Pol. J. Environ. Stud. Vol. 31, No. 4 (2022), 3373-3385
DOI: *10.15244/pjoes/145606* ONLINE PUBLICATION DATE: 2022-04-07

Original Research

Revealing the Impact of Socio-Economic Metrics on the Air Quality on Northeast China Using Multivariate Statistical Analysis

Mei Yang[1]*, Weixin Wang[1,2], Yuze Li[3], Yuwei Du[4], Feng Tian[5]**

[1]East Asian Disciplines, Pai Chai University, Daejeon 35345, South Korea
[2]Shandong Vocational Animal Science and Veterinary College, Weifang 261000, China
[3]Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ, Scotland
[4]School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China
[5]Shandong College of Information Technology, Weifang 261000, China

Received: 24 July 2021
Accepted: 7 January 2022

Abstract

Urban air pollution is an important issue facing China in the midst of rapid urbanization and economic development. To investigate the regional air quality characteristics and its drivers in Northeast China, this paper compares the spatial and temporal characteristics of air quality between cities and analyzes the influence of socioeconomic variables by using statistical analysis methods and geographic model. The results show that the air quality index (AQI) showed a downward trend in time and decreased from southwest to northeast in space. The duration of heavy pollution condition was not only controlled by the distribution of pollutant concentrations, but also influenced by the topography. Based on the mean concentrations of the 6 pollutants, 37 cities were divided into 4 categories by cluster analysis, reflecting the levels and characteristics of pollution. The level of industrialization was the most important cause for air quality, followed by the size of the city and the degree of economic development. The AQI predicted by geographic weighted regression model (GWR) showed a lower goodness of fit in developed cities, indicating that the factors controlling air quality are more complex in these regions. The influence of different socioeconomic metrics on AQI showed large spatial differences. AQI was more sensitive to variations in socioeconomic metrics in less developed small and medium-sized cities. This study provides a theoretical basis for revealing the causes of urban air pollution and formulating pollution control measures in Northeast China.

Keywords: atmospheric pollution, geographic weighted regression model, industrialization, spatial heterogeneity

*e-mail: qansy@163.com
**e-mail: sdwfwwx@163.com

Introduction

Air pollutants are characterized by fast propagation speed, strong diffusion ability and wide influence range, and thus it is difficult to perform good treatment within a short time [1-4]. The continuous deterioration of air quality will become a major threat to human health [5, 6]. In recent decades, air pollution is no longer limited to the local scale, and regional, national and even global air pollution is becoming more and more common [7, 8]. Since 1980s, a large amount of energy has been consumed in China due to the rapid growth of economy, and a great amount of air pollutants has been produced and emitted. Previous studies have shown that the annual emissions of CO_2 and SO_2 in China are the highest in the world. In addition to the traditional air pollution caused by single pollutant types, air pollution problems with complex sources such as haze and photochemical smog have become increasingly prominent in China [9]. In recent years, large-scale and persistent air pollution incidents represented by haze have occurred frequently in China, which has had a great negative impact on residents' quality of life and economic development [10]. How to accelerate the improvement of ambient air quality and effectively control air pollution is one of the most important challenges of China's effort on environmental protection at present. In Northeast China, the problem of air pollution is severe as the population is densely distributed, the number and size of industrial and mining enterprises are large, and the heating period relying on coal burning in winter is long.

The temporal and spatial evolution of air quality is a complex dynamic process under the joint influence of natural factors (e.g., meteorology and terrain conditions) and socio-economic factors (e.g., economy, population and land) [11-16]. Since the beginning of 21st century, the global air pollution has gradually intensified with the accelerated economic growth, urban expansion and population explosion [16-18]. Particularly, air pollution caused by anthropic factors such as population agglomeration, traffic congestion and excessive energy consumption has become increasingly prominent during the process of urbanization [13, 19, 20]. Therefore, a comprehensive exploration on the temporal and spatial variation of urban air quality and its socio-economic drivers is not only conducive to scientific understanding of air pollution characteristics, but also provides reference for the formulation of regional prevention and control measures. In previous studies, the influence of socio-economic factors on atmospheric environment has been widely reported in different spatio-temporal scales. For example, the nonlinear relationship between economic growth and atmospheric environment is analyzed based on the Environmental Kuznets Curve [21, 22]; Buehn, et al. (2013) investigated the temporal and spatial variation of air quality across 122 countries in the past 20 years, and found that economic growth shows closely relation with air quality. The energy consumption and industrialization are important factors leading to deteriorated urban air quality, as they are accompanied with the emission of pollutants, especially the consumption of coal. In the maturity developing stage with high energy efficiency, the air pollution can be greatly reduced [23]. The spatial distribution of population and roads has also been proved to be significantly related to air quality [24, 25]. The increase of population is generally associated with the increased scale of productive activities, motor vehicle emissions, industrial pollution and construction dust, which will be another important factor intensifying the air pollution [26].

Northeast China is a base of heavy industry, where resource-based enterprises with high energy consumption and high pollution are concentrated, and the atmospheric environment is seriously polluted. In some areas, the harmful gases produced by the massive combustion of coal, oil and natural gas in industrial production are directly emitted into the atmosphere without treatment, causing serious air pollution, thus leading to a series of environmental and human health problems. In recent years, the environmental treatment projects led by government and environmental protection policies, the transformation of economic production mode and the optimization of industrial structure have made the air quality significantly improved in some regions, but the regional air pollution problem is still severe. In Northeast China, the distribution of socio-economic factors varies greatly in space. In some cities, heavy industry such as heavy machinery, steel, petrochemical industry and equipment manufacturing still works as the pillar of economy. Economic recession and population loss happen in these cities because of the over-capacity, but the air pollution is severe. In other cities, industry transformation has been completed and the air pollution has been mitigated, leading to gradual population inflow. Because of the pronounced spatial heterogeneity, the relationship between economic structure, social characteristics and air pollution is complex in Northeast China. In some studies, the sources and temporal evolution of pollutants in Northeast China have been investigated [27]. However, to our knowledge, there is still few research to systematically analyze the present situation, spatial heterogeneity of air quality and its socio-economic causes in Northeast China by using multivariate statistical methods, which is of great significance for regional environmental protection, sustainable economic development and residents' health.

Based on the monitoring data of air quality and socio-economic indicators in Northeast China from 2015 to 2019, this paper aims to determine the dynamic mechanism and spatial differences of socio-economic controls on air quality in Northeast China. The specific objectives of the present study are as follows: first, describe the characteristics of spatio-temporal variation of air pollution and socio-economic conditions; and then, determine the dynamic mechanism and spatial differences of socio-economic controls on air quality in Northeast China.

Methods and Materials

Study Area

Northeast China, including provinces of Heilongjiang, Jilin and Liaoning, is the geographical unit with the highest latitude in China. The geographical location of northeast China and the distribution of cities are shown in Fig. 1. Northeast China is rich in mineral resources, mainly including coal, iron ore and oil. Coal resources are mainly distributed in central and western Liaoning and northeast Heilongjiang, iron ore resources are mainly concentrated in Anshan and Benxi cities, and oil resources are mainly distributed in Songnen Plain and the middle and lower reaches of Liaohe River. In the early days of the new China, Northeast China formed an industrial system dominated by heavy industries such as steel, machinery, oil and coal relying on the abundant resources, and became an important heavy industry and commodity grain base in China, playing an important role in promoting economic and social development. Since 1990s, the growth of traditional industries has stagnated due to the adjustment of national economic layout, making the population and economy in Northeast China have presented a decreasing trend. The serious ecological and environmental problems have occurred due to the extensive economic growth pattern at the expense of resource consumption and environmental damage, which have become one of the important reasons that constrain the economic development in Northeast China.

Description of Data

Air Quality Index (AQI) is a comprehensive index proposed by the Ministry of Ecology and Environment of China, which is used to synthetically quantify air quality and inform the public about levels of air pollution. It is calculated based on 6 types of pollutants namely $PM_{2.5}$, $PM1_0$, SO_2, NO_2, CO and O_3. The smaller the value, the better the air quality. According to the Chinese Ambient Air Quality Standards (CAAQS, GB3095-2012), AQI is divided into 6 categories according to their level of hazard to human health. The value below 50 is determined as Class I, that between 51 and 100 is determined as Class II, that between 101 and 150 is determined as Class III, that between 151 and 200 is determined as Class IV, that between 201 and 300 is determined as Class V and that above 301 is determined as Class VI. Among the categories, Class I and Class II are considered

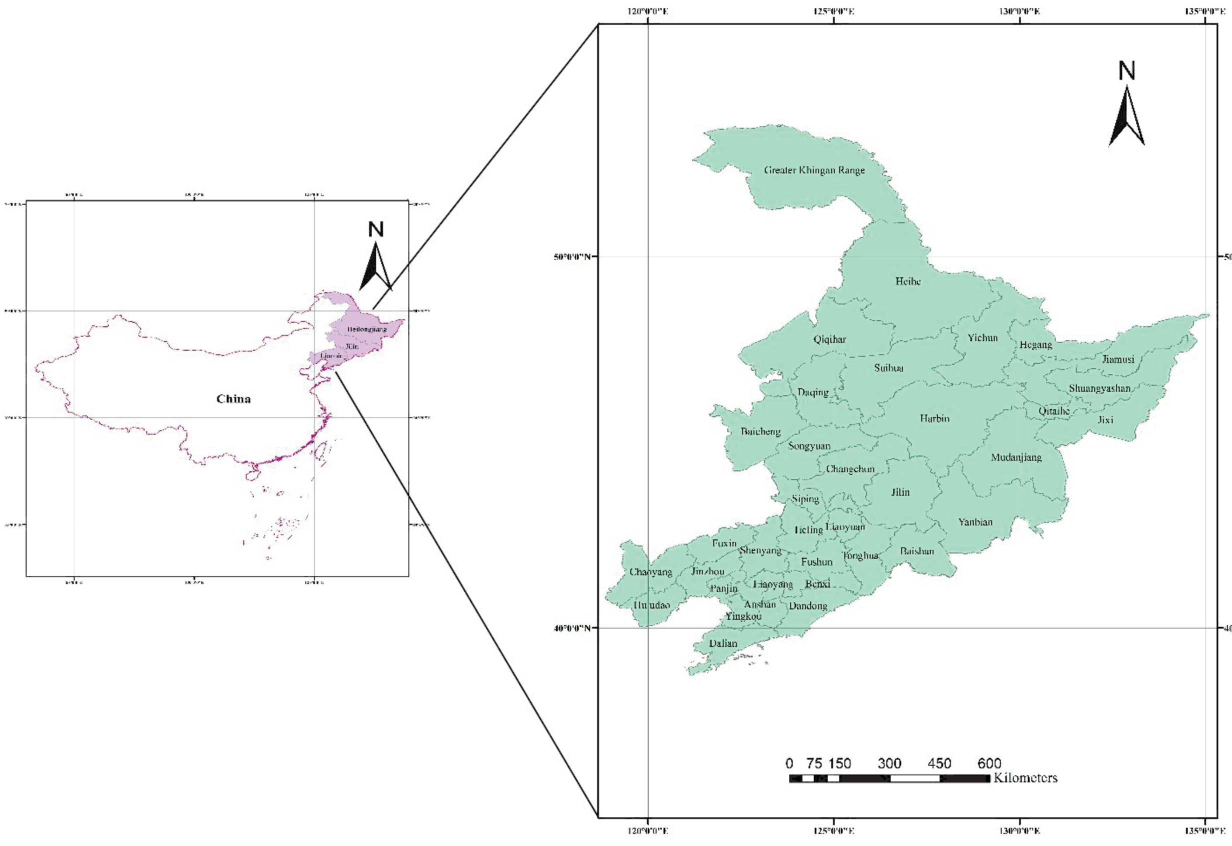

Fig. 1. Location of study area and the distribution of cities in Northeast China.

Table 1. Overview of the socio-economic and air quality variables.

Type of parameters	Metrics	Unit
Socio-economic indicators	Population (POP)	Millon
	Gross domestic product (GDP)	$\times 10^9$ RMB
	Per capita GDP (PERGDP)	RMB
	Urban land area (ULA)	km^2
	Ratio of urban land area (RULA)	%
	Green land area (GLA)	km^2
	Ratio of green land area (RGLA)	%
	Ratio of primary industry (RoI_1)	%
	Ratio of secondary industry (RoI_2)	%
	Ratio of tertiary industry (RoI_3)	%
Air quality indicators	AQI	/
	$PM_{2.5}$	$\mu g/m^3$
	PM_{10}	$\mu g/m^3$
	SO_2	$\mu g/m^3$
	NO_2	$\mu g/m^3$
	CO	$\mu g/m^3$
	O_3	$\mu g/m^3$

to have little impact on human health, Class III and Class IV will affect the heart and respiratory system of sensitive groups, and Class V and Class VI will generally endanger human health. In the present study, AQI data comes from the data interface of the data center of the Ministry of Ecology and Environment of China (http://datacenter.mee.gov.cn/websjzx/queryIndex.vm).

According to the results of previous studies regarding the relationship between air quality and socio-economic parameters [28, 29], 10 factors are selected as potential controls to be considered on the air quality (shown in Table 1). The socio-economic data comes from the China City Statistical Yearbook of 2019.

Statistical Methods for Global Analysis

Cluster Analysis

Cluster analysis is a multivariate statistical method to classify targets based on proximity or similarity, which can be used to reveal the inherent characteristics between them. Hierarchical clustering is the most commonly used analysis method, which gradually form higher clusters by measuring the similarity between points by Euclidean distance. In this study, based on the average values of AQI and six air pollutants (i.e., $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3) from 2015 to 2019, hierarchical clustering method is used to reveal the differences of air pollution characteristics among cities. The clustering results are represented by dendrogram, which reflects the homogeneity within groups and the differences between groups. Cluster analysis is realized by IBM SPSS Statistics 22.

Pearson Correlation Analysis

Pearson correlation coefficient (r) is a real number between [-1, 1], which is a parameter used to measure the correlation between two variables x and y. The larger the absolute r value between variables, the stronger the correlation. When the r value equal to 0, there is no correlation between the parameters. Generally recognizing, it is considered to have a strong correlation between variables when the r value is higher than 0.6, a moderate correlation when r is between 0.3 and 0.6, and a weak correlation when r is between 0.1 and 0.3. In this study, Pearson correlation coefficient matrix is used to reflect the correlation between AQI and various socio-economic indicators, and Two-tail test is used to identify whether the correlation between variables is significant. Correlation matrix is drawn by Hiplot data visualization analysis online platform (https://hiplot.com.cn).

Principal Component Analysis

Principal component analysis (PCA) is a method to transform multiple original variables into independent variable sets (principal components, PC) containing as much information as possible. It can be used to decrease the influence of collinearity among variables, reduce the dimension of data set and determine the interaction between different variables. In this study, PCA is used to quantify the contribution of different socio-economic indicators to AQI variation, and explore the potential factors to explain the air quality in Northeast China. KMO and Bartlett's sphericity test is employed to determine the applicability of data used for PCA [30]. PCA is performed by IBM SPSS Statistics 22.

Spatial Statistical Models

Spatial interpolation is a method for estimating the features of unknown points through discrete known points and generate continuous surfaces, which originates from the similar related laws in geography. Kriging interpolation is a regression algorithm for spatial modeling and prediction of stochastic processes based on covariance function, which has been widely used in geographic information and meteorological research. Due to the limited number of monitoring stations in this study, the continuous change of pollution index in space cannot be observed directly. The ordinary Kriging interpolation method is used to transform the distribution of AQI in cities into a continuous spatial pattern.

Based on the spatial distribution of socio-economic variables, a spatial regression model is established by using GWR model to reflect the spatial heterogeneity of AQI and its response to the change of driving factors. Compared with the global regression model based on the least square method, GWR model considers the spatial relationship as a weight during the operation, and the relationship between independent and dependent variables changes with the spatial position, which can be used for scenario analysis considering the spatial heterogeneity of multiple variables. The mathematical equation is as follows:

$$y_i = \beta_{i0} + \sum_{k=1}^{m} \beta_{ik} x_{ik} + \varepsilon_i$$

where y represents the predicted AQI values; β_{i0} represents the intercept parameter of the i-th city; m represents the number of independent variables; β_{ik} represents the regression coefficient of the k-th variable; x_{ik} represents the value of the k-th variable in the i-th city and ε represents random error. The spatial change of prediction results is performed by establishing a local model of each variable, and regressing each factor falling within the bandwidth. Optimized bandwidth is determined based on the minimum AICc. Due to the deficiency of many social and economic indicators in the published yearbooks, cities of Yanbian and Great Khingan are not included in GWR model to avoid errors. The spatial visualization of Kriging interpolation model and GWR model is performed by ArcGIS 10.2.1.

Results and Discussion

Temporal Variation and Spatial Pattern of Air Pollutants

Temporal Comparison of AQI between 2015 and 2019

All the detected pollutants show pronounced downward trend in 2019 when comparing with those in 2015. For $PM_{2.5}$, the average values in Liaoning, Jilin and Heilongjiang provinces are 43.61, 40.05 and 38.81, respectively; for PM_{10}, the average values in Liaoning, Jilin and Heilongjiang provinces are 74.97. 68.77 and 55.33, respectively; for SO_2, the average values in Liaoning, Jilin and Heilongjiang provinces are 27.08, 19.23 and 14.39, respectively; for NO_2, the average values in Liaoning, Jilin and Heilongjiang provinces are 29.09, 26.67 and 20.32, respectively; for CO, the average values in Liaoning, Jilin and Heilongjiang provinces are 1.01, 0.93 and 0.68, respectively; for O_3, the average values in Liaoning, Jilin and Heilongjiang provinces are 62.93, 57.70 and 53.20, respectively. All the pollutants show similar pattern within Northeast China, showing that the behaviors of them are similar. Therefore, AQI is used as a comprehensive indicator to represent the general characteristics of pollutants.

The comparison of air quality between 2015 and 2019 is presented by the box-plot of mean AQI values in Fig. 2. In 2015 and 2019, the variation range and average value of AQI among the provinces in Northeast China are significantly different. In 2015, the order of AQI average is Liaoning (85.22) > Jilin (83.03) > Heilongjiang (64.09). Compared with 2015, although the order

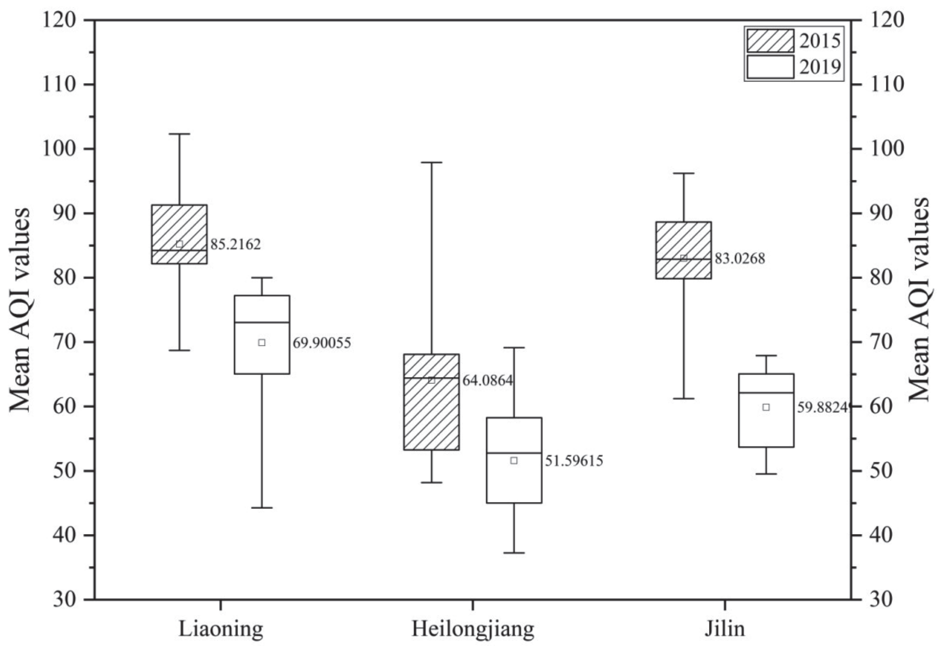

Fig. 2. Box-plot showing the range of AQI in 2015 and 2019. The box presents the 25 % and 75 % percentile, and the whisker presents the maximum and minimum values. The square within the box shows the mean value.

of mean AQI didn't change, that of provinces of Liaoning, Jilin and Heilongjiang decreased to 69.90, 59.88 and 51.60 respectively in 2019, with the largest decline of up to 27.9 % existing in Jilin province. The significant decrease of AQI value reflects the improvement of air quality in Northeast China, and indicates the positive impact of government-led environmental treatment projects and published environmental policies on air quality during the study period.

Spatial Patterns for Air Quality

Cluster analysis is preformed to determine the similarity of air pollutants between cities (Fig. 3). According to the results of cluster analysis, when the relative distance is selected as 20, 37 cities in Northeast China can be divided into 4 clusters, showing significant differences in the degree of air quality degree and the characteristics of pollutants in Northeast China. Cluster 1 only includes Benxi City located in the east of Liaoning Province, reflecting that there are significant differences in pollutant characteristics between this city and others. According to the mean concentrations of the 6 pollutants in cities of Northeast China (shown in Table 2). Among the 6 types of pollutants, the average CO concentration of Benxi City from 2015 to 2019 (1.51 μg/m^3) is much higher than that of all cities in the study area (0.88 μg/m^3), which may be related to the local pillar of iron and steel industry and the huge coking coal and steel output. Previous studies have shown that industrial processes related to iron and steel production are important factors related to regional CO concentration, and there is a strong correlation between the control of CO emissions and the improvement of combustion efficiency of iron and steel enterprises [31]. Cluster 2 contains three cities of Shenyang, Jinzhou and Huludao, representing the areas with most serious air pollution in Northeast China. All 6 pollutants in these 3 cities show significantly higher concentrations than the regional average. All three cities are located in Liaoning Province, adjacent to Beijing-Tianjin-Hebei Economic Circle, with high level of regional economic development, developed industry, dense population and high pollutant emissions. In addition, this region is connected with the North China Plain, where air pollution is the most serious across the country, the migration of pollutants may also be one of the reasons for the decline of air quality. Cluster 3 contains three cities: Heihe, Daxinganling and Yichun. These cities are all located in the mountainous areas in the north of Heilongjiang Province, which belongs to the highest-latitude area in China. Due to the low population density and the high proportion of rural and woodland area, the ecological environment in this area is good, and the air quality is slightly disturbed by human activities. The average AQI of all the three cities is below 50, which is the lowest in Northeast China and belongs to Class I, which also reflects the very low degree of pollution. Cluster 4 contains the remaining 30 cities, showing moderate air pollution.

The distribution patterns of AQI in Northeast China in 2015 and 2019 are spatially visualized by ordinary Kriging interpolation model (shown in Fig. 4 a, b). In 2015 and 2019, the spatial variation of AQI showed a decreasing trend from southwest to northeast. The central part of Liaoning Province is the region with the highest AQI, showing the most serious air pollution. This is because this region is characterized by densely distributed heavy industry enterprises, and is close to the North China Plain where air pollution is the most serious in China, and thus is easily affected by the migration of pollutants. In the eastern part of Jilin Province and the northern part of Heilongjiang Province, the AQI shows lower concentration, indicating that the air quality in these areas is better. This is likely to be related to the sparse population and agriculture as the main industry. Compared with 2015, the high value of AQI decreases in 2019, and the area of high pollution zone decreases, especially in the south of Heilongjiang and the middle of Jilin, which reflected the effectiveness of air pollution control and the promotion of energy saving and emission reduction measures in this area. The distribution of percentage of Class V & VI (representing severe polluted days) and Class I & II (representing good-air quality days) is shown in Fig. 4(c, d), respectively. The distribution of the percentage of Class V & VI is generally consistent with that of AQI, and that of Class I & II shows the opposite trend to AQI distribution, reflecting the positive correlation between pollution duration and pollutant concentration. Although Harbin does not show the highest concentration of AQI, the proportion of its severe pollution period is the highest. This may be due to the hilly and mountainous terrain around it, which is not conducive to the diffusion of pollutants, resulting in a long time of being affected by seriously polluted air. The coastal areas in eastern and southern Liaoning is characterized by developed industries and dense population, but its AQI is lower among the cities in this province and the proportion of Class I & II is higher, which is likely because of the air circulation between land and sea accelerating the discharge of pollutants.

Socio-Economic Controls on AQI Identified by Statistical Analysis

Pearson Correlation Matrix, which shows the relationship between the daily average AQI value and the selected socio-economic metrics in northeast China, is shown in Fig. 5. AQI is significantly correlated with population, the proportion of urban land and the contribution of primary, secondary and tertiary industries to GDP at a confidence level of 0.05. The strongest positive correlation (r = 0.55) appears between AQI and urban land area, indicating that urban expansion is an important reason for the increase of pollutant concentration and the deterioration of air

Table 2. Daily mean values of AQI and concentrations of 6 air pollutants (in μg/m³) during the period of 2015~2019.

Province	City	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	CO	O_3
Liaoning	Anshan	79.42	38.76	66.88	20.58	25.37	0.88	58.47
	Benxi	72.31	44.10	78.25	30.69	33.50	1.51	50.64
	Chaoyang	71.62	39.11	73.15	27.37	21.31	1.32	61.16
	Dalian	67.17	36.71	65.10	18.48	28.10	0.83	77.07
	Dandong	62.32	36.86	62.74	23.46	23.00	1.10	57.27
	Fushun	77.84	45.69	81.20	23.80	31.86	1.03	59.90
	Huludao	81.30	47.08	84.19	40.17	33.78	1.26	67.71
	Fuxin	75.16	41.09	81.49	34.01	25.04	0.93	63.45
	Liaoyang	79.40	46.49	82.51	24.07	29.25	1.20	61.60
	Panjin	73.35	40.89	65.52	21.92	27.05	0.93	71.63
	Shenyang	82.67	50.90	88.17	36.52	38.64	0.96	58.79
	Wafangdian	71.09	45.23	58.28	24.58	25.49	0.59	51.70
	Tieling	79.93	47.00	83.31	18.87	29.81	0.76	59.08
	Yingkou	77.68	43.22	71.14	17.56	28.84	0.85	79.02
	jinzhou	84.05	51.08	82.63	44.08	35.34	1.05	66.49
Heilongjiang	Harbin	79.11	50.86	78.57	24.67	39.79	0.96	49.59
	Hegang	55.89	30.71	57.05	8.33	13.65	0.74	58.74
	Heihe	43.91	21.36	40.05	16.70	13.73	0.53	54.32
	Jixi	53.95	29.67	56.18	12.82	18.64	0.79	48.91
	Mudanjiang	62.43	36.15	66.20	12.09	24.16	0.66	48.71
	Qitaihe	65.99	39.12	67.03	12.98	23.42	0.66	57.23
	Qiqihar	59.98	32.73	59.71	20.06	20.45	0.71	53.41
	Shuangyashan	56.11	33.62	55.99	12.02	17.34	0.72	53.52
	Suihua	60.18	35.22	59.43	15.21	21.19	0.62	53.65
	Yichun	40.10	20.02	35.79	7.70	12.57	0.47	47.00
	Daqing	58.99	34.17	56.29	12.90	23.77	0.64	61.59
	Greater Khingan Range	43.19	19.41	38.23	23.04	15.82	0.69	48.71
	Jiamusi	53.15	30.46	48.74	8.55	19.69	0.68	56.25
Jilin	Jilin	72.89	43.19	71.83	17.93	26.61	0.85	68.42
	Liaoyuan	70.17	43.43	60.57	18.79	26.23	1.03	60.06
	Siping	77.12	45.58	81.75	19.33	30.48	0.91	60.35
	Songyuan	69.10	34.90	72.80	12.17	20.18	0.82	60.63
	Tonghua	64.01	36.55	66.21	23.78	29.44	1.13	47.45
	Yanbian	54.64	29.73	47.38	12.51	20.69	0.75	55.86
	Changchun	75.98	45.20	78.09	21.55	36.20	0.89	57.53
	Baicheng	69.04	40.52	69.46	22.54	26.37	0.91	57.91
	Baishan	68.61	41.36	70.81	24.44	23.84	1.13	51.11
Total mean		67.29	38.60	66.56	20.71	25.42	0.88	58.24

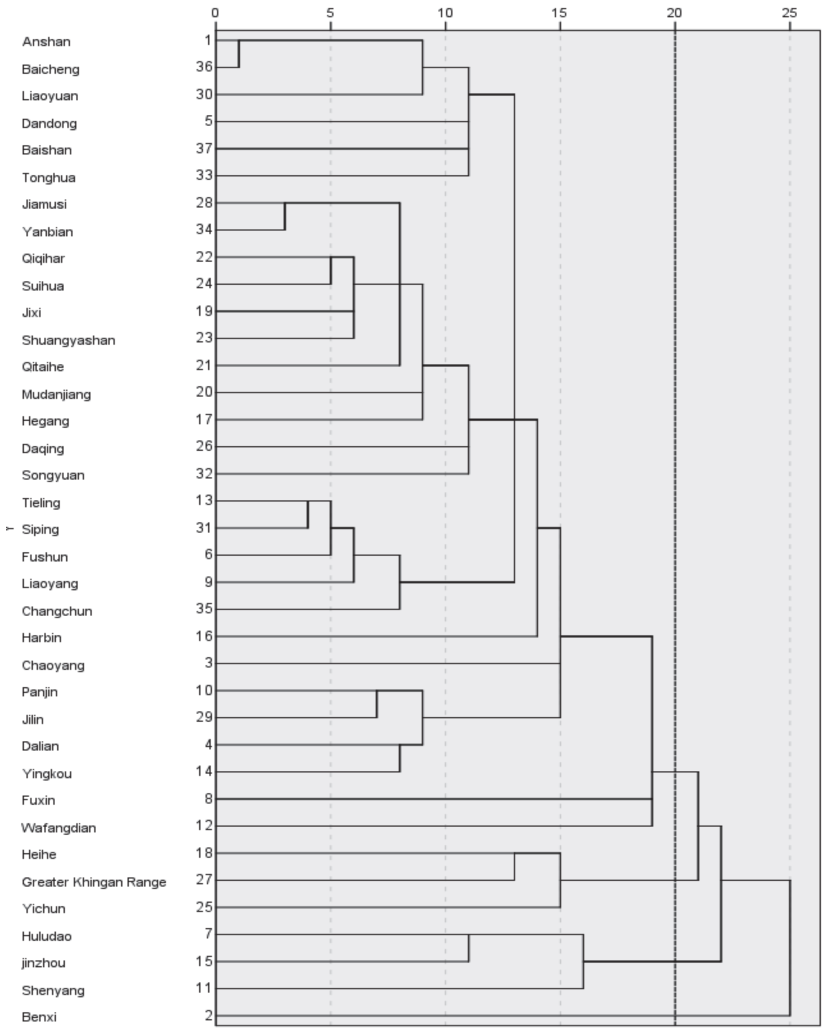

Fig. 3. Dendrogram exhibiting the similarities of air pollution metrics among the 37 cities in Northeast China. X-axis means the rescale distance cluster combine.

quality. AQI shows the strongest negative correlation with the economic contribution of the primary industry, indicating that agriculture will not cause the deterioration of air quality (or has a slight impact). Moderate positive correlation is also observed between AQI and population (r = 0.38), economic proportion of secondary industry (r = 0.44) and tertiary industry (r = 0.44), suggesting that air pollution shows an increasing trend in densely populated, industrialized and urbanized cities. There is a weak and insignificant positive correlation between AQI and GDP as well as per capita GDP, reflecting that air pollution mainly depends on industrial structure rather than economic level. This also explains that in Northeast China, most resource-based cities are facing the dual pressures of economic recession and environmental deterioration.

Interestingly, AQI shows a weak and insignificant positive correlation with urban green land area (r = 0.22) and green space ratio (r = 0.01), which is counter-intuitive. Previous studies showed that urban green land has marked blocking, filtering and adsorption effects on smoke and dust, and is able to remove a large number of air pollutants [32, 33]. In Northeast China, the area of urban green land is positively correlated with population, GDP, per capita GDP, urban area and the economic contribution of secondary industry, indicating that the area of urban green space is larger in areas with concentrated population, large urban size and high degree of economic development. Our result shows that the positive effect of urban green land on reducing air pollution may be covered up by other factors that promote pollutant emission, especially in developed large cities.

PCA can be used for variables that do not contribute significantly in the original variable data set, thus simplifying the data structure. To better understand the control of different socio-economic characteristics on air quality, PCA is employed to determine the correlation between socio-economic indicators. The KMO of selected socio-economic indicators is 0.615, and the P value of Bartlett's sphere test is less than 0.001, indicating that the data set is suitable for

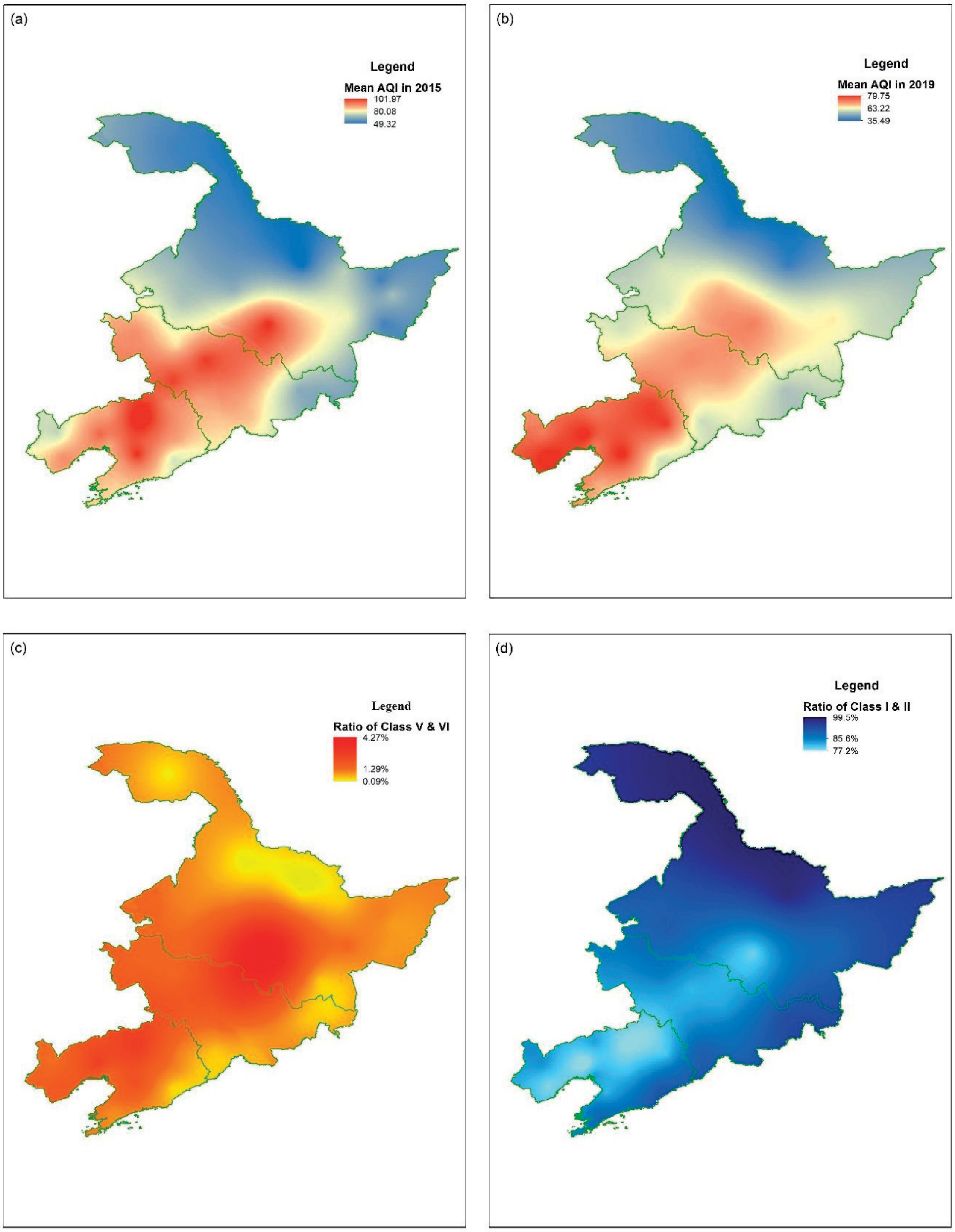

Fig. 4. Spatial pattern of mean AQI values in (a) 2015 and (b) 2019, and the distribution of percentage of (c) Class V & VI and (d) Class I and II performed by Kriging Spatial Interpolation.

performing PCA. Based on eigenvalue (>1), three principal components were selected, accounting for 80.56% of the total variance. The eigenvalue of PC1 is 4.900, which explains 49.00% of the total variance. The eigenvalue of PC2 is 1.895, which explains 18.95% of the total variance. The eigenvalue of PC3 is 1.261, which explains 12.61% of the total variance.

According to the empirical values reported in previous studies, the absolute load values of variables of >0.75, 0.75-0.50 and 0.49-0.30 can be classified as "strong", "medium" and weak correlation in PCA, respectively [34]. Our results show that PC1 is strongly correlated with GDP, per capita GDP, urban land area and green land area, and moderately correlated with population and economic contribution of primary, secondary and tertiary industries, reflecting the city scale. PC2 is strongly related to the economic contribution of the secondary industry and moderately

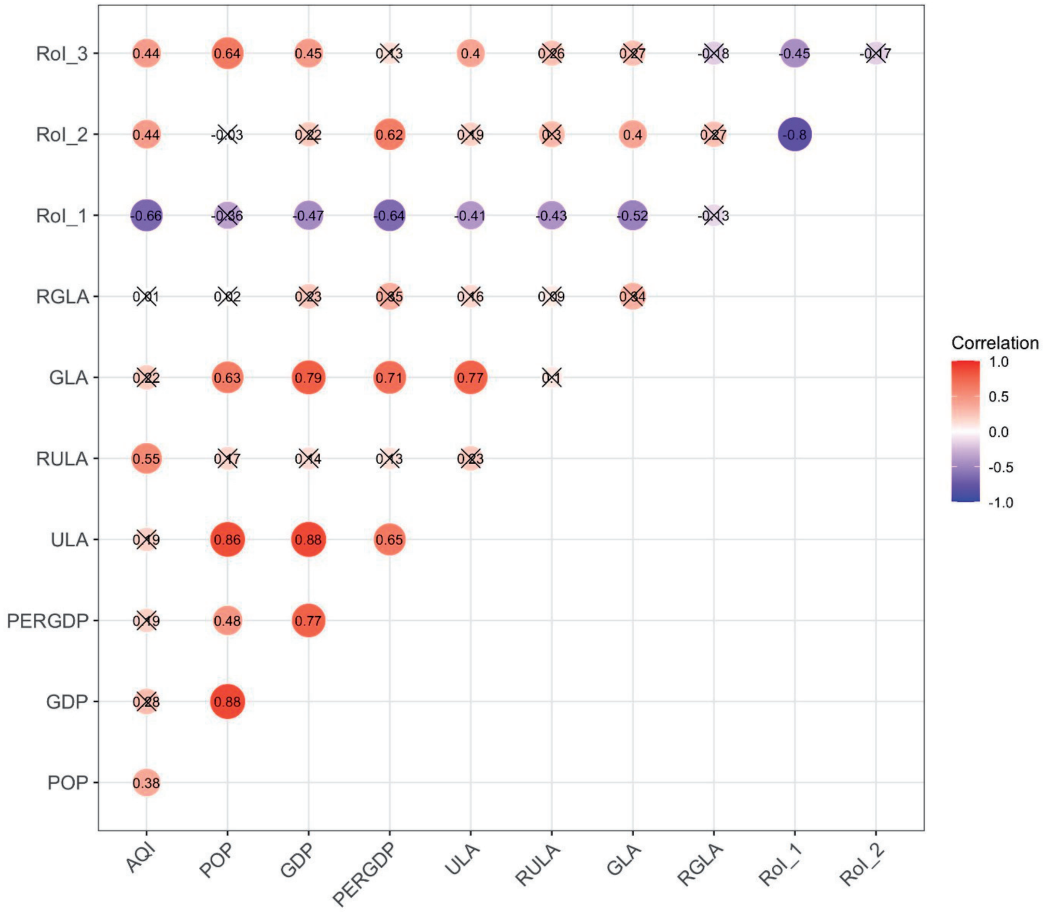

Fig. 5. Pearson's Correlation matrix presents the relationship between AQI and socio-economic parameters.

related to the population, which represents the degree of industrialization. PC3 is moderately related to the economic contribution of the tertiary industry and the proportion of urban area, which represent the level of economic development. Combining the result of PCA and the correlation between variables shown in the Pearson correlation matrix, it can be shown that the air quality is dominantly controlled by the degree of industrialization, followed by the level of economic development and city scale.

Analysis of the Spatial Variability of Main Drivers based on GWR Model

GWR model can be used as an instrument to evaluate the response of AQI to the changes of various variables. According to the results of Pearson correlation matrix and PCA, population, ratio of urban land area, GDP, ratio of urban green land area and economic contribution of secondary industry are selected as explanatory variables of the spatial distribution of AQI. The result of GWR model is shown in Fig. 6. By comparing the predicted AQI with the measured results (Fig. 6a, b), it can be seen that the spatial characteristics of the observed AQI can be mostly reproduced by using GWR model. In the west of Liaoning Province, the predicted AQI is slightly underestimated, while in the west of Heilongjiang, it is overestimated. The local R^2 value ranges from 0.48 to 0.67, and a decreasing trend from southwest to northeast can be observed. The highest R2 value appears in the eastern part of Heilongjiang, showing that the variables selected can better explain the AQI variation in this area. However, the poor performance of the model is observed in the southwest of Liaoning Province (R^2<0.5), indicating that the variation of AQI is more complicated and driven by other natural or anthropogenic factors in areas with developed economy, concentrated population and high urbanization level.

A positive correlation between population and AQI can be observed across the whole region, and the regression coefficient shows little difference among different cities (Fig. 6d), indicating that population aggregation promotes the deterioration of air quality, and the impact of population growth on air quality deterioration is relatively stable in space. The regression coefficient between the proportion of urban area and AQI increases from south to north (Fig. 6e), which is consistent with the spatial trend of urban scale and economic level, indicating that urbanization imposes a more significant impact on air quality in small and medium-sized cities. The regression coefficient between AQI and GDP is low (Fig. 6f), with slight spatial variation, indicating that GDP is not the main variable driving AQI change. Except a small area in the west of Liaoning province, there is a positive correlation between the ratio of green land and AQI (Fig. 6g), suggesting that the positive effect of urban greening on air quality is easily concealed by other variables related to industrialization and economic development level. This is consistent with the results of PCA in the previous chapter. A positive correlation is observed between AQI and the economic contribution of the secondary industry, and the regression coefficient presents an increasing trend from southwest to northeast, indicating that energy-consuming industries relying on resource exploitation serve as an important source leading to the deterioration of atmospheric environment in Northeast China. The pollutants emitted during the process of rapid industrial development are the main causes of atmospheric pollutants, and also represent the pressure brought by industrialization on the improvement of atmospheric environment. In the agglomeration areas of small and medium-sized cities in the east of Heilongjiang Province, the population density and economic development level are low with high economic contribution of agriculture, and the air pollution is relatively slight. However, the predicted AQI based on socio-economic metrics in these areas shows the largest local R^2 value, and high regression coefficient is observed between AQI and population, GDP, ratio of green land and economic contribution of secondary industry. This suggests that in small and medium-sized cities with underdeveloped economy and low urbanization, air quality is more sensitive to economic development, urban expansion and improvement of industrial level.

To summarize, the air quality in Northeast China is greatly dependent on socio-economic factors. The aggregation of population leads to more crowded traffic and heavier emission of atmospheric pollutants, which is thus negatively related to air quality. GDP is not a dominant driving factor on air quality in Northeast China, which may be because the developing patterns vary between cities. Whereas, the economic structure (i.e., proportion of primary, secondary and tertiary industry) is a very important control on air quality. In city areas, the level of pollution is mainly controlled by the pillar industries. Areas with large economic contribution from heavy industries, such as petrochemical industry, thermal power production and steel industry, is characterized by dramatic pollutant emission and thus much prone to be exposed to hazardous pollutants. Although previous studies showed that the increased area of green land has positive influence on air quality improvement [17, 24], such effect is very insignificant as the emitted atmospheric pollutants far outweighs the absorption capacity of vegetation in Northeast China. Therefore, the adjustment of industrial structure can be regarded as the most effective way to optimize the air quality in Northeast China.

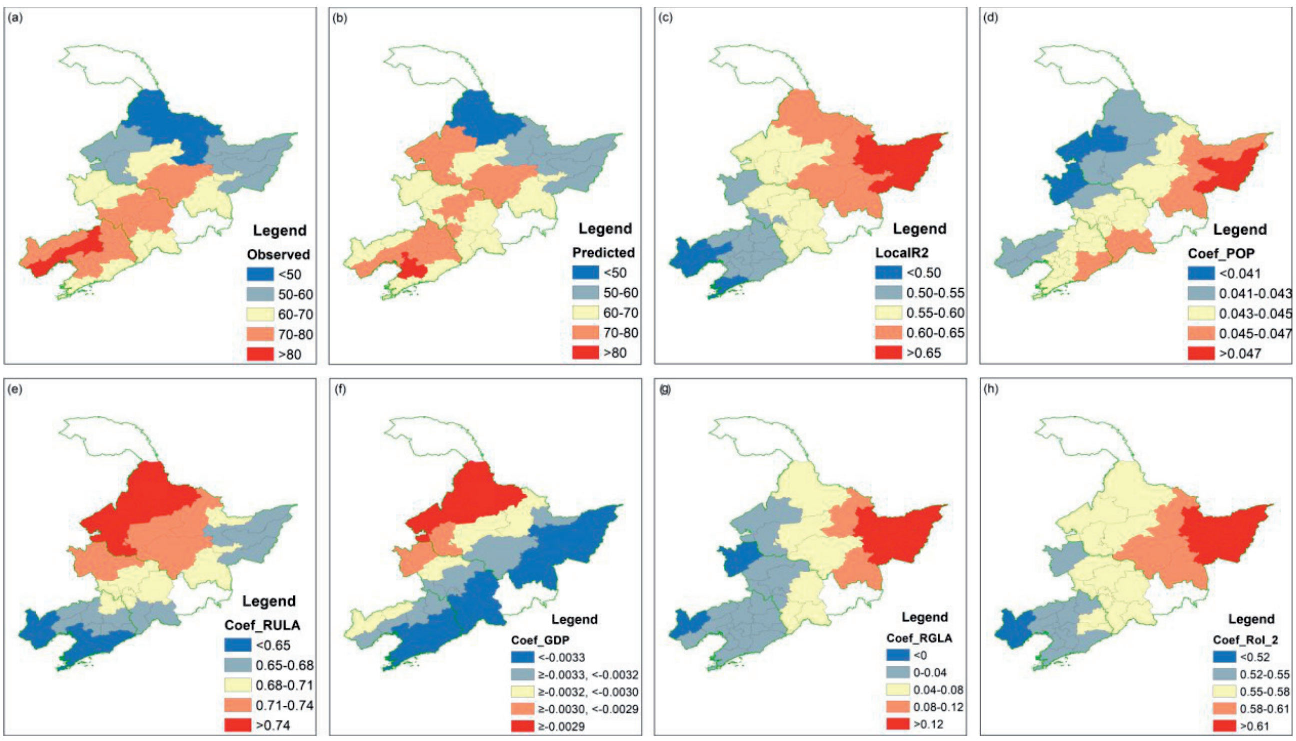

Fig. 6. The comparison between (a) observed and (b) GWR-predicted AQI patterns, (c) the local coefficients of determination R2, and (d-h) the spatial distribution of coefficient of explanatory variables on AQI.

Conclusions

Based on the records of air pollutants and socio-economic metrics of cities in Northeast China from, this paper analyzes the temporal changes and spatial patterns of air quality by combining mathematical statistics with geospatial models, and identifies the driving mechanisms of different types of variables on the evolution of air quality. Compared with 2015, the average and extreme values of AQI in all provinces decreased significantly in 2019, showing the reduction of air pollution. The cities in Northeast China are divided into four categories according to the mean value of 6 pollutant types, including CO enrichment group, seriously polluted group, medium polluted group and good quality group. Air pollution shows a decreasing spatial trend from southwest to northeast direction, and the duration of severe pollution depends not only on the concentration of pollutants but also on topography. The proportion of urban area, population and the economic contribution of the primary, secondary and tertiary industries are significantly related to AQI. The air pollution in Northeast China is mainly controlled by the degree of industrialization, followed by the level of economic development and city scale. On the whole, GWR model can reflect the spatial characteristics of AQI and the driving factors on air quality variation, but its prediction of air quality is poor in the southwest of Liaoning Province, reflecting that the driving factors in this region are more complex. The impacts of different socio-economic indicators on AQI show spatial differentiation. In areas with small city scale, sparsely population and underdeveloped economy, air quality is more sensitive to the changes of socio-economic metrics. This study indicates that the multivariate statistical method can provide information for the comprehensive evaluation of air quality, and provide insights for environmental protection, industrial structure adjustment and sustainable economic development with adaption to local conditions.

Acknowledgments

The authors would like to express their sincere thanks to the State Statistics Bureau and Ministry of Ecology and Environment of China for the published open access data.

Conflict of Interest

The authors declare no conflict of interest.

References

1. BAO J.Z., YANG X.P., ZHAO Z.Y., WANG Z.K., YU C.H., LI X.D. The Spatial-Temporal Characteristics of Air Pollution in China from 2001-2014. Int. J. Env. Res. Pub. He., **12**, 12, **2015**.
2. MAO M., SUN H.F., ZHANG X.L. Air Pollution Characteristics and Health Risks in the Yangtze River Economic Belt, China during Winter. Int. J. Env. Res. Pub. He., **17**, 24, **2020**.
3. SONG H., ZHUO H.M., FU S.Z., REN .LJ. Air pollution characteristics, health risks, and source analysis in Shanxi Province, China. Environ. Geochem. Hlth., **43**, 1, **2021**.
4. SONG R., YANG L.M., LIU M.Y., LI C., YANG Y.R. Spatiotemporal Distribution of Air Pollution Characteristics in Jiangsu Province, China. Adv. Meteorol., **2019**, **2019**.
5. GURJAR B.R., JAIN A., SHARMA A., AGARWAL A., GUPTA P., NAGPURE A.S., et al. Human health risks in megacities due to air pollution. Atmos. Environ., **44**, 36, **2010**.
6. ALMETWALLY A.A., BIN-JUMAH M., ALLAM A.A. Ambient air pollution and its influence on human health and welfare: an overview. Environmental Science and Pollution Research, **27**, 20, **2020**.
7. FOWLER D., PYLE J.A., SUTTON M.A., WILLIAMS M.L. Global Air Quality, past present and future: an introduction. Philos. T. R. Soc. A., **378**, 2183, **2020**.
8. SHADDICK G., SALTER J.M., PEUCH V.H., RUGGERI G., THOMAS M.L., MUDU P., et al. Global Air Quality: An Inter-Disciplinary Approach to Exposure Assessment for Burden of Disease Analyses. Atmosphere, **12**, 1, **2021**.
9. CHAN C.K., YAO X. Air pollution in mega cities in China. Atmos. Environ., **42**, 1, **2008**.
10. WANG Y.S., YAO L., WANG L.L., LIU Z.R., JI D.S., TANG G.Q., et al. Mechanism for the formation of the January 2013 heavy haze pollution episode over central and eastern China. Sci. China Earth. Sci., **57**, 1, **2014**.
11. WEI F.L., LI S., LIANG Z., HUANG A.Q., WANG Z., SHEN J.S., et al. Analysis of Spatial Heterogeneity and the Scale of the Impact of Changes in $PM_{2.5}$ Concentrations in Major Chinese Cities between 2005 and 2015. Energies, **14**, 11, **2021**.
12. XU Y., YING Q., HU J.L., GAO Y., YANG Y., WANG D.X., et al. Spatial and temporal variations in criteria air pollutants in three typical terrain regions in Shaanxi, China, during 2015. Air. Qual. Atmos. Hlth., **11**, 1, **2018**.
13. CETIN M. The effect of urban planning on urban formations determining bioclimatic comfort area's effect using satellitia imagines on air quality: a case study of Bursa city. Air. Qual. Atmos. Hlth., **12**, 10, **2019**.
14. CETIN M. Climate comfort depending on different altitudes and land use in the urban areas in Kahramanmaras City. Air. Qual. Atmos. Hlth., **13**, 8, **2020**.
15. CETIN M. Using GIS analysis to assess urban green space in terms of accessibility: case study in Kutahya. Int. J. Sust. Dev. World, **22**, 5, **2015**.
16. CETIN M. Evaluation of the sustainable tourism potential of a protected area for landscape planning: a case study of the ancient city of Pompeipolis in Kastamonu. Int. J. Sust Dev. World, **22**, 6, **2015**.
17. CETIN M., ONAC A.K., SEVIK H., SEN B. Temporal and regional change of some air pollution parameters in Bursa. Air Qual. Atmos. Hlth., **12**, 3, **2019**.
18. CETIN M., SEVIK H., SAAT A. Indoor Air Quality: The Samples of Safranbolu Bulak Mencilis Cave. Fresen Environ. Bull., **26**, 10, **2017**.
19. FANG C.L., WANG Z.B., XU G. Spatial-temporal characteristics of $PM_{2.5}$ in China: A city-level perspective analysis. Journal of Geographical Sciences, **26**, 11, **2016**.

20. CETIN M. SEVIK H. Change of Air Quality in Kastamonu City in Terms of Particulate Matter and CO_2 Amount. Oxid. Commun., **39**, 4, **2016**.
21. BEN SAAD M. The Effect of Economic Complexity on Air Pollution: Another approach to the Environmental Curve of Kuznets. Rev. Dev., 46, **2017**.
22. TSUJIMOTO M. Economic Growth and Air Pollution in the Persian (Arabian) Gulf States: Environmental Kuznets Curve Hypothesis. J .Jpn. I. Energy, **97**, 1, **2018**.
23. BUEHN A., FARZANEGAN M.R. Hold your breath: A new index of air pollution. Energ. Econ., **37**, **2013**.
24. PATTON A.P., PERKINS J., ZAMORE W., LEVY J.I., BRUGGE D., DURANT J.L. Spatial and temporal differences in traffic-related air pollution in three urban neighborhoods near an interstate highway. Atmos. Environ., **99**, **2014**.
25. ZHAO J.J., CHEN S.B., WANG H., REN Y., DU K., XU W.H., et al. Quantifying the impacts of socio-economic factors on air quality in Chinese cities from 2000 to 2009. Environ. Pollut., **167**, **2012**.
26. CHEN J.D., WANG B., HUANG S., SONG M.L. The influence of increased population density in China on air pollution. Science of the Total Environment, **735**, **2020**.
27. MA Y.J., ZHAO H.J., DONG Y.S., CHE H.Z., LI X.X., HONG Y., et al. Comparison of Two Air Pollution Episodes over Northeast China in Winter 2016/17 Using Ground-Based Lidar. J Meteorol Res-Prc, **32**, 2, **2018**.
28. LIU H.M., FANG C.L., ZHANG X.L., WANG Z.Y., BAO C., LI F.Z. The effect of natural and anthropogenic factors on haze pollution in Chinese cities: A spatial econometrics approach. J. Clean. Prod., **165**, **2017**.
29. LIANG L.W., WANG Z.B. Control Models and Spatiotemporal Characteristics of Air Pollution in the Rapidly Developing Urban Agglomerations. Int. J. Env. Res. Pub. He., **18**, 11, **2021**.
30. HELENA B., PARDO R., VEGA M., BARRADO E., FERNANDEZ J.M., FERNANDEZ L. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. Water. Res., **34**, 3, **2000**.
31. ZHENG B., CHEVALLIER F., CIAIS P., YIN Y., DEETER M.N., WORDEN H.M., et al. Rapid decline in carbon monoxide emissions and export from East Asia between years 2005 and 2016. Environ. Res. Lett., **13**, 4, **2018**.
32. YANG L.K., XING Y.G., JONES P. Exploring the potential for air pollution mitigation by urban green infrastructure for high density urban environment. Iop. C. Ser. Earth. Env., **227**, **2019**.
33. YILMAZ S., SEZEN I., SARI E.N. The relationships between ecological urbanization, green areas, and air pollution in Erzurum/Turkey. Environ. Ecol. Stat., **2021**.
34. BU H.M., TAN X., LI S.Y., ZHANG Q.F. Water quality assessment of the Jinshui River (China) using multivariate statistical techniques. Environmental Earth Sciences, **60**, 8, **2010**.