

FIXED EFFECTS PANEL COUNT MODEL FOR MODELING DENGUE INCIDENCE

WAN FAIROS WAN YAACOB¹, NIK NUR FATIN FATIHAH SAPRI, AND YAP BEE WAH

ABSTRACT. This study focused on modelling the dengue incidence using Fixed Effects Negative Binomial model on a panel of 9 districts in Selangor, Malaysia covering the period of 2013 to 2017. We examined various climatic factors associated to dengue incidence. Among the factors considered were rainfall, temperature, humidity and its lags of 7 days, 14 days, 21 days and 28 days of climatic factors. Various model specifications were estimated including Fixed Effects Poisson and Fixed Effects Negative Binomial models. Results revealed that there were significant positive relationships between temperature, temperature of lag 7 days and rainfall amount of lag 28 days with dengue incidence rate. However, humidity of lag 21 days was found to be negatively related with dengue incidence rate. The findings of the study can be used for early warning systems by obtaining the information on temperature, humidity and rainfall with its lag of 7, 21 and 28 days respectively.

1. INTRODUCTION

There are numerous methods of statistical model analysis have been used to examine the link between climate variables and dengue fever count including Poisson, Negative Binomial model, logistic, multivariate regression and autoregressive model [1-3]. The Poisson models are widely used in modelling count data type particularly in dengue cases. However, if we have a combination of

¹corresponding author

2010 *Mathematics Subject Classification.* 62J12.

Key words and phrases. Dengue Incidence, Fixed Effects, Negative Binomial, Poisson.

cross section and time series count data (panel count data), the appropriate panel count models to be used are fixed effects and random effects of Poisson or Negative Binomial model. This model can handle the assumption of independent (or at least uncorrelated) observations in the standard Poisson GLMs which may not be valid. Most of the literature on panel count model included fixed effects and random effects model which began with the seminal work of Hausman et al. [4].

A study done by Lowe [5] applied a fixed effects Negative Binomial of GLM for dengue cases in smaller regions of Brazil identified 3-month averaged precipitation and temperature (lag 2 months) and Oceanic Nino Index (lag 6 months) as important climatic covariates in predicting the dengue outbreak in Brazil. Furthermore, Chandrakantha [6] revealed negative binomial regression model showed a significant positive relationship between monthly dengue incidence and the rainfall amount of lag 2 months Sri Lanka. Findings by Ahmed et al. [7] showed that minimum temperature, maximum temperature, humidity and windspeed at lag 3 weeks significantly affected the occurrence of dengue in Karachi. Meanwhile, in the context of dengue study in Malaysia, Yaacob et al. [8] found that NB GLM at lag 14 days and lag 21 days were significant to the incidence of dengue in Putrajaya.

The aim of this study is to examine the association between climatic factors and dengue incidence using panel count model of fixed effects negative binomial regression model. Using this panel count model, significant climatic factors were used in determining the modelling of daily dengue count in the districts of Selangor. This finding can be used for early warning systems in districts of Selangor.

2. METHODOLOGY

2.1. Overview of Panel Count Model. Suppose that we have panel count data for i district and each district observe a total of T times. Let η_{it} be the count variable for individual i at time t . Then the expected value of η_{it} is linked to asset of regressors by:

$$(2.1) \quad \begin{aligned} E(\eta_{it}) &= \tilde{\lambda}_{it} = \exp(d_i + x'_{it}\beta) = \alpha_i \lambda_{it} \\ i &= 1, 2, \dots, N \\ t &= 1, 2, \dots, T, \end{aligned}$$

where d_i is district specific dummies, $\alpha_i = e^{d_i}$ is the individual specific effect, and x_{it} is a vector of regressors. The Fixed Effects Poisson GLM model can be used to estimate the model parameter via conditional maximum likelihood method. However, fitting fixed effects Poisson regression model may also result to overdispersion. Hence, to explicitly allow for overdispersion in panel count model, the fixed effects negative binomial regression model is utilized. In this paper, the fixed effects negative binomial regression model is illustrated using the number of dengue cases which are daily figures for individual district with defined climatic characteristics. When cross sectional heterogeneity exists, the fixed effects model is more appropriate. Based on the fixed effects negative binomial formulation of Hausman et al. [1], the dengue model is expressed as $\log \lambda_{it} = \alpha_i + \beta x_{it}$ $i = 1, 2, \dots, n$ and $t = 1, 2, \dots, T$ where λ_{it} is the expected number of dengue in district i in day t , α_i is the fixed effects associated with i , and β is vectors of parameters to be estimated for the vector of explanatory variables x_{it} . Here, the number of dengue, y_{it} for a given time period, t is assumed to follow a negative binomial distribution with parameters $\alpha_i \lambda_{it}$ and ϕ_i , where $\lambda_{it} = \exp(x'_{it}\beta)$ gives y_{it} as mean $\alpha_i \lambda_{it} / \phi_i$ and variance $(\alpha_i \lambda_{it} / \phi_i) \times (1 + \alpha_i / \phi_i)$. This model allows the variance to be greater than the mean. The parameter α_i is the individual-specific fixed effects and the parameter ϕ_i is the negative binomial overdispersion parameter which can take on any value and varies across individuals. The negative binomial mass function is given by:

$$(2.2) \quad f(y_{it}) = \frac{\Gamma(\lambda_{it} + y_{it})}{\Gamma(\lambda_{it}) \Gamma(y_{it} + 1)} \left(\frac{1}{1 + \frac{\alpha_i}{\phi_i}} \right)^{\lambda_{it}} \left(\frac{\frac{\alpha_i}{\phi_i}}{\alpha_i} \right)^{y_{it}}.$$

2.2. Goodness of Fit Model. There were 4 different time lag effects being considered in the study which included lag of 7 days, 14 days, 21 days and 28 days. To select appropriate prediction model fit, each model with and without lag of climatic factors was evaluated using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), deviance and (pseudo-R²). The AIC equation is given as follow:

$$(2.3) \quad AIC = -2 \ln(L) + 2k,$$

where L is the log likelihood of the fitted model and k is the number of parameters in the fitted model. Meanwhile, BIC is computed as follows:

$$(2.4) \quad BIC = -2\ln(L) + k \log(n),$$

where n is the sample size. In conventional generalized linear modelling with fixed effects, the deviance, D is an important measure that should be put into consideration [9]. Deviance is the difference between the log-likelihood of the fitted model and saturated model which can be specified as follows:

$$(2.5) \quad D = 2[\ln(\hat{L}) - \ln(L)]$$

where $\ln(\hat{L})$ is the log-likelihood for saturated model and $\ln(L)$ is the log-likelihood for fitted model. Model which is correctly specified will be close to unity, 1. If the value of D is close to $n - p$ (degrees of freedom), the model is considered as adequate. However, if $D > n - p$, then the data is overdispersed which means negative binomial model can be considered to fix such problem.

Since the study proposes using 'linear' model, the proportion of variation in the response variable can be explained by explanatory variables, R^2 which is a typical measure used as a model adequacy. As suggested in the study of Lowe [9], the R^2 is based on decomposition of the deviance which is termed as 'pseudo- R^2 ':

$$(2.6) \quad D' = D + \hat{D},$$

where D' is the deviance in the intercept-only model (null model), D is the deviance of fitted model and \hat{D} is explained deviance. Thus, the pseudo- R^2 , (R_D^2) is specified as follows:

$$(2.7) \quad R_D^2 = 1 - \frac{D'}{\hat{D}},$$

where pseudo- R^2 measures the reduction in deviance due to inclusion of explanatory variables. The interpretation of R_D^2 is similar to common R^2 at which the values closer to 1 indicate a better fit. The next section of this paper is the description of the data used in the study.

2.3. The Data. In 2019, population in Selangor is estimated at 663 million [9]. According to iDengue [10], Selangor is recognized as dengue endemic region as it consistently contributes the highest number of dengue for the past 10 years

compared to other states in Malaysia. Figure 1 presents the annual trend of dengue incidence rate per 100,000 population by district in Selangor. Dengue count (DC) were obtained from Vector Borne and Infectious Diseases Sector (VBIDS), Ministry of Health (MOH) Malaysia located in Putrajaya through appropriate procedure which refers to the daily number of dengue cases in 9 districts of Selangor namely Gombak, Hulu Langat, Hulu Selangor, Klang, Kuala Langat, Kuala Selangor, Petaling, Sabak Bernam and Sepang that comprises a period between 1st January 2013 to 31st December 2017. This study also used several climatic factors (average temperature, relative humidity and the amount of rainfall) with lag time effects (lag 7, 14, 21 and 28 days) as the explanatory variables. The lag time effect was taken into consideration in the study as biologically, the eggs take a certain period of time to emerge into an adult mosquito and the influence of climatic surrounding is expected to be visible within one or two months later. The source of climatic variables was collected from Climate Data Online (<https://power.larc.nasa.gov/data-access-viewer/>). The list of variables used in the study are summarized and tabulated in Table 1.

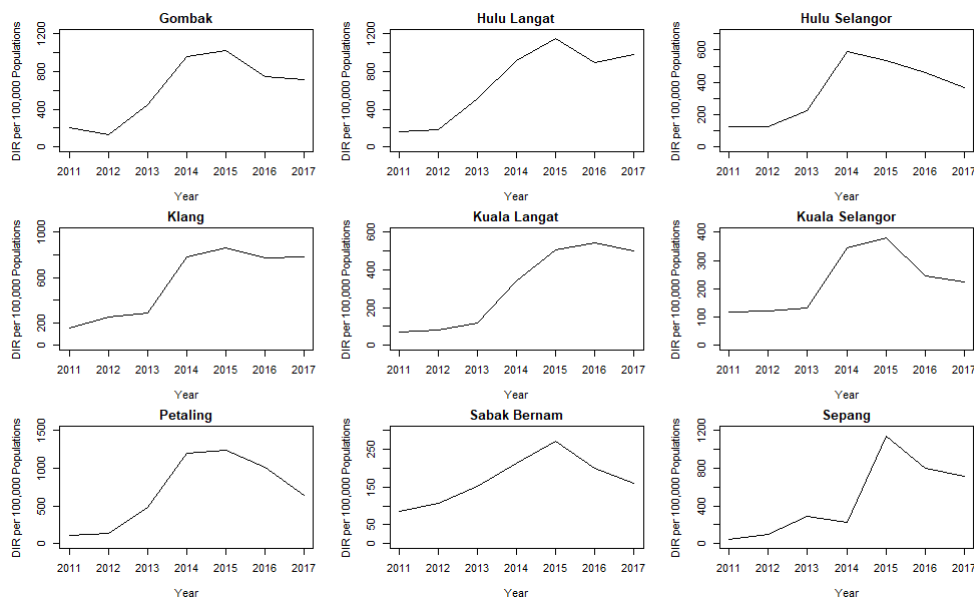


FIGURE 1. Annual Dengue Incidence Rate per 100, 000 Populations

TABLE 1. List of Variables

Variables	Variable Label	Measurement Units
Temperature	Temp	°C
Rainfall	Rain	mm
Relative Humidity	Humid	%

3. RESULTS AND DISCUSSIONS

Several fixed effects negative binomial models were developed and fitted under the study with several selections of climatic covariates. Then, the significance climatic covariates were identified for the purpose of modelling daily dengue count in Selangor. The models were developed for 5 years' time period (1st January 2013 to 31st December 2017). Table 2 summarizes the parameter estimates for fixed effects negative binomial model analyzed by districts of Selangor.

Table 2 provides a series of fixed effects negative binomial model and fixed effects Poisson model analyzed for 9 districts in Selangor. It can be seen that the fixed effect negative binomial model was preferred over the fixed effect Poisson model. Note that, the factor effect for day is not reported in the table. The baseline district is Gombak (reported) and the baseline day is Day 1 (not reported). At 0.05 level of significance, variable temperature, temperature of lag 7 days, humidity of lag 21 days and rainfall of lag 28 days were found to be statistically significant. It was also found that there was a statistically significant positive relationship between temperature and temperature of lag 7 days with dengue incidence rate. Besides that, a significant negative relationship was found between humidity of lag 21 days and dengue incidence rate while a significant positive relationship was found between the rainfall amount of lag 28 days and dengue incidence rate. These findings suggested that warm, humid and the amount of rainfall had promoted the mosquitoes' development, the biting rate activity and creating mosquitoes' breeding sites. This finding can be used for early warning systems by obtaining the information on temperature, humidity and rainfall with its lag of 7, 21 and 28 days respectively.

Meanwhile, Figure 2 reports the ability of the fixed effects negative binomial model to predict dengue incidence rate in 9 districts of Selangor. It can be seen that the model was able to fit correctly in some districts with very high

dengue incidence rate such as Gombak, Hulu Langat, Klang and Petaling within 2014 to 2017. However, the dengue incidence rate was under-predicted in Hulu Selangor when serious epidemic occurred in 2014 and 2015. A very high dengue incidence rate in Kuala Langat was spotted to be under-predicted of high dengue incidence rate and low dengue incidence rate in 2016 and 2017 respectively (see Figure 2). LR Test: $-2(\text{LLPoisson} - \text{LLNegative Binomial}) = -2(-86906.12 - (-48675.52)) = 76461.2$

TABLE 2. Fixed Effects NB and Poisson GLM

	Full Model (FENB)	FE NB	FE Poisson
Constant	-1.326e+01*** (8.774e-01)	-1.280e+01*** (4.465e-01)	-1.200e+01*** (1.074e-01)
Temp	3.853e-02* (1.650e-02)	3.600e-02** (1.185e-02)	3.527e-02*** (2.692e-03)
Temp_lag7	4.334e-02* (1.688e-02)	3.789e-02** (1.185e-02)	1.863e-02*** (2.683e-03)
Temp_lag14	2.161e-02 (1.704e-02)		
Temp_lag21	-6.018e-03 (1.691e-02)		
Temp_lag28	-5.054e-03 (1.645e-02)		
Humid	1.140e-03 (2.809e-03)		
Humid_lag7	2.679e-03 (2.855e-03)		
Humid_lag14	-8.259e-04 (2.890e-03)		
Humid_lag21	-7.998e-03 (2.863e-03)	-7.921e-03*** (1.950e-03)	-7.685e-03*** (4.416e-04)
Humid_lag28	-3.254e-03 (2.817e-03)		
Rain	5.319e-04 (6.042e-04)		
Rain_lag7	6.545e-04 (6.043e-04)		
Rain_lag14	1.165e-03 (5.998e-04)		
Rain_lag21	5.933e-04 (6.032e-04)		
Rain_lag28	1.316e-03* (5.964e-04)	1.183e-03* (5.882e-04)	5.133e-04*** (1.481e-04)
Hulu Langat	1.477e-01*** (3.058e-02)	1.436e-01*** (3.055e-02)	1.405e-01*** (7.135e-03)
Hulu Selangor	-5.448e-01*** (3.567e-02)	-5.587e-01*** (3.466e-02)	-5.389e-01*** (1.559e-02)
Klang	-4.086e-01*** (5.704e-02)	-3.645e-01*** (4.754e-02)	-2.926e-01*** (1.149e-02)
Kuala Langat	-8.208e-01*** (4.040e-02)	-8.052e-01*** (3.743e-02)	-7.363e-01*** (1.559e-02)
Kuala Selangor	-1.312e+00*** (5.168e-02)	-1.285e+00*** (4.496e-02)	-1.208e+00*** (1.982e-02)
Petaling	-1.249e-01* (5.423e-02)	-7.818e-02* (4.676e-02)	-2.730e-02* (1.065e-02)
Sabak Bernam	-1.640e+00*** (6.070e-02)	-1.606e+00*** (5.382e-02)	-1.530e+00*** (3.065e-02)
Sepang	-4.707e-01*** (5.149e-02)	-4.380e-01*** (4.553e-02)	-3.657e-01*** (1.478e-02)
Deviance	19342	19344	124228
R_D^2	0.177424	0.176822	0.183173
AIC	98114.91	98107.04	174566.2
BIC	101012.6	101012.4	177471.6
Likelihood Value		-48675.52	-86906.12

4. CONCLUSIONS

Dengue incidence is alarming in Selangor for the past 10 years. The purpose of this study was to model the relationship between climatic factors and dengue



FIGURE 2. Observed and Model Fit DIR (Selangor)

incidence in Selangor. Panel count data was used which pooled 9 districts of Selangor for daily dengue data of 5 years (1st January 2013 to 31st December 2017). In order to cater the overdispersion problem, fixed effects negative binomial model was chosen over fixed effects Poisson model in predicting the dengue incidence rate in Selangor. The findings of the study revealed that the climatic factors of current temperature, temperature at lag 7 days, humidity at lag 21 days and rainfall at lag 28 days were found to be statistically significant in predicting the dengue incidence in Selangor. Furthermore, the spatial map produced in the study was also able to reveal the dengue endemic districts or can be known as dengue hotspot areas which were Klang, Kuala Langat, Gombak, Hulu Langat, Petaling and Sepang. Therefore, the findings obtained in the study are hopefully able to give important information especially to the local

authorities and public health institution to take necessary steps to safeguard the community from dengue outbreak.

ACKNOWLEDGMENT

We would like to express our appreciation and record our gratitude to the Ministry of Higher Education (MOHE) and Universiti Teknologi MARA for financial support under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2016/STG06/UITM/02/2). The authors are also grateful to the Editor, the Associate Editor and anonymous referees for their insightful comments and suggestions.

REFERENCES

- [1] R. TULADHAR, A. SINGH, A. VARMA, D.K. CHOUDHARY: *Climate Factors Influencing Dengue Incidence in an Epidemic Area of Nepal*, BMC Res Notes., **12**(131) (2019). <https://doi.org/10.1186/s13104-019-4185-4>
- [2] N.C. HIM, T.C. BAILEY, D.B. STEPHENSON: *Climate Variability and Dengue Incidence in Malaysia*, 27th International Workshop on Statistical Modelling, **2** (2012), 435–440. <https://iopscience.iop.org/article/10.1088/1742-6596/995/1/012003/pdf>
- [3] S. ATIQUE, S.S. ABDUL, C.H. HSU, T.W. CHUANG: *Meteorological Influences on Dengue Transmission in Pakistan*, Asian Pacific Journal of Tropical Medicine, **9**(10) (2016), 954–961. <http://dx.doi.org/10.1016/j.apjtm.2016.07.033>
- [4] J.A. HAUSMAN, B.H. JUY HALL, Z. GRILICHES: *Econometric Models for Count Data with an Application to the Patents - R&D Relationship*, Econometrica, **52**(4) (1984), 909–938.
- [5] R. LOWE : *Spatio-temporal Modelling of Climate Sensitive Disease Risk: Towards an Early Warning System for Dengue in Brazil*, Ph.D. thesis, University of Exeter, United Kingdom, (2010).
- [6] L. CHANDRAKANTHA : *Statistical Analysis of Climate Factors Influencing Dengue Incidences in Colombo, Sri Lanka: Poisson and Negative Binomial Regression Approach*, International Journal of Scientific and Research Publications, **9**(2) (2019), 133–144. <http://dx.doi.org/10.29322/IJSRP.9.02.2019.p8616>.
- [7] S.A. AHMED, J.S. SIDDIQI, Q. SABAH, A.A. SIDDIQUI: *Analysis of Climate Structure with Karachi Dengue Outbreak*, Journal of Basic and Applied Science. **11**(2015), 544–552.
- [8] W.F.W. YAACOB, W.W. AZAKI, M.A. LAZIM, Y.B. WAH: *Modelling Dengue Fever (DF) and Dengue Hemorrhagic Fever (DHF) Outbreak Using Poisson and Negative Binomial Model*, World Academy of Science, Engineering and Technology, **38** (2010), 903–908.
- [9] DEPARTMENT OF STATISTICS MALAYSIA: *Malaysia at a Glance*. (2019). Available: <http://www.dosm.gov.my>
- [10] IDENGUE.: *Statistik Denggi*, (2019), <https://idengue.mysa.gov.my/>

FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES
UNIVERSITI TEKNOLOGI MARA CAWANGAN KELANTAN, KAMPUS KOTA BHARU
LEMBAH SIREH, 15050 KOTA BHARU, KELANTAN, MALAYSIA
Email address: wnfairos@uitm.edu.my

FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES
UNIVERSITI TEKNOLOGI MARA, .
45400 SHAH ALAM, SELANGOR, MALAYSIA.
Email address: fatinfatihah26@yahoo.com

FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES
UNIVERSITI TEKNOLOGI MARA, .
45400 SHAH ALAM, SELANGOR, MALAYSIA.
Email address: beewah@tmsk.uitm.edu.my