



HHS Public Access

Author manuscript

Atmos Environ (1994). Author manuscript; available in PMC 2021 July 29.

Published in final edited form as:

Atmos Environ (1994). 2020 February 15; 223: . doi:10.1016/j.atmosenv.2020.117267.

Application of land use regression to assess exposure and identify potential sources in PM_{2.5}, BC, NO₂ concentrations

Jing Cai^{a,b,1}, Yihui Ge^{a,1}, Huichu Li^a, Changyuan Yang^a, Cong Liu^a, Xia Meng^a, Weidong Wang^a, Can Niu^c, Lena Kan^d, Tamara Schikowski^e, Beizhan Yan^f, Steven N. Chillrud^f, Haidong Kan^{a,*}, Li Jin^{g,h,**}

^aSchool of Public Health, Key Lab of Public Health Safety of the Ministry of Education and Key Lab of Health Technology Assessment of the Ministry of Health, Fudan University, Shanghai, China

^bShanghai Key Laboratory of Meteorology and Health, Shanghai meteorological service, shanghai, China

^cKey Laboratory of Medicinal Chemistry and Molecular Diagnosis, College of Public Health, Hebei University, Baoding, 071002, China

^dSchool of Public Health, University of California, Berkeley, USA

^eLeibniz Research Institute for Environmental Medicine, Düsseldorf, Germany

^fDivision of Geochemistry, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA

^gState Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai, China

^hCMC Institute of Health Sciences, Taizhou, Jiangsu Province, China

Abstract

*Corresponding author. 130 Dong-An Road, Shanghai, 200032, China. haidongkan@gmail.com (H. Kan). **Corresponding author. 220 Han Dan Road, Shanghai, 200438, China. lijn@fudan.edu.cn (L. Jin).

CRedit authorship contribution statement

Jing Cai: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Yihui Ge:** Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Huichu Li:** Investigation, Writing - review & editing. **Changyuan Yang:** Investigation, Writing - review & editing. **Cong Liu:** Investigation, Writing - review & editing. **Xia Meng:** Investigation, Writing - review & editing. **Weidong Wang:** Investigation, Writing - review & editing. **Can Niu:** Investigation, Writing - review & editing. **Lena Kan:** Writing - review & editing. **Tamara Schikowski:** Writing - review & editing. **Beizhan Yan:** Writing - review & editing. **Steven N. Chillrud:** Funding acquisition, Writing - review & editing. **Haidong Kan:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Supervision, Writing - review & editing. **Li Jin:** Conceptualization, Data curation, Investigation, Supervision, Writing - review & editing.

¹These authors contributed equally to this work.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.atmosenv.2020.117267>.

Background: Understanding spatial variation of air pollution is critical for public health assessments. Land Use Regression (LUR) models have been used increasingly for modeling small-scale spatial variation in air pollution concentrations. However, they have limited application in China due to the lack of spatially resolved data.

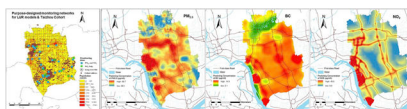
Objective: Based on purpose-designed monitoring networks, this study developed LUR models to predict fine particulate matter (PM_{2.5}), black carbon (BC) and nitrogen dioxide (NO₂) exposure and to identify their potential outdoor-origin sources within an urban/rural region, using Taizhou, China as a case study.

Method: Two one-week integrated samples were collected at 30 PM_{2.5} (BC) sites and 45 NO₂ sites in each two distinct seasons. Samples of 1/3 of the sites were collected simultaneously. Annual adjusted average was calculated and regressed against pre-selected GIS-derived predictor variables in a multivariate regression model.

Results: LUR explained 65% of the spatial variability in PM_{2.5}, 78% in BC and 73% in NO₂. Mean (\pm Standard Deviation) of predicted PM_{2.5}, BC and NO₂ exposure levels were 48.3 (\pm 6.3) $\mu\text{g}/\text{m}^3$, 7.5 (\pm 1.4) $\mu\text{g}/\text{m}^3$ and 27.3 (\pm 8.2) $\mu\text{g}/\text{m}^3$, respectively. Weak spatial correlations (Pearson $r = 0.05$ – 0.25) among three pollutants were observed, indicating the presence of different sources. Regression results showed that PM_{2.5}, BC and NO₂ levels were positively associated with traffic variables. The former two also increased with farm land use; and higher NO₂ levels were associated with larger industrial land use. The three pollutants were correlated with sources at a scale of 5 km and even smaller scales (100–700m) were found for BC and NO₂.

Conclusion: We concluded that based on a purpose-designed monitoring network, LUR model can be applied to predict PM_{2.5}, NO₂ and BC concentrations in urban/rural settings of China. Our findings highlighted important contributors to within-city heterogeneity in outdoor-generated exposure, and indicated traffic, industry and agriculture may significantly contribute to PM_{2.5}, NO₂ and BC concentrations.

Graphical Abstract



Keywords

Land use regression model; Air pollution; Spatial variation; Exposure assessment

1. Introduction

Exposure to air pollution, including ambient fine particulate matter (PM_{2.5}) and nitrogen dioxide (NO₂), has been associated with increasing morbidity and mortality (Chen et al., 2017). Research has further shown black carbon (BC), an important component of PM_{2.5}, typically comprising 5–20% of the PM_{2.5} mass, has been linked to various health end points (Bell et al., 2009; Cowell et al., 2015; Spira-Cohen et al., 2011).

Exposure estimation of air pollution is a critical component of health effects studies (Pope et al., 2002; Zhao et al., 2019). Given that it is not feasible to assess individual-level exposure, exposure estimation for population-based studies can be a large challenge. Early studies mostly used data from fixed monitoring sites as surrogates, but monitoring sites typically have limited geographic coverage and difficulty to capture the intra-urban variability. To address such challenges, recent studies used more sophisticated methods such as land use regression (LUR) model (Eeftens et al., 2012a), kriging and inverse distance weighing (IDW) interpolation (Ramos et al., 2016), dispersion models, chemical transport models (Liu et al., 2010) and satellite-derived data (e.g., aerosol optical depth) (van Donkelaar et al., 2010; Zou et al., 2016; Di et al., 2016). Of these methods, LUR model combines the ability to effectively capture small-scale variability with a relatively low demand for data input (Hoek et al., 2008; Briggs et al., 1997). Also, it has been applied to identify potential sources of air pollution (Niu et al., 2018). Predictor variables included in LUR model provide causative information reflecting pollutant sources.

Given the aforementioned advantages, LUR method has been used to estimate various air pollutants exposure worldwide (Eeftens et al., 2012a; Beelen et al., 2013; Moore et al., 2007; Ross et al., 2007; Sangrador et al., 2008; Lee et al., 2015, 2017). However, the application of LUR model in China is limited to just a few cities where relatively dense routine networks exist (Huang et al., 2017; Liu et al., 2016; Chen et al., 2010). A representative monitoring network is crucial for developing and calibrating LUR models. Monitoring data used for LUR model development were often collected from routine or purpose-designed networks (Eeftens et al., 2012a; Moore et al., 2007; Lee et al., 2017; Tang et al., 2013). Purpose-designed monitoring allows investigators to determine the number and types of monitoring sites (e.g. street level, traffic, background), while routine monitoring is designed for regulatory purpose rather than measuring human exposure (Hoek et al., 2008). In China, the development of LUR model for PM is compromised by the lack of high spatial resolution monitoring data in different time periods, regions and pollutants: until the end of 2012, there were no monitoring data on PM_{2.5} at a national level, and many regions of China still lack monitoring data to enable meaningful modelling of small-scale variability of outdoor air pollution. Additionally, routine monitoring networks in most Chinese cities do not measure components of interest (e. g., soot, metals). Therefore, due to the lack of spatially resolved data on particle component concentrations, long-term effects of PM component have not yet been studied in China.

The Taizhou Air cohort, which was nested in the Taizhou Longitudinal study (Wang et al., 2009), aimed to estimate cardiovascular effects of air pollution. In this study, we developed LUR model using a purpose-designed monitoring network to predict PM_{2.5}, BC and NO₂ for the Taizhou Air cohort study. We further discussed out regression results on suggesting potential sources, model performance on explaining the spatial variance and cross-validation, as well as issues in sampling campaign and the applicability of LUR model in epidemiological studies of air pollution in China.

2. Materials and methods

2.1. Study area

The greater Taizhou region is located in the north bank of Yangtze River in eastern China, and has three municipal districts (Hailing, Gaogang and Jiangyan) and three satellite cities (Taixing, Xinghua, Jingjiang). This region covers a total area of 5787 km², most of which is plains with altitudes ranging from 0 to 8 m. The Taizhou region serves as a prototype of urban-rural mixed area of China as it not only has several high-density downtowns but nearly 69% of the land (3993 km²/5787 km²) is used for agriculture and 0.32% is for industry, indicating various potential emission sources of air pollution.

The Taizhou Longitudinal Study recruited over 0.1 million residents, and 99.7% of them resided in areas of Hailing, Gaogang, Jiangyan, and Taixing (Fig. 1). Correspondingly, we selected these four districts/satellite cities (a total area of 2621 km²) as our study area for purpose-designed monitoring and LUR model development.

2.2. Air pollution measurements

2.2.1. Monitoring site selection—By the end of 2016, there were only four national-standard air quality monitoring stations for PM_{2.5} and NO₂ in the study area and no station measured BC. To compensate for the lack of monitoring data, we performed study-specific sampling in additional sites, including 30 PM_{2.5} (BC) sites and 45 NO₂ sites across the study area. These sampling sites were divided into regional background sites (located in villages near the city), urban background sites (located at the city-center) and local-source sites (e.g. traffic and agricultural emission). The number (proportion) of these 3 types of sites was determined based on the following principles: 1) sampling sites should be broadly distributed proportional to the distribution of cohort participants; 2) these sites should be sufficient to capture the anticipated spatial variation of air pollutants in this area; 3) the selection of local-source sites should reflect the diversity of air pollution potential sources in the study area (ESCAPE Study manual, <http://www.escapeproject.eu/manuals/>). Due to the different types of land use, the local-source sites were categorized into traffic, residential, agriculture and industrial sites. A reference site was chosen at an urban background location where measurements were recorded over an entire year to reflect temporal variability of concentrations.

2.2.2. Monitoring and analysis—The Taizhou region has distinct summer and winter but very short spring and autumn. Therefore, our sampling campaign was conducted in summer (July–September) and winter (November–January). In each season, we collected 2 integrated air pollutant samples, each for 7 consecutive days, in each sampling site. Seven-day integrated measurements obtained at the reference site was operated year-round from July 2015 to August 2016.

Ambient PM_{2.5} samples were collected using 37-mm Teflon filters (225–8303, SKC Inc., PA, USA) with Harvard Impactor (Air Diagnostics and Engineering, Inc., Harrison, ME, USA) and sampling pump (Legacy, SKC Inc., PA, USA). To avoid overloading, we used an intermittent sampling strategy with 15 min of active sampling (flow rate: 10L/min) in every

2 h. All filters were pre-conditioned at a constant air temperature of $20\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$ and constant relative humidity (RH) of 50% ($\pm 5\%$) for 48 h prior to weighing. $\text{PM}_{2.5}$ filters were weighed before and after sampling, and the mass of $\text{PM}_{2.5}$ in each sample was calculated as the difference in filter weights. BC mass were then measured from the Teflon filter using the multi-wavelength optical measurements via optical equipment including a balanced deuterium tungsten halogen light source (DH-2000-BAL), an integrating sphere (ISP-50-8-R), a labmade filter holder, and an Ocean Optics USB4000-VIS-NIR fiber optic spectrometer (Yan et al., 2011).

NO_2 samples were collected using the Ogawa passive sampler (Ogawa & Co. USA Inc., Pompano Beach, FL). Nitrite content in the exposed filters was then extracted and quantified in spectrometer under 545 nm using the Saltzman method. All sample collection and processing were conducted according to the manufacturer's instructions (Ogawa., 1998).

Due to a limited number of samplers, 10 $\text{PM}_{2.5}$ (BC) sampling sites and 15 NO_2 sites were measured simultaneously each time. Given that air pollutants have a substantial temporal variation, adjustment for this variability is essential (Eeftens et al., 2012b). For each site a seasonal average concentration was calculated by averaging two 7-day measurements and then correcting for temporal variation using measurements obtained from the reference site (Cyrus et al., 2012). The difference between the concentration for a specific two-week sampling period and the annual average at the reference site was subtracted from each measurement. The adjusted annual average concentration of each sampling site was then calculated by averaging two seasonal average concentrations.

2.2.3. Quality control—Field blanks and duplicate samples for $\text{PM}_{2.5}$ (BC) and NO_2 were collected in each 7-day sampling period with the same procedure of corresponding normal sample collection, except that field blank samples were collected without an active pump for PM and were not exposed to the air for NO_2 when in the field. The limit of detection (LOD) was calculated as 3 times the standard deviation (SD) of measured field blanks. The precision of our measurements was estimated by the coefficient of variation (CV) from the duplicate samples. For data quality control during the sample collection, we excluded $\text{PM}_{2.5}$ (BC) samples if (1) the sampling pump achieved less than 75% of the programmed schedule; or (2) the flow rate failed to maintain at 10.0 L/min (± 0.5 L/min) from the start to the end of sampling. Additional QC measures were implemented throughout our BC analysis. We ran calibration filters and field blank samples for every 10 samples and one sample was measured twice to ensure the consistency.

2.2.4. Air pollution data from air quality monitoring stations—We also collected hourly concentrations data (between July 2015 and August 2016) of $\text{PM}_{2.5}$ and NO_2 from four national-standard air quality monitoring stations which were operated by Taizhou Environmental Monitoring Center (TEMC). $\text{PM}_{2.5}$ concentrations were measured using the Tapered Element Oscillating Microbalance (TEOM) and NO_2 using the chemiluminescence method. All ambient measurements were operated under the China National Quality Control ((HJ/T 193–2005) and (GB3095–2012)).

To compare modeling prediction with routine monitoring data, we calculated daily average of TEMC PM_{2.5} and NO₂ derived from hourly data, and calculated annual averages derived from daily data to match with model prediction. Following the criteria of the Ambient Air Quality Standards of China, daily and annual average calculation require at least 75% valid data coverage (e.g. at least 18 valid hourly data points available for daily average calculation).

2.3. GIS predictor variables

As presented in Table 1, we included variables of local land use, road networks, population counts, emission inventory and distance to the Yangtze River for each sampling site in our geographic predictor data-base. Circular buffers around the sampling sites were generated with varying radii. The GIS layers were then intersected with the circular buffers and the corresponding values of each layer within buffers at each site was calculated. Calculations were performed using ArcGIS 10.3 (ESRI, Redlands, CA).

Population counts, land use and emission inventory were area-integrated variables. Ambient population density (persons/km²) was calculated according to the sixth census data of Taizhou. Land use data were obtained from the Landsat TM5 dataset of 2015 with a resolution of 30 m. Six categories of land use in the study area were included, namely urban area, industrial area, rural area, farm land area, green space and water area. Annual (2012) NO_x and PM_{2.5} gridded emission data (1 km × 1 km resolution) of major sources (including agriculture, industry, resident, power plants, and traffic) were derived from Multi-resolution Emission Inventory for China (MEIC) (<http://www.meicmodel.org/dataset-meic.html>).

Road networks and distance to the Yangtze River were distance-based variables. Given that traffic density data was not available, we obtained the digital road network data with a resolution of 1:50,000 from National Geomatics Center of China (<http://ngcc.sbsm.gov.cn>). We divided road networks into four classes according to the roughly estimated traffic density and permission of heavy trucks. Highways and arterial roads were classified as the first-class roads. County-standard roads, village-standard roads and other types of roads were, respectively, ranked as the second-, the third- and the fourth-class roads. The length of the four types of roads within each buffer and distance to the nearest road of each class were calculated for each sampling site. Additionally, the distance to the nearest road of each type was calculated. We also measured the straight-line distance between a sampling site to the nearest coastline of the Yangtze River in the land use map of Taizhou.

2.4. Model construction and diagnosis

Stepwise-multiple linear regression method was applied to construct LUR models for PM_{2.5}, BC and NO₂. Detailed information has been described elsewhere (Eeftens et al., 2012a). In brief, all predictor variables were included as candidate independent variables. The adjusted annual average of PM_{2.5}, BC and NO₂ were used as dependent variables. Thereafter, model construction started by including predictor variables with the highest adjusted explained variance (adjusted R²) in univariate regressions. The remaining candidate variables were selected into the model if they met the following criteria: 1) the adjusted R² of the model increased by at least 1% to that of the previous model; 2) the coefficient of the variables in

the model were significant (with two-sided p value <0.05), and were in the right direction of their expected effect.

For model diagnosis, the Cook's Distance was calculated to detect the influential data points for model validity. Observations with Cook's Distance greater than 1 would be excluded and then the model would be re-evaluated. Variance Inflation Factor (VIF) was used to further identify the multi-collinearity between predictor variables. Variables with VIFs larger than 3 would be removed and the model would be re-evaluated. Leave-one-out-cross-validation (LOOCV) was then employed to evaluate the overall model performance. R^2 and root mean squared error (RMSE) between model prediction and observations for all sampling sites were calculated to assess the model fit. We also inspected the residual distribution of model fit. All data analysis was performed using R software (Version 2.15.3).

2.5. Mapping

The predicted annual average concentration surfaces of pollutants were created according to the final LUR models. We then divided the study area into 10708 500 m × 500 m grids and applied the LUR models to derive predicted pollutants concentration at the centroid of each grid as the average level of each grid. Where the regression equations produced negative estimates (−2% for all 3 parameters), grid cell values were set to zero; where estimates exceeded a quantification limit (QL) defined as 150% of the maximum observed concentrations (1.5% for all 3 parameters), grid cell values were set to QL (Henderson et al., 2007; Amini et al., 2014). At last, Kriging interpolation pollution maps were drawn based on the predicted values.

3. Results

3.1. Measurements and predictors

We finally selected 2 regional background sites, 2 urban background sites, 11 traffic sites, 11 urban residential sites, 14 rural residential sites, 3 agricultural sites, 1 industrial site and 1 background reference site. Except for 1 rural residential site that had to be taken down due to an unexpected construction occurred during the sampling period, we successfully collected samples from 29 $PM_{2.5}$ (BC) sites and 44 NO_2 sites. As shown in Fig. 1, the distribution of sampling sites effectively covered the residential area of cohort participants. We eventually collected 175 qualified samples for $PM_{2.5}$ (BC) and 234 for NO_2 . LODs were 9.0 μg for $PM_{2.5}$ and 36.2 μg for NO_2 . No BC was detected in the field blank samples, so we used an LOD of 1.4 ng/mm^2 according to a previous study (Yan et al., 2011). The average CV in duplicate samples were 7.0% for $PM_{2.5}$, 6.5% for BC and 6.9% for NO_2 , indicating good reproducibility.

Table 2 presents the statistic description of averaged annual concentrations after adjusting for temporal variability. The mean (\pm SD) concentrations of $PM_{2.5}$ were 52.9 (\pm 6.1) $\mu g/m^3$, which was 5 times above the annual average in the World Health Organization (WHO) air quality guideline (10 $\mu g/m^3$), while the annual mean NO_2 level [mean (\pm SD) = 25.9 (\pm 5.6) $\mu g/m^3$] was below the WHO guideline (40 $\mu g/m^3$). The mean (\pm SD) of BC was 8.3 (\pm 1.2) $\mu g/m^3$ and accounted for 15.6% of $PM_{2.5}$ in average. There were no significant differences

between urbanized area and rural areas ($p > 0.1$) for $PM_{2.5}$, BC and NO_2 levels. Comparable annual levels during the study period were observed at four air quality monitoring stations, the range of which was 51–59 $\mu\text{g}/\text{m}^3$ for $PM_{2.5}$ and 28–34 $\mu\text{g}/\text{m}^3$ for NO_2 .

Table 2 also shows that our sampling sites dispersed in the study area with significant diversity in the land use area, traffic and population density. For example, population counts within the 500m buffers in our sampling sites ranged from 284 to 21,018. The average length (\pm SD) of the first-class road and the third-class road within 3000m were 35,459 (\pm 30,254) m and 102,592 (\pm 87,390) m, respectively. Agricultural land use within 3000m buffer varied from 548,199 m^2 to 2,506,809 m^2 .

3.2. Land use regression models

Table 3 presents the predictor variables that were finally included into the LUR models for $PM_{2.5}$, BC and NO_2 , respectively. Overall, the VIF values differed between variables but were all less than 3, which suggested a relatively low collinearity between predictor variables. Analysis of Cook's distance identified measurements from 1 $PM+NO_2$ site as influential data points for their corresponding model construction. Measurements from this site were extremely high which could be affected by point source and this site was thus eventually discarded. The final models were constructed based on 27 and 42 sites for $PM_{2.5}$ (BC) and NO_2 , respectively.

For $PM_{2.5}$, a substantial fraction (65%) of the measured spatial variability was explained by 4 GIS predictor variables: length of the first-class road (3000m buffer), agricultural land area (2000m buffer), water area (5000m buffer), and distance to the nearest forth-class road. We found an association between the larger length of the first-class road and agricultural land area with higher $PM_{2.5}$ concentrations; while the associations between water area and distance to the nearest forth-class road with $PM_{2.5}$ concentrations were negative. Contribution values (regression coefficient $\beta \times \text{IQR}$) of these 4 variables suggest that traffic and agricultural emission contributed most to $PM_{2.5}$ concentrations in our study area. For example, annual $PM_{2.5}$ increased by 3.7 $\mu\text{g}/\text{m}^3$ for each IQR (34,186m) increase in length of the first-class road (3000m buffer).

Model adjusted R^2 was higher for BC (78%) than for $PM_{2.5}$. Five predictors entered the BC model, including length of the first-class road (500m buffer), length of the third-class road (100m buffer), water area (4000m buffer), industrial area (4000m buffer), and agricultural land area (2000m buffer). Except the water area, all other predictors were positively associated with BC values. Similar to $PM_{2.5}$, traffic and agricultural emission were two primary influencing factors on BC. For example, each IQR increase in length of the first-class road within 500 m buffer (1596.22m) and agricultural land area within 2000 m buffer (5.69 km^2) corresponded to 0.52 $\mu\text{g}/\text{m}^3$ and 0.90 $\mu\text{g}/\text{m}^3$ increase in BC concentrations, respectively.

For NO_2 , the local traffic variables coupled with variables related to industrial emission with 5000 m buffer yielded 73% of explained variability in NO_2 concentrations, including length of the first-class road (700 m buffer), length of the third-class road (100 m buffer), distance to the nearest first-class road and emission of NO_x from traffic (5 km buffer). The result that

80% (4/5) of the selected variables were traffic-related, suggests substantial contribution of traffic emission to ambient NO₂ concentrations. Specifically, each IQR increase (2772 m) in length of the first-class road within a 700-m buffer was associated with 2.46 µg/m³ increase in NO₂.

3.3. Model evaluations

Residual distribution of the final LUR models showed no apparent bias in model fitting. LOOCV R² were 0.56, 0.70 and 0.66 for PM_{2.5}, BC and NO₂, respectively (Table 3). The differences between the model R² and LOOCV R² less than 10% indicated good robustness of our models (Beelen et al., 2013). Cross-validation RMSE were 3.12 µg/m³ for PM_{2.5}, 0.51 µg/m³ for BC and 3.02 µg/m³ for NO₂, suggesting the predicted values coincided well with the measured values. The even distribution of the predicted value and measured value along the 1:1 line in Fig. 2 further supported a considerable agreement between the measurements and prediction.

3.4. Mapping

Fig. 3a, b and c show maps predicting annual concentrations of PM_{2.5}, BC and NO₂, respectively. The predicted mean (±SD) was 48.3 (±6.3) µg/m³ for PM_{2.5}, 7.5 (±1.4) µg/m³ for BC and 27.3 (±8.2) µg/m³ for NO₂. There were several high concentration hotspots across the city, suggesting large within-city gradients found in these three pollutants. For PM_{2.5}, the middle-western area with more intensive traffic networks and larger coverage of agriculture land use had higher concentrations than the eastern and northern areas. Mapping also revealed clearly elevated NO₂ and BC concentrations along the main road and fell sharply as distance to the road network increased, indicating traffic is an influential source for BC and NO₂ in the study area. In addition, BC and NO₂ increased in areas corresponding to the agriculture land use, and the three peak spots along the Yangtze River reflected emissions from three local industrial zones. Note that considering the limited number of sampling sites, our monitoring network mainly covered the areas where cohort distributed only. Therefore, it should be in caution when interpreting the prediction where spatial variation of air pollution may not be captured by our sampling network, because bias might occur.

Of note, as shown in Fig. 3, distinct spatial distribution among three air pollutants were observed. All three pollutants were across 10708 500 m×500 m grids positively correlated; however, their spatial correlations were weak. The Pearson r was 0.25 (p < 0.001) between PM_{2.5} and BC, 0.05 (p < 0.001) between PM_{2.5} and NO₂, and 0.15 (p < 0.001) between BC and NO₂.

4. Discussion

Although LUR models have been used widely for air pollution exposure assessment worldwide (Eeftens et al., 2012a; Beelen et al., 2013; Moore et al., 2007; Lee et al., 2017; Tang et al., 2013; Amini et al., 2014; Wu et al., 2015a; Meng et al., 2015; Clougherty et al., 2013), this modeling approach still has had limited application in China due to the lack of spatially resolved data on air pollution. Based on a purpose-designed monitoring network,

we developed land use regression models in Taizhou area where an air pollution cohort study was located. Our models could explain 65%, 78% and 73% of the variability in PM_{2.5}, BC and NO₂ in the study area. The concentration surface retrieved from the models demonstrated large within-city heterogeneity in PM_{2.5}, BC and NO₂ exposure, highlighting that ignoring the spatial variation in these air pollutants may result in great exposure misclassification. Our study also suggested traffic, industry and agriculture as the major sources of PM_{2.5}, NO₂ and BC concentrations, which may have implications for developing air pollution abatement strategies to protect public health.

Many prior epidemiological studies used data from routine monitoring stations as surrogates of exposure, which often ignored the intra-urban variability of air pollutants due to the lack of enough geographic coverage (Ryan and LeMasters, 2007). In Taizhou, the spatial variable coefficient (CV) across four TEMC air quality monitoring stations (6% for PM_{2.5} and 7% for NO₂), were much lower than prediction of LUR models (11% for PM_{2.5} and 35% for NO₂). This finding confirmed that compared to limited number of monitoring stations, LUR model has stronger capacity to capture the spatial variability of air pollutants. In addition, our LUR models developed explained substantial spatial variation of air pollution for Taizhou areas. The percentage explained variation (R²) of our LUR models were comparable to the models in previous studies (Eeftens et al., 2012a; Beelen et al., 2013; Lee et al., 2015; Clougherty et al., 2013). The adjusted explained variance of BC (78%) and NO₂ (73%) models in this study was higher than that of New York, USA (65% for BC and 67% for NO₂) (Clougherty et al., 2013) and comparable to that of Catalunya, Spain (75% for BC and 71% for NO₂), but lower than that of Munich/Augsburg, Germany (91% for BC and 86% for NO₂) (Eeftens et al., 2012a; Beelen et al., 2013). Such differences could be explained by the original variability in the measured concentrations, quality of the predictor variables, and the geographic and socioeconomic characteristics of the study area (Hoek et al., 2008).

Interestingly, our mapping results showed weak spatial correlations among three air pollutants (Pearson $r = 0.09-0.23$); while high spatial correlations were reported in some previous studies (Clougherty et al., 2013; Li et al., 2018). For example, Clougherty et al. (2013) constructed LUR models to predict wintertime street-level air pollution concentrations of New York City and found strong correlations among three pollutants (Pearson $r > 0.70$) (Clougherty et al., 2013). The inconsistency may be due to that these studies mainly focused on urban settings with relatively single sources (traffic and/or heating). Consistent with our findings, poor spatial correlations also have been observed in some other Chinese cities, such as Shanghai (Liu et al., 2016) and Nanjing (Huang et al., 2017), indicating these pollutants may have significantly different sources.

Our model predictors provided further explanation for the observed weak spatial correlations. Consistent with previous studies, traffic-related variables were the most common predictors for three pollutants, suggesting traffic may be an importance local source of air pollution. Relatively high ratio of BC over PM_{2.5} (16%) confirmed the significant contribution of traffic. Prior LUR models mainly included road length, traffic intensity and traffic load (Eeftens et al., 2012a; Beelen et al., 2013; Moore et al., 2007; Clougherty et al., 2013). However, we used length of roads and the distance to the nearest road as proxies due to unavailability of traffic flow/density data in our study. These variables have also been

applied in other studies to reflect traffic density and generated successful prediction (Liu et al., 2016; Meng et al., 2015; Lee et al., 2014). Besides traffic, water and agricultural land use remained in our PM_{2.5} and BC models and industrial land use remained in BC and NO₂ models. Similar results found in Taipei and Shanghai studies, which reflects dilution effect of rivers/water on PM_{2.5} and BC pollution. Farm land use, however, was seldom selected as a significant predictor (Wu et al., 2015a). One possible explanation is that most previous studies were conducted in urban areas with little agriculture; while our study area is mixed with urban and rural areas covering almost 70% of farm land use. The positive effect of farm land use suggested that agricultural activities, such as straw burning (Wu et al., 2015a, 2018), may contribute greatly to ambient PM_{2.5} and BC. In general, the predicting variables in our models indicate necessity in controlling traffic, industrial and agricultural emissions, and protecting local water spaces so as to reduce air pollution in our study area.

In our final LUR models, more variables with larger buffer (especially for >1 km) were found in PM_{2.5} and NO₂ models, whereas BC model was relatively more sensitive to variables with smaller buffers. Similar conclusions were also drawn by Henderson et al. (2007) and Wu et al. (2015) as BC and NO maps showed more pronounced small scale spatial contrast compared to PM_{2.5} and NO₂ maps (Henderson et al., 2007; Wu et al., 2015a). This is most likely due to the difference of spatial heterogeneity between pollutants. NO₂ is a typical secondary pollutant; while both primary and secondary emission can significantly contribute to PM_{2.5} concentrations (Dai et al., 2018; Lonati et al., 2008). NO₂ and PM_{2.5} showed gradual decay and less ability to reflect small scale variability of traffic emissions as much as primary pollutants (Karner et al., 2010). In contrast, BC as a well-known primary pollutant was characterized by an important decreasing trend with increasing distance to sources. Karner et al. (2010) reported a rapid drop of BC concentration within 150 m to traffic emissions. Variables with larger buffer size for PM_{2.5} and NO₂ models may suggest the significant contribution from secondary sources in this region.

Our site selection followed many of the best practices from Europe and North America outlined for LUR modelling (ESCAPE Study manual, <http://www.escapeproject.eu/manuals/>). However, we also found that site selection and sampling in locations such as Taizhou was likely to be more challenging. Firstly, the Taizhou area has more diversified emission sources. In addition to common urban sources (such as traffic), industry and agricultural activities were also important contributors in our study area. Specifically, straw burning prevailed in Taizhou, where 69.7% of the land use were farm land and 67.5% of the population lived in rural area. Wu et al. (2015, 2018) confirmed straw burning accounted for 8.0% of pollutant sources of PM_{2.5} (Wu et al., 2015b, 2018). For rural-urban mixed areas, agricultural sites should be included in the monitoring network. Secondly, our PM_{2.5} mean concentrations were 1.8–29.3 times higher than levels in European cities (Eeftens et al., 2012a). The ESCAPE protocol proposed a sampling period of 2 continuous weeks in each season, but such a sampling period may cause overloading under high background pollution levels, especially for our winter sampling campaign. Consequently, we adjusted our sampling period to two one-week intervals, although it resulted in a higher workload for fieldwork. Alternatively, a lower duty cycle on the pump could help to fix this issue. Overall, European and North American cities are likely to have lower pollution and population densities, as well as fewer small-scale dispersed pollution sources than Chinese cities.

Therefore, their strategy of site selection and sampling may not be fully transferable to other study areas. Additionally, it is important to notice that self-designed sampling campaigns could be money- and time-consuming and it often can only afford short-term measurements (e.g. days to weeks) (Eeftens et al., 2012a; Beelen et al., 2013). Like many other studies, non-contemporaneous measurements of sites were conducted may cause that temporal variability of air pollution could hardly be addressed. To overcome this limitation, recent studies found combining LUR models with satellite-derived data (e.g., aerosol optical depth) has becoming an effective method (van Donkelaar et al., 2010; Di et al., 2016).

A few previous studies have attempted to apply LUR models in Chinese cities with limited number of routine monitoring sites (Liu et al., 2016; Wu et al., 2015a). However, it has been documented that compared to the purpose-designed monitoring network, the routine monitoring network was likely to be of low reliability due to its biased estimate of exposure (Hoek et al., 2008). We are one of the few that developed LUR models for exposure assessment of air pollution based on specific-monitoring networks in China. Additionally, one significant benefit from the purpose-designed monitoring network is that it allowed us to measure black carbon, which was not monitored in the routine monitoring network, with finer spatial resolution. Lastly, our models explained a major portion of annual average of PM_{2.5}, NO₂ and BC variability, and the LOOCV analysis implied stable performance. Hence, this study may add experience on LUR model development for air pollution studies in China.

However, our study has limitations. Firstly, 22%–35% of the variability remains unexplained by our models, and there are some other influential factors to be explored in future studies. For example, we had no traffic data for trucks or cars. Secondly, the number of sampling sites reported in previous studies with comparable model performance ranged from 14 to 155 (Eeftens et al., 2012a; Lee et al., 2015; Clougherty et al., 2013). Although there are no strict rules for a minimum required number of sites, 40–80 monitoring sites for LUR models were recommended by Hoek et al. (2008) (Hoek et al., 2008). However, due to limited budgets and sampling devices, we only have 27/42 valid monitoring sites for PM/NO₂ model development. Finally, like many earlier studies (Beelen et al., 2013; Liu et al., 2016; Meng et al., 2015), our LUR models treat relationships between air pollutants and predictor variables linear, which may not be adequate for all variables. That may be partly the reason why the validation R² of PM_{2.5}, BC and NO₂ were only 0.56, 0.70 and 0.66 in this study. Besides, LUR model can only predict ambient concentrations, while personal activities and microenvironment characteristics can introduce additional variations in exposure. For example, people may spend most of their time indoors during winter, thus exposure misclassification can still occur. Further studies may be needed to improve these issues.

5. Conclusions

This study applied LUR models in a greater metropolitan area of China based on a purpose-designed monitoring network. The LUR models developed from this study captured substantial spatial variation of PM_{2.5}, BC and NO₂ in the great Taizhou area, and suggested that traffic, industrial and agricultural activities were the main sources for local air pollution

in that area. Our results may provide evidence for air pollution regulation, and help improve air pollution exposure assessment for the Taizhou Air Cohort.

Acknowledgement

This work was supported by the Ministry of Science and Technology of the People's Republic of China (2016YFC0206202), National Natural Science Foundation of China (91543114, 81502774, 91643205 and 51808178), China Medical Board Collaborating Program (16-250), National Institute of Environmental Health Sciences, United States (P30-09089) and the Shanghai Key Laboratory of Meteorology and Health, China (QXJK201703).

This study does not involve experimental animals.

Abbreviations

BC	black carbon
CV	coefficient of variation
ESCAPE	Study The European Study of Cohorts for Air Pollution Effects study
IDW	inverse distance weighing
IQR	interquartile range
LOD	limit of detection
LOOCV	Leave-one-out-cross-validation
LUR	land use regression
MEIC	Multi-resolution Emission Inventory for China
NO	nitrogen monoxide
NO₂	nitrogen dioxide
NO_x	nitrogen oxide
PM_{2.5}	fine particulate matter, particulate matter less than 2.5 μm in diameter
QL	quantification limit
RMSE	root mean squared error
VIF	Variance Inflation Factor

References

- Amini H, et al., 2014. 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, 343–353. [PubMed: 24836390]
- Beelen R, et al., 2013. Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe - the ESCAPE project. *Atmos. Environ* 72, 10–23.
- Bell ML, et al., 2009. Hospital admissions and chemical composition of fine particle air pollution. *Am. J. Respir. Crit. Care Med* 179 (12), 1115–1120. [PubMed: 19299499]

- Briggs DJ, et al., 1997. Mapping urban air pollution using GIS: a regression-based approach. *Int. J. Geogr. Inf. Sci* 11 (7), 699–718.
- Chen L, et al., 2010. A land use regression for predicting NO₂ and PM₁₀ concentrations in different seasons in Tianjin region, China. *J. Environ. Sci* 22 (9), 1364–1373.
- Chen RJ, et al., 2017. Fine particulate air pollution and daily mortality A nationwide analysis in 272 Chinese cities. *Am. J. Respir. Crit. Care Med* 196 (1), 73–81. [PubMed: 28248546]
- Clougherty JE, et al., 2013. Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: the New York City Community Air Survey (NYCCAS). *J. Expo. Sci. Environ. Epidemiol* 23, 232. [PubMed: 23361442]
- Cowell WJ, et al., 2015. Associations between prenatal exposure to black carbon and memory domains in urban children: modification by sex and prenatal stress. *PLoS One* 10 (11), e0142492. [PubMed: 26544967]
- Cyrys J, et al., 2012. Variation of NO₂ and NO_x concentrations between and within 36 European study areas: results from the ESCAPE study. *Atmos. Environ* 62, 374–390.
- Dai QL, et al., 2018. Chemical nature of PM_{2.5} and PM₁₀ in Xi'an, China: insights into primary emissions and secondary particle formation. *Environ. Pollut* 240, 155–166. [PubMed: 29734076]
- Di Q, et al., 2016. Assessing PM_{2.5} exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol* 50 (9), 4712–4721. [PubMed: 27023334]
- Eeftens M, et al., 2012a. Development of land use regression models for PM_{2.5}, PM_{2.5} absorbance, PM₁₀ and PM_{coarse} in 20 European study areas; results of the ESCAPE project. *Environ. Sci. Technol* 46 (20), 11195–11205. [PubMed: 22963366]
- Eeftens M, et al., 2012b. Spatial variation of PM_{2.5}, PM₁₀, PM_{2.5} absorbance and PM_{coarse} concentrations between and within 20 European study areas and the relationship with NO₂ - results of the ESCAPE project. *Atmos. Environ* 62, 303–317.
- Henderson SB, et al., 2007. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ. Sci. Technol* 41 (7), 2422–2428. [PubMed: 17438795]
- Hoek G, et al., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ* 42 (33), 7561–7578.
- Huang L, Zhang C, Bi J, 2017. Development of land use regression models for PM_{2.5}, SO₂, NO₂ and O₃ in Nanjing, China. *Environ. Res* 158, 542–552. [PubMed: 28715783]
- Karner AA, Eisinger DS, Niemeier DA, 2010. Near-roadway air quality: synthesizing the findings from real-world data. *Environ. Sci. Technol* 44 (14), 5334–5344. [PubMed: 20560612]
- Lee JH, et al., 2014. Land use regression models for estimating individual NO_x and NO₂ exposures in a metropolis with a high density of traffic roads and population. *Sci. Total Environ* 472, 1163–1171. [PubMed: 24377679]
- Lee JH, et al., 2015. LUR models for particulate matters in the Taipei metropolis with high densities of roads and strong activities of industry, commerce and construction. *Sci. Total Environ* 514, 178–184. [PubMed: 25659316]
- Lee M, et al., 2017. Land use regression modelling of air pollution in high density high rise cities: a case study in Hong Kong. *Sci. Total Environ* 592, 306–315. [PubMed: 28319717]
- Li H, et al., 2018. Short-term exposure to fine particulate air pollution and genome-wide DNA methylation: a randomized, double-blind, crossover trial. *Environ. Int* 120, 130–136. [PubMed: 30081103]
- Liu XH, et al., 2010. Understanding of regional air pollution over China using CMAQ, part I performance evaluation and seasonal variation. *Atmos. Environ* 44 (20), 2415–2426.
- Liu C, et al., 2016. A land use regression application into assessing spatial variation of intra-urban fine particulate matter (PM_{2.5}) and nitrogen dioxide (NO₂) concentrations in City of Shanghai, China. *Sci. Total Environ* 565, 607–615. [PubMed: 27203521]
- Lonati G, Giugliano M, Ozgen S, 2008. Primary and secondary components of PM_{2.5} in Milan (Italy). *Environ. Int* 34 (5), 665–670. [PubMed: 18207243]
- Meng X, et al., 2015. A land use regression model for estimating the NO₂ concentration in shanghai, China. *Environ. Res* 137, 308–315. [PubMed: 25601733]

- Moore DK, et al., 2007. A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA. *J. Environ. Monit* 9 (3), 246–252. [PubMed: 17344950]
- Niu Y, et al., 2018. Fine particulate matter constituents and stress hormones in the hypothalamus-pituitary-adrenal axis. *Environ. Int* 119, 186–192. [PubMed: 29960262]
- Ogawa NO, 1998. NO₂, NO_x and SO₂ Sampling Protocol Using the Ogawa Sampler [cited 2018].
- Pope CA 3rd, et al., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc* 287 (9), 1132–1141.
- Ramos Y, et al., 2016. Spatio-temporal models to estimate daily concentrations of fine particulate matter in Montreal: kriging with external drift and inverse distance-weighted approaches. *J. Expo. Sci. Environ. Epidemiol* 26 (4), 405–414. [PubMed: 26648248]
- Ross Z, et al., 2007. A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmos. Environ* 41 (11), 2255–2269.
- Ryan PH, LeMasters GK, 2007. A review of land-use regression for characterizing intraurban air models pollution exposure. *Inhal. Toxicol* 19, 127–133. [PubMed: 17886060]
- Sangrador JLT, et al., 2008. A land use regression model for predicting PM_{2.5} in Mexico city. *Epidemiology* 19 (6). S259–S259.
- Spira-Cohen A, et al., 2011. Personal exposures to traffic-related air pollution and acute respiratory health among Bronx schoolchildren with asthma. *Environ. Health Perspect* 119 (4), 559–565. [PubMed: 21216722]
- Tang R, Blangiardo M, Gulliver J, 2013. Using building heights and street configuration to enhance intraurban PM₁₀, NO_x, and NO₂ land use regression models. *Environ. Sci. Technol* 47 (20), 11643–11650. [PubMed: 24001269]
- van Donkelaar A, et al., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ. Health Perspect* 118 (6), 847–855. [PubMed: 20519161]
- Wang X, et al., 2009. Rationales, design and recruitment of the Taizhou longitudinal study. *BMC Public Health* 9, 223. [PubMed: 19589173]
- Wu JS, et al., 2015a. Applying land use regression model to estimate spatial variation of PM_{2.5} in Beijing, China. *Environ. Sci. Pollut. Control Ser* 22 (9), 7045–7061.
- Wu YC, Chen Wang, Lei Chen, Ying Zhu, Xi-xi, 2015b. Variation characteristics of the air quality in Taizhou during winter. *Environmental Monitoring and Forewarning* 7 (6), 47–50.
- Wu YC, Ying Chen, Jun Wang, 2018. Lei Characteristics of air Quality in Taizhou during the Perid of the summer Harvest and autumn harvest. *The Administration and Technique of Environmental Monitoring* 30 (1), 31–35.
- Yan BZ, et al., 2011. Validating a nondestructive optical method for apportioning colored particulate matter into black carbon and additional components. *Atmos. Environ* 45 (39), 7478–7486.
- Zhao Y, et al., 2019. Particulate air pollution exposure and plasma vitamin D levels in pregnant women: a longitudinal cohort study. *J. Clin. Endocrinol. Metab* 104 (8), 3320–3326. [PubMed: 30896756]
- Zou B, et al., 2016. High-resolution satellite mapping of fine particulates based on geographically weighted regression. *IEEE Geosci. Remote Sens. Lett* 13 (4), 495–499.

HIGHLIGHTS

- Lack of spatially resolved air pollution data limits LUR model application in China.
- We are one of the few building LUR models upon specific-monitoring network in China.
- PM_{2.5}, BC and NO₂ models explain a large fraction of concentration variability.
- We add experience on air pollution exposure assessment for population-based studies.

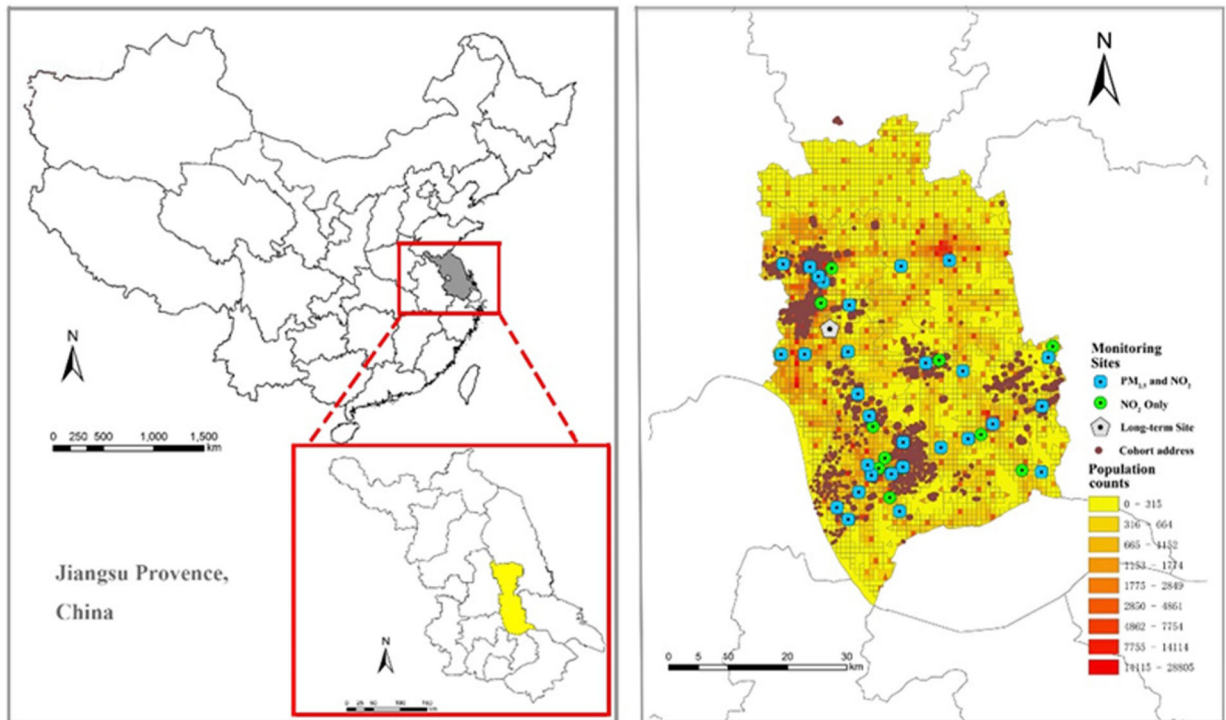


Fig. 1. Locations of the study area.

The left side map shows the location of Taizhou and the right side map shows the distribution of monitoring sites and Taizhou Cohort.

Abbreviations: PM 2.5, particulate matter with aerodynamic diameter $2.5 \mu\text{m}$ (fine particulate matter); NO₂, nitrogen dioxide.

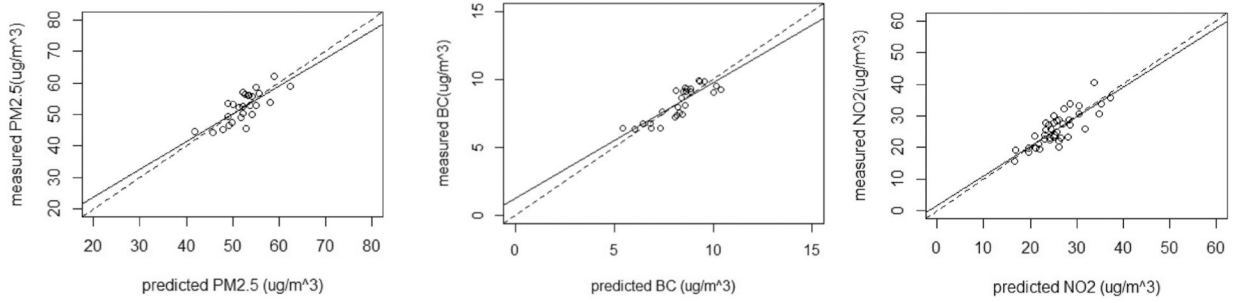


Fig. 2. Results of leave-one-out cross validation for $\text{PM}_{2.5}$, BC and NO_2 LUR models: predicted concentrations (x-axis) against measured concentrations (y-axis) for $\text{PM}_{2.5}$, BC and NO_2 models.

Abbreviations: $\text{PM}_{2.5}$, particulate matter with aerodynamic diameter $2.5 \mu\text{m}$ (fine particulate matter); BC, black carbon; NO_2 , nitrogen dioxide.

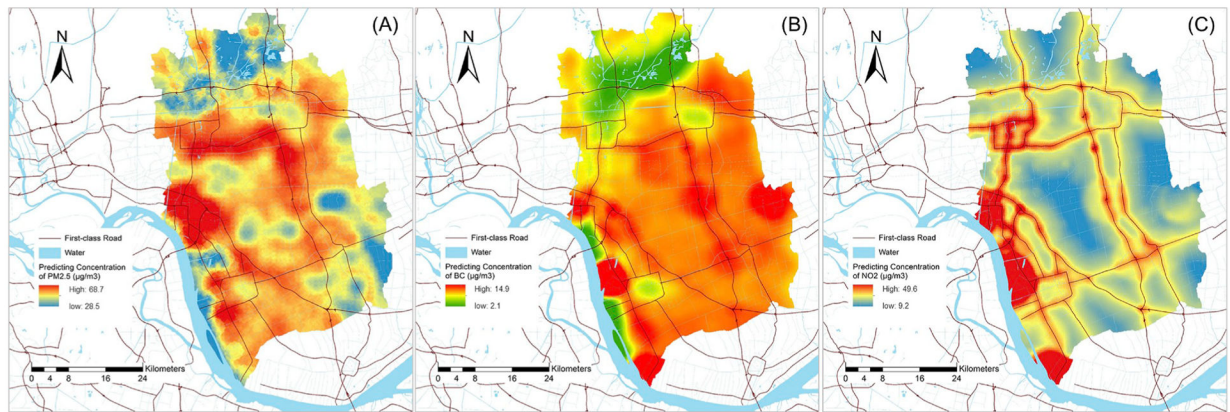


Fig. 3. LUR model prediction surfaces:

Estimated annual PM_{2.5} (A), BC (B) and NO₂ (C) concentration from the final land use regression models.

Abbreviations: PM_{2.5}, particulate matter with aerodynamic diameter 2.5 µm (fine particulate matter); BC, black carbon; NO₂, nitrogen dioxide

Table 1

Potential variables with units, defined buffer sizes and priori defined directions of the effect.

GIS dataset	Predictor variable	Unit	Buffer size (radius in meters)	Prior direction
Population	Population count	Person	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+
Traffic ^a	Length of the first-class roads in a buffer	m	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+
	Length of the second-class roads in a buffer	m	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+
	Length of the third-class roads in a buffer	m	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+
	Length of the forth-class roads in a buffer	m	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+
	Distance to the nearest first-class roads	m	NA	-
	Distance to the nearest second-class roads	m	NA	-
	Distance to the nearest third-class roads	m	NA	-
	Distance to the nearest forth-class roads	m	NA	-
Land use	Urban area	m ²	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+
	Industrial area	m ²	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+
	Farmland area	m ²	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+
	Rural residential area	m ²	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+
	Green space area	m ²	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	-
	Water area	m ²	100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	-
Coastline	Distance to the nearest coastline	m	NA	-
Emissions	Emissions from agriculture	t	1000, 5000, 10000	+
	Emissions from industry	t	1000, 5000, 10000	+
	Emissions from resident	t	1000, 5000, 10000	+
	Emissions from power plants	t	1000, 5000, 10000	+
	Emissions from transportation	t	1000, 5000, 10000	+

^aThe first-class roads consist of highways, national roads and provincial roads. The second-class roads are county roads and the third-class roads are village roads. The forth-class roads include other lower-class roads.

Table 2

Statistic description of measured air pollutants and potential predictor variables^a

Variable	Mean	SD	Minimum	P25	Median	P75	Maximum
PM _{2.5} (µg/m ³) (n = 28)	52.9	6.1	42.6	49.0	53.0	56.2	68.8
BC (µg/m ³) (n = 28)	8.3	1.2	6.4	7.3	8.6	9.3	9.9
NO ₂ (µg/m ³) (n = 43)	25.9	5.6	15.5	22.4	24.9	28.7	40.3
Population count							
500 m	4241	–	284	836	1442	5528	21018
3000 m	96527	–	10041	26923	47105	104400	416947
Length of first-class road (m)							
500 m	1048	1982	0	0	0	1596	8047
3000 m	35459	30254	0	13449	26391	47657	127752
Length of third-class road (m)							
500 m	5266	3934	0	1786	5390	7712	14498
3000 m	102592	87390	4135	28552	77710	161996	289990
Distance to the nearest first-class road (m)	1357	1567	12	308	895	1667	7447
Distance to the nearest third-class road (m)	150	197	0.5	55	73	172	1000
Farmland area (m ²)							
500 m	263177	244757	0	6168	242740	423513	785191
3000 m	15658107	7171735	548198	11562163	19093172	20817572	25068093
Emission of NO _x from traffic (t)							
5000 m	57	75	0	4	16	113	249
10000 m	133	118	3	35	98	167	402

^aThe descriptive statistics of potential predictor variables were based on 43 monitoring sites. Abbreviations: PM_{2.5}, particulate matter with aerodynamic diameter 2.5 µm; NO₂, nitrogen dioxide; BC, black carbon; SD, standard deviation. P25: the 25th percentile; P75: the 75th percentile.

Table 3

Description of developed LUR models for PM_{2.5}, BC and NO₂

Model	Variable	β^f	SE ^b	IQR × β^f	p-value	VIF ^d	Global statistics
PM _{2.5} (µg/m ³)	Intercept	46.04	1.70		<0.001	NA	Adjusted R ² = 0.65
	Length of first-class road (3000m buffer), m ²	1.09 × 10 ⁻⁴	2.03 × 10 ⁻⁵	3.74	<0.001	1.280	RMSE = 3.12 µg/m ³
	Farmland area (2000m buffer), m ²	7.68 × 10 ⁻⁷	2.04 × 10 ⁻⁸	4.37	0.001	1.163	LOOCV R ² = 0.56
	Water area (5000m buffer), m ²	-1.04 × 10 ⁻⁶	2.01 × 10 ⁻⁷	0.38	<0.001	1.731	
	Distance to the nearest forth-class road, m	-7.19 × 10 ⁻³	2.94 × 10 ⁻³	-1.72	0.02	1.321	
BC (µg/m ³)	Intercept	6.09	0.30		<0.001	NA	Adjusted R ² = 0.78
	Length of first-class road (500m buffer), m	2.96 × 10 ⁻⁴	6.09 × 10 ⁻⁵	0.52	<0.001	1.189	RMSE = 0.51 µg/m ³
	Length of third-class road (100m buffer), m	1.08 × 10 ⁻³	3.83 × 10 ⁻⁴	-0.38	0.010	1.126	LOOCV R ² = 0.70
	Water area (4000m buffer), m ²	-4.40 × 10 ⁻⁷	6.52 × 10 ⁻⁸	-0.14	<0.001	1.200	
	Industrial area (4000m buffer), m ²	4.44 × 10 ⁻⁶	9.83 × 10 ⁻⁷	0.17	<0.001	1.130	
NO ₂ (µg/m ³)	Farmland area (2000m buffer), m ²	2.17 × 10 ⁻⁷	3.49 × 10 ⁻⁸	0.90	<0.001	1.114	
	Intercept	22.39	0.94		<0.001	NA	Adjusted R ² = 0.73
	Length of first-class road (700m buffer), m	8.78 × 10 ⁻⁴	1.48 × 10 ⁻⁴	2.46	<0.001	1.287	RMSE = 3.02 µg/m ³
	Length of third-class road (100m buffer), m	4.67 × 10 ⁻³	1.38 × 10 ⁻³	1.77	0.002	1.115	LOOCV R ² = 0.66
	Distance to the nearest first-class road, m	-1.25 × 10 ⁻³	2.75 × 10 ⁻⁴	-1.67	<0.001	1.381	
Industrial area (5000m buffer), m ²		1.25 × 10 ⁻⁵	2.54 × 10 ⁻⁶	1.96	<0.001	1.296	
	Emission of NO _x from traffic (5000m buffer), t	1.60 × 10 ⁻²	6.21 × 10 ⁻³	1.90	0.014	1.164	

Abbreviations PM_{2.5}, particulate matter with aerodynamic diameter 2.5 µm; NO₂, nitrogen dioxide; BC, black carbon.

^a β^f is the regression coefficient of each predictor variable.

^b SE is the abbreviation of standard error estimate of each predictor variable.

^c IQR is the abbreviation of the interquartile range of each predictor variable, the contribution of each variable to predicted concentrations ($\beta \times$ inter-quartile range (IQR)).

^d VIF is the abbreviation of Variance Inflation Factor.