Pol. J. Environ. Stud. Vol. 31, No. 6 (2022), 5633-5647
DOI: *10.15244/pjoes/152165* ONLINE PUBLICATION DATE: 2022-09-26

Original Research

Spatial Patterns, Drivers and Heterogeneous Effects of PM2.5: Experience from China

Xufeng Cui[1], Weige Huang[2], Wei Deng[1], Chengye Jia[3]*

[1] School of Business Administration, Zhongnan University of Economics and Law,182 Nanhu Ave., Wuhan 430073, China
[2]Wenlan School of Business, Zhongnan University of Economics and Law, 182 Nanhu Ave., Wuhan 430073, China
[3]School of Economics, Shandong University of Finance and Economics, NO.40 Shungeng Rd, Jinan, Shandong, 250014, China

Received: 21 April 2022
Accepted: 16 July 2022

Abstract

PM2.5 not only affects air visibility, but also can enter the lungs and blood through the respiratory tract, causing important damage to the human respiratory system, cardiovascular and cerebrovascular system. Infants, children, the elderly, patients with cardiovascular disease and chronic lung disease have become more sensitive to it, and has become an important factor endangering public health. Identifying PM2.5 spatial and temporal characteristics and influencing factors can provide key information for urban atmospheric environmental governance and public health improvement. Previous studies have only explored the influencing factors of PM2.5, while ignoring which is the more important factors. Firstly, this study explores the spatial-temporal evolution characteristics and spatial correlation characteristics of PM2.5 distribution in 285 cities in China. Then, a selection model--- Bayesian model average method is applied to identify which variables are more likely to affect PM2.5 in China. We find that (1) From 2000 to 2018, the high-value concentration areas of PM2.5 distribution in China were mainly distributed in the central and eastern regions, and showed the trend of "moving eastward". (2) Among all variables in this study, population density, utilization rate of common industrial solid wastes, per capita Gross Domestic Product (GDP), proportion of secondary industry to GDP, centralized treatment rate of sewage treatment plant and industrial emission of sulfur dioxide are the most important drivers to predict PM2.5 in China. In addition, we found that the relationship between the selected variables and PM2.5 tends to change over time. In addition, we also show that the influence of selected variables on PM2.5 depends on the distribution of PM2.5, that is, there is a heterogeneous effect.

Keywords: PM2.5, Industrial Structure, Government Governance, Economic Development, Bayesian Model Averaging

*e-mail: chengye.jia@sdufe.edu.cn

Introduction

With rapid speed of urbanization and industrialization, PM2.5 has become a serious issue in China due to its negative impact on environment [1-2] and human health [3-4]. For instance, PM2.5 is a kind of pollutant with fine particles which can be inhaled into the lungs, destroy the alveolar wall and dysfunction the lung and consequently lead to serious respiratory and cardiovascular diseases [5-7]. Therefore, it is crucial to study the important causing variables in order to reduce pollution and create environment conducive to human health.

There are lots of researches focusing on the driving variables of PM2.5 concentration from different perspectives. From social and economic perspective, many studies find that urbanization [8], income or per capita Gross Domestic Product (GDP) [9], energy intensity [10], coal consumption [11], transportation [12], population density [13-15], industrial structure [16-18] and so on, have significant impact on PM2.5 concentration. Moreover, [19] use dynamic spatial panel models to analyze the impact of foreign direct investment on PM2.5 concentration. [20] examines the relationship between R&D and PM2.5. In view of natural science, lots of literature focus on the impacts of meteorological conditions on PM2.5 concentration. Those conditions including temperature, rainfall, pressure, wind, planetary boundary layer and so on, are closely related to the PM2.5 concentration. For instance, [21] studies the effects of meteorological conditions on PM2.5 and finds that temperature and rainfall affect PM2.5 negatively and pressure have a positive impact on the PM2.5 concentration. [22] shows that increase in the speed of wind can effectively reduce PM2.5 concentration. [23] studies the interaction between planetary boundary layer and PM2.5 concentration. After figuring out the driving variables above, a number of researches have concentrated on finding effective ways of reducing PM2.5 emission such as improving energy efficiency [24] and using clean energy. In addition, another effective way in controlling PM2.5 concentration is government governance which includes increasing green coverage rate [25], domestic garbage harmless disposal rate [26], output value of products made from waste gas, waste water & solid wastes and controlling industrial emission of waste gas.

There are many variables which seem to influence PM2.5. However, some variables are not necessarily important in predicting PM2.5. In this paper, we apply Bayesian Model Averaging (BMA) to study the impacts of industrial structure, urban governance and economic development on PM2.5 concentration. More specifically, we first select important variables using BMA which is able to alleviate the model uncertainty problem since we have no idea in advance which variables should be included in the model. Therefore, we need to select important variables and then study the impacts of the variables selected on PM2.5.

The rest of the paper is organized as follows. The section 2 introduces the methods used in spatial analysis and variable selection, the section 3 describes the data, the section 4 is the analysis result, and the section 5 is the main conclusions.

Material and Methods

Geospatial Analysis

Spatial Autocorrelation Model

In this study, the spatial autocorrelation model is used to test the spatial distribution of PM2.5 concentration values in 285 cities in China and the degree of correlation between PM 2.5 concentration values and neighboring cities [27-28]. This model includes global autocorrelation model and local autocorrelation model.

Global autocorrelation model is mainly used to explore the spatial correlation degree of the whole region, which is generally expressed by Global Moran's *I*; Local autocorrelation model is mainly used to explore the degree of spatial heterogeneity between local regions, which is generally expressed by Local Moran's *I* [29], and their formulas are as follows.

Global Moran' *I*:

$$\mathrm{I} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{W_{ij}}(\mathrm{x_i}-\bar{\mathrm{x}})(\mathrm{x_j}-\bar{\mathrm{x}})}{S^2\sum_{\mathrm{i}=1}^{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{W_{ij}}}, I \in [-1,1] \tag{1}$$

Local Moran' *I*:

$$\begin{cases} I_i = X_i \times \sum_{j=1}^{n} W_{ij} \times X_j \\ X_i = (x_i - \bar{x}) / \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ X_j = (x_j - \bar{x}) / \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 \end{cases} \tag{2}$$

Among them, I represents the degree of spatial correlation, X_i and X_j are the PM2.5 concentration values in the study area, $\bar{X}$ is the average value of PM2.5 concentration in all areas, W_{ij} is the spatial weight matrix, and S_2 is the variance of attribute values. S'_i and S'_j are the standardized values of PM2.5 concentration in the study area. Moran's I is between [-1,1], and the closer the result is to -1, the greater the spatial difference in the study area; the closer the result is to 1, the smaller the spatial difference in the study area.

*Getis-Ord Gi**

Getis-ord Gi* can be used to explore the hot and cold spots of PM2.5 concentration distribution in the study area [30]. Although the spatial correlation model mentioned above can reflect the spatial correlation and difference of PM2.5 concentration distribution, it is difficult to accurately reflect the aggregation between specific high or low values, and this method can identify the hot and cold spots of different levels of attribute values. The formula is as follows:

$$G_i^* = \frac{\sum_{j=1}^{n} W_{ij}X_j - \overline{X}\sum_{j=1}^{n} W_{ij}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} W_{ij}^2 - \left(\sum_{j=1}^{n} W_{ij}\right)^2\right]}{n-1}}} \quad (3)$$

Among them, X_j is the concentration value of PM2.5 in *j* area, W_{ij} is the spatial weight, *n* is the total number of samples, and *S* is the standard deviation of samples. When $Gi^*\geq 1.96$, it is a high value; $Gi^*\leq -1.96$ is a low value.

Bayesian Model Averaging (BMA)

This section introduces Bayesian Model Averaging (BMA) which addresses model uncertainty problem [31]. Consider the following regression with y being outcome variable, X_l the covariates, α_l the constant and β_l the coefficients for the model indexed by *l*.

$$y = \alpha_l + X_l\beta_l + \epsilon \quad \epsilon \sim N(0, \sigma^2 I) \quad (4)$$

Assume there exist K variables in X with $X_l \in X$,, we will have models which is all possible combinations of X. We intend to figure out which variables are more important and should be included into the model. To address this problem, we need to estimate the model weights. That is the posterior model probabilities (PMP):

$$p(M_l|y, X) = \frac{p(y|M_l, X)p(M_l)}{p(y|X)} \quad (5)$$

p(y|X) is constant over all models. Thus, $p(M_l|y, X)$ is proportional to the marginal likelihood of the model $p(y|M_l, X)$ (the probability of the data given the model M_l) times a prior model probability $p(M_l)$ - that is, how probable the researcher thinks model M_l before looking at the data. This leads to the PMPs and thus the model weighted posterior distribution for any statistics θ such as α and β.

$$p(\theta|y, X) = \sum_{l=1}^{2^K} p(\theta|M_l, y, X)p(M_l\,|y, X) \quad (6)$$

The model prior $p(M_l)$ reflects researcher's prior beliefs. In this paper, we use a uniform prior probability which represents our lack of prior knowledge about the models.

Table 1. Definition of variables.

Variables Definition		
1	PM2.5	Average exposure to PM2.5
2	ciswur	Common industrial solid wastes utilized ratio
3	ctrstp	Centralized treatment rate of sewage treatment plant
4	dgaca	Domestic garbage harmless disposal rate
5	gcrca	Green covered rate of completed area
6	ieno	Industrial emission of nitrogen oxides
7	iesdio	Industrial emission of sulfur dioxide
8	pss	Proportion of service sector in GDP
9	psi	Proportion of secondary industry in GDP
10	gpc	GDP per capita
11	pd	Population density

Results and Discussion

Data Description

PM2.5 are obtained from the Atmospheric Composition Analysis Group at Dalhousie University[1]. Other variables studied in this paper are obtained from the China City Statistical Yearbook[2]. Our dataset includes 285 cities ranging from 2000 to 2018.

The variables studied are shown in Table 1.

Industrial Structure

We use the proportion of secondary industry in GDP (psi) and the proportion of service sector in GDP (pss) to represent industrial structure. In China, manufacture industry is the main sector of secondary industry and contribute greatly to the PM2.5 concentration [32, [33]. With the development of economy, service sector gradually replaces the energy intensive industry and thus the PM2.5 concentration decreases [17].

Government Governance

Following [27], government governance can be categorized into two groups-self-governance and public governance. We add common industrial solid wastes utilized ratio (ciswur) and centralized treatment rate of sewage treatment plant (ctrstp) into the self-governance group. Both variables have negative impact on PM2.5

[1] The original link to the data source of PM2.5 is invavid any more. Therefore, we provide new links to the dataset used in this paper. The researchers should be able to download the dataset from either of the two links by registering an account:

(1) GoogleDrive: https://drive.google.com/file/d/1uY70NrhM1uOiOqTCwU7U0K9F3NKIl7Lg/view?usp=sharing;

(2) BaiduDrive: https://pan.baidu.com/s/1eStmbu8jWNL06uKdO6MJZQ?pwd=tcij

[2] https://www.yearbookchina.com/

concentration. Moreover, we put green covered rate of completed area (gcrca) and domestic garbage harmless disposal rate (dgaca) into the group of public governance. [25] shows that increase in gcrca can reduce PM2.5 concentration.

Economic Development

GDP represents the level of economic development GDP per capita (inc) and population density (pd) can represent the level of economic development. The higher the inc and pd, the higher economic activities which cause air pollution [26]. Besides, it is common to us that more industrial activities mean more industrial emission of waste gas such as nitrogen oxides (NO_x) and sulfur dioxide (SO_2) and thus more PM2.5 concentration. The more industrial enterprises above the designated size are, the higher the degree of urban economic development is.

PM2.5 Concentrations

The all years sample ranges from 2000 to 2018. We also separate the sample into two subsamples which are the 2000-2010 and 2011-2018 samples. Fig. 1 shows the geographical distribution of average annual PM2.5 concentrations of the cities for the all years sample and the difference between two subsamples. The bottom right panel shows that PM2.5 increases in some cities especially in the north but decreases in some cities mostly located in the south. Table 2 summarizes the statistics of the variables in the all years sample and the differences between two subsamples. In the second table, SD in the sixth column is the standard deviation of differences in the mean values for each city in the 2000-2010 sample and the 2011-2018 sample. We can see that the mean difference in PM2.5 between two subsamples is -1.51 ug/m^3 and the standard deviation is 0.349, implying that mean difference is statistically significant by t test and Fig. 9 shows PM2.5 decreases over time after the year 2011.

Spatial Pattern of PM2.5

Spatial Trend Analysis

Through the trend analysis tool of ArcGIS 10.8 software, the spatial distribution of PM2.5 in 285 cities in China in 2000, 2010 and 2018 was analyzed from three perspectives, as shown in Fig. 2, where the Z axis is the concentration value of PM2.5, and the X and Y directions are due east and due north respectively. It can be seen from the figure that in 2000, the spatial distribution of PM2.5 in China showed the characteristics of "medium-high-low" from east to west. In 2010 and 2020, the spatial distribution of PM2.5 in China is characterized by "high in the east and low in the west". On the whole, from 2000 to 2018, the spatial difference of PM2.5 distribution in 285 cities in China from 2000 to 2020 is obvious, showing a pattern of "high in the east and low in the west".

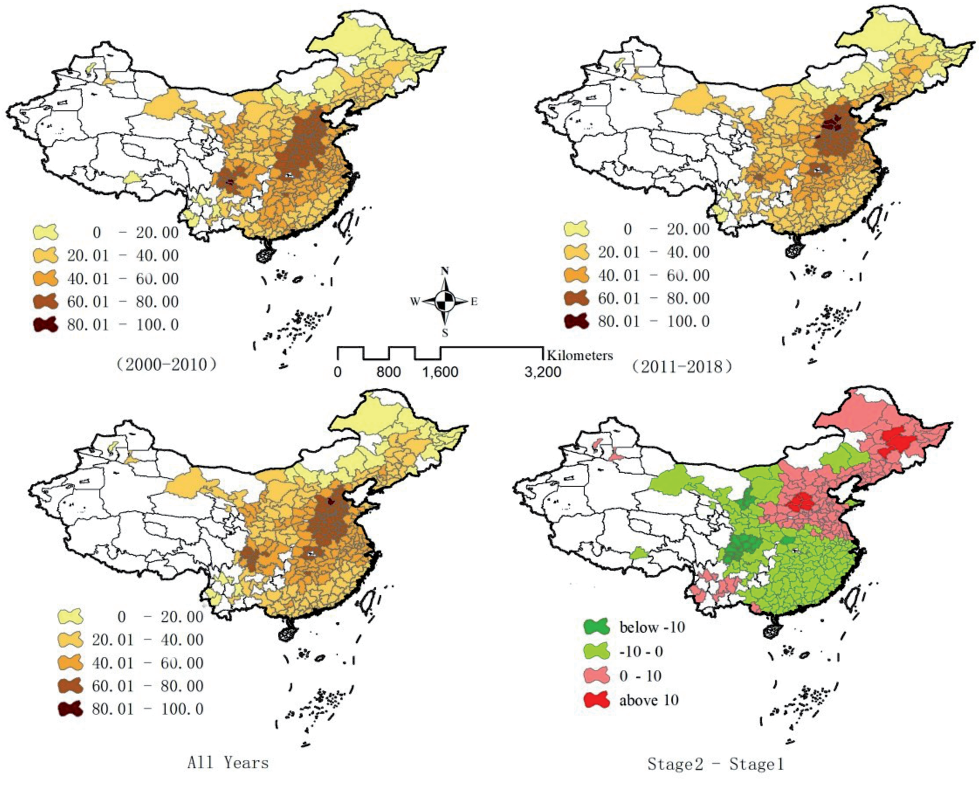

Fig. 1. Geographical distribution of average annual PM2.5 concentrations of the cities.

Table 2. Summary statistics.Temperature, Dir – Wind direction, Sp – Wind speed).

All years					
	Min	Mean	Median	Max	SD
PM2.5	4.53	42.77	40.26	80.24	17.26
ciswur	11.37	78.22	83.44	136.91	18.23
ctrstp	33.03	68.24	69.93	92.73	10.53
dgaca	39.07	81.29	83.63	111.05	12.87
gcrca	5.46	35.77	36.79	129.55	8.12
ieno	341.44	31667.19	22657.50	489813.85	43107.07
iesdio	207.86	60077.60	47138.62	558991.94	55582.78
pss	16.26	37.32	36.31	71.48	7.59
psi	17.64	47.53	48.11	83.23	9.71
gpc	6364.07	33197.90	27433.31	128025.31	21555.51
pd	5.08	421.04	340.21	2442.45	320.12
Differences					
	Min	Mean	Median	Max	SD
PM2.5	-21.02	-1.51	-1.69	13.27	5.90
ciswur	-127.97	3.47	3.16	72.16	16.79
ctrstp	-2.52	35.48	35.74	87.74	15.32
dgaca	-16.25	22.77	19.52	88.52	19.91
gcrca	-13.90	6.78	6.45	26.25	6.47
ieno	-73753.92	14553.21	3150.25	1003678.90	76804.44
iesdio	-172741.12	-6034.99	-4178.91	145776.25	33788.14
pss	-10.56	3.45	3.38	22.05	4.97
psi	-15.11	1.39	1.33	23.32	6.83
gpc	6544.43	30207.76	26070.46	129697.22	17261.30
pd	-900.77	18.75	11.67	612.05	85.40

Spatial Correlation Analysis

(1) Global spatial autocorrelation. Table 1 is the estimated value of global Moran's I index of PM2.5 distribution in 285 cities in China from 2000 to 2018. The global Moran's I index all passed the 10% significance test and was positive, which indicated that there was a significant positive spatial correlation in PM2.5 distribution, and there was an agglomeration phenomenon in the spatial distribution. From the evolution trend, the global Moran's I index shows a trend of continuous increase, which indicates that the spatial agglomeration degree of PM2.5 distribution among cities in the study area is constantly strengthening.

(2) Local spatial autocorrelation. Select 2000, 2010 and 2018 as time points to calculate local spatial autocorrelation, and divide 285 cities in China into four types: high-high(H-H), high-low(H-L), low-high(L-H)

Table 3. Global Moran's I estimated value.

Year	2000	2002	2004	2006	2008	2010	2012	2014	2016	2018
Moran's I	0.434	0.579	0.572	0.605	0.603	0.602	0.591	0.614	0.63	0.609
Z	12.7416	16.4331	16.149	16.9737	16.9622	16.807	16.5891	17.0207	17.4238	17.0454
P	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

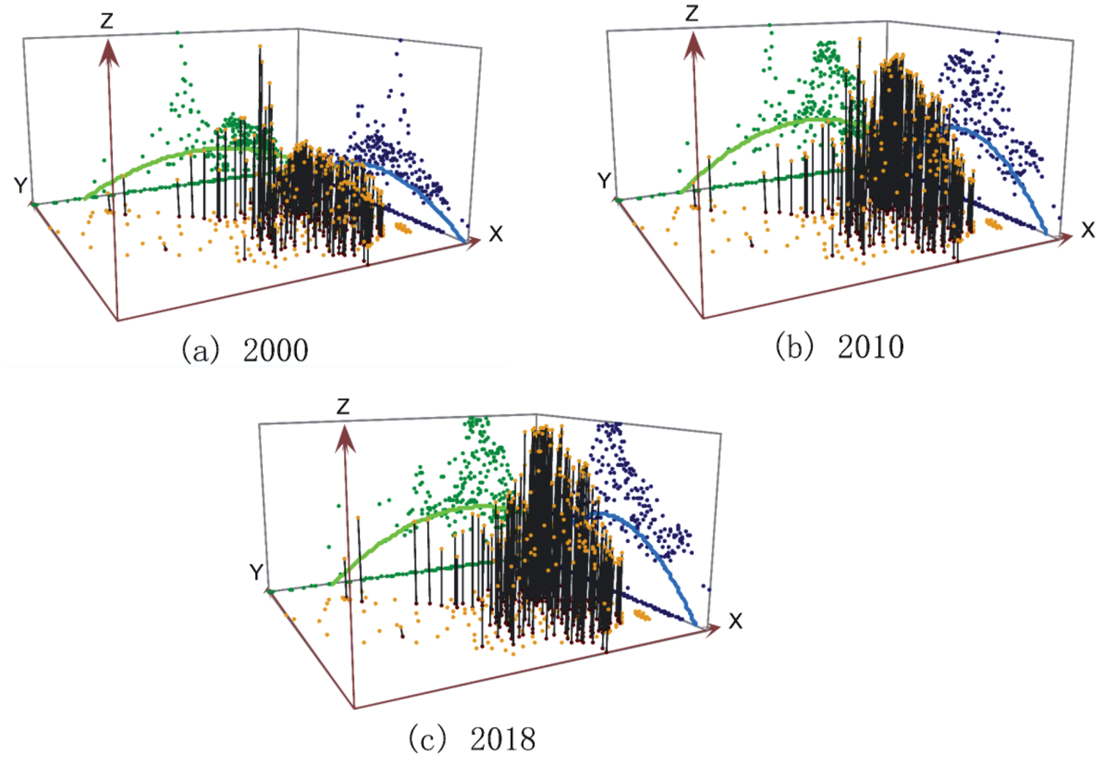

Fig. 2. Spatial trend analysis of PM2.5.

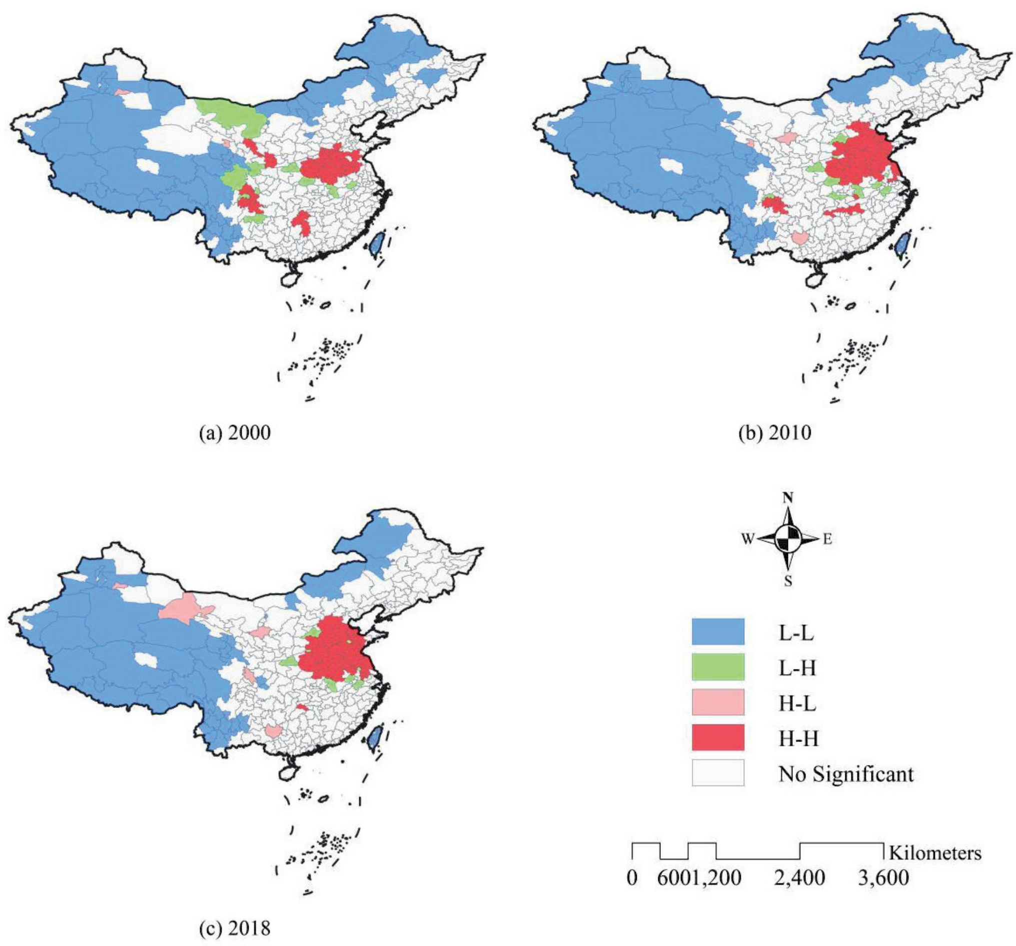

Fig. 3. LISA agglomeration analysis of PM2.5.

and low-low(L-L), and then summarize their spatial distribution and evolution characteristics. As shown in Fig. 3.

The "high-high" type, the concentration of PM2.5 in this type of city is higher than that in neighboring cities, showing a significant positive correlation. During the research period, the distribution of this type of cities in Henan Province, Shandong Province, Anhui Province and Jiangsu Province was relatively stable, and showed the trend of concentration and expansion. Some of them are scattered in Sichuan, Gansu and Hunan provinces, showing the trend of attenuation and contraction.

The "high-low" type, this type of city has a higher PM2.5 concentration, while its neighboring cities have a lower PM2.5 concentration, showing a negative correlation of "high self and low periphery". During the research period, this type of cities showed a trend of scattered distribution and slow increase in space and quantity, mainly distributed in Jiuquan City, Xining City, Urumqi City, Wuzhong City, Mianyang City and Hechi City.

The "low-high" type, this type of city has a lower PM2.5 concentration, while the neighboring cities have a higher PM2.5 concentration, showing a negative correlation of "self-low and peripheral high". During the research period, this type of cities showed a trend of moving westward in space, mainly distributed in Shangluo, Huanggang, Longnan, Bijie, Alxa League and Aba Tibetan and Qiang Autonomous Prefecture.

The "low-low" type, the concentration of PM2.5 in this type of city itself and neighboring cities is low, showing a positive correlation. During the research period, the number of cities of this type showed a decreasing trend, mainly distributed in Xinjiang Uygur Autonomous Region, Tibet Autonomous Region, Yunnan Province and Inner Mongolia Autonomous Region.

*Getis-Ord Gi**

The Getis-Ord Gi* index of PM2.5 distribution in 285 cities in China in 2000, 2010 and 2018 was calculated by ArcGIS 10.8 software, and the analysis results were divided into hot spots, sub-hot spots, sub-cold spots and cold spots, and the evolution map of cold spots of PM2.5 distribution was drawn according to this, as shown in Fig. 4.

In terms of quantity, the cold and hot spots of PM2.5 distribution in Chinese cities generally showed a decreasing trend. The total number decreased by 37, among which, the number of hot spots*** decreased

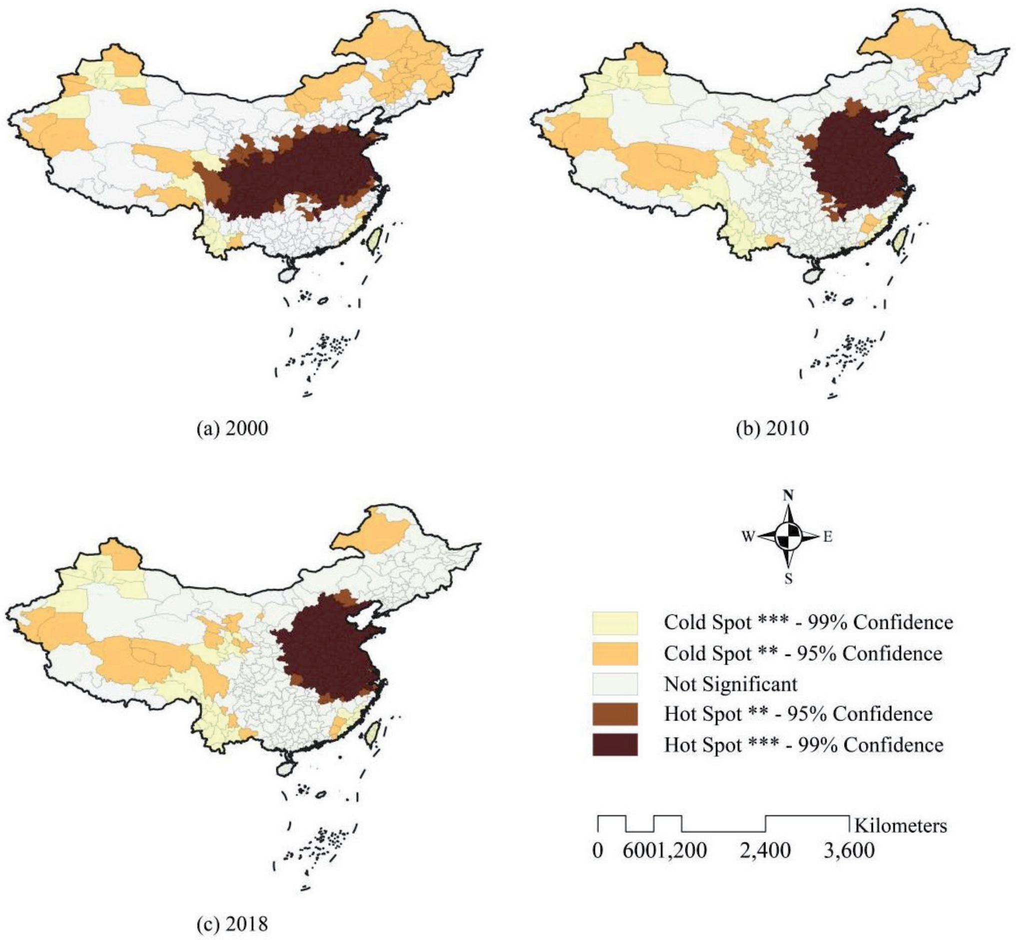

Fig. 4. Hot spots evolution of PM2.5.

Table 4. Regression results using variables selected by BMA.

	Dependent variable:		
	PM2.5		
	(All years)	(2000-2010)	(2011-2018)
pd	0.356***	0.424***	0.413***
	(0.053)	(0.053)	(0.053)
ciswur	0.320***	0.211***	0.218***
	(0.053)	(0.052)	(0.050)
pss		−0.264***	
		(0.047)	
gpc	−0.394***	−0.271***	−0.263***
	(0.055)	(0.055)	(0.049)
psi	0.318***		0.290***
	(0.052)		(0.047)
ctrstp	0.183***	0.231***	0.194***
	(0.050)	(0.053)	(0.048)
iesdio	0.141***	0.163***	
	(0.050)	(0.051)	
Constant	3.455***	6.665***	3.701***
	(0.419)	(0.308)	(0.482)
Observations	285	285	285

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5. Regression results using all variables studied.

	Dependent variable:		
	PM2.5		
	(All years)	(2000-2010)	(2011-2018)
ciswur	0.302***	0.216***	0.244***
	(0.056)	(0.056)	(0.052)
ctrstp	0.174***	0.212***	0.182***
	(0.053)	(0.059)	(0.050)
dgaca	0.035	0.019	−0.016
	(0.055)	(0.059)	(0.050)
gcrca	0.016	0.026	0.0003
	(0.050)	(0.052)	(0.050)
ieno	−0.016	0.030	−0.003
	(0.052)	(0.066)	(0.049)
iesdio	0.158***	0.134*	0.113**
	(0.056)	(0.070)	0.053
pss	−0.090	−0.219***	0.015
	(0.073)	(0.068)	(0.084)
psi	0.240***	0.080	0.288***
	(0.083)	(0.084)	(0.083)
gpc	−0.361***	−0.321***	−0.289***
	(0.069)	(0.071)	(0.065)
pd	0.367***	0.414***	0.404***
	(0.056)	(0.055)	(0.055)
Constant	4.055***	6.073***	3.726***
	(0.743)	(0.614)	(0.859)
Observations	285	285	285

Note: *p<0.1; **p<0.05; ***p<0.01

from 126 to 121, the number of hot spots ** decreased from 39 to 11, the number of Cold Spot** decreased from 26 to 9, and the number of Cold Spot*** increased from 48 to 64. The decreasing hot spots are mainly distributed in Sichuan Province and its surrounding cities, while the increasing cold spots are mainly distributed in Heilongjiang Province, Jilin Province, Inner Mongolia Autonomous Region, Tibet Autonomous Region and Yunnan Province.

Spatially, the distribution of PM2.5 hot spots in Chinese cities shows obvious regional differences, mainly reflected in "the central and eastern parts are hot and the surrounding areas are cold". Hot spots mainly include Shanxi Province, Hebei Province, Jiangsu Province, Anhui Province, Shandong Province, Henan Province and Hubei Province, while cold spots mainly include Yunnan Province, Tibet Autonomous Region, Xinjiang Uygur Autonomous Region and Heilongjiang Province. In reality, the terrain of hot spots is mainly plain, flat and open, with better livable and development conditions, while the terrain of cold spots is mainly plateau, with rugged terrain, and the population growth and economic development are relatively lagging behind. After consulting the data, it is found that the population density, industrial development level and per capita GDP level in hot spots are higher than those in cold spots, which also confirms the objectivity and scientificity of the selection of influencing factors in BMA analysis.

In addition, during the research period, the hot spots showed an obvious trend of eastward-moving, and showed a more concentrated feature. It mainly shows that the region with Sichuan Province as the core has withdrawn from the hot spots, and the hot spots are concentrated in Hubei Province, Anhui Province, Jiangsu Province, Shandong Province, Hebei Province and Henan Province as a whole, and the scattered sub-hot spots in the surrounding areas are decreasing simultaneously.

Important Factors

Table 4 shows regression results using variables selected by BMA. Although all samples share the sample selected variables such as pd, ciswur, gpc and ctrstp, the variables selected are different when the

Table 6. Regression results using variables selected by Lasso.

	Dependent variable:		
	PM2.5		
	(All years)	(2000-2010)	(2011-2018)
ciswur	0.309***	0.211***	0.218***
	(0.054)	(0.052)	(0.050)
ctrstp	0.185***	0.231***	0.194***
	(0.050)	(0.053)	(0.048)
iesdio	0.150***	0.163***	
	(0.050)	(0.051)	
pss	−0.085	−0.264***	
	(0.071)	(0.047)	
psi	0.245***		0.290***
	(0.080)		(0.047)
gpc	−0.349***	−0.271***	−0.263***
	(0.067)	(0.055)	(0.049)
pd	0.372***	0.424***	0.413***
	(0.055)	(0.053)	(0.053)
Constant	4.164***	6.665***	3.701***
	(0.728)	(0.308)	(0.482)
Observations	285	285	285

Note: *p<0.1; **p<0.05; ***p<0.01

samples are different. To be specific, BMA selects pd, ciswur, gpc, psi, ctrctp and iesdio for the all year sample, pd, ciswur, pss, gpc, ctrctp and iesdio for 2000-2010 sample and pd, ciswur, gpc and ctrctp for 2011-2018 sample. This suggests that the importance of variables for PM2.5 could change over time. We can see pss is important in estimating PM2.5 concentration for the 2000-2010 sample but is not for the 2011-2018 sample. Oppositely, psi and iesdio are important for the 2000-2010 sample but not for the 2011-2018 sample. gpc has negative impacts on PM2.5 for all samples and pss also negatively correlates with PM2.5 in the 2000-2010 sample. And all other selected variables have positive impacts on PM2.5.

Table 5 shows the regression results using all variables studied in this research. We find that the variables not selected by BMA are statistically insignificant, implying BMA does a good job in variable selection. To compare variables selected by BMA with other approaches, we show selection results using Lasso (Least absolute shrinkage and selection operator, Table 6) and Bagging (Fig. 5). As shown in Tables 3 and 5, BMA and Lasso select almost the same variables except pss. We also show results using Bagging which is based on decision tree since this method can capture

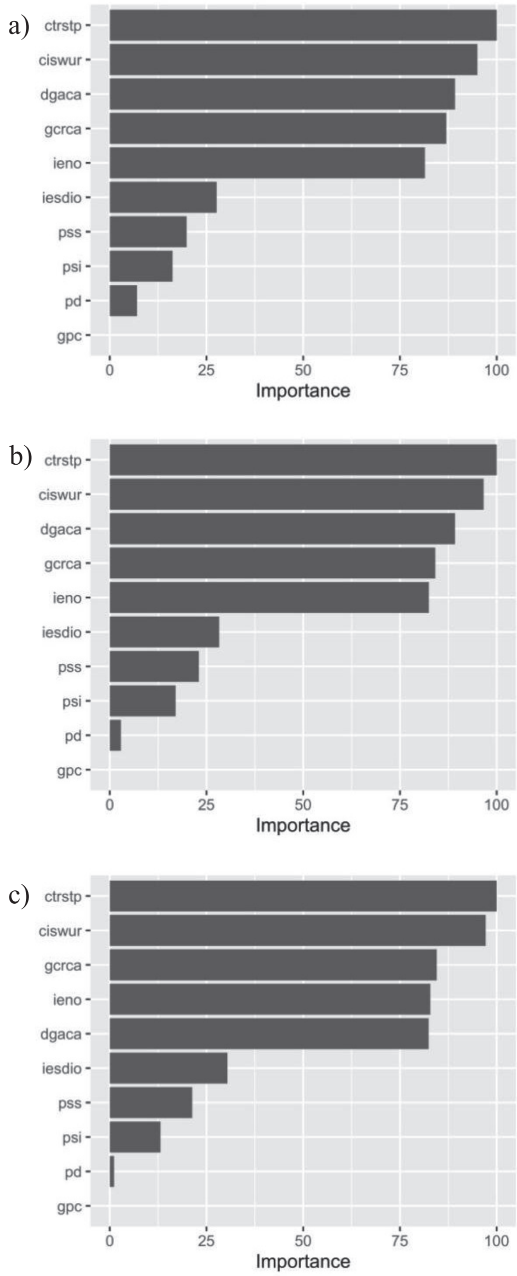

Fig. 5. Feature importance by Bagging: a) All years, b) 2000-2010, c) 2011-2018.

the non-linear relationship between dependent and independent variables without strict assumptions such as conditional independence. Fig. 5 shows that the top five important variables are ctrstp, ciswur, dgaca, gcrca and ieno where ctrstp is the most important one. Compared with the variables selected by BMA, however, we can see that dgaca, gcrca and ieno are not selected by BMA.

Heterogeneous Effects

In the section, we apply quantile regression to study the heterogeneous effects of variables selected by BMA on PM2.5 concentration in different samples.

Table 7. Quantile regression estimation results for different samples.

All years									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Intercept	-1.00	1.25*	4.41***	4.99***	4.79***	5.01***	5.32***	5.60***	5.69***
pd	0.22	0.36***	0.52***	0.57***	0.56***	0.56***	0.53***	0.49***	0.41***
ciswur	0.61***	0.4***	0.22***	0.18***	0.20***	0.20***	0.17***	0.20***	0.23***
gpc	-0.49***	-0.48***	-0.50***	-0.45***	-0.42***	-0.36***	-0.31***	-0.28***	-0.31***
psi	0.54***	0.51***	0.19***	0.15***	0.21***	0.2***	0.17***	0.16***	0.18***
ctrstp	0.38***	0.26***	0.18***	0.15***	0.14***	0.11**	0.13***	0.10*	0.10**
iesdio	0.27*	0.16*	0.14***	0.12***	0.11***	0.07	0.05	0.05	0.05
2000-2010									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Intercept	5.76***	6.23***	6.01***	6.2***	6.45***	6.56***	7.17***	7.34***	7.56***
pd	0.42***	0.41***	0.48***	0.47***	0.51***	0.59***	0.55***	0.47***	0.37***
gpc	-0.46***	-0.39***	-0.42***	-0.33***	-0.34***	-0.30***	-0.21***	-0.19***	-0.15***
ctrstp	0.46***	0.42***	0.30***	0.21***	0.18***	0.14***	0.07*	0.03	0.04
ciswur	0.40***	0.26**	0.21***	0.23***	0.21***	0.19***	0.15***	0.16***	0.19***
pss	-0.53***	-0.42***	-0.18***	-0.15***	-0.15***	-0.13***	-0.16***	-0.14***	-0.15***
iesdio	0.20	0.17	0.18**	0.19***	0.15***	0.11***	0.08*	0.08***	0.06*
2011-2018									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Intercept	1.45	2.71***	3.87***	4.06***	4.40***	4.92***	5.18***	5.94***	5.97***
pd	0.36**	0.5***	0.59***	0.57***	0.66***	0.70***	0.65***	0.60***	0.67***
gpc	-0.28**	-0.35***	-0.35***	-0.32***	-0.29***	-0.27***	-0.23***	-0.23***	-0.24***
ciswur	0.22	0.18**	0.15**	0.14***	0.11**	0.12***	0.11**	0.11***	0.07***
psi	0.41***	0.45***	0.31***	0.26***	0.20***	0.17***	0.14***	0.14***	0.14***
ctrstp	0.29**	0.18**	0.16**	0.20***	0.21***	0.17***	0.17***	0.11***	0.13***

The results are shown in Table 7 and Figs 6 to 8. Table 7 shows that almost every coefficient at different quantiles is statistically significant and we reject the null hypothesis that all coefficients across the whole quantiles are equal by conducting anova test for different samples. This implies that the heterogeneous effects of selected variables exist. For all years sample, we realize that population density (pd) has the greatest impact on PM2.5 concentration at each quantile (except 10th and 20th quantiles). GDP per capita (gpc) is the only factor which is negatively correlated with PM2.5. Also, we realize psi is the most important variable below 20th quantile when the level of PM2.5 concentration is very low. For 2000-2010 sample, pd is still the most important variable affecting PM2.5 concentration positively at each quantile (except 10th and 20th quantiles). In the opposite, pss and gpc have negative impacts on PM2.5 across all quantiles. When the quantile is below 0.2, the impact of proportion of service sector in GDP (pss) on PM2.5 is larger than gpc and centralized treatment rate of sewage treatment plant (ctrstp) affects PM2.5 positively more than pd does. In 2011-2018 sample, pd is the most important variable to raise the PM2.5 concentration at each quantile (expect 10th quantile). gpc is the only factor which tend to decrease the PM2.5 concentration. At 10^{th} quantile, psi increases PM2.5 more than pd does.

Figs 6 to 8 visualize the heterogeneous effects for all years, 2000-2010, 2011-2018 samples separately. The gray area represents the 95% confidence band at each quantile. As we can see, most gray areas do not cover the zero value in the vertical line, implying the statistically significance of the coefficients for variables at different quantiles. We can see that GDP per capita (gpc) is negatively correlated with PM2.5 but the magnitude of impacts tends to decrease over time for

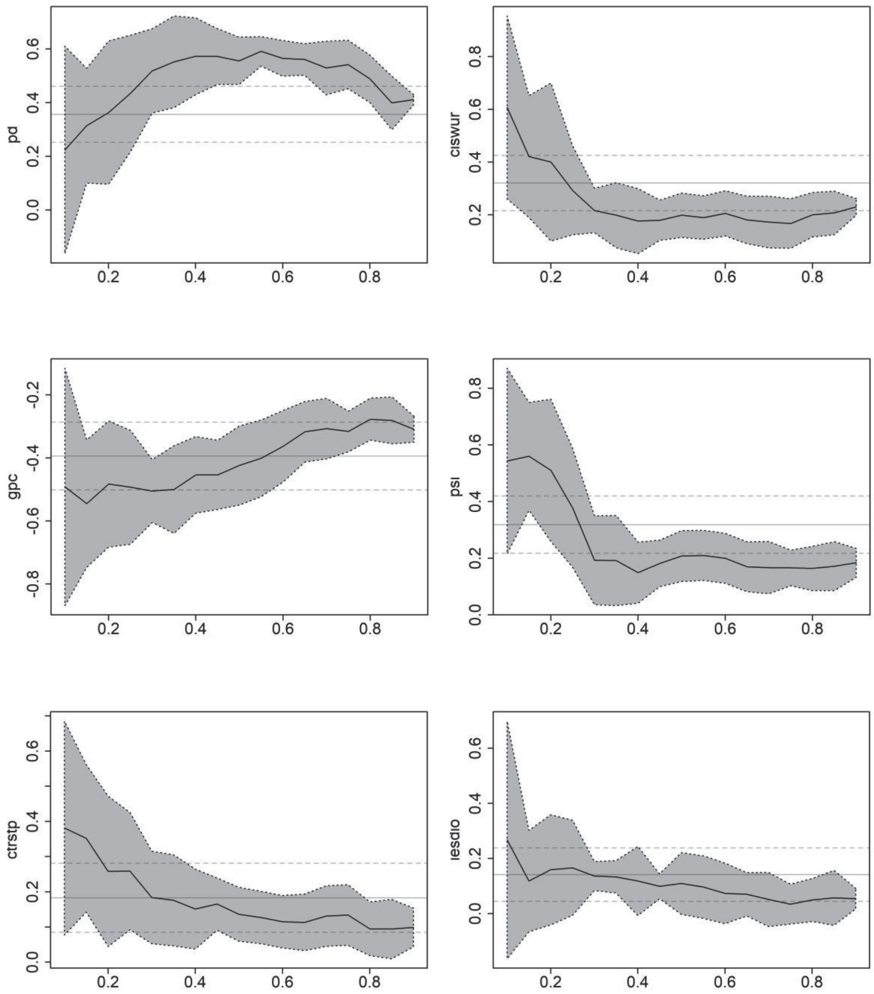

Fig. 6. Quantile effects (All years).
Note: Quantile regression estimates with 95% confidence intervals for the impact of influential variables on PM2.5 concentrations. The vertical axes show the estimated value of coeffcients of variables over the PM2.5 concentrations' distribution. The horizontal axes depict the quantile levels of PM2.5 concentration. The grey horizontal solid lines represent the corresponding OLS estimations with their 95% confidence interval (grey dashed line).

all samples. Also, population density (pd) is positively correlated with PM2.5 and the impacts increase with quantiles. The impacts of ciswur, psi, ctrstp and iesdio on PM2.5 tend to decrease with quantiles.

In sum, we conclude that effects of variables selected by BMA on PM2.5 concentration are heterogeneous.

Discussion

In this study, 285 cities in China were taken as the research objects, and the temporal and spatial characteristics and important factors of PM2.5 distribution were discussed, which deepened the related research of PM2.5 in China and provided some decision-making reference for relevant departments.

Firstly, it is the possible marginal contribution of this study. In the past, most of the researches on influencing factors of PM2.5 distribution stopped at "which factors may affect the distribution of PM2.5 in the study area", but in fact there are many factors affecting PM2.5, and this study uses Bayesian model average method to explore which factors are more important in predicting PM2.5 and their influence on PM2.5. Specifically, in this study, firstly, the Bayesian model average method, which can alleviate the uncertainty of the model, is used to select important variables, because we don't know which variables should be included in the model in advance, and then the influence of selected variables on PM2.5 is studied.

Secondly, according to the analysis results, some policy optimization suggestions are put forward, combined with the spatial and temporal distribution characteristics of PM2.5 in 285 cities in China, such as "the east is high and concentrated, and the west is low and scattered", as well as the important influencing factors such as population density, general industrial solid waste utilization rate, etc., we think that the optimization measures that can be taken

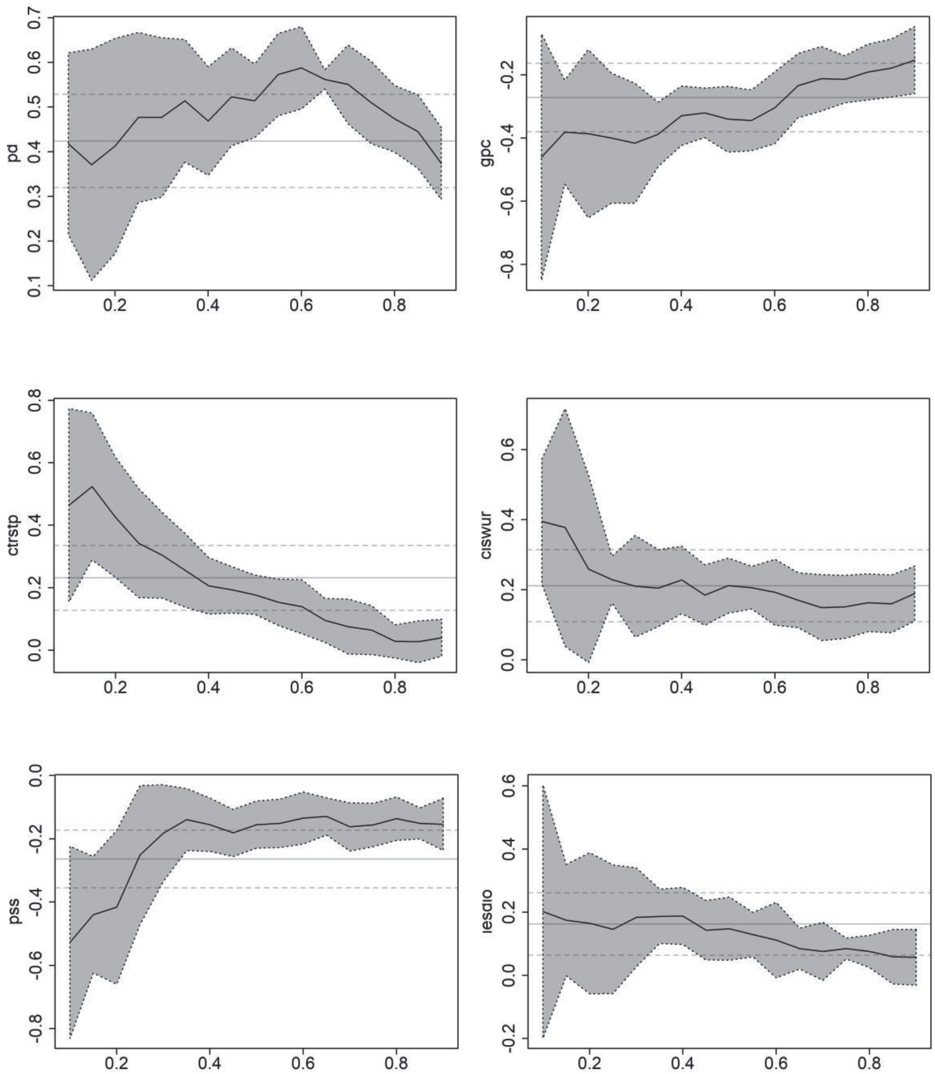

Fig. 7. Quantile effects (2000-2010).
Note: Quantile regression estimates with 95% confidence intervals for the impact of influential variables on PM2.5 concentrations. The vertical axes show the estimated value of coeffcients of variables over the PM2.5 concentrations' distribution. The horizontal axes depict the quantile levels of PM2.5 concentration. The grey horizontal solid lines represent the corresponding OLS estimations with their 95% confidence interval (grey dashed line).

in the future are: (1) Optimize the construction of public transportation system, especially in areas where PM2.5 is high and concentrated (Hubei, Anhui, Jiangsu, Henan, Hebei, Shandong), these areas have high population density, high per capita income level, relatively high private car travel rate, and automobile exhaust emission is also one of the known important sources of PM 2.5. Only when the public transportation system is more convenient, more perfect and closer to life can the demand for private car travel be relatively reduced; (2) Improve the coverage rate of domestic waste sorting treatment. At present, the implementation rate of domestic waste sorting treatment in China is still low, which will lead to the increase of unnecessary waste incineration, which in turn will lead to the increase of PM2.5 concentration; (3) Update the industrial solid waste treatment technology, the wrong treatment of industrial solid waste will also lead to the increase of PM2.5 concentration, so the related treatment technology should be innovated to improve the conversion rate of industrial solid waste into industrial raw materials; (4) Promote the production and utilization of clean energy vigorously The use and production of traditional energy sources (such as raw coal, thermal power, etc.) will lead to the generation of PM2.5. Now China's development has entered a new stage, and energy sources such as wind power and photovoltaic have been put into production initially. In the future stage, more attention should be paid to the production and diversified utilization of clean energy to make the energy utilization structure more environment-friendly.

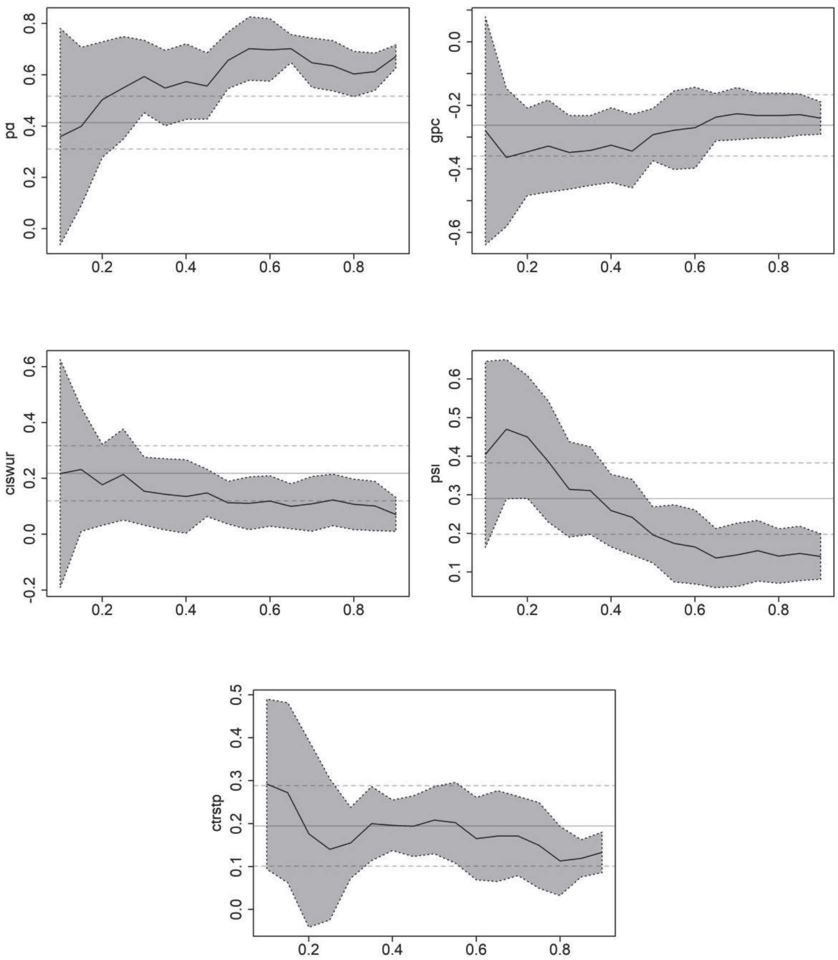

Fig. 8. Quantile effects (2011-2018).
Note: Quantile regression estimates with 95% confidence intervals for the impact of influential variables on PM2.5 concentrations. The vertical axes show the estimated value of coefficients of variables over the PM2.5 concentrations' distribution. The horizontal axes depict the quantile levels of PM2.5 concentration. The grey horizontal solid lines represent the corresponding OLS estimations with their 95% confidence interval (grey dashed line).

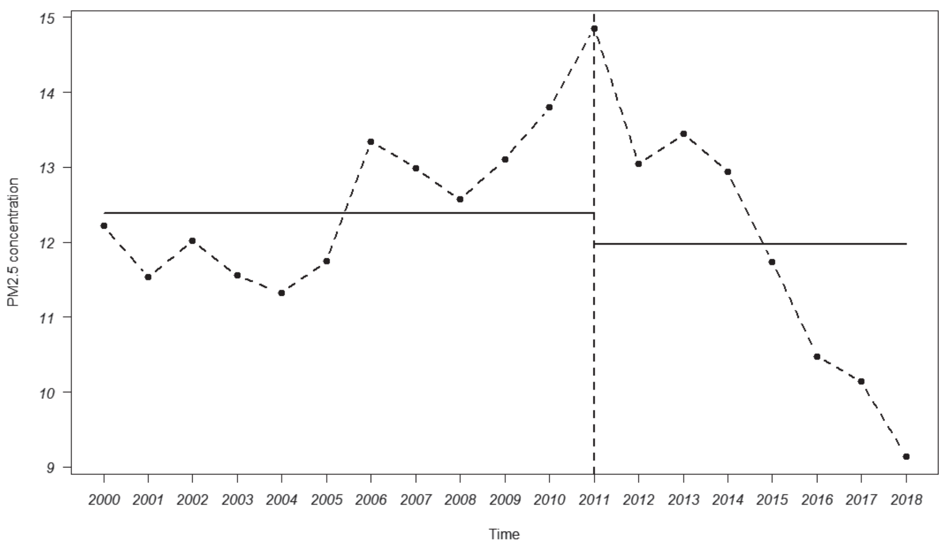

Fig. 9. The time trend of PM2.5 concentration.
Note: The solid lines represent the mean values of PM2.5 concentration in the sample 2000-2010 and 2011-2018 respectively. The black point connected by dashed lines is the total sum of PM2.5 concentration divided by 1000 across 285 cities in each year. The vertical dashed line represents the year 2011 which divides the total sample into two subsamples.

Conclusions

In this study, firstly, the spatial analysis method was used to explore the spatial-temporal evolution characteristics and spatial correlation characteristics of PM2.5 distribution in 285 cities in China. Secondly, BMA method was used to explore which variables have more important influence on PM2.5 concentration in China. The main conclusions of this study are as follows:

(1) The overall distribution of PM2.5 concentration in 285 cities in China showed the spatial characteristics of "high in the east and low in the west" and showed the trend of "eastward-moving"; In terms of spatial correlation, the spatial correlation degree of PM2.5 distribution is relatively high, showing the spatial characteristics of "hot in the east and cold in the west", and the agglomeration trend in the east is obvious.

(2) Among all the variables studied, population density, utilization rate of general industrial solid waste, per capita GDP, proportion of secondary industry to GDP, centralized treatment rate of sewage treatment plant and industrial emission of sulfur dioxide are the most important drivers to predict PM2.5 concentration in China.

(3) Sample separation and quantile regression were applied to explore the effects of independent variables on PM2.5 concentration, and display a heterogenous pattern – the effects change over time and across the distribution of PM2.5 concentration.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, Zhongnan University of Economics and Law (CN) (Grant number: 2722022BY014) and Hubei Provincial Department of Education Project of Philosophy and Social Sciences (CN) (No. 21G015) .

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. KAN H., CHEN R., TONG S. Ambient air pollution, climate change, and population health in China. Environment international, **42**, 10, **2012**.
2. TAI A.P., MICKLEY L.J., JACOB D.J. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. Atmospheric environment, **44** (32), 3976, **2010**.
3. HUEGLIN C., GEHRIG R., BALTENSPERGER U., GYSEL M., MONN C., VONMONT H. Chemical characterisation of PM2.5, PM10 and coarse particles at urban, near-city and rural sites in Switzerland. Atmospheric Environment, **39** (4), 637, **2005**.
4. ZHANG X., FAN C., REN Z., FENG H., ZUO S., HAO J., LIAO J., ZOU Y., MA L. Maternal PM2.5 exposure triggers preterm birth: a cross-sectional study in Wuhan, China. Global health research and policy, **5**, 1, **2020**.
5. LI L., YANG J., SONG Y., CHEN P., OU C. The burden of COPD mortality due to ambient air pollution in Guangzhou, China. Scientific reports, **6** (1), 1, **2016**.
6. XING Y., XU Y., SHI M., LIAN Y. The impact of PM2.5 on the human respiratory system. Journal of thoracic disease, **8** (1), E69, **2016**.
7. SAMOLI E., ANALITIS A., TOULOUMI G., SCHWARTZ J., ANDERSON H. R., SUNYER J., BISANTI L., ZMIROU D., VONK J.M., PEKKANEN J., GOODMAN P, PALDY A., SCHINDLER C., KATSOUYANNI K. Estimating the exposure–response relationships between particulate matter and mortality within the APHEA multicity project. Environmental health perspectives, **113** (1), 88, **2005**.
8. XU B., LIN B. Regional differences of pollution emissions in China: Contributing factors and mitigation strategies. Journal of Cleaner Production, **112**, 1454, **2016**.
9. JI X., YAO Y., LONG X. What causes PM2.5 pollution? Cross-economy empirical analysis from socioeconomic perspective. Energy Policy, **119**, 458, **2018**.
10. XU B., LUO L., LIN B. A dynamic analysis of air pollution emissions in China: Evidence from nonparametric additive regression models. Ecological indicators, **63**, 346, **2016**.
11. DE RIDDER K., LEFEBRE F., ADRIAENSEN S., ARNOLD U., BECKROEGE W., BRONNER C., DAMSGAARD O., DOSTAL I., DUFEK J., HIRSCH J., INTPANIS L., KOTEK Z., RAMADIER T., THIERRY A., VERMOOTE S., WANIA A., WEBER C. Simulating the impact of urban sprawl on air quality and population exposure in the German Ruhr area. Part II: Development and evaluation of an urban growth scenario. Atmospheric environment, **42** (30), 7070, **2008**.
12. HAO Y., LIU Y. The influential factors of urban PM2.5 concentrations in China: a spatial econometric analysis. Journal of Cleaner Production, **112**, 1443, **2016**.
13. DING Y., ZHANG M., QIAN X., LI C., CHEN S., WANG W. Using the geographical detector technique to explore the impact of socioeconomic factors on PM2.5 concentrations in China. Journal of Cleaner Production, **211**, 1480, **2019**.
14. LUO J., DU P., SAMAT A., XIA J., CHE M., XUE Z. Spatiotemporal pattern of PM2.5 concentrations in mainland China and analysis of its influencing factors using geographically weighted regression. Scientific reports, **7** (1), 1, **2017**.
15. MA Q., CAI S., WANG S., ZHAO B., MARTIN R.V., BRAUER M., COHEN A., JIANG J., ZHOU W., HAO J., FROSTAD J. Impacts of coal burning on ambient PM2.5 pollution in China. Atmospheric Chemistry and Physics, **17** (7), 4477, **2017**.
16. CHEN X. Spatiotemporal mapping and multiple driving forces identifying of PM2.5 variation and its joint management strategies across China. Journal of Cleaner Production, **250**, 119534, **2020**.
17. JI X., YAO Y., LONG X. What causes PM2.5 pollution? Cross-economy empirical analysis from socioeconomic perspective. Energy Policy, **119**, 458, **2018**.
18. YANG D., WANG X., XU J., XU C., LU D., YE C., WANG Z., BAI L. Quantifying the influence of natural

and socioeconomic factors and their interactive impact on PM2.5 pollution in China. Environmental Pollution, **241**, 475, **2018**.
19. CHENG Z., LI L., LIU J. Identifying the spatial effects and driving factors of urban PM2.5 pollution in China. Ecological Indicators, **82**, 61, **2017**.
20. WU W., ZHANG M., DING Y. Exploring the effect of economic and environment factors on PM2.5 concentration: a case study of the Beijing-Tianjin-Hebei region. Journal of Environmental Management, **268**, 110703, **2020**.
21. ZHAO D., CHEN H., SUN X., SHI Z. Spatio-temporal variation of PM2.5 pollution and its relationship with meteorology among five megacities in China. Aerosol and Air Quality Research, **18** (9), 2318, **2018**.
22. LIU Z., SHEN L., YAN C., DU J., LI Y., ZHAO H. Analysis of the Influence of Precipitation and Wind on PM2.5 and PM10 in the Atmosphere. Advances in Meteorology 2020, **2020**.
23. MIAO Y., LI J., MIAO S., CHE H., WNAG Y., ZHANG X., ZHU R., LIU S. Interaction between planetary boundary layer and PM2.5 pollution in megacities in China: A Review. Current Pollution Reports, **5** (4), 261, **2019**.
24. FU Z., LI R. 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, **2020**.
25. CHEN M., DAI F., YANG B., ZHU S. Effects of neighborhood green space on PM2.5 mitigation: Evidence from five megacities in China. Building and Environment, **156**, 33, **2019**.
26. SHI T., ZHANG W., ZHOU Q., WANG K. Industrial structure, urban governance and haze pollution: Spatiotemporal evidence from China. Science of The Total Environment, **742**, 139228, **2020**.
27. BAO H., SHAN L., WANG Y., JIANG Y., LEE C., CUI X. How does local real estate investment influence neighborhood PM2.5 concentrations? A spatial econometric analysis. Land, **10**, 518, **2021**.
28. MORAN P.A.P. Notes on continuous stochastic phenomena. Biometrika, **37**, 17, **1950**.
29. CUI X., HUANG S., LIU C., ZHOU T., SHAN L., ZHANG F., CHEN M., LI F., DE VRIES W.T. Applying SBM-GPA Model to Explore Urban Land Use Efficiency Considering Ecological Development in China. Land, **10** (9), 912, **2021**.
30. TADESSE S., ENQUESELASSIE F., HAGOS S. Spatial and space-time clustering of tuberculosis in Gurage Zone, Southern Ethiopia. PloS ONE, **13** (6), e0198353, **2018**.
31. HOETING J.A., MADIGAN D., RAFTERY A.E., VOLINSKY C.T. Bayesian model averaging: a tutorial. Statistical science, **14** (4), 382, **1999**.
32. GUAN D., SU X., ZHANG Q., PETERS G.P., LIU Z., LEI Y., HE K. The socioeconomic drivers of China's primary PM2. 5 emissions. Environmental Research Letters, **9** (2), 024010, **2014**.
33. FANG F. A study of financial risks of listed manufacturing companies in China. Journal of financial risk management, **5** (4), 229, **2016**.