Enhanced Knowledge in Sciences and Technology Vol. 2 No. 1 (2022) 103-111
© Universiti Tun Hussein Onn Malaysia Publisher's Office

EKST

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

Dengue Epidemic and Meteorological Factors in Central Region of Malaysia Using Generalised Poisson Model

Anis Yasmin Mohamad Yasir[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.2022.02.01.012
Received 02 January 2022; Accepted 01 February 2022; Available online 1 August 2022

Abstract: *Aedes aegypti* mosquitoes are responsible for the transmission of dengue. Despite dengue being the most rapidly spreading vector-borne disease, there has been insufficient discussion of the latest dengue incidence rate (DIR) model to use to predict future outbreaks. The inclusion of trivial elements in the DIR prediction model might result in an incorrect output. In this study, the aim is to identify the monthly trend of DIR, to build a model of DIR, and then to explore which model best fits the data. Petaling, Hulu Langat, Gombak, and Klang are the four provinces that were studied from 2010 to 2015. Variables such as population, dengue cases, population density, DIR and rainfall were considered in this study. These data were collected over the course of several months from the Department of Statistics Malaysia (DOSM), the Department of Environment (DOE), and the Ministry of Health (MOH). Methods for this study included plotting weekly DIRs against time, Poisson GLMs, and statistical tests to select models, such as Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC) and deviance (D). By looking at D, AIC, and BIC, Cluster 2 is the best model as it has the lowest value. The significant variables obtained such as population density (β_1), locality number (β_2), rainfall lag 1 (β_3) and rainfall lag 2 (β_4). Clustering technique was to produce two sets of data, as well as the ability of merging meteorological and non-climatic parameters in modelling dengue incidence rate in Selangor.

Keywords: Dengue, Dengue Incidence Rate, Poisson GLMs, Clustering

1. Introduction

Female *Ae. aegypti* are the species of mosquito that transmit dengue viruses to humans. Dengue fever (DF) is the most dangerous mosquito-transmitted epidemic [1] and it is also the fastest spreading. In Malaysia, several dengue control projects have been implemented, including the general people and competent governments. The first to record DF in Malaysia was in 1902 [2]. Following the disclosure of the case, there were epidemics of dengue shock syndrome (DSS) and dengue haemorrhagic fever

*Corresponding author: norziha@uthm.edu.my

(DHF) and in 1962, and DF cases were legally noticeable in 1971[3][4]. All health practitioners are now required to report every case of dengue illness to a neighbouring local medical center within 24 hours of diagnosis [5]. The Ministry of Health Malaysia has continuously documented an increase in annual dengue cases as surveillance techniques have advanced, notably since 1980, emphasizing DF as a serious public health concern in the heavily urbanized states of Selangor [6].

Despite dengue being the most rapidly spreading vector-borne disease, there has been insufficient discussion of the latest DIR model to use to predict future outbreaks. The inclusion of trivial elements in the DIR prediction model might result in an incorrect output. Few modelling studies have been undertaken in Malaysia on the association between DIR and meteorological variables and other factors. The challenge is establishing a realistic model that explains the link between DIR and both climatic and non-climatic elements that may impact the development of EWS in various Malaysian states or districts.

These are questions that were considered in this research such as what is the monthly trend of DIR for certain districts in Petaling, Hulu Langat, Gombak and Klang, is there a link between the possible variables and the dengue incidence rate as determined by the Generalized Poisson Model and non-climatic factors, such as intrinsic seasonality in dengue incidence, demographic factors, and so on get muddled.

The research objectives for this research are to determine the monthly trend of DIR in Petaling, Hulu Langat, Gombak and Klang. The second objective is to explore the relationship of climatic and non-climatic factors of dengue cases at the Petaling, Hulu Langat, Gombak and Klang. The last objective is to build a generalized poisson regression in DF in central Petaling, Hulu Langat, Gombak and Klang.

2. Materials and Methods

2.1 Materials

Table 1 show the variables used in this study such as population, dengue cases, population density, DIR, and rainfall. The Department of Statistics Malaysia (DOSM) collected data on the population, areas, and household size in Petaling, Hulu Langat, Gombak, and Klang from 2010 to 2015. The population density was calculated by dividing the number of people in each state or district by the total area of the states or districts. Following that, the climatic dataset was used to identify possible risk variables that might impact DIR in Petaling, Hulu Langat, Gombak, and Klang. The climatic variable, such as rainfall, were collected from the Malaysia Meteorology Department (MMD).

Table 1: Data Description

Variables	Description
Population	The population data for Petaling, Hulu Langat, Gombak and Klang.
Dengue cases	The number of dengue cases.
Population Density	The population density is calculated by dividing the population by the area per square kilometre.
DIR	DIR is computed by dividing the number of dengue cases in a region by the number of people in that region per 100,000 people.
Rainfall	The number of rainfalls in Petaling, Hulu Langat, Gombak and Klang.

2.2 Data Cleaning

The missing data must be dealt with and controlled as the initial stage in starting the study. Missing data can weaken the statistical power of a study and provide skewed estimates, resulting in erroneous results [7]. In addition to adding new variables, lag data is the process of modifying existing variables. Each variable will be created with seven lags based on the original variable of potential factors.

2.3 Plotting the Monthly DIR Trend

The dengue incidence rate (DIR) is define as the number of reported cases, in the area, during the time period is divided by the total estimated population of the area for the year in the time period falls [8]. The DIR can be calculated using the Eq.1 where y_{st} represent the number of new dengue cases in the area during the time period. It is then divided by the estimation of the total population of the area, P_{st} .

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

2.3 Clustering Data by District

There are two clusters within the dataset. The first cluster contains district data with a mean monthly DIR over 100 cases per 100,000 populations, while the second cluster contains data with a mean monthly DIR under 100 cases per 100,000 populations. Due to the high DIR values recorded between January 2010 and August 2015, the districts of Petaling and Hulu Langat are considered Cluster A in this study, while Cluster B refers to the low DIR values in Selangor districts which are Klang and Gombak [9].

Table 2: The division of the district in Selangor to divide the data into two clusters

Cluster A	Cluster B
Petaling	Klang
Hulu Langat	Gombak

2.4 Covariates selection

The model's linearity is assessed by studying the scatter plot of dengue incidence rates and potential variables for each site [10]. If the scatterplot shows that the plot is random and devoid of any pattern or trend, the dataset is assumed to be linear [11]. For linear data, the Generalised Linear Model (GLM) as Eq.2 is commonly utilised. The Poisson GLM with a log link and population is described as in Eq.3.

$$p(y|\theta_i, \emptyset) = \exp\left[\frac{(y\theta_i - b(\theta_i))}{a(\emptyset)} + c[y, \emptyset]\right] \qquad Eq.2$$

$$\log \mu_i = \log p_i + \log \rho_i = \ log\, p_i + \ \beta_0 + \sum_{j=1}^{p} \beta_j\, x_{ji} \qquad Eq.3$$

where,

ρ_i = DIR at certain time

x_{ji}= Predictor of DIR at certain time and place

$\log p_i$= Population size

2.5 Comparing AIC, BIC and Deviance

The AIC, BIC, and Deviances from Poisson GLM were compared in order to select the best model. It can also be viewed as a measure of whether relative information has been lost from the model [12].

The AIC value is regarded as the estimator of the error predictor. As long as the AIC value is the lowest, there will be the best model since the losses are minimal [12]. AIC formula is shown in Eq.4. This makes BIC a powerful way to analyse linear regression data since it can provide the much needed help in fitting a good model [12]. BIC values, as shown in Eq.5 also produce a great model when it reaches the lowest value. Model selection is also influenced by deviation [12]. The least deviation indicates that the model did not diverge far from its nature, and therefore was addressed as in Eq.6. The deviation D, which is the difference between the fitted and saturated model log likelihoods, is a critical metric which should be considered.

$$AIC = -2\ln(L) + 2k \qquad Eq. 4$$

$$BIC = -2\ln(L) + klog(n) \qquad Eq. 5$$

$$Deviance = 2\left[\ln(\widehat{L}) - \ln(L)\right] \qquad Eq. 6$$

3. Results and Discussion

There are 612 data with 20 variables in data set from 2010 to 2015. The dplyr library's na.omit() function offers a straightforward approach to exclude missing observations.

3.1 Plotting Monthly DIR

Figure 1 shows the results of the assumption made about the monthly DIR trend for the dataset were utilized and the monthly DIR trend versus monthly times were analyse. All districts were projected to see rising trends in the monthly assumption. There may be differences in the DIR trends in different districts due to varying areas, meteorological circumstances, and geographical facts.

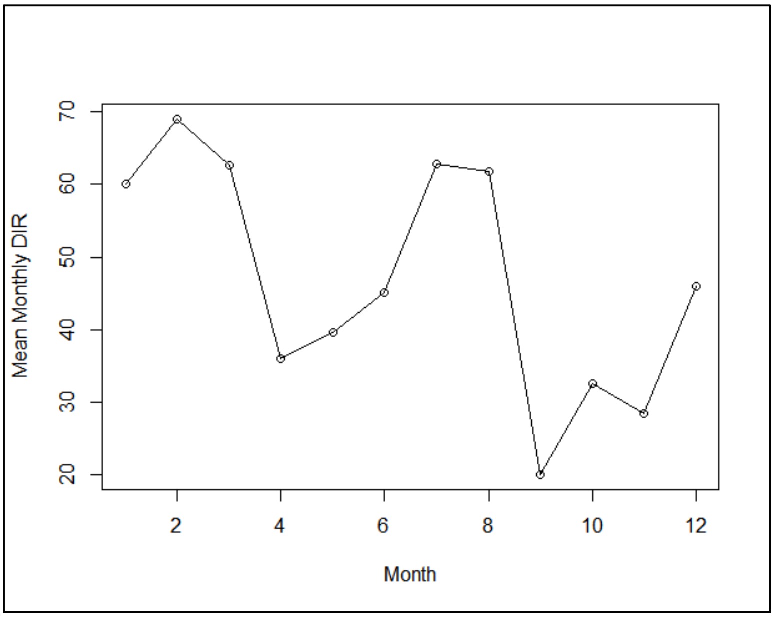

Figure 1: Average monthly DIR values for Selangor from 2010 to 2015

Figure 1 shows the monthly data set provides insight into seasonality in Selangor. From January to December, DIR does not appear to show much of an upward trend. February and July mark two of the highest DIR values, while April and September are when the lowest values are recorded. Due to Selangor's location on the west coast of Malaysia, it experiences its heaviest rainfall during October and November. A monsoon season may be responsible for this and occurs between the months of May and September, along with the Northeast monsoon between November and March. The incidence of dengue increases after heavy rains specifically in urban areas since static rainwater creates good breeding conditions for mosquitoes.

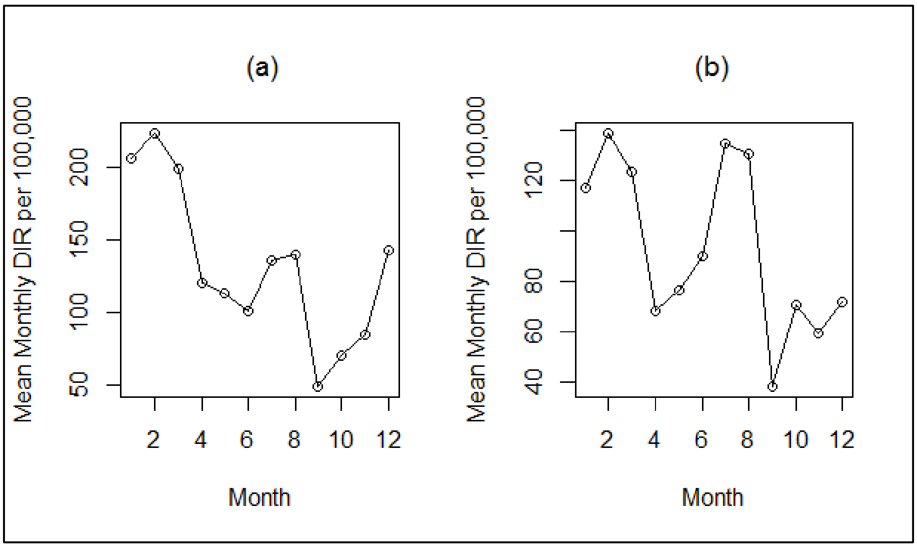

Figure 2: Mean monthly DIR for Cluster A: (a) Petaling and (b) Hulu Langat

Figure 2 shows the mean monthly DIR for Cluster A is substantially higher than the mean monthly DIR in Cluster B (Figure 4). This can be noticed by comparing the y-axes. Cluster A has a greater average monthly DIR compared to Cluster B. Cluster A consists of Petaling and Hulu Langat, both of which had a DIR of greater than 100.

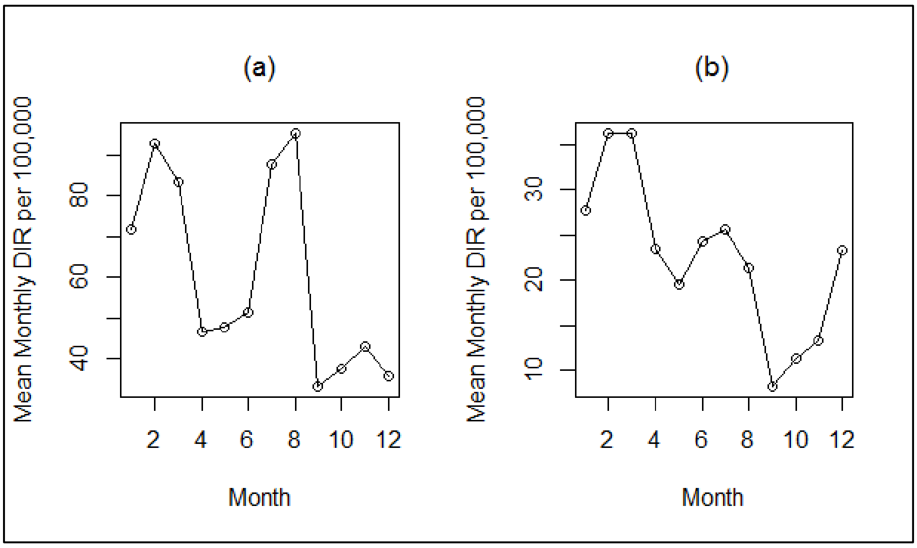

Figure 3: Mean monthly DIR for Cluster B: (a) Gombak and (b) Klang

Figure 3 shows the mean monthly DIR for Cluster B that consists of Klang and Gombak, both of which had a DIR of less than 100. All districts in Cluster B appear to display a similar mean monthly DIR pattern. All districts' peak periods for mean monthly DIR coincide around January to March and July to November. In Klang, the growth began in May and fluctuated until September, which can be seen in Figure 3, where the trend tends to be more substantial from September to December. In Gombak, the growth began in April and was quite stable until June, when it started to rise and peaked in August.

This is mostly due to population increase in metropolitan areas such as Cluster A, Petaling, and Hulu Langat, which are located surrounding Malaysia's capital, Kuala Lumpur, and are enjoying major economic growth. It is common in all vector-borne diseases to have seasonal patterns in incidence rate, since all vectors have life cycles and are dependent on local climate conditions and adequate temperatures to breed.

3.2 Dengue Incidence and Demographic Data

Figure 4 presents the relationship between the log of monthly DIR and population density in Selangor state for the 5 years period from 2010 until 2015. The plot demonstrates high levels of variability as was expected. However, there is some evidence that states with very high population densities tend to have higher DIR. According to Figure 4, the relationship between the population density and log of DIR is not necessarily linear and positive, and the scatter plots clearly demonstrate the complexity of this relationship. The infectious nature of DIR also provides an indirect influence of demographics, both population size and density - larger populations serve as more hosts for the virus, while denser populations increase opportunities for transmission. This infectious effect can be influenced by comparing DIR with DIR from the previous months.

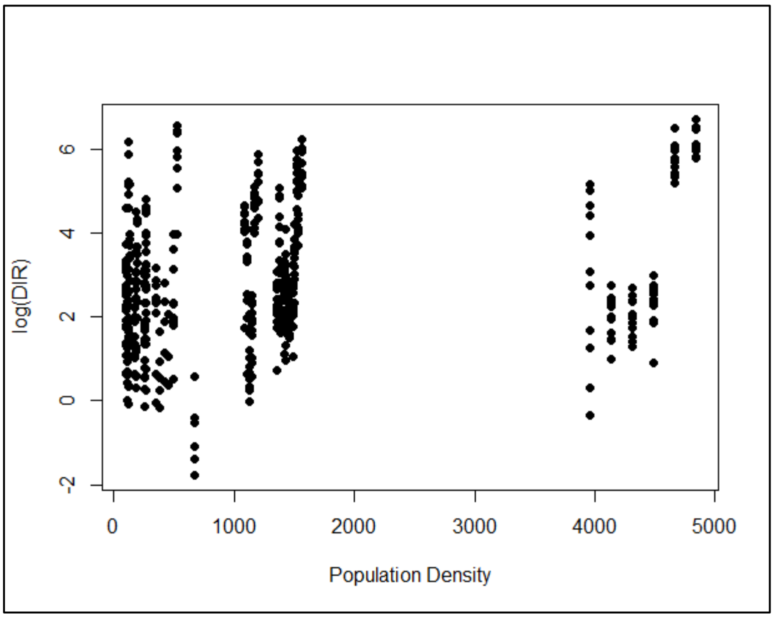

Figure 4: Relationship between log monthly DIR and population density

Figure 5 shows the relationship between log monthly DIR and locality number, which indicates that the trend of the plots between logs monthly DIR and number of localities is exponential. Figure 5 show that there is a similar pattern in both clusters, meaning that an increase in DIR value also occurred when increasing one unit of the locality. Dengue symptoms tend to be more prevalent in higher numbers of localities. The reason for this is the increased density of settlements due to the increasing number of localities. Mosquito transmission rates rise as a result of inter-individual contact and migration.

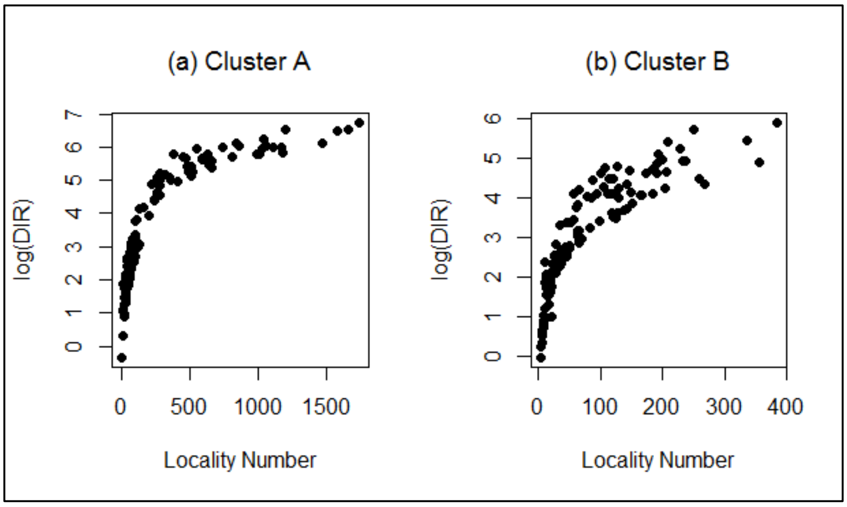

Figure 5: Relationship between log monthly DIR and locality number

3.3 Dengue Incidence and Climate Data

There are many previous studies focusing on climatic factors and dengue incidence that are either universal or particular to Malaysia as a whole. Malaysia's changing climate may have a major impact on dengue incidence. The monsoon season is known to contribute highly to the increase in dengue incidents.

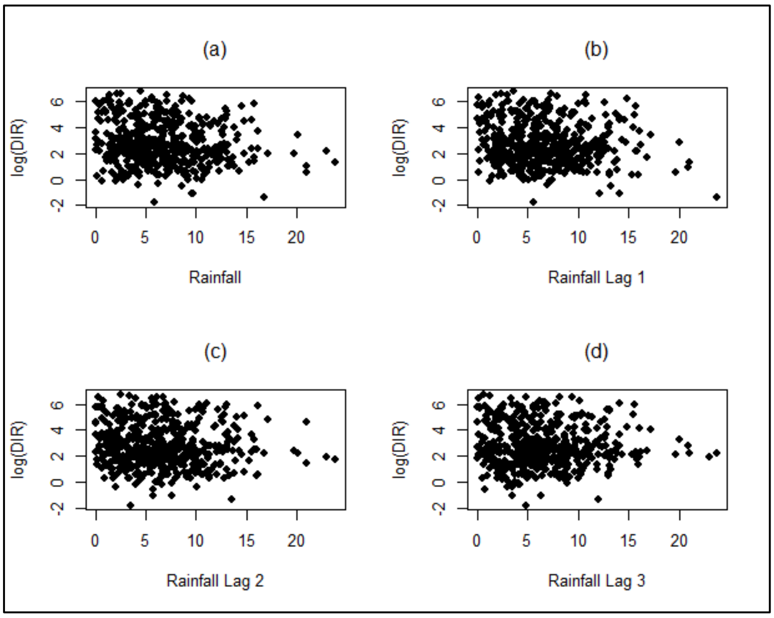

Figure 6: Relationship between log monthly DIR and rainfall

Figure 6 indicates the correlation between the logarithm of monthly DIR and average monthly rainfall with a lag of one to three months in the Petaling, Hulu Langat, Gombak and Klang districts from 2010 to 2015. These plots reveal almost identical patterns for both clusters. With climate factors taken into account, dengue cases can be modelled more accurately. A more frequent occurrence of rain may be followed by more rain throughout the month if it continues to rain. The interaction between the variables will be taken into consideration in future research.

3.4 Poisson GLM and Comparing AIC, BIC and Deviance

The model in Eq. 7 found that a variety of variables had a negative influence on DIR in Petaling, Hulu Langat, Gombak, and Klang. One of the socioeconomic characteristics that had a negative impact on this model was population density. According to this concept, if one unit of population density in a state increases, the DIR decreases. Meanwhile, in Model Cluster A, rainfall lag 1 and rainfall lag 3 were the climatic characteristics that had a negative influence on DIR. The model utilised lag effects developed earlier in the data processing procedure for each climate possibility. The lagged values given by may need to be adjusted in order to take into account dynamic epidemic behaviour, therefore up to 7 months of lagged data was used to create the model.

Table 3: Comparisons of Deviance (D), Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) using Poisson GLM

Model	Deviance	AIC	BIC
1 (Cluster A)	119309	120419	120437
2 (Cluster B)	**11574**	**12518**	**12535**

Accordingly, Table 3 shows a comparison of the Deviance (D) values, the Akaike Information Criterion (AIC) values, and the Bayesian Information Criterion (BIC) values for each cluster. By looking at D, AIC, and BIC, Cluster 2 is the best model as it has the lowest value. This study proved to utilize the best model as shown in Eq.7.

$$DIR = \exp(-0.002298\beta_1 + 0.00924\beta_2 + 0.007101\beta_3 - 0.01075\beta_4) \qquad Eq.7$$

Factors affecting climate and non-climatic factors can be accounted for in Eq.3.1. This study reveals a rise in DIR for the locality number (β_2) when variables other than climate are studied. As an example, variable district was used. After exponentiation of the coefficient value of district Petaling, it was discovered that the value of DIR in Petaling is predicted to rise by 67.6% when compared to the district of Hulu Langat, which serves as a control variable in this model. Nevertheless, there are also other variables like population density (β_1) that show a decreasing value of DIR despite the climate changes. Climate conditions have a huge impact on DIR in Selangor. In the lagged values from lag 1 month, the average monthly amount of rainfall shows positive impact on DIR.

4. Conclusion

This study highlights the use of a clustering technique to produce two sets of clustering group, as well as the ability of merging meteorological and non-climatic parameters in modelling dengue incidence rate in Selangor. A district-level clustering approach was suggested to aid the public health authority in identifying areas where dengue is endemic and taking appropriate time to decide how to manage it. From exploring this dataset, it becomes clear that there are two clusters based on the value of average DIR. Cluster A consists of Petaling and Hulu Langat, where the DIR value exceeds 100 per 100,000 populations, while Cluster B consists of Gombak and Klang, where the DIR value is less than 100 per 100,000 populations. All the objectives of this study are achieved. Plotting the DIR amounts in Petaling, Hulu Langat, Gombak, and Klang was the first objective. DIR values are highest in February and July, while the lowest are in April and September. Following that, this study discovered that the probable meteorological factors that exhibit a significant association to DIR in Selangor include the average monthly quantity of rainfall from the current month up to three months' lag. The number of localities and population density are other non-climatic variables that affect the DIR. The key contribution of this work is the use of a clustering approach to generate two groups of data, as well as the possibility of combining meteorological and non-climatic factors in modelling dengue incidence rate in Selangor, specifically in Petaling, Hulu Langat, Gombak, and Klang. This could provide more work in the future for modelling dengue cases, and possibly other infectious diseases as well.

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] World Health Organization et al., Dengue: guidelines for diagnosis, treatment, prevention and control. World Health Organization, 2009.

[2] F.M.T. Skae, "Dengue fever in Penang," British Medical Journal, vol. ED-2(2185), pp. 1581-1583, 1902.

[3] A. Rudnick et al., "Mosquito-borne haemorrhagic fever in Malaya," British Medical Journal, vol. ED-1(5445), pp. 1269-1272, 1965.

[4] S. Poovaneswari, "Dengue situation in Malaysia," Malaysian Journal Pathology, vol. ED-15(1), pp. 3-7, 1993.

[5] H. Narwani et al., "A review of dengue fever incidence in Kota Bharu, Kelantan, Malaysia during the years 1998-2003," The Southeast Asian Journal of Tropical Medicine and Public Health, vol. ED-36(5), pp. 1179-1186, 2005.

[6] K.T. Ang, et al., "Role of primary care providers in dengue prevention and control in the community," Medical Journal Malaysia, vol. ED-65(1), pp. 58-62, 2010.

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

[8] H. Hassan, "Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia." Geospatial health, pp. 21-25, 2012.

[9] N. Him et al., "Dengue Incidence Rate Clustering by District in Selangor," International Journal of Engineering & Technology, vol. ED-7(4.30), pp. 416-418, 2018, doi:http://dx.doi.org/10.14419/ijet.v7i4.30.22349

[10] A Daniel Keim et al.,"Generalized scatter plots," Information Visualization, vol. ED-9(4), pp. 301–311, 2010, doi:10.1057/ivs.2009.34

[11] M. Friendly and D. Denis, "The early origins and development of the scatterplot," Journal of the History of the Behavioral Sciences, vol. ED-41(2), pp. 103-130, 2005.

[12] J.J. Dziak et al., "Sensitivity and specificity of information criteria," Briefings in bioinformatics, vol. ED-21(2), pp.553-565, 2020.