



Mapping Leprosy Distribution with Geographically Weighted Bivariate Zero Inflated Poisson Regression Method

Siti Masliyah Lubis, Henny Pramoedyo, Suci Astutik

Abstract: Geographically Weighted Bivariate Zero Inflated Poisson regression modelling has been developed to evaluate overdispersion and spatial heterogeneity in factors the number of PB Leprosy and MB Leprosy Cases in North Sumatera Province in 2017. The modelling results shows there are 25 different models for each district. PB Leprosy cases are mostly influenced by the percentage of poor people and the percentage of healthy houses and MB Leprosy cases are influenced by percentage of poor people, percentage of clean and healthy life behavior, the ratio of medical personnel and the percentage of healthy houses.

Keywords: Leprosy, Overdispersion, Geographically Weighted Bivariate Zero Inflated Poisson Regression.

I. INTRODUCTION

Leprosy is a disease Leprosy is a chronic skin infection caused by *Microbacterium leprae*. Indonesia has been designated by WHO as one of the areas affected by leprosy. Leprosy is an infectious disease that causes very complex problems. According to World Health Organizations (WHO) leprosy is divided into two types namely PB (Pausi Bacillary) and MB (Multi Bacillary). The purpose of the classification of leprosy is to determine the treatment regimen, prognosis, complications and operational planning.

North Sumatera Province is a region with low leprosy cases, but it is very worrying because it continues to increase. This is evident from existing data recorded in 2011-2015 which continues to increase, as many as 170 new cases in 2011, 179 new cases in 2012, 175 new cases in 2013, 188 new cases in 2014, 216 new cases of leprosy in 2015 [6].

Data on the number of leprosy cases is a count data, which can be approached with the Poisson model. Poisson regression is used to analyze the count data, but there are assumptions that must be met in the use of poisson regression, which is the similarity of mean and variance (equidispersion).

In this study, there is a pair of overdispersion count data (variance values greater than mean) containing zero inflated values so that the assumptions of mean and variance are not fulfilled then it can be modeled with Bivariate Zero Inflated Poisson Regression (BZIPR), BZIPR models produce estimated values of parameters that are global for all locations, assuming that each location has the same characteristics even though the Characteristics in each region are very likely to affect the number of events in the region as well as events that are distributed poisson. Differences in these characteristics are influenced by several factors such as natural or geographical conditions, socio-cultural, economic and others, so that to overcome the occurrence of diversity can be used spatial data analysis, namely the Geographically Weighted Bivariate Zero Inflated Poisson Regression method.

II. MATERIAL AND METHOD

Study Area

In this study, which is used as an observation unit is a district in North Sumatra Province. Observation units were 25 districts.

Data Collection

The research variables used in the study were obtained from the publication of the Health Profile of North Sumatra Province [7], consisting of two response variables (Y), namely the number of Pausi Bacillary (PB) leprosy cases, the number of Mausii Bacillary leprosy cases (MB) and four predictor variables (X), namely factors related to the cause and transmission of leprosy, namely the percentage of poor people, the percentage of households behaving clean and healthy, the ratio of medical personnel, the percentage of healthy homes.

III. STATISTICAL METHOD

Breusch Pagan Test (BP-Test)

Breusch Pagan Test is conducted to see the diversity between locations due to a location having different structure and characteristics from other locations [1].

With the following hypothesis:

$$H_0: \sigma^2_{(u_1, v_1)} = \sigma^2_{(u_2, v_2)} = \dots = \sigma^2_{(u_n, v_n)} = \sigma^2$$

$$H_1: \text{there is at least one } \sigma^2_{(u_i, v_i)} \neq \sigma^2$$

To test this hypothesis, the Breusch-Pagan Test statistic is used, namely:

Manuscript published on 30 September 2019

* Correspondence Author

Siti Masliyah Lubis*, Master Student of Statistics Departement, Brawijaya University, Malang, Indonesia. Email: sitimasliyahlubis@gmail.com

Henny Pramoedyo, Lecturer of Statistics Departement, Brawijaya University, Malang, Indonesia. Email: pramoedyohp@yahoo.com

Suci Astutik, department, Lecturer of Statistics Departement, Brawijaya University, Malang, Indonesia. Email: suci_sp@ub.ac.id

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

$$BP = \left(\frac{1}{2}\right) h^T Z(Z^T Z)^{-1} Z^T h + \left(\frac{1}{2}\right) \left[\frac{\varepsilon^T W \varepsilon}{\sigma^2}\right] \sim \chi^2_{(P+1)} \quad (1)$$

Decion criteria: H_0 rejected $BP > \chi^2_{(P+1)}$ meaning that they have different variations or spatial heterogeneity.

Weighting and Optimum Bandwidth Determination

The choice of weighting is very important in the parameter estimator because it can provide the results of parameter estimators that are different from each observation location. Adaptive Gaussian kernel weighting function [5], namely:

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{h_i^2}\right) \quad (2)$$

Euclidean distance (d_{ij}) between location $-i$ and location $-j$ with the equation as follows:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (3)$$

The method used to determine the optimum bandwidth is used the Cross Validation (CV) method as follows:

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{-i}(h))^2 \quad (4)$$

Bandwidth is said to be optimum if a minimum CV is obtained from the Golden Section search technique.

Geographically Weighted Bivariate Zero Inflated Poisson Regression

PB leprosy and MB leprosy case count data that have correlations, have a mostly zero value and there is a spatial heterogeneity using the Geographically Weighted Bivariate Zero Inflated Poisson Regression method. GWBZIPR Model [8].

$$\ln(\hat{\mu}_1) = e^{x_i \beta(u_i, v_i)}; \ln(\hat{\mu}_2) = e^{x_i \beta(u_i, v_i)}; \text{logit}(\hat{\pi}_i) = x_i \gamma(u_i, v_i) \quad (5)$$

Distribution of Bivariate Zero Inflated Poisson Regression [4]:

$$f(Y_1, Y_2) = \begin{cases} (1-\pi) + \pi e^{-(\mu_1+\mu_2)} (1 + \alpha(1 - e^{-\mu_2^c})) & , (y_1, y_2) = (0, 0) \\ \frac{\pi e^{-(\mu_1+\mu_2)} \mu_1^{y_1} \mu_2^{y_2}}{y_1! y_2!} (1 + \alpha(e^{-y_1} - e^{-\mu_1^c})(e^{-y_2} - e^{-\mu_2^c})) & , (y_1, y_2) \neq (0, 0) \end{cases} \quad (6)$$

where: $c = 1 - \frac{1}{e}$
 $\mu_1 = e^{x^T \beta_1}$ dan $\mu_2 = e^{x^T \beta_2}$

$$\pi = \frac{e^{x^T \gamma}}{1 + e^{x^T \gamma}} \text{ dan } (1 - \pi) = \frac{1}{1 + e^{x^T \gamma}}$$

Estimating the parameters of the GWBZIPR model using the Maximum Likelihood Estimation (MLE) method is then solved by the Newton-Raphson iteration method.

Partial testing is also needed to determine the significance of each β and γ parameter with the hypothesis for testing are as follows:

(i) $\beta(u_i, v_i)$ Parameter

$$H_0: \beta_{j1}(u_i, v_i) = \beta_{j2}(u_i, v_i) = \dots = \beta_{jk}(u_i, v_i) = 0, \quad j = 1, 2$$

$$H_1: \beta_{jl}(u_i, v_i) \neq 0, \quad j = 1, 2; l = 1, 2, \dots, k$$

With test statistics, namely:

$$t = \frac{\hat{\beta}_{jl}(u_i, v_i)}{SE(\hat{\beta}_{jl}(u_i, v_i))} \quad (7)$$

(ii) γ Parameter

$$H_0: \gamma_1(u_i, v_i) = \gamma_2(u_i, v_i) = \dots = \gamma_k(u_i, v_i) = 0$$

$$H_1: \gamma_l(u_i, v_i) \neq 0, \quad l = 1, 2, \dots, k$$

With test statistics, namely:

$$(8)$$

$$t = \frac{\hat{\gamma}_i(u_i, v_i)}{SE(\hat{\gamma}_i(u_i, v_i))}$$

Rejected H_0 , if $|Z_{hit}| > Z_{\alpha/2}$

IV. RESULT AND DISCUSSION

Data Description

The response variable used was the number of PB leprosy cases and MB leprosy cases in North Sumatra Province. Administratively, North Sumatra Province consists of 33 Regencies / Cities. In this study, the observation unit was conducted in the regencies in North Sumatra Province, this was due to differences between districts and cities in terms of health facilities, economy, education, healthy behavior, sanitation and environmental conditions (District / City health profile in 2017).

Response Variable Correlations

GWBZIPR analysis involves two response variables, the two response variables must be interconnected or correlated. To see whether there is a relationship between the response variables the following hypotheses are tested:

$$H_0: \rho = 0, \text{ there is no correlation between } Y_1 \text{ dan } Y_2$$

$$H_1: \rho \neq 0 \text{ there is a correlation between } Y_1 \text{ dan } Y_2$$

Correlation checks are performed using the Pearson Correlation test, the results obtained are the correlation coefficient values for PB leprosy response variable (Y_1) with MB leprosy response variable (Y_2) of 0.576 with a P-value of 0.000, using a significance level (α) of 5% , p-value <0.05, then reject H_0 , meaning that there is a significant correlation between PB leprosy response variable (Y_1) and MB leprosy response variable (Y_2).

Multicollinearity test

Multicollinerity test is performed by Variance Inflation Factors (VIF) test, if the VIF > 10 then there is multicollinearity between the predictor variables, the oppsite. The VIF value of each predictor variable can be seen in Table 1.

Table 1. VIF Values of Predictors

	X_1	X_2	X_3	X_4
VIF	1.26	1.04	1.17	1.34

Table 1 shows that the VIF value of each predictor variable is smaller than 10 so that it can be concluded that there is no multicollinearity between the predictor variables, meaning that there is no correlation between the predictor variables.

Breusch Pagan Test

Breusch Pagan Test see the diversity between locations due to a location that has a different structure and relationship parameters with other locations. Breusch-Pagan test results obtained p-value **0.010** < α (0.05) meaning that there is diversity or spatial heterogeneity in the data on the number of PB leprosy and MB leprosy cases between districts in the study.

Geographically Weigthed Bivariate Zero Inflated Poisson Regression Modelling

The significance test of the Geographical Weigthed Bivariate Zero Inflated Poisson Regression model is partially carried out to determine what factors are significantly influential in each region.



Based on the results of the analysis, obtained significant variables that are different for each district both for PB leprosy and MB leprosy cases. Significant variables in each district that have been grouped can be seen in Table 2 below:

Table 2. Grouping each district based on significant variables

District		Significant Variable
PB Leprosy	MB Leprosy	
Middle Tapanuli, Dairi, Langkat, Humbang Hasundutan, North Nias	Mandailing Natal, South Tapanuli, North Tapanuli, Asahan, Dairi, Karo, Deli Serdang, Langkat, South Nias, Padang Lawas, North Padang Lawas, North Nias	X_1, X_2, X_3, X_4
-	Humbang Hasundutan	X_1, X_2, X_3
South Tapanuli	-	X_1, X_2, X_4
Nias, Samosir	-	X_1, X_2, X_4
Karo, Batubara	Middle Tapanuli, South Labuhan Batu	X_2, X_3, X_4
-	Labuhan Batu, Batubara	X_1, X_2
-	Simalungun	X_1, X_3
Toba Samosir	Serdang Bedagai	X_2, X_3
-	-	X_2, X_4
Labuhan Batu, South Nias, Serdang Bedagai, West Nias	-	X_1
-	Nias, Samosir	X_2
North Tapanuli, Asahan	North Labuhan Batu	X_3
Mandailing Natal, North Labuhan Batu	Toba Samosir	X_4

The following is a mapping of Regency groupings based on significant variables in Table 3, illustrated in the map of the Province of North Sumatera:

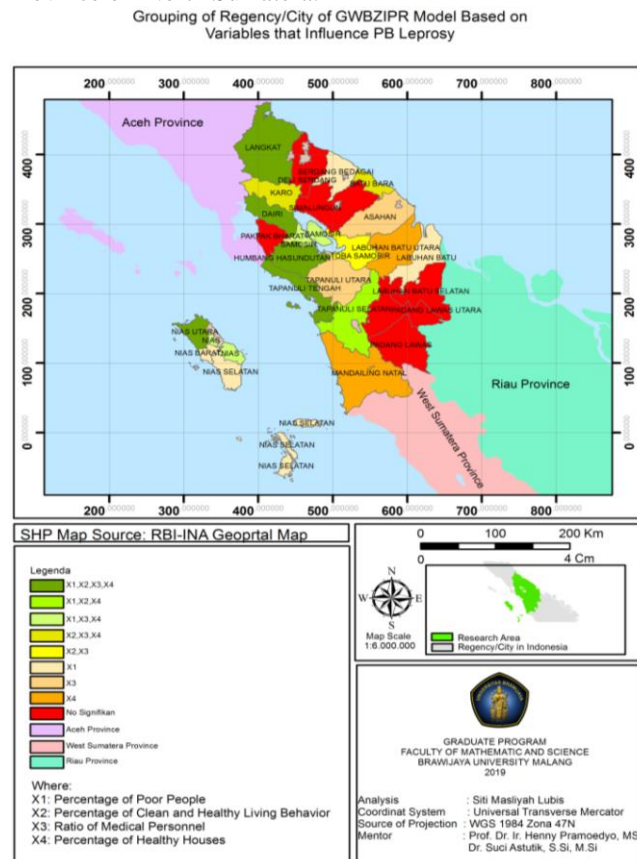


Figure 1: Regency grouping is based on variables which are significant in PB leprosy cases

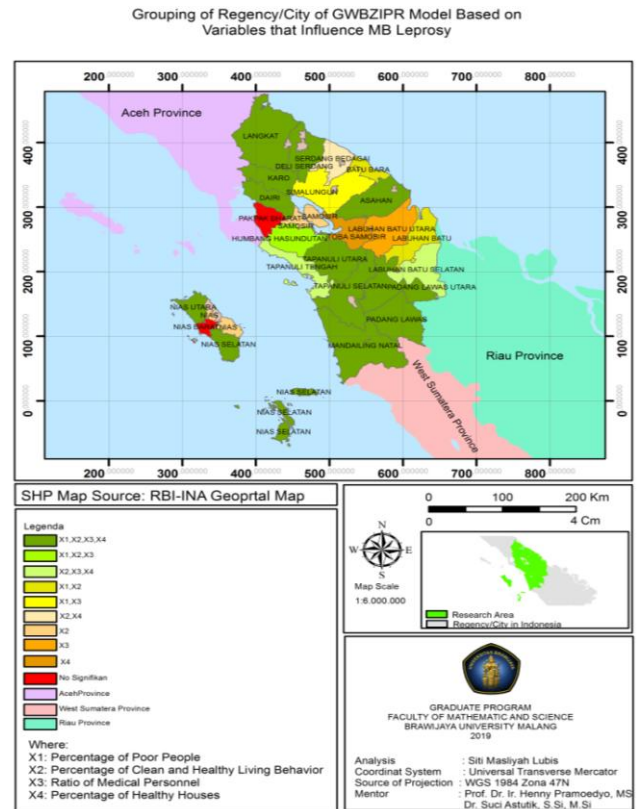


Figure 2: Regency grouping is based on variables which are significant in MB leprosy cases

Based on partial testing of parameters, for example, testing parameters will be presented at the first research location, Langkat Regency. it can be concluded that there are 4 significant variables in the GWBZIPR model, so that the model for Langkat Regency is as follows:

- Poisson State model for $\hat{\mu}_1$ (PB Leprosy)

$$\hat{\mu}_1 = \exp(22429.65 + 9.228291 X_1 - 151.067 X_2 - 1020.1 X_3 - 158.069 X_4)$$
- Poisson State Model for $\hat{\mu}_2$ (MB Leprosy)

$$\hat{\mu}_2 = \exp(260.6553 + 0.014344 X_1 + 0.148691 X_2 - 0.02441 X_3 - 0.00867 X_4)$$
- Zero State Model

$$\logit(\hat{\pi}) = (-0.02003 - 114.789 X_1 + 0.279538 X_2 + 2.7857 X_3 - 0.8215 X_4)$$

V. CONCLUSION

The GWBZIPR Modeling on data on the number of PB leprosy and MB leprosy cases in North Sumatra Province in 2017 has 25 different models for each district, with predictor variables that significantly influence PB leprosy and MB leprosy which are different for each district in North Sumatra Province. PB Leprosy cases are mostly influenced by the percentage of poor people and the percentage of healthy houses and MB Leprosy cases are influenced by percentage of poor population, percentage of clean and healthy life behavior, the ratio of medical personnel and the percentage of healthy houses.

REFERENCES

1. Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Netherlands: Kluwer Academic Publishers.
2. Barriga, G.D.C., & Louzada, F. 2014. The Zero Inflated Conway-Maxwell-Poisson Distribution: Bayesian Inference, Regression Modeling and Influence Diagnostic. *Statistical Methodology*. 21, 23-24.
3. Cameron, A.C., & Trivedi, P.K. 2013. *Regression Analysis of Count Data Second Edition*. USA: Cambridge University Press.
4. Krishna, P.M., & Tukaram, S.D. 2011. Bivariate Zero-Inflated Power Series Distribution. *Applied Mathematics*. 2: 824-829
5. Leung, Y., Mei, C.L. dan Zhang, W.X. 2000. Statistical Test for Spatial Nonstationary Based on the Geographically Weighted Regression Model. *Environment and planning A: Economy and Space*. 32: 9-32.
6. Ministry of Health of the Republic of Indonesia. 2015. *Leprosy Infodatin*. Jakarta
7. North Sumatra Provincial Health Service. 2018. *Health Profile of North Sumatra Province 2017*. Medan: Provincial Government of North Sumatera
8. Pangulimang, J., Puhadi., & Sutikno. 2016. Parameter Estimation and Hypotesis Testing Geographically Weighted Bivariate Zero-Inflated Poisson. *Implementation and Education of Mathematics and Sciences*. Proceeding of 3rd International Conference on Research.

AUTHORS PROFILE



Siti Masliyah Lubis S.Si. Master Student of Statistics Departement, Brawijaya Universiy, Malang, Indonesia.



Prof. Dr. Ir. Henny Pramodyo, MS Lecturer of Statistics Departement, Brawijaya University, Malang, Indonesia.



Dr. Suci Astutik, S.Si., M.Si Lecturer of Statistics Departement, Brawijaya University, Malang, Indonesia. words.