Letters in Economic Research Updates
ISSN: 3068-9236

PRIME OPEN ACCESS

Volume 1, Issue 2

Research Article

Date of Submission: 01 April, 2025
Date of Acceptance: 20 August, 2025
Date of Publication: 28 August, 2025

Spatial Econometrics Models Applied to Environmental Pollution. A Systematic Review

Manuel Canaveral[1*], Juliano Marangoni[1], Leonardo Emmendorfer[2] and Cassius Oliveira[3]

[1]Institute of Mathematics, Statistics and Physics, Federal University of Rio Grande, Brazil

[2]Department of Electric Energy Processing, Federal University of Santa Maria, Brazil

[3]Institute of Economic, Federal University of Rio Grande, Brazil

***Corresponding Author:**
Manuel Canaveral, Institute of Mathematics, Statistics and Physics, Federal University of Rio Grande, Brazil.

Citation: Canaveral, M., Marangoni, J., Emmendorfer, L., Oliveira, C. (2025). Spatial Econometrics Models Applied to Environmental Pollution. A Systematic Review. *Lett Econ Res Updates, 1*(2), 01-22.

Abstract
The interest in spatial analysis has been growing in recent years, mainly due to communication technology advances, economic globalization, and the development of new statistical methods and computational tools. This article aims to contribute to the dissemination of spatial statistical models applied in econometrics, by presenting some basic theoretical aspects and a literature review of articles that address the socio-economic drivers that lead to environmental pollution. Three spatial regression models are reviewed here: the spatial lag model (SLM) also known in the literature as SAR, the spatial error model (SEM), and the spatial Durbin model (SDM). A literature search was conducted using specific terms of interest in eight databases, from 1996 to February 2021, where 22 articles were considered for analysis. The results showed that most articles studied environmental problems in China. The most used exploratory spatial analysis model was Moran Index and the most used explanatory spatial analysis models were SDM and SLM.

Keywords: Spatial Econometrics, Literature Review, Socioeconomic Drivers, Environmental Pollution

Introduction
Among the major environmental challenges that potentially threaten humanity over the forthcoming decades, one should mention climate change, biodiversity loss, unsustainable fishery, emerging pathogens, deforestation, land usage issues, just to name a few. Solving those emerging crises requires us to be able to address increasingly complex problems with large amounts of multidimensional data. Those issues demand sophisticated statistical tools along with the application of complex quantitative models [1].

Under this novel paradigm one should address questions concerning the extent of the impact, the existence of an agglomeration pattern, the occurrence of environmental externalities spilling over geographically to other regions, the factors related to a given phenomenon, among other questions. Given the inherent complexity of the environment-society interrelationships, progress towards the solution of those challenges demands the adoption of increasingly complex statistical tools.

The spatiality of the phenomena under consideration is another important element to be addressed. As Fotheringham et al., (2000) point out, it is increasingly recognized that under a multidisciplinary context, most data are spatial [2]. This refers to the fact that most variables have an implicit geographic location, which implies data are two-dimensional, dimension of the variable *per se* and a spatial dimension.

For example, in the case of socioeconomic variables, some express their spatial dimension implicitly, because their geographical location is not immediately observable or attributable to a specific place, but they can be easily assigned to geographical locations of different scales (municipalities, provinces, countries, etc.), here we could mention GDP, unemployment rate, CO2 emissions, deforestation rate, etc. Other variables, mainly environmental, but also socioeconomic, express their spatial dimension explicitly, because they occurred in specific immediately observable locations, for example, greenhouse gases and water pollutants measurements collected at specific stations or points, as well as cases of crime, traffic accidents also observable at specific points.

Then analyzing data from different sources and with different characteristics can result in spatially dependent observations [3]. In addition, processes, whether environmental, biological or socioeconomic, are often spatially or temporally related [4]. This creates the need for spatial statistical modeling techniques. Therefore, these techniques have been widely used in recent years in the social sciences in general, and in applied environmental research in particular [5,6].

This paper aims to contribute to the dissemination of spatial statistical data modeling, presenting some theoretical background and a literature review related to socioeconomic drivers related to environmental pollution, focusing on three main spatial regression models. The remainder of the paper is organized as follows: Section 2 describes the theoretical framework. The spatial regression models presented here are the spatial lag model (SLM), the spatial error model (SEM), and the spatial Durbin model (SDM). The literature search methodology is presented in Section 3. Section 4 provides the results and the conclusions are given in Section 5.

Spatial Econometrics

We use the term "Spatial econometrics" as a short way to denote "frequentist spatial statistical models applied in econometrics". The increasingly higher interest in spatial analysis in recent years is due to the advances in communication technologies, economic globalization, and the development of new statistical and econometric tools, among which spatial econometrics (SE) stands out [7]. The difference between conventional econometrics (CE) and SE is that the former usually neglect interactions that might occur between observations on geographic level. CE assumes no autocorrelation or independence, i.e., there is no covariance between the residuals of different spatial units [8]. This can be represented by the variance-covariance matrix of the residuals ε, with $\varepsilon = y - X\beta$. Where y is an $nx1$ vector of the dependent variable, X is a non-stochastic matrix of regressors of size nxk, β is the vector of parameters, and n is the size of a cross-sectional sample of spatial units.

$$\mathbb{E}(\varepsilon\varepsilon') = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I \qquad (1)$$

The variances are given by the elements of the main diagonal and the covariances by the elements outside the main diagonal, and the covariances are equal to zero, because the residual of the individual i is not related to the residual of the individual j, that is, they are independent. However, the spatial interactions of the observations violate this independence assumption, since at the geographic level, those might be autocorrelated or spatially dependent.

For example, in environmental sciences, the state of a given ecosystem in one region is not detached from the state of ecosystems in neighboring regions. If air or water quality is low in region i, it is very likely that in the other regions near i it is also low. This might result either from the inner similarities between regions that share socioeconomic patterns that negatively impact ecosystems or due to the negative effects from environmental externalities that spillover to neighboring regions. Spatially dependence among observations is precisely the main interest of SE, unlike CE majorly focuses on temporal dependence [9].

The choice of a statistical model must be taken considering that there are potential spatial effects that will affect the performance of the model. If we use the classical Ordinary Linear Model (OLM), commonly estimated using the Ordinary Least Squares (OLS) method, in a sample of n observations with spatial location (explicit or not), the results could be affected by significant biases [10]. Because the OLM assumes independence and spatial homogeneity. Therefore, it is necessary to select models that incorporate spatial effects.

There are two main spatial effects, spatial heterogeneity and spatial dependence [11]. The first is related to the spatial differentiation of geographic units, i.e., different impacts or coefficients for each spatial unit of the sample [12]. This is the main difference with CE, since in OLM the coefficients will be constants over all sample units. One of the main estimation techniques to control for spatial heterogeneity effects is Geographically Weighted Regression (GWR), which is an extension of OLS estimation to the case of geo-referenced data [13]. GWR in its scalar form can be expressed as:

$$y_i = \beta_0 + \sum_{j=1}^{k} x_{ij}\beta_{ij}[lat_i, lon_i] + \varepsilon_i \quad (2)$$

Where y_i represents the response variable in spatial unit i with i=1,2,3…n, x_{ij} represents the observation i of regressor j, with j=1,2,3…k, β_0 is the intercept term, β_{ij} are the parameters to be estimated that depend on the latitude and longitude coordinates, and ε_i is a normally distributed error term associated with location i.

Since GWR fits a regression equation for each spatial unit i, it will have a β_{ij} associated with each i and each regressor j. While at MCRL β is the same for the entire vector of spatial units. For example, if we model environmental pollution as a function of population growth, for n regions, using a simple log-linear regression model, as:

$$\ln Y = \beta_0 + \beta_1 \ln X + \varepsilon \quad (3)$$

Where Y is an $nx1$ vector representing environmental pollution, β_0 = ln A, is the intercept, β_1 is the slope, X is an $nx2$ matrix representing population growth and a column of ones. If β_1=0.2 the OLS interpretation would be: For every 1% increase in population, environmental pollution increases by an average of 0.2% for all n regions equally, even if we have that x number of regions are more densely populated than others. This naturally exacerbates pollution, and assuming an average increase of 0.2% for all regions is an unrealistic result.

If we are interested in modeling heterogeneity, we must ask ourselves if the parameters of the model would be different for each of the spatial location in the sample. For the example above, the question would be: Does the environmental pollution of all regions react in the same way to increases in population? Most likely not, so we need to know what the differential effect is for each region. Therefore, we apply a model of spatial heterogeneity.

The spatial dependency effect is also called Spatial autocorrelation, in this document they will be used as synonyms. Spatial autocorrelation is the correlation of a variable with itself and arises when the value of a variable in a location is related to its value in other locations in space [14]. It can be expressed as:

$$Corr[X(s_i);\ X(s_j)];\ \forall i \neq j \quad (4)$$

This means the correlation of a variable X in the location S_i concerning other neighborhood units S_j. Spatial autocorrelation can be either positive or negative. Positive spatial autocorrelation (PSA) occurs when a variable x at one location i exerts a contagion effect on neighboring regions j, which raises clusters of spatial units with geographic proximity. Negative spatial autocorrelation (NSA) occurs when a variable x at one location i exerts a repulsive effect on neighboring regions j, which is related to the dissimilarity between spatial units that are geographically close to each other. For example, suppose a municipality m located in a region i decides to implement environmental flexibility policies that eliminate certain restrictions to the entry of mining investments, which generates economic growth. If other municipalities near i, motivated by the mining economic benefits, also decide to implement environmental flexibility policies, this configures a PSA scenario. On the contrary, if the neighbors observe that mining environmental costs are greater than benefits and decide to restrict mining by implementing environmentally conscious policies, this results in NSA scenario. Figure 1 illustrates those two contrasting cases.

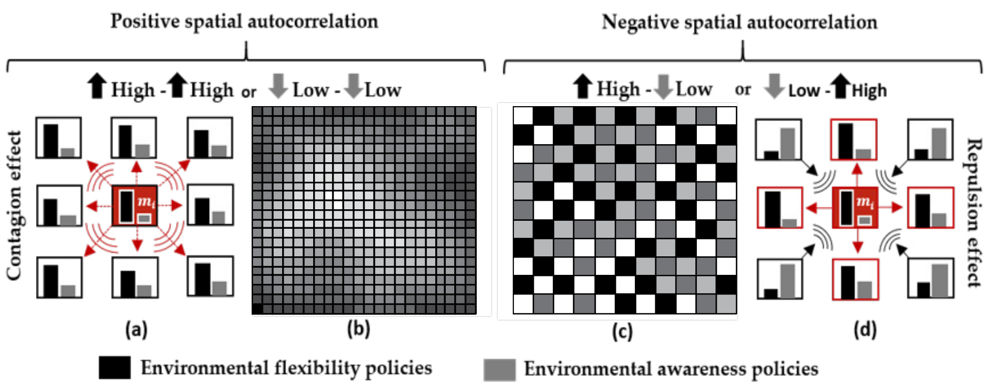

Figure 1: Positive and Negative Spatial Autocorrelation Scenarios

(a) shows a PSA scenario where the contagion or spillover effect occurs. Here m_i and its neighbors presents high values of the environmental flexibilization policies variable, the same happens with the environmental awareness policies variable with low-low correlations. (b) shows how on a macro level spatial clusters are formed. (d) illustrates how the repulsion effect prevents the cluster formation, therefore, on a macro level, a mosaic-style pattern might arise, as shown in (c). As a result, NSA arise from higher values correlated with lower values and vice-versa.

It is important to mention that NSA is less intuitive and more difficult to imagine than PSA, for this reason NSA has been less studied in empirical research (Griffith, 2019). This spatial pattern of close neighbors with different values may be due to the characteristics of each spatial unit, since these present different characteristics and, therefore, different values depending on the measurement variable (see Figure 2). There can even be PSA on one variable and NSA on another variable with the same neighborhood.

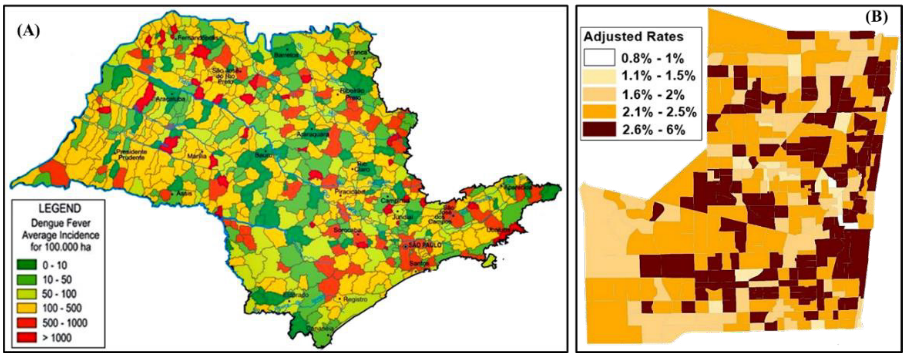

Figure 2: Examples of negative spatial autocorrelation
(A) Source: adapted from Brum-Bastos et al., (2016) and represents the average incidence of dengue for the municipalities of the state of São Paulo, Brazil. For the 2000 to 2014 period.
(B) Source: adapted from Hu, (2020) and represents the spatial pattern of age-adjusted breast cancer rates at the census tract level in Broward County, Florida, USA. For the 2000 to 2010 period.

Although several factors can produce NSA, this spatial phenomenon occurs mainly in situations of competition of scarce resources, for example, geographic space, volume of clients, privileges, food, etc [15-18]. Several competition examples can be given. In criminology, criminal gangs constantly compete for territory to sell illicit substances. In ecology, the effect of competition is reflected in the reduction of growth of individuals that are in a more unfavorable competitive situation [19]. For example, two species of microscopic algae, *Asterionella* and *Synedra* found in the same area and compete for food (silica). Synedra species consume silica until exhausted, this prevents the growth of Asterionella [20].

Spatial autocorrelation study the geographical distribution of the values of a variable for different spatial units. Spatial dependence tries to explain how this functional relationship is occurring. When data show spatial dependence, it means that observations from one region i depend on observations from another region j [21,22]. This can be formally expressed as:

$$y_i = f(y_j) + \varepsilon_i, \ \forall i \neq j \text{ with j= (1,2, 3, ……, n)} \qquad (5)$$

This means that y in the region i is explained not only by internal conditions in i but also by the values of the same variable y in neighboring regions j. Equation 7 helps to understand the spatial dependence phenomenon, but in practice, it is unfeasible, since given a sample n spatial units, the spatial dependence of each of the i regions with each of the remaining j regions must be estimated. For example, if n= 5 one should specify the following system of equations:

$$\begin{aligned}
y_1 &= \beta_{21}y_2 + \beta_{31}y_3 + \beta_{41}y_4 + \beta_{51}y_5 + \varepsilon_1 \\
y_2 &= \beta_{12}y_1 + \beta_{32}y_3 + \beta_{42}y_4 + \beta_{52}y_5 + \varepsilon_2 \\
y_3 &= \beta_{13}y_1 + \beta_{23}y_2 + \beta_{43}y_4 + \beta_{53}y_5 + \varepsilon_3 \\
y_4 &= \beta_{14}y_1 + \beta_{24}y_2 + \beta_{34}y_3 + \beta_{54}y_5 + \varepsilon_4 \\
y_5 &= \beta_{15}y_1 + \beta_{25}y_2 + \beta_{35}y_3 + \beta_{45}y_4 + \varepsilon_5
\end{aligned} \qquad (6)$$

The first row tells us that in a 5-region system, region 1 depends on regions 2, 3, 4, and 5 plus an error term ε_1, and so on. $\beta_{21} y_2$ measures the impact or influence that the value of the dependent variable in region 2 has on the value of the dependent variable in region 1. The problem is that given a sample of n observations in a cross-section there are $(n^2 - n) / 2$ parameters to be estimated (symmetry) being impossible to perform the estimation. This problem is caused by the multi-directionality of spatial dependence. That is, spatial dependence can occur in several directions across space due to interdependence between spatial units [23]. This is the main reason why spatial dependence cannot be treated by CE since CE deals with unidirectional effects in time, where the past explains the present. For example, Figure 3 represents the geographical distribution of the 5-region system with multidirectional flows in space.

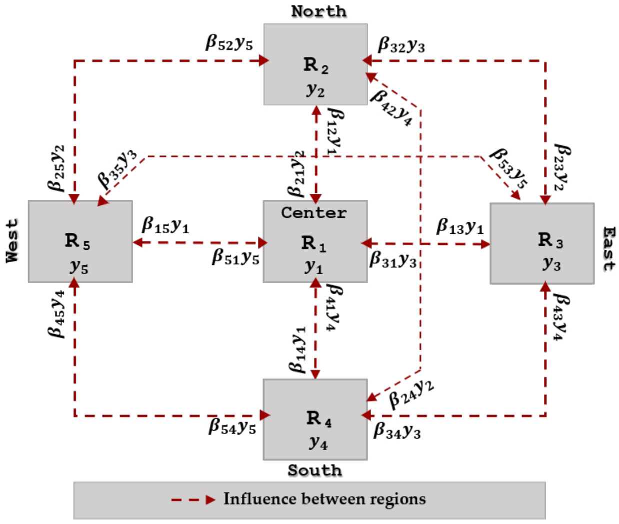

Figure 3: Multi-directionality of spatial dependence.

The red dashed lines show the influence that the y_i value in a region i has on the value of the dependent variable of the other regions.

The multi-directionality of spatial dependence is the influence that one region has on another and vice versa. This interdependence of regions (all regions versus all regions) is what generates the multiple parameters to be estimated. Therefore, restrictions should be imposed on the way observations interact in space, thus reducing the number of parameters. This is achieved by constructing the spatial weights matrix (W), which controls the multi-directionality problem. This matrix makes it possible to operationalize the relationships in space, incorporating into the general linear model the interdependence relationships of the spatial units. This makes the model cease to be classical and take on the spatial characteristic. The matrix W also represents the reduction of interdependencies between regions. The concept is aligned to Waldo Tobler's first law of geography which states that "*Everything is related to everything else, but near things are more related than distant things*" [24].

W is an *nxn* matrix of non-negative values and zeros on the main diagonal (no one is a neighbor of himself). The main characteristics of this matrix are as follows:

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix} \quad (7)$$

Positive (defined positive or non-singular): all its elements w_{ij} that represent the connectivity (adjacency or closeness) between regions i and j are positive. Symmetric ($w_{ij} = w_{ji}$ is an assumption to be checked): it means that the value of element w_{12} is the same as element w_{21}. For example, if 1= Rio Grande city and 2= Pelotas city, the spatial dependence between Rio Grande and Pelotas is the same as between Pelotas and Rio Grande. Non-stochastic: means that it is a matrix of numerical values that are not random, i.e., the values are fixed (exogenous) given by a spatial configuration that is specific in advance.

The specification of a spatial weights matrix is one of the most important and conflicting aspects in spatial econometric analysis because different matrices produce different results [25]. There are several ways to construct the W matrix, however, the most commonly used approach is based on the binary contiguity of spatial units (first-order). This method uses 0 and 1 values to represent the degree of interaction between regions. For instance, $w_{ij} = 1$ might correspond to spatial units that share a common border, while $w_{ij}=0$ might represent noncontiguous areas.

With the W matrix, spatial autocorrelation can be quantified by means of spatial indicators. These are statistics that help identify whether clustering or dispersion occurs in the data. Usually are differentiated by the scope or scale of analysis in "Global" and "local" measures [26]. Global measures analyze all spatial units in the sample taken together to determine the occurrence of clusters, or random distribution, the main methods include Moran's I, Geary's C and Getis-Ord's G and G* statistics. In global measures one of the most widely used is Moran's I [27-29,7]. Formally expressed as:

$$Global\ I = \frac{N}{\sum_{ij}^{N} w_{ij}} \frac{\sum_{ij}^{N} w_{ij}(y_i - \overline{y})(y_j - \overline{y})}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{8}$$

With $i \neq j$, where y_i is the interest variable in region i, $\overline{y}$ is the variable mean value, N is the number of observations, w_{ij} is the spatial weight. The Moran index varies between -1 and 1. A value close to 1 indicates the presence of clusters, and a value close to -1 indicates spatial dispersion in the data. If the value is 0, there is no spatial autocorrelation.

Local spatial autocorrelation is associated with the occurrence of clusters located in specific areas of the territory that concentrate higher or lower values (hot or cold spots) of a variable compared to its mean value [30]. Local measures arises because global statistics hide the variation in spatial associations that may exist within local areas due to spatial heterogeneity [11]. Therefore, local measures analyzed each spatial unit separately, obtaining a statistic for each region of the sample. Local methods including local Moran's I, local Geary's C, and local G and G* statistics. Local Moran's I is expressed formally as:

$$Local\ I = \frac{(y_i - \overline{y}) \sum_{j=1}^{N} w_{ij}(y_j - \overline{y})}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{9}$$

The specifications are the same as in the global Moran index.

A Taxonomy of Spatial Regression Models

After the spatial autocorrelation analysis confirms the spatial dependence of the geographic elements, it is necessary to establish a spatial statistical linear model to evaluate the effect and spatial interaction of these elements. In this section, we present spatial autoregressive models, which are used for cross-sectional data.

Spatial dependence can be incorporated into the basic linear regression model through two different approaches: substantive spatial dependence, also known as the Spatial Lag Model (SLM), or residual spatial dependence, also known as the Spatial Error Model (SEM). Another model of interest for this paper is the Spatial Durbin Model (SDM), which is a special case of SLM.

The SLM also called Spatial Autoregressive Regression Mixed Model is used when an endogenous variable of a regression model is spatially correlated. Dependence is considered as an additional regressor under the form of a spatially lagged dependent variable (Wy). This is the most basic model of spatial dependence, which is suitable for expressing situations where the values of a variable depend systematically on the geographic location of that variable. It is formally expressed as:

$$\begin{gathered} y = \rho W y + X\beta + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 I) \end{gathered} \tag{10}$$

Where y is a vector of $nx1$ dependent variables. X is a matrix of nxk exogenous explanatory variables, and β is a vector of $kx1$ regression parameters. Wy is a vector of spatial lags of dimension $nx1$, which result from the product of the spatial weights matrix W and the vector of dependent variables y, and reflect the weighted average of the values of y in neighborhood i. ρ is a scalar parameter reflecting the strength of spatial dependence. When the scalar parameter ρ takes the value zero the spatial lag model simplifies to the classic OLM. Finally, we assume that the vector ε of residuals with dimension $nx1$ contains independent, normally distributed terms with zero mean and constant variance.

The Spatial Error Model also called linear regression model with spatially autoregressive disturbances is used when the spatial autocorrelation is present only in the disturbance term, and is mostly considered to be a nuisance which needs to be eliminated [31]. The SEM is the most widely adopted specification when the basic linear regression model is ineffective in explaining a spatially autocorrelated phenomenon. The existence of certain factors or variables not explicitly considered in the model transfers to the error terms the configuration of clustering of values present in the endogenous variable [30]. For the SEM the usual OLM ($y = X\beta + \varepsilon$) is complemented by a term ($\lambda W\varepsilon$) which represents the spatial structure in the error term [32]. It is expressed formally as:

$$\begin{aligned} &y = X\beta + \varepsilon \\ &\varepsilon = \lambda W\varepsilon + \mu \\ &\mu \sim N(0, \sigma^2 I) \end{aligned} \tag{11}$$

Where μ is a white noise term and λ is the autoregressive parameter reflecting the intensity of interdependencies in the perturbation term.

The Spatial Durbin Model is a spatial case of the SLM because the SDM consists of an SLM augmented by spatially lagged explanatory variables [33]. That is, the SDM presents a structure of spatial autocorrelation in the dependent variable and in the independent ones. It is expressed formally as:

$$y = \rho Wy + X\beta + WX\theta + \varepsilon \qquad (12)$$
$$\varepsilon \sim N(0, \sigma^2 I)$$

The equation 12 tells us that y depends of the average value of the dependent variable of its neighbors (Wy), plus k internal factors contained in matrix X, plus the average value of the k factors of its neighbors (WX), plus a vector of residuals.

For example, if y is air pollution in each spatial unit. Thus, i states that air pollution in region 1 might affect pollution in region 2, and viceversa. If X contains a population density, the variable WX would indicate that density of region 1 would affect air pollution in region 2.

Since SDM includes the spatially lagged dependent variable Wy (neighbor dependent variables), the explanatory variables matrix X (proper regional factors), and the spatially lagged explanatory variables WX (neighbor regional factors), it is more general and has a more effective explanatory power than SLM and SEM. Therefore, SDM occupies an interesting position in the field of spatial econometrics [34]. The Table 1 summarize the spatial models of this revision and the classic OLM.

Model	Model name	Matrix formula	Assumptions about the error	Value of the parameters ρ, θ and λ
OLM	Ordinary Linear Model	$y = X\beta + \varepsilon$	$\varepsilon \sim N(0, \sigma^2 I)$	Spatial units are spatially independent and $\rho = 0,\ \theta = 0,\ \lambda = 0$.
SLM	Spatial Lag Model	$y = \rho Wy + X\beta + \varepsilon$	$\varepsilon \sim N(0, \sigma^2 I)$	The spatial dependence is in the dependent variable and $\rho \neq 0,\ \theta = 0,\ \lambda = 0$.
SEM	Spatial Error Model	$y = X\beta + \varepsilon$; $\varepsilon = \lambda W\varepsilon + \mu$	$u \sim N(0, \sigma^2 I)$	The spatial dependence is in the error term and $\rho = 0,\ \theta = 0,\ \lambda \neq 0$.
SDM	Spatial Durbin Model	$y = \rho Wy + X\beta + WX\theta + \varepsilon$	$\varepsilon \sim N(0, \sigma^2 I)$	The spatial dependence is in y and matrix X, and $\rho \neq 0,\ \theta \neq 0,\ \lambda = 0$

Table 1: Summary of the Spatial Models of the Study and OLM Model

Bibliographic Search Method

A structured search of articles in eight databases was carried out to identify existing publications on the topic of socioeconomic factors influencing environmental pollution. The databases consulted were: Taylor & Francis Online – Journals; PMC (PubMed Central); SpringerLink; OneFile (GALE); Social Sciences Citation Index (Web of Science); Directory of Open Access Journals (DOAJ); Scopus (Elsevier); Science Citation Index Expanded (Web of Science). The search terms were: (socioeconomic drivers) AND (environmental pollution index) AND "spatial econometric", the search was performed by title or keywords or abstract, restricting to the papers in the article. More precisely:

"Socioeconomic drivers" was used to find articles that address the influence of socioeconomic factors on environmental pollution. "Environmental pollution index" was used to find articles expressing environmental pollution in terms of indexes. Finally, "spatial econometric" was used to find articles that apply spatial econometric methods to identify the spatial relationship between socioeconomic drivers and environmental pollution. The time of the search was from 1996 to February 2021, resulting in 68 articles found. After a review of the subject matter and methods applied, 22 articles were retained. Only articles applying SLM, SEM, and SDM spatial regression models were selected. In addition, the construction of indices to describe environmental pollution was included in their methodological procedures and the specification of econometric models addressed the analysis of the influence of socioeconomic factors on environmental pollution.

Results

The systematic review results are presented through the descriptive analysis of three sub-themes: internal, external characteristics, and general summary of the articles, which are listed in Figure 4.

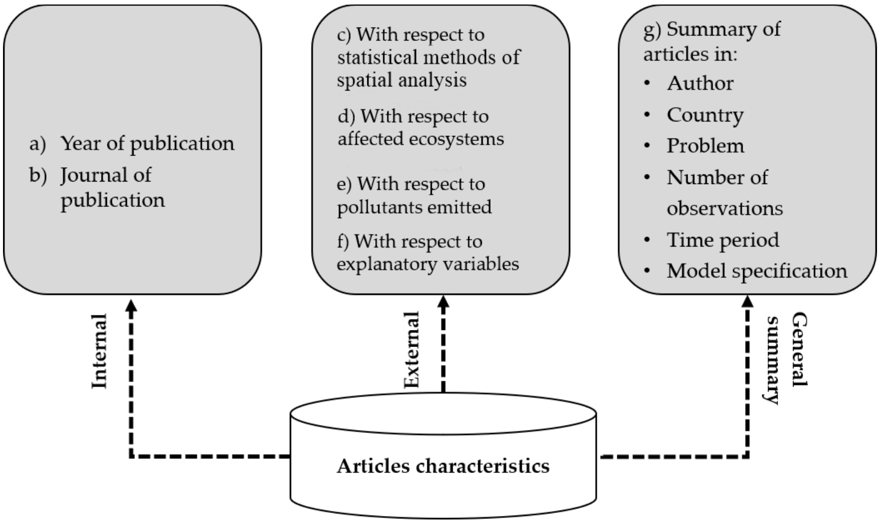

Figure 4: Results Presentation Diagram

Year of Publication

The publications start from 2014, concentrating in 2019 and 2020 with 6 and 10 articles respectively, as shown in Figure 5.

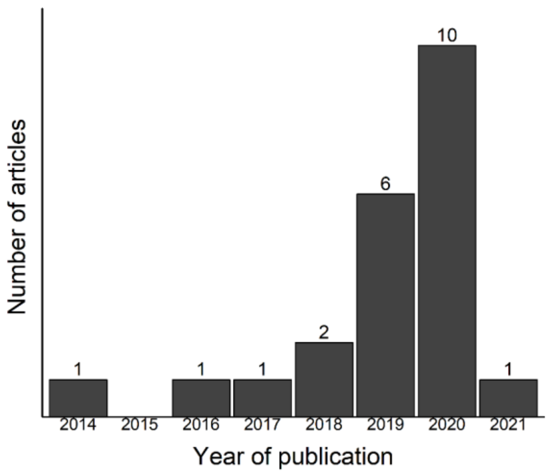

Figure 5: Year of Publication of The Articles

The intention of researchers to carry out studies on environmental quality taking into account spatial relationships has shown an evident increase in recent years. This may be due to the marked presence of environmental problems and the need to apply statistical inference techniques with a spatial approach.

Journal of Publication

Concerning the journals (Table 2), the International Journal of Environmental Research & Public Health was the journal with the highest number of publications and is among the 4 journals with the highest impact factor of the 12 journals.

Academic Journals	Impact Factor	Frequencies
Atmosphere	2.397 (2019)	1
Carbon Management	1.897 (2019)	1
Complexity	2.591 (2018)	1
Discrete Dynamics in Nature and Society	0.970 (2019)	2
Environment, Development, and Sustainability	2.191 (2019)	1
Environmental Economics and Policy Studies	1.430 (2018)	1
Environmental Science and Pollution Research	3.306 (2019)	2
International Journal of Environmental Research & Public Health	3.127 (2019)	8
Journal of Geographical Sciences	3.457 (2019)	1
Public Library of Science	2.740 (2019)	1
Remote Sensing	4.509 (2019)	2
Spatial Economic Analysis	1.231 (2017)	1
Sum		22

Table 2: The Number of Publications by Journals and Impact Factor

Statistical Methods of Spatial Analysis

Table 3 shows the different statistical methods of spatial analysis used by the 22 articles. The articles used the spatial models: SLM, SEM and SDM grouped in (a), and used the spatial indicators: GMI and LMI grouped in (b). (c) refers to articles that used only one spatial model, (d) refers to the articles that used all spatial models together, (e) refers to articles that used only one spatial indicator, and (f) to those that used all statistical methods of spatial analysis together.

In (a) 15 articles adopted SLM and SDM, and 10 articles used SEM. In (b) GMI was used in 15 articles, which was combined with LMI in 5 of the 15 articles. In 10 articles GMI was used as the single model to explore spatial autocorrelation. In (c) 7 articles opted for the SDM and 4 for the SLM. None study used the SEM as the only explanatory model. In (d) 7 articles used all 3 explanatory models together. In (e) only the GMI was used as the only exploratory model by 10 articles. In (f) only two articles used the 3 explanatory models and the 2 exploratory models together.

Articles	**(a)**			**(b)**		**(c)**			**(d)**	**(e)**		**(f)**
	SLM	SEM	SDM	GMI[1]	LMI[2]	Only SLM	Only SEM	Only SDM	Used SLM-SEM-SDM	Only GMI	Only LMI	All
1			x	x	x			x				
2	x	x	x						x			
3	x	x	x	x					x	x		
4			x					x				
5	x	x		x						x		
6	x	x	x						x			
7	x	x		x						x		
8	x					x						
9			x	x				x		x		
10			x					x				
11	x					x						
12	x		x	x						x		
13	x	x	x	x	x				x			x
14	x	x										
15			x	x				x		x		
16			x	x				x		x		
17			x	x	x			x				
18	x	x	x	x					x	x		
19	x	x	x	x	x				x			x
20	x			x		x				x		

21	x	x	x	x					x	x		
22	x			x	x	x						
Sum	**15**	**10**	**15**	**15**	**5**	**4**		**7**	**7**	**10**	**0**	**2**

[1] Global Moran Index; [2] Local Moran Index.

Table 3: Spatial Analysis Models Used by the Articles

SDM and GMI were the most used models. The reason why SDM was adopted more often is, possibly, due to it being a further complete model that includes spatial lags in the dependent variable, as well as in the independent variables. Therefore, this model contains a greater explanatory richness, in terms of analysis of the spatiality between the variables. GMI was used more often because the global test is usually the most frequent choice among researchers. In all the articles, the different models were combined. This is to compare the results of the different methods given the same database.

Ecosystems Affected

This refers to the main elements that are environmental pollution receptors according to the problem of each article. Figure 6 shows that the main pollution receiver of pollution is air with total of 16 articles. In 4 other articles, air pollution is also a factor intended to be analyzed, along with other affected elements such as water and land. This is because most of the articles come from researchers in China, and air pollution is one of the main environmental problems today in China [35].

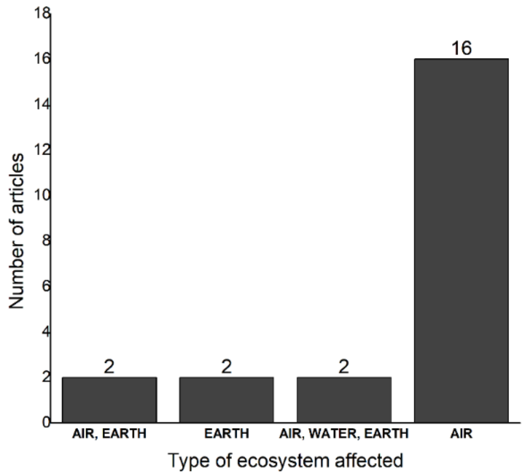

Figure 6: The Number of Articles According to the Type of Ecosystem Affected

Pollutants Emitted

Of the 22 articles analyzed in this review, we found that the main ecosystem affected is the air. Therefore, the main pollutants emitted of this review are those that produce atmospheric pollution such as particulate materials of 2.5 ($PM_{2.5}$) and 10 micrometers (PM_{10}), nitrogen oxides (NO_x, NO_2), carbon (CO_2), and sulfur dioxides (SO_2), and greenhouse gases (GHG). Figure 7 shows the pollutants studied by the articles considered. The dark bars are 16 articles that investigated air pollution, disaggregated by pollutant. The red bar indicates that 2 articles addressed environmental pollution in a comprehensive manner (COMP), i.e., effects on air, water, and land simultaneously.

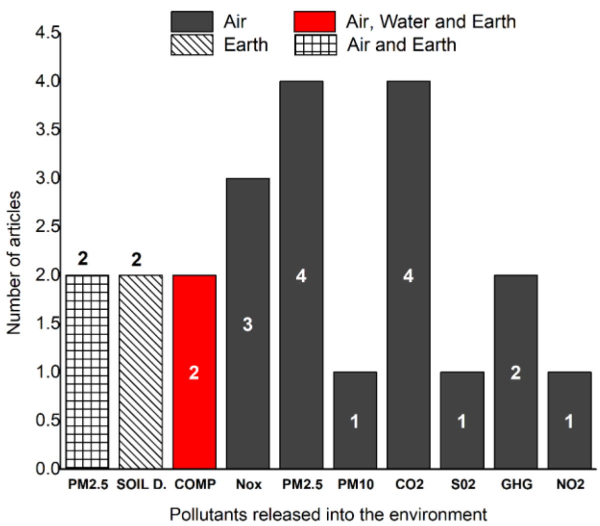

Figure 7: The Number of Articles According to the Main Pollutants Emitted

The diagonal striped bar refers to the effects on land, caused by soil degradation (SOIL D.) with 2 articles. Finally, the bar with squares indicates air and land contamination due to $PM_{2.5}$, since land use activities such as construction generate large amounts of this pollutant and this affects soil and air.

Explanatory Variables Considered

The econometric models used different and numerous explanatory variables. Figure 8 shows the central theme or axes of the variables most used by the models. On the GDP axis, the articles mainly used GDP per capita to represent economic development/growth. In this axis, 20 of the 22 articles, used GDP as one of several socioeconomic drivers explaining environmental pollution.

Foreign direct investment (FDI) was used to refer specifically to the foreign capital inflow per GDP unit, being used by 11 articles. The industry axis was used to refer mainly to secondary industry output as a GDP proportion, referred to in most articles as industrial structure. It was also used to refer to the number of employees in the secondary industry.

The industrial sector, unlike the primary sector (extractivism, agriculture, fishing, etc.) and the tertiary sector (services), focuses on transformation processes that require the use of fossil fuels and, therefore, have significant environmental impacts. Industrial sector importance is reflected in the 19 articles that used this variable as an explanatory factor for pollution.

The population axis was used to represent population density (inhabitants/ km^2), population urbanization or urbanization rate (urban population/total population), and population or population growth per se. This axis captures everything that has to do with population, 13 articles used this type of variable, because there is a direct relationship between pollution and population.

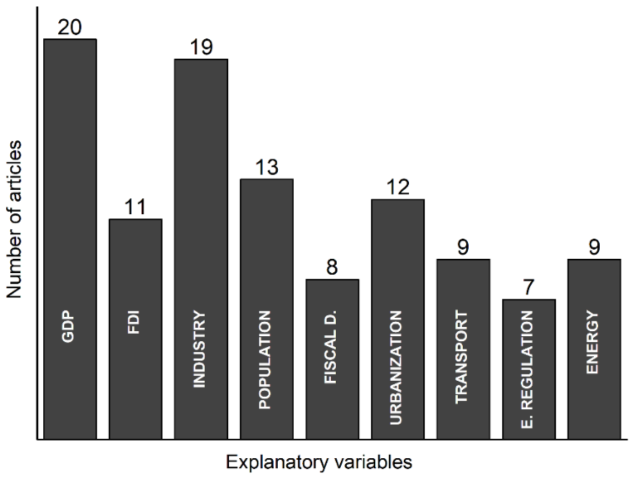

Figure 8: The Number of Articles According to Socioeconomic Drivers Used as Explanatory Variables

The fiscal decentralization axis (Fiscal D.) refers to everything related to public finances. In the 8 articles that used this variable, it was represented as: technical level or technological progress (science and education spending/total public spending), intergovernmental transfers (government transfers/local revenues), and as fiscal decentralization per se, expressed as local spending percentage over total spending. Fiscal decentralization refers to the fact of taking powers away from the central government and assigning them to local governments with their respective financing (i.e., resources transfer to cover the powers). There is no widely accepted consensus on the type of relationship between fiscal decentralization and environmental pollution. However, it is recognized that it has an important impact on the decisions of local governments in terms of socioeconomic and environmental policies.

The urbanization axis, unlike the population axis, which focuses on surface areas, was used in 12 articles, usually referred to as territorial urbanization or urbanized area (urban area/total area). Urbanization was also expressed as the average number of employees at work. The use of this variable type is because it has a direct relationship with environmental pollution.

The transportation axis comprises everything related to human mobilization means, in the 9 articles it was used as transportation intensity expressed as urban road area per capita to measure automobile exhaust emissions, motor vehicle ownership, rail mileage measured by unit area average rail mileage within a given region, road density and as transportation per se. This variable was mainly used in articles addressing air pollution, given the direct relationship between transportation and air pollutants.

Environmental regulation is everything related to environmental policies and investment in environmental protection. This variable was used in 7 of the 22 articles, which tells us that policies to reduce pollution were not variables widely used in the articles. The relationship between environmental policies and pollution is negative.

Concerning the energy axis, this was represented by energy structure (coal consumption/total consumption of fossil fuels), energy intensity or efficiency (energy consumption/GDP), social energy consumption. These types of variables were used by 9 articles, mainly related to CO_2, GHG, and NO_x emissions. Figure 9 refers to the number of times an article used a given explanatory variable to explain a given pollutant. For example, the GDP variable was used 4 times to explain NO_x and 7 times to explain PM.

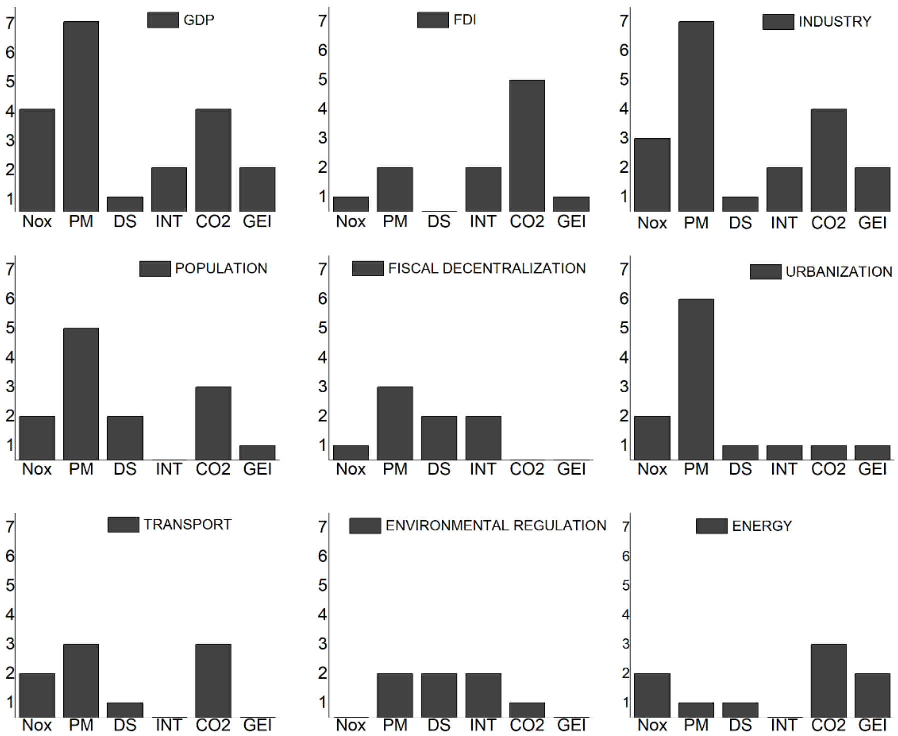

Figure 9: Disaggregation of the Number of Explanatory Variables by Pollutant

Variables related to GDP, industry, population, and urbanization were most used to explain PM. In summary, the graph above shows how the different socioeconomic drivers used to explain various types of environmental pollution are distributed.

General Summary of the Articles

Table 4 presents a summary analysis of each article in the study. The following characteristics are addressed here: the authors, the country under study, the problem, and the specification of the econometric model. The objectives of each article, in general, are to evaluate the impact of certain socioeconomic factors on certain environmental pollutants that are the subject of the problem in the articles.

Most of the articles are studies about environmental problems affecting China. Concerning the problem addressed, they are mostly related to severe levels of air pollution as a product of high rates of economic growth and social development.

AUTHOR / COUNTRY / PROBLEM	ECONOMETRIC MODEL SPECIFICATION
M. Liu, (2021) - CHINA In China, accelerated industrialization and urbanization processes, together with impressive economic and social performance, exacerbate natural resources consumption and the emission of pollutants. Thus, deteriorating environmental quality.	$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X_1} + \boldsymbol{X_2} + \boldsymbol{X_3} + \boldsymbol{X_4} + \boldsymbol{X_5})$; T = 2003 al 2012; N=272; A=10; SDM; GMI; LMI $\boldsymbol{Y}$ =Environmental pollution index, $\boldsymbol{X_1}$= Fiscal Decentralization, $\boldsymbol{X_2}$=Economic Growth (GDP per capita), $\boldsymbol{X_3}$= Foreign Direct Investment (FDI/GDP), $\boldsymbol{X_4}$=Industrial Structure (Industrial GDP/GDP), $\boldsymbol{X_5}$= Population Density (Population/$/Km^2$).
Ge et al., (2018) - CHINA Increased air pollutants greater acceleration in China, nitrogen oxides (NOx), as product of rapid economic growth and urbanization.	$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X_1} + \boldsymbol{X_2})$; T = 2010 al 2015; N=30; A=6; SLM; SEM; SDM $\boldsymbol{Y}$ = Nitrogen oxides (NO_X), $\boldsymbol{X_1}$ = Income (GDP per capita) $\boldsymbol{X_2}$ = Urbanization (Urban population and Total population).
Diao et al., (2018)- CHINA Atmospheric pollution and poor governance result from industrialization and urbanization.	$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X_1} + \boldsymbol{X_2} + \cdots + \boldsymbol{X_6})$; T = 2006 al 2015; N= 31; A= 10; SLM; SEM; SDM; GMI $\boldsymbol{Y}$ = NO_X emissions, $\boldsymbol{X_1}$ = Economic development, $\boldsymbol{X_2}$ = Industrial structure, $\boldsymbol{X_3}$= Energy efficiency, $\boldsymbol{X_4}$ = Urbanization, $\boldsymbol{X_5}$ = Transport, $\boldsymbol{X_6}$ = Population.
Fan et al., (2019) - CHINA Rapid urbanization in China not only promotes rapid urban population expansion and economic agglomeration but also aggravates haze pollution.	$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X_1} + \boldsymbol{X_2} + \boldsymbol{X_3} + \boldsymbol{X_4})$; T= 2001 al 2016; N= 342; A= 16; SDM $\boldsymbol{Y}$ = Haze pollution ($\boldsymbol{PM}_{2.5}$), $\boldsymbol{X_1}$ = Urban scale (Total population), $\boldsymbol{X_2}$ = Urban agglomeration (GDP), $\boldsymbol{X_3}$ = GDP per capita, $\boldsymbol{X_4}$ = Industrial structure.

Shao et al., (2020) - CHINA The temporal and spatial concentration of $PM_{2.5}$ in Hebei as a product of intensive land use.	$\boldsymbol{Y = f(X_1 + X_2 + X_3 + X_4)}$; T= 2001 al 2016; N= 11; A= 4; SLM; SEM; GMI $\boldsymbol{Y}$= Concentrations of $PM_{2.5}$, $\boldsymbol{X_1}$= Intensity of land use, $\boldsymbol{X_2}$= Land use structure, $\boldsymbol{X_3}$= Land input level, $\boldsymbol{X_4}$= Land output benefit.
Liu & Dong, (2019) - CHINA Serious haze pollution resulting from the coexistence of industrialization and urbanization in China, with the main components of haze being fine particles and inhalable particles, which can pose a serious threat to human health.	$\boldsymbol{Y = f(X_1 + X_2 + X_3 + X_4 + \cdots + X_8)}$; T= 2008 a 2016; N= 30; A=9; SLM; SEM; SDM $\boldsymbol{Y}$ =Concentrations of PM_{10}, $\boldsymbol{X_1}$= Industrial transfer, $\boldsymbol{X_2}$ = Population density, $\boldsymbol{X_3}$ =GDP per capita, $\boldsymbol{X_4}$ = Energy (coal consumption/total energy consumption), $\boldsymbol{X_5}$ =Local vehicle usage (km^2 road / km^2 city), $\boldsymbol{X_6}$ = Environmental regulation (Industrial pollution investment / industrial GDP), $\boldsymbol{X_7}$ = Industrial structure (Industrial GDP / GDP), $\boldsymbol{X_8}$ = Weather conditions (precipitation and relative humidity).
Zhong et al., (2020) **EMERGING ECONOMIES** Carbon dioxide intensity ($\boldsymbol{CO_2}$) related to energy.	$\boldsymbol{Y = f(X_1 + X_2 + \cdots + X_5)}$ T= 1995 a 2011; N= 39; A=17; SLM; SEM; SDM; GMI; LMI $\boldsymbol{Y_1}$= GHG incorporated in import trade, $\boldsymbol{Y_2}$ = GHG incorporated in export trade, $\boldsymbol{X_1}$= Population size, $\boldsymbol{X_2}$= GDP, $\boldsymbol{X_3}$= Energy intensity, $\boldsymbol{X_4}$= Percentage of clean energy/energy consumption, $\boldsymbol{X_5}$= Percentage of Industrial GDP.

Wang et al., (2022) - CHINA China's rapid industrialization and urbanization coupled with high annual emissions of pollutants have caused the air quality of Chinese cities to deteriorate significantly, threatening the public health and well-being of urban residents.	$Y = f(X_1 + X_2 + X_3 + \cdots + X_{10})$; T= 2014 al 2017; N= 388; A=4; SLM Y = Concentration of $PM_{2.5}$, X_1= GDP per capita, X_2 = FDI, X_3 = Population density, X_4 = Urbanized area, X_5 = Urban greening rate, X_6 = Number of own vehicles, X_7 = Percentage of industrial GDP, X_8 = Total energy consumption, X_9 = Science and technology expenditure to GDP, X_{10} = Soot emissions.
Zheng et al., (2019) - CHINA Large-scale urban construction, urban population aggregation, and the rapid growth in the number of motor vehicles accompanied by urban development have led to the deterioration of air quality in urban areas in China.	$Y = f(X_1 + X_2 + X_3 + \cdots + X_6)$; T= 2018 al 2019; N=334; A=2; SLM; GMI; LMI Y = Nitrogen Dioxide (NO_2), X_1 = Population, X_2 = GDP, X_3 = Industrial GDP, X_4 = Social energy consumption X_5 = Total energy consumption, X_6 = Ownership of motor vehicles.
Ronaghi et al., (2020) **OPEP COUNTRIES** According to the World Bank, energy-related CO_2 emissions from developing countries will be 127% higher in the world's most developed economies by 2040. Due to their generally stronger economic growth and continued use of fossil fuels and since oil is the most prominent fossil fuel. OPEC countries are the largest emitters of CO_2.	$Y = f(X_1 + X_2 + X_3 + X_4 + X_5 + X_6)$; T= 2006 al 2015; N= 13; A= 13; SDM Y = Carbon dioxide emissions (CO_2) X_1 = GDP growth (annual %), X_2 = Exports as a percentage of GDP X_3 = Imports as a percentage of GDP, X_4 = Inflation rate (annual percentage) X_5 = Employment (percentage of total work), X_6 = Governance.

Ren et al., (2019) - CHINA With the reform and opening up, the Chinese economy has achieved rapid development; however, the interdependence between economies also involves the interaction of energy consumption and carbon emissions. World Bank data indicated that China has become the largest carbon emitter in the world since 2008.	$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X_1} + \boldsymbol{X_2} + \boldsymbol{X_3} + \cdots + \boldsymbol{X_7})$; T= 2004 a 2016; N= 284; A= 13; SLM; SEM; SDM; GMI $\boldsymbol{Y} = CO_2$ emissions, $\boldsymbol{X_1}$= Economic development (GDP per capita), $\boldsymbol{X_2}$ = Energy efficiency, $\boldsymbol{X_3}$ = Industrial structure (% employed in secondary industry), $\boldsymbol{X_4}$ = Energy structure (% coal consumption), $\boldsymbol{X_5}$ = Population density, $\boldsymbol{X_6}$ = Traffic facilitation ($km^2\ road/\ km^2$ city), $\boldsymbol{X_7}$ = FDI.
Gómez-Antonio et al., (2016) SPAIN Over the past 20 years, rates of conversion to residential land use have far exceeded population growth rates on the continent [20% vs. 6%]. As a result, the amount of urbanized land consumed per person during this period has more than doubled, leading to the formation of new peripheral cities around traditional urban centers and scattered residential developments located on the urban periphery.	$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X_1} + \boldsymbol{X_2} + \cdots + \boldsymbol{X_{25}})$; T= 1990 al 2000; N= 3131; A=11; SLM $\boldsymbol{X_1}$= Population growth, $\boldsymbol{X_2}$= Vehicles per household, $\boldsymbol{X_3}$= % of population between 25 and 45 years old, $\boldsymbol{X_4}$= % of population over 65 years old, $\boldsymbol{X_5}$= Average number of children per household, $\boldsymbol{X_6}$= % of residents with a high degree of education, $\boldsymbol{X_7}$= Poverty level approximated by % of resident population without schooling, $\boldsymbol{X_8}$= % Immigrants, $\boldsymbol{X_9}$= % Employed in manufacturing, $\boldsymbol{X_{10}}$= % Retail Employees, $\boldsymbol{X_{11}}$= % Employed in other services, $\boldsymbol{X_{12}}$= Maximum average temperature, $\boldsymbol{X_{13}}$= Minimum average temperature, $\boldsymbol{X_{14}}$= Average precipitation, $\boldsymbol{X_{15}}$= % Open space, $\boldsymbol{X_{16}}$= % Sports and leisure facilities, $\boldsymbol{X_{17}}$= Water availability (%), $\boldsymbol{X_{18}}$= Inland waters (%), $\boldsymbol{X_{19}}$= Terrain roughness index, $\boldsymbol{X_{20}}$= Elevation range (km), $\boldsymbol{X_{21}}$ = Road density, $\boldsymbol{X_{22}}$ = Distance to road (km), $\boldsymbol{X_{23}}$ = Distance to city center, $\boldsymbol{X_{24}}$= Property tax revenues, $\boldsymbol{X_{25}}$= Intergovernmental transfers as % of local revenues.

Han et al., (2020) - CHINA Rapid urbanization in China causes major impacts on land use and scarcity of land resources becoming a constraint to sustainable urban development.	$Y = f(X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8)$; T= 2010 a 2016; N= 287; A= 7; SLM; SEM; GMI Y = Urban land use efficiency, X_1 =Economic agglomeration (population density), X_2 = Industrial structure (GDP tertiary industry/GDP secondary industry), X_3 = Economic development (GDP per capita), X_4 = Government intervention (financial expenditure per capita), X_5 = Investment in science and education (science and education expenditure / fiscal expenditure), X_6 = Environmental governance (Gas treatment, wastewater treatment, and waste disposal fees), X_7 = Energy consumption intensity (Unit of GDP per energy consumption), X_8 = Marquetization of land
Chen, (2016) - CHINA Serious environmental problems in China are the result of economic development and high levels of industrial consumption and emissions. This pollution is accompanied by negative externalities in the form of overflow effects in pollutant discharges.	$Y = f(X_1 + X_2 + X_3 + X_4 + X_5 + X_6)$; T= 2003 al 2017; N= 31; A= 15; SDM; GMI Y = Environmental pollution index, X_1= Fiscal Decentralization, X_2= Environmental regulation, X_3 = GDP per capita, X_4 = Industrial structure, X_5= FDI, X_6 = Urbanization.
Long et al., (2020) - CHINA Resource scarcity and increasingly severe environmental pollution in China.	$Y = f(X_1 + X_2 + X_3 + X_4 + X_5)$; T = 2008 a 2017; N= 11; A= 10; SDM; GMI Y = Efficiency of industrial green technology innovation X_1= Economic development, X_2 = Environmental regulation, X_3 = Government support for technological innovation (Government funding of R&D funds), X_4 = FDI, X_5 = Industrial structure.

Yue et al., (2020) - CHINA With the rapid growth of economic development, the imbalance of energy supply and demand poses a critical challenge to our country's energy security. Meanwhile, the increasing and excessive energy consumption lead to the greenhouse effect and air pollution, greatly threatening the survival and development of human beings.	$Y = f(X_1 + X_2 + X_3 + X_4)$; T= 1995 a 2016; N= 354; A= 22; SDM; GMI Y = Energy intensity (total energy consumption / GDP), X_1 = Economic growth (GDP), X_2 = Urbanization rate (urban population / total population), X_3 = Industrial structure (industrial GDP / GDP), X_4 = FDI
Gan et al., (2021) - CHINA With sustained economic development, the ecological environment of China is becoming increasingly fragile and the problem of haze pollution is becoming more and more prominent, which has affected the normal life of human beings and the stable development of society.	$\mathbf{Y} = \mathbf{f}(\mathbf{X_1} + \mathbf{X_2} + \mathbf{X_3} + \cdots + \mathbf{X_8})$; T= 1998 a 2016; N= 287; A= 19; SDM; GMI; LMI $\mathbf{Y}$ = Haze measurement index ($\mathrm{PM}_{2.5}$), $\mathbf{X_1}$ = Economic development ($\mathrm{Ln}_{\mathrm{GDP}}$), $\mathbf{X_2}$ = Population density, $\mathbf{X_3}$ = Industrial structure, $\mathbf{X_4}$ = Urbanization (average number of employees at work), $\mathbf{X_5}$ = Technical level (% science expenditure in public finances), $\mathbf{X_6}$ =FDI, $\mathbf{X_7}$ = Market scale (total retail sales of social consumer goods), $\mathbf{X_8}$= Transportation intensity (urban road area per capita).
Xie et al., (2019) - CHINA Over the past four decades, urbanization in China has rapidly increased from 17.92% in 1978 to 58.52% in 2017. This rapid urbanization is accompanied by urban population agglomeration, urban land use, and serious industrial emissions leading to high ambient air pollution.	$\mathbf{Y} = \mathbf{f}(\mathbf{X_1} + \mathbf{X_2} + \mathbf{X_3} + \mathbf{X_4} + \mathbf{X_5})$; T= 2006 a 2016; A=11; SLM $\mathbf{Y}$ = $\mathrm{PM}_{2.5}$, $\mathbf{X_1}$ = Urbanized population, $\mathbf{X_2}$ = Population density, $\mathbf{X_3}$ = GDP per capita $\mathbf{X_4}$ = Technological progress (% financial expenditure on science and education), $\mathbf{X_5}$ = Industrial structure.

Cheng et al., (2020) - CHINA The Yangtze River Delta [YRD] is a rapidly urbanizing region in China that experienced severe haze pollution in recent decades.	$\mathbf{Y} = \mathbf{f}(\mathbf{X_1} + \mathbf{X_2} + \cdots + \mathbf{X_4})$; T=2002 a 2017; N= 41; A= 16; SLM; SEM; SDM; GMI $\mathbf{Y} = PM_{2.5}$, $\mathbf{X_1}$ = Population urbanization, $\mathbf{X_2}$ = Territorial urbanization (% urban built-up area), $\mathbf{X_3}$ = Economic urbanization (regional GDP per capita), $\mathbf{X_4}$ = Secondary industry, $\mathbf{X_5}$ = Vegetation cover (vegetation area/total area), $\mathbf{X_6}$ = Precipitation, $\mathbf{X_7}$ = Wind speed.
Jiang et al., (2020) - CHINA China has already become the world's largest coal consumer since 2010. Consequently, pollutants such as sulfur dioxide [SO_2] have been emitted enormously for years, which has led China to be the world's largest emitter of pollutants. In particular, over the past two decades, most Chinese cities have still suffered from the severity and extent of SO_2 pollution.	$\mathbf{Y} = \mathbf{f}(\mathbf{X_1} + \mathbf{X_2} + \cdots + \mathbf{X_6})$; T= 2008 a 2017; N= 30; A= 10; SLM; SDM; GMI Y= Per capita Sulfur Dioxide Emissions (SO_2) $\mathbf{X_1}$= GDP, $\mathbf{X_2}$= FDI, $\mathbf{X_3}$= Secondary industry, $\mathbf{X_4}$= The commerce, $\mathbf{X_5}$= Fossil fuel combustion, $\mathbf{X_6}$= Environmental regulation.
Lv et al., (2021) - CHINA Chinese scale-driven economic growth mode has achieved spectacular success in recent decades. At the same time, rapid economic development has given rise to enormous challenges related to environmental issues, such as rising energy consumption and energy-related CO_2 emissions.	$\mathbf{Y} = \mathbf{f}(\mathbf{X_1} + \mathbf{X_2} + \mathbf{X_3} + \mathbf{X_4})$; T= 1998 a 2015; N=31; A=18; SLM; SEM $\mathbf{Y}$ = Carbon Dioxide Emissions (CO_2) $\mathbf{X_1}$= Energy structure (% coal consumption/total energy consumption), $\mathbf{X_2}$= Industrial structure (% of industrial GDP/total GDP), $\mathbf{X_2}$ = Urbanization level (urban population and total population), $\mathbf{X_3}$= FDI, $\mathbf{X_4}$ = Rail mileage (unit area average rail mileage within a given region).

Cheng et al., (2014) - CHINA	
Sustainable development has been seriously challenged by global climate change due to carbon emissions. As a developing country, China pledged to reduce its carbon intensity by 40% to 45% below the 2005 level by 2020. The realization of this target depends not only on the substantial transition of the society and economy at the national scale but also the action and participation of energy-saving and emission reduction at the provincial scale.	$\mathbf{Y} = \mathbf{f}(\mathbf{X_1} + \cdots + \mathbf{X_8})$; T= 1997 a 2010; N=30; A= 14; SLM; SEM; SDM; GMI; LMI Y= Carbon intensity, $\mathbf{X_1}$ = Total population, $\mathbf{X_2}$ = GDP per capita, $\mathbf{X_3}$ = Energy intensity (energy consumption/GDP), $\mathbf{X_4}$ = Energy structure (% coal consumption/total energy consumption), $\mathbf{X_5}$ = Industrial structure, $\mathbf{X_6}$ = Urbanization rate (urban population/total population), $\mathbf{X_7}$ = Foreign trade openness ((Import - Export) / GDP), $\mathbf{X_8}$ = FDI.

Table 4: A general Summary of Articles from the Literature Review

The model specification is presented in functional form, $Y = f(X)$. Certain characteristics are also presented such as T which refers to the time period, N is the sample size and A is the number of years of the study. The statistical methods of spatial analysis used in each study are also mentioned here.

Conclusions

Spatial econometric techniques tend to be increasingly used in the empirical research field, providing advances towards the knowledge of the spatial interdependence relationships between socioeconomic drivers that directly and indirectly influence the environment quality.

This literature review presents the basic theoretical aspects of spatial econometrics, as well as an overview of publications in the area, highlighting applications of those techniques in the study of environmental pollution problems and presenting several articles that applied the most widely used methods of spatial econometrics (spatial models and spatial indicators) to try to answer how socioeconomic drivers impact the environment, discovering how these impacts occur from a spatial perspective. Spatial phenomena are relevant in environmental sciences, as much as in economic-business sciences, health sciences, among others. Moving towards appropriate methods of spatial estimation improves the quality of the results.

As environmental problems become more acute, as is the case with severe air pollution problems in China, researchers should be able to apply diverse types of methods to promote the search for answers which could help to implement appropriate public policies for pollution mitigation. Among the main limitations of this review, it could be mentioned that we only rely on reviewing statistical methods for spatial analysis from a frequentist perspective. Bayesian methods form a large body of theory and have also been developed for spatial analysis [36-57].

Author Contributions

M.C. designed the study, analyzed the articles, and wrote the manuscript. L.E. revised the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Touchon, J. C., & McCoy, M. W. (2016). The mismatch between current statistical practice and doctoral training in

ecology. Ecosphere, 7(8), e01394.
2. Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). Quantitative geography: Perspectives on spatial data analysis. Sage.
3. Anselin, L. (2001). Spatial Effects in Econometric Practice in Environmental and Resource Economics. American Journal of Agricultural Economics, 83.
4. Gaspard, G., Kim, D., & Chun, Y. (2019). Residual spatial autocorrelation in macroecological and biogeographical modeling: A review. Journal of Ecology and Environment, 43(1), 19.
5. Goodchild, M. F., Anselin, L., Appelbaum, R. P., & Harthorn, B. H. (2000). Toward spatially integrated social science. International Regional Science Review, 23(2), 139-159.
6. Bockstael, N. E. (1996). Modeling economics and ecology: The importance of a spatial perspective. American Journal of Agricultural Economics, 78(5), 1168-1180.
7. Bohórquez, I. A., & Ceballos, H. V. (2008). Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. Ecos de Economía: A Latin American Journal of Applied Economics, 12(27), 9-2.
8. Uriel, E. (2013). El modelo de regresión simple: Estimación y propiedades. Introducción a la econometría, 1-49.
9. Pérez Pineda, J. A. (2006). Econometría espacial y ciencia regional. Investigación económica, 65(258), 129-160.
10. LeSage, J. P. (2004). Lecture 1: Maximum likelihood estimation of spatial regression models. Spatial Econometrics Course 2006.
11. Fotheringham, A. S. (2009). "The problem of spatial autocorrelation" and local spatial statistics. Geographical analysis, 41(4), 398-403.
12. Yrigoyen, C. C. (2004). Modelos de heterogeneidad espacial. University Library of Munich, Germany, Tech. Rep.
13. Herrera Gómez, M., & Cid, J. C. (2015). Fecundidad, determinantes socioeconómicos e interacciones sociales. Un análisis de heterogeneidad espacial para la Argentina.
14. Vayá, V., & Moreno, S. (2000). TECNICAS ECONOMETRICAS PARA EL TRATAMIENTO DE DATOS ESPACIALES: LA ECONOMETRIA ESPACIAL. Edicions Universitat Barcelona.
15. Griffith, D. A. (2019). Negative Spatial Autocorrelation: One of the Most Neglected Concepts in Spatial Statistics. Stats, 2(3), Art. 3.
16. Brum-Bastos, V., Azevedo, T., Sallum, M., & Demšar, U. (2016, julio 2). Geostatistical spatial modelling of dengue fever across São Paulo state, Brazil.
17. Hu, L. (2020). Negative Spatial Autocorrelation and its Impacts on Georeferenced Data Analyses: With Case Studies of Cancer Incidences.
18. Biswas, S. R., MacDonald, R. L., & Chen, H. Y. (2017). Disturbance increases negative spatial autocorrelation in species diversity. Landscape Ecology, 32(4), 823-834.
19. Horlent, M., Arturi, M., Cellini, J. M., Casal, D. P., Buus, J. M., & Goya, J. F. (2003). Crecimiento y competencia intraespecífica en Celtis tala en el este de Buenos Aires (Argentina). Agriscientia, 20, 79-84.
20. Márquez, A. (2021). Competencia interespecífica: Qué es, características y ejemplos. ecologiaverde.com.
21. Anselin, L. (1988). Spatial Econometrics: Methods and Models. Springer Netherlands.
22. LeSage, J. P. (1999). The theory and practice of spatial econometrics (Vol. 28). Citeseer.
23. Yrigoyen, C. C. (2002). Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales
24. Tobler, W. (2004). On the first law of geography: A reply. Annals of the Association of American Geographers, 94(2), 304-310.
25. Benjanuvatra, S. (2012). Essays on Spatial Econometrics [PhD Thesis]. University of York.
26. Fischer, M. M., & Getis, A. (2010). Handbook of applied spatial analysis: Software tools, methods and applications. Springer.
27. Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. Biometrika, 37(1/2), 17-23.
28. Geary, R. C. (1954). The contiguity ratio and statistical mapping. The incorporated statistician, 5(3), 115-146.
29. Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. Geographical analysis, 24(3), 189-206.
30. Yrigoyen, C. C. (2003). Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales. Dirección General de Economía y Planificación.
31. Anselin, L., & Rey, S. (1991). Properties of tests for spatial dependence in linear regression models. Geographical analysis, 23(2), 112-131.
32. Kissling, W. D., & Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecology and Biogeography, 17(1), 59-71.
33. Borrego Sánchez, J. Á. (2018). Modelos de regresión para datos espaciales.
34. Mur, J., & Angulo, A. (2006). The Spatial Durbin Model and the Common Factor Tests. Spatial Economic Analysis, 1(2), 207-226.
35. Liu, M. (2021). The spatial clustering analysis of COVID-19 and its associated factors in mainland China at the prefecture level. Science of The Total Environment, 145992-145992.
36. Chen, Y. (2016). Spatial Autocorrelation Approaches to Testing Residuals from Least Squares Regression. PLOS ONE, 11(1), e0146865.
37. Cheng, L., Zhang, T., Chen, L., Li, L., Wang, S., Hu, S., Yuan, L., Wang, J., & Wen, M. (2020). Investigating the impacts of urbanization on PM2. 5 pollution in the yangtze river delta of China: A spatial panel data approach. Atmosphere, 11(10), 1058.
38. Cheng, Y., Wang, Z., Ye, X., & Wei, Y. D. (2014). Spatiotemporal dynamics of carbon intensity from energy consumption

in China. Journal of Geographical Sciences, 24(4), 631-650.
39. Diao, B., Ding, L., Su, P., & Cheng, J. (2018). The spatial-temporal characteristics and influential factors of NOx emissions in China: A spatial econometric analysis. International journal of environmental research and public health, 15(7), 1405.
40. Fan, Q., Yang, S., & Liu, S. (2019). Asymmetrically spatial effects of urban scale and agglomeration on haze pollution in China. International Journal of Environmental Research and Public Health, 16(24), 4936.
41. Gan, T., Yang, H., Liang, W., & Liao, X. (2021). Do economic development and population agglomeration inevitably aggravate haze pollution in China? New evidence from spatial econometric analysis. Environmental Science and Pollution Research, 28(5), 5063-5079.
42. Ge, X., Zhou, Z., Zhou, Y., Ye, X., & Liu, S. (2018). A spatial panel data analysis of economic growth, urbanization, and NOx emissions in China. International journal of environmental research and public health, 15(4), 725.
43. Gómez-Antonio, M., Hortas-Rico, M., & Li, L. (2016). The causes of urban sprawl in Spanish urban areas: A spatial approach. Spatial Economic Analysis, 11(2), 219-247.
44. Han, X., Zhang, A., & Cai, Y. (2020). Spatio-econometric analysis of urban land use efficiency in China from the perspective of natural resources input and undesirable outputs: A case study of 287 cities in China. International Journal of Environmental Research and Public Health, 17(19), 7297.
45. Jiang, L., Zhou, H., & He, S. (2020). The role of governments in mitigating SO2 pollution in China: A perspective of fiscal expenditure. Environmental Science and Pollution Research, 27(27), 33951-33964.
46. Liu, Y., & Dong, F. (2019). How industrial transfer processes impact on haze pollution in China: An analysis from the perspective of spatial effects. International journal of environmental research and public health, 16(3), 423.
47. Long, R., Guo, H., Zheng, D., Chang, R., & Na, S. (2020). Research on the measurement, evolution, and driving factors of green innovation efficiency in Yangtze river economic belt: A super-sbm and spatial durbin model. Complexity, 2020.
48. Lv, K., Cheng, Y., & Wang, Y. (2021). Does regional innovation system efficiency facilitate energy-related carbon dioxide intensity reduction in China? Environment, Development and Sustainability, 23(1), 789-813.
49. Ren, Y., Ren, X., & Hu, J. (2019). Driving factors of China’s city-level carbon emissions from the perspective of spatial spillover effect. Carbon Management, 10(6), 551-566.
50. Ronaghi, M., Reed, M., & Saghaian, S. (2020). The impact of economic factors and governance on greenhouse gas emission. Environmental Economics and Policy Studies, 22(2), 153-172.
51. Shao, J., Ge, J., Feng, X., & Zhao, C. (2020). Study on the relationship between PM2. 5 concentration and intensive land use in Hebei Province based on a spatial regression model. PLoS One, 15(9), e0238547.
52. Vilalta, C. J. (2005). Cómo enseñar autocorrelación espacial. Economía, Sociedad y Territorio, 18, 323-333.
53. Wang, S., Zhang, M., Huang, X., Hu, T., Li, Z., Sun, Q. C., & Liu, Y. (2022). Urban-regional disparities in mental health signals in Australia during the COVID-19 pandemic: A study via Twitter data and machine learning models. Cambridge Journal of Regions, Economy and Society, rsac025.
54. Xie, W., Deng, H., & Chong, Z. (2019). The spatial and heterogeneity impacts of population urbanization on fine particulate (PM2. 5) in the Yangtze River Economic Belt, China. International journal of environmental research and public health, 16(6), 1058.
55. Yue, Y., Tian, L., Yue, Q., & Wang, Z. (2020). Spatiotemporal variations in energy consumption and their influencing factors in China based on the integration of the DMSP-OLS and NPP-VIIRS nighttime light datasets. Remote Sensing, 12(7), 1151.
56. Zheng, Z., Yang, Z., Wu, Z., & Marinello, F. (2019). Spatial variation of NO2 and its impact factors in China: An application of sentinel-5P products. Remote Sensing, 11(16), 1939.
57. Zhong, Z., Zhang, X., & Gao, W. (2020). Spatiotemporal evolution of global greenhouse gas emissions transferring via trade: Influencing factors and policy implications. International journal of environmental research and public health, 17(14), 5065.

Foot Notes

[1]Spatial autocorrelation is a weaker expression of spatial dependence, relating only to the first moments of the joint distribution of a variable (Vayá & Moreno, 2000). Spatial autocorrelation analyzes are exploratory, where an indicator is used to detect and quantify the phenomenon. While spatial dependence analyzes are explanatory. That is, it seeks to explain why these dependence relationships occur at the geographical level, based on an underlying theoretical context (Vilalta, 2005)