Geographically Weighted Negative Binomial Regression for Modeling Maternal Deaths in East Java Province

Calista Deva Salfatah[1], and Purhadi[2*]
[1,2]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
*Corresponding author: purhadi@its.ac.id

Received: 01 August 2025
Revised: 25 May 2026
Accepted: 3 June 2026

ABSTRACT – The number of maternal deaths in Indonesia increased from 3,572 in 2022 to 4,460 in 2023. East Java, as one of the provinces with the largest population in Indonesia, still faces challenges in reducing the number of maternal deaths, with 499 cases recorded, placing it in second place with a contribution of 11.19% of the total national cases. This study aims to model the number of maternal deaths in East Java using the Geographically Weighted Negative Binomial Regression (GWNBR) method, which considers spatial aspects and overdispersion of data. The GWNBR model, both with and without exposure, was applied using the Adaptive Bisquare, Gaussian, and Tricube Kernel weighting functions. The exposure variable represented by the number of pregnant women to adjust for inter-regional risk. Based on the model goodness-of-fit (AICc), the GWNBR model with exposure using the Adaptive Gaussian Kernel weighting function was the best model. This model identified four groups of districts/cities based on the similarity of significant predictor variables. Significant variables across all regions are the ratio of community health centers and hospitals per 100,000 population, while the percentage of obstetric complication treatment and postpartum family planning participants are only significant in some regions.
Keywords – GWNBR, Maternal Deaths, NBR, Overdispersion, Spatial.

I. INTRODUCTION

In realizing Indonesia Emas 2045, the transformation towards high-quality human resources begins with the family. Mothers play an important role in managing the household, educating children, and maintaining the health of family members. Monitoring the health status of mothers is very important because the Maternal Mortality Rate (MMR) reflects the level of health and welfare in a region. The MMR is the ratio of deaths of women due to complications during pregnancy, childbirth, and the postpartum period per 100,000 live births in a region during a certain period, excluding other factors such as incidental events [1]. The number of maternal deaths is directly related to the MMR, where the higher the number of maternal deaths, the higher the MMR in a region. Factors influencing maternal mortality require serious attention to enable appropriate preventive measures.

According to the 2020 Population Census, Indonesia's MMR decreased by 45%, from 346 in 2010 to 189 deaths per 100,000 live births in 2020 [2]. This figure is nearly in line with the target set in the 2020–2024 National Medium-Term Development Plan (RPJMN) of 183 per 100,000 live births. However, efforts to accelerate the decline in MMR still need to be made to achieve the SDGs target of 70 per 100,000 live births by 2030. This urgency is even more apparent with the increase in the number of maternal deaths in Indonesia from 3,572 cases in 2022 to 4,460 cases in 2023 [1].

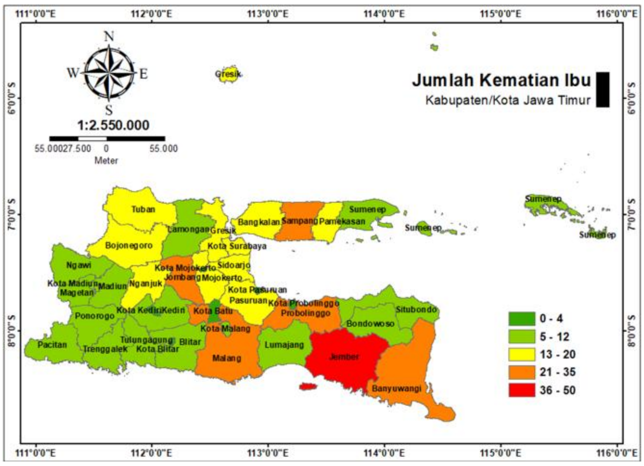

Figure 1 The Distribution of maternal deaths in East Java in 2023

East Java, one of the provinces with the largest population in Indonesia, still faces major challenges in reducing the number of maternal deaths. The latest data from the East Java Provincial Health Office shows that the MMR is recorded at 93.34 per 100,000 live births [1]. The number of maternal deaths in East Java in 2022 and 2023 remained stagnant at 499 cases, with the province ranking second in Indonesia with a contribution of 11.19% of the total 4,460 cases. The data on the number of maternal deaths in East Java is discrete and shows overdispersion, where the variance is greater than the mean, so Poisson regression is not suitable because the assumption of equidispersion is not met. To overcome this, this study uses Negative Binomial Regression (NBR). The exposure variable in this study is the number of pregnant women to adjust the level of maternal mortality risk in each district/city in East Java so that the inter-regional analysis is more proportional. Figure 1 shows the uneven distribution of cases, supported by the Breusch-Pagan test, which indicates spatial heterogeneity.

Previous studies, including Aini in 2020, also revealed spatial variation due to differences in characteristics across locations [3]. To account for this spatial heterogeneity, the Geographically Weighted Negative Binomial Regression (GWNBR) method is applied. GWNBR analysis can be used to create thematic maps illustrating the patterns of significant predictor variables in districts/cities in East Java based on latitude and longitude coordinates. This visualization clarifies the distribution of variable influences, identifies variations in causes of maternal mortality, and highlights differences in characteristics between regions. Therefore, policies and planning tailored to the characteristics of each region are needed to ensure the effectiveness and efficiency of programs to reduce the number of maternal deaths.

II. LITERATURE REVIEW

A. Poisson Regression and Overdispersion

The Poisson distribution models the count of independent random discrete events in a specific time or space interval [4]. For i-th observation ($i = 1,2,\dots,n$), the general form of the Poisson regression model with an exposure variable is shown in Equation (1).

$$\begin{aligned}\mu_i &= q_i \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_j x_{ij} + \cdots + \beta_k x_{ik}) \\ &= q_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})\end{aligned} \quad (1)$$

For the Poisson regression model without exposure, the value of $\boldsymbol{q_i} = \mathbf{1}$. The Poisson regression model assumes that $\boldsymbol{E(Y)} = \boldsymbol{Var(Y)}$ or equidispersion. If $\boldsymbol{Var(Y)} > \boldsymbol{E(Y)}$, overdispersion occurs, which can lead to inaccurate parameter estimates [5]. To check for overdispersion, Equation (2) is used.

$$\hat{\theta} = \frac{D_{PR}(\hat{\boldsymbol{\beta}})}{n-k-1} = \frac{2\left(\ln L(\widehat{\Omega}_{PR}) - \ln L(\widehat{\omega}_{PR})\right)}{n-k-1} \quad (2)$$

where $\ln L(\widehat{\Omega}_{PR}) = \sum_{i=1}^{n}\left[y_i \mathbf{x}_i^T\hat{\boldsymbol{\beta}} - \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}}) - \ln(y_i!)\right]$ and $\ln L(\widehat{\omega}_{PR}) = \sum_{i=1}^{n}\left[y_i\hat{\beta}_{0\omega} - \exp(\hat{\beta}_{0\omega}) - \ln(y_i!)\right]$. The estimate $\hat{\theta} > 1$ indicates overdispersion and $\hat{\theta} < 1$ indicates underdispersion. Meanwhile, $\hat{\theta} = 1$ indicates equidispersion [6].

B. Negative Binomial Regression (NBR)

Negative Binomial Regression (NBR) is one of the models commonly used to analyze count data, especially when overdispersion occurs in the response variable (Y). The probability function of the Negative Binomial distribution is formulated in Equation (3).

$$p(y;\mu,\theta) = \frac{\Gamma(y+\theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y+1)}\left(\frac{1}{1+\theta\mu}\right)^{\theta^{-1}}\left(\frac{\theta\mu}{1+\theta\mu}\right)^{y} \quad (3)$$

where $y = 0, 1, 2, \dots;\ \mu > 0;\ \theta > 0$. The dispersion parameter is denoted by θ, while $\Gamma(.)$ is the gamma function [7]. The expected value of the NBR model is $E(Y) = \mu$ and its variance is $Var(Y) = \mu + \theta\mu^2$ [8]. The parameters of the NBR model are estimated using the Maximum Likelihood Estimation (MLE) method. If there are n observations, the NBR model with an exposure variable for the i-th observation ($i = 1,2,\dots,n$) is shown in Equation (4).

$$\begin{aligned}\mu_i &= q_i \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_j x_{ij} + \cdots + \beta_k x_{ik}) \\ &= q_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})\end{aligned} \quad (4)$$

For the NBR model without an exposure variable, the value of q_i is assumed to be $q_i = 1$. Parameter significance testing is performed simultaneously and partially. Simultaneous testing is conducted using the Maximum Likelihood Ratio Test (MLRT) with the following hypothesis.

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

H_1: at least one $\beta_j \neq 0;\ j = 1, 2, \dots, k$

The simultaneous test statistic is given by Equation (5).

$$D_{NBR}(\hat{\boldsymbol{\beta}}, \hat{\theta}) = 2\left(\ln L(\widehat{\Omega}_{NBR}) - \ln L(\widehat{\omega}_{NBR})\right) \quad (5)$$

The rejection region for H_0 is $D_{NBR}(\hat{\boldsymbol{\beta}}, \hat{\theta}) > \chi^2_{(\alpha,k)}$ or $p_{value} < \alpha$, meaning that at least one parameter has a significant effect on the response variable in the model. If the simultaneous test rejects H_0, a partial test is conducted based on the following hypothesis.

$H_0: \beta_j = 0$

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

The partial test statistic is expressed in Equation (6).

$$Z_{j;NBR} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (6)$$

with $SE(\hat{\beta}_j) = \sqrt{\widehat{\text{var}}\,(\hat{\beta}_j)}$. The rejection region for H_0 is $|Z_{j;NBR}| > Z_{\left(\frac{\alpha}{2}\right)}$ or $p_{value} < \alpha$, meaning that the parameter has a significant effect on the response variable.

C. Spatial Heterogeneity Testing

Spatial heterogeneity testing is important to determine whether each observation location has unique characteristics. This test uses the Breusch-Pagan test statistic with the following hypothesis [9].

$\mathrm{H}_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$

H_1: at least one $\sigma_i^2 \neq \sigma^2$; $i = 1,2, \ldots, n$

The Breusch-Pagan test statistic is stated in Equation (7).

$$BP = \left(\frac{1}{2}\right)\mathbf{f}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{f} \tag{7}$$

The rejection region for H_0 is $BP > \chi^2_{(\alpha;k)}$ or $p_{value} < \alpha$, which means that spatial heterogeneity exists, so that the variances between locations are different.

D. Spatial Weighting Matrix

Weighting functions are divided into two types, namely fixed and adaptive kernel [10]. The spatial weighting functions used in this study are Adaptive Bisquare, Gaussian, and Tricube Kernels, which are formulated in Equations (8), (9), and (10).

1. Adaptive Bisquare Kernel Function

$$w_{ii^*} = \begin{cases} \left(1 - \left(\frac{d_{ii^*}}{h_i}\right)^2\right)^2 & , d_{ii^*} \leq h_i \\ 0 & , d_{ii^*} > h_i \end{cases} \tag{8}$$

2. Adaptive Gaussian Kernel Function

$$w_{ii^*} = \exp\left[-\frac{1}{2}\left(\frac{d_{ii^*}}{h_i}\right)^2\right] \tag{9}$$

3. Adaptive Tricube Kernel Function

$$w_{ii^*} = \begin{cases} \left(1 - \left(\frac{d_{ii^*}}{h_i}\right)^3\right)^3 & , d_{ii^*} \leq h_i \\ 0 & , d_{ii^*} > h_i \end{cases} \tag{10}$$

where $d_{ii^*} = \sqrt{(u_i - u_{i^*})^2 + (v_i - v_{i^*})^2}$ is the Euclidean distance and h_i is the bandwidth at location i-th.

E. Geogrographically Weighted Negative Binomial Regression (GWNBR)

GWNBR is an extension of the non-spatial NBR model that allows for spatial variation in parameters, resulting in varying local parameters at each observation location [11]. The GWNBR model with an exposure variable is formulated in Equation (11).

$$\begin{gathered} Y_i \sim NB\left[q_i \exp\left(\textstyle\sum_{j=0}^{k} \beta_j(u_i, v_i)\, x_{ij}\right), \theta(u_i, v_i)\right] \\ E(Y_i) = q_i \exp\left(\sum_{j=0}^{k} \beta_j(u_i, v_i)\, x_{ij}\right) \end{gathered} \tag{11}$$

where $x_{i0} = 1$; $i = 1, 2, \ldots, n$. In the GWNBR model without an exposure variable, the value of q_i is assumed to be $q_i = 1$.

Model similarity testing aims to evaluate whether there are significant differences between the NBR (global) and GWNBR (local) models, particularly in the regression coefficients produced with the following hypothesis.

$\mathrm{H}_0: \beta_j(u_i, v_i) = \beta_j$; $i = 1,2, \ldots, n$ and $j = 0, 1, 2, \ldots, k$

H_1: at least one $\beta_j(u_i, v_i) \neq \beta_j$

This test statistic is formulated in Equation (12).

$$F_{hit} = \frac{D_{NBR}(\hat{\boldsymbol{\beta}}, \hat{\theta}) / \mathrm{df}_{NBR}}{D_{GWNBR}(\hat{\boldsymbol{\beta}}(u_i, v_i), \hat{\theta}(u_i, v_i)\ ; i = 1,2, \ldots, n) / \mathrm{df}_{GWNBR}} \tag{12}$$

where $\mathrm{df}_{NBR} = k$ and $\mathrm{df}_{GWNBR} = nk$. The rejection region for H_0 is $F_{hit} > F_{(\alpha;\, \mathrm{df}_{NBR};\, \mathrm{df}_{GWNBR})}$ or $p_{value} < \alpha$, which means there is a significant difference between the two models.

The GWNBR model parameters are estimated using the MLE method, through the Newton-Raphson iteration process with the following steps [12].

1. Determine the initial estimate values for each parameter.
   $\hat{\boldsymbol{\delta}}_{(0)}(u_{i^*}, v_{i^*}) = \left[\hat{\theta}_{(0)}(u_{i^*}, v_{i^*}) \quad \hat{\beta}_{0(0)}(u_{i^*}, v_{i^*}) \quad \hat{\beta}_{1(0)}(u_{i^*}, v_{i^*}) \quad \cdots \quad \hat{\beta}_{k(0)}(u_{i^*}, v_{i^*})\right]^T$, where $\hat{\theta}_{(0)}(u_{i^*}, v_{i^*}) = \hat{\theta}_{(0)}$.
2. Construct the gradient vector $\mathbf{g}$.
   $\boldsymbol{g}^T\left(\hat{\boldsymbol{\delta}}_{(m)}(u_{i^*}, v_{i^*})\right)_{(k+2)\times 1} = \left[\frac{\partial \ln L(.)}{\partial \theta(u_{i^*}, v_{i^*})} \quad \frac{\partial \ln L(.)}{\partial \beta_0(u_{i^*}, v_{i^*})} \quad \frac{\partial \ln L(.)}{\partial \beta_1(u_{i^*}, v_{i^*})} \quad \cdots \quad \frac{\partial \ln L(.)}{\partial \beta_k(u_{i^*}, v_{i^*})}\right]$, where $L(.) = L(\boldsymbol{\beta}(u_{i^*}, v_{i^*}), \theta(u_{i^*}, v_{i^*}))$.

3. Construct the symmetric Hessian matrix $\mathbf{H}$.

$$\mathbf{H}\left(\hat{\boldsymbol{\delta}}_{(m)}(u_{i^*},v_{i^*})\right)_{(k+2)\times(k+2)} = \begin{bmatrix} \frac{\partial^2 \ln L(.)}{\partial(\theta(u_{i^*},v_{i^*}))^2} & \frac{\partial^2 \ln L(.)}{\partial\theta(u_{i^*},v_{i^*})\partial\beta_0(u_{i^*},v_{i^*})} & \cdots & \frac{\partial^2 \ln L(.)}{\partial\theta\partial\beta_k(u_{i^*},v_{i^*})} \\ & \frac{\partial^2 \ln L(.)}{\partial(\beta_0(u_{i^*},v_{i^*}))^2} & \cdots & \frac{\partial^2 \ln L(.)}{\partial\beta_0(u_{i^*},v_{i^*})\partial\beta_k(u_{i^*},v_{i^*})} \\ & & \ddots & \vdots \\ Simetris & & & \frac{\partial^2 \ln L(.)}{\partial(\beta_k(u_{i^*},v_{i^*}))^2} \end{bmatrix}.$$

4. Substitute the value of $\hat{\boldsymbol{\delta}}_{(0)}(u_{i^*},v_{i^*})$ into each element of the vector $\mathbf{g}$ and matrix $\mathbf{H}$ to obtain $\mathbf{g}\left(\hat{\boldsymbol{\delta}}_{(m)}(u_{i^*},v_{i^*})\right)$ and $\mathbf{H}^{-1}\left(\hat{\boldsymbol{\delta}}_{(m)}(u_{i^*},v_{i^*})\right)$.
5. Perform iteration starting from $m = 0$ with the equation $\hat{\boldsymbol{\delta}}_{(m+1)}(u_{i^*},v_{i^*}) = \hat{\boldsymbol{\delta}}_{(m)}(u_{i^*},v_{i^*}) - \mathbf{H}^{-1}\left(\hat{\boldsymbol{\delta}}_{(m)}(u_{i^*},v_{i^*})\right)\mathbf{g}\left(\hat{\boldsymbol{\delta}}_{(m)}(u_{i^*},v_{i^*})\right)$.
6. Repeat steps 2 to 5 until convergence is reached. The iteration will stop if $\left\|\hat{\boldsymbol{\delta}}_{(m+1)}(u_{i^*},v_{i^*}) - \hat{\boldsymbol{\delta}}_{(m)}(u_{i^*},v_{i^*})\right\| < \varepsilon$, where $\varepsilon > 0$ is a small threshold (e.g., 0.00001).
7. If convergence is reached at the M-th iteration, the parameter estimate $\hat{\boldsymbol{\delta}}(u_{i^*},v_{i^*}) = \left[\hat{\theta}_{(M)}(u_{i^*},v_{i^*}) \quad \hat{\boldsymbol{\beta}}^T_{(M)}(u_{i^*},v_{i^*})\right]^T$ is obtained.
8. Obtain the covariance matrix using the formula $\widehat{Cov}\left(\hat{\boldsymbol{\delta}}(u_{i^*},v_{i^*})\right) \underset{n\to\infty}{\cong} -\mathbf{H}^{-1}\left(\hat{\boldsymbol{\delta}}(u_{i^*},v_{i^*})\right)$.
9. Compute the standard error of $\hat{\beta}_j(u_{i^*},v_{i^*})$ as $SE\left(\hat{\beta}_j(u_{i^*},v_{i^*})\right) = \sqrt{\widehat{\mathrm{var}}(\hat{\beta}_j(u_{i^*},v_{i^*}))}$, where $\widehat{\mathrm{var}}(\hat{\beta}_j(u_{i^*},v_{i^*}))$ is the $(j+1)$ diagonal element of $\widehat{Cov}\left(\hat{\boldsymbol{\delta}}(u_{i^*},v_{i^*})\right)$.
10. Based on the central limit theorem, the partial test statistic is expressed as $Z_j = \frac{\hat{\beta}_j(u_{i^*},v_{i^*}) - \beta_j(u_{i^*},v_{i^*})}{SE\left(\hat{\beta}_{j(u_{i^*},v_{i^*})}\right)}$ which follows the standard normal distribution.
11. This iteration process is performed at locations $i^* = 1, 2, \dots, n$.

The hypothesis for testing the significance of parameters simultaneously in the GWNBR model is as follows.

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

H_1: at least one $\beta_j(u_i,v_i) \neq \beta_j;\ i = 1,2,\dots,n$

The simultaneous test statistic is given by Equation (13).

$$D_{GWNBR} = 2\left(\ln L(\hat{\Omega}_{GWNBR}) - \ln L(\hat{\omega}_{GWNBR})\right) \qquad (13)$$

The rejection region for H_0 is $D_{GWNBR} > \chi^2_{(\alpha,\mathrm{df}_{GWNBR})}$. If H_0 is rejected, the next step is the partial test.

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

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

The partial test statistic is expressed in Equation (14).

$$Z_{j;GWNBR} = \frac{\hat{\beta}_j(u_i,v_i)}{SE\left(\hat{\beta}_j(u_i,v_i)\right)} \qquad (14)$$

The rejection region for H_0 is $|Z_{j;GWNBR}| > Z_{\left(\frac{\alpha}{2}\right)}$.

F. *Maternal Mortality*

Maternal mortality is defined as the death of a woman during pregnancy, childbirth, and the postpartum period (42 days after delivery), excluding incidental causes. The causes of maternal mortality are generally divided into two categories: direct (obstetric complications) and indirect (pre-existing diseases present before or during pregnancy that are not directly related to obstetric complications) [13]. According to McCarthy and Maine's theory, causes of maternal mortality are categorized into three determinants: immediate determinants (directly influencing maternal mortality), intermediate determinants (influencing immediate determinants), and remote determinants (not directly influencing maternal mortality but important to consider) [14].

III. METHODOLOGY

A. *Data Sources*

The data used in this study is secondary data obtained from BPS in the form of East Java poverty data for 2023 and the East Java Provincial Health Office in the form of the East Java Provincial Health Profile for 2023, which includes the number of maternal deaths and suspected contributing factors. The observation unit covers 29 districts and 9 cities in East Java.

B. Research Variables

This study uses one response variable (Y), six predictor variables (X), and one exposure variable (q).

Table 1 Research variables

Variable	Description
Y	Number of maternal deaths
X_1	Percentage of obstetric complication treatment
X_2	Percentage of deliveries in health care facilities
X_3	Ratio of community health centers and hospitals per 100,000 population
X_4	Percentage of postpartum family planning participants
X_5	Percentage of K6 health service coverage for pregnant women
X_6	Percentage of poor population
q	Number of pregnant women

C. Analysis Steps

The following are the stages of analysis in this study.

1. Conduct descriptive analysis.
2. Conduct data pre-processing.
3. Visualize the distribution of each research variable.
4. Test the distribution of the response variable using the Kolmogorov-Smirnov method.
5. Detect multicollinearity among predictor variables by examining correlation and VIF values.
6. Model the NBR with and without exposure as formulated in Equation (4), including the following steps.
   a. Estimate the parameters using the MLE method.
   b. Test the significance of the parameters (simultaneous and partial tests).
7. Conduct the Breusch-Pagan test using Equation (7) to identify spatial heterogeneity.
8. Model the GWNBR with and without exposure as formulated in Equation (11), including the following steps.
   a. Calculate the Euclidean distance between observation locations.
   b. Determine the optimal bandwidth for each location using the Cross-Validation (CV) method.
   c. Compute the weighting matrix using the Adaptive Kernel functions in Equations (8), (9), and (10).
   d. Estimate the parameters using the MLE method.
   e. Test the significance of the parameters (simultaneously and partially).
   f. Test the similarity between the GWNBR and NBR models using Equation (12).
   g. Evaluate the GWNBR model according to the AICc value.
9. Select the best GWNBR model.
10. Interpret the best model.
11. Formulate conclusions and recommendations.

IV. RESULTS AND DISCUSSIONS

A. Descriptive Statistics

East Java has an area of 47,803.39 km^2 and the second largest population in Indonesia. Each district/city in this province has varying maternal health conditions. Various factors that influence the number of maternal deaths can be analyzed through data covering health services and socioeconomic conditions.

Table 2 Descriptive statistics of research variables

Variable	Mean	Variance	Minimum	Maximum
Y	12.97	104.67	0.00	50.00
X_1	97.95	396.54	20.00	154.00
X_2	92.49	41.14	78.40	101.40
X_3	3.82	2.05	2.26	9.03
X_4	59.88	275.27	17.70	96.60
X_5	78.04	123.18	54.70	97.60
X_6	10.29	18.67	3.31	21.76
q	12.97	104.67	0.00	50.00

Table 2 shows that the average number of maternal deaths in East Java in 2023 was 12.97 cases. Jember District recorded the highest number with 50 cases, while Kediri City, Madiun City, and Mojokerto City reported no cases of maternal deaths. The variance in the number of maternal deaths was 104.67, indicating data diversity.

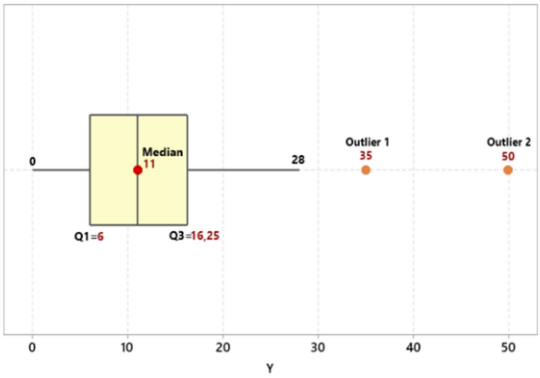

Figure 2 Boxplot of the number of maternal deaths

Figure 2 shows a right-skewed distribution, which means that most districts/cities have a number of maternal deaths concentrated at lower values.

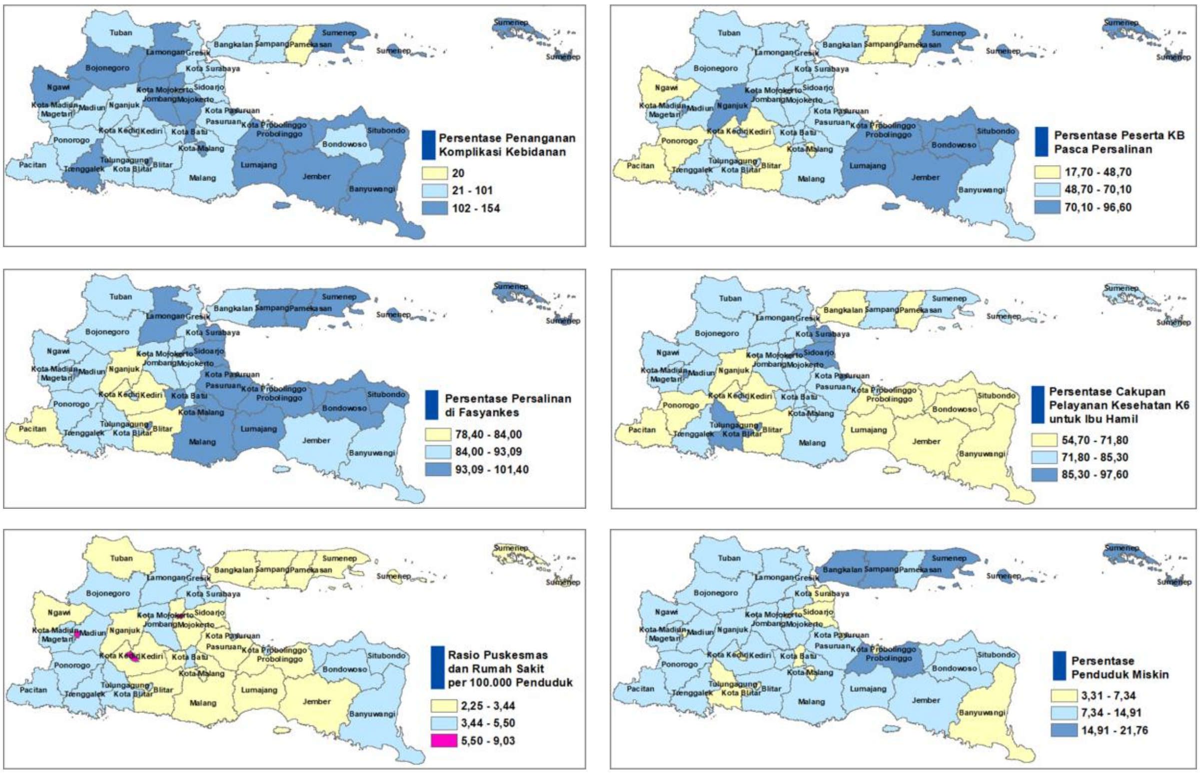

Figure 3 Thematic maps $X_1, X_2, X_3, X_4, X_5, X_6$

Figure 3 illustrates the distribution of each variable, which overall shows that some areas with adjacent locations tend to be included in the same group.

B. Data Distribution Testing

Distribution testing on the response variable aims to determine the appropriate probability distribution for the data. This test uses the Kolmogorov-Smirnov method. The analysis results obtained a *p-value* of 0.284, which is greater than the 10% significance level, thus failing to reject H_0. Therefore, the number of maternal deaths in East Java in 2023 is distributed according to a negative binomial distribution.

C. Overdispersion

The next step is to check for overdispersion based on the Poisson regression model estimation results. A deviation value of 119.98 with a degree of freedom of 31 was obtained, so the ratio of deviation to degrees of freedom is 3.8704, which is greater than 1. This indicates that the number of maternal deaths in East Java exhibits overdispersion. In addition, Table 2 shows that $Var(Y) > E(Y)$. This condition does not comply with the equidispersion assumption of the Poisson distribution. Therefore, the Negative Binomial Regression (NBR) model is more appropriate to use.

D. Multicollinearity Examination

Multicollinearity can be detected by examining the correlation coefficients and the Variance Inflation Factor (VIF) values between predictor variables.

Table 3 VIF of predictor variables

Variable	X_1	X_2	X_3	X_4	X_5	X_6
VIF	1.713	1.599	1.363	1.983	1.816	1.547

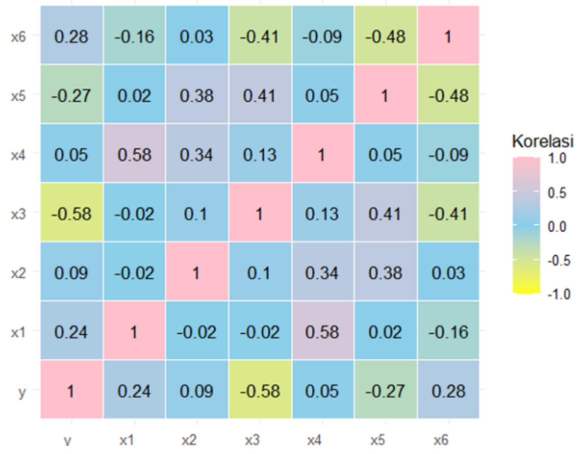

Figure 4 Heatmap of correlation matrix between variables

Correlation is a measure that indicates the strength of the relationship between two variables. The higher the absolute value, the stronger the relationship between the two variables. Table 3 shows that all predictor variables have a VIF < 2 or $R^2 < 0.5$, so there are no cases of multicollinearity. Therefore, all predictor variables can be used in the next stage of the analysis.

E. NBR Modeling of the Number of Maternal Deaths Without Exposure

Modeling the number of maternal deaths globally using the NBR method, where the influencing factors are uniform in each district/city. NBR modeling without exposure used a θ value of 5.4335, resulting in an AICc value of 251.6698. The parameter estimation results of this model are presented in Table 4. Parameter significance tests were performed simultaneously and partially.

The simultaneous test aims to evaluate whether the predictor variables simultaneously have a significant effect on the response variable with the following hypothesis.

$\mathrm{H}_0: \beta_1 = \beta_2 = \ldots = \beta_6 = 0$

H_1: at least one $\beta_j \neq 0; \; j = 1, 2, \ldots, 6$

Based on Equation (5), the value of D_{NBR} is 39.965, and $\chi^2_{(0,1;6)}$ is 10.645. With a significance level of 10%, the decision is to reject H_0 because $D_{NBR} > \chi^2_{(0,1;6)}$. This means that at least one predictor variable has a significant effect on the response variable in the model.

Table 4 NBR model parameter estimation without exposure

Parameter	Estimate	SE $(\hat{\beta}_j)$	Z_j	*P-value*
β_0	3.467670	1.709172	2.029	0.0511*
β_1	0.007440	0.006243	1.192	0.2424
β_2	0.011545	0.018925	0.610	0.5463
β_3	-0.650913	0.118684	-5.484	5.35×10^{-6}*
β_4	-0.003461	0.008192	-0.422	0.6756
β_5	-0.004275	0.011807	-0.362	0.7197
β_6	0.005703	0.027180	0.210	0.8352

*) significant with α=10%

After rejecting H_0 in the simultaneous test, a partial test is continued to determine which predictor variables are significant to the response variable with the following hypothesis.

$\mathrm{H}_0: \beta_j = 0$

$\mathrm{H}_1: \beta_j \neq 0; \; j = 1, 2, \ldots, 6$

The partial test is performed by comparing the test statistic $|Z_{j;NBR}|$ from Equation (6) with the critical value $Z_{0,05} = 1.645$. If $|Z_{j;NBR}| > Z_{0,05}$, the decision made is to reject H_0, indicating that the variable has a significant effect on the response variable in the model. Table 4 shows that X_3 (ratio of community health centers and hospitals per 100,000 population) is partially significant for the number of maternal deaths in districts/cities in East Java in 2023. The NBR model without exposure formed is as follows.

$$\hat{\mu} = \exp(3.467670 + 0.007440X_1 + 0.011545X_2 - 0.650913X_3 - 0.003461X_4 - 0.004275X_5 + 0.005703X_6)$$

F. NBR Modeling of the Number of Maternal Deaths with Exposure

NBR modeling with exposure used a θ value of 35.9019, resulting in an AICc value of 223.4322. The parameter estimation results of this model are presented in Table 5. Parameter significance tests were performed simultaneously and partially. The following is the hypothesis for the simultaneous test.

H_0: $\beta_1 = \beta_2 = \ldots = \beta_6 = 0$
H_1: at least one $\beta_j \neq 0$; $j = 1, 2, \ldots, 6$

Based on Equation (5), the value of D_{NBR} is 42.396, and $\chi^2_{(0,1;6)}$ is 10.645. With a significance level of 10%, the decision is to reject H_0 because $D_{NBR} > \chi^2_{(0,1;6)}$. This means that at least one predictor variable has a significant effect on the response variable in the model.

Table 5 NBR model parameter estimation with exposure

Parameter	Estimate	SE $(\hat{\beta}_j)$	Z_j	*P-value*
β_0	-2.892508	1.076404	-2.687	0.01148*
β_1	0.005842	0.003689	1.584	0.12344
β_2	0.007835	0.012126	0.646	0.52295
β_3	-0.042181	0.078612	-0.537	0.59539
β_4	-0.004239	0.005284	-0.802	0.42848
β_5	-0.011528	0.007282	-1.583	0.12354
β_6	0.035272	0.016908	2.086	0.04529*

*) significant with α=10%

After rejecting H_0 in the simultaneous test, a partial test is continued with the following hypothesis.

H_0: $\beta_j = 0$
H_1: $\beta_j \neq 0$; $j = 1, 2, \ldots, 6$

The partial test is performed by comparing the test statistic $|Z_{j;NBR}|$ from Equation (6) with the critical value $Z_{0,05} = 1.645$. If $|Z_{j;NBR}| > Z_{0,05}$, the decision made is to reject H_0, indicating that the variable has a significant effect on the response variable in the model. Table 5 shows that X_6 (percentage of poor population) is partially significant for the number of maternal deaths in districts /cities in East Java in 2023. The NBR model with exposure formed is as follows.

$$\hat{\mu} = q\exp(-2.892508 + 0.005842X_1 + 0.007835X_2 - 0.042181X_3 - 0.004239X_4 - 0.011528X_5 + 0.035272X_6)$$

G. Spatial Heterogeneity Testing

GWNBR modeling can only be applied if the data indicates spatial heterogeneity, so a Breusch-Pagan test is conducted with the following hypothesis.

H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_{38}^2 = \sigma^2$
H_1: at least one $\sigma_i^2 \neq \sigma^2$; $i = 1, 2, \ldots, 38$

Based on Equation (7), the BP value is 14.79 and the *p-value* is 0.02195. With a significance level of 10%, the decision made is to reject H_0 because $BP > \chi^2_{(0,1;6)}$ or *p-value* $< \alpha$. This indicates spatial heterogeneity, meaning that there are differences in characteristics between districts /cities in East Java, thus the GWNBR modeling can be applied.

H. GWNBR Modeling of the Number of Maternal Deaths

The GWNBR model is an extension of the non-spatial NBR model that allows for spatial variation in parameters. This model is designed to handle overdispersion and spatial heterogeneity in data by generating different local parameters for each district/city. Weighting in this model is based on geographic coordinates, including latitude (u) and longitude (v). The following are several stages of analysis in GWNBR modeling.

1) Model Similarity Testing

The hypothesis in the model similarity test is as follows.

H_0: $\beta_j(u_i, v_i) = \beta_j$; $i = 1, 2, \ldots, 38$ dan $j = 0, 1, 2, \ldots, 6$
H_1: at least one $\beta_j(u_i, v_i) \neq \beta_j$; $j = 0, 1, 2, \ldots, 6$

The model similarity test was conducted by comparing the test statistic F_{hit} from Equation (12) with the critical value $F_{(0,1;6;228)} = 1.80012$. The F_{hit} value for each GWNBR model, both with and without exposure, for each type of weighting function are shown in Table 6.

Table 6 Model comparison test statistic for NBR and GWNBR

Model		Weights	F_{hit}
GWNBR	Without Exposure	*Adaptive Bisquare Kernel*	3.6619
		Adaptive Gaussian Kernel	3.1618
		Adaptive Tricube Kernel	3.9855
	With Exposure	*Adaptive Bisquare Kernel*	7.9099
		Adaptive Gaussian Kernel	4.0855
		Adaptive Tricube Kernel	9.6533

Table 6 shows that the $F_{hit} > F_{(0,1;6;228)}$ for all models, leading to the decision to reject H_0 for all models. Therefore, it can be concluded that there are significant differences between the regression coefficients of the GWNBR model without exposure, using Adaptive Bisquare, Gaussian, or Tricube Kernel weighting functions, and the NBR model without exposure. The same conclusion applies to the GWNBR model with exposure, where there are significant differences in the regression coefficients compared to the NBR model with exposure.

2) Selection of the Best Model

The selection of the best model for the number of maternal deaths is based on the AICc value. Table 7 presents the AICc calculation results for each GWNBR model.

Table 7 AICc value of GWNBR model

Model		Weights	AICc
GWNBR	Without Exposure	*Adaptive Bisquare Kernel*	291.0891
		Adaptive Gaussian Kernel	289.5868
		Adaptive Tricube Kernel	292.9274
	With Exposure	*Adaptive Bisquare Kernel*	216.2796
		Adaptive Gaussian Kernel	216.2370
		Adaptive Tricube Kernel	217.8558

The model with the lowest AICc value was considered the best because it demonstrated the optimal balance between model fit and complexity. The smallest AICc values for the GWNBR model, both without and with exposure, were obtained using the Adaptive Gaussian Kernel weighting, at 289.5868 and 216.2370, respectively. Thus, the GWNBR model with exposure and the Adaptive Gaussian Kernel are the best models for analyzing the number of maternal deaths in East Java in 2023.

3) GWNBR Modeling with Exposure

The GWNBR model with exposure involves the exposure variable (q), which is the number of pregnant women in each district/city. Parameter significance tests were performed simultaneously and partially. The following is the hypothesis for the simultaneous test.

H_0: $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_6(u_i, v_i) = 0$

H_1: at least one $\beta_j(u_i, v_i) \neq \beta_j$; $i = 1,2, \ldots ,38$

Based on Equation (13), the value of $D_{GWNBR;Gauss}$ is 394.3331 and $\chi^2_{(0,1;228)}$ is 255.759. With a significance level of 10%, the decision made is to reject H_0 because $D_{GWNBR;Gauss} > \chi^2_{(0,1;228)}$. This indicates that at least one predictor variable has a significant influence on the response variable in the model.

After rejecting H_0 in the simultaneous test, a partial test is continued to determine which predictor variables are significant to the response variable with the following hypothesis.

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

H_1: $\beta_j(u_i, v_i) \neq 0$; $i = 1,2, \ldots ,38$ dan $j = 1,2, \ldots ,6$

The partial test is conducted by comparing the test statistic $|Z_{j;GWNBR}|$ from Equation (14) with $Z_{0,05} = 1{,}645$. If $|Z_{j;GWNBR}| > Z_{0,05}$,, then reject H_0 which means the variable has a significant effect on the response variable in the model. Significant variables in each region are grouped into four groups for the Adaptive Gaussian Kernel weighting, as shown in Table 8.

Table 8 Significant variables in the GWNBR model with exposure

Group	Variable	District/City	Total
1	X_1, X_3	Tulungagung, Blitar, Kediri, Mojokerto, Jombang, Nganjuk, Kota Kediri, Kota Blitar, Kota Malang, Kota Mojokerto, Kota Batu	11
2	X_1, X_3, X_4	Sidoarjo, Kota Surabaya	2
3	X_3	Pacitan, Ponorogo, Trenggalek, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Pamekasan, Sumenep, Kota Probolinggo, Kota Madiun	20
4	X_3, X_4	Pasuruan, Gresik, Bangkalan, Sampang, Kota Pasuruan	5
Total			38

Figure 5 presents a thematic map of the grouping districts/cities based on significant variables from Table 8 in the GWNBR model with exposure and Adaptive Gaussian Kernel weighting. Figure 5 shows that Surabaya and Sidoarjo are included in the second group in the spatial analysis of the number of maternal deaths. This may be influenced by similarities in demographic, social, or economic aspects, considering that both are part of the Gerbangkertosusila metropolitan area with populations of 2,936.8 thousand in Surabaya and 2,389.1 thousand of people in Sidoarjo, respectively, in 2023 [1].

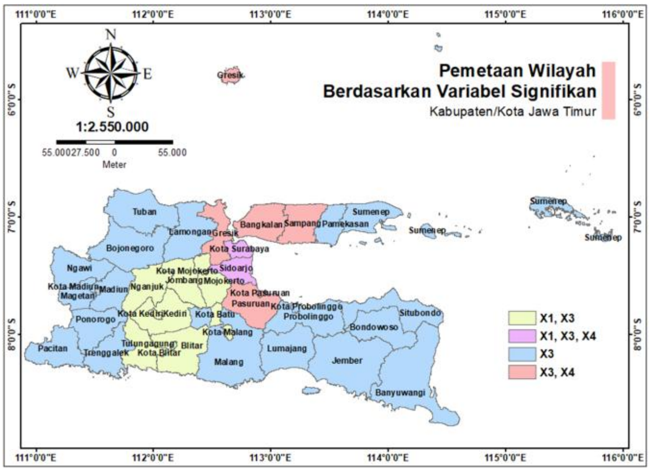

Figure 5 Regional mapping based on the GWNBR model with exposure

Surabaya, as the capital of East Java, is a major destination for migration, both from within and outside the province. This urbanization phenomenon is influenced by economic, educational, and employment factors, making Surabaya a rapidly growing economic center. As a big city with a more diverse economic sector, Surabaya provides more job opportunities with competitive wages. Surabaya's population density reaches 8,986 per km^2, making it the most densely populated area in East Java, which means that residential land is increasingly limited.

This condition has encouraged some people who work in Surabaya to choose to live in neighboring areas, such as Sidoarjo, which offers more affordable housing prices and living costs. Although Gresik is also close to Surabaya, Sidoarjo provides more accessible residential areas for Surabaya workers. Its superior transportation infrastructure and geographical proximity make Sidoarjo a attractive option for those seeking affordable housing with easy access to Surabaya's economic center. Population mobility between Sidoarjo and Surabaya is very high, allowing Sidoarjo residents to work and commute in Surabaya. Conversely, Gresik tends to develop as an independent economic region with more industrial areas and large factories.

In terms of economy, the minimum wages in Surabaya and Sidoarjo in 2023 are Rp 4,525,479.19 and Rp 4,518,581.85, respectively. Meanwhile, Gresik has a minimum wage of Rp 4,522,030.51, which is almost the same as Surabaya [15]. However, the percentage of poor population in Gresik (10.96%) is significantly higher than in Sidoarjo (5.00%) and Surabaya (4.65%). This indicates greater economic inequality in Gresik, resulting in more diverse levels of social welfare across social strata. Gresik relies more heavily on the industrial sector, as shown by the Gresik Integrated Industrial Estate (JIIPE), which indicates that income is more concentrated in the industrial sector, with wages tending to be lower than the more developed trade and service sectors in Surabaya and Sidoarjo.

Although Gresik has a higher ratio of community health centers and hospitals per 100,000 population (3.72) than Sidoarjo (2.30), unequal economic conditions can hinder the optimal utilization of maternal health facilities, such as for the treatment of obstetric complications. Therefore, Surabaya and Sidoarjo share more similar characteristics in terms of economic well-being and population mobility, making the two regions more integrated in maternal health services.

Table 9 shows the estimated parameters of the GWNBR model using exposure and Adaptive Gaussian Kernel weighting for the 37th observation location, namely the City of Surabaya.

Table 9 GWNBR model parameter estimation with exposure in Surabaya

Parameter	Estimate	SE $\left(\hat{\beta}_j(u_i, v_i)\right)$	Z_j	*P-value*
$\beta_0(u_{37}, v_{37})$	0.3046232	4.2915888	0.0710	0.94340
$\beta_1(u_{37}, v_{37})$	-0.0326400	0.0187995	-1.7819	0,.7476*
$\beta_2(u_{37}, v_{37})$	-0.0028286	0.0553729	-0.0511	0.95926
$\beta_3(u_{37}, v_{37})$	-0.6940744	0.1518879	-5.3975	6.76×10^{-8}*
$\beta_4(u_{37}, v_{37})$	0.0503431	0.0248878	2.0951	0.03616*
$\beta_5(u_{37}, v_{37})$	0.0540427	0.0361839	1.5187	0.12883
$\beta_6(u_{37}, v_{37})$	-0.0000866	0.0684038	-0.0013	0.99899

*) significant with α=10%

With a significance level of 10%, the decision based on Table 9 is to reject H_0 on X_1, X_3, X_4 because $|Z_{j;GWNBR}| > Z_{(0,05)}$. It can be concluded that X_1 (percentage of obstetric complication treatment), X_3 (ratio of community health centers and hospitals per 100,000 population), and X_4 (percentage of postpartum family planning participants) have a partial significant effect on the number of maternal deaths in Surabaya City in 2023.
The GWNBR model with exposure variables using Adaptive Gaussian Kernel weighting for Surabaya City is as follows.

$$\hat{\mu}_{37} = q_{37}\exp(0.3046232 - 0.0326400X_{1;37} - 0.0028286X_{2;37} - 0.6940744X_{3;37} + 0.0503431X_{4;37} + 0.0540427X_{5;37} - 0.0000866X_{6;37})$$

Based on the results of the model analysis and significant variables in Surabaya, it can be concluded that each one-unit increase in the percentage of obstetric complication treatment (X_1) will reduce the number of maternal deaths by exp(-0.0326400) ≈ 0.96789 times the previous value, assuming other variables remain constant. The sign of the X_1 coefficient in the model does not match the sign of the correlation between X_1 and Y, which is positive. However, this is in line with the actual situation, as a higher percentage of obstetric complication treatment leads to a lower risk of maternal deaths.

Similarly, each one-unit increase in the ratio of community health centers and hospitals per 100,000 population (X_3) will reduce the number of maternal deaths by exp(-0.6940744) ≈ 0.49954 times the previous value, assuming other variables remain constant. The sign of the X_3 coefficient in the model is the same as the sign in the correlation between X_3 and Y, which is negative. This is also in line with the actual situation because the higher the ratio of health centers and hospitals per 100,000 population, the lower the risk of maternal deaths.

Conversely, each one-unit increase in the percentage of postpartum family planning participants (X_4) will increase the number of maternal deaths by exp(0.0503431) ≈ 1.05163 times the previous value, assuming other variables remain constant. The sign of the X_4 coefficient in the model is the same as the sign in the correlation between X_4 and Y, which is positive. However, this contradicts the actual situation, as a higher percentage of postpartum family planning participants should theoretically lead to a decrease in maternal deaths.

The positive relationship between the percentage of postpartum family planning participants (X_4) and the number of maternal deaths contradicts the theoretical expectations, which should show a negative relationship. This implies that a higher percentage of postpartum family planning participants should decrease maternal deaths. This can be explained by several factors. First, the sign of the X_4 coefficient in the model shows a positive relationship, which matches its correlation with the response variable (Y). Second, the use of annual data is insufficient to capture long-term dynamics because the impact of postpartum family planning programs typically becomes evident several years after implementation. In other words, the effects of effective postpartum family planning program are not fully reflected in the annual data, especially if the program started mid-year. As a result, the expected negative relationship between X_4 and Y is not immediately apparent when analyzing annual data.

Third, Figure 4 shows a positive dependency between variable X_4 and X_1 (percentage of obstetric complication treatment). In the GWNBR model with exposure, the X_1 coefficient is negative, which aligns with the theory suggesting a protective effect on maternal mortality after controlling for other variables. The positive coefficient for X_4 in the model indicates that postpartum family planning programs are more commonly applied to high-risk mothers, including those who have experienced obstetric complications. Mothers who experience obstetric complications tend to require more attention and care after childbirth. One form of care provided is the postpartum family planning program, which aims to regulate birth spacing and maintain maternal health. Although implemented among the high-risk mothers, this does not directly mean that postpartum family planning increases maternal mortality. On the contrary, high-risk groups are more likely to receive family planning programs, which could potentially reduce maternal mortality in the long term, with the program's full effectiveness becoming apparent in subsequent years.

V. CONCLUSIONS AND SUGGESTIONS

A. Conclusions

The conclusions that can be drawn from the results of the analysis and discussion of the research are as follows.

1) In 2023, there were 499 cases of maternal deaths in East Java Province, with Jember District having the highest number of cases at 50. In contrast, Kediri City, Madiun City, and Mojokerto City reported no cases of maternal deaths. The variance in the number of maternal deaths was 104.67, indicating data heterogeneity.
2) The GWNBR model with exposure, using Adaptive Gaussian Kernel weighting, was found to be the best model for modeling maternal deaths in East Java Province in 2023. This model identified four groups of districts/cities based on similarities in the predictor variables that had a significant effect on the number of maternal deaths. The significant variable across all regions is the ratio of community health centers and hospitals per 100,000 population (X_3), while the percentage of obstetric complication treatment (X_1) and the percentage of postpartum family planning participants (X_4) are only significant in some regions.

B. Suggestions

Based on the results of GWNBR modeling of the number of maternal deaths in East Java in 2023, it is recommended that significant variables in each district/city be used as a basis for consideration by the government, particularly the East Java Provincial Health Office, in formulating more focused and effective health policies to reduce the number of maternal deaths. For future research, the use of predictor variable data from the previous year can be considered to accommodate temporal effects, given the time lag of program effects, as shown in the percentage of postpartum family planning participants. In addition, it is recommended to include additional variables that have the potential to affect the number of maternal deaths in East Java so that the resulting model better reflects the actual conditions. The addition of these variables also allows for a more comprehensive analysis using Mixed Geographically Weighted Negative Binomial Regression (MGWNBR), which can accommodate spatial variations in the influence of predictor variables as well as significant variables across all locations.

REFERENCES

[1] Kementerian Kesehatan RI, *Profil Kesehatan Indonesia 2023*. Jakarta: Kementerian Kesehatan RI, 2024.
[2] Badan Pusat Statistik, *Mortalitas di Indonesia: Hasil Long Form Sensus Penduduk 2020*. Jakarta: Badan Pusat Statistik, 2023.
[3] Q. Aini, *Penaksiran Parameter dan Pengujian Hipotesis Geographically Weighted Bivariate Zero Inflated Generalized Poisson Regression*. Surabaya: Departemen Statistika FSAD ITS, 2020.
[4] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability & Statistics for Engineers & Scientists*, 9th ed. Boston: Pearson Education, 2012.
[5] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 2nd ed. London: Chapman and Hall, 1989.
[6] F. Famoye, J. T. Wulu, and K. P. Singh, “On the Generalized Poisson Regression Model with an Application to Accident Data,” *Journal of Data Science*, vol. 2, no. 3, pp. 287-295, 2004.
[7] W. Greene, “Functional Form and Heterogeneity in Models for Count Data,” *Foundations and Trends in Econometrics*, vol. 1, no. 2, pp. 113-218, 2007.
[8] J. M. Hilbe, *Negative Binomial Regression*, 2nd ed. Cambridge, United Kingdom: Cambridge University Press, 2011.
[9] T. S. Breusch and A. R. Pagan, “A Simple Test for Heteroscedasticity and Random Coefficient Variation,” *Journal of The Econometric Society*, vol. 47, no. 5, pp. 1287-1294, 1979.
[10] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley, 2002.
[11] R. E. Caraka and H. Yasin, *Geographically Weighted Regression (GWR): Sebuah Pendekatan Regresi Geografis*. Yogyakarta, Indonesia: Mobius, 2017.
[12] A. R. Silva and T. C. Rodrigues, “Geographically Weighted Negative Binomial Regression-Incorporating Overdispersion,” *Statistics and Computing*, vol. 24, no. 5, pp. 769–783, 2014.
[13] World Heatlh Organization, *The Global Health Observatory: Maternal Deaths*, 2024. [Online]. Available: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/4622. [Accessed: Dec. 2024].
[14] J. McCarthy and D. Maine, “A Framework for Analyzing the Determinants of Maternal Mortality,” *Family Planning*, vol. 23, no. 1, pp. 23-22, 1992.
[15] Badan Pusat Statistik, *Upah Minimum Kabupaten/Kota di Jawa Timur (Rupiah) Tahun 2023*, 2024. [Online]. Available: https://probolinggokab.bps.go.id/id/statistics-table/2/Mjg4IzI=/upah-minimum-kabupaten-kota-di-jawa-timur. [Accessed: May 2025].