Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2026, 2026:39

https://doi.org/10.28919/cmbn/9883

ISSN: 2052-2541

COMPARISON OF GEOGRAPHICALLY WEIGHTED AND MIXED GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION MODELS FOR TUBERCULOSIS INCIDENCE

SITI CHOIROTUN AISYAH PUTRI[1], ASWI ASWI[2,*], RULIANA[2]

[1]Postgraduate Student in Statistics, Makassar State University, Makassar, Indonesia

[2]Statistics Department, Makassar State University, Makassar, Indonesia

Abstract: Tuberculosis (TB) remains a major public health concern characterized by spatial variation and overdispersion in case counts. In small-area analyses, differences in local characteristics may generate spatial heterogeneity in the relationships between risk factors and TB cases. Conventional Poisson regression is often inadequate because the equidispersion assumption is frequently violated and the model does not account for spatially varying effects. This study aims to model the number of TB cases in Makassar City in 2024 across 15 districts by comparing Geographically Weighted Negative Binomial Regression (GWNBR) and Mixed Geographically Weighted Negative Binomial Regression (MGWNBR). Seven predictor variables were examined: population density, health workforce, health facility, Immunodeficiency Virus (HIV)/Acquired Immune Deficiency Syndrome (AIDS), access to safe water, clean and healthy behavior practices (CHB), and malnutrition prevalence. Diagnostic tests indicated the presence of overdispersion and spatial heterogeneity, thereby justifying the application of spatial count regression models. Model comparison based on deviance and the Akaike Information Criterion (AIC) demonstrated that MGWNBR provided a superior fit. The CHB variable exhibited a global effect, whereas the remaining predictors showed spatially varying effects. These findings suggest that determinants of TB cases are not spatially homogeneous, and that a mixed modeling framework more effectively captures local dynamics while maintaining model parsimony.

*Corresponding author

E-mail address: aswi@unm.ac.id

Received March 17, 2026

The results support the development of geographically targeted TB control strategies in small-area settings.

Keywords: overdispersion; spatial heterogenity; GWNBR; MGWNBR; TB.

2020 AMS Subject Classification: 62J12.

1. INTRODUCTION

Tuberculosis (TB) remains the leading cause of death from a single infectious agent and ranks among the top ten causes of death worldwide. Caused by *Mycobacterium tuberculosis*, TB primarily affects the lungs and is transmitted through airborne droplets when an infected individual coughs, sneezes, or speaks [1]. Despite global control efforts, TB continues to pose a serious public health challenge, particularly in developing countries. Indonesia ranks second globally after India, contributing approximately 10% of total TB cases worldwide [2]. In 2022, an estimated 10.7 million TB cases were reported globally, while in Indonesia the number of cases increased from 677,464 in 2022 to 821,200 in 2023 [3].

At the provincial level, South Sulawesi ranked seventh nationally in 2022, with 24,209 confirmed cases and treatment coverage of 69%. In 2023, cases increased to 25,761, while treatment coverage declined to 55% [4,5]. Makassar City, a densely populated urban area, represents one of the highest TB-burden areas in the province, with 5,418 reported cases [6]. The substantial number of cases and interregional variation, shaped by population density, socioeconomic conditions, environmental factors, and access to health services underscore the importance of spatial analytical approaches to produce more accurate and context-specific inferences.

TB cases data are typically modeled as count data, making Poisson regression a common initial approach [7], [8]. However, Poisson regression assumes equidispersion, where the variance equals the mean [9]. In practice, this assumption is often violated due to unobserved heterogeneity and spatial variability in risk, resulting in overdispersion that may bias standard errors and lead to misleading inferences [10]. Negative Binomial Regression (NBR) addresses this limitation by incorporating a dispersion parameter that allows the variance to exceed the mean [11]. Nevertheless, NBR is a global model that assumes constant covariate effects across spatial units, which may be unrealistic in epidemiological contexts where relationships vary geographically [12].

To accommodate spatial heterogeneity, Geographically Weighted Regression (GWR) was developed to allow regression coefficients to vary across locations [13]. For count data, this approach was extended to Geographically Weighted Poisson Regression (GWPR) [14], and subsequently to Geographically Weighted Negative Binomial Regression (GWNBR), which

integrates spatially varying parameter estimation with overdispersion handling [15]. While GWNBR captures local variation in detail by estimating all parameters locally, treating every parameter as spatially varying may introduce estimation instability and complicate interpretation, particularly when certain predictors are theoretically or empirically global in nature.

Mixed Geographically Weighted Negative Binomial Regression (MGWNBR) was introduced to address this limitation by allowing both global and local parameters within a unified framework [16]. This approach provides greater flexibility by accommodating spatial heterogeneity without imposing a fully local structure on all predictors. By integrating global and local components, MGWNBR aims to balance spatial sensitivity with model parsimony.

Previous studies have applied Poisson and Negative Binomial regression to address overdispersion in TB case data in Indonesia and West Java [10,17]. Spatial heterogeneity has been incorporated using GWNBR [12,15], while Bayesian spatial models have been employed to estimate TB relative risk in South Sulawesi [18]. MGWNBR has subsequently been proposed to simultaneously identify global and local effects [19].

However, empirical studies comparing GWNBR and MGWNBR, particularly at the city level with a limited number of spatial units, remain scarce. When the number of observation units is small, specifying all parameters as local may lead to unstable estimates, making the choice between fully local and mixed parameter structures a critical methodological consideration.

Therefore, this study compares the performance of GWNBR and MGWNBR in modeling TB case counts in Makassar City in 2024. The comparison evaluates whether a mixed parameter specification improves model fit relative to a fully local approach. The findings are expected to contribute methodologically to spatial count data modeling, particularly in small-area contexts with limited spatial units.

2. PRELIMINARIES

2.1. Data

This study used secondary data obtained from the Makassar City Health Office. The research variables consisted of response variables, predictor variables, and observation location coordinates. To provide clear operational boundaries, all variables used in this study are presented in Table 1.

Table 1. Research Variables

Variable	Operational Definition	Unit
Y (Number of TB cases)	Total number of TB cases recorded in each region	Count
X_1 (Population density)	A measure that indicates the level of concentration or pressure of the population on a region	%
X_2 (Health workers)	The proportion of health workers to the population in a region, reflecting the level of availability of health care resources	%
X_3 (Health facilities)	The proportion of health care facilities to the population in a region, reflecting the level of availability of health care facilities	%
X_4 (Immunodeficiency Virus (HIV)/Acquired Immune Deficiency Syndrome (AIDS))	The proportion of HIV/AIDS cases to the population in a given area during a specific period, reflecting the prevalence of HIV/AIDS	%
X_5 (Access to safe water)	The proportion of households that use water sources that meet health safety standards (do not contain microorganisms) relative to the total number of households in a region	%
X_6 (Clean and healthy living behavior, CHB)	The proportion of households that meet the CHB indicators according to the established criteria to the total number of households in a region	%
X_7 (Malnutrition)	The proportion of individuals identified as suffering from malnutrition relative to the total population in a region	%

2.2. Negative Binomial Regression for Count Data

The number of TB cases in each region is count data that is generally modeled using Poisson regression [11]. However, in epidemiological data, overdispersion is often found, which is when the variance exceeds the mean value. This condition causes the assumption of equidispersion in Poisson regression to not be met, making the inference less reliable. Detection can be done through the Variance Mean Ratio (VMR) [20], Pearson dispersion, deviance dispersion, or the Cameron–Trivedi test [12].

To address this, Negative Binomial Regression (NBR) is used, which is a mixture of Poisson–

Gamma and estimated using maximum likelihood [21]. If $Y \sim \text{NB}(\mu, \theta)$ the probability mass function of the negative binomial distribution is:

$$f(y, \mu, \theta) = \frac{\Gamma\left(y + \frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right) y!}\left(\frac{1}{1+\theta\mu}\right)^{1/\theta}\left(\frac{\theta\mu}{1+\theta\mu}\right)^{y}, y = 0,1,2,\ldots$$

with $\mu > 0$ as the mean parameter and $\theta > 0$ as the dispersion parameter. This model extends the Poisson model by adding θ, resulting in a variance of $\text{Var}(Y_i) = \mu_i + \theta\mu_i^2$. When $\theta = 0$, the NB model reduces to the Poisson model.

Mathematically, the NBR model is expressed in the log link function in equation (1):

$$\ln(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}, i = 1,2,\ldots,n \quad (1)$$

where $\boldsymbol{x}_i$ is the vector of explanatory variables and $\boldsymbol{\beta}$ is the vector of regression parameters.

2.3. Spatial Heterogeneity and Geographically Weighted Regression (GWR)

The relationship between risk factors and TB cases may vary across regions due to differences in social conditions, environmental characteristics, and access to healthcare services. This phenomenon is referred to as spatial heterogeneity, indicating that the strength and direction of associations are not constant across geographic locations [22]. It is detected using the Breusch–Pagan (BP) test [13] with the following hypothesis:

H_0 : no heterogeneity$(\sigma_1^2 = \cdots = \sigma_i^2 = \sigma^2)$,

H_1 : there is heterogeneity (at least one $\sigma_i^2 \neq \sigma^2$ exists).

The test statistic is formulated in equation (2):

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

where $f_i = \frac{\varepsilon_i^2}{\sigma^2} - 1$ and H_0 are rejected $BP > \chi^2_{\alpha;p}$ or p-value$< \alpha$.

Because global regression models assume constant parameters, Geographically Weighted Regression (GWR) was developed to allow parameters to vary according to location, with the general form of equation (3):

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta}(u_i, v_i) \quad (3)$$

where (u_i, v_i) is the coordinate of region i and $\boldsymbol{\beta}(u_i, v_i)$ is the local parameter. Estimation is performed through distance-based weighting [13], one of which is an adaptive kernel with different bandwidths for each location [23]. The Adaptive Gaussian Kernel weighting function is written in equation (4):

$$w_i(u_i, v_i) = \exp\left[-\left(\frac{d_{ij}}{b_i}\right)^2\right] \quad (4)$$

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (5)$$

where u is the *longitude* coordinate and v is the *latitude* coordinate. Meanwhile, b is the optimal *bandwidth* determined using the *Cross Validation* (CV) method with the following equation (6):

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(b_i)]^2 \quad (6)$$

where $\hat{y}_{\neq i}(b_i)$ is the estimated value of y_i obtained by excluding the observation at location (u_i, v_i) from the estimation process (i.e., leave-one-out cross-validation). The optimal bandwidth is selected by minimizing the cross-validation (CV) criterion. The choice of kernel bandwidth determines the degree of spatial localization in the model and directly influences the bias–variance trade-off. A smaller bandwidth increases model locality but may produce unstable parameter estimates, whereas a larger bandwidth enhances stability at the expense of reduced sensitivity to local variation [24].

2.4. Geographically Weighted Negative Binomial Regression (GWNBR)

GWNBR is an extension of Negative Binomial regression that integrates the concept of geographic weighting. This model is used to address two problems simultaneously, namely overdispersion and spatial heterogeneity [23]. The GWNBR model is written in equation (7):

$$\ln\big(\mu(u_i, v_i)\big) = \boldsymbol{x}_i^T \boldsymbol{\beta}(u_i, v_i)\,, i = 1,2,\dots,n \quad (7)$$

Parameter estimation in the GWNBR model is performed using the maximum likelihood method with Newton–Raphson iteration. Then, simultaneous parameter significance testing in the GWNBR model is performed using the deviance test [25].

2.5. Mixed Geographically Weighted Negative Binomial Regression (MGWNBR)

In practice, not all variables have spatially varying effects. Some factors may be relatively homogeneous across the entire region. Therefore, the MGWNBR model was developed by combining global and local parameters into the model [16]. The MGWNBR model with $Y_i \sim NB(\mu_i, \theta)$ is written in equation (8):

$$\ln\big(\mu(u_i, v_i)\big) = \boldsymbol{x}_{iL}^T \boldsymbol{\beta}(u_i, v_i) + \boldsymbol{x}_{iG}^T \boldsymbol{\gamma}\,, i = 1,2,\dots,n \quad (8)$$

where $\boldsymbol{x}_{iL}$ denotes local variables, $\boldsymbol{x}_{iG}$ denotes global variables, and $\boldsymbol{\gamma}$ is the global parameter.

2.6. Parameter Significance Testing

Significance testing of parameters in the GWNBR and MGWNBR models includes simultaneous and partial tests. Simultaneous tests aim to determine whether there is at least one parameter that significantly affects the response variable with the hypothesis:

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

$H_1: \beta_j(u_i, v_i) \neq 0; j = 1,2, \ldots, p$

The statistic used is the Maximum Likelihood Ratio Test (MLRT) [19], with the deviation in equation (9):

$$D(\hat{\beta}) = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = 2 \left(\ln L(\hat{\Omega}) - L(\hat{\omega}) \right) \quad (9)$$

where $L(\hat{\Omega})$ is the likelihood of the full model, and $L(\hat{\omega})$ is the likelihood of the restricted model.

Reject the null hypothesis H_0 if $D(\hat{\beta}) \geq \chi^2_{p,\alpha}$.

Next, a partial test is performed using the Wald test to determine the significant parameters in each region with the hypothesis:

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

$H_1: \beta_j(u_i, v_i) \neq 0; j = 1,2, \ldots, p$

The test statistic used is the Wald test defined in equation (10):

$$Z_{score} = \frac{\hat{\beta}_j(u_i, v_i)}{se\left(\hat{\beta}_j(u_i, v_i)\right)} \quad (10)$$

Where $se\left(\hat{\beta}_j(u_i, v_i)\right)$ is standard error of the local parameter estimator $\hat{\beta}_j(u_i, v_i)$ at location i.

Reject H_0 if $|Z_{score}| \geq Z_{\alpha/2}$, indicating that the parameter has a significant effect on the response variable at the relevant location.

2.7. Model Selection Criteria

The best model is selected by evaluating the model's fit to the data using the Akaike Information Criterion (AIC). AIC is used to compare multiple models and select the model with the least information loss. A model with a lower AIC value is considered better [7]. The AIC value is determined using equation (11):

$$AIC = -2 \log(L) + 2k \quad (11)$$

Where L is the likelihood of the model, $\log(L)$ is the log-likelihood, and k is the number of parameters in the model (including the intercept).

3. MAIN RESULTS

3.1. Descriptive Analysis

As an initial step, a descriptive analysis was conducted to obtain an overview of the characteristics of the research variables, both response variables and explanatory variables. A summary of the descriptive statistics is presented in Table 2.

Table 2. Descriptive Statistics

Variable	Min.	Max.	Mean	Variance	SD	CoV	Category
Y	52.00	774.00	387.27	56219.35	237.11	0.61	Height
X_1	1.44	13.95	6.67	19.04	4.36	0.65	Height
X_2	2.00	20.00	7.40	21.28	4.61	0.62	Height
X_3	1.52	11.11	3.76	5.15	2.27	0.60	Height
X_4	0.22	12.90	6.67	14.74	3.84	0.58	Moderate
X_5	66.20	99.42	81.12	83.21	9.12	0.11	Low
X_6	20.01	90.96	74.78	389.57	19.74	0.26	Low
X_7	0.23	30.57	6.67	55.97	7.48	1.12	Height

As shown in Table 2, the average number of TB cases Y was 387.27 with a fairly wide range, namely 52 to 774 cases. The variance reached 56219.35 and the standard deviation was 237.11, resulting in a coefficient of variation (CoV) of 0.61, which is classified as high. This value indicates that the distribution of TB cases across districts in Makassar City has a strong level of heterogeneity. For the explanatory variables, X_1, X_2, and X_3 have a coefficient of variation above 0.60, indicating high interregional variation. The variable X_4 is in the moderate category, while X_5 and X_6 show relatively low variation. The variable X_7 has the highest CoV (1.12), indicating a very large distribution imbalance compared to other variables.

The high variation in the response variable and most of the explanatory variables provides an initial indication that the relationship between variables is likely not homogeneous across all regions. This indication is then reinforced through spatial visualization in Figure 1.

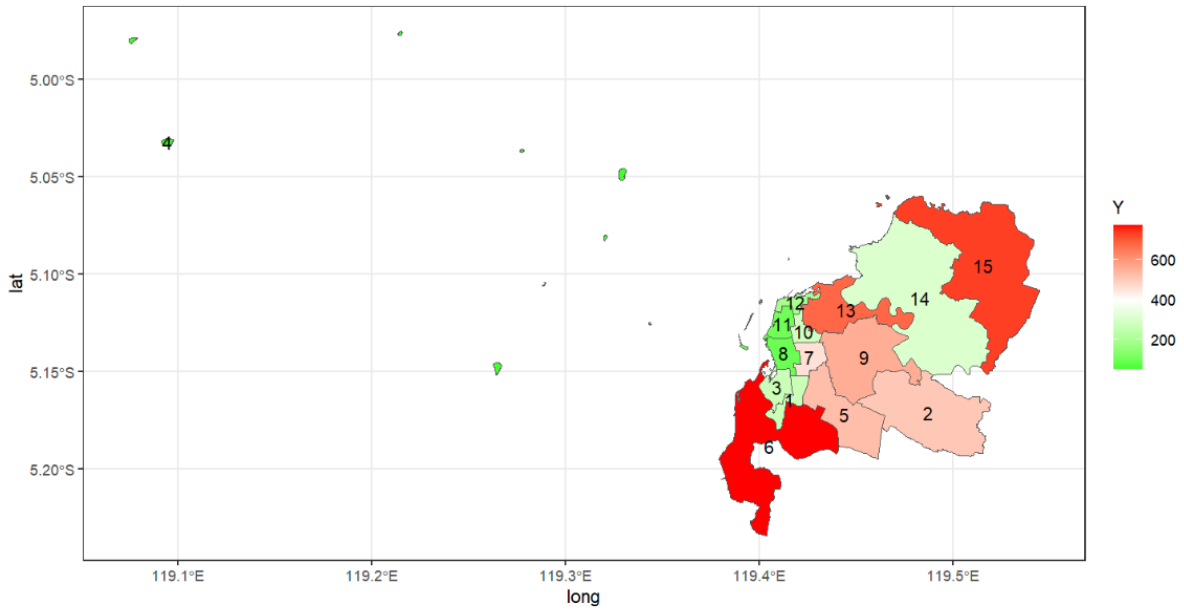

Figure 1. Map of TB Case Distribution in Makassar City

Figure 1 shows the spatial distribution of TB cases in Makassar City based on 15 districts. The areas with the highest number of cases are marked in red, namely Tamalate District (ID=6), followed by Biringkanaya District (ID=15) and Tallo District (ID=13). The areas with the lowest number of cases are marked in green, namely Wajo District (ID=11), followed by Ujung Pandang District (ID=8), and finally the Sangkarrang Islands (ID=4). The pattern of color intensity differences between regions shows that the distribution of cases is not uniform. Although a formal spatial test will be conducted in the next stage, this visualization indicates that TB risk factors may work differently between districts. These descriptive findings form the basis for proceeding to the global modeling stage.

3.2. Testing Assumptions

To ensure that the estimated model meets the statistical prerequisites and spatial data characteristics, tests for multicollinearity, overdispersion, and spatial heterogeneity were conducted before proceeding to the modeling stage.

3.2.1. Multicollinearity Testing

Multicollinearity testing was conducted to ensure that there was no high correlation between independent variables that could interfere with the stability of parameter estimation. The results of the testing using the Variance Inflation Factor (VIF) are presented in Table 3.

Table 3. Multicollinearity Test Results

Variable	VIF
X_1	2.32
X_2	2.25
X_3	1.76
X_4	3.52
X_5	1.99
X_6	2.11
X_7	1.89

As shown in Table 3, all VIF values are below 5. This indicates that there is no high correlation between independent variables that could interfere with the stability of the model. Thus, all variables are suitable for inclusion in the model.

3.2.2. Overdispersion Testing

Overdispersion testing was conducted to identify whether the variance of the case count data exceeded its mean value, which could indicate that the basic count data model with the assumption of equidispersion was not appropriate for use. The results are presented in Table 4.

Table 4. Overdispersion Diagnostic Test Results

Indicator	Value
VMR	145.1696
Pearson Dispersion	19.4129
Deviance Dispersion	19.5422

As shown in Table 4, the VMR value of 145.17 indicates that the variance is far above the average. The Pearson dispersion value (19.41) and deviance dispersion (19.54), which are well above one, indicate very strong overdispersion. This diagnostic test is also supported by the dispersion test results, which produce a Z_{score} value 2.545 with p-value = 0.005 < 0.05, thus rejecting the assumption of equidispersion. This finding explains that overdispersion causes the estimated standard error to be smaller than it should be, thereby increasing the probability of spurious significant conclusions. Therefore, an alternative model capable of accommodating overdispersion, namely the NBR model, is used.

3.2.3. Spatial Heterogeneity Testing

Spatial testing aims to detect spatial heterogeneity in the data. The BP test results show a statistical value of 18.5040 with a p-value of 0.0099 (< 0.05). This result indicates that the null hypothesis stating that there is no heterogeneity is rejected. In other words, there is significant spatial

heterogeneity, which shows that the residual variance is not constant between regions and the relationship between the explanatory variables and the number of TB cases may differ in each region. This result is the basis for using a location-based modeling approach to accommodate local parameter variations.

3.3. Spatial Count Regression Modeling

As a follow-up to the assumption testing results, modeling was continued with a spatial-based count regression approach that allows for local parameter estimation. In this process, an adaptive Gaussian kernel weighting function was used to capture the influence of geographical proximity between regions, with optimally determined bandwidth. Bandwidth and spatial weighting information are presented in Table 5.

Table 5. Optimum Euclidean Distance Bandwidth and Spatial Weighting

Region (j)	Bandwidth	Region 1		⋯	Region 15	
		d_{ij}	W_{ij}	⋯	d_{ij}	W_{ij}
1	0.0382	0.0000	1.0000	⋯	0.1212	0.0000
2	0.0756	0.0719	0.6356	⋯	0.0805	0.0000
⋮	⋮	⋮	⋮	⋱	⋮	⋮
⋮	⋮	⋮	⋮	⋱	⋮	⋮
14	0.0645	0.0850	0.0000	⋯	0.0367	0.8505
15	0.1006	0.1212	0.0000	⋯	0.0000	1.0000

As shown in Table 5, each region has a different optimal bandwidth, reflecting the model's ability to adjust the level of estimation locality according to the spatial structure of the data. A small bandwidth indicates a strong influence from the nearest region, while a large bandwidth involves a wider coverage of neighbors. Euclidean distance (d_{ij}) represents the proximity between regions, and spatial weights (W_{ij}) are calculated using an adaptive Gaussian kernel, where weights decrease with increasing distance. This variation in bandwidth confirms the existence of spatial heterogeneity that can be accommodated through this approach.

3.3.1. GWNBR Modeling

Based on GWNBR modeling, the spatial heterogeneity of TB cases in Makassar City is reflected in the differences in the combination of significant variables in each district. The determination of the significance of the parameters was carried out through a partial test by comparing the $|Z_{score}|$ value with the Z_{table} value at a significance level of 5%, which is 1.96. A parameter is considered significant if $|Z_{score}| > 1.96$, which is equivalent to a p-value < 0.05. This criterion is used consistently in all interpretations of local parameters in the GWNBR model. The grouping of

districts based on significant variables from the GWNBR model is presented in Table 6.

Table 6. Grouping of Districts Based on Significant Variables from the GWNBR Model

Group	Variable	District
1	X_5 (Access to Safe Water) X_6 (CHB)	Mamajang, Mariso, Rappocini, Tamalate, Makassar, Ujung Pandang, Bontoala, Wajo, Ujung Tanah
2	X_1 (Population Density) X_5 (Safe Water Access) X_6 (CHB)	Panakukkang
3	X_1 (Population Density) X_3 (Health Facilities) X_5 (Safe Water Access) X_6 (CHB)	Sangkarrang
4	X_2 (Health Workers) X_4 (HIV/AIDS Cases) X_5 (Safe Water Access) X_6 (CHB)	Manggala
5	X_1 (Population Density) X_2 (Health Workers) X_4 (HIV/AIDS Cases) X_5 (Safe Water Access) X_6 (CHB)	Tallo
6	X_1 (Population Density) X_2 (Health Workers) X_4 (HIV/AIDS Cases) X_6 (CHB) X_7 (Malnutrition)	Tamalanrea, Biringkanaya

As shown in Table 6, there are six district groups based on the combination of variables that meet the criteria $|Z_{score}| > Z_{table}$. The variable X_6 (CHB) emerges as the most consistently significant variable across regions. This indicates that clean and healthy living behaviors have a relatively dominant spatial contribution in explaining the variation in TB cases. Conversely, other variables such as X_1 (Population Density), X_2 (Health Workers), X_3 (Health Facilities), X_4 (HIV/AIDS Cases), X_5 (Access to Safe Water) and X_7 (Malnutrition) show different patterns of significance between districts, confirming the existence of variations in local risk structures. The grouping of districts based on GWNBR modeling is visualized in Figure 2.

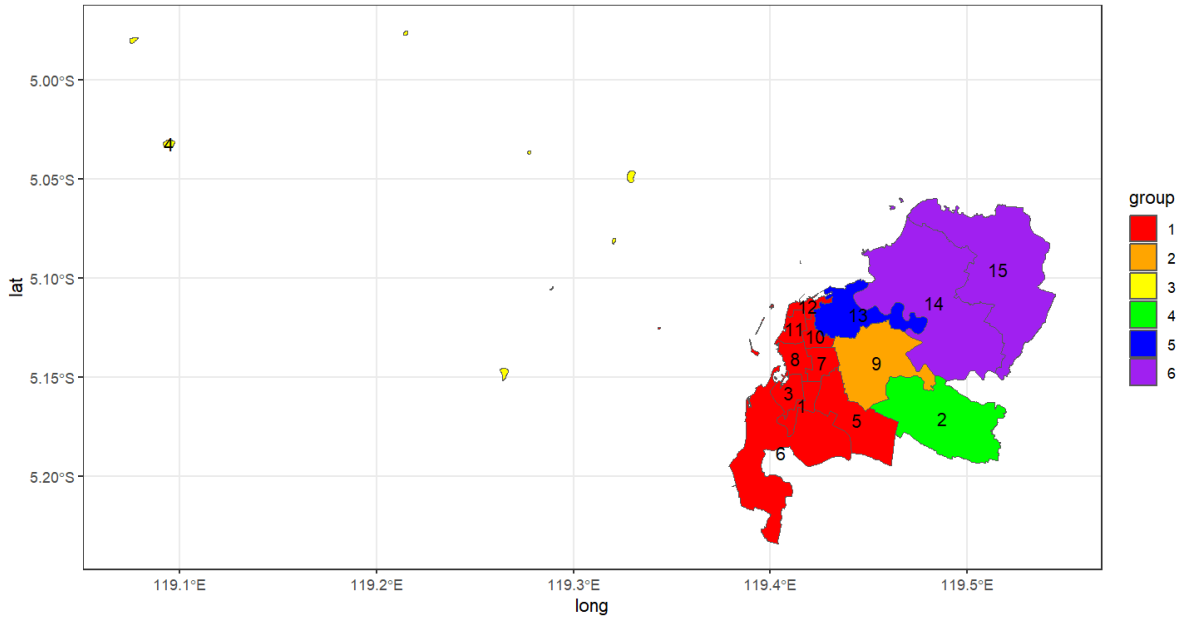

Figure 2. Map of District Grouping Based on GWNBR Modeling

Figure 2 shows six groups of Districts in Makassar City based on the combination of significant variables from the GWNBR model, with each color on the map representing the similarity of variable structures that significantly influence TB cases in the area. Spatially, it can be seen that the western and southern parts of the city are dominated by Group 1 (red), which shows fewer significant variable combinations compared to other groups. This region tends to have a simpler pattern of determinants, so that the factors affecting TB cases are more focused on specific variables.

Meanwhile, in the eastern and northeastern parts of the city, there are groups with more diverse combinations of significant variables (e.g., Groups 5 and 6). This pattern indicates that the factors influencing TB cases in these areas are more complex and are not only influenced by a single dimension but by several interrelated conditions. The more significant variables in a group, the greater the likelihood that the dynamics of TB cases in that area are influenced by the interaction of various social factors, density, health services, and environmental conditions.

3.2.2. MGWNBR Modeling

The MGWNBR model was developed by setting variable X_6 (CHB) as a global parameter, while other variables were estimated locally. The setting of X_6 as a global parameter was based on several considerations. First, the coefficient of variation (CoV) value of X_6 was relatively low, indicating minimal spatial variation. The Wald test results also did not show significant parameter

differences between regions, so the effect of X_6 was considered spatially constant and more appropriately modeled as a global parameter in the MGWNBR framework.

As in GWNBR, parameter significance is determined by comparing the $|Z_{score}|$ with $Z_{table} = 1.96$. This approach allows for the identification of variables whose effects are globally stable while maintaining local flexibility for other variables. The grouping of districts based on significant variables from the MGWNBR model is presented in Table 7.

Table 7. Clustering of Districts Based on Significant Variables from the MGWNBR Model

Group	Variable	District
1	X_1 (Population Density) X_4 (HIV/AIDS Cases) X_5 (Access to Safe Water) X_7 (Malnutrition)	Mamajang, Tamalate
2	X_1 (Population Density) X_3 (Health Facilities) X_4 (HIV/AIDS Cases) X_5 (Access to Safe Water) X_7 (Malnutrition)	Mariso, Rappocini, Makassar, Ujung Pandang
3	X_1 (Population Density) X_2 (Health Workers) X_3 (Health Facilities) X_4 (HIV/AIDS Cases) X_5 (Access to Safe Water) X_7 (Malnutrition)	Bontoala, Wajo, Ujung Tanah
4	X_1 (Population Density) X_3 (Health Facilities) X_4 (HIV/AIDS Cases) X_5 (Access to Safe Water) X_6 (CHB) X_7 (Malnutrition)	Sangkarrang
5	X_1 (Population Density) X_2 (Health Workers) X_3 (Health Facilities) X_4 (HIV/AIDS Cases) X_5 (Access to Safe Water) X_6 (CHB) X_7 (Malnutrition)	Manggala, Panakukkang, Tallo, Tamalanrea, Biringkanaya

As presented in Table 7, the grouping of districts in the MGWNBR model shows a different pattern compared to GWNBR. Almost all districts have significant variables X_1, X_4, X_5, and X_7 based on the criteria $|Z_{score}| > 1.96$. This indicates that when X_6 is treated as a global parameter, the structure of the local significance of other variables becomes clearer and more distributed. The grouping of districts based on MGWNBR modeling is visualized in Figure 3.

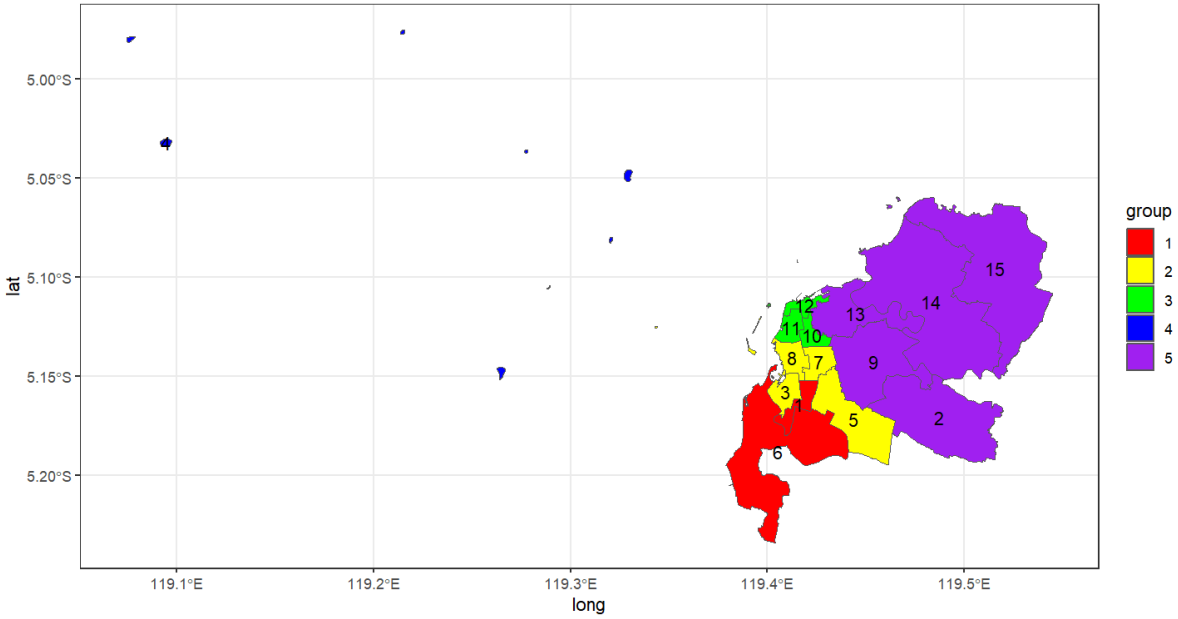

Figure 3. Map of District Clustering Results Based on MGWNBR Modeling

Figure 3 shows the results of district clustering in Makassar based on a combination of significant variables in the MGWNBR model, with each color representing the similarity of statistically significant TB determinant structures in the region. Spatially, it can be seen that the eastern and northeastern areas of Makassar City are dominated by Group 5 (purple), which includes districts with a fairly complex number of significant variables. This means that the increase in TB cases in this area is not only influenced by one or two factors, but by a combination of several conditions at once. Therefore, efforts to address this issue in this region should not be carried out partially, but should involve various sectors such as health, environment, and social sectors in order to achieve more effective results.

In contrast, the western and southern areas of the city tend to fall into Groups 1 and 2 (red and yellow), which have fewer significant variables. This indicates that in these areas, the determinants of TB are relatively more focused on specific variables, without involving all variables simultaneously. This pattern indicates that spatial heterogeneity occurs not only in the magnitude

of the coefficients but also in the structure of the variables significance.

Overall, the pattern in Figure 3 confirms that the MGWNBR approach is able to capture sharper spatial heterogeneity. The relatively structured cluster distribution, with a dominance of complex groups in the east and simpler groups in the west, shows that TB risk factors in Makassar City are not randomly distributed, but rather follow specific geographical patterns. This finding has important implications that TB control strategies should be tailored to the characteristics of local determinants in each district, rather than being applied uniformly across the entire region.

3.4. Selection of The Best Model

To determine the best model for explaining data variation, a comparison was made between the GWNBR and MGWNBR models. The evaluation was conducted using the deviance and AIC criteria. The results of this comparison are presented in Table 8.

Table 8. Model Fit Test Results

Criteria	GWNBR	MGWNBR
Deviance	176.095	29.993
AIC	200.559	172.133

Based on Table 8, the MGWNBR model produces a deviance value of 29.993, which is much smaller than GWNBR at 176.095. A lower deviance value indicates that the level of model mismatch with the data is smaller, so that MGWNBR is able to represent data patterns better. Similar results are also shown by the AIC value, where MGWNBR (172.133) has a lower value than GWNBR (200.559). Since models with lower AIC values are considered more optimal in balancing model fit and parameter complexity, this significant difference provides strong evidence that MGWNBR is superior in modeling data.

The substantial difference in deviance and AIC values indicates that the mixed approach can significantly improve model fit without adding excessive complexity. This finding suggests that not all parameters need to be treated as local parameters. The combination of local and global parameters in MGWNBR actually produces a model that is more efficient in capturing spatial variation.

Furthermore, in settings with a relatively small number of spatial units, specifying all parameters as local may lead to unstable coefficient estimates due to limited effective observations contributing to each local fit. A mixed specification provides a more balanced alternative, as it allows spatially varying effects to be captured while retaining global parameters for covariates that do not exhibit substantial geographic variation. By constraining selected effects to remain global,

the model reduces unnecessary complexity and improves estimation stability. Accordingly, MGWNBR can be considered not only statistically advantageous in terms of model fit, but also conceptually appropriate for small-area data structures where fully local parameterization may lead to overfitting and reduced interpretability.

As an illustration, Panakkukang District has a value of θ of 104.0517 with MGWNBR parameter estimates presented in Table 9.

Table 9. MGWNBR Model Parameter Estimation Results for Panakkukang District

Parameter	Estimate	$\lvert Z_{score}\rvert$	p-value	Description
$\hat{\beta}_0(u_i, v_i)$	3.4403	649.090	0.000	Significant
$\hat{\beta}_1(u_i, v_i)$	-0.0213	-7.228	0.000	Significant
$\hat{\beta}_2(u_i, v_i)$	-0.0119	-3.279	0.001	Significant
$\hat{\beta}_3(u_i, v_i)$	-0.0867	-4.578	0.000	Significant
$\hat{\beta}_4(u_i, v_i)$	0.1426	15.098	0.000	Significant
$\hat{\beta}_5(u_i, v_i)$	0.0132	51.363	0.000	Significant
$\hat{\gamma}_6$	0.0076	2.503	0.012	Significant
$\hat{\beta}_7(u_i, v_i)$	0.0101	11.294	0.000	Significant

Based on Table 9, the MGWNBR model for Panakkukang District is written as:

$$\ln(\mu_{Panakukkang}) = 3.4403 - 0.0213X_1 - 0.0119X_2 - 0.0867X_3 + 0.1426X_4 + 0.0132X_5 + 0.0076X_6 + 0.0101X_7$$

Based on the partial test with the criteria $|Z_{score}| > Z_{table} = 1.96$ at a significance level of 5%, all independent variables (X_1 to X_7) show a significant effect on the number of TB cases. This indicates that the structure of TB determinants in Panakkukang District is multifactorial, where demographic, health service, environmental, and social factors jointly contribute to case variation. The coefficients for X_1 (Population Density), X_2 (Health Workers), and X_3 (Health Facilities) are negative, indicating that an increase in these variables correlates with a decrease in the expected number of TB cases, assuming other variables remain constant. Substantively, these results may reflect that areas with higher population density or better availability of health services have more effective detection and control systems. However, this interpretation must be made with caution because causal relationships cannot be directly inferred from regression models. Conversely, the variables X_4 (HIV/AIDS Cases), X_5 (Access to safe water), X_6 (CHB), and X_7 (Malnutrition) have positive coefficients, indicating that increases in these variables correlate with an increase in

the number of TB cases. Epidemiologically, the positive relationship between HIV/AIDS and TB cases is consistent with the literature, given that HIV is a major risk factor for TB. Meanwhile, the positive relationship between safe water access and CHB may reflect regional characteristics with better reporting systems or the presence of other factors that are not fully controlled in the model.

4. CONCLUSIONS

Based on the results of the analysis and model comparison, it can be concluded that the MGWNBR approach provides better performance than GWNBR in modeling the number of TB cases in Makassar City. The advantage of MGWNBR shows that not all explanatory variables require local parameter modeling. The combination of local and global parameters is able to capture spatial variations more proportionally, resulting in more stable and interpretable estimates. In the context of a relatively small observation area, the mixed approach becomes more relevant because it can reduce the potential for estimation instability that often arises in models with all parameters being local. Thus, MGWNBR is not only statistically superior but also more conceptually appropriate in representing the spatial characteristics of TB case data at a small regional scale.

ACKNOWLEDGMENTS

The author gratefully acknowledges financial support from the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia through the 2025 Master's Thesis Research (PTM) scheme (Contract Nos. 084/C3/DT.05.00/PL/2025 and 2882/UN36.11/TU/2025). Appreciation is extended to the Makassar City Health Office for providing the data, to the Postgraduate Program of Makassar State University especially the Statistics Study Program and to the supervising lecturers for their guidance. The author also wishes to thank Nurul Aulya Bakri, Wanda Yudi, and Dewi Aprilia Wardani Syam, postgraduate students in Statistics at Universitas Negeri Makassar, for their valuable assistance in collecting TB data.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] WHO, Tuberculosis, 2025. https://www.who.int/news-room/fact-sheets/detail/tuberculosis (Accessed March 3, 2025).

[2] WHO, Global Tuberculosis Report 2024, World Health Organization, Geneva, 2024.

[3] WHO, Global Tuberculosis Report 2023, World Health Organization, Geneva, 2023.

[4] Direktorat Jenderal Pencegahan dan Pengendalian Penyakit, Laporan Program Penanggulangan Tuberkulosis Tahun 2022, Kementerian Kesehatan RI, Jakarta, 2023.

[5] Direktorat Jenderal Pencegahan dan Pengendalian Penyakit, Laporan Program Penanggulangan Tuberkulosis Tahun 2023, Kementerian Kesehatan RI, Jakarta, 2024.

[6] Rismayanti, Muh. Arman Nyomba, Aliyyah Ansariadi, Alika Tasya Devana, Analisis Determinan Tuberculosis Di Kota Makassar, Media Publ. Promosi Kesehat. Indones. 6 (2023), 290-295. https://doi.org/10.56338/mppki.v6i2.3038.

[7] P.K. Dunn, G.K. Smyth, Generalized Linear Models with Examples in R, Springer New York, 2018. https://doi.org/10.1007/978-1-4419-0118-7.

[8] D. Handayani, A.F. Artari, W. Safitri, W. Rahayu, V.M. Santi, Count Regression Models for Analyzing Crime Rates in the East Java Province, J. Phys.: Conf. Ser. 2123 (2021), 012028. https://doi.org/10.1088/1742-6596/2123/1/012028.

[9] E. Espinoza, U. Saputri, F. Hafiz Fadilah, D. Devianto, Modeling the Count Data of Public Health Service Visits with Overdispersion Problem by Using Negative Binomial Regression, J. Phys.: Conf. Ser. 1940 (2021), 012021. https://doi.org/10.1088/1742-6596/1940/1/012021.

[10] R. Yotenka, A. Banapon, Modelling the Number of Tuberculosis (TB) Cases in Indonesia Using Poisson Regression and Negative Binomial Regression, in: Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019), Atlantis Press, Paris, France, 2020. https://doi.org/10.2991/assehr.k.201010.007.

[11] A. Agresti, Foundations of Linear and Generalized Linear Models, Wiley, 2015.

[12] D.R. Putri, M. Fathurahman, S. Suyitno, Pemodelan Jumlah Kasus Tuberkulosis Paru Di Indonesia Dengan Geographically Weighted Negative Binomial Regression, EKSPONENSIAL 15 (2024), 49. https://doi.org/10.30872/eksponensial.v15i1.1303.

[13] R.E. Caraka, H. Yasin, Geographically Weighted Regression (GWR): Sebuah Pendekatan Regresi Geografis, Mobius, Indonesia, 2017.

[14] H. Khaulasari, R. Antonius, Model Mixed Geographically Weighted Poisson Regression Dengan Pembobot Fungsi Kernel Fixed Bi-Square Pada Penderita Tuberculosis di Surabaya, Statistika 7 (2019), 71-83.

[15] C. Nisa, M. Nur Aidi, I.M. Sumertajaya, Geographically Weighted Negative Binomial Regression Modeling of Tuberculosis Cases with Distribution Evaluation, Int. J. Sci. Res. Sci. Eng. Technol. (2020), 279-285. https://doi.org/10.32628/ijsrset1207473.

[16] M.L.S. Putera, L. Wahyunita, F. Yusup, Spatial Modelling of COVID-19 Confirmed Cases in Kalimantan, Indonesia: How Neighborhood Matters?, Walailak J. Sci. Technol. 18 (2021), 22120.

[17] H. Zeanova, P. Taniwan, R.A. Putri, D.Y. Faidah, Analisis Faktor Penyebab Penyakit TBC di Jawa Barat

Menggunakan Regresi Binomial Negatif, J. Ilm. Pendidik. Mat. Mat. Stat. 5 (2024), 2284-2302. https://doi.org/10.46306/lb.v5i3.

[18] A. Aswi, S. Sukarna, N. Nurhilaliyah, Pemetaan Kasus Tuberkulosis Di Provinsi Sulawesi Selatan Tahun 2020 Menggunakan Model Bayesian Spasial BYM Dan Leroux, J. Math. Comput. Stat. 4 (2021), 114-123. https://doi.org/10.35580/jmathcos.v4i2.32755.

[19] A.A. Nurfajrin S, N. Sunusi, E.T. Herdiani, Modeling Mixed Geographically Weighted Negative Binomial Regression on the Number of Tuberculosis Cases in South Sulawesi, Commun. Math. Biol. Neurosci. 2023 (2023), 126. https://doi.org/10.28919/cmbn/8267.

[20] C. Nisa, M.N. Aidi, I.M. Sumertajaya, Kajian Variance Mean Ratio Pada Simulasi Sebaran Data Binomial Negatif, Indones. J. Stat. Appl. 4 (2020), 615-626. https://doi.org/10.29244/ijsa.v4i4.689.

[21] P.C. Ambarwati, I. Indahwati, M.N. Aidi, Kajian Simulasi Overdispersi Pada Regresi Poisson Dan Binomial Negatif Terboboti Geografis Untuk Data Balita Gizi Buruk, Indones. J. Stat. Appl. 4 (2020), 484-497. https://doi.org/10.29244/ijsa.v4i3.684.

[22] H. Yasin, B. Warsito, A.R. Hakim, Regresi Spasial (Aplikasi dengan R), Team WADE, 2020.

[23] N. Delvia, M. Mustafid, H. Yasin, Geographically Weighted Negative Binomial Regression Untuk Menangani Overdispersi Pada Jumlah Penduduk Miskin, J. Gaussian 10 (2021), 532-543. https://doi.org/10.14710/j.gauss.v10i4.33106.

[24] P.F. Utami, A. Rusgiyono, D. Ispriyanti, Pemodelan Semiparametric Geographically Weighted Regression Pada Kasus Pneumonia Balita Provinsi Jawa Tengah, J. Gaussian 10 (2021), 250-258. https://doi.org/10.14710/j.gauss.v10i2.30945.

[25] H. Yasin, I. Suryani, P. Kartikasari, Graphical Interface of Geographically Weighted Negative Binomial Regression (GWNBR) Model Using R-Shiny, J. Phys.: Conf. Ser. 1943 (2021), 012155. https://doi.org/10.1088/1742-6596/1943/1/012155.