Computers, Environment and Urban Systems 75 (2019) 35–48

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Analysis of spatial and seasonal distributions of air pollutants by incorporating urban morphological characteristics

Ye Tian[a], Xiaobai Yao[a,*], Liding Chen[b]

[a] Department of Geography, University of Georgia, Athens, GA, USA
[b] State Key Lab of Urban Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences (CAS), Beijing, China

ARTICLE INFO

Keywords:
Air pollutants
Morphological characteristics
Spatial and seasonal distribution
Spatiotemporal regression
Spatial stratefied heterogeneity

ABSTRACT

Due to the worldwide trend of industrialization and urbanization, air pollutants were emitted heavily on a global scale particularly in developing countries, which produces adverse effects on human health by causing health problems such as respiratory and lung diseases. Many regression models based on land use types and urban fabrics have been built to analyze the spatiotemporal distribution of air pollutants, however, few of them examined the relationship between urban morphological characteristics and the distribution of air pollutants in a megacity. This study investigates such relationships for six types of air pollutants (PM2.5, PM10, SO_2, NO_2, O_3, and CO) and a composite AQI (Air Quality Index) based on hourly data at 35 monitoring stations in Beijing in 2016, with morphological characteristics (Morphological building index), meteorological factors (Land Surface Temperature, LST), land use (vegetation, road length, gas station and industry point data), and population distribution data. We also analyzed the results with spatiotemporal regression and SSH (Spatial Stratified Heterogeneity) models respectively. According to the spatiotemporal regression model, the morphological building index (MBI) shows a strong correlation with the dispersion of PM2.5 ($R^2 = 0.81$) and AQI ($R^2 = 0.80$) in the warm season and this finding was reinforced through the Leave-one-out-cross-validation (LOOCV) analysis. From the SSH analysis, the road length in a large proximal region impacts air pollutants the most, especially for O_3; and population density significantly affects PM 2.5, AQI, SO_2, and NO_2 in the cold season. From an integrated interpretation, distance to nearest industry impacts the spatial distribution of NO_2 in cold season, while it impacts that of PM2.5 and AQI in both warm and cold seasons. The research finds that these two models supplement each other well and together help to give us a better understanding of how air quality is affected in the urban landscape.

1. Introduction

Since the rapid urbanization and industrialization, China has experienced deteriorating air condition, especially in large cities during the past two decades. For Chinese cities, the emissions of NO_2 (nitrogen dioxide) and SO_2 (sulfur dioxide) in 2010 were as 2.5 times and 1.5 times the values in 1990, respectively (Mendenhall, Sincich, & Boudreau, 1996). Among them all, the most common air pollutant is PM2.5 in recent years and a lot of attention has been paid to it (Wang, Hu, Chen, Chen, & Xu, 2013, Yang et al., 2017, Yuan, Liu, Castro, & Pan, 2012). In Beijing, the annual average concentration of PM2.5 was around 70–100 μg/m^3, which were two to three times higher than the level 1 Interim Target (35μg/m^3) assigned by WHO (Cheng et al., 2013). Besides, these emissions have adversely impacted air quality at regional, national, and even global scales and played an essential role in global climate change (Grimm et al., 2008). In addition, air pollutants have been damaging the environment due to their chemical properties. For example, the acidification of soil, lakes, and rivers are resulted from NO, SO_2, and Ammonia, which caused the loss of plant life, the habitat of animals, and the reduction of crop yield (Rao, Rajasekhar, & Rao, 2014). Moreover, annually, almost 3.7 million people die prematurely because of outdoor air pollution around the world (WHO, 2014). In China, air pollution has become the fourth greatest risk factor in all deaths (Fang, Liu, Li, Sun, & Miao, 2015). Many epidemiological researchers have demonstrated that air pollution leads to a variety of health problems by long-term exposure to air pollutants (Brunekreef & Holgate, 2002), which produces the considerable medical cost for individuals and economic loss for the government by the reduction of productivity.

Air pollutants are released from a variety of sources, such as

* Corresponding author.
E-mail addresses: Ye.Tian@uga.edu (Y. Tian), xyao@uga.edu (X. Yao), Liding@rcees.ac.cn (L. Chen).

https://doi.org/10.1016/j.compenvurbsys.2019.01.003
Received 30 April 2018; Received in revised form 29 November 2018; Accepted 9 January 2019

industry and transportation, and are driven by various social economic factors. Many researchers believe air pollution is closely related to the urban land use patterns and optimization of spatial planning could improve the air quality in the long run. Thus, the relationship between the spatial land use and land cover and the distribution of air pollution attracts our attention. The conception of LUR (Land Use Regression) was first introduced and termed regression mapping (Briggs et al., 1997) and was applied in the SAVIAH (Small Area Variation in Air Quality and Health) project. The development and accessibility of GIS data and techniques contributed a lot to the popularity of LUR models. Ross et al. (2007) predicted PM2.5 in New York City and surrounding counties by LUR models, integrated with traffic density and census data. The model interpreted 60% or more of the variation of PM2.5 concentration with prediction errors below 10% (Ross, Jerrett, Ito, Tempalski, & Thurston, 2007). Meng et al. combined population, length of major roads, agricultural land area, and the number of the industry sites to the LUR model and explained a large part of the variability of NO_2 concentration in Shanghai, which outperformed the interpolation methods of Kriging and IDW (Meng et al., 2015). Liu, Henderson, Wang, Yang, and Peng (2016) used LUR model to interpret the variances of NO_2 and PM2.5 concentrations in Shanghai, and found anti-correlation with coastal regions and correlation with industry and highway intensity (Liu et al., 2016). Zheng et al. analyzed the land use patterns with the spatiotemporal distribution of multiple air pollutants (i.e. O_3, SO_2, NO_2, and CO) to find their relationships (Zheng et al., 2017). To improve the temporal granularity of air pollution monitoring data, Anand et al. built a mixed-effects LUR method to model daily NO_2 concentrations in Hong Kong from 2005 to 2015 with satellite datasets, which realized the daily mapping of ambient surface NO_2 (Anand & Monks, 2017). When it comes to the global scale, Larkin et al. created the first global NO_2 LUR model to find the spatial variability of NO_2 concentration. The model explained 54% of the annual change of NO_2 and continental R^2 ranges from 0.42 to 0.67 (Larkin et al., 2017).

Some studies constructed multiple LUR models for different scenarios or purposes. For example, Wu et al. constructed eight LUR models to explain diurnal, seasonal and annual spatial changes of PM2.5 concentration in Beijing (Wu et al., 2015). Huang et al. explored the relationship of fifty-nine potential variables (e.g. land use, traffic and industry emission, and population density) with four air pollutants (i.e. PM2.5, SO_2, NO_2, and O_3) through LUR models based on national monitoring network, and the variance of the four pollutants could be explained to a certain degree (Huang et al., 2017a). Yang et al. set four LUR models to estimate the air pollution concentration through ground-based measurements, remote sensing data, air quality model, and other spatial inputs, and found the best model to explain the NO_2 and PM2.5, respectively (Yang et al., 2017). Yang et al. developed seasonal LUR models to simulate the change of PM2.5 in the urban area, and these models had a good fit and explained the variation of spatial distribution of PM2.5 concentration well (Yang et al., 2017). However, such multiple LUR models may not always show good performance. For instance, Muttoo et al. stated that seasonal models did not show clear differences for measuring NOx with similar R^2 values, and this was due to the high correlation between seasonal measurements for each of the monitoring sites (Muttoo et al., 2018).

Even though the spatial land use patterns are related to the distribution of air pollutants, the accuracy of the final results depends heavily on the accuracy of classification of land use types and most of them interpreted the results based on LUR model only. Besides, the urban morphology, as the three-dimension form of a set of buildings and urban shapes (Chen, 2013), has significant influences on the concentration of air pollution (Cárdenas Rodríguez, Dupont-Courtade, & Oueslati, 2016). Bereitschaft et al. quantified urban form by preexisting sprawl indices and spatial metrics and analyzed their relationship with air pollution among 86 U.S. cities (Bereitschaft & Debbage, 2013). Yuan et al. quantified the impact of urban morphological parameters, urban permeability, and building geometries, on the dispersion of air pollution by CFD (Computational fluid dynamics) approach in the high-density urban regions (Yuan, Ng, & Norford, 2014). Rodríguez et al. proved that urban morphology produced significant influences on pollution centration and found that higher concentration of NO_2 and PM10 was related to fragmented and highly constructed cities and the higher concentration of SO_2 was related to densely populated cities (Cárdenas Rodríguez et al., 2016). She et al. (2017) analyzed the correlation between urban form which was described by six spatial metrics and group-based measurements of six air pollutants in the largest metropolitan zone, Yangtze River Delta. The study demonstrated that urban form did affect the urban air quality (She et al., 2017). Lu et al. developed the geographically weighted regression model to analyze the relationship between urban form and the density of NO_2 and SO_2 derived by satellite data, and the results showed that urban form produced significant effects on the air quality in urban areas of China (Lu & Liu, 2016).

As for street-level analysis, Maignant evaluated the dispersion of air pollution through the MISKAM (Mikroskaliges Klima und Ausbreitungsmodell) model integrating with urban morphology, buildings volume and roughness, and climatology factors in a street canyon (Maignant, 2006). Edussuriya et al. found the linkage between urban morphology and air quality, and the final results showed that six morphological variables significantly explained the variance of air pollution at the street level (Edussuriya, Chan, & Ye, 2011). Shen et al. investigated the connection of street morphology or canyons to dispersion of air pollution for six cities around the world. It was concluded that an open central street would greatly improve the air quality due to a larger vertical exchange of air flow via the street roof (Shen, Gao, Ding, & Yu, 2017). Shi et al. used vehicle-based mobile measurements and regression models to estimate the spatial distribution of PM2.5 and PM10, and found the most decisive factor of urban morphology was the frontal area index on street-level air quality in the central area of Hong Kong (Shi et al., 2016a).

However, these previous studies emphasized those building morphology parameters (e.g. sky view factor (Shi et al., 2016b, Silva & Monteiro, 2016), and building orientation (Lu & Liu, 2016)) or urban form indices (e.g. frontal area index (Ghassoun & Löwner, 2017, Lu & Liu, 2016, Shi et al., 2016a)) that are largely affected by meteorological factors such as wind direction and wind speed, or by the detailed building fabrics, which is more suitable for small research areas with clear 3D building maps and accurate group-based field measurements. For a megacity like Beijing, this process would be time-consuming and need too much human interference. This current study proposes to investigate morphological characteristics at various spatial levels based on data extracted from high-resolution (HR) images. Such urban form characteristics has not been investigated before in statistical approaches to analyzing the spatiotemporal distribution of air pollution in megacities. Another shortcoming in many of the previous LUR models is the assumption of linear relationship. Because any type of models (e.g. LUR model) may have its own limitations due to its assumption (e.g. linearity) and internal mechanism, the current study will apply multiple models to supplement each other in order to provide findings for a better understanding.

This study aims to answer three questions for the study area: (1). What is the potential utility of morphological information for estimating the spatiotemporal distribution of air pollution? (2). Are there seasonal differences in the distributions of air pollutants and if so how are these differences distributed spatially? (3). To what extent does each factor of interest affects the spatial and seasonal distributions of air pollutants? The innovation of the research lies in four aspects: i) it is the one of the first studies that incorporates urban morphological characteristic, which is not sensitive to the accuracy of land use classification and is easy for general adoption, as an essential predictor for air pollution distribution. We combine morphological, meteorological,

land use and socioeconomic variables together to analyze the distribution of air pollutants' concentrations; ii) the research is designed carefully in consideration of possible sensitivity of parameter choices, thus the approach is generally applicable to study other cities; iii) parallel analyses of spatial distribution of seven air pollution indexes are conducted simultaneously for both warm and cold seasons, which helps to capture the changing dynamics of air pollutants throughout the year; iv) the use of both spatiotemporal regression and SSH models provides a more comprehensive understanding of how air pollutants concentrations are affected by different factors.

2. Research area

The research area is Beijing, the capital city of China. The metropolis covers a total area of approximately 16,410 km^2 according to statistics in 2010 by the Beijing statistics bureau. There are 35 air quality monitoring stations distributed across the entire area. These stations are classified into four categories, including 12 urban environmental evaluation sites, 16 suburban environmental evaluation sites, 5 traffic pollution monitoring sites, and 2 regional background control sites.

Since the urban functional zone was proved to be a suitable spatial scale to explore the effects of urban land use on air pollution (Yang et al., 2017), this study recognizes four functional areas, namely core functional area (i.e. Dongcheng and Xicheng districts), urban function extension area (i.e. Chaoyang, Fengtai, Shijingshan, and Haidian districts), new urban development area (i.e. Fangshan, Tongzhou, Shunyi, Changping, and Daxing districts), and ecological conservation area (Mentougou, Huairou, Pinggu, Miyun, and Yanqing districts). The spatial locations of the monitoring sites and the division of urban functional zones are shown in Fig. 1.

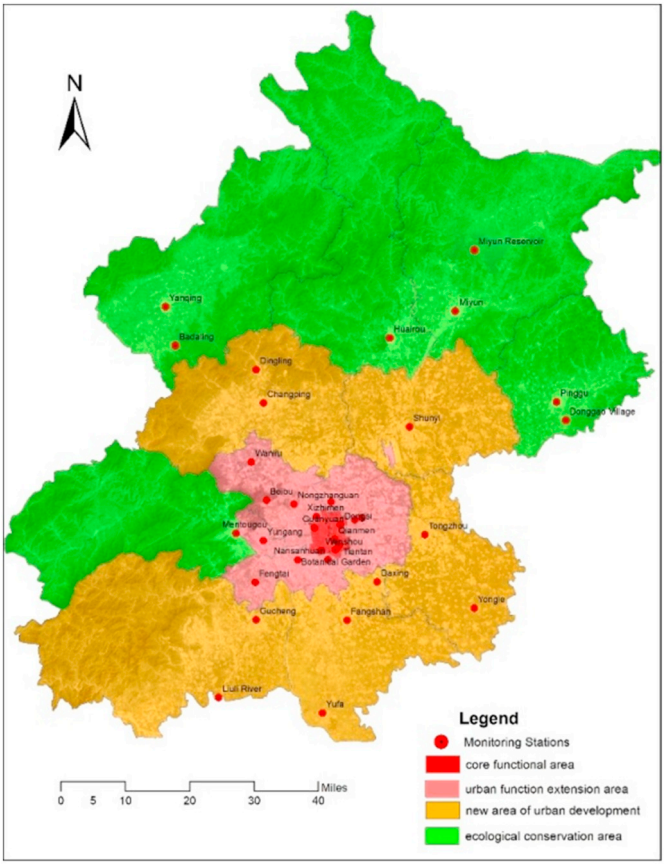


Fig. 1. Spatial distribution of monitoring stations and urban functional zones.

3. Data collection and preprocessing

3.1. Air pollution data

The air pollution data are acquired from 35 monitoring stations in Beijing. They are based on hourly measurements of six types of air pollutants (PM2.5, PM10, SO2, NO2, O3, and CO) and an overall Air Quality Index (AQI) in 2016. AQI is a measure of air quality with the maximum concentration among the abovementioned six air pollutants (She et al., 2017). During data preprocessing, the hourly data are summarized to monthly averages.

3.2. Morphological characteristics

The morphological characteristics in the research area is mainly quantified with the MBI (Morphological Building Index) proposed by Huang et al.(Huang & Zhang, 2011), as the index has been successfully applied in various urban regions (Huang & Zhang, 2012, Wang et al., 2016). The reason of choosing MBI can be illustrated as following: Firstly, it is hard to distinguish building shadows on low spatial resolution images, because buildings usually have clear shadows in the high spatial resolution images, such as ZY-3 images used in this study. Then, the morphological operations could exaggerate the spectral difference between the buildings and their corresponding shadows, which is differential morphological profile (DMP). Finally, MBI is constructed with the logic that buildings have larger DMP values if they show higher local contrast in all four directions (buildings and its shadow have clear contrast concerning spectral values). Thus, MBI is able to convert the physical properties of buildings (e.g., brightness, size, and contrast) to morphological features (e.g., reconstruction, granulometry and directionality). Therefore, MBI represents the probability of the presence of buildings. Because built-up density is closely associated with the distribution of air pollutants (She et al., 2017, Yang et al., 2017, Zheng et al., 2017), it is reasonable to use MBI as a candidate predictor for the distribution of air pollutants.

In order to calculate MBI, the optical high-resolution (HR) images from the Chinese No. 3 Resources Satellite (ZY-3) were utilized. The following summarizes the major steps of calculating of MBI in this study. Please refer to the literature (Huang & Zhang, 2011, Wang et al., 2016) for more details.

Step (1) Top-hat reconstruction:

White top-hat (*W_TH*) transformation is defined as the difference between the original image and its morphological opening operation:

$$W\_TH(d,s) = b - \gamma_b^{re}(d,s) \quad (1)$$

where b is the maximum value among all multispectral bands and $\gamma_b{}^{re}$ is the result of opening-by-reconstruction from the brightness image. And s and d represent the length and direction of a linear structural element (*SE*), respectively.

Step (2) Directional $\overline{W\_TH}$:

The multidirectional information of *W-TH* is computed by averaging four directions of the *se*:

$$\overline{W\_TH}(s) = \underset{d}{mean} W\_TH(s) \quad (2)$$

Step (3) Granulometry:

Granulometry indicates the scale and size of an object in imagery. Therefore, the differential morphological profile (DMP) is described as:

$$DMP_{\overline{W\_TH}}(d,s) = \overline{W\_TH}(d, s + \triangle s) - \overline{W\_TH}(d,s) \quad (3)$$

where $s_{min} \le s \le s_{max}$ and $\triangle s$ are the intervals of granulometry.

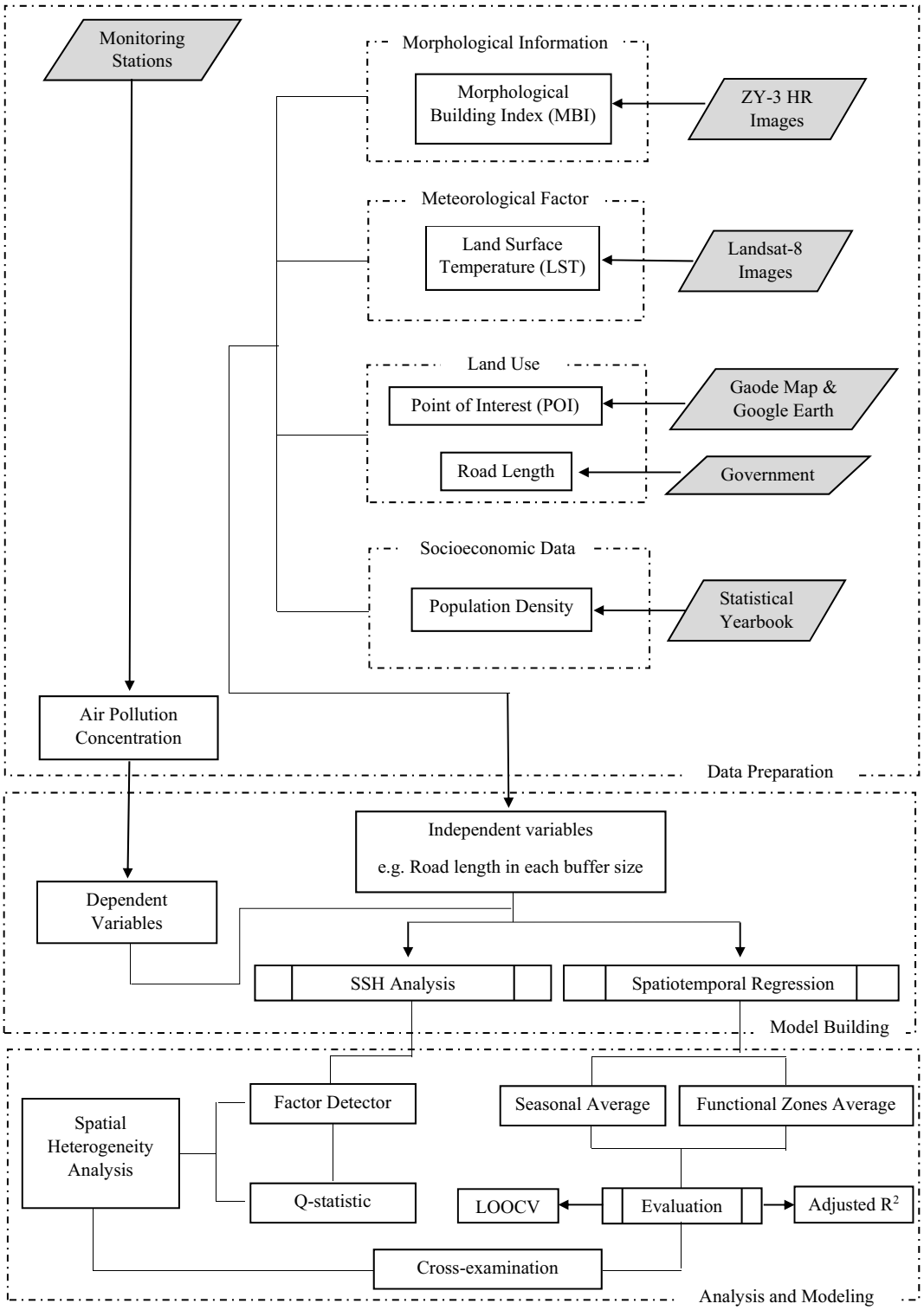


Fig. 2. Flowchart of the research design.

Step (4) MBI:

The MBI is obtained by averaging $DMP_{\overline{W\_TH}}$:

$$\begin{cases} MBI = \dfrac{\sum_{d,s} DMP_{\overline{W\_TH}}(d,s)}{D \times S} \\ S = (s_{max} - s_{min})/\triangle s + 1 \end{cases} \quad (4)$$

where D and S represent the amount of directionality and the scale of profiles, respectively.

3.3. Land use data

3.3.1. Traffic data

This study uses road lengths by road type to surrogate for traffic density. The method has been applied in previous studies and was proven appropriate (Brauer et al., 2003; Henderson, Beckerman,

Table 1
Candidate variables used in the study.

Type	Potential variable	Buffer size (m)	Effect direction
Morphological information	MBI	100,300,500,1000,2000,3000,5000	positive
Meteorological data	LST	100,300,500,1000,2000,3000,5000	positive
Socioeconomic data	Population density	NA	positive
Land use	NDVI	100,300,500,1000,2000,3000,5000	negative
	Road density	100,300,500,1000,2000,3000,5000	positive
	POI industry	NA	positive
	POI gas station	100,300,500,1000,2000,3000,5000	positive

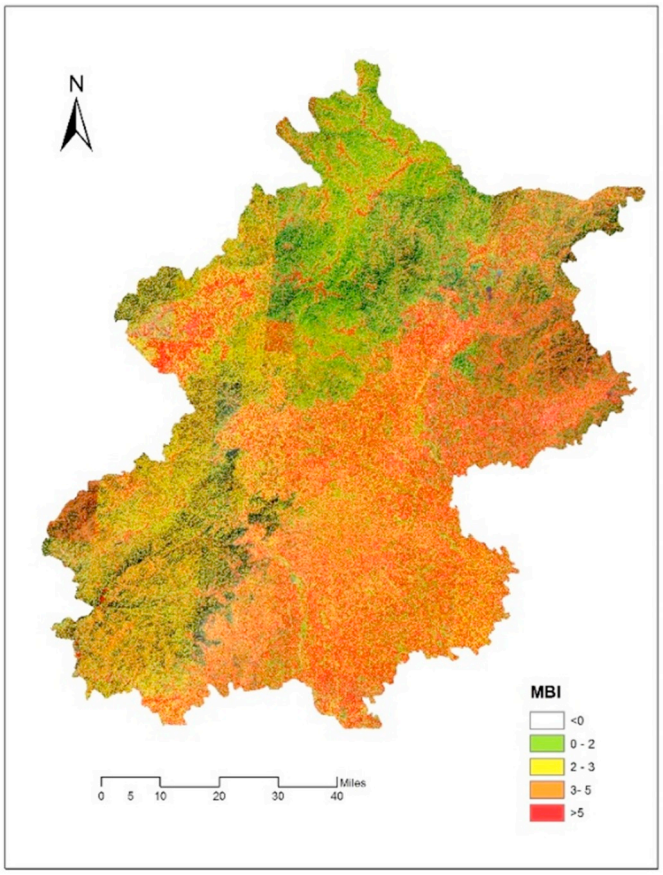


Fig. 3. MBI value in the research area.

Jerrett, & Brauer, 2007; Hoek et al., 2008; Madsen et al., 2007). The data was obtained from Beijing institute of Surveying and Mapping (http://www.bism.cn/). Four types of road types are classified in the dataset, namely national highway, provincial highway, municipal highway, and county highway. Road length within a buffer area of each monitoring station is calculated. Previous study (Meng et al., 2015) has indicated reasonable buffer sizes ranging from 100 m, 300 m, 500 m, 1000 m, 2000 m, 3000 m, and 5000 m, and thus we tested with all of them in order to choose the best sizes.

3.3.2. POI data

Since the distribution of air pollutants is influenced by the pollution source, this study considers two types of point data, gas stations which can release air pollutants due to oil leakage and industry sites which may produce pollutants throughout the production process. Specifically, 1062 gas stations, 31 industry sites are collected in the entire research area. These data were obtained from Gaode map and Google earth.

3.3.3. Vegetation

The vegetation data is mainly evaluated from NDVI calculated from Landsat-8 downloaded from USGS throughout the 2016.

3.4. Meteorological information

According to (Zheng et al., 2017), land surface temperature (LST), which influences and intensity of particles mobility, and built-up density were both related to concentrations of particulate matters, SO_2, NO_2, and CO. Therefore, we calculated the LST in each month of 2016 from landsat-8 images and integrated them into the final statistical model.

3.5. Socioeconomic data

The population density was obtained from the statistical yearbook of Beijing in 2016 (http://www.bjstats.gov.cn/nj/main/2016-tjnj/zk/indexeh.htm), which has direct relationship with distribution of air pollutants (Meng et al., 2015, Muttoo et al., 2018, Wu et al., 2015, Yang et al., 2017)

4. Methods

Spatial autocorrelation and spatial heterogeneity are two fundamental characteristics of geographical data and processes. Supported by the First law of Geography (Tobler, 1970), spatial autocorrelation refers to the situation when a variable correlates with itself through space. In a general sense, spatial heterogeneity is about the uneven distribution of traits, events, relationship or spatial variation of properties across a region (Wang et al., 2016). But particularly, it concerns the spatial non-stationarity of relationship which means the same stimulus may provoke a different response in different space (Fotheringham, 2009). Similarly, temporal heterogeneity concerns such non-stationarity over time.

The research considers both spatiotemporal autocorrelation and spatial heterogeneity. At the beginning of analysis, all candidate variables went through the process of step-wise regression for a coarse selection of significantly related variables for the estimation of air pollution indexes. Then two parallel approaches were taken to model the relationships between the variables and air pollutants' concentrations. The first approach is spatiotemporal regression, while the second is the spatial stratified heterogeneity (SSH) analysis. There are three reasons why the research design takes two parallel modeling approaches. First of all, the spatiotemporal regression model is based on the assumption of linear relationship, however we also want to explore whether non-linear relationships exist between dependent and independent variables (Wang & Xu, 2017). Secondly, each of the two approaches focuses on one of the two fundamental laws of spatial data. The spatiotemporal regression explicitly considers spatial and temporal autocorrelations, while the SSH analysis takes care of spatial heterogeneity. Thus the parallel applications of two approaches will provide a more comprehensive view of the relationships. Thirdly, the comparison of results from the two different approaches will allow us to cross-examine the identified relationship. Fig. 2 shows the research design.

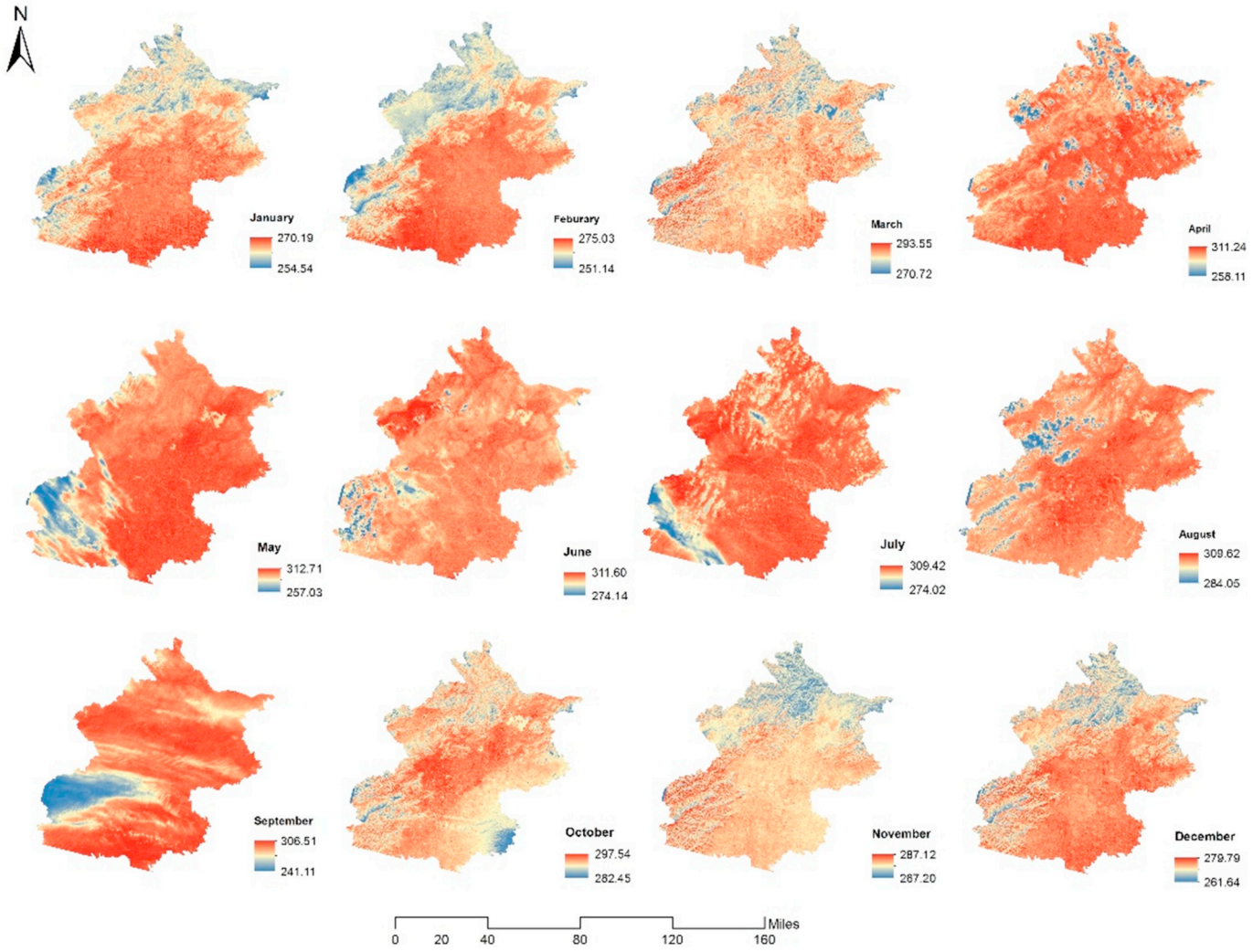


Fig. 4. LST during the year of 2016.

4.1. Spatiotemporal regression

4.1.1. Coarse selection of significant variables

To find significant variables among candidates in an objective and automatic way, the step-wise regression modeling approach was used for a coarse selection. This approach combines backward elimination and forward selection. It regresses multiple variables while simultaneously removing those that are not significant.

4.1.2. Spatiotemporal regression modeling

Based on the factors above, a spatially and temporally weighted regression was conducted to analyze the associations between these factors and spatiotemporal distributions of air pollutants. The regression equation is described as follows:

$$Y_i = \beta_{0i} + \beta_{1i}X + \beta_{2i}WX + \varepsilon i \quad (5)$$

where the subscript of i indicate the air pollutant i. $Y_{i,}$ is the concentration of air pollutants i; X indicates the dependent variables: *LST* (land surface temperature), *NDVI*, *MBI* (morphological building index calculated from the HR images), *POP* (population density), *ROAD* (road lengths), *DIS* (distance to the nearest industry), and *GAS* (number of gas stations) in certain buffer areas. W is the spatiotemporal matrix, and ε is the error item.

Most of the variables, namely *LST, NDVI, MBI, ROAD, and GAS*, are calculated within a proximal area around each monitoring station j. Here a buffer around each station is created to represent the proximal area. The results can be sensitive to the chosen buffer size. Therefore, multiple buffer sizes are chosen in a wide range between 100 m and 5000 m for these variables. Table 1 shows all the candidate variables. Following (Guo et al., 2017), the spatiotemporal weights W were obtained from the multiplication of spatial weights and temporal weights. In this study, the spatiotemporal weights were calculated when we analyzed the distribution of air pollutants in warm and cold seasons.

4.1.3. Model evaluation

A set of statistics were used to test the multicollinearity, spatial autocorrelation of the residuals, stationarity and model bias. These statistics include the Variance Inflation Factors (VIFs), Moran's I, Koenker Statistics, and Jarque-Bera Statistics, which were calculated in the diagnostic tests, respectively. If VIF is greater than 3 (redundancy) for any variable in the model, then the variable would be removed from it (Meng et al., 2015). The Moran's I is used to test the spatial autocorrelation of the model's residuals. The p-values of Koenker Statistics and Jarque-Bera Statistics are used to evaluate the stationarity and model bias respectively.

Because there were a small number (35) of monitoring stations in the study area, multifold cross-validation may not be plausible. Thus the model performance was tested by the Leave-one-out-cross-validation (LOOCV) method (Meng et al., 2015). The final model was fitted to 34 stations and dependent variables of the remaining station were

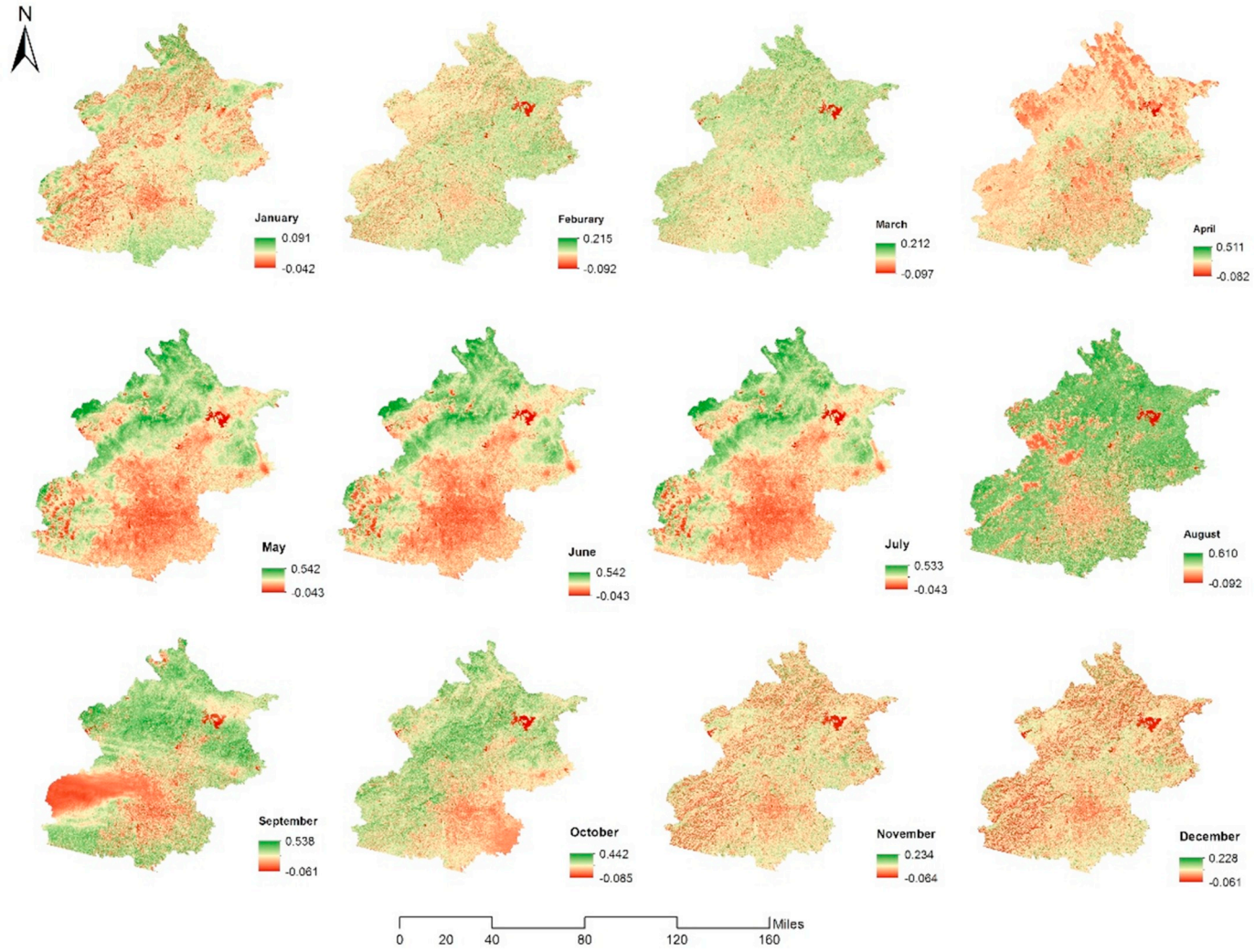


Fig. 5. NDVI during the year of 2016.

estimated by the fitted model. Such procedure would repeat 35 times and the overall R^2 and the root mean squared error (RMSE) for all sites were computed to show the model performance. Fig. 2 is the flowchart of this study.

4.2. SSH (spatial stratified heterogeneity) analysis

The spatial stratified heterogeneity (SSH) model has been applied in many research fields, including air pollutant $PM_{2.5}$ cause analysis (Lou, Liu, Li, & Li, 2016), health risk assessment (Wang et al., 2010), surface modeling (Wang, Christakos, & Hu, 2009), and more. An open-source program, Geodetector, is available for use. There are four detectors in the Geodetector, and this study uses the factor detector in order to identify the significance and strength of associations between pairs of candidate factors (variables) and air pollutant indexes. The program calculates a so-called q-statistic to test the significance of each association. The value of q is a ratio ranging from 0 to 1, where 0 means no association between the dependent variable and a predictor, while 1 means that the dependent variable is perfectly associated with the predictor. The q-statistic can be calculated with Eq. (6) (Wang et al., 2016):

$$q = 1 - \frac{SSW}{SST},\ where\ SSW = \sum_{h=1}^{L} N_h \sigma_h^2,\ SST = N\sigma^2 \tag{6}$$

where N and σ^2 are the number of units and the variance in the study area which is composed of L strata (h = 1,2….L), respectively. N_h is the number of units and $\sigma_h$is the variance in stratum h. Larger value of q means larger spatial heterogeneity in the study area. Another benefit of the q-statistic is that it measures the association between X and Y, either linearly or nonlinearly.

The format of p can be transformed so that it can satisfy the non-central F-distribution.

$$F = \frac{N-L}{L-1}\frac{q}{1-q} \sim F(L-1, N-L; \lambda) \tag{7}$$

$$\lambda = \frac{1}{\sigma^2}\left(\sum_{h=1}^{L} \overline{Y_h}^2 - \frac{1}{N}\left(\sum_{h=1}^{L} \sqrt{N_h}\overline{Y_h}\right)^2\right) \tag{8}$$

where λ, $\overline{Y_h}$ are the non-central parameter and mean value in stratum h, respectively. Then, we could use the *p*-value of q-statistic testifies that whether they have significant differences of variances in different strata.

5. Results and discussions

5.1. MBI

The final result of MBI is shown in Fig. 3. An MBI value can be interpreted as probability of buildings. High MBI values are found in the southeastern part of the research area. Indeed, this area of flatland is

Fig. 6. Road distribution in the research area.

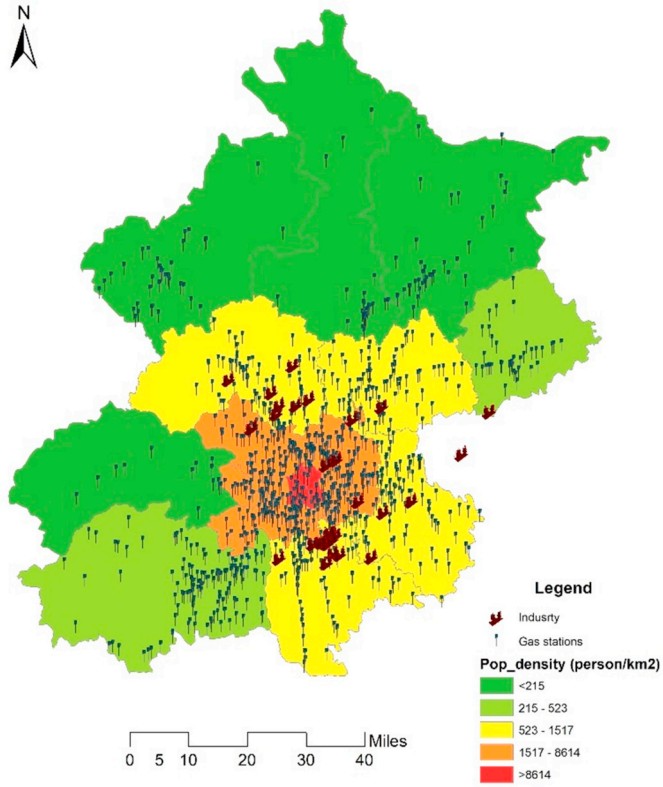


Fig. 7. Population density and POI of the industry and gas station.

highly suitable for constructing buildings and does see high building densities.

5.2. LST and NDVI

Both LST (in Fahrenheit degree) and NDVI were calculated from the Landsat-8, Level 1 TP, OLI_TIRS imagery in the year 2016. The results are displayed in Figs. 4 and 5. It can be seen from Fig. 4 that urban areas have relatively higher values of LST than suburban areas especially the mountainous regions in the northwestern part of research area. Such regions also have lower NDVI value according to Fig. 5. This phenomenon was partly due to the high coverage of imperious surface in the central urban regions. Besides, seasonal changes of NDVI could be clearly observed, which would potentially affect the spatiotemporal distribution of air pollutants. Since the mean value of LST and NDVI in certain buffer size around the monitoring station is necessary, the result is less influenced by some images that have a relatively higher degree of cloud (e.g. May, July, and September) and show abnormal distribution of LST to some extent.

5.3. Road lengths

Fig. 6 shows the four types of roads in the study area. The national highways connect the city with external areas and they are mostly distributed outside the core functional areas with low density. In fact, these highways typically have high traffic flows as they have more lanes and higher speed limits. The provincial expressways incorporate the main ring roads around central regions, and then they stretch outside, and connect with national highways. As for municipal freeway and county roads, they cover more areas with relatively low capacity but high density. In this study, we calculated the total lengths of all four types of roads in each buffer zone.

5.4. POI data and population density

Fig. 7 shows the distribution of industry points and gas stations in different population density regions. It shows that most industrial sites are located in the urban function extension area, probably due to the convenience and good accessibility to the core functional area. The gas stations are more spread out and are distributed mainly along the roads.

5.5. Spatial and Seasonal distribution of Air Pollutants

Given that air pollution may change dramatically from month to month (Jiang, Wang, Tsou, & Fu, 2015), we analyzed the temporal change of each pollution index on the monthly base. Fig. 8 displays the monthly change of each air pollution index in all stations. As shown in the figure, $PM_{2.5}$, PM_{10}, SO_2 and AQI enjoy relatively low values from April to September, whereas CO shows the reverse situation. Besides, NO_2 fluctuates less drastically than the others. Therefore, we conduct the dichotomic research by dividing the time period into cold (October to March) and warm seasons (April to September). Furthermore, some places have higher values of air pollutants, such as Liuli River, Yufa (the most two southern stations), and Daxing, than other regions (Miyun Reservoir). The global trend shows that most air pollutants concentrate more in the southern part than in the northern part of the study area and more in the cold season than the warm season.

Fig. 9 displays the average value of each air pollution index in both warm and cold seasons. The surfaces are fitted by kriging which is a geostatistical procedure that generates an estimated surface from point-based value. It is shown in the figure that most air pollutants concentrated in the southern part of the area except for CO. The reason might be related to high population density and heavy traffic in the southern area. This finding is in accordance with other studies (Guo et al., 2017b). Besides, from the warm season to the cold season, concentrations of most air pollutants increased except for CO. It is

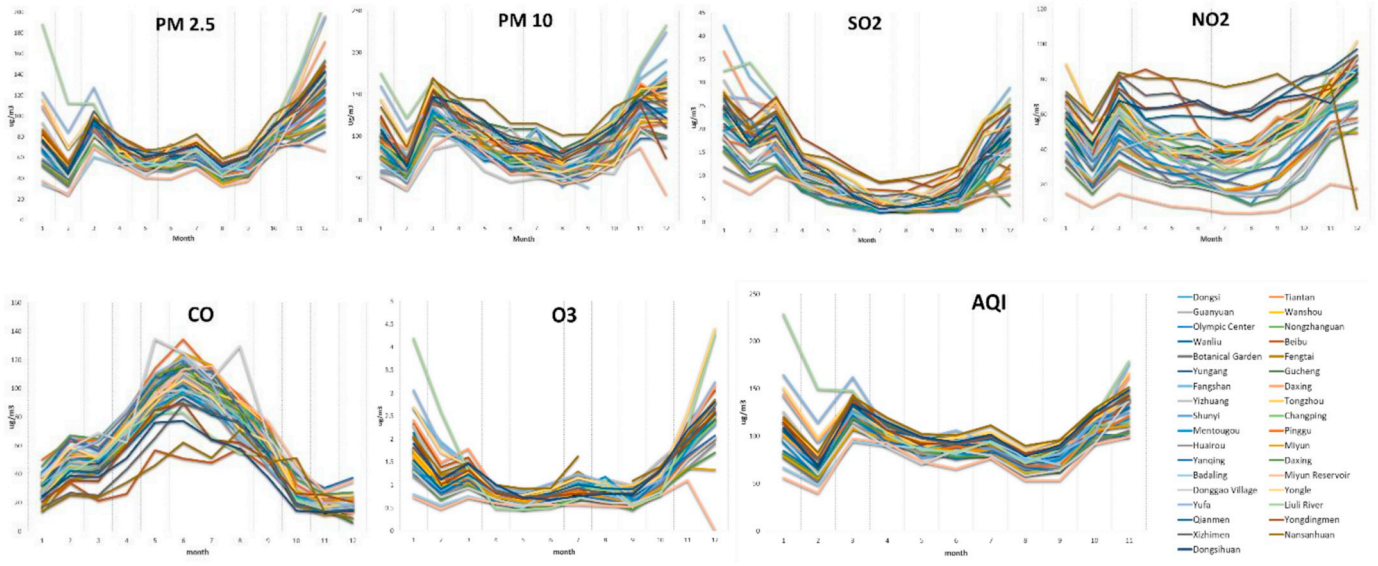


Fig. 8. Monthly change of air pollution index.

reasonable considering huge energy consumption in the winter for heating. CO shows a totally different trend with all other air pollutants, more detailed investigation of reasons requires further analysis.

5.6. Spatiotemporal regression modeling results

In order to incorporate temporal weights in the final model, the weighted mean was calculated in both warm and cold seasons. According to the time-series analysis, in the warm season, the data in June and July have the highest weights, followed by May and August, while April and September have the lowest weights. The situation is similar in the cold season. The model has 7 dependent variables with 37 candidate predictors for both seasons respectively. To select the most effective variables and to reduce the chance of multicollinearity, a step-wise regression was conducted. Since the spatiotemporal weight is obtained by multiplying the spatial weight and temporal weight matrix (Guo et al., 2017a), the spatiotemporal regression models could be calculated by incorporating temporal weight into the GWR model.

5.6.1. Final selected variables and spatiotemporal regression models

Even though the step-wise process is useful to select statistically significant variables, the coefficients of predictors need to be consistent with the effect direction listed in Table 1. After filtering potential predictors through step-wise regression and considering correct effect direction, a regression model is obtained for each of the 7 air pollution indexes (dependent variables). The 7 regression models are listed in Table 2. Note that all the statistically significant predictors at the significance level of 0.05 are listed for each model. These predictors will be used for the spatiotemporal regression analysis in the next stage.

Taking a closer examination of the 7 models, several findings are observed. When it comes to warm season, MBI within 1000 m buffer size (MBI_1000) have shown statistical significance for the $PM_{2.5}$, PM10, SO_2, NO_2, and O_3 concentrations and the overall AQI. The models for PM2.5 and AQI have higher R^2 (0.81 and 0.80) which suggest about 80% of the variations in these two indexes can be explained by the respective models. As for the cold season, MBI appear to be less significant, whereas LST_2000 and NDVI_100 are both significantly related to PM2.5, PM10, and AQI, and NO_2 concentrations. The reason why MBI did not show a strong association with any air pollutant concentrations in cold season might be related to the fact that Beijing suffered higher level of air pollutants partly due to the district heating during this period. High concentrations of air pollutants may be more closely associated with meteorological factors (e.g. aerosol optical depth (AOD)), and industry emission (e.g. combustion of a large amount of charcoal). It is noteworthy that the direction of effects on CO is counter-intuitive since its spatiotemporal change is dissimilar with all other pollutants shown in Fig. 9, and such situation also happened in the previous study(Zheng et al., 2017).

Furthermore, the influence of LST is interesting. The concentration of air pollutants is influenced by LST in small proximal areas in warm season, whereas in large proximal areas in cold season. Since the high temperature would accelerate the flow of air, and the warm season enjoys drastic air convection. As a result, even the smaller area of LST would produce obvious effects on the distribution of air pollutants in the warm season. Obviously, population density does not show any correlation with these air pollutants in both seasons, which may attribute to the relatively coarse spatial resolution of the data (administrative region based not grid based). Overall, we can safely draw the conclusion that MBI has an obvious influence on the concentration of air pollutants within a certain period.

5.6.2. Model evaluation

In this part, the values of Moran's I, Koenker (BP) Statistic, VIF of each selected predictor, Jarque-Bera Statistic are used to evaluate the presence/absence of spatial autocorrelation, stationarity, and multicollearity, and normal distribution with in a model. If the *p*-value of Jarque-Bera Statistic shows statistically significant (e.g. PM10, and SO_2 in warm season), the spatiotemporal regression model could reduce the non-stationarity or heteroscedasticity to a large degree.

As can be seen from Table 3, all selected predictors have their VIF values smaller than 3, which means the final selected model has no sign of multicollinearity. The *p*-values of Moran'I values for all models are greater than 0.05, which suggests that the models' residuals are not spatial autocorrelated. Besides, all of the variables show statistically non-significant Koenker (BP) Statistic, indicating that the residuals are all normally distributed without bias. As for Jarque-Bera Statistic, a majority of them are statistically non-significant except for PM_{10} and SO_2 in the warm season and $PM_{2.5}$ in the cold season. The statistics also suggests that the distributions of these pollutants are spatially and seasonally nonstationary, which demonstrates the importance of spatiotemporal regression model in estimating their distributions.

5.6.3. Cross-validation

The LOOCV cross-validation process was conducted in R and results are displayed in Table 4. According to the table, all models show satisfactory results. We paid particular attention to the influence of

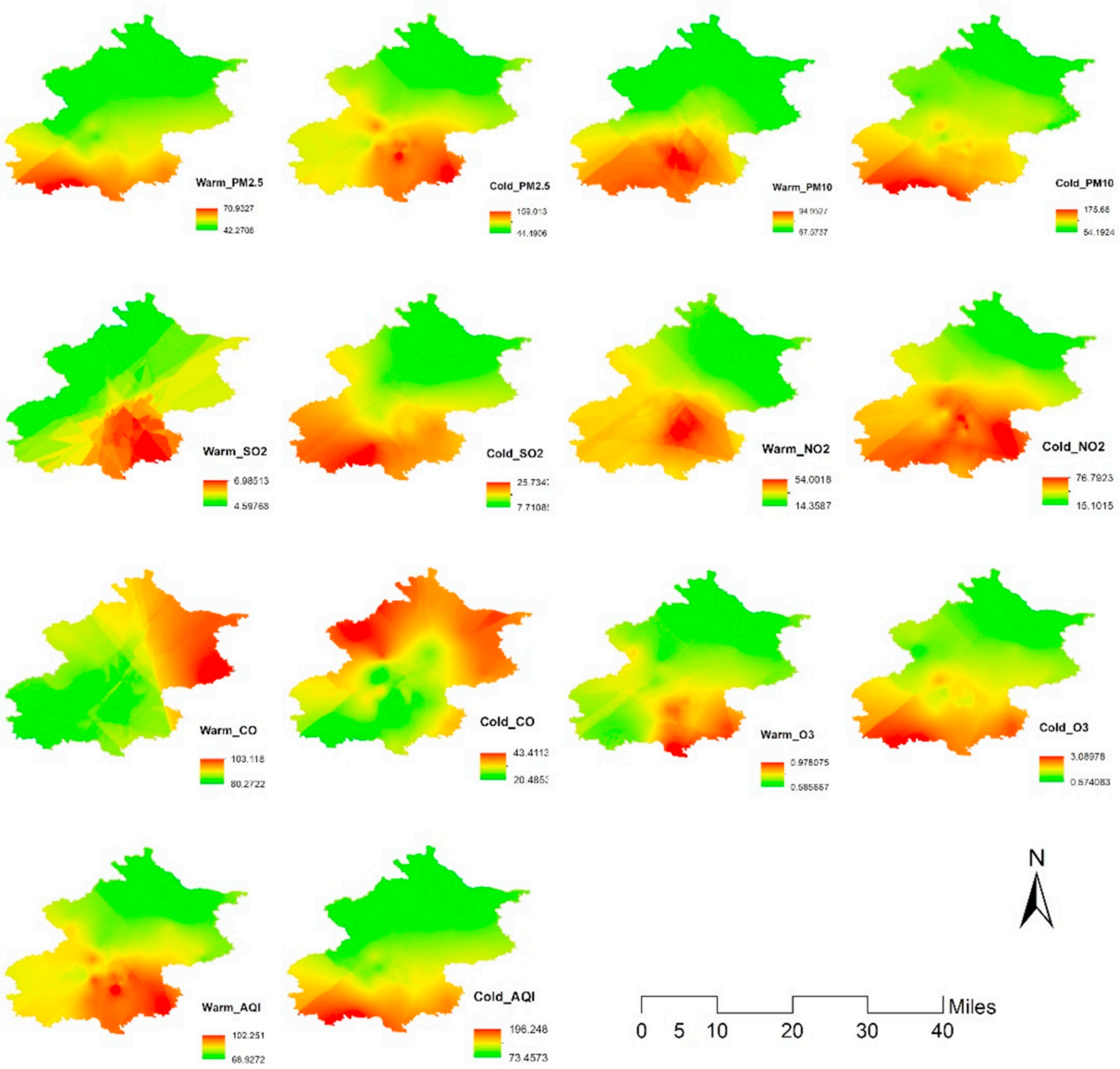


Fig. 9. Spatiotemporal change of air pollution in the study area.

Table 2
Summary of step-wise regression results.

Pollutant type	Regression model	Adjusted R^2	P-value of global F-test
Warm season			
PM2.5	PM2.5 = −147.96–72.0.98 × NDVI_100*** - 0.17 × DIS** + 0.76 × LST_100** + 1.45 × MBI_1000***	0.81	2.39e-09***
PM10	PM10 = 112.86–182.35 × NDVI_100***	0.61	1.44e-07***
AQI	AQI = −217.76–81.19 × NDVI_100*** − 0.16 × DIS** + 1.00 × LST_100*** + 1.67 × MBI_1000***	0.80	6.01e-09***
SO_2	SO_2 = 6.28 + 0.06 × GAS_5000* + 0.36 × MBI_1000°	0.28	0.02*
NO_2	NO_2 = 64.46 + 0.55 × GAS_5000** - 141.77 × NDVI_100** + 2.89 × MBI_1000°	0.58	1.57e-05***
CO	NA		
O_3	O_3 = −0.66–1.74 × NDVI_300*** +0.01 × LST_100° + 0.03 × MBI_1000**	0.55	4.14e-05***
Cold season			
PM2.5	PM2.5 = −2622.57–244.70 × NDVI_100** + 9.93 × LST_2000***	0.60	1.56e-06***
PM10	PM10 = −1962.06–305.65 × NDVI_100** + 7.60 × LST_2000***	0.49	5.81e-05***
AQI	AQI = −2874.77–291.97 × NDVI_100** + 10.97 × LST_2000***	0.59	2.35e-06***
SO_2	SO_2 = 19.41 + 8.58 × ROAD_100* - 55.79 × NDVI_100°	0.50	5.93e-03***
NO_2	NO_2 = −1285.16–0.73 × DIS*** − 215.14 × NDVI_100*** + 4.97 × LST_2000**	0.76	1.30e-08***
CO	CO = 752.69–0.23 × GAS_5000** − 2.63 × LST_2000*** + 77.35 × NDVI_100*	0.57	1.61e-06***
O_3	O_3 = −60.48 + 0.25 × LST_2000***	0.45	2.51e-05***

where NDVI_100 means the mean value of NDVI in 100 m buffer size, LST_100 means the mean value of LST in 100 m buffer size, and the rest can be interpreted in the same manner. *** Correlation is significant at the 0.001 level (2-tailed); ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed); ° Correlation is significant at the 0.1 level (2-tailed).

Table 3
Summary of model evaluation.

Pollution type	Significant predictors	VIF	P-value of Koenker (BP) Statistic	P-value of Moran's I	*P*-value of Jarque-Bera Statistic
Warm season					
PM2.5	NDVI_100	1.30	0.34	0.92	0.60
	DIS	1.18			
	LST_100	1.06			
	MBI_1000	1.13			
PM10	NDVI_100	NA	0.21	0.76	0.045*
AQI	NDVI_100	1.30	0.33	1.00	0.51
	DIS	1.18			
	LST_100	1.06			
	MBI_1000	1.13			
SO_2	GAS_5000	1.04	0.22	0.60	0.014*
	MBI_1000	1.04			
NO_2	GAS_5000	1.35	0.70	0.051	0.38
	NDVI_100	1.41			
	MBI_1000	1.09			
O_3	NDVI_300	1.50	0.71	0.92	0.29
	MBI_1000	1.38			
	LST_5000	1.36			
Cold season					
PM2.5	LST_2000	1.00	0.16	0.09	0.009**
	NDVI_100	1.00			
PM10	LST_2000	1.00	0.58	0.95	0.40
	NDVI_100	1.00			
AQI	LST_2000	1.00	0.23	0.12	0.03*
	NDVI_100	1.00			
SO_2	Road_100	1.14	0.93	0.89	0.91
	NDVI_100	1.14			
NO_2	DIS	1.18	0.30	0.74	0.59
	NDVI_100	1.00			
	LST_2000	1.18			
CO	GAS_5000	1.16	0.47	0.39	0.85
	LST_2000	1.01			
	NDVI_100	1.14			
O_3	LST_2000	NA	0.46	0.44	0.66

Table 4
Summary of LOOCV results.

Pollution type	R-squared	RMSE	R-squared	RMSE
	Warm season		Cold season	
PM2.5	0.72	4.31	0.66	17.62
PM10	0.55	12.30	0.58	20.59
AQI	0.71	5.15	0.61	20.31
SO_2	0.51	1.75	0.53	4.75
NO_2	0.63	12.69	0.65	10.93
CO	NA	NA	0.63	5.35
O_3	0.59	0.09	0.57	0.41

morphological characteristic on indexes of air pollutants. We found that the prediction accuracy of PM2.5 (R^2 = 0.72), AQI (R^2 = 0.71), and NO_2 (R^2 = 0.63) in the warm season that enjoy strong correlation with MBI through the LOOCV. This confirmed the influence of MBI on air pollutants concentrations.

5.6.4. Standardized residuals in different urban functional zones

To compare the relative prediction strength across different urban functional zones, values of standardized residuals were calculated for all zones respectively. The absolute values of them are shown in Fig. 10. The core functional area has the least standardized residuals for almost all pollutants in all seasons, which may partly due to the relatively high density of monitoring stations (one station for every 15.38 km^2) in this area. The largest standardized residual of AQI is found in the new area of urban development throughout the year. This is probably due to the fact that the majority of industrial sites are located is built in this zone (24/37) and it is difficult to acquire real-time emission data. Moreover, this zone is directly adjacent to another major metropolis (Tianjin) with heavy industry and its potential influence is not considered in the study.

During the warm season, the model produces the largest error for PM_{10}, SO_2, and NO_2 in the zone of urban functional extension area. This may be attributed to the lack of some meteorological data, such as wind speed and wind direction, that would directly affect the dynamics of air pollution. As for the cold season, the new area of urban development and the ecological conservation area present the top two largest residuals for all pollutants except CO, which may be caused by the discrete spatial distribution of monitoring stations and complicated topographic conditions. The two regions include large areas of farmland, lakes, and mountains where the Great Wall of China is located.

5.7. SSH modeling results

The q-statistic values and corresponding *p*-values are calibrated and only the statically significant q-statistics are reported in Table 5. The significance level is set at 0.05.

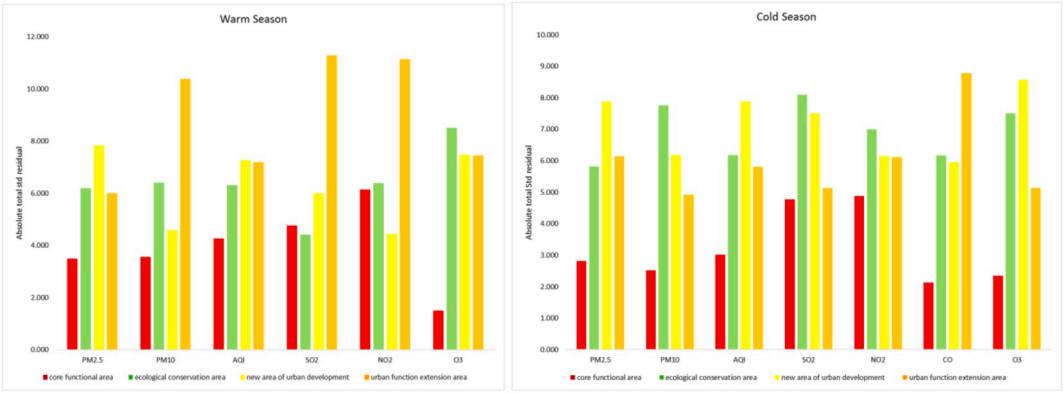
It is interesting to find that ROAD_5000 impacts all air pollutants in the table in the cold season and O_3 in both seasons. The q-statistic values related to ROAD_5000 are all higher than 0.99, which mean almost perfect associations between ROAD_5000 and the respective air pollutants. This suggests that traffic in a very large proximal area (buffer size of 5 km) has significant impact on the concentrations of most air pollutants, especially in the cold season. It is an important finding in supplement to other findings from the spatiotemporal regression modeling results. In addition, population density (POP) is found to be significantly associated with $PM_{2.5}$, AQI, SO_2, and NO_2 in the cold season. This finding accords with the fact that more heating is needed in regions of larger population density during the cold season and thus would produce PM 2.5, SO_2, and NO_2. As for the variable of distance to the nearest industry site (DIS), it only shows significant association with $PM_{2.5}$ and AQI in the cold season.

5.8. Comparisons and discussions

Results of two modeling approaches are not exactly the same but are generally in accordance. For instance, the SSH analysis found population density to be a significant predictor in the cold season, which is not identified by the spatiotemporal regression models. However, it is noticed that the spatiotemporal regression did find the urban morphological characteristic index (MBI) to be a significant predictor for most pollutants. High density of buildings usually indicates high population density (Wang et al., 2016) and high demand of energy consumption, which would lead to the worse air conditions especially in urban areas (Silva, Oliveira, & Leal, 2016). Because MBI considers building index, which is highly correlated with population density, the population density might be dropped in the process of stepwise regression to avoid multicollinearity. With this consideration, the results of the two modeling approaches are not contradictory. The variable of DIS is found to be a significant predictor for $PM_{2.5}$ and AQI in the warm season and NO_2 in the cold season based on the spatiotemporal regression models, while the SSH analysis added similar message in the warm season.

More importantly, results of the two approaches are also complementary to each other. With the combined findings from both modeling approaches, distance to industry is found to be a significant predictor in both seasons. Thus the message is strengthened that industry emission affects the air quality throughout the year. While the spatiotemporal regression models captured other relationships in detail, SSH analysis added the important finding of the significant impact of traffic on air pollution.

Several variables, namely MBI, LST, NDVI, Road density and POI of gas stations, were constructed for a proximal area around each sample point. The analysis result can be sensitive to the predefined size of the proximal area. To test the sensitivity and to find the real scope of influence, seven different buffer radii (100, 300, 500, 1000, 2000, 3000,



(a) (b)

Fig. 10. Comparison of standardized residuals in each functional zone in the warm season (a) and the cold season (b).

Table 5
Summary of q-statistic.

Pollutant	Independent variable	q-statistic	p-value
Warm season			
O_3	ROAD_5000	0.996126	0
Cold season			
PM2.5	DIS	0.91361	0.030382
	POP	0.797056	0.044715
	ROAD_5000	0.991886	0
PM10	Road_5000	0.999857	0
AQI	DIS	0.912245	0.049241
	POP	0.807308	0.049926
	ROAD_5000	0.992438	0
SO_2	POP	0.785184	0.049594
	ROAD_5000	0.998024	0
NO_2	POP	0.786147	0.02862
	ROAD_5000	0.996314	0
O_3	ROAD_5000	0.997391	0

5000 m) were chosen and tested for each of these variables. Results show that the scope of influence differs among these predictive variables.

5.9. Limitations

The study has several limitations. First of all, the study is limited by data availability, suitability, and quality. Although the data source of air quality is expected to report hourly data of all seven types of indexes at all 35 monitoring stations, there were often missing data at random time and locations. Besides, remotely sense images used in the study have a relatively higher degree of cloud (see Figs. 4 and 5), which may lead to errors in LST and NDVI calculation. Other datasets, such as nighttime light data (Zhang & Hu, 2017) and aerosol optical depth from MODIS (Moderate Resolution Imaging Spectro radiometer) (Chen et al., 2018; Feng, Zou, & Tang, 2017), were used in the study. But their coarse spatial resolutions, 1 km and 3 km respectively, may not be appropriate for a urban region with high densities of buildings and population. Moreover, some meteorological factors, such as relative humidity, wind speed, wind direction, precipitation and air pressure, are highly relevant. However, such data in station-based scale were not available and thus had to be left out in the study.

Secondly, there are also limitations in the techniques used in the study. For the spatiotemporal regression modeling, even though stepwise regression is effective for prescreening of statistically significant variables, some internal flaws are noteworthy: i) as a “greedy search” method, it would not consider all possible combinations among all potential variables, and thus some significant groups may be subjectively ignored; ii) it only considers the main effect terms, not interaction terms or high-order terms of predictors (e.g. quadratic term) among variables. Moreover, the study emphasized distinguishes two seasons for comparison. Thus each spatiotemporal regression model would only have six sets of monthly air pollution data respectively. This may limit the usefulness of the temporal weight matrix. As for SSH model, we only used the factor detector to examine the associations between an independent variable and an air pollution index. Other three detectors of the SSH analysis could also be useful.

6. Conclusions and future research directions

This study investigates the relationships between air pollution and various factors in the urban landscape including socioeconomic, urban form, and morphological characteristics. The research design combines two parallel modeling approaches, the spatiotemporal regression and social stratified heterogeneity (SSH) analysis. In the experiment, the urban morphological characteristic variable, MBI, is proven to be a statistically significant factor for the estimations of air pollution. In addition, traffic density, distance to industry, density of gas stations, and population density are also found to be significant contributing factors.

Comparing with the conventional LUR model (Huang et al., 2017b, Larkin et al., 2017, Yang et al., 2017), this study makes two innovative improvements. First, it includes morphological characteristics in the modeling process. Even though a street-level analysis of air quality also considers urban morphology (Edussuriya & Chan, Shi et al., 2016, Shi, Xie, Fung, & Ng, 2018), it requires detailed field survey data and is less feasible or cost-effective than the zone-based analysis in this study. The second innovative improvement is the consideration of spatial autocorrelation, temporal autocorrelation, and spatial heterogeneity by the use of two modeling approaches. The spatiotemporal regression modifies classical regression by adding spatiotemporal weight matrices in the equation. The SSH analysis considers spatial heterogeneity in the process of identify associations among variables. In addition, the two approaches also consider both linear and non-linear relationships respectively. The use of two modeling strategies not only allow us to consider spatiotemporal autocorrelation and spatial heterogeneity in the same study, it also provides basis for cross-examination of findings.

Findings from the study can have far-reaching implications to urban environment policy and urban design practices for the case study city. Among other findings, results of this study suggests that air pollutants' concentrations in any location of a city are significantly influenced by road density and urban morphology index in a large proximal area of the location (i.e. ROAD_5000, MBI_1000). Moreover, distance to industry also has significant impacts. Thus, urban policy may need to control further expansion and/or intensification of road network and

buildings. Planning practitioners may also carefully choose the location of industry to keep them in distance from the populated area. The study provides evidences that can be used as references for future informed urban design and planning policies to achieve more ecologically sustainable urban development.

There are plenty of rooms for future research efforts. First, research can be conducted on finer spatial and temporal scales. It remains unknown how concentrations of air pollutants vary on different days of a week and at different times of a day. Future studies may also choose a finer spatial granularity, for instance, at the level of street canopy, with considerations of relevant meteorological factors. Research findings at various spatial and temporal scales will help us to gain fuller understanding of air pollution in urban environments. Another research avenue is to investigate the topic of interest for different types of cities. Furthermore, a comparative study of findings from different cities will be interesting and informative. The third direction of future research is making more use of modern geospatial technology to collect and analyze data in innovative ways. For instance, field data collection with mobile devices and other handheld instruments can be used more. Moreover, interactions among different variables can be further analyzed in the modeling process. Finally, different air pollutants have distinct causes and mechanisms of affecting air quality. For instance, SO_2 primarily comes from industrial emissions, such as soot and dust. NO_2 and CO are mostly produced by traffic emission. $PM_{2.5}$ and PM_{10} may be caused by combustion of coal used for heating during winter. Therefore, these air pollutants may better be analyzed separately in further research.

Acknowledgements

This work was partially supported by the State Key Laboratory of Urban and Regional Ecology, China.

References

Anand, J. S., & Monks, P. S. (2017). Estimating daily surface NO2 concentrations from satellite data – A case study over Hong Kong using land use regression models. *Atmospheric Chemistry and Physics, 17*(13), 8211–8230.

Bereitschaft, B., & Debbage, K. (2013). Urban form, air pollution, and CO_2 emissions in large US metropolitan areas. *The Professional Geographer, 65*(4), 612–635.

Brauer, M., Hoek, G., van Vliet, P., Meliefste, K., Fischer, P., Gehring, U., ... Lewne, M. (2003). Estimating long-term average particulate air pollution concentrations: Application of traffic indicators and geographic information systems. *Epidemiology, 14*(2), 228–239.

Briggs, D. J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., ... Van Der Veen, A. (1997). Mapping urban air pollution using GIS: A regression-based approach. *International Journal of Geographical Information Science, 11*(7), 699–718.

Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The Lancet, 360*(9341), 1233–1242.

Cárdenas Rodríguez, M., Dupont-Courtade, L., & Oueslati, W. (2016). Air pollution and urban structure linkages: Evidence from European cities. *Renewable and Sustainable Energy Reviews, 53*, 1–9.

Chen, Y. (2013). A set of formulae on fractal dimension relations and its application to urban form. *Chaos, Solitons & Fractals, 54*, 150–158.

Chen, Z.-Y., Zhang, T.-H., Zhang, R., Zhu, Z.-M., Ou, C.-Q., & Guo, Y. (2018). Estimating PM 2.5 concentrations based on non-linear exposure-lag-response associations with aerosol optical depth and meteorological measures. *Atmospheric Environment, 173*, 30–37.

Cheng, S., Lang, J., Zhou, Y., Han, L., Wang, G., & Chen, D. (2013). A new monitoring-simulation-source apportionment approach for investigating the vehicular emission contribution to the PM2. 5 pollution in Beijing, China. *Atmospheric Environment, 79*, 308–316.

Edussuriya, P., Chan, A., & Ye, A. (2011). Urban morphology and air quality in dense residential environments in Hong Kong. Part I: District-level analysis. *Atmospheric Environment, 45*(27), 4789–4803.

Edussuriya, P. S., & Chan, A. (2016). *Analysis of urban morphological attributes and street level air pollution in high-density residential environments in Hong Kong.*

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.

Feng, H., Zou, B., & Tang, Y. (2017). Scale-and region-dependence in landscape-PM2. 5 correlation: Implications for urban planning. *Remote Sensing, 9*(9), 918.

Fotheringham, A. S. (2009). "The problem of spatial autocorrelation" and local spatial statistics. *Geographical Analysis, 41*(4), 398–403.

Ghassoun, Y., & Löwner, M.-O. (2017). Land use regression models for total particle number concentrations using 2D, 3D and semantic parameters. *Atmospheric Environment, 166*, 362–373.

Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., & Briggs, J. M. (2008). Global change and the ecology of cities. *Science, 319*(5864), 756–760.

Guo, Y., Tang, Q., Gong, D.-Y., & Zhang, Z. (2017a). Estimating ground-level PM2. 5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sensing of Environment, 198*, 140–149.

Guo, Y., Tang, Q., Gong, D.-Y., & Zhang, Z. (2017b). Estimating ground-level PM 2.5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sensing of Environment, 198*, 140–149.

Henderson, S. B., Beckerman, B., Jerrett, M., & Brauer, M. (2007). Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science & Technology, 41*(7), 2422–2428.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment, 42*(33), 7561–7578.

Huang, L., Zhang, C., & Bi, J. (2017a). Development of land use regression models for PM2.5, SO2, NO2 and O3 in Nanjing, China. *Environmental Research, 158*, 542–552.

Huang, L., Zhang, C., & Bi, J. (2017b). Development of land use regression models for PM2. 5, SO2, NO2 and O3 in Nanjing, China. *Environmental Research, 158*, 542–552.

Huang, X., & Zhang, L. (2011). A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery. *Photogrammetric Engineering & Remote Sensing, 77*(7), 721–732.

Huang, X., & Zhang, L. (2012). Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5*(1), 161–172.

Jiang, W., Wang, Y., Tsou, M. H., & Fu, X. (2015). 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, 10*(10), e0141185.

Larkin, A., Geddes, J. A., Martin, R. V., Xiao, Q., Liu, Y., Marshall, J. D., ... Hystad, P. (2017). Global land use regression model for nitrogen dioxide air pollution. *Environmental Science & Technology, 51*(12), 6957–6964.

Liu, C., Henderson, B. H., Wang, D., Yang, X., & Peng, Z. R. (2016). A land use regression application into assessing spatial variation of intra-urban fine particulate matter (PM2.5) and nitrogen dioxide (NO2) concentrations in City of Shanghai, China. *Science Total Environment, 565*, 607–615.

Lou, C.-R., Liu, H.-Y., Li, Y.-F., & Li, Y.-L. (2016). Socioeconomic drivers of PM2. 5 in the accumulation phase of air pollution episodes in the Yangtze River Delta of China. *International Journal of Environmental Research and Public Health, 13*(10), 928.

Lu, C., & Liu, Y. (2016). Effects of China's urban form on urban air quality. *Urban Studies, 53*(12), 2607–2623.

Madsen, C., Carlsen, K. C. L., Hoek, G., Oftedal, B., Nafstad, P., Meliefste, K., ... Brunekreef, B. (2007). Modeling the intra-urban variability of outdoor traffic pollution in Oslo, Norway—A GA2LEN project. *Atmospheric Environment, 41*(35), 7500–7511.

Maignant, G. (2006). Measurements of air pollution due to the traffic in Bonaparte Street in Paris. *Air Pollution XIV,* 307–314.

Mendenhall, W., Sincich, T., & Boudreau, N. S. (1996). *A second course in statistics: regression analysis. Vol. 5* Prentice Hall Upper Saddle River, New Jersey.

Meng, X., Chen, L., Cai, J., Zou, B., Wu, C. F., Fu, Q., ... Kan, H. (2015). A land use regression model for estimating the NO2 concentration in Shanghai, China. *Environmental Research, 137*, 308–315.

Muttoo, S., Ramsay, L., Brunekreef, B., Beelen, R., Meliefste, K., & Naidoo, R. N. (2018). Land use regression modelling estimating nitrogen oxides exposure in industrial South Durban, South Africa. *Science Total Environment, 610-611*, 1439–1447.

Rao, N. V., Rajasekhar, M., & Rao, G. C. (2014). Detrimental effect of air pollution, corrosion on building materials and historical structures. *American Journal of Engineering Research, 3*(03), 359–364.

Ross, Z., Jerrett, M., Ito, K., Tempalski, B., & Thurston, G. (2007). A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmospheric Environment, 41*(11), 2255–2269.

She, Q., Peng, X., Xu, Q., Long, L., Wei, N., Liu, M., ... Xiang, W. (2017). Air quality and its response to satellite-derived urban form in the Yangtze River Delta, China. *Ecological Indicators, 75*, 297–306.

Shen, J., Gao, Z., Ding, W., & Yu, Y. (2017). An investigation on the effect of street morphology to ambient air quality using six real-world cases. *Atmospheric Environment, 164*, 85–101.

Shi, Y., Lau, K. K., & Ng, E. (2016b). Developing street-level PM2.5 and PM10 land use regression models in high-density Hong Kong with urban morphological factors. *Environmental Science & Technology, 50*(15), 8178–8187.

Shi, Y., Lau, K. K.-L., & Ng, E. (2016a). Developing street-level PM2. 5 and PM10 land use regression models in high-density Hong Kong with urban morphological factors. *Environmental Science & Technology, 50*(15), 8178–8187.

Shi, Y., Xie, X., Fung, J. C.-H., & Ng, E. (2018). Identifying critical building morphological design factors of street-level air pollution dispersion in high-density built environment using mobile monitoring. *Building and Environment, 128*, 248–259.

Silva, L. T., & Monteiro, J. P. (2016). The Influence of Urban Form on Environmental Quality within a Medium-sized City. *Procedia Engineering, 161*, 2046–2052.

Silva, M., Oliveira, V., & Leal, V. (2016). *Urban morphology and energy: Progress and prospects.*

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(sup1), 234–240.

Wang, J., Hu, Z., Chen, Y., Chen, Z., & Xu, S. (2013). Contamination characteristics and possible sources of PM10 and PM2. 5 in different functional areas of Shanghai, China. *Atmospheric Environment, 68*, 221–229.

Wang, J., & Xu, C. (2017). Geodetector: Principle and prospective. *Acta Geographica*

Sinica, 72(1), 116–134.
Wang, J.-F., Christakos, G., & Hu, M.-G. (2009). Modeling spatial means of surfaces with stratified nonhomogeneity. *IEEE Transactions on Geoscience and Remote Sensing, 47*(12), 4167–4174.
Wang, J. F., Li, X. H., Christakos, G., Liao, Y. L., Zhang, T., Gu, X., & Zheng, X. Y. (2010). Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *International Journal of Geographical Information Science, 24*(1), 107–127.
Wang, J.-F., Zhang, T.-L., & Fu, B.-J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators, 67*, 250–256.
Wang, S., Tian, Y., Zhou, Y., Liu, W., & Lin, C. (2016). Fine-scale population estimation by 3D reconstruction of urban residential buildings. *Sensors, 16*(10), 1755.
WHO (2014). *Ambient (outdoor) air quality and health.* WHO Media Centre313.
Wu, J., Li, J., Peng, J., Li, W., Xu, G., & Dong, C. (2015). Applying land use regression model to estimate spatial variation of PM(2).(5) in Beijing, China. *Environmental Science Pollution Research International, 22*(9), 7045–7061.
Yang, H., Chen, W., & Liang, Z. (2017). Impact of land use on PM2.5 pollution in a representative City of Middle China. *International Journal of Environmental Research and Public Health, 14*(5).
Yang, X., Zheng, Y., Geng, G., Liu, H., Man, H., Lv, Z., ... de Hoogh, K. (2017). Development of PM2.5 and NO_2 models in a LUR framework incorporating satellite remote sensing and air quality model data in Pearl River Delta region, China. *Environmental Pollution, 226*, 143–153.
Yuan, C., Ng, E., & Norford, L. K. (2014). Improving air quality in high-density cities by understanding the relationship between air pollutant dispersion and urban morphologies. *Building and Environment, 71*, 245–258.
Yuan, Y., Liu, S., Castro, R., & Pan, X. (2012). PM2. 5 monitoring and mitigation in the cities of China. *Environmental Science & Technology,* 3627–3628.
Zhang, X., & Hu, H. (2017). Improving satellite-driven PM2.5 models with VIIRS nighttime light data in the Beijing–Tianjin–Hebei Region, China. *Remote Sensing, 9*(9).
Zheng, S., Zhou, X., Singh, R., Wu, Y., Ye, Y., & Wu, C. (2017). The spatiotemporal distribution of air pollutants and their relationship with land-use patterns in Hangzhou City, China. *Atmosphere, 8*(6).