

# Time series analysis of dengue cases in Central Luzon: comparison of stochastic models using climate variables as predictors

*Angel R. Ang, Kevin Q. Nabong, & Mary C. Martin\**

*Department of Mathematics and Physics, College of Science,  
& Research Center for the Natural and Applied Sciences  
University of Santo Tomas, 1015 Manila, Philippines*

---

This study compares multiple stochastic models generated in order to identify which of them best fits the data. Monthly dengue cases from 2003 to 2013 in Central Luzon were used to generate the model. Seasonal Autoregressive Integrated Moving Average (SARIMA) is used to predict dengue cases (RMSE = 0.442) based on past dengue outbreaks with climate variables as predictors. It shows that the predictive variables, individually, brings about significant changes with the predicted dengue occurrence, lowering its RMSE to 0.402, and 0.403. Via cross correlation, it is found that the minimum temperature has the highest correlation to dengue occurrence ( $r = 0.6332$ ) with RMSE = 0.402. The root mean square error is then used to compare and determine which among the stochastic models best forecasts dengue incidence.

**Keywords:** time series analysis, dengue, Central Luzon, stochastic model, climate variables

---

## INTRODUCTION

Today, dengue ranks as the world's most lethal mosquito-borne viral disease. It is also the most rapidly spreading mosquito-borne viral disease in the world [1]. Increasing amount of evidence suggests that this infectious disease is under-recognized and inadequately represented in terms of estimation of disease burden and cost. Under-recognition of dengue's impact in the population could have harmful effects, since accurate quantification is critical for officials to set health priorities

and make plans to implement disease control measures based on disease burden. Dengue is a threat to nearly half the world's population, and is a pressing public health priority in many countries in Asia and Latin America where epidemics occur [2]. The Asia-Pacific bears 75% of the current global dengue disease burden [1].

The National Epidemiology Center of the Philippines' Department of Health reports a total of 59,943 dengue cases from January 1 to September 6, 2014. Of the total cases, 8.01% came from Region III, Central Luzon [3].

---

\*To whom correspondence should be addressed  
marycmar@yahoo.com

Human health is one of the most vital sectors, which will be severely affected by climate change. Incremental increases in temperatures and rain regimes could trigger a number of adverse impacts; in particular, the outbreak and spread of water-based and vector-borne diseases leading to higher morbidity and mortality. The rise in global temperatures poses strong risks for natural systems and human societies. Around the world, countries are already facing negative impacts of global warming, changes in rainfall patterns, and increasing number of extreme weather events like cyclones, flooding, and drought. But only a few countries are as exposed and vulnerable to climate change as the Philippines.

Climate influences dengue ecology by affecting vector dynamics, agent development, and mosquito/human interactions. Although these relationships are known, the impact climate change will have on transmission is unclear. Links between climate and diseases with various modes of transmission (vector-, water-, food-, soil- and airborne) have been identified with the strongest associations being between climate and mosquito-borne diseases [4, 5].

There have been several studies in various countries to assess claims on the relationship between climate and dengue using different statistical tools such as correlation coefficients, linear univariate regression, linear multivariate regression, and time series analysis. Studies in Thailand [6, 7], Bangladesh [8], Singapore [9] and Trinidad and Tobago [10] revealed that increased temperature is correlated with dengue incidence. Rainfall is correlated also with dengue incidence from studies in Indonesia [11], Trinidad [10] and Thailand [7].

Time series analysis had been used in developing Autoregressive Integrated Moving Average (ARIMA) models to predict dengue outbreaks. Among the countries who developed

ARIMA models for dengue assessment are Bangladesh [8], Brazil [12], China [13], and French West Indies [14].

A study using ARIMA models was done in the Philippines for the National Capital Region (NCR) [15]. Results show that climatic variables such as temperature, relative humidity, and rainfall are significantly correlated to dengue incidence.

This study aims to generate a prediction model of dengue outbreaks in the Central Luzon Region based on past dengue occurrences and climate variables. A comparison is done for the generated stochastic models using different climate variables as predictor so as to obtain the best model. Specifically it aims to identify the climate variables which are of significant relationship to dengue and quantify the strength of the relationship between dengue occurrence and individual climate variables.

## EXPERIMENTAL

**Settings.** Central Luzon is located north of the nation's capital, Manila, and is inhabited by a population of 11,218,177 [16]. In an area of 21,543 km<sup>2</sup>, it has been the center of attention by the country's health department due to its alarming susceptibility and incidence of the dengue fever. According to the disease surveillance report of the DOH-Central Luzon conducted on 18<sup>th</sup> September 2014, 131 barangays in Central Luzon have been listed to have clustering of dengue cases [17].

**Data collection.** Monthly cases of dengue fever from year 2003 to 2013 were gathered from the Department of Health (DOH). Also, monthly data of Climate variables (i.e. humidity, rainfall, minimum and maximum temperature) were obtained from the Philippine Atmospheric, Geophysical, and Astronomical Service Administration (PAGASA).

Both dengue cases and climate variables were provided in Microsoft Excel file formats, so it was best to analyze them using the same program. In addition, SAS JMP 11 was used in comparison and contrasts of graphs and other visuals.

**Data processing and analysis.** The Box-Jenkins ARIMA models were used and compared with Neural Network in forecasting dengue incidence. These methods were also applied to identify which of the variables is the most significant. More precisely, the one which yields a lower RMSE will be favored.

The Box-Jenkins Seasonal Modeling. Data transformations made the data stationary, namely logarithmic transformation and 1<sup>st</sup> order Seasonal Differencing. Then, appropriate SARIMA models have been fitted to the data. The results from SARIMA are then compared to MATLABR2014's Neural Networks.

**RESULTS**

**SARIMA method.** The graphs of dengue cases and the climatic variables were observed to have similar trends and pattern. This feature implies that all of the variables exhibit a common cyclical pattern. The dengue incidence had an irregularity between months 100 and 120. This behavior suggests an inconsistent variance in the series.

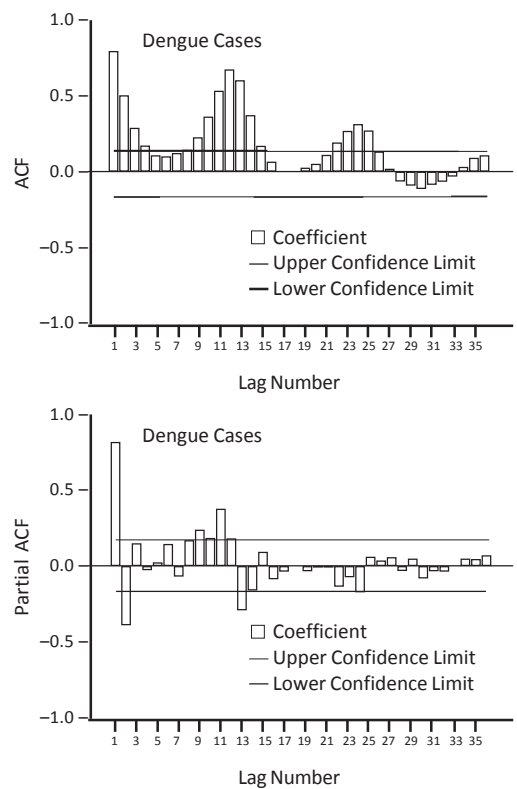
Further analysis showed that the minimum and maximum temperature and relative humidity displayed a constant mean and variance over time. Also, among the climate variables, the plot of the relative humidity had the strongest fluctuations in a single cycle.

Due to the inconsistency of the variance within the dengue series, the data could be transformed using its natural logarithmic equivalent,  $y^* = \ln y$ , where  $y^*$  is now the transformed value of the original data,  $y$ . This transformation enabled an easier comparison within and among

variables. For these same reasons, logarithmic transformation was also applied to the variable amount of rainfall.

After applying logarithmic transformation to the dengue cases, the resulting series incidentally exhibited a linear trend with a positive slope. Considering the nature of the data, a seasonal differencing was applied. A traditional tool to distinguish stationarity within a series is the Autocorrelation Function, ACF. If the time series is stationary, the ACF will die out quickly, otherwise, the series is nonstationary. The results shown in Fig. 1 indicates a nonstationary time series.

A conditional correlation used is the Partial Autocorrelation Function PACF. Behaviors of the ACF and PACF coefficients in Fig. 2 imply



**Figure 1.** Plot of the ACF coefficients for the logarithm of the number of the dengue cases

a Seasonal Autoregressive Integrated Moving Average (SARIMA) model. Identifying a model requires inspection in both nonseasonal ( $p, d, q$ ) and seasonal level ( $P, D, Q$ ) of the ACF and PACF. The criteria of the ARIMA model can be identified through inspecting at what lag the ACF/PACF cut off.

The stationarized dengue series exhibits, at a nonseasonal level, an ACF that dies down quickly and a PACF that cuts off after lag 3. This suggests an AR (3) model or an ARIMA (3, 0, 0). At seasonal level, the ACF cuts off at lag 1 and the PACF dies down quickly. This suggests an SMA (1).

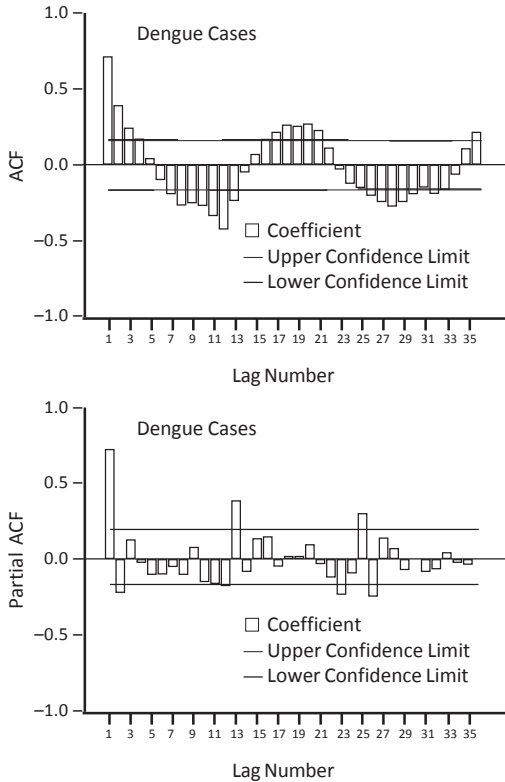


Figure 2. ACF and PACF plots of the dengue time series after log transformation and 1<sup>st</sup> order seasonal differencing.

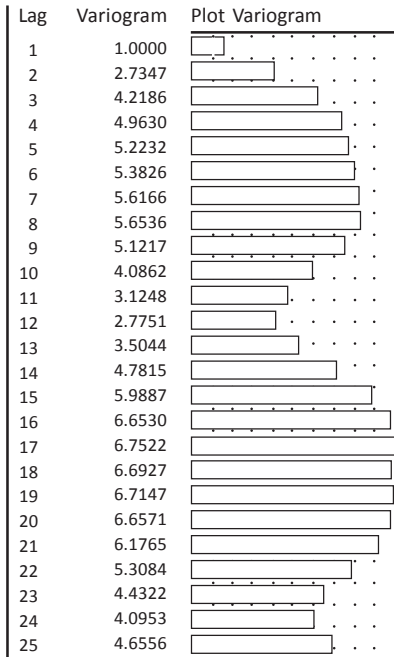
To further justify the stationarity of the transformed series, a more specific tool designed to determine stationarity alone is used. This tool, the Variogram, was defined for both stationary and (many) nonstationary processes, unlike the autocorrelation. If a process is stationary, the variance of differences between observations far enough apart will stabilize and the variogram will flatten out and come to a stationary asymptote. A variogram of the number of dengue cases before and after transformation is illustrated in Fig. 3 and Fig. 4.

**Criterion Filling – Identification of  $(p, 0, q)(P, 1, Q)$ .** ACF and PACF provided a tentative set of values for  $(p, d, q)(P, D, Q)$  which is still subject to verification since some other values can be associated to the analysis of ACF and PACF. Akaike’s Information Criterion (AIC) was applied to choose which among the models best fit the data. The model with the lowest AIC was chosen. The Bayesian Information Criterion (BIC) was also observed, to ensure the sensitivity of the model to its number of parameters. Table 1 shows the list of AIC and BIC of other possible SARIMA models.

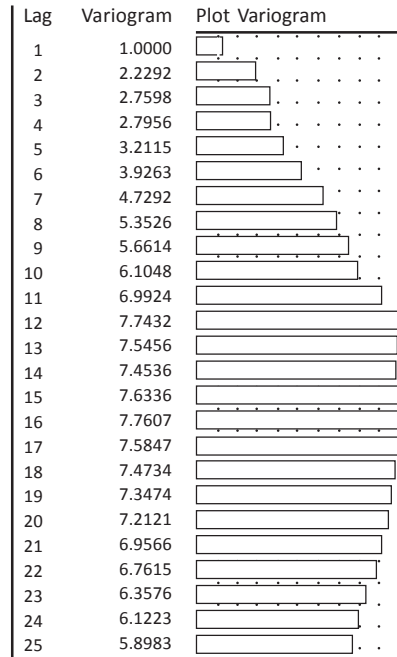
Table 1. Models of seasonal ARIMA Climate variables with estimates

Seasonal ARIMA Model	AR	MA	SMA	AIC	BIC
Maximum Temperature $(1,0,1)(0,1,1)_{12}$	0.75816408	0.43838571	1.00000*	258.806	267.169
Minimum Temperature $(1,0,1)(0,1,1)_{12} + c$	0.42928644	-0.0515006	0.80122	145.5031	156.653
Relative Humidity $(1,0,1)(0,1,1)_{12}$	0.54627886	0.10677508	0.84246	533.3763	541.739
Rainfall $(1,0,0)(0,1,1)_{12} + c$	0.11500734	-	1.00000*	320.2837	328.646

## Time series analysis of dengue cases in Central Luzon



**Figure 3.** Plot of the logarithmic transformation of dengue series before a 1<sup>st</sup> order seasonal differencing



**Figure 4.** Plot of the logarithmic transformation of dengue series after a 1<sup>st</sup> order seasonal differencing

**Table 2. Summary of the Ljung-Box Statistics of the dengue cases**

Climate Variable	Ljung-Box (Q18)	
	Statistics	Sig.
Maximum Temperature	15.15	0.368
Minimum Temperature	15.77	0.327
Relative Humidity	15.213	0.364
Rainfall	13.47	0.49

All climate variables were significantly related to the dengue cases, since the patterns or behaviour that may have been missed by the model after applying it was not significant, as measured by the Ljung-Box statistic - having a p-statistic (Sig.) greater than 0.05 as shown in Table 2.

According to the cross correlation coefficients, as illustrated in Table 3, the minimum temperature at lag 3 when treated as a time series by itself shows the highest similarity to the dengue series.

**Table 3. Correlation coefficients of the climatic variables to the number of dengue cases**

Climate Variable	Correlation
Lag (Maximum Temperature, 4)	0.4691
Lag (Minimum Temperature, 3)	0.6332
Lag (Relative Humidity, 1)	0.5827
Lag (Rainfall, 1)	0.4895

Figure 5 shows the graph of dengue incidence when the best fit model (SARIMA (3,0,0) (0,1,1)<sub>12</sub>) is applied. The graph evidently exhibits a similar trend with the actual data. It can also be observed that the application of the external regressor, minimum temperature, posed significant positive changes in the graph, as the Ljung-Box statistics initially determined.

**Neural network.** Neural Network belongs to a family of statistical learning algorithms inspired by the way information are processed

by the biological nervous system such as the brain.

The MATLAB neural time series tool (*ntstool*) was used with dengue series as the target output and climactic variables as the input data. All three default training algorithms - Levenburg-Marquardt, Bayesian Regularization, and Scaled

**Table 4. MSE and RMSE of climate variables as inputs and dengue series as target data, with Levenburg-Marquardt as training algorithm**

Regressor	Levenburg-Marquardt	
	MSE	RMSE
None	0.2624	0.5123
Maximum Temperature	0.2452	0.4952
Minimum Temperature	0.1578	<b>0.3973</b>
Relative Humidity	0.2997	0.5475
Rainfall	0.3151	0.5613

**Table 5. MSE and RMSE of climate variables as inputs and dengue series as target data, with Bayesian Regularization as training algorithm**

Regressor	Bayesian Regularization	
	MSE	RMSE
None	0.2588	0.5087
Maximum Temperature	0.2699	0.5195
Minimum Temperature	0.1110	<b>0.3332</b>
Relative Humidity	0.1252	<b>0.3539</b>
Rainfall	0.1630	<b>0.4037</b>

**Table 6. MSE and RMSE of climate variables as inputs and dengue series as target data, with Scaled Conjugate Gradient as training algorithm**

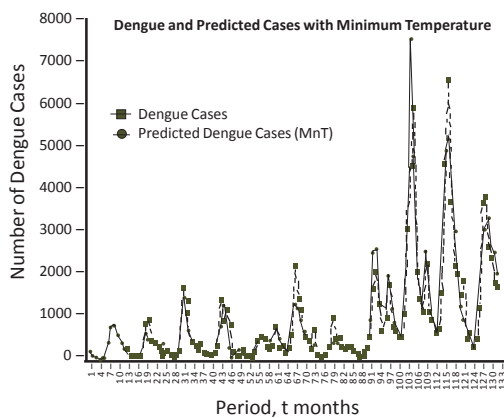
Regressor	Scaled Conjugate Gradient	
	MSE	RMSE
None	0.3960	0.6292
Maximum Temperature	0.2795	0.5287
Minimum Temperature	0.3346	0.5784
Relative Humidity	0.3836	0.6193
Rainfall	0.4549	0.6745

Conjugate Gradient - in MATLAB were applied in generating the desired output layer, yielding three forecast results per variable per algorithm.

Within the 15 forecast columns from Microsoft Excel, only the ones with the lowest RMSE compared to Seasonal ARIMA were chosen. Table 4, Table 5, and Table 6 present a summary of the MSE and RMSE of the output of the three training algorithms for the various climate variables input.

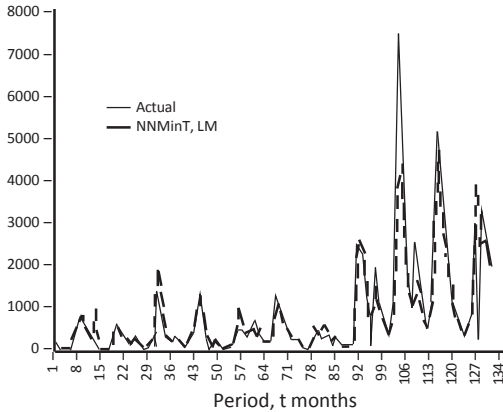
Predictions chosen from the Neural Network are: *minimum temperature* as input value with Levenburg-Marquardt, also known as damped least squares used to solve non-linear least squares problems, as training algorithm. On the other hand, *minimum temperature, rainfall amount, and relative humidity* were the input values with Bayesian Regularization as their training algorithm.

The use of Scaled Conjugate Gradient as training algorithm yielded high RMSE and this was not used for comparison with seasonal ARIMA.

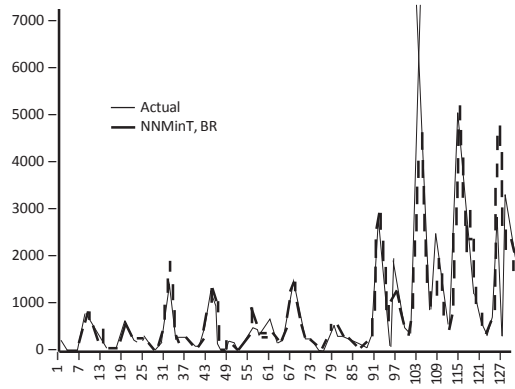


**Figure 5.** Plotted time series of actual and predicted dengue cases with Minimum Temperature as an external regressor after the application of SARIMA  $(3, 0, 0)(0, 0, 1)_2 + c$  in Central Luzon from January 2003 to December 2013

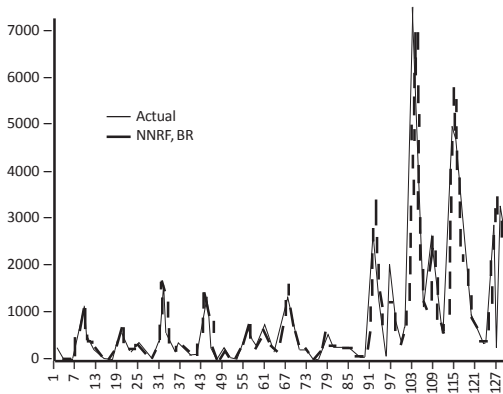
## Time series analysis of dengue cases in Central Luzon



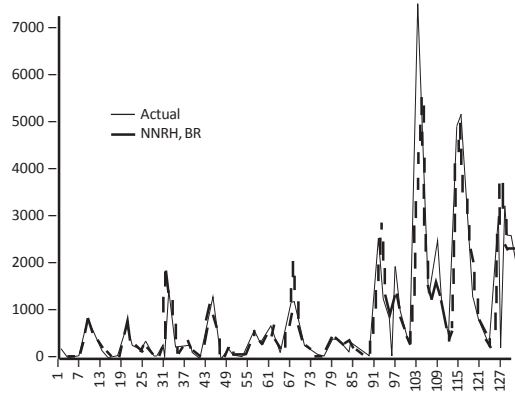
**Figure 6.** Actual and predicted dengue cases of Neural Network with minimum temperature using Levenburg-Marquardt as training algorithm.



**Figure 7.** Graph of actual and predicted values of Neural Network with minimum temperature using Bayesian Regularization as training algorithm.



**Figure 8.** Graph of actual and predicted values of Neural Network with rainfall using Bayesian Regularization as training algorithm.



**Figure 9.** Graph of actual and predicted values of Neural Network with relative humidity using Bayesian Regularization as training algorithm.

The prediction figures from neural network algorithms are shown in Fig. 6 to Fig. 9. Using minimum temperature as input values, the Levenburg-Marquardt training algorithm predicts a peak incidence of dengue in 2012 (Fig. 6). Actual cases show a high incidence between 2010 and 2011.

For the Bayesian Regularization algorithm, a peak dengue incidence was predicted between 2010 and 2011, using minimum temperature (Fig. 7), rainfall (Fig. 8) and relative humidity

(Fig. 9) as the input values. These predictions are close to the actual cases.

## CONCLUSION

Cross-correlation, along with Ljung-Box statistic, revealed a significant degree of relationship between the climate variables and dengue occurrence. With the climatic variables treated as an independent time series to and with dengue cases, it revealed that the maximum temperature, minimum temperature, relative

humidity, and rainfall has a strength of 0.4691, 0.6332, 0.5827, and 0.4895, respectively. Evidently, minimum temperature has the highest correlation with dengue cases.

The best fit SARIMA model was found to be  $(3,0,0),(0,0,1)_{12}$  with minimum temperature lag-3 as an external regressor. Dengue incidence correlates with maximum temperature in Central Luzon lag-4, relative humidity lag-1, rainfall lag-1.

Previous studies indicated correlation of dengue with minimum temperature in NCR lag-2 [15] and with average temperature lags 3-4 in Thailand [7]. In Taiwan, dengue is significantly correlated with maximum temperature lag 1-4, minimum temperature lag 1-3, relative humidity lag-1, and rainfall lag-2 [18].

Neural Network algorithms have a more powerful forecasting potential, yielding 4 RMSE values lower than the RMSE generated by the best fit SARIMA model.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA) for providing us with the data on the climatic variables and the Department of Health (DOH) and its regional offices for the data on the number of dengue cases.

## REFERENCES

- [1] WHO ASEAN Workshop on Priority Action For Dengue Prevention and Control Report 2011 July 12.
- [2] Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, Hunsperger E, Kroeger A, Margolis HS, Martinez E, Nathan MB, Pelegrino JL, Simmons C, Yoksan S, & Peeling RW. Dengue: a continuing global threat. *Nat. Rev. Microbiol.* **2010**; 8:S7-S16.
- [3] World Health Organization. *WPRO dengue situation updates*. Retrieved March 4, 2015 from [http://www.wpro.who.int/philippines/areas/communicable\\_diseases/dengue/continuation\\_dengue\\_area\\_page/en/](http://www.wpro.who.int/philippines/areas/communicable_diseases/dengue/continuation_dengue_area_page/en/)
- [4] Morin CW, Comrie AC, & Ernst K. Climate and Dengue Transmission: Evidence and Implications. *Environ. Health Perspect.* **2013**; 121:1264-1272.
- [5] Patz J, Martens W, Focks D, & Jetten T. Dengue fever epidemic potential as projected by general circulation models of global climate change. *Environ. Health Perspect.* **1998**; 106:147-153.
- [6] Kongnuy R & Pongsumpun P. Mathematical Modeling for Dengue Transmission with the Effect of Season. *Int. J. Math. Comput. Phys. Elec. Comp. Eng.* **2011**; 5:255-259.
- [7] Nakhapakorn K & Tripathi NK. 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.
- [8] Choudhury MZ, Banu S, & Islam MA. Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis. *Dengue Bulletin* **2008**; 32:29-37.
- [9] Burattini MN, Chen M, Chow A, Coutinho FA, Goh KT, Lopez LF, Ma S, & Massad E. Modelling the control strategies against dengue in Singapore. *Epidemiol. Infect.* **2007**; 136:309-319.
- [10] Wegbreit J. In: Arlinghaus SL & Drake WD (Eds.) The Possible Effects of Temperature and Precipitation on Dengue Morbidity in Trinidad and Tobago: A Retrospective Longitudinal Study. In *Population-Environment Dynamics: Issues and Policy* (University of Michigan, **1997**).
- [11] Corwin AL, Larasati RP, Bangs MJ, Wuryadi S, Arjoso S, Sukri N, Listyaningsih E, Hartati S, Namursa R, Anwar Z, Chandra S, Loho B, Ahmad H, Campbell JR, & Porter KR. Epidemic dengue transmission in southern Sumatra, Indonesia. *Trans. R. Soc. Trop. Med. Hyg.* **2001**; 95:257-265.
- [12] Luz PM, Mendes BVM, Codeco CT, Struchiner CJ, & Galvani AP. Time Series Analysis of Dengue Incidence in Rio de Janeiro, Brazil. *Am. J. Trop. Med. Hyg.* **2008**; 79:933-939.
- [13] 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**; 9:395-400.
- [14] Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, Girdary L, & Marrama L. Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors. *BMC Infect. Dis.* **2011**; 11:166-179.



## Time series analysis of dengue cases in Central Luzon

- [15] Dela Cruz AC, Lubrica JA, Punzalan BV, & Martin MC. Forecasting Dengue Incidence in the National Capital Region, Philippines: Using Time Series Analysis with Climate Variables as Predictors. *Acta Manilana* **2012**; 60:19–26.
- [16] Republic of the Philippines. *2015 Census and Housing Population*. (National Statistics Office, **2015**).
- [17] Manaysay FV. DOH declares outbreak in Sta. Rita, *Sun Star Pampanga*, **2014** September 19.
- [18] Wu PC, Guo HR, Lung SC, Lin CY, & Su HJ. Weather as an effective predictor for occurrence of dengue fever in Taiwan. *Acta Trop*. **2007**; 103:50–57.