Warm months are April and May.

Data collection. The model in this study was generated using the monthly data from January 2005 to December 2010. The monthly data from January 2011 to September 2011 were used for validating the model. The data on the climate variables were gathered from the Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA), particularly from the Science Garden, Quezon City weather station. On the other hand, the data on the number of dengue cases in NCR were gathered from the Department of Health (DOH).

Processing and data analysis. The predictive model for dengue incidence in this study was generated using the Box-Jenkins methodology for ARIMA model. ARIMA models were popularized by George Box and Gwilym Jenkins in the early 1970s. This method of forecasting is different from most methods because it does not assume any particular pattern in the historical data of the series to be forecasted. It uses an interactive approach of identifying a possible model from a general class of models. This, specifically, is the Box-Jenkins Methodology for ARIMA Models. The Box-Jenkins methodology refers to a set of procedures for identifying, fitting, and checking ARIMA models with time series data. Forecasts follow directly from the form of the fitted model. The basis of Box-Jenkins approach to modeling time series consists of three phases: (1) identification, (2) estimation and testing, and (3) application. If necessary, the initial model is modified and the process is repeated until the residual indicate no further modification is necessary. At this point the fitted model can be used for forecasting.

In the identification phase of this study, the data were prepared by performing a log-transform to stabilize the variance. This transformation is particularly useful in time series analysis. Then, a first seasonal differencing was applied to the dengue incidence data to achieve *stationarity*. Stationarity of a series is an important phenomenon because it can influence its behavior. For example, the term ‘shock’ is used frequently to indicate an unexpected change in the value of a variable (or error). For a stationary series, a shock will gradually die away. That is, the effect of a shock during time ‘t’ will have a smaller effect in time ‘t+1’, a still smaller effect in time ‘t+2’, and so on [13].

When the series is stationary, the time series plot shows no change in the mean over time, and there is no obvious change in the variance over time. This is essential so that other correlation structure present in the series can be seen before proceeding with model building.

In this study, stationarity was achieved by taking the first seasonal difference of the dengue incidence data. A seasonal difference is the difference between an observation and the corresponding observation from the previous year, as described by the equation

$$y'_t = y_t - y_{t-s}$$

where s is the length of the season. The seasonally differenced series represents the change in dengue incidence between months of consecutive years.

The next step for the Box-Jenkins approach is the estimation and testing phase. The orders of the moving average (MA) and autoregressive (AR) parameters were identified by the analysis of the autocorrelation function (ACF) as a check for seasonality and partial autocorrelation (PACF). Since there are more than one plausible models identified, different methods were used to determine which of them is preferred. The model with the lowest Akaike Information Criterion (AIC) was selected. AIC is a goodness of fit measure used to assess which of two ARIMA models is better, when both have acceptable residuals. Usually, the model with the smallest AIC will have residuals, which resemble white noise. Occasionally, it might be necessary to adapt a model with not quite the smallest AIC value, but with better-behaved residuals. Before using the model for forecasting, it must be checked for adequacy. It is adequate if the residuals left over after fitting the model are simply white noise.

Lastly, the model generated was applied and used to forecast future dengue incidence. The predicted data for January 2011 to June 2011 were compared with the observed data in order to validate the model. The error was analyzed using the root mean square error (RMSE).

After the univariate model was selected, the multivariate models including external regressors could be elaborated by using the climatic variables to improve the predictive power of the model. Cross-correlation analyses between the data of dengue incidence and climatic variables were performed. The predictive power of the models was estimated using the RMSE, wherein a significant decrease in the RMSE denotes an improvement of the model. The Wilcoxon Signed Rank test was used to determine whether the improvements in RMSE are statistically significant.

All the analyses, computations and graphs were made using the statistical software JMP and IBM SPSS 20.

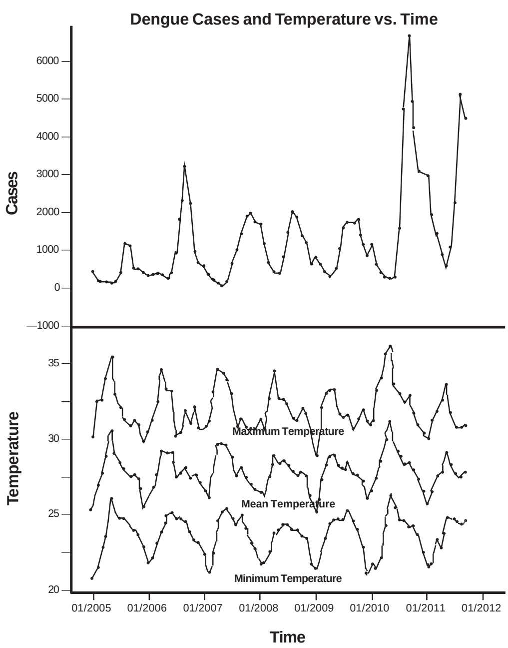
RESULTS

The time series plot of the number of dengue cases in Fig. 2 shows major dengue outbreaks in September 2010 and August 2011, which is not clearly attributable to the changes in the temperature.

For the graph of the relative humidity (Fig. 3), it can be observed that there are some similar patterns on the spikes of the graph of dengue cases and relative humidity, indicating a good correlation between them.

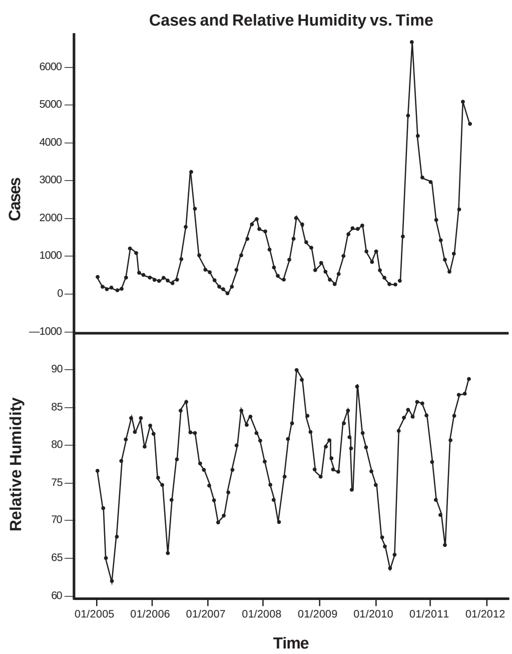
In the time series graph of the amount of rainfall (Fig. 4), there are some significant similarities with the trends in the dengue incidence.

The three climatic variables namely, temperature (minimum, mean, and maximum), relative humidity, and amount of rainfall, are the three climate variables considered for this study. An annual seasonality is identified for all these variables.



Sources: Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) and Department of Health, Philippines

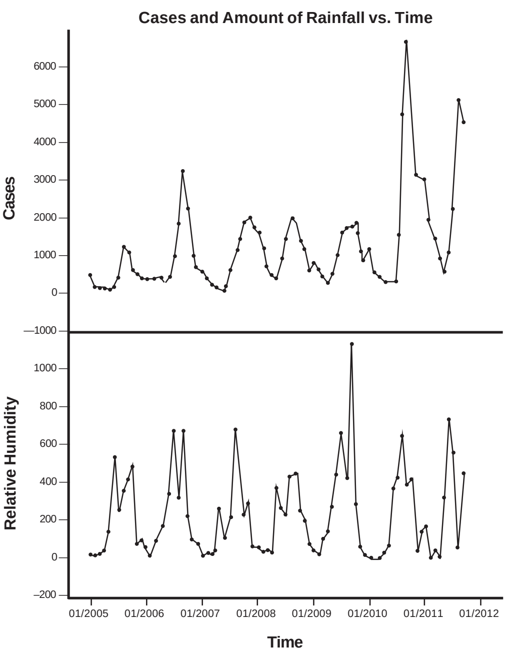
Figure 2. Number of dengue cases and temperature trends from January 2005 to September 2011.



Sources: Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) and Department of Health, Philippines

Figure 3. Number of dengue cases and relative humidity trends from January 2005 to September 2011

To stabilize the variance, a log transformation was done for the number of dengue cases. A first order seasonal differencing was applied to the natural logarithm of the number of dengue cases since the data is seasonal and non-stationary. Then, the data were checked for stationarity using the time series, ACF and PACF plots. The plots of ACF and PACF confirmed the need to use a seasonal ARIMA or SARIMA model with non-seasonal (p, d, q) and seasonal (P, D, Q) parameters (Fig. 5). After the model was selected, the parameters were estimated (Table 1) using the maximum likelihood procedure in the statistical software JMP.



Sources: Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) and Department of Health, Philippines.

Figure 4. Number of dengue cases and amount of rainfall trends from January 2005 to September 2011

In order to include the climatic variables, it is necessary to determine how the climatic variables are related to the number of dengue cases, so cross correlation analyses were performed. The residual cross correlation analysis showed which lagged values of the climatic variables are most positively related to the number of dengue cases within a period of 6 months. The climatic variables correlated to dengue incidence were included one by one, to test their influence on the model. Some of the observed correlations are moderate since they are measured after an adjustment on the trend and the seasonality. However, they are statistically significant, as indicated by the correlation coefficients, which means that the test is powerful.

Table 2 shows the improvements done by the inclusion of the climatic variables. The improvements were measured through the difference in the RMSE. Minimum temperature and relative humidity gave the best results while rainfall gave the least.

The Wilcoxon Signed ranked test was used to show that all improvements in the predictions of the univariate model through the inclusion of climatic variables were statistically significant, with $r < 0.05$.

The best model selected is the SARIMA (1, 0, 1) (0, 1, 1)12 with minimum temperature lag-2 as an external regressor. The predictions for the 2

Table 1. Parameter estimates of the Dengue SARIMA (1, 0, 1) (0, 1, 1) 12 Model

Parameter	Estimate	Standard Error	*t*-Stat.	ρ-Value
AR1, 1	0.825559	0.107219	7.7	<.0001*
MA1, 1	–0.37714	0.137686	–2.74	0.0082*
MA2, 12	0.921466	0.921105	1	0.3213

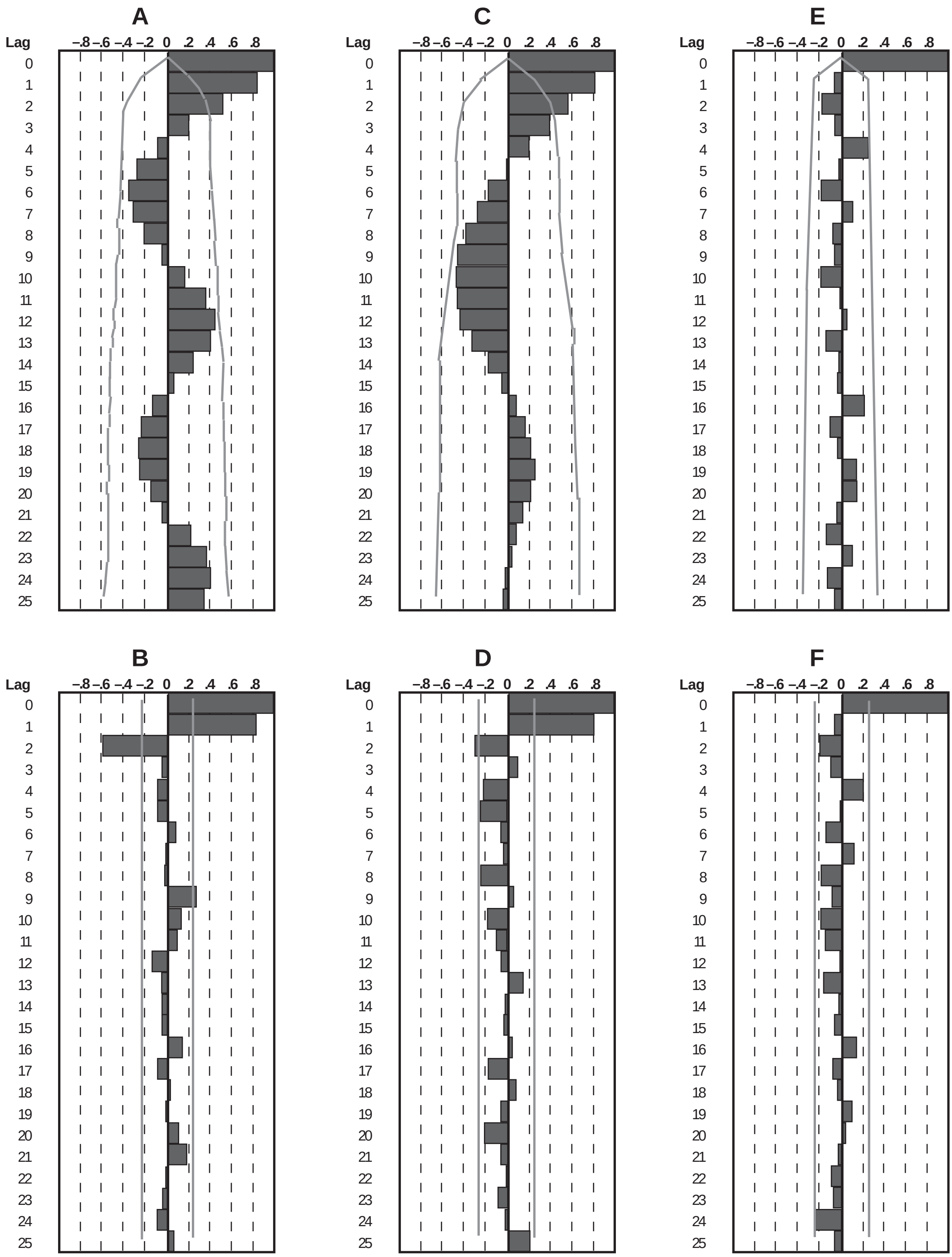


Figure 5. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. A and B Original dengue incidence; C and D Dengue incidence after differencing; E and F Residuals after applying a SARIMA (1, 0, 1) (0, 1, 1)12 model. The X-axis gives the value of the correlation coefficient and, the Y-axis gives the number of lags. Blue lines indicate 95% confidence interval.

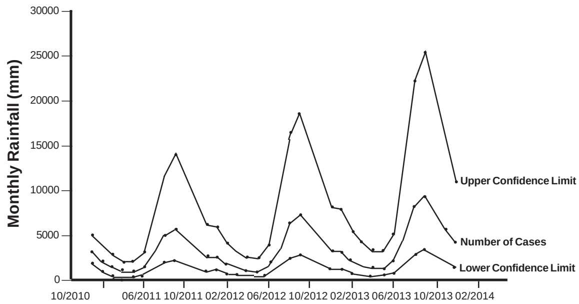


Figure 6. Forecasted dengue incidence from 2011 to 2014 using the SARIMA (1,0,1)(0,1,1)12 model with minimum temperature as predictor

Table 2. Model comparison through RMSE and AIC

Model	Coefficients	p-Value	AIC	RMSE
Univariate	–	–	45.678	0.3365
Minimun Temperature Lag_2	0.0106	0.0012	39.858	0.2576
Mean Temperature Lag_3	0.0084	0.0024	40.7065	0.2600
Maximum Temperature Lag_3	0.0073	0.0024	40.6477	0.2594
Relative Humidity Lag_2	0.0558	0.0014	40.6646	0.2579
Rainfall Lag_1	0.0372	0.0149	44.5386	0.2988

following months were the best compromise for helping the health authorities to take measures to mitigate transmission, morbidity, and mortality. Figure 6 presents the predictions of dengue incidence using the model, including its upper and lower confidence intervals.

The delayed effect of minimum temperature on dengue incidence could be explained by the climatic factors which influence indirectly the incidence of dengue through the effect on the life-cycle of the vector. Depending on the respective lag between the biological cycle and the clinical symptoms, the lag between the weather data and the incidence data will differ.

Conclusions

Climatic variables such as temperature, relative humidity, and rainfall are significantly correlated to dengue incidence. Minimum temperature is the best predictor that would improve the model, with a lag of two months. This is due to the fact that the lifespan of mosquitoes is affected by temperature, which consequently affects the level of dengue transmission.

Acknowledgment

We thank the Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA) for providing us with the data on the climatic variables used in this study. We also express our sincere gratitude to the Department of Health (DOH) for providing us with the data on the number of dengue cases.

References

[1] World Health Organization. *Dengue haemorrhagic fever; diagnostic, treatment, prevention and control* (2nd Edition). (Geneva, **1997**)

[2] World Health Organization. *Dengue: Guidelines for diagnosis, treatment, prevention and control* (New Edition). (Geneva, **2009**).

[3] Rocklov J & Wilder-Smith A. Climate change and vector-borne infections Comment on 'Modeling the impact of global warming on vector-borne infections' by Eduardo Massad, Francisco Antonio Bezerra Coutinho, Luiz Fernandes Lopez and Daniel Rodrigues da Silva. *Physics of Life Reviews* **2011**; 8:204–5.

[4] The Burden of Dengue. *Lorenzo* (**2006**, September 26).

[5] Asian Development Bank. *Accounting for Health Impacts of Climate Change*, **2011**.

[6] Kongnuy R & Pongsumpun P. Mathematical Modeling for Dengue Transmission with the Effect of Season. *International Journal of Biological and Life Sciences* **2011**; 7(3).

[7] Wegbreit J. The Possible Effects of Temperature and Precipitation on Dengue Morbidity in Trinidad and Tobago: A Retrospective Longitudinal Study. In: Arlinghaus SL & Drake WD (Eds.) *Population-Environment Dynamics: Issues and Policy*. (University of Michigan, **1997**)

[8] Choudhury MZ, Banu S, & Islam MA. Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis. *Dengue Bulletin* **2008**; 32.

[9] Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, Girdary L, *et al.* Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors. In: Gharbi *et al. BMC Infectious Diseases* **2011**; 11(166).

[10] Lu L, Lin H, Tian L, Yang W, Sun J, & Qiyong Liu. Time series analysis of dengue fever and weather in Guangzhou, China. *BMC Public Health* **2009**.

[11] Luz PM, Mendes BVM, Codeco CT, Struchiner CJ, & Galvani AP. Time Series Analysis of Dengue Incidence in Rio de Janeiro, Brazil. *American Journal of Tropical Medicine* **2008**; 933–9.

[12] World Bank. *World Development Report — Development and Climate Change*. (Washington, DC: The World Bank, **2010**).

[13] Mohamed IE. *Time Series Analysis Using SAS.* (NorthEastSASUsersGroup, **2008**).