Master's Thesis

The Effect of Natural and Socioeconomic Factors on Air Pollution Change in China: A Spatial Panel Analysis

53618015

Lina Ren

Graduate School of Fisheries and Environmental Sciences

Nagasaki University

Supervisor: Ken'ichi Matsumoto

February, 2020

Abstract:

In recent years, with the rapid growth of China's economy and the high level of energy consumption, the problem of air pollution has become one of the most important environmental issues to be resolved. In the past literature, we found that (i) few systematic studies using panel data to combine natural factors and socioeconomic factors, and (ii) most of them are concentrated in national or part of areas. In order to analyze the spatial effect and effective factors of air pollution, this paper takes SO_2 and NO_X emissions as the research object of air pollution, and used spatial econometric model to analyzed the panel data of 31 provinces in China from 2011 to 2017. First, we analyzed the spatial distribution and spatial autocorrelation of two air pollutants through maps and Moran's I. We found that air pollution has significant and strong spatial autocorrelation. Then, we selected the fixed-effect spatial lag model through Hausman test and Lagrange Multiplier test. Finally, we analyzed the effect of natural and socioeconomic factors on air pollution through fixed-effect spatial lag model. The results show that heating degree days, population, added value from secondary industry, and urbanization rate had positive and statistically significant impact on the air pollution emissions, while cooling degree days, per capita GDP, population density and relative humility had significant negative effects. Precipitation had no significant effect. By analyzing the effect factors of air pollution, we have put forward related suggestions such as strengthening regional cooperation, vigorously developing the tertiary industry, and promoting industrial upgrading to improve air pollution.

Keywords: Sulphur dioxide emissions, Nitrogen oxides emissions, natural factors, socioeconomic factors, spatial autocorrelation, spatial econometric models

Contents

List of Figures and Tables

1. Introduction

1.1 Background

Over the past few decades, with economic growth rapidly and large-scale urbanization, China's harmful gas emissions have increased significantly due to the upsurge in energy use. As a result, pollution problem has become increasingly serious. Li and Zhang, (2014) pointed out that the persistent large-scale air pollution not only hindered China's economic development, but also adversely affected daily life and people's health. The way to effectively control discharging of air pollutants and to improve urban ambient air quality has proved to be important goals of social and economic transformation of development in China (Wang et al., 2014).

The Chinese government has been trying to alleviate the problem of air pollution. In China's Twelfth Five-Year Plan (2011–2015), control of nitrogen oxides (NO_X) emissions, sulfur dioxide (SO_2), and other major particulate matter was implemented nationwide, and NO_X emissions were first be incorporated into the constraint indicator system (Ding et al., 2017). Although the plan has been amended many times, the situation of severely polluted air has not been effectively improved. According to the reference in 2017, 99 of the 338 (29.3%) prefecture-level cities and above nationwide met ambient air quality standards. Similarly, a 2013 report by the Asian Development Bank and Tsinghua University scholar states that seven out of the world's ten worst air quality cities are located in China (Huang, 2018). Less than five of China's 500 cities meet air quality standards set by the World Health Organization (WHO) (Huang 2018). Therefore, air pollution has been one of the most concerned and anticipated problems for a long time.

1.2 Literature review

Recent studies on air pollution have focused on pollutant emissions (SO_2 and NO_X), and pollutant concentrations (particulate matter (PM) and air quality index (AQI). The influencing factors of air pollution are extensive and complex, and many researchers have worked to discover the causes of changes in air pollutant emissions in China. Table 1 summarizes the literature on the factors affecting air pollution, which mainly includes natural factors and socioeconomic factors. About natural factors, Requia et al., (2019) employed generalized additive models to estimate weather-associated changes in $PM_{2.5}$ composition in the US during 1988–2017. They found that wind speed and relative humidity were associated with the most $PM_{2.5}$ components during both warm and cold seasons. Feng et al., (2019) aimed to identify the dominant variable of the air pollution using global datasets of fine $PM_{2.5}$ concentrations, precipitation, and air temperature. The results show that air pollution is mainly negatively correlated with precipitation and positively correlated to temperature in tropical, arid, and temperate regions. However, the conditions are much more complex in cold regions. Chen et al., (2016) used Spearman-Rank analysis and the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise method to study the influencing factors of $PM_{2.5}$ in Nanjing, China from 2013 to 2015. They found that $PM_{2.5}$ exhibits a reversed relation with wind speed, relative humidity, and precipitation. Although temperature had a positive association with $PM_{2.5}$ in most months, a negative correlation was observed during the entire period.

About socioeconomic factors, Chang et al., (2018) decomposed SO_2 and NO_X emission variation into the change of socioeconomic driving factors in China using the logarithmic mean Divisia index method. They showed that energy intensity and economic growth have a large positive impact on the increasing of SO_2 and NO_X emissions. Peng et al., (2018), based on a noncompetitive (import) input-output model for China from 1995 to 2009, showed that economic expansion played an important role in accelerating air pollution emissions, and fixed capital formation is the main driver of air pollutant emissions, followed by household consumption and exports. Zeng et al., (2019) employed a spatial econometric model to empirically test the effects of two types of energy policies on China's emissions of major air pollutants, namely PM_{10}, $PM_{2.5}$, and SO_2 emissions. The results offer evidence that provincial emission reduction policies have positive impacts on reduction of PM_{10}, whereas provincial renewable energy policies have positive impacts on the reduction of SO_2 and $PM_{2.5}$. Hu et al., (2019) used the structural equation modelling to quantify the contributing effects of various driving forces of air pollution in 2015 in prefecture-level cities of China. The results showed that industrial scale, city size, and residents' activities have a significant impact on NO_X pollution. Hu et al., (2019) estimated the linkages among total SO_2 emissions, total GDP and energy efficiency using China's provincial panel data from 2002 to 2015. The analysis shows that GDP has a positive impact on total SO_2 emissions in the short run and energy efficiency has a significant negative effect on emissions in the long run. Wang et al., (2016) used the Stochastic Impacts by Regression on Population, Affluence and Technology regression model to analyze the relationship between socioeconomic factors (economic growth, income, and urbanization) and SO_2 emissions in China, and indicated an inverted U-shaped curve relationship between economic level and SO_2 emissions. This suggests an environmental Kuznets curve in SO_2 emissions. Fu and Li (2020) used spatial econometric models and geographic and temporal weighted regression model to analyze the relationship between global socioeconomic factors and $PM_{2.5}$ in the global scale from 2000–2014. The results suggested that renewable energy consumption ratio, per capita GDP, per capita CO_2 emission, urban population ratio, and fossil fuel consumption ratio were major factors responsible for the global $PM_{2.5}$ pollution. Ryou et al., (2018) selected estimated sources of PM_{10} and $PM_{2.5}$ contributions performed for 2000–2017 in South Korea using Positive Matrix Factorization and Chemical Mass Balance. They found that secondary aerosol and motor vehicle contributed highly to PM_{10} and $PM_{2.5}$, while the contribution of combustion/industry was high for PM_{10}. Zhao et al., (2019) also used a multiple linear regression model analysis to evaluate the relationship between $PM_{2.5}$ concentration and socioeconomic factors from 2015 to 2016. They found that population density and the share of the secondary industry were key factors in controlling air pollution. Zhao et al., (2018) used a panel data model (2004–2012) to quantify the effective factors of ambient $PM_{2.5}$ concentrations in five hot spots in China. The results show that GDP and private cars are important positive factors for $PM_{2.5}$ concentration. Other studies indicated that urbanization (Hao et al., 2020) and the number of vehicles (Xu et al., 2019) have a positive impact on air pollution. McCarty and Kaza, (2015) investigate the relationship between urban spatial structure and air quality in the United States. Controlling for demographic variables and economic activity, they found a strong relationship between the type and pattern of economic development and pollutant levels. Although these studies helped us to understand the causes of the reductions in air pollution, they focused only on socioeconomic factors but do not consider natural factors.

There are also comprehensive studies considering of both natural and socioeconomic factors. Yang et al., (2017) used a series of global regression models (ordinary least squares model, spatial lag model, and spatial error model) and local geographic weighted regression models to process 2014 data for 113 major cities in China. The results indicated that precipitation exerts a significant effect on SO_2 reduction. Both emission and temperature factors were found to aggravate SO_2 concentrations, although no such significant correlation was found in relation to wind speed. Liu et al., (2017) quantitatively estimated the contribution and space spillover of different natural and socioeconomic factors to the AQI of 289 prefecture-level cities in 2014. They found that urbanization, urban population aggregation and industrialization had a significant positive impact on the AQI. The spillover effect of car density is also positive significant. Except for temperature is insignificant, all of natural factors had a negative impact on AQI. Han et al., (2019) utilized local regression models to explore the main influential factors on AQI in China. They found that there are spatial differences in the effects of different factors on the AQI.

Table 1 Summary of the factors effecting air pollution in the literature.

Factors	Literature	Region	Period	Method	Variables
Natural factors	Feng et al. （2019）	Global	1998–2015	Geographically Weighted Regression; Tropical Rainfall Measuring Mission; anomaly interpolation approach	Dependent variable: $PM_{2.5}$ concentrations Effect factors: precipitation, temperature
	Chen et al. (2016)	Nanjing in China	2013.4.1–2015.12.31	Spearman-Rank analysis; Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method	Dependent variable: $PM_{2.5}$ concentrations Effect factors: temperature, wind speed, relative humidity, precipitation
Socioeconomic factors	Peng et al. (2018)	China	1995–2009	structural decomposition analysis	Dependent variable: SO_X emissions Effect factors: economic expansion, fixed capital formation, household consumption and exports
	Zeng et al. (2019)	31 provinces	2003–2016	Ordinary Least Squares regression (OLS) , spatial autoregressive model (SAR) and spatial error model (SEM)	Dependent variable: PM_{10}, $PM_{2.5}$, and SO_2 emissions Effect factors: emission reduction policies; renewable energy policies
	Zhao et al. (2018)	85 cities (1996) and 200 cities (2000) in China	1996, 2000	Structural equation model	Dependent variable: SO_2 and NO_X concentration Effect factors: household electricity consumption, civilian vehicles, urban built-up area, resident population, secondary industry GDP, tertiary industry GDP, power generation, and urban heated area
	Fu and Li 2020	Global	2000–2014	geographical and temporal weight regression (GTWR)	Dependent variable: $PM_{2.5}$ concentration Effect factors: energy consumption, renewable energy consumption ratio, fossil fuel consumption ratio, forest area, population density, per capita GDP, per capita CO_2 emission, fertilizer consumption, urban population ratio, and high technology

					export
	Zhao et al. (2019)	269 cities in China	2015–2016	Ordinary Least Squares regression (OLS)	Dependent variable: $PM_{2.5}$ concentration Effect factors: urban size (population, size of built-up areas, and gross domestic product (GDP)), population density, share of secondary industry, total amount of private vehicles, per capita disposable income
	Zhao et al. (2018)	5 hot spots (Yangtze River Delta, Bohai Rim, Pan–Pearl River Delta, Central region, and Western region) in China	2004–2012	The coefficient of divergence (COD) and stochastic Impacts by regression on Population, Affluence and Technology (STIRPAT) model	Dependent variable: $PM_{2.5}$ concentration Effect factors: GDP, total population, vehicles, private cars, taxi
	Hao et al. (2016)	73 cities in China	2013	spatial Lag model (SLM), spatial Error Model (SEM)	Dependent variable: Air Quality Index (AQI) and $PM_{2.5}$ concentrations Effect factors: GDP per capita, industrial structure, vehicle population, population density
Natural and socioeconomic factors	Yang et al. (2017)	113 cities in China	2014	OLS, SLM, and SEM, local, geographic weighted regression (GWR) model	Dependent variable: SO_2 concentration Natural factors: precipitation, temperature, wind speed, relative humidity Socioeconomic factors: SO_2 emissions (the sum of industrial and household emissions)
	Liu et al. (2017)	289 cities in China	2014	spatial Durbin model (SDM)	Dependent variable: Air Quality Index (AQI) Natural factors: temperature, precipitation, atmospheric pressure, relative humidity, wind speed, elevation Socioeconomic factors: NDVI, green ratio, total population, urban

				population, GDP, urban land, industry, population density, per capita GDP, car density
Han et al. (2019)	152 cities in China	2016	global and local regression model	Dependent variable: Air Quality Index (AQI) Natural factors: temperature, precipitation, atmospheric pressure, wind speed, elevation, green ratio Socioeconomic factors: secondary industry GDP, industrial structure, population density, per capita GDP, urbanization rate, number of civil vehicles, traffic mileage

To improve air quality, you must first determine the composition of the major pollutants in air pollution. From the antecedent literature, we can see that many studies are about (SO_2 and NO_X) emissions and fine PM concentrations. In this article we define air pollution as SO_2 and NO_X emissions. The specific reasons are as follows. (1) Although PM contains a mixture of different particulate components, most air pollutants come from power generation and industrial processes. SO_2 and NO_X are the sources of air pollution. For the SO_2, although China has achieved significant reduction since 2007 from 36.6 Mt to 8.4 Mt in 2016, it is still the world second largest emitter (Hu et al., 2019). China emits about 25% of the world's NO_X (Cui et al., 2013). The sources of NO_X mainly include waste gases generating in burning of fossil fuels and producing of explosives, dyes, nitric acid, and nitrogenous fertilizer (Lee et al., 1977; Cui et al., 2013). Zhang and Crooks, (2012) found that over 70% of SO_2 emissions is derived from industrial point sources, including thermal power stations. (2) There are few studies on the emissions of various pollutants, which can make up for the research gaps here. In summary, in this study we take SO_2 and NO_X emissions as the dependent variables of air pollution, and explore the natural and socioeconomic factors of air pollution.

Regarding the research model, although some previous studies (Zhao et al., 2019; Zhao et al., 2018) focused on identifying the drivers of air pollution, they tended to address the factors that affect the changes in air pollutant concentration levels over time. Considering that regional air pollutant discharge has the characteristics of spatial spillover and spatial diffusion, we choose two main models in the spatial econometric model: the spatial lag model (SLM) and the spatial error model (SEM). The SLM and SEM in the spatial econometric model capture the spatial by allowing the regression model parameters to vary with spatial and has been applied in many studies (Zhao et al., 2018; Fu and Li 2020; Liu et al., 2017; Zeng et al., 2019).

1.3 Purpose

Through the review of the literature, we can know that most research on air pollution focuses on socioeconomic factors and ignores natural factors. Only a few people have conducted comprehensive factors research using single-year data in a part of cities. Although the city-level research will be more specific, considering the long-term and availability of the data, we selected the provincial-level data for 2011–2017.

The purpose of this study is to explore the following questions. What are the spatial effect and effective factors of air pollution? This paper combines natural factors and socioeconomic factors, through comparative study of the models, spatial econometric model was used to analyze the nine effect factors. The results of effect factors are of great significance for China to improve air pollution from various factors.

The remainder of this paper is organized as follows. In the next section, variable selection of the data and the spatial econometric models are introduced. Then, the result and discussion of spatial autocorrelation, model selection and effecting factors of air pollution are presented respectively. After analyzing the results, the conclusion and policy implications are presented.

2. Data and methodology

In this chapter, we introduce the selection of influencing factors of SO_2 and NO_X emissions, and explain the spatial econometric model and estimation method. The estimation method is the process of selecting the appropriate model by Hausman test and Lagrange Multiplier test.

2.1 Variable selection

By reviewing the literature, we learned that not only natural factors can affect air pollution, but also socioeconomic factors are also influencing factors of air pollution. Considering the data availability, we choose nine factors. Table 2 provides the detailed description of the selected factors.

Table 2 Definition and statistical description of influencing factors of air pollution.

Factors	Variables	Units	Definition	Mean	SD	Min	Max
Air pollution	SO_2	ton	Annual sulfur dioxide emissions in each province	561809	398404.2	3463	1827397
	NO_X	ton	Annual nitrogen oxide emissions in each province	624496	424650.9	30154	1801138
Natural factors	*HDD*	℃	Heating degree days of a major city in each province	2295	1380.9	45.4	5465.6
	CDD	℃	Cooling degree days of a major city in each province	171.2	166.1	0	685.5
	PRE	mm	Annual precipitation of major cities in each province	942.5	565.9	148.8	2939.7
	RHU	%	Annual average relative humidity of major cities in each province	65.3	11.9	33.5	84.58
Socioeconomic factors	*POP*	10^4 people	Permanent population at the end of the year	4399	2769.8	303	11169
	PD	km^2/person	Population density	2797	1164.1	515	5821
	URB	%	Urbanization rate	55.58	13.4	22.73	89.61
	PCGDP	yuan/person	Per capita GDP	50239	23517.4	16413	128994
	SDA	10^9 yuan	Added value from secondary industry	100058.9	8299.7	208.8	39654.9

Many studies have revealed that natural factors affect air pollution. For example, Meng et al., (2008) suggested that strong temperature inversions influence vertical distribution of SO_2 and NO_2 concentrations over urban Beijing. Li et al., (2014) confirmed that temperature is negatively correlated with air pollution index. Therefore, it is important to consider temperature as a factor of air pollution. This study selected the daily temperature. However, the temperature itself is not appropriate to capture hot summer and cold winter with linear models. Therefore, degree days (heating degree day (*HDD*) and cooling degree day (*CDD*)) were calculated using the daily temperature. *HDD* refers to the cumulative daily temperatures below a base temperature in a year, while *CDD* refers to the cumulative daily average temperatures above a base temperature. Degree days are defined as the difference between the daily mean temperature and a given reference temperature. In calculations, the reference temperature is considered to be a comfortable temperature for humans, and it varies across countries and regions. Here, according to the reference, 18°C and 26°C are used as the reference temperatures for *HDD* and *CDD*, respectively. Eq. (1) and (2) are used to calculate the degree days.

$$HDD = \sum_{m=1}^{p} rd(18 - T_m) \tag{1}$$

$$CDD = \sum_{m=1}^{p} rd(T_m - 26) \tag{2}$$

where p is the number days in the year. T_m is the daily mean temperature for day m, and rd is equal to $(18 - T_m)$ or $(T_m - 26)$ if T_m is lower than 18 or higher than 26 and is equal to 0 otherwise.

Bai et al., (2019) and Li et al., (2014) found that precipitation and humidity were negatively correlated with air quality. Furthermore, Yang et al., (2017) revealed that precipitation maintained a negative relationship with SO_2 concentration levels. Precipitation has a dissolving effect on air pollutants, with greater precipitation increasing the wet deposition of pollutants. Liu et al., (2019) showed that precipitation is beneficial to reduce $PM_{2.5}$ concentration, but no significant causality was found with relative humidity. Therefore, we included *RHU* and *PRE* as one of the effect factors in our models. In other studies, topography (Hester and Harrison 2009), precipitation (Li et al., 2014), relative humidity

(Whiteman et al., 2014) , wind speed (Liu et al., 2019), sunshine duration (Statheropoulos et al., 1998), and road trees (Tong et al., 2016) are also found as important natural factors that influence air pollution. Due to this study used provincial-level annual data, we did not use wind speed, sunshine duration, roads trees, topography, and other factors in this study.

As for socioeconomic factors, five variables were selected based on the literature. Human activities have a positive impact on SO_2 emissions, and these effects did vary in areas (Yang et al., 2017). Zhao et al., (2018) showed that the resident population had both direct and indirect effects on urban air quality. Therefore, the *POP* may have a large impact on air pollution.

About population density, its impact on air pollution is somewhat complicated. On the one hand, higher population density would lead to higher degree of urbanization and industrialization, which will positively affect air quality (Hao et al., 2016). On the other hand, high population density makes it possible for the intensive use of energy, which reduces total emissions of pollutants and is therefore beneficial to the environment (Hao et al., 2016). Huang (2018) used the population density as a control variable in the panel spatial Durbin model and pointed out that higher population density can reduce SO_2 emissions. Therefore, we used *PD* as one of the independent variables to analyze the relationship between population density and pollutant emissions.

In addition, Bai et al., (2019) takes the Yangtze River Economic Belt as a study area, and they showed that pollution and urbanization factors were positively correlated with air pollution. Similarly, Han et al., (2014) confirmed that urbanization exerted a significant effect on $PM_{2.5}$ concentrations in urban areas and surrounding regions. Therefore, *URB* is one of the important factors affecting air pollution.

According to the literature, Brajer et al., (2008) provided evidence of an N-shaped relationship between SO_2 and income. And Fu and Li (2020) indicated that per capita GDP was the main factor causing global $PM_{2.5}$ pollution. Therefore, this article used *PCGDP* as one of the independent variables of air pollutions.

In other studies, Jiao et al., (2017) showed that the development of the secondary industry has stimulated the use of coal and greatly increased the concentration level of $PM_{2.5}$. Among the various influencing factors of NO_X pollution, the GDP of the secondary industry has the greatest impact (Hou et al., 2018).

Therefore, this paper introduces the added value of the *SDA* as an explanatory variable.

2.2 Data

This article explores the effects of two air pollutants emissions from natural and socioeconomic factors. In this study, considering the availability and comprehensiveness of the data, we adopted the annual data of the 31 provinces from 2011 to 2017. According to the prior literature (Hu et al., 2019; Cui et al., 2013; Lee et al., 1977; Zhang and Crooks, 2012), we selected SO_2 emissions and NO_X emissions[1] as the dependent variables of air pollutants. The data for degree days is calculated based on the temperature data[2]. The annual average precipitation and relative humidity were taken from the reference. Regarding the natural factors of the independent variables, due to the difficulty in obtaining provincial data, this paper selects a major city[3] as representatives. The data for socioeconomic factors were taken from the provincial data of the reference.

2.3 Methods

Spatial econometrics is a quantitative method to analyze the spatial interaction and spatial structure of economic activities. To better study the spatial dependence over the years, we have selected provincial data for the period 2011-2017. This study used an analytical framework shown in Figure 1 to determine the spatial autocorrelation and influencing factors of the dependent variables (i.e., SO_2 and NO_X emissions). The spatial autocorrelation test mainly tests the spatial aggregation characteristics of air

---

[1] SO_2 and NO_X emissions were extracted from the reference.

[2] Temperature data is from the weather station reference.

[3] The selected major cities are as follows: Hefei (Anhui), Beijing, Chongqing, Fuzhou (Fujian), Yuzhong (Gansu), Guangzhou (Guangdong), Nanning (Guangxi), Guiyang (Guizhou), Haikou (Hainan), Shijiazhuang (Hebei), Harbin (Heilongjiang), Zhengzhou (Henan), Wuhan (Hubei), Changsha (Hunan), Nanjing (Jiangsu), Nanchang (Jiangxi), Changchun (Jilin), Shenyang (Liaoning), Hohhot (Inner Mongolia), Yinchuan (Ningxia), Xining (Qinghai), Jinghe (Shaanxi), Jinan (Shandong), Baoshan (Shanghai), Taiyuan (Shanxi), Wenjiang (Sichuan), Tianjin, Urumqi (Xinjiang), Lhasa (Xizang), Kunming (Yunnan), Hangzhou (Zhejiang).

pollutant emissions through Moran's I and scatter plot. To explore the influencing factors, we need to select the appropriate spatial panel model through the Hausman test and the Lagrange Multiplier (LM) test.

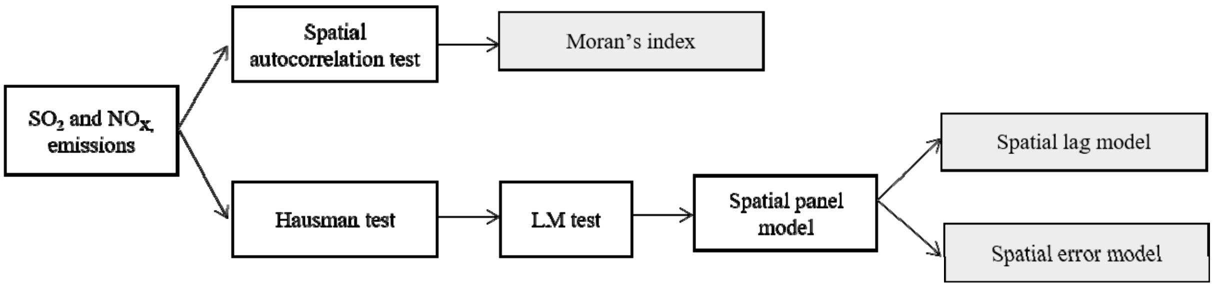


Figure 1. The framework of this study.

2.3.1 Spatial autocorrelation analysis

Spatial dependence is widespread for human geographical phenomena. Considering spatial dependence, regional air pollutant discharge has the characteristics of spatial spillover and spatial diffusion, and it have a great impact on air pollution of neighboring areas. Here we select the most commonly used Moran's I to measure if the data have spatial correlation. The Moran's I is defined as Eq. (3):

$$I = \frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n} W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S^2\left(\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\right)} = \frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n} W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} W_{ij}\right)\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (3)$$

where n is the number of spatial units indexed by *i* and *j*, W_{ij} is a matrix of spatial weight with zeroes on the diagonal (refer to Eq. (7) for details), x_i and x_j refers to the observations of spatial location *i* and *j*, and $\bar{x}$ refers to the mean of *x*. The value of Moran's I ranges from -1 to 1. When the value is close to -1, the spatial distribution will show a discrete trend. In contrast, when the value is close to 1, clustering trends appear in the spatial distribution. If the value is close to 0, it means no correlation. The results of Moran's I are not only displayed numerically, but also through Moran scatter plots. As shown in Figure 2, the X axis of the Moran scatter plot represents the standardized (mean: 0 and standard deviation: 1) observation, and the Y axis represents the spatial lag variables of the standardized observation. The scatter plot is decomposed into four quadrants, and all quadrants exhibit spatial autocorrelation. The first quadrant indicates not only the province but also the surrounding provinces also have high pollutant emissions. The second quadrant indicates although the province's pollutant emissions are low, the surrounding provinces are high. Third quadrant indicates not only the province but also the surrounding provinces have low pollutant emissions. Fourth quadrant indicates that although the province's pollutant emissions are high, the surrounding provinces have low pollutant emissions. Therefore, if most provinces show positive spatial autocorrelation in the first and third quadrants, the Moran's value is close to 1. On the contrary, if most provinces in the second and fourth quadrants indicate spatial negative spatial autocorrelation, and Moran's value approaches -1. In addition, the slope of the linear smooth line in the figure is consistent with the Moran's I value.

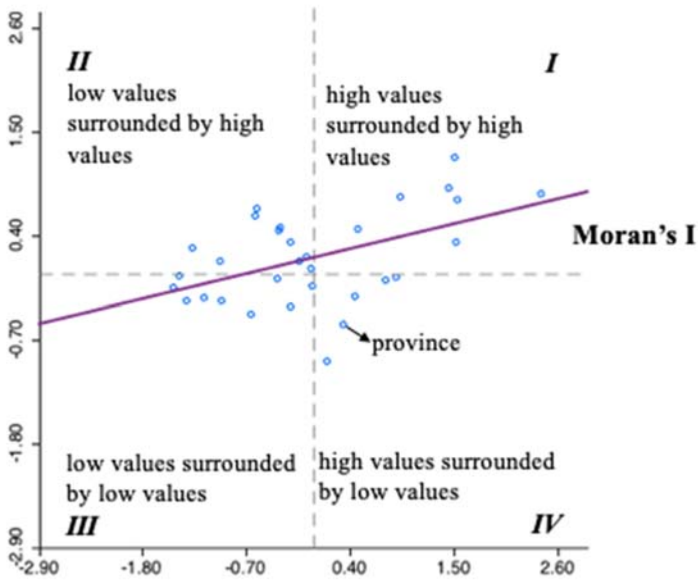


Figure 2. An example of Moran scatter plot.

2.3.2 Spatial panel models

The spatial econometric model mainly analyzes the interaction and interdependence of spatial regions.

First, we consider a simple pooled linear regression model with space specific effects but without spatial interaction effects (Elhorst, 2010). The simple pooled linear regression model can be written as Eq. (4):

$$y_{kt} = x_{kt}\beta + u_k + \varepsilon_{kt} \tag{4}$$

where k is an index for the cross sectional dimension (spatial units), with k =1,…,N, and t is an index for the time dimension (time periods), with t=1,…,T. y_{kt} is the dependent variable at k and t, x_{kt}represents the independent variables at k and t, and β represents the regression coefficient. ε_{kt} is an independently and identically distributed error term for k and t with zero mean and variance σ^2, while u_k denotes a spatial specific effect. The standard reasoning behind spatial specific effects is that they control for all space specific time invariant variables whose omission could bias the estimates in a typical cross-sectional study.

Fixed effect models and random effect models are often used for panel data analysis. In the random effect models, the individual specific effect is a random variable that is uncorrelated with the explanatory variables. In the fixed effects models, the individual specific effect is a random variable that is allowed to be correlated with the explanatory variables.

The spatial econometric model can effectively solve the spatial dependence problem. In order to examine and measure the possible effects, two main types of spatial econometric models are used in this study: the SLM and SEM**.** SLM can be interpreted that the dependent variable depends on the dependent variable observed in adjacent units and a set of observed local features. The SLM can be expressed as Eq. (5):

$$y_{kt} = \delta \sum_{j=1}^{N} w_{ij}{}' y_{jt} + x_{kt}\beta + \mu_k + \varepsilon_{kt} \tag{5}$$

where δ is the spatial autoregressive coefficient and $w_{ij}{}'$ is the spatial weight matrix after row-standardized of w_{ij} (refer to Eq. (9) for the details).

The SEM considers that the dependent variable depends on a set of observed local characteristics and that the error terms are correlated across space. The SEM can be written as Eq. (6):

$$\begin{aligned} y_{kt} &= x_{kt}\beta + \mu_i + \phi_{kt} \\ \phi_{kt} &= \lambda \sum_{j=1}^{N} w_{ij}{}'\phi_{kt} + \varepsilon_{kt} \end{aligned} \tag{6}$$

where $\sum_{j=1}^{N} w_{kj}\,\phi_{kt}$ denotes the interaction effects among the disturbance terms of the different units and λ refers to the spatial autocorrelation coefficient. ϕ_{kt} reflects the spatially autocorrelated error term.

2.3.3 Spatial weight matrix

A spatial weight matrix is necessary when implementing spatial panel analysis. It provides spatial structure information between adjacent areas and how they interact with each other. How to choose the spatial weight matrix is the premise for data analysis. There are two types of spatial weight matrix, which are based on contiguity and on distance. Here we use a contiguity spatial weight matrix. The spatial weight matrix is defined as **W** with elements W_{ij} indicating whether observations i and j are spatially close. If units i and j (not equal i) are neighbors, the spatial weight is 1, otherwise 0. W_{ij} can be written as Eq. (7)

$$W_{ij} = \begin{cases} 1 & if\ \ i\ \ is\ \ contiguous\ \ to\ \ j \\ 0 & otherwise \end{cases} \tag{7}$$

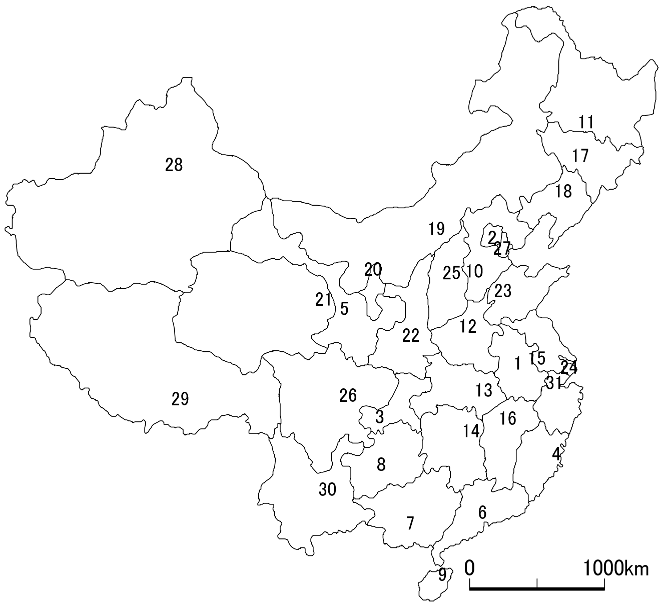
For example, spatial weight matrix based on contiguity where units 2 and 3 and units 3 and 4 are neighbors, but when units 2 and 4 are not neighbors, W_{ij} can be written as Eq. (8).

$$W_{ij} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{8}$$

The spatial weight matrix needs to be "row-standardized," which means that the weights need to sum up to one on each row. The row standardization is needed because in a weighted average formula, the total weights must be 1. The spatial weight matrix based on Eq. (6) can be written as Eq. (9)

$$W_{ij}' = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{9}$$

In order to study irregularly shaped area data, we need to create a spatial weight matrix. This study used neighboring information from 31 provinces in China (Figure 3 and Table 3).



No.	Province	No.	Province
1.	Anhui	17	Jilin
2.	Beijing	18	Liaoning
3.	Chongqing	19	Inner Mongolia
4.	Fujian	20	Ningxia
5.	Gansu	21	Qinghai
6.	Guangdong	22	Shaanxi
7.	Guangxi	23	Shandong
8.	Guizhou	24	Shanghai
9.	Hainan	25	Shanxi
10.	Hebei	26	Sichuan
11.	Heilongjiang	27	Tianjin
12.	Henan	28	Xinjiang
13.	Hubei	29	Xizang
14.	Hunan	30	Yunnan
15.	Jiangsu	31	Zhejiang
16.	Jiangxi		

Figure 3. The geographical locations of Chinese provinces.

Table 3 Geographical proximity information of 31 provinces in China.

Number	province	adjacent information	number	province	Adjacent information
1	Anhui	12,13,15,16,23,31	17	Jilin	11,18,19
2	Beijing	10,27	18	Liaoning	10,17,19
3	Chongqing	8,13,14,22,26	19	Inner Mongolia	5,10,11,17,18,20,22,25
4	Fujian	6,16,31	20	Ningxia	5,19,22
5	Gansu	19,20,21,22,26,28	21	Qinghai	5,26,28,29
6	Guangdong	4,7,9,14,16	22	Shaanxi	3,5,12,13,19,20,25,26
7	Guangxi	6,8,9,14,30	23	Shandong	1,10,12,15
8	Guizhou	3,7,14,26,30	24	Shanghai	15,31
9	Hainan	6,7	25	Shanxi	10,12,19,22
10	Hebei	1,12,18,19,23,25,27	26	Sichuan	3,5,8,21,22,29,30
11	Heilongjiang	17,19	27	Tianjin	2,10
12	Henan	1,10,13,22,23,25	28	Xinjiang	5,21,29
13	Hubei	1,3,12,14,16,22	29	Xizang	21,26,28,30
14	Hunan	3,6,7,8,13,16	30	Yunnan	7,8,26,29
15	Jiangsu	1,23,24,31	31	Zhejiang	14,15,16,24
16	Jiangxi	1,4,6,13,14,31			

2.3.4 Hausman test and Lagrange Multiplier test

It is important to choose the suitable model after establishing the panel regression model. Here we perform a series of Hausman test to identify the presence of endogeneity in the explanatory variables, so as to effectively estimate random effects and fixed effects (Huang, 2018).

After deciding to choose a random effects model or a fixed effects model, SLM or SEM need to be further selected. Based on the literature (Zeng et al., 2019; Anselin, 2005), here we choose the most popular LM test. This test is a general principle of testing the parameter assumptions in the likelihood framework. The hypothesis under test is expressed as one or more constraints on the values of parameters. To perform the LM test, only estimation of the parameters subject to the restrictions is required.

The spatial regression model selection process is shown in Figure 4. It contains four LM tests as follows: LM-lag, LM-error, robust LM-lag and robust LM-error. In LM test, when the standard versions (LM-Lag or LM-error) are both significant, the robust versions of the statistical tests are conducted. When they are not, the properties of the robust versions may no longer hold. In the process of spatial regression decision, we need to consider the LM-error and LM-Lag tests. When both LM tests are not significant, stop the test. If only one of the LM-error and LM-lag tests is significant, the corresponding spatial regression model is chosen. When both the LM-error and LM-lag tests are significant, we need to conduct further robust tests. From the robust LM-error test and the robust LM-lag test, we need to choose the one which is most significant. Usually either robust LM-error or robust LM-lag make significance, or one of them is more significance than the other. In the latter case, we will choose the most significance spatial regression model. In the case both are highly significant, further checks are required for specifications errors.

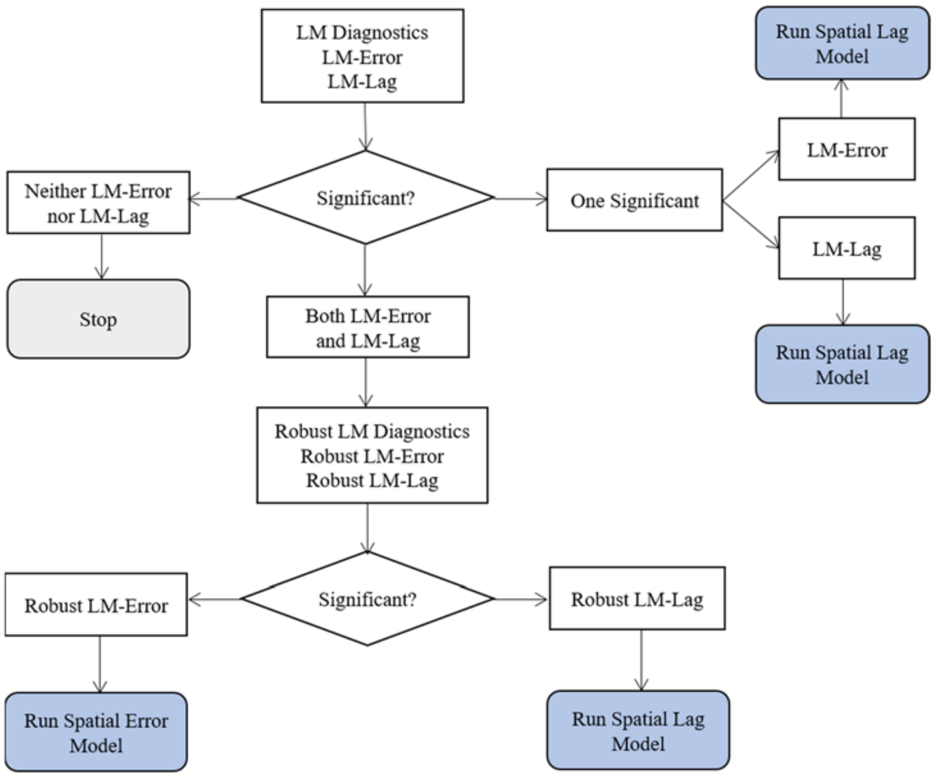


Figure 4. Spatial regression decision process.

2.3.5 Software

In this study, we mainly used two kinds of software. One is GeoDa for spatial regression analysis. In the study, we used GeoDa software to make a spatial description map of China and calculated Moran's I value.

Because GeoDa cannot analyze spatial panel models, we also used R with the splm package, for the model analysis. The software is used for the Hausman test, the LM test, and analysis of the influencing factors of SO_2 and NO_X.

3. Results and discussion

In this chapter, we have discussed mainly in two things. One is to describe the spatial effect of air pollution from the temporal and spatial distribution of air pollution and the results of spatial autocorrelation. The other is to use the appropriate model from the model test results to analyze the influencing factors of air pollution.

3.1 Temporal and spatial distribution of air pollution

According to reference, the total national emissions of SO_2 decreased from 22.18 Mt in 2011 to 8.75 Mt in 2017. And the total emissions of NO_x emissions decreased from 24.04 Mt in 2011 to 12.59 Mt in 2017. We can know that national pollutant emissions have been gradually reduced from 2011 to 2017. In order to observe the air pollutant emissions in recent years from a spatial perspective, this study used GeoDa to map the changes in air pollutant emissions (Figure 5). Observing by region, we can know that the emissions of air pollutant in the northern and central areas, such as Inner Mongolia, Hebei, Henan, Shanxi, Shandong, and Liaoning, were higher than southern in 2011 and 2014. However, by 2017, the regions with high pollutant emissions changed from central and northern to northern. Han et al., (2019) pointed out that northern and central areas are densely populated, with a high car density and persistent traffic congestion. Cheng et al., (2017) also suggested that coal was mainly used for central heating in winter in the northern region, and the utilization rate of filter coal is low. At the same time, environmental dust removal equipment used by utility companies is inefficient. These caused SO_2 generated by burning coal to be directly discharged into the air, thereby exacerbating local urban pollution. Regarding the Beijing-Tianjin-Hebei (BTH) region, severe air pollution has aroused great concerns of the Chinese government. The study (Zhu et al., 2017) have pointed out that increasing the burden on neighboring provinces also threatens the local environment. Pollutant emissions in the south and southwest were lower than other areas. This is because the areas with high pollutant emissions were gradually shifted from the southwest and south (Sichuan, Guizhou, Hunan, and Guangdong) to the north of China (Shanxi, Liaoning, Hebei, and Shandong).

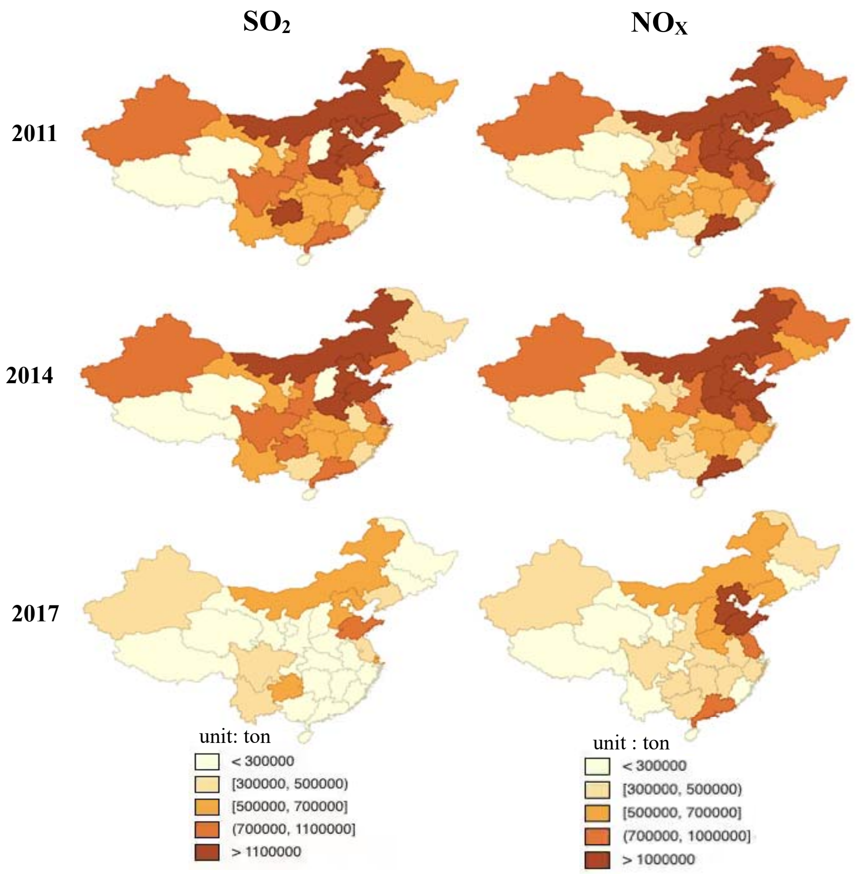


Figure 5. Spatial distribution of two air pollutants with annual emissions.

3.2 Spatial autocorrelation

To test the spatial correlation of air pollutants in China, we use the Moran's I statistics. Table 4 lists the Moran's I statistics for SO_2 emissions and NO_X emissions in china from 2011 to 2017. As shown in the table, the Moran's I values were positive and statistically significant at the 10% level for almost every year. These results indicate that there is a spatial positive autocorrelation in SO_2 and NO_X. The positive values indicate that areas with high air pollutants emissions tend to cluster together, and regions with low emissions tend to cluster together.

The spatial clustering and heterogeneity can be further seen from the Moran scatter plots (Figure 6). The most provinces were located in the first and third quadrants, meaning that the emissions of SO_2 and NO_X exhibit positive spatial autocorrelation. Therefore, to improve air pollution, the neighboring provinces should strengthen cooperation and formulate local control measures. Specifically, from the scatter plot of the two air pollutant emissions, we can find the following points. First, there was a positive spatial autocorrelation in the first and third quadrants for SO_2 emissions in 2011, including Liaoning, Inner Mongolia, Hebei, Shanxi, Shaanxi, Henan, Shandong, Shanghai, Zhejiang, Jiangxi, Fujian, Hainan, Guangxi, Yunnan, Qinghai and Tibet. In 2017, expect for Shanghai, Fujian, and Guangxi, there was no change in the total number of provinces with positive correlations. Second, Heilongjiang, Liaoning, Inner Mongolia, Hebei, Shanxi, Shaanxi, Henan, Shandong, Jiangsu, Anhui, Zhejiang, Gansu, Xinjiang, Qinghai, Tibet, Sichuan, Yunnan, Guizhou, Chongqing, Hunan, and Guangxi were distributed in the first and third quadrants for NO_X emissions in 2011. By 2017, except for Sichuan, Chongqing, Hunan and Guangxi, the total number of provinces with positive correlations has not changed. From these results, we can find that areas with high pollutant emissions (Hebei, Henan, Tianjin and Shandong Province) and areas with low pollutant emissions (Jiangsu, Zhejiang, Sichuan, Yunnan, Hainan, Guangdong and Guangxi) were located in the first and third quadrants.

Table 4 Moran's I statistics for China's provincial SO_2 and NO_X emissions from 2011 to 2017.

	SO_2			NO_X		
Year	**Moran's I**	**z-value**	**p-value**	**Moran's I**	**z-value**	**p-value**
2011	0.242	2.333	0.018	0.280	2.650	0.011
2012	0.193	1.916	0.042	0.267	2.544	0.013
2013	0.187	1.859	0.046	0.249	2.402	0.019
2014	0.176	1.763	0.050	0.248	2.403	0.018
2015	0.193	1.919	0.042	0.260	2.502	0.014
2016	0.117	1.325	0.097	0.155	1.606	0.055
2017	0.096	0.112	0.136	0.185	1.878	0.042

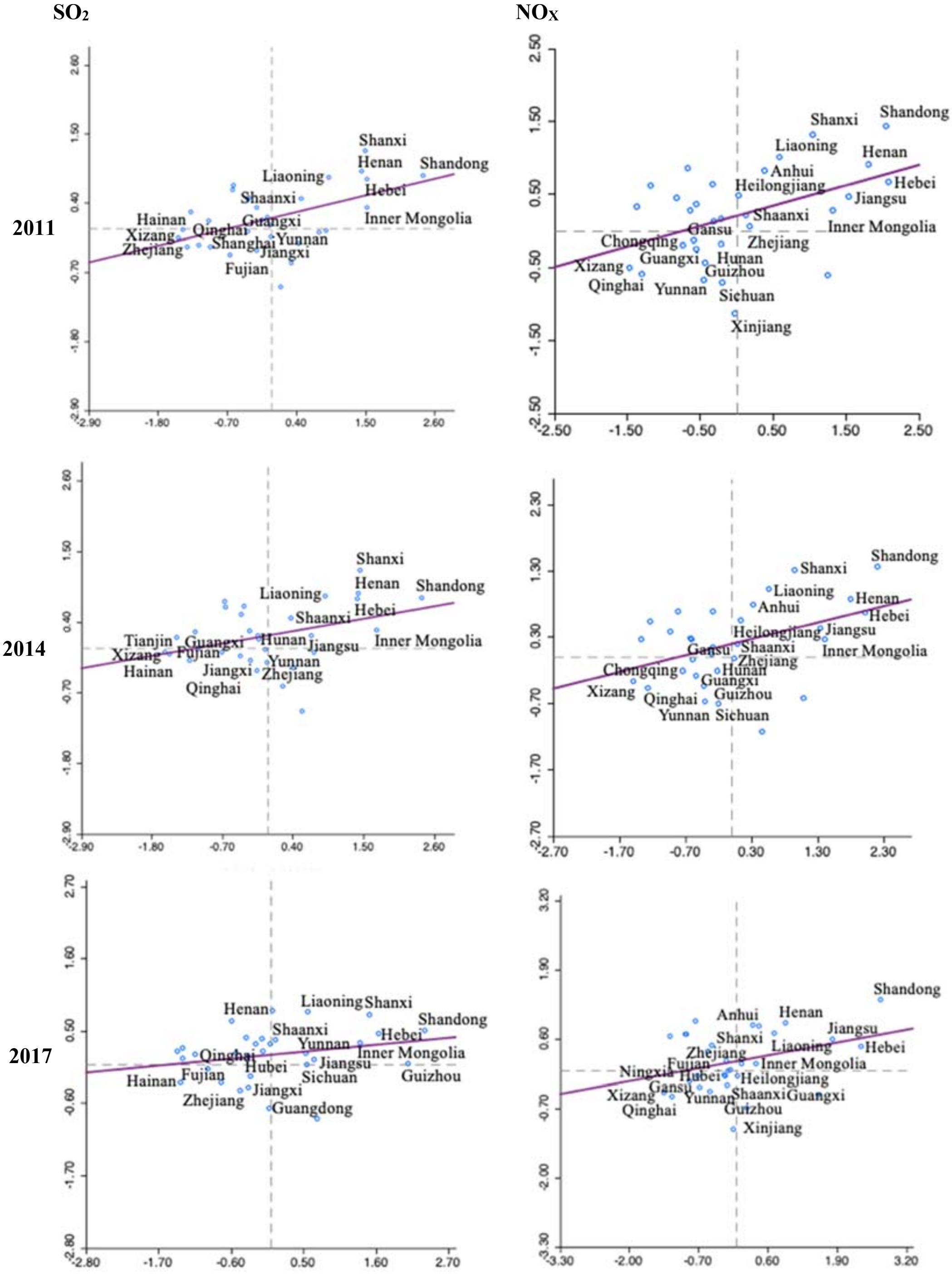


Figure 6. Moran scatter plots of provinces with two air pollutant emissions.

3.3 Hausman test and LM test for model selection

To select the model (fixed-effect or random-effect models), we conducted the Hausman test. The test showed that the P values for SO_2 and NO_X were 0.0000542 and 0.005068, respectively. The models for both pollutants passed the significance level of 5%. Therefore, the fixed-effect model was selected.

We then conducted the LM test. Table 5 shows that the LM-lag and LM-error tests for SO_2 and NOx were close to 0, meaning that all models passed the significance level of 1%. Therefore, to further determine the models, we conducted the robust LM tests. The P value for the robust LM-error test passed the significance level of 1%, while the robust LM-error test did not. Therefore, the sample data are more suitable with the SLM models.

Table 5. The result of LM test.

	SO_2	**NO_X**
SLM test	2.20×10^{-16}	2.20×10^{-16}
SEM test	2.20×10^{-16}	2.20×10^{-16}
robust SLM test	4.93×10^{-8}	2.23×10^{-8}
robust SEM test	0.006358	1.019×10^{-7}

3.4 The effective factors of air pollution change

From the Hausman tests and LM tests, the fixed-effect SLM was selected for all specifications. Moreover, in the fixed-effect models, we considered two model specifications: individual fixed effect and two-way fixed-effect (both individual and time period). Table 6 presents the results for both the individual fixed-effect and two-way fixed-effect estimations.

For natural factors, we found that the coefficients for *HDD* is positive and significant. The *CDD* of SO_2 is negative and significant at the 5% level. And *RHU* were negative and statistically significant on the air pollutants emissions, which are consistent with the findings of previous studies (Bai et al., 2019; Li et al., 2014; Yang et al., 2017). In contrast, *PRE* has no significant effect on air pollutant emissions. Therefore, natural factors play an important role in the progress of air pollution and have also been confirmed from Zhang et al., (2015). They also point out that air pollution varies widely in different regions. Therefore, we also need to take effective measures to reduce air pollution in accordance with local actual conditions. In addition, we need to take into account our findings when determining predictions and pollution control measures related to air pollution, so as to improve the accuracy of prediction of air pollution under different natural factors and provide effective measures to reduce pollution.

Table 6 Results of SLM models of two air pollutants.

Pollutant	**SO_2**		**NO_X**	
Variable	Individual FE	Two-way FE	Individual FE	Two-way FE
HDD	33.240	54.815*	63.852***	67.214**
	(1.229)	(1.790)	(2.750)	(2.564)
CDD	-502.817***	-400.565**	84.075	67.473
	(-2.687)	(-1.978)	(0.525)	(0.389)
PRE	-97.381	-64.266	-85.684	-70.039
	(-1.467)	(-0.933)	(-1.511)	(-1.188)
RHU	-3965.412	-5121.993*	-7899.231***	-8377.583***
	(-1.510)	(-1.867)	(-3.519)	(-3.574)
POP	64.885***	63.118***	65.910***	50.837***
	(3.150)	(2.792)	(3.716)	(2.627)
PCGDP	-8.073***	-6.514***	-8.993***	-8.702***
	(-3.408)	(-2.633)	(-4.443)	(-4.111)
SDA	13.210*	12.144	22.960***	26.233***
	(1.836)	(1.585)	(3.732)	(4.002)
URB	7216.526**	3749.085	9699.814***	8768.969***
	(2.011)	(0.960)	(3.157)	(2.624)
PD	-80.444***	-89.757***	-48.232***	-57.665***
	(-3.942)	(-4.708)	(-3.055)	(-3.534)
lambda	1.210	0.872	1.118	0.806

Notes: Robust standard errors in parentheses; *** $p<0.01$. ** $p<0.05$. * $p<0.1$.

For socioeconomic factors, the results show that the *POP* is positive and statistically significant at the 1% level for SO_2 and NO_X. This suggests that SO_2 and NO_X emissions increase as population increases. It is consistent with the findings of Liu et al., (2017). Furthermore, the *URB* is also positive and statistically significant with NO_X, which means that as the *URB* increases, NO_X emissions also increase. At present, China is still in the stage of rapid urbanization and the urban population is growing rapidly. Rapid urbanization is associated with more traffic, greater population size, more industry, and more energy consumption, all of which lead to higher levels of air pollution (Zhao et al., 2018). However, the effect of *PD* for SO_2 and NO_X emissions were negative and statistically significant, which is consistent with the previous finding (Hao et al., 2016). We can find the answer from the literature (Cheng et al., 2017). Because the increase in *PD* can also produce the agglomeration effect by not only improving the utilization efficiency of public transportation and resources, but also by pollution agglomeration facilities, which then alleviates air pollution to some extent. Contrariwise, *PD* has a significant positive impact on air pollution both through the scale effect and the agglomeration effect (Cheng et al., 2017). From the viewpoint of scale effect, the higher the urban *PD*, the higher the housing, electricity, and transportation demand; all three are direct causes of air pollution (Cheng et al., 2017). In addition, high *PD* is not conducive to the diffusion of pollutants, and this indirectly aggravates the air pollution. From the estimates in this study, the externalities of the agglomeration effect have been fully exploited. These require that in the process of urban agglomeration construction, China should allow those positive externalities of population agglomeration that improve the efficiency of resources utilization and the environment to fully play themselves out, thereby effectively buffering the scale effect of population agglomeration on pollutant emissions (Cheng et al., 2017).

In addition, *SDA* also plays a significant positive role in SO_2 and NO_X emissions. The higher the added value of the secondary industry, the more SO_2 and NO_X will be emitted. China is now at a stage of accelerated industrialization and urbanization, and energy consumption in secondary industry is much higher than that in other sectors. What is more serious is that most of the SO_2 and NO_X emissions come from the combustion of fossil fuels and industrial processes, and the large amount of pollutants produced by the consumption of fossil fuels are directly discharged into the air. These are also important reasons for aggravating air pollution, and it is suggested that China not only accelerate the adjustment and upgrade of industrial structure, but also promote the development of green industries.

Finally, the estimated effect of *PCGDP* on SO_2 and NO_X were negative and statistically significant at the 1% level. Wang et al., (2016) pointed out that this has a lot to do with China's increased investment in reducing emissions of SO_2 in flue gas and investment in methods to replace coal or high sulfur coal.

And he also found evidence in support of an inverted U-shaped curve relationship between economic growth and SO_2 emissions. Therefore, in order to balance the relationship between environmental protection and economic development, effective economic leverage measures need to be implemented nationwide.

4. Conclusion and policy implication

Along with China's rapid economic growth and high levels of energy consumption, the problem of air pollution has become one of the most important environmental problems to be solved urgently. In this study, we have innovated in using panel data to combine natural factors and socioeconomic factors. We aim to analyze the spatial effect and effective factors of air pollutant emissions to make better policy decisions from varies factors. We considered two types of air pollutant: SO_2 and NO_X. Using panel data of 31 provinces during the period from 2011 to 2017 to construct a spatial panel model to analyze the effect of natural (*HDD, CDD*, *PRE*, *RHU*) and socioeconomic factors (*POP*, *PD*, *URB*, *PCGDP*, *SDA*) on air pollution in China. We find that regional cooperation and optimizing industries have a great role in solving environmental problems. The primary conclusions are as follows. The emissions of two pollutants were on a downward trend. However, highly polluted areas were transferred from the south and southwest areas to north area. Nevertheless, air pollution has a significant and strong spatial autocorrelation. Furthermore, from the fixed-effect SLM, we found that *HDD, POP*, *SDA* and *URB* had positive and statistically significant impact on the air pollution emissions, while *CDD, PCGDP*, *PD* and *RHU* had significant negative effects. *PRE* had no significant effect.

Having identified the key influencing factors and spatial effects of air pollution, we need to adopt more efficient air pollution prevention and control strategies to promote sustainable development. Based on the findings of this study, we propose the following two policy measures to reduce air pollution.

1. *Strengthening regional cooperation*: The spatial autocorrelation of the two air pollutant emissions shows that each province and its neighbors could influence each other. Due to the geographical location and economic development of each province, the policy goals and capabilities of each province are also different. Therefore, we cannot only rely on one province for decrease air pollution. When considering the issue of mitigation of air pollution, the provinces should strengthen regional cooperation and joint governance. For example, Song et al., (2020) indicate that the expansion of the core area of air pollution joint prevention and control in Beijing-Tianjin-Hebei (BTH) region and surrounding areas is conducive to further improving regional air quality.

Specifically, while strengthening joint prevention and control between regions, we also need to establish an innovative pollution control and responsibility model, focusing on environmental planning and environmental legislation in urban agglomeration planning.

2. *Optimizing the industrial*: As the secondary industry has a positive impact on pollutant emissions, optimizing the industrial structure will be particularly important. The production and consumption of energy, particularly fossil fuels are the predominant source of air pollution, the national energy control policy could affect the amount of air pollutant emissions. On the one hand, China should raise the environmental standards, environmental regulations, and pollutant discharge requirements of manufacturing companies, limit the rapid growth of energy-intensive and pollutant-intensive manufacturing, and gradually phase out backward production methods. At the same time, China should vigorously develop a circular economy and encourage the development of low carbon environmental protection industries. China should also establish a scientific, efficient, clean, and sustainable coal supply system. Through independent research, technology development and introduction, it should continuously improve the level of equipment technology and reduce the proportion of coal as the main energy source. The government should also vigorously develop new and renewable energy sources, strengthen the development and utilization of natural gas, and increase the proportion of the secondary industry.

As a developing country, China has been plagued by air pollution for a long time. The results of this study can not only provide important insights to Chinese domestic policymakers, but also can provide valuable reference for other developing countries. While developing countries take measures to reduce pollutant emissions in socioeconomic terms, they also need to consider the effects of natural factors. At the same time, developing countries is now in the stage of accelerating industrialization and urbanization, and the energy consumption of the secondary industry is an important reason for exacerbating pollutant emissions. It is suggested that developing countries such as China should not only accelerate the adjustment and upgrade of industrial structure, but also promote the development of green industries.

There are still some deficiencies in this study. Although we used SLM in spatial econometric analyses, we have not analyzed spatial spillover effects between provinces. Other spatial econometric models may provide new insights for studying air pollution. In future research, we can continue to collect greater range of data to improve the study and more driving forces should be included in the model, such as topography and climate characteristics.

Acknowledgements

I would first like to thank my supervisor, Prof. Ken'ichi Matsumoto. Whenever I have questions about the content and article writing of my research, he always answers patiently and gives academic advice. He consistently allowed this paper to be my own work, and took me to the right the direction with inspirational guidance. Without his guidance, I would not have been able to complete my dissertation. Second, Secondly, I would also like to thank my two associate supervisors, Prof. Kensuke Katayama and Prof. Yuki Yamamoto, who also provided a lot of support for my thesis creation process. Finally, I want to thank the students in the laboratory for providing me with a good learning atmosphere.

Appendix

The Hausman test, LM test and spatial econometric models estimation in R are as follows:

```
library(plm)

library(splm)

library(readxl)

#DATA: 2011-2017

data1 <- read_excel("Desktop/data1.xlsx")

View(data1)

pdata<-pdata.frame(data1,index=c("NUM","YEAR"))

#model specification

fm1<-SO2~POP+PCGDP+SDA+URB+PD+HDD+CDD+PRE+RHU

###Spatial panel

#Spatial weights

#SWM

W<-matrix(0,31,31)

W[1,1:31]<-

c(0,0,0,0,0,0,0,0,0,0,0,0.166666667,0.166666667,0,0.166666667,0.166666667,0,0,0,0,0,0,0.16666666
7,0,0,0,0,0,0,0,0.166666667)

W[2,1:31]<-c(0,0,0,0,0,0,0,0,0,0.5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.5,0,0,0,0)

W[3,1:31]<-c(0,0,0,0,0,0,0,0.2,0,0,0,0,0.2,0.2,0,0,0,0,0,0,0,0.2,0,0,0,0.028571429,0,0,0,0,0)

W[4,1:31]<-

c(0,0,0,0,0,0.333333333,0,0,0,0,0,0,0,0,0.333333333,0,0,0,0,0,0,0,0,0,0,0,0,0.333333333)

W[5,1:31]<-

c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.166666667,0.166666667,0.166666667,0.166666667,0,0,0,0.166
666667,0,0.166666667,0,0,0)

W[6,1:31]<-c(0,0,0,0.2,0,0,0.2,0,0.2,0,0,0,0,0.2,0,0.2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
```

```
W[7,1:31]<-c(0,0,0,0,0,0.2,0,0.2,0.2,0,0,0,0,0.2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.2,0)
W[8,1:31]<-c(0,0,0.2,0,0,0,0.2,0,0,0,0,0,0,0.2,0,0,0,0,0,0,0,0,0,0,0.2,0,0,0,0,0.2,0)
W[9,1:31]<-c(0,0,0,0,0,0.5,0.5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
W[10,1:31]<-
c(0,0.142857143,0,0,0,0,0,0,0,0,0,0.142857143,0,0,0,0,0,0.142857143,0.142857143,0,0,0,0.14285714
3,0,0.142857143,0,0.142857143,0,0,0,0)
W[11,1:31]<-c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.5,0,0.5,0,0,0,0,0,0,0,0,0,0,0)
W[12,1:31]<-
c(0.166666667,0,0,0,0,0,0,0,0,0.166666667,0,0,0.166666667,0,0,0,0,0,0,0,0,0.166666667,0.16666666
7,0,0.166666667,0,0,0,0,0,0)
W[13,1:31]<-
c(0.166666667,0,0.166666667,0,0,0,0,0,0,0,0,0.166666667,0,0.166666667,0,0.166666667,0,0,0,0,0,0.
166666667,0,0,0,0,0,0,0,0,0)
  W[14,1:31]<-
c(0,0,0.166666667,0,0,0.166666667,0.166666667,0.166666667,0,0,0,0,0.166666667,0,0,0.166666667
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
W[15,1:31]<-c(0.25,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.25,0.25,0,0,0,0,0,0,0.25)
W[16,1:31]<-
c(0.166666667,0,0,0.166666667,0,0.166666667,0,0,0,0,0,0,0.166666667,0.166666667,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0.166666667)
W[17,1:31]<-
c(0,0,0,0,0,0,0,0,0,0,0.333333333,0,0,0,0,0,0,0.333333333,0.333333333,0,0,0,0,0,0,0,0,0,0,0,0)
W[18,1:31]<-
c(0,0,0,0,0,0,0,0,0,0.333333333,0,0,0,0,0,0,0.333333333,0,0.333333333,0,0,0,0,0,0,0,0,0,0,0,0)
W[19,1:31]<-
c(0,0,0,0,0.125,0,0,0,0,0.125,0.125,0,0,0,0,0,0.125,0.125,0,0.125,0,0.125,0,0,0.125,0,0,0,0,0,0)
```

```
W[20,1:31]<-
c(0,0,0,0,0.333333333,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.333333333,0,0,0.333333333,0,0,0,0,0,0,0,0)
W[21,1:31]<-c(0,0,0,0,0.25,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.25,0,0.25,0.25,0,0)
W[22,1:31]<-
c(0,0,0.125,0,0.125,0,0,0,0.125,0,0,0,0.125,0,0,0,0,0,0.125,0.125,0,0,0,0,0.125,0.125,0,0,0,0,0)
W[23,1:31]<-c(0.25,0,0,0,0,0,0,0,0,0.25,0,0.25,0,0,0.25,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
W[24,1:31]<-c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.5)
W[25,1:31]<-c(0,0,0,0,0,0,0,0,0,0.25,0,0.25,0,0,0,0,0,0,0.25,0,0,0.25,0,0,0,0,0,0,0,0,0)
W[26,1:31]<-
c(0,0,0.142857143,0,0.142857143,0,0,0.142857143,0,0,0,0,0,0,0,0,0,0,0,0,0.142857143,0.142857143,
0,0,0,0,0,0,0.142857143,0.142857143,0)
W[27,1:31]<-c(0,0.5,0,0,0,0,0,0,0,0.5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
W[28,1:31]<-
c(0,0,0,0,0.333333333,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.333333333,0,0,0,0,0,0,0.333333333,0,0,0)
W[29,1:31]<-c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.25,0,0,0,0,0.25,0,0.25,0,0.25,0)
W[30,1:31]<-c(0,0,0,0,0,0,0.25,0.25,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.25,0,0,0.25,0,0)
W[31,1:31]<-c(0.2,0,0,0.2,0,0,0,0,0,0,0,0,0,0,0.2,0.2,0,0,0,0,0,0,0,0,0.2,0,0,0,0,0,0,0)
W
SWM= mat2listw(W)
#Test
#Hausman: Choose Fixed effects model or Random effect model
test3<-sphtest(fm1, data = pdata, index=NULL,listw = SWM, spatial.model = "error", method = "GM")
test3
#LMtest
LMLag<-slmtest(fm1,data=pdata,listw=mat2listw(W), test="lml")
LMLag
```

```
LMRLag<-slmtest(fm1,data=pdata,listw=mat2listw(W), test="rlml")
LMRLag
LMerror<-slmtest(fm1,data=pdata,listw=mat2listw(W), test="lme")
LMerror
LMRerror<-slmtest(fm1,data=pdata,listw=mat2listw(W), test="rlme")
LMRerror
#Fixed effects model SLM
#Individual
model75<-spml(fm1,data=pdata,listw=SWM,model="within",effect="individual",
                      spatial.error="none",lag=TRUE)
summary(model75)
#Two-way
model76<-
spml(fm1,data=pdata,listw=SWM,model="within",effect="twoways",spatial.error="none",lag=TRUE)
summary(model76)
#NOX：fm2 instead of fm1 loop
fm2<-NOX~POP+PCGDP+SDA+URB+PD+HDD+CDD+PRE+RH
```

References

Anselin, L., & Anselin, L. (2005). Exploring Spatial Data with GeoDa. Geography. http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf

Bai, L., Jiang, L., Yang, D. yang, & Liu, Y. bin. (2019). Quantifying the spatial heterogeneity influences of natural and socioeconomic factors and their interactions on air pollution using the geographical detector method: A case study of the Yangtze River Economic Belt, China. Journal of Cleaner Production, 232, 692–704. https://doi.org/10.1016/j.jclepro.2019.05.342

Chinese Ministry of Ecology and Environmental Protection. (2017). Bulletin of the State of China's Eco-Environment. http://www.gov.cn/guoqing/2019-04/09/content_5380689.htm

Chang, M., Zheng, J., Inoue, Y., Tian, X., Chen, Q., & Gan, T. (2018). Comparative analysis on the socioeconomic drivers of industrial air-pollutant emissions between Japan and China: Insights for the further-abatement period based on the LMDI method. Journal of Cleaner Production, 189, 240–250. https://doi.org/10.1016/j.jclepro.2018.02.111

Chen, T., He, J., Lu, X., She, J., & Guan, Z. (2016). Spatial and temporal variations of PM 2.5 and its relation to meteorological factors in the urban area of Nanjing, China. International Journal of Environmental Research and Public Health, 13(9). https://doi.org/10.3390/ijerph13090921

Cheng, Z., Li, L., & Liu, J. (2017). Identifying the spatial effects and driving factors of urban PM2.5 pollution in China. Ecological Indicators, 82(June), 61–75. https://doi.org/10.1016/j.ecolind.2017.06.043

Chinese Ministry of Construction. Design Standard for Energy Efficiency of Residential Buildings in Hot Summer and Cold Winter Zone JGJ 134-2001 (in Chinese)

Cui, S., Shi, Y., Groffman, P. M., Schlesinger, W. H., & Zhu, Y. G. (2013). Centennial-scale analysis of the creation and fate of reactive nitrogen in China (1910-2010). Proceedings of the National Academy of Sciences of the United States of America, 110(6), 2052–2057. https://doi.org/10.1073/pnas.1221638110

Ding, L., Liu, C., Chen, K., Huang, Y., & Diao, B. (2017). Atmospheric pollution reduction effect and regional predicament: An empirical analysis based on the Chinese provincial NOx emissions. Journal of Environmental Management, 196(x), 178–187. https://doi.org/10.1016/j.jenvman.2017.03.016

Elhorst, J. P. (2010). Handbook of Applied Spatial Analysis. In Handbook of Applied Spatial Analysis. https://doi.org/10.1007/978-3-642-03647-7

Feng, H., Zou, B., Wang, J., & Gu, X. (2019). Dominant variables of global air pollution-climate interaction: Geographic insight. Ecological Indicators, 99(December 2018), 251–260. https://doi.org/10.1016/j.ecolind.2018.12.038

Fu, Z., & Li, R. (2020). The contributions of socioeconomic indicators to global PM2.5 based on the hybrid method of spatial econometric model and geographical and temporal weighted regression. Science of The Total Environment, 703, 135481. https://doi.org/10.1016/j.scitotenv.2019.135481

GeoDa. Introducing GeoDa 1.14 https://geodacenter.github.io

Han, L., Zhou, W., Li, W., & Li, L. (2014). Impact of urbanization level on urban air quality: A case of fine particles (PM2.5) in Chinese cities. Environmental Pollution, 194, 163–170. https://doi.org/10.1016/j.envpol.2014.07.022

Han, X., Li, H., Liu, Q., Liu, F., & Arif, A. (2019). Analysis of influential factors on air quality from global and local perspectives in China. Environmental Pollution, 248, 965–979. https://doi.org/10.1016/j.envpol.2019.02.096

Hao, Y., & Liu, Y. M. (2016). The influential factors of urban PM2.5 concentrations in China: A spatial econometric analysis. Journal of Cleaner Production, 112, 1443–1453. https://doi.org/10.1016/j.jclepro.2015.05.005

Hao, Y., Zheng, S., Zhao, M., Wu, H., Guo, Y., & Li, Y. (2020). Reexamining the relationships among urbanization, industrial structure, and environmental pollution in China—New evidence using the dynamic threshold panel model. Energy Reports, 6, 28–39. https://doi.org/10.1016/j.egyr.2019.11.029

Hester, R.E., Harrison, R.M., 2009. Air Quality in Urban Environments. Royal Society of Chemistry.

Hu, B., Li, Z., & Zhang, L. (2019). Long-run dynamics of sulphur dioxide emissions, economic growth, and energy efficiency in China. Journal of Cleaner Production, 227, 942–949.

https://doi.org/10.1016/j.jclepro.2019.04.170

Huang, J. T. (2018). Sulfur dioxide (SO_2) emissions and government spending on environmental protection in China - Evidence from spatial econometric analysis. Journal of Cleaner Production, 175, 431–441. https://doi.org/10.1016/j.jclepro.2017.12.001

Jiao, J., Han, X., Li, F., Bai, Y., & Yu, Y. (2017). Contribution of demand shifts to industrial SO_2 emissions in a transition economy: Evidence from China. Journal of Cleaner Production, 164, 1455–1466. https://doi.org/10.1016/j.jclepro.2017.07.060

Li, L., Qian, J., Ou, C. Q., Zhou, Y. X., Guo, C., & Guo, Y. (2014). Spatial and temporal analysis of Air Pollution Index and its timescale-dependent relationship with meteorological factors in Guangzhou, China, 2001-2011. Environmental Pollution, 190, 75–81. https://doi.org/10.1016/j.envpol.2014.03.020

Li, M., & Zhang, L. (2014). Haze in China: Current and future challenges. Environmental Pollution, 189(2014), 85–86. https://doi.org/10.1016/j.envpol.2014.02.024

Liu, H., Fang, C., Zhang, X., Wang, Z., Bao, C., & Li, F. (2017). The effect of natural and anthropogenic factors on haze pollution in Chinese cities: A spatial econometrics approach. Journal of Cleaner Production, 165, 323–333. https://doi.org/10.1016/j.jclepro.2017.07.127

Liu, Q., Wang, S., Zhang, W., Li, J., & Dong, G. (2019). The effect of natural and anthropogenic factors on PM 2.5: Empirical evidence from Chinese cities with different income levels. Science of the Total Environment, 653, 157–167. https://doi.org/10.1016/j.scitotenv.2018.10.367

McCarty, J., & Kaza, N. (2015). Urban form and air quality in the United States. Landscape and Urban Planning, 139, 168–179. https://doi.org/10.1016/j.landurbplan.2015.03.008

Meng, K., Xu, X., Cheng, X., Xu, X., Qu, X., Zhu, W., … Zhao, Y. (2018). Spatio-temporal variations in SO_2 and NO_2 emissions caused by heating over the Beijing-Tianjin-Hebei Region constrained by an adaptive nudging method with OMI data. Science of the Total Environment, 642(2), 543–552. https://doi.org/10.1016/j.scitotenv.2018.06.021

National Bureau of Statistics of China. Annual data by province (2011-2017).

http://data.stats.gov.cn/easyquery.htm?cn=E0103

National Bureau of Statistics of China. China statistical Yearbook (2011-2017). http://www.stats.gov.cn/tjsj/ndsj/

Ooi Tatsuo. (2015). Spatial pattern analysis of local tourists using Moran's I statistic. 245–263. https://www.hosei.ac.jp/toukei/shuppan/g_shoho47-oi.pdf (in Japanese)

Peng, J., Zhang, Y., Xie, R., & Liu, Y. (2018). Analysis of driving factors on China's air pollution emissions from the view of critical supply chains. Journal of Cleaner Production, 203, 197–209. https://doi.org/10.1016/j.jclepro.2018.08.219

Qingfeng Zhang; Robert Crooks. (2013). Toward an Environmentally Sustainable Future: Country Environmental Analysis of the People's Republic of China. In Asian Development bank. Retrieved from https://www.adb.org/sites/default/files/publication/29943/toward-environmentally-sustainable-future-prc.pdf

Requia, W. J., Jhun, I., Coull, B. A., & Koutrakis, P. (2019). Climate impact on ambient PM2.5 elemental concentration in the United States: A trend analysis over the last 30 years. Environment International, 131(February), 104888. https://doi.org/10.1016/j.envint.2019.05.082

Ryou, H. gon, Heo, J., & Kim, S. Y. (2018). Source apportionment of PM10 and PM2.5 air pollution, and possible impacts of study characteristics in South Korea. Environmental Pollution, 240, 963–972. https://doi.org/10.1016/j.envpol.2018.03.066

Song, Y., Li, Z., Yang, T., & Xia, Q. (2020). Does the expansion of the joint prevention and control area improve the air quality? - Evidence from China's Jing-Jin-Ji region and surrounding areas. Science of the Total Environment, 706, 136034. https://doi.org/10.1016/j.scitotenv.2019.136034

Statheropoulos, M., Vassiliadis, N., & Pappa, A. (1998). Principal component and canonical correlation analysis for examining air pollution and meteorological data. Atmospheric Environment, 32(6), 1087–1095. https://doi.org/10.1016/S1352-2310(97)00377-4

Tong, Z., Baldauf, R. W., Isakov, V., Deshmukh, P., & Max Zhang, K. (2016). Roadside vegetation barrier designs to mitigate near-road air pollution impacts. Science of the Total Environment, 541, 920–927. https://doi.org/10.1016/j.scitotenv.2015.09.067

Smith, T. E. (2009). Spatial weights matrices. Geographical Analysis. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.2009.00768.x/full

Wang, S., Xing, J., Zhao, B., Jang, C., & Hao, J. (2014). Effectiveness of national air pollution control policies on the air quality in metropolitan areas of China. Journal of Environmental Sciences (China), 26(1), 13–22. https://doi.org/10.1016/S1001-0742(13)60381-2

Wang, Y., Han, R., & Kubota, J. (2016). Is there an Environmental Kuznets Curve for SO_2 emissions? A semi-parametric panel data analysis for China. Renewable and Sustainable Energy Reviews, 54, 1182–1188. https://doi.org/10.1016/j.rser.2015.10.143

Wang, Y., Han, R., & Kubota, J. (2016). Is there an Environmental Kuznets Curve for SO2 emissions? A semi-parametric panel data analysis for China. Renewable and Sustainable Energy Reviews, 54, 1182–1188. https://doi.org/10.1016/j.rser.2015.10.143

Weather station (2011.01.01-2017.12.31). (www.meteomanz.com)

Whiteman, C. D., Hoch, S. W., Horel, J. D., & Charland, A. (2014). Relationship between particulate air pollution and meteorological variables in Utah's Salt Lake Valley. Atmospheric Environment, 94, 742–753. https://doi.org/10.1016/j.atmosenv.2014.06.012

Xu, W., Sun, J., Liu, Y., Xiao, Y., Tian, Y., Zhao, B., & Zhang, X. (2019). Spatiotemporal variation and socioeconomic drivers of air pollution in China during 2005–2016. Journal of Environmental Management, 245(November 2018), 66–75. https://doi.org/10.1016/j.jenvman.2019.05.041

Yang, X., Wang, S., Zhang, W., & Yu, J. (2017). Are the temporal variation and spatial variation of ambient SO_2 concentrations determined by different factors? Journal of Cleaner Production, 167, 824–836. https://doi.org/10.1016/j.jclepro.2017.08.215

Zeng, J., Liu, T., Feiock, R., & Li, F. (2019). The impacts of China's provincial energy policies on major air pollutants: A spatial econometric analysis. Energy Policy, 132(1037), 392–403. https://doi.org/10.1016/j.enpol.2019.05.052

Zhang, H., Wang, Y., Hu, J., Ying, Q., & Hu, X. M. (2015). Relationships between meteorological parameters and criteria air pollutants in three megacities in China. Environmental Research, 140, 242–254. https://doi.org/10.1016/j.envres.2015.04.004

Zhao, S., Liu, S., Hou, X., Cheng, F., Wu, X., Dong, S., & Beazley, R. (2018). Temporal dynamics of SO_2 and NO_X pollution and contributions of driving forces in urban areas in China. Environmental Pollution, 242(X), 239–248. https://doi.org/10.1016/j.envpol.2018.06.085

Zhao, X., Zhou, W., Han, L., & Locke, D. (2019). Spatiotemporal variation in PM2.5 concentrations and their relationship with socioeconomic factors in China's major cities. Environment International, 133(September), 105145. https://doi.org/10.1016/j.envint.2019.105145

Zhu, L., Gan, Q., Liu, Y., & Yan, Z. (2017). The impact of foreign direct investment on SO_2 emissions in the Beijing-Tianjin-Hebei region: A spatial econometric analysis. Journal of Cleaner Production, 166, 189–196. https://doi.org/10.1016/j.jclepro.2017.08.032