DOI: 10.1111/tgis.12580

RESEARCH ARTICLE

Transactions in GIS WILEY

Examining the influences of air quality in China's cities using multi-scale geographically weighted regression

A. Stewart Fotheringham[1] | Han Yue[2] | Ziqi Li[1]

[1]School of Geographical Sciences & Planning, Arizona State University, Tempe, Arizona

[2]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

Correspondence
Han Yue, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, No.129 Luoyu Road, Wuhan, China, 430079.
Email: hanyue.geo@gmail.com

Funding information
The US National Science Foundation, Grant/Award Number: 1758786; the China Scholarship Council, Grant/Award Number: 201806270199

Abstract

This study evaluates the influences of air pollution in China using a recently proposed model—multi-scale geographically weighted regression (MGWR). First, we review previous research on the determinants of air quality. Then, we explain the MGWR model, together with two global models: ordinary least squares (OLS) and OLS containing a spatial lag variable (OLSL) and a commonly used local model: geographically weighted regression (GWR). To detect and account for any variation of the spatial autocorrelation of air pollution over space, we construct two extra local models which we call GWR with lagged dependent variable (GWRL) and MGWR with lagged dependent variable (MGWRL) by including the lagged form of the dependent variable in the GWR model and the MGWR model, respectively. The performances of these six models are comprehensively examined and the MGWR and MGWRL models outperform the two global models as well as the GWR and GWRL models. MGWRL is the most accurate model in terms of replicating the observed air quality index (AQI) values and removing residual dependency. The superiority of the MGWR framework over the GWR framework is demonstrated—GWR can only produce a single optimized bandwidth, while MGWR provides covariate-specific optimized bandwidths which indicate the different spatial scales that different processes operate.

1 | INTRODUCTION

China became the world's second largest economy in 2010 after only 30 years of rapid economic evolution since the reform of its economy in 1978. However, this rapid urbanization and industrial process has triggered a series of environmental problems, among which air pollution is one of the most serious (Fang, Liu, Li, Sun, & Miao, 2015). Air pollution not only inflicts damage on ecosystems and affects climate change, but also impacts human health by triggering various medical problems, such as lung cancer, asthma, cardiovascular disease, and respiratory infections (Brauer et al., 2012; Pope III, 2000). Ambient air pollution has become the fourth greatest risk factor in all deaths in China (Matus et al., 2012). Air quality degradation in China has also led to enormous financial losses (Crane & Mao, 2015). Consequently, it is important to study the distribution patterns and driving forces of ambient air pollution in China in order to propose prevention and control measures (Hao & Liu, 2016; Zhou, Chen, & Wang, 2018).

1.1 | Variables affecting air quality

Air pollution is formed by a complex set of mechanisms, and various factors have been demonstrated to have an impact on it. One important category of factors covers meteorological conditions. For example, a study in Guangzhou, China reported a negative relationship between air pollutant level and precipitation, because during precipitation events the accumulated air pollutants absorb more water and fall to the ground (Li et al., 2014). A significant negative relationship between air pollution level and daily amount of precipitation was also observed in Birmingham, UK (Vardoulakis & Kassomenos, 2008). Wind velocity also affects air quality, because high winds increase horizontal mixing and are helpful in dispersing and diluting air pollutants (Dawson, Adams, & Pandis, 2007; Hu et al., 2013; Kleeman, 2007; Li et al., 2014). However, under certain circumstances, wind may blow soil and road dust into the air, which will increase air pollution levels (Vardoulakis & Kassomenos, 2008).

Socio-demographic characteristics are another critical type of impact factor. Previous research suggests the existence of a positive relationship between population density and air pollution, which is obvious because people need to consume energy to support their life and production, and greater energy consumption implies more waste gas (Zhou et al., 2018). China is in the process of rapid urbanization, which has raised environmental problems (Wang, Fang, & Wang, 2016). In addition to population density, economic development also has an impact on air quality. However, the relationship between air quality and economic development is not clear. Some researchers have suggested an inverted U-shaped relationship between economic development and air quality (Hao & Liu, 2016), with cities having higher per capita gross regional product (GRP) experiencing better air quality because a high degree of affluence improves awareness and the ability to pay for environmental protection (Wang & Fang, 2016; Zhou et al., 2018). Equally, cities with low GRP might experience a low level of air pollution because of little industrial development. The industrial structure of a city is usually measured by the contribution of secondary industries to GRP, which in China cover traditional manufacturing, construction industry manufacturing, and the production and supply of electric power, gas, and water. These are all high energy-consuming sectors and a high share of secondary industries has been widely demonstrated to cause worse ambient air quality (Wang & Fang, 2016; Wang, Zhou, Wang, Feng, & Hubacek, 2017). Electric energy in China is mainly generated by burning coal, which results in carbon emissions and dust pollution, and a larger amount of electricity consumption is then correlated with more adverse air quality (Yu, 2017). Industrial soot or dust emitted from factories to the air during fuel combustion will also aggravate pollution (Remoundaki et al., 2012; Wang & Fang, 2016).

Built environment factors also affect air quality. It is commonly recognized that the density of traffic vehicles will increase air pollution as automobile exhausts contain various pollutant chemicals such as sulfur dioxide, nitrogen oxides, and carbon compounds. High road density normally indicates intense transportation, so density of roads has been shown to be correlated with high levels of air pollution (Shao, Li, Cao, & Yang, 2016; Zhou et al., 2018). Conversely, a high density of green spaces is thought to be negatively correlated with air pollution. Trees can improve ambient air quality in three respects: first, they can capture atmospheric pollutants such as particulate matter onto leaf surfaces or absorb them into the tree directly; second, trees can dilute the concentration of air pollutants by changing air temperatures and releasing volatile organic compounds; and third, trees can reduce power consumption and consequent pollutant emissions from power plants (Nowak, Hirabayashi, Bodine, & Hoehn, 2013).

1.2 | Methods used to investigate the determinates of air quality

As with many other geographical phenomena, ambient air pollution exhibits an autocorrelated and nonstationary distribution over space (Lu, Xu, Yang, & Zhao, 2017; Wang, Zhou et al., 2017; Zhou et al., 2018). Apart from air pollution itself, the processes which generate air quality may also be heterogeneous, that is to say, air quality is in part a function of location. For example, an impact factor of air quality, such as wind velocity, may not contribute to air pollution to the same extent in all cities. In some places high wind velocity can mitigate pollution by dispersing air pollutants; in places near deserts, however, high wind velocity will exacerbate pollution by carrying dust into air. Such subtle disparities create challenges for modeling air quality.

Various methods have been proposed to examine the underlying mechanisms of air quality. Wang, Zhou, et al. (2017) used cross-sectional linear regression models to estimate the impacts of urban population, urban area, urban second industry share, population density, and GDP per capita on PM2.5 concentrations. Wang, Liu, Zhou, Hu, and Ou (2017) used a two-stage least-squares method to estimate the impacts of socioeconomic factors, urban form, and transportation networks on pollutant emissions in China's four megacities. Hao and Liu (2016) and Zhang, Wang, and Zhang (2016) used non-spatial models (ordinary least squares, OLS) and spatial models (spatial error regression, SER and spatial lag regression, SLR) to determine the directions and strength of the impacts of a range of socioeconomic factors on PM2.5 concentrations in China. They found that non-spatial models were ineffective because they failed to take the spatial effects into account. However, all these models ignored the potential spatial heterogeneity of the relationships between air quality and impact factors. To account for spatial nonstationarity in the determinants of air pollution, geographically weighted regression (GWR) has been used by several researchers. For example, Hu et al. (2013), and Wang and Fang (2016) used GWR to estimate the determinants of the concentration of air pollutants in North America, China, and the Bohai Rim Urban Agglomeration in China, respectively. All their results show that GWR outperforms global regression models, and can detect spatial nonstationarity within the processes.

However, the GWR methods used in previous studies have two major limitations. Firstly, even when potential spatial heterogeneity of the processes generating air pollution is taken into consideration by using GWR, researchers have usually ignored the multiple testing issue so that the significance of the local parameter estimates produced by GWR is questionable (da Silva & Fotheringham, 2016). Secondly, the GWR models produce a single optimized bandwidth for all variables which assumes that all the factors affect air quality at the same spatial scale. This is a questionable assumption given that different processes may affect air quality at different spatial scales.

This study addresses all the issues identified from previous research on air quality. Firstly, we use a recently proposed local model—multi-scale geographically weighted regression (MGWR)—to obtain a set of optimal covariate-specific bandwidths in which each bandwidth indicates the spatial scale at which a particular factor impacts air quality (Fotheringham, Yang, & Kang, 2017). Secondly, we use a newly developed correction method for inference in GWR/MGWR to solve the multiple testing issues, so as to obtain reliable local parameter estimates (da Silva & Fotheringham, 2016; Yu et al., 2019). And thirdly, we include a spatial lag variable in the framework to allow for spatial dependency in the distribution of air quality. We compare the results of calibrating the new model with results from both GWR and OLS models.

The objective of this article is to identify various factors of air quality in China and examine if these impacts are stable over space, and if they are not stable, at what spatial scale do they vary? The empirical case study involves cities in China, a country faced with severe air pollution. One year of air quality data, as well as a set of meteorological data, socio-demographic data, and built environment data, are used.

2 | DATA PREPARATION

2.1 | The study area and definition of air quality

The Chinese government uses the air quality index (AQI), a dimensionless composite index, to measure atmospheric pollution levels (Ministry of Ecology & Environment of the People's Republic of China, 2012). AQI integrates a range of air pollutant measures which are recorded by monitoring stations installed in each city. A higher AQI means a higher level of air pollution and worse air quality (Fang et al., 2015). Air quality is divided into six categories based on the AQI values, as shown in Table 1. Each day the government releases the AQI of each city to provide guidance for citizens' outside activity and whether protective measures are warranted. In this study, we derive the annual mean air quality of each city by averaging its daily AQI recorded by all the monitoring stations that are located in this city across a whole year. Data are available for 231 cities in 2016. The dependent variable is then the annual mean AQI in each city and the variation of average air quality across cities is the subject of interest—why are average AQIs larger in some cities than others?

A map of the spatial distribution of annual mean AQIs in 2016 is shown in Figure 1. Among 231 cities, only 22 cities (9.52%) have good ambient air quality (annual mean AQI ≤ 50), most of which are located in the southeast coastal areas, such as Haikou, Shenzhen, and Fuzhou. 169 cities (73.16%) have moderate ambient air quality (50 < annual mean AQI ≤ 100). The remaining 40 cities (17.32%) have lightly polluted ambient air quality (100 < annual mean AQI ≤ 150), most of which are located in the north, especially the core cities in the Beijing–Tianjin–Hebei (BTH) region, such as Beijing, Tianjin, and Shijiazhuang. The map clearly demonstrates the concentration patterns and spatial heterogeneity of the air quality in China.

2.2 | Independent variables

Ambient air quality has been demonstrated to be affected by various factors. Considering previous research and data availability, we selected 11 factors which could be divided into three categories based on their similarities, as shown in Table 2. The meteorological data are provided by the National Meteorological Information Center (http://data.cma.cn). There are 824 national meteorological stations throughout the country, each recording a

TABLE 1 Six levels of air quality indicated by AQI

AQI values	Air quality condition	Potential impacts on health
AQI ≤ 50	Good	Air quality is satisfactory and there is little to no pollution
51 ≤ AQI ≤ 100	Moderate	Air quality is acceptable, but some pollutants may have a weak impact on a few very sensitive people
101 ≤ AQI ≤ 150	Lightly polluted	The air is unhealthy for sensitive groups and slight irritations may occur
151 ≤ AQI ≤ 200	Moderately polluted	The symptoms of susceptible people are further aggravated, and the heart and respiratory system of healthy people may be affected
201 ≤ AQI ≤ 300	Heavily polluted	Symptoms of patients with heart disease and lung disease are significantly increased, their exercise tolerance is reduced, and symptoms are common in healthy people
AQI ≥ 301	Severely polluted	Exercise tolerance in healthy people is reduced, there are obvious strong symptoms, and some diseases may appear

range of daily meteorological information. We obtained the annual mean meteorological data for each city by averaging the daily meteorological information recorded by those stations located in that city. The socio-demographic factors and the built environment factors of each city are derived from the China City Statistical Yearbook (National Bureau of Statistics of China, 2017). Multicollinearity does not pose a problem, because all the variance inflation factors (VIFs) reported in relation to the variables are below 6.

3 | METHODOLOGIES

We compare the performance of three types of model: classic OLS; the commonly used GWR; and the recently proposed MSGWR. In all three model types we also investigated the effect of adding a spatial lag term to the model to account for spatial dependency in the AQI data.

As the covariates have different units of measurement (e.g., mm, m/s) and large disparities in magnitude (e.g., .01, 14,055,500), we took their logged forms and standardized these logged forms to have mean of 0 and variance of 1, so as to make the parameter estimates independent of units and easy to compare. Logging covariates can also reduce nonlinearities in the original relationships (Fotheringham & Park, 2018).

3.1 | Ordinary least squares

The classic OLS model can be formulated in terms of air quality as

$$\log Y_i = \beta_0 + \sum_j \beta_j \log X_{ij} + \epsilon_i \tag{1}$$

where i denotes a city, Y_i denotes the annual mean AQI of city i, X_{ij} is the jth explanatory variable of city i, β^* are unknown parameters to be estimated which measure the association between air quality and covariates *ceteris paribus*, and ϵ_i is a random error component.

Spatial data are almost always distributed with some degree of positive spatial dependency such that observations near each other tend to be more similar than observations further apart (Anselin & Bera, 1998). This spatial dependency violates the assumption of common regression models that observations are independent of each other, and failing to consider this effect will lead to an overestimation of the significance of estimates (Clifford & Richardson, 1985). One common method to address this issue is to include a lagged form of the dependent variable in the regression model as another independent variable. The model is then termed "spatial lag regression" (Anselin & Bera, 1998). The spatially lagged term for the dependent variable is calculated as:

$$\mathrm{lag}(Y_i) = \sum_{k=1, k \neq i}^{n} \frac{Y_k}{d_{ik}} \tag{2}$$

where $\mathrm{lag}(Y_i)$ is the spatially lagged value of Y_i for city i, k is one of the other cities, n is the number of all the cities, and d_{ik} is the distance between city i and city k, which is limited to a threshold t (500 km in our study). Adding the variable to the OLS model give us an OLS model with spatially lagged dependent variable (OLSL).

3.2 | Geographically weighted regression

GWR extends global regression modeling by allowing local rather than global parameters to be estimated (Fotheringham, Brunsdon, & Charlton, 2002), and is formulated as:

$$\log Y_i = \beta_0(u_i, v_i) + \sum_j \beta_j(u_i, v_i) \log X_{ij} + \epsilon_i \tag{3}$$

TABLE 2 Description statistics of variables

Category	Variable	Definition	Unit	Mean	*SD*	Min	Max	VIF[a]
Dependent variable	AQI	Annual mean of daily AQI	–	76.54	20.52	33.15	130.36	–
Meteorological factors	PRE	Annual mean daily precipitation	mm	3.37	1.88	.28	7.61	1.42
	WIND	Annual mean daily wind velocity	m/s	22.01	6.99	9.47	53.22	1.31
Socio-demographic factors	POPT	Total population	10,000 persons	473.39	334.92	20.25	3,371.84	2.84
	POPD	Population density	persons/km^2	457.69	343.55	10.25	2,501.14	4.17
	POPU	Proportion of urban population	%	36.61	23.44	4.68	100	3.42
	PCGRP	Per capita GRP	yuan	52,141.48	30,215.42	10,987	207,163	3.29
	SEGRP	Secondary industry as a percentage of GRP	%	47.03	8.94	19.74	71.45	1.49
	ELEC	Annual electricity consumption	10,000 kWh	1,069,372	1,709,767	24,800	14,055,500	5.41
	DUST	Volume of industrial soot (dust) emission	ton	52,577.19	153,803.2	935	1,859,866	1.59
Built environment factors	ROAD	Area of city paved roads as a percentage of urban area	%	1.22	1.24	.01	6.88	2.88
	GREEN	Green-covered area as a percentage of urban area	%	39.09	6.74	2.71	51.44	1.38
Spatial lag	LAG[b]	Spatial lag variable	–	1,134.67	664.19	28.46	2,629.47	2.15

[a]The VIFs are reported based on the logged forms of the variables.
[b]This variable will be explained in detail later.

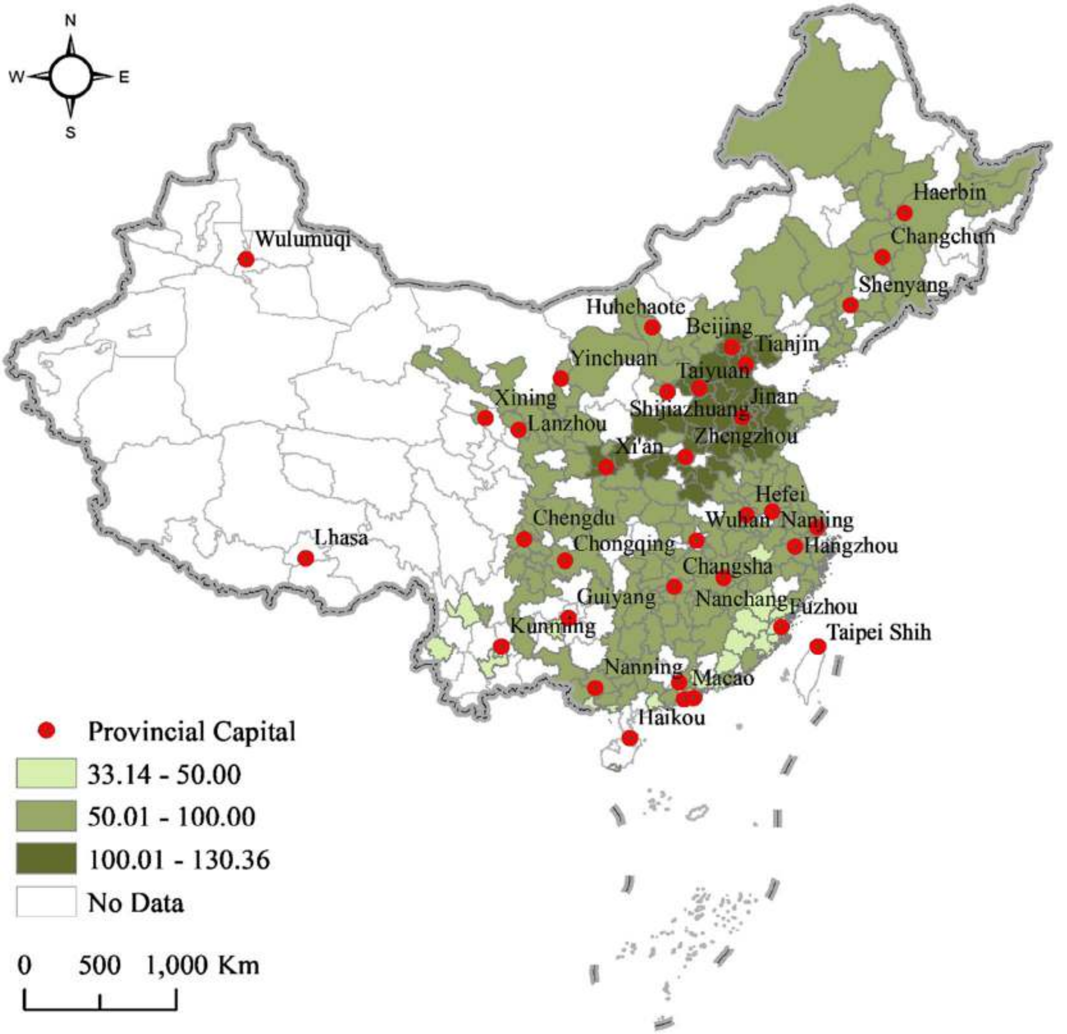


FIGURE 1 Spatial distribution of annual mean AQI in eastern China in 2016

where (u_i,v_i) represents the centroid coordinates of city i. The location-specific parameter estimates allow the relationship between covariates and air quality to vary between cities. The estimation is realized through "borrowing" data from nearby locations using specific distance-weighting functions such that data near the regression point are assigned larger weights than data farther away (Fotheringham et al., 2002). In this study we used the commonly used adaptive bi-square kernel function as the distance-weighting function:

$$w_{ij} = \begin{cases} \left[1-\left(\frac{d_{ij}}{b}\right)^2\right]^2 & \text{if } d_{ij} < b \\ 0 \text{ otherwise} & \end{cases} \tag{4}$$

where w_{ij} is the weight between city i and city j, d_{ij} is the distance between city i and city j, and b is a critical distance from regression location i to its Mth nearest neighbor. M is the optimal number of nearest neighbors, determined by minimizing the corrected Akaike information criterion (Fotheringham et al., 2002).

A second GWR model is formulated by adding the same spatial lag variable defined in Equation (2) to the model's covariates. This model is termed geographically weighted regression with lagged dependent variable (GWRL).

3.3 | Multi-scale geographically weighted regression

Although GWR captures any spatial heterogeneity in relationships, it does so under the assumption that all such relationships vary at the same spatial scale across all covariates. MGWR is a significant improvement on GWR because it relaxes the "same spatial scale" assumption and allows covariate-specific bandwidths to be optimized. It is formulated as (Fotheringham et al., 2017):

$$\log Y_i = \beta_{bw0}(u_i, v_i) + \sum_j \beta_{bwj}(u_i, v_i) \log X_{ij} + \epsilon_i \tag{5}$$

where bw_* is the specific optimal bandwidth used in the calibration of the *th conditional relationship. MGWR thus allows different processes to operate at different spatial scales by deriving separate bandwidths for the conditional relationships between the response variable and different predictor variables.

MGWR is calibrated using a back-fitting algorithm as described in Fotheringham et al. (2017). The back-fitting process is initialized with GWR parameter estimates. Based on these initial values, the calibration process works in an iterative manner and during each iteration, all local parameter estimates and optimal bandwidths are evaluated. Iteration terminates when the difference between the parameter estimates from successive iterations converges to a specified threshold (we selected 1e-5 in this study). The details of the process can be found in Fotheringham et al. (2017) and Oshan, Li, Kang, Wolf, and Fotheringham (2019).

Again, we created a separate MGWR model by adding the spatial lag term in Equation (2) to the covariates and term this model multiscale GWR with lagged dependent variable (MGWRL). We used the MGWR1.0 software to undertake all calibrations (https://sgsup.asu.edu/sparc/mgwr).

4 | RESULTS AND DISCUSSION

We compare the performance of six models of air quality across Chinese cities based on several criteria. These models are:

1. OLS
2. OLS with spatially lagged dependent variable (OLSL)
3. GWR
4. GWR with spatially lagged dependent variable (GWRL)
5. MGWR
6. MGWR with spatially lagged dependent variable (MGWRL).

4.1 | Model comparison regarding goodness-of-fit

The six models were compared in terms of their ability to replicate the observed air quality indices using four metrics: residual sum of squares (RSS); mean absolute error (MAE); corrected Akaike information criterion (AICc); and adjusted R-squared value (R^2). The results are shown in Table 3. For RSS, MAE, and AICc, lower values indicate better replication of the known air quality indices across the 231 locations; for R^2, higher values indicate better model fit. All four goodness-of-fit measures suggest the following.

TABLE 3 Performance of six models

	RSS	MAE	AICc	R^2
OLS	79.85	.46	438.85	.65
OLSL	48.50	.37	325.97	.79
GWR	23.45	.26	340.40	.89
GWRL	21.79	.25	273.71	.91
MGWR	18.35	.22	237.99	.92
MGWRL	18.15	.22	198.09	.93

1. MGWR outperforms GWR and GWR outperforms OLS, even accounting for the extra parameters in the local models. At least some of the relationships being modeled vary spatially and this nonstationarity is modeled by both GWR and MGWR but not by OLS. The MGWR models outperform their equivalent GWR versions by allowing covariate-specific bandwidths to be optimized rather than producing a single average bandwidth applied to all relationships.
2. Adding the spatial lag variable increases the goodness-of-fit of all three models although the differences are very small for MGWR, indicating that this model form appears to account for any spatial dependency in the error terms better than GWR or OLS.
3. MGWR without a spatial lag outperforms both the OLS and GWR models with spatial lag terms, again suggesting that MGWR alone without a spatial lag can account for spatial error dependencies. We return to this point below.

4.2 | Model comparison regarding residuals

The aspatial distributions of the residuals for the six models are displayed using a box-and-whisker plot in Figure 2 and confirm the conclusions drawn from the discussion of goodness-of-fit measures above. The MGWR models produce the lowest residuals and the two global models produce the highest residuals. The addition of the spatial lag variable reduces error variance considerably in the global model but less obviously in the two local models.

However, a more important criterion of the residuals from each of the six models is the degree to which they are spatially autocorrelated. A tenet of regression is that the residuals should be independent of each other and therefore randomly distributed in space. The spatial autocorrelations of the six sets of residuals—as measured by Moran's *I*—are shown in Table 4, where it can be seen that only the MGWR models produce residuals which exhibit no spatial autocorrelation. Both OLS models exhibit the strongest levels of positive autocorrelation and although this level is severely reduced in both equivalent GWR models, the level of spatial autocorrelation is still significant. Consequently, although GWR does reduce the problem of spatially autocorrelated residuals found in OLS models, the restriction of a single bandwidth does not remove the problem completely. Only the residuals from the two MGWR models have no significant spatial pattern. Interestingly, although the addition of the spatial lag term reduces the problem of spatially autocorrelated residuals in both OLS and GWR, it does not eliminate it completely, as happens in MGWR. The spatial patterning of the residuals from all six models is shown in Figure 3.

As can be seen from Figure 3, the spatial clustering patterns of the residuals of both OLS and OLSL are obvious: positive residuals are clustered in center areas while negative residuals are mostly found in peripheral areas. The residuals from GWR and GWRL are relatively scattered, but some clusters can also be found, such as clusters of positive residuals in the BTH region and clusters of negative residuals in the Shandong Peninsula. The residuals from MGWR and MGWRL, however, show random patterns and there are no obvious clusters. Consequently, we conclude that global models are not effective at reducing residual dependency in this case, even when a spatial

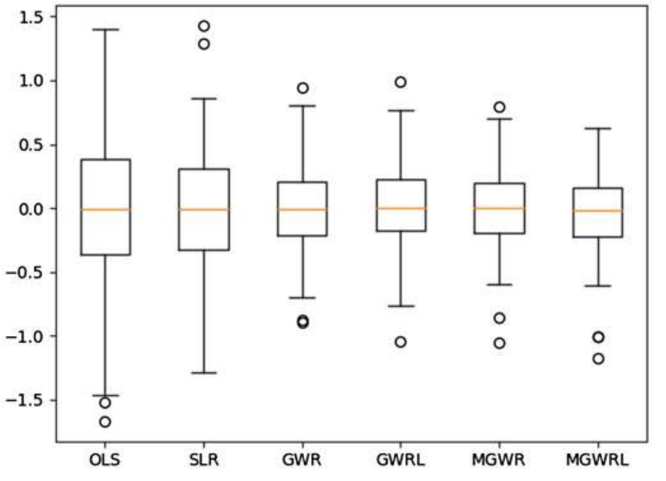


FIGURE 2 Box plots of residuals

TABLE 4 Spatial autocorrelation of residuals

	Moran's *I* index	Expected index	*Z* score	*p* Value	Pattern
OLS	0.286	−0.004	16.377	<.001	Clustered
OLSL	0.136	−0.004	8.184	<.001	Clustered
GWR	0.059	−0.004	3.763	<.001	Clustered
GWRL	0.034	−0.004	2.233	.026	Clustered
MGWR	−0.007	−0.004	−0.178	.859	Random
MGWRL	−0.014	−0.004	−0.554	.580	Random

lag is included. GWR can reduce but cannot eliminate residual dependency; only MGWR effectively eliminates residual dependency.

4.3 | Model comparison regarding optimized bandwidths and scale

The most significant improvement of MGWR models is that they not only allow the parameter estimates to vary over space, but also produce individual optimal bandwidths for the conditional relationships between the response variable and each predictor variable, which allows the spatial variation of different processes to be modeled at different spatial scales. As all the variables are standardized to have mean = 0 and standard deviation = 1, the optimal bandwidths deduced by GWR and MGWR models are direct indicators of the spatial scale at which the individual conditional relationship between AQI and each covariate varies (Fotheringham et al., 2017).

The blue histograms in Figure 4 indicate the optimal bandwidths for each covariate generated by the MGWR model; orange histograms indicate the standard deviations of parameter estimates. The solid and dotted black horizontal lines, respectively, denote the single optimal bandwidth obtained by GWR and the average of the 12 bandwidths obtained by MGWR. It is obvious that processes modeled by MGWR operate at different spatial scales. A variable with a large bandwidth affects the dependent variable at a large scale; in other words, the influence is similar across space with small heterogeneity, so the standard deviation of parameter estimates is small. In contrast, a variable with a small bandwidth affects the dependent variable at a local scale, so the standard deviation of the local parameter estimates is large. From this point of view, with the optimal bandwidth of PRE (precipitation), intercept, POPD (population density), and GREEN (green-covered area) being 44, 44, 48, and

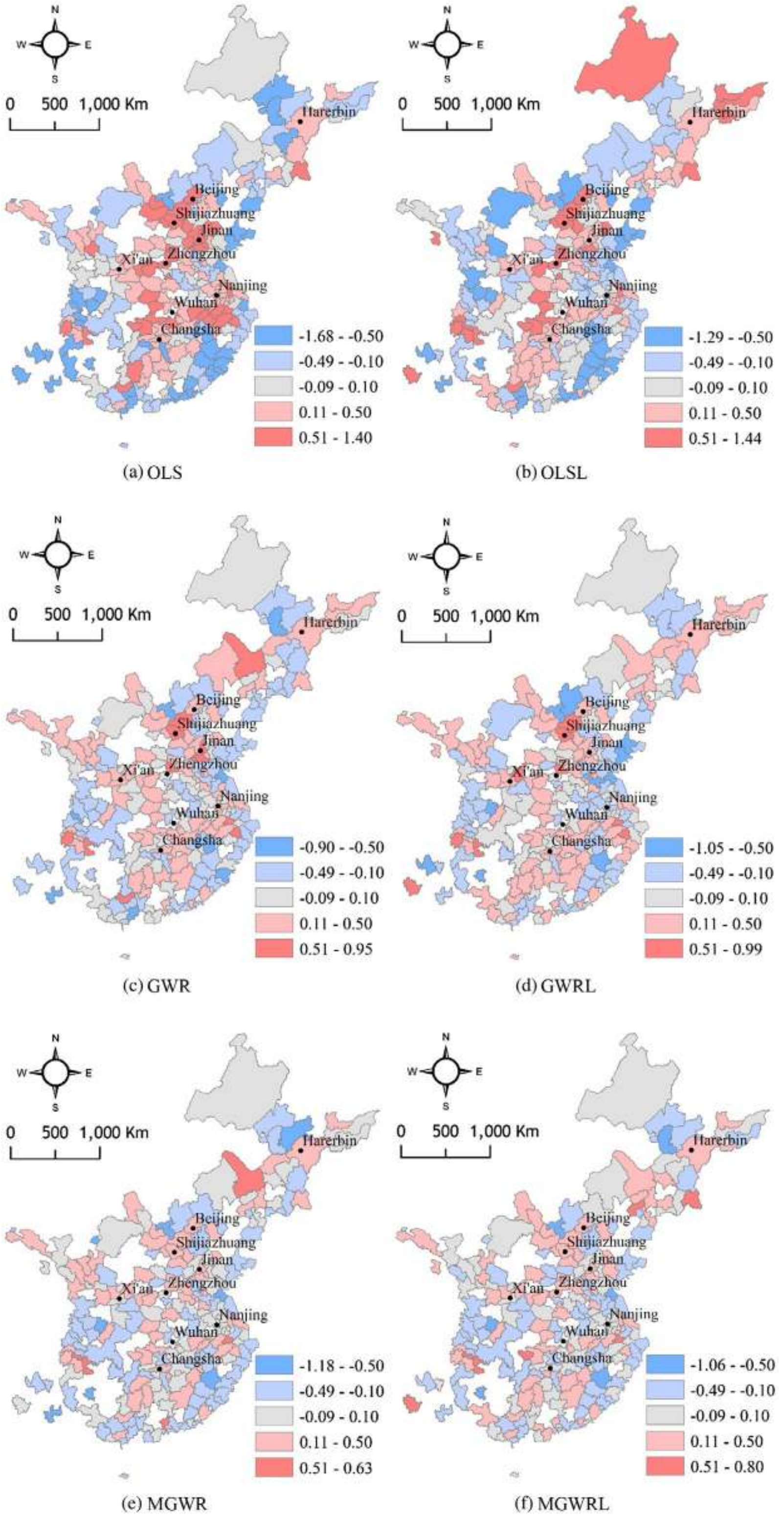


FIGURE 3 Spatial distribution of residuals (cities with no data are not shown)

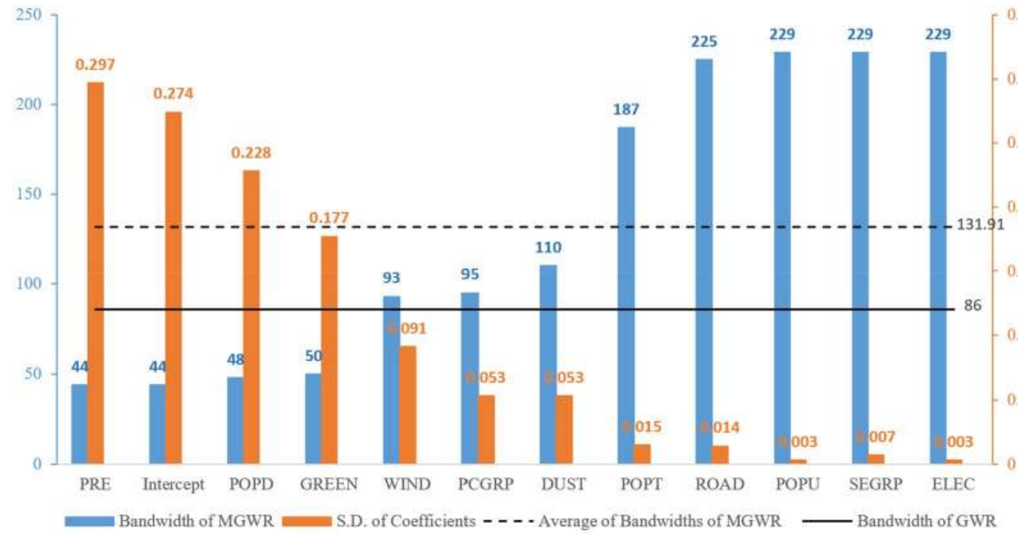


FIGURE 4 Optimal bandwidths generated by MGWR and GWR and standard deviations of parameter estimates of MGWR

50 nearest neighbors, respectively, these three variables affect air quality at relatively local scales, and their parameter estimates have relatively large differences over space. The relationships between WIND (wind velocity), PCGRP (per capita GRP), DUST (dust emission), and air quality exhibit spatial nonstationarity but the processes vary at broad regional scales, with the optimal bandwidth being 93, 95, and 110 nearest neighbors, respectively. The other variables affect air quality at global scales, as their optimal bandwidths are close to the maximum possible number of neighbors which is 230.

GWR, however, generates a single optimal bandwidth of 86 nearest neighbors for all the variables, which approximates the average of the 11 optimal bandwidths obtained by MGWR (131.91). This single bandwidth assumes that all the variables affect air quality at the same regional scale, which seems highly restrictive. The bandwidth produced by GWR can be understood as a weighted average of the different degrees of spatial heterogeneity of the 11 separate processes, with the weighting being a function of the explanatory ability of each relationship in the local model (Fotheringham et al., 2017).

4.4 | Model comparison regarding parameter estimates

Global models generate only one parameter estimate for each covariate at all locations, assuming the relationship between this covariate and the dependent variable is stationary over space. In order to acquire a comprehensive understanding of the influences of various factors on AQI, we list the parameter estimates associated with each covariate in OLS and OLSL. However, local models generate individual parameter estimates for each variable at each location. The abundant information generated by local models presents a challenge for displaying the results. As MGWR models are more effective than GWR models and MGWR achieves similar performance to MGWRL, we only focus on the local estimates of MGWR. Summary statistics for the parameter estimates from GWR, GWRL, and MGWRL are listed in the Appendix.

4.4.1 | Results from OLS and OLSL

The parameter estimates generated by OLS and OLSL are listed in Table 5. The coefficients of the meteorological variables indicate that PRE (precipitation) and WIND (wind velocity) are significantly negatively associated with

TABLE 5 Parameter estimates for the regression of air quality generated by OLS and OLSL

Variables	OLS	OLSL
Intercept	.000	.000
PRE	−.702***	−.681***
WIND	−.112*	−.071*
POPT	.098	.076
POPD	.471***	.100
POPU	.002	.071
PCGRP	−.035	−.040
SEGRP	.140**	.055
ELEC	−.148	−.062
DUST	.219***	.143***
ROAD	.128	.141**
GREEN	.058	.011
LAG	–	.541***

Note: Significant at
*.05; **.01; ***.001 level.

AQI. Both models generate positive but nonsignificant coefficients for three population variables—POPT (total population), POPD (population density), and POPU (urban population). One exception is that the OLS parameter estimate for POPD (population density) is statistically significant. PCGRP (per capita GRP) and ELEC (electricity consumption) have negative but nonsignificant influences on AQI. OLS generates a significant positive coefficient for SEGRP (secondary industry), while the coefficient of SEGRP (secondary industry) generated by OLSL is not significant. DUST (dust emission) appears to have a positive effect on AQI, as indicated by both the OLS and OLSL calibrations. The results of OLS and OLSL suggest the positive influence of ROAD (ratio of roads) on AQI, although this influence is not significant in OLS. Both models generate positive but nonsignificant parameter estimates for GREEN (green-covered area). The significant positive coefficient of LAG generated by OLSL indicates the existence of spatial dependency—a city's air quality is not independent of the air quality in nearby cities.

4.4.2 | Results from MGWR

Unlike global models, MGWR generates local parameter estimates which reflect possible spatial heterogeneity in the processes affecting air quality. Table 6 lists summary statistics of the local parameter estimates generated by MGWR and they are displayed in full in Figure 5. The second column of Table 6 shows the minimum (min), maximum (max), and mean (mean) values of the local parameter estimates for each covariate; the third column indicates a classification of coefficients based on *t* tests, adjusted for multiple hypothesis testing (da Silva & Fotheringham, 2016), including the proportion of significant coefficients ($p \le .05$), the proportion of significant positive coefficients to significant coefficients (+), and the proportion of significant negative coefficients to significant coefficients (−).

The results show that the local intercepts are significantly different from zero for 65.8% of cities and that 98.03% of these are positive while 1.97% are negative. The former indicates an elevated rate of air pollution even accounting for the variables in the model, while the latter indicates a reduced level of air pollution given the conditions existing in these cities. The local parameter estimates from PRE (precipitation) are significant for two-thirds of the cities and in each case higher precipitation levels are associated with low air pollution. The local parameter estimates for WIND (wind velocity) are significant in just over 20% of the cities and in each case WIND (wind

TABLE 6 Parameter estimates for the regression of air quality using MGWR

Variables	MGWR coefficients			Percentage of cities by significance (95% level) of *t* test		
	Min	Max	Mean	*p* ≤ .05 (%)	+ (%)	− (%)
Intercept	−.293	.846	.353	65.80	98.03	1.97
PRE	−.999	.088	−.438	66.67	.00	100.00
WIND	−.249	.109	−.053	20.78	.00	100.00
POPT	−.008	.054	.010	.00	.00	.00
POPD	−.109	.799	.264	49.78	100.00	.00
POPU	−.018	−.002	−.009	.00	.00	.00
PCGRP	−.131	.084	−.045	.00	.00	.00
SEGRP	.124	.155	.136	100.00	100.00	.00
ELEC	.049	.058	.053	.00	.00	.00
DUST	−.048	.151	.048	6.06	100.00	.00
ROAD	.041	.086	.065	.00	.00	.00
GREEN	−.521	.406	.002	5.63	.00	100.00

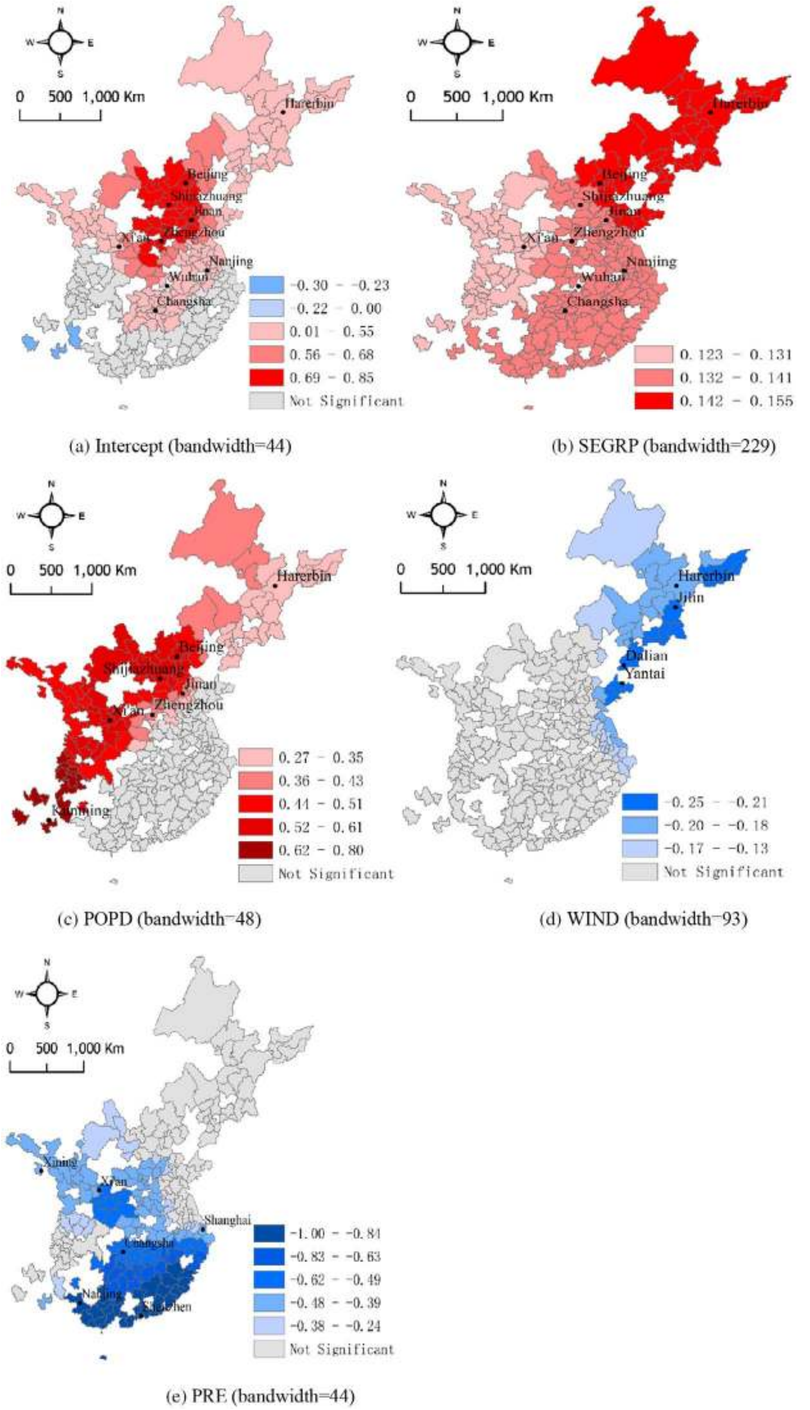
velocity) associates negatively with the air pollution index. The local parameter estimates from POPD (population density) are significant for half of the cities and in each case higher population densities are associated with high air pollution. The local estimates for SEGRP (secondary industry) are significantly positive for every city, indicating a robust and relatively uniform relationship between a city's secondary industry development and air pollution levels. The local parameter estimates for POPT (total population), POPU (urban population), PCGRP (per capita GRP), ELEC (electricity consumption), and ROAD (ratio of roads) are all insignificant, presenting fairly strong evidence that these variables have little impact on air quality. Two other variables, DUST (dust emission) and GREEN (green-covered area), have very limited impact in a handful of cities. Where they are significant, increasing values of DUST (dust emission) generate poorer air quality whereas increasing levels of GREEN (green-covered area) generate better levels of air quality. The spatial patterns of the locally significant parameter estimates are discussed below.

Local estimates of the intercept

We can see from Figure 5a that the parameter estimates for the intercept are significantly positive in the north of the country focused on the BTH region and a very small number of cities in the southwest have significant negative parameter estimates. The estimates of the local intercepts are of interest because they indicate elevated levels of pollution (significantly positive) or lower levels of pollution (significantly negative), given the covariates in these cities. Elevated levels of air pollution in the BTH region may be caused by the people living in northern China using more dirty fuels such as coal for heating than people in southern cities, where the climate is not as harsh. In addition, in rural areas surrounding some developed regions such as Beijing, the use of household biofuel (e.g., crop residuals and wood) for heating and cooking, plus straw burning by farmers to increase the productivity of fields, all might increase ambient air pollutants (Zhang & Cao, 2015).

Local estimates of the parameters associated with SEGRP (secondary industry)

The local parameter estimates of SEGRP in Figure 5b are significantly positive across the whole study area and are very uniform, ranging from .123 to .155 with a standard deviation of .007. The optimized bandwidth of SEGRP is 229 (see Figure 4), indicating that SEGRP affects air quality at a global scale; in other words, SEGRP has a similar influence on air quality in all cities.



(a) Intercept (bandwidth=44)

(b) SEGRP (bandwidth=229)

(c) POPD (bandwidth=48)

(d) WIND (bandwidth=93)

(e) PRE (bandwidth=44)

FIGURE 5 The spatial distribution of MGWR local coefficients (cities with no data are not shown)

As a frequently used proxy for industrial structure, the contribution of secondary industry to GRP in China in 2016 was 39.8%, which is much larger than that of developed economies such as America (18.88%), the European Union (24.5%), and Japan (26.8%) (National Bureau of Statistics of China, 2017). Secondary industries such as steel production, automobile manufacturing, and chemical production are typically heavy polluters of the atmosphere and where these industries are concentrated, air pollution is generally worse. The northeast region contains the heavy industry base of China and economic development here depends heavily on the consumption of coal and crude oil. Vast quantities of energy consumption lead to emissions of air pollutants such as sulfur dioxide and carbon compounds, which explains why SEGRP is significantly positively related to AQI in every city, but especially so in northeast China. The negative impact of secondary industry on air quality in China has been widely demonstrated in previous studies (Cheng, Li, & Liu, 2017; Lu et al., 2017; Wang, Zhou et al., 2017; Zhou et al., 2018).

Local estimates of the parameter estimates associated with POPD (population density)

Figure 5c shows the spatial variation of the local parameter estimates associated with the variable POPD. The standard deviation of POPD's coefficients is .228, which is large; the optimized bandwidth of POPD is 48, which is small (see Figure 4). Both indicate that POPD affects air quality at a very local scale.

Previous studies have suggested a negative relationship between population density and air quality (Zhou et al., 2018). Here we support these findings but only in the northwestern part of the study region; there is no significant association between population density and air quality in the southeast part of the country. These differences can be understood from two perspectives. First, cities in the north and northeast have rich coal reserves and residents of these areas are more dependent on coal because it is much cheaper than cleaner energies such as natural gas (Hao, Liu, Weng, & Gao, 2016). In addition, the north and west regions are colder so that people need to burn coal for heat, especially in the winter. Coal is a dirty energy because various air pollutants (such as sulfur dioxide, nitrogen oxides, and fine particles) are directly emitted into the atmosphere during the process of burning (Song et al., 2007). Cities which experience cold winters and have large populations burn more coal and air quality suffers. Second, cities in east and south China are more developed than cities in the west. These cities with advanced economies demonstrate no obvious relationship between population density and air quality, which agrees with the ecological modernization theory and environmental Kuznets curve theory. Both theories advocate that environment problems first increase from low to medium development stages, then decrease with further development (Stern, 2004). A similar pattern was found in a recent study: increased population density promotes CO_2 emissions in less developed areas but reduces CO_2 emissions in more developed areas (Liu, Gao, & Lu, 2017). People in developed areas are more environmentally conscious and the governments have a greater ability to pay for environment protection (Wang & Fang, 2016; Zhou et al., 2018).

Local estimates of the parameters associated with WIND (wind velocity)

The spatial distribution of the local parameter estimates associated with WIND is presented in Figure 5d. The optimized bandwidth of these local estimates suggests that WIND affects air quality at a regional scale and is clearly only of importance in affecting air quality levels in the northeastern part of the region, and particularly in coastal cities. High winds are thought to benefit air quality because they disperse and dilute air pollutants (Dawson et al., 2007; Han, Zhou, & Li, 2016; Hu et al., 2013; Kleeman, 2007; Li et al., 2014; Lu et al., 2017). We confirm the negative relationship between wind speeds and air pollution levels in the northeastern and eastern coastal areas but elsewhere the relationship is not significant. Cities with stronger winds are mostly found in north and northeast China, as well as eastern coastal districts (He & Kammen, 2014). In the northwest, the beneficial effects of wind dispersion may be offset by sand and dust particles being mixed into the atmosphere from the Gobi Desert (Vardoulakis & Kassomenos, 2008).

Local estimates of the parameter estimates associated with PRE (precipitation)

Figure 5e presents the spatial distribution of parameter estimates of PRE. The optimized bandwidth of PRE is small (44), suggesting that PRE has a local effect on air quality. PRE is significantly negatively associated with AQI, primarily in the southeast. This pattern shares some similarities with the typical spatial distribution of rainfall amounts in China: some parts of south China are located in tropical and subtropical zones, weather in these places is wet and humid because the frequent southeastern winds carry moisture from the Pacific Ocean. These districts receive the greatest amount of rainfall in China. Previous studies have demonstrated that rain can "wash off" the air pollutants and purify ambient air (Buchholz, Junk, Krein, Heinemann, & Hoffmann, 2010; Li et al., 2014; Vardoulakis & Kassomenos, 2008). This explains the strong negative relationship between precipitation and AQI in these areas, especially in the coastal cities. The influence of precipitation weakens towards the northwest inland because the amount of moisture in the air reduces gradually and rainfall becomes less. Other cities demonstrate no significant relationship between precipitation and AQI, probably because rainfall amounts are much less. It would appear that rainfall amounts have to be quite large to have a major impact on reducing air pollution levels.

5 | CONCLUSIONS

Spatial data may result for spatially nonstationary processes (i.e., the processes generating the associations between variables may not be constant over space, as is traditionally assumed). Additionally, the scale at which each covariate impacts a dependent variable may vary across covariates: some variation in relationships might vary at a very local scale whilst others may vary at a more regional scale and some may be invariant to location. Therefore, to improve our understanding of geographical processes, both spatial heterogeneity and scale differences regarding processes should be taken into consideration when conducting spatial analysis and modeling. This study considers both these issues in evaluating the influences on air quality in China using a recently proposed model—MGWR. The results suggest that MGWR provides more reliable information on the processes affecting air quality than either OLS or GWR: the model not only achieves higher goodness-of-fit but also performs better at alleviating residual autocorrelation.

The most significant contribution of this study is the determination of respective bandwidths for each covariate, which indicates the scale at which the association between that covariate and air quality varies over space. This is achieved in MGWR by relaxing the single-bandwidth assumption of the traditional GWR model, allowing covariate-specific bandwidths to be optimized. The results suggest the existence of scale differences between the processes affecting air quality. For example, secondary industries as a contributor to GRP affect the air quality at a global scale (i.e., the local parameter estimates associated with this variable are similar across space, whereas the effect of precipitation on air quality varies locally). Traditional GWR models, which fail to distinguish between such scale differences, should be replaced by MGWR as demonstrated in Figure 4.

Finally, MGWR seems to account for spatial dependency in the error terms better than GWR and OLS. Comparing OLS, GWR, and MGWR with and without a spatial lag term suggests that while both OLS and GWR benefit from the addition of the lag variable, MGWR is probably preferred to MGWR with a lag because the residual autocorrelation is virtually the same and the former does not suffer from the lag variable producing bias in the remaining parameter estimates.

In summary, increasing attention is being given to air quality problems in China because air pollution levels are causing problems to human health. Spatial statistical models provide useful methods for understanding the determinants of air pollution by taking spatial dependency and spatial heterogeneity into consideration. This study demonstrates the superiority of a new approach—MGWR—in accounting for these two spatial effects, as well as the scale differences between the impacts of various covariates.

ACKNOWLEDGMENTS

We acknowledge financial support from the US National Science Foundation (Grant No. 1758786) and the China Scholarship Council (Grant No. 201806270199).

ORCID

Han Yue https://orcid.org/0000-0002-5869-7424

REFERENCES

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. E. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237–289). New York, NY: Marcel Dekker.

Brauer, M., Amann, M., Burnett, R. T., Cohen, A., Dentener, F., Ezzati, M., ... Thurston, G. D. (2012). Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. *Environmental Science & Technology*, *46*(2), 652–660.

Buchholz, S., Junk, J., Krein, A., Heinemann, G., & Hoffmann, L. (2010). Air pollution characteristics associated with mesoscale atmospheric patterns in northwest continental Europe. *Atmospheric Environment*, *44*(39), 5183–5190.

Cheng, Z., Li, L., & Liu, J. (2017). Identifying the spatial effects and driving factors of urban PM2.5 pollution in China. *Ecological Indicators*, *82*, 61–75.

Clifford, P., & Richardson, S. (1985). Testing the association between two spatial processes. *Statistics & Decisions*, *2*(Suppl.), 155–160.

Crane, K., & Mao, Z. (2015). *Costs of selected policies to address air pollution in China*. Santa Monica, CA: RAND Corporation.

da Silva, A. R., & Fotheringham, A. S. (2016). The multiple testing issue in geographically weighted regression. *Geographical Analysis*, *48*(3), 233–247.

Dawson, J. P., Adams, P. J., & Pandis, S. N. (2007). Sensitivity of PM2.5 to climate in the eastern US: A modeling case study. *Atmospheric Chemistry & Physics*, *7*(16), 4295–4309.

Fang, C., Liu, H., Li, G., Sun, D., & Miao, Z. (2015). Estimating the impact of urbanization on air quality in China using spatial regression models. *Sustainability*, *7*(11), 15570–15592.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Hoboken, NJ: John Wiley & Sons.

Fotheringham, A. S., & Park, B. (2018). Localized spatiotemporal effects in the determinants of property prices: A case study of Seoul. *Applied Spatial Analysis & Policy*, *11*(3), 581–598.

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247–1265.

Han, L., Zhou, W., & Li, W. (2016). Fine particulate (PM2.5) dynamics during rapid urbanization in Beijing, 1973–2013. *Scientific Reports*, *6*, 23604.

Hao, Y., & Liu, Y. (2016). The influential factors of urban PM2.5 concentrations in China: A spatial econometric analysis. *Journal of Cleaner Production*, *112*, 1443–1453.

Hao, Y., Liu, Y., Weng, J.-H., & Gao, Y. (2016). Does the environmental Kuznets curve for coal consumption in China exist? New evidence from spatial econometric analysis. *Energy*, *114*, 1214–1223.

He, G., & Kammen, D. M. (2014). Where, when and how much wind is available? A provincial scale wind resource assessment for China. *Energy Policy*, *74*, 116–122.

Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G., Estes, S. M., ... Liu, Y. (2013). Estimating ground-level PM2.5 concentrations in the southeastern U.S. using geographically weighted regression. *Environmental Research*, *121*, 1–10.

Kleeman, M. J. (2007). A preliminary assessment of the sensitivity of air quality in California to global change. *Climatic Change*, *87*(Suppl. 1), S273–S292.

Li, L., Qian, J., Ou, C., Zhou, Y., Guo, C., & Guo, Y. (2014). Spatial and temporal analysis of air pollution index and its timescale-dependent relationship with meteorological factors in Guangzhou, China, 2001–2011. *Environmental Pollution*, *190*, 75–81.

Liu, Y., Gao, C., & Lu, Y. (2017). The impact of urbanization on GHG emissions in China: The role of population density. *Journal of Cleaner Production*, *157*, 299–309.

Lu, D., Xu, J., Yang, D., & Zhao, J. (2017). Spatio-temporal variation and influence factors of PM2.5 concentrations in China from 1998 to 2014. *Atmospheric Pollution Research*, *8*(6), 1151–1159.

Matus, K., Nam, K. M., Selin, N. E., Lamsal, L. N., Reilly, J. M., & Paltsev, S. (2012). Health damages from air pollution in China. *Global Environmental Change*, *22*(1), 55–66.

Ministry of Ecology & Environment of the People's Republic of China. (2012). *Technical regulation on ambient air quality index (on trial)* (Vol. HJ 633). Beijing, China: Author.

National Bureau of Statistics of China. (2017). *China statistical yearbook*. Beijing, China: China Statistics Press.

Nowak, D. J., Hirabayashi, S., Bodine, A., & Hoehn, R. (2013). Modeled PM2.5 removal by trees in ten U.S. cities and associated health effects. *Environmental Pollution*, *178*, 395–402.

Oshan, T., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). MGWR: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, *8*, 269.

Pope III, C. A. (2000). Review: Epidemiological basis for particulate air pollution health standards. *Aerosol Science & Technology*, *32*(1), 4–14.

Remoundaki, E., Papayannis, A., Kassomenos, P., Mantas, E., Kokkalis, P., & Tsezos, M. (2012). Influence of Saharan dust transport events on PM2.5 concentrations and composition over Athens. *Water, Air, & Soil Pollution*, *224*(1), 1373.

Shao, S., Li, X., Cao, J., & Yang, L. (2016). China's economic policy choices for governing smog pollution based on spatial spillover effects. *Economic Research Journal*, *9*(9), 73–88.

Song, Y., Tang, X., Xie, S., Zhang, Y., Wei, Y., Zhang, M., … Lu, S. (2007). Source apportionment of PM2.5 in Beijing in 2004. *Journal of Hazardous Materials*, *146*(1), 124–130.

Stern, D. I. (2004). The rise and fall of the environmental Kuznets curve. *World Development*, *32*(8), 1419–1439.

Vardoulakis, S., & Kassomenos, P. (2008). Sources and factors affecting PM10 levels in two European cities: Implications for local air quality management. *Atmospheric Environment*, *42*(17), 3949–3963.

Wang, S., Fang, C., & Wang, Y. (2016). Spatiotemporal variations of energy-related CO_2 emissions in China and its influencing factors: An empirical analysis based on provincial panel data. *Renewable & Sustainable Energy Reviews*, *55*, 505–515.

Wang, S., Liu, X., Zhou, C., Hu, J., & Ou, J. (2017). Examining the impacts of socioeconomic factors, urban form, and transportation networks on CO_2 emissions in China's megacities. *Applied Energy*, *185*, 189–200.

Wang, S., Zhou, C., Wang, Z., Feng, K., & Hubacek, K. (2017). The characteristics and drivers of fine particulate matter (PM2.5) distribution in China. *Journal of Cleaner Production*, *142*, 1800–1809.

Wang, Z., & Fang, C. (2016). Spatial–temporal characteristics and determinants of PM2.5 in the Bohai Rim urban agglomeration. *Chemosphere*, *148*, 148–162.

Yu, C. (2017). *China electric power yearbook*. Beijing, China: China Electric Power Press.

Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., & Wolf, L. J. (2019). Inference in multiscale geographically weighted regression. *Geographical Analysis*.

Zhang, H., Wang, Z., & Zhang, W. (2016). Exploring spatiotemporal patterns of PM2.5 in China based on ground-level observations for 190 cities. *Environmental Pollution*, *216*, 559–567.

Zhang, Y., & Cao, F. (2015). Is it time to tackle PM2.5 air pollutions in China from biomass burning emissions? *Environmental Pollution*, *202*, 217–219.

Zhou, C., Chen, J., & Wang, S. (2018). Examining the effects of socioeconomic development on fine particulate matter (PM2.5) in China's cities using spatial regression and the geographical detector technique. *Science of the Total Environment*, *619–620*, 436–445.

How to cite this article: Fotheringham AS, Yue H, Li Z. Examining the influences of air quality in China's cities using multi-scale geographically weighted regression. *Transactions in GIS*. 2019;23:1444–1464. https://doi.org/10.1111/tgis.12580

APPENDIX

TABLE A1 Parameter estimates for the regression of air quality using GWR

	GWR coefficients			Percentage of cities by significance (95% level) of *t* test		
Variables	Min	Max	Mean	*p* ≤ .05 (%)	+ (%)	− (%)
Intercept	−.727	1.150	.414	73.16	89.94	10.06
PRE	−1.565	−.059	−.559	74.03	.00	100.00
WIND	−.360	.201	−.078	30.30	18.57	81.43
POPT	−.267	.382	−.002	8.66	100.00	.00
POPD	−.269	.681	.303	59.74	100.00	.00
POPU	−.295	.319	−.002	7.79	100.00	.00
PCGRP	−.478	.365	−.069	13.42	6.45	93.55
SEGRP	−.047	.297	.134	17.32	100.00	.00
ELEC	−.432	.638	.033	9.52	50.00	50.00
DUST	−.183	.377	.083	11.69	100.00	.00
ROAD	−.154	.386	.096	18.61	100.00	.00
GREEN	−.426	.504	−.006	5.63	7.69	92.31

TABLE A2 Parameter estimates for the regression of air quality using GWRL

	MGWR coefficients			Percentage of cities by significance (95% level) of *t* test		
Variables	Min	Max	Mean	*p* ≤ .05 (%)	+ (%)	− (%)
Intercept	−.676	.538	.082	50.65	71.79	28.21
PRE	−1.094	−.141	−.593	93.94	.00	100.00
WIND	−.221	.105	−.029	6.06	.00	100.00
POPT	−.188	.265	−.006	3.03	100.00	.00
POPD	−.233	.436	.125	14.72	100.00	.00
POPU	−.278	.229	−.020	9.52	.00	100.00
PCGRP	−.298	.327	−.021	9.96	8.70	91.30
SEGRP	−.076	.232	.075	.00	.00	.00
ELEC	−.249	.442	.089	14.72	100.00	.00
DUST	−.073	.166	.058	.87	100.00	.00
ROAD	−.108	.304	.070	14.72	100.00	.00
GREEN	−.253	.159	−.087	12.12	.00	100.00
LAG	.229	1.076	.691	95.67	100.00	.00

TABLE A3 Parameter estimates for the regression of air quality using MGWRL

	MGWR coefficients			Percentage of cities by significance (95% level) of *t* test		
Variables	Min	Max	Mean	$p \le .05$ (%)	+ (%)	− (%)
Intercept	−.048	.131	.058	.00	.00	.00
PRE	−.996	.079	−.535	83.12	.00	100.00
WIND	−.091	.057	−.030	.00	.00	.00
POPT	.013	.044	.034	.00	.00	.00
POPD	−.216	.420	.118	25.97	100.00	.00
POPU	−.005	.022	.003	.00	.00	.00
PCGRP	−.016	.012	−.005	.00	.00	.00
SEGRP	.061	.080	.071	79.65	100.00	.00
ELEC	.041	.059	.048	.00	.00	.00
DUST	.053	.066	.059	.00	.00	.00
ROAD	−.044	.141	.052	10.82	100.00	.00
GREEN	−.174	.058	−.038	.00	.00	.00
LAG	.257	1.070	.628	96.54	100.00	.00