© 2019 IJSRSET | Volume 6 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : https://doi.org/10.32628/IJSRSET196124

Modelling the Number of Cases of Dengue Hemorragic Fever with Mixed Geographically Negative Binomial Regression in West Java Province

Zulfanita Dien R[1], Asep Saefuddin[2], Anik Djuraidah[3]
[1,2,3]Department of Statistics, Bogor Agricultural University, Bogor, West Java, Indonesia

ABSTRACT
Dengue hemorrhagic fever (DHF) is an infectious disease caused by the dengue virus of the genus Flavivirus, which is transmitted by the bite of the Aedes Aegypti. Different regional demographics cause the number of DHF cases to differ in each region following by environmental conditions in the area. The model applied is GWNBR (Geographically Weigthed Negative Binomial Regression) due to count data outcome affected by geographical effect. In some instances not all in the GWNBR model have spatial effects, sometimes the estimate parameter are constant, so the GWNBR model can be developed using a mixed model to become MGWNBR. Determination of global and local parameters using the confidence interval. This study aims to analyze the factors that influence the number of dengue cases in West Java Province in 2015 using the MGWNBR approach. Based on the comparison of AIC values, the MGWNBR model has a smaller AIC value compared to the negative binomial regression model. The variables that significant globally are population density (X_1) and health worker (X_2)The variables that significant locally are number of health facilities (X_3) PHBS (X_4) and healthy homes (X_5)

Keywords: Dengue hemorrhagic fever, GWNBR, MGWNBR

I. INTRODUCTION

Dengue hemorrhagic fever (DHF) is an infectious disease caused by the dengue virus of the genus *Flavivirus*, family of *Flaviviridae* which is transmitted by Aedes Aegypti (Kemenkes 2010). DHF is one of the main health problems in Indonesia. West Java Province is the province that has the highest number of dengue cases in Indonesia, which is 21237 cases with 14 cases of death (Ministry of Health 2015). The number of these cases increased by 14.6 percent from the previous year which was as much as 18116 (Kemenkes 2014). Dengue cases in West Province are increasing every year, prevention efforts need to be done by knowing the factors that influence the development of the disease.

The number of DHF cases is one example of a count event so that the right modeling is used to determine the factors that influence the development of DHF cases namely Poisson regression. Poisson regression is one of the statistical methods that is used to find out the relationship between the variables in the form of data with one or more explanatory variables. However, sometimes in the application of Poisson regression overdispersion phenomena are found. Overdispersion occurs when the variance value is greater than the average value. If the Poisson model with overdispersion is still used in the data it will result in the occurrence of bias parameters (Evadianti and Purhadi 2014), so that one alternative modeling solution that can be used is the negative binomial distribution. The binomial negative model can overcome the problem of overdispersion in Poisson regression because it has dispersion parameters that can explain the variability in the data (Hilbe 2011).

the spread of DHF tends to cluster (Ruliansyah *et al.* 2017). According to Sukendra and Kusuma (2016) infectious diseases are not influenced by administrative boundaries so it can be said that the spread of dengue disease transmitted through Aedes Aegypti can spread to areas to areas affected by DHF. Different regional demographics cause the number of DHF cases to differ in each region following the environmental conditions in the area. The right analysis is used to determine the factors that influence the development of dengue disease by taking into account the geographical location and have a response variable in the form of count data, namely negative geographical binomial weighted regression analysis (GWNBR).

GWNBR is a weighted statistical method with variable variables in the form of count data used to estimate parameters that have spatial diversity so that each observation location has different estimator values. In some cases not all estimators in the GWNBR model have spatial effects, sometimes the explanatory variables are constant. Nakaya *et al.* (2005) developed a mixed model so that there are two parameters, namely local and global. So that the GWNBR model can be developed using a mixed model into mixed GWNBR. The parameter estimation in the GWNBR mixture combines parametric methods to predict global parameters and nonparametric estimates to predict local parameters. There are two methods that can be used to determine global and local parameters, namely the interval of confidence (Pongoh 2015) and the linear model of coregionalization (Mar'ah et al. 2017). The purpose of this study was to analyze the factors that influence the increase in the number of dengue cases in West Java Province in 2015 with the mixed GWNBR approach.

II. METHODS AND MATERIAL

The data used for this study are secondary data obtained from the West Java Provincial Statistics Agency and the West Java Provincial Health Office in 2015. The unit of analysis of this study is as many as 27 districts / cities in West Java Province.

Table 1: Response and Explanatory Variables

Variables	Explanation
Y	Number of DHF Cases
X_1	Population density
X_2	Number of Health Workers
X_3	Number of Health Facilities
X_4	Percentage of Healthy and Clean Houses
X_5	Percentage of Healthy Houses

Analysis and modeling in this study were assisted by statistical programs, R (3.4.1). The steps used to analyze data on the number of dengue cases in infants in West Java Province in 2015 using the mixed GWNBR approach are as follows:

1. Describe the characteristics of the number of dengue cases in West Java Province in 2015

2. Estimating a negative binomial regression model
   Binomial negative regression is one of the statistical methods used to overcome the overdispersion in the Poisson distribution. The parameter estimation is a negative binomial regression model with the following formula:
   $$ln(\mu_i) = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} \quad (1)$$
3. Multicollinearity Test
   Multicollinearity testing can be done by looking at the VIF value (variance inflation factor) to see

whether or not there is multicollinearity. The VIF formula is as follows:

$$VIF = \frac{1}{1-R_k^2} \quad (2)$$

with R_k^2 is the coefficient of determination of the explanatory variable for $k = 1,2,\dots,n$. If the VIF value is> 10, the assumption of multicollinearity is not fulfilled.

4. spatial dependencies Test

   Testing spatial dependencies using Moran I. The hypothesis used to test spatial dependencies is as follows:

   $H_0 : I = 0$

   $H_1 : I \neq 0$

   The Moran I test statistics are as follows:

   $$Z_{hit} = \frac{I-E(I)}{\sqrt{var(I)}} \quad (3)$$

   The calculation of Moran I is based on a normal distribution, so for decision making that is if the value of Moran I is more than $z\alpha/_2$ so it will reject H_0.

5. Spatial heterogeneity Test

   Testing of spatial heterogeneity can be done using the Breusch-Pagan test (Anselin 1988). The hypothesis used for testing spatial heterogeneity is as follows:

   $H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_j = 0$

   $H_1 : at\ least\ one\ \sigma_j \neq 0$

   with statistic test as follows:

   $BP = \left(\frac{1}{2}\right) f'Z(Z'Z)^{-1}Z'f \sim \mathrm{X}^2(p)$

   with $f_i = \left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right)$

   Decision making in the Pagan Breusch test is if the value of $BP > \mathrm{X}^2_{(\alpha;p)}$ so it will reject H_0

6. Estimating the mixed GWNBR model

   a. Determine the optimum smoothing parameters and spatial weight

   One method that can be used to select the optimum smoothing parameter is cross validation.

   $CV(h) = \sum_{i=1}^{n}(y_i - \widehat{y_{\neq \iota}}(h))^2$

   with , n adalah number of observation, y_i is the location response variable i, $\widehat{y_{\neq \iota}}(h)$ is the estimated value of observation of the location of i whose value is obtained without involving the observation of the location of the i itself.

Fotheringham *et al.* (2002) state that there are two spatial weighting, namely a fixed spatial kernel and an adaptive spatial kernel. The spatial weighting function aims to estimate the parameter values of each observation location.

i. Gaussian Kernel Function

$$w_{ij} = exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right]$$

with d_{ij} is the distance between locations, distance used is euclidean distance, $d_{ij} = \sqrt{\sum_{k=1}^{m}\left(x_{ik} - x_{jk}\right)^2}$ and b is optimum smoothing parameter.

ii. Function of Kernel Bisquare

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2, jika\ d_{ij} < b \\ 0, jika\ d_{ij} \geq b \end{cases}$$

b. Estimate the GWNBR model

   Estimating parameters of the GWNBR model uses the following formula

   $$\boldsymbol{\beta}^{(l+1)}(u_i,v_j) = \left(\boldsymbol{X'W}(u_i,v_j)\boldsymbol{A}(u_i,v_j)^{(l)}\boldsymbol{X}\right)^{-1}\boldsymbol{X'W}(u_i,v_j)\boldsymbol{A}(u_i,v_j)^{(l)}\boldsymbol{z}(u_i,v_j)^{(l)}$$

After getting the GWNBR model then look for the confidence interval for each parameter using the following formula:

$$\beta_j \pm t_{\alpha/2} s\sqrt{c_{jj}}$$

Criteria is that if the confidence interval has a percentage above 65 percent then the variable can be used as a global parameter and if the percentage is less than 65 percent then the variable can be used as a local parameter.

c. Predict the mixed GWNBR model parameters
Algorithm for estimating the mixed GWNBR model as follows:

i. Initiation γ (which comes from a negative binomial regression model)
ii. Using GWNBR for errors adjusted as estimating local parameters and calculating smoother **S**
iii. Use the formula as follows:

$$\boldsymbol{\gamma}^{(l+1)} = \left(\boldsymbol{X}'_{par}\boldsymbol{A}^{(l)}\boldsymbol{X}_{par}\right)^{-1}\boldsymbol{X}'_{par}\boldsymbol{A}^{(l)}\left(\boldsymbol{I}-\boldsymbol{S}^{(l)}\right)\boldsymbol{z}^{(l)}$$

to get global parameters.

iv. Iteration is carried out until the convergence is reached.
v. Choose the best model with the AIC criteria
The selection of the best model uses the Akaike Information Criteria with the following formula:

$$AIC = -2log_e\left(L(\hat{\theta}|data)\right) + 2K$$

with $log_e\left(L(\hat{\theta}|data)\right)$ is the maximum possible maximum estimation value, θ is an unknown parameter, K is the number of parameters estimated, and n is the number of observations

7 Interpret the results of the best models

III.RESULTS AND DISCUSSION

West Java Province is the province that has the highest number of dengue cases in Indonesia. Districts that have the highest number of dengue cases are Bandung Regency with a total of 3640 cases, while Districts which have the least number are Pangandaran Regency with 29 cases. The map of the number of dengue cases in West Java Province is shown in Figure 1.

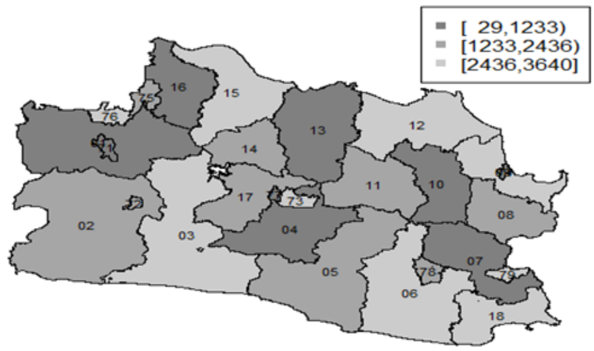

Figure 1: Number of dengue cases in West Java in 2015

Districts that have the highest number of DHF cases is Kabupaten Bandung, have the highest population density of 14750.45 inhabitants / km^2, while Districts with the lowest number of DHF cases is Pangandaran Regency, have the lowest population density of 386.62 inhabitants / km^2.

Based on the estimation results, the negative binomial regression model is obtained by estimating the parameters presented in Table 2.

Table 2 : Estimator parameters for negative binomial regression models

Parameter	Estimate	P-value
β_0	5.063	<2e-16
β_1	-2.05 x 10^{-6}	0.953
β_2	0.18	1.83 x 10^{-4}
β_3	0.031	2.41e-9
β_4	0.011	0.31
β_5	-0.018	0.013
Deviance = 28.277 Df = 21 AIC = 397.5		

Based on the estimation results for negative binomial regression in Table 2 above, simultaneous and partial testing can be carried out to model the number of dengue cases in West Java Province. Simultaneous testing of the parameter estimators modeled the number of dengue cases in West Java Province using 5% level , so that the values of ${\chi_{(4;0.05)}}^2$ is 9.48. Value of χ^2 is less than $D(\bar{\beta})$ which is 28.277, so that there is enough proof to state that there is at least one explanatory variable that significant 5% level. Partial testing of parameter estimators modeled the number of DHF cases in West Java Province with 5% level obtained values $z_{(0.05/2)}$ is 1.96. The value of z Table is less thani z_{hitung} in each explanatory variable, so that there is proof evidence to state that the explanatory variable significant at 5% level. The explanatory variables that have a significant effect are health workers, the number of health facilities, and healthy homes.

Multicolinerity testing is done by calculating the VIF value (variance inflation factor) in each explanatory variable. If the VIF value is> greater than 10, it can be assumed that there is a correlation between the explanatory variables. In Table 3 it can be seen that the VIF value in the explanatory variable is less than 10 so it can be concluded that between explanatory variables are not correlated for each variables.

Table 3: VIF values for explanatory variables

Explanatory Variables	VIF Values
X1	2.93
X2	2.98
X3	1.33
X4	1.46
X5	1.49

Spatial dependency is indicated by the existence of interrelationships between regions. Testing of spatial dependencies is done by calculating the value of the Moran Index. Testing of spatial heterogeneity (spatial diversity) is done by calculating the value of Breusch-Pagan. The results of testing the spatial assumptions are presented in Table 4.

Table 4: Testing of spatial assumptions

Dependency Test		Heterogeneity Test	
Moran Index	-0.63	BP values	104.61
$z_{statistic}$	-6.67	$\chi^2_{(5;0.05)}$	11.070
$z_{0.05/2}$	1.96		

GWNBR produces different regression parameter values at each observation location. The first step taken in GWNBR modeling is to determine the kernel weighting function used to produce the best model with the model selection criteria using the AIC value.

Table 5 : Selection of the best kernel weighting functions

Model	AIC
GWNBR Fix *Bisquare*	309.95
GWNBR Fix *Gauss*	350.70
GWNBR Adaptive *Bisquare*	255.40
GWNBR Adaptive G*auss*	339.82

Based on Table 5, it can be seen that GWNBR Adaptive Bisquare model has the smallest AIC value. The results of the model suitability test using the F test produce a calculated F value of 16.28 with a level of 5% resulting in a value $F_{(0.05;21;21)}$ amounting to 2.08 resulting in the conclusion that there is enough proof to state that there is a difference between the negative binomial regression model and the GWNBR model.

The GWNBR model has explanatory variables that affect each district / city. The map of explanatory variables for each Regency / City is presented in Figure 2. The closeness between regions between districts / cities affects the explanatory variables that influence the region.

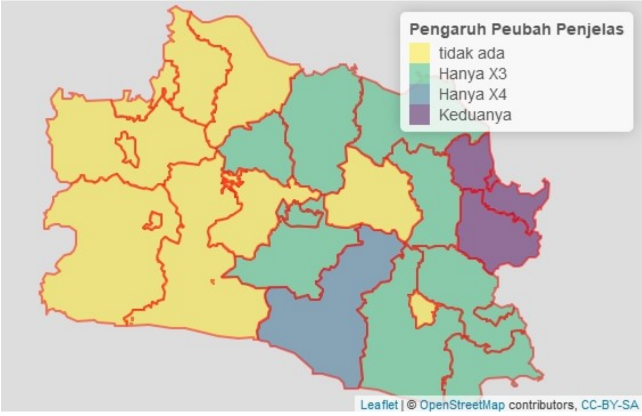

Figure 2 : The sigfinicant variable in the GWNBR model

The selection of global and local parameters can be done using the confidence interval. The selection criteria using a confidence interval, if the percentage of the confidence interval is more than 65%, can be grouped into global parameters. Variable X_1,X_2 each has a percentage of more than 65%, 85.18% and 74.07%, so that these variables can be grouped into global variables. Variables X_3, X_4, X_5 each has a percentage of 29.62%, 62.96% and 62.96% so that it can be grouped into local variables. The division of global and local variables is used to estimate the parameters in the Mixed Geographical Binomial Negative Regression model(GWNBR).

Mixed GWNBR is a model that combines global and local parameters in one model. The mixed GWNBR model has different parameter estimation values in each field. Comparison of the AIC values of the negative binomial regression model and the mixed GWNBR model based on the AIC values are presented in Table 6.

Table 6 : Comparison of the AIC values of the negative binomial regression model and the mixed GWNBR model

Model	AIC
Negative binomial regression	397.50
Mixed GWBNR	286.57

The mixed GWNBR model has an AIC value smaller than the negative binomial regression model so that the spatial model is better used. In addition, the mixed GWNBR model has more variables that influence each Regency / City than the GWNBR model. The map of explanatory variables that influence the mixed GWNBR model is presented in Figure 3.

One of the factors that influence the incidence of DHF (Kusuma and Sukendra 2016; Muliansyah and Baskoro 2016) is population density. Uncontrolled aspects of mobility and urbanization are factors that increase population, especially in urban areas and industrial estates. This condition has resulted in the creation of a conducive environment for the breeding of dengue vectors so that it will cause an increase in the number of dengue cases (Muliansyah and Baskoro 2016). This is because urban areas and industrial estates have a settlement layout that coincides with each other so that it will facilitate the transmission process (Ashlilihah *et al.* 2015). Another aspect that affects the incidence of DHF is clean and healthy living behavior and healthy homes (Candra 2010; Ashlilihah *et al.* 2016). According to Respati *et al.* (2017) said that basic sanitation can be used to determine risk factors for dengue cases.

Based on the variables that influence the GWNBR and mixed GWBNR models, it can be explained that the variables in the mixed GWNBR model provide more extensive information than the GWNBR model. The influential variables in the mixed GWNBR model not only pay attention to the variables in each region but also consider variables that influence globally.

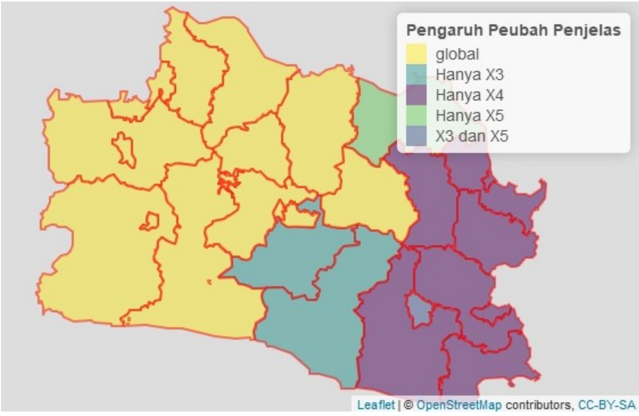

Figure 3: The sigfinicant variable in the mixed GWNBR model

IV.CONCLUSION

Based on the results of the study it can be seen that the mixed GWNBR model provides a good model of the number of dengue cases in West Java Province in 2015 compared to the negative binomial regression model. The variables that influence the mixed GWNBR model are population density (X_1) and health workers (X_2) influential globally. Variable number of health facilities (X_3), PHBS (X_4), and healthy houses (X_5) influential local.

V. REFERENCES

[1] [BPS] Badan Pusat Statistik Provinsi Jawa Barat. 2016. Provinsi Jawa Barat Dalam Angka. Pemerintah Provinsi Jawa Barat.

[2] [DINKES] Dinas Kesehatan Provinsi Jawa Barat. 2015. Profil Kesehatan Provinsi Jawa Barat. BPS Provinsi Jawa Barat.

[3] Evadianti E, Purhadi. 2014. Pemodelan Jumlah Kematian Ibu di Jawa Timur dengan Pendekatan Geographically Weigthed Negative Binomial Regression. *Jurnal Seni dan Sains ITS*. Vol 3 (2): 2337-3539.

[4] Fotheringham A, Brunsdon C, Charlton M. 2002. *Geographically Weighted Regression*. Chichester (UK) : John Wiley & Sons Ltd.

[5] Hilbe MJ. 2011. *Negative Binomial Regression:2nd*. New York (USA): Cambridge University Press.

[6] [Kemenkes] Kementrian Kesehatan Indonesia. 2014. Profil Kesehatan Indonesia 2014. Jakarta : Kemetrian Kesehatan Republik Indonesia.

[7] [Kemenkes] Kementrian Kesehatan Indonesia. 2015. Profil Kesehatan Indonesia 2015. Jakarta : Kemetrian Kesehatan Republik Indonesia.

[8] Kusuma AP, Sukendra DM. 2016. Analisis Spasial Kejadian DBD Berdasarkan Kepadatan Penduduk. *Unnes Journal of Public Health. 5(2016).*

[9] Mar'ah Z, Djuraidah A, Wigena AH. 2017. Pemodelan Regresi Terboboti Geografi Semiparametrik dengan Model Koregionalisasi. *International Journal and Sciences: Basic and Applied Research.* 34(2017):178-186.

[10] Muliansyah, Baskoro T. 2016. Analisis Pola Sebaran Demam Berdarah Dengue Terhadap Penggunaan Lahan Dengan Pendekatan Spasial di Kabupaten Banggai Provinsi Sulawesi Tengah Tahun 2011-2013. *Journal of Information Systems for Public Health*. 1(2016):47-54.

[11] Nakaya T, Fotheringham AS, Brunsdon C. 2005. Geographically Weigthed Poisson for disease association mapping. *Statist. Med*. 24(2005): 2696-2717.

[12] Pongoh F. 2015. Regresi Terboboti Geografis dan Regresi Terboboti Geografis Campuran. [tesis]. Bogor (ID) : Institut Pertanian Bogor.

[13] Respati T, Raksanegara A, Djuhaeni H, Sofyan A, Agustian D, Faridah L, Sukandar H. 2017. Berbagai Faktor yang Mempengaruhi Kejadian Demam Berdarah Dengue di Kota Bandung. *ASPIRATOR.*9(2017):91-96

[14] Ruliansyah A, Yuliasih Y, Ridwan W, Kusnandar AJ. 2017. Analisis Spasial Sebaran Demam Berdarah Dengue di Kota Tasikmalaya Tahun 2011 – 2015. A*SPIRATOR*. 9(2):85-90.

Cite this article as :

Zulfanita Dien R, Asep Saefuddin, Anik Djuraidah, "Modelling The Number of Cases of Dengue Hemorragic Fever With Mixed Geographically Negative Binomial Regression in West Java Province", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), ISSN : 2456-3307, Volume 6 Issue 1, pp. 71-77, January-February 2019. Available at doi : https://doi.org/10.32628/IJSRSET196124