DOI: https://doi.org/10.34123/jurnalasks.v16i2.802

POLITEKNIK STATISTIKA STIS

Available online at https://jurnal.stis.ac.id/

Jurnal Aplikasi Statistika & Komputasi Statistik
Vol. 16 No. 2, December 2024
e- ISSN: 2615-1367, p-ISSN:2086-4132

Spatial Dependencies in Environmental Quality: Identifying Key Determinants

Omas Bulan Samosir[1], Rafidah Abd Karim[2], M. Irfan Fauzi[3], Sarni Maniar Berliana[4*]

[1]Universitas Indonesia, Depok, Indonesia, [2]Academy of Language Studies, Universiti Teknologi MARA Perak Branch Tapah Campus, Malaysia, [3]BPS-Statistics Pesisir Selatan Regency, Indonesia, [4]Politeknik Statistika STIS, Jakarta, Indonesia
**Corresponding Author:* sarni@stis.ac.id

ARTICLE INFO

Article history:
Received 4 September, 2024
Revised 15 November, 2024
Accepted 19 November, 2024
Published 31 December, 2024

Keywords:
Natural Environment; Queen Contiguity; Spatial Analysis; Spatial Autoregressive Regression

Abstract

Introduction/Main Objectives: Environmental quality is essential to human development because it reflects the condition of our natural surroundings. **Background Problems:** Understanding the determinants of environmental quality is crucial for Indonesia as it helps identify the key factors influencing environmental quality. **Novelty:** Spatial models offer detailed, location-specific insights but require extensive data and computational resources, while non-spatial models provide a broader overview with simpler data requirements but may miss important spatial nuances. This study seeks to identify the determinants of environmental quality in regencies and municipalities on Java Island, incorporating spatial effects into the analysis. **Research Methods:** The dependent variable is environmental quality index. The independent variables are GRDP in industrial sector, GRDP in agricultural sector, urban population rate, population density, and poverty rate. We applied spatially lag regression model using contiguity spatial weight matrix. **Finding/Results:** This study shows the spatially lag regression model outperforms the OLS model. GRDP in the industrial sector, GRDP in the agricultural sector, urban population rate, and population density have negative effects, suggesting the increases in these variables were associated with lower environmental quality. About 40%–44% of each variable's effect on environmental quality is due to spatial spillover effects.

1. Introduction

Environmental quality is essential to human development and well-being because it reflects the condition of our natural surroundings and their capacity to support life [1]. The environment has changed significantly over the past century due to rapid urbanization, industrialization, and population growth. This has resulted in a number of problems, such as deforestation, air and water pollution, and biodiversity loss. In addition to disrupting ecosystems, these changes have put human health at serious risk. For example, air pollution from vehicle and industrial exhausts leads to cardiovascular issues and respiratory problems, while contaminated water sources can cause a variety of waterborne illnesses.

Wildlife is also put at risk by the destruction of natural habitats, which reduces biodiversity—a crucial component of ecological resilience and balance [2].

Several of the Sustainable Development Goals (SDGs), which reflect a global commitment to safeguarding and enhancing the natural environment [3], are closely linked to environmental quality. SDG 6 (Clean Water and Sanitation) emphasizes the importance of ensuring access to clean water and sanitation, as well as their sustainable management for all. Likewise, SDG 11 (Sustainable Cities and Communities) seeks to minimize the adverse environmental effects of urban areas, such as waste management and air pollution. SDG 13 (Climate Action) stresses the need to integrate climate action into national plans and calls for urgent measures to address climate change and its impacts. Additionally, SDGs 14 (Life Below Water) and 15 (Life on Land) focus on the protection of marine and terrestrial ecosystems, respectively. Collectively, these SDGs underscore the critical role that environmental quality plays in achieving sustainable development and improving global living standards.

The Indonesian government places significant importance on environmental quality, as evidenced by a range of key regulations aimed at addressing this issue. Central to this framework is Law No. 32/2009 on Protection and Management of the Environment, which was amended by Law No. 6 of 2023 following the enactment of Government Regulation in Lieu of Law No. 2 of 2022 on Job Creation. This law emphasizes sustainable development and environmental harm reduction, outlining the core principles, objectives, and strategies for environmental governance. It requires the preparation of Environmental Protection and Management Plans (RPPLH) and Environmental Impact Assessments (AMDAL) for projects that may significantly affect the environment. Additionally, the law defines the roles and responsibilities of key government agencies, such as the Ministry of Environment and Forestry (MoEF), in ensuring the enforcement and ongoing maintenance of environmental regulations.

Environmental Protection and Management Regulation No. 22/2021 provides detailed regulations on various aspects of environmental quality, complementing the broader environmental law. This regulation covers topics such as waste management, the preservation of terrestrial and marine ecosystems, and the quality of water and air. It outlines procedures for monitoring and reporting environmental performance and sets requirements for waste treatment, emissions control, and effluent discharge. To ensure adherence to environmental standards, the regulation also stipulates administrative penalties for non-compliance, including fines and the revocation of licenses. Together with other laws and regulations, this framework creates a robust legal structure designed to safeguard Indonesia's natural resources and promote sustainable development.

The MoEF is dedicated to fostering sustainable development that improves the welfare of the Indonesian people and contributes to a more advanced nation. The MoEF's main responsibilities include developing and enforcing policies related to environmental protection, forest conservation, and climate change mitigation. The ministry manages several programs focused on reducing deforestation, conserving peatlands, and promoting biodiversity. Additionally, the MoEF works on encouraging sustainable forest management, controlling pollution, and addressing issues related to hazardous and toxic substances [4]. A key tool used by the MoEF to assess and monitor the environment across Indonesia is the Environmental Quality Index (EQI). This index offers a comprehensive assessment of environmental health by evaluating factors such as air quality, water quality, land conditions, and seawater quality. The EQI is essential for the MoEF to identify areas needing improvement and to guide targeted actions to improve environmental quality [5].

Understanding the factors that determine the Environmental Quality Index (EQI) is essential for Indonesia, as it helps pinpoint the key elements affecting environmental quality. By examining these factors, policymakers and environmental experts can create focused strategies to tackle specific challenges. For example, studies have indicated that variables such as poverty, slum conditions, sanitation, and income inequality (as measured by the Gini ratio) have a significant impact on the EQI [6]. Recognizing these determinants enables more effective interventions, such as strengthening regulations and enforcement for businesses and empowering local communities, which can lead to better environmental management [7]. A deeper understanding of these factors also offers valuable insights for reducing environmental exposures, ultimately helping to mitigate adverse health outcomes [8].

Previous research on the factors influencing environmental quality has been quite extensive. For example, a study by [9] analyzed environmental quality determinants in 198 countries from 1990 to 2018 using panel quantile regression, finding a link between economic activities and carbon emissions. Another study by [10], conducted between 2000 and 2018 in 54 countries part of the Belt and Road Initiative (BRI), employed spatial econometric techniques to explore factors contributing to environmental degradation caused by economic activities. Meanwhile, [11] examined the factors influencing varying levels of environmental quality in Indonesia, particularly comparing Java to other islands, using canonical discriminant analysis. While the second study applied spatial modeling, the first and third studies utilized non-spatial models. Spatial models offer more detailed, location-specific insights but demand extensive data and computational resources. In contrast, non-spatial models provide a broader overview with simpler data needs, though they may overlook spatial variations and fail to capture spatial dependencies, potentially resulting in less accurate policy recommendations. This study, therefore, seeks to identify the determinants of environmental quality in regencies and municipalities on Java Island, incorporating spatial effects into the modeling of the relationship between independent variables and environmental quality.

2. Material and Methods

2.1. Data

The data used are secondary data sourced from the BPS-Statistics Indonesia and the MoEF. The dependent variable is EQI for regencies/municipalities on Java Island in 2021 except for Kepulauan Seribu. We removed this regency since we applied spatial area approach using contiguity spatial weight matrix. Thus, 118 regencies/municipalities are included in the study. The independent variables are Gross Regional Domestic Product (GRDP) in industrial sector (in natural logarithm), GRDP in agricultural sector (in natural logarithm), percentage of urban population, population density, and poverty rate. We employed several software or packages for data processing, including Rstudio version 2024.04.2 to obtain the regression models, GeoDa version 1.20.0.10 to obtain the spatial weight matrix, and QGIS version 3.36.2 to create thematic maps.

2.2. Model and Analysis Step

Spatial analysis, as defined by [12], refers to the quantitative study of phenomena that occur within a given space. According to [13], it involves examining how human activities and the physical environment vary across space, or, in other words, how these activities change with distance from specific reference points or objects of interest. The concept of "spatial analysis" is broad and includes several key components: (a) processing spatial data through geographic information systems (GIS); (b) conducting descriptive and exploratory analyses of spatial data; (c) applying statistical methods to assess the potential for drawing conclusions; and (d) creating models to predict outcomes and identify relationships within a spatial context [14].

According to Tobler's first law of geography, objects that are located near each other tend to share similar characteristics and are more likely to interact than those that are farther apart. As a result, spatial analysis requires the application of specific values or functions to define what is considered "near," "far," or "neighbor" for a set of spatial objects. In this study, the neighborhoods used to calculate the spatial weights matrix and regression model are defined through the contiguity adjacency matrix, which is the simplest way to define neighbors. The most common contiguity methods include: (1) the rook definition, where neighbors are defined as areal units that share a common edge; (2) the bishop definition, where neighbors are those sharing a common vertex; and (3) the queen definition, which combines both the rook and bishop methods, treating any object that shares either a common edge or vertex as a neighbor [15].

Most spatial analyses follow a standard approach that starts with a non-spatial linear regression model, and then determines whether spatial interaction effects should be included in this baseline model. The non-spatial linear regression model is expressed as [16]

$$\mathbf{y} = \alpha \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \varepsilon \tag{1}$$

In equation, $\mathbf{y}$ represents a vector of the dependent variable for each unit in the sample with the size of $n \times 1$ $(i = 1, \ldots, n)$, $\mathbf{1}_n$ is a vector of ones corresponding to the constant term parameter α to be estimated, $\mathbf{X}$ is a matrix of independent variables with the size of $n \times k$, $\boldsymbol{\beta}$ is a vector of unknown parameters to be estimated with the size of $k \times 1$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ is a vector of error terms, where ε_i is assumed to be independent and identically distributed with a mean of zero and variance σ^2. The linear regression model is often called the OLS model since it is usually calculated using Ordinary Least Squares (OLS).

A comprehensive model that includes all types of interaction effects can be written as

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \alpha\mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{u}, \tag{2}$$

$$\mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}.$$

In the model , $\mathbf{Wy}$ represents the interaction effects on the dependent variable, $\mathbf{WX}$ reflects the interaction effects between the independent variables, and $\mathbf{Wu}$ represents the interaction effects between the error terms of the different units, ρ is the spatial autoregressive coefficient, while λ the spatial autocorrelation coefficient. Both $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are vectors of unknown parameters to be estimated with a dimension of $k \times 1$. Meanwhle, $\mathbf{W}$ is a nonnegative weight matrix with the size of $n \times n$ that defines the spatial arrangement or structure of the locations in the sample.

When $\lambda = 0$ and $\boldsymbol{\theta} = \mathbf{0}$, the model in simplifies to

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \alpha\mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} \tag{3}$$

The simplified model in is known as the spatial lag model or spatially autoregressive model (SAR model), where the spatial lag of y_i for each location i is computed as the weighted sum of the dependent variable values of neighboring locations

$$[\mathbf{Wy}]_i = \sum_{j=1}^{n} w_{ij} y_j = w_{i1}y_1 + w_{i2}y_2 + \text{K} + w_{in}y_n \tag{4}$$

Here w_{ij} represents the weight between the i^{th} and j^{th} location, which is stored in the spatial weights matrix $\mathbf{W}$.

When when $\rho = 0$ and $\boldsymbol{\theta} = \mathbf{0}$, the model in becomes:

$$\mathbf{y} = \alpha\mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{5}$$

Where $\mathbf{u}$ is defined in . The model in is referred to as the spatial error model (SEM model). The SEM model addresses spatial dependencies by incorporating of a spatial autoregressive error term [13].

In this case, the values of the dependent variable y at each location are influenced by the stochastic error ε of neighboring locations, as determined by the filter $(\mathbf{I} - \lambda\mathbf{W})^{-1}$.

Before applying model or , it is essential to first assess whether the data show spatial dependence. The first step in a spatial area analysis is to test for spatial dependence without considering independent variables. Moran's I was used to tests for spatial autocorrelation in the data in this study. The formula of Moran's I is presented as [17]

$$I = \left[\frac{n}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right] \times \left[\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}\right] \tag{6}$$

When the data exhibit positive autocorrelation, this means nearby locations will have similar values, resulting in a Moran's I with positive value. Conversely, negative autocorrelation implies that the neighboring locations have dissimilar values, resulting in a negative Moran's I value. A score of 0.3 or greater, or of -0.3 or less, suggests a strong autocorrelation. To assess the significance of Moran's I, a p-value and a *z*-score are computed based on the null hypothesis of no spatial dependence (i.e., complete spatial randomness). If $p > 0.05$, we fail to reject the null hypothesis, suggesting the spatial distribution of the values could be due to random chance. However, if $p < 0.05$, we reject the null hypothesis and conclude that spatial dependence is present.

When spatial dependence exists in the data, choosing the correct spatial dependence model becomes essential. As outlined in Table 1, we used a set of diagnostic tools known as Lagrange multiplier (LM) test statistics, which include four different tests. If either the LM lag or LM error test is statistically significant, we proceed with the corresponding model. In situations where both tests are significant, we then check the results of the corresponding robust tests. If only one of the robust tests is significant, we select the model associated with the significant test. However, if both robust tests are significant, we choose the model with the higher value of the robust statistic.

Table 1. Lagrange multiplier diagnostics for spatial dependence.

Name of diagnostic test	Detects	Hypothesis
LM lag	Spatial lag effect	$H_0: \rho = 0$ vs. $H_1: \rho \neq 0$†
LM error	Spatial error effect	$H_0: \lambda = 0$ vs. $H_1: \lambda \neq 0$†
Robust LM lag	Spatial lag effect	$H_0: \rho = 0$ vs. $H_1: \rho \neq 0$†
Robust LM error	Spatial error effect	$H_0: \lambda = 0$ vs. $H_1: \lambda \neq 0$†

† Reject H_0 when *p*-value < α

The spatial multiplier, which links the independent variables to the dependent variable, complicates the interpretation of effects in a spatial lag model compared to a non-spatial model (or spatial error model). In a non-spatial model, the effect of a change in an independent variable is consistent across all locations, regardless of which specific location undergoes the change. In contrast, a spatial lag model shows that the impact of independent variables varies by location, due to the differences in neighboring units for each location [18]. The spatial regression model captures the feedback between locations through spatial lag terms, such as $\mathbf{Wy}$, which create interactions where changes in independent variables at one location, say location j, can influence the dependent at another location, i [19]. In a spatial lag model, the effect of an independent variable consists of two components: a direct effect, which reflects the local impact on location i, and an indirect or spillover effect, which is mediated through the spatial multiplier. Thus, the total effect of an independent variable is the sum of both the direct and indirect effects [20].

Table 2. Non-spatial diagnostics.

Name of diagnostic test	Detects	Hypothesis/Value
Jarque-Bera test	Non normality	† H_0 : the errors are normally distributed H_1 : the errors are not normally distributed
Breusch-Pagan test	Heteroscedasticity	† H_0 : constant variance of errors H_1 : non-constant variance of errors
Variance Inflator Factor (VIF)	Multicollinearity	$VIF = 1-4$ (No evidence of collinearity) $VIF = 4-10$ (Additional analysis is required) $VIF > 10$ (Indicates strong collinearity)

† Reject H_0 when *p*-value < α

Non-spatial diagnostics were also conducted to check whether the assumptions of normality and homoscedasticity are met, as well as the absence of multicollinearity. A detailed discussion of these is not the focus of this study. Readers can get a detailed discussion of these in [21]. The non-spatial diagnostics carried out in this study are presented in Table 2.

2.3. Model Evaluation

For models' evaluation we used two measures, which are Akaike Information Criterion (AIC) and likelihood ratio test (LRT) to determine the best-fitting model, whether it is the OLS model or the spatial regression model. The AIC is calculated as follows [22]

$$AIC = -2\log\left(L\left(\hat{\mathbf{\Theta}}\right)\right) + 2k \tag{7}$$

where $\hat{\Theta}$ is a vector of estimated parameters, $L(\hat{\Theta})$ is the likelihood function of the estimated model parameters, k is the number of estimated parameters, and n is the number of observations. A model is considered to have better performance when its AIC is lower.

We also calculated a likelihood ratio test to compare the spatial regression model as a more complex model to the OLS model as the simpler model. A higher likelihood score is always obtained by adding more parameters. Nevertheless, there comes a time when a model's fit to a given dataset can no longer be significantly improved by adding more parameters. Following the work of [23], the hypotheses for LRT in this study can be defined as follows

H_0 : the OLS model is equivalent to the spatial regression model

H_1 : the spatial regression model outperforms the OLS model

The test statistic for testing the hypotheses stated in is:

$$G^2 = -2\ln\left(\frac{L(\text{spatial regression model})}{L(\text{OLS model})}\right) \tag{8}$$

where $L(\text{spatial regression model})$ is the loglikelihood function under the spatial regression model and $L(\text{OLS model})$ is the loglikelihood function under the OLS model. This test statistic follows a Chi-square distribution with degree of freedom $v_2 - v_1$, where v_1 and v_2 are the number of estimated parameters in the OLS model and in the spatial regression model, respectively. The null hypothesis is rejected when $G^2 > \chi^2_{\alpha, v_2 - v_1}$. If the null hypothesis is rejected, then we should choose the spatial regression model.

3. Result and Discussion

The EQI distribution across Java Island's regencies/municipalities shown in Figure 1. The interval classes were divided based on natural breaks. The EQI value ranges from 41.26 in Bekasi to 72.24 in Batu City, with an average of 59.97. Adjacent regencies/municipalities that have relatively similar EQI scores tend to be clustered together, as shown in Figure 1. In general, the eastern side of Java Island has generally higher environmental quality than the western side, particularly in the regencies/municipalities that are part of Greater Jakarta.

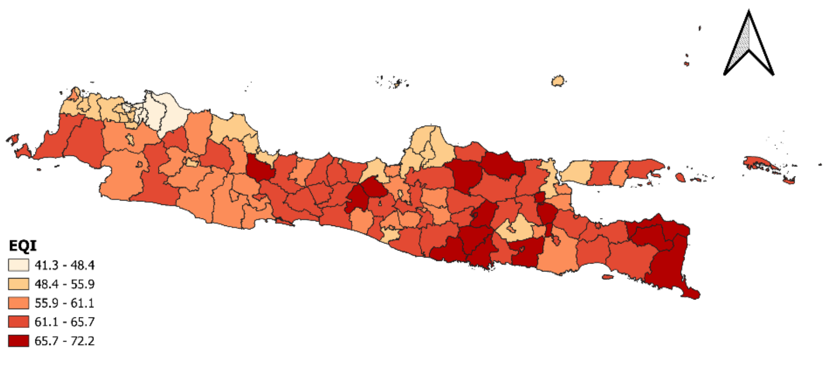

Figure 1. Environmental Quality Index by regencies/municipalities on Java Island.

The non-spatial diagnostics that we performed to assess if our model satisfies the classic assumptions are presented in Table 3. The normality assumption using the Jarque-Bera test gives p=0.3596 that we cannot reject the null hypothesis. Meanwhile, the Breusch-Pagan test for testing the homoscedasticity assumption gives p=0.1255, so we also cannot reject the hypothesis that the errors have constant variance. Each independent variable's *VIF* value is less than five, indicating the lack of multicollinearity. Therefore, each classic assumption now holds true. The estimated parameters of the OLS model for EQI are presented in Table 4.

All independent variables significantly affect EQI except poverty rate as shown in Table 4. All independent variables have a negative relationship to EQI, meaning that a one-unit increase in the independent variable will decrease EQI by the corresponding regression coefficient. Based on the smallest *p*-value, population density seems to have the strongest effect on EQI, followed by GRDP in industry and urban population rate.

As previously stated, to support the application of spatial regression models, we have to assess if our data reveal spatial autocorrelation. First, we need to compute the spatial weight matrix. In this study, we apply the queen contiguity definition to define a location's neighbors. Since there was the Suramadu bridge that connected Surabaya City and Bangkalan, we modified the neighbor for these two locations, where previously these two locations were not considered as neighbors based on the shapefile used for the analysis. Because there is no connection between Kepulauan Seribu and the Java Island's mainland, we therefore removed this location from the analysis. We have ensured that each location in our study has at least one neighbor, as shown in Figure 2.

Table 3. Results of non-spatial diagnostics.

Name of diagnostic test	Statistic	*p*-value
Jarque-Bera test	2.0458	0.3596
Breusch-Pagan test	8.6135	0.1225
	Independent variable	Value
VIF	GRDP in industrial	1.6346
	GRDP in agricultural sector	4.5700
	Percentage of urban population	3.0646
	Population density	3.2033
	Poverty rate	1.1429

Table 4. Coefficients of the OLS model for EQI on Java Island, 2021.

Coefficients	Estimate	Std. Error	*p*-value
Intercept	88.020337	4.6650	< 2e-16
GRDP in industrial sector	-1.067489	0.3676	0.0044
GRDP in agricultural sector	-1.552635	0.5818	0.0088
Percentage of urban population	-0.067834	0.0253	0.0086
Population density	-0.000861	0.0002	0.0000
Poverty rate	-0.000017	0.0072	0.9982

A location can have as many as 11 neighbors at most, with the average number being 4.42. Bogor has the highest number of neighboring areas, including Bekasi, Bekasi City, Bogor City, Cianjur, Depok City, Karawang, Lebak, Purwakarta, Sukabumi, Tangerang, and South Tangerang. On the other hand, 14 locations have only one neighbor, namely Blitar City, Bogor City, Cilegon City, Cirebon City, Kediri City, Magelang City, Malang City, Mojokerto City, Pasuruan City, Probolinggo City, Salatiga City, Serang City, Sukabumi City, and Sumenep. Except for Sumenep, all of these are cities within larger regencies. Sumenep, located on Madura Island, has Pamekasan as its only neighboring area.

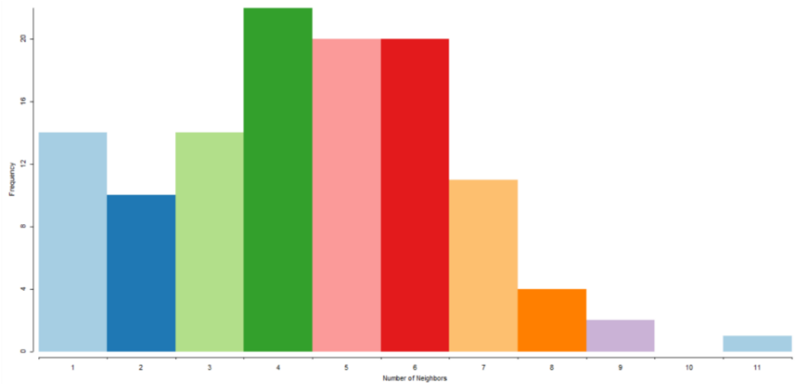

Figure 2. Number of neighbors based on queen contiguity definition.

The Moran's *I* for EQI, computed using the queen contiguity definition is 0.5752 with a *z*-score of 8.7454 (p=0.0000). The value indicates positive spatial autocorrelation, meaning that locations with higher EQI values tend to be surrounded by other locations with high EQI values, and similarly for lower EQI areas. The Moran's *I* distribution, shown in Figure 3, reveals that the observed value indicated by the green line departs significantly from the reference distribution, which contains no Moran's *I* value greater than 0.5752 under 999 permutations. This strongly suggests that the observed spatial autocorrelation is genuine rather than implying that it happened by chance, supporting the use of a spatial regression model over an OLS model for analyzing the determinants of EQI.

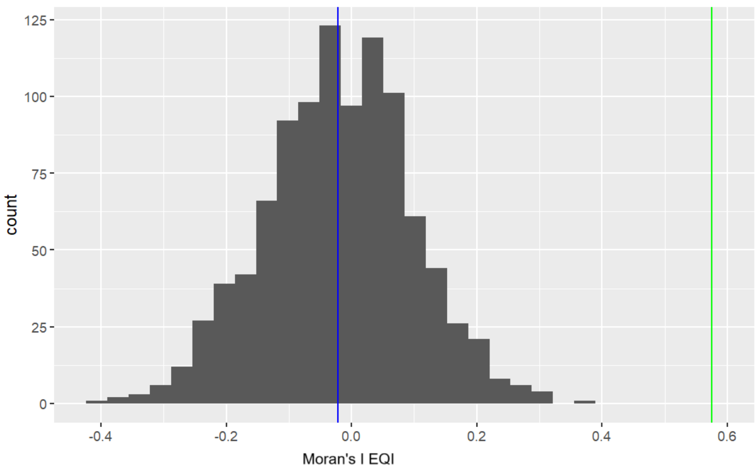

Figure 3. Moran's *I* of EQI under 999 permutations.

The next phase involves choosing the best spatial regression model for the dataset, which requires performing spatial diagnostics as previously mentioned above. The LM diagnostic test results, shown in Table 5, reveal that both the LM lag and the LM error tests are significant at 5%. Consequently, we turn to the robust LM test, which shows that the robust LM lag is markedly significant. This indicates that the correct model to use is the spatial lag model, also referred to as the spatially autoregressive model (SAR model).

Table 5. Results of LM diagnostics for spatial dependence.

Name of diagnostic test	Statistic	*p*-value
LM lag	24.83292	0.0000006252
LM error	20.40707	0.0000062598
Robust LM lag	4.67573	0.03059
Robust LM error	0.24989	0.61716

Table 6. Coefficients of the SAR model for EQI on Java Island, 2021.

Coefficients	Estimate	Std. Error	*p*-value
Intercept	51.4980	7.4930	0.0000
GRDP in industrial sector	-0.6298	0.3151	0.0456
GRDP in agricultural sector	-1.2274	0.5027	0.0146
Percentage of urban population	-0.0636	0.0216	0.0033
Population density	-0.0005	0.0002	0.0018
Poverty rate	-0.0023	0.0061	0.7047
ρ	0.4813	0.0816	0.0000

The results from the SAR model are reported in Table 6. The autoregressive parameter (ρ), is statistically significant, suggesting that spatial spillover effects play a substantial role in the relationship between the independent variables and EQI. The positive value of value is consistent with the positive Moran's I statistic. As shown, all independent variables, except for the poverty rate, have a significant effect on EQI. All variables have negative coefficients, suggesting that increases in these independent variables are associated with a decrease in EQI. As noted earlier, the total effect of each independent variable includes both a direct effect at the location and an indirect spillover effect due to spatial dependencies. The direct, indirect, and total effects of the independent variables along with their significances are presented in Table 7. All independent variables, except for the poverty rate, show significant direct, indirect, and total effects on EQI except for the poverty rate. Approximately 40%–44% of the effect of each independent variable on EQI is attributed to spatial spillover effects.

Table 7. Direct, indirect, and total effectc in SAR model for EQI on Java Island, 2021.

Variable	Direct	*p*-value	Indirect	*p*-value	Total	*p*-value
GRDP in industrial sector	-0.6718	0.0346	-0.5423	0.0702	-1.2141	0.0377
GRDP in agricultural sector	-1.3092	0.0140	-1.0570	0.0432	-2.3662	0.0166
Percentage of urban population	-0.0679	0.0012	-0.0548	0.0246	-0.1227	0.0033
Population density	-0.0005	0.0019	-0.0004	0.0102	-0.0010	0.0015
Poverty rate	-0.0025	0.6587	-0.0020	0.6615	-0.0045	0.6559

Table 8 shows that the SAR model outperforms the OLS model in this study. The positive autocorrelation in the data makes the SAR model an appropriate choice for analyzing the determinants of EQI, as confirmed by the spatial diagnostics in Table 5. The *AIC* value for the SAR model is lower than that of the OLS model, indicating a better fit model. Additionally, the likelihood ratio test (LRT) yields a *p*-value of 0.0000, leading to the rejection of the hypothesis of no difference. The lower *AIC* value and the significant LRT result support the superiority of the SAR model. Furthermore, the standard errors of the SAR model estimators are generally smaller than those of the OLS model. These findings suggest that the SAR model provides a more accurate and reliable fit for analyzing the factors affecting EQI in the regencies and municipalities on Java Island.

Table 8. Model evaluation.

Measure	OLS	SAR
AIC	699.0427	675.3000
Log-likelihood	-342.5213	-329.6500
Likelihood ratio = -25.743		
Degree of freedom = 1		
p-value = 0.0000003901		

This study shows that for every one percent increase in GRDP in the industrial sector at location i will have a direct impact on decreasing the EQI on its own location by 0.6718%. Industrial activities are often associated with increased pollution and environmental degradation due to emissions from manufacturing processes, waste generation, and resource extraction. Studies have shown that regions with higher industrial GRDP tend to experience a decline in environmental quality, primarily due to the release of pollutants such as carbon emissions (CO2), sulfur air pollution (SO2), and airborne particles [24]. For instance, a study on newly industrialized countries (NICs) found that industrialization-driven economic growth frequently degrades environmental quality, as increased industrial output results in larger ecological footprints [25]. Based on Table 7, it is also known that for every one percent increase in GRDP in the industrial sector of all neighbors of location i will decrease the EQI at location i by 0.5423%. Thus, EQI will decrease by 1.2141% in total for every one percent increase in GRDP in the industrial sector.

The EQI at location i is directly influenced by the GRDP in the agriculture sector, with every one percent increase in this variable resulting in a 1.3092% decrease in EQI. Research indicates that the agricultural sector often has a positive and significant relationship with environmental quality [26, 27]. This is because, when agricultural activities are managed sustainably, practices like crop rotation, organic farming, and conservation farming, which boost soil health and lower pollution, can improve environmental quality. However, the degree of intensity and type of farming practices employed can alter this relationship. The overuse of chemical pesticides and fertilizers, for instance, can cause soil degradation and water contamination, which have an adverse effect on the EQI. The GRDP in the agricultural sector indirectly reduces the EQI at location i by 1.0570% for every one percent increase of all neighbors of location i. So, in total, a one percent increase in this variable will reduce EQI by 2.3662%.

As the urban population rate increases one percent at location i, it impacts directly to decrease the EQI at the same location by 0.0679 units. As the number of people living in cities increases, so does the demand for resources like water, energy, and land, which frequently leads to environmental degradation. High population densities in urban areas can exacerbate pollution of the air and water, increase waste production, and impose a burden on infrastructure and public services. A study on urban growth and population density changes in China shows that rapid urbanization frequently leads to higher levels of air pollution and traffic congestion, which have a substantial negative impact on urban dwellers' quality of life [28]. A different study by [29] found that the unplanned expansion of cities known as "urban sprawl" may result in the loss of natural environments and green spaces, which lowers the quality of the surrounding environment. A study by [30] also shows that the air pollution rises in connection with the urban population. The effect of an increase in urban population from all location i's neighbors will cause a 0.0548-unit decrease in the EQI at location i. This variable's total impact on EQI decline as its increase is 0.1227 units.

There is a statistically significant impact of population density at location i on EQI at that location, both directly and indirectly. The EQI will decrease by 0.0005 and 0.0004 units, respectively, for each one unit rise in population density. This finding is consistent with a study by [30] that shows rising population density will result in higher air pollution. High population density often leads to increased pollution levels as more people generate more waste and emissions. Another study conducted in Indonesia discovered that higher population density correlates with increased air and water pollution, which negatively affects the EQI [31]. High-population density urban areas typically have higher rates of energy consumption, industrial activity, and vehicles—all of which worsen the degradation of the environment. This study shows the total effect of a one-unit increase in population density on the decrease in EQI was 0.001 unit.

4. Conclusion

When spatial dependence is present in the data, using a spatial regression model is essential to appropriately model the spatial structure. This study shows that the SAR model provides a more accurate fit than the OLS model by accounting for spatial dependencies, leading to more reliable parameter estimates and smaller standard errors. The findings indicate that the industrial and agricultural GRDP, urban population rate, and population density all have negative effects on EQI, with increases in these variables being associated with declines in environmental quality. Moreover, around 40%–44% of the effect of each variable on environmental quality is due to spatial spillover effects.

To improve environmental quality, policies should promote sustainable industrial and agricultural practices, manage urban population growth, and address population density issues. Given the significant spatial spillover effects, it is crucial to promote regional collaboration to handle environmental challenges collectively, such as joint pollution control efforts and coordinated land-use planning to ensure that improvements in one area benefit neighboring regions as well.

Ethics approval

Approval for ethics was not required for this study.

Competing interests

The authors confirm that there are no conflicts of interest.

Funding

No external funding was provided for this study.

Underlying data

The data supporting the results of this study can be requested from the corresponding author.

Credit Authorship

Omas Bulan Samosir: Writing – original draft, Writing – review and editing, Supervision. **Rafidah Abd Karim**: Writing – original draft, Writing – review and editing. **M. Irfan Fauzi**: Data processing. **Sarni Maniar Berliana**: Model conceptualization, Writing – original draft, Writing – review and editing.

References

[1] P. J. Campbell, A. MacKinnon and C. R. Stevens, "The Natural Environment," in *An introduction to global studies*, West Sussex, United Kingdom, Wiley-Blackwell, 2010, pp. 122-160.

[2] United Nations Environment Programme, "Healthy environment, healthy people," UNEP, 2016.

[3] Department of Economic and Social Affairs, "The 17 Goals," United Nations, [Online]. Available: https://sdgs.un.org/goals. [Accessed 25 August 2024].

[4] Ministry of Environment and Forestry, Regulation of the Minister of Environment and Forestry No. 16 of 2020 about the Strategic Plan 2020-2024, Jakarta: Ministry of Environment and Forestry, 2020.

[5] Ministry of Environment and Forestry, Indeks Kualitas Lingkungan Hidup Indonesia 2022 [The 2022 Indonesia Environmental Quality Index], Jakarta: Ministry of Environment and Forestry, 2022.

[6] A. R. Noormalitasari and A. Setyadharma, "Determinants of Environment Quality Index In Indonesia," *Efficient: Indonesian Journal of Development Economics*, vol. 4, no. 2, pp. 1174-1187, 2021.

[7] P. F. Butarbutar, C. F. Ananda and F. Prasetyia, "The determinants of environmental quality in Indonesia," *Journal of Indonesian Applied Economics*, vol. 11, no. 1, pp. 27-39, 2023.

[8] US EPA Research, "EPA Researchers Release Updates to Environmental Quality Index," United States Environmental Protection Agency, 18 May 2021. [Online]. Available: https://www.epa.gov/sciencematters/epa-researchers-release-updates-environmental-quality-index. [Accessed 30 August 2024].

[9] M. Wang, N. Arshed, M. Munir, S. F. Rasool and W. Lin, "Investigation of the STIRPAT model of environmental quality: a case of nonlinear quantile panel data analysis," *Environment, Development and Sustainability,* vol. 23, p. 12217–12232, 2023.

[10] Q. Chen, G. R. Madni and A. A. Shahzad, "The usage of spatial econometric approach to explore the determinants of ecological footprint in BRI countries," *PLoS ONE,* vol. 18, no. 10, pp. 1-14, 2023.

[11] A. Pujiati, T. Nurbaeti and N. Damayanti, "What are the factors that determine differing levels of environmental quality? Evidence from Java and other islands in Indonesia," *Management of Environmental Quality,* vol. 34, no. 2, pp. 290-307, 2022.

[12] L. Anselin, "What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis," University of California, Santa Barbara, 1989.

[13] F. Wang, Quantitative Methods and Socio-Economic Applications in GIS, New York: CRC Press, 2014.

[14] D. O'Sullivan and D. J. Unwin, Geographic Information Analysis, 2 ed., Hoboken, New Jersey: John Wiley & Sons, 2010.

[15] L. M. Scott and M. V. Janikas, Spatial Statistics in ArcGIS, M. M. Fischer and A. Getis, Eds., Springer-Verlag Berlin Heidelberg, 2010.

[16] J. P. Elhorst, Spatial Econometrics: From Cross-Sectional Data to Spatial Panels, New York: Springer Heidelberg, 2014.

[17] P. A. P. Moran, "Notes on Continuous Stochastic Phenomena," *Biometrika,* vol. 37, no. 1/2, pp. 17-23, 1950.

[18] M. D. Ward and K. S. Gleditsch, "Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace," *Political Analysis*, vol. 10, no. 3, pp. 244-260, 2002.

[19] J. P. LeSage and M. M. Fischer, "Spatial Econometric Methods for Modeling Origin Destination Flows," *SSRN,* p. 25, 2008.

[20] J. P. LeSage and R. K. Pace, "Spatial Econometric Models," in *Handbook of Applied Spatial Analysis*, M. M. Fischer and A. Getis, Eds., Berlin Heidelberg, Springer-Verlag, 2010, pp. 355-376.

[21] D. N. Gujarati, Basic Econometrics, 4 ed., New York: McGraw Hill , 2004.

[22] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory,* Budapest, 1973.

[23] Q. H. Vuong, "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica,* vol. 57, no. 2, pp. 307-333, 1989.

[24] M. Demiral, Ö. Haykır and E. D. Aktekin-Gök, "Environmental pollution effects of economic, financial, and industrial development in OPEC: comparative evidence from the environmental Kuznets curve perspective," *Environment, Development and Sustainability,* 2023.

[25] C. Karaduman, "The effects of economic globalization and productivity on environmental quality: evidence from newly industrialized countries," *Environmental Science and Pollution Research,* vol. 29, p. 639–652, 2022.

[26] I. Ummi, Rusdarti and H. Y. , "Relationship of the GRDP Sectors with Environmental Quality Index in Indonesia 2012-2017," *Journal of Economic Education,* vol. 8, no. 2, pp. 152 - 158, 2019.

[27] S. Saghaian, H. M. and M. Mohammadi, "The Effects of Agricultural Product Exports on Environmental Quality," *Sustainability,* vol. 14, pp. 1-11, 2022.

[28] H. Lu, Z. Shang, Y. Ruan and L. Jiang, "Study on Urban Expansion and Population Density Changes Based on the Inverse S-Shaped Function," *Sustainability,* vol. 15, no. 13, pp. 1-19, 2023.

[29] H. Zhang, "The Impact of Urban Sprawl on Environmental Pollution: Empirical Analysis from Large and Medium-Sized Cities of China," *International Journal of Environmental Research and Public Health,* vol. 18, no. 16, pp. 1-19, 2021.

[30] L. Han, W. Z. W. Li and Y. Qian, "Urbanization strategy and environmental changes: An insight with relationship between population change and fine particulate pollution," *Science of The Total Environment,* vol. 642, pp. 789-799, 2018.

[31] A. N. Wafiq and Suryanto, "The Impact of Population Density and Economic Growth on Environmental Quality: Study in Indonesia," *Jurnal Ekonomi & Studi Pembangunan*, vol. 22, no. 2, pp. 301-312, 2021.