ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 14, Nomor 1, Tahun 2025, Halaman 42 - 53
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

COMPARISON OF NEGATIVE BINOMIAL SAR, SEM, AND SARMA METHODS IN MODELING THE NUMBER OF MALNUTRITION CASES AMONG TODDLERS IN CENTRAL JAVA

Andi Rosilala[1*], Siti Hadijah Hasanah[2]
[1,2] Department of Statistics, Faculty of Science and Technology, Universitas Terbuka
***e-mail** : arosilala@gmail.com

DOI: 10.14710/j.gauss.14.1.42-53

Article Info:
Received: 2024-12-03
Accepted: 2025-02-17
Available Online: 2025-02-20

Keywords:
negative binomial; overdispersion; spatial regression; malnutrition.

Abstract: Malnutrition among toddlers remains a critical public health issue in Indonesian, with high rates of stunting, wasting, and underweight. The 2023 Indonesian Health Survey in Central Java reports 20.7% stunting, 7.1% wasting, 14.4% underweight, and 4.2% overweight, closely reflecting national trends. This study employs spatial analysis to examine malnutrition patterns at Central Java, finding overdispersion in the data (mean = 69.8; variance = 1033.4), which supports the use of the Negative Binomial model over Poisson. Moran's I and Getis-Ord G tests confirm spatial dependence. Among the spatial models tested—SAR, SEM, and SARMA—the SARMA-NB model was the most suitable, with optimal fit measures (AIC = 299.37, MAD = 9.74, MAPE = 15.88, RMSE = 12.77). Significant predictors include the percentage of children with complete immunization, exclusive breastfeeding rates, per capita meat consumption, access to sanitation, health insurance with contribution assistant coverage, poverty levels, low birth weight incidence, and inadequate housing. The findings emphasize the importance of spatial analysis in understanding malnutrition risk factors.

1. INTRODUCTION

Indonesia faces a severe malnutrition crisis with three key issues: undernutrition, micronutrient deficiencies, and rising obesity. Undernutrition affects 1 in 5 toddlers, with 1 in 12 experiencing wasting, and over 460,000 toddlers suffering from severe wasting, increasing their risk of stunting. Micronutrient deficiencies, particularly anemia, are critical, with 1 in 6 mothers underweight and 280,000 babies born with low birth weight, indicating poor maternal nutrition. Meanwhile, obesity is rising, with nearly one million toddlers and 1 in 5 schoolchildren overweight, posing future health risks (UNICEF, 2022). The National Medium-Term Development Plan (RPJMN) aims to reduce stunting to 14% and wasting to 7% by 2024 (Bappenas, 2021), but recent trends suggest these targets may be difficult to achieve. In Central Java, 2,443,282 child malnutrition cases were reported in 2021, with Brebes, Cilacap, and Semarang City recording the highest numbers, and 15 districts exceeding the provincial average of 70 cases per 1,000 toddlers. According to the 2023 Indonesia Health Survey (SKI), the province's malnutrition prevalence is close to the national average: stunting (20.7% vs. 21.5%), wasting (7.1% vs. 8.5%), underweight (14.4% vs. 15.9%), and overweight (4.2% vs. 4.2%) (Ministry of Health, 2023).

A spatial approach is highly relevant in analyzing the distribution of malnutrition, as it is influenced by factors such as economic conditions, healthcare services, and community lifestyles, which are localized and interconnected across regions (Usada, Wanodya, and Trisna, 2021). Various studies have demonstrated the effectiveness of spatial methods in investigating the determinants of malnutrition. Fatmah and Sutanto (2014) compared OLS, SAR, and SEM models in East Java, identifying low birth weight, iron tablet consumption, and clean and healthy living behavior (PHBS) as significant factors in the SAR model, while immunization, low birth weight, and PHBS were significant in the SEM model. Ramadani, Rahmawati, and Hoyyi (2013) employed the Spatial Durbin Model (SDM) in Central Java, finding low birth weight, access to clean water, and healthy housing as influential variables. Mustika and Sulistyawan (2019) revealed exclusive breastfeeding, healthy housing, and active community health posts (posyandu) as significant in the SEM model for East Java. Rohimah, Nuraidi, and Djuraidah (2011) applied the SAR Poisson model in East Java, linking malnutrition to slum areas, land use structures, and per capita GDP. Lastly, Amelia, Djuraidah, and Anisa (2023) used Negative Binomial (NB) models within SAR, SLX, and SDM frameworks for stunting in West Java, with the SAR-NB model performing best. This model highlighted exclusive breastfeeding, food facilities, and low birth weight as the primary determinants.

Based on the six studies mentioned, this research utilizes 11 predictors to explain malnutrition cases in Central Java, where data exploration reveals overdispersion, indicated by a variance exceeding the mean, making a standard Poisson model inadequate (Hilbe, 2011, in Wahyuni, 2011). To address this, the NB model is initially selected due to its ability to handle overdispersion by introducing an additional parameter for data variability. However, malnutrition is also influenced by spatial dependencies between regions and according to the Lagrange Multiplier (LM) test confirms that these data exhibit to the consideration of SAR, SEM, and SARMA models. Consequently, these three models will be compared to determine the most suitable spatial regression approach. By identifying the best model, this study aims to provide empirical evidence to guide policymakers in designing more effective interventions, optimizing resource allocation, enhancing nutrition programs, and improving healthcare accessibility in high-risk areas.

2. LITERATURE REVIEWS

Negative Binomial Regression is used to analyze count data with overdispersion, where the variance exceeds the mean (Hilbe, 2011 in Wahyuni, 2011). It models the relationship between a response variable Y (non-negative integers) and predictor variables $X_1, X_2, \dots, X_k$ as $y_i = x_i'\beta + \varepsilon_i$ with expected value $\mu = x_i'\beta$. Since Y_i must remain non-negative, a link function is used to transform the linear predictor $x_i'\beta$. The canonical link function is complex, but the log link $\ln(\mu_i) = x_i'\beta$ is often employed for simplicity. The model is expressed as: $\ln\{E[Y_i|x_i]\} = x_i'\beta$, giving $\mu_i = \exp(x_i'\beta)$.

Unlike linear regression, NB models define the response variable's distribution rather than the error term. Overdispersion is detected by examining the deviance or Pearson Chi-Square divided by degrees of freedom, with values greater than 1 indicating its presence. The deviance is calculated as:

$$D = 2\sum_{i=1}^{n}\left[y_i\, ln\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)\right]$$

D = Deviance

y_i = The actual observed value of the response for the i-th

$\hat{y}_i$ = The estimated value of the response for the i-th

While the Pearson Chi-Square value is:

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i}$$

X^2 = The Pearson Chi-Square value

y_i = the observed value for the i-th observation

$\hat{y}_i$ = the predicted or fitted value for the i-th observation

Negative Binomial Spatial Regression models count data while incorporating spatial dependencies. The general equation of spatial regression is: (Djuraidah and Anisa, 2023).

$$\boldsymbol{y} = \rho \boldsymbol{W}_1 \boldsymbol{y} + \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{W}_2 \boldsymbol{X} \gamma + \boldsymbol{\mu}$$

- y: Response variable vector $(n \times 1)$, representing counts at each location.
- $\rho W_1 y$: Autoregressive term where ρ indicates the influence of neighboring locations, and W_1 $(n \times n)$ is the spatial weight matrix for y.
- $X^* \beta^*$: Predictor matrix $\big(n \times (p+1)\big)$ with coefficients $\beta^* \big((p+1) \times 1\big)$, capturing the direct effects of predictors.
- $W_2 X_\gamma$: Spatial component where $(n \times n)$ accounts for spatial effects of predictors, and γ measures spatial influence on predictors.

To incorporate the NB Regression component into the spatial regression model, we can express as follow:

$$y_i \sim NB(\mu_i, \theta)$$

$$\ln(\mu_i) = \rho \boldsymbol{W_1} \boldsymbol{y} + \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{W_2} \boldsymbol{X} \gamma$$

Where the response variable y_i captures overdispersion using the dispersion parameter θ and log link function ensures $\mu_i > 0$, making it suitable for count data.

The error term (μ) captures spatial dependence as:

$$\mu = \lambda \boldsymbol{W_3} \mu + \varepsilon, \varepsilon \sim N(0, \sigma^2 I)$$

- λ: Spatial autoregressive coefficient for residuals.
- W_3: Spatial weight matrix for residuals $(n \times n)$.
- ε: Error vector with mean 0 and constant variance σ^2.

SAR (Spatial Autoregressive) is a spatial regression model that incorporates spatial dependence in the response variable. The general equation for the spatial model, when $\rho \neq 0$ and $\lambda = 0$ becomes the SAR model. According to Anselin (1988) as cited in Djuraidah and Anisa (2023), the spatial model is defined as follows:

$$\boldsymbol{y} = \rho \boldsymbol{W}_1 \boldsymbol{y} + \boldsymbol{X} \boldsymbol{\beta} + \varepsilon$$

Where y is the response variable. W_1 is the spatial weight matrix for y. β is the regression coefficient vector. ε is the residual. ρ is the spatial autoregressive lag coefficient. The equation for the NB Spatial Autoregressive (SAR) model is as follows (Glaser, 2017, in Amelia, Djuraidah, and Anisa, 2023).

$$\mu_i = e^{(\rho \boldsymbol{w}'_{1i} \boldsymbol{y} + \boldsymbol{x}_i \boldsymbol{\beta})}$$

SEM (Spatial Error Model) is a spatial model that arises due to spatial effects on the error term. In the equation above, when $\rho = 0$ and $\lambda \neq 0$, the model becomes the SEM, and the spatial model is defined as follows (Anselin, 1988, in Djuraidah and Anisa, 2023):

$$\boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \mu$$

$$\mu = \lambda \boldsymbol{W} \mu + \varepsilon$$

Using the log-linear function for the mean μ, the NB SEM can be expressed as (Glaser, 2017):

$$\boldsymbol{y} = \boldsymbol{X\beta} + \mu$$
$$\mu = \lambda \boldsymbol{W}\mu + \varepsilon \gg (\boldsymbol{I} - \lambda \boldsymbol{W})^{-1}\varepsilon$$
$$\mu_i = e^{(\boldsymbol{x_i\beta} + (\boldsymbol{I} - \lambda \boldsymbol{w})^{-1}\varepsilon)}$$

SARMA (Spatial Autoregressive Moving Average) is a combination of the SAR and SEM spatial models. In the equation, when $\rho \neq 0$ and $\lambda \neq 0$, the SARMA model is formed (Sari, Hayati, and Wahyuningsih, 2020). The general form of SARMA model is as follows:

$$\boldsymbol{y} = \rho \boldsymbol{W}_1 \boldsymbol{y} + \boldsymbol{X\beta} + \mu$$
$$\mu = \lambda \boldsymbol{W\mu} + \varepsilon$$
$$\mu = (\boldsymbol{I} - \lambda \boldsymbol{W})^{-1}\varepsilon$$

To satisfy the characteristics of the NB distribution, the expectation of the response variable y is expressed on a log scale as follows:

$$\mu_i = e^{\left(\rho \boldsymbol{w'_{1i}y} + \boldsymbol{x_i\beta}\right) + ((\boldsymbol{I} - \lambda \boldsymbol{w})^{-1}\varepsilon)}$$

Where $y_i \sim NB(\mu_i)$

The form of parameter estimation for the SAR, SEM, and SARMA regression models is obtained using the maximum likelihood method (Wahyuni, 2011; Sauddin, Auliah, and Alwi, 2020; & Sari, Hayati, and Wahyuningsih, 2020). The steps for interpreting the parameters are as follows:

1. Forming the likelihood function.

$$L(\beta, \lambda, \alpha) = \prod_{i=1}^{n} f(y_i; \mu_i; \alpha)$$

With the probability (density) function of the NB distribution $f(y_i; \mu_i; \alpha)$ given by:

$$f(y_i; \mu_i; \alpha) = \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{y_i!\,\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\frac{1}{\alpha}}$$

2. Forming the log function of the obtained likelihood function.

$$L(\beta, \lambda, \alpha) = ln\, L\,(\beta, \lambda, \alpha)$$
$$= \sum_{i=1}^{n} \left[\ln \Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln y_i! - \ln \Gamma\left(\frac{1}{\alpha}\right) + y_i \ln\left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right) + \frac{1}{\alpha}\ln\left(\frac{1}{1 + \alpha\mu_i}\right)\right]$$

Where:

$\mu_i = e^{(\rho \boldsymbol{w'_{1i}y} + \boldsymbol{x_i\beta})}$ for SAR;

$\mu_i = e^{(\boldsymbol{x_i\beta} + (\boldsymbol{I} - \lambda \boldsymbol{w})^{-1}\varepsilon)}$ for SEM;

$\mu_i = e^{\left(\rho \boldsymbol{w'_{1i}y} + \boldsymbol{x_i\beta}\right) + ((\boldsymbol{I} - \lambda \boldsymbol{w})^{-1}\varepsilon)}$ for SARMA

The estimation of parameters β, λ, and α in the SAR, SEM, and SARMA NB models using the Maximum Likelihood (ML) method requires complex numerical optimization due to the spatial components in the error term and the NB distribution. In computational implementation, the `spatialreg` or `spdep` packages in R do not directly support the NB distribution. Therefore, it is necessary to combine these packages with other approaches for NB estimation, such as using the `glm.nb` function.

3. RESEARCH METHOD

This study uses a quantitative approach with secondary data from 11 predictor variables to explain malnutrition cases in 29 regencies and 6 cities in Central Java province in 2021.

Table 1. Research Variables

Variable	Definition	Source
y	The number of malnourished toddlers (per 1000)	Central Java Statistics
x_1	The percentage of toddlers with complete basic immunization	Health Profile of Central Java
x_2	The percentage of toddlers under 6 months receiving exclusive breastfeeding	Central Java Statistics
x_3	The percentage of Standardized Food Processing Units	Health Profile of Central Java
x_4	The average weekly per capita consumption of chicken, beef, and goat meat (in kg).	Central Java Statistics
x_5	The percentage of households with access to adequate sanitation.	Health Profile of Central Java
x_6	The average monthly per capita expenditure on food. (100K Rupiah)	Central Java Statistics
x_7	The percentage of the population covered by the National Health Insurance with Contribution Assistance	Central Java Statistics
x_8	The percentage of the population below the poverty line	Central Java Statistics
x_9	The percentage of low birth-weight infants	Central Java Statistics
x_{10}	The percentage of stunted toddlers	Health Profile of Central Java
x_{11}	The number of uninhabitable houses	The Department of Public Housing and Settlement Areas of Central Java.

The analysis of malnutrition case data was conducted using the R software with the following modeling analysis steps:

1. Data Exploration: Present a summary of descriptive statistics for all variables, and visualize the response variable using distribution maps.
2. Multicollinearity Detection: Assess predictor variables using Variance Inflation Factor (VIF) to identify multicollinearity.
3. Overdispersion Test for overdispersion: Examine deviance and Pearson's chi-square divided by degrees of freedom. Values > 1 indicates overdispersion.
4. Regression Modeling: Apply Poisson regression if no overdispersion is detected. Use NB regression if overdispersion is present.
5. Spatial Effect Testing:
   a. Residual Homogeneity: Use the Breusch-Pagan test; p-value > 0.05 indicates homogeneous residual variance.
   b. Spatial Dependency: Conduct Lagrange Multiplier tests with spatial weight matrices to choose models SAR, SEM, or SARMA.
6. Spatial Dependency (Autocorrelation) Effect Testing: Use Moran's I to test global and local spatial autocorrelation. Apply Geary's C to identify hotspots and coldspots and its cluster.
7. Spatial NB Regression Parameter Estimation: Estimate parameters for the spatial NB regression model based on spatial effect tests.

8. Model Fit Evaluation: Select the model with the smallest AIC. Validate fit using MAD, MAPE, and RMSE metrics.
9. Interpret results from the best model and draw conclusions based on the findings.

4. RESULT AND DISCUSSION

Data exploration was carried out by generating descriptive statistics summaries, thematic maps of malnutrition distribution, VIF values for multicollinearity, and overdispersion analysis.

Table 2. Descriptive statistic of the variables

Variable	Min	Median	Mean	Max	SD	CV
y	7.00	68.00	69.80	133.00	32.15	46.06
x_1	0.40	0.90	0.89	1.39	0.17	19.41
x_2	0.19	0.25	0.26	0.39	0.05	18.24
x_3	0.37	0.72	0.72	0.97	0.12	16.78
x_4	0.00	0.05	0.05	0.07	0.01	31.22
x_5	0.41	0.89	0.84	0.98	0.13	14.95
x_6	36.38	52.32	52.93	74.03	8.44	15.95
x_7	0.30	0.44	0.44	0.65	0.07	16.96
x_8	0.05	0.11	0.11	0.18	0.04	31.22
x_9	0.10	0.14	0.14	0.20	0.03	20.77
x_{10}	0.02	0.09	0.09	0.23	0.05	50.28
x_{11}	1,816.00	40,480.00	48,776.06	245,022.00	46,928.85	96.21

In the Table 2, the response variable y has a mean of 69.8 with a coefficient of variation (CV) of 46.06%, indicating a high level of data dispersion. Predictor variables x_1 through x_5 have relatively small means and SD, suggesting stable data distribution. Meanwhile x_{11} has a CV 96.21%, indicating very high dispersion. The variable x_{10} has a small mean with a low SD, indicating that the data is concentrated around its average value.

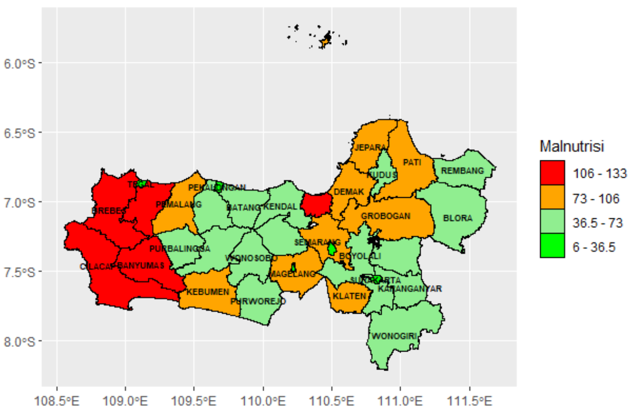

Figure 1. The thematic map of malnutrition clusters in Central Java for 2021

Figure 1 illustrates the clustering of malnutrition cases, with the highest concentrations (indicated in red) observed in western regions such as Brebes and Cilacap. Conversely, areas with the lowest prevalence (indicated in green) are distributed across most other regions.

Table 3. VIF value for each variable

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
VIF	1.27	1.41	1.72	1.44	2.00	2.36	1.79	2.34	1.43	2.28	1.39

Table 3 shows that the model is free from multicollinearity, as all VIF values are below the threshold of 5, allowing the regression analysis to proceed without any issues.

Table 4. Overdispersion Testing Results

Test	Statistic Test	df	Statistic Test/df
Chi-sq Pearson	306.79	23	13.34
Deviance	305.94	23	13.30

Table 4 shows that both the Pearson chi-square and deviance values per degree of freedom exceed 1, indicating a violation of the equidispersion assumption required for Poisson regression. This is further supported by the mean of the response variable y (69.80) and its variance (1033.40), confirming the presence of overdispersion. To reinforce this finding, additional visual assessments, the Cullen & Frey plot and Q-Q plot, can be utilized.

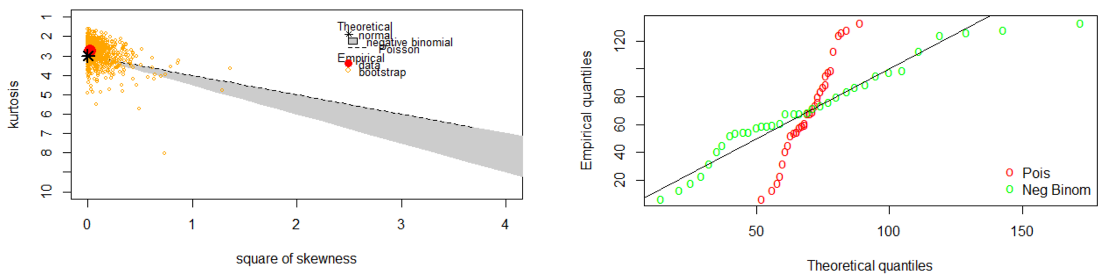

Figure 2. Cullen & Frey Graph and Q-Q Plot

According to Figure 2, the Cullen and Frey plot suggests the data resemble a NB distribution, showing kurtosis between 2 and 4 and skewness squared below 1, indicating overdispersion. The Q-Q plot confirms this, with the data aligning closely with the NB (green points) and deviating from the Poisson distribution (red points), further supporting the presence of overdispersion. Simultaneous testing in the Poisson and NB regression models yielded G-test values of 248.3 and 23.09, both exceeding the critical value of $\chi^2(0.05, 11) = 19.67$, indicating statistical significance for at least one predictor. In the Poisson model, predictors x_1, x_3, x_4, x_5, x_8, x_{10}, and x_{11} were significant, while in the NB model, only x_4 and x_{11} were significant at $\alpha = 0.05$.

Table 5. Poisson and NB regression estimates

Model	Coefficient						AIC
	β_0	β_1	β_2	β_3	β_4	β_5	
Pois	6.33*	-0.67*	0.45	-0.58*	-10.62*	-1.11*	537.72
NB	7.23*	-0.78	0.42	-0.79	-16.60*	-1.15	348.02
	β_6	β_7	β_8	β_9	β_{10}	β_{11}	
Pois	1.02e-03	-0.49	2.89*	0.13	-2.08*	3.70e-06*	537.72
NB	-4.64e-03	-1.23	4.63	0.42	-1.91	4.29e-06*	348.02

*) significant at the $\alpha = 0.05$

From the table 5, the NB model is more appropriate as it exhibits a smaller AIC, indicating a better fit for the data characterized by overdispersion. In the NB model, the significant predictors are average meat consumption per capita per week and the number of uninhabitable houses. The equation for the NB regression model is as follows:

$$\ln(\hat{\mu}) = 7.23 - 0.78x_1 + 0.42x_2 - 0.79x_3 - 16.60x_4 * -1.15x_5 - 0.00464x_6 - 1.23x_7 + 4.63x_8 + 0.42x_9 - 1.91x_{10} + 4.299e06x_{11} *$$

The spatial effect tests include heteroscedasticity and spatial dependence. The Breusch-Pagan test for spatial heteroscedasticity gave a statistic of 10.994, below the critical value of $\chi^2_{(0.05;11)}$ = 19.675, indicating no significant heterogeneity and no need for a geographically weighted NB regression. The LM test and Moran's Index test revealed a Moran's Index of 2.320 with a p-value of 0.010, indicating significant spatial dependence in the malnutrition cases among children under five in Central Java.

To demonstrate spatial dependence, a Moran's I test was conducted along with a scatter plot that divides the region into hotspot (H-H), coldspot (L-L), and other quadrants, which can be visualized on a thematic map below.

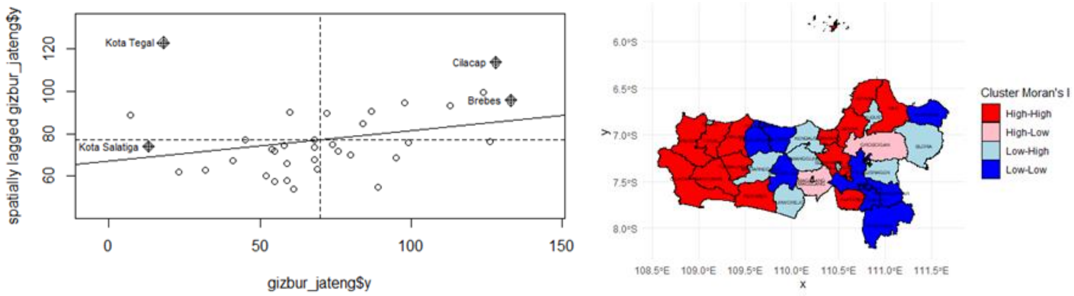

Figure 3. The results of the Moran I plot and the Moran I cluster map

Based on Figure 3, the H-H (hotspot) region consists of Cilacap and Brebes, which exhibit a positive autocorrelation with a high number of malnutrition cases surrounded by similar areas. The H-L region indicates a local anomaly, such as Tegal City, which has low cases but is surrounded by areas with high cases. The L-L (coldspot) region, such as Salatiga City and several southeastern areas, has low cases surrounded by similar regions, while the L-H region shows a local anomaly with high cases surrounded by areas with low cases, such as around Grobogan Regency.

After confirming the existence of spatial dependence, the LM test is conducted to determine the appropriate spatial model using various weight matrices such as Queen Contiguity, k-nearest neighbor (KNN), inverse distance weight (IDW), and negative exponential.

Table 6. The four weight matrices for the LM test

Weight Matrices	LM Test	Statistic	p-value	Weight Matrices	LM Test	Statistic	p-value
	RSlag	0.49	0.48		RSlag	2.69	0.10
	RSerr	3.15	0.08		RSerr	0.00	0.97
Queen	AdjRSlag	8.12	0.00**	KNN	AdjRSlag	8.64	0.00**
	AdjRSerr	10.78	0.00**		AdjRSerr	5.95	0.01
	SARMA	11.27	0.00**		SARMA	8.64	0.01*
	RSlag	4.36	0.04*		RSlag	0.02	0.89
	RSerr	5.59	0.02*		RSerr	0.64	0.42
IDW	AdjRSlag	0.11	0.74	Exponential	AdjRSlag	6.77	0.00**
	AdjRSerr	1.34	0.25		AdjRSerr	7.38	0.00**
	SARMA	5.70	0.06		SARMA	7.40	0.03*

According to Table 6, from all the weight matrices, the LM test results that are significant at α = 0.05 include spatial lag, spatial error, and SARMA (a combination of lag and error). Therefore, the models that will be used in this study are SAR, SEM, and SARMA.

SAR NB adds spatial dependence to the response variable, while SEM NB addresses spatial effects on the error term. If both lag and error dependencies are present, the model used is SARMA NB, with the parameter ρ representing the lag effect on the dependent variable and λ representing the lag effect on the residuals. The estimated parameters for the spatial NB regression model are presented in the following table:

Table 7. The estimation of the parameters for the spatial NB regression

NB Model	Coefficient						
	β_0	β_1	β_2	β_3	β_4	β_5	β_6
SAR	6.63	-0.63	-0.29	-0.57	-20.77*	-0.85	-0.01
SEM	6.97	-1.17*	2.42	-0.55	-16.81*	-1.91*	0.01
SARMA	5.42	-0.84*	2.28*	0.02	-21.87*	-1.28*	-2.11e-03
	β_7	β_8	β_9	β_{10}	β_{11}	ρ	λ
SAR	-1.89	3.48	2.9	0.39	4.65e-06*	0.01*	-
SEM	-1.58	6.80*	-1.64	-3.41	5.05E-06	-	-0.02*
SARMA	-2.78*	5.69*	3.11*	0.52	5.64e-06*	0.01*	-0.03*

*) significant for the α = 0.05

The selection of the best model for predicting malnutrition cases in Central Java is based on the model with the smallest AIC value, the significance of ρ and λ, as well as the error matrix values, including MAD, MAPE, and RMSE.

Table 8. The indicators for selecting the best spatial NB model

NB Model	AIC	ρ (LagY)	λ (LagResidual)	MAD	MAPE	RMSE
NB	348.02	-	-	22.29	39.94	28.52
SAR	344.62	0.01*	-	18.6	34.43	24.74
SEM	338.33	-	-0.02*	20.73	33.15	24.94
SARMA	299.37	0.01*	-0.03*	9.72	15.88	12.77

Based on Table 8, the SARMA-NB model is the best model as it has the lowest AIC (299.37) and better MAD, MAPE, and RMSE values compared to the other models, with the following equation:

$$
\begin{aligned}
\ln(\hat{\mu}) = 5.42 + 0.01394 \sum_{j=1}^{n} w_{i,j} Y_j * -0.84X_{1i}^* + 2.28X_{2i}^* + 0.02X_{3i} - 21.87X_{4i}^* \\
- 1.28X_{5i}^* - 0.002X_{6i} - 2.78X_{7i}^* + 5.69X_{8i}^* + 3.11X_{9i}^* + 0.52X_{10i} \\
+ 5.642e06X_{11i}^* - 0.02971 \sum_{j=1}^{n} w_{i,j}\varepsilon_j *
\end{aligned}
$$

The coefficient ρ = 0.01394 indicates that for every 1 unit increase in the number of malnutrition cases in neighboring areas (lag), the number of malnutrition cases in the region increases by approximately exp(0.01394) or 1.4%. The coefficient λ = -0.02971 means that if a region is surrounded by other areas, the residuals of that region are corrected by the residuals of its neighboring areas by exp(-0.02971) or 2.9% of the spatial weight value of the surrounding region. The model identifies key factors influencing malnutrition in Central Java in 2021 at α = 0.05: complete immunization rate (x_1), exclusive breastfeeding (x_2), meat consumption (x_4), sanitation access (x_5), national health insurance with contribution assistant coverage (x_7), poverty (x_8), low birth weight (x_9), and uninhabitable houses (x_{11}).

Golgher and Voss (2016) in Djuraidah (2020) provide a method for interpreting spatial regression coefficients by examining both direct and indirect effects, as follows:

Table 9. Effect on SARMA NB Model

Effect	**Variable**							
	x_1	x_2	x_4	x_5	x_7	x_8	x_9	x_{11}
Direct	-0.84	2.28	-21.87	-1.28	-2.78	5.69	3.11	0.00
Indirect	-0.01	0.03	-0.31	-0.02	-0.04	0.08	0.04	0.00
Total	-0.85	2.31	-22.18	-1.29	-2.82	5.77	3.15	0.00

According to table 9, the direct effects for each variable are not significantly different from the estimated parameter values in the model. Additionally, the indirect effects generated are relatively small. This is due to the small value of the lag coefficient (ρ) in the SAR NB model. The direct effect of variable x_1 is -0.84, meaning that if the percentage of complete basic immunization rate in a region increases by one percent, the average number of malnutrition cases in that region will decrease by 56.83% per 1000 children, holding other predictors constant. Meanwhile, the indirect effect of variable x_1 is -0.01, meaning that if the complete basic immunization rate increases by one percent in a region, the average number of malnutrition cases in another region will decrease by 1.2% per 1000 children, holding other predictors constant. The same interpretation applies to the rest of variables. Visually, the thematic maps of the direct and indirect effects can be seen in the image below.

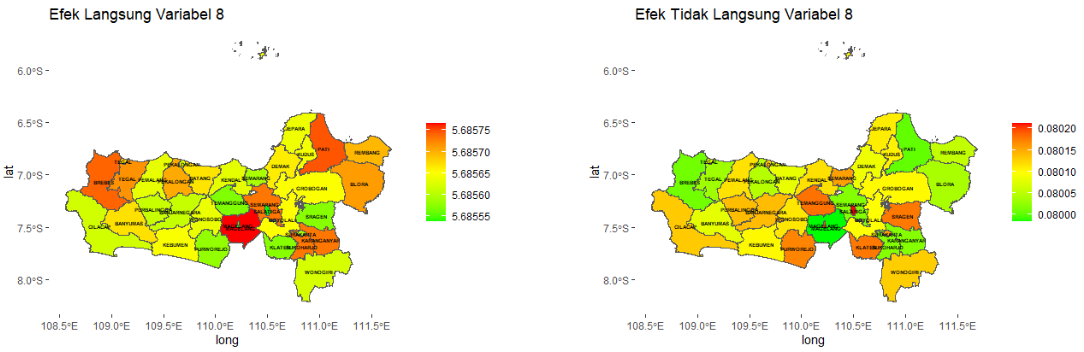

Figure 4. Thematic Map for direct effect and indirect effect on variable x_8

As shown in Figure 4, areas marked in red, such as Magelang and Semarang Regency, have the highest direct effect values, indicating that an increase in x_8 (the percentage of the population below the poverty line) significantly raises malnutrition cases in these regions. In contrast, areas in green, such as Temanggung and Klaten, show lower direct effects, meaning x_8 has a smaller influence on malnutrition. The indirect effect reflects how changes in x_8 in one area impact neighboring regions; for example, Semarang significantly affects Kendal and Salatiga, suggesting that rising poverty in Semarang may contribute to increased malnutrition in surrounding areas due to factors like migration or regional policies. In the northern coastal region (Pantura), covering Tegal, Brebes, and Cirebon, indirect effects vary based on proximity and socio-economic relationships, with Tegal and Brebes more strongly influenced by surrounding poverty levels, while Cirebon experiences a smaller impact.

5. CONCLUSION

This study models 11 factors influencing malnutrition cases in Central Java in 2021, finding that the NB model is more appropriate than the Poisson model due to overdispersion. When a spatial approach was applied, the data showed spatial dependence in variance. Using five weighting matrices, three spatial models were identified and compared for their performance: SAR, SEM, and SARMA. Based on model fit tests, the best spatial regression model is SARMA NB. The variables significantly affecting malnutrition cases in Central Java are complete basic immunization rates (x_1), exclusive breastfeeding (x_2), meat consumption per capita per week (x_4), sanitation access (x_5), national health insurance with contribution assistant coverage (x_7), poverty (x_8), low birth weight (x_9), and uninhabitable houses (x_{11}). To reduce malnutrition in Central Java, improving access to quality health services, promoting exclusive breastfeeding, and enhancing sanitation facilities are essential. Furthermore, addressing economic disparities and increasing education on nutrition can help mitigate malnutrition risks. Lastly, strengthening collaboration between local and regional authorities will ensure more effective interventions in high-risk areas. Future studies should incorporate time-series data with panel data to explore both spatial and temporal dependencies using models. This approach will provide deeper insights into malnutrition trends and improve long-term policy recommendations and forecasts.

REFERENCES

Amelia, M., Djuraidah, A., dan Anisa, R. (2023). "Pendugaan Faktor – Faktor yang Memengaruhi Kasus Stunting di Jawa Barat Tahun 2021 Menggunakan Regresi Spasial Binomial Negatif". Jurnal Matematika, Statistika, dan Komputasi. Volume 20, Nomor 1.

Anselin L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Academic Publisher.

Djuraidah, A. 2020. Monograph Penerapan dan Pengembangan Regresi Spasial dengan Studi Kasus pada Kesehatan, Sosial, dan Ekonomi. Bogor: IPB Press.

Djuraidah, A. dan Anisa R. (2023). "Analisa Statistika untuk Data Geospasial". Bogor: IPB Press.

Fatmah, I. N. dan Sutanto, H. T. (2013). "Pemodelan Faktor – Faktor yang Mempengaruhi Balita Gizi Buruk di Jawa Timur dengan Pendekatan Regresi Spasial". MATHunesa. Volume 2 Nomor 3.

Glaser, S., (2017). *A Review of Spatial Econometric Models for Count Data*. Hohenheim Discussion Papers in Business, Economics and Social Sciences, Volume 19.

Golgher, A. B., dan Voss, P. R. (2016). *How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects*. Spatial Demography, Volume 4.

Hilbe, J. M., (2011). *Negative Binomial Regression* Second Edition. New York: Cambridge University Press.

Kementerian Kesehatan RI. "Survei Kesehatan Indonesia 2023". https://www.badankebijakan.kemkes.go.id/hasil-ski-2023/

Keputusan Menteri Perencanaan Pembangunan Nasional/Kepala Badan Perencanaan Pembangunan Nasional Tentang Penetapan Rencana Aksi Nasional Pangan dan Gizi Tahun 2021-2024.

Mustika, R. dan Sulistyawan, E. (2019). "Spasial Error Model untuk Balita Gizi Buruk Di di Provinsi Jawa Timur Tahun 2016". Jurnal Riset dan Aplikasi Matematika Volume 3 Nomor 1.

Pawitan, Y. (2001). *In all Likelihood: Statistical Modelling and Inference Using Likelihood.* Oxford: Clarendon Press.

Ramadani, I. R., Rahmawati, R., dan Hoyyi, A. (2013). “Analisis Faktor-Faktor yang Mempengaruhi Gizi Buruk Balita di Jawa Tengah dengan Metode Spatial Durbin Model. Jurnal Gaussian. Volume 2 Nomor 4.

Rohimah, S. R., Nuraidi, M., dan Djuraidah, A. (2011). “Model Spasial Autoregresif Poisson untuk Mendeteksi Faktor-Faktor yang Berpengaruh terhadap Jumlah Penderita Gizi Buruk di Provinsi Jawa Timur”. SNS, Volume 2.

Sari, D. N. E., Hayati, M. N., dan Wahyuningsih, S. (2020). “Model Spatial Autoregressive Moving Average (SARMA) pada Data Jumlah Kejadian Demam Berdarah Dengue (DBD) di Provinsi Kalimantan Timur dan Tengah Tahun 2016”. Jurnal EKSPONENSIAL Volume 11 Nomor 1.

Sauddin, A., Auliah, N., dan Alwi, W. (2020). “Pemodelan Jumlah Kematian Ibu di Provinsi Sulawesi Selatan Menggunakan Regresi Binomial Negatif”. Jurnal Matematika dan Statistika serta Aplikasinya Volume 8 Nomor 2.

Usada, K. N., Wanodya, S. K., dan Trisna, N. (2021). “Analisis Spasial Gizi K Analisis Spasial Gizi Kurang Balita di K ang Balita di Kota Tangerang Tahun 2019”. Jurnal Biostatistik, Kependudukan, dan Informatika Kesehatan. Volume 2 Nomor 1.

Wahyuni, W. (2011). “Penaksir Parameter Model Regresi Binomial Negatif pada Kasus Overdispersi”. Depok: Universitas Indonesia.

UNICEF. “Mengatasi tiga beban malnutrisi di Indonesia” https://www.unicef.org/indonesia/id/gizi?gad_source=1&gclid=CjwKCAjwyfe4BhAWEiwAkIL8sKzUXUhJarlsVgK-YVBsMWOMrC-IwekSLUX7YLXl1Jrc-JZrvzrZDhoCN4UQAvD_BwE