JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 9, No. 2, April 2025, pp. 600-614

Spatial Clustering Regression in Identifying Local Factors in Stunting Cases in Indonesia

Ummul Auliyah Syam[1*], Anik Djuraidah[1], Utami Dyah Syafitri[1]
[1]Department of Statistics, IPB University, Indonesia
ummulsyam@apps.ipb.ac.id

Article History:
Received : 18-02-2025
Revised : 17-04-2025
Accepted : 20-04-2025
Online : 29-04-2025

Keywords:
Cluster Analysis;
Geographically Weighted Regression;
Spatial Analysis;
Spatial Clustering Regression;
Stunting.

ABSTRACT

Stunting is a significant health problem in Indonesia with high spatial disparities between regions. This study applies the Spatial Clustering Regression (SCR) method to analyze spatial patterns and identify local factors influencing stunting. SCR is a method that combines spatial regression and clustering analysis simultaneously using a k-means clustering-based formulation and a penalty likelihood function motivated by the Potts model to encourage similar clustering in adjacent locations with regression parameter estimation done locally in areas that have similar characteristics. This quantitative study uses secondary data from the Central Bureau of Statistics in 2022 covering 510 districts/cities, with one response variable (percentage of stunting) and seven explanatory variables reflecting socioeconomic, health, and infrastructure conditions. The results show that SCR divides the region into four spatial clusters characterized by different local factors. Cluster 1 has the lowest percentage of stunting that is influenced by access to clean water, sanitation, and education, Cluster 2 by poverty rate, number of public health centers, access to clean water, and education, Cluster 3 by poverty and nutrition of pregnant women, and Cluster 4 is the most vulnerable area with the highest stunting rate with a significant influential factor which is access to sanitation. The SCR approach allows for easier and more in-depth interpretation of results than other spatial methods such as GWR, as it can capture complex spatial patterns in the form of regional clusterings. These results provide a strong basis for formulating region-specific intervention policies, such as poverty alleviation and sanitation improvement in Cluster 4, strengthening health services in Cluster 2, developing education and nutrition programs in Cluster 3, and maintaining and improving nutrition consumption in Cluster 1.

Crossref
https://doi.org/10.31764/jtam.v9i2.29803

——————————— ◆ ———————————

A. INTRODUCTION

Stunting is a growth disorder in children resulting from insufficient psychosocial stimulation, recurrent infections, and inadequate nutrition. Prolonged malnutrition can contribute to stunting, causing a child's height-for-age to fall more than two standard deviations below the median of normal child growth standards (WHO, 2015). It reflects the cumulative effects of chronic undernutrition during the first 1,000 days of life, is also associated with lack of education, poverty, and poor health, and indicates a poor quality of life that negatively impacts human capital (Anggraini & Romadona, 2020). Stunting is part of Sustainable Development Goal 2.2, which aims to eliminate all forms of malnutrition, including stunting and wasting, in children under the age of five by 2030 and reduce the prevalence of stunting by 40% by 2025. Based on data from the Indonesian Nutritional Status Survey, the national stunting rate declined from 24.4% in 2021 to 21.6% in 2022. Indonesia must ensure an annual reduction of 3.8% to achieve the target of 14% by 2024 (Ministry of Health, 2022).

Differences in geographical conditions, socio-economic conditions, and people's lifestyles can cause the factors that cause stunting in one region to another (Nursyamsiyah et al., 2021). Therefore, mapping stunting cases and the risk factors that cause them in each region is very important to know. Factors that vary in each location indicate spatial heterogeneity in stunting data that must be considered when building the model. Based on research conducted by Azis & Aswi (2023) t shows that spatial effects are present in stunting data. Ignoring spatial effects can produce biased and inconsistent estimators. Research related to the application of spatial models for stunting cases is very diverse, among others, has been conducted by Fitri et al. (2024) using the conditional autoregressive regression method, Djara et al. (2022) using spatial autoregressive with instrument variables and s-estimator, Eryando et al. (2022) using the spatial autoregressive model and Muche et al. (2021) using geographically weighted regression (GWR).

Spatial regression approaches in public health phenomena have widely accommodated spatial effects. Geographically weighted regression (GWR) and spatially varying coefficient (SVC) models have become popular methods in modeling spatial heterogeneity to capture the influence of explanatory variables on response variables in a region (Sugasawa & Murakami, 2021). However, this method has limitations in capturing complex spatial structures, especially when the relationship between response variables and explanatory variables changes in clusters at the boundary but is homogeneous within the region (Li & Sang, 2019). In addition, GWR also tends to produce extreme coefficient estimates caused by spatially varying coefficients that complicate interpretation because the spatial pattern of clusters is difficult to identify (Zhong et al., 2022). Meanwhile, spatial Bayesian methods can overcome some of these drawbacks, but they are not computationally efficient for large-scale data (Louzada et al., 2021).

The Spatial Clustering Regression (SCR) method addresses these limitations through a novel approach to spatial regression with spatially localized coefficients that combines spatial regression modelling and clustering analysis. SCR assumes that all geographic locations can be divided into a finite number of clusters with locations within the same cluster having the same regression coefficients. The results obtained using the clustering technique will be easier to interpret than GWR (Sugasawa & Murakami, 2021). In contrast to previous studies that separate the stages of clustering and spatial regression modelling, such as Lee et al. (2017), Nicholson et al. (2019), Zhong et al. (2022), and Sukmawati et al. (2021), SCR performs regression parameter estimation and clustering simultaneously. This method was introduced by Sugasawa & Murakami (2021) through a k-means based clustering formulation with a penalty function motivated by the Potts model to encourage the formation of geographically close spatial clusters. Simulation study results show that this method performs better than GWR.

One of the main advantages of SCR over previous methods is its ability to explicitly identify spatial clusters using a spatially-based penalty likelihood approach, making the resulting model more concise and easier to interpret. By grouping regions based on similar regression characteristics, SCR makes it easier to understand spatial patterns and develop region-specific policies according to local factors for public health issues such as stunting that have spatial heterogeneity effects. Based on this, this study aims to conduct SCR modelling to identify clustered spatial patterns in district/city level stunting data and identify local factors that significantly affect the percentage of stunting cases in Indonesia. The results of the study can provide precise spatial mapping and information related to local factors that affect stunting in each cluster so that it can be used as a basis for formulating more targeted and area-based stunting reduction policies.

B. METHODS

This study utilizes data on stunting cases in Indonesia, obtained from the Central Statistics Agency for the year 2022, covering 510 districts/cities as observation units. The dataset includes one response variable and seven explanatory variables, as outlined in Table 1. The selection of explanatory variables is based on previous research identifying factors associated with stunting.

Table 1. Variables Used in the Study

Variable	Description	Unit	Reference
x, y	Geographic coordinates of district/city (longitude and latitude)	Degrees (WGS84)	
Y	Stunting	Percent	
X_1	Poverty rate	Persen	(Ali, 2021)
X_2	Number of public health centers	Unit	(Rizal & van Doorslaer, 2019)
X_3	Nutrition of pregnant women	Calories	(Muhamad et al., 2023)
X_4	Access to sanitation	Percent	(Vaivada et al., 2020)
X_5	Access to clean water	Percent	(Cameron et al., 2021)
X_6	Gross Regional Domestic Product (GRDP)	Billion Rupiah	(Beal et al., 2018)
X_7	Average years of schooling	Year	(Anggraini & Romadona, 2020)

This study uses the SCR method which accounts for spatial heterogeneity in the model parameters by explicitly introducing a clustering parameter. SCR is computationally effective and can identify spatial clusters while simultaneously estimating regression parameters within each cluster, resulting in better estimates under the assumption that neighboring locations with similar characteristics have similar regression parameters. Data analysis was conducted using R Studio software, and the analysis process followed the following steps:

1. Exploring the data to determine the characteristics of the data on the percentage of stunting in districts/cities in Indonesia in general. Data exploration consisted of:
   a. Thematic maps to determine the distribution pattern of the percentage of stunting in districts/cities in Indonesia in 2022 using the choropleth map
   b. Calculate the Pearson correlation coefficient between the response variable and each explanatory variable to identify their potential relationship.

2. Determine the spatial weighting matrix to identify the relationship between neighboring regions. The weight matrix used in this study is based on distance, namely the k-nearest neighbor. Suppose d_{ij} is the distance between the center of the location-i and location-j with $i \neq j$ which is calculated based on the following Euclidean distance.

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \quad (1)$$

where u_i is the latitude coordinate at the i-th location, u_j is the latitude coordinate at the j-th location, v_i is the longitude coordinate at the i-th location and v_j is the longitude coordinate at the j-th location (Djuraidah, 2020). The distance d_{ij} is then ordered $d_{ij(1)} \leq d_{ij(2)} \leq \cdots \leq d_{ij(n-1)}$. K-nearest neighbor of location- i is $N_k(i) = \{j(1), j(2), \dots, j(k)\}$ with the value of $k = 1, \dots, n-1$. For the location of the nearest neighbor of the location-i the value is given as follows (Djuraidah, 2020):

$$w_{ij} = \begin{cases} 1, & j \in N_k(i) \\ 0 & other \end{cases} \quad (2)$$

3. Checking the multicollinearity assumption for all explanatory variables using the Variance Inflation Factor (VIF) as follows:

$$VIF = \frac{1}{1 - R_i^2}$$

where R_i^2 is the coefficient of determination of the i-th explanatory variable. A VIF value greater than five indicates multicollinearity in the explanatory variables (Akinwande et al., 2015).
4. Identify the effect of spatial dependency using Moran's I statistics (Djuraidah, 2020). Moran's I is formulated as follows:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j \neq i=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (3)$$

where x_i is the observed value of location-i, $(i = 1, \dots, n)$, x_j is the observed value of location-j, $(j = 1, \dots, n)$, $\bar{x}$ is the average observed value of all locations and w_{ij} is the nearest neighbor weight element between location-i and location-j.
5. Identifying the effects of spatial heterogeneity with the Breusch-Pagan (BP) test (Kholifia et al., 2021). The BP test statistic is as follows:

$$BP = \frac{1}{2} \boldsymbol{f}^T \boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{f} \quad (4)$$

where

$$f_i = \left(\frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2} - 1 \right) \quad (5)$$

$\hat{\varepsilon}_i^2$ is the residual for the i-th location from estimating the regression parameters by ordinary least square, $\hat{\sigma}^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$, and $\boldsymbol{X}$ is a matrix of explanatory variables of size

$n \times (p+1)$. Reject H_0 if the value of BP is greater than $\chi^2_{(p-1)}$ where p is the number of regression parameters.

6. Estimating the parameters of the SCR model. According to Sugasawa & Murakami (2021), the steps are as follows:
   a. Determine the number of clusters to be evaluated $G = (3, 4, 5, \ldots 15)$. The number of cluster ranges was used because it already reflects moderate clusters (Sugasawa & Murakami, 2021).
   b. Determine the initial values of the model parameters $\theta_{(0)}$ and the initial cluster membership variables $g_{(0)}$.
   c. Update the current parameter values $\theta_{(k)}$ and $g_{(k)}$, as follows:
      1) Update the cluster parameter θ_g for each cluster $g = 1, 2, 3, \ldots, G$ using Equation (6) to obtain the new θ_g coefficient value:

$$\theta_g^{(k+1)} = \arg\max_{\theta_g} \sum_{i=1}^{n} I\left(g_i^{(k)} = g\right) \log f\left(y_i | x_i; \theta_g\right) \tag{6}$$

      2) Update the i-th cluster member g_i for each cluster $g = 1, 2, 3, \ldots, G$. For each location -i determine the new cluster membership g_i based on equation (7):

$$g_i^{(k+1)} = \arg\max_{g \in \{1,\ldots,G\}} \left\{ \log f(y_i | x_i; \theta_g^{(k+1)} + \phi \sum_{j=1; j\neq 1}^{n} w_{ij} I\left(g = g_j^{(k)}\right) \right\} \tag{7}$$

      Observant districts/cities initially in a cluster are then updated to enter a new cluster. The criteria for updating the cluster members is based on the objective function value of the largest district/city observation $Q(\theta, g) = \sum_{i=1}^{n} logf\left(y_i | x_i; \theta_{g_i}\right) + \emptyset \sum_{i<j} w_{ij} I\left(g_i = g_j\right)$ is in the g-th cluster.

   d. Repeat step (c) until convergence is achieved. The iteration process stops when the difference between the current and updated values falls below the tolerance threshold of $\varepsilon = 10^{-6}$. The convergence formula is expressed as in Equation (8):

$$\Delta^{(k+1)} = \frac{|Q_k(\theta, g) - Q_{k+1}(\theta, g)|}{|Q_{k+1}(\theta, g)|} \tag{8}$$

   where $\Delta^{(k+1)}$ is the difference between the value of the $(k+1)$ iteration and the k iteration, $Q_k(\theta, g)$ is the k iteration penalty likelihood objective function value, dan $Q_{k+1}(\theta, g)$ is the $(k+1)$ iteration penalty likelihood objective function value.

7. Determining the optimum number of clusters G uses the information criterion as in Equation (9).

$$IC(G) = -2\sum_{i=1}^{n} \log f\left(y_i \middle| x_i; \hat{\theta}_{\hat{g}_i}\right) + c_n \dim(\theta) \tag{9}$$

   where $c_n = \log n$ which refers to the BIC-type criterion as a constant that depends on the sample size n and $\dim(\theta)$ denotes the dimension θ which depends on G. The selection of the value of G corresponds to $\hat{G} = \arg\min_{G\in\{G_1,\dots,G_L\}} IC(G)$ where $G_1, \dots, G_L$ are the candidates of G.
8. Interpret the model based on the optimum number of clusters, which was conducted by testing the effect of each explanatory variable on the response variable in each spatial cluster using the t-test. This test aims to identify variables that significantly affect stunting in each cluster. Furthermore, the interpretation is strengthened by examining the characteristics of each cluster based on the mean values of the explanatory variables.

C. RESULT AND DISCUSSION

1. Data Exploration

The thematic map in Figure 1 illustrates the distribution of stunting percentages across districts/cities in Indonesia, revealing a clear spatial pattern. The highest prevalence of stunting is predominantly clustered in the eastern regions of Indonesia, particularly in Papua, West Papua, and East Nusa Tenggara. Darker colors on the map indicate a high percentage of stunting, while lighter colors indicate a low percentage of stunting. In contrast, areas in western Indonesia, such as Sumatra, Java, and Bali, mostly have a lower percentage of stunting. This pattern reflects significant spatial disparities in stunting rates, which socioeconomic differences, access to health facilities, and quality of infrastructure in each region may influence.

The percentage of stunting (Y) in districts/cities in Indonesia ranged from 4,80 percent to 54,40 percent, with an average of 24,67 percent and a median of 24,30 percent. This shows that most districts/cities have stunting levels close to the national average of 21,5 percent (Ministry of Health, 2024), although there are areas with very high percentages above 50 percent, such as Asmat, Yahukimo, Arfak Mountains, Nduga, and Tolikara. The largest percentage of stunting in Indonesia is in Asmat District at 54,40 percent, and the lowest is Surabaya City with 4,80 percent, as shown in Figure 1.

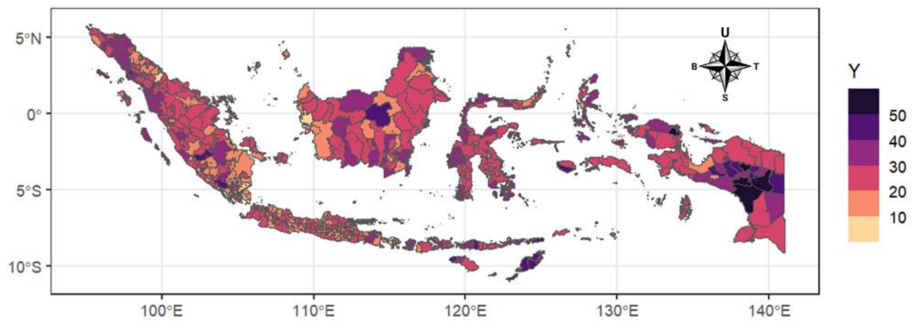

Figure 1. Thematic map of the percentage of stunting in Indonesia (non-scale map)

The correlation plot in Figure 2 shows the relationship between the response and explanatory variables. The poverty rate (X_1) has a positive correlation with stunting (Y) of $r = 0{,}52$, indicating that areas with higher poverty rates tend to have a greater prevalence of stunting. In contrast, the variables of access to sanitation (X_4), access to clean water (X_5), and average years of schooling (X_7) have a negative correlation with the stunting of $r = -0{,}47, r = -0{,}40$, ($r = -0{,}39$) respectively, indicating that an increase in sanitation access, clean water access, and average years of schooling can decrease stunting. Meanwhile, the variable number of public health centers (X_2), nutrition of pregnant women (X_3), and GRDP (X_6) have a very weak correlation with Y, as shown in Figure 2.

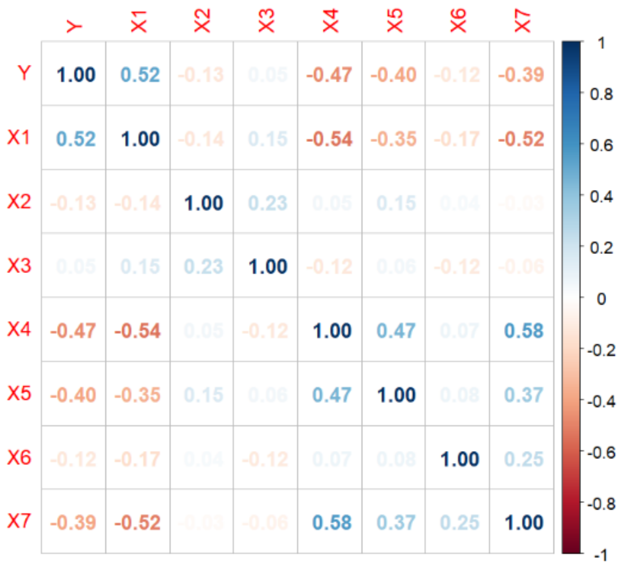

Figure 2. Correlation Plot between Response Variables and Explanatory Variables

Variation between regions may affect the low correlation because the factors contributing to stunting do not have a uniform influence in all regions. Each region has different social, economic, and infrastructural characteristics, which causes the relationship between explanatory variables and stunting to vary. A globally calculated correlation would summarize the entire relationship in one average value without considering that the relationship may be more substantial in some regions. In contrast, it could be very weak or insignificant in other regions. For example, the poverty rate (X_1) has a positive correlation with stunting overall, but the effect may be more substantial in areas with limited health services and poor nutrition, such as rural or remote areas. Conversely, in urban areas with better health facilities, the impact of poverty on stunting may be more minor due to more effective social intervention programs and food assistance. Thus, the low correlation between variables does not mean that these factors do not affect stunting but indicates that the relationship pattern differs between regions.

2. Identification of Multicollinearity

The multicollinearity is identified to verify that there is no strong linear correlation among the explanatory variables in the regression model. A VIF value greater than five indicates multicollinearity in the explanatory variables.

Table 2. VIF Value of Explanatory Variables

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7
VIF Value	1,66	1,13	1,14	1,91	1,36	1,11	1,75

The multicollinearity test results in Table 2 using the Variance Inflation Factor (VIF) show that all explanatory variables (X_1 to X_7) have VIF values less than 5. This value indicates no significant multicollinearity problem among the explanatory variables in the regression model.

3. Spatial Weight Matrix

Before identifying spatial effects, a spatial weight matrix is formed to explain the spatial relationship between observed locations. The spatial weight matrix used in this study uses the k-nearest neighbor (KNN) method based on distance information. The KNN weighting matrix uses parameter k or the number of different neighbors. The optimal parameter selection is determined by evaluating the Moran Index test statistics and its corresponding significance level. Table 3 presents the evaluation results for the KNN weight matrix parameters. Based on Table 3, it is observed that the parameter k with the highest Moran Index value and a statistically significant p-value is at $k = 3$. Therefore, the analysis is performed using the KNN spatial weighting matrix with $k = 3$.

Table 3. Summary of KNN Spatial Weight Matrix Parameter k Evaluation Results

Parameter k	Moran Index	p-value
3	0,38	< 2,2e-16
5	0,34	< 2,2e-16
7	0,32	< 2,2e-16
9	0,29	< 2,2e-16

2. Identification of Spatial Effects

When modeling spatial area data, attention must be paid to spatial effects between locations, namely spatial heterogeneity and spatial dependency. Spatial heterogeneity indicates a different error variance at each observed location. Meanwhile, spatial dependency shows that one observation location and another have a strong relationship. The effect of heterogeneity is evaluated using the Breusch-Pagan test. The test results in Table 4 show a p-value smaller than $\alpha = 0{,}05$ so reject H_0 which means there is a spatial heterogeneity effect. Testing for global spatial dependence is done using the Moran Index with a 3-nearest neighbor weight matrix. The Moran Index test results for the response variables show a p-value less than $\alpha = 0.05$ and a positive Moran Index value, indicating spatial dependence among the observed locations. The test results show a spatial heterogeneity effect and a spatial dependency effect between observed locations so that spatial regression can be used to measure stunting data in Indonesia.

Table 4. Spatial Effect Test Results

Test	Test Statistics	p-value	Conclusion
Breusch-Pagan	38,06	2, 952e-06	There is spatial heterogeneity
Indeks Moran	0,38	< 2, 2e-16	There is a spatial dependency

3. Spatial Clustering Regression

The analysis of the SCR model begins by identifying the number of potential clusters (G) that can be obtained from the data. The analysis considered the number of clusters ranging from $G = 3, 4, 5, \dots, 15$ to determine the optimal number that minimizes the information criterion value. The results showed that the lowest information criterion value, 3507.80, was obtained when G = 4, indicating that four clusters effectively classified the Indonesian regions based on the regression parameters. The selection of the number of clusters is an important step in the analysis because if there are too few clusters, the model is not flexible enough to win the local diversity in the data, while if there are too many clusters, the model becomes complex, as shown in Figure 3.

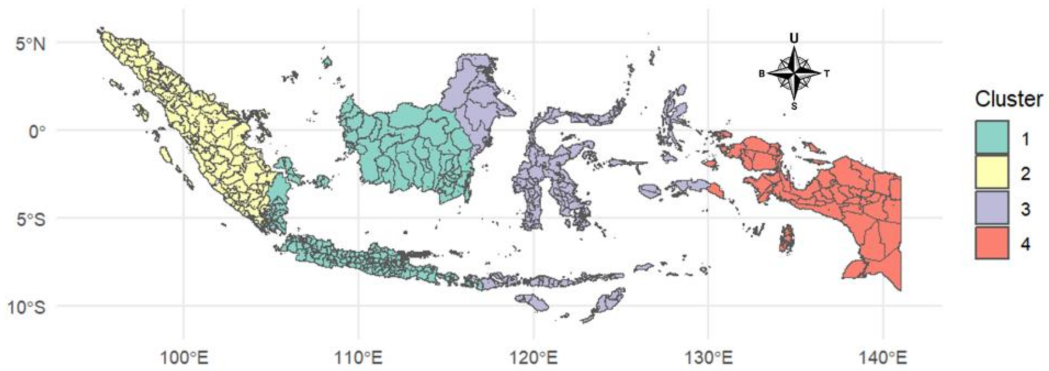

Figure 3. Spatial Cluster Map (non-scale map)

Different colors on the map (Figure 3) reflect different clusters. Cluster 1 covers Java, Bali, West Kalimantan, Central Kalimantan, South Kalimantan, South Sumatra, Lampung and Bangka Belitung Islands. Cluster 2 dominates most of Sumatra Island. Cluster 3 covers the islands of Sulawesi, North Kalimantan, East Kalimantan, West Nusa Tenggara, East Nusa Tenggara, and North Maluku. Meanwhile, cluster 4 covers Papua, West Papua, and Maluku. The spatial map shows homogeneous geographical patterns within each cluster, enabling the implementation of region-based policies according to the dominant factors affecting stunting.

4. Identification of Factors Affecting Stunting from Each Cluster

The factors affecting the percentage of stunting were identified based on the previously formed clusters. Each cluster has a different regression coefficient, allowing for variation in significant variables. Identification of significant variables was conducted using the t-test with the t value calculated using the $t = \hat{\beta}_{ig}/se_{\hat{\beta}_{ig}}$ where $i = 1, 2, \dots 7$ and $g = 1, 2, 3, 4$. $se_{\hat{\beta}_{ig}}$ is the standard deviation of the regression model parameters for each g-th cluster. Significant variables were determined by comparing the p-value of each cluster with the 5% significance level. The estimated regression parameters β in Table 5 and the map of estimated regression parameters in Figure 4 for each cluster show significant differences between regions.

Table 5. Significant Effect Regression Coefficients of Each Cluster

Variable	**Regression Coefficient of Each Cluster**							
	$\boldsymbol{g=1}$		$\boldsymbol{g=2}$		$\boldsymbol{g=3}$		$\boldsymbol{g=4}$	
Intercept	48,67	***	28,76	***	44,67	***	45,23	***
X_1	0,02		0,40	**	0,36	***	0,09	
X_2	-0,01		-0,05		0,14	*	0,21	
X_3	$2{,}38 \times 10^{-4}$	*	$5{,}40 \times 10^{-4}$	*	$1{,}03 \times 10^{-4}$		$2{,}12 \times 10^{-5}$	
X_4	-0,10	*	-0,06		-0,07		-0,16	*
X_5	-0,14	***	-0,02		-0,08	*	-0,06	
X_6	-0,58		-0,30		-0,15		0,37	
X_7	-1,09	**	-0,24		-1,28	**	-0,76	

* significant at $\alpha = 0{,}05$, ** significant at $\alpha = 0{,}01$, ***significant at $\alpha = 0{,}001$

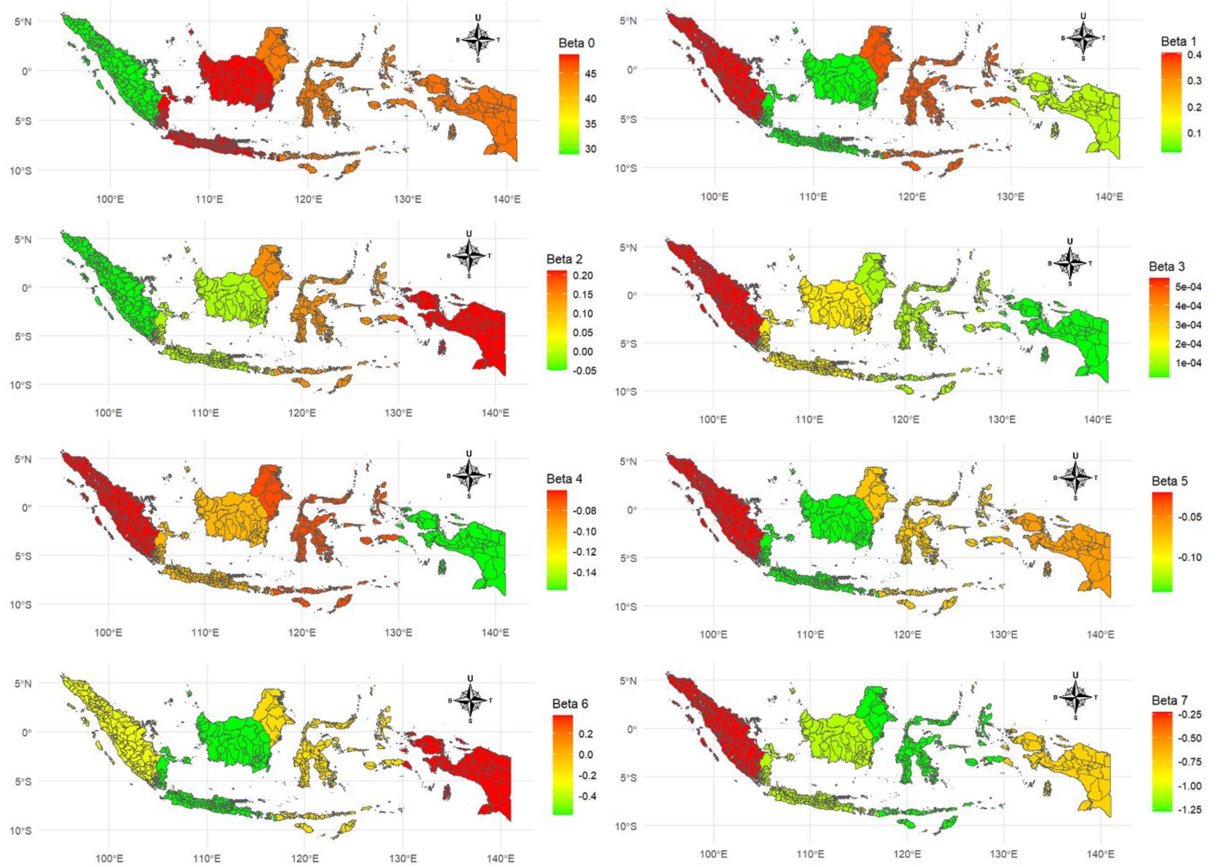

Figure 4. Estimated Regression Coefficients of Explanatory Variables (non-scale map)

Based on Table 5, Cluster 1 (Java, Bali, West Kalimantan, Central Kalimantan, South Kalimantan, South Sumatra, Lampung, and Bangka Belitung Islands) has the lowest percentage of stunting, with the main contributing factors being average years of schooling (X_7), access to clean water (X_5), and access to sanitation (X_4), which contribute to reducing stunting. However, nutrition of pregnant women (X_3) has a positive effect, indicating a less balanced consumption pattern. Cluster 2 (most of Sumatra Island) shows that poverty rate (X_1) has a positive effect on stunting, while improved nutrition of pregnant women (X_3) has not been able to reduce stunting effectively. Cluster 3 (Sulawesi Island, North Kalimantan, East Kalimantan, West Nusa Tenggara, East Nusa Tenggara, and North Maluku) has poverty rate (X_1) and the number of public health centers (X_2) positively affecting stunting, suggesting that although the number of public health centers is increasing, gaps in the quality of health services may still be a challenge. In contrast, access to clean water (X_5) and average years of schooling (X_7) contributed to decreased stunting. Cluster 4 (Papua, West Papua, and Maluku) shows that access to sanitation (X_4) has a significant negative effect on stunting, suggesting that improving sanitation infrastructure may help reduce stunting, especially in areas with limited basic infrastructure.

The characteristics of each cluster can also be identified based on the mean of each variable as presented in Table 6 and the boxplot in Figure 5. Based on Table 6, the characteristics of Cluster 1 show relatively better conditions than the other clusters with the lowest percentage of stunting (20.31%). The areas in this cluster have the lowest poverty rate (8.61%) and very good access to sanitation and clean water (81.50% and 88.91%). In addition, the number of public health centers is relatively high (24.77), and the average number of years of schooling is at a medium level (8 years). The GRDP in this region is also relatively high, although not the highest. These conditions indicate that areas in Cluster 1 have better access to health facilities, adequate basic infrastructure, and more stable socio-economic factors, resulting in lower stunting rates, as shown in Table 6.

Table 6. Characteristics of Clusters Based on Average Variables

Variable		Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Average	$g = 1$	**20.31**	**8.61**	**24.77**	8152.81	**81.50**	**88.91**	0.48	8.33
	$g = 2$	24.49	10.54	14.87	2805.95	76.04	83.64	**1.81**	**9.08**
	$g = 3$	28.38	12.22	15.86	4576.74	80.62	85.13	1.26	8.66
	$g = 4$	33.16	26.66	14.11	**13245.24**	52.68	73.90	0.50	7.10

Cluster 2 has a medium percentage of stunting (24.49%), with a higher poverty rate (10.54%) than Cluster 1. Access to sanitation and clean water is also quite good (76.04% and 83.64%), although lower than in Cluster 1. However, fewer public health centers (14.87) may affect people's access to health services. The GRDP in this region is the highest (1.81 billion rupiah), indicating better economic conditions. The average years of schooling is also higher (9 years), suggesting a better level of education. Cluster 2 has a fairly good economic condition but still faces challenges in accessing health services, improving the quality of health services and nutrition for pregnant women could be a priority intervention. The following Boxplot of Response Variables from Each Spatial Cluster, as shown in Figure 5.

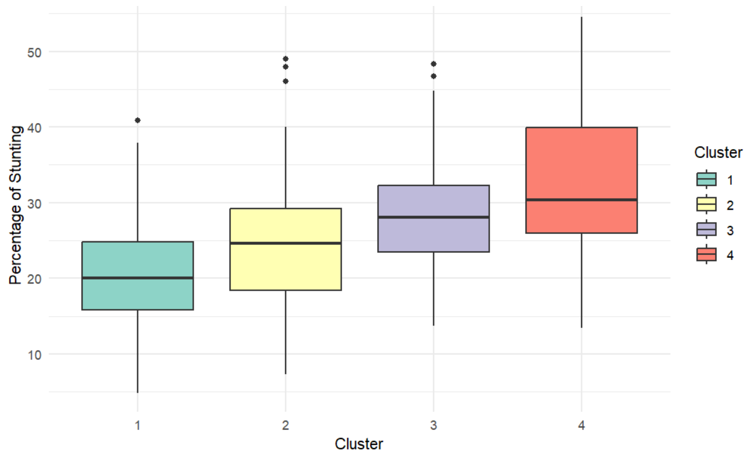

Figure 5. Boxplot of Response Variables from Each Spatial Cluster

Cluster 3 has a higher percentage of stunting (28.38%) compared to Clusters 1 and 2. The poverty rate in this area is higher (12.22%), and the number of health centers is limited (15.86), which can impact the lack of access to health services. Access to sanitation and clean water is still quite good (80.62% and 85.13%), and the average years of schooling are quite high (9 years). However, the GRDP in this region is lower than in Cluster 2. The combination of vulnerable economic factors and limited health facilities means that the stunting rate in this region remains high, despite good education and sanitation factors.

Cluster 4 has the highest percentage of stunting (33.16%) and the highest poverty rate (26.66%). It also has the most limited access to sanitation and clean water (52.68% and 73.90%). The low number of public health centers (14.11) further worsens access to health services for the community. The average years of schooling are also the lowest compared to other clusters (7 years), indicating that education remains a significant challenge. Although the nutrition of pregnant women is high (13,245 calories), the quality of nutrition consumed may not be balanced, which still results in high stunting rates. With difficult economic conditions, lack of access to sanitation, and limited health services, areas in Cluster 4 require more intensive interventions, particularly in improving basic infrastructure, health services, and nutrition education.

Cluster 1 is the best-conditioned region in terms of socio-economics, sanitation infrastructure, and health, resulting in a low percentage of stunting. In contrast, Cluster 4 is the most vulnerable, with significant economic, sanitation, and education challenges. Clusters 2 and 3 are at a moderate level, with several factors still needing improvement, especially in the quality of health and education services. Therefore, area-based policies need to be tailored to the characteristics of each cluster, with the primary focus on improving sanitation, equitable distribution of health facilities, and improving the quality of education and nutrition of pregnant women in areas with high stunting rates.

The results of this study align with the research of Eryando et al. (2022), which showed spatial autocorrelation and spatial inequality in the prevalence of stunting between districts/cities in Indonesia. The study used the Spatial Autoregressive (SAR) model and identified factors that significantly affect stunting, such as poverty and access to sanitation. Furthermore, Alam et al. (2021) emphasized the importance of inter-regional differences in handling stunting by considering spatial influences using the Geographically Weighted Logistic Regression (GWLR) method to prioritize area-based interventions. The results also show that poverty, access to sanitation, and access to clean water significantly affect stunting in Indonesia. Thus, the findings of this study not only strengthen evidence from previous studies on the importance of socio-economic factors and infrastructure in stunting prevention, but also offer interpretive advantages through SCR.

D. CONCLUSION AND SUGGESTIONS

The SCR model proved effective in identifying clustered spatial patterns in stunting data in Indonesia by dividing the region into four clusters based on uniform regression parameters. Compared to other spatial methods such as GWR, SCR has the advantage of producing stable and interpretative regression parameters, and facilitating policy formulation through regional clustering maps. SCR offers a more strategic approach to identifying dominant local factors and setting the direction of more effective and targeted stunting intervention policies. Future research can evaluate the performance of SCR by using other spatial weight matrices to capture clustered spatial patterns in the data. In addition to stunting, SCR can also potentially be applied to other spatial phenomena, such as poverty, food security, access to health services, and education, to support more effective spatial-based policies.

The analysis shows that each cluster shows different local factors affecting stunting. Cluster 1 has the lowest average percentage of stunting (20.31%), with significant decreasing factors being access to clean water, access to sanitation, and average years of schooling. In contrast, the nutrition of pregnant women has a positive effect, suggesting the possibility of unbalanced calorie consumption. Cluster 2 has an average stunting percentage of 24.49%, with the poverty rate as a significant positive factor, and the nutrition of pregnant women also has a positive effect. Cluster 3 recorded an average stunting rate of 28.38%, with poverty and the number of health centers having a positive impact, while clean water and average years of schooling played a role in reducing stunting. Cluster 4 has the highest average stunting percentage (33.16%) and the highest poverty rate (26.66%), with access to sanitation as the only significant factor in reducing stunting. These findings emphasize the importance of a region-based approach in formulating stunting prevention policies. Interventions should be tailored to the characteristics of each cluster, such as poverty alleviation, improving basic infrastructure, health services, and nutrition education in Cluster 4, improving the quality of health services in Cluster 2, and strengthening education and nutrition programs in Cluster 3. While Cluster 1 has the lowest stunting rate, attention needs to be focused on nutrition education for pregnant women.

ACKNOWLEDGEMENT

We sincerely express our gratitude to the Central Statistics Agency for granting access to the data utilized in this research. We deeply acknowledge their commitment to making statistical data accessible for scientific and policy-related studies.

REFERENCES

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open Journal of Statistics*, *05*(07), 754–767. https://doi.org/10.4236/ojs.2015.57075

Alam, F. K., Widyaningsih, Y., & Nurrohmah, S. (2021). Geographically weighted logistic regression modeling on stunting cases in Indonesia. *Journal of Physics: Conference Series*, *1722*(1), 1-8. https://doi.org/10.1088/1742-6596/1722/1/012085

Ali, A. (2021). Current Status of Malnutrition and Stunting in Pakistani Children: What Needs to Be Done? *Journal of the American College of Nutrition*, *40*(2), 180–192. https://doi.org/10.1080/07315724.2020.1750504

Anggraini, Y., & Romadona, N. F. (2020). Review of Stunting in Indonesia. *Proceedings of the International Conference on Early Childhood Education and Parenting 2019*, *454*, 281–284. https://doi.org/10.2991/assehr.k.200808.055

Azis, A. A., & Aswi, A. (2023). Spatial Clustering of Stunting Cases in Indonesia: a Bayesian Approach. *Communications in Mathematical Biology and Neuroscience*, *2023*(28), 1–11. https://doi.org/10.28919/cmbn/7898

Beal, T., Tumilowicz, A., Sutrisna, A., Izwardy, D., & Neufeld, L. M. (2018). A review of child stunting determinants in Indonesia. *Maternal & Child Nutrition*, *14*(4), 1-10. https://doi.org/10.1111/mcn.12617

Cameron, L., Chase, C., Haque, S., Joseph, G., Pinto, R., & Wang, Q. (2021). Childhood stunting and cognitive effects of water and sanitation in Indonesia. *Economics and Human Biology*, *40*(100944), 1-38. https://doi.org/10.1016/j.ehb.2020.100944

Djara, V. A. D., Andriyana, Y., & Noviyanti, L. (2022). Modelling the Prevalence of Stunting Toddlers Using Spatial Autoregressive With Instrument Variable and S-Estimator. *Communications in Mathematical Biology and Neuroscience*, *2022*(29), 1–23. https://doi.org/10.28919/cmbn/7234

Djuraidah, A. (2020). *Monograph Penerapan dan Pengembangan Regresi Spasial dengan Studi Kasus pada Kesehatan, Sosial, dan Ekonomi*. PT Penerbit IPB Press.

Eryando, T., Sipahutar, T., Budhiharsana, M. P., Siregar, K. N., Nur Aidi, M., Minarto, Utari, D. M., Rahmaniati, M., & Hendarwan, H. (2022). Spatial analysis of stunting determinants in 514 Indonesian districts/cities: Implications for intervention and setting of priority. *Geospatial Health*, *17*(1055), 1-7. https://doi.org/10.4081/gh.2022.1055

Fitri, D. J., Djuraidah, A., & Wijayanto, H. (2024). Bayesian Conditional Negative Binomial Autoregressive Model: A Case Study of Stunting on Java Island in 2021. *Communications in Mathematical Biology and Neuroscience*, *2024*(18), 1–17. https://doi.org/10.28919/cmbn/8281

Kholifia, N., Rahardjo, S., Muksar, M., Atikah, N., & Afifah, D. L. (2021). Spatial analysis of factors influencing Gross Regional Domestic Product (GRDP) in East Java: A spatial durbin error model analysis. *Journal of Physics: Conference Series*, *1918*(4), 1–7. https://doi.org/10.1088/1742-6596/1918/4/042044

Lee, J., Gangnon, R. E., & Zhu, J. (2017). Cluster detection of spatial regression coefficients. *Statistics in Medicine*, *36*(7), 1118–1133. https://doi.org/10.1002/sim.7172

Li, F., & Sang, H. (2019). Spatial Homogeneity Pursuit of Regression Coefficients for Large Datasets. *Journal of the American Statistical Association*, *114*(527), 1050–1062. https://doi.org/10.1080/01621459.2018.1529595

Louzada, F., Do Nascimento, D. C., & Egbon, O. A. (2021). Spatial statistical models: An overview under the bayesian approach. *Axioms*, *10*(4), 1–38. https://doi.org/10.3390/axioms10040307

Ministry of Health. (2022). *Buku Saku Hasil Survey Status Gizi Indonesia (SSGI) Tahun 2022*. Kementerian Kesehatan RI. https://repository.badankebijakan.kemkes.go.id/id/eprint/4855

Ministry of Health. (2024). *Profil Kesehatan Indonesia 2023*. Kementerian Kesehatan RI. https://kemkes.go.id/id/profil-kesehatan-indonesia-2023

Muche, A., Melaku, M. S., Amsalu, E. T., & Adane, M. (2021). Using geographically weighted regression analysis to cluster under-nutrition and its predictors among under-five children in Ethiopia: Evidence from demographic and health survey. *PLOS ONE*, *16*(5), 1–30. https://doi.org/10.1371/journal.pone.0248156

Muhamad, Z., Mahmudiono, T., Abihail, C. T., Sahila, N., Wangi, M. P., Suyanto, B., & Binti Abdullah, N. A. (2023). Preliminary Study: The Effectiveness of Nutrition Education Intervention Targeting Short-Statured Pregnant Women to Prevent Gestational Stunting. *Nutrients*, *15*(19), 1–11. https://doi.org/10.3390/nu15194305

Nicholson, D., Vanli, O. A., Jung, S., & Ozguven, E. E. (2019). A spatial regression and clustering method for developing place-specific social vulnerability indices using census and social media data. *International Journal of Disaster Risk Reduction*, *38*(101224), 1-14. https://doi.org/10.1016/j.ijdrr.2019.101224

Nursyamsiyah, Sobrie, Y., & Sakti, B. (2021). Faktor-faktor yang berhubungan dengan kejadian stunting pada anak usia 24-59 bulan. *Jurnal Ilmu Keperawatan Jiwa*, *4*(6), 611–622. https://journal.ppnijateng.org/index.php/jikj

Rizal, M. F., & van Doorslaer, E. (2019). Explaining the fall of socioeconomic inequality in childhood stunting in Indonesia. *SSM - Population Health*, *9*(100469), 1–10. https://doi.org/10.1016/j.ssmph.2019.100469

Sugasawa, S., & Murakami, D. (2021). Spatially clustered regression. *Spatial Statistics*, *44*(100525), 1-16. https://doi.org/10.1201/9781439891117

Sukmawati, S., Djuraidah, A., & Wigena, A. H. (2021). Spatial Clustered Regression Analysis of 2017 Getis Score Indonesian Malaria Prevalence Data. *Journal of Physics: Conference Series*, *1863*(1), 1-8. https://doi.org/10.1088/1742-6596/1863/1/012043

Vaivada, T., Akseer, N., Akseer, S., Somaskandan, A., Stefopulos, M., & Bhutta, Z. A. (2020). Stunting in childhood: An overview of global burden, trends, determinants, and drivers of decline. *American Journal of Clinical Nutrition*, *112*(2), 777S-791S. https://doi.org/10.1093/ajcn/nqaa159

WHO. (2015). *Stunting in a nutshell*. World Health Organization. https://www.who.int/news/item/19-11-2015-stunting-in-a-nutshell

Zhong, X., Wang, P., & Zhang, H. (2022). Spatial homogeneity pursuit of regression coefficients for hand, foot an mouth disease in Xinjiang Uygur Autonomous Region in 2018. *Scientific Reports*, *12*(1), 1–10. https://doi.org/10.1038/s41598-022-26003-6