

Article

Not peer-reviewed version

---

# Spatial Distribution Characteristics of Urban Air Quality and the Spatial Heterogeneity of Driving Factors: A Case Study of Beijing

---

[Zhixiong Tan](#) , Haili Wu , Qingyang Chen , [Jiejun Huang](#) \*

Posted Date: 11 June 2024

doi: 10.20944/preprints202406.0694.v1

Keywords: air quality index; multi-scale geographically weighted regression; multi-source geographic data; spatial autocorrelation analysis; spatial heterogeneity



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# Spatial Distribution Characteristics of Urban Air Quality and the Spatial Heterogeneity of Driving Factors: A Case Study of Beijing

Zhixiong Tan, Haili Wu, Qingyang Chen and Jiejun Huang \*

School of Resources and Environmental Engineering, Wuhan University of Technology Wuhan, 430070, China; ZhixiongTan@whut.edu.cn

\* Correspondence: hjj@whut.edu.cn

**Abstract:** Urban air pollution is a critical global environmental issue, necessitating an understanding of the spatiotemporal characteristics and driving factors for effective environmental protection and urban planning. This study introduced an advanced air quality assessment system that integrates environmental, socio-economic, and urban layout factors, addressing gaps in traditional models that often overlook the impact of urban spatial structures. Analyzing air quality data from Beijing's main urban area (2016-2020) alongside multi-source geographic data, the research develops a comprehensive evaluation system, incorporating 14 key driving factors. Employing Geographically Weighted Regression (GWR) and Multi-scale Geographically Weighted Regression (MGWR), the study quantitatively assessed the influence and spatial heterogeneity of these factors. Findings revealed an annual improvement in air quality, with a U-shaped seasonal pattern and significant spatial clustering (Global Moran's  $I = 0.922$ ). The MGWR model, in particular, provided a superior fit over GWR, effectively capturing the spatial variability of factors. Variables such as NDVI, economic output (GDP), and humidity space adjustment capability (HSAC) showed significant positive spatial impacts on air quality, while population density (POP), temperature (TEMP), and road density (RD) exhibited negative effects. These insights enhance the understanding of air pollution dynamics and aid in refining urban planning strategies.

**Keywords:** air quality index; multi-scale geographically weighted regression; multi-source geographic data; spatial autocorrelation analysis; spatial heterogeneity

## 1. Introduction

Cities, as areas of high population and activity concentration, are not only the political, economic, and cultural centers of regions but also pivotal for national sustainable development[1]. However, with rapid economic growth and accelerated urbanization, an increasing number of cities in China are facing severe environmental pollution challenges, such as rising greenhouse gas emissions, intensifying urban heat island effects, frequent extreme high-temperature events, and the degradation of urban ecosystems[2–5]. Among these, urban air pollution in China is particularly severe and has become a globally recognized environmental issue[6,7]. The exceeding concentrations of pollutants such as particulate matter, nitrogen dioxide, and ozone pose major challenges to improving urban air quality, significantly threatening the health of urban residents and sustainable development[8]. Concurrently, as public concern over the travel environment grows, there is an increasing desire to understand the spatiotemporal distribution patterns and causative factors of air quality in cities[9,10]. Therefore, monitoring the periodic changes in urban air quality and analyzing the causes of air pollution formation are crucial for protecting the living environment of urban residents and managing air pollution.

The monitoring of urban air pollution has incorporated the Air Quality Index (AQI) as a numerical indicator of air quality status and its potential health impacts[11]. AQI serves as a straightforward method for the public to understand the level of air pollution, converting concentrations of various pollutants in the air into a standardized index value[12]. Guided by the

World Health Organization (WHO), the AQI is designed to provide a quantifiable method to indicate current air quality and potential health impacts[13]. The higher the AQI value, the more severe the air pollution and the greater the potential health risks[14–16]. The classification standards for the AQI are shown in Table A1. AQI Classification Standards.. Studying the spatiotemporal characteristics and driving factors of AQI is crucial for assessing urban air quality and mitigating air pollution[17–21]. Numerous studies have quantitatively analyzed the spatiotemporal distribution and trends of AQI in China from different regional perspectives. At the national scale, researchers analyzing AQI changes across the country have identified regions with consistently high AQI primarily in North China, the Yangtze River Delta, and the Pearl River Delta[5,22,23]. At the urban scale, researchers have focused on areas or urban agglomerations with advanced urbanization and industrialization processes, identifying spatiotemporal patterns and influencing factors of air quality indices[24–27]. Recent studies, targeting smaller spatial scales, such as specific cities or their main urban areas, employ higher spatiotemporal resolution to more finely explore the spatial distribution patterns and heterogeneity of air quality within urban areas[28,29].

In urban areas, researchers have leveraged the high spatiotemporal resolution advantages of satellite remote sensing data to conduct many studies on the factors affecting air quality[30,31]. Initially, meteorological factors such as temperature, precipitation, relative humidity, and wind speed were introduced to explain the variations in China's Air Quality Index and air pollutants[31–33]. Currently, the most classic indicator system consists of factors from both natural environments and socio-economic indicators. From a natural perspective, vegetation indices such as the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI) are used to reflect the impact of the natural environment on air quality[34]. These products are positively correlated with improvements in air quality and health status due to vegetation's ability to absorb and block air pollutants[35,36]. Socio-economic indicators include road network density, Normalized Difference Built-up Index (NDBI) [37,38], nighttime light (NTL), population density, per capita GDP, and other factors related to resident activities and energy consumption. These studies effectively mine the specific sources of regional air pollution from both natural and socio-economic perspectives, but they also have limitations, such as neglecting differences in urban internal layouts and not recognizing the driving mechanisms of urban form, industrial structure, and other urban layout attributes on the urban air environment[37–39]. Therefore, it is crucial to select differentiated urban layout indicators on top of natural and socio-economic impacts to deepen the research on factors influencing urban air quality.

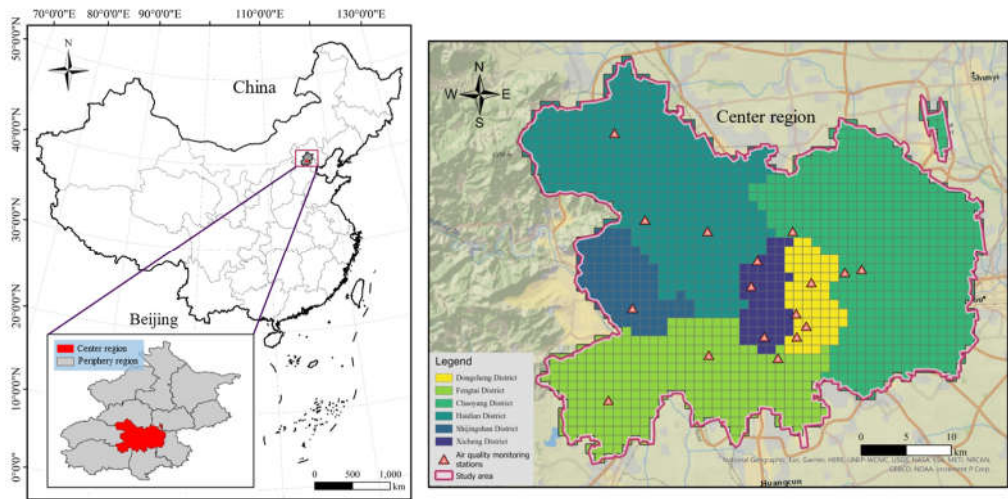
To unveil the driving mechanisms behind urban air quality influencing factors, it is necessary to select appropriate spatiotemporal regression models for modeling and analyzing the drivers of air pollution. Early research on air quality time series utilized linear and nonlinear models to predict urban air quality, comparing the abilities of different models to forecast pollutant concentrations using meteorological factors and air quality monitoring data[40]. Currently, many scholars have explored spatial models based on geographic data, such as Inverse Distance Weighting (IDW) based on spatial autocorrelation and Land Use Regression models (LUR)[41]. Global and local Moran's I can systematically measure the spatial autocorrelation of Air Quality Index (AQI) values at the urban level in China, while Ordinary Least Squares (OLS), Spatial Autoregression (SAR), and Geographically Weighted Regression (GWR) quantitatively estimate the comprehensive impacts and spatial variations of urbanization on air quality[5]. The Geographically and Temporally Weighted Regression (GTWR) model can explain discontinuous AQI values and capture spatiotemporal non-stationarities that GWR does not consider[28]. Most models are limited by the collinearity among variables, which restricts their explanatory power and robustness, and fail to effectively identify scale-varying dominant influencing factors. The Multi-scale Geographically Weighted Regression (MGWR) model considers the multi-scale spatial effects between air quality and various variables, providing deeper insights into the spatial distribution of air quality and the spatial mechanisms of influencing factors. This model is particularly suited for exploring complex environmental issues and formulating precise environmental management strategies.

This paper addresses the gap in existing research that lacks an understanding of the driving factors of urban air pollution from the perspective of urban spatial layout. By integrating time series data from air quality monitoring stations with multi-source geographic information data, a comprehensive indicator system for urban air pollution research is proposed. Utilizing the Multi-scale Geographically Weighted Regression (MGWR) model, the study investigates the spatiotemporal characteristics of air quality in the central region of Beijing and the spatial heterogeneity and intensity of its multiple influencing factors.

2. Study Area and Data

2.1. Study Srea

The study selected the central urban districts of Beijing, including Dongcheng, Xicheng, Chaoyang, Haidian, Shijingshan, and Fengtai districts. This area is the most densely populated and economically active region in Beijing, occupying less than 10% of the total area of Beijing but housing more than half of the city's permanent population. Due to its unique geographical location and high level of urbanization, it has become an ideal area for studying the spatiotemporal heterogeneity of urban air pollution. In the study, a 1 km × 1 km grid sampling was conducted across these six districts, ultimately obtaining 1382 samples (Figure 1), providing foundational data for subsequent analysis.



**Figure 1.** Geographical location of the study area and distribution of monitoring sites. (2019 China Map-Map Approval No. GS (2019) 1822).

2.2. Data Sources and Pre-Processing

The data used in the study primarily consists of station monitoring data, remote sensing product data, and open-access geographic big data (Table 1). Below is an introduction to the data and its preprocessing steps.

**Table 1.** Data Sources and description.

Type	Name	Resolution		Source
		Temporal	Spatial	
Raster	minimum temperature	2016-2020 monthly	2.5 minutes (~21 km2 at the equator)	WorldClim ( <a href="https://www.worldclim.org/">https://www.worldclim.org/</a> )
	maximum temperature			
	precipitation			
	NDVI	2016-2020 yearly	250 m	The Land Processes Distributed Active Archive Center (LP DAAC) ( <a href="https://lpdaac.usgs.gov/">https://lpdaac.usgs.gov/</a> )



	population		100 m	WorldPop ( <a href="https://www.worldpop.org/">https://www.worldpop.org/</a> )
	RDLS	2014	1 km	Global Change Research Data Publishing & Repository ( <a href="https://www.geodoi.ac.cn/">https://www.geodoi.ac.cn/</a> )
	Luojia 1-01 NPP-DNB product	2018	130 m	Luojia-1 satellite official website ( <a href="http://59.175.109.173:8888/index.html">http://59.175.109.173:8888/index.html</a> )
	OSM	2020	\	OpenStreetMap ( <a href="https://www.openstreetmap.org/">https://www.openstreetmap.org/</a> )
	POI		\	Gaode Map open platform
Vector	AQI		\	Beijing Municipal Ecological and Environmental Monitoring Center ( <a href="https://www.bjmemc.com.cn/">https://www.bjmemc.com.cn/</a> )
		2016-2020 hourly		National Centers for Environmental Information ( <a href="https://www.ncei.noaa.gov/">https://www.ncei.noaa.gov/</a> )
	wind speed		\	
Statistics	per_GDP	2016-2020 yearly	\	Beijing Municipal Bureau of Statistics ( <a href="https://tjj.beijing.gov.cn/">https://tjj.beijing.gov.cn/</a> )

2.2.1. Station Monitoring Data

The station monitoring data spans from 2016 to 2020, including hourly air quality monitoring data from 35 stations in Beijing and daily ground meteorological monitoring data from 400 stations across the country.

2.2.2. Remote Sensing Product

The remote sensing imagery products include global monthly maximum temperature, minimum temperature, and precipitation products from 2016 to 2020, average NDVI products, China population distribution data, the 2014 China topographic relief kilometer grid dataset, and the Luojia 1-01 NPP-DNB product. The Luojia 1-01 NPP-DNB product can reflect the intensity and distribution of human activities, making it a powerful tool for analyzing and estimating energy consumption levels. The study selected three cloud-free scenes from the NPP-DNB product for October 2018 and conducted preprocessing, which includes the radiance conversion formula as follows:

$$L = 10^{-10} \cdot DN^{3/2}$$

(1)

where  $L$  is the radiance value after absolute radiometric calibration, expressed in units of, and  $W/(m^2 \cdot sr \cdot \mu m)$  is the image grayscale value.

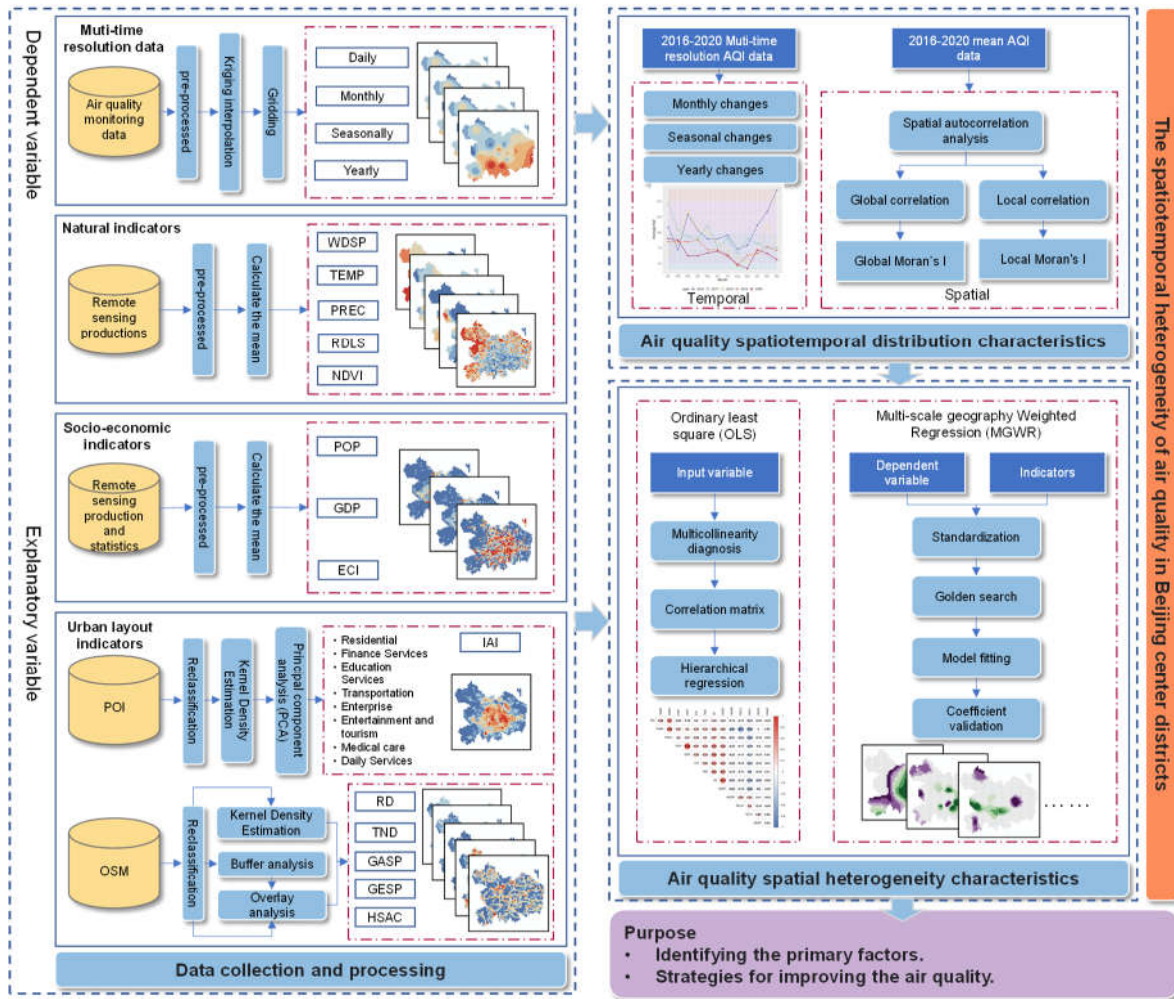
2.2.3. OSM and POI Data

The 2020 China Shapefile vector format OSM was obtained from Geofabrik's free download server (<https://download.geofabrik.de/>), and the data has been trimmed and topologically corrected. Additionally, 367,914 Points of Interest (POIs) within the study area were acquired through the Gaode open platform, categorized into eight classes: commercial residential, financial institutions, science and education culture, transportation facilities, corporate enterprises, leisure and entertainment, medical care, and lifestyle services.

3. Methods and Models

The research framework of this article (Figure 2), encompasses the following aspects: (1) Constructing an indicator system for urban air quality influencing factors based on natural environment, socio-economic, and urban layout dimensions; (2) Conducting multi-temporal scale studies of air quality trends to analyze and compare the trends of AQI at different time scales, and using spatial autocorrelation to analyze the overall spatial distribution characteristics of AQI; (3) Employing Geographically Weighted Regression (GWR) and Multi-scale Geographically Weighted

Regression (MGWR) models to quantitatively analyze the impact strength and spatial heterogeneity of air quality driving factors, providing refined references for regional development.



**Figure 2.** Research Framework.

### 3.1. IndicatorConstruction Module

Considering the natural environment, socio-economic factors, and urban spatial layout comprehensively, an evaluation indicator database for air quality driving factors at the urban scale has been constructed.

The Ordinary Kriging interpolation method is employed to interpolate discrete monitoring data, resulting in the distribution data of AQI as the dependent variable. The Ordinary Kriging interpolation estimates that if there are  $n$  known points  $x_i (i = 1, 2, \dots, n)$  around a point, then the estimated value at point  $x_0$  can be calculated using equation (2):

$$Z(X_0) = \sum_{i=1}^n \beta_i \cdot z(x_i) \quad (2)$$

$$\sum_{i=1}^n \beta_i = 1 \quad (3)$$

where  $\beta_i$  is the weight assigned to  $z(x_i)$  in the distribution. The weights are calculated using the Lagrange multipliers method, and the sum of the weights equals 1 (as stated in equation (3));  $z(x_i)$  is the observed value at the known location  $x_i$ ;  $Z(x_0)$  is the estimated value at the estimation location  $x_0$ ;  $n$  is the number of actual observed values used to estimate  $Z(x_0)$ .

Natural environmental factors are composed of meteorological elements, topographic features, and vegetation cover. Meteorological factors include average temperature (TEMP), precipitation (PREC), and wind speed (WDSP). Wind speed data are also obtained by Ordinary Kriging interpolation, producing annual average values for 2016-2020. Topographic features are represented by relief degree of land surface (RDLS), which describes the urban surface morphology. Vegetation cover is characterized by the Normalized Difference Vegetation Index (NDVI).

Socio-economic characteristics are comprised of population distribution (POP), distribution of Gross Domestic Product (GDP), and energy consumption intensity (ECI). Nightlight remote sensing data represent energy consumption intensity.

Urban spatial layout is composed of the Industrial Agglomeration Index (IAI), Traffic Network Density (TND), Residential Density (RD), Green Space Ratio (GESP), Grey Space Ratio (GASP), and Humidity Space Conditioning Ability (HSAC). Residential Density is used to identify areas of concentrated population and relatively sparse areas. The Industrial Agglomeration Index reflects the spatial concentration of industries within the city. Based on the selected eight categories of POI data, urban functional areas are identified through kernel density analysis. Subsequently, Principal Component Analysis (PCA) is applied to these complex functional area classification data to derive key dimensions, resulting in the Industrial Agglomeration Index, which facilitates an intuitive understanding and analysis of urban industrial distribution. Traffic Network Density reflects the aggregation of roads and railways within the study unit, and the calculation formula is as follows:

$$RD = (rail_{len} + road_{len})/grid_{area} \quad (4)$$

where  $rail_{len}$  is the total length of roads within the study unit,  $road_{len}$  is the total length of railways within the study unit, and  $grid_{area}$  is the area of the study unit.

The greenness space percentage (GESP) and grayness space percentage (GASP) represent the greenness space and floor area percentage in each grid, respectively. The urban humidity space is characterized by water bodies, and the humidity space is analyzed by multilevel buffer zones, and the buffer radii of 50m, 100m and 200m are set sequentially, corresponding to the air quality regulation coefficients of 0.65, 0.25 and 0.1, respectively. Finally, the humidity space regulation capacity of each unit is calculated according to the following equation:

$$HSAC = \sum_{i=1}^n x_i \cdot \beta_i \quad (5)$$

where  $\beta_i$  is the adjustment factor, and  $x_i$  is the area of the buffer corresponding to  $\beta_i$ .

### 3.2. Spatial Distribution Feature Generation

Spatial autocorrelation is categorized into global autocorrelation and local autocorrelation. Global autocorrelation can be characterized by the Global Moran's I statistic to characterize the overall spatial distribution of air quality. The formula is as follows:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{i,j} z_i z_j}{S_0 \sum_{i=1}^n z_i^2} \quad (6)$$

where  $n$  is equal to the total number of elements,  $S_0$  is the aggregation of all spatial weights,  $z_i$  is the deviation between the attribute of element  $i$  and its mean, and  $w_{i,j}$  is the spatial weight between elements  $i$  and  $j$ . The value  $> 0$  indicates that there is a positive correlation in spatial data, and the value  $< 0$  indicates that there is a negative correlation in spatial data. When the value is close to 0, there is no significant correlation.

Spatial autocorrelation is categorized into global autocorrelation and local autocorrelation. Global autocorrelation can be characterized by the Global Moran's I statistic to characterize the overall spatial distribution of air quality. The formula is as follows:

$$I_i = \frac{Z_i}{S^2} \sum_{j \neq i}^n w_{i,j} Z_j \quad (7)$$

where  $I_i$  is the local Moran's I of region  $i$ ,  $S^2$  is the variance of all observations,  $Z_i$  is the normalized observation,  $w_{i,j}$  is the spatial weight between element  $i$  and  $j$ ,  $Z_j$  is the standardized score of the attribute value of position  $j$ , and  $n$  is the synthesis of all regions in the study area.

### 3.3. Spatial Heterogeneity Feature Generation

Spatial variation in the impact of different influencing factors on the AQI was analyzed using GWR and MGWR. GWR is a local modeling tool based on the optimization of global regression models, where each sampling location has its own parameter with the following equation:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^n \beta_j(u_i, v_i) x_{ij} + \varepsilon_i \quad (8)$$

where  $(u_i, v_i)$  the coordinates of point  $i$  in space,  $\beta_j(u_i, v_i)$  is the regression coefficient of each variable at point  $i$ ,  $\beta_0(u_i, v_i)$  is a constant term, and  $\varepsilon_i$  is the regression residual of the point  $i$ , and  $n$  is the number of independent variables.

MGWR is an improved algorithm based on GWR, which can solve the problem that the GWR model defaults that all the influencing factors play a constant role on the spatial scale in the analysis of spatial heterogeneity, so that we can study the relationship between the influencing factors at different scales. The equation is as follows:

$$y_i = \beta_{bw0}(u_i, v_i) + \sum_{j=1}^n \beta_{bwj}(u_i, v_i) x_{ij} + \varepsilon_i \quad (9)$$

Each regression coefficient  $\beta_{bwj}$  is obtained based on a local regression process, and has different bandwidth settings. Where  $bw0$  is the intercept bandwidth,  $bwj$  is the bandwidth of the variable  $j$ , and the remaining parameters are consistent with equation (8).

In order to evaluate the effect of these two regression models, the coefficient of determination ( $R^2$ ), the corrected Akaike information criterion (AICc) and the residual sum of squares (RSS) were used as evaluation indicators.

## 4. Results and Discussion

### 4.1. Dominant Impact Indicators

The air quality of the monitoring area is measured by the average value of the historical data of each monitoring station in 5 years (Figure 3). The results show that the air quality in Fengtai District is the worst, and the air quality in Shijingshan District and Haidian District is the best. The area of Chaoyang District in the study area is large, the difference of site monitoring results is small, and the air quality is at a medium level. The eastern and western urban areas are the core areas of Beijing. The AQI level shows a trend of high in the north and south and low in the middle, that is, the central area has excellent air quality, while the air quality at the junction of the peripheral urban areas is generally lower.



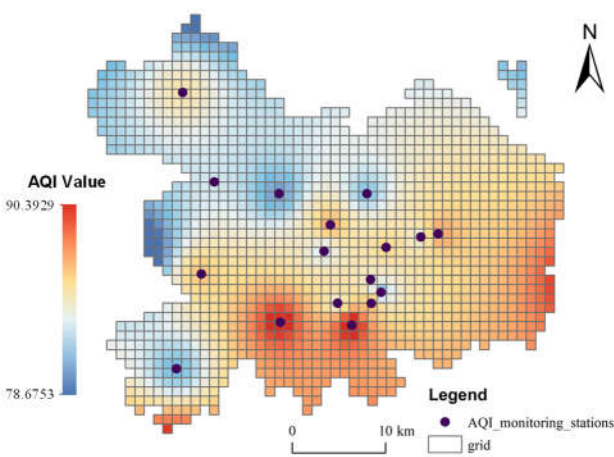


Figure 3. Spatial distribution of average AQI from 2016 to 2020.

The correlation analysis matrix (Figure 4) shows that the correlation between POP and GDP is 0.85, and the correlation between IAI and GASP is 0.75, both of which show strong positive correlation. The correlation between TEMP and RDLS is -0.73, showing a strong negative correlation. Through multicollinearity diagnosis, it is further verified whether these highly correlated indicators need to be eliminated. The diagnosis results Figure 4. Correlation analysis matrix of indicators.

Table 2 show that the VIF of all indicators is less than 5, which indicates that there is no covariance relationship between the indicators and can explain their impact on air quality separately.

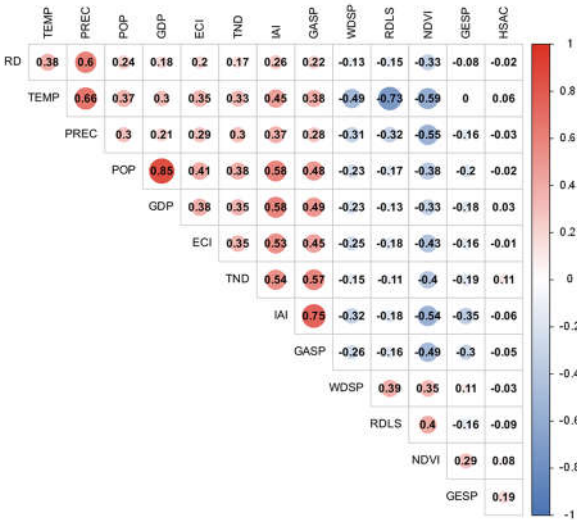


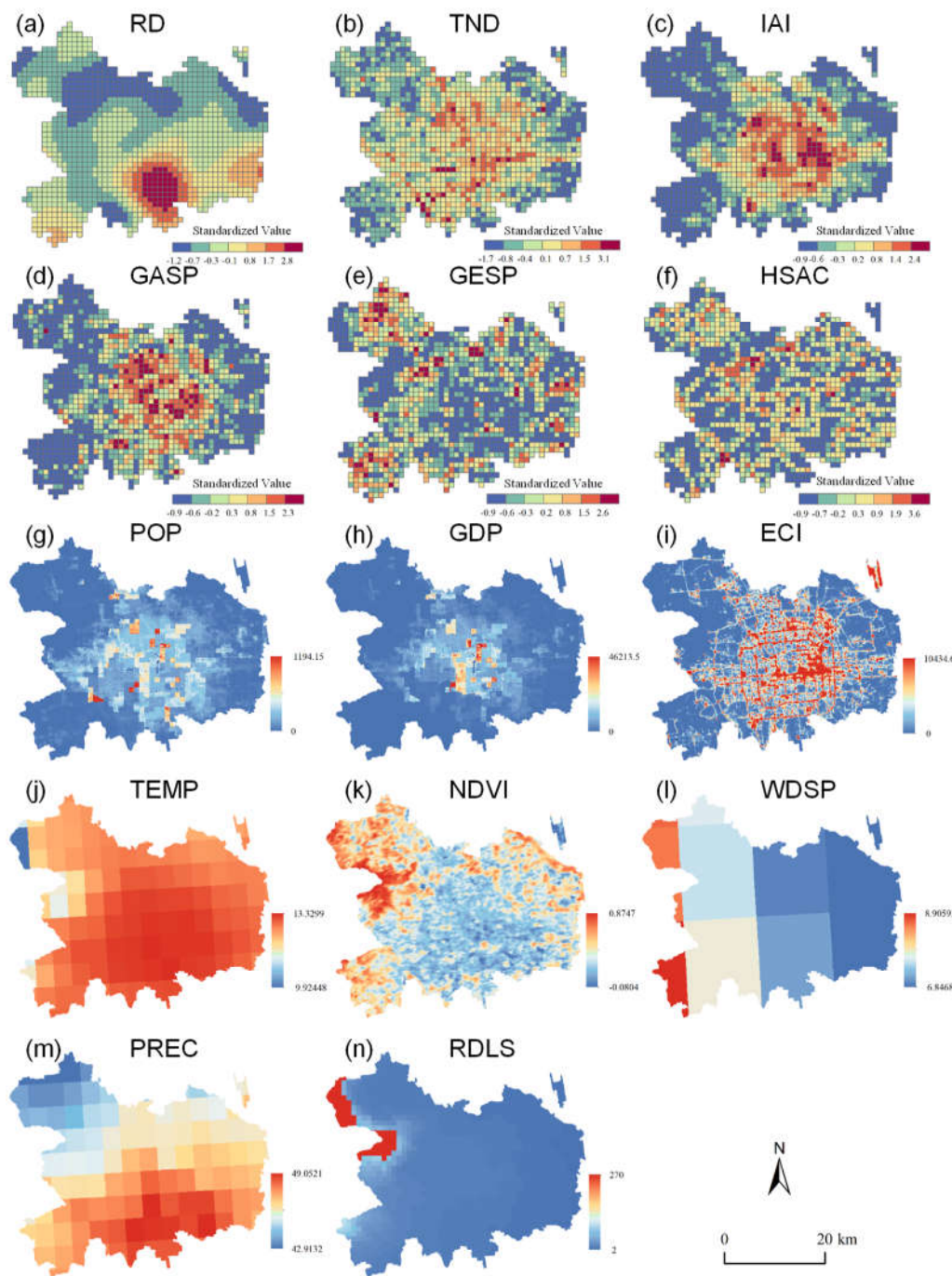
Figure 4. Correlation analysis matrix of indicators.

Table 2. AQI explanatory variable index system and multicollinearity diagnostic results.

Indicator Category	Indicator	Abbreviation	Description	Unit	Collinearity Statistic	
					Tolerance	VIF
Natural indicators	Wind Speed	WDSP	Mean wind speed from 2016 to 2020	.1 knots	0.720	1.390
	Temperature	TEMP	Mean temperature from 2016 to 2020	℃	0.215	4.659
	Precipitation	PREC	Mean precipitation from 2016 to 2020	mm	0.360	2.780

Socioeconomic indicators	Relief Degree of Land Surface	RDLS	Comprehensive characterization of altitude and surface incision in 2014	\	0.383	2.613
	Normalized Difference Vegetation Index	NDVI	Mean NDVI from 2016 to 2020	\	0.466	2.147
	Energy Consumption Intensity	ECI	Characterization using nighttime light remote sensing data in 2018	\	0.670	1.494
	Population	POP	Total population distribution	10 <sup>4</sup> people	0.245	4.090
	Gross Domestic Product	GDP	Mean GDP from 2016 to 2020	10 <sup>4</sup> yuan	0.248	4.034
	Industry Aggregation Index	IAI	Reflect the agglomeration of urban industries	\	0.304	3.295
	Transportation network density	TND	The ratio of road network length to unit area in each grid	m/km <sup>2</sup>	0.589	1.696
Urban layout indicators	Residential density	RD	Reflect the gathering situation of residential areas	\	0.618	1.617
	Green space percent	GESP	Green space area to unit ratio	%	0.736	1.358
	Gray space percent	GASP	Building area to unit ratio	%	0.381	2.627
	Humidity space adjustment capability	HSAC	Reflects the ability of wetlands to purify the air quality	\	0.907	1.103

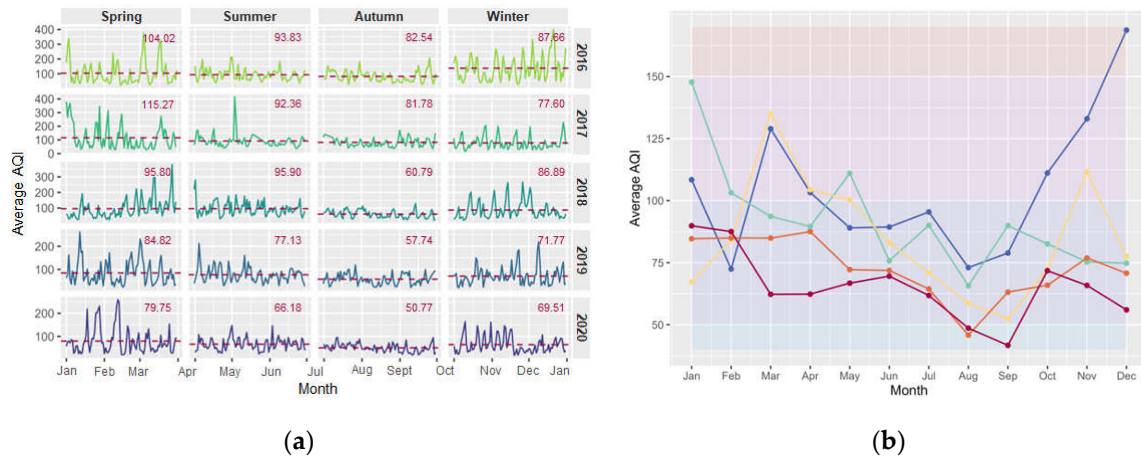
The spatial distribution of each variable reflects the spatial distribution heterogeneity of the driving factors (Figure 5). As an indicator of urban spatial pattern, Figure 5a-f show the most obvious spatial heterogeneity.



**Figure 5.** Spatial distribution of AQI variables.

4.2. Temporal Characteristics of AQI

In the five years from 2016 to 2020, the annual AQI of the study area showed an overall downward trend (Figure 6a). AQI followed a U-shaped variation between seasons (Figure 6b), and the AQI in spring was generally higher. Subsequently, in summer, the increase in rainfall usually helps to reduce pollutants in the air, resulting in a decrease in AQI values. In autumn, AQI remained at a low level, but in winter, with the increase of heating demand, coal pollution caused AQI value to rise again.

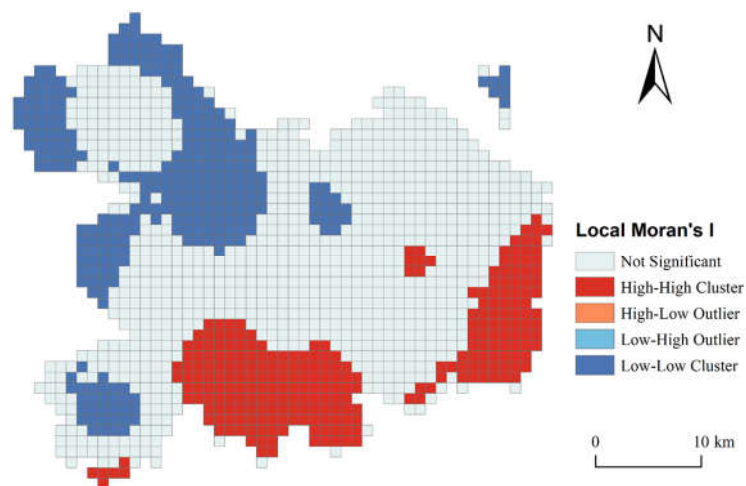


**Figure 6.** AQI multi-time scale variation characteristics. (a) The multi-panel map of AQI over time from 2016 to 2020, each panel represents the change and mean of AQI in the season of the year. (b) The monthly average AQI line chart of 2016-2020 and the corresponding air quality grade of each month.

The seasonal variation of monthly average AQI values from 2016 to 2020 highlights the difference in monthly air quality (Figure 6b). At the turn of summer and autumn, the AQI value is usually lower, which may be due to the more precipitation during this period, which helps to clean the air. From 2016 to 2018, air quality was mainly concentrated in the range of moderate and mild pollution. From 2019 to 2020, air quality is mainly distributed between good and medium grades, showing a significant increase in the proportion of air quality improvement.

#### 4.3. Spatial Characteristics of AQI

The global Moran's  $I = 0.922$ , indicating that there is a significant positive spatial autocorrelation in AQI in the study area. Through local spatial autocorrelation analysis, cold and hot spots are obtained (Figure 7). The population in the hot spot area is concentrated and distributed, the coal consumption for heating in winter is large, and the power consumption for high temperature in summer is large, resulting in a large number of air pollutant emissions. In addition, the plain terrain is conducive to the spread of air pollutants between regions, which eventually leads to the accumulation of high AQI values. Multiple continuous grids in Haidian District in the northwest of the study area belong to the cold spot area, and the AQI value is low, showing low-low cluster and high local correlation. The main reason is that Haidian District is located in a number of large national parks, the northwest has a high greening rate and strong air purification capacity.



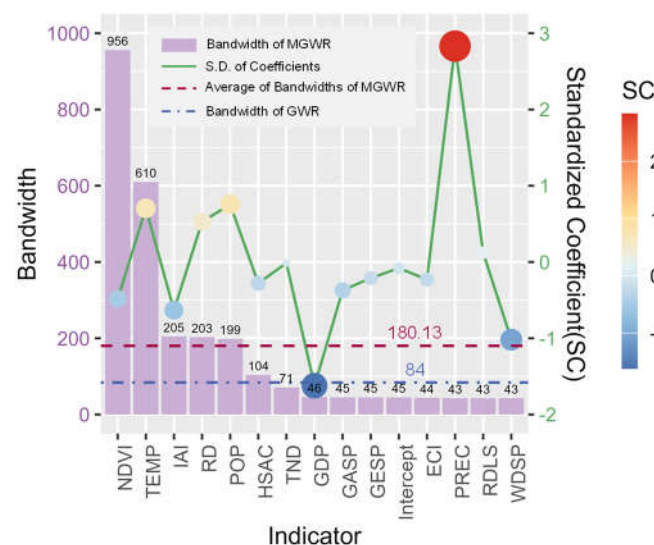


**Figure 7.** Local spatial association index (LISA).

#### 4.4. Spatial Heterogeneity Analysis of AQI Influencing Factors

The multi-scale geographically weighted regression (MGWR) model shows better performance than the traditional geographically weighted regression (GWR) model. Specifically, the MGWR model has significantly higher coefficient of determination ( $R^2 = 0.985$ ) and adjusted coefficient of determination (Adjusted  $R^2 = 0.975$ ), which means that the model has significant explanatory power for almost all explanatory variables. In addition, the MGWR model has lower Akaike information criterion correction value ( $AICc = 104.817$ ) and residual sum of squares ( $RSS = 20.318$ ), indicating that it has higher accuracy and lower error in fitting data, and can better capture and explain the local variability of spatial data.

The optimal bandwidth of each variable generated by the MGWR and GWR models, and the standardized parameter estimation corresponding to the MGWR model variables (Figure 8), where the optimal bandwidth of the GWR model is 84, and the average of the 15 bandwidths of the MGWR model is 180.13. MGWR simulates the effects of variables on different spatial scales, resulting in different optimal bandwidths. The larger the bandwidth, the greater the overall impact of the variable on the air quality in space, the higher the degree of spatial stability, and the smaller the spatial heterogeneity. On the contrary, variables with smaller bandwidth have a significant impact on air quality in a local range, with a high degree of spatial non-stationarity and high spatial heterogeneity. The results show that NDVI and TEMP have large bandwidth and low spatial heterogeneity. The bandwidth of IAI, RD and POP are slightly higher than the average bandwidth, indicating that the spatial heterogeneity of these three variables is at a medium level. The nine variables of HSAC, TND, GDP, GASP, GESP, ECI, PREC, RDLS, and WDSP have small bandwidths and significant heterogeneity in the impact on air quality.

**Figure 8.** The optimal bandwidth generated by MGWR and GWR and the standardized parameter estimates corresponding to each variable in the MGWR model.

The standardized coefficient (SC) can reflect the degree of influence of variables on air quality. From the standardized correlation coefficient, the SC values of TND and RDLS are close to 0, indicating that these two variables have no significant effect on air quality.  $PREC > POP > TEMP > RD > 0.5$ , indicating that there is a significant positive relationship between these four variables and AQI, indicating that these four variables are important factors that aggravate air pollution.  $GDP < WDSP < IAI < -0.5 < NDVI < GASP < HSAC < ECI < GESP < 0$ , showing a negative relationship, that is, the value of AQI decreases with the increase of variable value, which has a positive effect on the improvement of air quality.

4.2.2. Spatial Heterogeneity Analysis of Influencing Factors

The local regression coefficient of each variable and its significant proportion (Table 3). The results show that at the level of  $p \leq 0.05$ , the parameter estimates of PREC, TEMP, POP and RD are significant in more than 85 % of the region, which is a significant positive correlation. The parameter estimates of NDVI, WDSP, GDP and IAI are significant in more than 70% of the region, which is a significant negative correlation. Among the remaining variables, the positive effects of RDLS, GASP and TND on AQI are higher than the negative effects, while the negative effects of GESP and HSAC are higher than the positive effects.

Table 3. Statistical characteristics of MGWR coefficients.

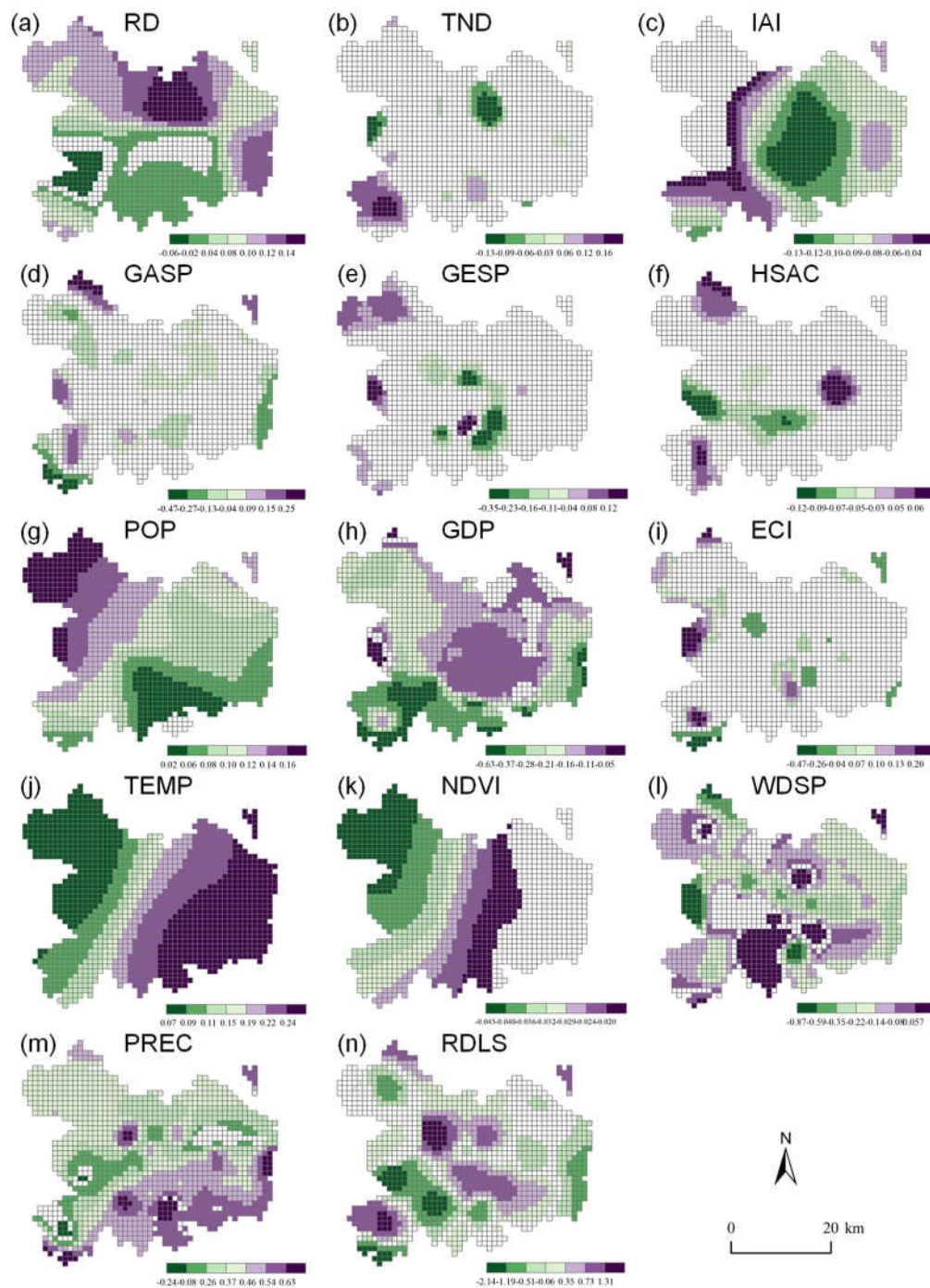
Indicator Category	Indicator	MGWR Coefficients			Percentage of grids by significance (95 % Level) of t-Test		
		Min	Max	Mean	$P \leq .05$ (%)	+ (%)	- (%)
NI	NDVI	-0.043	-0.010	-0.026	70.55	0	100
	PREC	-0.242	0.877	0.404	93.49	99.31	0.69
	RDLS	-2.143	2.137	0.053	70.98	58.71	41.29
	TEMP	0.069	0.257	0.173	100	100	0
	WDSP	-0.872	0.468	-0.126	85.53	13.11	86.89
SEI	POP	0.020	0.209	0.108	98.77	100	0
	ECI	-0.466	0.286	0.015	20.11	74.46	25.54
	GDP	-0.634	0.373	-0.178	91.90	2.05	97.95
ULI	GESP	-0.346	0.188	-0.009	25.11	47.84	52.16
	HSAC	-0.119	0.082	-0.004	26.34	47.80	52.20
	GASP	-0.464	0.435	0.003	33.72	54.94	45.06
	TND	-0.140	0.200	0.009	17.44	63.07	36.93
	RD	-0.065	0.186	0.076	87.84	95.30	4.70
	IAI	-0.126	0.034	-0.072	76.70	0	100

The spatial distribution of the normalized coefficients of the MGWR model variables (Figure 9) shows the different characteristics of the spatial heterogeneity of the variables. In the urban spatial layout index: the residential density (Figure 9a) has the most significant positive correlation, the negative area of the regression coefficient is concentrated in the south of Shijingshan District and the northwest of Fengtai District, and the positive area is concentrated in Haidian District and Chaoyang District. The border area and the eastern part of Chaoyang District. The traffic network density TND (Figure 9b) has a negative significance in the northwest of Chaoyang District and a positive significance in the west of Fengtai District, indicating that although the current road traffic network density in Fengtai District is low, the traffic pollution emissions are large. IAI is the most significant negative correlation variable (Figure 9c), and this correlation shows a trend of weakening from the center of Dongcheng District and Xicheng District to the surrounding areas, indicating that the current industry in the study area is relatively reasonable. Figures d-e represent the heterogeneity of the impact of urban green space, gray space and humidity space on air quality, and reflect the effect of urban architecture and ecological environment configuration on air quality. On the whole, the positive effect of GASP in the study area is greater than the negative effect. This may be because the concentrated distribution of buildings forms a local heat island effect, which increases the concentration of air pollutants and is not easy to diffuse. GESP and HSAC showed the opposite effect with GASP. Increasing urban green space and wetland is still an important means to purify urban air.

In the socio-economic indicators, POP (Figure 9g) showed a significant positive correlation, showing an increasing trend from the southeast to the northwest of the six districts of the city. GDP (Figure 9h) shows a significant negative correlation, and the correlation is generally manifested as an increasing trend from the middle of the study area to the surrounding areas. The overall performance

of ECI (Figure 9i) is positively correlated, indicating that under current conditions, reducing the current energy (especially electrical energy) consumption intensity will help improve air quality.

Among the natural environment indicators, TEMP (Figure 9j) and NDVI (Figure 9k) showed significant positive and negative correlations, respectively, and the positive correlation was stronger than the negative correlation. The positive correlation shows an increasing trend from west to east, because the vegetation distribution has a regulating effect on the surface temperature. The influence of WDSP on AQI is negatively correlated as a whole. The area with high wind speed is helpful to the diffusion ability of pollutants and reduce the accumulation of pollutants. PREC (Figure 9m) shows a strong positive correlation as a whole, which may be related to the research scale. This study considers the average effect in a long time series, ignoring the relationship between rainfall and AQI over time. RDLS (Figure 9n) has no significant effect on AQI, because pollutants are not easy to gather in undulating flat areas. However, the spatial distribution of the regression coefficient of this variable has certain heterogeneity, which may be caused by the distribution of buildings.



**Figure 9.** Spatial distribution of MGWR standardized coefficients.

#### 4.5. Strategies and Suggestions

In view of the comprehensive strategy of improving air quality in the study area, we can start from the three key dimensions of natural environmental protection, social and economic sustainable development, and urban spatial layout optimization.

In terms of the natural environment, we attach importance to the role of wind speed (WDSP) in promoting the diffusion of air pollutants, and through the rational design of urban planning, such as building ventilation corridors, optimizing building layout, and enhancing air circulation capacity. At the same time, by increasing urban vegetation coverage (NDVI), it not only regulates urban temperature, but also reduces the heat island effect, thereby improving air quality. In terms of socio-economic characteristics, reducing energy consumption intensity (ECI) and balancing the relationship between population density (POP) and economic growth (GDP) are crucial. The optimization of urban spatial layout needs to pay attention to the impact of industrial activities (IAI) and traffic network (TND) on air quality. By strictly controlling industrial emissions, optimizing traffic structure, encouraging the use of public transport, and increasing urban green space (GESp), improving gray space (GASP) and humidity space (HSAC), it can not only effectively reduce air pollution sources, but also provide a more livable environment by enhancing the city's ecosystem services.

## 5. Conclusions

This paper analyzes the spatial and temporal distribution characteristics of AQI in the study area, and explores the relationship and spatial heterogeneity of the driving factors affecting urban air quality from three aspects: natural environment, social economy and urban spatial layout based on multi-source data. The main conclusions of the study are as follows: (1) The spatial and temporal distribution of air quality in the study area shows certain regularity. In terms of time distribution, the overall air quality shows a trend of increasing year by year, and shows a U-shaped change rule on the seasonal scale, that is, spring and winter are the high-value seasons of AQI, while summer and autumn are low. In terms of spatial distribution, the high-value aggregation area of AQI is distributed in the eastern part of Chaoyang District and Fengtai District, which is related to the strong human activities in the region. The low-value aggregation areas are distributed in the central and northern parts of Haidian District and the western part of Shijingshan District. The possible reason is that the regional natural environment is relatively superior. (2) The fitting effect of MGWR model is better than that of GWR. The bandwidth of MGWR model can dynamically reflect the spatial effect of variables, which is suitable for mining the dominant factors affecting air quality and their spatial effect differences from multivariate. The standardized parameter estimates obtained from the MGWR model reflect that the influence of TND and RDLS in the variables used in the study is low. Among the variables with a significant positive relationship with AQI, the degree of influence is :  $PREC > POP > TEMP > RD > 0.5$ . Among the variables showing a negative relationship with AQI, the degree of influence is as follows :  $GDP < WDSP < IAI < -0.5 < NDVI < GASP < HSAC < ECI < GESp < 0$ . (3) Identify the intensity and spatial heterogeneity of driving factors. Among the natural environment indicators, the spatial heterogeneity of NDVI and TEMP is higher than that of PREC, RDLS and WDSP. In the socio-economic indicators, the spatial heterogeneity of POP, ECI and GDP is similar. Among the urban spatial layout indicators, the heterogeneity of GESp, HSAC, TND and RD is similar, and higher than that of GASP and IAI. The strong spatial heterogeneity variables with significant positive effects on air quality are NDVI, GDP, and HSAC, and the strong spatial heterogeneity variables with significant negative effects are POP, TEMP, and RD. Therefore, in the process of urban planning, it is necessary to coordinate the relationship between population and development quality, natural environment and building environment from the perspective of regional differences, so as to improve air quality.

However, this study has certain limitations, which are reflected in the analysis of the heterogeneity of driving factors on the missing time scale. In the future research, we will deeply



explore the spatial and temporal characteristics of driving factors in the process of spatial and temporal dynamic change, which will help each region to formulate fine management countermeasures scientifically.

**Author Contributions:** Conceptualization, Zhixiong Tan and Haili Wu; Methodology, Zhixiong Tan, Haili Wu and Jiejun Huang; Software, Qingyang Chen; Validation, Haili Wu; Formal analysis, Qingyang Chen; Resources, Jiejun Huang; Data curation, Zhixiong Tan and Haili Wu; Writing—original draft, Zhixiong Tan and Haili Wu; Writing—review & editing, Zhixiong Tan, Haili Wu and Jiejun Huang; Visualization, Qingyang Chen; Supervision, Jiejun Huang. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the by the National Natural Science Foundation of China (No. 41071104) and the National Innovation and Entrepreneurship Training Program for College Students (No. S202310497275).

**Data Availability Statement:** Data derived from public domain resources.

**Conflicts of Interest:** The authors declare no conflicts of interest.

Appendix A

Table A1. AQI Classification Standards.

AQI Value	Classification	Air Quality Level
0~50	Level 1	Good
51~100	Level 2	Moderate
101~150	Level 3	Lightly polluted
151~200	Level 4	Moderately polluted
201~300	Level 5	Heavily polluted
>300	Level 6	Seriously polluted

References

1. Cohen, B. Urbanization in Developing Countries: Current Trends, Future Projections, and Key Challenges for Sustainability. *Technol. Soc.* **2006**, *28*, 63–80.
2. Li, R.; Cui, L.; Li, J.; Zhao, A.; Fu, H.; Wu, Y.; Zhang, L.; Kong, L.; Chen, J. Spatial and Temporal Variation of Particulate Matter and Gaseous Pollutants in China during 2014–2016. *Atmos. Environ.* **2017**, *161*, 235–246.
3. Zhou, D.; Zhao, S.; Zhang, L.; Sun, G.; Liu, Y. The Footprint of Urban Heat Island Effect in China. *Sci. Rep.* **2015**, *5*, 11160.
4. Horton, R.M.; Mankin, J.S.; Lesk, C.; Coffel, E.; Raymond, C. A Review of Recent Advances in Research on Extreme Heat Events. *Curr. Clim. Change Rep.* **2016**, *2*, 242–259.
5. Fang, C.; Liu, H.; Li, G.; Sun, D.; Miao, Z. Estimating the Impact of Urbanization on Air Quality in China Using Spatial Regression Models. *Sustainability* **2015**, *7*, 15570–15592.
6. Chan, C.K.; Yao, X. Air Pollution in Mega Cities in China. *Atmos. Environ.* **2008**, *42*, 1–42.
7. Huang, R.-J.; Zhang, Y.; Bozzetti, C.; Ho, K.-F.; Cao, J.-J.; Han, Y.; Daellenbach, K.R.; Slowik, J.G.; Platt, S.M.; Canonaco, F. High Secondary Aerosol Contribution to Particulate Pollution during Haze Events in China. *Nature* **2014**, *514*, 218–222.
8. Wang, T.; Xue, L.; Brimblecombe, P.; Lam, Y.F.; Li, L.; Zhang, L. Ozone Pollution in China: A Review of Concentrations, Meteorological Influences, Chemical Precursors, and Effects. *Sci. Total Environ.* **2017**, *575*, 1582–1596.
9. Xu, W.; Sun, J.; Liu, Y.; Xiao, Y.; Tian, Y.; Zhao, B.; Zhang, X. Spatiotemporal Variation and Socioeconomic Drivers of Air Pollution in China during 2005–2016. *J. Environ. Manage.* **2019**, *245*, 66–75.
10. Pak UnJin, P.U.; Ma Jun, M.J.; Ryu UnSok, R.U.; Ryom KwangChol, R.K.; Juhyok, U.; Pak KyongSok, P.K.; Pak ChanIl, P.C. Deep Learning-Based PM2. 5 Prediction Considering the Spatiotemporal Correlations: A Case Study of Beijing, China. **2020**.
11. Hu, J.; Ying, Q.; Wang, Y.; Zhang, H. Characterizing Multi-Pollutant Air Pollution in China: Comparison of Three Air Quality Indices. *Environ. Int.* **2015**, *84*, 17–25.
12. Zhang, J.; Cui, K.; Wang, Y.-F.; Wu, J.-L.; Huang, W.-S.; Wan, S.; Xu, K. Temporal Variations in the Air Quality Index and the Impact of the COVID-19 Event on Air Quality in Western China. *Aerosol Air Qual. Res.* **2020**, *20*, 1552–1568.

13. World Health Organization *WHO Global Air Quality Guidelines: Particulate Matter (PM<sub>2.5</sub> and PM<sub>10</sub>), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*; World Health Organization, 2021; ISBN 92-4-003422-6.
14. Shen, F.; Ge, X.; Hu, J.; Nie, D.; Tian, L.; Chen, M. Air Pollution Characteristics and Health Risks in Henan Province, China. *Environ. Res.* **2017**, *156*, 625–634.
15. Du, X.; Chen, R.; Meng, X.; Liu, C.; Niu, Y.; Wang, W.; Li, S.; Kan, H.; Zhou, M. The Establishment of National Air Quality Health Index in China. *Environ. Int.* **2020**, *138*, 105594.
16. Liao, T.; Jiang, W.; Ouyang, Z.; Hu, S.; Wu, J.; Zhao, B.; Wang, B.; Wang, S.; Sun, Y. Evaluation of the Health Risk of Air Pollution in Major Chinese Cities Using a Risk-Based, Multi-Pollutant Air Quality Health Index during 2014–2018. *Air Qual. Atmosphere Health* **2021**, *14*, 1605–1617.
17. Xu, L.; Zhou, J.; Guo, Y.; Wu, T.; Chen, T.; Zhong, Q.; Yuan, D.; Chen, P.; Ou, C. Spatiotemporal Pattern of Air Quality Index and Its Associated Factors in 31 Chinese Provincial Capital Cities. *Air Qual. Atmosphere Health* **2017**, *10*, 601–609.
18. Zhang, X.; Gong, Z. Spatiotemporal Characteristics of Urban Air Quality in China and Geographic Detection of Their Determinants. *J. Geogr. Sci.* **2018**, *28*, 563–578.
19. Jiang, W.; Wang, Y.; Tsou, M.-H.; Fu, X. Using Social Media to Detect Outdoor Air Pollution and Monitor Air Quality Index (AQI): A Geo-Targeted Spatiotemporal Analysis Framework with Sina Weibo (Chinese Twitter). *PloS One* **2015**, *10*, e0141185.
20. Tan, S.; Xie, D.; Ni, C.; Zhao, G.; Shao, J.; Chen, F.; Ni, J. Spatiotemporal Characteristics of Air Pollution in Chengdu-Chongqing Urban Agglomeration (CCUA) in Southwest, China: 2015–2021. *J. Environ. Manage.* **2023**, *325*, 116503.
21. Shi, K.; Chen, Y.; Li, L.; Huang, C. Spatiotemporal Variations of Urban CO<sub>2</sub> Emissions in China: A Multiscale Perspective. *Appl. Energy* **2018**, *211*, 218–229.
22. Wu, C.; Hu, W.; Zhou, M.; Li, S.; Jia, Y. Data-Driven Regionalization for Analyzing the Spatiotemporal Characteristics of Air Quality in China. *Atmos. Environ.* **2019**, *203*, 172–182.
23. Fotheringham, A.S.; Yue, H.; Li, Z. Examining the Influences of Air Quality in China's Cities Using Multi-scale Geographically Weighted Regression. *Trans. GIS* **2019**, *23*, 1444–1464.
24. Xu, W.; Tian, Y.; Liu, Y.; Zhao, B.; Liu, Y.; Zhang, X. Understanding the Spatial-Temporal Patterns and Influential Factors on Air Quality Index: The Case of North China. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2820.
25. Miao, L.; Liu, C.; Yang, X.; Kwan, M.-P.; Zhang, K. Spatiotemporal Heterogeneity Analysis of Air Quality in the Yangtze River Delta, China. *Sustain. Cities Soc.* **2022**, *78*, 103603.
26. Lu, X.; Yao, T.; Fung, J.C.; Lin, C. Estimation of Health and Economic Costs of Air Pollution over the Pearl River Delta Region in China. *Sci. Total Environ.* **2016**, *566*, 134–143.
27. Yuan, J.; Wang, X.; Feng, Z.; Zhang, Y.; Yu, M. Spatiotemporal Variations of Aerosol Optical Depth and the Spatial Heterogeneity Relationship of Potential Factors Based on the Multi-Scale Geographically Weighted Regression Model in Chinese National-Level Urban Agglomerations. *Remote Sens.* **2023**, *15*, 4613.
28. Guo, Y.; Tang, Q.; Gong, D.-Y.; Zhang, Z. Estimating Ground-Level PM<sub>2.5</sub> Concentrations in Beijing Using a Satellite-Based Geographically and Temporally Weighted Regression Model. *Remote Sens. Environ.* **2017**, *198*, 140–149.
29. Wang, Q.; Feng, H.; Feng, H.; Yu, Y.; Li, J.; Ning, E. The Impacts of Road Traffic on Urban Air Quality in Jinan Based GWR and Remote Sensing. *Sci. Rep.* **2021**, *11*, 15512.
30. Wang, Z.; Ma, P.; Zhang, L.; Chen, H.; Zhao, S.; Zhou, W.; Chen, C.; Zhang, Y.; Zhou, C.; Mao, H. Systematics of Atmospheric Environment Monitoring in China via Satellite Remote Sensing. *Air Qual. Atmosphere Health* **2021**, *14*, 157–169.
31. Zheng, S.; Wang, J.; Sun, C.; Zhang, X.; Kahn, M.E. Air Pollution Lowers Chinese Urbanites' Expressed Happiness on Social Media. *Nat. Hum. Behav.* **2019**, *3*, 237–243.
32. Lin, B.; Zhu, J. Changes in Urban Air Quality during Urbanization in China. *J. Clean. Prod.* **2018**, *188*, 312–321.
33. Zhao, S.; Yu, Y.; Yin, D.; He, J.; Liu, N.; Qu, J.; Xiao, J. Annual and Diurnal Variations of Gaseous and Particulate Pollutants in 31 Provincial Capital Cities Based on in Situ Air Quality Monitoring Data from China National Environmental Monitoring Center. *Environ. Int.* **2016**, *86*, 92–106.
34. Zhu, Y.; Wang, J.; Meng, B.; Ji, H.; Wang, S.; Zhi, G.; Liu, J.; Shi, C. Quantifying Spatiotemporal Heterogeneities in PM<sub>2.5</sub>-Related Health and Associated Determinants Using Geospatial Big Data: A Case Study in Beijing. *Remote Sens.* **2022**, *14*, 4012.
35. Sun, Z.; Zhan, D.; Jin, F. Spatio-Temporal Characteristics and Geographical Determinants of Air Quality in Cities at the Prefecture Level and above in China. *Chin. Geogr. Sci.* **2019**, *29*, 316–324.
36. Grzędzicka, E. Is the Existing Urban Greenery Enough to Cope with Current Concentrations of PM<sub>2.5</sub>, PM<sub>10</sub> and CO<sub>2</sub>? *Atmospheric Pollut. Res.* **2019**, *10*, 219–233.
37. Li, F.; Zhou, T.; Lan, F. Relationships between Urban Form and Air Quality at Different Spatial Scales: A Case Study from Northern China. *Ecol. Indic.* **2021**, *121*, 107029.

38. Liu, J.; Ding, W. Spatial and Temporal Coupling Characteristics of Industrial Structure Optimization and Air Quality in Chinese Cities and Multi-Scale Driver Analysis. *Environ. Sci. Pollut. Res.* **2023**, *30*, 83888–83902.
39. Rao, Y.; Wu, C.; He, Q. The Antagonistic Effect of Urban Growth Pattern and Shrinking Cities on Air Quality: Based on the Empirical Analysis of 174 Cities in China. *Sustain. Cities Soc.* **2023**, *97*, 104752.
40. Singh, K.P.; Gupta, S.; Kumar, A.; Shukla, S.P. Linear and Nonlinear Modeling Approaches for Urban Air Quality Prediction. *Sci. Total Environ.* **2012**, *426*, 244–255.
41. Zhang, L.; Tian, X.; Zhao, Y.; Liu, L.; Li, Z.; Tao, L.; Wang, X.; Guo, X.; Luo, Y. Application of Nonlinear Land Use Regression Models for Ambient Air Pollutants and Air Quality Index. *Atmospheric Pollut. Res.* **2021**, *12*, 101186.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.