Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2025, 2025:66
https://doi.org/10.28919/cmbn/9217
ISSN: 2052-2541

EXPANDED SPATIAL DURBIN MODEL FOR ANALYZING STUNTING PREVALENCE IN JAVA ISLAND

ANNISA NUR FALAH[1,*], YUDHIE ANDRIYANA[2], I GEDE NYOMAN MINDRA JAYA[2], BERTHO TANTULAR[2], EDY MARYADI[3]

[1]Post Doctoral Program, Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia

[2]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia

[3]Research Center for Artificial Intelligence and Cyber Security, National Research and Innovation Agency (BRIN), Bandung 40135, Indonesia

Abstract: Stunting is a chronic health problem that impacts children's physical and cognitive development, especially in developing countries like Indonesia. There are several exogenous factors that influence the prevalence rate of stunting. Therefore, when examining spatial aspects, there is a possibility that the spatial dependences are occured not only on the response variable but also on the exogenous variables. Thus, a model is required to consider these spatial dependences. The Expanded Spatial Durbin Model (ESDM) can be used to predict stunting prevalence influenced by exogenous variables. We used the Euclidean method to determine the inverse distance weight matrix, and the Moran Index test was applied to identify autocorrelation. By incorporating spatial effects and relevant exogenous variables, the model can provide more precise estimates of stunting prevalence in different regions. This modelling technique is very effective for increasing the accuracy of stunting prediction, considering exogenous factors such as malnutrition prevalence and human development index (HDI) where all variables contain spatial dependencies. For the study, we

*Corresponding author
E-mail address: annisa15046@mail.unpad.ac.id
Received February 27, 2025

choose Java Island in Indonesia as one of regions with a highest population density and significant stunting rates. Using the proposed technique, we found that the malnutrition variable has much stronger effect on the stunting prevalance than the HDI variable.
Keywords: stunting; spatial Durbin model; model expansion; RShiny.
2020 AMS Subject Classification: 93A30.

1. INTRODUCTION

In Indonesia, stunting remains one of the unresolvable nutritional issues. Stunting is a condition of failure to thrive in children under the age of five due to chronic malnutrition and recurrent infections, especially during the first 1,000 days of life, from fetus to 23 months old [1] [2]. Stunting indicates that suboptimal nutrition intake not only affects growth, but also other important body functions, such as brain development and the immune system [3]. Stunting will have long-term effects, such as a decline in cognitive, mental, intellectual, and physical development [4]. In the end, stunting can inhibit economic growth, increase poverty and widen inequality [3]. The WHO states that when the prevalence of stunting hits 20% or more, it becomes an issue [5]. Stunted toddlers are 22.9% common worldwide, and their nutritional state is responsible for 2.2 million of all baby deaths, according to World Health Organization (WHO) data from 2020 [6]. With 7,547,000 children, Indonesia had the highest number of stunted children worldwide, ranking fifth out of 81 nations, while the incidence of stunted toddlers in Southeast Asia reached 33.8% in 2011 [7]. The government has designated Accelerated stunting elimination as a national priority program. The government's goal in the national mid-term development plan (RPJMN) 2020–2024 is to lower the stunting rate to 14% by 2024 [1].

Java island has the highest number of stunted toddlers, with 4,353,000 of them suffering from stunting, according to the World Bank's 2017 National Team for the Acceleration of Poverty Reduction (TNP2K) map of the country's toddler stunting population [8]. The government will have a more challenging time reaching Indonesia's stunting reduction goal due to the rise in stunting incidence on Java Island. Stunting can be reduced by understanding the primary causes of the high stunting rate in the area. The condition of malnourished children and the human development index (HDI) impact the prevalence of stunting. Tiopan Sipahutar (2021) conducted a study to find stunting hotspot areas in all districts/cities in Indonesia. According to the results, three provinces on Java island were among the 14 provinces with a high distribution of stunting cases. There is a correlation between the surrounding area and the prevalence of stunting in

Indonesian districts and cities [9]. Spatial analysis in the context of stunting still needs to be widely used in Indonesia. It has yet to be used as a decision support system in policies or programs at the national and regional levels. Due to its ability to accommodate various spatial dependency structures, econometric spatial modelling has become widely used in epidemiological studies [10].

One of the econometric spatial models is the Expanded Spatial Durbin Model (ESDM) [11][12]. The ESDM is an effective analytical tool for predicting stunting prevalence because it considers the influence of variables present in a particular area and their spatial effects on surrounding areas. In the context of stunting, ESDM allows analysis of internal factors in an area and how those factors in other areas contribute to stunting prevalence in the target area. For example, if one area has a high prevalence of stunting, factors from neighbouring regions that are similar or influence each other, such as malnutrition and the human development index (HDI), may exacerbate stunting in that area. Model parameters can be estimated using R, an open-source program. For people accustomed to using it, software R can be relatively easy, but it will be challenging for others who are not. A lot of academics have sought to improve the R package's usability. One of R's visualization tools, Shiny, makes building interactive web apps straight from R easy. The application allows a user to quickly and easily visualize data [5]. This study analyses ESDM of stunting prevalence in Java Island using RShiny.

2. MATERIALS AND METHOD

In this paper, we analyze the prevalence of stunting in Java island, Indonesia, involving two exogenous variables. Since the data consists of spatial information, then we need to consider the spatial dependence structure which may occurred between spaces. The dependence is not only occurred in the response variable, but also the exogenous variables. Hence, we propose to implement a Spatial Durbin Model (SDM). However, the classical SDM is not able to measure the effects of each area and hence we expand the proposed technique using Casetti's approach.

2.1 Spatial Dependency

Spatial dependency occurs due to dependencies in regional data. Spatial dependency appeared based on Tobler's law I (1979): "Everything is related to everything else, but near things are more related than distant things"; everything is related to one another, but near things have more influence than distant things. This law is the pillar of the study of spatial econometrics developed by Anselin (1988). The linkages that occur between regions can be both positive and negative. Positive linkages occur if areas with similar characteristics surround an area. Conversely, a negative linkage occurs if a region is surrounded by other regions that do not share the same characteristics. Therefore, if there is a spatial dependency effect, the most appropriate solution is the area approach.

There are two reasons why observation values at a location/region are not mutually independent of observations at other locations, namely [13]:

1. Observation data collection related to spatial units such as subdistricts, districts/cities, and so on has measurement errors. This happens if the original data information from the sampling process cannot be accurately described.
2. The influence of socio-demographic dimensions of the region and regional economic activity in modeling. Regional science is based on the idea that location and distance are important forces for understanding human geography and economic activity. The interaction of economic activities of neighboring regions tends to be stronger than those of distant regions.

2.2. Inverse Distance Weight Matrix

In this study, the inverse distance weight matrix is used to represent the true distance between sites. The latitude and longitude of the observed location's center point are used to determine the distance between two locations. The Euclidean distance is used to measure, d_{ij}, the distance between location *i* and location *j* [14], as follows:

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

where

u_i : the latitude of the ith location, $i = 1, 2, 3, ..., N$,

v_i : the longitude ith location,

u_j : the latitude of the jth location, $j = 1, 2, 3, ..., N$,

v_j : the longitude of the jth location.

The Euclidean distance calculations are converted to kilometers using a conversion factor of 111.319, that is, $\left|d_{ij}\right| \times 111.319$. This conversion factor corresponds to the distance represented by one degree of longitude [15]. The actual distance between observation locations is then weighted as the inverse distance, computed using the following equation [16]:

$$w_{ij} = \begin{cases} \dfrac{1}{d_{ij}}, i \neq j \\ 0 \quad , i = j \end{cases}. \tag{2}$$

Next, if the sum of the distance weights of a row in the inverse distance weight matrix is not equal to 1 then the distance weights must be standardized [17] to obtain $\sum_{j=1}^{N} w_{ij} = 1, \ \forall \ i = 1, 2, 3, \ldots, N,$ where:

$$w_{ij}^{*} = \frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}}, \ \forall \ i = 1, 2, 3, \ldots, N. \tag{3}$$

2.3. Moran Index

One approach to determine the presence of spatial dependencies and check the spatial relationship or correlation between locations is to conduct a spatial autocorrelation test using Moran's Index statistic. Spatial autocorrelation estimates the correlation between observation values associated with locations on the same variable. If there is a systematic pattern in the distribution of a variable, then there is spatial autocorrelation. Spatial autocorrelation explains the dependency of spatial data between one location and another based on a measure of proximity or intersection [18].

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(x_i - \bar{x} \right) \left(x_j - \bar{x} \right)}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}. \tag{4}$$

The hypothesis formulation in this test is as follows:

$H_0 : I = I_0$ There is no spatial autocorrelation between locations, vs.

$H_1 : I \neq I_0$ There is spatial autocorrelation between locations,

with $I_0 = E(I)$.

The test statistic employed is expressed below:

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}} \sim N(0,1), \tag{5}$$

with

$$E(I) = -\frac{1}{n-1}, \tag{6}$$

$$Var(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{\left(n^2 - 1 \right) S_0^2} - \left[E(I) \right]^2, \tag{7}$$

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\ , \qquad S_1 = \frac{1}{2}\sum_{1\neq j}^{n}\left(w_{ij}+w_{ji}\right)^2\ , \qquad S_2 = \sum_{i\neq j}^{n}\left(\sum_{i=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2,$$

where

I : the Moran Index value,

n : the number of observation locations,

x_i : the value of the observation variable at the *i*th location,

x_j : the value of the observation variable at *j*th location,

$\bar{x}$: the average of the number of variables,

w_{ij} : an element of the standardized weight matrix between locations i and j.

The Moran Index ranges between [−1, 1]. A negative value indicates negative spatial autocorrelation, while a positive value signifies positive spatial autocorrelation [19].

Decision Rule:

The null hypothesis is rejected based on the following decision rule:

Reject H_0 at the significance level α if $-Z_{score} \le -Z_{\alpha/2}$ or $Z_{score} \ge Z_{\alpha/2}$.

2.4 Expanded Spatial Durbin Model (ESDM)

The ESDM is a spatial regression model that has a form like the Spatial Autoregressive which has a spatial lag in the response variable. While the ESDM is characterized by a spatial lag in the exogenous variables [18]. The ESDM has the following equation form [11][12]:

$$\boldsymbol{y} = \rho\mathbf{W}\boldsymbol{y} + \alpha\mathbf{1}_n + \mathbf{XZJ}\boldsymbol{\beta}_0 + \mathbf{W}\tilde{\mathbf{X}}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \text{ with } \boldsymbol{\varepsilon} \overset{iid}{\sim} N(\mathbf{0},\sigma^2\mathbf{I}) \tag{8}$$

Letting $\mathbf{A} = \mathbf{XZJ}$, it follows that:

$$\boldsymbol{y} = \rho\mathbf{W}\boldsymbol{y} + \alpha\mathbf{1}_n + \mathbf{A}\boldsymbol{\beta}_0 + \mathbf{W}\tilde{\mathbf{X}}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \text{ with } \boldsymbol{\varepsilon} \overset{iid}{\sim} N(\mathbf{0},\sigma^2\mathbf{I}) \tag{9}$$

$$\boldsymbol{y} = \rho\mathbf{W}\boldsymbol{y} + \boldsymbol{U}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{10}$$

where:

$\boldsymbol{U} = \begin{bmatrix}\mathbf{1}_n & \mathbf{A} & \mathbf{W}\tilde{\mathbf{X}}\end{bmatrix}$,

$\boldsymbol{\delta} = \begin{bmatrix}\alpha \\ \boldsymbol{\beta}_0 \\ \boldsymbol{\theta}\end{bmatrix}$,

with:

$\boldsymbol{y}$: vector of dependent variables of size $(n\times 1)$,

$$\boldsymbol{y} = \begin{pmatrix} y(s_1) \\ y(s_2) \\ \vdots \\ y(s_n) \end{pmatrix}$$

$\tilde{\mathbf{X}}$: matrix of independent variables of size $(n \times k)$,

$$\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1k} \\ \tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2k} \\ \cdots & \cdots & \ddots & \vdots \\ \tilde{x}_{n1} & \tilde{x}_{n2} & \cdots & \tilde{x}_{nk} \end{pmatrix}$$

$\mathbf{X}$: matrix of independent variables of size $(n \times nk)$,

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & x_{21} & \cdots & x_{2k} & 0 & \vdots & \cdots & \vdots \\ \vdots & \ddots & \vdots & 0 & \ddots & 0 & \ddots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & x_{n1} & \cdots & x_{nk} \end{pmatrix}$$

ρ : spatial lag coefficient of the dependent variable,

α : constant parameter,

$\mathbf{W}$: spatial weight matrix of size $(n \times n)$,

$$\mathbf{W} = \begin{pmatrix} 0 & w_{12} & \dots & w_{1n} \\ w_{21} & 0 & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & 0 \end{pmatrix}$$

$\mathbf{Z}$: location information that contains elements Z_{xi}, Z_{yi} with $i = 1, \dots, n$ representing the latitude and longitude of each observation, of size $(nk \times 2nk)$,

$$\mathbf{Z} = \begin{pmatrix} Z_{x1} \otimes \mathbf{I}_k & Z_{y1} \otimes \mathbf{I}_k & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & Z_{xn} \otimes \mathbf{I}_k & Z_{yn} \otimes \mathbf{I}_k \end{pmatrix}$$

$\mathbf{J}$: expansion of the identity matrix of size $(2nk \times 2k)$,

$$\mathbf{J} = \begin{pmatrix} \mathbf{I}_k & 0 \\ 0 & \mathbf{I}_k \\ \vdots & \vdots \\ 0 & \mathbf{I}_k \end{pmatrix}$$

$\boldsymbol{\beta}$: matrix of size $(nk \times 1)$ contains parameter estimators for all explanatory k variables at each observation,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1(s_1) \\ \beta_1(s_2) \\ \vdots \\ \beta_k(s_n) \end{pmatrix}$$

$\boldsymbol{\beta}_0$: parameter expressed by $\boldsymbol{\beta}_{latitude}, \boldsymbol{\beta}_{longitude}$ of size $(2k \times 1)$,

$$\boldsymbol{\beta}_0 = \begin{pmatrix} \boldsymbol{\beta}_{latitude} \\ \boldsymbol{\beta}_{longitude} \end{pmatrix}$$

$\boldsymbol{\theta}$: spatial lag parameter vector of covariate variable of size $(k \times 1)$,

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{pmatrix}$$

$\otimes$: Kronecker product,

$\boldsymbol{\varepsilon}$: error vector of size $(n \times 1)$,

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon(s_1) \\ \varepsilon(s_2) \\ \vdots \\ \varepsilon(s_n) \end{pmatrix}$$

$\mathbf{s_i}$: location matrix with $i = 1, \dots, n$.

$$\boldsymbol{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix}$$

2.5 Parameter estimation using Maximum Likelihood

The random error variable in the ESDM is assumed to follow a normal distribution. Consequently, parameter estimation in this model adopts the SAR-X estimation method, employing the Maximum Likelihood Estimation (MLE) technique [20]. Equation (13) can be expressed as follows [11]:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \boldsymbol{U}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \text{ with } \boldsymbol{\varepsilon} \overset{iid}{\sim} N(\mathbf{0}, \sigma^2\mathbf{I}),$$

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \rho\mathbf{W}\boldsymbol{y} - \boldsymbol{U}\boldsymbol{\delta}. \tag{11}$$

The probability density function used is expressed as follows:

$$f(\mathbf{y}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \boldsymbol{U}\boldsymbol{\delta})^T(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \boldsymbol{U}\boldsymbol{\delta})}{2\sigma^2}\right]. \tag{12}$$

The likelihood function of the dependent variable $\boldsymbol{y}$ is formulated as follows:

$$\begin{aligned} L(\rho, \boldsymbol{\delta} \mid \mathbf{y}) &= f(\mathbf{y} \mid \rho, \boldsymbol{\delta}) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \boldsymbol{U}\boldsymbol{\delta})^T(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \boldsymbol{U}\boldsymbol{\delta})}{2\sigma^2}\right]. \end{aligned} \tag{13}$$

Furthermore, the log-likelihood function is obtained as:

$$\begin{aligned} \ln L(\rho, \boldsymbol{\delta} | \boldsymbol{\varepsilon}) &= \ln\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \boldsymbol{U}\boldsymbol{\delta})^T(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \boldsymbol{U}\boldsymbol{\delta})}{2\sigma^2}\right] \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \boldsymbol{U}\boldsymbol{\delta})^T(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \boldsymbol{U}\boldsymbol{\delta})}{2\sigma^2}. \end{aligned} \tag{14}$$

The estimation parameters ρ and $\boldsymbol{\delta}$ obtained by maximizing the log likelihood function are given as follows:

$$\rho = \left((\mathbf{W}\mathbf{y})^T \mathbf{W}\mathbf{y}\right)^{-1} (\mathbf{W}\mathbf{y})^T (\mathbf{y} - \boldsymbol{U}\boldsymbol{\delta}) \tag{15}$$

$$\boldsymbol{\delta} = \left(\boldsymbol{U}\boldsymbol{U}^T\right)^{-1} \boldsymbol{U}^T (\mathbf{y} - \rho\mathbf{W}\mathbf{y}) \tag{16}$$

2.6. Mean Absolute Percentage Error (MAPE)

To evaluate the model’s performance, the Mean Absolute Percentage Error (MAPE) is calculated as follows:

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\boldsymbol{y}(s_i) - \hat{\boldsymbol{y}}(s_i)}{\boldsymbol{y}(s_i)}\right|\right) \times 100\%, \tag{17}$$

with

$y(s_i)$: the values in the actual data at the location s_i,

$\hat{y}(s_i)$: the values in the prediction data at the location s_i,

n : the number of observation locations.

According to Lawrence's criteria (2009) [21], MAPE values can be categorized as follows:

Table 1. MAPE Score Scale

Scale MAPE	Accuracy Score
$\leq$ 10%	Very accurate prediction
10 < MAPE $\leq$ 20%	Good prediction
20 < MAPE $\leq$ 50%	Reasonable prediction
> 50%	Inaccurate prediction

3. Results and Discussion

This study used data from the Central Bureau of Statistics and basic health research reports 2018. The units of analysis are 119 districts/cities on Java Island. All data used in this study are taken from the public domain. The variables employed were the prevalence of stunting, malnutrition, and the human development index (HDI).

Table 2. Response variable and exogenous variables

Variables	Notation
The prevalence of stunting	$\boldsymbol{y}$
The prevalence of malnutrition	$\boldsymbol{X}_1$
The Human Development Index (HDI)	$\boldsymbol{X}_2$

In this study, the ESDM method is used to analyze spatial relationships and predict stunting prevalence, as described in the flowchart in Figure 1, which includes the stages of data collection, the formation of an inverse distance weight matrix, identification of spatial autocorrelation using the Moran Index, estimation of ESDM parameters through the Maximum Likelihood Estimation (MLE) method, to visualization and interpretation of prediction results.

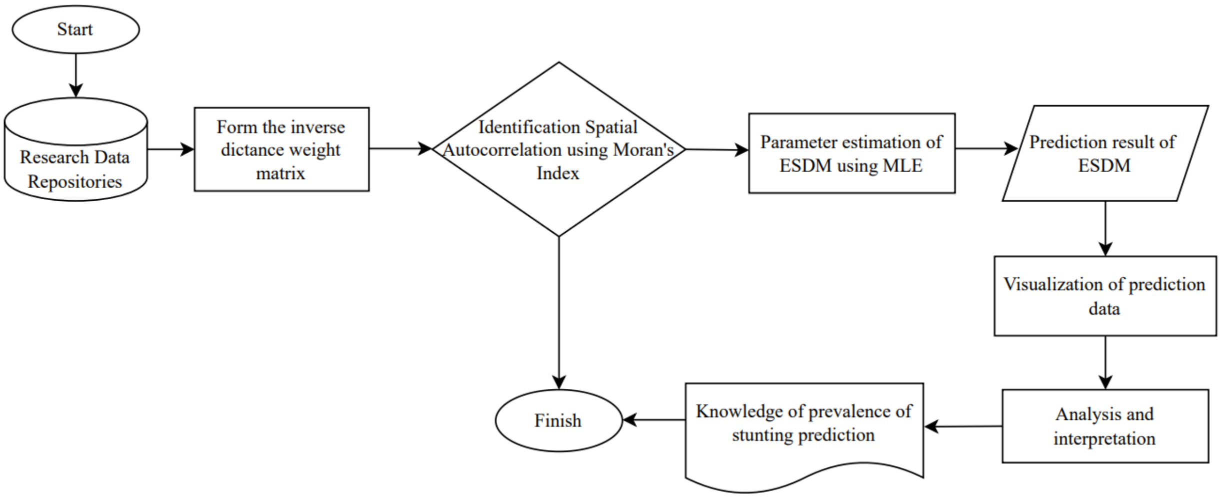

Figure 1. ESDM flowchart

The prediction procedure was implemented by developing an RShiny web application for the Expanded Spatial Durbin Model (ESDM). RShiny's development results are publicly available since it is constructed straight from R software. Because R is a free program with an open source code, anyone who wants to can share and enhance it [22]. This application is accessible via the following URL: https://andriyanafalah.shinyapps.io/SDM-Expansion/. The application has six menues, such as the model description, import data saved in CSV format, construct vector and matrix based on the data, prediction result, download data, and created by. Figure 1 illustrates the dashboard contained in RShiny from ESDM which consists of six dashboards including Description of Model, Import Data, Vector and Matrix, Result of Prediction, Download Data, and Created by button.

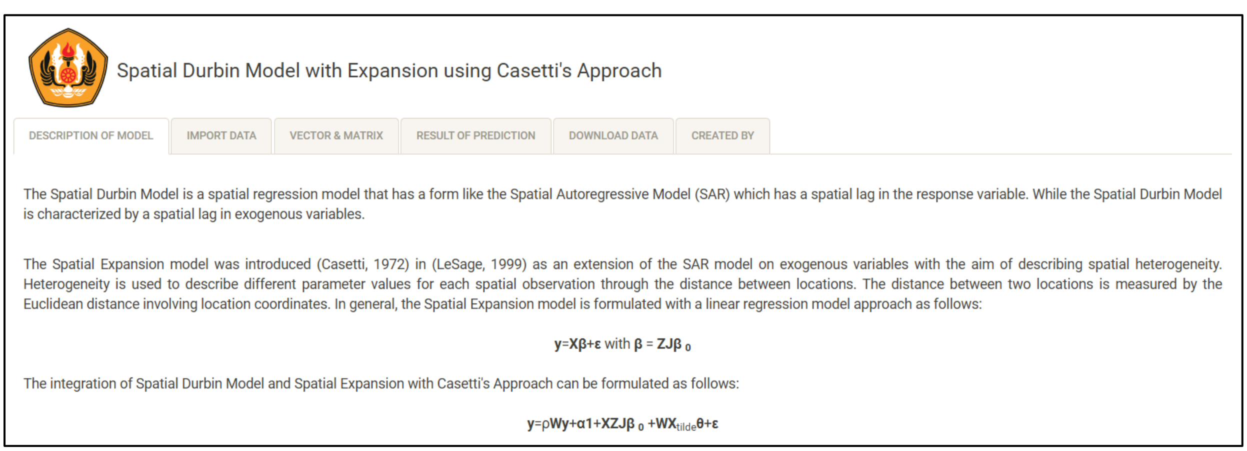

Figure 2. Dashboard for Description of the Model

As can be seen in Figure 2, we firslty introduce the proposed model in the landing page of the dashboard. We provide several menues to import or download the data. For the analysis, we can choose the Vector & Matrix button where the appearance is provided in Figure 3 below.

Figure 2. Dashboard of Data Analysis

The Moran's Index was computed to assess the presence of spatial autocorrelation between the observation sites. Equation (4) was used to calculate the Moran's Index for each variable using the Inverse Distance Weight Matrix. The results of these calculations are presented in Table 3.

Table 3. The result of the spatial autocorrelation test

No	Variables	I	$E(I)$	$Var(I)$	$p-value$
1	$\boldsymbol{y}$	0.313	-0.008	0.002	4.911×10^{-12} *
2	$\boldsymbol{X}_1$	0.187	-0.008	0.002	1.659×10^{-5} *
3	$\boldsymbol{X}_2$	0.313	-0.008	0.002	5.345×10^{-12} *

*Significant $\alpha = 0.05$

Based on Table 3, the p-values are all less than the significance level $(\alpha = 0.05)$ so the null hypothesis is rejected, which means that there is positive spatial autocorrelation on each variable. The high stunting area will be surrounded by high areas as well, and low stunting area will be surrounded by low areas as well. The high malnutrision area will be surrounded by high areas as well, and low malnutrision area will be surrounded by low areas as well. The high human development index area will be surrounded by high areas as well, and low human development index area will be surrounded by low areas as well.

The estimation of prediction parameters in the ESDM was conducted via the RShiny web application. An estimated $\hat{\rho}$ value of 0.999 was obtained, producing an optimum spatial lag with a positive value $(\hat{\rho} > 0)$ and indicative of a spatial lag dependence. This $\hat{\rho}$ value signifies the influence of adjacent locations within Java Island on prevalence stunting prediction data. It further highlights the spatial autocorrelation of prevalence stunting on Java island, indicating that districts and cities with high prevalence stunting tend to be spatially clustered with others exhibiting similar high prevalence stunting levels, or with minimal variation. The results of the parameter estimate calculation $\widehat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\theta}}$ are shown in Table 4.

Table 4. Parameter estimated value

Coefficient	Parameter Estimated Value		
		$\widehat{\boldsymbol{\beta}}_0$	$\hat{\boldsymbol{\theta}}$
$\boldsymbol{X_1}$ (malnutrition)	$\hat{\beta}_{latitude}$	0.134	11.969
	$\hat{\beta}_{longitude}$	-0.036	
$\boldsymbol{X_2}$ (HDI)	$\hat{\beta}_{latitude}$	0.011	−0.227
	$\hat{\beta}_{longitude}$	-0.007	

The estimate $\widehat{\boldsymbol{\beta}}_0$ measures the direct impact of the exogenous variables on prevalence stunting within the same region, while the estimate $\hat{\boldsymbol{\theta}}$ captures the spillover effects of the exogenous variables on prevalence stunting. Based on the estimate $\widehat{\boldsymbol{\beta}}_0$, we can derive the estimate $\hat{\boldsymbol{\beta}}$, from which individual parameter estimates were derived for each exogenous variable across the 119 districts and cities. The ESDM produces different parameter estimates for each exogenous variable at each location due to the expansion of the exogenous variable matrix involving latitude and longitude information at each location. Stunting prevalence estimations and categories is shown in Table 5 and visualization in 119 districts/cities in Java island is shown in Figure 4.

Table 5. Stunting Prevalence Estimations and Categories

No.	Locations	Stunting Prevalence	Category	No.	Locations	Stunting Prevalence	Category
1	Sleman	14.7	low	61	Temanggung	30.26	medium
2	City Jakarta Barat	14.84	low	62	Ponorogo	30.5	medium
3	City Jakarta Selatan	16.23	low	63	Tegal	30.59	medium
4	City Bekasi	16.75	low	64	Situbondo	30.66	medium
5	City Yogyakarta	16.93	low	65	Tuban	30.76	medium
6	City Jakarta Timur	18.37	low	66	City Cirebon	31.18	high
7	City Jakarta Utara	18.95	low	67	Sukoharjo	31.33	high

No.	Locations	Stunting Prevalence	Category	No.	Locations	Stunting Prevalence	Category
8	City Tangerang	19.07	low	68	Wonosobo	31.49	high
9	City Tangerang Selatan	19.85	low	69	Malang	31.74	high
10	City Tegal	20.72	medium	70	Banyuwangi	32.01	high
11	City Mojokerto	20.86	medium	71	Banyumas	32.02	high
12	City Cimahi	21.06	medium	72	Serang	32.04	high
13	City Bandung	21.74	medium	73	Cilacap	32.1	high
14	Kulon Progo	22.65	medium	74	Sumedang	32.22	high
15	Bantul	22.89	medium	75	Madiun	32.44	high
16	City Depok	23.21	medium	76	Gunung Kidul	32.51	high
17	Tangerang	23.23	medium	77	Pangandaran	32.71	high
18	City Cilegon	23.32	medium	78	Bogor	32.86	high
19	Purworejo	23.33	medium	79	Blora	32.86	high
20	City Malang	23.42	medium	80	Karawang	33.11	high
21	City Jakarta Pusat	23.51	medium	81	Sragen	33.16	high
22	Banjarnegara	24.09	medium	82	Ciamis	33.39	high
23	City Serang	24.62	medium	83	Cianjur	33.51	high
24	Semarang	24.87	medium	84	Jepara	33.57	high
25	Batang	25.28	medium	85	Pemalang	33.68	high
26	Kendal	25.48	medium	86	Cirebon	33.71	high
27	City Blitar	25.59	medium	87	Tasikmalaya	33.8	high
28	City Tasikmalaya	25.73	medium	88	Kebumen	33.9	high
29	Rembang	25.73	medium	89	Magelang	33.95	high
30	Demak	26.1	medium	90	Indramayu	33.99	high
31	Tulungagung	26.16	medium	91	Lumajang	34.01	high
32	Pacitan	26.33	medium	92	City Salatiga	34.24	high
33	Bekasi	26.37	medium	93	Sumenep	34.34	high
34	Sidoarjo	27.05	medium	94	City Kediri	34.63	high
35	Kepulauan Seribu	27.13	medium	95	Garut	34.64	high

No.	Locations	Stunting Prevalence	Category	No.	Locations	Stunting Prevalence	Category
36	Gresik	27.16	medium	96	City Pasuruan	34.88	high
37	Blitar	27.36	medium	97	Bojonegoro	34.91	high
38	City Magelang	27.55	medium	98	Pekalongan	35.12	high
39	City Bogor	27.79	medium	99	Bandung	35.21	high
40	City Banjar	27.89	medium	100	Lamongan	35.49	high
41	Wonogiri	27.9	medium	101	Majalengka	36.62	high
42	Karanganyar	28.02	medium	102	Bandung Barat	36.69	high
43	City Batu	28.33	medium	103	Pati	37.9	high
44	Boyolali	28.51	medium	104	Bondowoso	37.97	high
45	City Surabaya	28.57	medium	105	Jember	38.31	high
46	Subang	28.64	medium	106	Brebes	38.53	high
47	Kuningan	28.67	medium	107	Grobogan	39.16	high
48	Purbalingga	28.88	medium	108	Pandeglang	39.47	high
49	City Sukabumi	28.99	medium	109	City Surakarta	39.5	high
50	City Madiun	29.06	medium	110	Pasuruan	39.7	high
51	Kediri	29.36	medium	111	Trenggalek	39.88	high
52	Kudus	29.38	medium	112	Probolinggo	39.9	high
53	Klaten	29.62	medium	113	Lebak	40.19	very high
54	Nganjuk	29.65	medium	114	Ngawi	40.47	very high
55	City Semarang	29.68	medium	115	Purwakarta	41.01	very high
56	City Pekalongan	29.73	medium	116	Sukabumi	41.35	very high
57	Jombang	29.77	medium	117	Bangkalan	41.87	very high
58	Mojokerto	29.91	medium	118	Pamekasan	44.12	very high
59	City Probolinggo	30.02	medium	119	Sampang	47.92	very high
60	Magetan	30.25	medium				

Based on Table 5, the stunting prevalence estimations in Java Island is 30.14% on average, with Sampang district having the most significant incidence (47.92%) and Sleman district having the lowest prevalence (14.7%). Only nine districts and cities fall into the low category (green zone) in Figure 4, with a less than 20% stunting frequency. In contrast, 110 other districts and cities throughout all provinces on Java island suffer from stunting issues. The proposed model yielded a MAPE value of 11.66%, indicating good prediction.

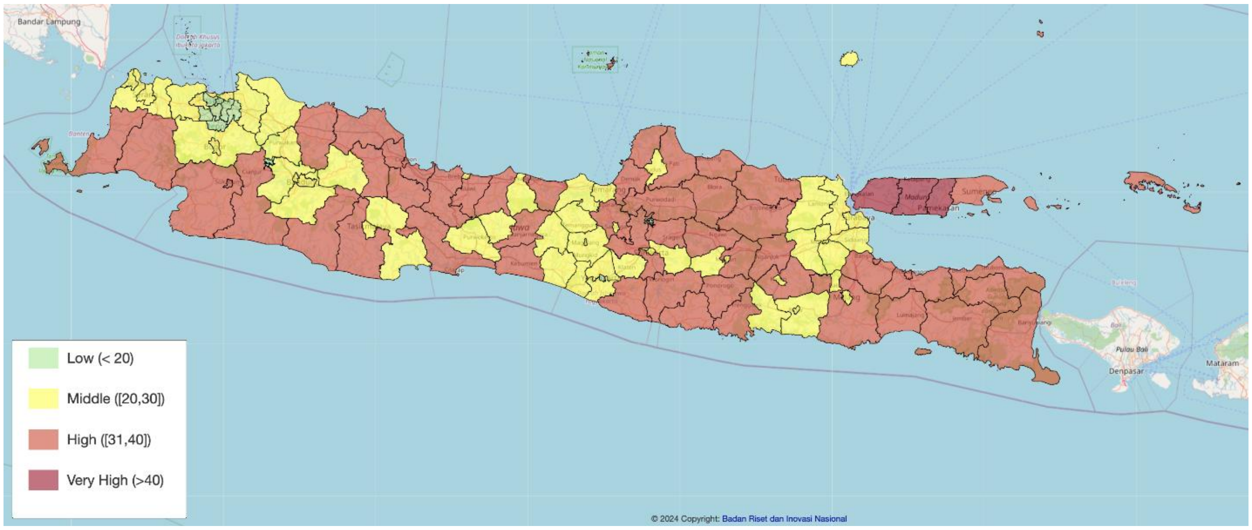

Figure 4. Map of the prevalence of stunting prediction in Java island

Based on Figure 4 shows the category levels (Low, Middle, High, Very High) grouped by color. A spatial pattern is visible, where areas with similar categories tend to cluster geographically. The “Very High” regions are concentrated in the eastern region. This suggests that areas in this category tend to influence or be influenced by neighboring areas in the same category. The “Low” regions are more dispersed in the western and central parts but still form a small group.

4. CONCLUSIONS

The prevalence of stunting in Java Island shows the existence of spatial autocorrelation in both the response variable and the exogenous variables. The spatial econometric model that can be used to analyze is the ESDM. The spatial dependency test results show positive spatial autocorrelation in stunting prevalence, malnutrition prevalence, and human development index (HDI). The R program, accessed for free at https://andriyanafalah.shinyapps.io/SDM-Expansion/, was used to estimate the model parameters. Malnutrition and human development index (HDI) variables significantly influence stunting in Java island. Therefore, stunting reduction can be accelerated by considering these two factors. The existence of spatial dependencies illustrates that the incidence of stunting in each district/city on Java Island is influenced by changes in causal factors in the area and the surrounding regions. Therefore, cooperation between districts/cities in Java Island is needed to achieve the stunting reduction target.

ACKNOWLEDGMENTS

The authors gratefully thank to Universitas Padjadjaran for supporting the research which is funded by Postdoctoral Research Grant scheme with the contract number 2413/UN6.3.1/PT.00/2024 and RDPD Research Grant scheme with the contract number 2096/UN.6.3.1/PT.00/2024.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] BPS, Laporan Indeks Khusus Penanganan Stunting 2018-2019, Badan Pusat Statistik, Jakarta, 2019.

[2] Sekretariat Wakil Presiden Republik Indonesia, 100 Kabupaten/Kota Prioritas untuk Intervensi Anak Kerdil (Stunting), TNP2K, Jakarta Pusat, 2017.

[3] BPS, Laporan Indeks Khusus Penanganan Stunting 2021-2022, Badan Pusat Statistik, Jakarta, 2022.

[4] I. Adedeji, C. John, S. Okolo, et al. Malnutrition and the Intelligence Quotient of Primary School Pupils in Jos, Nigeria, Br. J. Med. Med. Res. 21 (2017), 1–13. https://doi.org/10.9734/BJMMR/2017/32504.

[5] G. Apriluana, S. Fikawati, Analisis Faktor-Faktor Risiko Terhadap Kejadian Stunting Pada Balita (0-59 Bulan) Di Negara Berkembang Dan Asia Tenggara, Media Penelit. Pengemb. Kesehat. 28 (2018), 247–256. https://doi.org/10.22435/mpk.v28i4.472.

[6] World Health Organization, World Health Statistics 2016: Monitoring Health for the SDGs, Sustainable Development Goals, World Health Organization, Geneva, 2016. https://iris.who.int/handle/10665/206498.

[7] C.R. Titaley, I. Ariawan, D. Hapsari, et al. Determinants of the Stunting of Children Under Two Years Old in Indonesia: A Multilevel Analysis of the 2013 Indonesia Basic Health Survey, Nutrients 11 (2019), 1106. https://doi.org/10.3390/nu11051106.

[8] Sutarto, D. Mayasari, R. Indriyani, Stunting, Faktor Resiko dan Pencegahannya, J. Agromedicine, 540 (2018), 243–243.

[9] T. Sipahutar, T. Eryando, M.P. Budhiharsana, et al. Finding Stunting Hotspot Areas In Seven Major Islands Using Spatial Analysis: for the Acceleration of Stunting Prevention in Indonesia, medRxiv (2021). https://doi.org/10.1101/2021.03.31.21254736.

[10] I.G.N.M. Jaya, Y. Andriyana, Analisis Data Spasial Perspektif Bayesian, Alqaprint Jatinangor, Bandung, 2020.

[11] Y. Andriyana, A.N. Falah, B.N. Ruchjana, A. Sulaiman, E. Hermawan, T. Harjana, D.L. Lim-Polestico, Spatial Durbin Model with Expansion Using Casetti's Approach: A Case Study for Rainfall Prediction in Java Island, Indonesia, Mathematics 12 (2024), 2304. https://doi.org/10.3390/math12152304.

[12] A.N. Falah, Y. Andriyana, B.N. Ruchjana, et al. An Expanded Spatial Durbin Model with Ordinary Kriging of Unobserved Big Climate Data, Mathematics 12 (2024), 2447. https://doi.org/10.3390/math12162447.

[13] J.P. LeSage, The Theory and Practice of Spatial Econometrics, University of Toledo, Toledo, (1999).

[14] G.Y. Lu, D.W. Wong, An Adaptive Inverse-Distance Weighting Spatial Interpolation Technique, Comput. Geosci. 34 (2008), 1044–1055. https://doi.org/10.1016/j.cageo.2007.07.010.

[15] E. Maria, E. Budiman, Haviluddin, M. Taruk, Measure Distance Locating Nearest Public Facilities Using Haversine and Euclidean Methods, J. Phys.: Conf. Ser. 1450 (2020), 012080. https://doi.org/10.1088/1742-6596/1450/1/012080.

[16] R. Rosmanah, V.A. D. Djara, Y. Andriyana, I.G.N.M. Jaya, The Spatial Econometrics of Economic Growth in Sumatera Utara Province, J. Math. Comput. Sci. 12 (2022), 128. https://doi.org/10.28919/jmcs/7283.

[17] Y. Notonegoro, Y. Andriyana, B.N. Ruchjana, Comparison of Distance-Based Spatial Weight Matrix in Modeling Internet Signal Strengths in Tasikmalaya Regency Using Logistic Spatial Autoregressive Model, Int. J. Data Netw. Sci. 8 (2024), 893–906. https://doi.org/10.5267/j.ijdns.2023.12.016.

[18] L. Anselin, Spatial Econometrics, Springer, Dordrecht, 1988.

[19] K. Kopczewska, Applied Spatial Statistics and Econometrics: Data Analysis in R, Routledge, 2021. https://doi.org/10.4324/9781003033219.

[20] A.N. Falah, B.N. Ruchjana, A.S. Abdullah, J. Rejito, The Hybrid Modeling of Spatial Autoregressive Exogenous Using Casetti's Model Approach for the Prediction of Rainfall, Mathematics 11 (2023), 3783. https://doi.org/10.3390/math11173783.

[21] K.D. Lawrence, R.K. Klimberg, S.M. Lawrence, Fundamentals of Forecasting Using Excel, Industrial Press, New York, 2009.

[22] L.D.F. Ramalho, W.R.D.C. Segundo, R-Shiny as an Interface for Data Visualization and Data Analysis on the Brazilian Digital Library of Theses and Dissertations (BDTD), Publications 8 (2020), 24. https://doi.org/10.3390/publications8020024.