Limits: Journal of Mathematics and Its Applications
E-ISSN: 2579-8936
P-ISSN: 1829-605X
Vol. 22, No. 1, Maret 2025, 43-66
DOI: http://dx.doi.org/10.12962/limits.v22i1.3372

Variables that Influence Urban Sprawl in DKI Jakarta, West Java and Banten Provinces in 2020

Azzahra Dhisa Khamila [1] *, Timbang Sirait [2]
[1,2]Jalan Otto Iskandardinata No. 64C, Jakarta Timur, DKI Jakarta, 13330 Indonesia
[1,2]Program Studi D-IV Statistika, Politeknik Statistika STIS
e-mail: [1]212011481@stis.ac.id, [2]timbang@stis.ac.id

Diajukan: 28 Maret 2024, Diperbaiki: 13 Desember 2024, Diterima: 13 Januari 2025

Abstrak

Provinsi DKI Jakarta, Jawa Barat dan Banten menjadi rumah bagi dua kawasan metropolitan besar di Indonesia yang saling terhubung. Akibatnya daerah tersebut memiliki tingkat urbanisasi yang tinggi yang bisa mengarah pada terjadinya urban sprawl. Urban Sprawl dapat menimbulkan berbagai dampak negatif utamanya terhadap lingkungan. Untuk itu perlu meminimalisir urban sprawl yang salah satu caranya adalah dengan menganalisis variabel-variabel yang memengaruhi urban sprawl. Beberapa penelitian mengenai analisis spasial urban sprawl telah banyak memanfaatkan data citra satelit, salah satunya menyatakan bahwa NDBI dapat menangkap pola, karakteristik, serta hal yang memengaruhi urban sprawl. Namun, penelitian yang memanfaatkan NDBI sebagai variabel pendekatan urban sprawl belum pernah dilakukan di Indonesia. Oleh karena itu, penelitian ini dilakukan dengan tujuan untuk menganalisis pengaruh variabel-variabel yang diduga memengaruhi urban sprawl di Provinsi DKI Jakarta, Jawa Barat dan Banten menggunakan analisis spasial. Hasilnya menunjukkan rata-rata nilai NDBI yang tinggi berada di wilayah perkotaan yang mayoritas berada di Provinsi DKI Jakarta. Variabel yang berpengaruh secara signifikan terhadap urban sprawl adalah persentase penduduk migran dan PDRB sektor tersier. Dengan berfokus pada variabel-variabel tersebut, pemerintah dapat mengambil kebijakan untuk meminimalisir dan mengendalikan urban sprawl yang terjadi di wilayahnya.

Kata Kunci: NDBI, Urban Sprawl, Analisis Spasial, Spatial Error Model

Abstract

DKI Jakarta, West Java and Banten provinces are the place of two large metropolitan areas in Indonesia that are interconnected. As a result, these areas have a high level of urbanization which can lead to urban sprawl. Urban Sprawl can cause various negative impacts, especially on the environment. Therefore, it is necessary to minimize urban sprawl, one of many ways is by analyzing the variables that affect urban sprawl. Several studies on spatial analysis of urban sprawl have been made extensively using satellite imagery data, one of them states that NDBI can capture patterns, characteristics and the causes of urban sprawl. However, research that utilizes NDBI as a variable approach for the urban sprawl has never been conducted in Indonesia. Therefore, this research was conducted with the aim of analyzing the effect of variables that indicated influence urban sprawl in the provinces of DKI Jakarta, West Java and Banten using spatial analysis. The results show that the average NDBI value is high in urban areas where the majority are in DKI Jakarta province. The variables that significantly influence urban sprawl are percentage of migrant population and tertiary sector of GRDP. By focusing on these variables, the government can make policies to minimize and control urban sprawl that occur in their area.

***Keywords**: NDBI, Urban Sprawl, Spatial Analysis, Spatial Error Model*

1 Introduction

Indonesia is the country with the fastest urban growth compared to cities in other Asian countries, with average urban growth reached 4.1 percent per year. In 2005, the percentage of urban population in Indonesia was 47.9 percent. The Ministry of Public Works and Public Housing estimates that by 2025 the number of urban residents in Indonesia will increase to 60 percent.

There are 10 metropolitan areas that are prioritized for accelerated development by the Regional Infrastructure Development Agency of the Ministry of Public Works and Public Housing. DKI Jakarta, West Java and Banten provinces are the place of two large metropolitan areas in Indonesia that are interconnected, namely Jakarta Metropolitan Area (Jabodetabek-Punjur) and Bandung Metropolitan Area (Greater Bandung or Bandung Basin), which are called Jakarta-Bandung Mega-Urban Region (JBMUR) [1]. Based on Indonesian Government Regulation no. 13 of 2017 concerning Amendments to Government Regulation no. 26 of 2008 concerning National Regional Spatial Planning, the two metropolitan areas are included in 76 National Strategic Areas (KSN) as economic growth areas. Apart from that, there are several new city developments in West Java and Banten Provinces. Based on BAPPEDA of West Java statement, since October 2015 the West Java Provincial government has formed three metropolitan areas, including Greater Bandung or Bandung Basin, Bodebekarpur, and Rebana or Greater Cirebon. Meanwhile in Banten Province, there is a metropolitan area called Serang Raya. Residential infrastructure development has also been carried out intensively for the construction of the New Maja Public City in Lebak Regency of Banten as an embodiment of the 2015-2019 RPJMN.

With the large number of urban areas, these three provinces have a high urban tourism attraction so that urbanization become more massive. The increase in population and economic activities will increase the demand for wider urban space, while the available urban space is fixed and limited [2]. This causes city development starting to spread in all directions in an unplanned manner, moving away from the city center towards the fringes of the city and even areas that had not yet been developed. As a result, there is an expansion of suburbanization, which then becomes a satellite city or buffer for the city center. This uncontrolled and unplanned expansion of city which mainly causes conversion of agricultural land and forest known as urban sprawl [3]. Another explanation, Bhatta [4] in his book describes the concept of sprawl as the emergence of unauthorized and unplanned development, usually in suburban areas or along main roads or urban fringes, especially the haphazard construction of housing, commercial, industrial areas or other inappropriate land uses, and progressively until the density is low. Further explanation by The

European Environment Agency (EEA), urban sprawl refers to the physical expansion of urban areas that occurs primarily from major cities into the surrounding vegetation lands.

There are various impacts that can be caused by massive urban sprawl, such as increasing the cost of living, traffic jams, reducing the amount of agricultural land, increasing water and air pollution, depleting groundwater resources, climate change, increasing city temperatures, loss of biodiversity, ecological disruption, and increased waste production [5]. This is proven by the BPS report which notes that in DKI Jakarta, the majority of changes in the function of rice fields were due to the construction of factories, housing and Covid-19 cemeteries [6]. Apart from that, changes in agricultural to non-agricultural land use, especially for industry and residential areas in Banten are among the highest compared to other provinces in Indonesia [7]. Also, there is a decrease in the area of agricultural land due to the very high intensity of development in the non-agricultural sector, such as residential and industrial areas [8]. This makes the provinces of West Java and Banten are among the 8 provinces experiencing a reduction in the area of raw rice fields [9].

According to the OECD, the main step in controlling land use in order to realize Sustainable Urban Development is to minimize urban sprawl. In order to minimize urban sprawl, it is necessary to know the things that influence it. According to The European Environment Agency (EEA) [10], several important factors that encourage urban sprawl include demographic, socioeconomic, political, technological and geophysical factors. However, basically the factors that cause urban sprawl vary depending on the level of development of a country or the structure of society [11]. Zhang & Pan [12] concluded that there are three main factors that influence urban sprawl, which are demographic, economic and government regulation factors. They stated that tertiary sector GDP, GRDP, and investment in real estate development, which are representating the economic factors contributed the greatest influence to urban sprawl. Meanwhile, demographic factor, which is represented by the total population at the end of the year, have the smallest influence compared to other variables used in the research. Another research conducted by Li & Li [13] which examined the spatiotemporal pattern of urban sprawl in China during the period 2006 to 2014 and identified the driving socioeconomic factors, stated that the growth rate of population density, the growth rate of GDP per capita, the percentage of secondary sector GDP, and the percentage of tertiary sector GDP has a significant positive effect on urban sprawl. In this research, it was stated that demographic factors contribute the greatest influence to urban sprawl, which is shown by a 1 percent increase in population density growth which will increase the Urban Sprawl Rate (USR) by 0.68%.

From a demographic perspective, it is known that population size and structure influence the size of the built-up area. This is because the larger the population, the more space is needed to accommodate everyone. However, the pattern that occurs in cities in developed countries in Europe shows that population growth continues to decline while the urban sprawl continues to increase. The decline in population growth is caused by increasingly severe environmental degradation that encourage residents looking for other better areas to live. Bontje [14] stated that large-scale migration then became the main cause of urban sprawl.

From an economic perspective, the total population output at the district/city level can be reflected in the Gross Regional Domestic Product (GRDP). Increasing GRDP in the tertiary sector is essential for economic development, providing employment opportunities and increasing population income. An increase in the tertiary sector in a region can lead to labor market agglomeration, which leads to high urban expansion [15]. Apart from the tertiary sector, the secondary sector also has a large influence on economic development. However, the negative impact of a high secondary sector in an area is the high rate of conversion of land, especially agriculture, into industrial and manufacturing land [16].

In terms of government regulations, the development and provision of communication and transportation facilities increases the population's dependence on private and public transportation. The increase in the number of vehicles is accompanied by an increase in residential development in suburban areas as well as environmental degradation due to high fuel consumption and the resulting pollution [17].

Research on urban sprawl generally involves spatial analysis using satellite imagery data because it can accurately reflect the spatial distribution of the economic and social conditions of society [18]. Various metrics have been developed as approaches to measure urban sprawl, such as using the percentage of built-up land [19], the sprawl index calculated based on the household ratio and building ratio [20], the Urban Sprawl Index (USI) derived from the area ratio and population ratio [13], among others. Additionally, there is a single indices that can reflect urban sprawl namely Normalized Difference Built-Up Index (NDBI). NDBI is an index that focuses on highlighting built-up areas which are usually of high in urban areas. NDBI is able to capture massive land changes caused by the emergence of many new cities or the conversion of green land into low-density built-up areas, which refers to urban sprawl [21]. Apart from that, NDBI can also capture urban sprawl patterns, characteristics of urban sprawl, as well as matters that related to the causes of urban sprawl [22]. Karanam [23] also highlighted that the NDBI method can produce land development maps with an accuracy exceeding 90 percent. This indicates that the NDBI metric, as a single indicator, can be utilized to measure urban sprawl.

Unfortunately, research on variables that influence urban sprawl using NDBI as an approach variable for urban sprawl has never been carried out in Indonesia, especially the three provinces with two large metropolitan areas, which are DKI Jakarta, West Java and Banten. This research aims to provide a comprehensive overview of the variables that potentially influence urban sprawl and to analyze their impact on urban sprawl in the provinces of DKI Jakarta, West Java and Banten through spatial analysis. The results of this research can be used as a reference for stakeholders in forming policies aimed at minimizing the incidence and impact of urban sprawl.

2 Research Method

2.1 Data, Tools, and Analysis Method

This research covers districts and cities in DKI Jakarta, West Java and Banten provinces, with a total of 40 districts and cities after excluding Kepulauan Seribu because it consists of many small islands separated from the main land area of Jakarta. The study utilizes multisource data collected in 2020, considering the availability of the latest data for all variables. The data collected includes NDBI, percentage of migrant population, secondary sector GRDP, tertiary sector GRDP, and percentage of vehicles.

NDBI is calculated using the reflectance ratio of the Shortwave Infrared (SWIR) and Near-Infrared (NIR) band channels following this equation.

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \qquad (1)$$

By using the Sentinel-2 satellite, the SWIR used is Band 11 while the NIR used is Band 8. NDBI data is collected and processed via Google Earth Engine (GEE) which can be exported in the form of raster data in Tag Image File Format (TIFF), then the average value was calculated with Zonal Statistics using Quantum GIS (QGIS) software. Zonal Statistics is operation to calculate statistics from raster cell values in a zone defined by another dataset [24]. The tool can calculate one or several statistics with the output is a single value from each zone in tabular form. The components required to perform zonal statistics are the input layer / zone layer and the value layer / raster layer . The input layer or zone layer is a layer that defines the shape, value and location of the zone. Meanwhile, the value layer or raster layer refers to raster data for which statistical values will be calculated. In this study, the input layer comprises vector data in shapefile format, representing administrative boundaries for each district and city within the provinces of DKI Jakarta, West Java, and Banten. On the other hand, the raster layer consists of raster data derived from NDBI, extracted using GEE and presented in TIFF format. The zonal statistics were used to calculate the average NDBI value for each district/city within the observation area. Once the average NDBI values were obtained, the data was exported in two formats, which are Comma Separated Value (CSV) and ESRI shapefile (SHP), to facilitate further analysis.

Migrant percentage data is the percentage of lifetime migrant population from the total population in a district/city obtained through the BPS publication of Migration Statistics of the 2020 Population Census Long Form from each district/city in the three provinces. Secondary sector GDP is the total gross regional domestic income (GRDP) in a district/city, especially in the secondary sector, which is the industrial sector that processes raw materials into goods that can be reprocessed. Secondary sector DPRB data is obtained through the official BPS website in each district/city by adding up the GRDP values in KBLI categories C to F. Meanwhile, tertiary sector GDP is the total gross regional domestic income (GRDP) in a district/city, especially in the tertiary sector, which is industrial sector whose products are in the form of services. Tertiary sector DPRB data is obtained through the official BPS website in each district/city by adding up the GRDP values in KBLI categories G to R,S,T,U. The percentage of vehicles per district/city is obtained by dividing the number of vehicles in each district/city by the number of vehicles in each corresponding province. The number of vehicles data is taken from the KORLANTAS POLRI.

In this research, several software were used to assist data processing are Q-GIS 3.28.3, Geoda 1.20.0.22 and R-Studio 4.2.1, adjusting the availability and features of each software to support descriptive and inferential analysis. Descriptive analysis aims to find out a comprehensive overview of the data used. To fulfill this objective, the researcher created a thematic map to see the grouping of districts/cities based on the value of each variable, which can be an indication that there is a spatial problem. Thematic maps were created using Q-GIS. In forming the thematic map, the formation of class intervals is carried out using the natural breaks method with a total of 5 classes, namely very low, low, medium, high and very high. The spatial distribution pattern of dependent variable will be checked through global spatial autocorrelation values (Global Moran’s I) and Local Spatial Autocorrelation Values (LISA). Meanwhile, inferential analysis aims to test the relationship between variables by seeing whether there is an influence of the independent variable on the dependent variable. To fulfill these objectives, researchers carried out the following series of stages .

1. Form a multiple linear regression with NDBI as the dependent variable, and percentage of migrant population, secondary sector GRDP, tertiary sector GRDP, and percentage of vehicles as independent variables. To overcome the large differences in unit size, a natural logarithm transformation process was carried out on all independent variables that were not in percentage form, so that the multiple linear regression equation formed was as follows.

$$NDBI_i = \beta_0 + \beta_1 Migrant_i + \beta_2 \ln(GRDPsecondary_i) + \beta_3 \ln(GRDPtertiary_i) + \beta_4 Vehicle_i + \varepsilon_i \quad (2)$$

Where :

$NDBI_i$	= average NDBI of district/city i
$Migrant_i$	= percentage of migrant population in district/city i
$GRDPsecondary_i$	= GDP of the secondary sector of district/city i
$GRDPtertiary_i$	= GRDP of the tertiary sector of district/city i
$Vehicle_i$	= percentage of vehicles in district/city i
ε_i	= error term i

2. Testing the classic assumptions of OLS regression includes the normality test, homoscedasticity test, non-multicollinearity test, and identification of spatial autocorrelation by observing the Moran's I value.
3. Selecting a spatial regression model using the Lagrange Multiplier test (LM-Test) and Robust Lagrange Multiplier. According to Anselin [25], the spatial regression model selection procedure follows the following chart.

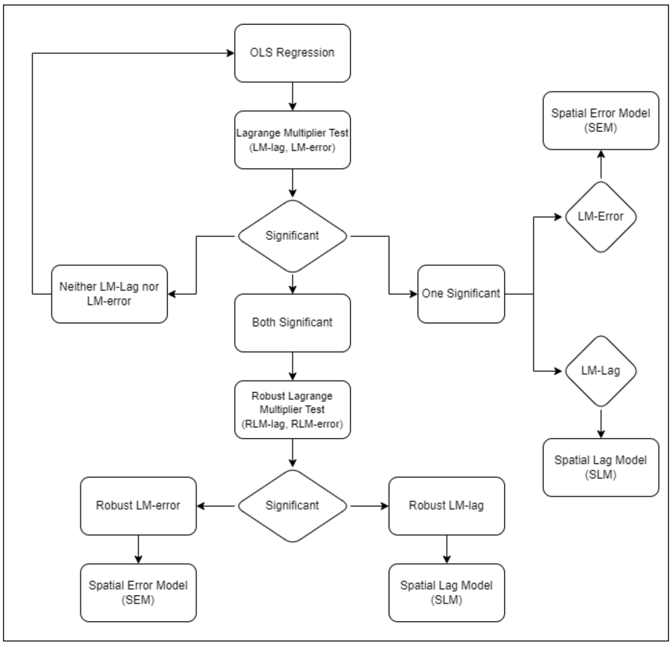


Figure 1. Spatial Analysis Flowchart

Based on the Figure 1, after carrying out OLS regression modeling, it is necessary to carry out Lagrange Multiplier (LM) testing to determine whether the spatial regression model is suitable for use. Through the LM test results, a statistical comparison of LM-lag and LM-error is carried out. If both LM-lag and LM-Error values are not significant, then the model used is OLS regression. If one of them is significant, then the model used is a spatial regression of significant LM statistics. However, if both are significant, then continue with the Robust LM test.

In practice, Robust LM will be significant in one of them, Robust LM-lag or Robust LM-error. In these conditions, the model used is a spatial regression of the Robust LM statistics which is significant.

4. Carry out parameter significance tests both simultaneously and partially, and continue with interpretation of the modeling results.
5. Compare the multiple linear regression model with the spatial regression model formed using the R-Square and AIC criteria.

2.2 Multiple Linear Regression

Regression analysis is a statistical tool that studies the relationship between two or more quantitative variables through a mathematical equation so that a variable can be predicted based on one or more other variables [26]. If a form of regression analysis uses more than one independent variable, it is called multiple linear regression. In the modeling, two variables are used, which are the independent variable or predictor variable and the dependent variable or response variable. An independent variable or predictor variable is a variable whose value is determined independently or a variable that is thought to influence the dependent variable. The general form of RLB can be written as follows.

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{3}$$

Where :

$\boldsymbol{Y}$ = vector of dependent variables

$\boldsymbol{X}$ = matrix of independent variables

$\boldsymbol{\beta}$ = vector of regression parameter coefficients

$\boldsymbol{\varepsilon}$ = error vector where $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I})$

In testing the significance of parameters, simultaneous and partial tests are carried out. The simultaneous test was carried out using the F test statistic with the following statistical hypothesis.

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_1 : at\ least\ there\ is\ one\ \beta_k \neq 0\ \ ; k = 1, \dots, p-1$$

Where k is number of variables and p is number of parameters. The F test statistics formula is as follows.

$$F^* = \frac{MSR}{MSE} = \frac{\frac{SSR}{p-1}}{\frac{SSE}{n-p}} \tag{4}$$

H_0 rejected if $F^* > F_{1-\alpha;p-1,n-p}$ where n is number of observations, or if the p-value is smaller than the significance level α used, which in this study is 5 percent.

In addition, a partial test was carried out using the t test statistic with the following statistical hypothesis.

$$H_0 : \beta_k = 0$$
$$H_1 : \beta_k \neq 0$$

With the t test formula as follows.

$$t^* = \frac{b_k}{se(b_k)} \tag{5}$$

H_0 rejected if $|t^*| > |t_{1-\frac{\alpha}{2};n-p}|$ or if the p-value is smaller than the significance level α used, which in this study is 5 percent.

In multiple linear regression, several assumptions must be met in order to produce accurate, unbiased and consistent estimates. These assumptions include normality, homoscedasticity, non-multicollinearity, and non-autocorrelation.

a. Normality

The normality test can be carried out in various ways, one of them is the Jarque-Bera normality test. The Jarque-Bera normality test utilizes calculations of skewness and kurtosis. The data will meet the assumption of normality if the skewness (S) value is 0 and the kurtosis (K) value is 3. The hypothesis used is as follows.

$$H_0 : S = 0 \; dan \; K = 3$$
$$H_1 : S \neq 0 \; dan \; K \neq 3$$

With Jarque-Bera normality test statistics as follows.

$$JB = n\left[\frac{\hat{S}^2}{6} + \frac{\left(\hat{K} - 3\right)^2}{24}\right] \tag{6}$$

Information :

n = number of observations

$\hat{S}$ = skewness scale

$\hat{K}$ = kurtosis scale

H_0 rejected if $JB > \chi^2_{\alpha;2}$ or if the p-value is smaller than the significance level α used, which in this study is 5 percent.

b. Homoscedasticity

One of the most widely used classical methods is the Breusch-Pagan test. In the $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i$ regression model, will be tested whether there is a linear relationship between the squared error ε_i with the independent variables. Therefore, the second equation

$(\varepsilon_i^2 = \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_p X_{pi} + u_i)$ is formed with the hypothesis $H_0\colon \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$. The test statistic used is as follows.

$$BP = \frac{1}{2} f'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'f \tag{7}$$

$$f_i = \left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right) \tag{8}$$

H_0 rejected if $BP > \chi^2_{\alpha,k}$, or if the p-value is smaller than the significance level α used, which in this study is 5 percent.

c. Non Multicollinearity

To test for multicollinearity, it can be done by calculating the Variance Inflation Factor (VIF). VIF is used to measure how much the estimated variance of the regression coefficient increases if the independent variables in the model are interconnected. The VIF calculation is as follows.

$$VIF_k = \frac{1}{1 - R_k^2} = \frac{1}{Tolerance} \tag{9}$$

Tolerance is also known as the inverse of VIF. The lower the tolerance, the greater the possibility of multicollinearity. A VIF with a value of more than 10 indicates that the regression coefficient is estimated to be weak due to high multicollinearity in the independent variables.

d. Non Autocorrelation

Autocorrelation commonly occurs in time series data, but basically it can also occur in cross-sectional data, namely spatial autocorrelation. Spatial autocorrelation is a condition when the value of a variable in one area has the same value as another adjacent area (locational similarity) [27]. As Tobler's first law of geography states that "Everything is related to everything else, but near things are more related than distant things", meaning that spatial dependence needs to be considered. One way to test whether there is a presence of spatial autocorrelation is using Moran’s I Test. The statistical hypothesis used in Moran's I Test is as follows:

$$H_0 : E(I) = 0$$

$$H_1 : E(I) \neq 0$$

The test statistics used are as follows.

$$Z^* = \frac{I - E(I)}{\sqrt{V(I)}} \tag{10}$$

With

$$I = \frac{N}{S_o}\left(\frac{\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}\right) \quad (11)$$

$$E(I) = \frac{tr(\boldsymbol{MW})}{n-k} \quad (12)$$

$$V(I) = \frac{tr(\boldsymbol{MWMW}') + tr(\boldsymbol{MW})^2 + \{tr(\boldsymbol{MW})\}^2}{(n-k)(n-k+2)} - [E(I)]^2 \quad (13)$$

Where

$\boldsymbol{\varepsilon}$ = vector of residuals from OLS regression

$\boldsymbol{W}$ = spatial weighting matrix

n = number of observations

k = number of independent variables

S_o = number of spatial weights ($\sum_i \sum_j w_{ij}$)

$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

H_0 will be rejected if $|Z^*| > Z_{\frac{\alpha}{2}}$ or if the p-value is less than the significance level. In identifying spatial autocorrelation, it is necessary to define "neighbors", which refers to other locations that have the potential to experience interaction with the location of observation. To address this, a measure known as the spatial weighting matrix is constructed to define the neighboring relationships for each observation.

Traditional approaches that rely on the spatial arrangement of each observation identify "neighbors" as locations that are related to the observed location based on the existence of intersecting boundaries (first-order contiguity) or based on a certain distance between observations (distance-based contiguity). In this research, we use the distance-based contiguity, using the assumption that short distances between observations will give rise to greater spatial interactions compared to longer distances between observations due to increased geographic barriers. One of several method that can be used to determine "neighbors" using distance is distance to all neighbors which assumes that all observations in an area are neighbors and have an influence on other areas.

2.3 Lagrange Multiplier Test

Once the presence of spatial autocorrelation in the model residuals is identified, the subsequent step is conduct the Lagrange Multiplier Test to specify spatial alternative model, whether a model with spatial lag dependence or a model with spatial error dependence. Lagrange Multiplier Test for LM-lag uses the following statistical hypothesis.

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

The test statistics used are as follows.

$$LM_\rho = \frac{\left(\frac{\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y}}{\sigma^2}\right)^2}{D+T} \tag{14}$$

With

$D = \frac{(\boldsymbol{WX\beta})'\boldsymbol{M}(\boldsymbol{WX\beta})}{\sigma^2}$

$T = tr[(\boldsymbol{W}' + \boldsymbol{W})\boldsymbol{W}]$

$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

H_0 rejected if $LM_\rho > \chi_1^2$ or if the p-value is smaller than the significance level. Meanwhile, the Lagrange Multiplier test for LM-Error uses the following statistical hypothesis.

$$H_0 : \lambda = 0$$
$$H_1 : \lambda \neq 0$$

The test statistics used are as follows.

$$LM_\lambda = \frac{\left(\frac{\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{\varepsilon}}{\sigma^2}\right)^2}{T} \tag{15}$$

H_0 rejected if $LM_\lambda > \chi_1^2$ or if the p-value is smaller than the significance level.

2.4 Robust Lagrange Multiplier Test

Robust testing is carried out if the LM-Lag and LM-Error tests both produce significant results. Robust LM testing for spatial lag follows the following formula.

$$RLM_\rho = \frac{\left(\frac{\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y} - \boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{\varepsilon}}{\sigma^2}\right)^2}{D} \tag{16}$$

Meanwhile, the Robust LM test for spatial error follows the following formula.

$$RLM_\lambda = \frac{D + T\left(\frac{\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{\varepsilon}}{\sigma^2} - \left(\frac{T}{D+T}\right)\frac{\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y}}{\sigma^2}\right)^2}{DT} \tag{17}$$

Both Robust LM tests for spatial lag and error will have a reject decision H_0 if the test statistical value is greater than χ_1^2 or has a p-value that is smaller than the significance level.

2.5 Spatial Regression

Spatial regression is an expanded form of simple regression, where the unit of observation is an area that includes spatial effects in measuring the relationship that occurs between the independent and dependent variables. The existence of spatial effects in the form of spatial dependence can be included in the linear regression model in two ways, which are the spatial lag model and the spatial error model. The basic spatial lag model or Spatial Autoregressive Model (SAR) has the following form.

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{18}$$

By $\boldsymbol{y}$ is an $n \times 1$ vector of the observed dependent variable, $\boldsymbol{W}\boldsymbol{y}$ denotes the spatially lagged dependent variable for the spatial weighting matrix $\boldsymbol{W}$. $\boldsymbol{X}$ is a $n \times k$ matrix of observations for each independent variable, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of error, and $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients. ρ is a parameter that will explain the strength of the spatial relationship y_i between $\sum_j W_{ij} y_j$. ρ is in the range [0,1) and $\rho = 0$ is a linear regression model.

Meanwhile, the spatial error model is a model that takes into account spatial dependence in the model error so that errors from different areas show spatial covariance. The spatial error model (SEM) has the following form.

$$\begin{gathered} \boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} = \lambda \boldsymbol{W} \boldsymbol{\varepsilon} + \boldsymbol{u} \end{gathered} \tag{19}$$

With λ is the spatial autoregression coefficient on the error and $\boldsymbol{u}$ is an uncorrelated and homoscedastic error. Reliance on spatial error is interpreted as the presence of noise that reflects spatial autocorrelation in model error or variables that are not included in the model.

2.6 Best Model Selection

The best model selection carried out using two criterias, which are R-square and AIC. To calculate the coefficient of determination or R-square, following the formula:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{20}$$

which has a value $0 \leq R^2 \leq 1$, if it is close to 1, the regression model used is more precise or better. Apart from that, the AIC value can also be calculated using a formula:

$$AIC_p = n \ln SSE_p - n \ln n + 2p \tag{21}$$

where the smaller the AIC value, the simpler the model used.

3 Result and Discussion

3.1 Descriptive Analysis

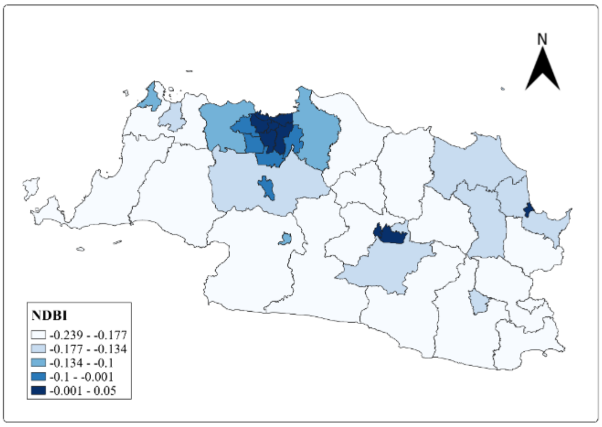
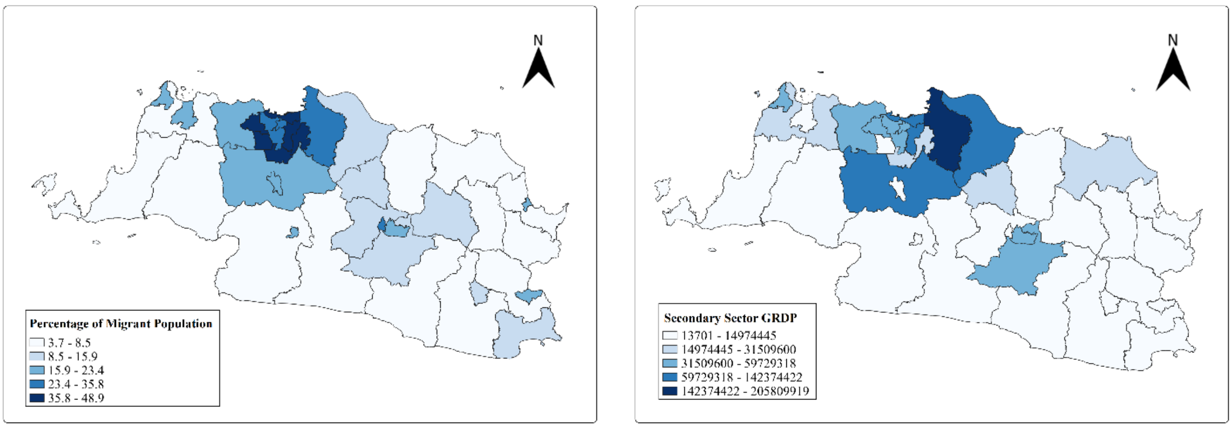
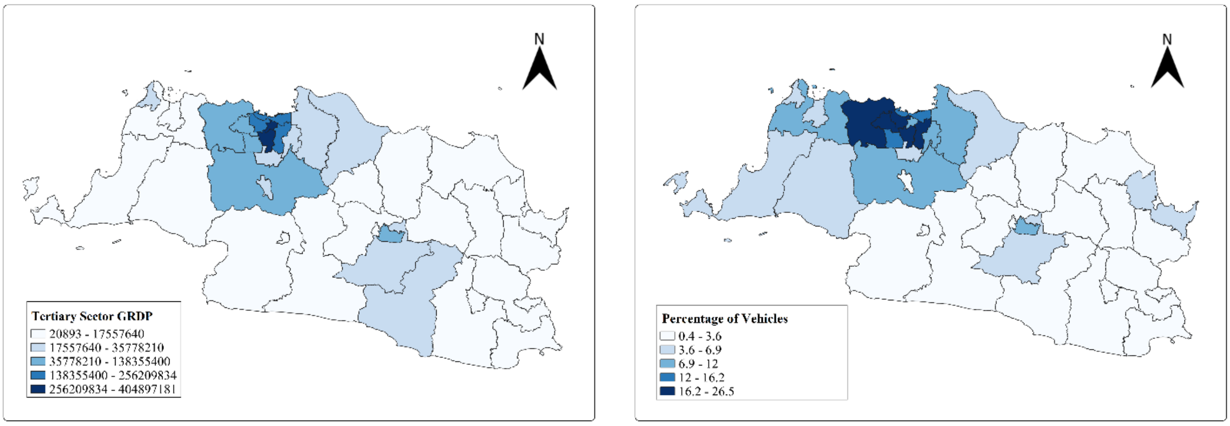


Figure 1. Map of NDBI distribution for Banten Province, DKI Jakarta and West Java in 2020

Figure 2 shows the average NDBI per district/city in Banten, DKI Jakarta and West Java Provinces. The darker the color of the area visible on the map, the higher the average NDBI value in that area. Conversely, the lighter the color of the area visible on the map, the lower the average NDBI value in that area. Groups of regions that are classified as having very high average NDBI values as indicated by the darkest colored areas on the map include North Jakarta, West Jakarta, Central Jakarta, East Jakarta, South Jakarta, Cimahi City, Bandung City and Cirebon City. Meanwhile, the regional group is classified as having a very low average NDBI value, which is indicated by quite a lot of light colored areas. The highest average NDBI value is in Bandung City, namely 0.05, while the lowest average NDBI value is in Pandeglang Regency, namely -0.239.



(a) (b)



(c) (d)

Figure 2. Map of migrant population (a), secondary sector GRDP (b), tertiary sector GRDP (c), and percentage of vehicles (d) distribution for DKI Jakarta, West Java and Banten Provinces in 2020

From the Figure 3, the darker the color visible on the map means the higher value of each variable in the area. On the other hand, the lighter the color on the map, the lower value of each variable in the area. The map (a) shows that the areas with the darkest color or those with a very high percentage of migrant residents, around 35.8 to 48.9 percent include North Jakarta, East Jakarta, Depok City, South Tangerang City, Tangerang City and Bekasi City. These cities also have an average NDBI value that is classified as very high. Meanwhile, Cimahi City, Bandung City and Cirebon City, which are included in the group that has a very high average NDBI value, are included in the group that has a high percentage of migrant residents. In addition, the majority of urban areas in the three provinces are classified into groups with high and medium percentages of migrant populations. On the other hand, there are quite a lot of light colored areas or those classified as areas with a low and very low percentage of migrant population, the majority of which are district administrative areas.

From map (b) can be seen that the area with the darkest color or which has a very high secondary sector GDP, namely 205.81 trillion rupiah, is Bekasi Regency. Apart from that, surrounding areas such as North Jakarta City, East Jakarta City, Bogor Regency and Karawang Regency are included in the group of areas with high secondary sector GRDP, around 59.73 to 142.37 trillion rupiah. Several regions that are classified as having high secondary sector GRDP are classified as regions with medium, low and very low average NDBI values.

From map (c) can be seen that the areas with the darkest color or those with very high tertiary sector GRDP, namely around 256.21 to 404.90 trillion rupiah include Central Jakarta City and South Jakarta City. Apart from that, surrounding cities such as North Jakarta City, East Jakarta City and West Jakarta City are included in the group of regions with high tertiary sector GRDP, around 138.35 to 256.21 trillion rupiah. The grouping of regions with tertiary sector GRDP and NDBI values tends to be the same, in other words, the majority of regions classified as having high tertiary sector GDP are also included in the group of regions classified as having high average NDBI values.

From map (d) can be seen that the areas with the darkest color or those with a very high percentage of vehicles, around 16.2 to 26.5 percent include Tangerang Regency, Tangerang City, West Jakarta City, South Jakarta City and East Jakarta City. Meanwhile, North Jakarta City and South Tangerang City are included in the group of areas with a high percentage of vehicles, around 12 to 16.2 percent. The grouping of regions with the percentage of vehicles and NDBI tends to be the same, in other words, the majority of regions that are classified as having a high percentage of vehicles are also included in the group of regions that are classified as having a high average NDBI value.

To identify spatial autocorrelation of the average NDBI value per district/city, Exploratory Spatial Data Analysis (ESDA) or a spatial data analysis method can be used which can capture interactions between adjacent regions through global spatial autocorrelation values (Global Moran's I) and autocorrelation values. local spatial (Local Indicator of Spatial Association (LISA)).

Table 1. Global Spatial Autocorrelation Value

No.	Variable	Coefficient
1	Global Moran's I (I)	0.3269
2	E[I] (I_0)	-0.0256
3	p-value*	0.0020

Note: *) The p-value was obtained by randomization with 999 permutations

Based on the results seen in Table 1, the Global Moran's I value of the average NDBI is 0.3269, which means there is positive spatial autocorrelation of the average NDBI variable. Positive autocorrelation indicates that the average NDBI value for neighboring districts/cities is more related or will be similar compared to districts/cities that are far apart. There are two ways to check the significance of Global Moran's I, namely by comparing the I value with I_0 and by looking at the p-value . Based on Table 1, it can be seen that the I value < I_0, which means the I value is significant. Apart from that, it can be seen that the p-value is 0.0020, which with a significance level of 5% ($\alpha = 0.05$) produces the same decision, namely the I value is significant because the p-value < α.

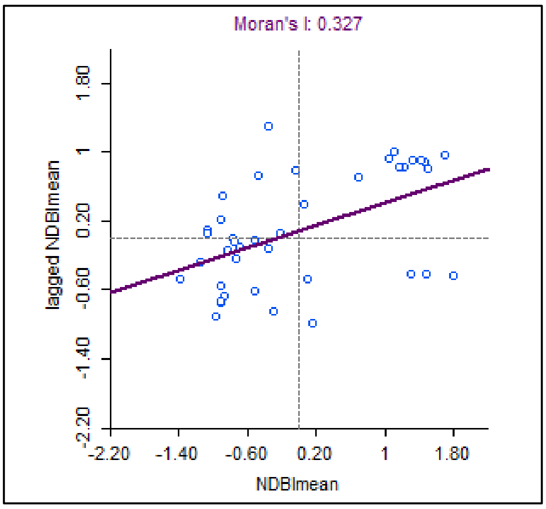


Figure 3. Moran Scatterplot Mean NDBI Values

The process of identifying spatial autocorrelation is continued by forming a Moran Scatterplot from the average NDBI values. Based on Figure 4, it can be seen that the majority of district/city observations are in quadrants 1 and 3 which are hotspot (High-High) and coldspot (Low-Low) areas. This means that districts/cities with high average NDBI values are surrounded by regions with districts/cities that also have high average NDBI values. On the other hand, districts/cities with low average NDBI values are surrounded by regions with districts/cities that also have low average NDBI values.

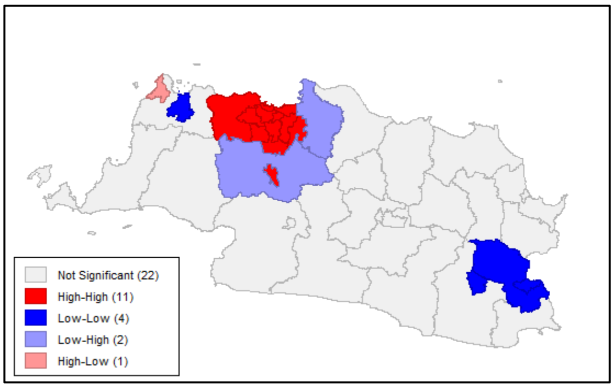


Figure 4. LISA Cluster Map Average NDBI Value

Identification of spatial autocorrelation can also be done by measuring the Local Indicator of Spatial Association (LISA). Figure 5 shows a clustering map of the average NDBI values in the Provinces of Banten, DKI Jakarta and West Java. As seen in the legend, areas colored dark red are districts/cities that are included in the High-High cluster. The average NDBI value in districts/cities belonging to this cluster is high due to the high average NDBI of their neighbors. There are 11 regions included in this cluster, including Tangerang Regency, Tangerang City, South Tangerang City, Depok City, West Jakarta City, North Jakarta City, Central Jakarta City, South

Jakarta City, East Jakarta City, Bekasi City, and Bogor City. On the other hand, areas colored dark blue are districts/cities that are included in the Low-Low cluster. The average NDBI value in districts/cities belonging to this cluster is low due to the low average NDBI of their neighbors. There are 4 areas included in this cluster, including Serang City, Tasikmalaya City, Ciamis Regency, and Banjar City.

3.2 Inferential Analysis

3.2.1 Multiple Linear Regression Modelling

Through OLS regression modeling, the following results were obtained.

Table 2. Summary of OLS regression parameter estimation results

Variable	Coefficient	P-value
Intercept	-0.4221	0.0000*
Percent of Migrant Population	0.0052	0.0000*
Ln(Secondary Sector GRDP)	-0.0122	0.1121
Ln(Tertiary Sector GRDP)	0.0243	0.0115*
Percentage of Vehicles	-0.0003	0.8489

Note: *) significant at α= 0.05

Based on the results in Table 2, modeling the average NDBI value using OLS regression can be written in the following equation.

$$\widehat{NDBI_i} = -0.4221 + 0.0052\,Migrant_i^* - 0.0122\,Ln(GRDPsecondary_i) + 0.0243\,Ln(GRDPtertiary_i)^* - 0.0003\,Vehicle_i \quad ;i = 1,2,\dots,40 \tag{22}$$

Information :

$\widehat{NDBI_i}$ = Average NDBI value of district/city i

$Migrant_i$ = Percentage of migrant population in the district/city i

$Ln(GRDPsecondary_i)$ = GDP growth rate in the secondary sector of district/city i

$Ln(GRDPtertiary_i)$ = GDP growth rate in the tertiary sector of district/city i

$Vehicle_i$ = Percentage of vehicles in district/city i

The results of this modeling were then tested for parameter significance simultaneously and partially. Simultaneous tests were carried out using the F test while partial tests were carried out using the t test. Based on the processing results, the F test statistical value was 31.1143 with a p-value of 0.0000, resulting in a reject decision H_0 at a significance level of 5%. This means that all independent variables together significantly influence the dependent variable.

Referring to Table 2, if tested partially, there are 2 independent variables that significantly influence the average NDBI, namely the percentage of migrant population and the tertiary sector

GRDP growth rate. From the OLS regression equation, an R-square value of 0.7805 is obtained. This means that 78.05 percent of the variation in the average NDBI can be explained by the independent variables used in the model. Meanwhile, the remaining 21.95 percent is explained by other variables not included in the modeling. Next, an OLS regression assumption test was carried out with the results as follows.

a) Normality

From the Jarque-Bera test which produced a p-value of 0.05566 which resulted in a failure to reject decision H_0. This means that at a significance level of 5% the model residuals are normally distributed.

b) Non-multicollinearity

Testing the non-multicollinearity assumption was carried out by checking the VIF values for all independent variables. Based on Table 3, the VIF value for all variables is less than 10. This means that there is no correlation between the independent variables used in the model so that the non-multicollinearity assumption is met.

Table 3. Checking the VIF value of all independent variables

Variable	VIF
Percent of Migrant Population	1.8021
Ln(Secondary Sector GRDP)	3.6243
Ln(Tertiary Sector GRDP)	4.0969
Percentage of Vehicles	2.1099

c) Homoscedasticity

Testing the homoscedasticity assumption was carried out using the Breusch-Pagan test which produced a p-value of 0.60739 which resulted in a failure to reject decision H_0. This means that the model residuals have a constant variance so that the homoscedasticity assumption is met.

3.2.2 Spatial Autocorrelation Testing

Table 4. Checking spatial autocorrelation in model residuals

No.	Variable	Coefficient
1	Global Moran's I (I)	0.2901
2	p-value*	0.0000

Note: *) The p-value was obtained by randomization with 999 permutations

Based on Table 4, it can be seen that the Global Moran's I value formed from the residual model is 0.2901, which means there is positive spatial autocorrelation. To test its significance, it can be done by looking at the value p-value. From Table 4, it can be seen that the p-value is 0.0000, which is less than the 5% significance level (α = 0.05), resulting in a reject decision H_0, which means there is spatial autocorrelation between locations.

Through the results of this check, it can be concluded that there is spatial autocorrelation in the model residuals, so the next step is to carry out spatial regression modeling which is preceded by Lagrange Multiplier (LM) testing.

3.2.3 Spatial Regression Modelling

The first step in determining the spatial regression that will be used to model the average NDBI, it is necessary to carry out the Lagrange Multiplier Test (LM-Test), the results of which are as follows.

Table 5. Lagrange Multiplier Test (LM-Test) Results

Spatial Dependency Test	Mark	p-value	Decision
Lagrange Multiplier (Lag)	2.0512	0.1521	Failed to Reject H_0
Robust LM (Lag)	0.1195	0.7295	Failed to Reject H_0
Lagrange Multiplier (Error)	7.2176	0.0072	Reject H_0
Robust LM (Error)	5.2859	0.0215	Reject H_0

Based on Table 5, the results show that the LM-Error test is significant as indicated by a p-value that is smaller than the 5% significance level. Because the LM-error is significant, there is no need to continue with Robust LM testing. Referring to this result, the spatial regression model that can be used is the Spatial Error Model (SEM). The results of modeling the average NDBI with SEM are written in the following equation.

$$\begin{gathered}\widehat{NDBI}_i = -0.4021 + 0.0056\, Migrant_i^* - 0.0118\, Ln(GRDPsecondary_i) \\ +0.0217\, Ln(GRDPtertiary_i)^* + 0.0010\, Vehicle_i + \varepsilon_i^* \\ \varepsilon_i^* = 0{,}5142 \sum_{j=1, i\neq j}^{40} w_{ij}\varepsilon_j \end{gathered} \tag{23}$$

Where w_{ij} is the spatial weighting matrix between districts/cities i and j, ε_i is the spatial autocorrelation residual of the district/city i, and ε_j is the spatial autocorrelation residual of neighbor j of a district/city.

Referring to the SEM modeling results above, it can be seen that the weighing variable λ has a positive sign and is significant at the 5% significance level, which means that there is a correlation between the average NDBI value in one district/city and the error from other adjacent districts/cities. The errors in question are other variables outside the research that have a spatial influence on the average NDBI value in a district/city.

Variables that influence the average of NDBI significantly showed by the symbol (*), which are percentage of migrant population and tertiary sector of GRDP. The variable percentage of migrant population has a significant positive influence on the average NDBI value. This means that the higher the percentage of migrant population will increase the average NDBI value, which means it will increase urban sprawl. In line with research conducted by Hasnine & Rukhsana [28] which states that high migration will increase the need for urban land which causes illegal, unplanned and uncontrolled growth called urban sprawl. Over years, migration was viewed positively because of its impact on economic development. But in fact, Bekele [29] states that migration accounts for 40 percent of urban population growth, which is exceeded rates of job creation and service provision. It can causing many problem, especially for the land use.

The GRDP growth rate variable in the tertiary sector has a significant positive influence on the average NDBI value. This means that the higher the GDP growth rate, which means that urban sprawl will increase. This is proven by the increase in industry in the tertiary sector, which requires the agglomeration of human resources, will cause the expansion of urban land by encroaching on agricultural land, especially those close to urban areas [30].

3.2.4 Comparison of OLS Regression and Spatial Regression

Through a series of processes that have been carried out, two models of average NDBI per district/city were obtained, namely the OLS regression model and the Spatial Error Model (SEM) spatial regression model. To determine the best model, two criteria are used, the largest R-square value and the smallest Akaike Information Criterion (AIC) value. The comparison table for the best model criteria is as follows.

Table 6. Criteria for determining the best model

Model	AIC	R-square
OLS regression	-132.997	0.8087
Spatial Error Model (SEM)	-136.492	0.8315

Based on Table 6, it can be seen that the SEM model has a smaller AIC and a larger R-square compared to the OLS regression model. From these results it can be concluded that SEM is more appropriate to use in modeling the average NDBI value in DKI Jakarta, West Java and Banten Provinces compared to OLS regression.

4 Conclusion

Based on descriptive analysis, it is indicated that the average NDBI value is high in urban areas in the three provinces, especially in DKI Jakarta Province, because of its status as the nation's capital which makes it a global city and economic center in Indonesia so that it has a dense built-up area. Meanwhile, based on inferential analysis, the results showed that the variables that had a significant influence on urban sprawl were the percentage of migrant population and the size of tertiary sector GRDP. The percentage of migrant population has a significant positive influence on urban sprawl, which means that increasing the percentage of migrant population will increase urban sprawl. Likewise, GRDP in the tertiary sector has a significant positive influence on urban sprawl, which means that an increase in the GRDP value of the tertiary sector in an area will increase urban sprawl. Meanwhile, the other two variables, secondary sector GRDP and percentage of vehicles, do not have a significant influence on urban sprawl in the three provinces.

5 References

[1] T. Firman and I. A. I. Dharmapatni, "The Emergence of Extended Metropolitan Regions in Indonesia: Jabotabek and Bandung Metropolitan Area," *Review of Urban & Regional Development Studies, 7(2),* pp. 167-188, 1995.

[2] A. Indrayati, R. Rijanta, L. Muta'ali and R. Rachmawati, "Built-Up Area Changes, Spatial Pattern and Urban Sprawling in Kedungsepur Metropolitan Area," *International Journal of Sustainable Development and Planning Vol. 18 No. 8,* pp. pp. 2541-2546, 2023.

[3] G. Galster, R. Hanson, M. R. Ratcliffe, H. Wolman, S. Coleman and J. Freihage, "Wrestling Sprawl to the Ground: Defining and measuring an elusive concept," *Housing Policy Debate 12(4),* pp. 681-717, 2001.

[4] B. Bhatta, Analysis od urban growth and sprawl from remote sensing data, Springer Science & Business Media, 2010.

[5] C. S. Chhangani, "Urban Sprawl and Its Impact: A Comprehensive Review in Indian Context," *International Journal of Creative Research Thoughts (IJCRT) Volume 11,* pp. a907-a912, 2023.

[6] BPS, "Padi jakarta, Menjaga Produktivitas di Lahan Terbatas," BPS Provinsi DKI Jakarta, Jakarta, 2022.

[7] DLHK Banten, "Dokumen Informasi Kinerja Pengelolaan Lingkungan Hidup Daerah Tahun 2020," Dinas Lingkungan Hidup dan Kehutanan Provinsi Banten, Banten, 2020.

[8] PPID Jawa Barat, "Laporan Kinerja Instansi Pemerintah Tahun 2020," PPID Provinsi Jawa Barat, Bandung, 2020.

[9] DPR RI, "Alih Fungsi Lahan Sebabkan Krisis Pangan," 18 October 2023. [Online]. Available: https://www.dpr.go.id/berita/detail/id/47110/t/Alih%20Fungsi%20Lahan%20Sebabkan%20Krisis%20Pangan.

[10] European Environment Agency (EEA), "Urban Sprawl in Europe," Publications Office of the European Union , Luxembourg, 2016.

[11] Z. Karakayaci, "The Concept of Urban Sprawl and Its Causes," *The Journal of International Social Research,* pp. 815-818, 2016.

[12] X. Zhang and J. Pan, "Spatiotemporal Pattern and Driving Factors of Urban Sprawl in China," *Land 10 (11)*, p. 1275, 2021.

[13] G. Li and F. Li, "Urban Sprawl in China: Difference and socioeconomic drivers," *Science of the Total Environment,* pp. 367-377, 2019.

[14] M. Bontje, "Dealing with deconcentration: Population deconcentration and planning response in polynucleated urban regions in north-west Europe," *Urban Studies 38(4),* pp. 769-785, 2001.

[15] B. Paramasivam and I. Arumugavelu, "Urban Sprawl," in *Sustainability in Urban Planning and Design*, IntechOpen, 2020, p. Chapter 6.

[16] R. Ernan, "Alih Fungsi Lahan dalam Perspektif Lingkungan Perdesan," *Lokakarya Penyusunan Kebijakan dan Strategi Pengelolaan Lingkungan Kawasan Perdesaan,* pp. 10-11, 2001.

[17] S. H. Hosseini and M. Hajilou, "Drivers of urban sprawl in urban areas of Iran," *Papers in Regional Sciences,* pp. 1137-1158, 2019.

[18] J. Wang, S. Qu, K. Peng and Y. Feng, "Quantifying Urban Sprawl and Its Driving Force in China," *Discrete Dynamics in Nature and Society* , pp. 1-14, 2019.

[19] Fuadina, L. N., Rustiadi, E., & Pravitasari, A. E, “Analisis Faktor-Faktor yang Mempengaruhi Urban sprawl di Kawasan Cekungan Bandung,” *TATALOKA 23(1),* pp. 105-114, 2021.

[20] Sari, D. P., Wartaman, A. S., & Luru, M. N, “The characteristic of Urban sprawl in Bekasi City, Indonesia,” *IOP Conference Series: Earth and Environmental Science (Vol. 737, No. 1, p. 012029),* IOP Publishing, 2021.

[21] F. S. Zahra, M. Widanirmala and J. A. Syahbana, "Pengaruh Urban Sprawl Terhadap Kualitas Air Tanah di Daerah Bagian Utara Kota Semarang," *Jurnal Riptek Vol. 14 (I),* pp. 44-49, 2020.

[22] M. Y. Yasin, M. M. Yusof and N. M. Noor, "Urban Sprawl Assessment Using Time-Series LUCL and NDVI Variation: A Case Study of Sepang, Malaysia," *Applied Ecology & Environmental Research 17(3),* 2019.

[23] Karanam, H. K, “STUDY OF NORMALIZED DIFFERENCE BUILT-UP (NDBI) INDEX IN AUTOMATICALLY MAPPING URBAN AREAS FROM LANDSAT TM IMAGERY,” *International Journal of Scientific Research and Review, Volume 7, Issue 1*, pp. 1-8, 2018.

[24] ArcGIS Pro, "How zonal statistics tools work," 2023. [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-zonal-statistics-works.htm.

[25] L. Anselin, Exploring Spatial Data with GeoDaTM: a workbook, Center for spatially integrated social science, 1963, 157, 2005.

[26] M. H. Kutner, C. J. Nachtsheim and J. Neter, Applied Linear Regression Models 4th edition, Boston: Irwin Inc., 2004.

[27] L. Anselin and A. K. Bera, "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," in *Handbook of Applied Economic Statistics*, NY, CRC Press, 1998, pp. pp. 237-289.

[28] M. Hasnine and Rukhsana, "An Analysis of Urban Sprawl and Prediction of Future Urban Town in Urban Area of Developing Nation: Case Study in India," *Journal of the Indian Society of Remote Sensing (48),* pp. 909-920, 2020.

[29] H. Bekele, Urbanization and Urban Sprawl, Stockholm, Sweden: Royal Institute of Technology, 2005.

[30] D. Chen, X. Lu, W. Hu, C. Zhang and Y. Lin, "How urban sprawl influences eco-environmental quality: Empirical research in China by using the Spatial Durbin Model," *Ecological Indicators,* pp. 1-13, 2021.