Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 1, 2025
DOI: 10.30829/zero.v9i1.25006
Page: 228-238

Geographically Weighted Negative Binomial Regression (GWNBR) Modeling of Tuberculosis (TB) in North Sumatra

[1] Afrila Ayunda
Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

[2] Ismail Husein
Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

[3] Prof Ulrich Faigle
Mathematisches Institut, Universität zu Köln, Köln, Germany

Article Info

Article history:

Accepted, 28 May 2025

Keywords:

GWNBR;
NBR;
North Sumatra;
Spatial Modeling;
Tuberculosis.

ABSTRACT

Tuberculosis (TB) remains a critical public health issue in North Sumatra, Indonesia, calling for accurate statistical techniques to understand its spatial distribution and related risk factors. This study compares three methods – Poisson Regression, Negative Binomial Regression (NBR), and Geographically Weighted Negative Binomial Regression (GWNBR) – to model TB cases across 33 districts and cities in 2022. An overdispersion test showed significant variance, making the Poisson model inappropriate. The NBR approach identified the number of medical staff as the only significant variable, yielding an AIC of 478.31. Meanwhile, a Breusch–Pagan test revealed significant spatial heterogeneity, supporting the use of GWNBR. This method captured spatially varying relationships between TB incidence and its covariates, yielding an AIC of 512.34 and offering more localized insights. These results underscore the value of spatially adaptive modeling techniques for analyzing disease patterns, providing evidence to guide targeted, area-specific public health policies and interventions.

Corresponding Author:

Afrila Ayunda,
Department of Mathematics,
Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: Afrila0703212066@uinsu.ac.id

1. INTRODUCTION

Tuberculosis (TB) remains a serious public health threat and is the second leading cause of death globally, after coronary heart disease[1][2] This translates to approximately 824,000 cases and 93,000 deaths each year, or roughly 11 people dying every hour [3][4]. In Indonesia, the disease is highly prevalent, with the country ranking second in the world for pulmonary TB cases, following India[5]. Its occurrence is influenced by a range of factors, including population density[6] and clean and healthy lifestyle habits (PHBS), which play an important role in prevention and control within communities.

To model the factors associated with TB incidence, statistical approaches for count data are required. In Poisson regression, the equidispersion assumption must be satisfied, meaning the mean and the variance are assumed to be equal. However, in many TB datasets, this assumption is violated due to overdispersion, where the variance is larger than the mean[7]. In such instances, the Negative Binomial (NB) model is generally preferred, as it better accommodates overdispersed data[8].

The Geographically Weighted Negative Binomial Regression (GWNBR) has emerged as an effective approach for addressing both overdispersion and spatial heterogeneity. GWNBR allows parameter estimates to vary across locations, making it especially suitable for spatial count data with significant variability across areas[9][10][11]. Studies conducted in Indonesia have demonstrated the strength of GWNBR for capturing spatially varied patterns in disease data, yielding more accurate and robust insights for public health interventions [12][13].

Compared to previous studies such as Prasenda et al. (2024), which were limited to analyzing TB incidence among children in 30 sub-districts in Bandung, this study makes several novel contributions. First, it broadens the geographical scope by covering 33 districts and cities across North Sumatra, a province characterized by more complex topographical, socio-economic, and infrastructural diversity. Unlike Bandung, which is largely urban and relatively homogenous in healthcare infrastructure, North Sumatra exhibits sharp contrasts between urban centers like Medan and remote areas such as Nias or Mandailing Natal. These disparities intensify the effects of spatial heterogeneity, making localized modeling not only more informative but also crucial for targeted policy-making. Second, this study incorporates a richer set of covariates, including population density, clean and healthy living behaviors (PHBS), waste management, access to safe drinking water, and the availability of medical personnel—factors that interact differently across districts due to uneven development and resource distribution.

Third, the application of the GWNBR model allows the detection of varying relationships between these predictors and TB incidence across space, overcoming the limitations of global models like Poisson or NB regression that assume spatial stationarity. This makes the model more responsive to local health conditions and better suited for informing district-level public health interventions in spatially heterogeneous provinces such as North Sumatra.

2. RESEARCH METHOD

2.1 Reasearch Design

This research adopts a quantitative design based on a spatial statistical framework to analyze the incidence of Tuberculosis (TB) across 33 districts and cities in North Sumatra in 2022. The methods used for modeling include Poisson Regression, Negative Binomial Regression (NBR), and Geographically Weighted Negative Binomial Regression (GWNBR).

2.2 Poisson Regression Model

Poisson regression is an extension of the global regression approach, where the response variable represents a count of cases and, therefore, can only take non-negative values[14][15]. Poisson Regression is a statistical approach frequently applied to model data that represent counts. This method is used to model the relationship between a dependent variable and one or more independent variables, based on the given assumption where the dependent variable Y is assumed to have a Poisson distribution [16].

$$Y_i \sim \text{Poisson}(\mu_i),\ \mu_i = \exp(X_i\beta) \quad (1)$$

Where:

Y_i: Number of TB cases in area i,

μ_i: Mean of Y_i,

X_i:Vector of covariates for area i,

β: Vector of regression coefficients.

2.3 Negative Binomial Regression Model

Negative Binomial Regression (NBR) assumes that the response variable Y_i follows a Negative Binomial distribution derived from a Poisson–Gamma mixture [[17]. This approach is used when the Poisson model is inappropriate due to overdispersion (i.e., when the variance is larger than the mean) [5], Under the NBR, Y_i is modeled as:

$$Y_i \sim NB(\mu_i, \theta),\ \mu_i = \exp(X_i\beta) \quad (2)$$

With the variance defined as:

$$Var(Y_i) = \mu_i + \theta\mu_i 2 \quad (3)$$

Where:

Y_i: Number of TB cases in area i,

μ_i: Mean of Y_i,

θ: dispersion parameter.

Assumptions of NBR:

NBR assumes that the count data arise from a Poisson process with an added Gamma-distributed error term, making it suitable for overdispersed data. This allows for more accurate estimation when the data's variance is larger than its mean, unlike Poisson regression.

2.4 Spatial Heterogeneity

There are two types of spatial testing: spatial heterogeneity and spatial dependence. To evaluate whether spatial heterogeneity exists within the data, the Breusch–Pagan test can be performed[18][19]. This test is used to assess whether a global model is adequate or if a spatially adaptive method, like GWNBR[20], is required to capture spatial heterogeneity in the data.

$$BP = \left(\frac{1}{2}\right) f^T Z\left(Z^T Z\right)^{-1} Z^T f \sim X^2_{(p)} \tag{4}$$

Where:

$f = (f_1, f_2, \ldots, f_n)^T$ with $f_i = \frac{e_i^2}{\sigma^2} - 1$

e_i = the error term for the i-th observation

σ^2 = thee error variance

Z = an $n \times (p+1)$ matrix containing a column of constants

the null hypothesis H_0 is rejected if the BP test statistic is greater than $X^2_{(p,a)}$ or if the p value is less than α. This means that the variances across locations are different.

2.5 Geographically Weighted Negative Binomial Regression Model

GWNBR estimation is a highly effective approach for modeling overdispersed count data that exhibit spatial heterogeneity[21][22]. This method produces location-specific parameter estimates, allowing each site to have its own set of regression coefficients[23][24].

$$Y_i \sim NB\left(\mu_i, \theta\right) \quad \mu_i = \exp\left(\beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) X_{ik}\right) \tag{5}$$

Where:

Y_i: number of TB cases in are i

μ_i: Mean of Y_i,

X_{ik}: Value of the k-th covariate in are i,

$\beta_0(u_i, v_i)$: location-dependent intercept,

$\beta_k(u_i, v_i)$: location-dependent coeffient for covariate k.

2.6 Model Selection Criteria

1. Akaike Information Criterion (AIC)

   The Akaike Information Criterion (AIC) is used to compare the quality of different statistical models by assessing both their fit and complexity. The model with the lowest AIC value is generally preferred, as it achieves an optimal balance between goodness of fit and parsimony[25].

   AIC formula:

$$AIC = 2p - 2 \ln L(\hat{\theta}) \tag{6}$$

Where:

$L(\hat{\theta})$: The likelihood value of the estimated model, derived from the log-likelihood.

p: The number of estimated parameters in the model.

2. McFadden's R-Squared

   In Poisson and Negative Binomial regression models, the traditional R^2 measure is not applicable. Instead, McFadden's R-Squared is used as an indicator of model fit, ranging from 0 to 1, where higher values reflect a better fit of the model to the data.

   McFadden's R-Squared formula:

$$R^2_{McFadden} = 1 - \frac{\ln L_{model}}{\ln L_{null}} \tag{7}$$

Where:

$\ln L_{model}$: The log-likelihood of the estimated model (including covariates).

$\ln L_{null}$: The log-likelihood of the null model (intercept-only).

2.7 Research Procedure

The following steps outline the analytical process for identifying the factors associated with the number of Tuberculosis (TB) cases in North Sumatra Province:

1. Collect the data on the total number of TB cases in North Sumatra for the year 2022.
2. Provide an overall description of the pulmonary TB case data, along with an explanation of the potential explanatory variables.
3. Check for signs of multicollinearity among the independent variables using the Variance Inflation Factor (VIF).
4. Perform Poisson Regression Modeling by:
   a. Estimating the parameters using the Maximum Likelihood Estimation (MLE) approach.
   b. Assessing the significance of the Poisson model parameters both globally, using the Maximum Likelihood Ratio Test (MLRT), and individually, using the Wald test.
5. Conduct an overdispersion test.
6. Build the Negative Binomial Regression model through the following steps:
   a. Estimate the parameters using the MLE method.
   b. Evaluate the significance of the model parameters, both globally via the MLRT and individually via the Wald test.
7. Test for spatial heterogeneity by applying the Breusch–Pagan test.
8. Develop the Geographically Weighted Negative Binomial Regression (GWNBR) model, which involves the following steps:
   a. Measure the Euclidean distances between observations based on their latitude and longitude coordinates.
   b. Identify the optimal bandwidth using a minimum value criterion.
   c. Calculate the adaptive bi-square kernel weighting scheme.
   d. Estimate the GWNBR model parameters using the MLE method.
   e. Compare the GWNBR model with the NBR model to assess parameter consistency.
   f. Evaluate the significance of the GWNBR parameters globally using the MLRT and individually using the Wald test.
   g. Choose the best-fit model by comparing NBR and GWNBR based on the AIC statistic.
   h. Interpret the results derived from the GWNBR model.
9. Formulate conclusions based on the results of the analyses conducted.

3. RESULT AND ANALYSIS

3.1 Characteristics of Tuberculosis Cases in North Sumatra in 2022

North Sumatra is a province in Indonesia with a high prevalence of tuberculosis (TB). The area consists of 33 regencies and cities, with the number of TB cases in 2022 showing considerable variation across these locations. Medan City recorded the highest count with 10,050 cases, followed by Deli Serdang Regency with 4,170 cases, and Langkat Regency with 1,927 cases. In contrast, the lowest numbers were observed in Pakpak Bharat Regency (117 cases), West Nias Regency (119 cases), and North Nias Regency (163 cases). The spatial distribution of TB cases across North Sumatra is presented in Figure 1.

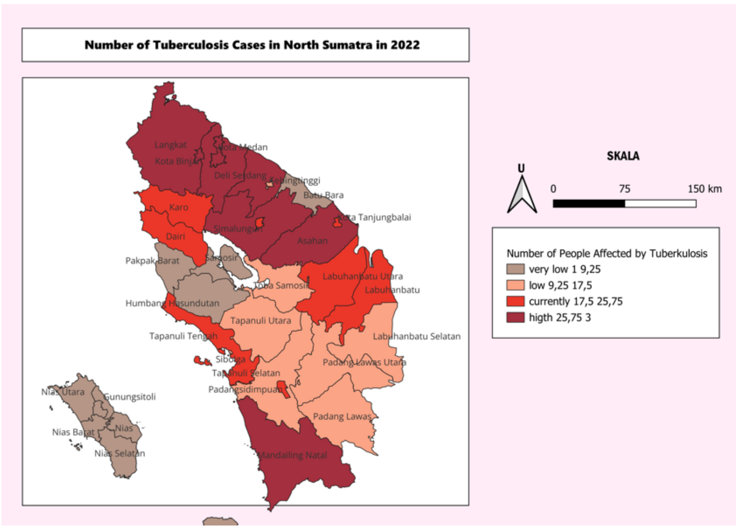

Figure 1. Distribution of Tuberculosis Cases in North Sumatra Province in 2022

As shown in Figure 1. Medan City recorded the highest number of TB cases, with a total of 10,050 cases, while Pakpak Bharat Regency had the lowest count, with only 117 cases.

3.2 Descriptive Analysis

Table 1. Descriptive Analysis

Variable	Maximum	Minimum	St.Deviation	Variance
Y	1	117	1786,561	3191801
X_1	9314	4482	2120,802	4,497802
X_2	83	0	26.2626	689.7289
X_3	37,4545	0	653,2394	4267218
X_4	23,714	0	6851,505	4694312
X_5	279	18	49.3921	2439.5852

Table 1. presents the descriptive statistics for the variables used in this study across the 33 districts and cities in North Sumatra for the year 2022. The results show a substantial variation between areas. The number of TB cases ranged from 117 to 1,786, indicating significant differences in disease distribution across the region. Similarly, other variables—including population density, household hygiene (PHBS), waste management, access to clean water, and the number of medical personnel—exhibit a wide range of values. This highlights the diverse conditions across North Sumatra, suggesting that such variability may play a role in the spatial distribution of tuberculosis cases.

3.3 Multicollinearity Test

Table 2. Correlation Coefficients Among Predictor Variables

	X_1	X_2	X_3	X_4
X_2	0.0615617399			
X_3	0.0008942562	0.37449311		
X_4	-0.0972435260	0.18337072	0.1951587204	
X_5	0.4984257033	-0.26524089	0.2797298474	0.07986768

Table 2. displays the correlation coefficients calculated among the predictor variables used in this study. In general, the correlation values remain below 0.95, suggesting that no strong relationships exist between the independent variables. This result indicates that multicollinearity is not a concern within the dataset, allowing all the predictors to be included in the model simultaneously without adversely affecting the reliability of the parameter estimates.

Table 3. Variance Inflation Factors (VIF) for the Predictor Variables

Predictor Variable	VIF
X_1	1.641522
X_2	1.692854
X_3	1.589282
X_4	1.107625
X_5	2.133815

The calculated VIF values for all the predictor variables are below the threshold of 10, indicating that no significant multicollinearity is present in the model. In other words, all the variables can be included simultaneously in the analysis without causing distortion or instability in the parameter estimates.

3.4 Poisson Regression Modeling

Table 4. Poisson Regression Model Parameter Estimates

	Estimation	Std.Error	Z.Value	Pr(>|z|)
(Intercept)	58,64	0.01591	368.601	< 2e-16 ***
X_1	-0,00001206	0,000004001	-3.014	0.00258 **
X_2	-0,004679	0,0003406	−13.736	< 2e-16 ***
X_3	0,000001885	0,00000007410	25.438	< 2e-16 ***
X_4	0,000009912	0,0000009614	10.310	< 2e-16 ***
X_5	0,01256	0,0001701	73.879	< 2e-16 ***

The Poisson regression results yielded a *null deviance* of 46,994.2 with 32 degrees of freedom, a *residual deviance* of 3,631.2 with 27 degrees of freedom, and an AIC value of 3,915. Based on these results, the Poisson regression model can be expressed as:

$$\mu_i = \exp(58{,}64 - 0{,}00001206x_{1_i} - 0{,}004679x_{2i} + 0{,}000001885x_{3i} + 0{,}000009912x_{4i} + 0{,}01256x_{5i}$$

From the parameter estimates presented in the table, it can be concluded that all the independent variables included in this study have an effect on the number of Tuberculosis cases across the districts and cities in North Sumatra.

3.5 Overdispersion

Prior to applying the Negative Binomial Regression, a dispersion test was conducted on the Poisson model by calculating the ratio of the deviance to its degrees of freedom (df):

$$\emptyset = \frac{\text{nilaidevience}}{\text{df}} \quad \emptyset = \frac{3631.2}{27} \quad \emptyset = 134.4873$$

The result, a dispersion value of approximately 134.4873, is far greater than 1, indicating the presence of overdispersion — a condition where the variance is significantly higher than the mean. This situation suggests that the Poisson model may not be suitable, as it can lead to inefficient estimates and higher error rates. As a result, the Negative Binomial Regression model was selected as a more appropriate approach for analyzing the factors associated with Tuberculosis cases in North Sumatra.

3.6 Negative Binomial Regression Modeling

Table 5. Negative Binomial Regression Model Parameter Estimates

	Estimation	Std.Error	z Value	Pr(>|z|)
(intercept)	56,71	1.901e-01	29.840	< 2e-16 ***
X_1	0,00004564	4.708e-05	0.969	0.332
X_2	-0,004640	3.862e-03	-1.201	0.230
X_3	0,000001898	1.500e-06	1.266	0.206
X_4	0,00001699	1.196e-05	1.421	0.155
X_5	0,001363	2.302e-03	5.919	3.23e-09 ***

Based on Table 5. it can be observed that among the five predictor variables included in the analysis, only X_5 (the number of medical personnel) has a statistically significant effect on the number of Tuberculosis cases in North Sumatra, with a p-value of 3.23e-09 (less than 0.05). The other variables were found to be statistically insignificant, as their p-values were greater than 0.05. The resulting Negative Binomial Regression Model is expressed as:

$$\mu_i = \exp(56{,}71 + 0{,}00004564x_{1_i} - 0{,}004640x_{2i} + 0{,}000001898x_{3i} + 0{,}00001699\, x_{4i} + 0{,}001363x_{5i}$$

The model yielded a *null deviance* of 189.782 with 32 degrees of freedom, a *residual deviance* of 34.096 with 27 degrees of freedom, and an AIC value of 478.31. These results indicate that the Negative Binomial Regression Model is well suited for analyzing the number of Tuberculosis cases, especially in the presence of overdispersion.

3.7 Spatial Effect Testing

Table 6. Spatial Heterogeneity and Dependence Test Results

Test	Statistic Value	Df	p-value	Conclusion
Breusc-Pagan	13.313	5	0.02062	Significant spatial heterogeneity detected (p < 0.05)
Moran's I	-0.0222	-	0.78660	No significant spatial autocorrelation detected (p > 0.05)

Based on the results shown in the table, the Breusch-Pagan test yielded a statistic value of 13.313 with a p-value of 0.02062, indicating significant spatial heterogeneity ($p < 0.05$). This result suggests that the residual variances are not constant across locations, implying a violation of the homoscedasticity assumption. Hence, a spatial modeling approach such as the Geographically Weighted Negative Binomial Regression (GWNBR) is required to properly accommodate this spatial heterogeneity.

Meanwhile, the Moran's I test resulted in a value of -0.0222 with a p-value of 0.78660, indicating no significant spatial autocorrelation ($p > 0.05$). In other words, the distribution of Tuberculosis cases does not exhibit a significant clustering pattern across the study area.

Due to the presence of spatial heterogeneity, a global modeling approach such as the Negative Binomial Regression is insufficient to fully capture the relationship between the predictor variables and the number of Tuberculosis cases across the entire study area. Therefore, it is necessary to proceed with a more advanced modeling approach, such as the Geographically Weighted Negative Binomial Regression (GWNBR), which is capable of capturing the variations in the effects of the predictors at each location.

3.8 Geographically Weighted Negative Binomial Regression (GWNBR) Modeling

Calculating Distance, Bandwidth, and Weights

In the GWNBR modeling process, the first step is to construct a spatial weighting matrix based on the distances between locations. These distances are calculated using the Euclidean formula, which is derived from the latitude and longitude coordinates of each district.

For example, the distance between South Tapanuli District (98.87, 2.55) and North Tapanuli District (99.02, 2.00) is calculated as:

$$d_{12} = \sqrt{(98.87\text{-}99.02)^2+(2.55\text{-}2.00)^2} = \sqrt{(\text{-}0.15)^2 + (0.55)^2} = \sqrt{0.0225+0.3025} = \sqrt{0.325} \approx 0.570$$

Since this value is in decimal degrees, it is then converted to kilometers by multiplying it by approximately 111.319 km (the average length of one degree). Thus:

$$0.570 \times 111.319 = 6.452 \text{ km}$$

After obtaining the distances between locations, the next step is to determine the bandwidth value, which serves as the spatial range for the weighting scheme. In this example, the bandwidth for South Tapanuli is set to 1.43, based on a calibration method such as cross-validation.

Once the bandwidth is defined, spatial weights can be calculated using the adaptive bisquare kernel. For instance, for South Tapanuli (i) and North Tapanuli (j), the weight is:

$$w_{ij} = \left[1\text{-}\left(\frac{0.570}{1.43}\right)^2\right]^2 = [1\text{-}(0.3986)^2]^2 = (1\text{-}0.1589)^2 = (0.8411)^2 \approx 0.7075$$

This result indicates that closer locations have a higher influence in the GWNBR model, allowing the method to capture spatial variations effectively.

Significance Testing of the GWNBR Model Parameters

Table 7. Parameter Estimation of the GWNBR Model

No.	Regency/City	Significant variable
1	Tapanuli Selatan	Intercept, X_1, X_5
2	Tapanuli Utara	Intercept, X_4
3	Tapanuli Selatan	Intercept, X_5
4	Nias	Intercept, X_5
5	Langkat	Intercept, X_5
6	Karo	Intercept, X_5
7	Deli Serdang	Intercept, X_5
8	Simalungan	Intercept, X_2, X_5
9	Asahan	Intercept, X_1, X_2X_3, X_5
10	Labuhan Batu	Intercept, X_1, X_2X_3
11	Dairi	Intercept, X_1, X_2X_5
12	Toba	Intercept, X_1, X_5
13	Mandailing Natal	Intercept, X_5
14	Nias Selatan	Intercept, X_5
15	Pakpak Barat	Intercept
16	Humbang Hasundutan	Intercept, X_5
17	Samosir	Intercept
18	Serdang Berdagai	Intercept
19	Batu Bara	Intercept
20	Padang Lawas Utara	Intercept
21	Padang Lawas	Intercept
22	Labuhan Batu Selatan	Intercept, X_2, X_3, X_5
23	Labuhan Batu Utara	Intercept
24	Nias Utara	Intercept, X_1, X_4
25	Nias Barat	Intercept, , X_1, X_4
26	Medan	Intercept
27	Pamatang Siantar	Intercept
28	Sibolga	Intercept, X_3
29	Tanjung Balai	Intercept
30	Binjai	Intercept, X_1, X_2, X_3, X_4, X_5
31	Tebing Tinggi	Intercept
32	Padang Sidempuan	Intercept
33	Gunung Sitoli	Intercept

The GWNBR estimation results (Table 7) reveal that the effects of the covariates vary across different districts. In urban areas such as Medan and Deli Serdang, variables like population density and waste management have a significant positive influence, indicating a higher TB incidence in areas with higher population densities and poorer waste management. Meanwhile, variables such as clean and healthy living behavior (PHBS), access to clean water, and availability of medical staff tend to have significant negative effects in certain areas, suggesting their role in reducing TB cases. These findings highlight the spatial heterogeneity of TB determinants and underscore the importance of tailored intervention strategies for each area.3.7.3. Optimal Model Selection.

Table 8. Optimal Model Selection.

Model	AIC value	Nilai McFadden's R-Squared
Regresi Poisson	3915.04	0.9174242
Regresi Binomial Negatif	478.3097	0.1160469
GWNBR	512.34	0.1500

Although the GWNBR model has a slightly higher AIC than the NBR, its ability to capture spatially varying parameter estimates provides more nuanced insights for targeted policy formulation. Its McFadden's R^2 value of 0.15 confirms an acceptable fit for a spatially adaptive count model.

Based on the results in Table 8, the lowest AIC value was obtained from the GWNBR model, which was 512.34. Subsequently, a plot will be created to examine the agreement between the estimated number of TB cases in North Sumatra and the actual observed values.

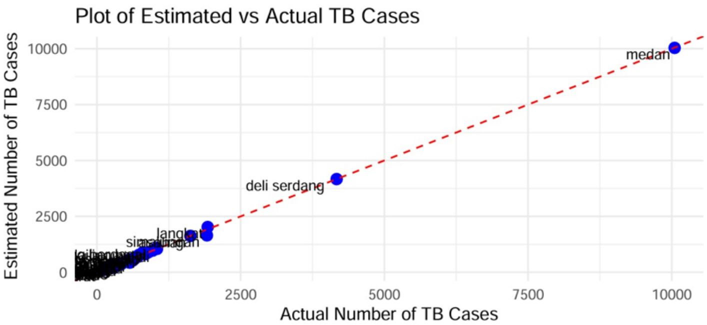

Figure 2. Scatter Plot of Estimated number versus Observed number.

As illustrated in Figure 2, the data points form a pattern that closely aligns with a diagonal line at a 45-degree angle, representing the ideal condition where the estimated number of TB cases matches the actual values. For example, when the actual number of TB cases in a district is 4,000, the predicted value using the GWNBR model is also approximately 4,000. This alignment indicates that the Geographically Weighted Negative Binomial Regression (GWNBR) method provides accurate predictions of TB case counts across districts in North Sumatra.

The effectiveness of the model is further supported by its Akaike Information Criterion (AIC) value of 512.34, which, despite being slightly higher than that of the global Negative Binomial model (AIC = 478.31), offers the crucial advantage of accounting for spatial heterogeneity in the data. This makes GWNBR more suitable for interpreting localized risk factors, especially in areas with highly variable conditions.

As an example of local model interpretation, we consider Medan City, for which the estimated model is:

$$\hat{\mu} = \exp\text{-}16.0032{+}0.0213X_1{+}0.00000041X_2{+}0.1981X_3{+}0.000321X_4{+}0.0059X_5)$$

This model suggests that increases in population density (X_1), waste management practices (X_3), and the proportion of medical personnel (X_5) are associated with higher TB incidence in Medan. This may indicate a reverse causality, where areas with more medical staff are also those with higher case burdens, rather than staff availability directly causing more cases.

In contrast, the model for Deli Serdang Regency is as follows:

$$\hat{\mu} = \exp\text{-}17.2085{+}0.0257X_1{+}0.00000032X_2{+}0.2186X_3{+}0.000275X_4{+}0.0101X_5)$$

Here, the negative coefficient for medical personnel (X_5) implies that greater healthcare access contributes to reducing TB cases. Meanwhile, variables such as population density and waste management continue to play a positive role in increasing TB incidence.

These contrasting results demonstrate that the influence of each factor is not uniform across all districts, highlighting the necessity of using a spatially adaptive model like GWNBR. This approach enables targeted public health strategies by identifying which variables drive TB transmission locally, ensuring that interventions are both efficient and context-specific.

Description of the Magnitude of Predictor Variable Effects

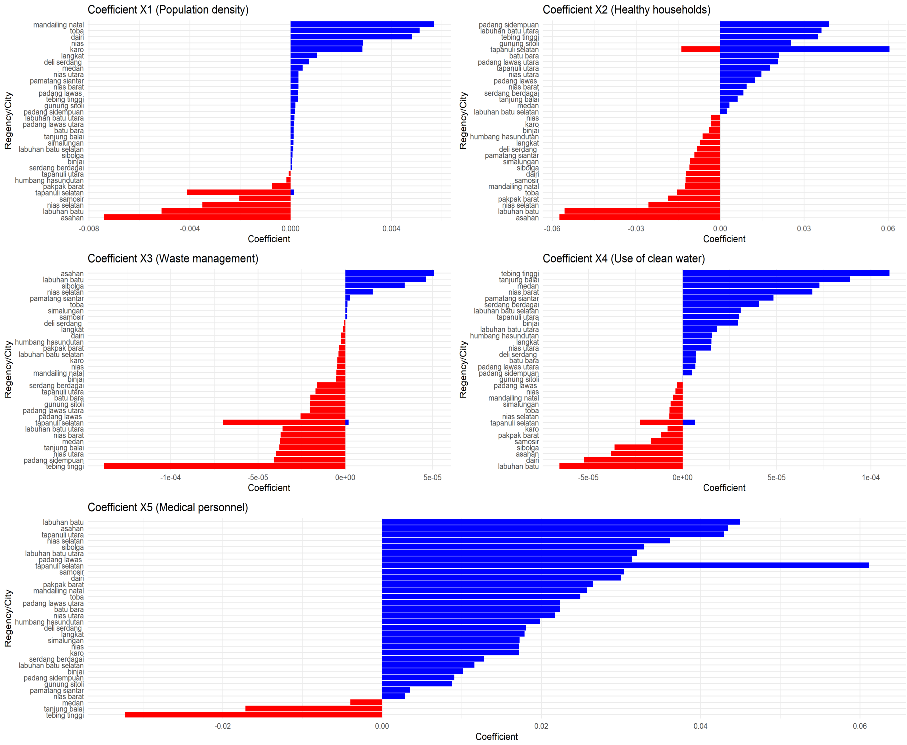

Figure 3. Distribution of Estimated Values by Regency/City in North Sumatra.

Figure 3 reveals a pattern of regionalization in the variable representing the percentage of households practicing clean and healthy living behavior (β_2), which falls within the range of −0.004128109 to −0.000215732. A similar spatial tendency is observed in the variable for access to clean drinking water (β_4), with estimated coefficients ranging from −0.000486057 to 0.0001445605. These findings suggest that these two variables have a relatively consistent negative association with TB incidence, particularly concentrated in the central and eastern regions of North Sumatra. This implies that healthier household practices and improved access to clean water are linked to lower TB case counts in those areas.In contrast, the variables for waste management (β_3) and availability of medical personnel (β_5) exhibit more dispersed patterns, without forming clear regional clusters. Their respective coefficient values range between −0.000201849 to 0.000213981 and −0.005851211 to −0.000152861, indicating that the influence of these predictors on TB incidence varies locally and does not follow a uniform regional trend.

Additionally, the population density variable (β_1) displays a positive spatial pattern, particularly in urban districts such as Medan and Deli Serdang, with coefficient values ranging from 0.00000485 to 0.00000694. This suggests that higher population density is associated with increased TB incidence, likely due to greater exposure and transmission risks in densely populated areas.Overall, the spatial distribution of local coefficient estimates highlights the importance of using a geographically weighted model like GWNBR. It effectively captures non-stationary relationships between predictors and TB cases, allowing for more context-specific interpretations and targeted public health interventions across districts in North Sumatra.

4. CONCLUSION

This study demonstrates the importance of spatially adaptive modeling in analyzing the distribution and determinants of Tuberculosis (TB) in North Sumatra. By applying the Geographically Weighted Negative Binomial Regression (GWNBR) model, the analysis captures local variations in how factors such as population density, clean and healthy living behaviors, waste management, water access, and medical personnel availability affect TB incidence.Although the GWNBR model had a slightly higher AIC compared to the global model, its strength lies in identifying localized risk patterns that global models often overlook. For example, urban areas like Medan and Deli Serdang require targeted interventions addressing environmental conditions, while rural districts like Langkat may benefit more from improving the distribution of healthcare workers. These spatially nuanced insights support the formulation of district-specific public health policies rather than one-size-fits-all strategies.

To strengthen disease control efforts, future studies should integrate spatio-temporal modeling and machine learning approaches to enhance prediction accuracy and enable early warning systems. Such advancements are especially critical for regions with limited resources, where timely and targeted interventions can significantly reduce the burden of TB.

5. REFERENCES

[1] T. Saifudin, M. Firmansyah, J. T. Victory, and M. Aisharezka, "Pemodelan Kasus Tuberkulosis di Indonesia dengan Metode GWPR Guna Mendukung SDGs 2030," vol. 21, no. 3, pp. 289–304, 2024, doi: http://dx.doi.org/10.12962/limits.v21i3.17041.

[2] H. Helmy, M. T. Kamaluddin, I. Iskandar, and Suheryanto, "Investigating Spatial Patterns of Pulmonary Tuberculosis and Main Related Factors in Bandar Lampung, Indonesia Using Geographically Weighted Poisson Regression," *Trop. Med. Infect. Dis.*, vol. 7, no. 9, 2022, doi: https://doi.org/10.3390/tropicalmed7090212.

[3] N. Siddalingaiah, K. Chawla, S. B. Nagaraja, and D. Hazra, "Risk factors for the development of tuberculosis among the pediatric population: a systematic review and meta-analysis," *Eur. J. Pediatr.*, vol. 182, no. 7, pp. 3007–3019, 2023, doi: https://doi.org/10.1007/s00431-023-04988-0.

[4] Dinas Kesehatan Sumatera Utara, "Profil Kesehatan Provinsi Sumatera Utara 2022," *Dinas Kesehatan Sumatera Utara*, vol. 2, pp. 1–466, 2022. Available at: https://dinkes.sumutprov.go.id/.

[5] D. R. Putri, M. Fathurahman, and S. Suyitno, "Pemodelan Jumlah Kasus Tuberkulosis Paru di Indonesia dengan Geographically Weighted Negative Binomial Regression," *EKSPONENSIAL*, vol. 15, no. 1, p. 49, May 2024, doi: https://doi.org/10.30872/eksponensial.v15i1.1303.

[6] M. Y. Darsyah, "Pemodelan Geographically Weighted Negative Binomial Regression (GWNBR) pada Kasus Malaria di Indonesia," *J. Litbang Edusaintech*, vol. 2, no. 2, pp. 149–164, 2021. Available: http://dx.doi.org/10.17509/ijost.v1i2.

[7] L. Zhang, J. Cheng, and C. Jin, "Spatial interaction modeling of OD flow data: Comparing geographically weighted negative binomial regression (GWNBR) and OLS (GWOLSR)," *ISPRS Int. J. Geo-Information*, vol. 8, no. 5, 2019, doi: https://doi.org/10.3390/ijgi8050220.

[8] B. B. Prasenda, M. D. Hermawan, M. Aisharezka, S. A. Tsauri, and N. Chamidah, "Pemodelan Jumlah Kasus Tuberkulosis pada Anak di Kota Bandung dengan Pendekatan Geographically Weighted Negative Binomial Regression," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 528–537, 2024, doi: https://doi.org/10.33379/gtech.v8i1.3881.

[9] A. A. Nurfajrin S, N. Sunusi, and E. T. Herdiani, "Modeling Mixed Geographically Weighted Negative Binomial Regression on the Number of Tuberculosis Cases in South Sulawesi," *Commun. Math. Biol. Neurosci.*, vol. 2023, pp. 1–13, 2023, doi: https://doi.org/10.28919/cmbn/8267.

[10] H. Yasin, I. Suryani, and P. Kartikasari, "Graphical interface of geographically weighted negative binomial regression (GWNBR) model using R-Shiny," *J. Phys. Conf. Ser.*, vol. 1943, no. 1, 2021, doi: https://doi.org/10.1088/1742-6596/1943/1/012155.

[11] M. J. T. L. Gomes, F. Cunto, and A. R. Silva, "Geographically weighted negative binomial regression applied to zonal level safety performance models," *Accid. Anal. Prev.*, vol. 106, pp. 254–261, 2017, doi: https://doi.org/10.1016/j.aap.2017.06.011.

[12] C. Nisa, M. N. Aidi, and I. M. Sumertajaya, "Geographically Weighted Negative Binomial Regression Modeling of Tuberculosis Cases with Distribution Evaluation," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 4099, pp. 279–285, 2020, doi: https://doi.org/10.32628/ijsrset1207473.

[13] A. O. Halim and N. Satyahadewi, "Factor Analysis on Poverty in Kalimantan Island with Geographically Weighted Negative Binomial Regression," *PIJMath*, vol. 4, no. 1, pp. 41–52, 2025, doi: https://doi.org/10.30985/pijmathvol4iss1pp41-52.

[14] F. W. R. Fadilah et al., "Geographically weighted negative binomial regression model to analysis of factors that influence on maternal mortality in Central Java Province," *AIP Conf. Proc.*, vol. 2202, pp. 1–6, 2019, doi: https://doi.org/10.1063/1.5141718.

[15] I. Suryani, H. Yasin, and P. Kartikasari, "Pemodelan Jumlah Kasus Demam Berdarah Dengue (Dbd) Di Jawa Tengah Dengan Geographically Weighted Negative Binomial Regression (GWNBR)," *J. Gaussian*, vol. 10, no. 1, pp. 136–148, 2021, doi: https://doi.org/10.14710/j.gauss.v10i1.29400.

[16] Z. Mar'ah, Z. Rais, and A. S. Haris, "Geographically Weighted Negative Binomial Regression (GWNBR) in Modeling the Risk Factors of Pneumonia Disease Among Toddlers in Central Sulawesi Province," *J. Stat. Its Appl. Teach. Res.*, vol. 5, no. 3, pp. 118–131, 2023, doi: https://doi.org/10.35580/variansiunm151.

[17] M. F. Rasyidin et al., "RAGAM: Journal of Statistics and Its Application," vol. 02, no. 02, pp. 1–15, Dec. 2023. Available at: https://jurnaluniv45sby.ac.id/index.php/JUMMA45/article/view/3166.

[18] E. Evadianti and Purhadi, "Pemodelan Jumlah Kematian Ibu di Jawa Timur dengan Geographically Weighted Negative Binomial Regression (GWNBR)," *J. Sains dan Seni ITS*, vol. 3, no. 2, pp. 182–187, 2014, doi: https://doi.org/10.12962/j23373520.v3i2.8128.

[19] J. Statistika, "Pemodelan dan Pemetaan Jumlah Kasus DBD di Kota Surabaya dengan GWNBR dan Flexibly Shaped Spatial Scan Statistic," vol. 4, no. 2, 2015, doi: https://doi.org/10.12962/j23373520.v4i2.11183.

[20] U. Menangani, O. Pada, and J. Penduduk, "GWNBR untuk menangani overdispersi pada jumlah penduduk miskin," vol. 10, pp. 532–543, 2021, doi: https://doi.org/10.20956/j.v19i1.21757.

[21] E. D. Safire and P. Purhadi, "Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Diabetes Melitus di Jawa Timur," *Inferensi*, vol. 6, no. 1, 2023, doi: https://doi.org/10.12962/j27213862.v6i1.12623.

[22] I. U. Utami, I. Setiawan, and D. Daniaty, "Modelling the number of HIV/AIDS in Central Sulawesi," *J. Phys. Conf. Ser.*, vol. 1763, no. 1, 2021, doi: https://doi.org/10.1088/1742-6596/1763/1/012046.

[23] K. Stunting et al., "Modeling GWNBR on Stunting Incidence in Malang," vol. 19, no. 1, pp. 163–171, 2022, doi: https://doi.org/10.20956/j.v19i1.21757.

[24] D. S. Rini, "Geographically Weighted Negative Binomial Regression untuk Jumlah Kasus Demam Berdarah Dengue Kabupaten/Kota Provinsi Bengkulu," vol. 1, pp. 736–744, 2018.

[25] R. GWNBR, "Kajian Efek Spasial pada Kasus Difteri dengan GWNBR," pp. 91–104, 2019, doi: https://doi.org/10.29244/IJSA.V3I1.185.