Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Sustainable Cities and Society 61 (2020) 102329

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Spatial distribution characteristics of $PM_{2.5}$ and PM_{10} in Xi'an City predicted by land use regression models

Li Han[a], Jingyuan Zhao[a,*], Yuejing Gao[a], Zhaolin Gu[b,*], Kai Xin[a], Jianxin Zhang[a]

[a] Chang An Univ., Coll Architecture, 161 Chang An Rd., Xian, 710061, Shaanxi, People's Republic of China
[b] Xi An Jiao Tong Univ., Sch Human Settlement & Civil Engn, Xian, 710049, Shaanxi, People's Republic of China

ARTICLE INFO

Keywords:
Land use regression
Xi'an City
Fine particulate matter ($PM_{2.5}$)
Respirable particulate matter (PM_{10})
Spatial distribution

ABSTRACT

$PM_{2.5}$ and PM_{10} could increase the risk for cardiovascular and respiratory diseases in the general public and severely limit the sustainable development in urban areas. Land use regression models are effective in predicting the spatial distribution of atmospheric pollutants, and have been widely used in many cities in Europe, North America and China. To reveal the spatial distribution characteristics of $PM_{2.5}$ and PM_{10} in Xi'an during the heating seasons, the authors established two regression prediction models using $PM_{2.5}$ and PM_{10} concentrations from 181 monitoring stations and 87 independent variables. The model results are as follows: for $PM_{2.5}$, R^2 = 0.713 and RMSE = 8.355 μg/m^3; for PM_{10}, R^2 = 0.681 and RMSE = 14.842 μg/m^3. In addition to the traditional independent variables such as area of green space and road length, the models also include the numbers of pollutant discharging enterprises, restaurants, and bus stations. The prediction results reveal the spatial distribution characteristics of $PM_{2.5}$ and PM_{10} in the heating seasons of Xi'an. These results also indicate that the spatial distribution of pollutants is closely related to the layout of industrial land and the location of enterprises that generate air pollution emissions. Green space can mitigate pollution, and the contribution of traffic emission is less than that of industrial emission. To our knowledge, this study is the first to apply land use regression models to the Fenwei Plain, a heavily polluted area in China. It provides a scientific foundation for urban planning, land use regulation, air pollution control, and public health policy making. It also establishes a basic model for population exposure assessment, and promotes the sustainability of urban environments.

1. Introduction

The rapid economic development and accelerated urbanization process in China have been accompanied by high energy consumption and excessive pollutant emission, this has caused serious air pollution issues and hindered the sustainable development of urban areas (Liu, Sun, & Feng, 2020; Ortolani & Vitale, 2016). As main air pollutants, $PM_{2.5}$ and PM_{10} are the most harmful to human health and thus, of the most concern to the general public. They are the focus of smog control in China at the current stage. The concentrations of $PM_{2.5}$ and PM_{10} are impacted by urban space morphology, land use layout, and adverse meteorological factors and are thereby likely to accumulate in cities (Jin et al., 2019). Long-term exposure to contaminated atmosphere increases the risk of cardiovascular and respiratory diseases (Barzeghar, Sarbakhsh, Hassanvand, Faridi, & Gholampour, 2020; Berman, Burkhardt, Bayham, Carter, & Wilson, 2019; Feng, Gao, Liao, Zhou, & Wang, 2016), Properly planned commuting routes can reduce human exposure to pollution (Ahmed, Adnan, Janssens, & Wets, 2020; Pilla & Broderick, 2015; Qiu et al., 2017). The variations in spatial distribution of urban air pollutants have become a widespread concern in several fields such as urban and rural planning, environmental science, and medicine (Son et al., 2018; Yang et al., 2020; Yuan, Song, Huang, Shen, & Li, 2019; Zou, Wilson, Zhan, & Zeng, 2009).

Land use regression (LUR) models have proven to be an effective method for predicting the spatial distribution of pollutants (Jerrett et al., 2005). LUR models work using pollutant concentration data collected at a limited number of monitoring stations in conjunction with characteristic variables such as land use information to evaluate pollutant concentrations in areas that lack monitoring stations. A main feature of LUR models is the correlation of land use characteristics to the concentrations of target pollutants, which can be used to determine the relationship between air pollutant concentrations and other geographic variables to simulate the spatial distribution of air pollutants in urban areas and identify the causes of pollutants to a certain extent

* Corresponding authors.
E-mail addresses: 2017041002@chd.edu.cn (L. Han), zjyqtt@chd.edu.cn (J. Zhao), 2016041002@chd.edu.cn (Y. Gao), guzhaoln@mail.xjtu.edu.cn (Z. Gu), 2018041003@chd.edu.cn (K. Xin), zhangjx@chd.edu.cn (J. Zhang).

https://doi.org/10.1016/j.scs.2020.102329
Received 24 January 2020; Received in revised form 2 June 2020; Accepted 8 June 2020
Available online 13 June 2020

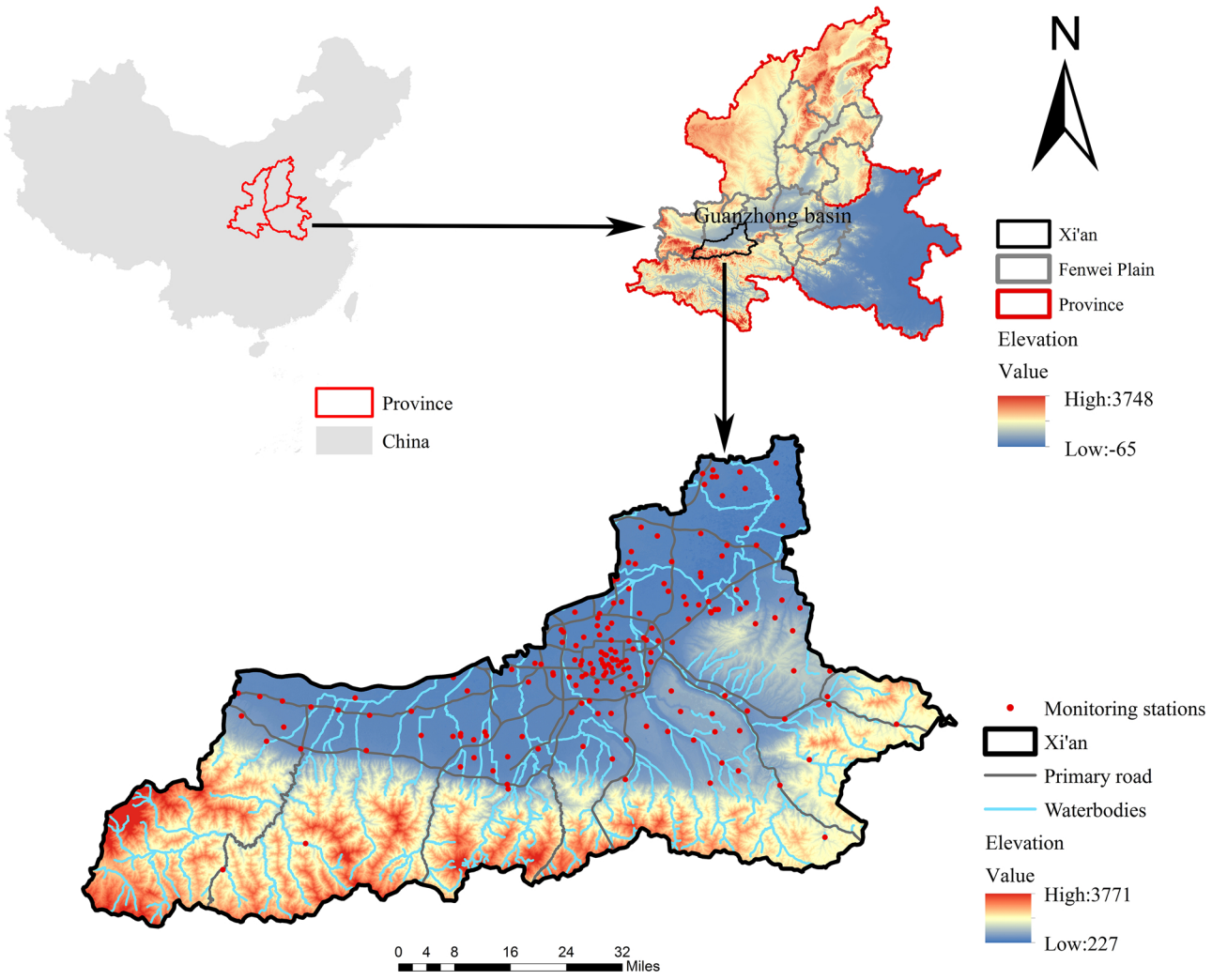

Fig. 1. Study area and distribution of monitoring stations.

(Henderson, Beckerman, Jerrett, & Brauer, 2007; Hoek et al., 2008). The interpretation of the relationship can guide urban land use adjustment. By adjusting the layout of industrial land, intensive land use is achieved to reduce pollutant emission, and ultimately accomplish sustainable urban land use.

One of the earliest LUR models was developed by Briggs et al. (1997) to predict NO_2 concentrations in three European cities and plot pollution maps. Owing to the simplicity of model construction and ease of data acquisition as well as improvement in the modeling technology and content and higher diversity in the allowed variable types, LUR models have also been applied to the prediction of air pollutants such as NO_2, PM_{10}, and $PM_{2.5}$ in Europe and North America (Allen, Amram, Wheeler, & Brauer, 2011; Briggs et al., 2000; Eeftens et al., 2012; Moore, Jerrett, Mack, & Kunzli, 2007; Song, Jia, Li, Tang, & Wang, 2019). Two early studies in 2010 (Chen, Bai et al., 2010, 2010b) predicted the concentrations of PM_{10}, NO_2, and SO_2 in two Chinese cities (Tianjin and Jinan) and plotted pollutant distribution maps. Since 2013, China has begun to gradually establish and improve its air quality monitoring network, an action accompanied by a growing body of LUR-model-based studies investigating the spatial distribution of air pollutants in Chinese cities such as Chengdu, Changsha, Beijing, and Shanghai, with a particularly large number of studies in the latter two cities (Ji, Wang, & Zhuang, 2019; Liu et al., 2015; Meng et al., 2015, 2016; Wu et al., 2015; Xiao, Wang, Wu, Fu, & Zhu, 2018). Compared with the research results of other cities, this study is the first to apply land use regression models to the Fenwei Plain, a region in China with severe air pollution. The number of monitoring stations used in this study is the greatest so far. The dependent variables are selected specifically during the heating seasons. It explains the reasoning behind the selection of buffer zones. It applies comprehensively all aspects of regression model diagnosis, cross-validation and verification of LUR applicability in different heating seasons. Hence, to a certain extent, it has enriched the application of land use regression models in the prediction of the spatial distribution of atmospheric pollutants.

Xi'an is an important western Chinese city located on the west of the rift basin of the Fenwei Plain, encompassing 11 cities; a plain subjected to smog in large areas and considered one of the most atmospherically polluted areas in China. The national air quality report released by the Ministry of Ecology and Environment of China revealed that between January and December 2018, Xi'an ranked 158 out of 169 cities in terms of air quality and was under severe air pollution. Xi'an has special natural conditions (i.e., unique topography and unfavorable meteorological conditions) with a special land use status, economic development level, and industrial layout. Thus, the air pollution is further exacerbated (Huang, Zhang, Tang, & Liu, 2015; Song et al., 2015). Therefore, a deep insight into the spatial distribution characteristics of the concentrations of $PM_{2.5}$ and PM_{10} during the heating seasons of Xi'an is the key to supporting air pollution control.

The purpose of this study is to establish a regression prediction model for $PM_{2.5}$ and PM_{10} in the heating seasons of Xi'an to test the applicability of the land use regression models, and to reveal the spatial distribution characteristics of these pollutants. Further, it aims to

analyze the relationship between the spatial distributions of $PM_{2.5}$ and PM_{10} and land use characteristics and therefore provides a scientific foundation for urban planning, land use regulation, air pollution control, and public health policy-making. It also establishes a basic model for population exposure assessment. The application of the land use regression models to areas requiring heating in winter and heavily polluted areas plays a positive role in achieving sustainable urban environment and promoting sustainable development in urban environments.

2. Materials and methods

The air quality data used in this study were the daily mean concentrations of $PM_{2.5}$ and PM_{10} at 181 air quality monitoring stations under the Xi'an Ecology and Environment Bureau. The concentration measurements were conducted during the winter heating season from November 15, 2018 to March 15, 2019, and the period-averaged concentrations of $PM_{2.5}$ and PM_{10} obtained from each monitoring station were used as the dependent variables, as depicted in Fig. 1. A total of 86 factors in the five categories of land use, road traffic facilities, socio-economic development, emission source, and geospatial information were considered as independent variable candidates. For a particular monitoring station, independent variable candidates were extracted in two ways: (1) with the monitoring station as the center, circular buffer zones at various distances were delineated using GIS, and the independent variable candidates were the length, number, or area within the buffer zones, Fig. 2 shows the length of the roads in the 3000 m buffer zone around the New Software City station; (2) the independent variable candidates were the distances from the monitoring stations to notable objects or the characteristic values of monitoring stations, Fig. 3 displays the distance from the New Software City station to the nearest highways. After the extraction of variables, correlation analysis was performed between the independent and dependent variables using SPSS software. Variable screening was performed based on the magnitude of the Pearson correlation coefficients, and the selected independent variables were included in multiple stepwise regression analyses using R software. The established models were subjected to cross-validation to test their generalizability using R software. The procedure followed in this study is depicted in Fig. 4.

2.1. Study area and its characteristics

Xi'an, the capital of Shaanxi Province and an important central city in western China, is located in the Guan zhong Basin in the middle of Yellow River watershed and has the largest variations in elevation

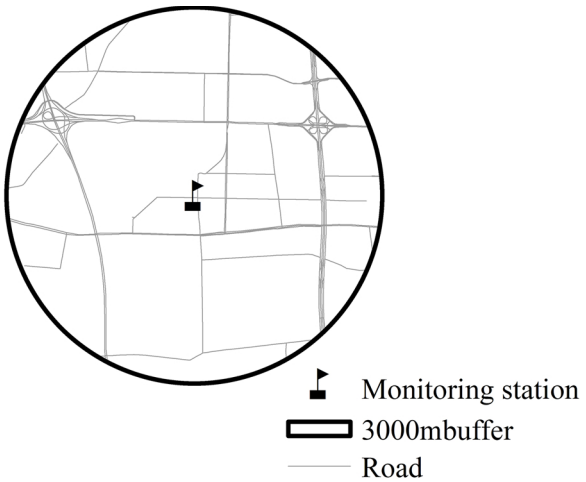

Fig. 2. Length of the roads in the 3000 m buffer zone around the New Software City station.

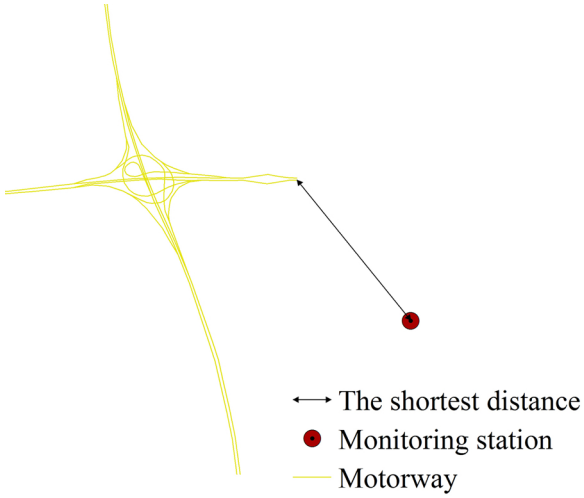

Fig. 3. Distance from the New Software City station to the nearest highways.

within its administrative areas among all Chinese cities. Xi'an is composed of 11 districts with two counties and has been entrusted the administration of the Xi Xian New District, having a total area of 10,752 km^2, approximately 204 km long in the east-west direction and approximately 116 km wide in the south-north direction. By the end of 2018, the resident population had reached 10,003,700, and the city's GDP in 2018 was 834.986 billion Chinese RMB. The prevailing winds in Xi'an are the northeasterly winds with low speed in Fig. 5. Thus, the meteorological conditions are not favorable for the diffusion of pollutants. A stagnation zone of anticyclonic airflow is prone to form owing to obstruction by mountains and the sinking of the leeward airflow. When pollution occurs, the pollutants accumulate mostly near the surface, where diffusion is difficult. The Fenwei Plain is located in a valley region, and any city in the plain is under the direct impact of pollutant emissions from other cities. In terms of the energy consumption structure, coal consumption accounts for a relatively high proportion of total energy consumption in the Fenwei Plain. Shaanxi and Shanxi Provinces are large coal producers as well as consumers. In particular, coal consumption is more centralized in the Fenwei Plain, accounting for nearly 90 % of its total energy consumption, far greater than the national average of 60 %.

2.2. Dependent variables

The daily mean concentrations of $PM_{2.5}$ and PM_{10} obtained from 181 air quality monitoring stations of Xi'an Ecology and Environment Bureau were used in this study, which were measured during the winter heating season from November 15, 2018 to March 15, 2019, with period-averaged concentrations at each station used as the dependent variables. As illustrated in Figs. 6 and 7, the concentrations of $PM_{2.5}$ and PM_{10} from the 181 stations display a normal distribution. The ean is the average value of the concentrations, and Std. Dev. is the standard deviation of the concentrations. Xi'an has cold weather in winter, and the local government implements central a heating system, which requires combustion of a substantial amount of coal and natural gas. This aggravates the $PM_{2.5}$ and PM_{10} pollution in the city. In addition, due to frequent calm days in winter plus the adverse terrain in Xi'an for pollution dispersion, the occurrence of heavy pollution is concentrated during this period, causing great harm to human health. Considering these factors, the authors select the $PM_{2.5}$ and PM_{10} concentration data of the heating season from November 15, 2018 to March 15, 2019.

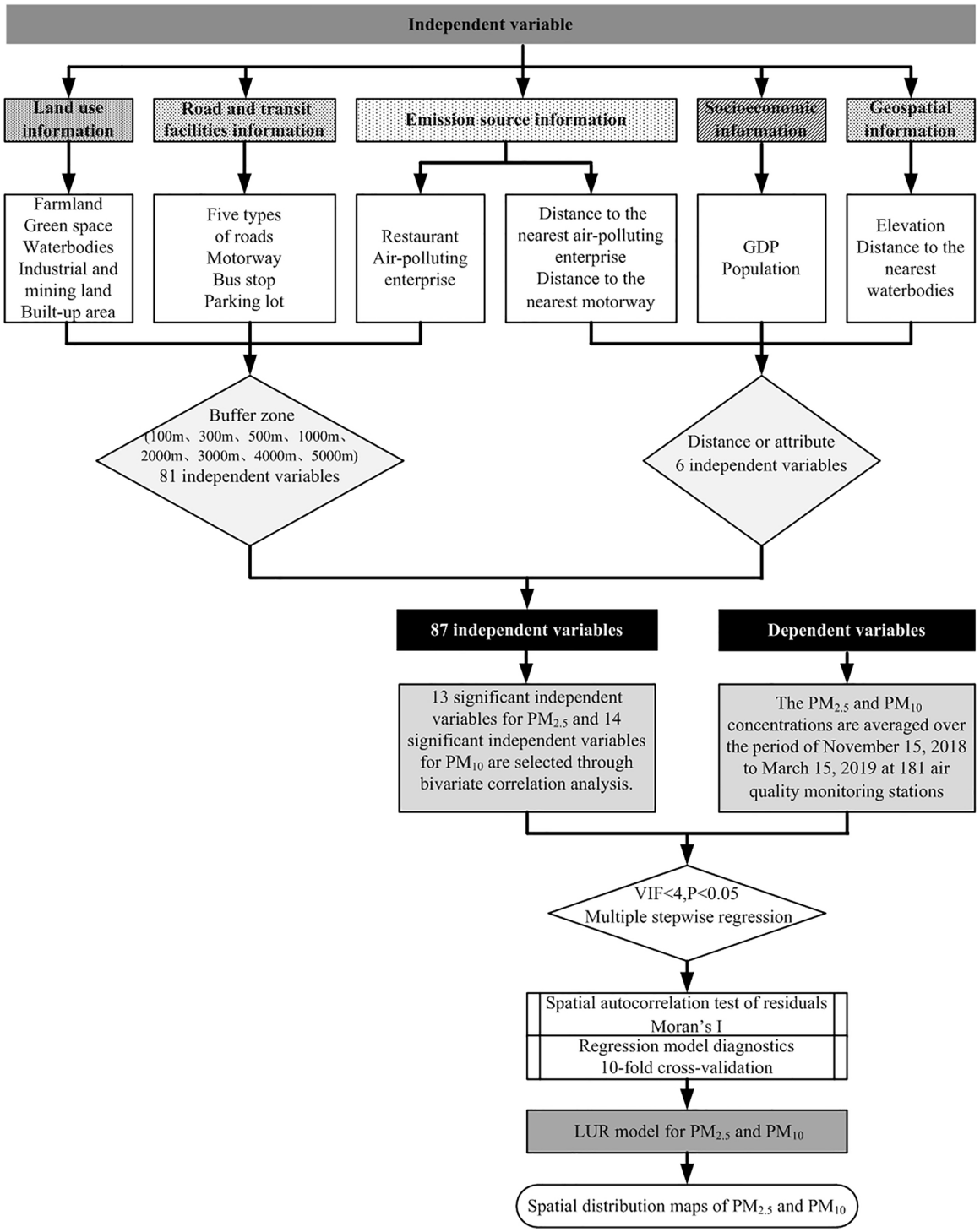

Fig. 4. Research flow chart.

2.3. Independent variables

2.3.1. Land use information

The land use data used in this study were derived from the remote sensing monitoring data of China's land use, built by the Chinese Academy of Sciences. The Landsat TM/ETM/OLI remote sensing images are the main data source. After going through processes such as image fusion, geometric correction, image enhancement and stitching etc., the land use types were classified into 6 first-level categories, 25 second-level categories and certain third-level categories according to the land use/cover classification system in China, via human-computer interactive visual interpretation. In this study, five types of land-use data were extracted, namely, those from farmland (paddy fields + dry land), green land (forest land + grassland), waterbodies (waterways + lakes + reservoirs + ponds), industrial and mining land (factories and mines + industrial areas + airports), and construction land (urban land + rural residential areas). The shape and size selection of a buffer zone was based on the diffusion range of the pollutants in the atmosphere and the impact of geographical elements on the pollutants. However, because of the complexity and uncertainty of atmospheric pollution diffusion, previous research results were usually referenced when determining a new buffer zone. In this study, the correlation

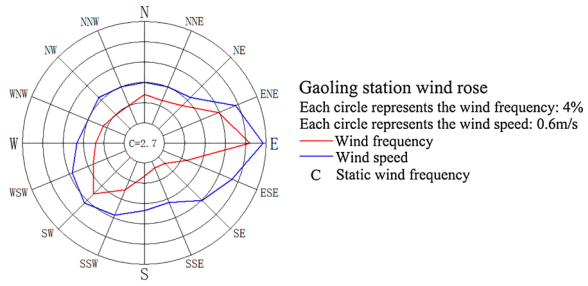

Fig. 5. Wind Rose of Gaoling Station from November 15, 2018 to March 15, 2019.

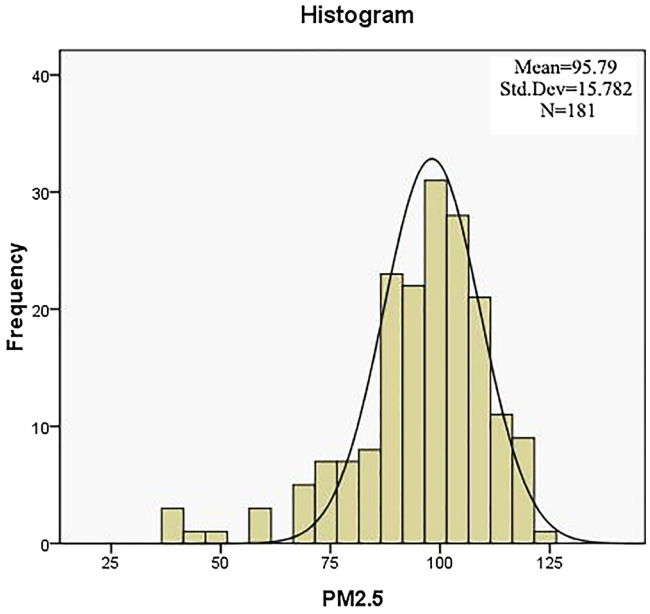

Fig. 6. Histogram of $PM_{2.5}$ concentration frequency distribution.

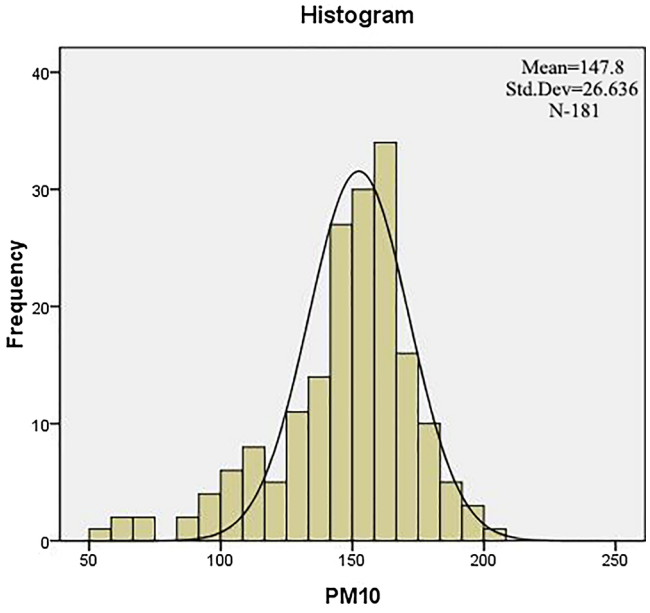

Fig. 7. Histogram of PM_{10} concentration frequency distribution.

coefficients of most independent variables increased with respect to pollutant concentrations as the buffer zones increased in size until the distance of 5000 m, a limiting distance beyond which the correlation coefficients would not increase. Thus, the maximum distance of buffer zones was set to 5000 m. GIS software was employed to establish station-centered circular buffer zones for each of the 181 monitoring stations at successive distances of 100, 300, 500, 1000, 2000, 3000, 4000, and 5000 m and then to extract the length or area data associated with each type of land use in each buffer zone.

2.3.2. Road traffic facility information

Road traffic facility information referred to road networks, parking lots, and bus stops. In this study, five types of road network data were extracted from OpenStreetMap, namely, motorways, primary roads, secondary roads, tertiary roads, and trunk lines. The road network data in a buffer zone were extracted using two metrics: extraction of the length of motorways in the buffer zone or extraction of the total length of the five types of roads in the buffer zone. The parking lot data and bus stop data were obtained from Gaode map, i.e., the number of parking lots and bus stops in each buffer zone were calculated.

2.3.3. Socioeconomic information

The socioeconomic information consisted of GDP and population data. The GDP data came from the kilogram grid dataset of the spatial distribution of China's GDP (GDP Grid China) constructed by the Chinese Academy of Sciences. It has a raster data form, with each raster representing the total GDP output value within the grid range (1 square kilometer) in the unit of ten thousand yuan/km^2. The GDP value of each grid with a monitoring site was extracted from GIS data. The population data were derived from the sixth national census and used to calculate the population of each abovementioned sub-district or town.

2.3.4. Emission source information

The emission sources consisted of air-polluting enterprises, restaurants, and motorways. Air-polluting enterprise information was obtained from the Monitoring Information Release Platform of Key Pollutant-discharging Enterprises in Shaanxi Province and the List of Key Pollutant-discharging Units released by Xi'an Ecology and Environment Bureau. There were 37 pollutant-discharging enterprises. The distance between each monitoring station and the nearest pollutant-discharging enterprise from it was calculated. Also the number of pollutant-discharging enterprises in each buffer zone was counted. Restaurant data were obtained from Gaode map by calculating the number of restaurants in each buffer zone. Motorway data were obtained by calculating the distance from each monitoring station to its nearest motorway.

2.3.5. Geospatial information

Geospatial information consisted of the elevation of each monitoring station and its distance to a water surface. Elevation data were obtained by extracting Digital Elevation Model (DEM) topographic data with a precision of 30 m. Considering the effect of water–land breeze on atmospheric pollution transport, the distance from each monitoring station to the nearest water bodies was calculated.

2.4. Model construction

A total of 87 independent variables were extracted. Bivariate correlation analysis was performed among the 87 independent variables and the dependent variables $PM_{2.5}$ and PM_{10} using SPSS 17.0 to obtain the Pearson correlation coefficients of each independent variable with respect to $PM_{2.5}$ and PM_{10}. Independent variables that are significantly correlated to the dependent variables ($P < 0.05$) were selected. In each category of independent variables, the variables with the largest Pearson correlation coefficient were selected, which led to the selection of 13 independent variables for $PM_{2.5}$(Table 1) and 14 for PM_{10} (Table 2). Multiple stepwise regression of $PM_{2.5}$ and PM_{10} was separately performed on these independent variables using R software to remove independent variables with a p-value > 0.01 while performing

Table 1
Independent variables for $PM_{2.5}$ LUR models.

Category	Sub-category	Buffer zone	Unit	Code	Mean	Standard deviation	Min	Max
Land use information	Green space	4000 m	m^2	GS-4000	5991739	9151113	0	45787914
	Industrial and mining land	5000 m	m^2	IM-5000	1796676	2565877	0	12036045
	Built-up area	1000 m	m^2	BA-1000	1639066	1146245	0	3141593
Road traffic information	Major road	5000 m	m	MR-5000	186615	121895	6738	417015
	Parking lot	5000 m	Number	PA-5000	1013	1548	0	5080
	Bus stop	4000 m	Number	BS-4000	98	108	0	354
Socioeconomic information	Population	No	Number	PO	58588	46813	2264	223840
Emission source information	Distance to the nearest air-polluting enterprise	No	m	DIS-PE	8076	7920	245	44126
	Distance to the nearest motorway	No	m	DIS-MO	4528	5094	21	43354
	Restaurant	5000 m	Number	RE-5000	3122	4489	0	14736
	Air-polluting enterprise	5000 m	Number	PE-5000	0.91	1.14	0	5
Geospatial information	Elevation	No	m	EL	467	154	345	1124
	Distance to nearest water bodies	No	m	DIS-WA	2716	2209	17	8796

collinearity diagnostics to remove independent variables with Variance Inflation Factor (VIF) > 4.

2.5. Model diagnostics

The models were subjected to regression diagnostics using R software. For regression diagnostics, the results of model fitting were presented in four plots: (1) a Residual-versus-Fitted plot to test the assumption that the independent variable in question was linearly correlated to the dependent variable in question; (2) a Normal Q-Q plot to test the normality of residuals; (3) a Scale-Location plot, intended to test the homoscedasticity assumption; and (4) a Residual-versus-Leverage plot to identify outliers, high-impact points, and high-leverage points.

2.6. Model cross-validation

The models were subjected to cross-validation using R software. Model validation was performed using the 10-fold cross-validation method aimed at testing the generalizability of the models. Thus, the samples in question were divided into 10 equal-sized subsamples, of which one subsample was retained as the validation group, whereas the remaining nine subsamples were used as the training group; repeating this process in turn for each subsample as the validation group finally led to a total of 10 prediction equations, whose R^2 values and RMSEs were averaged. The equations for RMSE and R^2 calculations are listed as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_{test}^{(i)} - \hat{y}_{test}^{(i)})^2}$$

$$R^2 = 1 - \frac{(\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2)/m}{(\sum_{i=1}^{m}(y^{(i)} - \bar{y})^2)/m}$$

In the RMSE equation, $y_{test}^{(i)}$ represents the monitoring concentration of the i-th monitoring station in the test set, $\hat{y}_{test}^{(i)}$ refers to the predicted concentration of the i-th monitoring station in the test set, and m is the number of monitoring stations in the test set. In the R^2 equation, $\hat{y}^{(i)}$ and $y^{(i)}$ represent the predicted and monitoring concentrations of the i-th monitoring station respectively in the training set, $\bar{y}$ represents the average concentration of the training set, and m indicates the number of monitoring stations in the training set.

2.7. Industrial land

The results section analyzed the relationship between the industrial land layout of the study area and the Regulatory Management Zone (RMZ) and the spatial distributions of $PM_{2.5}$ and PM_{10} in the heating seasons in Xi'an. The definition of industrial land in urban and rural planning is applied here, which refers to land for production shops, warehouses and auxiliary facilities of industrial and mining enterprises, including land for special railways, docks and auxiliary roads, and parking lots. The information of industrial land layout was from the central urban area land use plan map from Xi'an's Urban Master Plan (2008–2020) and the remote sensing monitoring data of China's land use.

Table 2
Independent variables for PM_{10} LUR models.

Category	Sub-category	Buffer zone	Unit	Code	Mean	Standard deviation	Min	Max
Land use information	Green space	5000 m	m^2	GS-5000	9814783	14460656	0	71238521
	Industrial and mining land	5000 m	m^2	IM-5000	1796676	2565877	0	12036045
	Built-up area	1000 m	m^2	BA-1000	1639066	1146245	0	3141593
	Water bodies	5000 m	m^2	WA-5000	2119489	3060745	0	14499236
Road traffic information	Motorway	5000 m	m	MO-5000	19906	21388	0	88843
	Major road	5000 m	m	MR-5000	186615	121895	6738	417015
	Bus stop	5000 m	Number	BS-5000	145	162	0	550
Socioeconomic information	Population	No	Number	PO	58588	46813	2264	223840
Emission source information	Distance to nearest air-polluting enterprise	No	m	DIS-PE	8076	7920	245	44126
	Distance to nearest motorway	No	m	DIS-MO	4528	5094	21	43354
	Restaurant	5000 m	Number	RE-5000	3122	4489	0	14736
	Air-polluting enterprise	5000 m	Number	PE-5000	0.91	1.14	0	5
Geospatial information	Elevation	No	m	EL	467	154	345	1124
	Distance to nearest water bodies	No	m	DIS-WA	2716	2209	17	8796

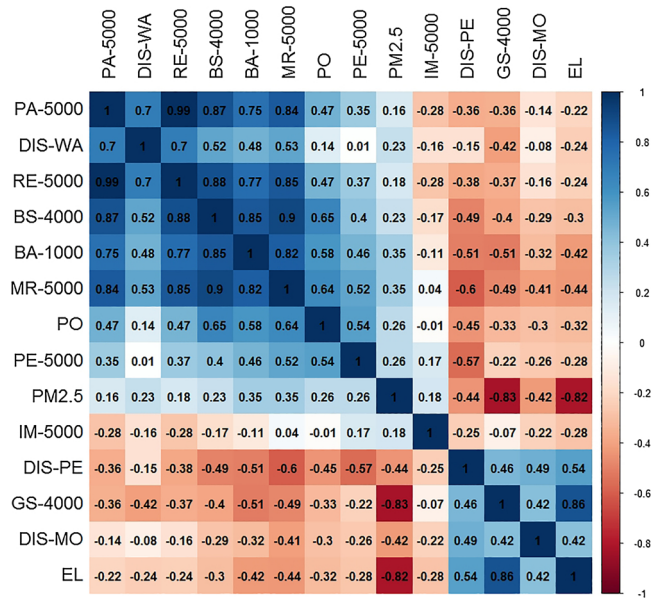

Fig. 8. The matrix diagram of variable correlation coefficient of $PM_{2.5}$ LUR model.

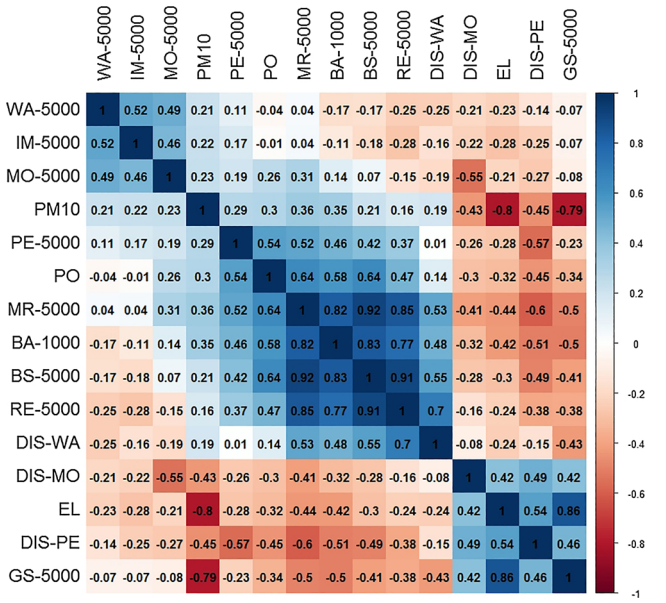

Fig. 9. The matrix diagram of variable correlation coefficient of PM_{10} LUR model.

3. Results

3.1. Significant independent variables

Bivariate correlation analysis indicated that 44 of the 87 independent variables were significantly correlated to $PM_{2.5}$ and 42 to PM_{10}. Among these 44 and 42 independent variables, those with the largest correlation coefficients in the respective categories were selected, ultimately leading to 13 and 14 independent variables selected for $PM_{2.5}$ and PM_{10}, respectively, which included 10 in common, as presented in Tables 1 and 2.

The independent variables demonstrated positive correlation to the concentration of $PM_{2.5}$, except for four independent variables, which demonstrated negative correlation, namely, GS-4000, reflective of the area of green spaces in buffer zones at a distance of 4000 m; DIS-PE, reflective of the distances from the stations to their respective nearest air-polluting enterprises; DIS-MO, reflective of the distances from the stations to their respective nearest motorways; and EL, reflective of the elevation of the stations, as depicted in Fig. 8. Similarly, GS-5000, DIS-PE, DIS-MO, and EL demonstrated negative correlation to the concentration of PM_{10}, as depicted in Fig. 9. The negative correlation coefficient r of GS-4000 with respect to $PM_{2.5}$ was the largest among the four negative correlation coefficients with respect to $PM_{2.5}$. The negative correlation coefficient r of GS-5000 with respect to PM_{10} was smaller than that of only EL, indicating that larger the green spaces in buffer zones, smaller were the concentrations of $PM_{2.5}$ and PM_{10} at the stations. The negative correlation coefficient r of EL with respect to $PM_{2.5}$ and PM_{10} was the second largest (after that of GS-4000) and the largest, respectively, which was attributed to the special terrain of Xi'an, i.e., stations with higher EL had larger green spaces. The correlation coefficient r of DIS-PE and DIS-MO with respect to the pollutants was large and significantly negative, indicating that closer the air-polluting enterprises and motorways, higher were the concentrations of $PM_{2.5}$ and PM_{10}.

3.2. LUR models

3.2.1. Multiple stepwise regression

Bivariate correlation analysis confirmed that three independent variables, GS-4000, RE-5000, and PE-5000, should be included in the $PM_{2.5}$ LUR model. The regression coefficient was significant at the $p < 0.05$ level, and the adjusted R^2 value (Adj-R^2) was 0.713, indicating that 71.3 % of the variance of the concentration of $PM_{2.5}$ was accounted for by the model. The model RMSE was 8.355 μg/m^3 and the VIF of each variable was less than 4, indicating that there was no multicollinearity among the three independent variables (Table 3). The observed and predicted concentrations of $PM_{2.5}$ were compared (Fig. 10).

For the PM_{10} LUR model, bivariate correlation analysis confirmed that four independent variables, GS-5000, MO-5000, BS-5000, and PE-5000, should be included in the model. The regression coefficient was significant at the $p < 0.05$ level, and the Adj-R^2 was 0.681, indicating that 68.1 % of the variance of the concentration of PM_{10} was accounted for by the model. The model RMSE was 14.842 μg/m^3, and the VIF of each variable was less than 4, indicating that there was no multicollinearity among the four independent variables (Table 4). The observed and predicted concentrations of PM_{10} were compared (Fig. 11).

3.2.2. Model evaluation

The model evaluation measures included regression diagnostics, cross-validation, and the spatial autocorrelation test. Regression diagnostics were performed by checking four plots, namely, a residual-versus-fitted plot, normal Q-Q plot, scale-location plot, and residual-versus-leverage plot. Figs. 12a and 13 a depict that there was no correlation between the residuals and fitted values, and the dependent and independent variables were linearly correlated. In Figs. 12b and 13 b all the data points of the plots fall on a straight line at an angle of 45°, indicating the fulfillment of the normality assumption. Figs. 12c and 13 c present a random distribution of data points around the horizontal line, indicating the fulfillment of the homoscedasticity assumption. Fig. 12d depicts that the $PM_{2.5}$ regression model was free of outliers, highly influential points, and high-leverage points. Fig. 13d indicates that for the PM_{10} model there was an outlier data point—sample 164, but it was not advisable to delete this outlier. In the multivariate LUR models for $PM_{2.5}$ and PM_{10}, the coefficient of multiple determination (Mul-R^2), Adj-R^2, and RMSE were all less than the respective averages in the 10-fold cross-validation, which indicated that there was no overfitting and underfitting and the LUR models had good generalizability. For the spatial autocorrelation test of $PM_{2.5}$ residuals, the p-value was greater than 0.05 (95 % confidence level) and the z-score was above the threshold of -1.65, indicating that the $PM_{2.5}$ residuals were randomly distributed in space without spatial clustering. For the PM_{10} residuals, the p-value of the spatial autocorrelation test was greater

Table 3
Results of multiple stepwise regression for $PM_{2.5}$.

	Estimate	Std error	*t* value	Pr (> |*t*|)	VIF	Result
Intercept	1.051e+02	1.074e+00	97.857	< 2e-16***		Multiple *R*-squared:
GS-4000	−1.494e-06	7.444e-08	−20.073	< 2e-16***	1.17	0.718
RE-5000	−6.894e-04	1.590e-04	−4.336	2.44e-05***	1.28	Adjusted *R*-squared: 0.713
PE-5000	1.964e+00	5.975e-01	3.288	0.00122**	1.17	RMSE : 8.355 μg/m^3

Note: ***, ** and * indicate significant levels of significance at 0, 0.001, and 0.01 respectively.

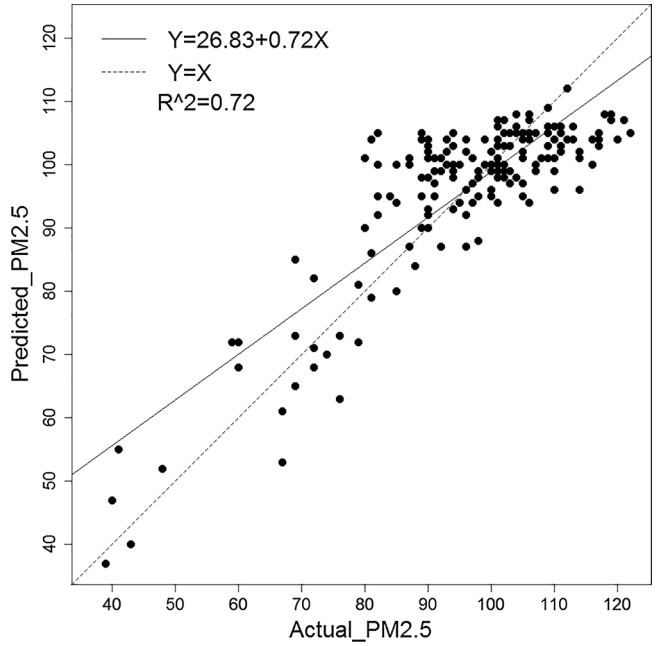

Fig. 10. Scatter plots showing observed and predicted $PM_{2.5}$ using LUR model.

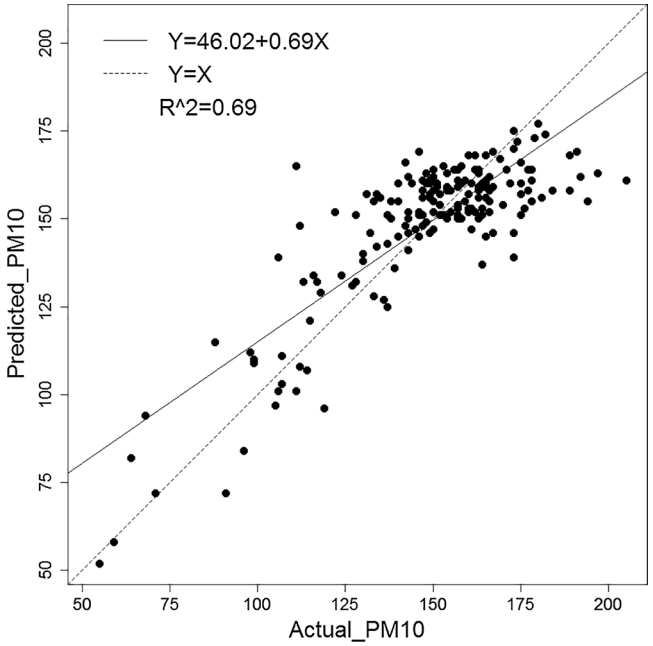

Fig. 11. Scatter plots showing observed and predicted PM_{10} using LUR model.

than 0.05 (95 % confidence level) and the *z*-score did not exceed the threshold of 1.65, indicating that the PM_{10} residuals were also randomly distributed in space without spatial clustering (Table 5).

3.2.3. Model verification for different heating seasons

To verify the applicability of the land use regression models in different heating seasons in Xi'an, the authors used the pollutant data from the recent heating season from November 15, 2019 to January 24, 2020 for verification. To exclude the impact of the novel coronavirus pandemic, only part of the data for the entire heating season of November 15, 2019-March 15, 2020 was selected. Affected by the pandemic, since Jan 24, 2020, most factories have been shut down, road traffic has declined sharply, and restaurants have been closed. The reduction in emission sources has a great impact on the distribution of pollutants. Compared with the model for the previous heating season, the independent variables of the $PM_{2.5}$ regression model of the recent heating season are the same as those of the previous heating season. Adjustable R^2 = 0.636, which is lower than the value of the previous season at 0.713; RMSE =0.759 μg/m^3 is, also lower than 8.355 μg/m^3, the value of the previous season, as shown in Table 6. Table 7 shows that the independent variables of the PM_{10} regression model of the recent heating season are the same as those of the previous heating season. Adjustable R^2 = 0.639, lower than 0.681, the value of the previous heating season; RMSE = 12.704 μg /m^3, also lower than the value from the previous heating season at 14.842 μg/m^3. The $PM_{2.5}$ and PM_{10} land use regression models for the recent heating season exhibit reduced accuracy in fitting and smaller errors, but they still have good prediction capabilities, proving the applicability of the land use regression models in different heating seasons.

3.3. Spatial distribution of $PM_{2.5}$ and PM_{10}

Xi'an was divided into a grid of 10,525 cells using ArcGIS, each of 1 km × 1 km area, and regression mapping was performed using the $PM_{2.5}$ and PM_{10} regression models. During regression mapping, the independent variables associated with a cell were extracted at the centroid of the cell and then substituted in the $PM_{2.5}$ and PM_{10} regression models to predict the concentrations of $PM_{2.5}$ and PM_{10} for the

Table 4
Results of multiple stepwise regression for PM_{10}.

	Estimate	Std error	*t* value	Pr (> |*t*|)	VIF	Result
Intercept	1.606e+02	2.282e+00	70.364	< 2e-16***		Multiple *R*-squared:
MO-5000	1.832e-04	5.352e-05	3.423	0.00077 ***	1.04	0.688
GS-5000	−1.520e-06	8.540e-08	−17.803	< 2e-16***	1.21	Adjusted *R*-squared: 0.681
BS-5000	−3.309e-02	8.159e-03	−4.056	7.49e-05***	1.39	RMSE:14.842 μg/m^3
PE-5000	3.635e+00	1.106e+00	3.286	0.00123**	1.26	

Note: ***, ** and * indicate significant levels of significance at 0, 0.001, and 0.01 respectively.

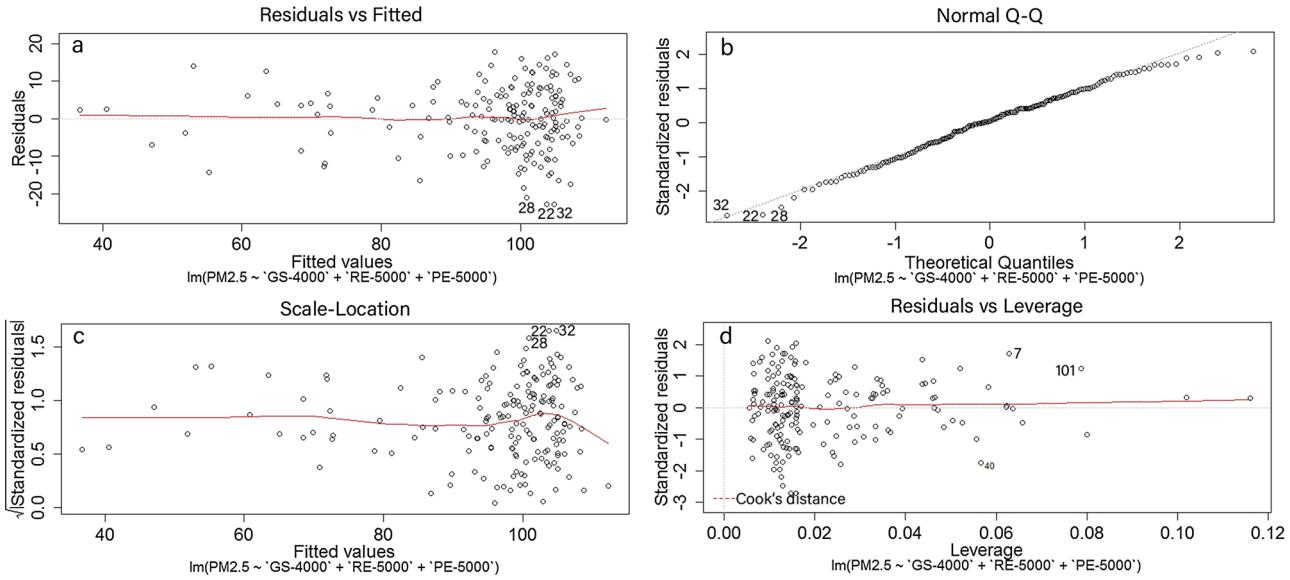

Fig. 12. Regression diagnostic plots of $PM_{2.5}$ LUR model.

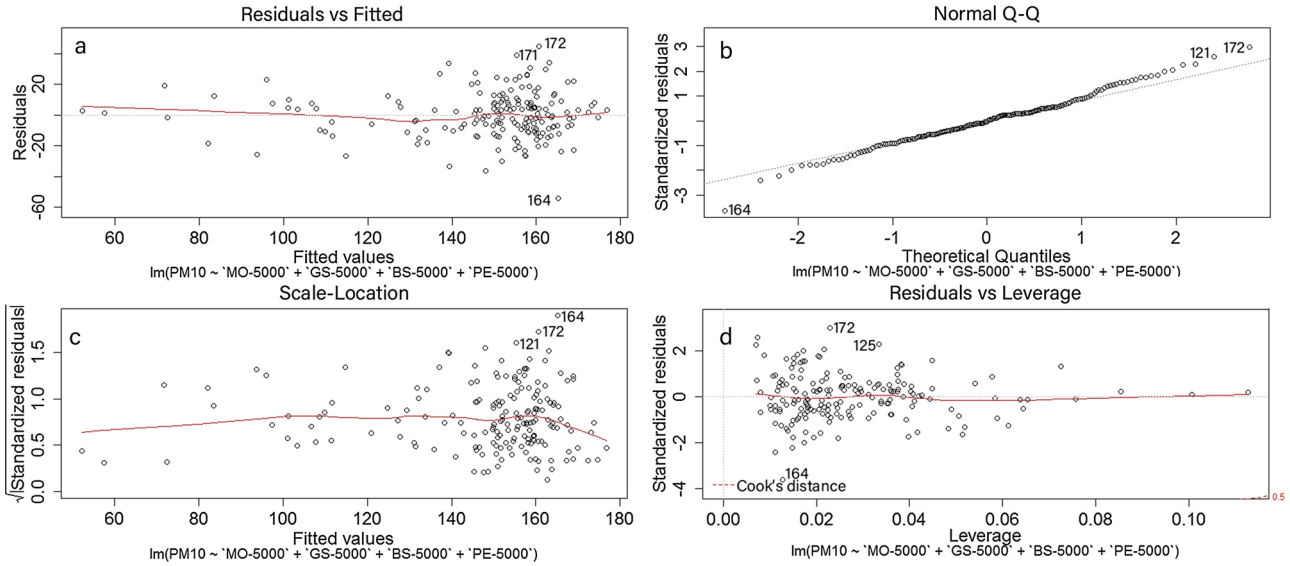

Fig. 13. Regression diagnostic plots of PM_{10} LUR model.

Table 5
Cross-validation of models and spatial autocorrelation test of residuals.

$PM_{2.5}$				PM_{10}			
	LUR	Cross-validation	Spatial Autocorrelation		LUR	Cross-validation	Spatial autocorrelation
Mul R^2	0.718	0.719	Moran I: -0.129	Mul R^2	0.688	0.691	Moran I: 0.589
Adj R^2	0.713	0.713	*z* score: -0.216	Adj R^2	0.681	0.682	*z* Score: 1.043
RMSE	8.355	8.660	*p*-value: 0.828	RMSE	14.842	15.609	*p*-value: 0.297

Table 6
Results of multiple stepwise regression for $PM_{2.5}$ in the new heating season.

	Estimate	Std.Error	t value	Pr(> | t |)	Vif	Result
Intercept	7.566e+01	9.906e-01	76.381	< 2e-16 ***		Muliple R-squared:
GS-5000	−7.405e-07	4.302e-08	−17.212	< 2e-16 ***	1.18	0.642
RE-5000	−6.666e-04	1.450e-04	−4.596	8.15e-06 ***	1.29	Adjusted R-squared:
PE-5000	1.083e+00	5.433e-01	1.994	0.0477 *	1.17	0.636
						Rmse:7.596 μg/m^3

Note: ***, ** and * indicate significant levels of significance at 0, 0.001, and 0.01 respectively.

Table 7
Results of multiple stepwise regression for PM_{10} in the new heating season.

	Estimate	Std.Error	t value	Pr(> | t |)	Vif	Result
Intercept	1.195e+02	2.129e+00	56.368	< 2e-16 ***		Muliple R-squared:
MO-5000	1.504e-04	5.483e-05	2.743	0.0067 **	1.04	0.647
GS-5000	−1.220e-06	5.384e-08	−22.659	< 2e-16 ***	1.21	Adjusted R-squared:
BS-5000	−2.935e-02	5.769e-03	−5.088	9.22e-07 ***	1.39	0.639
PE-5000	1.810e+00	7.489e-01	2.4167	0.0167 *	1.26	Rmse:12.704 μg/m^3

Note: ***, ** and * indicate significant levels of significance at 0, 0.001, and 0.01 respectively.

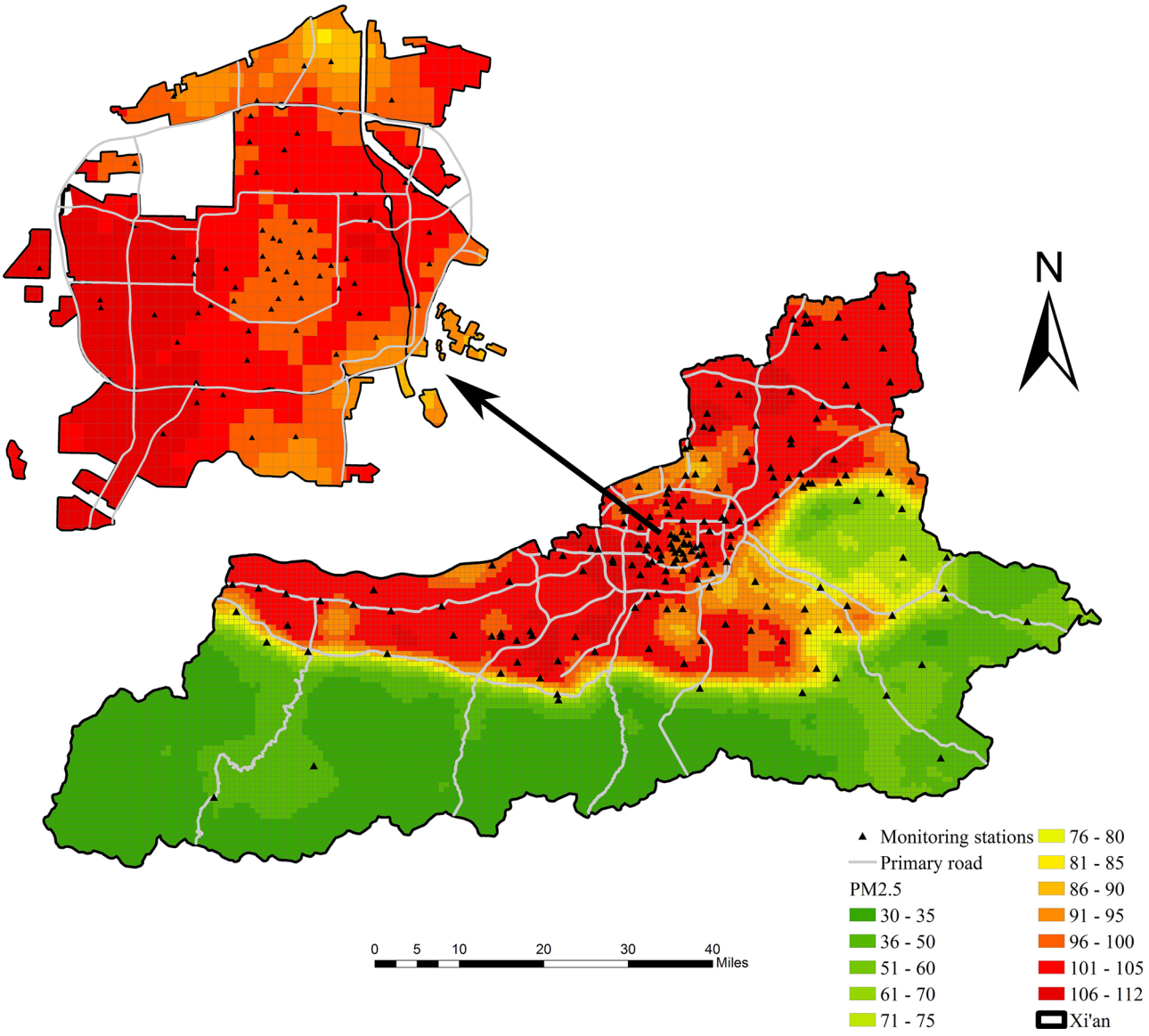

Fig. 14. Spatial distribution map of $PM_{2.5}$ in urban areas of Xi'an and regulatory management zone.

cell. The spatial distribution maps (Figs. 14 and 15) of the concentrations of $PM_{2.5}$ and PM_{10} in Xi'an were generated using the visualization function of ArcGIS, which presented significant differences between the spatial distributions of the concentrations of $PM_{2.5}$ and PM_{10} in Xi'an.

4. Discussion

4.1. Comparison with other studies

The correlation coefficients of the green-space area with respect to the concentrations of $PM_{2.5}$ and PM_{10} were the largest in each buffer zone, and the coefficients increased with the increase in the area of the buffer zone, which indicated that with increase in green space, the negative correlation between the pollutant concentrations and the green spaces was more significant. The independent variable shared by the $PM_{2.5}$ and PM_{10} land use regression prediction models of Xi'an is the number of polluting enterprises with a buffer area of 5000 m. Moreover, the results indicated that while green-space area was the most critical independent variable in both the models, it exerted its largest impact at various buffer-zone distances. In addition to the green-space area, the $PM_{2.5}$ LUR model also included the number of restaurants in the 5000-m-distance buffer zone as an independent variable,

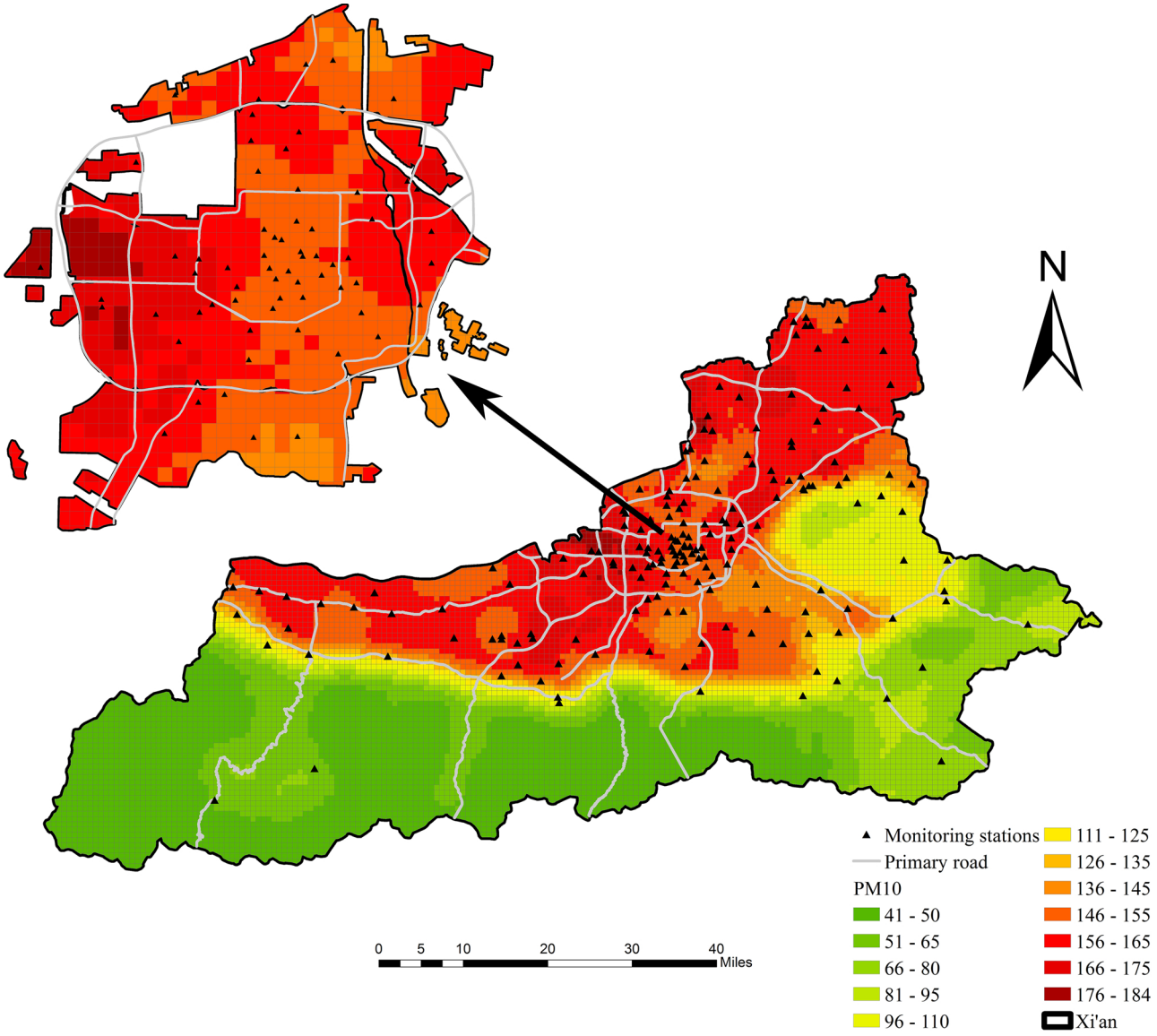

Fig. 15. Spatial distribution map of PM_{10} in urban areas of Xi'an and regulatory management zone.

whereas the PM_{10} LUR model included the number of bus stops and the length of motorways in the 5000-m-distance buffer zone as two independent variables. The reason green-space area was the most critical independent variable in both models was that Xi'an has a unique terrain and urban morphology. That is to say, the forest coverage of Xi'an is as high as 48 %, with a large area of green spaces and vegetation in the Qinling and Li Mountains. There is a strong negative correlation between the concentration of pollutants and the green-space area of the buffer zone and a significant positive correlation between the green-space area of the buffer zone and its elevation with a correlation coefficient of 0.86. The green-space area in a buffer zone around a station with a high elevation tends to be big. It indirectly reflects the strong negative correlation between the elevation of the monitoring station and the concentration of pollutants. Comparison with the studies of other cities in Table 8 indicates that there are similarities among Beijing, Czech-Poland and Xi'an in that a great number of mountains and green-space areas exist in the territories of these study areas. These researches also contain the independent variable of green-space area, which shows that the pollutant distributions are all impacted by the mountains and green space (Bitta, Pavlíková, Svozilík, & Jančík, 2018; Ji et al., 2019; Wu et al., 2015). Those studies, except for that in Beijing by Wu et al. (2015) did not include industry-related independent variables. Other studies included the independent variable of pollution emission or industrial land area. In this study, two industry-related independent variables were selected, namely the area of industrial land and the distance from air-polluting enterprise, and the distance from air-polluting enterprises was finally included in the model. Moreover, studies in Beijing and the other Chinese cities of Shanghai and Tianjin also included road length as an independent variable (Chen, Bai et al., 2010; Liu, Henderson, Wang, Yang, & Peng, 2016; Meng et al., 2016; Wu et al., 2015), consistent with the present study. The Adj-R^2 values of the $PM_{2.5}$ and PM_{10} LUR models of Xi'an were 0.713 and 0.681, respectively, lower than the Adj-R^2 values of 0.877 and 0.81, respectively, obtained in LUR models of Shanghai (Liu et al., 2016) and Beijing (Ji et al., 2019) but higher than 0.43–0.65 and 0.54, respectively, for Beijing in another study (Wu et al., 2015) and Hong Kong (Lee et al., 2017). Therefore, the Adj-R^2 value obtained in this study was considered to represent a moderate level compared with other reports. This study used data observed at 181 monitoring stations—the largest number and density of monitoring stations among studies of this type. However, the higher density of monitoring stations in this study did not lead to improvement in fitting accuracy compared with other studies, suggesting that fitting accuracy may not be simply dependent on the number and density of monitoring stations.

Table 8
Studies on LUR models for other cities.

Author	City	Research object	Variables entered into the model	Independent variables	Result	Monitoring period	Buffer zone	Number of stations	Area of study zone
This study 2019	Xi'an	$PM_{2.5}$	Green space_4000 m Restaurant_5000m Distance to the nearest air-polluting enterprise	Land use, road, traffic facilities, emission source, geospatial information, socioeconomic information	Adjusted R^2: 0.713 RMSE: 8.355	2018.11.15–2019.03.15	100 m,300 m, 500 m,1000 m, 2000 m,3000 m, 4000 m,5000m	181	10,752 km^2
		PM_{10}	Green space_5000m Motorway_5000m Bus stop_5000m Distance to the nearest air-polluting enterprise		Adjusted R^2: 0.681 RMSE:14.842				
Liu et al. (2016)	Shanghai	$PM_{2.5}$	Long Distance to the coast High way intensity_300 m Waterbody area_500 m Industrial land area_300	Land use, population density, road networks, distance from monitors to the ocean, major air pollution sources, longitude, latitude	Adjusted R^2: 0.877 RMSE:194.59	One year	100 m,300 m, 500 m,1000 m, 3000 m	35	5512 km^2
Meng et al. (2016)	Shanghai	PM_{10}	Distance to the coast Emission_7000m Green space_1000 m Major road_5000m	Land use, aerosol optical depth, meteorological data	Adjusted R^2: 0.80 RMSE: 4.2	2008	100 m,300 m, 500 m,1000 m, 2000 m,3000 m, 5000m	28	6300 km^2
Wu et al. (2015)	Beijing	$PM_{2.5}$	Major road length Water area_500 m Natural vegetation area_3000 m Crop area_3000 m	Street network, land cover, population density, catering services distribution, bus stop density, intersection density	Adjusted R^2: 0.43–0.65 RMSE: 6.5 − 19.1	2013.03–2014.03	100 m,200 m, 300 m,500 m, 750 m,1000 m, 2000 m,3000 m, 5000m	35	16,411 km^2
Ji et al. (2019)	Beijing	$PM_{2.5}$	Industrial-mining-warehouse land_200 m Average temperature Forest-grassland _500 m Wetland_3000 m	Land use, road, terrain, Population, meteorological factors	Adjusted R^2: 0.81 RMSE: 15.7	Heating season	100 m,200 m, 300 m,1000 m, 2000 m,3000 m, 4000 m,5000m	35	16,411 km^2
Lee et al. (2017)	Hong Kong	$PM_{2.5}$	Expressway_25m Distance to Shenzhen Car park density_1000 m Car park density_25m Government_100m Industrial_25m	Annual average traffic density, road length, traffic loading, urban build-up, land use, point value variables, point feature, value extracted at point, distance	Adjusted R^2: 0.54 RMSE: 4.0	2014.04.24-2014.05.15	50 m,100 m, 200 m,300 m, 500 m,1000 m, 1500 m,2000 m, 3000 m,4000 m, 5000m	80	None
Chen, Bai et al. (2010)	Tian Jin	PM_{10}	Major roads_2000 m Residence_2000 m Population density Distance to sea Wind speed	Major road, land use, population, meteorological variables, distance to sea	Adjusted R^2: 0.72	2006 Heating season	500 m,1000 m, 1500 m,2000 m	30	11,920 km^2
Bitta et al. (2018)	Czech–Polish border	PM_{10}	Emissions from industrial sources_2000 m Emissions from domestic heating_2000 m Forested land_1000 m	Pollution source, land use	Adjusted R^2: 0.65 RMSE: 8.34	None	100 m,200 m, 500 m,1000 m, 2000 m	27	8832 km^2
	Sabzevar	$PM_{2.5}$	Industrial_500 m Distance to religion/	Traffic surrogates, land use, urban morphology, distance variables,	Adjusted R^2: 0.68	2017.04.20-2018.03.06	100 m,200 m, 300 m,400 m, 500 m,750 m, 1000 m,1250 m, 1500	26	None

(*continued on next page*)

Table 8 (*continued*)

Author	City	Research object	Variables entered into the model	Independent variables	Result	Monitoring period	Buffer zone	Number of stations	Area of study zone
Miri, Ghassoun, Dovlatabadi, Ebrahimnejad, and Löwner (2019))			cultural land use	population density,	RMSE:		m,1750 m, 2000 m,2250 m,		
			Maximum height of building_200 m	geographic location of monitoring stations	3.33		2500 m,2750 m, 3000 m		
		PM_{10}	Other land use area_100m		Adjusted R^2:				
			Distance to religion		0.70				
			/cultural land use		RMSE:				
			Distance to bus terminal		8.99				
			Distance to urban facility land use						

4.2. Spatial distribution characteristics of pollutants in Xi'an

The concentrations of $PM_{2.5}$ and PM_{10} had a strong correlation, with the Pearson correlation coefficient as high as 0.864. Accordingly, the concentrations of $PM_{2.5}$ and PM_{10} in Xi'an presented basically the same trend in spatial distribution except for in some local regions and each pollutant demonstrated an overall decreasing trend of concentration from the north to the south. The concentrations of $PM_{2.5}$ and PM_{10} in the Qinling Mountains south of Xi'an and Li Mountains east of Xi'an were significantly lower than those in other areas, which was attributed to the fact that the reduction of pollutant emission sources and increase in vegetation have a mitigating effect on the concentrations of $PM_{2.5}$ and PM_{10}. You et al. (2016) performed the MODIS aerosol inversion to derive the spatial distribution of the concentration of PM_{10} in Xi'an during 2011–2013, finding a high concentration of PM_{10} accumulated in the west and northeast of the main urban district of Xi'an, consistent with the results of this study. However, You et al. (2016) did not observe the accumulation of high-concentration PM_{10} in the High-tech Industry Development Zone (HIDZ) southwest of the main urban district of Xi'an but observed it in the area east of the main urban district, which is significantly different from the observations in this study. This inconsistency may be attributed to the various data acquisition times and data types used by You et al. (2016), i.e., MODIS AOD data, compared with those used this study.

Outside the Qinling and Li Mountains, the concentrations of $PM_{2.5}$ and PM_{10} were closely related to the layout of industrial land use and the locations of air-polluting enterprises. Fig. 16 presents the layout of industrial land use and the locations of air-polluting enterprises in Xi'an. In particular, a large amount of industrial land existed in the HIDZ southwest of the main urban district and the Huyi District (HYD) and Caotang Science & Technology Zone (CSTZ) southwest of the regulatory management zone (RMZ), and a large number of industrial enterprises were present in the Fengdong New City (FDNC) west of the RMZ, leading to significantly higher concentrations of $PM_{2.5}$ and PM_{10} in these areas compared with other areas. The Lintong District (LTD), Gaoling District (GLD), and Yanliang District (YLD) in the northeast of the RMZ were home to the Lintong District of Modern Industry (LTDMI), Gaoling District of Equipment Industry (GLDEI), and Yanliang District of Aviation Industry (YLDAI), respectively, with significantly higher concentrations of $PM_{2.5}$ and PM_{10} compared with other areas. Huang et al. (2015) confirmed that the accumulation of $PM_{2.5}$ in Xi'an was closely related to industrial production. Wang et al. (2014) reported that as high as 58 % of the concentrations of $PM_{2.5}$ in Xi'an in the heavy pollution months were caused by industrial activities. Song et al. (2015) revealed that transportation and industrial emissions were the main sources of $PM_{2.5}$ in Xi'an. However, the correlation coefficients between the areas of industrial and mining land use and the concentrations of $PM_{2.5}$ and PM_{10} were quite small, and the correlation was not statistically significant, which may be attributed to the possibility that the information collected on industrial and mining land use was not truly reflective of the actual situation.

4.3. Spatial distribution characteristics of pollutants in RMZ of Xi'an

The spatial distribution characteristics of $PM_{2.5}$ and PM_{10} were investigated in the RMZ, which is larger than the main urban district. Fig. 17 presents the layout of industrial land use, water systems, green spaces, and roads, all of which were factors related to the distribution of the concentrations of $PM_{2.5}$ and PM_{10}. The highest concentrations of $PM_{2.5}$ and PM_{10} were observed in the area named WREWTR, which lies between the West Urban Ring Motorway and the West Third Ring Road, and in the HIDZ area. Both of these areas and their adjacent spaces contain a large industrial land and a large number of air-polluting enterprises, which also exist in the area north of the RMZ (NRMZ). However, there was no accumulation of high concentrations of $PM_{2.5}$ and PM_{10} in NRMZ, which may be attributed to the presence of large

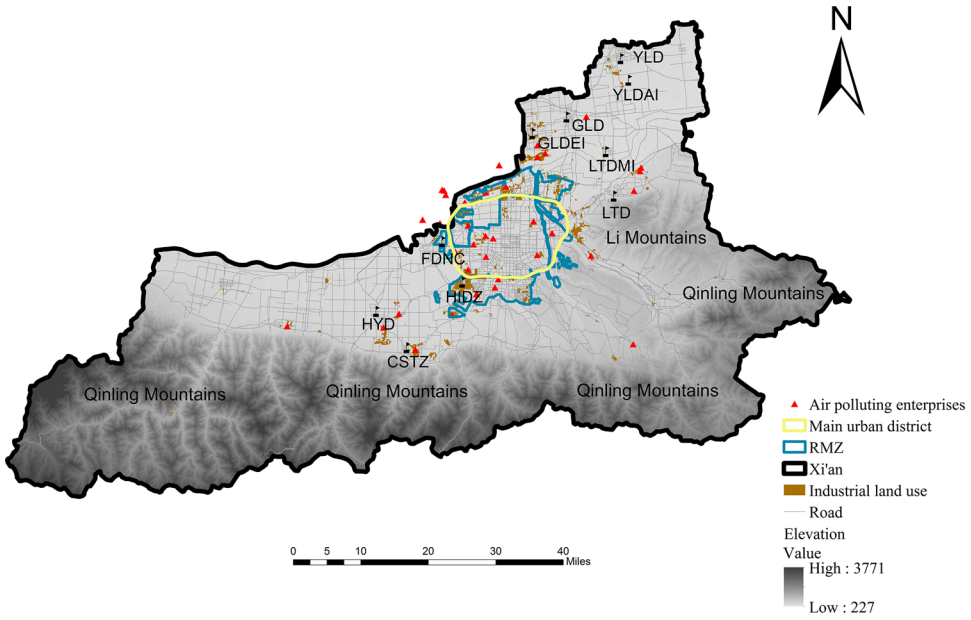

Fig. 16. Layout of industrial land use and positions of air-polluting enterprises in Xi'an.

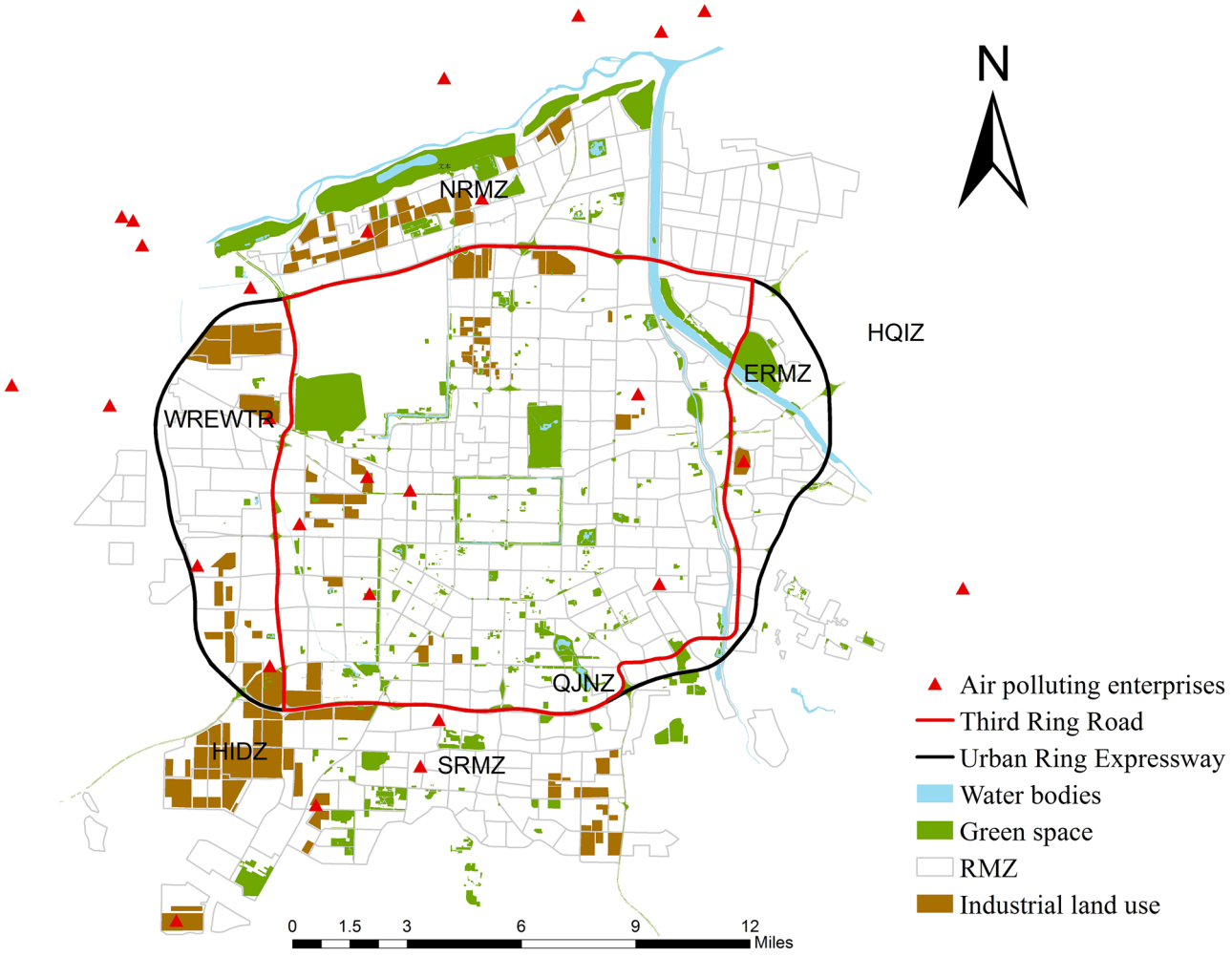

Fig. 17. Layout of industrial land use, water systems, green spaces, and ring roads in RMZ of Xi'an.

water bodies and green spaces around the area due to its proximity to the Wei River. On one hand, the large water bodies can generate water–land breeze through a mechanism similar to sea–land breeze (Bouchlaghem, Mansour, & Elouragini, 2007; Zhu & Zhou, 2019), which, coupled with the open terrain and low building density of NRMZ, facilitates the diffusion of the pollutants. On the other hand, vegetation, green spaces, and wetlands can improve the removal efficiency of $PM_{2.5}$ and PM_{10} (Amini Parsa, Salehi, Yavari, & van Bodegom, 2019; Feng, Zou, & Tang, 2017; Selmi et al., 2016; Zhu & Zeng, 2018). As depicted in Fig. 12, another high-concentration area of $PM_{2.5}$ and PM_{10} was ERMZ, which lies on the east of RMZ, in close proximity to the Hongqing Industrial Zone (HQIZ), which is home to a large number of industrial enterprises. The concentrations of $PM_{2.5}$ and PM_{10} in most areas within the Second Ring Road were generally lower than those in the surrounding areas. The southern part of the RMZ (SRMZ) and the Qujiang New Zone (QJNZ) in the southeastern part also had significantly lower concentrations of $PM_{2.5}$ and PM_{10} than those in surrounding areas.

Vehicular emissions serve as a substantial source of $PM_{2.5}$, leading to rapidly increasing concentration of $PM_{2.5}$. A study by Dai et al. (2018) demonstrated that motor vehicles provided a continuously growing contribution to the concentrations of $PM_{2.5}$ during 2006–2014 in Xi'an. However, except for the WREWTR and HIDZ most areas in the RMZ did not demonstrate the accumulation of high concentrations of $PM_{2.5}$ and PM_{10}, indicating that industrial emissions were still the main contributor to the concentrations of $PM_{2.5}$ and PM_{10} in Xi'an. Wang et al. (2019) reported that motor vehicles accounted for 11.13 % of the concentrations of $PM_{2.5}$ in Xi'an in the winter of 2017, significantly lower than the contribution of 51.02 % through the combustion of coal.

4.4. Limitations

Owing to the limited time available for data collection, this study focused on the heating season of Xi'an for investigating the spatial distribution of $PM_{2.5}$ and PM_{10}, thus being unable to address the distribution in other seasons and, thereby, making it impossible to understand inter-seasonal differences in the spatial distribution of $PM_{2.5}$ and PM_{10}. The land-use data employed in this study were mainly retrieved from remote sensing images as the main data source. However, these images were generated in an earlier time period than the pollution data, and, therefore, a certain degree of inter-period differences were noticed in land use, possibly introducing errors into the results of correlation analysis between the independent and dependent variables. Owing to the lack of corresponding meteorological monitoring data at the monitoring stations, this study did not consider the impact of meteorological factors such as wind direction, wind speed, air temperature, and humidity on the concentrations of $PM_{2.5}$ and PM_{10}. Further, the accuracy of model fitting can be improved using pollutant concentration data observed under lower wind speeds.

5. Conclusions

This study is the first to apply land use regression models to the Fenwei Plain, a heavily polluted area in China, and analyzes 87 factors from five categories of information including land use information, road traffic facility information, socioeconomic information, geospatial information, and emission source information in Xi'an. The correlation between the factors and the $PM_{2.5}$ and PM_{10} concentrations of 181 monitoring stations were investigated. Based on the correlation analysis results of independent variables and dependent variables, the land use regression prediction models of $PM_{2.5}$ and PM_{10} in Xi'an were established. Cross-validation confirms that the model exhibits good performance in spatial prediction and generalization, and the verification with various time periods proves the applicability of LUR models in different heating seasons. The model successfully predicted the $PM_{2.5}$ and PM_{10} concentrations in 10,525 grids, revealing the variations in the spatial distribution of $PM_{2.5}$ and PM_{10} concentrations in Xi'an.

Within the city limit of Xi'an, the $PM_{2.5}$ and PM_{10} concentrations generally show a trend with high values in the north and low values in the south. The concentrations of $PM_{2.5}$ and PM_{10} in the Qinling Mountains in the south and Li Mountains in the east are significantly lower than those in other regions. The spatial distribution of $PM_{2.5}$ and PM_{10} is closely related to layout of industrial land and the location of polluting enterprises. The spatial distribution of $PM_{2.5}$ and PM_{10} in the Regulatory Management Zone (RMZ) further confirms that large area of green space can effectively reduce the concentrations of $PM_{2.5}$ and PM_{10}. The RMZ has large traffic flow however only a small part of the zone seems to have high concentration of pollutants, indicating that industrial emissions are the main source of $PM_{2.5}$ and PM_{10} in Xi'an, followed by traffic emissions. This study enriches the application of land use regression models in the prediction of the spatial distribution of atmospheric pollutants and provides a scientific foundation for urban planning, land use regulation, air pollution control, and public health policy making. It presents a basic model for population exposure assessment and promotes the application of land use regression models in areas requiring heating in winter and heavily polluted areas. It plays a positive role in achieving sustainability of urban environment and promoting sustainable development in urban areas.

Declaration of Competing Interest

All authors declare no conflict of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (51678058); the Natural Science Foundation of Shaanxi Province (2019JM-475).

References

Ahmed, S., Adnan, M., Janssens, D., & Wets, G. (2020). A route to school informational intervention for air pollution exposure reduction. *Sustainable Cities and Society, 53*. https://doi.org/10.1016/j.scs.2019.101965.

Allen, R. W., Amram, O., Wheeler, A. J., & Brauer, M. (2011). The transferability of NO and NO_2 land use regression models between cities and pollutants. *Atmospheric Environment, 45*, 369–378. https://doi.org/10.1016/j.atmosenv.2010.10.002.

Amini Parsa, V., Salehi, E., Yavari, A. R., & van Bodegom, P. M. (2019). Analyzing temporal changes in urban forest structure and the effect on air quality improvement. *Sustainable Cities and Society, 48*. https://doi.org/10.1016/j.scs.2019.101548.

Barzeghar, V., Sarbakhsh, P., Hassanvand, M. S., Faridi, S., & Gholampour, A. (2020). Long-term trend of ambient air PM_{10}, $PM_{2.5}$, and O_3 and their health effects in Tabriz city, Iran, during 2006–2017. *Sustainable Cities and Society, 54*. https://doi.org/10.1016/j.scs.2019.101988.

Berman, J. D., Burkhardt, J., Bayham, J., Carter, E., & Wilson, A. (2019). Acute air pollution exposure and the risk of violent behavior in the United States. *Epidemiology, 30*, 799–806. https://doi.org/10.1097/ede.0000000000001085.

Bitta, J., Pavlíková, I., Svozilík, V., & Jančík, P. (2018). Air pollution dispersion modelling using spatial analyses. *ISPRS International Journal of Geo-Information, 7*. https://doi.org/10.3390/ijgi7120489.

Bouchlaghem, K., Mansour, F. B., & Elouragini, S. (2007). Impact of a sea breeze event on air pollution at the Eastern Tunisian Coast. *Atmospheric Research, 86*, 162–172. https://doi.org/10.1016/j.atmosres.2007.03.010.

Briggs, D. J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., et al. (1997). Mapping urban air pollution using GIS: A regression-based approach. *International Journal of Geographical Information Science, 11*, 699–718. https://doi.org/10.1080/136588197242158.

Briggs, D. J., de Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S., et al. (2000). *A regression-based method for mapping traffic-related air pollution: Application and testing in four contrasting urban environments.* 253, - 167.

Chen, L., Bai, Z., Kong, S., Han, B., You, Y., Ding, X., et al. (2010). A land use regression for predicting NO_2 and PM_{10} concentrations in different seasons in Tianjin region, China. *Journal of the Environmental Sciences, 22*, 1364–1373. https://doi.org/10.1016/s1001-0742(09)60263-1.

Chen, L., Du, S. Y., Bai, Z. P., Kong, S. F., You, Y., Han, B., et al. (2010). Application of land use regression for estimating concentrations of major outdoor air pollutants in Jinan, China. *Journal of Zhejiang University-Science A, 11*, 857–867. https://doi.org/10.1631/jzus.A1000092.

Dai, Q., Bi, X., Liu, B., Li, L., Ding, J., Song, W., et al. (2018). Chemical nature of $PM_{2.5}$ and PM_{10} in Xi'an, China: Insights into primary emissions and secondary particle formation. *Environmental Pollution, 240*, 155–166. https://doi.org/10.1016/j.envpol.

2018.04.111.
Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., et al. (2012). Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PMcoarse in 20 European study areas; results of the ESCAPE project. *Environmental Science & Technology, 46*, 11195–11205. https://doi.org/10.1021/es301948k.
Feng, H., Zou, B., & Tang, Y. (2017). Scale- and region-dependence in landscape-$PM_{2.5}$ correlation: Implications for urban planning. *Remote Sensing, 9*. https://doi.org/10.3390/rs9090918.
Feng, S., Gao, D., Liao, F., Zhou, F., & Wang, X. (2016). The health effects of ambient $PM_{2.5}$ and potential mechanisms. *Ecotoxicology and Environmental Safety, 128*, 67–74. https://doi.org/10.1016/j.ecoenv.2016.01.030.
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*, 2422–2428. https://doi.org/10.1021/es0606780.
Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., et al. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment, 42*, 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057.
Huang, P., Zhang, J., Tang, Y., & Liu, L. (2015). Spatial and temporal distribution of $PM_{2.5}$ pollution in Xi'an City, China. *International Journal of Environmental Research and Public Health, 12*, 6608–6625. https://doi.org/10.3390/ijerph120606608.
Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., et al. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology, 15*, 185–204. https://doi.org/10.1038/sj.jea.7500388.
Ji, W., Wang, Y., & Zhuang, D. (2019). Spatial distribution differences in $PM_{2.5}$ concentration between heating and non-heating seasons in Beijing, China. *Environmental Pollution, 248*, 574–583. https://doi.org/10.1016/j.envpol.2019.01.002.
Jin, J. Q., Du, Y., Xu, L. J., Chen, Z. Y., Chen, J. J., Wu, Y., et al. (2019). Using Bayesian spatio-temporal model to determine the socio-economic and meteorological factors influencing ambient $PM_{2.5}$ levels in 109 Chinese cities. *Environmental Pollution, 254*, 113023. https://doi.org/10.1016/j.envpol.2019.113023.
Lee, M., Brauer, M., Wong, P., Tang, R., Tsui, T. H., Choi, C., et al. (2017). Land use regression modelling of air pollution in high density high rise cities: A case study in Hong Kong. *The Science of the Total Environment, 592*, 306–315. https://doi.org/10.1016/j.scitotenv.2017.03.094.
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 ($PM_{2.5}$) and nitrogen dioxide (NO_2) concentrations in City of Shanghai, China. *The Science of the Total Environment, 565*, 607–615. https://doi.org/10.1016/j.scitotenv.2016.03.189.
Liu, W., Li, X., Chen, Z., Zeng, G., León, T., Liang, J., et al. (2015). Land use regression models coupled with meteorology to model spatial and temporal variability of NO_2 and PM_{10} in Changsha, China. *Atmospheric Environment, 116*, 272–280. https://doi.org/10.1016/j.atmosenv.2015.06.056.
Liu, X., Sun, T., & Feng, Q. (2020). Dynamic spatial spillover effect of urbanization on environmental pollution in China considering the inertia characteristics of environmental pollution. *Sustainable Cities and Society, 53*. https://doi.org/10.1016/j.scs.2019.101903.
Meng, X., Chen, L., Cai, J., Zou, B., Wu, C. F., Fu, Q., et al. (2015). A land use regression model for estimating the NO_2 concentration in Shanghai, China. *Environmental Research, 137*, 308–315. https://doi.org/10.1016/j.envres.2015.01.003.
Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., et al. (2016). Estimating ground-level PM_{10} in a Chinese city by combining satellite data, meteorological information and a land use regression model. *Environmental Pollution, 208*, 177–184. https://doi.org/10.1016/j.envpol.2015.09.042.
Miri, M., Ghassoun, Y., Dovlatabadi, A., Ebrahimnejad, A., & Löwner, M.-O. (2019). Estimate annual and seasonal PM_1, $PM_{2.5}$ and PM_{10} concentrations using land use regression model. *Ecotoxicology and Environmental Safety, 174*, 137–145. https://doi.org/10.1016/j.ecoenv.2019.02.070.
Moore, D. K., Jerrett, M., Mack, W. J., & Kunzli, N. (2007). A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA. *Journal of Environmental Monitoring: JEM, 9*, 246–252. https://doi.org/10.1039/b615795e.
Ortolani, C., & Vitale, M. (2016). The importance of local scale for assessing, monitoring and predicting of air quality in urban areas. *Sustainable Cities and Society, 26*, 150–160. https://doi.org/10.1016/j.scs.2016.06.001.
Pilla, F., & Broderick, B. (2015). A GIS model for personal exposure to PM_{10} for Dublin commuters. *Sustainable Cities and Society, 15*, 1–10. https://doi.org/10.1016/j.scs.2014.10.005.
Qiu, Z., Xu, X., Song, J., Luo, Y., Zhao, R., Zhou, B. X. W., et al. (2017). Pedestrian exposure to traffic PM on different types of urban roads: A case study of Xi'an, China. *Sustainable Cities and Society, 32*, 475–485. https://doi.org/10.1016/j.scs.2017.04.007.
Selmi, W., Weber, C., Rivière, E., Blond, N., Mehdi, L., & Nowak, D. (2016). Air pollution removal by trees in public green spaces in Strasbourg city, France. *Urban Forestry & Urban Greening, 17*, 192–201. https://doi.org/10.1016/j.ufug.2016.04.010.
Son, Y., Osornio-Vargas, A. R., O'Neill, M. S., Hystad, P., Texcalac-Sangrador, J. L., Ohman-Strickland, P., et al. (2018). Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters. *The Science of the Total Environment, 639*, 40–48. https://doi.org/10.1016/j.scitotenv.2018.05.144.
Song, W., Jia, H., Li, Z., Tang, D., & Wang, C. (2019). Detecting urban land-use configuration effects on NO_2 and NO variations using geographically weighted land use regression. *Atmospheric Environment, 197*, 166–176. https://doi.org/10.1016/j.atmosenv.2018.10.031.
Song, Y. Z., Yang, H. L., Peng, J. H., Song, Y. R., Sun, Q., & Li, Y. (2015). Estimating $PM_{2.5}$ concentrations in Xi'an City using a generalized additive model with multi-source monitoring data. *PLoS One, 10*, e0142149. https://doi.org/10.1371/journal.pone.0142149.
Wang, D., Hu, J., Xu, Y., Lv, D., Xie, X., Kleeman, M., et al. (2014). Source contributions to primary and secondary inorganic particulate matter during a severe wintertime $PM_{2.5}$ pollution episode in Xi'an, China. *Atmospheric Environment, 97*, 182–194. https://doi.org/10.1016/j.atmosenv.2014.08.020.
Wang, Y., Li, S., Wang, M., Sun, H., Mu, Z., Zhang, L., et al. (2019). Source apportionment of environmentally persistent free radicals (EPFRs) in $PM_{2.5}$ over Xi'an, China. *The Science of the Total Environment, 689*, 193–202. https://doi.org/10.1016/j.scitotenv.2019.06.424.
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 and Pollution Research International, 22*, 7045–7061. https://doi.org/10.1007/s11356-014-3893-5.
Xiao, K., Wang, Y., Wu, G., Fu, B., & Zhu, Y. (2018). Spatiotemporal characteristics of air pollutants (PM_{10}, $PM_{2.5}$, SO_2, NO_2, O_3, and CO) in the Inland Basin City of Chengdu, Southwest China. *Atmosphere, 9*. https://doi.org/10.3390/atmos9020074.
Yang, J., Shi, B., Shi, Y., Marvin, S., Zheng, Y., & Xia, G. (2020). Air pollution dispersal in high density urban areas: Research on the triadic relation of wind, air pollution, and urban form. *Sustainable Cities and Society, 54*. https://doi.org/10.1016/j.scs.2019.101941.
You, W., Zang, Z., Zhang, L., Zhang, M., Pan, X., & Li, Y. (2016). A nonlinear model for estimating ground-level PM_{10} concentration in Xi'an using MODIS aerosol optical depth retrieval. *Atmospheric Research, 168*, 169–179. https://doi.org/10.1016/j.atmosres.2015.09.008.
Yuan, M., Song, Y., Huang, Y., Shen, H., & Li, T. (2019). Exploring the association between the built environment and remotely sensed $PM_{2.5}$ concentrations in urban areas. *Journal of Cleaner Production, 220*, 1014–1023. https://doi.org/10.1016/j.jclepro.2019.02.236.
Zhu, C., & Zeng, Y. (2018). Effects of urban lake wetlands on the spatial and temporal distribution of air PM_{10} and $PM_{2.5}$ in the spring in Wuhan. *Urban Forestry & Urban Greening, 31*, 142–156. https://doi.org/10.1016/j.ufug.2018.02.008.
Zhu, D., & Zhou, X. (2019). Effect of urban water bodies on distribution characteristics of particulate matters and NO_2. *Sustainable Cities and Society, 50*. https://doi.org/10.1016/j.scs.2019.101679.
Zou, B., Wilson, J. G., Zhan, F. B., & Zeng, Y. (2009). Spatially differentiated and source-specific population exposure to ambient urban air pollution. *Atmospheric Environment, 43*, 3981–3988. https://doi.org/10.1016/j.atmosenv.2009.05.022.