Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2025, 2025:98

https://doi.org/10.28919/cmbn/9312

ISSN: 2052-2541

ANALYSIS OF MATERNAL MORTALITY RATE USING MULTIVARIATE ADAPTIVE GENERALIZED POISSON REGRESSION SPLINE

NANANG SETIA, ANNA ISLAMIYATI*, ERNA TRI HERDIANI

Department of Statistics, Hasanuddin University, Makassar, 90245, Indonesia

Abstract: Maternal mortality is an essential indicator for assessing public health levels and the quality of healthcare services in a country. In South Sulawesi Province, maternal mortality reached 198 cases in 2021, increasing from 133 cases in 2020. This high mortality rate requires urgent attention to identifying contributing factors. Maternal mortality data is discrete categorical data that follows a Poisson distribution and experiences overdispersion. This nonlinear and random data pattern makes it more suitable for nonparametric analysis. The appropriate method for these data characteristics is Multivariate Adaptive Generalized Poisson Regression Splines (MAGPRS). This study aims to identify the factors influencing maternal mortality in South Sulawesi Province using the MAGPRS method. The best model was obtained with a combination of basis function (BF) of 20, Maximum Interaction (MI) of 3, and Minimum Observation (MO) of 2, yielding a Generalized Cross-Validation (GCV) value of 0.000046683 and an R-Square of 99.30%, indicating high model accuracy. The findings reveal that the most influential factors include the percentage of pregnant women in the K4 program (X_1), those receiving Td3 immunization (X_5), postpartum mothers receiving vitamin A (X_4), deliveries in healthcare facilities (X_3), and pregnant women in the K1 program (X_2).

Keywords: MAGPRS; MARS; maternal mortality; overdispersion; Poisson.

2020 AMS Subject Classification: 62J12, 92C50.

*Corresponding author

E-mail address: annaislamiyati701@gmail.com

Received April 21, 2025

1. INTRODUCTION

Regression analysis is a statistical technique used to examine the relationship between predictor and response variables. The response variable can be either continuous or discrete. Poisson regression is one of the most employed methods for modeling discrete response variables, which evaluates the association between predictor and response variables under the assumption that the response variable follows a Poisson distribution [1]. However, Poisson regression relies on the assumption of equidispersion, meaning that the variance and mean of the response variable must be equal. In real-world applications, this assumption is frequently violated, particularly when the variance surpasses the mean (overdispersion) or, conversely, when it falls below the mean (underdispersion) [2]. To address these limitations, Generalized Poisson Regression (GPR) has been developed as an alternative method that can effectively account for both overdispersion and underdispersion, making it a more adaptable approach than traditional Poisson regression [3]. Previous research has demonstrated the efficacy of GPR in modeling count data with overdispersion [4-6].

Although GPR offers greater flexibility, it remains a parametric method. It may be unsuitable when the relationship between the predictor and response variables is complex and does not follow a predefined pattern. Nonparametric regression techniques, particularly splines, provide a viable alternative. Splines have been widely applied in various statistical analyses, including confidence interval estimation [7], multi-response longitudinal splines [8], penalized splines [9], bivariate splines [10], and principal component splines [11]. Additionally, some estimators leverage basis functions to determine optimal knot points in multivariate regression, notably the Multivariate Adaptive Regression Splines (MARS) method [12]. MARS has several advantages over other regression methods. MARS can identify and model complex nonlinear relationships between the response and predictor variables without requiring any assumptions [13]. This makes it highly effective in capturing intricate relationships that traditional parametric models cannot explain. Moreover, MARS is particularly useful for high-dimensional modeling, especially when the number of predictor variables ranges between 3 and 20 [12-13]. This method reduces the risk of overfitting by using Generalized Cross-Validation (GCV) as a model selection criterion [14].

A practical example of count data following a Poisson distribution in the healthcare sector is the Maternal Mortality Rate (MMR). Maternal mortality is a crucial public health indicator and remains a significant concern in Indonesia [1]. Despite advancements in healthcare services, MMR remains high in Indonesia [1], [15]. In South Sulawesi, maternal mortality increased from 133

cases in 2020 to 198 cases in 2021 [16], highlighting the urgent need to identify its key contributing factors. Given the complexity of maternal mortality, this study integrates Generalized Poisson Regression (GPR) within the MARS framework, leading to the development of the Multivariate Adaptive Generalized Poisson Regression Splines (MAGPRS) method. This approach is selected due to the uncertainty in the relationship between maternal mortality and its predictors, which exhibit nonlinear patterns. Additionally, since the response variable is count-based and follows a Poisson distribution, and exhibits overdispersion. Although previous studies have explored MAGPRS [17-18], limited research has specifically focused on identifying the key determinants of maternal mortality. Therefore, this study aims to apply MAGPRS in modeling nonlinear data patterns characterized by overdispersion, particularly in maternal mortality in South Sulawesi. The optimal model is selected based on the GCV criterion, prioritizing the model with the lowest GCV value to ensure high predictive accuracy. By enhancing the understanding of MMR risk factors, this research contributes to developing more effective public health policies to reduce maternal mortality.

2. PRELIMINARIES

Poisson Distribution Test

To assess whether the observed data follows a Poisson distribution, the Kolmogorov-Smirnov test can be utilized. The hypotheses for this test are structured as follows:

H_0: The data Y follows a Poisson distribution

H_1: The data Y does not follow a Poisson distribution

The Kolmogorov-Smirnov method is evaluated using the D-statistic, which is defined as follows:

$$D = max|F_0(y) - S_n(y)| \quad (1)$$

Where

D : The maximum absolute deviation between $F_0(y)$ dan $S_n(y)$,

$F_0(y)$: The cumulative relative frequency function of the theoretical distribution under H_0,

$S_n(y)$: The cumulative frequency distribution of the observed sample.

With the rejection criteria, H_0 is rejected if $D > D_{tabel}$ (the critical value from the Kolmogorov-Smirnov table) or if $p < \alpha$ [19].

Overdispersion Test

In Poisson regression modeling, one of the key assumptions that must be satisfied is equidispersion, meaning that the mean of the response variable (Y) should be equal to its variance. However, in

real-world applications, this assumption is often violated, leading to a phenomenon known as the overdispersion condition, where the variance of the response variable exceeds its mean [2]. Overdispersion in Poisson regression can significantly impact model interpretation, particularly in parameter estimation. It can result in underestimated standard errors, potentially leading to incorrect conclusions about the statistical significance of regression coefficients.

The hypotheses for testing overdispersion are formulated as follows [20]:

H_0: No overdispersion is present

H_1: Overdispersion is present

Test Statistic:

$$\phi = \frac{D^2}{df}\ ; \qquad D^2 = 2\ \sum_{i=1}^{n}\left\{y_i\ ln\left(\frac{y_i}{\mu_i}\right)\right\} \tag{2}$$

with ϕ is the dispersion parameter, D^2 is the deviance value, and $df = n - p$, where n is the number of observations, and p is the number of parameters, including the intercept.

Decision Criteria:

Reject H_0 if $\phi > 1$, indicating the presence of overdispersion in the data.

Generalized Poisson Regression (GPR)

The Generalized Poisson Regression (GPR) model has a structure similar to the Poisson regression model. However, in this model, the random component is assumed to follow a Generalized Poisson (GP) distribution. In other words, this model can be applied to discrete data that follow a Poisson distribution without requiring the equidispersion assumption. Therefore, in addition to the parameter μ, the GPR model also includes θ as a dispersion parameter [21]. Let $y = 0, 1, 2, \ldots$ be the response variable. The probability mass function (PMF) of the Generalized Poisson (GP) distribution is defined as follows [22]:

$$f(y, \mu; \theta) = \left(\frac{\mu}{1+\theta_\mu}\right)^y \frac{(1+\theta_y)}{y!}\, exp\left\{\frac{-\mu(1+\theta_y)}{1+\theta_\mu}\right\} \tag{3}$$

The mean and variance of the GPR model are given by:

$E(y) = \mu$ and $var(y) = \mu\left(1 + \theta_\mu\right)^2$

The GPR model is formulated as:

$$ln(\mu_i) = \boldsymbol{x_i^T \beta}$$

$$\mu_i = exp\left(\boldsymbol{x_i^T \beta}\right) \tag{4}$$

Multivariate Adaptive Regression Spline (MARS)

Multivariate Adaptive Regression Splines (MARS) is a nonparametric regression method first introduced by Friedman (1991). MARS is a more complex multivariate nonparametric approach than spline regression, as it utilizes recursive partitioning to construct continuous regression function estimates. This method is well-suited for high-dimensional data, mainly when the number of predictor variables ranges between 3 and 20 and effectively handles discontinuities within the data.

The MARS model is represented by the following equation [23]:

$$f(\tilde{\boldsymbol{x}}) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{Km} [S_{km}\left(x_{v(k,m)} - t_{km}\right)]_+ = a_0 + \sum_{m=1}^{M} a_m B_m(\tilde{\boldsymbol{x}}) \quad (5)$$

Where

a_0 : constant term of the regression function,

a_m : coefficient of the of the m-th basis function, where $m = 1, 2, \dots, M$,

M : maximum number of basis functions,

k_m : interaction degree of the m-th basis function,

S_{km} : a value of +1 or -1 indicating the relative position of the data point concerning the knot. $S_{km} = +1$, if the data point is to the right of the knot ($x_{v(k,m)} \geq t_{km}$), and $S_{km} = -1$ if the data point is to the left of the knot ($x_{v(k,m)} < t_{km}$),

$x_{v(k,m)}$: predictor variable,

t_{km} : knot position for the predictor variable $x_{v(k,m)}$.

There are several essential aspects to consider when using the MARS model [12], [23]:

a. Knots. Specific points within the predictor variable domain divide the data into multiple subregions.

b. Basis Functions (BF). Basis functions describe the relationship between predictor and response variables. Typically, polynomial basis functions with continuous derivatives at each knot are used. Friedman recommends setting the maximum number of basis functions between 2 to 4 times the number of predictor variables.

c. Maximum Interaction (MI). This parameter defines the highest level of interaction among predictor variables. The commonly used values for MI are 1, 2, or 3. If MI > 3, the resulting model may become overly complex and difficult to interpret.

d. Minimum Observation (MO). MO refers to the minimum number of observations per region or distance between knots. Typical values for MO are 0, 1, 2, or 3.

The optimal model MARS is determined based on the smallest Generalized Cross-Validation (GCV) value and the highest R-Square value. The GCV formula is defined as [24]:

$$GCV = \frac{MSE}{\left[1-\frac{C(\widetilde{M})}{n}\right]^2} = \frac{\frac{1}{n}\left[(\tilde{\boldsymbol{y}}-\hat{f}(\tilde{\boldsymbol{x}}))^T(\tilde{\boldsymbol{y}}-\hat{f}(\tilde{\boldsymbol{x}}))\right]}{\left[1-\frac{C(\widetilde{M})}{n}\right]^2} \quad (6)$$

where $C(\widetilde{M}) = \text{trace}\,(\mathbf{B}(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}) + 1$.

Here, $\mathbf{B}$ is the matrix of basis functions and $\tilde{\boldsymbol{y}}$ represents the observed values.

Multivariate Adaptive Generalized Poisson Regression Spline (MAGPRS)

The Multivariate Adaptive Generalized Poisson Regression Splines (MAGPRS) model extends the MARS framework by incorporating GPR. This integration allows for greater flexibility in handling count data that exhibits overdispersion or underdispersion. The general form of the MAGPRS model is as follows [17]:

$Y_i \sim GP(\mu, \theta)$

$$\boldsymbol{f}(\tilde{\boldsymbol{x}}) = \mathbf{ln}\,(\boldsymbol{\mu}) = \boldsymbol{\alpha_0} + \sum_{\mathbf{m=1}}^{\mathbf{M}} \boldsymbol{\alpha_m} \prod_{\mathbf{k=1}}^{\mathbf{K_m}} [\mathbf{s_{km}}(\mathbf{x_{v(k,m)i}} - \mathbf{t_{km}})]_+$$

$$\mu = \exp\left(\alpha_0 + \sum_{m=1}^{M} \alpha_m B_{mi}(\tilde{\mathbf{x}})\right) = \exp(\mathbf{B}\tilde{\boldsymbol{\alpha}}) \quad (7)$$

MAGPRS parameters are estimated using the Weighted Least Squares (WLS) approach. The error term is defined as [17]:

$$\varepsilon = Y - E(Y) = Y - \mu = Y - \exp(\mathbf{B}\tilde{\boldsymbol{\alpha}}) \quad (8)$$

The variance weight of Y is expressed in the form of a diagonal matrix $\mathbf{W}$ of size $n \times n$, where its diagonal elements:

$$w_{ii} = \frac{1}{\text{var}(Y_i)} \quad (9)$$

Based on the Generalized Poisson distribution assumption, the variance of Y_i is given by:

$$var(Y_i) = \mu_i(1 + \theta\mu_i)^2 \quad (10)$$

Thus, the diagonal elements of $\mathbf{W}$ are formulated as:

$$w_{ii} = \frac{1}{\mu_i(1+\theta\mu_i)^2} \quad (11)$$

By minimizing the weighted sum of squared errors (Weighted Sum of Squares Error), the equation is obtained as:

$$\boldsymbol{S}(\tilde{\boldsymbol{\alpha}}) = (\tilde{\boldsymbol{y}} - \mathbf{B}\tilde{\boldsymbol{\alpha}})^{\boldsymbol{T}}\mathbf{W}(\tilde{\boldsymbol{y}} - \mathbf{B}\tilde{\boldsymbol{\alpha}}) \quad (12)$$

By setting the first derivative concerning α to zero, the estimated MAGPRS parameters are obtained as:

$$\widehat{\widetilde{\boldsymbol{\alpha}}} = (\mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{W}\tilde{\boldsymbol{y}} \tag{13}$$

The estimated MAGPRS model is expressed as [17]:

$$\hat{f}(\tilde{\boldsymbol{x}}) = ln(\hat{\mu}) = \mathbf{B}\widehat{\widetilde{\boldsymbol{\alpha}}} = \mathbf{B}(\mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{B})^{-1}\mathbf{B}^{T}\mathbf{W}\tilde{\boldsymbol{y}} \tag{14}$$

The significance of predictor variables in the MAGPRS model is assessed through the Maximum Likelihood Ratio Test (MLRT). This test determines whether predictor variables collectively influence the response variable through the model's basis functions.

The hypotheses for this test are formulated as follows [17]:

$\mathbf{H_0}: \boldsymbol{\alpha_1} = \ldots = \boldsymbol{\alpha_M} = \mathbf{0}$ (No significant effect of predictor variables)

$\mathbf{H_1}$: at least one $\boldsymbol{\alpha_m} \neq \mathbf{0}$, where $\boldsymbol{m} = \mathbf{1}, \ldots, \boldsymbol{M}$ (at least one predictor variable significantly influences the response)

The test statistic is given by:

$$\hat{G}^2 = -2\,ln\,\Lambda = 2(ln\,L\,(\widehat{\boldsymbol{\Omega}}) - ln\,L\,(\widehat{\boldsymbol{\omega}}))$$

$$\hat{G}^2 = 2\begin{bmatrix} \sum_{i=1}^{n} y_i(\mathbf{B}\widehat{\boldsymbol{\alpha}}) - \sum_{i=1}^{n} y_i\,ln(\hat{\alpha}_0) + \sum_{i=1}^{n} y_i \frac{ln(1+\hat{\theta}\,exp(\hat{\alpha}_0))}{ln(1+\hat{\theta}\,exp(\mathbf{B}\widehat{\boldsymbol{\alpha}}))} \\ -\sum_{i=1}^{n} exp(\mathbf{B}\widehat{\boldsymbol{\alpha}})(1+\hat{\theta}y_i)(1+\hat{\theta}\,exp(\mathbf{B}\widehat{\boldsymbol{\alpha}}))^{-1} \\ +\sum_{i=1}^{n} exp(\hat{\alpha}_0)(1+\hat{\theta}y_i)(1+\hat{\theta}\,exp(\hat{\alpha}_0))^{-1} \end{bmatrix} \tag{15}$$

The test statistic $\widehat{\boldsymbol{G}}^{\mathbf{2}}$ follows a chi-square ($\boldsymbol{X}^{\mathbf{2}}$) distribution: $\widehat{\boldsymbol{G}}^{\mathbf{2}} \sim \boldsymbol{X}^{\mathbf{2}}{}_{\boldsymbol{\alpha},\boldsymbol{v}}$ where $\boldsymbol{v} = \boldsymbol{n}(\widehat{\boldsymbol{\Omega}}) - \boldsymbol{n}(\widehat{\boldsymbol{\omega}})$. The null hypothesis ($\mathbf{H_0}$) is rejected if $\widehat{\boldsymbol{G}}^{\mathbf{2}} > \boldsymbol{X}^{\mathbf{2}}{}_{\boldsymbol{\alpha},\boldsymbol{v}}$, indicating that the predictor variables significantly affect the response variable.

In addition to simultaneous testing, a t-test determines which specific predictor variables significantly impact the response variable through the model's basis functions. The hypotheses for this partial test are as follows [17]:

$\mathbf{H_0}: \boldsymbol{\alpha_m} = \mathbf{0}$ (The predictor variable has no significant effect).

$\mathbf{H_1}: \boldsymbol{\alpha_m} \neq \mathbf{0}$, where $\boldsymbol{m} = \mathbf{1}, \ldots, \boldsymbol{M}$ (The predictor variable has a significant effect).

The test statistic is given by:

$$\boldsymbol{z} = \frac{\hat{\boldsymbol{\alpha}}_{\boldsymbol{m}}}{\boldsymbol{se}(\hat{\boldsymbol{\alpha}}_{\boldsymbol{m}})} \quad \text{where} \quad \boldsymbol{se}(\hat{\boldsymbol{\alpha}}_{\boldsymbol{m}}) = \sqrt{\left(\frac{\sum_{i=n}^{n}(y_i - \hat{y}_i)^2}{\boldsymbol{n-k-1}}\right)\boldsymbol{C_{mm}}} \tag{16}$$

$\boldsymbol{C_{mm}}$ is the diagonal element of the covariance matrix $(\mathbf{B}^{\mathbf{T}}\mathbf{W}\mathbf{B})^{-\mathbf{1}}$, and $\boldsymbol{k}$ is the number of estimated parameters, excluding the intercept.

The null hypothesis ($\mathbf{H_0}$) is rejected if $|z| > z_{\alpha/2}$ or $p - value < \alpha$. Rejection of the null hypothesis indicates that the corresponding all selected basis functions in the MAGPRS model significantly contribute to explaining variations in the response variable.

Research Method

The data utilized in this study are secondary data obtained from the South Sulawesi Provincial Health Office. The study focuses on all 24 regencies and cities within South Sulawesi Province as the research population. The analysis employs maternal mortality data from the year 2021 as the response variable (Y), while the predictor variables (X) include: the percentage of pregnant women participating in the K4 program (X_1), the percentage of pregnant women participating in the K1 program (X_2), the percentage of deliveries occurring in healthcare facilities (X_3), the percentage of postpartum mothers receiving vitamin A supplementation (X_4), and the percentage of pregnant women receiving the Td3 immunization (X_5), as presented in Table 1.

Table 1. Research Variables

Variable	Definition	Scale
Number of maternal deaths (Y)	The total number of maternal deaths occurring in each district/city in South Sulawesi Province.	Ratio
Percentage of pregnant women participating in the K4 program (X_1)	The number of pregnant women who received standardized K4 antenatal care in a given area during a specific period, divided by the total number of pregnant women in that area and period, multiplied by 100 percent.	Ratio
Percentage of pregnant women participating in the K1 program (X_2)	The number of pregnant women who received standardized K1 antenatal care in a given area during a specific period, divided by the total number of pregnant women in that area and period, multiplied by 100 percent.	Ratio
Percentage of deliveries in healthcare facilities (X_3)	The number of deliveries assisted by qualified maternal health professionals (obstetricians, general practitioners, and midwives) in a given area during a specific period, divided by the total number of deliveries in that area and period, multiplied by 100 percent.	Ratio
Percentage of postpartum mothers receiving vitamin A (X_4)	The number of postpartum mothers who received vitamin A supplementation in a given area during a specific period, divided by the total number of postpartum mothers in that area and period, multiplied by 100 percent.	Ratio
Percentage of pregnant women receiving Td3 immunization (X_5)	The number of pregnant women who received the Td3 immunization in a given area during a specific period, divided by the total number of pregnant women who received Td3 in that area and period, multiplied by 100 percent.	Ratio

The data analysis method used in this study is the Multivariate Adaptive Generalized Poisson Regression Spline (MAGPRS). The estimation of MAGPRS parameters is conducted using the Weighted Least Squares (WLS) approach. The analysis process begins with presenting descriptive statistics of maternal mortality and its suspected contributing factors, followed by creating scatterplots between the response variable and each predictor variable to observe data distribution patterns. Next, an overdispersion test is conducted, and the MAGPRS model is applied by determining the optimal combination of the maximum number of basis functions (BF), maximum interaction (MI), and minimum observation (MO). Model estimation is performed to obtain the best model based on the smallest Generalized Cross Validation (GCV) value. Subsequently, model significance testing is carried out simultaneously (F-test) and partially (t-test), followed by interpretation of the best MAGPRS model and analysis of the relative importance of each predictor variable. The final step involves concluding the analysis results.

3. MAIN RESULTS

Descriptive Statistics of the Data

Descriptive statistics for each research variable are presented in Table 2.

Table 2. Descriptive Statistics of Research Variables

Variable	Min	Max	Mean	Variance
Y	1	17	8.25	21.065
X_1	71.1	110.1	87.662	99.510
X_2	85.7	119.7	99.387	73.603
X_3	71.5	115.3	94.892	90.092
X_4	75	115.3	95.229	87.044
X_5	4.5	41.4	14.291	75.833

The descriptive statistics in Table 2 indicate that the average maternal mortality across the 24 districts/cities in South Sulawesi was 8.25 cases, with a variance of 21.065, reflecting significant regional disparities. The average percentage of pregnant women completing the K4 program (X_1) was 87.66%, with notable variation (variance = 99.51). For the k1 program (X_2), the average participation was 99.39%, indicating high engagement, though disparities remained (variance = 73.60). The average percentage of deliveries in healthcare facilities (X_3) was 94.89%, with a variance of 90.09, suggesting uneven access or utilization across regions. Postpartum vitamin a supplementation (X_4) had an average coverage of 95.23% (variance = 87.04), and Td3 immunization coverage among pregnant women (X_5) was the lowest, averaging 14.29%, with

considerable variation (variance = 75.83). These results highlight the need to address regional inequalities in maternal healthcare services.

Assumption Testing and Data Pattern Analysis

Before conducting data analysis, certain assumptions must be satisfied, specifically that the data follows a Poisson distribution and that overdispersion needs to be assessed. To determine whether the response variable Y adheres to a Poisson distribution, the Kolmogorov-Smirnov test is applied. At a significance level of $\alpha = 0.05$, the analysis results yielded a p-value of 0.1173. Since the $p - value > 0.05$, the H_0 is accepted, indicating that the maternal mortality data from South Sulawesi Province in 2021 conforms to a Poisson distribution. Therefore, Poisson regression can be used for further analysis.

Before applying Poisson regression, the dataset must be tested for overdispersion by examining the dispersion parameter ϕ. Overdispersion is indicated when $\phi > 1$, while $\phi < 1$ suggests no overdispersion. This can be assessed using the deviance test, with the results summarized in Table 3.

Table 3. Overdispersion Test

Variable	Deviance/df
Y	$\frac{52.763}{18} = 2.931$

Based on the results shown in Table 3, the dispersion value obtained using R-Studio is 2.931 ($\phi > 1$), which indicates the presence of overdispersion in the maternal mortality data. Therefore, the standard Poisson regression model is unsuitable for analyzing this dataset. As an alternative, the Generalized Poisson Regression (GPR) model is recommended, as it is specifically designed to accommodate overdispersion in count data.

Furthermore, a scatterplot analysis explores the relationship patterns between predictor and response variables. A parametric regression model can be applied if the data points exhibit a structured pattern consistent with parametric distribution assumptions. This indicates that a predefined mathematical function can effectively describe the relationship between variables. Conversely, a nonparametric regression model is considered more appropriate if the data points appear randomly scattered or do not follow a discernible pattern, as it offers greater flexibility in capturing complex and nonlinear relationships. Researchers can determine the most suitable modeling approach by assessing the scatterplot, ensuring accurate data representation. The relationship between the predictor variables and the dependent variable is illustrated in Figure 1.

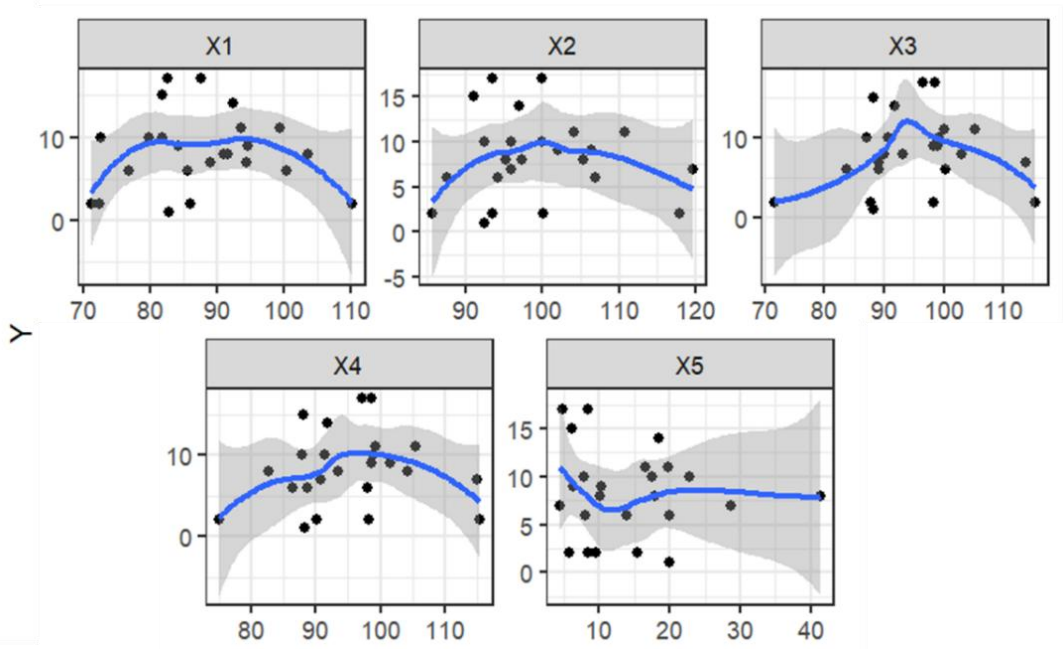

Figure 1. Scatterplot of Maternal Mortality and Predictor Variables

Based on Figure 1, the relationship between Y (Maternal Mortality) and the predictor variables (X_1 to X_5) does not exhibit a clear pattern. Data distribution tends to be random and does not follow a specific mathematical relationship. Therefore, a nonparametric regression model is recommended for analyzing the relationship patterns between the variables in this study.

Maternal Mortality Modeling Using MAGPRS

In MAGPRS modeling, the selection of the optimal model is based on the minimum Generalized Cross-Validation (GCV) value, which is determined by testing various combinations of Basis Functions (BF), Maximum Interaction (MI), and Minimum Observations (MO). In this study, the maximum number of basis functions (BF) was set to 10, 15, and 20, following Friedman's (1991) recommendation that suggests the maximum number of basis functions should range between two to four times the number of predictor variables. Since this study includes five predictor variables, these guidelines were applied accordingly. Additionally, the maximum interaction level (MI) was configured at 1, 2, and 3, while the minimum observation values (MO) were set at 0, 1, 2, and 3, also adhering to Friedman's (1991) principles. The results of the best-performing MAGPRS model are summarized in Table 4.

Table 4. Best MAGPRS Model Results

Basis Functions (BF)	Maximum Interaction (MI)	Minimum Observations (MO)	GCV	R^2
10	1	0	0.003506657	0.475368
10	1	1	0.001914733	0.713537
10	1	2	0.002432436	0.636083
10	1	3	0.001987111	0.702708
10	2	0	0.003184339	0.523591
10	2	1	0.002027753	0.696628
10	2	2	0.001743826	0.739106
10	2	3	0.001422453	0.787187
10	3	0	0.003184339	0.523591
10	3	1	0.002251423	0.663164
10	3	2	0.001743826	0.739106
10	3	3	0.001422453	0.787187
15	1	0	0.002567296	0.615906
15	1	1	0.00108379	0.837854
15	1	2	0.001434976	0.785313
15	1	3	0.001253528	0.812459
15	2	0	0.001850665	0.723122
15	2	1	0.001215474	0.818153
15	2	2	0.001020974	0.847252
15	2	3	0.00121208	0.818661
15	3	0	0.001850665	0.723122
15	3	1	0.001847683	0.723568
15	3	2	0.000488187	0.926962
15	3	3	0.001109191	0.834054
20	1	0	0.002531112	0.62132
20	1	1	0.000670799	0.899642
20	1	2	0.00078352	0.882777
20	1	3	0.000689574	0.896833
20	2	0	0.00122357	0.816941
20	2	1	0.00027402	0.959004
20	2	2	0.000558366	0.916463
20	2	3	0.000538742	0.919399
20	3	0	0.00122357	0.816941
20	3	1	0.001039079	0.844543
20	**3**	**2**	**0.000046683**	**0.993016**
20	3	3	0.000638848	0.904422

Based on Table 4, a total of 36 different models were evaluated in this study, each utilizing various combinations of Basis Functions (BF), Maximum Interaction (MI), and Minimum Observations (MO). The optimal model was identified with the parameters BF = 20, MI = 3, and MO = 2, yielding a Generalized Cross-Validation (GCV) value of 0.000046683 and an R-squared value of 99.30%, indicating a high level of model accuracy.

The MAGPRS model for maternal mortality in South Sulawesi is as follows:

$$\hat{\mu} = \begin{bmatrix} 2.335 - 0.079 * \mathrm{h}(99.3 - \mathrm{X}_1) - 0.130 * \mathrm{h}(\mathrm{X}_1 - 99.3) \\ -0.019 * \mathrm{h}(16.5 - \mathrm{X}_5) + 3.753 * \mathrm{h}(\mathrm{X}_5 - 16.5) \\ +0.027 * \mathrm{h}(99.3 - \mathrm{X}_1) * \mathrm{h}(\mathrm{X}_3 - 89) - 0.007 * \mathrm{h}(99.3 - \mathrm{X}_1) * \mathrm{h}(89 - \mathrm{X}_3) \\ -0.032 * \mathrm{h}(99.3 - \mathrm{X}_1) * \mathrm{h}(\mathrm{X}_4 - 88.8) - 0.019 * \mathrm{h}(99.3 - \mathrm{X}_1) * \mathrm{h}(88.8 - \mathrm{X}_4) \\ +2.726 * \mathrm{h}(82.6 - \mathrm{X}_1) * \mathrm{h}(\mathrm{X}_5 - 16.5) - 0.218 * \mathrm{h}(\mathrm{X}_1 - 82.6) * \mathrm{h}(\mathrm{X}_5 - 16.5) \\ -0.236 * \mathrm{h}(99.3 - \mathrm{X}_1) * \mathrm{h}(\mathrm{X}_5 - 16.5) + 0.010 * \mathrm{h}(99.3 - \mathrm{X}_1) * \mathrm{h}(16.5 - \mathrm{X}_5) \\ -0.022 * \mathrm{h}(82.6 - \mathrm{X}_1) * \mathrm{X}_2 * \mathrm{h}(\mathrm{X}_5 - 16.5) + 0.00039 * \mathrm{h}(99.3 - \mathrm{X}_1) \\ * \mathrm{h}(\mathrm{X}_3 - 89) * \mathrm{h}(\mathrm{X}_4 - 98.1) + 0.0107 * \mathrm{h}(99.3 - \mathrm{X}_1) * \mathrm{h}(\mathrm{X}_3 - 89) \\ * \mathrm{h}(98.1 - \mathrm{X}_4) + 0.00077 * \mathrm{h}(99.3 - \mathrm{X}_1) * \mathrm{h}(\mathrm{X}_4 - 88.8) * \mathrm{h}(16.5 - \mathrm{X}_5) \\ +0.0031 * \mathrm{h}(99.3 - \mathrm{X}_1) * \mathrm{h}(88.8 - \mathrm{X}_4) * \mathrm{h}(16.5 - \mathrm{X}_5) \end{bmatrix}$$

The basis functions (BF) are defined as follows:

$\mathrm{BF1} = \mathrm{h}(99.3 - \mathrm{X}_1)$, $\mathrm{BF2} = \mathrm{h}(\mathrm{X}_1 - 99.3)$, $\mathrm{BF3} = \mathrm{h}(16.5 - \mathrm{X}_5)$, $\mathrm{BF4} = \mathrm{h}(\mathrm{X}_5 - 16.5)$, $\mathrm{BF5} = \mathrm{h}(\mathrm{X}_3 - 89)$, $\mathrm{BF6} = \mathrm{h}(89 - \mathrm{X}_3)$, $\mathrm{BF7} = \mathrm{h}(\mathrm{X}_4 - 88.8)$, $\mathrm{BF8} = \mathrm{h}(88.8 - \mathrm{X}_4)$, $\mathrm{BF9} = \mathrm{h}(82.6 - \mathrm{X}_1)$, $\mathrm{BF10} = \mathrm{h}(\mathrm{X}_1 - 82.6)$, $\mathrm{BF11} = \mathrm{X}_2$, $\mathrm{BF12} = \mathrm{h}(\mathrm{X}_4 - 98.1)$, and $\mathrm{BF13} = \mathrm{h}(98.1 - \mathrm{X}_4)$.

Significance Testing of MAGPRS Parameters

The significance of MAGPRS parameters is assessed through both simultaneous and partial tests. The simultaneous test determines whether the predictor variables, as a group, exert a statistically significant influence on the response variable. The hypotheses for the simultaneous test are as follows:

$\mathrm{H}_0: \alpha_1 = \ldots = \alpha_M = 0$

$\mathbf{H_1}$: At least one $\boldsymbol{\alpha_m \neq 0}$, for $\boldsymbol{m = 1, \ldots, M}$

The test statistic is computed as follows:

$$\hat{G}^2 = 2\left(ln\,L\left(\hat{\Omega}\right) - ln\,L\left(\hat{\omega}\right)\right)$$

$$= 2\big((-118.0145) - (-412.6294)\big)$$

$$= 2(-118.0145 + 412.6294)$$

$$= 2(294.6149)$$

$= 589.2298$

The computed G-statistic is compared with the chi-square critical value. The decision rule states that H_0 is rejected if $\hat{G}^2 > X^2{}_{\alpha,v}$. Based on the analysis, the computed G-value is 589.2298 and the chi-square critical value $X^2_{(0.05;17)}$ is 27.587. Since $589.2298 > 27.58711$, H_0 is rejected. This indicates that at least one predictor variable significantly affects the response variable.

After establishing significance in the simultaneous test, a partial test is conducted to assess the significance of each predictor variable individually. The hypotheses for the partial test:

$\mathbf{H_0}: \boldsymbol{\alpha_m = 0}$

$\mathbf{H_1}: \boldsymbol{\alpha_m \neq 0, m = 1, \ldots, M}$

Decision Criterion: Reject H_0 if $|z| > z_{\alpha/2}$ or $p - value < \alpha$.

The test statistic is calculated as follows:

$$|z| = \frac{\hat{\alpha}_m}{se(\hat{\alpha}_m)} \quad \text{for} \quad |z_{BF1}| = \frac{\hat{\alpha}_{BF1}}{se(\hat{\alpha}_{BF1})} = \frac{-0.1303}{5.30\times10^{-16}} = |-2.45 \times 10^{14}| = 2.45 \times 10^{14}$$

The complete results of the partial significance test are summarized in Table 5.

Table 5. Results of MAGPRS Parameter Significance Testing Using Partial Test

Parameter	Estimate	Standard Error (SE)	z	P-value
(Intercept)	2.336	4.37×10^{-15}	5.33×10^{14}	2×10^{-16}
BF1	-0.1303	5.30×10^{-16}	-2.45×10^{14}	2×10^{-16}
BF2	-0.0795	4.90×10^{-16}	-1.62×10^{14}	2×10^{-16}
BF3	-0.0197	7.49×10^{-16}	-2.63×10^{13}	2×10^{-16}
BF4	3.754	2.21×10^{-14}	1.69×10^{14}	2×10^{-16}
BF1 * BF5	0.0271	9.66×10^{-17}	2.80×10^{14}	2×10^{-16}
BF1 * BF6	-0.0078	8.13×10^{-17}	-9.70×10^{13}	2×10^{-16}
BF1 * BF7	-0.0325	1.03×10^{-16}	-3.13×10^{14}	2×10^{-16}
BF1 * BF8	-0.0196	8.77×10^{-17}	-2.24×10^{14}	2×10^{-16}
BF9 * BF4	2.726	1.57×10^{-13}	1.73×10^{13}	2×10^{-16}
BF10 * BF4	-0.2189	1.31×10^{-15}	-1.67×10^{14}	2×10^{-16}
BF1 * BF4	-0.2368	1.35×10^{-15}	-1.75×10^{14}	2×10^{-16}
BF1 * BF3	0.01056	6.50×10^{-17}	1.62×10^{14}	2×10^{-16}
BF9 * BF11 * BF4	-0.0228	1.57×10^{-15}	-1.45×10^{13}	2×10^{-16}
BF1 * BF5 * BF12	0.00039	2.97×10^{-18}	1.33×10^{14}	2×10^{-16}
BF1 * BF5 * BF13	0.01073	2.37×10^{-17}	4.52×10^{14}	2×10^{-16}
BF1 * BF7 * BF3	0.00077	4.56×10^{-18}	1.69×10^{14}	2×10^{-16}
BF1 * BF8 * BF3	0.00316	1.62×10^{-17}	1.94×10^{14}	2×10^{-16}

Based on Table 5, with a significance level of $\alpha = 0.05$, the critical value of $z - table$ is $z_{0.025} = 1.96$. Since all the $|z|$ values of all basis functions are $> z_{0.025}$, leading to the rejection of H_0. This indicates that all coefficients in the regression model are statistically significant. Therefore, each selected basis function in the MAGPRS model makes a meaningful contribution to the response variable.

Importance Level of Predictor Variables

The importance level of each predictor variable is presented in Table 6.

Table 6. Importance Level of Predictor Variables

Variable	Importance Level (%)
X_1	100.00
X_5	88.80
X_4	73.70
X_3	57.10
X_2	41.80

Based on Table 6, the importance levels of predictor variables in maternal mortality vary. The percentage of pregnant women participating in the K4 program (X_1) is the most influential variable, with an importance level of 100.00%, indicating that this factor plays the most dominant role in determining maternal mortality. The percentage of pregnant women receiving Td3 immunization (X_5) follows with an importance level of 88.80%, while the percentage of postpartum mothers receiving vitamin A (X_4) has 73.70%, suggesting that both variables also have a significant impact. Meanwhile, the percentage of deliveries occurring in healthcare facilities (X_3) has an importance level of 57.10%, and the percentage of pregnant women participating in the K1 program (X_2) has 41.80%, indicating that while they still contribute, their influence is relatively lower compared to the other variables.

4. CONCLUSION

The findings of this study indicate that all basis functions in the optimal model significantly impact maternal mortality. The best model was identified with a parameter combination of BF = 20, MI = 3, and MO = 2, yielding a Generalized Cross-Validation (GCV) value of 0.000046683 and an R-squared value of 99.30%, indicating a high level of model accuracy. Among the predictor variables, the most influential factor affecting maternal mortality is the percentage of pregnant women participating in the K4 program (X_1), followed by the percentage of pregnant women receiving Td3 immunization (X_5), the percentage of postpartum mothers receiving vitamin A (X_4),

the percentage of deliveries occurring in healthcare facilities (X_3) and the percentage of pregnant women participating in the K1 program (X_2). These findings highlight the significance of maternal healthcare programs in reducing maternal mortality by emphasizing the importance of prenatal care, institutional childbirth, postpartum nutritional support, and immunization programs in improving maternal health outcomes.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] N. Ikhsani, E.T. Herdiani, A. Islamiyati, Geographically Weighted Bivariate Poisson Inverse Gaussian Regression Modeling with the Berndt-Hall-Hall-Hausman Algorithm on Maternal and Neonatal Mortality Data, Commun. Math. Biol. Neurosci. 2024 (2024), 126. https://doi.org/10.28919/cmbn/8900.

[2] M.R. da Rato, R.D. Guntur, Generalized Poisson Regression Modeling on the Number of Infant Deaths in East Nusa Tenggara Province in 2022, J. Stat. 17 (2024), 779-788.

[3] A. Prahutama, undefined. Sudarno, Modelling Infant Mortality Rate in Central Java, Indonesia Use Generalized Poisson Regression Method, J. Phys.: Conf. Ser. 1025 (2018), 012106. https://doi.org/10.1088/1742-6596/1025/1/012106.

[4] D.R.S. Saputro, A. Susanti, N.B.I. Pratiwi, The Handling of Overdispersion on Poisson Regression Model with the Generalized Poisson Regression Model, AIP Conf. Proc. 2326 (2021), 020026. https://doi.org/10.1063/5.0040330.

[5] J.U. Ibeji, T. Zewotir, D. North, L. Amusa, Modelling Fertility Levels in Nigeria Using Generalized Poisson Regression-Based Approach, Sci. Afr. 9 (2020), e00494. https://doi.org/10.1016/j.sciaf.2020.e00494.

[6] M.A. Haris, P.R. Arum, Negative Binomial Regression and Generalized Poisson Regression Models on the Number of Traffic Accidents in Central Java, BAREKENG: J. Ilmu Mat. Ter. 16 (2022), 471-482. https://doi.org/10.30598/barekengvol16iss2pp471-482.

[7] A. Islamiyati, R. Raupong, A. Kalondeng, U. Sari, Estimating the Confidence Interval of the Regression Coefficient of the Blood Sugar Model Through a Multivariable Linear Spline with Known Variance, Stat. Transit. New Ser. 23 (2022), 201-212.

[8] A. Islamiyati, Spline Longitudinal Multi-Response Model for the Detection of Lifestyle- Based Changes in Blood Glucose of Diabetic Patients, Curr. Diabetes Rev. 18 (2022), e171121197990. https://doi.org/10.2174/1573399818666211117113856.

[9] A. Islamiyati, A. Kalondeng, M. Zakir, S. Djibe, U. Sari, Detecting Age Prone to Growth Retardation in Children Through a Bi-Response Nonparametric Regression Model with a Penalized Spline Estimator, Iran. J. Nurs. Midwifery Res. 29 (2024), 549-554. https://doi.org/10.4103/ijnmr.ijnmr_342_22.

[10] A. Islamiyati, F. Fatmawati, N. Chamidah, Changes in Blood Glucose 2 Hours After Meals in Type 2 Diabetes Patients Based on Length of Treatment at Hasanuddin University Hospital, Indonesia, Rawal Med. J. 45 (2020), 31-34.

[11] A. Islamiyati, A. Kalondeng, N. Sunusi, M. Zakir, A.K. Amir, Biresponse Nonparametric Regression Model in Principal Component Analysis with Truncated Spline Estimator, J. King Saud Univ. - Sci. 34 (2022), 101892. https://doi.org/10.1016/j.jksus.2022.101892.

[12] M. Bekar Adiguzel, M.A. Cengiz, Model Selection in Multivariate Adaptive Regressions Splines (MARS) Using Alternative Information Criteria, Heliyon 9 (2023), e19964. https://doi.org/10.1016/j.heliyon.2023.e19964.

[13] N. Isra, A. Islamiyati, E.T. Herdiani, Implementation of Differential Evolution Algorithm for MARS Modeling Optimization on Indonesia Composite Index Data, Commun. Math. Biol. Neurosci. 2025 (2025), 38. https://doi.org/10.28919/cmbn/9168.

[14] Z. Qayyum, T. Mehmood, L.A. Al-Essa, Prediction of Inhibition Activity of Dihydrofolate Reductase Inhibitors with Multivariate Adaptive Regression Splines, IEEE Access 11 (2023), 50595-50604. https://doi.org/10.1109/access.2023.3272231.

[15] F. Fitriayu, A. Islamiyati, E.T. Herdiani, Regional Classification Based on Maternal Mortality Rate Using a Robust Semiparametric Geographically Weighted Poisson Regression Model, Commun. Math. Biol. Neurosci. 2024 (2024), 131. https://doi.org/10.28919/cmbn/8783.

[16] Dinas Kesehatan Provinsi Sulawesi Selatan, Profil Kesehatan Provinsi Sulawesi Selatan, Dinas Kesehatan Provinsi Sulawesi Selatan, Makassar, 2012.

[17] S. Hidayati, B.W. Otok, undefined. Purhadi, Parameter Estimation and Statistical Test in Multivariate Adaptive Generalized Poisson Regression Splines, IOP Conf. Ser.: Mater. Sci. Eng. 546 (2019), 052051. https://doi.org/10.1088/1757-899x/546/5/052051.

[18] A.Y. Kartini, M. Ishlahuddin, Multivariate Adaptive Generalized Poisson Regression Splines Untuk Pengembangan Model Prediksi Produksi Padi di Kabupaten Bojonegoro, J. Math. Educ. Sci. 5 (2022), 147-156. https://doi.org/10.32665/james.v5i2.562.

[19] K. Dewi, N. Gusriani, K. Parmikanti, Factors Affecting the Number of Infant Morality Cases in West Java for the 2019–2020 Period Using Generalized Poisson Regression (GPR), Eksakta: Berk. Ilm. Bid. MIPA 24 (2023), 259–270.

[20] F. Suryadi, S. Jonathan, K. Jonatan, M. Ohyver, Handling Overdispersion in Poisson Regression Using Negative Binomial Regression for Poverty Case in West Java, Procedia Comput. Sci. 216 (2023), 517-523. https://doi.org/10.1016/j.procs.2022.12.164.

[21] P. McCullagh, J.A. Nelder, Generalized Linear Models, Chapman & Hall, London, 1989.

[22] F. Famoye, On the Generalized Poisson Regression Model with an Application to Accident Data, J. Data Sci. 2 (2021), 287-295. https://doi.org/10.6339/jds.2004.02(3).167.

[23] J.H. Friedman, C.B. Roosen, An Introduction to Multivariate Adaptive Regression Splines, Stat. Methods Med. Res. 4 (1995), 197-217. https://doi.org/10.1177/096228029500400303.

[24] A.H. Naser, A.H. Badr, S.N. Henedy, K.A. Ostrowski, H. Imran, Application of Multivariate Adaptive Regression Splines (mars) Approach in Prediction of Compressive Strength of Eco-Friendly Concrete, Case Stud. Constr. Mater. 17 (2022), e01262. https://doi.org/10.1016/j.cscm.2022.e01262.