Enhanced Knowledge in Sciences and Technology Vol. 2 No. 1 (2022) 112-121

EKST

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/ekst
e-ISSN : 2773-6385

Visualization and Negative Binomial Model of Dengue Incidence Rate in the Central Region of Malaysia

Siti Raisah Harun[1], Norziha Che Him[1*]

[1]Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology,
Universiti Tun Hussein Onn Malaysia, 84600 Pagoh, Johor, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/ekst.2022.02.01.013
Received 02 January 2022; Accepted 01 February 2022; Available online 1 August 2022

Abstract: Dengue fever (DF) is a mosquito-borne viral illness caused by four categories of dengue virus (DENV), DENV-1, DENV-2, DENV-3, and DENV-4, all of which are members of the Flaviviridae family that mostly found in the species of Aedes aegypti and Aedes albopictus species. Humid and warm condition is a suitable condition for the breed of the mosquitoes and their lifestyle. The lack of study on the interaction between DIR and risk variables such as climatic variation and socioeconomic conditions is a problem, especially in the development of early warning and response system (EWARS) for dengue outbreak, which is thought to be crucial for increasing dengue outbreak awareness. The aim of this study is to visualize the trend of the dengue incidence rate (DIR), to obtain the significant climate variables that related to the DIR and to explore the model that best represents the data. The data of DIR from 2001 to 2009 specifically in the central region of Malaysia; Selangor and Kuala Lumpur were utilized. In achieving the objectives, scatter plot was used to visualize the relationship of the climate variables and DIR while Poisson GLM and Negative Binomial GLM model were used to model the data then statistical test such as AIC, BIC and deviance were utilized for the model selection. The monthly trend of the DIR in Selangor and Kuala Lumpur shows a seasonal trend while Negative Binomial GLM model is the best model for both Selangor and Kuala Lumpur.

Keywords: Dengue, Dengue Incidence Rate, Poisson GLM, Negative Binomial GLM.

1. Introduction

Dengue fever (DF) is a climate-sensitive mosquito-borne viral disease [1]. DENV-1, DENV-2, DENV-3, and DENV-4 are four kinds of dengue virus (DENV) from the Flaviviridae family [2]. Dengue fever is spread to humans by female mosquitoes that was infected with the dengue virus, which are usually found in the Aedes aegypti and Aedes albopictus species. [3] and [4] stated that a person

**Corresponding author: norziha@uthm.edu.my*

with DF may experience symptoms such as a sudden high fever, severe headache, rash, muscle and joint problems, which may appear three to fourteen days after infection and last for two to seven days to recover. However, it could develop to more serious DF known as dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) since both have higher rate of fatality and complications of the disease.

Mosquitoes prefer humid and warm climates, which are suitable for their breed and lifestyle. As a result, more mosquitos will reproduce, which will lead to more dengue cases. Previous research by [5] has shown that local meteorological factors, as well as non-climatic epidemiologic and ecologic variables, have a role in the DIR trend. Climate fluctuation in temperature and precipitation is said to be the most effected in DF, among other environmental factors, in the cause of the rise in DIR [6]. Apart from weather-related temperature changes, socioeconomic factors play an important role in dengue fever spread. Income, age, population density, drainage, water supply, and sanitation are all socioeconomic factors that may affect the DIR.

The lack of research on the relationship between DIR and risk variables such as meteorological variation and socioeconomic situations is an issue, particularly in the development of early warning and response system (EWARS), which is considered critical to increase the responsiveness of dengue outbreaks. Furthermore, the link between all climatic conditions and DIR is ambiguous and undefined. Not only that, additional risk variables including socioeconomic situations have not been well investigated in relation to dengue fever. Dengue fever cases, on the other hand, have risen steadily throughout the year, becoming an endemic that has a worldwide impact in a variety of ways.

Dengue fever has grown unmanageable and continues to spread due to a lack of public awareness about the disease. Any illness or health condition may be prevented if everyone, including the general public, is aware of the disease's risks and causes. In Malaysia, the lack of information accessible at the local geographic scale makes it difficult to stick to the schedule while researching meteorological conditions and dengue outbreaks.

The research questions that were considered in this research were the dengue incidence rate in the central region of Malaysia, the possible factors of the dengue incidence rate especially in the central region of Malaysia, the visualization of the dengue incidence rate in the central region of Malaysia and to what extend can the model used describe the relationship between the dengue incidence rate and the potential factors.

The objectives of this research are to obtain and visualize the DIR of the Selangor with the variables obtained using R software, to explore the possible factors of DIR in Selangor and to model the DIR and the possible factors by using Negative binomial GLM. In addition, the Poisson GLM and the Negative binomial GLM model will be compared by choosing the best model that could represent the data.

2. Materials and Methods.

2.1 Materials

This study uses monthly data of dengue cases of Selangor and Kuala Lumpur from 2001 to 2009. The data of the population was obtained from Department of Statistics Malaysia (DOSM). The climate variables were obtained from Department of Irrigation and Drainage Malaysia (DDIM), Malaysian Meteorological Department (MMD), National Centre for Environmental Prediction (NCEP) and National Centre for Atmospheric Research (NCAR). The monthly number of dengue cases collected from Ministry of Health (MOH) and State Health Department of Selangor.

Table 1: Data Description

Variables	Description
State	State used which is Selangor and Kuala Lumpur
Region	The region of each state
Year	Year; 2001 to 2009
Month	Month data since monthly trend will be observed
DF	Number of dengue fever cases
DHF	Number of dengue hemorrhagic fever cases
Death	Number of deaths
Latitude	The latitude of the states
Longitude	The longitude of the states
Population	Number of populations in the states
Nino4	ENSO values; the temperature of sea surface
Rain	Number of rainfalls
Precipitation	The precipitation of the weather
Temperature	Temperature obtained
Humidity	The humidity in each state
Days	Number of rainy days in Selangor and Kuala Lumpur
DIR	Dengue incidence rate

2.2 Calculation of Dengue Incidence Rate

The dengue incidence rate was first calculated in the study. Dengue Incidence Rate (DIR) could be defined as a number of confirmed dengue cases of DF and DHF that divided to a number of populations of the state for the particular year and month [7]. The DIR is calculated per 100,000 persons at the risk by using formula below:

$$DIR = \frac{y_{st}}{p_{sj}} \times 100{,}000 \qquad Eq. 1$$

where the y_{st} is the number of confirmed dengue cases in state in month while p_{sj} is the population in the state per year.

2.3 Data Cleaning and Lag Variable Process

Data cleaning were conducted to clean out the missing data from the actual data. The missing data can lower the power of statistics and resulting in the bias of the outcomes and thus the outcomes will be inaccurate and defective [8]. Then, the process of adding new lag variables where each variable that will be used as the descriptive variable were lagged from 1-month lag to 7-month lag. Lag variables are commonly employed as a reliable technique for eliminating residual autocorrelation and modelling dynamic data-generating processes [9].

2.4 Visualisation of DIR

The visualization of the data of DIR in each state was done accordingly to see the trends in the DIR from 2001 to 2009 in Selangor and Kuala Lumpur. Visualisation is the technique used to analyse the data to see the trend more clearly and create more different ways in making effective decision making [10]. Visualisation allows researchers to filter the variables that they needed and can reduced the amount of unnecessary data [11]. In this section, the visualization of the DIR with states; Selangor and Kuala Lumpur were used as the background context and will not represent the main modelling.

2.4 Covariate Selection

Then, variables that have significant relationship with the independent variable was chosen among the large number of other variables. In this process, the visualization in the form of scatter plot were used to identify the possible relationship of monthly DIR with the climate and non-climate variables. Scatter plots are one of the most effective and widely used visual data exploration tools since it may be used to determine the correlation between two points, clusters, variables, or massive quantity of multidimensional data [12]. The climate variables may consist of temperature, humidity, ENSO values, number of rainy days and rainfall.

2.5 Generalised Linear Model

The model of Poisson GLM and Negative Binomial GLM model were built for both Selangor and Kuala Lumpur after significant variables were chosen based on scatter plot. GLM is also known as identifying two components and the response variable is belong to the exponential distribution family [13]. The Poisson model was used to model dengue counts data. The count data, on the other hand, is very susceptible to overdispersion. Because the Poisson distribution assumes that the mean and variance values are equal, the presence of overdispersion might have an impact in the outcome. To deal with overdispersion, Poisson GLM should be enlarged to a model called Negative-binomial GLM, which is capable of handling overdispersion. According to [14], we may convert the model to a negative binomial model to adjust to the overdispersion. Thus, to counteract the presence of overdispersion, the Negative Binomial GLM was used.

Two model each was built since both state of Selangor and Kuala Lumpur were observed. The models then were written following the equation below:

$$y \sim P[\mu_I] = P[p_i, \rho_i] \qquad Eq. 2$$

$$y \sim NB\ [\mu_I, \theta_1] = P[p_i, \rho_i, \theta_1\ ] \qquad Eq. 3$$

$$\log \mu_i = \log p_i + \log \rho_i = \ log\, p_i + \ \beta_0 + \sum_{j=1}^{p} \beta_j\, x_{ji} \qquad Eq. 4$$

where p_i represents the population size, DIR represented by the symbol ρ_i and for Negative Binomial, θ_1 represents the mass probability function and x_{ji} represents the predictors variable that significant that had been selected based on the visualization of the scatter plots.

2.6 Model Selection

Since both Poisson and Negative-Binomial model were done in this study, a model selection was done by comparing the value of AIC, BIC and deviance. AIC or Akaike Information Criterion is value that considered as the estimator of the error predictor and estimates the relative information lost from the model. BIC or Bayesian Information Criterion is a method of scoring and selecting a model. The lesser the AIC value, the better the model since the information loss are lesser. However, the values of AIC are inconsistent since the when n is quite large, non-vanishing on selecting the unneeded model [15]. The model with the lowest AIC, BIC, and deviation value is the better model to be utilised according to the results.

$$AIC = \ -2\ln(L) + 2k \qquad Eq. 5$$

$$BIC = \ -2\ln(L) + klog(n) \qquad Eq. 6$$

where L is the log likelihood of the model, k is the number of factors in the model and n is the sample size. In the other hand, deviance explained the difference of the fitted model log likelihood,

$\ln(\widehat{L})$ and the saturated model likelihood $\ln(L)$ that can be seen by the Eq. 2.7 below. Hence, it will be required to be included in the statistical tests.

$$Deviance = 2\left[\ln(\widehat{L}) - \ln(L)\right] \qquad Eq. 7$$

3. Results and Discussion

A discussion of the results was held when the data analysis procedure was achieved using R software and following the approaches outlined in the methodology. After calculating the DIR value, the missing data then were deleted and left only 1260 data from 1297 data available. Then, the processing of adding the lag variables resulting in the addition of variables from 27 variables to 84 variables.

3.1 Trend of DIR

Both trend of DIR annually and monthly were visualized for Selangor and Kuala Lumpur from 2001 to 2009. Year 2008 recorded the highest DIR in Selangor while year 2006 recorded the highest DIR in Kuala Lumpur. Meanwhile, the lowest DIR documented was on 2001 for both Selangor and Kuala Lumpur.

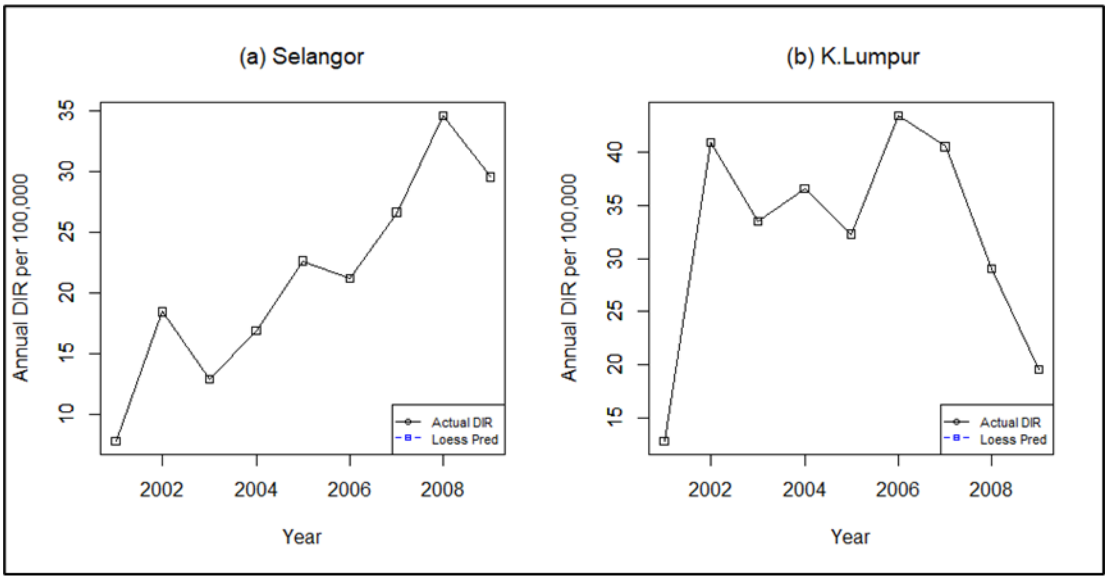

Figure 1: Annual DIR per 100,000 populations in (a) Selangor and (b) Kuala Lumpur

Generally, the annual DIR movement in Selangor shows an increasing trend throughout the year 2001 to 2009. In Figure 1, an annual DIR per 100,000 populations can be seen clearly in Selangor that illustrates an increasing trend of DIR from 2001 to 2002. However, the DIR declines with an approximate of 39% decrement in 2003 and continue rising until 2005. In 2006, a slight decrease in DIR can be shown. Next, the DIR starts increasing greatly up to peak in 2008 with the number of 21,262 cases which is the highest number of cases recorded among the year 2001 to 2009. However, the DIR slightly decrease in subsequent year, 2009.

Meanwhile, the annual DIR per 100,000 populations in Kuala Lumpur shows a sharp increase from 2001 to 2002 with a 63% increment of the DF and DHF cases recorded. Then, the DIR shows a minor decrement in the year 2003 followed an increment in 2004. Then, the DIR decreased in 2005. Then, in 2006, the DF and DHF reported cases recorded the highest with a total of 7804 cases and average DIR of 43.4 per 100,000 populations. Nonetheless, the trend of the DIR started to decrease sharply from with a two-fold decrement from 2006 to 2009 in Kuala Lumpur.

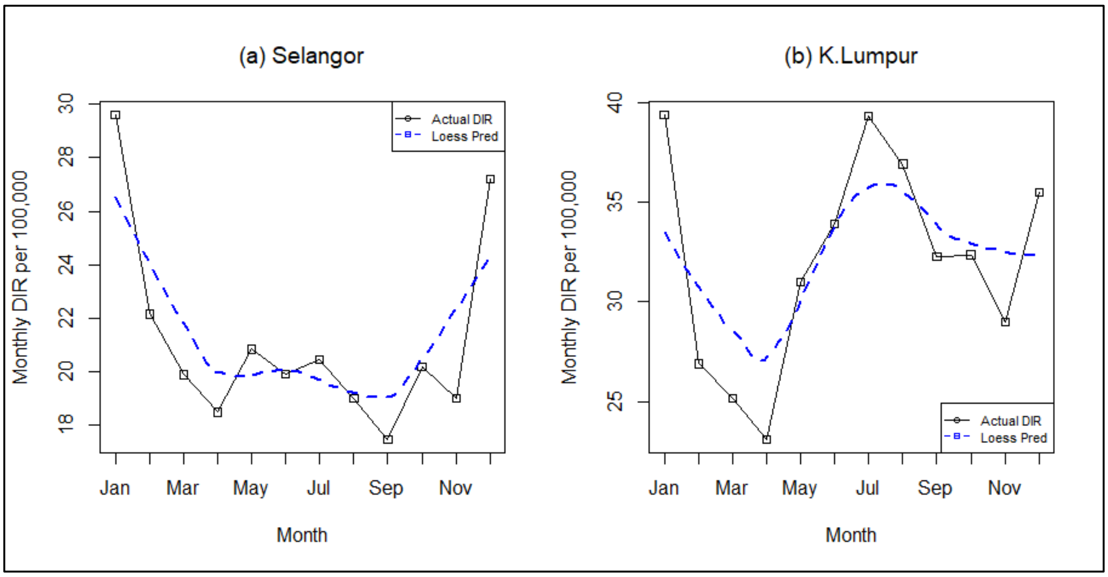

Figure 2: Monthly DIR per 100,000 populations in (a) Selangor and (b) Kuala Lumpur

Generally, the pattern of DIR in both Selangor and Kuala Lumpur based on Figure 2 above were quite similar which is a seasonal pattern. Seasonal variations in incidence rates are common in all vector-borne diseases, according to [16], because of the vector's life cycles and their need on the local climate for breeding sites and an appropriate temperature. The DIR per 100,000 people in Selangor plummeted from January to April, then began to climb in May before dropping slightly in June. Then, in July, a little increase of around 5%, followed by a drop in DIR the following month, August through September. Following a spike of almost 18 percent in October, Selangor's mean monthly DIR fell somewhat in November. The DIR, on the other hand, climbed substantially in December. The Northeast monsoon, which occurs between November and March, may have an impact on the monthly DIR pattern in Selangor. Considering Selangor and Kuala Lumpur are located in Malaysia's middle area, they may be impacted by the Northeast monsoon. The Northeast monsoon delivers heavy rain and stormy seas to the exposed coastlines of Southwestern Sarawak and northern and North-eastern Sabah, as well as flooding in some part of Peninsula's eastern [17]. Due to the high temperatures and humidity during the Northeast monsoon, it will be a suitable condition for the breeding of the Aedes mosquitoes.

For the first four months, the monthly DIR per 100,000 populations in Kuala Lumpur has been rapidly declining. The DIR then climbed dramatically in the following months, from April to July. The Southwest monsoon, which runs from May to September, may be the reason for the peak of DIR in the middle months. The Southwest monsoon mostly impacts Sabah's Southwestern coastal strip, which is prone to floods. Although neither peninsular nor insular Malaysia are in the tropical cyclone (typhoon) belt, their shores are periodically subjected to storms, which bring torrential rainstorms [17]. A declined in DIR can be seen in August to September and started to increase in November to December. Based on Figure 2, the high DIR recorded in both early month and latest month of Kuala Lumpur may be cause by the Northwest monsoon.

3.2 Model Framework

The previous significant variables that were chosen based on the scatter plot were then included in the model framework. The significant variables that were chosen are Nino4, precipitation, humidity, rainy days lag 2 months, rainy day lag 3 months, rainfall and rainfall lag 3 months for Selangor and Nino4, precipitation, precipitation lag 3 months, humidity, rainy days lag 2 months, rainy days lag 3 months, rainfall and rainfall lag 2 months were chose for Kuala Lumpur. Then, the significant variables were modelled by using Poisson and Negative Binomial GLM.

Table 2: Summary of Poisson GLM Model of Selangor

Coefficients	Covariates	Estimate	Pr(>|z|)
α_{PS}	Intercepts	4.0530	**0.0000**
$\boldsymbol{\beta_1}$	**NINO4**	-0.0336	**0.0000**
$\boldsymbol{\beta_2}$	**Precipitation**	0.0171	**0.0000**
β_3	Humidity	-0.0015	0.2860
$\boldsymbol{\beta_4}$	**Rainy days Lag 2**	-0.0371	**0.0000**
$\boldsymbol{\beta_5}$	**Rainy days Lag 3**	0.0490	**0.0000**
$\boldsymbol{\beta_6}$	**Rainfall**	-0.0018	**0.0000**
$\boldsymbol{\beta_7}$	**Rainfall Lag 3**	-0.0007	**0.0000**

Table 3: Summary of Negative Binomial GLM Model of Selangor

Coefficients	Covariates	Estimate	Pr(>|z|)
α_{NBS}	Intercepts	1.2334	0.6140
β_1	NINO4	0.0505	0.4860
β_2	Precipitation	0.0283	0.1960
β_3	Humidity	0.0025	0.9000
β_4	Rainy days Lag 2	0.0306	0.2040
β_5	Rainy days Lag 3	-0.0203	0.3200
β_6	Rainfall	-0.0005	0.3690
$\boldsymbol{\beta_7}$	**Rainfall Lag 3**	-0.0020	**0.0000**

In the Poisson model of Selangor, the significant climate variables were modelled. This model has a strong connection due to the calculated p-value that is lesser than 0.05 significance level. Besides, all the value of each variable computed in the model shows a statistically significant relationship by having p-value less than 0.05 except for variable humidity. Hence, the relationship between DIR and the climate variables of Nino4, precipitation, rainy day lag 2 months, rainy days lag 3 months, rainfall and rainfall lag 3 months are statistically significant at 0.05 significant level.

Table 3 above shows a Negative Binomial model of Selangor. The variables were the same with the previous Poisson model of Selangor. The Negative Binomial model of Selangor above are not statistically significant due to the higher *p*-value than 0.05 significance level. Only rainfall lag 3 months have a very strong relationship with the DIR compared to the other variables since the p-value is lesser than 0.05 significance level. Hence, variable rainfall lag 3 months can be concluded to have a strong relationship with the DIR in Selangor.

Table 4: Summary of Poisson GLM Model of Kuala Lumpur

Coefficients	Covariates	Estimate	Pr(>|z|)
α_{PKL}	Intercept	-5.949	**0.0000**
$\boldsymbol{\beta_1}$	**Rainy days Lag 2**	-0.0862	**0.0000**
$\boldsymbol{\beta_2}$	**Rainy days Lag 3**	0.0534	**0.0000**
$\boldsymbol{\beta_3}$	**Rainfall**	0.0004	**0.0000**
$\boldsymbol{\beta_4}$	**Rainfall Lag 2**	-0.0011	**0.0000**
$\boldsymbol{\beta_5}$	**Humidity**	-0.0099	**0.0000**
$\boldsymbol{\beta_6}$	**Precipitation**	0.0216	**0.0000**
$\boldsymbol{\beta_7}$	**Precipitation Lag 3**	-0.0120	**0.0000**
$\boldsymbol{\beta_8}$	**NINO4**	-0.3726	**0.0000**

Table 5: Summary of Negative Binomial GLM Model of Kuala Lumpur

Coefficients	Covariates	Estimate	Pr(>|z|)
α_{NBKL}	Intercept	-6.3416	**0.0109**
$\boldsymbol{\beta_1}$	**Rainy days Lag 2**	-0.1001	**0.0001**
$\boldsymbol{\beta_2}$	**Rainy days Lag 3**	0.0614	**0.0097**
β_3	Rainfall	0.0002	0.7830
β_4	Rainfall Lag 2	-0.0010	0.0548
β_5	Humidity	-0.0018	0.9271
β_6	Precipitation	0.0179	0.5290
β_7	Precipitation Lag 3	-0.0081	0.7233
$\boldsymbol{\beta_8}$	**NINO4**	0.3627	**0.0000**

Table 4 above shows a Poisson model of Kuala Lumpur. Significant climate variables that were chosen previously consists of rainy days lag 2 months, rainy days lag 3 months, rainfall, rainfall lag 2 months, humidity, precipitation, precipitation lag 3 months and Nino4. The model is statistically significant due to the p-value lesser than 0.05 significance level of all climate variables in the model. As a conclusion, rainy days lag 2 months, rainy days lag 3 months, rainfall, rainfall lag 2 months, humidity, precipitation, precipitation lag 3 months and Nino4 shows a very strong relationship with DIR in Kuala Lumpur since all the p-values computed is less than 0.05 significance level.

The Negative binomial model of Kuala Lumpur is statistically significant because the overall p-value of the model is lesser than 0.05 significance level. Next, variables rainy days lag 2 and 3 months along with Nino4 shows a strong association with the DIR in Kuala Lumpur since the p-value is lesser than 0.005 significance level. Next, based on the result, Nino4 have the most positive relationship with DIR considering the estimate value is 0.3627 which is the highest.

3.3 Model Selection

Both Negative Binomial and Poisson model that have been developed then undergone a statistical test to choose the best model between them. The statistical test used were AIC, BIC and deviance to observe which of the model has lesser variability. After the evaluation of the model, the results then were shown in the Table 6 and Table 7.

Table 6: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviance for Model Poisson GLM Selangor and Model Negative Binomial GLM Selangor

Statistical Test	Model Poisson GLM Selangor	**Model Negative Binomial Selangor**
AIC	21799	**1576**
BIC	21820	**1600**
Deviance	20883	**109**

Based on Table 6, between Poisson and Negative Binomial model of Selangor, Negative Binomial computed the lesser value in all AIC, BIC and deviance. The AIC number of the Negative Binomial model of Selangor is 1576 while the Poisson model is 21799. Meanwhile, the BIC value of the Negative Binomial model of Selangor is 1600 compared to the Poisson model of Selangor with a value of 21820. In addition, the deviance of the Negative Binomial model of Selangor is relatively low which is 109 compared to the Poisson model of Selangor which is 20883. As a conclusion, for Selangor, the best model that represents the relationship of DIR and the climate variables is Negative Binomial model since this model have lesser variance.

Table 7: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviance for Model Poisson GLM Kuala Lumpur and Model Negative Binomial GLM Kuala Lumpur.

Statistical Test	Model Poisson GLM Kuala Lumpur	**Model Negative Binomial GLM Kuala Lumpur**
AIC	9802	**1410**
BIC	9826	**1436**
Deviance	8926	**108**

For Kuala Lumpur, the best model is Negative Binomial model due to the AIC, BIC and deviance values which are 1410, 1436 and 108 respectively. The variance of the Negative Binomial model of Kuala Lumpur is relatively low than the Poisson model of Kuala Lumpur with the value of 9802 for AIC, 9826 for BIC and 8926 for the deviance

4. Conclusion

All the objectives of the study had been fulfilled and were discussed in detail. The illustration of the DIR trend in both Selangor and Kuala Lumpur was plotted in accordance with the objectives. While the tendency is explained by the monsoon, it can be argued that climatic variability had an influence on the DIR in Selangor and Kuala Lumpur from 2001 to 2009. All 84 variables contained in the data, including the lagged variables data, were shortlisted as significant variables. The significant variables were identified by looking at the trend on the scatter plot. Nino4, precipitation, humidity, rainy days lag 2 months, rainy days lag 3 months, rainfall, and rainfall lag 3 months are the main climatic factors for Selangor, according to the findings. Based on the Poisson GLM model of Selangor, all selected climate variables were said to strongly correlated with the DIR except humidity while the Negative Binomial model of Selangor computed that only rainfall lag 3 months affect the DIR in Selangor from 2001 to 2009. Furthermore, the Poisson model of Kuala Lumpur indicates that all climate variables that were selected have strong association with the DIR in Kuala Lumpur from 2001 to 2009. On the other hand, the Negative Binomial model of Kuala Lumpur concluded that only rainy days lag 2, rainy days lag 3 months and Nino4 have strong association with the DIR. Last but not least, the model selection objective which is the last objective of this research indicates that for both state of Selangor and Kuala Lumpur, Negative Binomial model is the best model since the AIC, BIC and deviance value is lower than the Poisson model.

Acknowledgement

The authors would like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support and to the reviewers for their beautiful remarks.

References

[1] Xuan, L.T.T., Van Hau, P., Thu, D.T. and Toan, D.T.T., “Estimates of meteorological variability in association with dengue cases in a coastal city in northern Vietnam: an ecological study,” Global Health Action, vol. 7, pp. 23119, 2014.

[2] Jamil, N.I, Che Him, N. and Hamdan, S.F.A., “Spatio-Temporal Distribution of Dengue Incidence Rate (DIR) in Selangor, Malaysia,” International Journal of Advanced Science and Technology, 29, pp. 2509-2516, 2020.

[3] World Health Organization, “Dengue and severe dengue,” 2021.

[4] Kularatne, S.A., “Dengue fever,” BMJ (Clinical research ed.); 351:h4661, 2015.

[5] Patz, J.A. and Olson, S.H., "Climate change and health: Global to local influences on disease risk," Annals of Tropical Medicine and Parasitology, 100(5), pp. 535–549, 2006.

[6] Watts, M. J., Kotsila, P., Mortyn, P. G., Sarto I Monteys, V., & Urzi Brancati, C., "Influence of socio-economic, demographic and climate factors on the regional distribution of dengue in the United States and Mexico," International journal of health geographics, 19(1), pp. 44, 2020

[7] Hafiz, H., Shamarina, S., and Nor, R. H., "Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia," Geospatial Health, 7(1), pp. 21–25, 2012.

[8] Kang, H., "The prevention and handling missing data," Korean J Anesthesiol, 64(5), pp. 402-406, 2013.

[9] Wilkins, Arjun., "To Lag or Not to Lag?: Re-Evaluating the Use of Lagged Dependent Variables in Regression Analysis," Political Science Research and Methods. 6, 2017.

[10] Gandhi, Parul & Pruthi, Jyoti, "Data Visualization Techniques: Traditional Data to Big Data. 2020.

[11] Cota, M. P, Rodríguez, M. D, Gonzalez - Castro, M. R. and Gonçalves, R. M. M., "Analysis of Current Visualization Techniques and Main Challenges for the Future," Journal of Information Systems Engineering & Management, 2017.

[12] Keim, Daniel A; Hao, Ming C; Dayal, Umeshwar; Janetzko, Halldor; Bak, Peter, "Generalized scatter plots," Information Visualization, 9(4), 301–311, 2010.

[13] Faraway, Julian, "Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models," Second Edition, 2016.

[14] Simon, P.W., Matthew, G.K. and Fred, L.M. (2003), "Statistical and econometric methods for transportation data analysis," Chapman and Hall/Crc. 2003.

[15] Dziak, J.J. Coffman, D.L. Lanza, S.T. Li, R., "Sensitivity and Specifying of Information Criteria," PeerJ Preprints. (4), 2017.

[16] Johansson, M. A., Dominici, F., and Glass, G. E., "Local and global effects of climate on dengue transmission in Puerto Rico," PLoS Neglected Tropical Diseases, 3(2), 2009

[17] Leinbach, T. R., Bee,. Ooi Jin, Lockard,. Craig A. and Ahmad,. Zakaria, "Malaysia," Encyclopedia Britannica. 2021