JOURNAL OF COMPUTATIONAL INNOVATION AND ANALYTICS
http://e-journal.uum.edu.my/index.php/jcia

How to cite this article:
Marigmen, J. L. D. C., & Addawe, R. C. (2022). Forecasting and on the influence of climatic factors on rising dengue incidence in Baguio City, Philippines. *Journal of Computational Innovation and Analytics, 1*(1), 43-68. https://doi.org/10.32890/jcia2022.1.1.3

FORECASTING AND ON THE INFLUENCE OF CLIMATIC FACTORS ON RISING DENGUE INCIDENCE IN BAGUIO CITY, PHILIPPINES

[1]Joseph Ludwin DC. Marigmen & [2]Rizavel C. Addawe
Department of Mathematics and Computer Science
College of Science, University of the Philippines Baguio,
Baguio City, Benguet 2600, Philippines

[1]Corresponding Author: jdmarigmen@up.edu.ph

Received: 20/9/2021 Revised: 24/11/2021 Accepted: 1/12/2021 Published: 27/1/2022

ABSTRACT

Dengue fever or dengue has been a concern for individuals living in Baguio City, Philippines. Every year, incidence counts rise during rainy seasons experienced from June to October. Several researches suggest that meteorological factors have great influence on the life, growth, and reproduction of dengue-carrying mosquitoes, resulting in higher dengue incidence in the area. With the continuing rise of dengue incidence in Baguio City, we aim to forecast dengue incidence in the area for the year 2019, starting from January until the end of the outbreak period in the area. Here, we use the *projections* package of R as it involves the serial interval distribution and R_t value of dengue incidence. We also aim to use multiple linear regression analysis to determine if meteorological factors have significant effects in the rise of dengue incidence in the city. With the inclusion of time-varying reproduction number and serial interval distribution of dengue, we projected that dengue incidence may reach up to 101 cases by June 16, 2021, and without further actions, cases may rise up to 529 cases by August 29, 2021. Based on the average two-year period, such increase is attributed to relative humidity and average temperature as these are the most significant factors associated with dengue incidence based on the MLR analysis. The highest and mean maximum temperatures remain as key meteorological variables that influence dengue incidence in the city. As early as possible, local officials are recommended to uphold proper safety and health procedures in preventing the spread of dengue in Baguio City.

Keywords: Dengue Fever; Climatic Factors; Forecasting; Multiple Linear Regression; Baguio City.

INTRODUCTION

Dengue fever (or dengue) is a mosquito-borne viral infection commonly spread by female *Aedes aegypti* mosquitoes in tropical and subtropical regions, especially urban environments (Cogan, 2021). In Baguio City, where rainy seasons are felt on June to October and dry seasons on the rest of the months in a year, dengue is a common illness among locals, where reported spikes in local dengue cases in the area from January to May 2016 compared to other months from 2010 to 2018 (Polonio, 2016). As a result, several protective measures and clean-up drives were conducted to destroy potential breeding sites of dengue-carrying mosquitoes and larvae.

Several notable researches were conducted to determine the effects of meteorological factors – humidity, precipitation, and temperature – to the growth of dengue incidence in a certain area. According to the study at Jakarta in Indonesia, the strong correlation between the population density of *Aedes aegypti* mosquitoes, temperature, and relative humidity indicated that the weather factors influence the growth and population of mosquitoes in the area (Sintorini, 2018). In another study in Thailand, the combination of humidity and temperature has become beneficial in the development, survival, length of the extrinsic incubation period, and competence of dengue-carrying mosquitoes (Campbell et al., 2013).

Similar studies were also conducted within the Philippines. In Metro Manila, a simple linear regression (SLR) analysis on the monthly climatic factors (temperature and rainfall) and dengue incidence data from 1996 to 2005 concluded that the dengue incidence is likely affected by changes in the amount of rainfall affecting mosquito populations as breeding grounds increase (Sia Su, 2008). Another correlation on the effect of climatic factors on laboratory-confirmed dengue and leptospirosis infections in the Philippines suggests that dengue fever and leptospirosis correlate with rainfall, relative humidity, and temperature (Sumi et al., 2016).

In another case study at Iligan City in Mindanao, researchers used multiple linear regression (MLR) analysis, Poisson regression, and random forest in developing a best-fitting model for dengue incidence in the area. Considering the monthly climatic factors - relative rainfall (x_1), maximum temperature (x_2), and average (relative humidity (x_3)- together with the monthly time period from (x_4)2008 to 2017 (Olmoguez et al., 2019), results show that the MLR model, having 18% accuracy percentage and 67.14% error result, has the form

$$\text{dengue cases} = 0.10x_1 - 9.83x_2 - 6.20x_3 + 0.72x_4 + 815.98 \quad (1)$$

However, they further conclude that the Random Forest performed better with a 73% accuracy percentage and 33.58% error result.

Several researches also involve forecasting using regression methods. In Pakistan, Sabir et al. (2018) applied SLR analysis to forecast the possible cumulative incidence of dengue in 2017 and 2018, utilising the yearly period as the independent variable. On the other hand, a case study in China by Guo et al. (2017) applied multiple regression algorithms to develop a dengue forecast model in the area.

In Baguio City, several other forecasting methods were also employed considering local dengue incidence data. Addawe et al. (2016) used Differential Evolution - Simulated Annealing (DESA) algorithm in acquiring the best-fitting forecasting models for young and adult populations. Another study by Magsakay et. al (2017) applied winsorization, square root, and logarithmic transformations to treat the outliers on age groups (ages 0 to 8 years old and 45 years old and above) who are not eligible for the dengue vaccine. Through the Univariate Box-Jenkins (UBJ-ARIMA) time series model, they acquired the best-fitting forecasting models for each age group.

Another notable characteristic to consider is the time-dependent reproduction number R_t.Compared to the basic reproduction number R_0, R_t incorporates the time-dependent variations in the transmission of diseases to the secondary cases from the corresponding primary case (Nishiura & Chowell, 2009). In the work of Nouvellet et al. (2017), they included the effect of R_t and the serial interval of some pathogens as factors to acquire an accurate forecast. As defined in Du et al. (2020), a serial interval is the time interval between the primary infected patient and its secondary infected patient. The corresponding probability density function is called the serial interval distribution w. With the influence of temperature, the serial interval of dengue is a combination of intrinsic incubation period (IIP), human-to-mosquito transmission period (HMTP), extrinsic incubation period (EIP), and mosquito-to-human transmission period (MHTP) (Siraj et al., 2017). To differentiate, IIP is the time difference between the patient's date of infection and the date of symptom onset; HMTP is the time between the conclusion of IIP and the date when the susceptible mosquito becomes infected; EIP is the time from ingesting the virus by the susceptible mosquito until it becomes infected; and MHTP is the time between the infected mosquito transfers the virus to a new host (Cogan, 2021; Siraj et al., 2017). In the study of Aldstadt et al. (2012) in Thailand, they concluded that the serial interval for dengue infection is most likely 15-17 days with a significant excess risk of illness that persists for 32-34 days.

In our previous studies, we conducted SLR analysis in determining the climatological factors affecting the dengue incidence in Baguio City within the two – year average period (Marigmen et al., 2021). Results showed that humidity is the main factor affecting the dengue incidence, followed by precipitation. Despite relatively high adjusted R-squared values, their difference from a perfect adjusted R-squared value (adjusted R-squared value is 1) indicates that there are other possible factors that affect the growth of dengue incidence in the area.

In this paper, we extend our previous study by conducting an MLR analysis on the dengue incidence from 2011 to 2018 to determine the said possible variables.

With the increase of dengue cases in the area, we also aim to conduct a dengue forecast using the *projections* package in R, a statistical software. The package requires the dengue serial interval distribution and computed R_t value. Using the computed R_t values and the serial interval distribution, we then forecast the dengue incidence in 2019, starting from January until the corresponding outbreak period. Further discussions on the package and methods are documented as part of the Methodology.

On the other hand, some limitations are discussed in this paper. We only considered the available monthly climatic factors from 2011 to 2018 in our MLR analysis. As for the forecast, due to the lack of published documents on the serial interval of dengue in the Philippines, we incorporated the findings of Aldstadt et al. (2012) in our analysis. In this paper, we are only using one forecasting method as part of our study.

Further discussions on the MLR analysis and forecasting are provided in the Theoretical Framework.

THEORETICAL FRAMEWORK

In this section, we discuss the theoretical framework behind forecasting and MLR analysis.

Given the daily incidence I, w, and Rt, we apply the renewal equation

$$I_t \sim Pois\left(R_t \sum_{s=0}^{t} I_{t-s}\omega_s\right) \tag{2}$$

in generating a branching process model and use it in our forecast (Nouvellet et al., 2017). In determining its accuracy to our available data, the renewal equation is applied to available historical data, where we estimate the trend of dengue incidence from previous years. Error analysis is also applied to determine the accuracy of the calculated estimates from the said data. Here, we apply the Root Mean Square Error (RMSE) to check the accuracy of the estimates. We emphasize here that the RMSE values must be low enough for the model to fit with the data. We also apply the Shapiro-Wilk W Test to check the normality of residuals. The residual is the difference between the estimated value from the equation and the actual value from the data.

To elaborate, the Shapiro-Wilk W Test is defined as the ratio of two estimates of the variance of a normal distribution based on a random sample of n observations (Royston, 1995; Gonzalez – Estrada & Cosmes, 2019). The residuals are normally distributed if the resulting p-value is greater than 0.05 (p-value > 0.05). Otherwise, the residuals are not normal. To attain the normality of residuals, we use the corresponding W statistic. Given the residuals i, the W statistic is defined as

$$W = \frac{\left(\sum_{i=1}^{24} a_i \hat{\epsilon}_i\right)^2}{\sum_{i=1}^{24} (\hat{\epsilon}_i - \overline{\epsilon})^2} \tag{3}$$

where $\overline{\epsilon}$ is the sample mean of the residuals, $\boldsymbol{a}_i = (a_1, a_2, \dots, a_n) = \frac{m^T v^{-1}}{(m^T v^{-1} v^{-1} m)^{1/2}}$ and $\boldsymbol{m} = (m_1, m_2, \dots, m_n)^T$ are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution, and v is the covariance matrix of the order statistics (Razali & Yap, 2011). The W statistic value lies between zero and one, and the larger the value indicates that the residuals are normally distributed.

Once the conditions for RMSE and the Shapiro-Wilk W Test are satisfied, we proceed with dengue forecasting in Baguio City for 2019, starting from January until the end of the outbreak period.

On the other hand, MLR is an extension of SLR where we consider more than one independent variable x_i and has the form

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n \tag{4}$$

where a_i's are the coefficients that need to determine and y is the dependent variable (Pearson, 2018).

As part of constructing best-fitting MLR models, we first conduct correlation analysis to determine the strength of relationship between two variables. In computing the correlation coefficient, we apply Pearson's Product Moment Correlation Coefficient r (Walpole et al., 2012). Here, we consider r to be strong if its absolute value is at least 0.70 (Schober et al., 2018). We also use the correlation results to check for collinearity or having high correlation between two independent variables x_i's to acquire a better fitting MLR model (Chatterjee & Simonoff, 2013).

In constructing an MLR model, we consider the following assumptions as discussed in Kleinbaum et al. (1988): (a) Existence of the MLR model such that for a combination of values of the independent variables x_i's, the dependent variable y is a random variable with a probability distribution having a finite mean and variance; (b) The y observations must be statistically independent of one another; (c) The mean value of y for each specific combination of x_i is a linear function of x_i; that is,

$$\mu_{y|x_1,x_2,\dots,x_k} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{5}$$

or

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{6}$$

where ϵ is the error component between the observed value y and the theoretical value $\mu_{y|x_1,x_2,\dots,x_k}$. Here, ϵ must be normally distributed with a zero mean and variance σ^2. For an acceptable estimate of ϵ, we apply the residual where

$$\hat{\epsilon} = y - \left(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k\right) \tag{7}$$

and $\hat{\beta}_i$'s are estimates of β_i's; (d) Homoscedasticity or the variance of y must be the same for any fixed combination of x_i's, that is,

$$\sigma^2_{y|x_1,x_2,\dots,x_k} = Var(y|x_1, x_2, \dots, x_k) \equiv \sigma^2; \tag{8}$$

(e) The variable y is normally distributed for any fixed combination of x_i's.

To elaborate, we use the Least Square Method in solving $\hat{\beta}_i$'s on each model. For the normality of the residual $\hat{\epsilon}$, we apply the Shapiro-Wilk *W* Test. While for the homoscedasticity of the model, we apply White's Test. White's Test compares the estimated variances of regression coefficients under homoscedasticity with the ones under heteroscedasticity (Jeong & Lee, 1999). Similar to the Shapiro-Wilk *W* Test, the MLR model is homoscedastic if its *p*-value is greater than 0.05. Otherwise, the model is heteroscedastic.

In addition to the aforementioned conditions, we also conduct additional error analysis on every MLR model. Here, we check the residual standard error, the multiple and adjusted *R*-squared, the *F* statistic and its *p*-value, and the *p*-value of each coefficient. Note that the coefficient is significant in the model if its corresponding *p*-value is less than 0.05. Since the multiple *R*-squared value ignores the issue of overfitting, we instead consider the adjusted *R*-squared value as it considers the issue of overfitting in the MLR model (Pearson, 2018).

The details and procedures for the application of these methods will be discussed in the methodology.

METHODOLOGY

For this paper, we use the daily dengue incidence data and the monthly climatic factors from 2011 - 2018 acquired from Baguio City Health Service Office and the Regional Department of Science and Technology - Philippine Atmospheric Geophysical and Astronomical Services Administration (DOST-PAGASA), respectively. We use the daily incidence data in conducting our forecast, whereas we use the monthly incidence data and the monthly climatic factors in conducting MLR analysis. For easier computation, we utilize various packages and functions from the statistical software R (Verzani, 2014).

Given the daily dengue incidence data from 2011 - 2018, we first determine the outbreak periods in Baguio City. The months having the highest monthly tallies of daily incidence in each year serve as the outbreak periods in the city.

Next, we determine the R_t values for each month until the outbreak period of each year. In computing R_t, we incorporate the *EpiEstim* package, a tool used to quantify transmissibility throughout an epidemic from the analysis of time series of incidence (Cori et al., 2013; Cori, 2020). From the package, we use the *estimate_R* function to calculate the R_t given the time series of incidence and the serial interval distribution. As mentioned earlier, we use the mean and median values of 16 days and 1.64 days, respectively (Siraj et al., 2017) for our serial interval distribution.

Once the outbreak periods and R_t are determined, we proceed with model fitting using the renewal equation. For convenience, we use the *projections* package in R (Jombart & Nouvellet, 2021). Within the package, we incorporate the *project* function to acquire the estimated model. In checking the model's normality of residuals, we apply the *shapiro.test* function to calculate the W statistic and the resulting

p-value. Once the model fits well with the data, we then proceed with forecasting dengue incidence in Baguio City.

In forecasting February to September 2019, the R_t values will be taken from averaging the eight values representing each month. In some cases, we will also treat the outlying values by replacing them with the calculated average of eight R_t values for that month. After conducting the forecast, we proceed with MLR analysis to determine the main factors that influence the growth of dengue incidence in Baguio City.

Aside from time in months, we consider three monthly climatic factors, namely relative humidity (RelHum) (unitless variable written in decimal form), precipitation (Prec) (measured in mm), and temperature (measured in ^{o}C), as the independent variables for our MLR model. We also emphasize that the temperature data has seven types, namely total maximum (TotalMaxTemp), mean maximum (MeanMaxTemp), total minimum (TotalMinTemp), mean minimum (MeanMinTemp), average (AveTemp), highest (HighTemp), and lowest (LowTemp). Note that the average temperature is computed as the average between the mean maximum and the mean minimum temperatures. Furthermore, each variable has ninety-six data points and are independent with one another. To avoid the issue of generating MLR models consisting mainly of different monthly temperature types, we restrict our MLR model to have one temperature type as the independent variable.

For an efficient statistical analysis, we utilize the *lm* function and the *summary* function to calculate the $\hat{\beta}_i$'s and the residuals of each MLR model, and to calculate the necessary parts of our error analysis. The normality of residuals will be checked through the use of *shapiro.test* function, and the homoscedasticity will be determined through the use of the *bptest* function from the *lmtest* package in calculating the resulting p-value (Zeileis & Hothorn, 2002).

RESULTS AND DISCUSSION

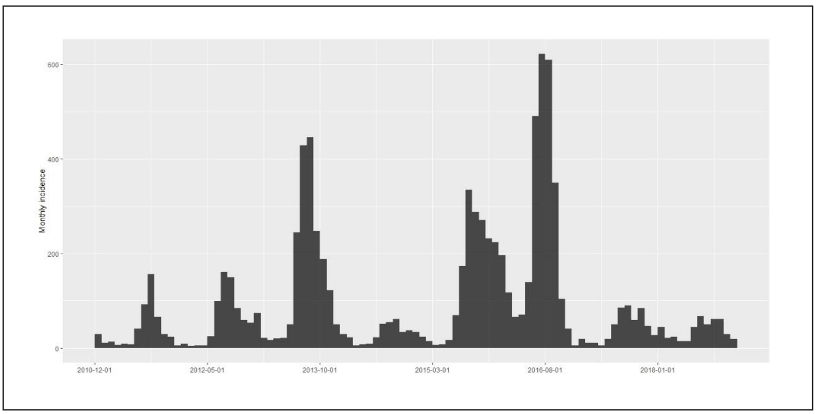
In this section, we discuss the results acquired from the model fitting and forecast using the branching process model and the MLR analysis. We also include an Exploratory Data Analysis (EDA) in MLR to explain the distribution and possible association of variables in our data frame. The accompanying results from EDA will be used in constructing our MLR model.

Forecasting

Figure 1 shows the monthly dengue incidence in Baguio City from 2011 – 2018. It can be observed that most of the monthly cases are recorded highest from July to September every year. Thus, we consider the third quarter of the year – the months of July, August, and September – as the dengue outbreak period in Baguio City.

Figure 1

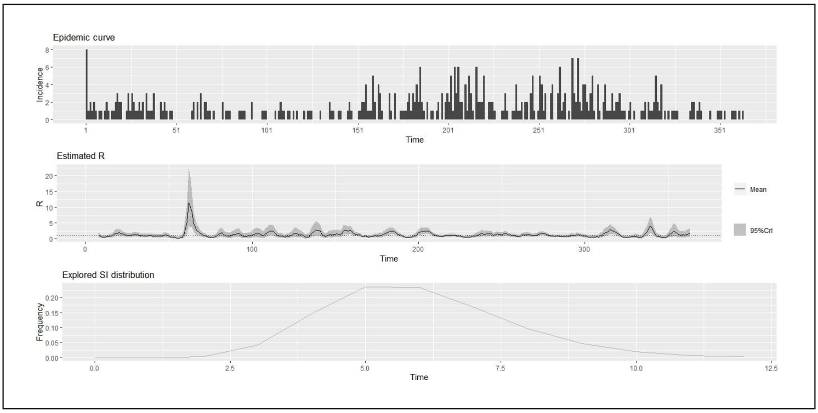
Monthly total dengue incidence in Baguio City from 2011 – 2018



Next, we calculate the latest R_t value for 2018. Figure 2 shows the result. It can be observed that the mean R_t value for the last week of 2018 is 1.71 having a 95% confidence interval of [0.63,3.32] and a standard deviation (SD) of 0.70. The mean R_t value implies that dengue incidence may still increase by almost 1.71 times in the succeeding weeks.

Figure 2

Epidemic curve, estimated R_t, and explored SI distribution of dengue incidence for 2018



Using the *projections* package in estimating the monthly dengue incidence for 2011 to 2018, Table 1 summarizes the results. All months, except for June and July, show high accuracy such that their absolute mean residuals range from 0–3 cases. On the other hand, the average mean residual for June shows that the estimated incidence is 9 cases larger compared to actual data. In addition, the average mean residual for July shows that the estimated incidence is around 16 cases larger compared to actual data. Meanwhile, normality p – values of February, April, June, July, August, and September show that the mean residual of the majority of the months are normally distributed. In addition, RMSE values are mostly low, indicating that the estimation on the dengue incidence in 2019 is accurate.

Table 1

Error analysis on the model estimation using projections package for January to September from 2011 to 2018

Year		January Estimate	February Estimate	March Estimate	April Estimate	May Estimate	June Estimate	July Estimate	August Estimate	September Estimate
2011	Mean R_t from previous month	0.71	3.48	1.58	2.71	0.53	3.40	1.46	1.40	0.43
	Mean Residual	0.15	1.49	0.67	1.14	-0.20	0.43	0.08	0.34	-0.76
	W Statistic	0.88	0.94	0.98	0.97	0.64	0.97	0.97	0.93	0.82
	Normality *p*-value	2.75E-03	0.10	0.76	0.57	1.96E-07	0.63	0.62	0.04	1.92E-04
	Type of normality	not normal	normal	normal	normal	not normal	normal	normal	not normal	not normal
	RMSE	0.80	2.16	0.85	1.80	0.55	1.62	1.76	3.38	1.98
2012	Mean R_t from previous month	0.63	0.63	0.66	11.03	6.02	2.07	3.75	1.19	0.94
	Mean Residual	-0.17	-0.08	-0.16	5.35	4.58	-0.24	34.26	1.84	1.44
	W Statistic	0.69	0.54	0.45	0.66	0.74	0.93	0.84	0.98	0.95
	Normality *p*-value	7.51E-07	2.09E-08	1.20E-09	4.18E-07	4.01E-06	0.06	3.42E-04	0.88	0.23
	Type of normality	not normal	not normal	not normal	not normal	not normal	normal	not normal	normal	normal
	RMSE	0.60	0.36	0.40	9.55	8.20	1.89	42.89	2.91	2.52
2013	Mean R_t from previous month	0.56	1.35	1.59	1.75	1.73	6.55	3.76	0.96	0.80
	Mean Residual	0.19	0.38	0.45	0.69	0.50	44.61	83.36	-2.99	1.96
	W Statistic	0.95	0.97	0.91	0.96	0.87	0.79	0.87	0.98	0.99
	Normality *p*-value	0.14	0.59	0.01	0.32	1.10E-03	5.13E-05	1.49E-03	0.84	0.98
	Type of normality	normal	normal	not normal	normal	not normal	not normal	not normal	normal	normal
	RMSE	0.75	0.87	1.18	1.16	1.94	68.16	108.85	5.64	3.89

(continued)

Year		January Estimate	February Estimate	March Estimate	April Estimate	May Estimate	June Estimate	July Estimate	August Estimate	September Estimate
2014	Mean R_t from previous month	1.49	1.03	0.35	1.94	1.32	1.46	2.29	1.34	0.70
	Mean Residual	1.79	-0.07	-0.04	0.48	0.37	-0.32	2.00	1.01	-1.17
	W Statistic	0.92	0.85	0.70	0.97	0.88	0.91	0.96	0.95	0.86
	Normality *p*-value	0.03	1.15E-03	1.11E-06	0.59	1.83E-03	0.01	0.26	0.20	9.73E-04
	Type of normality	not normal	not normal	not normal	normal	not normal	not normal	normal	normal	not normal
	RMSE	2.20	1.12	0.40	0.69	0.77	1.06	3.30	1.91	2.43
2015	Mean R_t from previous month	0.39	0.42	0.81	1.32	0.88	5.00	1.52	1.49	0.61
	Mean Residual	-0.58	-0.41	0.01	0.07	-0.39	12.61	0.24	-1.58	-4.95
	W Statistic	0.85	0.76	0.62	0.73	0.80	0.92	0.93	0.97	0.96
	Normality *p*-value	4.33E-04	1.98E-05	8.49E-08	4.66E-06	6.20E-05	0.03	0.04	0.61	0.28
	Type of normality	not normal	not normal	not normal	not normal	not normal	not normal	not normal	normal	normal
	RMSE	1.08	0.85	0.55	0.65	0.96	16.26	4.75	6.29	6.38
2016	Mean R_t from previous month	0.15	0.59	1.03	0.47	1.59	2.18	1.23	1.26	0.67
	Mean Residual	-5.98	-1.50	2.29	-1.65	1.49	11.39	6.53	9.58	-3.06
	W Statistic	0.96	0.96	0.94	0.93	0.96	0.98	0.96	0.95	0.95
	Normality *p*-value	0.28	0.30	0.10	0.05	0.34	0.55	0.25	0.20	0.14
	Type of normality	normal	normal	normal	normal	normal	normal	normal	normal	normal
	RMSE	6.78	2.72	2.79	2.61	2.83	14.65	11.82	15.58	5.72
2017	Mean R_t from previous month	0.57	0.20	0.95	3.31	2.93	2.35	1.38	0.52	0.87
	Mean Residual	-0.54	-0.33	-0.04	2.15	13.07	1.54	0.43	-2.22	0.48
	W Statistic	0.66	0.65	0.76	0.88	0.66	0.97	0.97	0.96	0.95
	Normality *p*-value	2.72E-07	5.31E-07	9.50E-06	3.43E-03	2.73E-07	0.46	0.60	0.36	0.19
	Type of normality	not normal	not normal	not normal	not normal	not normal	normal	normal	normal	normal
	RMSE	1.22	0.80	0.67	2.79	23.82	2.45	2.37	3.07	1.53

(continued)

Year		January	February	March	April	May	June	July	August	September
		Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
2018	Mean R_t from previous month	1.82	2.07	0.13	0.78	1.60	2.64	0.66	3.66	0.90
	Mean Residual	1.18	3.55	-0.74	-0.12	0.84	1.51	-1.55	20.04	-1.14
	W Statistic	0.95	0.95	0.84	0.86	0.98	0.95	0.92	0.87	0.93
	Normality p-value	0.15	0.19	2.81E-04	1.05E-03	0.71	0.16	0.02	1.72E-03	0.05
	Type of normality	normal	normal	not normal	not normal	normal	normal	not normal	not normal	normal
	RMSE	1.74	4.24	1.11	0.67	1.14	2.58	2.41	25.17	2.27

We incorporate the residual values as part of our 2019 forecasting. Table 2 tabulates the forecasting results for January to September 2019.

Table 2

Forecasted dengue incidence in Baguio City for January – September 2019.

Date	January	February	March	April	May	June	July	August	September
1	1	3	5	9	13	48	127	304	317
2	2	4	5	8	15	52	147	304	302
3	2	4	5	8	17	57	169	309	289
4	2	5	5	8	19	62	193	324	285
5	2	5	5	7	20	67	216	346	289
6	1	4	5	7	21	69	234	372	298
7	1	4	4	7	20	72	251	399	313
8	1	4	4	7	19	72	262	422	329
9	1	4	4	7	19	71	269	441	343
10	2	5	4	7	19	71	275	459	359
11	2	5	4	7	19	70	278	470	371
12	2	5	5	7	19	70	281	479	379
13	2	6	5	7	19	72	285	483	383
14	2	6	5	7	20	76	287	479	377
15	2	6	5	8	21	86	283	461	356
16	2	6	5	11	22	101	272	430	321
17	3	5	5	13	24	120	262	396	281
18	4	5	5	15	26	139	266	374	250
19	4	5	5	15	28	153	281	367	229
20	4	5	4	15	31	166	306	375	220
21	4	5	4	15	32	176	335	394	220
22	3	5	4	14	33	184	363	418	225
23	3	5	4	14	33	189	387	444	234
24	3	5	4	14	32	191	405	469	244
25	4	5	4	13	32	192	419	490	255
26	4	6	4	13	31	192	428	506	263
27	4	6	4	13	31	192	435	521	271
28	5	6	4	14	31	195	439	527	274
29	5		4	15	32	204	443	529	275
30	5		4	16	34	220	442	519	267
31	6		4		35		436	503	

Using the calculated R_t values and the mean and standard deviation of the serial interval distribution, it is projected that dengue cases may increase up to 529 incidences in a day, as reflected from August

29, 2021. It is also observed that dengue incidence is projected to dramatically increase starting from mid-June 2019 as dengue-carrying mosquitoes may start to infect at least 100 individuals in a day. Such values imply that a possible dengue outbreak may resurface in the city.

In a report by Agoot (2019), dengue cases in Baguio City show a downward trend as there are 210 reported cases from January 1 – July 13, 2019 compared to the same period last 2018. This resulted from the continued efforts of the locals in preventing the spread of the disease within their locality, and is further intensified with the cooperation of the local government in promoting and preventing the spread of dengue.

Once new data is available, the forecasting will be updated based from the newly – calculated R_t values.

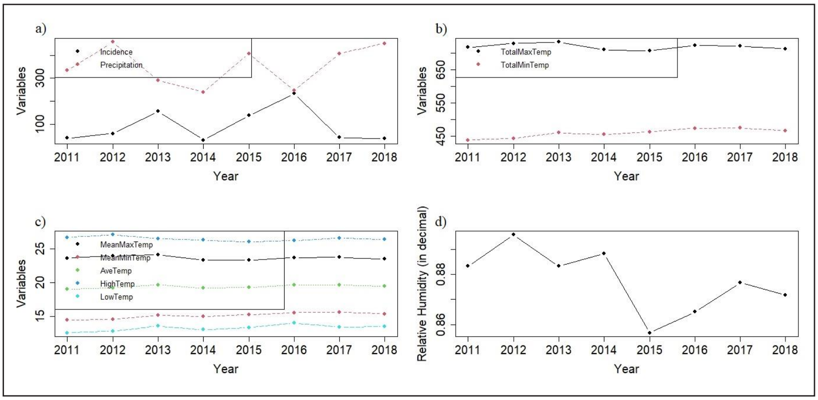
To fully understand the possible influences that further affect the current dengue incidence, we proceed with MLR analysis.

Multiple Linear Regression Analysis

Figure 3 shows the average yearly trend of dengue incidence and meteorological factors in Baguio City from 2011 to 2018. In Figure 3a, the trends show that precipitation has an inverse influence on the incidence. Observe that in 2016, the incidence reached its peak despite low precipitation. It can be seen from Figures 3b and 3c that Baguio City experiences high temperatures relative to other years, more evidently in terms of Total Minimum and Total Maximum. This has been noted in Polonio (2016) that the hot temperature gives an advantage to the lifespan and reproduction of dengue-carrying mosquitoes.

Figure 3

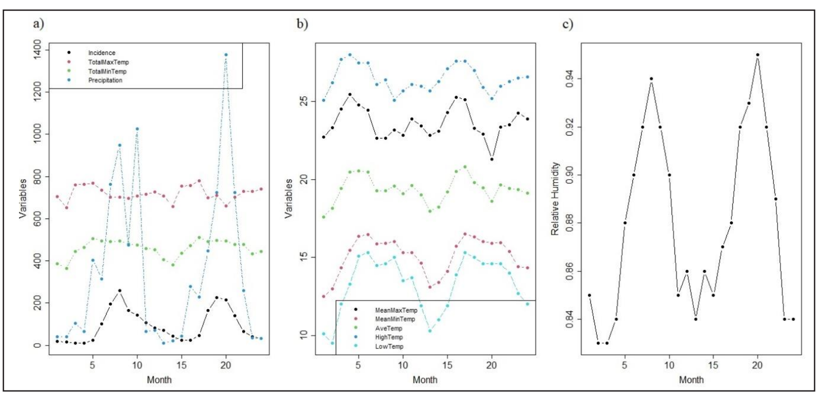
Average yearly trend of dengue incidence and meteorological factors in Baguio City from 2011 to 2018



In Figure 3d, relative humidity follows a generally increasing pattern in Baguio City for every two years. We use this pattern to analyze the influences of the period and the environmental factors on the incidence. As a result, we modify our monthly data into a two-year average monthly, resulting in twenty-four data points for each variable. Figure 4 shows the resulting two-year average monthly trend of dengue incidence and meteorological factors in Baguio City from 2011 to 2018.

Figure 4

Two-year average monthly trend of dengue incidence and meteorological factors in Baguio City from 2011 to 2018

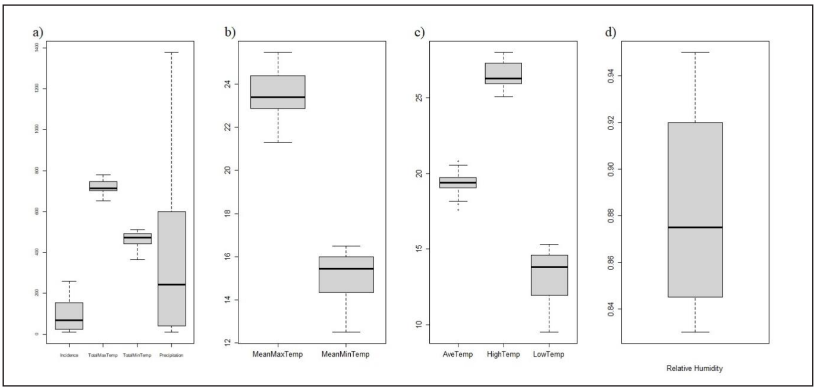


In Figures 4a and 4c, trends show that high incidence is recorded on the 8th and 20th months, both in August, in a two-year period along with precipitation and relative humidity. Furthermore, in Figures 4a and 4b, average temperature, mean maximum temperature, highest temperature, and total maximum temperature show a decline in those months. We use this information as we construct our MLR models. For the normality distribution, Figure 5 shows the boxplot for each data.

In Figures 5a – 5d, the data distributions are shown to be not normally distributed.

Figure 5

Two-year average monthly boxplots of dengue incidence and meteorological factors in Baguio City from 2011 to 2018



However, according to Li et al. (2012), it is more accurate to check the normality of the residuals or the conditional normality of the dependent variable rather than the dependent variable itself is normally distributed. Hence, despite the non-normality of the dependent variable, we can still proceed with our experiment on MLR models.

Table 3 shows the correlation coefficient matrix using Pearson's correlation. Strong correlations are indicated in green cells. Cells with NA entries indicate that the corresponding correlation coefficient is not applicable in our study.

It can be observed that Relative Humidity and Precipitation have high correlation with Incidence. This means that one of the two climatic factors has a high influence on the incidence. However, by collinearity, Relative Humidity and Precipitation cannot be both independent variables in one model.

Table 3

Correlation matrix between two variables. Green cell indicates strong correlation, and N/A indicates Not Applicable in the study

	Incidence	Time	Total Max Temp	Mean Max Temp	Total Min Temp	Mean Min Temp	AveTemp	High Temp	Low Temp	RelHum	Prec
Incidence	1.00										
Time	0.18	1.00									
TotalMaxTemp	-0.46	-0.01	1.00								
MeanMaxTemp	-0.66	-0.10	N/A	1.00							
TotalMinTemp	0.54	0.23	N/A	N/A	1.00						
MeanMinTemp	0.53	0.22	N/A	N/A	N/A	1.00					
AveTemp	-0.02	0.09	N/A	N/A	N/A	N/A	1.00				
HighTemp	-0.47	-0.11	N/A	N/A	N/A	N/A	N/A	1.00			
LowTemp	0.56	0.24	N/A	N/A	N/A	N/A	N/A	N/A	1.00		
RelHum	0.89	0.23	-0.32	-0.51	0.70	0.71	0.20	-0.31	0.75	1.00	
Prec	0.83	0.14	-0.34	-0.57	0.60	0.58	0.07	-0.35	0.60	0.88	1.00

Establishing the possible MLR models given the correlation table and applying the necessary error analysis, we acquire four models - Incidence vs. RelHum + AveTemp (Model 1), Incidence vs. RelHum + HighTemp (Model 2), Incidence vs. RelHum + MeanMaxTemp (Model 3), Incidence vs. Prec + MeanMaxTemp (Model 4).

All models exhibit relatively high *R*-squared values and small *p*-values, indicating that they are highly correlated with dengue incidence. There are other models that have high *R*-squared values like Incidence vs. RelHum + TotalMaxTemp and Incidence vs. RelHum + MeanMaxTemp + Time, but the *p*-value of one of their coefficients are greater than 0.05, thus making it insignificant for the model.

Model 3 shows to be the best-fitting model as it exhibits the highest adjusted *R*-squared value, indicating that the model shows a very strong correlation with 83.83%.

Table 4

Model estimation and error analysis of Model 1

Coefficients	Estimate	Standard Error	*p*-value
RelHum	1893.91	184.55	1.23e-09
AveTemp	-18.46	8.33	0.04
(Intercept)	-1215.87	205.00	6.91e-06
Residual standard error: 33 on 21 degrees of freedom			
Multiple *R*-squared: 0.83	Adjusted *R*-squared: 0.82		
F statistic: 52.67 on 2 and 21 degrees of freedom		*p*-value: 6.56e-09	

Table 5

Model estimation and error analysis of Model 2

Coefficients	Estimate	Standard Error	*p*-value
RelHum	1676.84	189.80	1.53e-09
HighTemp	-19.22	8.38	0.03
(Intercept)	-873.53	316.39	0.01
Residual standard error: 32.78 on 21 degrees of freedom			
Multiple *R*-squared: 0.84	Adjusted *R*-squared: 0.82		
F statistic: 53.53 on 2 and 21 degrees of freedom		*p*-value: 5.70e-09	

Table 6

Model estimation and error analysis of Model 3

Coefficients	Estimate	Standard Error	p-value
RelHum	1520.30	198.81	1.68e-09
MeanMaxTemp	-21.61	7.56	0.01
(Intercept)	-734.02	307.74	0.03
Residual standard error: 31.10 on 21 degrees of freedom			
Multiple R-squared: 0.85	Adjusted R-squared: 0.84		
F statistic: 60.63 on 2 and 21 degrees of freedom		p-value: 1.89e-09	

Table 7

Model estimation and error analysis of Model 4

Coefficients	Estimate	Standard Error	p-value
Prec	0.13	0.03	7.56e-05
MeanMaxTemp	-19.22	10.50	0.05
(Intercept)	-21.91	253.96	0.04
Residual standard error: 41.32 on 21 degrees of freedom			
Multiple R-squared: 0.74	Adjusted R-squared: 0.71		
F statistic: 29.80 on 2 and 21 degrees of freedom		p-value: 7.37e-07	

Next, we look at the normality of the residuals and the homoscedasticity of the four models. Table 8 summarizes the results.

Table 8

Normality of the residuals and homoscedasticity of the model

Model	Mean of Residuals	Normality of residuals			Homoscedasticity of the model	
		W statistic	p-value	Remark	p-value	Remark
Model 1	1.18e-15	0.98	0.82	Normal	0.87	Homoscedastic
Model 2	2.96e-15	0.97	0.68	Normal	0.17	Homoscedastic
Model 3	-2.52e-15	0.97	0.70	Normal	0.39	Homoscedastic
Model 4	5.18e-16	0.96	0.42	Normal	6.85e-04	Heteroscedastic

Model 4 is heteroscedastic since its p-value is less than 0.05. This model does not qualify as a good model. Each of the three remaining has at least 81% accuracy based on their adjusted *R*-squared values. Each model has a significant correlation coefficient value of at least 0.90. Their respective *W* statistic is at least 0.970, satisfying the assumptions of normality of model residuals. Model 1 has the largest *W* statistic, normality *p*-value, and homoscedastic p-value among the three models. These *p*-values demonstrate that Model 1 is homoscedastic, with normally distributed residuals with a mean of 0.

From the three models, relative humidity is the most significant independent variable. Following the assumptions of MLR analysis, Model 1 is the best MLR model, implying that relative humidity and average temperature largely influence dengue incidence in Baguio City.

Despite that, Model 2 and Model 3 may still be utilized. Variables such as highest temperature and mean maximum temperature are considered significant alternatives to average temperature since according to Sintorini (2017) and Campbell, et al. (2013), the temperature influences the incidence regardless its type. Henceforth, these three models are relevant in describing the influence of dengue cases in Baguio City.

CONCLUSION

Dengue incidence in Baguio City continues to infect more individuals as possible breeding sites for dengue-carrying mosquitoes develop with time. These have been brought about by continuous fluctuation of climatic factors - precipitation, relative humidity, and temperature - in the area. Using MLR analysis, relative humidity and temperature, either average, highest, or mean maximum temperature, are shown to be significant factors in the occurrence of dengue cases in Baguio City within a two-year average timespan. In particular, relative humidity and average temperature are the variables that show the strongest influence on dengue incidence in Baguio City.

Without proper protocols in preventing the spread of dengue, it is forecasted that by June 16, 2019, dengue incidences are projected to reach 101 cases in a day. And if the current situation continues without providing certain local actions, dengue incidence may rise to 529 cases by August 29, 2019. It is recommended to alert the local officials regarding this projection so that proper safety and health protocols must be established in the locality to prevent the further spread of the disease. Once we have acquired new data for 2019, the projection will be updated based on the newly – calculated R_t values.

Since we only consider weather-related factors, we recommend considering other possible factors such as geographic or population factors in the MLR model. Dengue cases and weather factors may be analyzed together as a time series to determine the trend in the area. We also recommend conducting other forecasting methods and regression algorithms to generate a more accurate dengue forecast in the area.

ACKNOWLEDGMENT

We acknowledge the Baguio City Epidemiology and Surveillance Unit and Regional DOST-PAGASA for the raw data on dengue cases and environmental factors, respectively. We also acknowledge the funding support provided by the University of the Philippines Baguio.

REFERENCES

Addawe, R., Addawe, J, and Magadia, J (2016). Optimization of seasonal ARIMA models using Differential Evolution – Simulated Annealing (DESA) Algorithm in forecasting dengue cases in Baguio City. *AIP Conference Proceedings*, 1776, 090021.

Agoot, L. (2019, July 23). Dengue cases in Baguio, Kalinga drop: DOH. *Philippine News Agency*. https://www.pna.gov.ph/articles/1075759.

Aldstadt, J, Yoon, I-K, Buddhari, D, Jarman, R, Thomas, S, Gibbons, R, Uppapong, A, Iamsirithaworn, S, Rothman, A, Scott, T, and Endy, T (2012). Space-time analysis of hospitalised dengue patients in rural Thailand reveals important temporal intervals in the pattern of dengue virus transmission. *Tropical Medicine & International Health: TM & IH*, 17(9), 1076–85.

Campbell, K., Lin, C. D., Iamsirithaworn, S., and Scott, T. W. (2013). The Complex Relationship between Weather and Dengue Virus Transmission in Thailand. *American Journal of Tropical Medicine and Hygiene*, 89(6), 1066–1080.

Chatterjee, S & Simonoff, J. S. (2013). *Handbook of Regression Analysis*. New Jersey: John Wiley & Sons, Inc.

Cogan, J. (2021, May 19). *Dengue and Severe Dengue.* World Health Organization. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.

Cori, A. (2020). EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves. *R package version 2.2-3*. https://CRAN.R-project.org/package=EpiEstim.

Cori, A., Ferguson, N., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology, 178*(9), 1505– 1512.

Du, Z., Xu, X., Wang, L., Cowling, B. J., & Meyers, L. A. (2020). Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Research Letter*, *26*(6), 1341–1343.

Gonzalez-Estrada, E. & Cosmes, W. (2019). Shapiro-Wilk test for skew normal distributions based on data transformations. *Journal of Statistical Computation and Simulation, 89*(17), 3258–3272.

Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J., Zhang, Q., Luo, G., Li, Z – H., He, J., Zhang, Y., & Ma, W. (2017). Developing a dengue forecast model using machine learning: A case study in China. *PLoS Neglected Tropical Diseases, 11*(10), e0005973.

Jeong, J. & Lee, K. (1999). Bootstrapped Whites test for heteroskedasticity in regression models. *Economics Letters*, *63*(3), 261–267.

Jombart, T. & Nouvellet, P. (2021). projections: Project Future Case Incidence. *R Package version 0.5.4.* https://CRAN.R-project.org/package=projections

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied Regression Analysis and Other Multivariable Methods* (2nd ed.). California: PWS-KENT Publishing Company.

Li, X., Wong, W., Lamoureux, E. L., & Wong, T. Y. (2012). Are Linear Regression Techniques Appropriate for Analysis When the Dependent (Outcome) Variable Is Not Normally Distributed? *Investigative Opthalmology & Visual Science*, *53*(6), 3082–3083.

Magsakay, C., De Vera, N., Libatique, C., Addawe, R., & Addawe, J. (2017). Treatment on outliers in UBJ-SARIMA models for forecasting dengue cases on age groups not eligible for vaccination in Baguio City, Philippines. *AIP Conference Proceedings, 1905*(1), 050028.

Marigmen, J., Balino, L., & Addawe, R. (2021). Analysis of Dengue Incidence in Baguio, Philippines. *AiP Conference Proceedings, 2423*, 070011.

Nishiura, H. & Chowell, G. (2009). The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends. In G. Chowell, J. M. Hyman, L. M. A. Bettencourt, & C. Castillo – Chavez (Eds.), *Mathematical and Statistical Estimation Approaches in Epidemiology* (pp. 103–121). New York: Springer.

Nouvellet, P., Cori, A., Garske, T., Blake, I., Dorigatti, I., Hinsley, W., Jombart, T., Mills, H., Nedjati-Gilani, G., van Kerkhove, M., Fraser, C., Donnelly, C., Ferguson, N., & Riley, S. (2017). A simple approach to measure transmissibility and forecast incidence. *Epidemics*, *22*, 29– 35.

Olmoguez, I. L. G., Catindig, M. A. C., Amongos, M. F. L., & Lazan, F. G. (2019). Developing a Dengue Forecasting Model: A Case Study in Iligan City. *International Journal of Advanced Computer Science and Applications*, *10*(9), 281– 286.

Pearson, R. K. (2018). *Exploratory Data Analysis Using R* (1st ed.). Florida: CRC Press.

Polonio, J. M. (2016, May 30). Dengue cases in Baguio City continue to rise. *Sunstar Philippines.* https://www.sunstar.com.ph/article/77365.

Razali, N. M. & Yap, B. W. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Royston, P. (1995). A Remark on Algorithm AS 181: The W-test for Normality. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, *44*(4), 547–551.

Sabir, M., Ali, Y., & Muhammad, N. (2018). Forecasting incidence of dengue and selecting best method for prevention. *Journal of Pakistan Medical Association*, *68*(9), 1383–1386.

Schober, P., Boer, C., & Schwarte, L. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.

Sia Su, G. L. (2008). Correlation of Climatic Factors and Dengue Incidence in Metro Manila, Philippines. *Ambio*, *37*(4), 292–294.

Sintorini, M. M. (2017, August 9 - 10). *The correlation between temperature and humidity with the population density of Aedes aegypti as dengue fever's vector*. The 4th International Seminar on Sustainable Urban Development, Jakarta, Indonesia.

Siraj, A., Oidtman, R., Huber, J., Kraemer, M., Brady, O., Johansson, M., & Perkins, A. (2017). Temperature modulates dengue virus epidemic growth rates through its effects on reproduction numbers and generation intervals. *PLoS Neglected Tropical Diseases, 11*(7), e0005797.

Sumi, A., Telan, E. F. O., Chagan-Yasutan, H., Piolo, M. B., Hattori, T., & Kobayashi, N. (2016). Effect of temperature, relative humidity and rainfall on dengue fever and leptospirosis infections in Manila, the Philippines. *Epidemiology and Infection*, *145*(1), 78–86.

Verzani, J. (2014). *Using R for Introductory Statistics* (2nd ed.). Florida: CRC Press.

Walpole, R. E., Myers, S. L., Myers, R. H., & Ye, K. (2012). *Probability and Statistics for Engineers and Scientists* (9th ed.). Massachusetts: Pearson Education, Inc.

Zeileis, A. & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News, 2*(3), 7–10.