

Original Research

Land Use Regression Models Using Satellite Aerosol Optical Depth Observations and 3D Building Data from the Central Cities of Liaoning Province, China

Jiping Gong^{1,2}, Yuanman Hu^{1*}, Miao Liu¹, Rencang Bu¹, Yu Chang¹,
Muhammad Bilal³, Chunlin Li¹, Wen Wu^{1,2}, Baihui Ren^{1,2}

¹State Key Laboratory of Forest and Soil Ecology, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110164, People's Republic of China

²University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

³Department of Land Surveying and Geo-Informatics, the Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Received: 12 October 2015

Accepted: 4 January 2016

Abstract

Land use regression (LUR) modeling is a promising method for assessing the spatial variation of air pollutant concentrations. We developed an LUR model for air pollutants (SO₂, NO₂, and PM₁₀) in the central cities of Liaoning Province using monitoring data collected during 2013. We evaluated whether the addition of annual satellite aerosol optical depth (AOD) observations and five canyon indicators (building height, building coverage ratio, floor area ratio, building shape coefficient, and high-rise building ratio) improved the LUR models. Out-of-sample “10-fold” cross validation was used to quantify the accuracy of the model predictions. Our results showed that the gross domestic product (GDP) and the distance to the nearest industrial emissions were the common variables for the models. Annual AOD demonstrated weak correlations with air pollutant concentrations because of its instantaneity, low resolution, and limited precision; however, it was useful for improving the coefficient of determination (R²) of the LUR models. The full models incorporating the annual AOD data and canyon indicators showed further improvement. The improvements of R² were 0.22, 0.19, and 0.39 for SO₂, NO₂, and PM₁₀, respectively, demonstrating that the consideration of canyon indicators could still be valuable and could be used in LUR models.

Keywords: air pollution, canyon indicators, aerosol optical depth (AOD), land use regression (LUR)

*e-mail: huym@iae.ac.cn,
landscapehym@163.com

Introduction

Land use regression (LUR) modeling is an important approach that has been applied successfully in many studies to predict long-term average pollutant concentrations at the intra-urban scale. The LUR model combines the monitoring of air pollution within a certain spatial scale and the development of stochastic models using GIS-collected predictor variables, while interpolation methods (e.g., kriging and inverse distance weighting) are used to provide predictions of the spatial distribution of pollutants [1]. The most common predictor variables in previous studies that are of interest to various traffic descriptions (e.g., intensity, congestion, road length, and distance) are population density, housing, land use, physical geography (e.g., elevation and distance to the sea), and meteorological data [2-4]. Further developments of LUR modeling have included greater focus on transferable models and predictor variables such as wind speed, wind direction, and emission data; further development of models were considered [5, 6].

Because of their extensive spatial coverage and ease of access, satellite aerosol optical depth (AOD) observations have been widely used to estimate the high-resolution distribution of ground-level PM, NO₂, or SO₂. Most previous research has focused on establishing relationships between monitoring values of hourly, daily, or monthly AOD and air pollutant concentrations to estimate surface air pollution concentrations using satellite-derived AOD data in urban areas [7-9]. However, the relationship between AOD and air pollutant concentrations is generally complex because of the dependence on meteorological conditions, aerosol vertical distribution, aerosol types, microphysical and chemical properties, and land use [10, 11]. Satellite-derived AOD is a measure of the light attenuation within the atmospheric column, which is affected by physiographic condition (e.g., humidity, vertical profile, and chemical composition), while air pollutant concentration is a measure of the dry particle mass near the surface; therefore, it should not be expected that these two parameters would be strongly correlated [12]. Furthermore, although satellites can extend the spatial coverage of observations, their spatial resolution is inadequate [13].

Street canyons in urban centers always have poor air quality. In the environments that surround buildings, air pollution is affected by a variety of factors, including local winds, traffic patterns, the built environment, and land use [14, 15]. In fact, the built environment characteristic reflects the effect of the street canyon on the dispersion of air pollution [16]. The existence of buildings in the street canyon might impede the flow of air, making air pollutants difficult to spread, and enhancing the turbulence characteristics within urban areas [17]. Several LUR studies have proved that the consideration of canyon indicators could improve predictions of air pollutant concentrations [18, 19]. In earlier 2D models of street canyons, the street aspect ratio (street height/street width, H/W) and the "skyview factor" are generally the most

common parameters incorporated; however, they cannot describe building geometries comprehensively [18, 20]. Here we describe an approach using interpretation of remote-sensing image to estimate building length, width, and height for all locations within a region of interest.

In this study, we calculated the indicators of the "urban canyon effect" using 3D building data and incorporated satellite-derived AOD observations to evaluate whether these canyon indicators improved the applicability of the LUR models. The development of method allows the incorporation of canyon indicators in LUR models of air pollution, and it can be used as a scientific basis for proposals of measures to control urban air pollution.

Material and Methods

Study Area

Liaoning Province is located in northeastern China (Fig. 1). Its northern and central parts are plains through which the Liaohe River flows. With the Bohai Sea and the Yellow Sea in the south, the eastern and western parts of the province are dominated by low hills and the southern part is dominated by Liaodong Peninsula. Under the control of a continental monsoonal climate within a warm temperate zone, the area has four distinct seasons and it receives considerable sunshine. The annual average temperature is between 6 and 11°C, decreasing inland from the coastal areas. The lowest temperatures, which occur in January (the coldest month), can reach -30°C, while the high temperatures can reach 35°C in July (the hottest month). The average annual rainfall is 500-1000 mm, decreasing from the east to the west. Precipitation occurs mainly in summer, accounting for 60-75% of the total [21].

The central cities in Liaoning Province comprise seven major industrial conurbations (Shenyang, Anshan, Fushun, Benxi, Yingkou, Liaoyang, and Tieling). The regional area and the population of the central cities account for 44% and 51% of the entire province, respectively, and their GDP and local fiscal revenue account for 63% and 54% of the entire province, respectively. Thus, this region is the largest and most economically developed core area in Liaoning Province and the entire northeast region of China. As the traditional base of heavy industry in China, the major activities within the province include metallurgical, mechanical, petrochemical, and construction industries. Unprecedented rapid urbanization is occurring in the central cities of Liaoning Province, causing their geographic expansion, an increase in the height of internal buildings, and increasingly serious environmental atmospheric pollution.

Air Pollution Data

In Liaoning Province, critical air pollutants such as PM_{2.5}, PM₁₀, SO₂, NO₂, NO_x, CO, and O₃ are monitored by air-quality monitoring networks. In this study, 24-hour

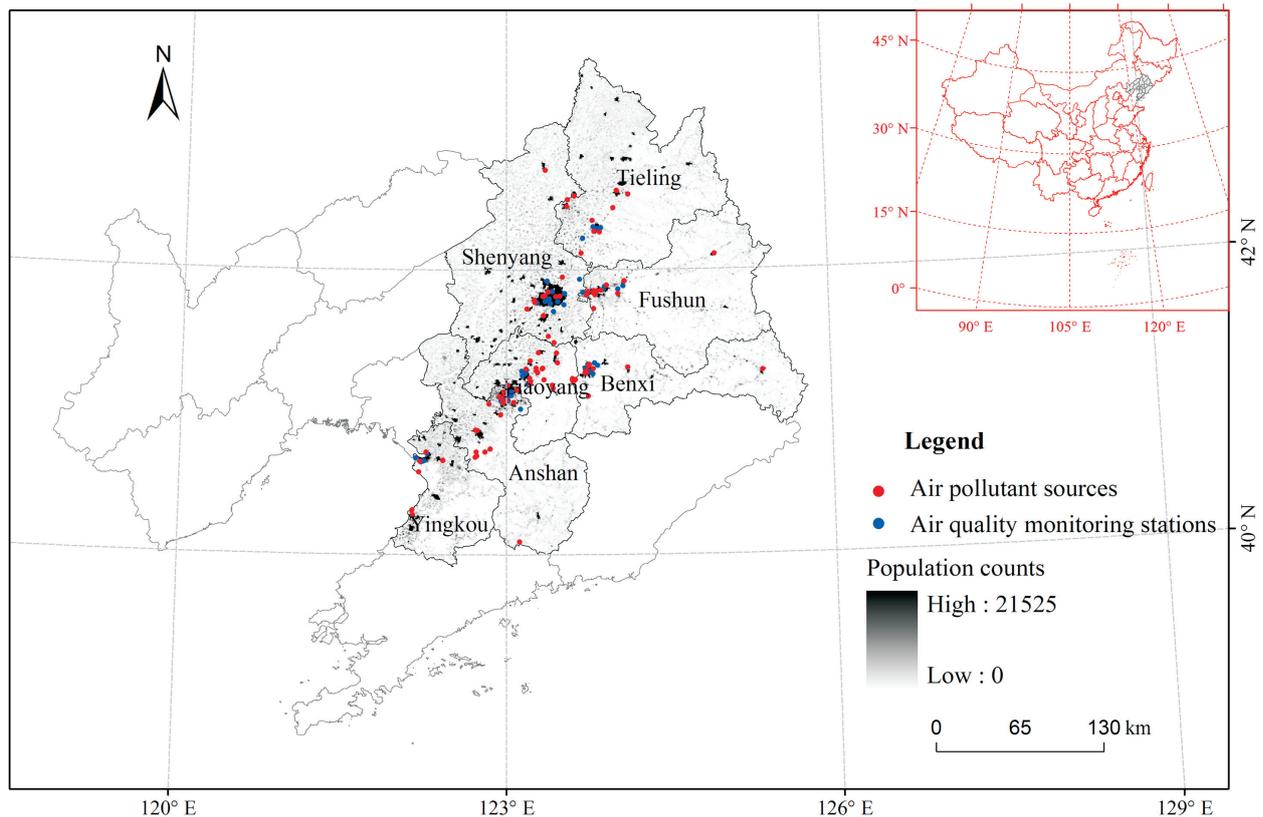


Fig. 1. Location of the central cities in Liaoning Province, distributions of air pollutant sources and air quality monitoring stations, and population counts.

concentrations of air pollutants (SO_2 , NO_2 , and PM_{10}) were calculated at 79 (2015) sites for monitoring air pollution, operated by the Ministry of Environmental Protection of the People's Republic of China. The air quality monitoring networks included regional background sites, urban background sites, and roadside sites. In this study, we analyzed annual SO_2 , NO_2 , and PM_{10} data obtained during 2013 from 34 urban background sites and roadside sites located in street canyons of seven major industrial conurbations (Fig. 1, Table 1).

Satellite Data

MODIS collection 5.1 Terra and Aqua AOD products obtained from NASA's Goddard Earth Sciences Distributed Active Archive Center were used in the analysis [22]. Terra passes over the equator at about 10:30 local time (LT) and Aqua overflies the equator at about 13:30 LT in the opposite direction, providing almost daily global AOD coverage [9]. The dimensionless AOD is retrieved based on sunlight attenuation by the aerosols within the vertical atmospheric column from the ground to the atmosphere's top. MODIS AOD dataset are retrieved using spectral radiance in the visible-near-infrared (0.47–2.1 μm) wavelengths at 3-km resolution at nadir [22]. In this study, the 2013 annual AOD dataset was calculated using the mean 0.55- μm AOD from the MOD04 aerosol

daily products. These AOD products were extracted from the MOD04 scientific "Corrected_Optical_Depth_Land" dataset, averaged over an area of 3×3 km. The application of MODIS AOD with linear approximations in the SO_2 , NO_2 , and PM_{10} estimations is discussed in sections 3.2 and 4.1.

Land Use

Canyon Indicator Variables

This paper proposes a novel methodology for the detection of the 3D information of buildings from high-resolution (0.61 m) QuickBird imagery [23, 24]. The framework of the proposed approach has two main steps. (1) Based on an object-oriented method, the building's shape is established using the technology of regional identity and feature measurement. (2) The building's height is estimated based on its shadow. First, the length of the building's shadow is measured and then the building's height information is calculated based on the relative geometric positions of the sun, satellite, building, and shadow. Fig. 2 presents illustrative schematics of two cases.

As shown in Fig. 2(a), when the sun and satellite are on the same side of the building, the building's height can be estimated using Eqs. (1)–(3):

Table 1. The location of the site.

Site Code	City	Longitude(°E)	Latitude(°N)	Site Code	City	Longitude(°E)	Latitude(°N)
21010095	Shenyang	123.36	41.78	210400053	Fushun	124.04	41.86
21010096	Shenyang	123.40	41.80	210400055	Fushun	123.81	41.85
21010097	Shenyang	123.48	41.78	210500405	Benxi	123.75	41.34
21010098	Shenyang	123.41	41.77	210500410	Benxi	123.73	41.30
21010099	Shenyang	123.43	41.85	210500420	Benxi	123.77	41.29
210100100	Shenyang	123.54	41.83	210500425	Benxi	123.80	41.27
210100101	Shenyang	123.28	41.77	210500435	Benxi	123.81	41.35
210100102	Shenyang	123.44	41.71	210600401	Yingkou	122.24	40.66
210100103	Shenyang	123.38	41.92	210600403	Yingkou	122.27	40.71
210100104	Shenyang	123.54	41.76	210700405	Yingkou	122.22	40.65
210300051	Anshan	123.04	41.12	210800409	Yingkou	122.17	40.68
210300052	Anshan	123.01	41.09	211000406	Liaoyang	123.15	41.26
210300053	Anshan	123.05	41.14	211000407	Liaoyang	123.18	41.27
210300054	Anshan	122.96	41.10	211000408	Liaoyang	123.14	41.29
210300080	Anshan	122.95	41.08	211200401	Tieling	123.81	42.30
210400051	Fushun	123.90	41.86	211200405	Tieling	123.88	42.30
210400052	Fushun	123.92	41.88	211200411	Tieling	123.85	42.29

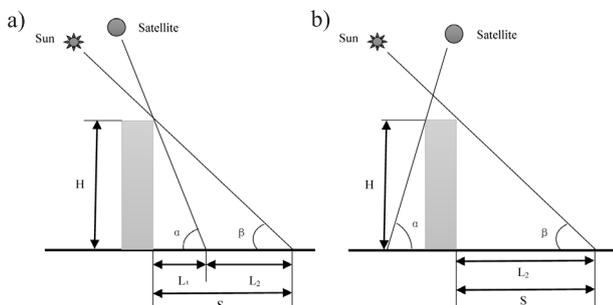
$$S = H/\tan\beta \tag{1}$$

$$L_2 = S - L_1 = H/\tan\beta - H/\tan\alpha \tag{2}$$

$$H = L_2 \times \tan\alpha \times \tan\beta / (\tan\alpha - \tan\beta) \tag{3}$$

As shown in Fig. 2(b), when the sun and satellite are located on opposite sides of the building (i.e., $S = L_2$ and $L_1 = 0$), the building's height can be estimated using Eqs. (4):

$$H = L_2 \times \tan\beta \tag{4}$$



H is a building's height
 S is the actual length of building's shadow
 L₁ is the sheltered length of building's shadow
 L₂ is the visible length of building's shadow
 α is satellite elevation angle
 β is solar azimuth angle

Fig. 2. Schematics of the method for estimating building heights.

Detailed data on the buildings of the seven central cities were available for 2010 by visual interpretation. Only the built-up area was considered and 153,735 building blocks were available for use with a relative error of <0.3 m and accuracy of 97% (Fig. 3). Five new canyon indicators

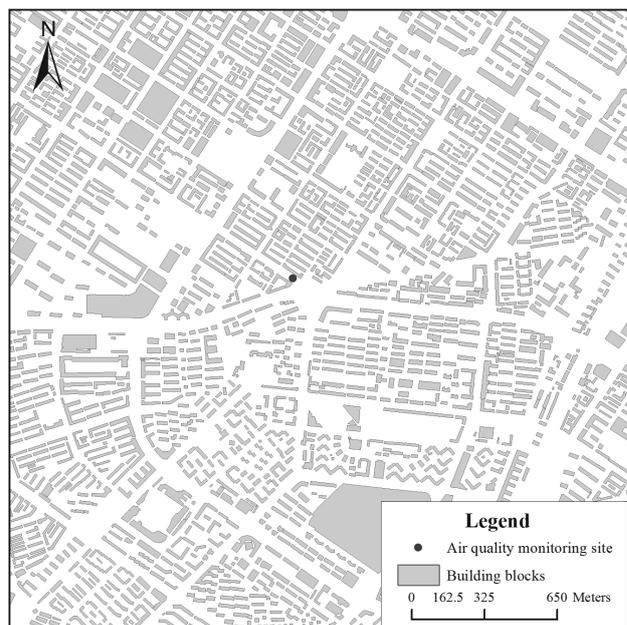


Fig. 3. Partial view of building blocks.

were established based on the aspects of height, density, shape, and volume of the buildings. We tested whether these canyon indicators improved pollutants' modeled level using LUR modelling.

Elevation

Digital elevation data were obtained through the SRTM 90-m Digital Elevation Database (<http://srtm.csi.cgiar.org/>). This database is produced by NASA and covers all the countries of the world.

Normalized Difference Vegetative Index (NDVI)

A satellite-derived NDVI, based on red and near-infrared (NIR) reflectances, is a temporal indicator of vegetation cover and its phenological state [25, 26]. Apart from spring's green-up periods and fall's senescence, the NDVI can generally be considered relatively constant on a monthly basis. Many other studies have found that regional NDVI is an effective index for the local means of air pollutant concentrations [27, 28]. In this study, 1-km MODIS monthly vegetation indexes products (MOD13A3) from 2013 that were obtained from the LAADS website (<http://ladsweb.nascom.nasa.gov>).

$$NDVI = (\rho_{NIR} - \rho_{RED}) / (\rho_{NIR} + \rho_{RED}) \quad (5)$$

Distance to Air Pollutant Point Source Emissions

In Liaoning Province, 166 principal point source emissions of air pollutants were obtained through the 2013 National Monitored Enterprise List (<http://www.ipe.org.cn/pollution/sources.aspx>), which includes large sources that emit >1500, >3000, and >800 tons of SO₂, NO_x, and dust, respectively, per year. The distances (m) from the air pollution monitoring sites to the nearest point source emissions were calculated using ArcGIS software.

Road Density Data

Road data were obtained from the Thematic Database for Human-Earth System (<http://www.data.ac.cn/>). We calculated the lengths of the total, primary, secondary, and connecting roads, and the road density of each buffer was calculated using ArcGIS software.

Population Density and GDP Density Data

Both the population density and the GDP density data were obtained from the Thematic Database for Human-Earth System (<http://www.data.ac.cn/>). These data were available at 1000-m resolution based on the 2010 Census and county government statistics provided by the National Bureau of Statistics of the People's Republic of China.

Table 2. Shapiro-Wilk normality test results (n = 34).

	SO ₂	NO ₂	PM ₁₀
Statistic value	0.960	0.990	0.947
Sig.	0.352	0.986	0.137

Modeling Approach

In this study, 10 buffers were generated from 200 to 2000 m around 34 monitoring sites (Table 3). This method was intended to capture both the more distant and proximal sources of variability in the air pollutant concentrations. Some predictor variables were calculated within each buffer (e.g., road lengths and canyon indicators) using either the sums or averages of the variables in each buffer. Other predictor variables were calculated at each monitoring site (e.g., GDP, population, percentage tree cover, elevation, distance to the nearest industrial emissions, and AOD value). Finally, we generated 86 potentially predictive variables in eight classes and 14 subclasses. The eight classes were population, economy, traffic variables, distance variables, vegetation cover, geographic location, satellite data, and canyon indicators.

It is important to choose the predictor variables for determining the LUR model. The detailed implementation processes of forward selection procedure by Su et al. [29] was followed in this study. An independent variable can be added to the LUR model when the matching conditions are satisfied: 1) its significance level is below 0.05; and 2) the variance inflation factors (VIFs) of all the predictor variables within the LUR model fall below 10 to avoid collinearity. We repeat 10-fold cross-validation (CV) three times to select all variables that met the above two conditions, keeping the variables that with the smallest cross-validated root mean square error [30].

We first built basic LU-only models based on 34 sites without the annual AOD data and canyon indicators. We then added annual AOD data and canyon indicators separately to these basic LU-only models to evaluate the additional predictive of the five canyon indicators' value. The models examined were: 1) LU-only model (basic form); 2) LU+AOD model (development form with additional AOD predictor variable); and 3) LU+AOD+3D model (development form with additional AOD and canyon indicators predictor variables).

Model Validation

The VIF is the reciprocal of tolerance, and both are primary measures of the degree of the i-th independent variable's multi-collinearity with the other independent variables in a LUR model. The larger the VIF value, the more serious the multi-collinearity. The empirical method shows that when $0 < VIF < 10$, there is no multi-collinearity [31]. If the VIF of any predictor was >10 , that model was considered unacceptable, the predictor was removed from

Table 3. Descriptive statistics of the input variables in the LUR models.

GIS dataset	Predictor variable	Variable subclass	Unit	Buffer size (radius in meters)	Prior direction
Population	Population in a buffer	POP	persons/m ²	-	+
Economy	Gross domestic product in a buffer	GDP	yuan/m ²	-	+
Traffic variables	Density of major roads in a buffer	RDA	m/m ²	200-2000	+
	Density of secondary and connecting roads in a buffer	RDB	m/m ²	200-2000	+
	Density of all roads in a buffer	RDC	m/m ²	200-2000	+
Distance Variables	Distance to the nearest Industrial emissions	DISTIE	m	-	-
Vegetation Cover	Mean Normalized Difference Vegetation Index	NDVI	-	-	-
Geographic Location	Elevation	ELEV	m	-	-
Satellite data	MODIS aerosol optical depth (AOD) products	AOD	-	-	+
Canyon indicators	Mean building height	BH	-	200-2000	+
	Building coverage ratio	BCR	-	200-2000	+
	Floor area ratio	FAR	-	200-2000	+
	Building shape coefficient	BSC	-	200-2000	+
	High-rise building ratio	HBR	-	200-2000	+
$BH = \sum_{j=1}^n H_{ij}$ $BCR = \sum F/A \times 100\%$ $FAR = \sum \left(\frac{H}{C} \times F\right) / A$ $BSC = (P \times H + F) / (F \times H)$ $HBR = N_h / N$	H is a building's height , F is a building's standing area, C is a constant which represents the average height of one storey of all the buildings (here C = 2.8 m), A is the total area of a site, P is a building's standing circumference N _h is number of buildings which height >24 m N is the total number of buildings 200 m buffer intervals ranging from 200 to 2000				

the eligible pool, and the process was restarted from the beginning of the step.

$$VIF = 1/(1 - R_i^2) \tag{6}$$

...where $1 - R_i^2$ is the tolerance and denotes the proportion of variance in the i-th predictor, which is correlated with the other predictors in the regression equation [32].

To analyze the sensitivity of the developed models we used a 10-fold CV approach to evaluate the capability of the model to predict concentrations of air pollutants for each of the monitoring sites within the study area. In the 10-fold CV, the dataset was repeatedly divided randomly into 90% (calibration) and 10% (held-out test) splits. We applied the fitted calibration model to estimate the concentrations of air pollutants for the held-out test set. This "out-of-sample" process was repeated 10 times to calculate the CV coefficient of determination (R^2) values and the root mean square error (RMSE) values.

Results

Air Pollution Data

The annual mean concentrations (standard deviation) of SO₂, NO₂, and PM₁₀ across the stations were 56.86 (29.73), 38.13 (9.78), and 97.83 (30.2) μg/m³, respectively (Fig. 4). The correlations between SO₂ and NO₂, SO₂ and PM₁₀, and NO₂ and PM₁₀ were 0.510 (P < 0.01), 0.658 (P < 0.01), and 0.587 (P < 0.01) in all cases, respectively (Table 4). Two outliers of SO₂ were found in the business district and industrial area of Shenyang, which is the most polluted city in Liaoning Province. Here we reserved the two outliers. Shapiro-Wilk test indicated that dependent variable (SO₂, NO₂, and PM₁₀) was normal distribution, significance level > 0.05 (Table 2); therefore, the data meet the requirements of the multiple linear regression model.

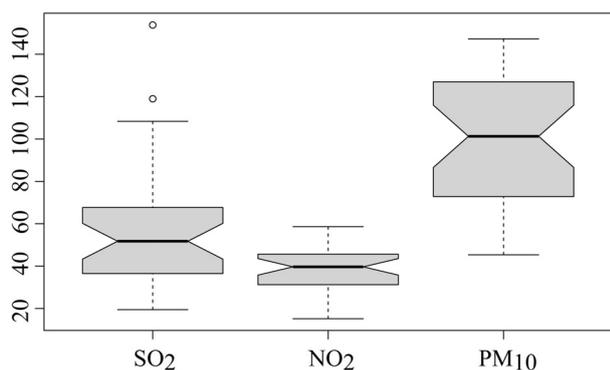


Fig. 4. Comparison of SO₂, NO₂, and PM₁₀ annual concentrations at sampling sites in the central cities of Liaoning Province in 2013.

Spatial Characteristics of Annual Mean AOD

Fig. 5 shows the spatial pattern of AOD of the central cities of Liaoning Province. Air pollutant concentrations show partial spatial consistency. The difference between the urban background and the roadside sites is very apparent. As expected, AOD values are higher in urban areas than the mountainous and hilly areas, especially in Shenyang, Liaoyang, Anshan, and Yingkou. Fig. 6 shows scatterplots and the relationships between the annual concentrations of SO₂, NO₂, and PM₁₀, and annual AOD values. The concentrations of the three air pollutants all appear to exhibit similar trends to AOD for the annual values; however, they are not predicted well by the values of the total AOD column. Although the R² values are very low, SO₂ has more significant correlation with AOD than have NO₂ and PM₁₀. The correlation is significant at level 0.01 on a statistical basis (Table 4), which implies that the AOD is valuable and that it could be used for improving the spatial R² of the LUR models.

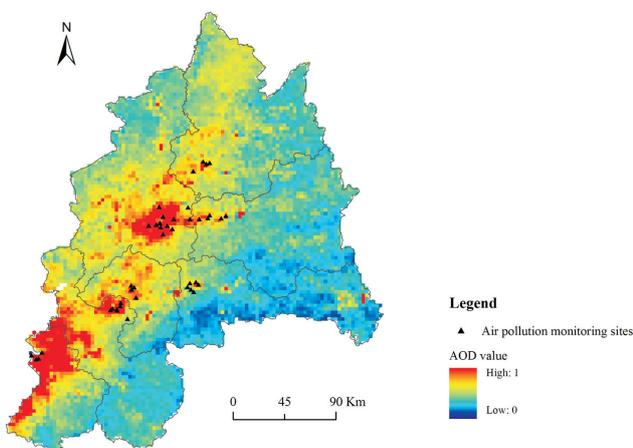


Fig. 5. AOD spatial distribution (3 × 3 km) in 2013 over the study area.

Table 4. Correlation between annual concentrations of SO₂, NO₂, and PM₁₀ and AOD values, Spearman’s rho (n = 34).

		SO ₂	NO ₂	PM ₁₀	AOD
SO ₂	Correlation Coefficient	1.000	0.510**	0.658**	0.560**
	Sig. (2-tailed)	-	0.002	0.000	0.001
NO ₂	Correlation Coefficient	-	1.000	0.587**	0.399*
	Sig. (2-tailed)	-	-	0.000	0.019
PM ₁₀	Correlation Coefficient	-	-	1.000	0.558**
	Sig. (2-tailed)	-	-	-	0.001
AOD	Correlation Coefficient	-	-	-	1.000
	Sig. (2-tailed)	-	-	-	-

**Correlation is significant at the 0.01 level (two-tailed).

Characteristics of Canyon Indicators

Statistical descriptions of the five canyon indicators were calculated for all 34 monitoring sites and they showed clear differences between the minima and maxima (Table 5), which implies that canyon type varied considerably. Deeper canyons had significantly higher mean building height (BH), floor area ratio (FAR), building shape coefficient (BSC), building coverage ratio (BCR), and high-rise building ratio (HBR) than low-to-moderate canyons. The correlations between the five different indicators had their own distinction. The correlation between BCR and FAR had high correlation coefficients. In contrast, BH and BSC had low correlation coefficients (Table 6).

Predictions of Concentrations of Air Pollutants Based on a Mixed-Effects Model

Table 7 presents the three different models and the performance of each model in terms of R², adjusted R², RMSE, and those values obtained from the 10-fold CV for SO₂, NO₂, and PM₁₀: the LU only model, LU+AOD model, and LU+AOD+3D model. Of the 36 LU variables generated, 3-4 were significantly predictive in each LU-only model. Among them, GDP and distance to the nearest industrial emissions (DISTIE) were the common variables for the models. As for the adjusted R² values for the entire year, LU-only models were 0.52 for SO₂, 0.42 for NO₂, and 0.34 for PM₁₀. As can be seen, adding the AOD term to the LU-only model partly improved the R² value, and the full model, incorporating both AOD and 3D, improved it further. The improvements of R² were 0.19, 0.22, and 0.39 for SO₂, NO₂, and PM₁₀, respectively.

Table 5. Distribution of derived canyon indicators (n = 34).

	BH200	BH400	BH600	BH800	BH1000	BH1200	BH1400	BH1600	BH1800	BH2000
min	3.00	4.17	4.32	3.89	4.01	4.09	3.97	4.25	4.19	4.31
mean	13.99	14.18	13.62	13.40	13.02	13.13	12.78	12.74	12.70	12.59
median	14.61	14.81	13.34	13.23	12.74	13.70	13.58	12.68	11.83	11.50
max	26.93	25.49	26.62	25.96	26.05	25.64	25.39	24.98	23.87	23.91
	FAR200	FAR400	FAR600	FAR8000	FAR1000	FAR1200	FAR1400	FAR1600	FAR1800	FAR2000
min	0.00	0.02	0.00	0.01	0.00	0.02	0.01	0.03	0.00	0.00
mean	0.95	0.99	0.87	0.83	0.77	0.80	0.75	0.72	0.66	0.62
median	0.51	0.90	0.80	0.63	0.49	0.68	0.65	0.58	0.51	0.45
max	3.76	3.04	2.40	2.46	2.58	2.59	2.52	2.39	2.29	2.25
	BSC200	BSC400	BSC600	BSC800	BSC1000	BSC1200	BSC1400	BSC1600	BSC1800	BSC2000
min	1.47	10.24	0.62	3.05	9.51	22.78	31.18	36.15	4.68	12.54
mean	15.92	74.79	158.92	218.11	303.37	455.12	587.80	661.10	881.29	1026.36
median	13.71	53.15	106.18	185.06	258.95	392.42	586.51	573.97	837.97	974.72
max	59.33	446.81	936.23	709.32	1026.49	1324.56	1596.40	1980.30	2977.93	3430.34
	BCR200	BCR400	BCR600	BCR800	BCR1000	BCR1200	BCR1400	BCR1600	BCR1800	BCR2000
min	0.36	1.29	0.13	0.48	0.15	0.18	0.41	0.92	0.02	0.06
mean	16.99	17.40	15.71	15.06	14.37	15.19	14.22	13.68	12.74	12.06
median	19.87	19.65	18.60	17.68	15.39	16.76	16.66	14.70	13.41	12.93
max	35.74	31.23	31.68	29.36	29.37	28.83	28.04	27.66	27.92	28.07
	HBR200	HBR400	HBR600	HBR800	HBR1000	HBR1200	HBR1400	HBR1600	HBR1800	HBR2000
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
mean	14.51	13.53	12.64	13.65	13.62	13.96	13.32	13.26	13.45	9.82
median	4.35	6.89	8.24	9.26	9.47	9.49	9.23	10.50	11.98	6.26
max	64.29	46.84	46.32	45.81	46.48	44.19	41.74	39.61	35.99	36.07

Model Stability

A final 10-fold CV was run for each of the final models. The results of the CV R^2 and RMSE (Table 7) indicate

that the derived LUR models are stable – especially the SO_2 models. For SO_2 , the 10-fold CV R^2 (RMSE) of the three models was 0.49 (21.1 $\mu\text{g}/\text{m}^3$), 0.51 (21.70 $\mu\text{g}/\text{m}^3$), and 0.65 (20.2 $\mu\text{g}/\text{m}^3$), respectively. For NO_2 , the 10-fold

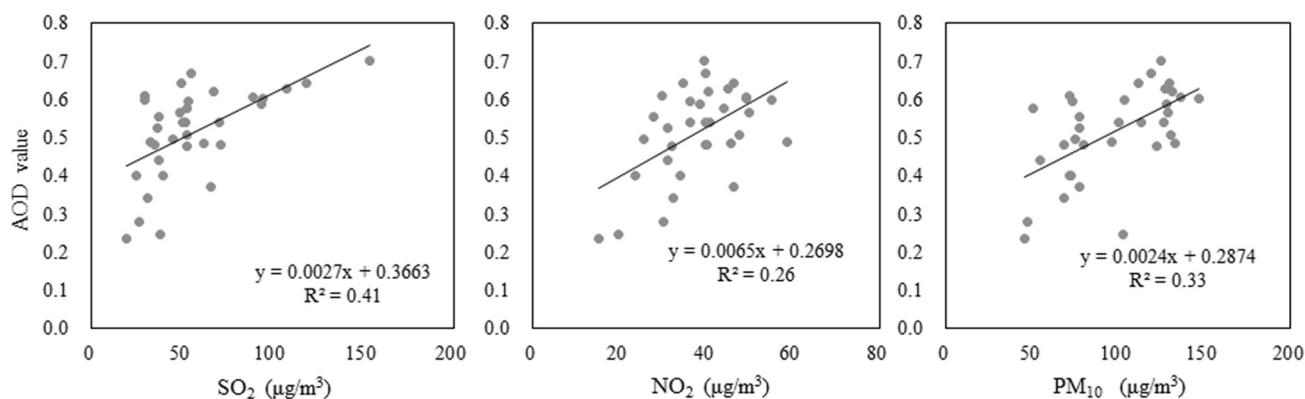


Fig. 6. Scatterplots and relationships between annual concentrations of SO_2 , NO_2 , and PM_{10} and MODIS AOD values in 2013.

Table 6. Correlation between the different canyon indicators, Spearman's rho (n = 308).

		BH	BCR	BSC	FAR	HBR
BH	Correlation Coefficient	1.000	0.475**	0.026	0.711**	0.747**
	Sig. (two-tailed)	-	0.000	0.649	0.000	0.000
BCR	Correlation Coefficient	-	1.000	0.268**	0.912**	0.371**
	Sig. (two-tailed)	-	-	0.000	0.000	0.000
BSC	Correlation Coefficient	-	-	1.000	0.281**	0.148**
	Sig. (two-tailed)	-	-	-	0.000	0.009
FAR	Correlation Coefficient	-	-	-	1.000	0.559**
	Sig. (two-tailed)	-	-	-	-	0.000
HBR	Correlation Coefficient	-	-	-	-	1.000
	Sig. (two-tailed)	-	-	-	-	-

**Correlation is significant at the 0.01 level (two-tailed).

CV R² (RMSE) of the three models was 0.38 (8.42 $\mu\text{g}/\text{m}^3$), 0.43 (8.57 $\mu\text{g}/\text{m}^3$), and 0.53 (9.90 $\mu\text{g}/\text{m}^3$), respectively. For PM₁₀, the 10-fold CV R² (RMSE) of the three models was 0.33 (26.3 $\mu\text{g}/\text{m}^3$), 0.33 (26.7 $\mu\text{g}/\text{m}^3$), and 0.55 (28.7 $\mu\text{g}/\text{m}^3$), respectively. According to significance level, the number of samples and observations and statistic values were compared to lower and upper critical values based on the Durbin-Watson test (Table 7), which indicated there was no autocorrelation in the residuals and proved the models were reliable.

Discussion

We built three LUR models for SO₂, NO₂, and PM₁₀ for the central cities of Liaoning Province using data from 34 urban monitoring sites. The intention was to evaluate whether the inclusion of annual AOD data and urban street canyon indicators could improve the spatially resolved LUR models. We have shown that annual AOD data has a weak correlation with air pollutant concentrations because of its instantaneity, low resolution, and limited precision; however, it is still valuable and it can be used to improve the R² of the LUR models. The use of canyon indicators provided better predictive power than the incorporation of AOD data.

Relationship of AOD and Air Pollution Concentrations

There is positive correlation between AOD and air pollutant concentrations, although this relationship is weak. The main reason for this, apart from random and systematic errors associated with satellite AOD retrievals, is that the AOD values are instantaneous (i.e., obtained at about 10:30 LT); satellite data have limited resolution, while the concentrations of air pollutants are average daily data. Second, the AOD is easily affected by meteorological and synoptic conditions [9, 33]. Because these factors

vary spatiotemporally, it is likely that the AOD-SO₂/NO₂/PM₁₀ relationships also vary spatiotemporally for different regions and seasons. In this study, we ignored the possible effects of these influencing factors on the AOD-SO₂/NO₂/PM₁₀ relationships.

Land Use Regression Models

LUR has been demonstrated as an appropriate technique for the modeling of air pollution in which the regressors often show a strong explanatory capability to the phenomena being modeled. Our basic LU-only models explained 40-57% of the spatial contrast without the inclusion of AOD and canyon indicators. The LU+AOD models were less accurate because the relatively coarse 3 × 3 km grid cell used for the AOD cannot always capture the concentration gradients near line or point sources. However, we found that the inclusion of the canyon indicators added significantly to the accuracy of the models. The effects were consistent for all air pollutant concentrations and in the expected direction.

It is important to note that our LUR models are associated with relatively small R² values (<0.8). Indeed, this is expected because our focus was only on the AOD and the canyon characteristics, and their effects on the concentrations of air pollutants. However, the determination of concentrations of air pollutants near roads includes other urban/regional effects, such as the presence of airports, distance to shorelines, land use characteristics, the presence of industrial sources, and restaurant emissions. Therefore, our model is not intended to be used for predictive purposes, but to demonstrate a method for capturing urban canyon determinants on air pollutant concentrations and to investigate the effects of a number of street-level characteristics that have not traditionally been used in LUR models.

Canyon indicators have been used recently to represent the effects of urban air pollution. The improvement of the model has obvious differences in different cases. Brauer

Table 7. Final LUR models for annual concentrations of SO₂, NO₂, and PM₁₀ in the central cities of Liaoning Province.

Response	Model	Equation	R ²	Adjusted R ²	CV R ²	RMSE	CV RMSE	Max VIF (variable)	p-value	Durbin-Watson
Annual SO ₂	LU only	$4.08e-02+7.74e-03 \times GDP-4.85e-02 \times NDVI-5.02e-06 \times DISTIE$	0.57	0.52	0.49	19.30	21.10	1.49(POP)	<0.001	2.718
	LU+AOD	$-3.01e-03+6.61e-03 \times GDP-1.55e-02 \times NDVI-4.17e-06 \times DISTIE+6.54e-02 \times AOD$	0.60	0.54	0.51	18.60	20.70	2.43(AOD)	<0.001	2.649
	LU+AOD+3D	$-4.38e-02+1.62e-03 \times GDP+5.48e-02 \times NDVI-3.52e-06 \times DISTIE+6.67e-02 \times AOD+2.73e-03 \times BCR2000+2.02e-02 \times FAR2000+7.78e-06 \times BSC2000+1.04e-03 \times HBR2000+4.10e-04 \times BH1000$	0.79	0.71	0.65	13.50	20.20	6.22(FAR2000)	<0.001	2.057
Annual NO ₂	LU only	$2.37e-02+1.02e-03 \times GDP+1.54e-03 \times POP+1.60e+00 \times RDA2000+7.65e-07 \times DISTIE$	0.49	0.42	0.37	6.90	8.42	1.93(GDP)	<0.001	1.835
	LU+AOD	$1.49e-02+6.26e-04 \times GDP+1.74e-03 \times POP+1.51e+00 \times RDA2000-8.61e-07 \times DISTIE+1.64e-02 \times AOD$	0.56	0.48	0.43	6.38	8.57	2.52(GDP)	<0.001	1.851
	LU+AOD+3D	$1.23e-02+2.24e-04 \times GDP+1.44e-03 \times pop+2.46e+00 \times RDA2000-4.21e-07 \times DISTIE+2.89e-02 \times AOD+4.79e-04 \times HBR2000+1.08e-04 \times BH600+1.20e-04 \times BCR400+1.14e-03 \times FAR400+3.35e-05 \times BSC200$	0.68	0.54	0.53	5.48	9.90	9.81(BCR400)	0.001	1.612
Annual PM ₁₀	LU only	$7.85e-02+7.18e-03 \times GDP-1.78e-04 \times ELEV-2.90e-06 \times DISTIE$	0.40	0.34	0.33	23.10	26.30	1.03(DISTIE)	0.002	1.813
	LU+AOD	$4.97e-02+5.10e-03 \times GDP-1.14e-04 \times ELEV-2.47e-06 \times DISTIE+6.21e-02 \times AOD$	0.43	0.35	0.33	22.50	26.70	2.16(AOD)	0.002	1.798
	LU+AOD+3D	$4.38e-02+6.80e-03 \times GDP-1.30e-04 \times ELEV-6.88e-07 \times DISTIE+7.25e-02 \times AOD+1.05e-03 \times HBR2000+5.25e-04 \times BH600+1.61e-03 \times BCR400+5.32e-03 \times FAR200+4.09e-04 \times BSC200$	0.79	0.67	0.55	13.70	28.70	4.26(BCR400)	0.001	1.985

et al. [34] indicate that the addition of canyon indicators derived from direct field observations to basic LUR-only models for $PM_{2.5}$ in Munich and Stockholm has made increases in R^2 of 4% and 6%, respectively. In our study, increases in R^2 of 19% and 39% (Table 7) indicate high-to-deep canyons in the central cities of Liaoning Province. Because of the differences between locations, the efficacy of canyon indicators differs substantially in different studies. However, we believe that urban morphology parameters are capable of not only explaining spatial distributions of urban air concentrations, but also play an important role in the selection of the predictor variable.

Many studies of street canyons and urban microscale pay attention to predominant meteorological situations, street textures, and the placement of air pollution monitoring devices because the interaction between the meteorology and the built environment causes variations – even within the same microenvironment, while these factors generally affect pollutant concentrations on a large scale. Weather conditions (e.g. wind speed, wind direction, and air temperature) in the surroundings could be changed by a building's shape. In this study, we have demonstrated the long-term effect of a building's shape on air pollutants without considering the short-term effect. We will consider how buildings affect air pollutant dispersion in a future work.

Limitations

The biggest limitation of this study is the small number of regulatory monitoring sites available for inclusion in the models. Unfortunately, the cost of a dedicated measurement campaign was too high, and therefore we used the 34 urban monitoring sites without consideration of the different effects between different canyon types, such as deep canyons and low-rise canyons.

Furthermore, the current approach relies on empirical parametrizations derived from high-resolution urban models, because the traditional method of surveying the area of interest to obtain the details of all the relevant buildings is expensive and time-consuming, and because remote sensing image are not always affordable.

Conclusions

Digital 3D building data are getting more and more available and they could be used to incorporate air pollution that is trapped by buildings to improve the accuracy of LUR air pollution models. This study has shown that incorporating canyon variables improves the performance of LUR models for the central cities in Liaoning Province, although the effect of incorporating annual AOD data is not that remarkable. The LUR model was found able to provide exposure assessments, and cohort studies on air pollution in Chinese cities are planned for the near future, although using our results for studies targeted at urban populations must be done with caution.

Acknowledgements

This research was co-funded by the National Natural Science Foundation of China (Nos. 41171155 and 41371198) and the “123” Project of the China Environment Protect Foundation (CEPF2013-123-1-8).

References

1. HOEK G., BEELEN R., DE HOOGH K., VIENNEAU D., GULLIVER J., FISCHER P., BRIGGS D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Environ*, **42**, 7561, **2008**.
2. AMINI H., TAGHAVI-SHAHRI S.M., HENDERSON S.B., NADDAFI K., NABIZADEH R., YUNESIAN M. Land use regression models to estimate the annual and seasonal spatial variability of sulfur dioxide and particulate matter in Tehran, Iran. *Sci Total Environ*, **488-489**, 343, **2014**.
3. CHUDNOVSKY A.A., KOUTRAKIS P., KLOOG I., MELLY S., NORDIO F., LYAPUSTIN A., WANG Y., SCHWARTZ J. Fine particulate matter predictions using high resolution Aerosol Optical Depth (AOD) retrievals. *Atmos Environ*, **89**, 189, **2014**.
4. KIM Y., GULDMANN J.-M. Land-use regression panel models of NO_2 concentrations in Seoul, Korea. *Atmos Environ*, **107**, 364, **2015**.
5. BERTAZZON S., JOHNSON M., ECCLES K., KAPLAN G. G. Accounting for spatial effects in land use regression for urban air pollution modeling. *Spatial and Spatio-temporal Epidemiology*, **14-15**, 9, **2015**.
6. LI X., LIU W., CHEN Z., ZENG G.M., HU C.M., LE N.T., LIANG J., HUANG G.H., GAO Z.H., LI Z., YAN W.F., HE X.X., LAI M.Y., HE Y.B. The application of semicircular-buffer-based land use regression models incorporating wind direction in predicting quarterly NO_2 and PM_{10} concentrations. *Atmos Environ*, **103**, 18, **2015**.
7. CHUDNOVSKY A.A., KOSTINSKI A., LYAPUSTIN A., KOUTRAKIS P. Spatial scales of pollution from variable resolution satellite imaging. *Environ Pollut*, **172**, 131, **2013**.
8. LI J., CARLSON B.E., LACIS A.A. How well do satellite AOD observations represent the spatial and temporal variability of $PM_{2.5}$ concentration for the United States? *Atmos Environ*, **102**, 260, **2015**.
9. LIN C.Q., LI Y., YUAN Z.B., LAU A.K.H., LI C.C., FUNG J.C.H. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level $PM_{2.5}$. *Remote Sens Environ*, **156**, 117, **2015**.
10. HAN Y., WU Y.H., WANG T.J., ZHUANG B.L., LI S., ZHAO K. Impacts of elevated-aerosol-layer and aerosol type on the correlation of AOD and particulate matter with ground-based and satellite measurements in Nanjing, southeast China. *Sci Total Environ*, **532**, 195, **2015**.
11. YOU W., ZANG Z.L., ZHANG L.F., LI Z.J., CHEN D., ZHANG G. Estimating ground-level PM_{10} concentration in northwestern China using geographically weighted regression based on satellite AOD combined with CALIPSO and MODIS fire count. *Remote Sens Environ*, **168**, 276, **2015**.
12. WANG J., CHRISTOPHER S.A. Intercomparison between satellite - derived aerosol optical thickness and $PM_{2.5}$ mass: implications for air quality studies. *Geophys Res Lett*, **30**, 2095, **2003**.

13. XIN J.Y., ZHANG Q., WANG L.L., GONG C.S., WANG Y.S., LIU Z.R., GAO W.K. The empirical relationship between the PM_{2.5} concentration and aerosol optical depth over the background of North China from 2009 to 2011. *Atmos Res*, **138**, 179, **2014**.
14. FARRELL W.J., DEVILLE CAVELLIN L., WEICHENTHAL S., GOLDBERG M., HATZOPOULOU M. Capturing the urban canyon effect on particle number concentrations across a large road network using spatial analysis tools. *Building and Environment*, **92**, 328, **2015**.
15. RICHMOND-BRYANT J., REFF A. Air pollution retention within a complex of urban street canyons: A two-city comparison. *Atmos Environ*, **49**, 24, **2012**.
16. KWAK K.H., BAIK J.J., RYU Y.H., LEE S.H. Urban air quality simulation in a high-rise building area using a CFD model coupled with mesoscale meteorological and chemistry-transport models. *Atmos Environ*, **100**, 167, **2015**.
17. HANG J., LI Y.G., SANDBERG M., BUCCOLIERI R., DI SABATINO S. The influence of building height variability on pollutant dispersion and pedestrian ventilation in idealized high-rise urban areas. *Building and Environment*, **56**, 346, **2012**.
18. EEFTENS M., BEEKHUIZEN J., BEELEN R., WANG M., VERMEULEN R., BRUNEKREEF B., HUSS A., HOEK G. Quantifying urban street configuration for improvements in air pollution models. *Atmos Environ*, **72**, 1, **2013**.
19. SU J.G., BRAUER M., BUZZELLI M. Estimating urban morphometry at the neighborhood scale for improvement in modeling long-term average air pollution concentrations. *Atmos Environ*, **42**, 7884, **2008**.
20. KR GER E.L., MINELLA F.O., RASIA F. Impact of urban geometry on outdoor thermal comfort and air quality from field measurements in Curitiba, Brazil. *Building and Environment*, **46**, 621, **2011**.
21. KONG S.F., DING X., BAI Z.P., HAN B., CHEN L., SHI J.W., LI Z.Y. A seasonal study of polycyclic aromatic hydrocarbons in PM_{2.5} and PM_{2.5-10} in five typical cities of Liaoning Province, China. *J Hazard Mater*, **183**, 70, **2010**.
22. CHU D.A., FERRARE R., SZYKMAN J., LEWIS J., SCARINO A., HAINS J., BURTON S., CHEN G., TSAI T., HOSTETLER C., HAIR J., HOLBEN B., CRAWFORD J. Regional characteristics of the relationship between columnar AOD and surface PM_{2.5}: Application of lidar aerosol extinction profiles over Baltimore-Washington Corridor during DISCOVER-AQ. *Atmos Environ*, **101**, 338, **2015**.
23. LEBOEUF A., BEAUDOIN A., FOURNIER R.A., GUINDON L., LUTHER J.E., LAMBERT M.C. A shadow fraction method for mapping biomass of northern boreal black spruce forests using QuickBird imagery. *Remote Sens Environ*, **110**, 488, **2007**.
24. PAN X.Z., ZHAO Q.G., CHEN J., LIANG Y., SUN B. Analyzing the variation of building density using high spatial resolution satellite images: the example of Shanghai City. *Sensors*, **8**, 2541, **2008**.
25. TUCKER C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens Environ*, **8**, 127, **1979**.
26. TUCKER C.J., SELLERS P.J. Satellite remote sensing of primary production. *Int J Remote Sens*, **7**, 1395, **1986**.
27. DADVAND P., RIVAS I., BASAGA A.X., ALVAREZ-PEDREROL M., SU J., DE CASTRO PASCUAL M., AMATO F., JERRET M., QUEROL X., SUNYER J., NIEUWENHUIJSEN M.J. The association between greenness and traffic-related air pollution at schools. *Sci Total Environ*, **523**, 59, **2015**.
28. XU L.Y., XIE X.D., LI S. Correlation analysis of the urban heat island effect and the spatial and temporal distribution of atmospheric particulates using TM images in Beijing. *Environ Pollut*, **178**, 102, **2013**.
29. SU J.G., JERRETT M., BECKERMAN B. A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. *Sci Total Environ*, **407**, 3890, **2009**.
30. PICARD R.R., COOK R.D. Cross-validation of regression models. *Journal of the American Statistical Association*, **79**, 575, **1984**.
31. KUTNER M.H., NACHTSHEIM C., NETER J. *Applied linear regression models*: McGraw-Hill/Irwin, **2004**.
32. O'BRIEN R.M. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, **41**, 673, **2007**.
33. MATEOS D., CACHORRO V.E., TOLEDANO C., BURGOS M.A., BENNOUNA Y., TORRES B., FUERTES D., GONZ LEZ R., GUIRADO C., CALLE A. Columnar and surface aerosol load over the Iberian Peninsula establishing annual cycles, trends, and relationships in five geographical sectors. *Sci Total Environ*, **518**, 378, **2015**.
34. BRAUER M., HOEK G., VAN VLIET P., MELIEFSTE K., FISCHER P., GEHRING U., HEINRICH J., CYRYS J., BELLANDER T., LEWNE M., BRUNEKREEF B. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology*, **14**, 228, **2003**.