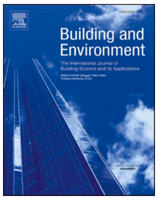
Building and Environment 188 (2021) 107479

Contents lists available at ScienceDirect

Building and Environment

journal homepage: http://www.elsevier.com/locate/buildenv



Assessing neighborhood variations in ozone and $PM_{2.5}$ concentrations using decision tree method

Ya Gao [a], Zhanyong Wang [b], Chao-yang Li [a], Tie Zheng [a], Zhong-Ren Peng [c,*]

[a] School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
[b] College of Transportation and Civil Engineering, Fujian Agriculture and Forestry University, Fuzhou, 350108, China
[c] International Center for Adaptation Planning and Design (iAdapt), School of Landscape Architecture and Planning, College of Design, Construction, and Planning, University of Florida, P.O. Box 115706, Gainesville, FL, 32611-5706, USA

ARTICLE INFO

Keywords:
Air pollution
Spatial variability
Decision tree model
Neighborhood scale
Urban form

ABSTRACT

Typical air pollution events involving ozone (O_3) and $PM_{2.5}$ occurred frequently in China, while the fine-scale pollution variation, especially at a neighborhood level (2 km*2 km), is complex and still not clear. To assess how urban form and meteorology influence neighborhood air pollution distribution, this study took the Minhang district in Shanghai, as experimental cases, and performed a neighborhood-scale investigation on O_3 and $PM_{2.5}$ by using mobile measurements. Both land-use regression model and decision tree model were used to examine the relationship between air pollutant concentration and influenced variables. As the decision tree model captured the linear and non-linear relationship between variables, it was demonstrated that explained more variations of O_3 and $PM_{2.5}$ concentrations than the LUR model. The results also showed that O_3 concentrations were mainly affected by meteorological factors while $PM_{2.5}$ concentrations were more heavily determined by background level and residential area. Both O_3 and $PM_{2.5}$ showed a significant correlation with air temperature, traffic volume, building height, and green space. Interestingly, green spaces were negatively correlated with the $PM_{2.5}$ variations, which was almost the opposite to that of O_3. With the superiority to the discrete observation, the decision tree model based concentration surfaces clearly revealed the heterogeneity of O_3 and $PM_{2.5}$ distributions. This study not only preliminarily identifies the impacts of land-use type and meteorological factors on the spatial patterns of O_3 and $PM_{2.5}$, but also provides a possible alternative method for assessing the neighborhood air pollution in the future.

1. Introduction

Ozone (O_3) and fine particulate matter ($PM_{2.5}$) are the two most important pollutants for current air pollution control in China, which have attracted extensive attention and huge amounts of efforts have been devoted to reducing their adverse impacts [1]. Studies have demonstrated even at a small scale, these pollutants could vary greatly in the urban environment due to the spatial variation of traffic and other emission sources [2]. Therefore, the risk of exposure to these pollutants is usually higher during personal activities in an urban microenvironment [3]. Furthermore, almost all existing researches on air pollution have been conducted at an urban, regional, national, or global scale, but evidence from the neighborhood scale is very scant. It is therefore essential and urgent to fill the gaps by predicting the high-resolution spatial patterns of air pollution and exploring the link between neighborhood environment and air pollution.

A few researchers have explored the relationship between air pollutant distribution with topographic features, economic development, and population density of areas in urban China [4,5]. These findings suggest that the spatial variation of air pollution in a city can be explained by various influencing factors with model R^2 values exceeding 50%. But, air pollution assessment of the entire city may be inaccurate if it is used directly in the analysis and evaluation of neighborhood-scale air pollutant exposure [6]. While averaging data in a city may comply with regulatory standards, health impacts are evident on a localized scale [7]. In this regard, the current city-scale research cannot be applied directly to assess exposure at a neighborhood scale due to the following reasons: low resolutions for urban form calculation, lack of micro-environments (e.g. on-site meteorology and traffic flows), and lack of comparisons among records within one community incurred by

* Corresponding author.
E-mail address: zpeng@ufl.edu (Z.-R. Peng).

https://doi.org/10.1016/j.buildenv.2020.107479
Received 20 June 2020; Received in revised form 17 October 2020; Accepted 19 November 2020
Available online 26 November 2020
0360-1323/© 2020 Elsevier Ltd. All rights reserved.

sparse monitor network [8–11]. Additionally, the Chinese Center for Disease Control and Prevention (CDC) has put forward the framework of environmental pollution and the disease monitoring system in a neighborhood scale. The framework and system were subsumed into the National Environmental and Health Action Plan, and cities like Shanghai, Nanjing, Qingdao, and Taiyuan became the first batch of pilot cities [12]. Therefore, a deep insight into the spatiotemporal evolution of air pollutants at the neighborhood scale (here is less than 2 km) will contribute to the improvement of air quality through effective measures.

Existing studies further suggest that it is important to clarify the dynamic spatial variation of air pollutants such as O_3 and $PM_{2.5}$, as this it would allow risk assessors to better predict exposures in locations with different meteorological, built, or geological features. The distributions of air pollutants are greatly affected by the surrounding built environment and meteorology which may change rapidly according to time and space [4,13]. Generally, influencing factors include static and dynamic factors. Static influencing factors can be categorized into two groups [14]: area-integrated and distance-based factors. Area-integrated variables include land-use types (residential, commercial, industrial, green space, bodies of water, etc.), demographics, and road networks. These features were often quantified in varied buffer distances of 5–10 km around a monitoring site [4]. Distance-based variables incorporate the distance from monitors to the sea, major air pollution sources, longitude, and latitude. Apart from static factors, there are certain dynamic factors such as traffic flow and meteorology. Su et al. [15] added wind speed, wind direction, and insolation in the traditional LUR to create a SA-LUR model to analyze the exposure of NO and NO_2. The SA-LUR model performed better than the traditional LUR in this regard. Researchers also proposed general models to quantify how traffic volume and meteorological conditions influence the level of air pollution. The well-built models revealed that the most important predictor variables were closely related to wind and traffic volume [16,17]. Moreover, these model-based studies pointed out that further research should consider more potential influential factors of air pollution at a finer spatiotemporal scale.

To date, a number of studies have analyzed the patterns of pollutant variations on the basis of a routine monitoring network, which is sufficient in terms of a city scale but not enough for the neighborhood scale. With the development of portable monitors, mobile monitoring is being used extensively to capture high-density data at more locations. This method provides a good chance to measure the fine-scale variability of outdoor pollutant concentrations at pre-selected sites or routes [18,19]. To make full use of this high-resolution data, various models such as deterministic models and empirical models have been applied to enhance the explanation of air pollution distribution. Deterministic models are based on numerical simulation technology, where various physical and chemical mechanisms of air pollutants from emission, transformation, diffusion, and transportation are integrated. Weather Research and Forecasting (WRF) and Community Multiscale Air Quality (CMAQ) are two widely used deterministic models [20]. In deterministic models, different combinations of macro-scale physical process parameterization schemes (e.g. planetary boundary layer schemes and land surface modeling) have significantly affected the accuracy of model results [21]. Hence, deterministic models have limited the analysis of air quality at micro scales.

Empirical models usually refer to statistical models that investigate mathematical relationships between predictor variables and air pollutants. Among empirical approaches, linear regression models and machine learning methods have received more attention. The land-use regression approach can estimate pollutant concentrations at unmonitored locations, by establishing a relationship between pollutant concentration measured at monitoring sites and influencing factors such as surrounding land use, traffic volume, and physical characteristics [13,22,23]. This method accurately describes the linear dependencies between variables. Yet, in an open and complex environment, both linear and nonlinear relationships appear among influencing factors and atmospheric pollutants. Linear regression models have thus failed to capture the complex interactions, especially the nonlinear relationships, among variables [24,25].

In recent years, machine learning methods such as Support Vector Machine [26], Multi-layer perceptron [27], and Sequence learning [21] have been applied to air pollution and epidemiology. These models take nonlinearity into account and have been proven to perform well, but they cannot rank the influencing variables based on their importance which can be compared to the relevance or R^2 in the LUR models. The decision tree model is resistant to these potential problems, and it can produce more accurate estimations of air pollutant concentrations [28]. Cole et al. [24] thought the decision tree model was often conducted in prediction analysis because of its increased accuracy and resistance to multi-collinearity and complex interaction problems as compared to linear regression. The decision tree model comes with the advantages of tree-based methods, which possess the ability to capture complex interactions and maintain low bias while alleviating the problem of uncertainty. Athanasiadis et al. [29,30] developed a novel land-use decision tree model to predict the elemental components of particulate matter with results being more accurate. What's more, decision trees as a supervised machine learning approach can not only predict the value of a target variable by learning decision rules inferred from the data feature [29,30], but also identify the relationships between responding variables and predictive variables [31]. Kunwar et al. [32] established tree ensemble models for seasonal discrimination and air quality prediction. These models performed better than support vector machine models, identifying fuel combustion and vehicular emissions as major air pollution sources. They can also capture complex interactions and maintain low bias by observing a system with the minimum human intervention [33]. Thus, the decision tree seems suitable to assess the spatial variations of air pollutant concentrations at small scales.

Our work is trying to use the decision tree model to assess spatial variation of air pollutants at a neighborhood scale and identify significant influencing factors. $PM_{2.5}$ and O_3 are selected as research objects because they exceed the standards in the air quality guidelines published by the World Health Organization, showing obvious distinctions in spatial distribution according to the previous studies. Taking the Minhang district in Shanghai as an experimental case, this study plans to achieve the following goals: 1) to illustrate the spatial variations of $PM_{2.5}$ and O_3 concentrations at a neighborhood scale; 2) to estimate the effects of different factors on the spatial distribution of $PM_{2.5}$ and O_3 concentrations; 3) to verify the effectiveness of the decision tree model for neighborhood air pollution prediction. In addition, high-resolution maps of air pollution are expected to visually assess subtle variations in exposure to air pollutants in the community of a city.

2. Data and methods

2.1. Study area

Shanghai is the largest metropolitan area in East China, with a population of more than 24.15 million in 2018, and more than 12 urban thousand residential neighborhoods. The experimental site was carried out at a neighborhood scale (2 km*2 km) in the Minhang District of Shanghai, with mobile platforms around 20 monitoring locations, as shown in Fig. 1. Emission sources from traffic emission, house cooking emission, and industrial operations are all prevalent in this district. The study area is located at 31.21° N and 121.30° E, 5 km west of the Hongqiao Transportation Hub, lying on a plain with nearly no elevation variation.

The urban form at the neighborhood scale is the morphology of urban socio-ecological systems. Not only does it serve people's immediate needs for life, work, and spirituality, but it is also tangible and directly appreciable by people in their daily lives. This study area is a typical residential neighborhood containing various elements: the south-north eight-lane Huaxiang & Jiamin Highway, a west-east four-

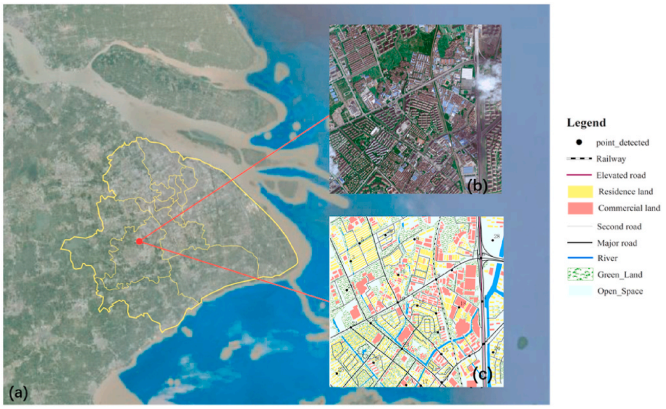


Fig. 1. Sketch map of the study area. (a–b) Google maps of the experimental field in the Minhang District, Shanghai; (c) GIS map of the Minhang study area.

lane Beiqing Artery across the north part, some other minor arterials such as Zhuxin Rd, Fanxing Rd, Beidi Rd, and Zidi Rd as well as some circuit branches spreading over the area. There is a residential area with 7-story buildings and rows of low houses. While some places tend to be crowded with people and vehicles while some places are surrounded by green spaces and bodies of water. There are also several industrial areas near the study area.

2.2. Instrumentation and measurements

Five sets of portable monitors were used to detect pollutants at different monitoring sites. Each set contains an ozone monitor, an aerosol monitor, a weather station and a video camera. The commercial portable ozone monitor (Model: POMTM, 2B Tech, USA) and the commercial portable aerosol monitor (Model: SidePakTM AM510, TSI, USA) were used in the experiment. Readings of the O_3 and $PM_{2.5}$ monitors were set with a time resolution of 10 s and 5 s, respectively. In addition to capturing air pollutant concentration (including O_3 and $PM_{2.5}$) observations, in situ measurements of meteorological parameters (including temperature, relative humidity, air pressure, solar radiation, wind speed and wind direction) were made by a Davis Vantage Weather Station with a time resolution of 1 s. During the measurement, the range of air wind speed and temperature sat as 1.5–4 m/s and 16 and 25 $^{\circ}C$, respectively. Traffic volumes were recorded separately by cameras in the proximity of each sampling site.

According to previous research, the number of monitors varies widely based on availability, and there is no rigorous methodology to determine the required number of monitoring sites. In this study, 20 monitoring sites were selected base on the availability of analytical and human activity data. The monitoring density is 0.2 km^2 in the study area. Experimental data was measured at each monitoring site, similar to that of Kumar et al. [34] and Gao et al. [11]. The first round of field campaign was conducted twice each day, i.e., morning (7 a.m.–12 p.m.) and afternoon (1 p.m.–8 p.m.). It lasted for 17 days from April 7th to May 10th, 2018, excluding rainy days. As we only have five sets of portable instruments to simultaneously measure twenty monitoring sites, all of the monitoring sites were divided into five groups according to the distance between each other to acquire a representative data set. Each group contains four monitoring sites. The samples were taken for 20 min at each monitoring site, and five monitoring sites of the different group were measured simultaneously [35]. It took about 60 min to complete one set at all the 20 monitoring sites. Sampling was done in twelve sets of measurements every day.

In order to clarify daily variations of hourly average $PM_{2.5}$ and O_3 concentrations, additional measurements were made continuously at 24 h a day for four days (May 13–16) in two sampling points (hereafter named the second round of field campaign). Daily maximum O_3 concentrations (hourly mean) on 12 days exceeded the threshold (100 ppb) of the ambient air quality standards class II in China during the first field campaign. Additionally, the maximum hourly mean O_3 concentration of 115.6 ppb occurred at LT 15:00 on 3 May, while the maximum daily mean O_3 concentration of 101.3 ppb occurred on 19 April. In contrast, $PM_{2.5}$ concentrations were in relatively low levels with only three exceedance day (where the daily $PM_{2.5}$ concentration exceeded $75\mu g/m^3$).

2.3. Development of predictor variables

Studies have shown that air pollution is often related to traffic conditions, topographic features, economic development, population density and local meteorology of the area [5,12]. Hence, a total of 38 independent variables in four categories concerning traffic emission, community form, meteorology and background concentration were considered in this study. As shown in Table 1, we subdivided each category by distance variables at different radii (50,150 and 300 m) through the ArcGIS 10.0 and Python 3.0 software. The background monitoring sites were co-located with continuous ambient monitoring sites operated by the Shanghai Environmental Monitoring Center, which neglected the impact of some factors including industrial plants, arterial roads and parking lots. Further information on the measurements of air pollution, meteorology and traffic volume, together with the process of data acquisition of community form and preparation of all the data, have been introduced in detail in our previous study [11].

2.4. Decision tree model

Among the factors that influence urban air pollution, some variables, such as human activities, are highly uncertain and difficult to measure in linear models. The decision tree model has an advantage in this aspect. The decision tree algorithm is widely used in fields of medical diagnosis, university evaluation and weather alerts, but is rarely concentrated on environmental pollution assessment. In addition, the decision tree model also ranks the inputs based on their importance which can be compared to the relevance or R^2 in traditional LUR models. The major algorithms used to build decision trees include ID3, C4.5 and CART (Classification and Regression Trees). C4.5 is used in this paper because

Table 1
Classification and description of each independent variable.

Category	Variable subcategories	Unit	Variable name
Transport land (12 variables)	Road network	m	roa_xx
	Subway	m	sub_xx
	Intersection	count	Inter_xx
	Public transportation stop density	count	stop_xx
Traffic volume (3 variables)	Traffic volume	vph	tv_xx
Community form (15 variables)	Green space	m^2	vege_xx
	Body of water	m^2	wat_xx
	Residential land	m^2	res_xx
	Commercial land	m^2	com_xx
	Building floor	m	flo_xx
Meteorology (6 variables)	Air temperature	$^{\circ}C$	at
	Relative humidity	%	rh
	Solar radiation	W/m^2	sr
	Air pressure	hPa	ap
	Wind direction	°	wd
	Wind speed	m/s	ws
Other (2variable)	Background concentration	ppb/$\mu g/m^3$	O_3_bc/ $PM_{2.5}$_bc

Note: xx = the circular buffer radii (meters), 50, 150, 300; Road network = the length of all roads in the buffer; Intersection = the number of all intersections in the buffer; Public transportation stop density = the number of bus and metro stops.

it is designed to handle continuous data and missing value, select attributes based on information gain ratio instead of information gain (information gain based model tends to select attributes having more variables), and use pre-pruning to keep the model from overfitting. Hence, this paper proposes a C4.5 decision tree construction algorithm that support both analysis of influencing factors and pollutant concentration prediction. The detailed procedures of decision tree model development are shown in Fig. 2.

Furthermore, to prevent the model from falling into the local minima, we've adjusted the information gain ratio of the features with more variables by incorporating the weighting factor into the model to increase their information entropy. The calculation of the weighting factor is calculated as shown below:

Suppose dataset S has |s| samples, A is one continuous attribute of this dataset, which divides the dataset into $\{A_1\ , A_{2\ ,}\ \ldots\ldots A_v\}$ corresponding to the values $\{a_1\ , a_{2\ ,}\ \ldots\ldots a_v\}$, where $|A_j|$ represents the number of samples when A equals to a_j and $|A_{ij}|$ represents the number of samples whose labels are i in subset A_j. Then the weighting factor is calculated from the function below:

$$\lambda = \frac{\sum_{j=1}^{v}\sum_{i=1}^{n}\left||A_{ij}| - |E_{ij}|\right|}{2^{*}|S|} \tag{1}$$

where $|E_{ij}|$ is the joint expectation of the subset A_j and the subset C_i corresponds to attribute A and label C, respectively. The calculation of $|E_{ij}|$ is shown below:

$$|E_{ij}| = \frac{|C_i| * |A_j|}{|S|} \tag{2}$$

The specific process is described as follows:

2.4.1. Categorize the continuous attributes

The algorithm firstly starts by sorting the attributes in descending order and creates potential splits from mid-points between adjacent values. Then the algorithm can evaluate the information gain of each potential split, followed by selecting the split with the largest information gain as the best split for the attributes.

For example, the probability of a_j is:

$$P(A_j) = \frac{|A_j|}{|S|} \tag{3}$$

In the subsetA_j, the probability of C_i is：

$$P(C_i|A_j) = \frac{|A_{ij}|}{|A_j|} \tag{4}$$

The information entropy of attribute A is：

$$\mathrm{Info_A}^{*}(S) = \sum_{j=1}^{v}[P(A_j) + \lambda] * \left[-\sum_{i=1}^{n} P(C_i|A_j)\log_2 P(C_i|A_j)\right] \tag{5}$$

The information entropy of the attribute represents the total uncertainty of labels in response to a particular attribute – the information used to create splits for the attribute. For example, the probability of label C_i is:

$$P(C_i) = \frac{|C_i|}{|S|} \tag{6}$$

The information entropy of dataset S is:

$$\mathrm{Info}(S) = -\sum_{i=1}^{n} P(C_i)\log_2 P(C_i) \tag{7}$$

The information entropy of a dataset represents the total uncertainty of labels in the dataset – the information used to split labels when building decision trees. Thus, the information gain of attribute A is:

$$Gain(A) = Info(S) - \mathrm{Info_A}^{*}(S) \tag{8}$$

2.4.2. Generate a decision tree

The algorithm calculates the information gain ratio of the best split as the gain ratio of the attribute, and selects the attribute wiyj the largest gain ratio as the test attribute to split the dataset. The initial decision tree can then be generated.

Let attribute A split the dataset S, and the generated information entropy is:

$$\mathrm{Split_A}(S) = = -\sum_{j=1}^{v} P(A_j)^{*}\log_2 P(A_j) = -\sum_{j=1}^{V}\frac{|A_j|}{|S|}\log_2\frac{|A_j|}{|S|} \tag{9}$$

The information gain ratio is:

$$\mathrm{GainRatio}^{*}(A) = \frac{|\mathrm{Info}(S) - \mathrm{Info}_A^{*}(S)|}{\mathrm{Split_A}(S)} \tag{10}$$

2.4.3. Pruning

The algorithm uses the error-based pruning technique to calibrate the initial decision tree and generate the optimal decision tree. The model can also select the optimal attribute of the current node in the attribute set, which corresponds to the total of 38 independent variables.

In order to maximize the benefits from data, we adopted the 10-flod cross-validation (10-CV) method instead of a training-and-testing method to test the fit of the decision tree model. Detailed information about 10-CV has been provided in previous studies [36]. The root mean error （RMSE）and the coefficient of determination (R^2) are calculated as the indicators of the model performance. In addition, using the final accurate and stable decision tree models, we divided the study area into a group of square grids at a spatial resolution of 100 m and estimated the grid-point-based predictions (including 400 points) of mean $PM_{2.5}/O_3$

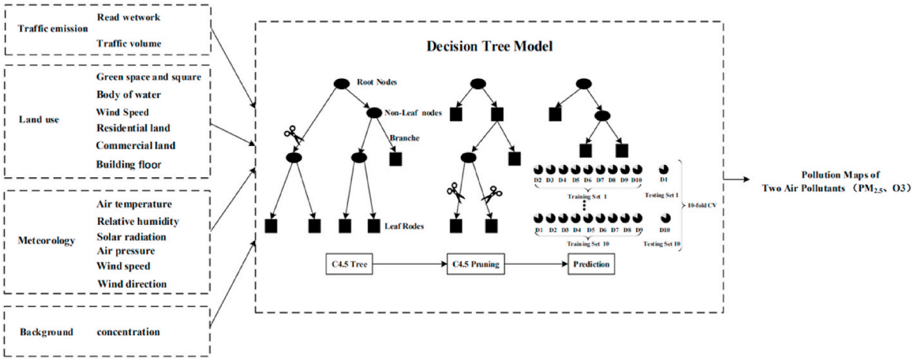


Fig. 2. Architecture of decision tree model.

concentrations. Consequently, the continuous pollution maps of $PM_{2.5}/O_3$ were created with the ArcGIS 10.0 software, better illustrating the fine variations of air pollutant concentrations.

3. Results

3.1. Temporal characteristics of O_3 and $PM_{2.5}$

As shown in Fig. 3a, the O_3 concentrations in the early morning showed a decreasing gradient. This can be explained by the fact that massive nitrogen oxides from urban traffic contributed to the removal of atmospheric O_3, and the clear-up rate markedly exceeded the secondary formation of photochemical reaction due to the lower solar radiation during the period. Then as the day advanced, the increased solar radiation substantially improved the formation of O_3. As a result, the near-ground O_3 concentrations accumulated rapidly. While in the afternoon, the solar radiation gradually attenuated and the generation rate of O_3 from photochemical reaction also exhibited a sharp decrease. The O_3 peak appeared during the period when the secondary formation rate equaled the clear-up rate. The O_3 concentrations then rapidly decreased. Especially in the evening, the secondary formation of O_3 would cease; However, the removal of O_3 frequently occurred due to its active involvement in the homogeneous and heterogeneous chemical reactions during the nighttime. Furthermore, the impact of dry deposition at ground level also resulted in a rapid reduction of O_3 concentrations. Therefore, the nocturnal O_3 usually showed a gradually decreasing trend.

The diurnal variations of the $PM_{2.5}$ concentrations exhibited significant bimodal distribution patterns (Fig. 3b). It is possible that the two peaks were associated with the changing characteristics of urban vehicular activities and meteorological factors across a day. The increasing vehicular emissions and stronger photochemical reactions during the morning rush hour contributed to a gradual increase in the $PM_{2.5}$ concentrations, and the $PM_{2.5}$ peak occurred before LT 9. During LT 11–15, the uplift in the atmospheric boundary layer (ABL) and the increased atmospheric turbulence both improved air pollutants to diffuse upwards, leading to a rapid decline in $PM_{2.5}$ concentrations. However, the $PM_{2.5}$ concentrations from LT 16 to LT 19 gradually increased and this trend mainly comes from the contraction of ABL, household emissions (e.g., cooking fumes) and the increase in vehicular exhausts during the evening rush hour. Near-ground $PM_{2.5}$ concentrations generally showed a same diurnal variation, as O_3 did, which demonstrates human activities and meteorological factors have significant impacts on the diurnal trends of both pollutants.

3.2. Model fitting and validation

In this study, the daily average O_3 and $PM_{2.5}$ concentrations were calculated for each monitoring site based on the quality-controlled data per 5 or 10 s and prepared as the dependent variables for the model. In order to correspond with the pre-processed daily air pollutant data, we also separately calculated the daily average values of micro-level meteorology and traffic volume. Based on previous findings, the variables with spatial scaling effects (e.g., community form and transport land) were extracted at a 50,150 and 300 m buffering radius. All these potential predictors were prepared as the modelling inputs. All independent variables were standardized and normalized before calculation. In the procedure of extracting GIS independent variables of different buffers, many zero values also appeared at a few monitoring sites that were within 50 m or 150 m of a subway length, body of water, public transportation stop, and so on. Independent variables with zero values may induce influential calculations. Generally, variables with more than five zero values were excluded from the decision tree model. Considering the varying physical and chemical properties, decision tree models were separately developed for O_3 (317 groups) and $PM_{2.5}$ (298 groups) datasets based on daily average concentrations at 20 monitoring sites.

The overall fit of the two models was assessed by using a summary of adjusted fitting R^2 and root mean square error (RMSE) between the observed air pollutant concentrations and the estimated ones. As shown in Table 2, the adjusted R^2 of the LUR model is 0.71 for O_3 and 0.58 for $PM_{2.5}$, and the corresponding adjusted R^2 of decision tree model is 0.85 and 0.76, respectively. This indicated that the prediction precision of decision tree models can be substantially improved when considering linear and nonlinear relationships between influencing factors and the air pollutant concentration. Fig. 4 shows the comparison of fitting and cross-validation results for decision tree models. From the results of the decision tree model, the adjusted R^2 values were found to be higher for O_3 (0.85), followed by $PM_{2.5}$ (0.76), which demonstrate that the predictive performance of decision tree models were stable during the whole study period. Correspondingly, the CV R^2 values were 0.82 for O_3 and 0.72 for $PM_{2.5}$ that are close to the model fitting parameters, suggesting the robustness of the model. All of the RMSE values for O_3 were less than 3.5 *ppb*, which also indicate that the predicted values fitted well with the measured values [37]. In contrast, the $PM_{2.5}$ models with a mean RMSE of 4.66 were slightly worse than the O_3 models. Additionally, the spatial autocorrelation of model residuals was calculated by Globe Moran's I in ArcGIS. The *p* values for models were greater than or equal to 0.05, which indicates weak spatial autocorrelation of the residuals.

3.3. Contributing factors to O_3 and $PM_{2.5}$ identified by decision tree models

All significant contributing factors are listed in Table 3. Six contributing factors, including solar radiation, relative humidity, air temperature, and green space, within 150 m buffer traffic volume with 300 m buffer, and building floor within 50 m buffer, can ultimately explain more than 80% of the variance of O_3 concentrations. This indicates that the variations of O_3 in the study area were mainly linked to meteorological variables. Compared to the O_3 model, there are more urban-related factors in the $PM_{2.5}$ model, such as transport land, residential land, building floor and green space. In both models, the

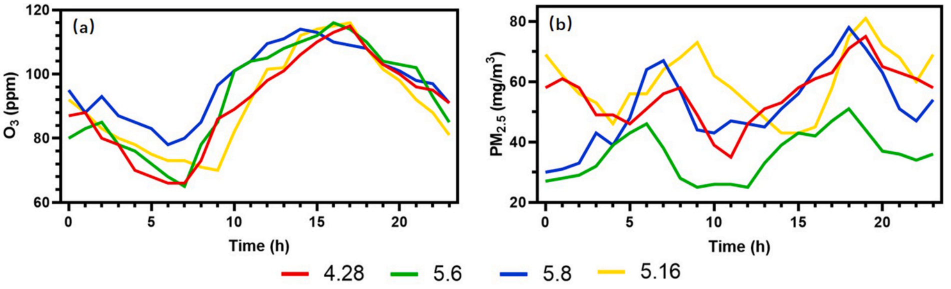


Fig. 3. Time series of air pollutant concentrations. (a) and (b) are the daily variations of hourly average O_3 and $PM_{2.5}$ concentrations for four days (the second round of field campaign), respectively. X-axes represents the 24 h of a day.

Table 2
Comparison of model fitting and cross-validation results for decision tree and LUR models.

Air pollutant	Model Fitting				Cross-Validation			
	Adj R^2		RMSE(ug/m^3)		Adj R^2		RMSE(ug/m^3)	
	GAM	LUR	GAM	LUR	GAM	LUR	GAM	LUR
O_3	0.85	0.71	3.39	7.57	0.82	0.61	2.44	7.85
$PM_{2.5}$	0.76	0.58	4.14	12.15	0.72	0.56	5.18	15.21

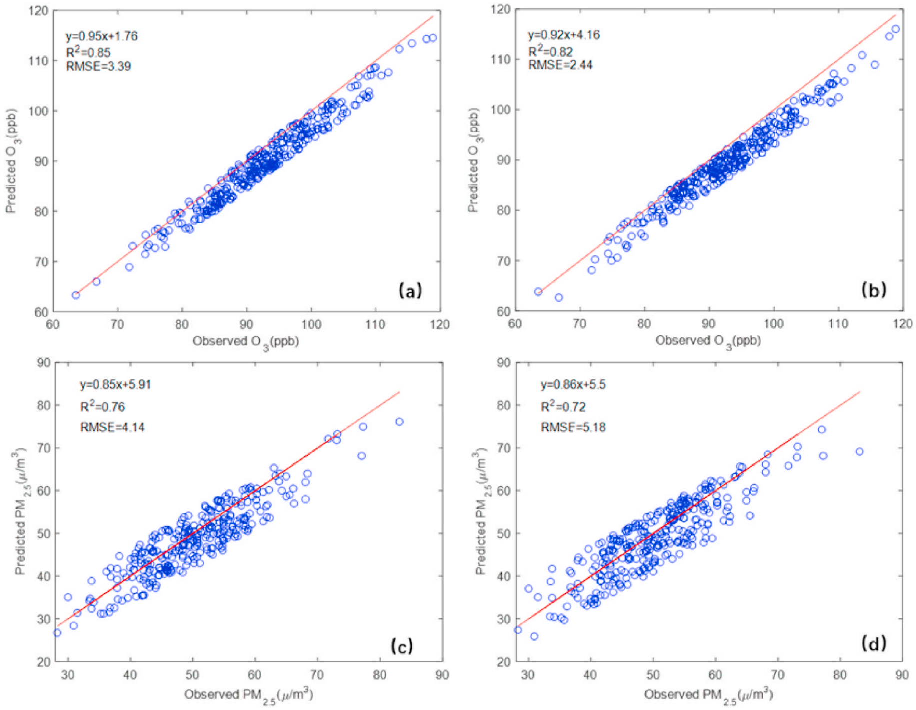


Fig. 4. Scatterplots of the fitting and validating results of decision tree models. (a) and (b) are the fitting results of decision tree models compared with the observations for O_3 and PM $_{2.5}$, respectively; (c) and (d) are the 10-fold validating results of the decision tree models compared the observation for O_3 and PM $_{2.5}$, respectively.

Table 3
The importance of predictor variables for O_3 and $PM_{2.5}$ estimation by decision tree models.

O_3	sr	rh	at	vege_300	tv_300	flo_50		
	0.329	0.248	0.18	0.138	0.102	6.81e-05		
$PM_{2.5}$	$PM_{2.5}$_bc	res_150	tv_300	roa_150	wd	vege_150	at	flo_50
	0.252	0.188	0.156	0.136	0.124	0.071	0.066	5.17e-05

sr_xx = solar radiation; rh = relative humidity; at = air temperature; vege_xx = green space; tv_xx = traffic volume; flo_xx = building height; $Pm_{2.5}$_bc = background concentration near the study area; roa_xx = road network; res_xx = residential land; wd = wind speed.

measured air pollutant concentrations are sensitive to the variations of air temperature (at), traffic volume (tv_xx), building height (flo_xx) and green space area (vege_xx). As described in the literature [38,39], $PM_{2.5}$ levels vary drastically at small scales when driven by the spatial patterns of traffic and other sources. O_3 is produced by a chemical reaction between sunlight and emission from vehicles. In addition, traffic-related pollutants cannot easily spread when dense high-rise residential buildings reach a certain height [11].

In all contributing variables, solar radiation was found to be the main variable in the O_3 model, which is consistent with its paramount role in photochemical reaction kinetics [40]. Furthermore, O_3 concentrations also exhibited strong diel variations due to the diurnal cycle of solar radiation (Fig. 5a). It is because that O_3 is the product of a photochemical reaction and the intensity of photochemical reaction is mainly affected by solar radiation intensity [41]. Fig. 5a shows the scatterplots between solar radiation and O_3. As the solar radiation intensity increases, O_3 accumulates rapidly through a photochemical process. The second significant variable was relative humidity (RH). As indicated in previous studies, low surface RH levels are frequently associated with subsided air parcels which are extremely dry. Subsided air parcels could also bring stagnant weather with high temperature and strong solar radiation. These weather conditions were typically favorable for the photochemical production and the accumulation of O_3 [42]. This is in consistency with our experimental data. In the seven day of O_3 pollution when daily O_3 concentrations exceeded the threshold (100 ppb), it were observed that the measured RH values were significantly lower (the average RH was 59.8%) than on other days (the average RH was 78.5%). The effect of air temperature on spatial distribution of O_3 arranged the 3rd order. Air temperature plays an important role in regulating the O_3 formation and diffusion processes. As shown in Fig. 5c, the response of

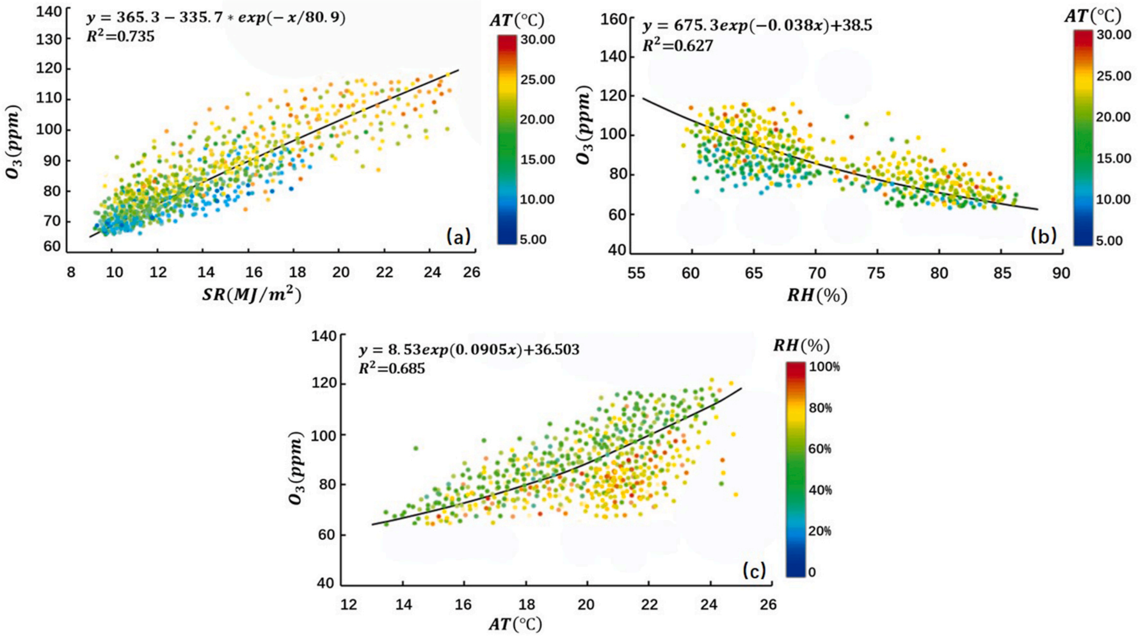


Fig. 5. Scatter plots between O_3 concentrations and meteorological factors selected by decision tree model, respectively, where (a) is solar radiation, (b) is relative humidity and (c) is air temperature.
SR = solar radiation AT = air temperature RH = relative humidity.

O_3 to the change in air temperature was positive, and the growth rate of O_3 concentrations became higher when the temperature continued to rise. The increase of air temperature with enhanced solar radiation often increases O_3 concentration [43].

The measured O_3 concentration is sensitive to the area of green space (Pearson correlation coefficient = 0.65, $p < 0.001$). Volatile organic compounds (VOCs) involving both anthropogenic (AVOCs) and biogenic (BVOCs) sources are favorable to the photochemical production of O_3. The VOCs' emission of vegetation accounts for over 90%. When NOx concentration is sufficient, some common BVOCs can form O_3 more easily than AVOCs emitted by the human being activities [44]. In contrast, green space has a reducing effect on $PM_{2.5}$ concentrations (Pearson correlation coefficient = −0.78, $p < 0.001$). On the one hand, there's a lack of significant emission sources in green spaces; on the other, plants can absorb atmospheric particulate through tiny openings in their leaves. This is consistent with previous researches that the concentration of $PM_{2.5}$ is much lower in vegetation coverage surfaces than in built-up urban areas [45].

As shown in Table 3, the background concentration was the most significant influencing factor in the $PM_{2.5}$ model. When conducting partial regression, there is a linear relationship (R^2 is 0.685) between the site $PM_{2.5}$ concentration and background concentration. In previous studies, a control station for clean air located far away from the city was always selected for the background concentration estimation [45]. However, in evaluating the personal exposure at a neighborhood scale, it would be inaccurate to adopt such a background value coming from other regions, as the concentrations in every region vary significantly from each other and in every day. The background monitoring site in this research was co-located with continuous ambient monitoring sites operated by the Shanghai Environmental Monitoring Center in the same district.

The variable of residential land area was the second significant variable in the $PM_{2.5}$ model. In recent years, with increased population and improved living standards, resident activities have become an important source of energy consumption and pollutant emissions. Buildings with improper distribution and high density in residential areas are leading to poor ventilation, thereby suppressing the diffusion of pollutants [46]. Meanwhile, air temperature and wind have also been found to play important roles in regulating the $PM_{2.5}$ formation and diffusion processes within the lower troposphere, which coincides with previous findings [17].

3.4. Neighborhood map of modeled O_3 and $PM_{2.5}$ concentrations

The well-built decision tree models with high R^2 were finally applied to a regular 100 m*100 m grid covering the entire study area to generate a continuous surface of pollutants concentrations. The results of the predicted concentration map across the study area are shown in Fig. 6. The study area was more polluted by O_3 than $PM_{2.5}$. The highest $PM_{2.5}$ concentrations appeared in residential areas where the population density is high, revealing that $PM_{2.5}$ was more heavily impacted by household activities within the second significant predictor variable of

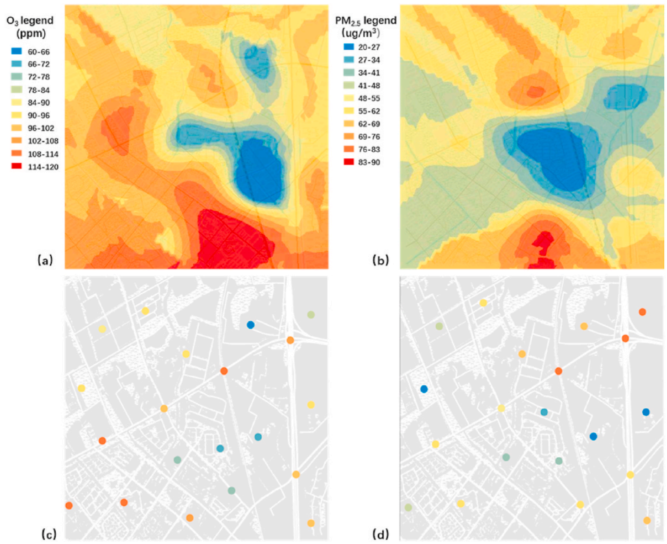


Fig. 6. Spatial distribution of modeled O_3 and $PM_{2.5}$ concentrations: (a) and (b) are the modeled air pollutant concentrations; (c) and (d) are the measured air pollutant concentrations.

residential land area. For the predicted concentration map of O_3, O_3 concentrations in areas with more green spaces were significantly higher than in other areas. But areas of green space, including the ecological park, show an obvious reducing effect on $PM_{2.5}$ concentrations. Although traffic volume was found to be a significant variable with a positive association in the two models, pollutants appeared to spread to surrounding areas due to the wind or the potential influence from the periphery of the study area [7].

4. Discussion

Air pollution variations and their relationship with land use have attracted much attention, but there's currently a lack of research on the neighborhood scale, especially in China [11,36]. Trust of the regular monitoring network in most of the previous studies is low due to its biased estimate for public exposure [13]. In order to acquire high dense data, this study has set a neighborhood scale parameter of 2 km*2 km. With such a parameter, we obtained detailed samples of pollutant concentrations that are closer to the actual personal exposure to air pollution. The monitoring network was specially designed with a high representation of areas with different characteristics.

Meanwhile, we considered potential factors, including road density, traffic volume, meteorology (i.e., temperature, pressure, solar radiation, wind speed and wind direction), and the forms of community on the basis of previous researches. All these predictor variables were measured by mobile platforms and the measuring buffers were from 50 to 300 m surrounding each measurement site [9]. It's difficult for traditional statistical models to capture the complex non-linear interactions among variables. In order to solve this problem and produce more accurate estimations of air pollutant concentrations, this study established decision tree models for modeling two typical air pollutants. The decision tree models were developed moderately to explain the variances of two key air pollutants. The adjusted explained variance of models was 0.85 for O_3 and 0.76 for $PM_{2.5}$. Values of the model R^2 in previous studies mainly concentrated between 0.6 and 0.8, with the lowest value of 0.49 and the highest value of 0.97, among which limited studies were conducted utilizing the regular monitoring network [13,22,37]. As for the overall fit of 10-fold cross-validation, the R^2 was 0.82 for O_3 and 0.72 for $PM_{2.5}$. In addition, when evaluated using an external dataset, the O_3 model performed slightly better than the $PM_{2.5}$ model. Thus, the models developed in this study achieved a high precision and better spatial resolution with a relatively simple procedure. This indicates that using decision tree models for a neighborhood-scale estimation of air pollution is feasible.

Although the two air pollutants adopted the same procedures for establishing decision tree models, the final models contained different variables, reflecting different contributing factors to the two pollutants. Meteorological variables, including solar radiation, relative humidity and air temperature, were found to be the primary influencing factors in the O_3 model [41]. Meanwhile, $PM_{2.5}$ was more heavily affected by the forms of community and traffic emission. Because small buffers of bodies of water or public transportation stops density had a lot of zero values, variables with more than five zero values were removed in the modeling process. Finally, the land use variables that remained in the O_3 models were green space and square within 150 m buffer and building floor within 50 m buffer. But in $PM_{2.5}$ models, there were residential land area within 150 m buffer, the length of road network within 300 m buffer, building floor within 150 m buffer and green space and square within 300 m buffer. We discovered that the land use variables of the same kind ended up in the models with different buffers. It indicates that the influence of land use types on air pollutants varied with different geographic space sizes.

It should be noted that green space area was negatively correlated with the $PM_{2.5}$ variations, which was almost opposite to that of O_3. O_3 is a secondary air pollutant produced by complex reactions between volatile organic compounds (VOCs) and nitrogen oxides (NOx), and the main source of VOCs is vegetation. Hence, green space can contribute to O_3 formation. In contrast, green space can reduce $PM_{2.5}$ concentrations. Additionally, instead of traffic influencing factors such as traffic volume and road length, which are sources of air pollution, the residential land area became the second influencing factor in the $PM_{2.5}$ model. This is likely due to the value of the residential land area that reflects the comprehensive situation of the residential style and road density to a certain extent [7]. The residential area on the concentration distributions of $PM_{2.5}$ would still need more exploration in future research.

To an extent, this study has revealed the spatial variation of $PM_{2.5}$ and O_3 concentrations at a neighborhood scale, although it should be emphasized that the results based on the proposed method are limited to a specific case and it is encouraged that it is applied to other different case. Further research is recommended to input new datasets to verify these models.

5. Conclusions

The decision tree model was introduced in this study to fit the daily pollution values of O_3 and $PM_{2.5}$. The model ranked their influencing factors at a neighborhood scale, such as air pollution background levels, meteorological parameters, traffic emissions, building heights, and land use type. Based on the models, the regression maps of the study area were drawn with an accuracy of 100 m*100 m. Compared to general studies, we have made several innovative contributions. The accuracy, effectiveness and time resolution of modeling spatial variation of air pollutants were improved significantly by 14%–21% by a decision tree model. The main focus of this paper is to identify the significant influencing factors at the neighborhood scale rather than at an urban, regional, or global scale, which is not common in current studies. The main findings of this study include:

(1) Significant daily variations in O_3 and $PM_{2.5}$ indicated the remarkable effects of human activity and meteorological factors on the diurnal trends of air pollutants. The curves of the average daily variations of O_3 concentration are similar, with one peak and one valley. There are often two peaks of $PM_{2.5}$ concentration, suggesting a significant contribution by traffic emissions to daytime $PM_{2.5}$ variations.

(2) Decision tree models were used to describe the evolution of surface O_3 and $PM_{2.5}$ concentrations as well as their relationships with influence factors at a neighborhood scale. In the case of Minhang district, Shanghai, China, the models performed well in simulating both pollutants, with CV R^2 values of 0.82 for O_3 and 0.72 for $PM_{2.5}$, respectively, indicating that spatial variations of air pollutants at a neighborhood scale can be well depicted by the decision tree models.

(3) Six contributing factors, including solar radiation, relative humidity, air temperature, green space within a 150 m buffer, traffic volume within a 300 m buffer, and building floor within a 50 m buffer, can ultimately explain more than 80% of the variance of O_3 concentrations. Compared to the O_3 model, there are more urban-related factors influencing the $PM_{2.5}$ model, including transport land, residential land, building floor, and green space. In general, O_3 concentrations were mainly affected by meteorological factors, while $PM_{2.5}$ was more heavily impacted by background concentration and residential area.

(4) Green space areas were negatively correlated with the $PM_{2.5}$ variations, which was almost opposite compared to that of O_3. O_3 is a secondary air pollutant produced by complex reactions between volatile organic compounds (VOCs) and nitrogen oxides (NOx), and the main sources of VOCs is vegetation. This suggests that appropriate green vegetation density, rather than large-scale increase of green vegetation, may greatly improve the neighborhood scale air quality.

In conclusion, this paper demonstrates the superiority of decision tree models in modeling spatial variations of O_3 and $PM_{2.5}$ concentrations at a neighborhood scale, providing a possible alternative for similar research in the future. These findings are also helpful for deeper understanding of the variation process of air pollutants at neighborhood scale, and may encourage urban planners to consider the impacts of land use design on air quality.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We acknowledge the funding received from the National Natural Science Foundation of China (No. 41701552), the National Planning Office of Philosophy and Social Science, China (No. 16ZDA048) and the Science and Technology Project of Guangzhou, China (No. 201803030032). We are also grateful to Meng-yi Jin, Xin-yi Song, Junjie-Li and Rui-feng Song from the Center for ITS and UAV Applications Research at Shanghai Jiao Tong University for their kind help in data collection. Finally, we would like to appreciate the editor and anonymous reviewers for their insightful comments on our work.

References

[1] J. Lelieveld, J.S. Evans, M. Fnais, D. Giannadaki, A. Pozzer, The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature 525 (7569) (2015) 367–371.

[2] T.M. de Kok, H.A. Driece, J.G. Hogervorst, J.J. Briede, Toxicological assessment of ambient and traffic-related particulate matter: a review of recent studies, Mutat. Res. 613 (2–3) (2006) 103–122.

[3] C.A. Pope, M. Ezzati, D.W. Dockery, Fine-particulate air pollution and life expectancy in the United States, new england, J. Med. 360 (4) (2009) 376–386.

[4] T. Wang, L. Xue, P. Brimblecombe, Y.F. Lam, L. Li, L. Zhang, Ozone pollution in China: a review of concentrations, meteorological influences, chemical precursors, and effects, Sci. Total Environ. 575 (2017) 1582–1596.

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

[6] L. Huang, C. Zhang, J. Bi, Development of land use regression models for PM2.5, SO2, NO2 and O3 in Nanjing, China, Environ. Res. 158 (2017) 542–552.

[7] J.J. Kim, S. Smorodinsky, M. Lipsett, B.C. Singer, A.T. Hodgson, B. Ostro, Traffic-related air pollution near busy roads: the east bay children's respiratory health study, Am. J. Respir. Crit. Care Med. 170 (5) (2004) 520–526.

[8] J. Richmond Bryant, C. Saganich, L. Bukiewicz, R. Kalin, Associations of PM2.5 and black carbon concentrations with traffic, idling, background pollution, and meteorology during school dismissals, Sci. Total Environ. 407 (2009) 3357–3364.

[9] H.A. Olvera, M. Garcia, W. Li, H. Yang, M.A. Amaya, O. Myers, S.W. Burchiel, M. Berwick, N.E. Pingitore Jr., Principal component analysis optimization of a PM2.5 land use regression model with small monitoring network, Sci. Total Environ. 425 (2012) 27–34.

[10] R.C. Abernethy, R.W. Allen, I.G. McKendry, M. Brauer, A land use regression model for ultrafine particles in Vancouver, Canada, Environ. Sci. Technol. 47 (10) (2013) 5217–5225.

[11] Y. Gao, Z. Wang, C. Liu, Z. Peng, Assessing neighborhood air pollution exposure and its relationship with the urban form, Build. Environ. 155 (2019) 15–24.

[12] D. Xu, L. Liu, X. Bao, W. Tian, J. Zhang, Monitoring and analysis of ozone pollution in neighborhood scale, J. Environ. Health 27 (2010) 266–267.

[13] G. Hoek, R. Beelen, K. de Hoogh, D. Vienneau, J. Gulliver, P. Fischer, D. Briggs, A review of land-use regression models to assess spatial variation of outdoor air pollution, Atmos. Environ. 42 (2008) 7561–7578.

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

[15] J.G. Su, M. Brauer, B. Ainslie, D. Steyn, T. Larson, M. Buzzelli, An innovative land use regression model incorporating meteorology for exposure analysis, Sci. Total Environ. 390 (2–3) (2008) 520–529.

[16] J. Richmond Bryant, L. Bukiewicz, R. Kalin, C. Galarraga, F. Mirer, A multi-site analysis of the association between black carbon concentrations and vehicular idling, traffic, background pollution, and meteorology during school dismissals, Sci. Total Environ. 409 (11) (2011) 2085–2093.

[17] Z. Wang, S. Zhong, H. He, Z. Peng, M. Cai, Fine-scale variations in PM2.5 and black carbon concentrations and corresponding influential factors at an urban road intersection, Build. Environ. 141 (2018) 215–225.

[18] K.Y. Kim, Y.S. Kim, Y.M. Roh, C.M. Lee, C.N. Kim, Spatial distribution of particulate matter (PM10 and PM2.5) in Seoul Metropolitan Subway stations, J. Hazard Mater. 154 (1–3) (2008) 440–443.

[19] H. Kamani, M. Hoseini, M. Seyedsalehi, Y. Mahdavi, J. Jaafari, G.H. Safari, Concentration and characterization of airborne particles in Tehran's subway system, Environ. Sci. Pollut. Control Ser. 21 (12) (2014) 7319–7328.

[20] S. Sharma, S. Chatani, R. Mahtta, A. Goel, A. Kumar, Sensitivity analysis of ground level ozone in India using WRF-CMAQ models, Atmos. Environ. 131 (2016) 29–40.

[21] H. Wang, X. Li, D. Wang, J. Zhao, H. He, Z. Peng, Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach, J. Clean. Prod. 253 (2020) 1–12.

[22] C. Li, Z. Wang, B. Li, Z. Peng, Q. Fu, Investigating the relationship between air pollution variation and urban form, Build. Environ. 147 (2019) 559–568.

[23] M. Khafaie, A. Ojha, S. Salvi, C. Yajnik, Methodological approach in air pollution health effects studies, Journal of Air Pollution and health 1 (3) (2016) 219–226.

[24] C. Brokamp, R. Jandarov, M.B. Rao, G. LeMasters, P. Ryan, Exposure assessment models for elemental components of particulate matter in an urban environment: a comparison of regression and random forest approaches, Atmos. Environ. 151 (2017) 1–11, 1994.

[25] D. Ruppert, The elements of statistical learning: data mining, inference, and prediction, J. Am. Stat. Assoc. 99 (466) (2004), 567-567.

[26] W. Lu, D. Wang, Ground-level ozone prediction by support vector machine approach with a cost-sensitive classification scheme, Sci. Total Environ. 395 (2–3) (2008) 109–116.

[27] W. Lu, D. Wang, Learning machines: rationale and application in ground-level ozone prediction, Appl. Soft Comput. 24 (2014) 135–141.

[28] A. Champendal, M. Kanevski, P. Huguenot, Air pollution mapping using nonlinear land use regression models, in: B. Murgante, S. Misra, A.M.A.C. Rocha, C. Torre, J. G. Rocha, M.I. Falcão, D. Taniar, B.O. Apduhan, O. Gervasi (Eds.), Computational Science and its Applications – ICCSA 2014, Springer International Publishing, Cham, 2014, pp. 682–690.

[29] I. Athanasiadis, K. Karatzas, P. Mitkas, Classification techniques for air quality forecasting, fifth ECAI workshop on binding environmental sciences and artificial intelligence, in: 17th European Conference on Artificial Intelligence, 2006, pp. 1–7. August.

[30] P. Pach, J. Abonyi, Association rule and decision tree based methods for fuzzy rule base generation, Proc. World Acad. Sci. Eng. Technol. 13 (2006) 45–50.

[31] K. Sachdeva, M. Hanmandlu, A. Kumar, Real life Applications of fuzzy decision tree, Int. J. Comput. Appl. 42 (2012) 24–28.

[32] K.P. Singh, S. Gupta, P. Rai, Identifying pollution sources and predicting urban air quality using ensemble learning methods, Atmos. Environ. 80 (2013) 426–437.

[33] T.O. Ayodele, Types of Machine Learning Algorithms, InTech2010.

[34] P. Kumar, P. Fennell, D. Langley, R. Britter, Pseudo-simultaneous measurements for the vertical variation of coarse, fine and ultrafine particles in an urban street canyon, Atmos. Environ. 42 (18) (2008) 4304–4319.

[35] Y. Gao, Z. Wang, Q. Lu, C. Liu, Z. Peng, Y. Yu, Prediction of vertical PM2.5 concentrations alongside an elevated expressway by using the neural network hybrid model and generalized additive model, Front. Earth Sci. 11 (2) (2017) 347–360.

[36] J. Rodríguez, A. Pérez, J. Lozano, Sensitivity analysis of k-fold cross validation in prediction error estimation, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 569–575.

[37] X. Meng, L. Chen, J. Cai, B. Zou, C. Wu, Q. Fu, Y. Zhang, Y. Liu, H. Kan, A land use regression model for estimating the NO2 concentration in Shanghai, China, Environ. Res. 137 (2015) 308–315.

[38] K. Miller, D. Siscovick, L. Sheppard, K. Shepherd, J. Sullivan, G. Anderson, J. Kaufman, Long-term exposure to air pollution and incidence of cardiovascular events in women, N. Engl. J. Med. 356 (2007) 447–458.

[39] W. Pan, H. He, Y. Xue, W. Lu, An environmental indicator: particulate characteristics on pedestrian pathway along integrated urban thoroughfare in Metropolis, Stoch. Environ. Res. Risk Assess. 32 (9) (2018) 2527–2536.

[40] W. Lu, D. Wang, Assessing the relative importance of surface ozone influential variables in regional-scale analysis, Atmos. Environ. 43 (22–23) (2009) 3621–3629.

[41] Z. Zhang, X. Zhang, D. Gong, W. Quan, X. Zhao, Z. Ma, S. Kim, Evolution of surface O3 and PM2.5 concentrations and their relationships with meteorological conditions over the last decade in Beijing, Atmos. Environ. 108 (2015) 67–75.

[42] A.O. Langford, J. Brioude, O.R. Cooper, C.J. Senff, R.J. Alvarez, R.M. Hardesty, B. J. Johnson, S.J. Oltmans, Stratospheric influence on surface ozone in the Los Angeles area during late spring and early summer of 2010, J. Geophys. Res.: Atmosphere 117 (D21) (2012) 1–17.

[43] X. Li, D. Wang, Q. Lu, Z. Peng, Q. Fu, X. Hu, J. Huo, G. Xiu, B. Li, C. Li, D. Wang, H. Wang, Three-dimensional analysis of ozone and PM2.5 distributions obtained by observations of tethered balloon and unmanned aerial vehicle in Shanghai, China, Stoch. Environ. Res. Risk Assess. 32 (5) (2018) 1189–1203.

[44] Q. Zhang, H. Li, Potential negative effects of urban green space on the atmospheric environment, Landsc. Des. 2 (2019) 4–11.

[45] L. Mo, X. Yu, Y. Zhao, F. Sun, N. Xia, H. Xia, Correlation analysis between urbanization and particle pollution in Beijing, Ecology and Environmental Sciences 5 (2014) 806–811.

[46] N. Feng, J. Ma, B. Lin, Y. Zhu, Impact of landscape on wind environment in residential area, J. Cent. South Univ. Technol. 16 (s1) (2009) 80–83.