International Journal of Engineering & Technology, 7 (4.30) (2018) 416-418

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Dengue Incidence Rate Clustering by District in Selangor

Norziha Che Him[1*], Nazeera Mohamad[1], Mohd Saifullah Rusiman[1], Kamil Khalid[1], Muhammad Ammar Shafi[1]

[1]Department of Mathematics & Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia 84600 Pagoh, Muar, Johor, Malaysia
**Corresponding author E-mail: norziha@uthm.edu.my*

Abstract

This study presents the used of Generalised Additive Model (GAM) in modelling Dengue Incidence Rate (DIR) with adopted clustering technique for districts in Selangor. This study identified a pattern for monthly observed dengue count and successfully select variables includes number of rainy days and amount of rainfall with time lags, number of locality and population density which significant to DIR in Selangor. Besides, this study found the districts divided into two clusters based on the value of mean DIR from January 2010 to August 2015. The first cluster consist of 6 districts of Selangor with value of mean DIR from 0 to 200 cases per 100,000 population. Meanwhile, there are 3 districts classified in the second cluster with value of mean DIR from 200 to 500 cases per 100,000 population. The Negative Binomial GAM then adopted in this study to able to handle the presence of overdispersion. In conclusion, clustering technique is one of the effective techniques to identify the different district with the higher potential of dengue risk.

Keywords: *Statistical modelling; Deviance; DIR; Negative binomial; Generalised Additive Model*

1. Introduction

Many infectious diseases are known to be carried by mosquitoes such as chikungunya, dengue, malaria and yellow fever. Dengue as the critical infectious disease since 1986 and Brazil was the first country confirmed with the outbreaks [1]. In Malaysia, DIR has grown severely and now significantly become a major public health concern to authority.

In 1902, [2] was the first person reported the dengue case in Malaysia started in Penang around December 1901 where the dengue cases then increased and expands year by year. This information then leads to the detection of the new area of dengue risk both in rural and urban [3].

Previous researchers reported on the influenced of climate towards DIR [4, 5, 6, 7, 18, 19]. Recent study in China have found duration of sunshine which had a positive relationship with dengue case meanwhile rainfall had a negative relationship towards DIR [8]. But Brazil, India and Malaysia have found a strong positive relationship between rainfall and DIR as reported by [6, 9, 10, 11, 18, 19]. In addition, [12, 18, 19] presents population density and housing condition influenced dengue fever in Ecuador and they also strongly recommended the inclusion of non-climatic factors in the future research.

Therefore, this paper developed a negative binomial GAM for monthly dengue count in Selangor by using a set of data which include climatic and non-climatic variables from January 2010 to August 2015. The clustering technique by each district in Selangor adopted to differentiate the values of DIR to which group of modelling setting.

2. Description of Data and Model Development

2.1. Description of data

In modelling dengue fever incidence in Selangor, monthly number of dengue cases for the period of six years, from January 2010 to August 2015 were obtained from the Ministry of Health Malaysia. There are 9 districts considered which are Gombak, Hulu Langat, Hulu Selangor, Klang, Kuala Langat, Kuala Selangor, Petaling, Sabak Bernam and Sepang. An annual population and population density in each district are obtained from Department of Statistics Malaysia (DOSM). As a record, there are 343,797 dengue cases reported in Selangor during this period of study.

Development of dengue model involving several climatic and non-climatic variables as explanatory variables. The monthly data for mean rainfall and number of rainy days were supplied by the Department of Irrigation and Drainage Malaysia (DDIM).

Dengue Incidence Rate (DIR) refers to the number of new cases of dengue fever which diagnosed in a certain-time period then divided by population per 100,000 [13, 18, 19]. Besides, number of localities refer to the total number of localities affected by dengue in each district of Selangor. Then, the population density is the measurement of population for each district in Selangor per unit area.

2.2. Model development

The increasing of monthly DIR in Selangor believed to relate to the confounding factors including climatic and non-climatic variables [3, 6, 14, 15, 16, 18, 19]. The explanatory data analysis shows the population density, the number of locality, the amount of rainfall with lag up to 3 months, the number of rainy days with lag 0 month to 3 months and the interaction between the amount of

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

rainfall and the number of rainy days shows some relationship with DIR (possibly positive or negative).

A generalised additive model (GAM) framework adopted by [17]. The response variable is the observed dengue cases, y_{dm}, where denotes the district and denotes as month, *m*. Due to the high variability in the monthly dengue counts, the data set, was assumed to follow the negative binomial distribution to handle the presence of overdispersion. The final model could be arranged such as in (1). The final model was referred from the previous study by [18].

$$y_{dm} \sim NegBin\left(\mu_{dm}=p_{dm}\upsilon_{dm},\phi\right) \quad (1)$$

$$\log\mu_{dm} = \log\left(p_{dm}\right)+\log\left(\upsilon_{dm}\right) = \log\left(p_{dm}\right)+\alpha+\sum_{k=1}^{8}\gamma_k z_{kdm}+\gamma_{15}z_{1dm}z_{5dm}+ \sum_{k=2}^{9}\beta_k x_{kdm}+\beta_{14}x_{14dm}+\beta_{16}x_{16dm}+f_d\left(x_{1dm}\right) \quad (2)$$

Here, y_{dm} represents the observed dengue cases for the district, $d\left(d=1,2,3,...,9\right)$ and month, $m\left(m=1,2,3,...,68\right)$ where we considered the μ_{dm} observed dengue cases to be negative binomial distributed where represent the mean value is given by the multiplication of population, p_{dm} and the DIR, υ_{dm} for a given district, d and month, m. The general $\beta_k x_{kdm}$ terms in the (2) has been divided into two different groups. First, the selected non-climatic covariates, $\sum_{k=2}^{9}\beta_k x_{kdm}$, which are referring to a factor of the year, five districts, the number of localities, population density. Meanwhile, $\beta_{14}x_{14dm}$, $\beta_{15}x_{15dm}$ and $\beta_{16}x_{16dm}$ refer to log DIR lagged 1, 2 and 3 months respectively. Secondly, the terms $\sum_{k=1}^{8}\gamma_k z_{kdm}$, represent the selected climatic covariates. The selected climatic covariates are the average amount of rainfall 0 until lag 3 months, an average number of rainy days at current up to lag of 3 months and $\gamma_{14}z_{1dm}$ represents the interaction between the average amount of rainfall and the average number of rainy days and $f_d\left(x_{1dm}\right)$ are smooth function of the calendar month, x_{1dm}

3. Result and discussion

Table 1: The division of the district in Selangor

Cluster 1	Cluster 2
Gombak	Hulu Langat
Hulu Selangor	Petaling
Klang	Sepang
Kuala Langat	
Kuala Selangor	
Sabak Bernam	

The dataset has been divided into two clusters, first cluster contains the district's data with the mean annual DIR from 0 to 200 cases per 100,000 population and the second cluster contains the mean annual DIR from 200 to 500 cases per 100,000 population. In this study, the district of Sabak Bernam, Kuala Selangor, Hulu Selangor, Gombak, Klang and Kuala Langat are considered as Cluster 1, due to the low value of DIR recorded from January 2010 to August 2015. Meanwhile, Cluster 2 refers to the high value of DIR recorded in Selangor with district Hulu Langat, Petaling and Sepang. Table 1 summarises the division of district based on the two clusters earlier.

Table 2: Comparisons of Deviance (D), *Akaike Information Criteria* (AIC) and *Bayesian Information Criteria* (BIC) using negative binomial GAM

Model	Deviance	*AIC*	*BIC*
A (Cluster 1)	484.4734	3246.222	3379.65
B (Cluster 2)	232.6857	2401.617	2481.91

Meanwhile, Table 2 shows the differences of Deviance (D), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for each cluster. The best model is Cluster 2 based on the smallest value of D, AIC and BIC. Therefore, the best potential model proved by this study such as in (3).

$$DIR = \exp\begin{pmatrix} 0.369\beta_2+0.004\beta_3-0.002\beta_4+3.478\beta_{11}- \\ 2.473\beta_{12}+0.004\beta_{14}+0.002\beta_{15}+0.0004\beta_{16}+ \\ 0.150\gamma_1+0.093\gamma_2+0.159\gamma_3+0.033\gamma_4-0.058\gamma_{5-} \\ 0.061\gamma_7-0.083\gamma_7-0.050\gamma_8-0.004\gamma_{15} \end{pmatrix} \quad (3)$$

In (3) can be summarised based on climatic and non-climatic factors. Focus on non-climatic variables, this study shows an increase in the value of DIR for the variables year (β_2), number of locality (β_3), district of Petaling (β_{11}) and DIR with lag 1 month (β_{14}), lag 2 months (β_{15}) and lag 3 months (β_{16}). To make it clearer, we take variables district as an example. After exponent the coefficient value of district Petaling, we found that the value of DIR is expected to increase by 67.6% in Petaling compared to the district of Hulu Langat (β_{13}) which act as a control variable in this model. Meanwhile, there are also several non-climatic variables that show decreasing in the value of DIR. The non-climatic variables are population density (β_4) and district of Sepang (β_{12}). Climatic variables also give significant effect towards DIR in Selangor. The average monthly amount of rainfall with the lagged values from current month (γ_1), lag 1 month (γ_2), lag 2 months (γ_3) and lag 3 months (γ_4) show positive influence towards DIR. For example, we take the variable mean amount of rainfall with the current month and the result shows the current month of rainfall is estimated to positively influence the value of DIR by 16.18%. Meanwhile, the average monthly number of rainy days with the lagged values show negative influence towards DIR. Also, the interaction between the average monthly of rainfall and number of rainy days (γ_{15}) show negative influence towards the value of DIR.

4. Conclusion

The main contribution of this study is the application of clustering process to create two clusters data and the potential of combination between climatic and non-climatic variables in modelling dengue incidence rate in Selangor. The clustering approach also proved by a study conducted in India [20]. They suggested by clustering by district able to help the public health authority to identify the dengue endemic zone and take appropriate time in decision making for disease management. The results from exploratory data analysis show that this study produced two clusters based on the value of average DIR in which Cluster 1 consists of six districts which recorded mean value of DIR from 0 to 200 cases per 100,000 populations meanwhile Cluster 2 consists of three districts which recorded mean value of DIR from 200 to 500 cases per 100,000 populations. Next, this study also found that the potential climatic variables which show significant relationship towards DIR in Selangor are the average monthly amount of rainfall and number of rainy days at the current month up to lagged 3 months. Meanwhile, the non-climatic variables that

significant towards the DIR are the year, month, population density and number of locality. The new approach that needs to be highlighted in this study is the application of clustering process which can be recognised as a new development in dengue modelling in Malaysia, specifically in Selangor. Therefore, it is hoped that this new potential model can help in providing more potential work for modelling dengue cases and perhaps other infectious diseases in the future.

Acknowledgement

The authors would like to express gratefully heartfelt thanks to the Universiti Tun Hussein Onn Malaysia and Office for Research, Innovation, Commercialization and Consultancy Management (ORICC) for the financial support under the TIER 1 research grant (U909).

References

[1] Barbosa GL, Donalisio MR, Stephan C, Lourenco RW, Andrade R, Arduino MDB & Lima VLC (2014), Spatial distribution of the risk of dengue and the entolomogical indicators in Sumaré, State of São Paulo, Brazil. Public Library of Science Neglected Tropical Disease 8(5), 1-9.

[2] Skae FMT (1902), Dengue fever in Penang. The British Medical Journal 2(2185), 1581-1582.

[3] Juni MH, Hayati KS, Cheng CM, Pyang GS, Abd Samad NH & Zainal Abidin ZS (2015), Risk behaviour associated with dengue fever among rural population in Malaysia. International Journal of Public Health and Clinical Sciences 2(1), 114-127.

[4] Wan Fairos WY, Wan Azaki WH, Mohamad Alias Y & Bee Wah Y (2010), Modelling dengue fever (DF) and dengue haemorrhagic fever (DHF) outbreak using Poisson and Negative Binomial model. International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering 4(2), 1-6.

[5] Lowe R, Bailey TC, Stephenson DB, Graham RJ, Coelho CA, Carvalho MS & Barcellos C (2011), Spatio-temporal modelling of climate-sensitive disease risk: towards an early warning system for dengue in Brazil. Computers & Geosciences 37(3), 371-381.

[6] Che Him N, Bailey TC & Stephenson DB (2012), Climate variability and dengue incidence in Malaysia, Proceedings of the 27th International Workshop on Statistical Modelling, Prague 2012, Volume II, 435-440.

[7] Morales I, Salje H, Saha S & Gurley ES (2016), Seasonal distribution and climatic correlates of dengue disease in Dhaka, Bangladesh. The American Journal of Tropical Medicine and Hygiene 9496, 1359-1361.

[8] Xiang J, Hanse A, Liu Q, Tong MX, Sun Y, Cameron S, Hanson-Easey S, Han GS, Williams C, Weinstein P & Bi P (2017), Association between dengue fever incidence and meteorological factors in Guangzhou, China, 2015-2014. Environmental Research 153, 17-26.

[9] Cheong YL, Burkart K, Leitao PJ & Lakes T (2013), Assessing weather effects of dengue disease in Malaysia. International Journal of Environmental Research and Public Health 10, 6319-6334.

[10] Ramachandran VG, Roy P, Das S, Mogha NS & Bansal AK (2016), Empirical model for estimating dengue incidence using temperature, rainfall and relative humidity: A 19-year retrospective analysis in East Delhi. Epidemiology Health 38, e2016052.

[11] Silva FD, Santos AM, Correa RDFCF & Caldas ADJM (2016), Temporal relationship between rainfall, temperature and occurrence of dengue disease in São Luís, Maranhão, Brazil. Ciencia & Saude Coletiva 21(2), 641-646.

[12] Stewart-Ibarra AM, Munoz AG, Ryan SJ, Ayala EB, Borbor-Cordova MJ, Finkelstein JL, Mejia R, Ordonez T, Recalde-Coronel GC & Rivero K (2014), Spatiotemporal clustering, climate periodicity and socio-ecological risk factors for dengue during an outbreak in Machala, Ecuador, in 2010. BioMed Central Infectious Diseases 14(1), 610.

[13] Hassan H, Shohaimi S & Hashim NR (2012), Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia. Geospatial Health 7(1), 21-25.

[14] Colon-Gonzalez FJ, Fezzi C, Lake IR & Hunter PR (2013), The effects of weather and climate change on dengue. Public Library of Science Neglected Tropical Diseases. 7(11), e2503.

[15] Ahmed SA, Siddiqi JS, Quaiser S & Kamal S (2015), Using PCA, Poisson and Negative Binomial Model to study the climate factor and dengue fever outbreak in Lahore. Journal of Basic & Applied Sciences 11, 8-16.

[16] Naim MR, Suhani M, Hod R, Hidayatulfathi O, Idrus S, Norzawati H, Hazrin H, Tahir A, Wen TH, King CC & Zainudin MA (2014), Spatio-temporal analysis for identification of vulnerability to dengue in Seremban district, Malaysia. International Journal of Geoinformatics 10(1), 31-38.

[17] Hastie T & Tibshirani R (1986), Generalized Additive Models, Statistical Science 3(1), 297-310.

[18] Che Him N (2015), Potential for using climate forecasts in spatio-temporal prediction of dengue fever incidence in Malaysia (Doctoral dissertation), Retrieved from https://ore.exeter.ac.uk/repository/handle/10871/23205

[19] Che-Him N, Kamardan MG, Rusiman MS, Sufahani S, Mohamad M and Kamaruddin NK (2018), Spatio-temporal modelling of dengue fever incidence in Malaysia. Journal of Physics Conference Series 995(1):012003, 1-8.

[20] Mutheneni SR, Mopuri R, Naish S, Gunti D and Upadhyayula SM (2018), Spatial distribution and cluster analysis of dengue using self-organizing maps in Andhra Pradesh, India, 2011-2013. Parasite Epidemiology and Control 3(1), 52-61.