Volume-6
Issue- I
June-2021

Rajarata University Journal

ISSN: 2362-0080
www.ruj.ac.lk/journals/

Category: Research Article

Meteorological Factors Affecting the Prediction of the Number of Dengue Cases Reported in Anuradhapura District

[1]De Silva SHMP & *[2]Nawarathna Lakshika S.

[1] *Department of Physical Sciences, Faculty of Applied Sciences, Rajarata University of Sri Lanka, Mihintale, Sri Lanka*

[2] *Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka*

ARTICLE DETAILS

Article History

Published Online: 30 June, 2021

Keywords

Dengue cases, Mean Absolute Percentage Error, Meteorological Factors, Trend Analysis

***Corresponding Author**

Email: *lakshikas@pdn.ac.lk*

ORCID:*https://orcid.org/0000-0001-9665-5471*

ABSTRACT

Dengue fever is a mosquito-borne viral infection rapidly spreading in the world by the vector *Aedes aegypti* mosquito. Day by day, new dengue cases are reported in Sri Lanka, and hence the risk of dengue exists throughout the country. Besides, the lowest amount of dengue cases in Sri Lanka was reported from the North Central province. Thus, this study focuses on the Anuradhapura district, the capital of the North Central Province. The main objectives of this study are to identify the weather factors affecting the number of dengue cases, investigate the variability and trends of reported dengue cases, identify the correlation between the number of dengue cases and weather factors reported in the Anuradhapura district. The weather data from 2010-2019 was collected from the Department of Meteorology-Sri Lanka, and the dengue cases data was collected from the Epidemiology Unit of the Ministry of Health. Bar charts, pie charts, line graphs and time series plots were used to visualize the data. The normality assumption of the data was checked using Anderson-Darling normality tests. Moreover, Pearson correlation tests were used to check the correlations between variables. The number of dengue cases and the rainfall were forecast for the upcoming years using a trend analysis method. Stepwise regression analysis was performed to identify the weather factors affecting the number of dengue cases and Akaike Information Criterion (AIC) values were used to select the best model. Further, Mean Absolute Percentage Error (MAPE) values were calculated to find the prediction accuracy of the proposed models.

1. Introduction

Dengue fever is a mosquito-borne viral infection called an arbovirus that rapidly spreads throughout the world through the vector Aedes mosquito. This fever mostly appears in tropical and subtropical regions globally and nearly 3.9 billion people live in these dengue endemic regions [1, 2]. Dengue fever was ranked as the most important mosquito-borne viral disease in the world in 2012. The largest dengue epidemic was reported in 2009, with 35,008 cases and 346 deaths, with an incidence rate of 170 per 100,000 populations. The first serologically confirmed dengue case was reported in 1962 and the first outbreak was reported in 1965 [3]. World Health Organization (WHO) estimates an annual total of 670,000 symptomatic cases in Sri Lanka. Epidemiology Unit of Sri Lanka said that more than 30,000 incidences were reported annually since 2012; however, the western part of the country is highly affected while the north and eastern part of the country is hardly affected. According to the Epidemiology Unit of Sri Lanka, the lowest dengue cases in Sri Lanka were reported from the North Central Province during the last ten years. Further, Sri Lanka is a tropical country with warm temperature and high humidity. From November to February, rain comes to the dry zones North & Eastern regions of Sri Lanka, while May to August brings plenty of rain to the Western and Southern regions. Mosquitos breeds in stagnant water; monsoon rains are often followed by two dengue summit periods; the main summit in June/July, while another one at the end of the year [4].

At present, the causes and affecting factors for dengue fever are unknown in Sri Lanka. Earlier studies revealed statistically significant associations between infectious disease and weather factors such as rainfall, temperature. Moreover, the effects of climate change on diseases such as Cholera, Malaria and plague have been recognized [5, 6]. Time series analysis is often used to study the relation and the trend of the weather factors and disease is most effective when data are collected over a long period. However, it is difficult to collect such meteorological and health daily data in developing countries [7, 8, 9]. Hence, most time series analyses use monthly or annual data.

Based on the research about a forecasting model for dengue incidence in the district of Gampaha, Sri Lanka, the researcher developed and validates three simple, accurate forecasting model using time series regression analysis according to monthly rainfall, rainy days, temperature, humidity, wind speed affected to the dengue incidences [10]. Proposed models were compared based on Akaike's information criterion (AIC), Bayesian information criterion (BIC) and residual analysis. From analysis, their selected model forecast correctly with mean absolute errors of 0.007 and 0.22 and root mean squared errors of 0.09 and 0.28 for the training and validation period [3]. However, no study has been performed to analyze the factors affecting the number of Dengue cases reported in the Anuradhapura District.

Therefore, this study is the pioneer study performed to find the number of dengue cases that can be reported and the behaviour of rainfall in the Anuradhapura district for the coming years and about usually how dengue cases and rainfall are reported monthly during the year. The study focused on identifying weather factors affecting the number of dengue cases, rainfall and whether the relation between them is negative or positive. The weather-based model and forecasting allow warning of impending dengue cases in advance and it could be used to manage limited public health resources effectively for patient management, vector scrutiny and involvement programs in the district. Further, annually forecast values and monthly variation pattern of rainfall will provide a knowledge base for better management of agriculture, irrigation and other water related activities in the Anuradhapura district.

This article is organized as follows. Section 2 describes the nature of the data set utilized for analysis, methods for predicting the number of dengue cases, rainfall and model selection and validation techniques used in this study. Section 3 outlines the statistical analysis and the key results derived. Section 4 concludes with a discussion. The model fitting and analysis is carried out by using R and Minitab statistical software.

2. Material and Methods

This study is based on the weather data in the Anuradhapura district collected from the Department of Meteorology, Sri Lanka. The data for 2010-2019 was considered in this study. The number of dengue cases reported in the Anuradhapura district was considered as the response variable where the data collected from the Epidemiology Unit of Ministry of Health and rainfall, maximum temperature, minimum temperature, average cloud amount, average wind speed, average sunshine hours, maximum humidity, minimum humidity are the independent variables which are considered in this study. Besides, the relationship between average rainfall in the Anuradhapura district and the weather factors was also examined. Table 1 describes the variables considered in this study.

Table 1: Description of variables, considered in this study

Variable	Variable name	Variable
X_1	Rainfall (mm)	Numerical
X_2	Maximum temperature(℃)	Numerical
X_3	Minimum temperature(℃)	Numerical
X_4	Average cloud	Numerical
X_5	Average wind	Numerical
X_6	Average sunshine hours	Numerical
X_7	Maximum humidity (%)	Numerical
X_8	Minimum humidity (%)	Numerical
X_9	Number of dengue cases	Numerical

All the variables were examined separately to identify their behaviour. The data were arranged in a tabular manner and each variable was represented by using bar charts, pie charts, and line graphs. Then the statistical relationship between each variable was obtained by using Pearson correlation analysis. The number of dengue cases and the rainfall were forecast for the upcoming years using the trend analysis method [11, 12]. The Anderson-Darling test was reformed to check the normality of each variable. Moreover, Stepwise regression analysis was performed to identify the weather factors affecting the number of dengue cases and Akaike Information Criterion (AIC) values were used to select the best model. When comparing two or more models, the model with the lowest AIC is selected as the best model. In addition, Multiple Linear Regression (MLR) analysis was used to identify the weather factors affecting the rainfall. Further, Mean Absolute Percentage Error (MAPE) values were measured to find the prediction accuracy of the proposed models. In regression, stepwise model selection with backward/forward direction includes the models in which the choice of explanatory (independent) variables is carried out by sequence procedure and AIC value is used to select the best model [13, 14]. A stepwise regression model is fitted to predict weather factors affecting the number of dengue cases reported in Anuradhapura district. When comparing two or more models, the model with the lowest AIC is selected as the best model.

Multiple linear regression is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of MLR is to model the linear relationship between the explanatory variables and the response variable [13]. A multiple linear regression (MLR) model was fitted to predict weather factors affecting for the average rainfall in the Anuradhapura district. MLR defined below basically specifies the relation of the response variable (*Y*) to function combination of explanatory variables (*X*) and unknown parameters (β).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon$$

where ε is the error term and β_0 is the intercept and β's are coefficients for selected variables. i.e. Y ≈ f (X, β). Further, the following hypothesis was performed to check the significance if the coefficient estimate.

$$H_o: \beta_i = 0 \text{ vs. } H_1: \beta_i \neq 0$$

H_0: The slope is equal zero. i.e., there is no significant linear relationship between the independent variable X and the dependent variable Y.

H_1: The slope is not equal to zero. i.e., there is a significant linear relationship between the independent variable X and the dependent variable Y, the slope will not equal zero.

Table 2: Pearson correlation analysis of variables with its corresponding p-value.

Variable	X1	X2	X3	X4	X5	X6	X7	X8
X2	-0.447							
	<0.001							
X3	-0.268	0.736						
	0.014	<0.001						
X4	0.339	-0.168	0.354					
	0.002	0.127	0.001					
X5	-0.228	0.404	0.681	0.343				
	0.037	<0.001	<0.001	0.001				
X6	-0.634	0.827	0.513	-0.42	0.363			
	<0.001	<0.001	<0.001	<0.001	0.001			
X7	0.663	-0.701	-0.697	-0.019	-0.664	-0.688		
	<0.001	<0.001	<0.001	0.866	<0.001	<0.001		
X8	0.694	-0.88	-0.535	0.333	-0.405	-0.885	0.84	
	<0.001	<0.001	<0.001	0.002	<0.001	<0.001	<0.001	
X9	-0.116	-0.167	-0.288	-0.283	-0.188	-0.071	0.111	0.073
	0.293	0.129	0.008	0.009	0.086	0.518	0.313	0.507

3. Results and Discussion

Time series plots of the monthly number of dengue cases and the variation of monthly average rainfall are shown in Figure 1. Most years, the highest number of cases were reported in July & December in the Anuradhapura district. Other than that, patients have been reported at least in January, February & December of some years. Besides, Rainfall is most abundant in the last quarter of the year in the Anuradhapura district. Most of these years have the highest rainfall in November except 2013 and 2016. The highest rainfall was reported in October 2010.

According to the Pearson correlation analysis with the corresponding p-value shown in Table 2, all the weather factors are weakly correlated with the number of dengue cases. When only maximum and minimum humidity increase, the number of dengue cases also increases, while other parameters have an inverse correlation. There is a strong positive correlation between rainfall with maximum humidity and minimum humidity, maximum temperature with minimum temperature and sunshine hours, minimum temperature with wind speed and sunshine hours, maximum humidity and minimum humidity. When the maximum and minimum humidity increase; maximum temperature, minimum temperature, sunshine hours, wind speed decrease.

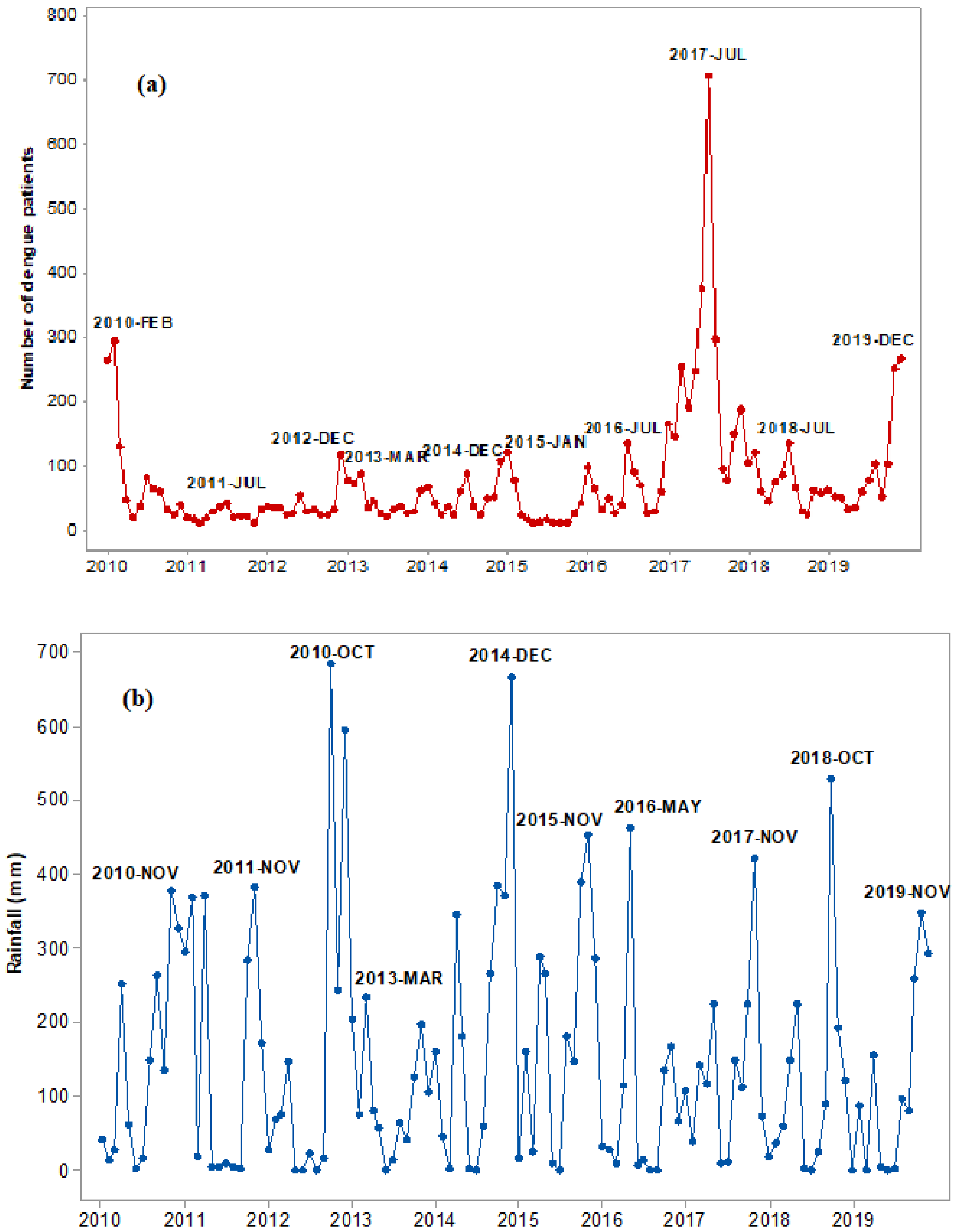

Figure 1: Time series plot of monthly reported (a) number of dengue cases and (b) rainfall from 2010-2019

Trend analysis plots of the average number of dengue cases and average rainfall are shown in Figure 2. No specific trend can be observed from 2010-2016. However, there is a declining trend in the annual average number of dengue cases. Also, dengue cases appear to be declining as rainfall decreases. The accuracy of the trend is 91.64%. In comparison, in 2017, 2018 and 2019, the average dengue cases were slightly higher than the predicted values.

An up and down trend from 2010-2016 in rainfall can be seen. The trend analysis plot shows a declining average annual rainfall for the upcoming years, with the accuracy of the trend being 73.62%. Besides, the years 2017, 2018 and 2019 average rainfall values approximate the predicted values and decrease.

The regression analysis was performed to find the best model for the number of dengue cases in the Anuradhapura district. As a result, the following model with minimum AIC value (AIC= -75.63) was selected as the best model.

$$\ln(\text{Number of dengue cases}) = 5.0128343 - 0.0011197 * X_1 - 0.0030648 * X_4^3 - 0.0137372 * X_6^2$$

The summary result of the model is given in Table 3. For this model, the p-value is 0.003935. Hence overall model and all model parameters are significant at a 5% significance level. Then the *ln* values of number of dengue cases in the Anuradhapura district is described by this model. Also 55.28% of the variation is explained by the above regression model and the obtained Adjusted R-squared is 53.11%. Then, the multiple linear regression analysis was performed to find the best model for rainfall in the Anuradhapura district. The fitted multiple linear regression model for rainfall is,

$$\text{Rainfall} = -4114.2952 + 118.1223 * X_2 - 94.6437 * X_3 + 0.614 * X_4^3 + 9.5796 * X_5 + 35.8176 * X_8$$

The Coefficient estimates of stepwise model and the corresponding p-values are given in Table 4.

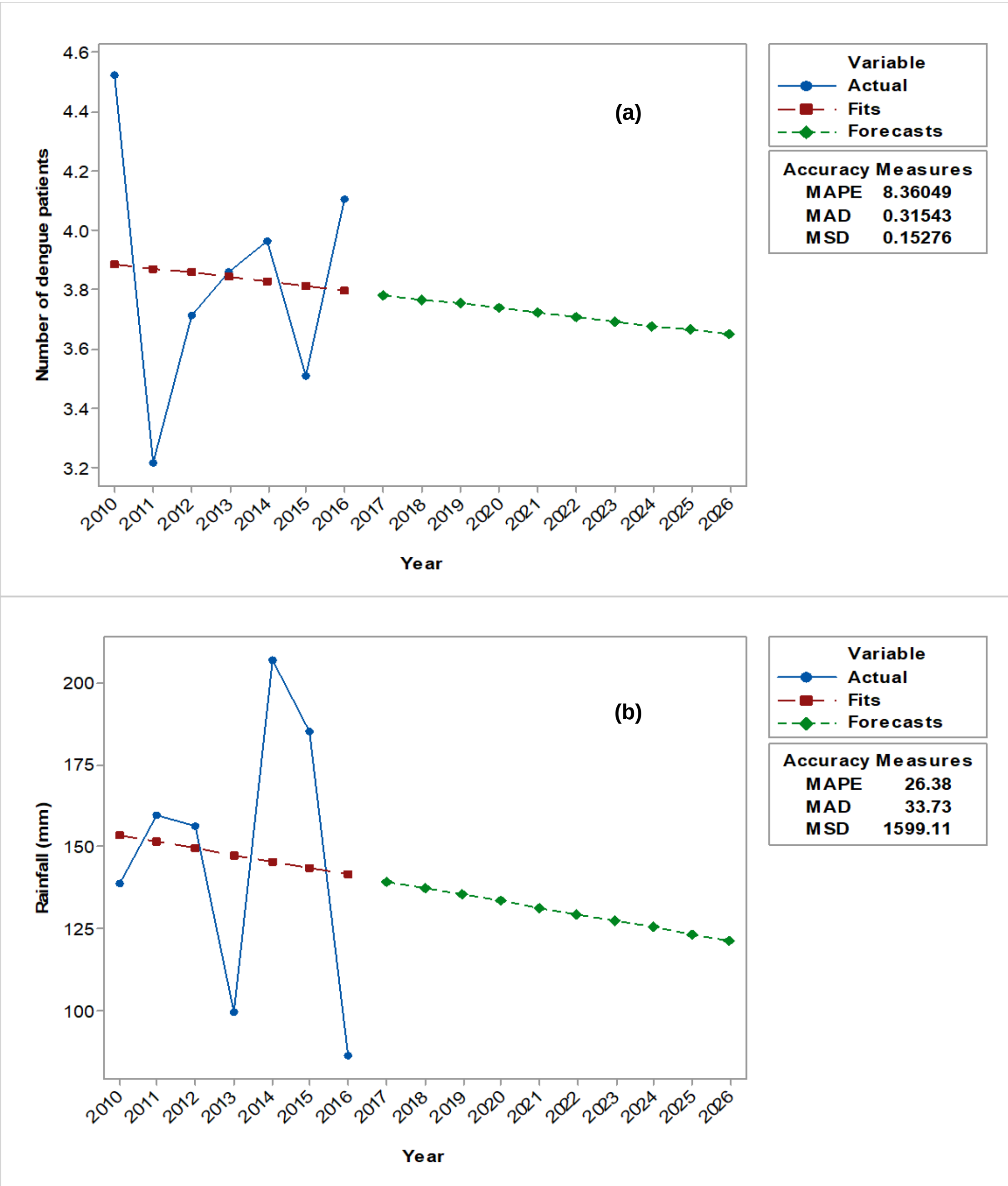

Figure 2: Trend analysis plot of (a) average number of dengue cases and (b) average rainfall.

Table 3: Coefficient estimates of stepwise model.

Coefficient	Estimate	Std. Error	t-value	p-value
Intercept	5.0128343	0.3679815	13.62	<0.0001
X_1	-0.0011197	0.0005333	-2.1	0.0389
X_4^3	-0.0030648	0.0011564	-2.65	0.0097

Moreover, Figure 2 shows the upcoming data for annual average rainfall and dengue cases forecasted for ten years using the trend analysis. The p-value of the MLR model is <0.0001. Hence overall model and all model parameters are significant at a 5% significance level. Therefore, the rainfall in the Anuradhapura district is well described by this model. Also, 83.67% of the variation is explained by the above regression model. Moreover, the adjusted R-squared is 81.6%.

Figure 3 shows the plot of actual & predicted values for both the MLR model and stepwise model. The observed values fairly go well with their predicted values through the identified regression model. Further, the fitted model has the MAPE = 0.51086/3*100 = 17.02872%. Hence, the prediction accuracy of the number of dengue cases in the Anuradhapura was 82.97127%. The observed values of annual average rainfall closely go with its predicted values through the identified regression model. The fitted model has the MAPE = 0.173466/3*100 = 5.782197%. Therefore, this model has the 94.2178% of rainfall prediction accuracy in the Anuradhapura district.

Table 4: Coefficient estimates of MLR model.

Coefficient	Estimate	Std.Error	T-value	p-value
Intercept	-4114.2952	630.0069	-6.531	<0.0001
X_2	118.1223	18.1547	6.506	<0.0001
X_3	-94.6437	22.3810	-4.229	<0.0001
X_4^3	0.6140	0.2359	2.603	0.0111
X_5	9.5796	4.0980	2.338	0.0220
X_8	35.8176	4.0349	8.877	<0.0001

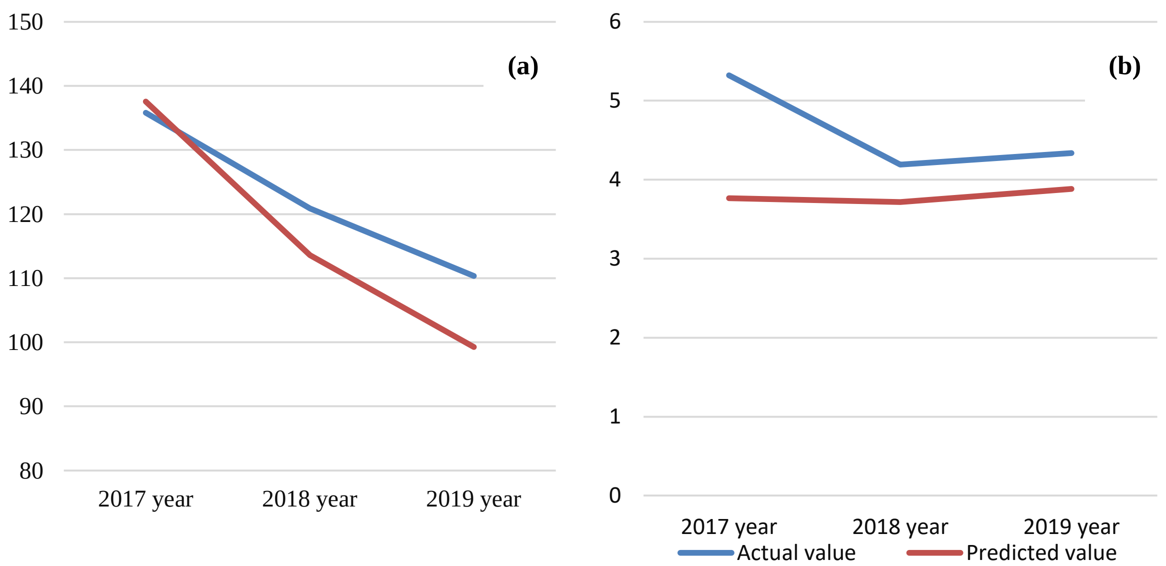

Figure 3: The plot of actual & predicted values for (a) MLR model and (b) stepwise mode

4. Conclusion

According to the results obtained, there was no pattern of dengue cases being reported in Anuradhapura. Typically, an average of 76 cases is reported monthly. It is found that all weather factors are weakly correlated with the number of dengue cases reported in the Anuradhapura district. The trend analysis reveals that there is not any specific trend in dengue cases. But with the 94.64% trend accuracy, there can be a decline in the number of dengue cases reported annually for upcoming years in the Anuradhapura district.

The average annual rainfall in the Anuradhapura district that spreads over the largest dry climatic area in Sri Lanka is approximately 140mm and varies between 0-680mm per month. The time series plot shows that the last quarter of the year, which is part of the Maha season, receives high rainfall. According to the trend analysis, there is an up and down trend in rainfall. With the 76.32% trend accuracy, a decline in the annual rainfall for upcoming years in the Anuradhapura district can be seen. However, it could also be a decline with ups and downs trend.

By stepwise regression analysis, a model for predicting the *ln* values of dengue cases reported in the Anuradhapura district was built using only three significant variables: rainfall, average cloud amount, and average sunshine hours at a 5% significance level. The best model was selected using a minimum AIC value of -75.63. The obtained model has an accuracy of 82.97%, and the variation of the dataset describes a 53.11% of adjusted R-squared value. This study belongs to the category of social or behavioural sciences related to humans. According to reviews of many researchers, low R-squared values can be expected with high accuracy in such studies.

Also, MLR analysis was performed to build a model for predicting the rainfall that can be occurred in the Anuradhapura district. Maximum and minimum temperature, average cloud amount, average wind speed, minimum humidity are the significant variables in the best model, selected at the 5% significance level. The obtained model has an accuracy of 94.22% and it describes almost all the variation of the dataset since it has 81.6% of the adjusted R-squared value. The results of data validation show that the model predicts well within its tolerance.

Recommendations

This study shows an excellent tendency to see a decline in the number of dengue cases reported in the Anuradhapura district in the future. Therefore, to improve the situation, the government should expand and enhance the existing dengue prevention programs. Further, it is essential to improve the tank system and introduce more water storage strategies to retain rainwater, minimize deforestation, and enact strict laws against it.

Limitations of the study

This research has been limited to the dengue cases and the weather factors reported in the Anuradhapura district, as these data were collected from the Anuradhapura weather station of the Department of Meteorology in Sri Lanka. Due to the limited meteorological data release, only eight weather factors are considered here, but other weather factors such as pressure, evaporation, soil temperature etc [7]. In addition to weather factors, there may be social or different kinds of factors that may play a role. Also, the Epidemiology Unit records the number of dengue cases in each district of Sri Lanka. In some cases, there may be some errors in the local data reporting in each district.

Acknowledgement

The authors are thankful to Mr. Sarath Anurasiri, Meteorologist of the Department of Meteorology, Sri Lanka, to collect the data.

References

1. Bhatt, S., Gething, P.W., Brady O.J., Messina J.P, Farlow A.W., Moyes C.L. *The global distribution and burden of dengue*. Nature 2013. 496(7446): 504-507. https://doi.org/10.1038/nature12060
2. Brad O.J., Gething P.W., Bhatt S., Messina J.P., Brownstein J.S., Hoen A.G. *Refining the global spatial limits of dengue virus transmission by evidence-based consensus.* PLoS Negl Trop Dis 2012. 6(8): e1760. https://doi.org/10.1371/journal.pntd.0001760
3. Withanage GP, Viswakula SD, Silva YIN, Hapugoda MD. *A forecasting model for dengue incidence in the district of Gampaha, Sri Lanka.* Parasites & Vectors 2018. 11(1), 262. https://doi.org/10.1186/s13071-018-2828-2
4. Jahanbakhsh S, Tadioni M, Salmanpoo R, Jahanbakhsh A, 2009. *The Relationship between air temperature by Heart Attack in the City of Ahar*. J Phys Geogr. 2009. 2(5): 29-37.
5. Egbendewe-Mondzozo A., Musumba M, McCarl BA, Wu X. *Climate change and vector-borne*

diseases: an economic impact analysis of malaria in Africa, Int J Environ Res Public Health 2011. 8(3):913-930. https://doi.org/10.3390/ijerph8030913

6. De Magny GC, Thiaw W, Kumar V, Manga NM, Diop B.M. *Cholera outbreak in Senegal in 2005: was climate a factor?* PLoS One 2012. 7(8): e44577. https://dx.doi.org/10.1371%2Fjournal.pone.0044577

7. Ehelepola NDB, Ariyaratne K, Jayaratne A. *The association between local meteorological changes and exacerbation of acute wheezing in Kandy, Sri Lanka*, Global Health Action, 2018.11(1), 1482998. https://doi.org/10.1080/16549716.2018.1482998

8. Ehelepola NDB, Ariyaratne K & Dissanayake WP. The correlation between local weather and leptospirosis incidence in Kandy district, Sri Lanka from 2006 to 2015, Global Health Action, 2019. 12(1), 1553283. https://doi.org/10.1080/16549716.2018.1553283

9. Mohammadi N, Kamali K, Abedinia Y, Ahadnejad M, Azari M, Medical Meteorology: The Relationship between Meteorological Parameters (Humidity, Rainfall, Wind, and Temperature) and Brucellosis in Zanjan Province. J Hum Environ Health Promot. 2016; 1(3):149-158. http://dx.doi.org/10.29252/jhehp.1.3.149

10. Honda Y, Ono M, Issues in health risk assessment of current and future heat extremes. *Glob Health Action*. 2009. 2(1), 2043. https://doi.org/10.3402/gha.v2i0.2043

11. Panda A, Sahu N. *Trend analysis of seasonal rainfall and temperature pattern in Kalahandi, Bolangir and Koraput districts of Odisha, India,* Atmos Sci Lett. 2019; 20(10): e932. https://doi.org/10.1002/asl.932

12. Longobardi A, Villani P, Trend analysis of *annual and seasonal rainfall time series in the Mediterranean area*. International Journal of Climatology, 2010. 30(10): 1538-1546. https://doi.org/10.1002/joc.2001

13. James G, Witten D, Hastie T, Tibshirani R, *An introduction to statistical learning with Applications in R*, 2013. New York: Springer-Verlag. https://doi.org/10.1007/978-1-4614-7138-7

14. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York, NY: Springer New York; 2009. https://link.springer.com/book/10.1007/978-0-387-84858-7

