Comparison of Geographically Weighted Generalized Poisson Regression (GWGPR) and Geographically Weighted Negative Binomial Regression (GWNBR) Methods in Determining Factors Affecting Tuberculosis Cases in Indonesia

Awaliyatul Uswah [(1)], Jose Rizal [(2)], Yulian Fauzi [(3)]

[1,2,3] Master of Statistics Study Program, University of Bengkulu, Bengkulu

WR. Supratman Street, Kandang Limun Village, Muara Bangkahulu District, Bengkulu City

e-mail: awaliawell359@gmail.com[(1)], jrizal@unib.ac.id[(2)], yulianfauzi@unib.ac.id[(3)]

ABSTRAK

Model Geographically Weighted Generalized Poisson Regression (GWGPR) dan Geographically Weighted Negative Binomial Regression (GWNBR) pada temuan penelitian ini menunjukkan hasil efektif dalam memodelkan data insiden tuberkulosis (TB) yang dicirikan oleh overdispersi dan heterogenitas spasial. Meskipun kedua model menghasilkan statistik kecocokan yang sebanding, seperti yang ditunjukkan oleh nilai Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) yang hampir identik, GWGPR menunjukkan sensitivitas yang lebih tinggi terhadap variabilitas regional, sebagaimana dibuktikan oleh pembentukan empat klaster provinsi yang berbeda berdasarkan variabel prediktor yang signifikan, dibandingkan dengan hanya dua klaster yang diidentifikasi oleh model GWNBR. Hal ini menunjukkan bahwa GWGPR dapat menawarkan pemahaman yang lebih berwarna tentang efek spasial dalam data epidemiologi. Pada beberapa variabel yaitu prevalensi merokok, kelembaban rata-rata tahunan, jumlah hari hujan, persentase penduduk yang melaporkan keluhan kesehatan, dan persentase penemuan dan pengobatan TB yang terbukti signifikan secara konsisten di semua provinsi dalam kedua pendekatan pemodelan. Hasil ini mendukung pentingnya teknik pemodelan yang terbobot secara geografis tidak hanya untuk meningkatkan akurasi prediktif tetapi juga untuk menginformasikan pengaruh kesehatan masyarakat berdasarkan wilayah. Dengan demikian, penggunaan model yang adaptif secara spasial seperti GWGPR dapat mendukung strategi pengendalian penyakit yang lebih terarah dan efektif dengan menyelaraskan respons kebijakan kesehatan dengan kebijakan penanggulangan TB di wilayah setempat.

Kata kunci: *AIC, BIC,* GWGPR, GWNBR, heterogenitas spasial, overdispersi, tuberkulosis

ABSTRACT

The findings of this study demonstrate that both the Geographically Weighted Generalized Poisson Regression (GWGPR) and Geographically Weighted Negative Binomial Regression (GWNBR) models are effective in modeling tuberculosis (TB) incidence data characterized by overdispersion and spatial heterogeneity. Although both models yield comparable fit statistics—as indicated by nearly identical Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values—GWGPR exhibits a higher sensitivity to regional variability, as evidenced by the formation of four distinct provincial clusters based on significant predictor variables, compared to only two clusters identified by the GWNBR model. This suggests that GWGPR may offer a more nuanced understanding of spatial effects in epidemiological data. Furthermore, several covariates; namely smoking prevalence, average annual humidity, number of rainy days, reported health complaints, and TB case detection and treatment coverage, emerged as consistently significant across all provinces in both modeling approaches. The recurrence of these variables across spatially disaggregated models highlights their fundamental role in influencing TB transmission dynamics at a national scale. Accordingly, the use of spatially adaptive models such as GWGPR can support more

targeted and effective disease control strategies by aligning health policy responses with the localized determinants of TB burden.
Keywords: *AIC, BIC, GWGPR, GWNBR, overdispersion, spatial heterogeneity, tuberculosis*

INTRODUCTION

Modeling count data is a common approach in statistical analysis when the dependent variable represents discrete, such as the number of disease cases. However, a frequent challenge in modeling count data is the presence of overdispersion, a condition in which the variance exceeds the mean [1]. This violates the key assumption of the Poisson regression model, which assumes equal mean and variance, leading to biased parameter estimates and inefficient inferences. To address this, extended count models such as Negative Binomial Regression (NBR) and Generalized Poisson Regression (GPR) have been developed [2]. These models introduce additional parameters to account for overdispersion and have been widely used in various applied contexts [1], [3].

Beyond overdispersion, spatial structure in the data presents another modelling challenge. In many real-world scenarios, observations are not independent but exhibit spatial correlation—values in one region may influence or resemble those in nearby regions. This is particularly relevant in public health, where disease spread and risk factors are often geographically patterned. To account for such spatial heterogeneity, extensions of GPR and NBR have been developed: Geographically Weighted Generalized Poisson Regression (GWGPR) and Geographically Weighted Negative Binomial Regression (GWNBR)**.** These geographically weighted models incorporate spatial coordinates (e.g., latitude and longitude) and allow model parameters to vary across locations, enabling more localized and context-sensitive analysis.

A relevant and pressing application of these spatial count models is in analysing tuberculosis (TB), a chronic infectious disease primarily affecting the lungs and caused by *Mycobacterium tuberculosis*. Transmission occurs through inhalation of airborne droplets, making the spread of TB highly sensitive to both environmental conditions and population density. Given its mode of transmission and the complex interaction between biological and ecological factors, TB incidence exhibits significant spatial variation. Previous studies have shown that environmental, climatic, and socioeconomic conditions vary markedly across Indonesian provinces, contributing to spatial heterogeneity in epidemic cases[1].

In light of these challenges, this study proposes a comparative spatial regression analysis using the GWGPR and GWNBR models to explore factors associated with TB incidence across Indonesia in 2021. Both models will incorporate adaptive bi-square kernel weighting to reflect spatial proximity. Model performance will be evaluated using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to determine the most suitable approach for addressing both modelling. This analysis aims to provide more accurate insights into TB determinants while supporting the development of geographically targeted public health interventions. the study intends to identify which approach better reflects the spatial dynamics of TB and to uncover key environmental and health-related predictors influencing its spread across regions.

METHOD

Before applying the GWGPR and GWNBR models, a multicollinearity diagnostic was conducted to ensure the independence of predictor variables. Multicollinearity occurs when two or

more independent variables are highly correlated, leading to unstable coefficient estimates and inflated standard errors. Identifying multicollinearity in independent variables based on the Variance Inflation Factor (VIF) value with the following equation [4], [5]:

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

The estimation of Poisson regression model parameters using the Maximum Likelihood Estimation (MLE) method by maximizing the likelihood function, which measures how likely it is to observe the given sample data under different parameter values. [6].

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{exp^{-\mu_i} {\mu_i}^{y_i}}{y_i!}$$

The assumption in poisson regression is that the mean and variance of the response variable are equal. Overdispersion occurs when the variance significantly exceeds the mean, which can invalidate the model's standard errors and lead to misleading inferences. Checking for overdispersion using the formula for the Pearson chi-square statistical test [7]:

$$\theta = \frac{\chi^2}{n-p}, \chi^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mu_i}$$

The parameter estimation of generalized Poisson regression using the MLE method with the likelihood function [8]:

$$L(\theta, \boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)}{1 + \theta exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)} \right)^{y_i} \frac{(1 + \theta y_i)^{y_i - 1}}{y_i!} exp\left[\frac{-exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)(1 + \theta y_i)}{1 + \theta exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)} \right]$$

The parameter estimation of the Negative Binomial Regression (NBR) model can be done using the MLE method. The likelihood function of NBR [9]:

$$L(\beta, \theta) = \prod_{i=1}^{n} \left(\frac{\Gamma(y_i + \theta^{-1})}{\Gamma(\theta^{-1}) + \Gamma(y_i + 1)} \left(\frac{1}{1 + \theta\mu_i} \right)^{\theta^{-1}} \left(\frac{\theta\mu_i}{1 + \theta\mu_i} \right)^{y_i} \right)$$

A statistically significant and positive Moran's I value would suggest that provinces with high (or low) TB counts tend to be near others with similar values, supporting the use of spatial modeling approaches. Moran's index test in testing spatial dependency of data, namely whether or not there is a spatial influence on the data [10].

$$Z(I) = \frac{I - E\,(I)}{\sqrt{Var(I)}}$$

To identify spatial heterogeneity, we assess whether the relationship between variables varies across different spatial locations. Spatial heterogeneity implies that the data-generating process differs across geographic regions, violating the assumption of global homogeneity in standard regression models.Breusch Pagan test in identifying spatial heterogeneity in data [10].

$$BP = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T f$$

Calculating the distance with the Haversine equation between observation locations based on geographical position. The Haversine equation can be formulated as follows [11] & [12]:

$$a = sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat_1) \cdot cos \cdot sin^2\left(\frac{\Delta long}{2}\right)$$

$$d_{ij} = 2r \cdot arc\sin\left(\sqrt{a}\right)$$

The optimal bandwidth for each observation in the geographically weighted models was determined using the cross-validation (CV) method [13]. This process identified the bandwidth that minimized the sum of squared prediction errors, thereby ensuring the best balance between local sensitivity and model generalizability. Optimum bandwidth for each observation location using Cross Validation (CV) can be written as follows [14]:

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

The selected bandwidth was then applied using an adaptive bisquare kernel, allowing each location to use a neighborhood size suited to its spatial context. Determining weights with the adaptive bisquare kernel function [15].

$$W_{ij}(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h_i}\right)^2\right)^2, & d_{ij} \leq h_i \\ 0, & d_{ij} > h_i \end{cases}$$

With d_{ij} calculated by two Haversine distance equations between location (u_i, v_i) to location (u_j, v_j) and h is the bandwidth .

Geographically Weighted Generalized Poisson Regression (GWGPR) parameter estimation is performed using the MLE method as follows:

$$L(\boldsymbol{\beta}(u_i, v_i)) = \prod_{i=1}^{n} f(y_i)$$
$$= \prod_{i=1}^{n} \left(\frac{\mu_i}{1+\theta\mu_i}\right)^{y_i} \frac{(1+\theta y_i)^{y_i-1}}{y_i!} exp\left[-\frac{\mu_i(1+\theta y_i)}{1+\theta\mu_i}\right]$$

Conducting Geographically Weighted Negative Binomial Regression (GWNBR) modeling. The GWNBR method can be formulated with [16]:

$$y_i \sim NB\left[t_i\left(\sum_k \beta_k(u_i, v_i)x_{ik}\right), \theta(u_i, v_i)\right]$$

The estimation of the GWNBR coefficient parameters was performed using the MLE method. The weighted natural logarithm function for the GWNBR model is:

$$ln\, L(\boldsymbol{\beta}(u_i, v_i), \theta_i | y_i x_i)$$
$$= \sum_{n=1}^{n} w_{ij}\,(u_i, v_i)\left\{ln\left[\frac{\Gamma(y_i + \theta^{-1})}{\Gamma(\theta^{-1}) + \Gamma(y_i + 1)}\right] - (y_i + \theta^{-1})ln(1+\theta\mu_i) + y_i ln(\theta\mu_i)\right\}$$

Simultaneous significance test using Maximum Likelihood Ratio Test. Hypothesis for the Poisson regression method, GPR, and NBR are as follows:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$

H_1 : there is at least one $\beta_j \neq 0, j = 1,2, \ldots, k$

Meanwhile, for GWGPR and GWNBR there are spatial elements (u_i, v_i), so the hypothesis is:

$H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_k(u_i, v_i) = 0$

H_1 : there is at least one $\beta_j(u_i, v_i) \neq 0, j = 1,2, \ldots, k$

Statistical test (Likelihood Ratio Test):

$$D(\hat{\beta}) = -2ln\left[\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right] = 2\ ln\left(L(\hat{\Omega}) - L(\hat{\omega})\right)$$

The function $L(\hat{\omega})$ the maximum likelihood value for a simple model without involving predictor variables and $L(\hat{\Omega})$ for a complete model. Reject H_0 if the value $D(\hat{\beta}) > \chi^2_{(\alpha,k)}$ which means that there is at least one parameter in the model that has a significant effect on the model. The value of $D(\hat{\beta})$ is the deviation, the smaller the value, the smaller the error produced by the model, so that the model becomes more precise.

Partial testing is carried out to determine which parameters have a significant effect on the model. In the Poisson, GPR, and NBR regression methods with the following test hypotheses:

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0, j = 1,2, \dots, k$

The test statistics used follow the z distribution, namely:

$$Z_{hit} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

Meanwhile, for GWGPR and GWNBR there are spatial elements (u_i, v_i), so the hypothesis is:

$H_0 : \beta_j(u_i, v_i) = 0$

$H_1 : \beta_j(u_i, v_i) \neq 0, j = 1,2, \dots, k$

Test statistics:

$$Z_{hit} = \frac{\hat{\beta}_j(u_i, v_i)}{se\left(\hat{\beta}_j(u_i, v_i)\right)}$$

The rejection criterion for the partial test hypothesis, namely reject H_0 if the value of $|Z| > Z_{\alpha/2}$ which means that the parameter has a significant effect on the model.

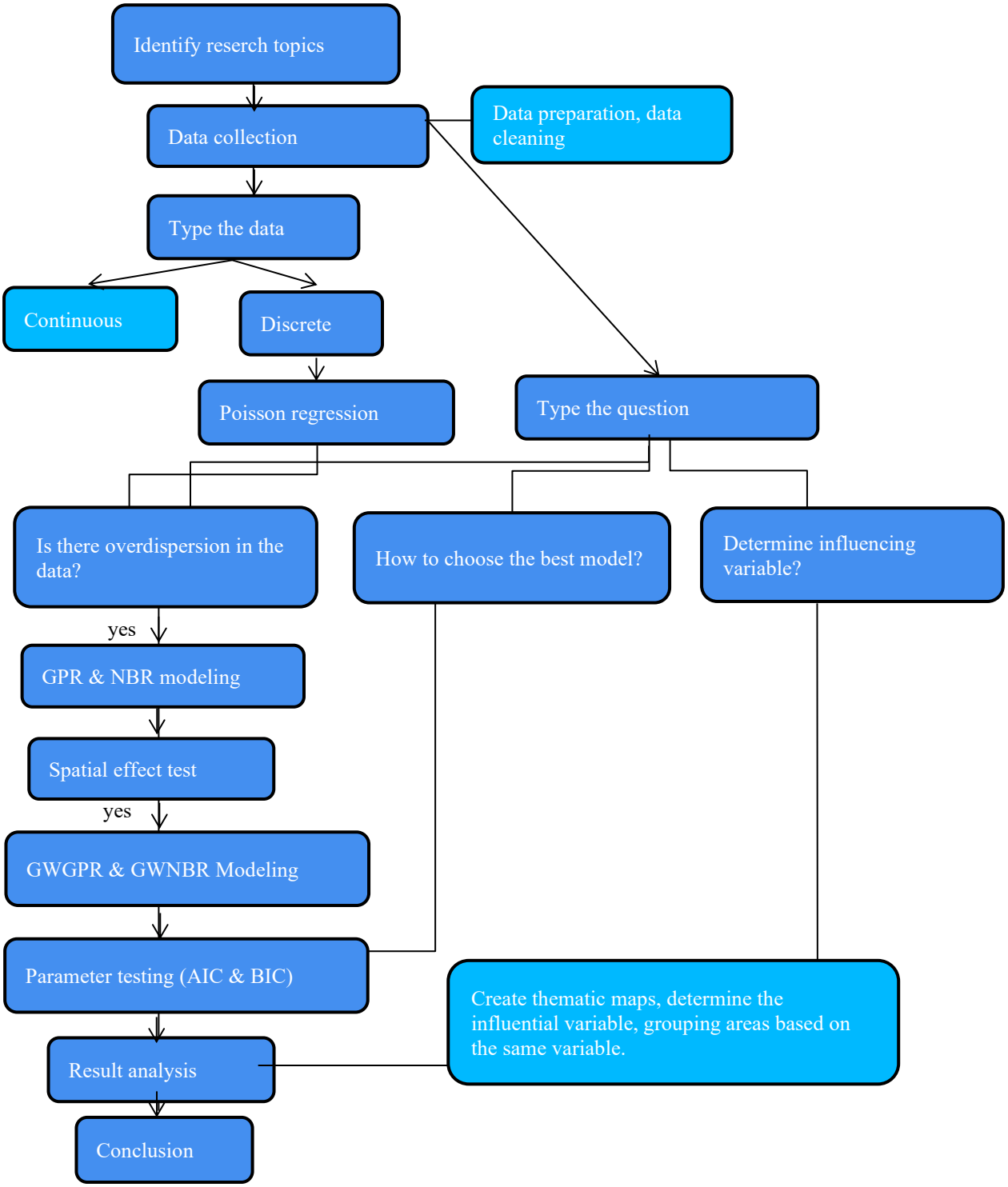
Comparing the goodness of fit values of the GWGPR and GWNBR models with two distances on the adaptive bi-square weighting based on the AIC and BIC criteria. According to [17] the equation for AIC is as follows:

$$AIC = -2logL(\hat{\beta}) + 2k$$

Meanwhile, the BIC formula is as follows [18]:

$$BIC = -2logL + log(n)k$$

RESEARCH FLOWCHART



Gambar 1. Flowchart of Research

RESEARCH VARIABLE

The variables used in this study are presented below.

Y: Number of positive tuberculosis (TB) cases
X_1: Percentage of smokers
X_2: Average annual temperature (℃)
X_3: Average annual humidity percentage
X_4: Average wind speed
X_5: Number of rainy days in a year
X_6: Average sunlight exposure
X_7: Population density
X_8: Percentage of population reporting health complaints
X_9: Percentage of TB detection and treatment
X_{10}: Percentage of poor population
X_{11}: Number of medical personnel
X_{12}: Number of healthcare workers
X_{13}: Number of community health centers (Puskesmas)

RESULT AND DISCUSSION

The number of TB cases in Indonesia in 2021 ranked second highest in the world after India. The trend of the number of TB cases has continued to increase since 2018-2021 [19]. The average number of TB cases in Indonesia in 2021 was 13,036 cases, with the highest cases in West Java Province, namely 103,253 cases, followed by Central Java Province with 47,430 cases and East Java with 47,398 cases, while the lowest cases were in North Sulawesi Province with 1,050 cases. The following is a graph of the number of TB cases in Indonesia:

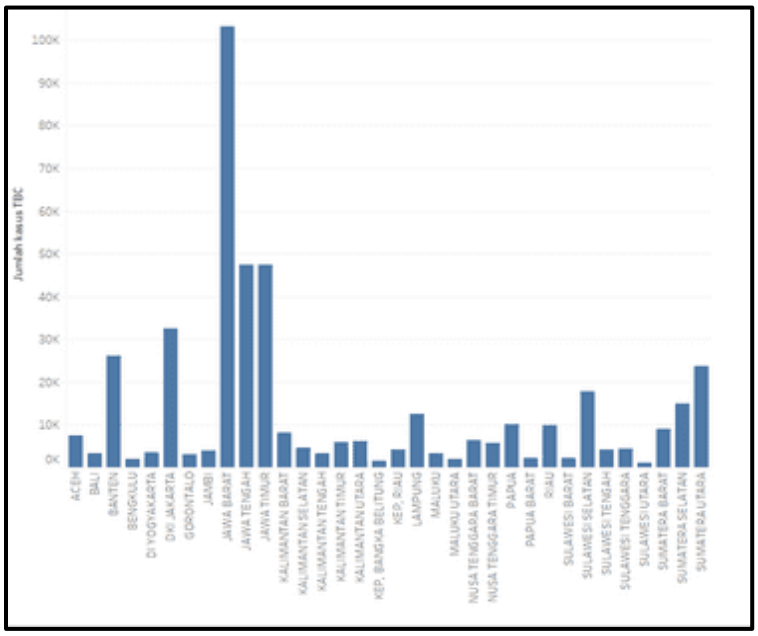


Figure 2. Graph of the number of TB cases per province in Indonesia in 2021

Multicollinearity testing is used to determine whether the independent variables meet the assumption that they are not correlated with each other so the model formed is a regression model in which there is no perfect relationship between the independent variables.

Table 1. VIF Values of Independent Variables

Variable	VIF value	Variable	VIF value
X_1	1.723	X_8	2.631
X_2	3.461	X_9	2.042
X_3	3.657	X_{10}	2.426
X_4	2.565	X_{11}	58.791
X_5	2.396	X_{12}	59.311
X_6	2.904	X_{13}	50.975
X_7	14.856		

Based on the VIF Value in Table 1, it is known that of the 13 variables tested, 9 of them have a VIF value < 10. Four variables produce VIF values > 10, namely X_7, X_{11}, X_{12}, X_{13}. While the variables that have VIF values < 10 namely X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_9 and X_{10}.

The result of the Pearson chi-square overdispersion test is 9536.552 which is greater than 1. So it is concluded that in the Poisson regression for TB case data in Indonesia in 2021, there was overdispersion. The occurrence of overdispersion in Poisson regression, the steps to handle it are to form a Generalized Poisson Regression and Negative Binomial Regression model.

Table 2. Estimated values of GPR model parameters

Parameter	Estimate	Std. Error	Z Value
$\hat{\beta}_0$	-4838	0,125	-38,804
$\hat{\beta}_1$	-0,014	0,038	-0,370
$\hat{\beta}_2$	-0,142	0,236	-0,600
$\hat{\beta}_3$	-0,083	0,056	-1,469
$\hat{\beta}_4$	-0,271	0,178	-1,518
$\hat{\beta}_5$	-0,007	0,005	-1,317
$\hat{\beta}_6$	-0,006	0,019	-0,309
$\hat{\beta}_8$	0,011	0,031	0,346
$\hat{\beta}_9$	0,057	0,014	4,116
$\hat{\beta}_{10}$	0,000	0,022	0,000
Devians	95.685		
AIC	691.59		

Table 3. Estimated values of the NBR model parameters

Parameter	Estimate	Std. Error	Z Value
$\hat{\beta}_0$	21,638558	7.1212245	3,039
$\hat{\beta}_1$	0,021245	0,044055	0,482
$\hat{\beta}_2$	-0,228496	0,162272	-1,408
$\hat{\beta}_3$	-0,082547	0,047744	-1,729
$\hat{\beta}_4$	-0,178384	0,182653	-0,977
$\hat{\beta}_5$	-0,007680	0,005102	-1,505
$\hat{\beta}_6$	0,001284	0,017675	-0,073
$\hat{\beta}_8$	0,014469	0,028613	0,506
$\hat{\beta}_9$	0,053197	0,010223	5,203
$\hat{\beta}_{10}$	-0,025263	0,025393	-0,995
Devians	95.685		
AIC	697.52		

The spatial effects tested as a condition for spatial regression are the spatial dependency test and spatial heterogeneity test. The results of spatial dependency test with Moran Index obtained $p\text{-}value = 0{,}000073$. With a significance level of $\alpha = 5\%$ the conclusion is rejected H_0 meaning that there is spatial dependency between locations, between observations of one location and other adjacent locations that influence each other.

Spatial heterogeneity test with Breusch Pagan test obtained $BP = 18{,}227$ and $p\text{-}value = 0{,}03263 < \alpha = 0{,}05$. The results of this value indicate that there is spatial heterogeneity or diversity between regions so that the parameters produced in each region can vary. Spatial heterogeneity conditions with overdispersion can be modeled using the GWGPR and GWNBR methods.

The estimation of GWGPR model parameters is obtained by including spatial weighting in its calculations using the Newton-Raphson iteration method. The estimation of GWGPR model parameters with adaptive Bisquare Kernel spatial weighting in each province obtained the value of Parameter estimation at each research location (u_i, v_i), where $i = 1, 2, 3, .., 4$.

Simultaneous parameter testing based on the deviation value is 11158, and a level greater than the value of $\chi^2_{(0.05,9)} = 16{,}919$. So reject H_0 hich means that there is at least one independent variable that has a significant effect on the dependent variable in each GWGPR model.

The partial test results produce different parameters in several provinces. Based on this difference, groups can be made where the number of TB cases that occur is influenced by the same variables.

Table 4. Grouping of provinces that have the same significant independent variables from the GWGPR model

No	Province	Significant Variables
1	Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, West Java, Central Java, DI Yogyakarta, East Java, Banten, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, South Kalimantan, East Kalimantan, Central Sulawesi, South Sulawesi, Southeast Sulawesi, West Sulawesi, Maluku, West Papua, Papua	$X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_9, X_{10}$
2	Central Kalimantan	$X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_9$
3	North Kalimantan	X_1, X_3, X_5, X_8, X_9
4	North Sulawesi, Gorontalo, North Maluku	$X_1, X_3, X_4, X_5, X_6, X_8, X_9, X_{10}$

The following is a visualization of the variable grouping in the GWGPR model:

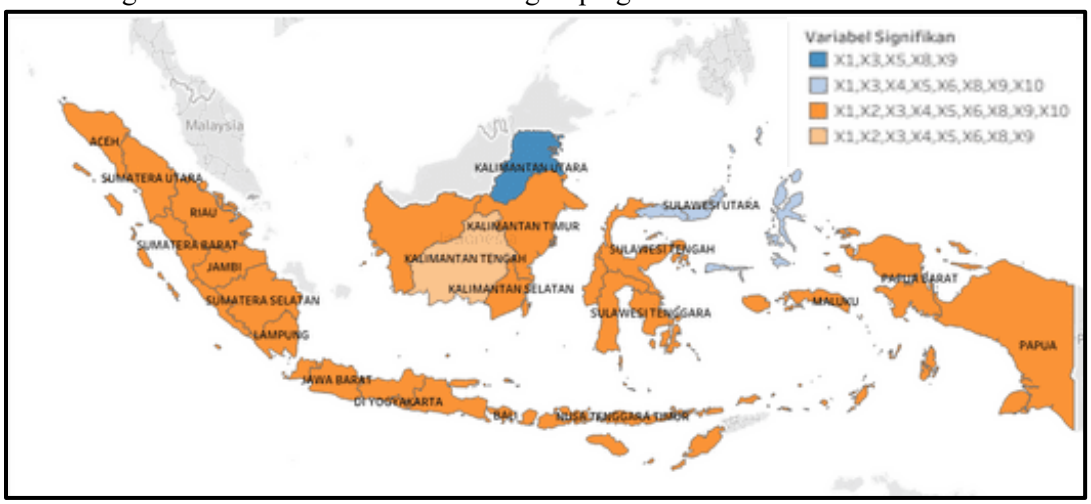


Figure 3. Grouping of provinces based on the significance of the same variables from the GWGPR model.

GWNBR parameter estimation uses the MLE method with the Newton-Raphson numerical iteration method, as well as geographic coordinate weighting of the region. The weighting components used are geographic location based on the spatial weighting matrix of the bandwidth and distance values of each location so that they have different parameters and describe the special or local properties of the model.

Simultaneous parameter testing is on the results of the deviation value $= 17549{,}5$ which is greater than the value of $\chi^2_{(0.05,9)} = 16{,}91898$. So reject H_0 which means that there is at least one independent variable that has a significant effect on the dependent variable in the GWNBR model. Grouping of provinces based on the similarity of independent variables that have a significant effect based on the partial test of the GWNBR model in the case of TBC.

Table 5. Grouping of provinces that have the same significant independent variables from the GWNBR model

No	Province	Significant Variables
1	Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, West Java, Central Java, DI Yogyakarta, East Java, Banten, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, South Kalimantan, East Kalimantan, Central Sulawesi, South Sulawesi, Southeast Sulawesi, West Sulawesi, Maluku, West Papua, Papua, Central Kalimantan, North Sulawesi, Gorontalo, North Maluku	$X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_9$
2	North Kalimantan	$X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_$

The following is a map visualization of the variable grouping in the GWNBR model:

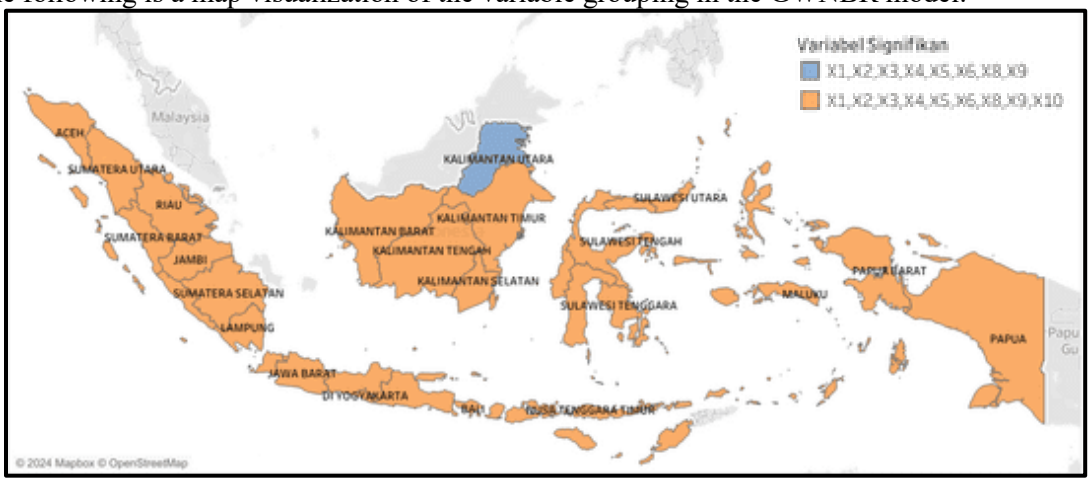


Figure 4. Grouping of provinces based on the same significant variables from the GWNBR model

The following is a comparison table of the model goodness of fit test results.

Table 6. Comparison AIC dan BIC model

No	Model	AIC	BIC
1	Poisson Regression	216152	216167.3
2	GPR	691.590	708.380
3	NBR	697.522	714.312
5	GWGPR	706.570	723.360
7	GWNBR	706.566	723.356

Based on Table 6, the smallest AIC and BIC values are the GPR model and the largest AIC and BIC values are the Poisson Regression. However, this model is less appropriate for use in this study because of the overdispersion conditions in the regression and there are spatial effects on the data. Therefore, the more appropriate models to use are the GWGPR and GWNBR models. Based on Table 8, the AIC and BIC values of the GWGPR and GWNBR models are not significantly different. Although both models yield comparable fit statistics, as indicated by nearly identical AIC and BIC values. GWGPR exhibits a higher sensitivity to regional variability, as evidenced by the formation of four distinct provincial clusters based on significant predictor variables, compared to only two clusters identified by the GWNBR model. This suggests that GWGPR may offer a more nuanced understanding of spatial effects in epidemiological data.

Variables X_2, X_4, X_6, X_{10} in several areas have significant differences in the partial test of the GWGPR and GWNBR models. Among them are variable X_2 (average annual temperature) in the provinces of North Kalimantan, North Sulawesi, Gorontalo, and North Maluku; variable X_4 (average wind speed) in North Kalimantan Province; variable X_6 (average sunshine) in North Kalimantan Province; variable X_{10} (percentage of poor population) in North Kalimantan and Central Kalimantan Provinces.

In several regions of Kalimantan Island, certain variables were found to be statistically insignificant, including the percentage of the population living in poverty (x_{10}). This can be attributed to the relatively low poverty rates in Kalimantan compared to other regions in Indonesia. In 2021, the national average poverty rate in Indonesia was 10.42%, while Central Kalimantan and North Kalimantan recorded significantly lower rates at 5.16% and 6.83%, respectively. Additionally, the number of tuberculosis (TB) cases across all provinces in Kalimantan was relatively low compared to the national average. The national average TB case count was 13.036, whereas the highest number of TB cases in Kalimantan was recorded in West Kalimantan (8,067 cases), and the lowest in Central Kalimantan (3,193 cases). Furthermore, other environmental variables such as average annual temperature (x_2), average humidity (x_3), average wind speed (x_4), and average solar radiation (x_6) were also found to be insignificant in several regions. These variables were generally associated with provinces that exhibited TB case counts below the national average, as reflected in the regional groupings based on the estimated coefficient values. This indicates that in areas with relatively low TB incidence, such variables may not play a dominant role in influencing TB prevalence.

This is an interesting finding from this final assignment, from several provinces producing the same variables, but in several other provinces, there are different significant variables. This is interesting to study further, where both models are theoretically quite significant in producing almost the same model goodness of fit test. However the empirical data found by the author is not enough evidence to support this theory.

The interpretation of the GWGPR and GWNBR models is based on spatially distributed estimates, visualized through thematic maps to highlight provincial-level patterns and characteristics. These spatial visualizations serve as references for policy-making, particularly in identifying variables with a higher estimated risk of influencing tuberculosis (TB) cases.

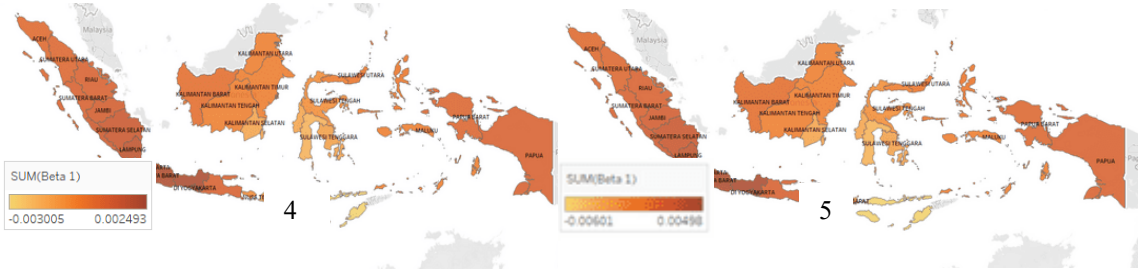


Figure 5. coefficient of the percentage variable of smokers in the GWGPR model
Figure 6. coefficient of the percentage variable of smokers in the GWNBR model

As shown in figure 5 & 6, darker brown shades on the maps represent provinces with higher estimated parameter values, while lighter shades indicate lower estimates. This color gradient underscores the spatial variability of regression coefficients across geographic regions. The variable representing the percentage of smokers shows that the darkest brown areas are primarily located in Sumatra and parts of Java, with the highest estimate found in the Special Region of Yogyakarta (DIY). The estimated coefficients for the smoking percentage in DIY are 0.002493 using GWGPR and 0.00498 using GWNBR. These estimates imply that a 1% increase in smoking prevalence is associated with an increase in TB cases by a factor of $\exp(0.002493) \approx 1.0025$ under GWGPR and

exp(0.00498) ≈ 1.005 under GWNBR, assuming other covariates remain constant. This result highlights the practical importance of tobacco control in mitigating TB incidence in high-risk provinces.

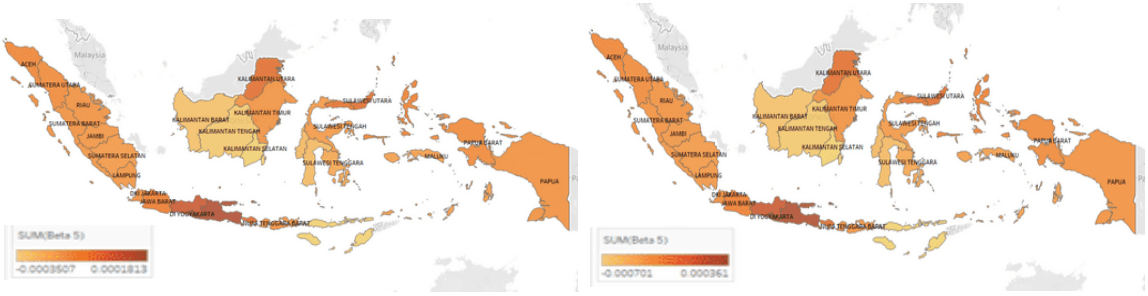


Figure 7. coefficient of the number of rainy days in the GWGPR model
Figure 8. coefficient of the number of rainy days in the GWNBR model

In the contrary, the variable for the number of rainy days (Figures 7 & 8) is predominantly light brown in areas such as Nusa Tenggara and parts of Kalimantan, indicating lower estimated values. The lowest estimates are found in East Nusa Tenggara, with coefficient values of -0.0003507 (GWGPR) and -0.000701 (GWNBR). These results suggest that an increase of one day in annual rainfall correlates with a decrease in TB cases by a factor of exp(-0.0003507) ≈ 0.9996 (GWGPR) and exp(-0.000701) ≈ 0.9993 (GWNBR), respectively—approximately equivalent to no meaningful change. These small effect sizes indicate that although the relationship is statistically detectable, the practical impact is minimal under these models.

Overall, the spatial heterogeneity captured by GWGPR and GWNBR models reveals that certain covariates exhibit both statistically and practically significant impacts on TB incidence, and that these effects vary geographically. These findings underscore the value of geographically weighted models in informing targeted, region-specific public health interventions.

This study is limited to data from the year 2021, and as such, the findings may not be generalizable to other time periods due to potential temporal variations in the influencing factors and disease patterns. Future research could explore the use of geographically weighted regression models that incorporate a broader range of social determinants, such as education level, housing conditions, and access to healthcare services, to enhance understanding of the spatial dynamics influencing TB incidence.

CONCLUSION

Based on the analysis conducted, several important conclusions can be drawn. The GWGPR (Geographically Weighted Generalized Poisson Regression) and GWNBR (Geographically Weighted Negative Binomial Regression) models are both applicable to modelling TB case data in Indonesia. The comparison of model fit statistics, particularly the AIC and BIC values, indicates no substantial difference between the two methods, suggesting that either model may be appropriately used to account for spatial variation in TB cases. However, GWGPR exhibits a higher sensitivity to regional variability, this suggests that GWGPR may offer a more nuanced understanding of spatial effects in epidemiological data

The analysis also revealed that the factors influencing TB incidence vary across provinces due to spatial heterogeneity and spatial dependency. Nevertheless, certain variables consistently showed a significant effect across all provinces in both the GWGPR and GWNBR models. These include the percentage of smokers, average annual humidity, number of rainy days per year, the percentage of the population reporting health complaints, and the percentage of TB detection and treatment coverage.

While the findings provide valuable insights into the spatial dynamics of TB in Indonesia, the study is limited by the use of cross-sectional data from 2021, which may not capture temporal trends or be generalizable across other years. Future research should consider longitudinal data and incorporate a broader set of social and environmental determinants. Furthermore, external validation with data from other time periods or regions would enhance the robustness and generalizability of the model outcomes.

REFERENCE

[1] N. Delvia, M. Mustafid, and H. Yasin, "Geographically Weighted Negative Binomial Regression Untuk Menangani Overdispersi Pada Jumlah Penduduk Miskin," *Jurnal Gaussian*, vol. 10, no. 4, pp. 532–543, 2021.

[2] M. Ririanti and R. D. Guntur, "Generalized Poisson Regression Modeling on the Number of Infant Deaths in East Nusa Tenggara Province in 2022," vol. 17, no. 2, pp. 779–788, 2024.

[3] F. Fitriani and M. Athoillah, "Penulisan Karya Tulis Ilmiah Bidang Sains Data," 2024.

[4] R. R. Hocking, *Methods and Applications of Linear Models*, Second Edi., vol. 39, no. 3. New York: John Wiley & Sons, Inc., 1997.

[5] R. K. Putri, M. Athoillah, and A. Haqiqiyah, "Analisis Faktor Yang Mempengaruhi Ketepatan Kelulusan Mahasiswa Dengan Algoritma Regresi Linear," *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 5, no. 2, pp. 671–680, 2024.

[6] R. E. Caraka and H. Yasin, *Geographically Weighted Regression Analysis*. Yogyakarta: Mobius, 2017.

[7] P. Mccullagh and J. A. Nelder, *Generalized Linear Models*, Second Edi., vol. 28, no. 1. London: Chapman and Hall, 1989.

[8] F. Famoye, "Restricted generalized poisson regression model," *Communications in Statistics - Theory and Methods*, vol. 22, no. 5, pp. 1335–1354, 2014.

[9] W. Greene, "Functional forms for the negative binomial model for count data," *Economics Letters*, vol. 99, no. 3, pp. 585–590, 2008.

[10] Breusch and Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, vol. 47, no. 5, pp. 1287–1294, 1979.

[11] B. D. Kifana and M. Abdurohman, "Great Circle Distance Methode for Improving Operational Control System Based on GPS Tracking System," *International Journal on Computer Science and Engineering (IJCSE)*, vol. 4, no. 04, pp. 647–662, 2012.

[12] Ö. G. Esenbuğa, A. Akoğuz, E. Çolak, B. Varol, and B. Erol, "Comparison of Principal Geodetic Distance Calculation Methods for Automated Province Assignment in Turkey,"

16th International Multidisciplinary Scientific GeoConference SGEM2016, Informatics, Geoinformatics and Remote Sensing, vol. 2, no. June, 2016.

[13] S. Alfiani and P. R. Arum, "Pemodelan Pertumbuhan Ekonomi di Jawa Barat Menggunakan Metode Geographically Weighted Panel Regression," *J statistika*, vol. 15, no. 2, pp. 219–227, 2022.

[14] A. Stewart Fotheringham, Chris Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. UK: John Wiley & Sons Ltd, 2015.

[15] A. Djuraidah, H. Djalihu, and A. M. Soleh, "Mixed Geographically and Temporally Weighted Autoregressive to Modeling the Levels of Poverty Population in Java in 2012-2018," *Journal of Physics: Conference Series*, vol. 1863, no. 1, 2021.

[16] M. Y. Darsyah, "Pemodelan Geographically Weighted Negative Binomial Regression (GWNBR) pada Kasus Malaria di Indonesia," *Jurnal Litbang Edusaintech*, vol. 2, no. 2, pp. 149–164, 2021.

[17] D. J. Briggs *et al.*, "Mapping urban air pollution using gis: A regression-based approach," *International Journal of Geographical Information Science*, vol. 11, no. 7, pp. 699–718, 1997.

[18] S. D. Oluwajana, P. Y. Park, and T. Cavalho, "Macro-level collision prediction using geographically weighted negative binomial regression," *Journal of Transportation Safety and Security*, vol. 14, no. 7, pp. 1085–1120, 2022.

[19] Sulistyo, R. D. Sagala, D. Asmoro, S. N. Rahma, B. Alisjahbana, and R. C. Koesoemadinata, *Annual Report National TB Program 2022*. Unicef Indonesia, 2022.