IN-COME
1st International Conference on Multidisciplinary Engagement

Analysis Spatial Heterogeneity of the Human Development Index in Central Java Using Bayesian Intrinsic Conditional Autoregressive Multinomial Logistic Regression

Chantika Bunga Tamiya[1*], Dewi Retno Sari Saputro[2], Purnami Widyaningsih[3]
[1,2,3]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sebelas Maret, Jl. Ir. Sutami No. 36, Surakarta 57126, Indonesia*

Corresponding Author.
**Email*: chantikabunga@student.uns.ac.id

Abstract: *The Human Development Index (HDI) is an important indicator used to measure human development achievements in terms of health, education, and standard of living. Disparities in HDI values across districts/cities in Central Java Province indicate the presence of spatial dependence, thus requiring an analytical approach that can accommodate spatial structure. This study aims to model HDI categories using a spatial Bayesian multinomial logistic regression (MLR) with an intrinsic conditional autoregressive (ICAR) prior. Spatial autocorrelation is tested using Moran's Index, while parameter estimation is conducted using the markov chain monte carlo (MCMC) method with the no-u-turn sampler (NUTS) algorithm. The model is evaluated using smoothed importance sampling leave-one-out cross-validation (PSIS-LOO). The results show the presence of positive spatial autocorrelation in the HDI distribution and spatial effect variation across regions, reflecting geographical relationships. Most predictor variables do not show significant effects as their 95% credible intervals still include zero; however, the direction of the coefficients indicates relationships between socio-economic factors and HDI categories. The spatial Bayesian MLR ICAR model is able to capture spatial dependence in the HDI data and demonstrates stable predictive performance. These findings provide important insights into understanding human development patterns and can serve as a basis for more targeted development policy formulation.*
Keywords: *Human Development Index, Spatial Bayesian, ICAR, Multinomial Logistic Regression*

1. INTRODUCTION

Indonesia's well-being is measured in terms of human development. Human development emphasizes efforts to expand people's life choices so that they can lead healthy, knowledgeable lives with a decent standard of living [1]. To measure development success, the United Nations Development Programme (UNDP) introduced the Human Development Index (HDI) as a composite indicator that reflects a country's average achievements across the three fundamental dimensions of human development: a long and healthy life, knowledge, and a decent standard of living.

The HDI score shows significant variation across regions. One province that continues to face disparities in human development is Central Java. Central Java ranks first among the provinces on Java Island for having the lowest HDI score, at 73.88 in 2024 [2]. Of the 35 regencies and cities, four regions are classified as very high, three as medium, and the rest as high. This situation indicates disparities in human development across regions. These differences in HDI scores are influenced by various factors, such as health, education, infrastructure, and socioeconomic conditions, which tend to be interrelated among neighboring regions (spatial dependence) [3]. Therefore, a spatial analysis approach is needed to identify patterns of interdependence among neighboring regions so that development policies can be more targeted.

The multinomial logistic regression (MLR) model is used to analyze the relationship between a categorical response variable and a number of predictor variables [4]. However, this model assumes that observations across regions are independent, whereas HDI data may contain spatial correlations. Ignoring spatial effects can lead to biased parameter estimates and reduce model accuracy [5]. To address this issue, a spatial

Bayesian approach with an Intrinsic Conditional Autoregressive (ICAR) prior is employed. This approach combines MLR with Bayesian inference to model spatial dependence through region-specific random effects influenced by neighboring regions [6]. Furthermore, the Bayesian framework can combine prior information with the likelihood to produce a posterior distribution that comprehensively describes the uncertainty of the estimates [7].

Previous research has shown that spatial Bayesian approaches can improve the accuracy of estimates for data with inter-regional correlations. Beltrán-Sánchez *et al.* [8] and Zhang *et al.* [9] demonstrated that incorporating a spatial component into Bayesian models yields more stable and accurate estimates compared to nonspatial models. However, the application of MLR models using a spatial Bayesian approach with an ICAR prior to model IPM categories at the district/city level remains limited, particularly in case studies in Indonesia.

Based on this, this study aims to determine an MLR model for HDI data in Central Java Province using the ICAR spatial Bayesian approach. This study is expected to contribute to the application of the MLR model using the ICAR spatial Bayesian approach in the analysis of HDI data in Central Java Province. The resulting model is expected to serve as a basis for understanding the spatial patterns of the HDI and the factors influencing them, thereby informing the formulation of more targeted and data-driven development policies.

2. METHOD

This study is an applied research that implements a MLR model with a spatial Bayesian ICAR approach on HDI data in Central Java Province. The methods used in this study include bibliometric visualization using VOSviewer software, MLR modeling, Bayesian approach, and parameter estimation. In addition, multicollinearity testing, spatial autocorrelation testing, and model evaluation are also conducted.

2.1 Visualization of Similarities Viewer (VosViewer)

In this study, the VOSviewer software was used to perform bibliometric analysis and visualize the development of research topics related to MLR and Bayesian spatial modeling [10]. VOSviewer is software designed to map and visualize the relationships between research topics based on co-occurrence extracted from Scopus metadata [11]. Using this tool, 53 documents sourced from journals over the past ten years were identified. The results of the bibliometric visualization of keyword co-occurrence are shown in Figure 1.

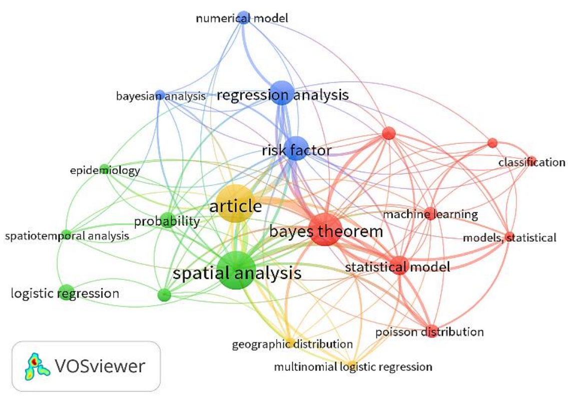


Figure 1. VosViewer visualization with title and abstract

As shown in Figure 1, the size of the circles indicates the frequency of keyword occurrence, while the connecting lines indicate the degree of association between keywords. The keywords "spatial analysis," "Bayes' theorem," and "regression analysis" have high frequencies of occurrence and strong associations, indicating that research on spatial modeling and Bayesian approaches in regression analysis has grown in the literature. However, the development of MLR models using a spatial Bayesian approach with the ICAR prior remains limited.

2.2 Multinomial Logistic Regression (MLR)

The MLR model is an extension of the binary logistic regression model used when the response variable has more than two nominal categories [4]. This model estimates the relative odds of a category relative to the reference category using the logit transformation [12]. Let Y_i be the response variable with m categories and the mth category as the reference; the MLR model is written as

$$log\left(\frac{P(Y_i=j)}{P(Y_i=m)}\right) = \alpha_j + \sum_{k=1}^{p} \beta_{jk} x_{ik}\,, \quad j = 1,2,\dots,m-1. \tag{1}$$

The probability of each category is expressed as

$$P(Y_i = j) = \frac{\exp\,(\alpha_j + \sum_{k=1}^{p} \beta_{jk} x_{ik})}{1 + \sum_{l=1}^{m-1} \exp\,(\alpha_l + \sum_{k=1}^{p} \beta_{lk} x_{ik})}$$

and for the reference category as

$$P(Y_i = m) = \frac{1}{1 + \sum_{l=1}^{m-1} \exp\left(\alpha_l + \sum_{k=1}^{p} \beta_{lk} x_{ik}\right)}.$$

The coefficient β_{jk} indicates the change in the log odds of an observation falling into category j compared to reference category m due to a one-unit change in the k predictor variable, assuming all other variables remain constant [13].

2.3 Bayesian Methods

The Bayesian method is used to estimate model parameters by combining the likelihood and the prior distribution via Bayes' theorem, thereby yielding a posterior distribution [14]. A Bayesian model consists of three main components: the likelihood, the prior, and the posterior distribution.

2.4.1 Likelihood

The response variable in region i is assumed to follow a categorical distribution with probabilities $p_{i1}, p_{i2}, \dots, p_{ik}$ such that $Y_i \sim Categorical\,(p_{i1}, \dots, p_{ik})$. The likelihood function for n regions is expressed as

$$L(\beta, u|y) = \prod_{i=1}^{n} \prod_{k=1}^{K} p_{ik}^{y_{ik}} \tag{2}$$

where y_{ik} is the k category indicator for the i region [15].

2.4.2 Prior

In this study, the regression parameters and the intercept are assumed to follow a normal distribution with a mean of zero and a variance of one, expressed as

$$\beta_{jk} \sim N(0,1), \qquad \alpha_k \sim N(0,1).$$

The spatial random effect u_i is expressed as

$$u_i \sim N(0, \sigma_u^2) \tag{3}$$

with the spatial standard deviation parameter following the distribution $\sigma_u \sim HalfNormal(1)$. The choice of prior aims to provide a weakly informative prior without dominating the information contained in the data [16]. In the Bayesian approach, the prior serves as an initial distribution that is updated by the data through the inference process to produce a posterior distribution of the parameter [17]. To ensure the validity of the parameters, a sum-to-zero constraint is applied to the spatial effects [18].

2.4.3 Posterior Distribution

The posterior distribution of the parameter is obtained via Bayes' theorem, which relates Equation (2) and Equation (3), and is thus expressed as

$$p(\alpha, \beta, u, \sigma_u \,|y) \,\propto\, L(y\,|\alpha, \beta, u)\, p(\alpha)\, p(\beta)\, p(u \mid \sigma_u)\, p(\sigma_u) \tag{4}$$

Since the posterior distribution does not have an analytically closed-form solution, parameter estimation is performed using an MCMC-based simulation method implemented in the PyMC software [19].

2.4 Markov Chain Monte Carlo (MCMC)

The MCMC method is a stochastic simulation technique for estimating parameters in Bayesian analysis by generating samples from the posterior distribution through the construction of Markov chains This approach is widely applied to various spatial statistical models to obtain efficient parameter estimates when the posterior distribution does not have a closed-form solution [20]. This method constructs a Markov chain whose stationary distribution matches the target distribution. One commonly used MCMC algorithm is the No-U-Turn Sampler (NUTS), which is an extension of Hamiltonian Monte Carlo (HMC) [21]. NUTS automatically determines the trajectory length during simulation, eliminating the need for manual specification of the number of steps.

Model parameter estimation was performed using MCMC with the aid of Python software via the PyMC library. The simulation process was run until the Markov chain reached a state of convergence. Chain convergence was verified using the Gelman–Rubin statistic $(\hat{R})$ where a value of $\hat{R}$ approaching 1 indicates that the Markov chain has converged. The resulting posterior samples $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(T)}$ are used as an approximation of the posterior distribution of the parameters [22].

2.5 Spatial and Prior Intrinsic Conditional Autoregressive (ICAR)

In spatial modeling, inter-area dependencies can be represented using a spatial random effect that follows an ICAR prior [23]. The ICAR model was first introduced by Besag and Mollié [7] to capture spatial autocorrelation based on the spatial neighborhood structure. The distribution of the spatial random effect u_j in area j is expressed as

$$E(u_j|u_{-j}) = \frac{1}{n_j}\sum_{k\in N(j)} u_k\,, \qquad Var(u_j|u_{-j}) = \frac{\sigma^2}{n_j} \tag{5}$$

where $N(j)$ is the set of neighbors of region j and n_j is the number of neighbors [24]. The ICAR prior can be expressed in the form of a multivariate normal distribution as

$$\mathrm{u} \sim N(0, {\sigma_u}^2 Q^{-1})$$

where the Q precision matrix is defined as

$$Q = D - W.$$

In this equation, the adjacency matrix $W = [w_{ij}]$. The element $w_{ij} = 1$ for regions i and j that are adjacent, and $w_{ij} = 0$ for regions that are not adjacent [18]. The matrix D is a diagonal matrix whose elements contain the number of neighbors for each region.

The matrix Q is known as the Laplacian matrix and describes the spatial dependency structure between regions [25]. This matrix is singular, so identification constraints such as the sum-to-zero constraint are required for the model to be well-defined [26].

2.6 Multicollinearity Test

A multicollinearity test is used to determine whether there is a strong linear relationship among the independent variables. One method for detecting multicollinearity is the Variance Inflation Factor (VIF), which is expressed as

$$VIF_i = \frac{1}{1 - R_i^2}, \qquad i = 1,2,\ldots,p \tag{6}$$

where R_i^2 is the coefficient of determination for the regression of the independent variable X_i on all other independent variables. If $VIF < 10$, there is no serious multicollinearity and conversely, if $VIF \geq 10$, high multicollinearity occurs, which can affect the stability of the model parameter estimates [27].

2.7 Spatial Autocorrelation Test

The Moran's I is used to detect the presence of global spatial autocorrelation, which is the relationship between the value of a variable in one area and the value of the same variable in a neighboring area [28]. This measure takes into account the elements of the spatial weight matrix $W = [w_{ij}]$ which describes the proximity relationships between areas. Moran's I is expressed as

$$\hat{I} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{7}$$

where n is the number of spatial units, x_i is the variable data for region i, x_j is the variable data for region j, $\bar{x}$ is the mean value of the HDI variable across all regions, and w_{ij} is the weight matrix element between regions i and j.

The Moran's I value is interpreted by comparing $\hat{I}$ with its expected value $\hat{I}_0$. If $\hat{I} > \hat{I}_0$the data exhibits positive autocorrelation. If $\hat{I} < \hat{I}_0$, the data exhibits negative autocorrelation, indicating that regions with different values are adjacent. The expected value of the Moran's Index under conditions where there is no spatial autocorrelation is written as $\hat{I}_0 = -\frac{1}{n-1}$.

2.8 Model Evaluation

The MLR model with a spatial Bayesian ICAR approach is evaluated using Pareto smoothed importance sampling leave-one-out cross-validation (PSIS-LOO). PSIS-LOO computes the expected log predictive density (ELPD), which reflects the model's ability to predict new data based on the posterior distribution through a leave-one-out cross-validation (LOO-CV) approach without re-estimating the model. According to Vehtari et al. [29], PSIS-LOO is defined as

$$\widehat{\mathrm{eldp}}_{\mathrm{LOO}} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \log\left(\frac{\sum_{\mathrm{s}=1}^{\mathrm{S}} \mathrm{w_i}^{(s)} p\left(\mathrm{y_i}\middle|\theta^{(s)}\right)}{\sum_{\mathrm{s}=1}^{\mathrm{S}} \mathrm{w_i}^{(s)}}\right) \tag{8}$$

where $\mathrm{w_i}^{(s)}$ denotes the importance sampling weights stabilized using a Pareto distribution, $\mathrm{y_i}$ represents the i-th observation, $\theta^{(s)}$ is the s-th sample from the posterior distribution of the model parameters, and S is the number of posterior samples. The Pareto diagnostic parameter k is obtained from the Pareto smoothing of the importance weights and is used to assess the quality of the estimation. A value of k < 0.7 indicates reliable estimation, whereas k ≥ 0.7 suggests potential instability in the computation.

2.9 Data and Analytical Steps

The data used in this study are presented in Table 1. These are secondary data obtained from the Central Java Provincial Statistics Agency and cover 29 regencies and 6 cities in Central Java Province. HDI values are categorized according to the Central Statistics Agency's classification, namely the low category (HDI < 60), medium category (60 ≤ HDI < 70), high category (70 ≤ HDI < 80), and very high category (HDI ≥ 80).

Table 1. Response and predictor variables in the research data

Variable	Description	Type	Category
Y	Human Development Index	Nominal	0 = low 1 = medium 2 = high 3 = very high
X_1	Life expectancy	Ratio	
X_2	Average years of schooling	Ratio	
X_3	Expected years of schooling	Ratio	
X_4	Per capita expenditure	Ratio	
X_5	Percentage of the poor population	Ratio	
X_6	Open unemployment rate	Ratio	
X_7	Healthcare workers per 10.000 population	Ratio	

The research steps are outlined as follows:

1. Categorizing the HDI response variable and conducting descriptive statistics to describe the data characteristics.
2. Detecting multicollinearity among predictor variables using VIF.
3. Testing spatial autocorrelation using Moran's Index based on the spatial weights matrix of regional adjacency.
4. Constructing an MLR model with a spatial Bayesian approach using the ICAR prior.
5. Estimating model parameters using the MCMC method with the NUTS algorithm and assessing convergence.
6. Finalizing the MLR model with a spatial Bayesian ICAR approach on HDI data in Central Java Province.
7. Evaluating the model using PSIS-LOO.

3. RESULTS AND DISCUSSION

This section presents the results of the HDI data analysis in Central Java Province using a spatial Bayesian MLR model with an ICAR prior. The analysis includes data categorization and descriptive statistics, multicollinearity testing, spatial autocorrelation testing, the spatial Bayesian ICAR MLR model, model implementation, model estimation, and model evaluation.

3.1. Data Categorization and Descriptive Statistics

To obtain an initial overview of the characteristics of the data used in this study, a descriptive analysis was conducted on the observed variables. The research data covers 35 districts/cities in Central Java Province in 2024, with the HDI category as the response variable. Based on the classification used, the study areas were divided into three HDI categories: medium, high, and very high; the low HDI category was not found in the observed data. The relationship between the predictor variables and the HDI categories was analyzed using a boxplot shown in Figure 1.

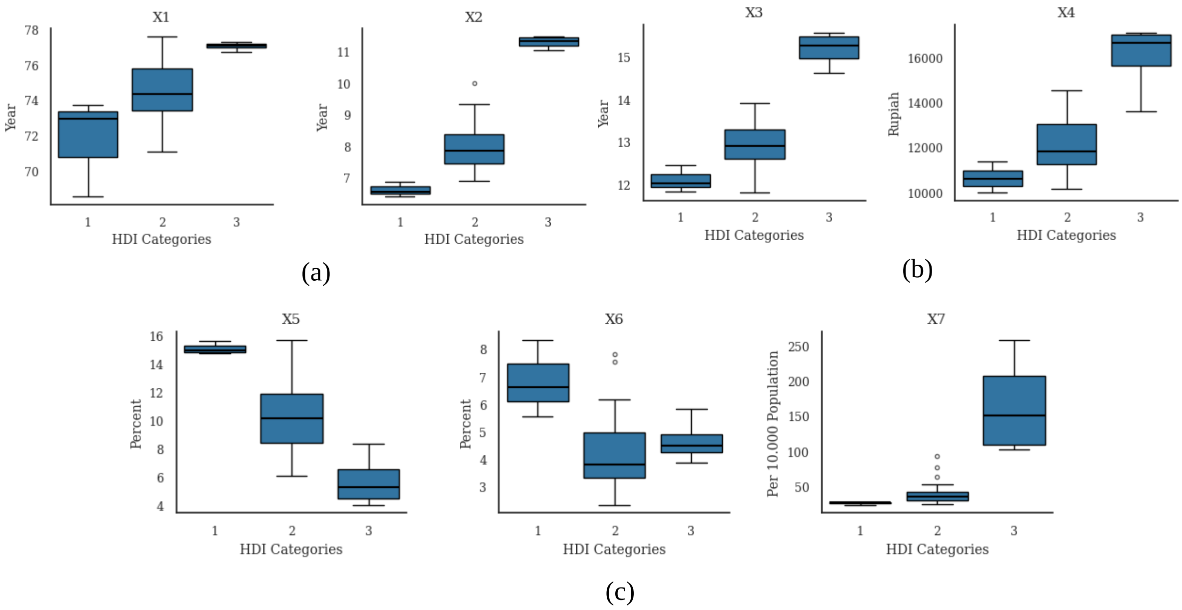


Figure 1. Boxplots of predictor variables by HDI category: (a) life expectancy (X_1) and average years of schooling (X_2); (b) expected years of schooling (X_3) and per capita expenditure (X_4); and (c) percentage of the poor population (X_5), open unemployment rate (X_6), and ratio of health workers per 10,000 population (X_7).

Based on Figure 1(a), variable X_1 shows an increase in the median from the medium HDI category to the very high HDI category. This indicates that regions with higher levels of human development have higher life expectancy. Variable X_2 also follows the same pattern, with the median in the very high HDI category being significantly higher than that of the other categories. This indicates that the level of education in a population has a strong correlation with an increase in the HDI value.

In Figure 1(b), variable X_3 shows that the highest median value is found in the "very high" HDI category. This indicates that regions with longer educational opportunities have higher levels of human development. Variable X_4 also shows that the level of economic well-being is positively correlated with an increase in the HDI.

Variable X_5 in Figure 1(c) shows an opposite pattern. The median percentage of the poor population is higher in the medium HDI category and decreases in higher HDI categories. This indicates that regions with lower poverty rates tend to have higher levels of human development. In general, variable X_6 shows a higher median in the medium HDI category compared to other categories. This suggests that the unemployment rate may be one of the factors related to the level of human development in a region. Additionally, variable X_7 reveals differences across HDI categories. The median in the very high HDI category is significantly higher than in the other categories. This indicates that regions with greater availability of healthcare personnel tend to have higher levels of human development.

3.2. Multicollinearity Test

The multicollinearity test is used to determine the presence of strong linear relationships among predictor variables. The multicollinearity test was conducted using the VIF, calculated based on Equation (6). A $VIF > 10$ indicates the presence of high multicollinearity. The VIF calculations for all initial predictor variables are shown in Table 2.

Table 2. VIF values for predictor variables

Variable	Description	VIF values
X_1	Life expectancy	4.18
X_2	Average years of schooling	18.60
X_3	Expected years of schooling	8.09
X_4	Per capita expenditure	3.14
X_5	Percentage of the poor population	2.04
X_6	Open unemployment rate	2.19
X_7	Healthcare workers per 10.000 population	3.58

Based on Table 2, variable X_2 has a $VIF > 10$ specifically 18.60 indicating a high degree of multicollinearity with the other predictor variables. This is because variable X_2 has a strong relationship with variable X_3 , both of which represent the education dimension in the formation of the HDI. To address the multicollinearity issue, variable X_2 was excluded from the model due to its high VIF value, and the VIF values for the remaining predictor variables were recalculated.

Table 3. VIF values resulting from variable reduction

Variable	Description	VIF values
X_1	Life expectancy	3.73
X_3	Expected years of schooling	4.19
X_4	Per capita expenditure	2.68
X_5	Percentage of the poor population	1.97
X_6	Open unemployment rate	2.10
X_7	Healthcare workers per 10.000 population	2.08

The results of the VIF recalculation in Table 3 show that all variables have $VIF < 10$, indicating that there is no serious multicollinearity in the model. Therefore, all six predictor variables were included in the modeling stage, and the education variable in the model is represented by the expected years of schooling variable, which captures the education dimension in the HDI category analysis.

3.3. Spatial Autocorrelation Test

The spatial autocorrelation test was conducted using the Moran's I index, as described in Equation (7), by examining the global spatial correlation between variable values in one area and those in adjacent areas based on the spatial weight matrix w_{ij}. Based on the calculation results, the Moran's Index value $\hat{I} = 0.22$ was obtained, with an expected value of $\hat{I}_0 = -0.029$, a *Z-score* of 2.121, and a *p-value* of 0.025.

The value $\hat{I} > \hat{I}_0$ indicates a tendency toward positive spatial autocorrelation, meaning that regions with high HDI values tend to be adjacent to regions that also have high HDI values, and conversely, regions with low HDI values tend to be adjacent to regions with low HDI values. Additionally, the *p-value* < 0.05 indicates that the spatial autocorrelation is significant. These results indicate the presence of spatial clustering in the distribution of the IPM across regions in Central Java Province, meaning that the assumption of independence between regions is not fully met.

3.4. MLR Model with Spatial Bayesian ICAR Approach

The model in Equation (1) is used as the basis for modeling the probability of HDI categories, which is further developed by considering spatial dependence across regions due to geographical factors. Spatial effects are incorporated into the model as regional random effects (u_i) to capture variation not explained by the predictor variables, particularly those related to geographic interactions. With the inclusion of spatial effects, the model in Equation (1) becomes

$$log\left(\frac{P(Y_i = j)}{P(Y_i = m)}\right) = \alpha_j + \sum_{k=1}^{p} \beta_{jk} x_{ik} + u_i\,, j = 1,2, \dots, m-1 \quad (9)$$

where Y_i represents the HDI category for the i-th region, x_{ik} denotes the predictor variables, α_j is the intercept, β_{jk} are the regression coefficients, and u_i represents the spatial random effect.

3.5. Model Implementation

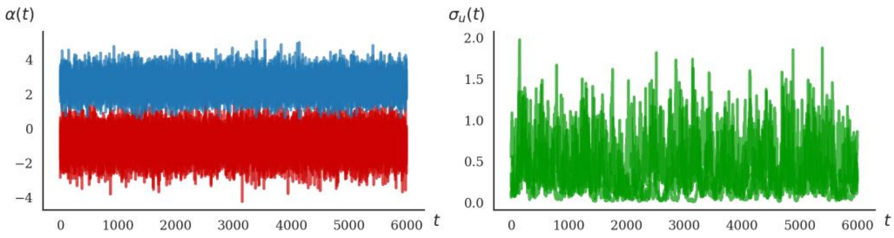
The model in Equation (9) is applied to HDI data in Central Java Province, consisting of 35 districts/cities, with the response variable defined as HDI categories and predictor variables selected based on multicollinearity testing. Spatial relationships are modeled using the W and L matrices within a Bayesian framework. The very high HDI category (category 3) is used as the reference, so parameter interpretation is conducted as a comparison relative to this category.

3.5.1. Estimation of Intercept and Coefficient Parameters

Parameter estimation in the model is carried out using a Bayesian approach through the MCMC method with the NUTS algorithm. The estimation process involves drawing samples from the posterior distribution in Equation (4). The resulting posterior samples are then used to draw conclusions regarding the model parameters.

a) Convergence Evaluation and Posterior Density of Model Parameters

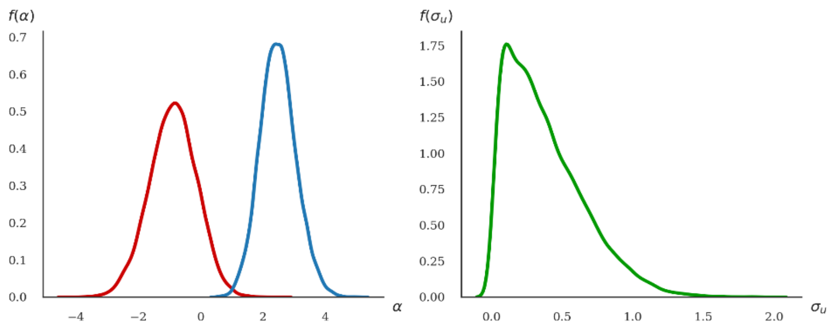
MCMC convergence evaluation is performed using trace plots to assess the stability of the Markov chains during the simulation process. The trace plot visualization is presented in Figure 2. Based on Figure 2(a), the trace plots of parameters α_1 and α_2 show stable fluctuations around their mean values throughout the iterations without any systematic trend. The parameter values are randomly distributed around their central values, indicating that the Markov chains have reached stationarity and the sampling process has converged.



(a) (b)

Figure 2. (a) Trace plots of parameters α_1 (red) and α_2 (blue); (b) trace plot of parameter σ_u (green)

In Figure 2(b), the trace plot of parameter σ_u also exhibits consistent random fluctuations without any particular trend. This indicates that the Markov chain for σ_u has converged and produces stable posterior samples. Further evaluation is conducted by examining the posterior density as shown in Figure 3.



(a) (b)

Figure 3. (a) Posterior density of parameters α_1 (red) and α_2 (blue); (b) posterior density of parameters σ_u (green)

Figure 3(a) shows that the posterior densities of parameters α_1 and α_2 are unimodal. Parameter α_1 has a mean of approximately $-0{,}88$ with a 95% credible interval in the range $[-2{,}29; 0{,}57]$, while parameter α_2 has a mean of approximately 2.49 with a 95% HDI in the range [1,36; 3,56]. In Figure 3(b), the posterior density of parameter σ_u lies entirely in the positive domain and is slightly right-skewed. The posterior mean of σ_u is 0.38 with a 95% HDI in the range [0,01; 0,88]. This indicates the presence of spatial effect variation across regions with a level of uncertainty within this range.

b) Estimation Results of Intercept and Regression Coefficients

The model parameters are estimated using 2024 HDI data along with seven predictor variables in Central Java Province. The estimation results, standard deviations, 95% credible intervals (HDI), $\hat{R}$ values, and effective sample size (ESS) are presented in Table 4.

Table 4. Estimation Results of the Spatial Bayesian ICAR MLR Model Parameters

Parameters	*Mean*	Standard Deviation	95% HDI	$\hat{R}$	*ESS*
α_1	-0.8826	0.7605	[-2.2931; 0.5679]	1.0003	16760.8377
α_2	2.4940	0.5795	[1.3647; 3.5634]	1.0001	16592.4586
$X_{1.1}$	-0.2929	0.7447	[-1.6702; 1.1187]	1.0001	18209.4831
$X_{1.3}$	-0.9794	0.7867	[-2.4652; 0.4924]	1.0001	17838.1685
$X_{1.4}$	-0.6853	0.8058	[-2.1942; 0.8059]	1.0001	16941.0816
$X_{1.5}$	0.9711	0.7293	[-0.4164; 2.3252]	1.0003	17737.9387
$X_{1.6}$	0.6766	0.6929	[-0.6497; 1.9509]	1.0000	17359.5373
$X_{1.7}$	-0.3091	0.8162	[-1.8214; 1.2101]	1.0002	17273.2970
$X_{2.1}$	-0.3238	0.6858	[-1.6095; 0.9639]	1.0000	17814.1139
$X_{2.3}$	-0.6315	0.6777	[-1.8747; 0.6599]	1.0002	17233.6431
$X_{2.4}$	-0.3982	0.6635	[-1.6781; 0.8151]	1.0001	18061.1092
$X_{2.5}$	-0.2028	0.6377	[-1.3665; 1.0183]	1.0002	16666.5437
$X_{2.6}$	-0.4924	0.6269	[-1.6486; 0.7002]	1.0003	17951.2702
$X_{2.7}$	-1.2278	0.7053	[-2.5570; 0.0709]	1.0002	16750.9814
σ_u	0.3793	0.2757	[0.0064; 0.8763]	1.0064	411.2164

Based on Table 4, parameter α_2 has a positive value with a 95% HDI that does not include zero, indicating that it serves as a clear separator between the high and very high categories. In contrast, parameter α_1 has an HDI interval that includes zero, suggesting no significant difference between the medium and very high categories.

For the predictor parameters, all 95% HDI intervals still include zero. In the comparison of category 1 to the reference category, the HLS and PPK variables show relatively stronger negative effects, indicating that increases in these variables tend to reduce the probability of being in the medium HDI category. Conversely, PPM and TPT show positive effects, implying that increases in poverty and unemployment raise the likelihood of being in that category. In the comparison of category 2 to the reference category, all coefficients are negative, with the strongest effect coming from the health workforce ratio. This suggests that improved availability of health personnel plays an important role in increasing the likelihood of regions achieving the very high HDI category. Education and unemployment variables also contribute in the same direction.

The spatial effect parameter indicates the presence of spatial dependence across regions, with σ_u equal to 0.3793 and a 95% HDI entirely in the positive domain, confirming the presence of spatial influence in the model. $\hat{R}$ values close to 1 for all parameters indicate that the MCMC process has converged and the parameter estimates are stable. Additionally, high ESS values for most parameters indicate good estimation precision, although the ESS for the spatial effect parameter is relatively lower but still within an acceptable range.

3.5.2. Estimation of Spatial Effect Parameters

To examine spatial variation not explained by the predictor variables in the model, the posterior values of the spatial effects (u_i) obtained from the spatial Bayesian model with an ICAR prior are mapped. The visualization of these spatial effects is presented in Figure 4.

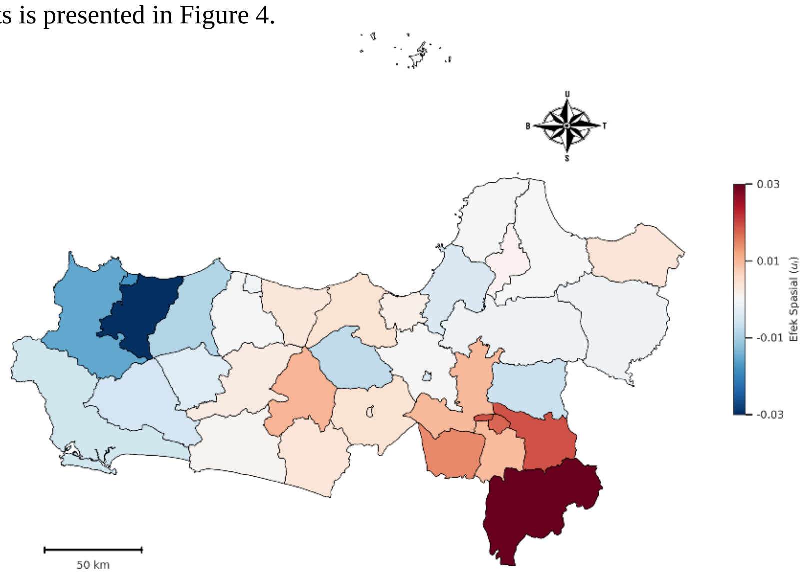


Figure 4. Spatial Effects Map (u_i) of the Spatial Bayesian ICAR Model on HDI in Central Java Province

Based on Figure 4, there is spatial variation in effects across districts/cities in Central Java Province after controlling for predictor variables in the model. The values of u_i represent the deviation of each region from the global mean, where positive values indicate a tendency toward relatively higher HDI, while negative values indicate the opposite. Regions with the highest positive spatial effects are Wonogiri Regency, followed by Karanganyar Regency and Surakarta City, suggesting the presence of additional factors beyond the model that contribute to higher HDI. Conversely, the lowest negative spatial effects are observed in Tegal Regency, Tegal City, and Brebes Regency, indicating relatively lower HDI levels. However, all 95% credible intervals (HDI) still include zero, implying that the spatial effects for each region are not statistically significant. Nevertheless, the variation in u_i values indicate the presence of spatial dependence structures across regions captured by the ICAR model.

3.6. Model Estimation

The parameter values in Table 4 are substituted into the model to obtain the final model representing the relationship between predictor variables and the response while accounting for spatial effects across regions. The spatial Bayesian ICAR MLR model for HDI data in Central Java Province, for category 1 relative to the reference category, is expressed as follows

$$log\left(\frac{P(Y_i=1)}{P(Y_i=3)}\right) = -0.8826 - 0.2929X_{1i} - 0.97945X_{3i} - 0.6853X_{4i} + 0.9711X_{5i} + 0.6766X_{6i} - 0.3091X_{7i} + u_i, \quad (10)$$

while for category 2 relative to the reference category, it is expressed as follows

$$log\left(\frac{P(Y_i=2)}{P(Y_i=3)}\right) = 2.4940 - 0.3238X_{1i} - 0.6315X_{3i} - 0.3982X_{4i} - 0.2028X_{5i} - 0.4924X_{6i} - 1.2278X_{7i} + u_i. \quad (11)$$

The value of u_i represents the spatial effect based on the ICAR prior, which accounts for spatial dependence across regions. Model interpretation is conducted using the odds ratio (OR). An OR > 1 indicates an increased likelihood relative to the reference category, while an OR < 1 indicates a decreased likelihood.

For category 1 relative to the reference category, the percentage of poor population (X_5) nd the open unemployment rate (X_6) increase the probability of being in the medium HDI category. In contrast, mean years of schooling (X_3), per capita expenditure (X_4), life expectancy (X_1), and the health workforce ratio (X_7) tend to reduce this probability. For category 2 relative to the reference category, all variables have OR < 1 indicating a tendency to decrease the likelihood of being in the high HDI category compared to the very high category. The health workforce ratio (X_7) has the strongest effect in driving the shift toward the highest HDI category. Overall, socio-economic variables such as poverty and unemployment increase the likelihood of being in lower HDI categories, while variables related to education, expenditure, and health workforce availability contribute to improvements toward higher HDI categories.

3.7. Model Evaluation

Model evaluation is conducted using PSIS-LOO to assess the predictive performance of the model on new data. Based on the calculation in Equation (8), the values obtained are $eldp_{loo}= -11,49$ and p_{loo} = 3, 70. The Pareto k diagnostic results show that all 35 observations fall within the good category k ≤ 0, 7 indicating that the PSIS-LOO estimation is stable and reliable. No Pareto k values greater than 0.7 are observed, suggesting that there are no observations with an extreme influence on the model.

4. CONCLUSION

Based on the results of the study, the MLR model with a spatial Bayesian ICAR approach applied to HDI data in Central Java Province indicates that most predictor variables are not statistically significant, as their 95% credible intervals still include zero, although the direction of the coefficients reflects relationships between socio-economic factors and HDI categories. In addition, the estimated spatial effects reveal variation across regions, indicating differences in regional characteristics that are not fully explained by the variables in the model. This finding is supported by the spatial autocorrelation test results, which show that neighboring regions tend to have similar HDI levels.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to all parties who contributed to this study, particularly in the development of the mathematical formulation, the selection of appropriate statistical models, as well as the estimation and validation procedures. Intensive scientific discussions regarding model assumptions, inferential consistency, and predictive performance evaluation have significantly strengthened the methodological validity of this research. The authors also acknowledge the support in data provision and computational resources, which enabled the efficient implementation of algorithms and numerical simulations. Such contributions have ensured that the results obtained are both reliable and reproducible.

REFERENCES

[1] United Nations Development Programme, *Human Development Report 1990: Concept and Measurement of Human Development*. New York, NY, USA: Oxford Univ. Press, 1990.
[2] Badan Pusat Statistik Provinsi Jawa Tengah, “Metode baru: Indeks pembangunan manusia menurut kabupaten/kota,” 2025. [Online]. Available: https://jateng.bps.go.id/id/statistics-table/2/ODMjMg==/-metode-baru--indeks-pembangunan-manusia-menurut-kabupaten-kota.html. Accessed: Aug. 25, 2025.
[3] L. Anselin, “Thirty years of spatial econometrics,” *Papers Reg. Sci.*, vol. 89, no. 1, pp. 3–25, 2010, doi: 10.1111/j.1435-5957.2010.00279.x.
[4] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.
[5] J. P. Cohen, “The broader effects of transportation infrastructure: Spatial econometrics and productivity approaches,” *Transp. Res. Part E Logist. Transp. Rev.*, vol. 46, no. 3, pp. 317–326, 2010, doi: 10.1016/j.tre.2009.11.003.

[6] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 2014, doi: 10.1201/b17115.

[7] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 2013.

[8] M. Á. Beltrán-Sánchez, M. A. Martinez-Beneito, and A. Corberán-Vallet, "Bayesian modeling of spatial ordinal data from health surveys," *Stat. Med.*, vol. 43, no. 21, pp. 4178–4193, 2024, doi: 10.1002/sim.10166.

[9] Z. Zhang, J. Zhang, J. Lu, and J. Tao, "Bayesian estimation of the DINA model with Pólya–Gamma Gibbs sampling," *Front. Psychol.*, vol. 11, pp. 1–15, 2020, doi: 10.3389/fpsyg.2020.00384.

[10] J. M. Saiz-Alvarez, "Innovation management: A bibliometric analysis of 50 years of research using VOSviewer and Scopus," *World*, vol. 5, no. 4, pp. 901–928, 2024.

[11] A. Kirby, "Exploratory bibliometrics: Using VOSviewer as a preliminary research tool," *Publications*, vol. 11, no. 1, p. 10, 2023.

[12] S. M. Ajeel, J. A. Haji, and B. H. Jahwar, "Using multinomial logistic regression to identify factors affecting platelet," *J. Univ. Duhok*, vol. 26, no. 2, pp. 47–56, 2023.

[13] J. S. Long and J. Freese, *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX, USA: Stata Press, 2014.

[14] R. van de Schoot et al., "Bayesian statistics and modelling," *Nat. Rev. Methods Primers*, vol. 1, no. 1, 2021, doi: 10.1038/s43586-020-00001-2.

[15] Y. Fong, H. Rue, and J. Wakefield, "Bayesian inference for generalized linear mixed models," *Biostatistics*, vol. 11, no. 3, pp. 397–412, 2010.

[16] D. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye, "Penalising model component complexity: A principled, practical approach to constructing priors," *Stat. Sci.*, vol. 32, no. 1, pp. 1–28, 2017, doi: 10.1214/16-STS576.

[17] D. R. S. Saputro, P. Widyaningsih, F. Handayani, and N. A. Kurdhi, "Prior and posterior Dirichlet distributions on Bayesian networks," *AIP Conf. Proc.*, vol. 1827, no. 1, 2017, doi: 10.1063/1.4979452.

[18] A. Freni-Sterrantino, M. Ventrucci, and H. Rue, "A note on intrinsic conditional autoregressive models for disconnected graphs," *Spat. Spatiotemporal Epidemiol.*, vol. 26, pp. 25–34, 2018, doi: 10.1016/j.sste.2018.04.002.

[19] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in Python using PyMC3," *PeerJ Comput. Sci.*, vol. 2, p. e55, 2016, doi: 10.7717/peerj-cs.55.

[20] D. R. S. Saputro, Y. K. Wardani, N. B. I. Pratiwi, and P. Widyaningsih, "Data simulation with Markov chain Monte Carlo, Gibbs sampling, and Bayes methods as parameter estimation of spatial bivariate probit regression model," *AIP Conf. Proc.*, vol. 2326, no. 1, 2021, doi: 10.1063/5.0040332.

[21] M. D. Hoffman and A. Gelman, "The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo," *J. Mach. Learn. Res.*, vol. 15, pp. 1593–1623, 2014.

[22] B. Carpenter et al., "Stan: A probabilistic programming language," *J. Stat. Softw.*, vol. 76, no. 1, pp. 1–32, 2017, doi: 10.18637/jss.v076.i01.

[23] A. Aswi et al., "Bayesian spatio-temporal conditional autoregressive localized modeling techniques for socioeconomic factors and stunting in Indonesia," *MethodsX*, vol. 15, 2025, doi: 10.1016/j.mex.2025.103464.

[24] E. M. Porter, C. T. Franck, and M. A. R. Ferreira, "Objective Bayesian model selection for spatial hierarchical models with intrinsic conditional autoregressive priors," *Bayesian Anal.*, vol. 19, no. 4, pp. 985–1011, 2024.

[25] G. Wu, "Fast and scalable variational Bayes estimation of spatial econometric models for Gaussian data," *Spat. Stat.*, vol. 24, pp. 32–53, 2018.

[26] M.-Z. Spyropoulou and J. Bentham, "Scaling priors for intrinsic Gaussian Markov random fields applied to blood pressure data," *Stat. Neerl.*, vol. 78, no. 3, pp. 491–504, 2023.

[27] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models*, 5th ed. New York, NY, USA: McGraw-Hill, 2005.

[28] Y. Chen, "Spatial autocorrelation equation based on Moran's index," *Sci. Rep.*, vol. 13, no. 1, p. 19296, 2023, doi: 10.1038/s41598-023-45947-x.

[29] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Stat. Comput.*, vol. 27, pp. 1413–1432, 2017.