SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 9603-9611
Open Access. Online & Print

DOI: 10.5281/zenodo.19981447

DEVELOPMENT OF NON-LINEAR SMALL AREA ESTIMATION USING THE MAGPRS APPROACH

Bambang Widjanarko Otok[1*], Shofi Andari[2], Rahmannuaji Satuhu[2], Hendra H Dukalang[2], Setia Ningsih[2], Ajriansyah[2], Rizky Jauhari Angwarmase[2], Septia D. P. Yasmirullah[3], Anang Kurnia[4]

[1,2]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya
[3]Data Science Technology Study Program, Airlangga University, Surabaya
[4]Department of Statistics, Institut Pertanian Bogor University, Bogor

Received: 15/03/2026
Accepted: 18/04/2026

Corresponding Author: Bambang Widjanarko Otok
(bambang_wo@its.ac.id)

ABSTRACT

Modeling of the number of poverties was conducted using a spatial approach via Multivariate Adaptive Generalized Poisson Regression Spline (MAGPRS) combined with Small Area Estimation (SAE), utilizing socioeconomic variables such as per capita GRDP, open unemployment rate, average years of schooling, and percentage of expenditure on food. The best model was obtained with parameters BF = 16, MI= 3, and MO= 3, with a GCV value of 1750.7859 and a coefficient of determination of 0.5891. Average years of schooling was identified as the most dominant variable, followed by the open unemployment rate, the percentage of expenditure on food, and per capita GRDP, with all variables significant both simultaneously and partially. Compared with a linear SAE baseline, the proposed MAGPRS-SAE model achieves higher Relative Efficiency (1.613 vs. 1.527), demonstrating superior variance reduction across small areas despite a marginal trade-off in point-prediction accuracy. These findings confirm that MAGPRS-SAE provides a methodologically rigorous and spatially adaptive tool for poverty mapping and offer actionable guidance for geographically targeted poverty alleviation policies aligned with Indonesia's Sustainable Development Goal commitment.

KEYWORDS: Poverty Population, Spatial Modeling, MAGPRS, SAE, Spatial Effects.

1. INTRODUCTION

Java is the center of economic and governmental activity in Indonesia, yet it also faces complex poverty issues. According to data from the Central Statistics Agency (BPS), Java is home to approximately 56% of Indonesia's total population (Badan Pusat Statistik, 2021), making it the most densely populated region in the country. This centrality makes Java a magnet for urbanization, creating a concentration of economic activity, but also making it vulnerable to inequality. Although the poverty rate in Java (around 10.34% in 2021) may be lower than in some provinces outside Java, the absolute number of poverty people is very large. Several provinces in Java, such as West Java and Central Java, are often recorded as having the highest number of poverty people at the national level, indicating that the issue of poverty in Java is one of sheer magnitude rather than merely a regional percentage.

Poverty in Java remains a complex and spatially clustered phenomenon, occurring not only in remote areas but also in agricultural and peri-urban regions. This indicates that poverty is influenced by diverse and location-specific factors, making spatial modeling essential for a more accurate understanding of its distribution. Spatial approaches enable the identification of interregional dependencies, where poverty in one area may be influenced by neighboring regions, thereby providing a stronger basis for designing targeted and effective development policies (Rasam & Shariff, 2019; Yanti & Rahardiantoro, 2021).

However, spatial data analysis presents methodological challenges, particularly due to violations of classical assumptions such as error independence and homogeneity. Spatial autocorrelation and heterogeneity frequently arise, leading to biased parameter estimates and unreliable inference if not properly addressed (Billé et al., 2016; Géniaux & Martinetti, 2017). In addition, poverty data derived from large-scale surveys often lack precision at smaller administrative levels due to limited sample sizes, necessitating the use of Small Area Estimation (SAE) techniques to improve estimation accuracy by integrating survey and auxiliary data (World Bank, 2005; Yuniarti, 2022).

To address these limitations, this study proposes a hybrid modeling framework that integrates Spatial Multivariate Adaptive Generalized Poisson Regression Splines (Spatial MAGPRS) with SAE. This approach combines the flexibility of nonparametric regression in capturing complex and nonlinear relationships with the ability to account for spatial dependence and small-area data limitations (Ramli et al., 2018). The proposed model is expected to provide more accurate and representative estimates of poverty distribution, while also offering valuable insights for policymakers in formulating targeted poverty alleviation strategies in Java (Yasmirullah et al., 2023).

By identifying groups of units with distinct characteristics in terms of poverty reduction, this model is expected to uncover previously hidden risk factors and provide a stronger foundation for the formulation of poverty reduction policies. This represents an important contribution to the advancement of scientific knowledge and the implementation of more effective poverty reduction strategies. Furthermore, the findings of this study are not only relevant for supporting national policies on the eradication of extreme poverty but also contribute to the achievement of the Sustainable Development Goals (SDGs), particularly SDG 1 (No Poverty), SDG 3 (Good Health and Well-being), and SDG 8 (Decent Work and Economic Growth). Therefore, this research not only holds academic value but also has significant practical implications for addressing one of the challenges of poverty in Indonesia, particularly on the island of Java.

2. THEORETICAL FOUNDATION AND HYPOTHESES DEVELOPMENT

2.1. *Multivariate Adaptive Regression Splines (MARS)*

MARS is a flexible, data-driven regression method well suited to high-dimensional problems, particularly when the number of predictor variables lies between 3 and 20 (Friedman, 1991). Unlike parametric methods that assume a fixed functional form (e.g., linear or quadratic), MARS lets the data determine the shape of the relationship between each predictor and the outcome. In plain terms, it automatically detects "breakpoints" (called knots) where the effect of a variable changes direction or slope. The general nonparametric form of the MARS model is given in Equation (1)

$$y_i = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})] + \varepsilon_i \quad (1)$$

Equation (1) can be rewritten in a more compact operational form as Equation (2), where the predicted outcome is expressed as a weighted sum of basis functions each basis function representing a piecewise linear segment of the predictor's effect.

$$\hat{f}(x) = a_0 + \sum_{i=1}^{M} f_i(x_i) + \sum_{i\le j=1}^{M} f_{ij}(x_i, x_j) + \sum_{i\le j\le k=1}^{M} f_{ijk}(x_i, x_j, x_k) + \cdots \quad (2)$$

For ease of interpretation, Equation (2) is condensed into the shorthand form in Equation (3), where each term captures the contribution of one or more predictor variables within a specific range of values.

$$\hat{f}(x) = a_0 + a_1 BF_1 + a_2 BF_2 + a_3 BF_3 + \ldots + a_M BF_M \quad (3)$$

Model selection in MARS is governed by the Generalized Cross Validation (GCV) criterion: the model with the lowest GCV value is preferred, as it strikes the best balance between fitting the data well and avoiding overfitting (Lasheras et al., 2018; Liu et al., 2019; López & Kholodilin, 2020). Intuitively, GCV penalises model complexity adding unnecessary basis functions increases the GCV even if it reduces the raw prediction error. The GCV formula is given in Equation (4).

$$GCV(M) = \frac{\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{f}_M(x_i)]^2}{[1 - \frac{C(M)}{n}]^2} \quad (4)$$

2.2. *Generalized Poisson Regression*

Standard Poisson regression assumes that the mean and variance of the count outcome are equal (equidispersion). In practice, count data such as the number of people in poverty often exhibit overdispersion (variance exceeding the mean) or underdispersion. The Generalized Poisson Regression (GPR) model was developed specifically to handle such violations (Melliana et al., 2013). It introduces an additional dispersion parameter θ which serves as a dispersion parameter (Wang & Famoye, 1997). The mathematical form of the Generalized Poisson model is expressed in the equation (5).

$$f(y_i;\mu,\theta) = \left(\frac{\mu}{1+\theta\mu}\right)^{y_i} \frac{(1+\theta y_i)^{(y_i-1)}}{y_i!} \exp\left(\frac{-\mu(1+\theta y_i)}{1+\theta\mu}\right) \quad (5)$$
$$y_i = 0,1,2,\ldots$$

where the value $\mu > 0$ and $-\infty < \theta < \infty$. Unlike standard Poisson regression where mean and variance are forced to be equal, the GPR allows them to differ controlled by the dispersion parameter. When the dispersion parameter equals zero, the GPR reduces to the standard Poisson model. The mean and variance expressions are given in Equation (6).

$$E(y_i) = \mu$$
$$Var(y_i) = \mu(1+\theta\mu)^2 \quad (6)$$

The full GPR model linking the expected count to predictor variables through a log link function is expressed as Equation (7). The log link ensures that predicted counts are always positive, which is essential when modelling quantities such as the number of people in poverty.

$$\mu(\mathbf{x}) = \exp(\mathbf{x}^T\boldsymbol{\beta}) \quad (7)$$

2.3. *Multivariate Adaptive Generalized Poisson Regression Splines (MAGPRS)*

MAGPRS integrates the two preceding frameworks: it applies MARS's adaptive spline structure to automatically detect nonlinear relationships, while using the Generalized Poisson distribution to properly model count outcomes with potential overdispersion. In essence, MAGPRS combines the pattern-detection power of MARS with the statistical appropriateness of GPR for count data. The general MAGPRS formulation is given in Equation (8)(Otok et al., 2019).

$$Y_i \sim GP(\mu_i, \theta)$$
$$\ln \mu_i = f(\mathbf{x}_i)$$
$$= a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)i} - t_{km})]_+ \quad (8)$$
$$= a_0 + \sum_{m=1}^{M} a_m B_{mi}(\mathbf{x}_i)$$
$$\mu = \exp(a_0 + \sum_{m=1}^{M} a_m B_{mi}(\mathbf{x}_i)) = \exp(\mathbf{B}\boldsymbol{\alpha})$$

1) MAGPRS Model Parameter Estimation

MAGPRS parameters are estimated using Weighted Least Squares (WLS). WLS is used rather than ordinary least squares because observations do not contribute equally to the fit regions with greater variance are downweighted to prevent them from unduly influencing the parameter estimates. Technically, the variance weights of the response variable are represented by a diagonal matrix $\mathbf{W}$. The $\mathbf{W}$ matrix is an $n \times n$ diagonal weighting matrix whose diagonal elements are $\frac{1}{w_i}$, where $\frac{1}{w_i} = \frac{1}{\mu_i(1+\theta\mu_i)^2}$ and n is the number of observations (Otok et al., 2019). The parameter estimates are then obtained by minimising the weighted sum of squared residuals, as shown in Equation (9) a standard optimisation step that ensures the model fits the data as closely as possible given the variance structure.

$$\psi = \boldsymbol{\varepsilon}^T \mathbf{W} \boldsymbol{\varepsilon}$$
$$\psi = \mathbf{y}^T \mathbf{W} \mathbf{y} - 2\exp(\boldsymbol{\alpha}^T \mathbf{B}^T) \mathbf{W} \mathbf{y} + \exp(\boldsymbol{\alpha}^T \mathbf{B}^T) \mathbf{W} \exp(\mathbf{B}\boldsymbol{\alpha}) \quad (9)$$

Minimising Equation (9) analytically by setting its first partial derivatives to zero yields the closed-form WLS estimator for the MAGPRS coefficients, given in Equation (10). This closed-form solution makes computation straightforward and guarantees a unique global minimum ψ ; $\boldsymbol{\alpha}$, which is given in Equation (10) below.

$$\hat{\boldsymbol{\alpha}}_{WLS} = (\mathbf{B})^{-1} \ln\left((\mathbf{B}^T \exp(\hat{\boldsymbol{\alpha}}^T \mathbf{B}^T) \mathbf{W})^{-1} \mathbf{B}^T \exp(\hat{\boldsymbol{\alpha}}^T \mathbf{B}^T) \mathbf{W} \mathbf{y} \right) \quad (10)$$

Substituting the estimated coefficients back into the model structure yields the fitted MAGPRS predictions, expressed as Equation (11). This equation is the operational model used to generate poverty estimates for each small area.

$$\hat{f}(x) = \exp(\mathbf{B}\hat{\boldsymbol{\alpha}}) = (\mathbf{B}^T \exp(\hat{\boldsymbol{\alpha}}^T \mathbf{B}^T) \mathbf{W})^{-1} \mathbf{B}^T \exp(\hat{\boldsymbol{\alpha}}^T \mathbf{B}^T) \mathbf{W} \mathbf{y} \quad (11)$$

2) Testing of MAGPRS Model Parameters

To confirm that the model as a whole is statistically meaningful, two levels of significance testing are applied. The simultaneous (global) test asks whether at least one predictor has a significant effect on poverty counts; it uses the Maximum Likelihood Ratio Test (MLRT), which compares the fitted model against a null model with no predictors. The partial test then examines each basis function individually to verify that every term retained in the model contributes significantly. The hypotheses for the simultaneous test are as follows.

$H_0 : \alpha_1 = \ldots = \alpha_M = 0$
H_1 : at least one $\alpha_m \neq 0, \mathrm{m} = 1, 2, \ldots, \mathrm{M}$
Test Statistics:

$$G^2 = -2\ln\Lambda = -2\ln\left(\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right) = -2\left(\ln L(\hat{\omega}) - \ln L(\hat{\Omega})\right) = 2\left(\ln L(\hat{\Omega}) - \ln L(\hat{\omega})\right) \quad (12)$$

The $\hat{G}^2$ test statistic follows a chi-squared distribution, so $\hat{G}^2 \sim \chi^2_{(\alpha,v)}$, where $v = n(\Omega) - n(\omega)$. The test statistic follows a chi-squared distribution. The null hypothesis is rejected meaning the model is deemed statistically significant when the test statistic exceeds the chi-squared critical value at the chosen significance level. If the global test is rejected, each basis function is then tested individually using a t-test (Equation 13–14) to confirm that no redundant terms remain in the model H_0 if $\hat{G}^2 \sim \chi^2_{(\alpha,v)}$. If H_0 is rejected in the simultaneous test, the next step is to perform a partial parameter test.

The partial test evaluates each basis function individually. The null hypothesis states that a specific basis function has no effect on poverty counts, while the alternative states that it does. The hypotheses are as follows.

$H_0 : \alpha_1 = \ldots = \alpha_M = 0$
$H_1 : \alpha_m \neq 0, \mathrm{m} = 1, 2, \ldots, \mathrm{M}$
Test Statistics:

$$T_{\mathrm{MAGPRS}} = \frac{\hat{\alpha}_m}{\mathrm{se}(\hat{\alpha}_m)} \quad (13)$$

where,

$$\mathrm{se}(\hat{\alpha}_m) = \sqrt{\left(\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}\right) C_{mm}} \quad (14)$$

where are the variance estimates for each basis function coefficient, extracted from the diagonal of the covariance matrix $(\boldsymbol{B}^T\boldsymbol{B})^{-1}$. Reject H_0 if $|T_{MAGPRS}| > t_{\alpha/2,v}$, where $v_2 = n - k$. A basis function is retained in the model if its t-statistic exceeds the critical value (with k degrees of freedom, where k is the number of basis functions in the MAGPRS model) or equivalently if its p-value , this means that all basis functions selected in the Multivariate Adaptive Generalized Poisson Regression Spline (MAGPRS) model have an effect on the response variable (Otok et al., 2019).

2.4. Small Area Estimation

Small Area Estimation (SAE) is a statistical technique for producing reliable estimates in geographic units or population subgroups where the available sample is too small to support direct, survey-based estimates (Fay & Herriot, 1979). In practical terms, a national household survey may interview only a handful of households in each district far too few to estimate local poverty rates accurately (Das & Chambers, 2024; Molina & Rao, 2010). SAE solves this by borrowing statistical strength from neighbouring areas and from auxiliary data sources (such as census records or administrative registers), combining them with the limited local sample to produce more precise sub-regional estimates.

The key advantage of SAE is its use of model-based inference: by relating small-area outcomes to predictor variables observed at a finer spatial scale, it extrapolates information from data-rich areas to data-sparse ones (Fay & Herriot, 1979). The quality of

SAE estimates is typically evaluated using the Mean Squared Error (MSE) and the Relative Efficiency (RE) the latter measuring how much more precise the SAE estimate is compared to the unaided direct survey estimate.

In the Indonesian context, SAE is particularly valuable because national surveys such as the National Socio-Economic Survey (Susenas) are designed to produce reliable estimates at the provincial level, not at the district or sub-district level. As regional autonomy has expanded, local governments increasingly require granular, area-specific statistics to plan and monitor development programmes. SAE bridges this data gap, enabling evidence-based policy at the sub-regional level.

2.5. Poverty

The central and local governments are constantly striving to promote regional development across all sectors in Indonesia with the aim of advancing the nation. One of the government's key indicators for measuring development progress is the decline in poverty rates across all regions of Indonesia. The Central Statistics Agency (BPS) calculates poverty rates using the poverty line derived from per capita household expenditure data obtained from the National Socio-Economic Survey (Susenas). Furthermore, from a macro perspective, poverty is not merely understood as a limitation of income but as a multidimensional condition influenced by health, energy, the environment, and socioeconomic factors (Marchetti et al., 2018; Tarozzi & Deaton, 2009). This perspective aligns with the World Bank (2018), which emphasizes that poverty must be analyzed across sectors to ensure that poverty alleviation strategies are more effective and sustainable.

The government has undertaken various efforts to address the issue of poverty. Among these, through the National Team for the Acceleration of Poverty Alleviation (TNP2K), the government has been and continues to strive to accelerate poverty alleviation through various programs implemented by relevant ministries and agencies.

3. METHODS

The data used are secondary data sourced from the Central Statistics Agency (BPS) in each province on the island of Java. Data collection was not limited to information available on the websites of the respective provincial BPS offices, as some variables for certain provinces were not fully available on those websites. The subjects of this study are the regencies and cities on the island of Java, resulting in a total of 119 study units. Table 1 below presents a list of the variables used in this study.

Table 1: Research Variables.

Variable	Name	Scale
Y	Number of Poverty Population	Ratio
X_1	GRDP per Capita (Constant Prices)	Ratio
X_2	Open Unemployment Rate	Ratio
X_3	Average Years of Schooling	Ratio
X_4	Percentage of Per Capita Expenditure on Food	Ratio

Based on Table 1, the variables employed in this study consist of one response variable, namely the number of poverty populations, and four predictor variables, including GRDP per capita at constant prices, open unemployment rate, average years of schooling, and the percentage of per capita expenditure on food. All variables are measured on a ratio scale, allowing for comprehensive quantitative analysis.

This study adopts a two-stage analytical framework integrating Multivariate Adaptive Generalized Poisson Regression Splines within a Small Area Estimation approach (MAGPRS-SAE). The first stage focuses on identifying the optimal nonparametric regression model using the MARS and MAGPRS approaches, while the second stage applies the Small Area Estimation (SAE) technique to improve estimation accuracy at the small area level. This sequential procedure enables the model to capture nonlinear relationships while addressing variability across areas.

The analytical steps implemented in this study are as follows:

1. Collecting data on the number of poverty populations as the response variable along with the associated predictor variables.
2. Conducting descriptive statistical analysis for both the response variable (Y) and predictor variables (X), including mean, standard deviation, minimum, and maximum values, to provide an initial overview of the data characteristics.
3. Performing an equidispersion test on the response variable to assess whether the variance equals the mean, which is a key assumption in Poisson-based modeling.
4. Determining the combination of model parameters in the MARS framework, including Basis Functions (BF), Maximum Interaction (MI), and Minimum Observation (MO), as follows: the number of basis functions is set between two to four times the number of predictors; the degree of interaction is

specified as 1, 2, and 3 to avoid excessive model complexity; and the minimum number of observations between knots is set at 0, 1, 2, 3, and 5.

5. Selecting the best MARS model based on the smallest Generalized Cross Validation (GCV) value.
6. Extending the selected MARS model into the MAGPRS framework by incorporating a spatial weighting matrix to account for spatial heterogeneity.
7. Identifying the optimal MAGPRS model based on the minimum GCV criterion.
8. Conducting parameter significance testing for the selected MAGPRS model, both simultaneously and partially, to evaluate the statistical relevance of the predictors.
9. Interpreting the MAGPRS model and analyzing the relative importance of predictor variables in explaining the response variable based on the GCV values.
10. Implementing Small Area Estimation (SAE) using a linear modeling approach as a baseline comparison.
11. Applying the SAE method based on the best-performing MAGPRS model to obtain more accurate estimates for small areas.
12. Interpreting and comparing the results obtained from the SAE models to derive meaningful insights regarding poverty estimation.

4. RESULT AND DISCUSSION

4.1. Descriptive Analysis

As an initial step, an exploratory data analysis was performed on all indicator variables included in this study, with the aim of providing a general understanding of the data characteristics.

Table 2: Descriptive Statistics.

Variable	Name	Mean	Min	Max	St. Dev
Y	Number of Poverty Population	111.24	3.49	446.8	72.6958
X_1	GRDP per Capita (Constant Prices)	48,172	14,130	510,015	63,034.99
X_2	Open Unemployment Rate	5.026	1.56	9.18	1.7988
X_3	Average Years of Schooling	8.767	5.08	12.12	1.6626
X_4	Percentage of Per Capita Expenditure on Food	50.83	36.69	65.01	6.5271

Table 2 summarises the descriptive statistics for all variables. The number of poverty people shows the widest spread across the 119 study units (standard deviation of 72.7 thousand, ranging from 3.49 to 446.8 thousand), indicating high spatial heterogeneity in poverty levels across Java's regencies and cities. In contrast, average years of schooling is the most homogeneous variable (standard deviation of 1.66 years), suggesting that educational attainment is more uniformly distributed across areas than economic or poverty outcomes. The large standard deviation of GRDP per capita (63,035 versus a mean of 48,172) further reflects the well-documented urban–rural economic divide in Java.

4.2. Multivariate Adaptive Regression Splines Modelling

The MARS model was selected through a systematic grid search over three hyperparameters basis functions (BF), maximum interaction (MI), and minimum observation (MO) evaluated by the Generalized Cross Validation (GCV) criterion and the coefficient of determination (R^2).

Table 3: Trial and Error Modelling MARS.

BF	MI	MO	GCV	R^2
16	**2**	**3**	**2642.228**	**0.4958**
16	3	3	2666.435	0.4912
16	3	1	2700.675	0.4846
⋮	⋮	⋮	⋮	⋮
8	1	0	3162.286	0.3965
8	2	0	3162.286	0.3965
8	3	0	3162.286	0.3965

As shown in Table 3, the optimal MARS configuration is BF = 16, MI = 2, MO = 3, yielding the lowest GCV of 2642.228 and an R^2 of 0.4958 the best trade-off between goodness-of-fit and model complexity across all candidate combinations.

4.3. Multivariate Adaptive Generalized Poisson Regression Splines Modelling

Building on the MARS framework, MAGPRS extends the approach within a Generalized Poisson structure to better accommodate the distributional characteristics of count response data. Model selection followed the same hyperparameter grid search (BF, MI, MO), guided by minimizing GCV and maximizing R^2.

Table 4: Trial and Error Modeling MAGPRS.

BF	MI	MO	GCV	R^2
16	**3**	**3**	**1750.786**	**0.5891**
16	3	1	1784.101	0.5813
16	2	1	1785.513	0.581
⋮	⋮	⋮	⋮	⋮

8	1	0	2232.491	0.4761
8	2	0	2232.491	0.4761
8	3	0	2232.491	0.4761

The optimal MAGPRS configuration (BF = 16, MI = 3, MO = 3) produced a GCV of 1750.786 and $R^2 = 0.5891$, representing a marked improvement over the MARS baseline (GCV 2642.228, $R^2 = 0.4958$). This reduction in GCV confirms that embedding the Generalized Poisson distribution substantially enhances the model's ability to capture the nonlinear structure inherent in small-area poverty counts.

Table 5: Simultaneous Test MAGPRS.

Test	F-statistic	df (Regression, Residual)	p-value	Decision
Simultaneous Test	8.42E+29	(14, 104)	0	Reject H_0

The simultaneous F-test (Table 5) returns an extremely large statistic with $p < 0.001$, confirming that the predictor variables collectively exert a significant effect on poverty counts. Partial testing at the 5% significance level (Table 6) further confirms that all 15 basis functions are individually significant ($p < 0.05$), validating the inclusion of each interaction term in the final MAGPRS model and reinforcing its capacity to represent complex, nonlinear covariate relationships.

Table 6: Partial Test MAGPRS.

Parameter	Estimate	Std. Error	t-value	p-value
Intercept	4.826192	5.39E-16	8.96E+15	0
BF4	-0.55123	4.94E-16	-1.12E+15	0
BF3	0.254267	2.44E-16	1.04E+15	0
BF2	0.307941	9.38E-16	3.28E+14	0
BF1	-0.1588	1.47E-16	-1.08E+15	0
BF2 × BF7	-0.15053	1.90E-16	-7.93E+14	0
BF2 × BF8	-0.0554	1.28E-16	-4.33E+14	0
BF4 × BF9	0.094441	1.94E-16	4.88E+14	0
BF4 × BF10	0.017698	2.62E-16	6.76E+13	0
BF6 × BF4 × BF9	-1.70E-06	1.80E-21	-9.47E+14	0
BF5 × BF4 × BF9	-7.43E-06	5.57E-21	-1.33E+15	0
BF6 × BF4	4.94E-06	5.92E-21	8.33E+14	0
BF5 × BF4	2.89E-05	3.28E-20	8.80E+14	0
BF12 × BF4 × BF10	-8.98E-07	3.64E-21	-2.47E+14	0
BF11 × BF4 × BF10	-0.00272	6.05E-18	-4.50E+14	0

$$\hat{f}(x) = 4.826192 - 0.158802 BF_1 + 0.3079411 BF_2 + 0.2542667 BF_3 - 0.5512344 BF_4 + 2.886175\times 10^{-5} BF_5 BF_4 + 4.93537\times 10^{-6} BF_6 BF_4 - 0.150531 BF_2 BF_7 - 0.0554 BF_2 BF_8 + 0.09444093 BF_4 BF_9 + 0.0176984 BF_4 BF_{10} - 0.002721057 BF_{11} BF_4 BF_{10} - 8.983494\times 10^{-7} BF_{12} BF_4 BF_{10} - 7.431452\times 10^{-6} BF_{13} BF_4 BF_9 - 1.704325\times 10^{-6} BF_{14} BF_4 BF_9$$

where the basis functions are as follows:

$$\begin{aligned} BF_1 &= \max(0, 6.95 - X_2) & BF_8 &= \max(0, 52.41 - X_4) \\ BF_2 &= \max(0, X_2 - 6.95) & BF_9 &= \max(0, X_4 - 40.79) \\ BF_3 &= \max(0, 9.34 - X_3) & BF_{10} &= \max(0, 40.79 - X_4) \\ BF_4 &= \max(0, X_3 - 9.34) & BF_{11} &= \max(0, 29057 - X_1) \\ BF_5 &= \max(0, 57160 - X_1) & BF_{12} &= \max(0, X_1 - 29057) \\ BF_6 &= \max(0, X_1 - 57160) & BF_{13} &= \max(0, 76301 - X_1) \\ BF_7 &= \max(0, X_4 - 52.41) & BF_{14} &= \max(0, X_1 - 76301) \end{aligned}$$

Furthermore, to determine the relative contribution of each predictor variable, the variable importance levels were calculated and are presented in Table 7.

Table 7: Importance Variable

Variable	Importance (%)
X3 (Average Years of Schooling)	100
X2 (Open Unemployment Rate)	68.6
X4 (Percentage of Per Capita Expenditure on Food)	59.4
X1 (GRDP per Capita (Constant Prices))	42.3

As shown in Table 7, average years of schooling (X3) emerges as the dominant predictor (importance = 100%), followed by open unemployment rate (X2, 68.6%), food expenditure share (X4, 59.4%), and GRDP per capita (X1, 42.3%). The primacy of education and employment over macroeconomic output suggests that human capital constraints rather than aggregate economic growth alone are the binding factors for poverty reduction at the small-area level in Java. These findings point to the need for place-sensitive interventions that prioritise schooling attainment and labour market access, particularly in areas where both indicators remain structurally weak.

4.4. Small Area Estimation

Table 8 summarizes the performance of the Linear SAE model applied to poverty estimation across Java.

Table 8: Model SAE Linier.

MSE	MAE	RMSE	R^2	RE
534.1397	8.2908	23.1115	0.9069	1.5269

The Linear SAE model achieves strong predictive accuracy ($R^2 = 0.9069$, RMSE = 23.11, MAE = 8.29), indicating a good overall fit. Its Relative Efficiency (RE = 1.527) reflects a moderate gain over direct survey estimates.

Table 9 presents the corresponding metrics for the proposed MAGPRS-SAE model.

Table 9: Model MAGPRS-SAE.

MSE	MAE	RMSE	R^2	RE
619.663	9.0334	24.893	0.8923	1.6127

Compared with the Linear SAE baseline, the MAGPRS-SAE model trades a modest reduction in point-prediction accuracy ($R^2 = 0.892$ vs. 0.907; RMSE = 24.89 vs. 23.11) for a higher Relative Efficiency (RE = 1.613 vs. 1.527). The elevated RE is the more policy-relevant metric in the SAE context: it quantifies the extent to which the model reduces estimation variance relative to direct survey estimates, and a higher value directly translates to more reliable poverty figures at the sub-district level. The MAGPRS-SAE model's superior RE therefore indicates that, despite its slightly larger prediction error on training data, it produces more stable and precise small-area estimates an outcome that is consistent with the model's ability to accommodate nonlinear covariate effects and spatial heterogeneity that the linear specification cannot capture.

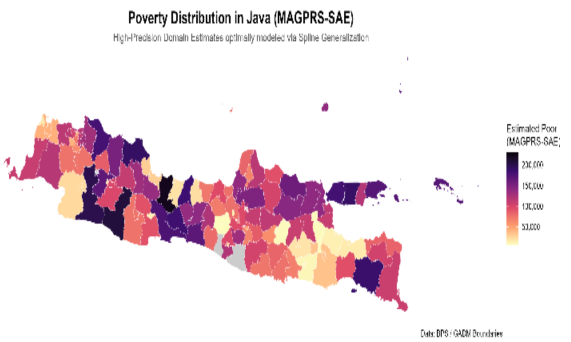

Figure 1: Poverty Distribution in Java using MAGPRS-SAE.

Figure 1 maps the estimated poverty distribution across Java's 119 regencies and cities. Poverty is spatially concentrated in the southern and eastern corridors of West Java and Central Java, while metropolitan centres Greater Jakarta (Jabodetabek), Bandung, Surabaya, and Yogyakarta record substantially lower poverty levels. This pronounced spatial clustering underscores the limitations of region-invariant models and validates the use of a spatially adaptive estimation framework. Notably, the MAGPRS-SAE estimates align closely with known socioeconomic disparities between Java's rural hinterlands and its urban economic cores, lending credibility to the model's spatial predictions and supporting the case for geographically targeted poverty alleviation policies.

5. CONCLUSION

This study demonstrates that the MAGPRS-SAE framework provides a robust, flexible approach to small-area poverty estimation in Java. By embedding a Generalized Poisson distribution within a nonparametric spline structure, the model effectively accommodates overdispersion in count data and captures nonlinear covariate interactions that conventional linear methods overlook. While the Linear SAE model attains marginally higher point-prediction accuracy ($R^2 = 0.907$), the MAGPRS-SAE model achieves superior Relative Efficiency (RE = 1.613), indicating a greater reduction in estimation variance across small areas the primary objective of SAE. Variable importance analysis highlights average years of schooling and open unemployment rate as the principal poverty determinants at the sub-regional level, underscoring the centrality of human capital in Java's poverty landscape. Collectively, these findings position MAGPRS-SAE as a methodologically sound and practically valuable tool for disaggregated poverty mapping, with direct implications for the design of spatially targeted development policies aligned with Indonesia's SDG commitments.

ACKNOWLEDGMENTS: This research is funded by the Indonesian Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science and Technology and managed under the EQUITY Program (Contract No. 4299/B3/DT.03.08/2025 & No 3029/PKS/ITS/2025).
The authors gratefully acknowledge the funding provider, Indonesian Endowment Fund for Education (LPDP). The authors also express their sincere appreciation for the academic support provided by the Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia.

REFERENCES

Badan Pusat Statistik. (2021). Hasil Sensus Penduduk 2020. Badan Pusat Statistik (BPS). https://www.bps.go.id

Billé, A., Benedetti, R., & Postiglione, P. (2016). A two-step approach to account for unobserved spatial heterogeneity. Spatial Economic Analysis, 12, 452–471. https://doi.org/10.1080/17421772.2017.1286373

Das, S., & Chambers, R. (2024). Small Area Poverty Estimation under Heteroskedasticity. Journal of Survey Statistics and Methodology, 12(2), 369–403. https://doi.org/10.1093/jssam/smad045

Fay, R. E., & Herriot, R. A. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association, 74(366), 269–277.

Friedman, J. H. (1991). Multivariate adaptive regression splines. In The Annals of Statistics (pp. 1–67).

Géniaux, G., & Martinetti, D. (2017). A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. Regional Science and Urban Economics. https://doi.org/10.1016/j.regsciurbeco.2017.04.001

Lasheras, J. S., González-Donquiles, C., Nieto, P., Moleón, J., Salas-Trejo, D., Gómez, S. L. S., Torre, A. J. M. D. L., González-Nuevo, J., Bonavera, L., Carballido-Landeira, J., & Juez, F. J. D. C. (2018). A Methodology for Detecting Relevant Single Nucleotide Polymorphism in Prostate Cancer with Multivariate Adaptive Regression Splines and Backpropagation Artificial Neural Networks. Neural Computing and Applications, 32, 1231–1238. https://doi.org/10.1007/s00521-018-3503-4

Liu, L., Zhang, S., Cheng, Y., & Liang, L. (2019). Advanced Reliability Analysis of Slopes in Spatially Variable Soils Using Multivariate Adaptive Regression Splines. Geoscience Frontiers. https://doi.org/10.1016/j.gsf.2018.03.013

López, F., & Kholodilin, K. A. (2020). The MARS Algorithm in the Spatial Framework: Non-Linearities and Spatial Effects in Hedonic Models. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3520600

Marchetti, S., Berkesewicz, M., Salvati, N., Szymkowiak, M., & Wawrowski, L. (2018). The Use of a Three-Level M-Quantile Model to Map Poverty at Local Administrative Unit 1 in Poland. Journal of the Royal Statistical Society Series A, 181(4), 1077–1104.

Melliana, A., Setyorini, Y., Eko, H., Rosi, S., & P. (2013). The Comparison of Generalized Poisson Regression and Negative Binomial Regression Methods in Overcoming Overdispersion. International Journal of Scientific \& Technology Research, 255–258.

Molina, I., & Rao, J. N. K. (2010). Small Area Estimation of Poverty Indicators. Canadian Journal of Statistics, 38(3), 369–385. https://doi.org/10.1002/cjs.10051

Otok, B. W., Hidayati, S., & P. (2019). Multivariate Adaptive Generalized Poisson Regression Spline (MAGPRS) on the Number of Acute Respiratory Infection Infants. Journal of Physics: Conference Series, 1397, 1–7.

Ramli, N., Latan, H., & Nartea, G. (2018). Why Should PLS-SEM Be Used Rather Than Regression? Evidence from the Capital Structure Perspective. In Partial Least Squares Structural Equation Modeling. Springer. https://doi.org/10.1007/978-3-319-71691-6_6

Rasam, A. R. A., & Shariff, N. M. (2019). Disease Mapping and Spatial Landscape Characterization of Tuberculosis Ecology in Malaysia. Abstracts of the ICA. https://doi.org/10.5194/ICA-ABS-1-1-2019

Tarozzi, A., & Deaton, A. (2009). Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas. Review of Economics and Statistics, 91(4), 773–792.

Wang, W., & Famoye, F. (1997). Modelling Household Fertility Decisions with Generalized Poisson Regression. Journal of Population Economics, 273–283.

World Bank. (2005). Guidelines to Small Area Estimation for Poverty Mapping. World Bank.

Yanti, Y., & Rahardiantoro, S. (2021). Identification of Factors Affecting Tuberculosis in West Java using Spatial Modeling. Proceedings of the International Conference on Statistics and Its Applications (ICSA), 9–17. https://doi.org/10.29244/ICSA.2019.PP9-17

Yasmirullah, S. D. P., Otok, B. W., Purnomo, J., & Prastyo, D. (2023). A Hybrid Model of Spatial Autoregressive-Multivariate Adaptive Generalized Poisson Regression Spline. Decision Science Letters. https://doi.org/10.5267/j.dsl.2023.7.004

Yuniarti, S. (2022). Small Area Estimation for Poverty Ratio in Java.