Science of the Total Environment 634 (2018) 77–86

Contents lists available at ScienceDirect

Science of the Total Environment

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

Development of land-use regression models for fine particles and black carbon in peri-urban South India

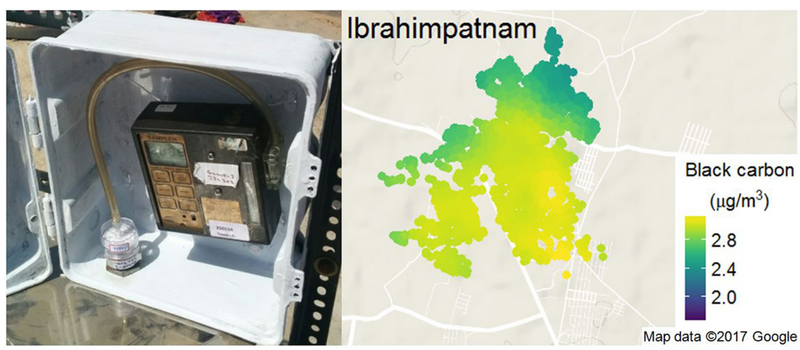
Margaux Sanchez [a,b,c,*], Albert Ambros [a,b,c], Carles Milà [a,b,c], Maëlle Salmon [a,b,c], Kalpana Balakrishnan [d], Sankar Sambandam [d], V. Sreekanth [e], Julian D. Marshall [e], Cathryn Tonne [a,b,c]

[a] *Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain*
[b] *Universitat Pompeu Fabra (UPF), Barcelona, Spain*
[c] *CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain*
[d] *Department of Environmental Health Engineering, Sri Ramachandra University (SRU), Chennai, India*
[e] *Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, United States*

HIGHLIGHTS

- We developed LUR models for $PM_{2.5}$ and black carbon in peri-urban South India.
- We derived predictors from local built-environment survey and satellite imagery.
- $PM_{2.5}$ and black carbon models reached 58% and 79% of explained variability.
- Data from local built-environment survey were relevant for black carbon model.
- We observed more spatial variability than typical values in other study areas.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 7 December 2017
Received in revised form 21 March 2018
Accepted 24 March 2018
Available online xxxx

Editor: P. Kassomenos

Keywords:
Black carbon
Exposure assessment
India
Land-use regression
Particulate matter
Peri-urban area

ABSTRACT

Land-use regression (LUR) has been used to model local spatial variability of particulate matter in cities of high-income countries. Performance of LUR models is unknown in less urbanized areas of low-/middle-income countries (LMICs) experiencing complex sources of ambient air pollution and which typically have limited land use data. To address these concerns, we developed LUR models using satellite imagery (e.g., vegetation, urbanicity) and manually-collected data from a comprehensive built-environment survey (e.g., roads, industries, non-residential places) for a peri-urban area outside Hyderabad, India. As part of the CHAI (Cardiovascular Health effects of Air pollution in Telangana, India) project, concentrations of fine particulate matter ($PM_{2.5}$) and black carbon were measured over two seasons at 23 sites. Annual mean (sd) was 34.1 (3.2) $\mu g/m^3$ for $PM_{2.5}$ and 2.7 (0.5) $\mu g/m^3$ for black carbon. The LUR model for annual black carbon explained 78% of total variance and included both local-scale (energy supply places) and regional-scale (roads) predictors. Explained variance was 58% for annual $PM_{2.5}$ and the included predictors were only regional (urbanicity, vegetation). During leave-one-out cross-validation and cross-holdout validation, only the black carbon model showed consistent performance. The LUR model for black carbon explained a substantial proportion of the spatial variability that could not be captured by simpler interpolation technique (ordinary kriging). This is the first study to develop a LUR model for ambient concentrations of $PM_{2.5}$ and black carbon in a non-urban area of LMICs, supporting the applicability of the LUR approach in such settings. Our results provide insights on the added value of manually-collected built-environment data to improve the performance of LUR models in settings with limited data availability. For

* Corresponding author at: Barcelona Institute for Global Health (ISGlobal), Doctor Aiguader, 88, 08003 Barcelona, Spain.
E-mail address: margaux.sanchez@isglobal.org. (M. Sanchez).

https://doi.org/10.1016/j.scitotenv.2018.03.308

both pollutants, LUR models predicted substantial within-village variability, an important feature for future epidemiological studies.

1. Introduction

Air pollution is a leading risk factor for mortality and morbidity worldwide. The World Health Organization estimated that 3 million deaths were attributable to ambient air pollution in 2012; 87% of these occurred in low- and middle-income countries (LMICs) (World Health Organization, 2016). Still, high-income countries (HICs) remain the focus of much of the current literature investigating the health effects of air pollution (Tonne et al., 2017). Epidemiological evidence based on populations in HICs may not apply to LMIC populations because of differences in exposure ranges, air pollution sources, age distributions, and baseline health status. Thus, there is a critical need for population-level estimates of long-term exposure that can be used for epidemiological purposes in LMIC settings (Ma et al., 2017), where the majority of air pollution and pollution-related health effects occur.

Land-use regression (LUR) is a modeling method widely used in epidemiological studies to estimate particulate matter and black carbon concentrations at fine spatial scale in urban areas of North America and Europe (Eeftens et al., 2012, 2016; Hoek et al., 2008; Montagne et al., 2015; van Nunen et al., 2017; Weichenthal et al., 2016; Wolf et al., 2017; Zhang et al., 2014). There have been some attempts to apply the methodology in China and India but this literature is scarce and is limited to urban areas (Huang et al., 2017; Saraswat et al., 2013; Wu et al., 2015, 2017). LUR models have shown reasonable performance in urban areas where much of the locally-emitted particulate matter is generated by road traffic (Karagulian et al., 2015), but conditions in peri-urban or rural areas in LMICs, such as India, may differ (e.g. higher contribution of domestic fuel use and open burning) (Paliwal et al., 2016). Previous studies have demonstrated the limited transferability of LUR prediction models to areas other than those in which they were developed (Patton et al., 2015; Wang et al., 2014). In Bangalore, India, increasing concentrations of $PM_{2.5}$ were observed with increasing proximity to roads in a middle-income neighborhood, a spatial pattern similar to what would be observed in HIC (Both et al., 2011). However, this was not the pattern observed in a low-income neighborhood, which was attributed by the authors to solid fuel use (Both et al., 2011), highlighting the complexity of spatial patterns of air pollution in LMICs. It is unknown how well LUR prediction models perform in less urbanized area of LMICs that showed different contribution of particulate matter sources. Challenges to the development of LUR models in these settings include limited availability of geographic information systems (GIS) data, sources of particulate matter emissions not well correlated with existing land use data, and lack of routine monitoring data.

To address the need for air pollution exposure assessment in LMICs, we developed LUR models to predict spatial variation of $PM_{2.5}$ and black carbon in a peri-urban area of South India in which local emission sources include household solid fuel use, local industries, and motor vehicles. We here advance the science of LUR modeling by coupling land-use data derived from satellite imagery and data derived from a built-environment survey, a data collection approach not previously employed in LUR, allowing us to overcome the limited access to GIS data often found in LMICs. To help future exposure assessment, we evaluated the added value of LUR modeling predictions as compared to ordinary kriging, an interpolation technique that does not require any additional geographic data. We contribute to the exposure assessment literature by studying a peri-urban environment in a LMIC for which there is scarce literature regarding the spatial variability of ambient air pollution and no literature on LUR as an approach to predict such variability.

2. Methods

2.1. Study area

The study area consists of 28 rural and peri-urban villages in the Southeast of Hyderabad, covering 8.2 km^2 in a 22 km × 35 km (i.e., 770 km^2) region (Fig. 1). The area between villages was not of interest as it includes mostly crops and agricultural lands with no or few inhabitants. All villages were included in the existing APCAPS (Andhra Pradesh Children and Parents' Study) cohort and CHAI (Cardiovascular Health effects of Air pollution in Telangana, India) project (Kinra et al., 2014; Tonne et al., 2017). Villages vary in terms of surface, population size, socioeconomic status, urbanization, and primary cooking fuel.

2.2. Sampling campaign ($PM_{2.5}$ and black carbon)

We identified 23 households located in 16 different villages as fixed sites for sampling. We selected sites to maximize contrasts in several variables expected to correlate with particulate matter: distance to primary roads, distance to the city of Hyderabad, distance to industry, 500-meter buffer household density, and village-level solid fuel use (Fig. 1). This was done to avoid extrapolation of the LUR models to values of predictors for which we had no measurements. The households located in the 12 villages that were not directly monitored were similar to the households located in the other 16 villages in terms of distance to primary roads, distance to the city of Hyderabad, household density in a 50-meter buffer, and village-level proportion of solid fuel use.

We measured 24-hour integrated gravimetric $PM_{2.5}$ concentrations at these 23 locations for a total of 21 days in two sessions: 11 nonconsecutive days during post-monsoon season (Sep–Oct 2015) and 10 nonconsecutive days during summer season (Mar–Apr 2016). Sampling was done every other day. All sites were sampled the same days. Monitors included pumps (model 224-PCMTX8, SKC Ltd., Dorset, UK) that drew air through a cyclone separator (cut point: 2.5 μm) attached to a cassette containing 37-mm filter (Emfab filter, Pallflex®). Monitors were placed on the households' rooftops. Filters were left on site and collected daily at noon. Filters were weighed pre- and post-exposure using the TAPHE (Tamil Nadu Air Pollution and Health Effects) study protocol which follows RTI (Research Triangle Institute) guidelines (Balakrishnan et al., 2015). Daily $PM_{2.5}$ concentrations were derived from filter mass after correction for mass accumulated on blank filters (season-specific correction using median blank weight). Of the 483 sampled filters (21 days at 23 sites), 13 experienced device malfunction (running time < 75% of the expected sampling duration or pump airflow < 20% of the expected value) and 5 showed unexplained weighing errors for $PM_{2.5}$ (post-weight < pre-weight). We imputed all missing values (4% of the data) using a linear combination of date and $PM_{2.5}$ measurements from an ambient background monitoring site (see below). The overall performance of the imputation model was fair (adjusted-R^2 = 0.61) and predicted and measured values correlated positively ($R_{Spearman}$ = 0.74).

Daily black carbon concentrations were derived from optical attenuation (880 nm) of the mass collected on the filters, using a Magee OT21 Sootscan Optical Transmissometer (Magee Scientific, Berkeley, California, USA). Negative concentrations were obtained for 41 filters (8% of the sample). We assumed these were due to concentrations below the lower end of our standard curve and we imputed them with the seasonal 5th percentile concentration: 0.21 μg/m^3 in summer season and 1.31 μg/m^3 in post-monsoon season. Days with device malfunction or

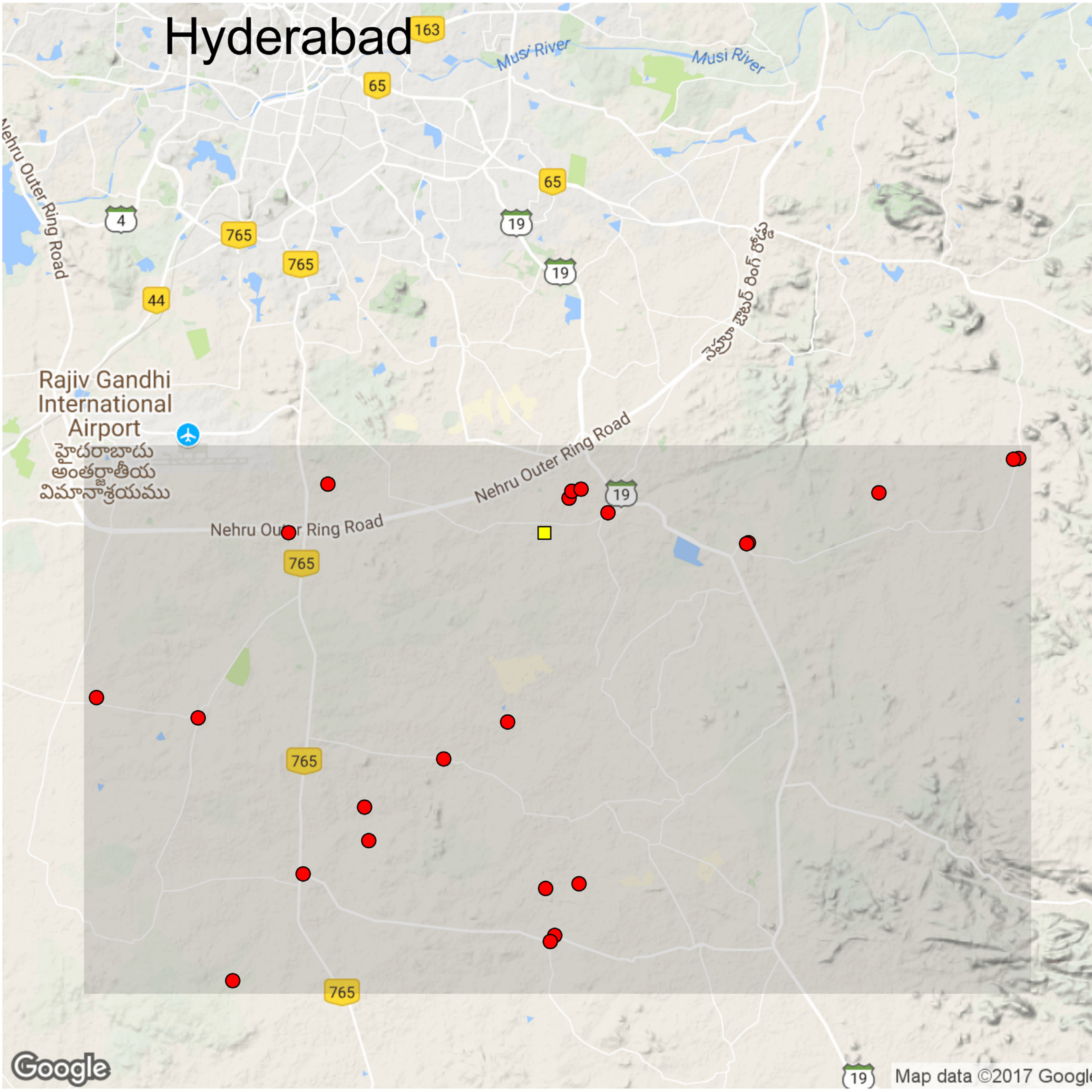


Fig. 1. The location of the study region (22 km × 35 km rectangle), the 23 households included in air pollution sampling (points), and the ambient background $PM_{2.5}$ monitor (square).

lost filters (n = 15) remained missing. Black carbon concentrations were thus available for 468 samples (97% of the attempted measurements).

We calculated the annual and seasonal concentrations of $PM_{2.5}$ and black carbon for each site by averaging daily concentrations. As we aimed to predict spatial (not temporal) variability and all sites were sampled the same days, no temporal adjustment was performed.

2.3. Background $PM_{2.5}$ monitoring

For the purpose of the CHAI project, continuous monitoring of $PM_{2.5}$ was implemented from 2015 on at one site in the North of the study area, far from traffic or other sources, thus being a good indication of background concentration (Fig. 1). Hourly concentrations were derived by beta-radiation attenuation using an e-BAM device (Model 9800, Met One, Grants Pass, OR). We calculated daily averages from noon to noon to correspond with the 24-h period of the gravimetric sampling. Concentrations were available for 19 days (of the expected 21 days) because of device malfunction for two days in summer season. We imputed the missing days using measurements from another background monitor located 13 km away using a linear combination of daily $PM_{2.5}$, relative humidity, and temperature. The overall performance of the imputation model was good (adjusted-R^2 = 0.87) and predicted and measured values correlated strongly ($R_{Spearman}$ = 0.96).

2.4. Geographic covariates

We selected 325 geographic variables with potential to predict spatial variability in particulate matter in our study area and calculated the values of each variable at all sampling sites. The selected variables related to different features (e.g., industries, households, roads, non-residential places, vegetation) and included indicators of proximity (distance to the nearest feature) as well as number or density of the feature within a buffer of a given radius. A total of eight buffer sizes were chosen to represent local variability (50 m to 300 m) as well as regional variability (up to 5 km) in $PM_{2.5}$ and black carbon (Supplementary Table S1).

2.4.1. Built-environment survey

As part of the built-environment survey performed for APCAPS in 2013, the following locations in the study area were manually geocoded with GPS devices: non-residential places inside villages (e.g., bus station, shops, temples) (Supplementary Table S1), industrial places with regular operation (according to industry type), and the main entrance of the Rajiv Gandhi International airport. Roads throughout the study area were manually mapped by tracing satellite imagery (importing data into the OpenStreetMap, www.openstreetmap.org) and confirmed through visits by the field team. Using APCAPS questionnaire data from 2012, we derived village-level indicators of liquefied petroleum gas use, solid fuel use, and car-, motorcycle-, and bicycle-ownership.

2.4.2. Satellite data

The night-time light intensity (NTLI) in 2012 was used as a village-level marker of economic activity and urbanicity. NTLI ranges from 0 (no light) to 63 and is derived from measurements of light emissions from persistent sources associated with human settlement provided by the Defense Meteorological Satellite Program satellites (Baugh et al., 2010; Elvidge et al., 1997). To assess the level of vegetation, we derived the Normalized Difference Vegetation Index (NDVI) and percentage of tree cover from the Landsat satellite data at 30-meter spatial resolution (U.S. Geological Survey, 2017). We calculated the NDVI for the "greenest" sampling season (post-monsoon, one image from October 2015), the least green sampling season (summer, one image from April 2015) and their average. The percentage of tree cover was derived from a sole composite image taken in 2015.

2.4.3. Pre-processing

We excluded 131 geographic covariates that were not sufficiently variable i.e., ≥75% of the sample had the same value. A total of 194 variables were considered in the present analysis (Supplementary Table S1).

2.5. Land-use regression

We developed LUR models to predict annual and seasonal (post-monsoon and summer) $PM_{2.5}$ and black carbon concentrations using the available geographic variables. We adapted the process standardized within the framework of the ESCAPE project (Eeftens et al., 2012). The average concentration per site was used as the dependent variable. A single linear regression was fit for each of the geographic predictors. The predictor with the highest adjusted-R^2 (explained variance) and with expected direction of effect was retained (Supplementary Table S1). The remaining geographic predictors were evaluated in turn in the model and the predictor with the highest increase in adjusted-R^2 was added if: a) it showed expected direction of effect, b) it did not change the direction of effect of previous variables, c) it was not highly correlated with previous variables ($R_{Spearman} < 0.70$), and d) the coefficients' variance were not affected by multi-colinearity (variance inflation factor < 4). The process was repeated until there were no variables which fit the criteria or the number of variables reached 6 (to limit the risk of over fitting) or the adjusted-R^2 increased by <1%. Variables with p-values larger than 0.1 were sequentially removed to obtain the final model i.e., all variables of the model have p-values ≤ 0.1.

Final LUR models were checked for normality of the residuals, spatial autocorrelation of the residuals (semi-variograms), heteroscedasticity of the residuals, and influential observations (Cook's D > 1). We checked whether and how model parameters changed when excluding influential observation(s). If removal did not cause large changes (10%) in coefficients, the model was retained. If removal caused large changes in coefficients, the model was rerun using all but the influential observation(s).

As one sensitivity analysis, we applied the LUR selection process without restricting the number of predictors to six. All other criteria were unchanged. As another sensitivity analysis, we applied the LUR selection process using solely the satellite-derived predictor variables (thus restricting the list to 36 predictors) to assess the added value of the built-environment survey predictors.

2.6. Ordinary kriging

Ordinary kriging uses the degree of spatial dependence between sampled sites to interpolate pollutant concentrations in unsampled locations over the gridded study area (Tyagi and Singh, 2013). Ordinary kriging requires no additional geographic data as inputs (as opposed to LUR) which makes it potentially attractive in settings with limited availability of GIS data. Ordinary kriging is specific to the spatial autocorrelation found in the study area, generally depicted by the semi-variogram (plotting the average squared difference between sampled locations according to distance between locations). In the present analysis, the semi-variogram for annual $PM_{2.5}$ could be fit with a Gaussian model, while the semi-variogram for annual black carbon could be fit with a spherical model. No ordinary kriging was performed for seasonal concentrations because no fit could be found. The spatial resolution of the gridded study area was fixed to 271 m.

2.7. Model evaluation

We compared the accuracy of modeling approach (ordinary kriging and LUR) on the basis of correlation ($R_{Spearman}$) and error (root-mean-squared error, RMSE) between predicted and observed values.

We performed a leave-one-out cross-validation to assess prediction performance of each modeling approach: each model was fitted on N-1 sites and the result was used to predict concentrations at the left-out site. This procedure was repeated N times. We reported correlation and RMSE between predicted and observed concentrations. For the LUR models, we also reported the overall level of fit (adjusted-R^2).

We additionally performed a cross-holdout validation to evaluate the LUR models (Wang et al., 2016). The cross-holdout validation is an extensive process that has been shown to more accurately reflect out-of-sample predictive ability of LUR models than classical leave-one-out cross-validation (Wang et al., 2016), in particular with small datasets. The cross-holdout validation successively built N "evaluation" models (with variable selection and so on) based on all N-1 site combinations. The procedure thus built 24 evaluation models for each $PM_{2.5}$ average (annual and seasonal) and 23 evaluation models for each black carbon average (annual and seasonal) with all different predictors and estimates. Pollutant concentrations were then predicted at the left-out sites. We reported correlation and RMSE between predicted and observed concentrations as well as adjusted-R^2 of and selected predictors in the evaluation models.

2.8. Model application

The final models (LUR and ordinary kriging) were used to predict $PM_{2.5}$ and black carbon annual concentrations at every household within the study area (n = 23,605 households located in 28 villages).

Table 1
Annual and seasonal measured concentrations of $PM_{2.5}$ and black carbon.

Pollutant	Time period	N sites	Measured concentrations ($\mu g/m^3$)[a]	IQR	Sd/mean
$PM_{2.5}$	Annual	24	34.11 (3.21) [25.68; 40.66]	3.35	0.09
	Post-monsoon	24	29.70 (3.73) [21.91; 36.83]	4.64	0.13
	Summer	24	38.96 (4.07) [29.83; 48.65]	5.34	0.10
Black carbon	Annual	23	2.73 (0.48) [1.85; 3.65]	0.70	0.17
	Post-monsoon	23	2.90 (0.69) [1.74; 4.56]	0.80	0.24
	Summer	23	2.56 (0.40) [1.98; 3.66]	0.54	0.15

Abbreviations: IQR: inter-quartile range; sd: standard deviation.
[a] Mean (standard deviation) [min; max].

3. Results

3.1. Description

The $PM_{2.5}$ annual mean (sd) across the 24 sites was 34.11 (3.21) $\mu g/m^3$ ranging from 25.68 to 40.66 [inter-quartile range: 3.35] (Table 1). The black carbon annual mean (sd) across the 23 sites was 2.73 (0.48) $\mu g/m^3$ ranging from 1.85 to 3.65 [0.70]. $PM_{2.5}$ concentrations were higher in summer season than in post-monsoon season, while the opposite pattern was observed for black carbon (Table 1). Black carbon had greater spatial contrast than $PM_{2.5}$; coefficients of variation (sd/mean) for annual concentrations were 17% and 9%, respectively. The correlation between the annual concentrations of $PM_{2.5}$ and black carbon was strong ($R_{Spearman} = 0.65$).

3.2. LUR models

Most of the developed LUR models included 6 variables (the limit we fixed) of which 2–4 were non-significant at 10% level and then removed. The final LUR models for annual and seasonal $PM_{2.5}$ and black carbon are presented in Table 2. The predictor explaining the most variance in annual $PM_{2.5}$ was tree coverage within 1000 m. All final LUR models for $PM_{2.5}$ included the village level of urbanicity (NTLI) and an indicator of green space (NDVI or tree coverage). Longitude and elevation were strongly negatively correlated in our study area ($R_{Spearman} = -0.84$) and appeared in annual and summer models for $PM_{2.5}$. The adjusted-R^2 for annual $PM_{2.5}$ was 0.58. The seasonal models for $PM_{2.5}$ showed less explained variance than the annual model. Overall, models for black carbon performed better than $PM_{2.5}$ models, with adjusted-R^2 ranging from 0.63 (post-monsoon black carbon) to 0.78 (annual black carbon). In all black carbon models, the most contributing predictor was primary or ring road indicator within 5000 m (Table 2). Tree coverage within 500–1000 m appeared to predict annual and post-monsoon black carbon concentrations, while annual and summer models both included indicators of energy supply places. No restriction on the number of included predictors led to the same final models except for summer $PM_{2.5}$, for which seven predictors were included instead of three (Supplementary Table S2). Excluding the predictors derived from the built-environment survey led to a general decrease in the variance explained in black carbon models, from 78% to 59% in annual concentrations, while no difference was observed for $PM_{2.5}$ (Supplementary Table S3). Diagnostic plots confirmed that our main models complied with underlying assumptions (Supplementary Figs. S1 and S2). Variograms of model residuals showed no patterns of semi-variance with distance, indicating no spatial autocorrelation in model residuals.

Prediction RMSE were about 50% of the sampled standard deviation for black carbon but higher (>60%) for $PM_{2.5}$ (Table 2). For both pollutants, almost all predictions fall within one standard deviation of the measured concentrations. Predicted concentrations for black carbon correlated strongly with measured concentrations in the annual model ($R_{Spearman} = 0.85$) and the seasonal models (0.69 for post-monsoon and 0.81 for summer). For $PM_{2.5}$, correlations between predicted and measured concentrations were lower (<0.70). The

Table 2
Final land-use regression models for $PM_{2.5}$ and black carbon.

Pollutant	Time period	Predictors included in the final land-use regression models	Variance explained (adjusted-R^2)	RMSE	Error *Mean (sd)* *[min;max]*	$R_{Spearman}$
$PM_{2.5}$	Annual	tree_1000 + NTLI + longitude + ndvimax_5000	0.58	1.86	0.00 (1.89) [−3.59;2.77]	0.66
	Post-monsoon	NTLI + hh_50 + ndviavg_5000	0.48	2.29	0.00 (2.34) [−6.21;4.11]	0.70
	Summer	elevation + tree_300 + NTLI	0.34	3.02	0.00 (3.08) [−6.96;3.72]	0.57
Black carbon	Annual	rrlen_5000 + tree_500 + distnrp_e1	0.78	0.20	0.00 (0.21) [−0.41;0.43]	0.85
	Post-monsoon	prlen_5000 + tree_1000	0.63	0.39	0.00 (0.40) [−1.05;0.69]	0.69
	Summer	rrden_5000 + nrp_e2_300 + nrp_p1_5000 + prlen_1000	0.65	0.21	0.00 (0.21) [−0.58;0.22]	0.81

Abbreviations: RMSE, root-mean-square error; sd, standard deviation. $R_{Spearman}$ represents correlation coefficients between measured and predicted concentrations. Variable names: tree coverage (tree_X), village-level night-time light intensity (NTLI), post-monsoon value for the Normalized Difference in Vegetation Index (ndvimax_X), annual average value for the Normalized Difference in Vegetation Index (ndviavg__), household density (hh_X), ring road length (rrlen_X), primary roads length (prlen_X), ring road density (rrden_X), distance to non-residential places related to wood/gas supply (distnrp_e1), and number of non-residential places related to petrol supply (nrp_e2_X) and to religious center (nrp_p1_X) with _X indicating the radius of the buffer.

correlation between the predicted annual concentrations of $PM_{2.5}$ and black carbon was strong ($R_{Spearman} = 0.61$) and similar to the correlation found between measurements.

3.3. LUR models evaluation

During leave-one-out cross-validation, black carbon models showed better robustness than $PM_{2.5}$ models (Table 3). Most black carbon predictions stayed in range of one standard deviation of measured concentration and predictions remained strongly correlated with measurements. For $PM_{2.5}$, correlations between predictions and measurements decreased (e.g., from 0.66 to 0.49 for annual model). During cross-holdout validation, adjusted-R^2 of the evaluation models built on N-1 sites were similar to or slightly larger than the adjusted-R^2 of the full-sites models (Table 3). The most-selected predictors in the evaluation models were consistent with those selected in the full-sites models, though variables in $PM_{2.5}$ models showed more diversity than in black carbon models. Influential observations were detected in 3 of the 24 evaluation models built for annual $PM_{2.5}$, while none were detected for annual black carbon. The predictive ability of the black carbon evaluation models was moderate, with overall correlations ranging from 0.23 to 0.59 between measurements and predictions (Table 3). In contrast, cross-holdout validation for $PM_{2.5}$ showed much increased RMSE (up to 4 times the full-sites model) and null correlations between measured and predicted values.

3.4. Application

Fig. 2 illustrates the annual concentrations of $PM_{2.5}$ and black carbon obtained from the LUR model to 23,605 households in the 28 villages of the study area. Corresponding concentrations from ordinary kriging are shown in the online supplement (Supplementary Fig. S3).

When comparing the predictions of the two methods (Fig. 3), larger within-village variability was observed when using LUR as compared to the ordinary kriging, particularly for black carbon (intra-class correlations were 0.72 and 0.99, respectively). The annual predictions for $PM_{2.5}$ obtained from LUR showed larger coefficient of variation than the predictions obtained from ordinary kriging (9% and 4%, respectively).

4. Discussion

To our knowledge, this is the first study to apply a LUR approach to a non-urban area of a LMIC. Our results show the feasibility of LUR in a setting with multiple local sources of particles and shed light on the spatial variability of $PM_{2.5}$ and black carbon concentrations in a peri-urban area of South India. The final LUR models for annual $PM_{2.5}$ and black carbon showed moderate to good explained variance (adjusted-R^2 of 58% and 79%, respectively), though only the black carbon model was robust through the validation process. The model for annual $PM_{2.5}$ included regional predictors retrieved from satellite imagery, while the black carbon model included both regional predictors and local predictors derived from a built-environment survey of the study area. Thus, our results provide insights on the added value of manually-collected built-environment data to increase the performance of the LUR model to predict black carbon concentrations in a peri-urban setting of LMICs. For both pollutants, LUR provided greater within-village variability in predictions as compared to ordinary kriging, an important feature for epidemiological studies relying on spatial variation in exposure.

Previous LUR models of $PM_{2.5}$ in urban areas have included primarily traffic-related predictors, either road length (Brokamp et al., 2017; Eeftens et al., 2012; Huang et al., 2017; Lee et al., 2017; Wolf et al., 2017; Wu et al., 2014, 2017) or traffic load/intensity (Dirgawati et al., 2016; Eeftens et al., 2012, 2016). None of the $PM_{2.5}$ models in our study included traffic-related features, likely because of the peri-urban nature of our study area. Nonetheless, night-time light intensity, an indicator of urbanicity, was included in all $PM_{2.5}$ models. This feature could relate to well-mixed traffic-related air pollution associated with urban areas as well as other emission sources (e.g., solid-fuel use) associated with rural areas. In the present work, all $PM_{2.5}$ models included an indicator for green spaces which is consistent with some previous LUR developed for cities that included indicators of green or natural spaces (Brokamp et al., 2017; Dirgawati et al., 2016; Eeftens et al., 2012, 2016; Hoek et al., 2011; Liu et al., 2016; Wu et al., 2015, 2017). Some previous LUR developed for $PM_{2.5}$ in cities included predictors of both small buffer sizes (≤100 m, generally traffic-related indicators) and large buffer sizes (e.g. population density, natural spaces, and industry) (Eeftens et al., 2012, 2016; Hoek et al., 2008; Wolf et al., 2017; Wu et al., 2014). In contrast, our $PM_{2.5}$ models mostly included predictors with large buffer sizes (village-level indicator or buffers ≥300 m),

Table 3
Evaluation of land-use regression models developed for $PM_{2.5}$ and black carbon using A) leave-one-out cross-validation and B) cross-holdout validation.

	Time period	A) Variance explained (adjusted-R^2)	A) RMSE	A) $R_{Spearman}$	B) Variance explained (adjusted-R^2)	B) Most included predictors during cross-holdout validation process (% of inclusion)	B) RMSE	B) $R_{Spearman}$
		Mean (sd) [min;max]	*Mean (sd) [min;max]*		*Mean (sd) [min;max]*		*Mean (sd) [min;max]*	
$PM_{2.5}$	Annual	0.58 (0.04) [0.40;0.65]	2.03 (1.35) [0.33;5.23]	0.49	0.56 (0.11) [0.32;0.81]	NTLI (58%), tree_1000 (54%), longitude (50%), ndvimax_5000 (50%)	7.48 (19.08) [0.60;96.24]	−0.06
	Post-monsoon	0.48 (0.05) [0.37;0.68]	2.21 (1.92) [0.08;7.99]	0.52	0.51 (0.11) [0.30;0.72]	ndvimax_5000 (54%), NTLI (42%), hh_50 (38%), tree_2000 (25%)	4.32 (2.73) [0.45;10.29]	−0.18
	Summer	0.34 (0.05) [0.15;0.44]	2.85 (2.47) [0.20;8.99]	0.34	0.53 (0.12) [0.18;0.78]	elevation (75%), NTLI (63%), tree_300 (63%), nrice_mill_1000 (58%), LPG (46%), nrp_e2_300 (33%), nbrick_kiln_work_2000 (29%), ndvimax_5000 (29%)	6.14 (7.30) [0.44;32.74]	−0.18
Black carbon	Annual	0.78 (0.02) [0.74;0.83]	0.20 (0.13) [0.01;0.54]	0.82	0.79 (0.04) [0.73;0.89]	tree_500 (87%), rrlen_5000 (83%), distnrp_e1 (78%)	0.30 (0.25) [0.01;0.90]	0.59
	Post-monsoon	0.63 (0.03) [0.55;0.74]	0.35 (0.29) [0.00;1.14]	0.60	0.67 (0.05) [0.60;0.80]	prlen_5000 (83%), LPG (52%), ndvimax_100 (26%), tree_500 (26%)	1.01 (2.20) [0.06;10.96]	0.39
	Summer	0.65 (0.04) [0.55;0.78]	0.23 (0.18) [0.02;0.67]	0.76	0.65 (0.04) [0.60;0.78]	rrden_5000 (78%), nrp_e2_300 (74%), nrp_p1_5000 (61%), prlen_1000 (57%)	0.36 (0.27) [0.02;1.14]	0.23

Abbreviations: RMSE, root-mean-square error; sd, standard deviation. $R_{Spearman}$ are correlation coefficients between measured and predicted concentrations. Variable names: tree coverage (tree_X), village-level night-time light intensity (NTLI), post-monsoon value for the Normalized Difference in Vegetation Index (ndvimax_X), annual average value for the Normalized Difference in Vegetation Index (ndviavg__), household density (hh_X), ring road length (rrlen_X), primary roads length (prlen_X), village-level liquid petroleum gas use (LPG), ring road density (rrden_X), number of rice mill industries (nrice_mill_X), number of operational brick kiln industries (nbrick_kiln_work_X), distance to non-residential place related to wood/gas supply (distnrp_e1), number of paved bus route (nrp_s8_X), and number of non-residential places related to petrol supply (nrp_e2_X) and to religious center (nrp_p1_X) with_X indicating the radius of the buffer.

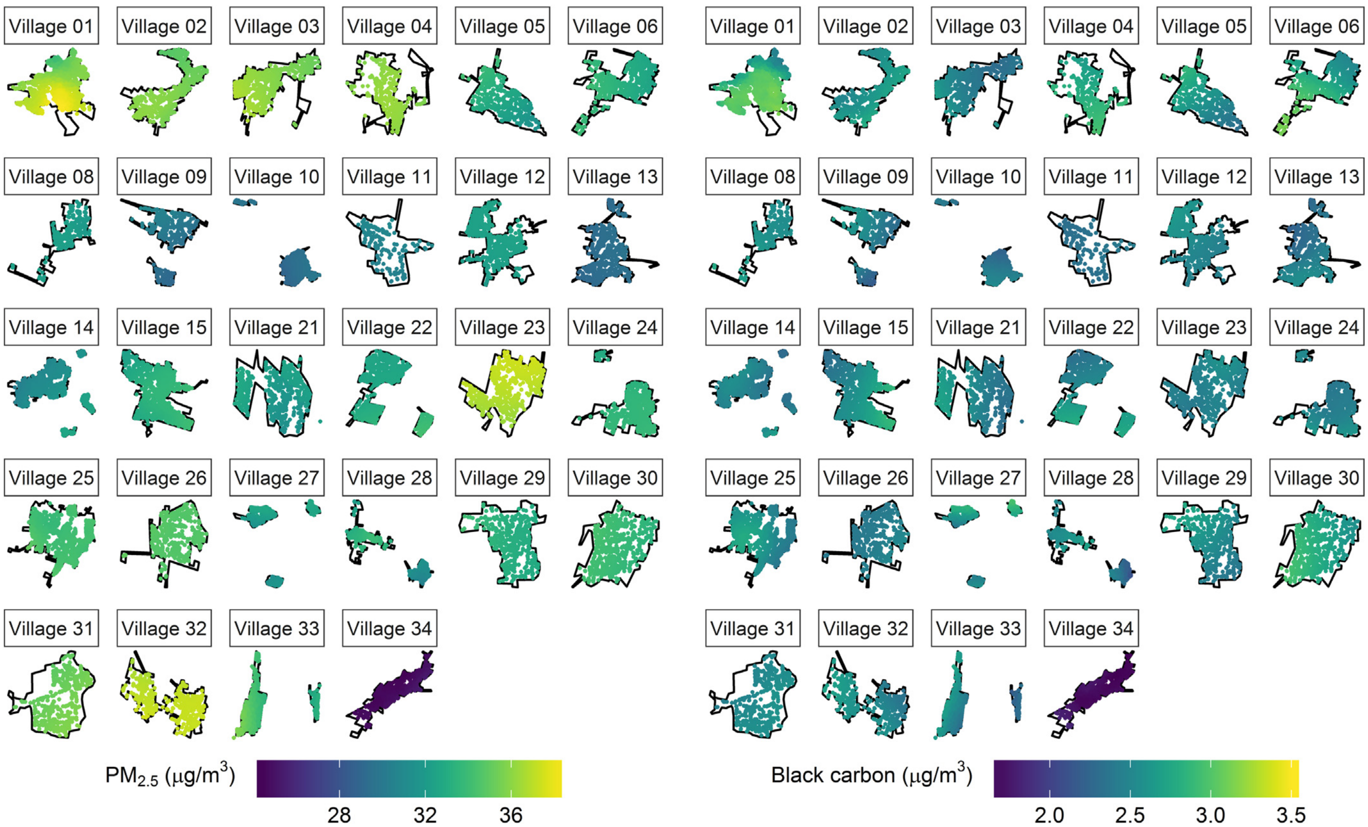


Fig. 2. Maps of predicted annual concentrations from the land-use regression models for 23,605 households of the study area for $PM_{2.5}$ (left) and black carbon (right). Only households inside villages are presented as the area between villages includes mostly crops and agricultural lands with no or few inhabitants. Village names are not displayed for confidentiality purposes.

reflecting predominantly regional source contributions. This is consistent with results from an ongoing analysis of the ambient background monitoring data in the study area which estimated that local sources contributed to <25% of the ambient $PM_{2.5}$, according to moving average subtraction method (Watson and Chow, 2001). The explained variance of our model for annual $PM_{2.5}$ was lower than previously observed in cities (≥70% in most case) (Dirgawati et al., 2016; Huang et al., 2017; Liu et al., 2016; Wolf et al., 2017; Wu et al., 2017) but was still similar to LUR in Beijing, China (Wu et al., 2015); Hong Kong (Lee et al., 2017); Cincinnati, USA (Brokamp et al., 2017); and some European

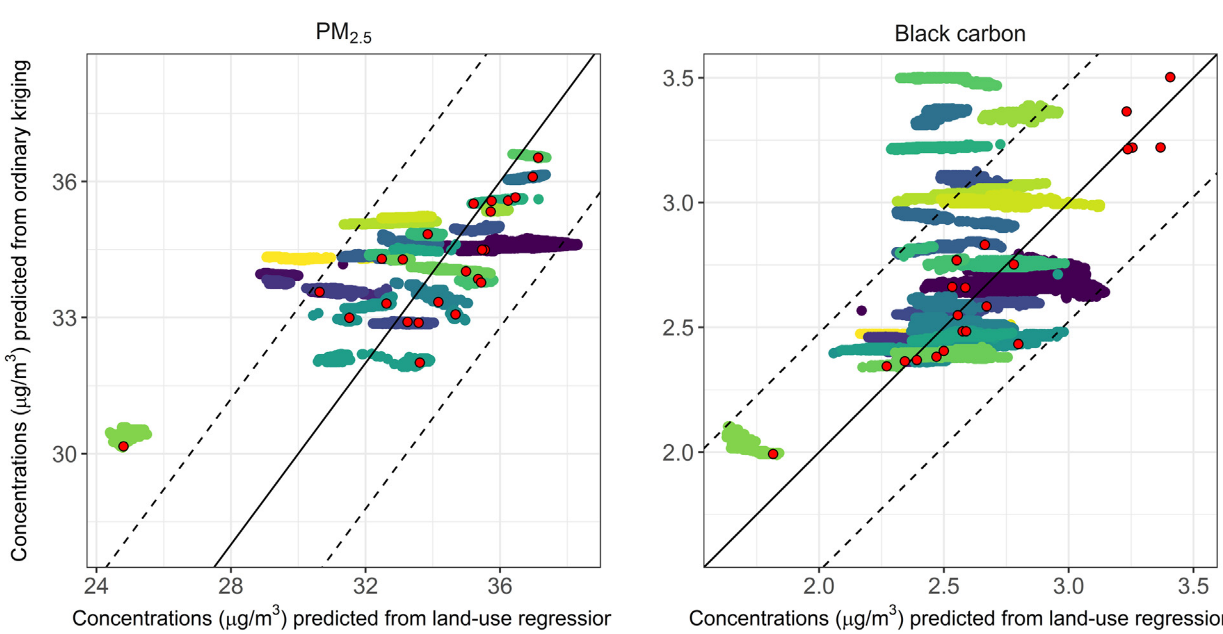


Fig. 3. Comparison of $PM_{2.5}$ and black carbon annual concentrations predicted by land-use regression and ordinary kriging at 23,605 households of the study area. Colors represent villages. Red points are sampled sites values. Solid line is identity. Dashed lines represent ±1 standard deviation of measurement. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

areas (Eeftens et al., 2012). The low predictive ability we found may relate to limited variability in measured concentrations, quality of the predictor variables, and complexity of sources.

Black carbon has been proposed as a useful indicator for the health effects of particulate matter, particularly in areas dominated by primary combustion such as in the present analysis (e.g., solid fuel use for cooking) (Janssen et al., 2011). Previous LUR models for black carbon (or $PM_{2.5}$ absorbance as proxy) have been developed in HICs and included traffic indicators, roads length or density, buildings or population density, industries, and water bodies (Dirgawati et al., 2016; Eeftens et al., 2012, 2016; Lee et al., 2017; Montagne et al., 2015; Wolf et al., 2017). In the present study, the most predictive variable for black carbon was an indicator of traffic but within a large buffer size. In our study area, indicators of ring/primary roads within 5000 m were highly and negatively correlated with village-level proportion of liquid petroleum gas use. Thus, the road length predictors selected here to predict black carbon may be representative of fuel use in a more local scale rather than solely be indicative of traffic within 5000 m. Other selected predictors of black carbon were mostly local, in particular locations related to energy supplies. Overall, our LUR models effectively captured the local spatial variability of black carbon, with a high explained variance of almost 80% for the annual model, which was comparable with (Dirgawati et al., 2016; Eeftens et al., 2012, 2016; Montagne et al., 2015; Wolf et al., 2017) or higher than (Dons et al., 2014; Lee et al., 2017) previously developed models.

We conducted extensive model evaluation including leave-one-out cross-validation and cross-holdout validation. The decrease in performance of our $PM_{2.5}$ models during leave-one-out cross-validation was in range with previously developed LUR (Dirgawati et al., 2016; Eeftens et al., 2012, 2016; Huang et al., 2017; Lee et al., 2017; Wolf et al., 2017; Wu et al., 2014). However, performance of our $PM_{2.5}$ models was way poorer during cross-holdout validation in terms of explained variance, influential observations, and predictive ability, which questioned the reliability and robustness of the models we developed. Cross-holdout validation has been shown to accurately reflect out-of-sample predictive ability of LUR models when using limited sample size (Wang et al., 2016). The relatively small sample size might be one explanation for poor performance (Basagaña et al., 2013) or the spatial variation of $PM_{2.5}$ in the area might not be well captured by the sampled sites. In contrast, our black carbon models were stable during cross-holdout validation, with high consistency in predictor selection and moderate to good performance of variance explained and predictive ability, supporting the reliability and robustness of the models.

There was limited similarity between the $PM_{2.5}$ and black carbon models. First, models for black carbon performed better than those for $PM_{2.5}$ as they explained more spatial variation and were more robust through evaluation process. The higher spatial variability in black carbon measurements might be one explanation. Consistent with our analysis, better performance for black carbon (or $PM_{2.5}$ absorbance) models than for $PM_{2.5}$ has been found in other locations (Dirgawati et al., 2016; Eeftens et al., 2012; Wolf et al., 2017). Second, our model for annual $PM_{2.5}$ only included regional-scale predictors that were all derived from satellite imagery (expect household density), while the model for annual black carbon included both local- and regional-scale predictors that were mostly retrieved from the built-environment survey. Some previous models for black carbon and $PM_{2.5}$ showed similar predictors for the same area (Lee et al., 2017; Saraswat et al., 2013; Wu et al., 2014). In the ESCAPE project, predictors for $PM_{2.5}$ and $PM_{2.5}$ absorbance were also similar, though the ones for $PM_{2.5}$ absorbance had larger buffer sizes (unlike our findings) (Eeftens et al., 2012). The observed discrepancy with the literature may in part be explained by the specificity of our peri-urban study area experiencing a mix of regional and very local sources of particulate matter. The considered predictors, mostly derived from built-environment survey, likely captured some sources that are usually not relevant for or available in most urban areas of HICs, such as energy use (e.g. biomass burning, liquid petroleum gas) and specific non-residential places (e.g. temples, shops).

Manually-collected geographic predictors derived from the built-environment survey were valuable to improve the variance explained by the LUR models for black carbon, but not $PM_{2.5}$. The built-environment survey data allowed us to consider culture-specific classes of predictors, such as distance to religious centers and energy supply places, which have not been considered previously in the LUR literature, either because they were unavailable or irrelevant. Other authors have highlighted the importance of culture- or site-specific land-use classes to predict particulate matter in Taipei, Taiwan, or Quebec, Canada (Smargiassi et al., 2012; Wu et al., 2017). The added value of built-environment survey data to LUR models of air pollution should be explored in other settings, particularly those with diverse local sources or limited GIS data.

Similar to other studies, ordinary kriging predictions here showed limited spatial variability, in particular within-village, as compared to those from LUR models (Alexeeff et al., 2015). Universal kriging, a two-step approach that combines LUR and ordinary kriging, has been proposed to improve predictions (Mercer et al., 2011; Young et al., 2016).We could not apply universal kriging because correlations were not found in the residuals of the LUR models, suggesting that our models already captured major predictors of $PM_{2.5}$ and black carbon spatial variations. The increased within-village variability observed using LUR models supports the use of LUR predictions for epidemiological studies on the long-term effect of air pollution in the study area.

One limitation of our study is the small number of monitoring sites included in the model development relative to the number of considered predictors, which could have decreased the reliability of the developed LUR models. Indeed, limited sample size has been shown to artificially increase the variance explained when too many predictors are offered to the model (Basagaña et al., 2012). Nonetheless, the monitor density in our study area was relatively high (23–24 monitors for an area of interest of 8.2 km^2) and the number of sites was comparable to previous studies in HICs. For several cities included in the ESCAPE project, the monitoring density was lower: ≤20 sites for larger area than ours (Eeftens et al., 2012). To minimize the risk of over fitting and collinearity, stringent criteria were applied during selection process – we used a supervised approach with *a priori* defined directions of effects, we considered only predictors with enough variability, we restricted predictor inclusion by checking correlation and variance inflation factor, and we limited the number of predictors in the final model. We also made an extensive effort to evaluate the robustness of the selected predictors and the performance of the models with leave-one-out cross-validation and cross-holdout validation. Despite extensive procedure, the lack of external dataset may have limited the evaluation of our models. Variability in measured concentrations could have limited the potential for prediction in the present analysis, though we selected sampling sites to maximize contrasts in pollutant concentrations. Another limitation of the present analysis was the village-level resolution of biomass cooking fuel use. Availability of data with better resolution may improve the interpretability or the performance of the model. However, a household-level indicator derived from comprehensive household survey did not show stronger crude association with our air pollution measurements. The temporal discrepancy between our sampling campaign (2015–2016) and the collection of the geographic variables (NTLI was derived for year 2012 and built-environment survey was performed in 2013) could have impacted the performance of our models. Though, we do not expect changes in sources or land use to be large over a few years. We used NTLI data from 2012 because it had been processed to achieve better spatial resolution; however we observed strong correlation between data we used and data concurrent with our sampling campaign (2015). Industry is mostly an informal sector in the study area with highly unpredictable operating time and duration, making the likely timing of emissions difficult to capture in the built-environment survey. This may partly explain why industries were not selected in our models. Changes in sources or levels of particulate matter associated with urbanization could limit the applicability of our models in the future, although substantial changes in sources or land use are not likely to happen within a decade.

The outdoor concentrations we measured and modeled in peri-urban South India exceed typical $PM_{2.5}$ and black carbon concentrations used for epidemiological purposes in the United States and Europe (Beelen et al., 2014; Chan et al., 2015; Kaufman et al., 2016; Lepeule et al., 2014; Rice et al., 2015), although they were substantially lower than concentrations in North India (Gautam et al., 2016; Sharma and Kulshrestha, 2014). In particular, the spatial variability of $PM_{2.5}$ and black carbon in our study area (as inter-quartile range or sd/mean) was consistently higher than typical values found in the literature (Beelen et al., 2014; Chan et al., 2015; Kaufman et al., 2016; Lepeule et al., 2014; Rice et al., 2015). Previous authors have highlighted the dearth of epidemiological cohort studies in regions with relatively high $PM_{2.5}$ exposures leading to uncertainty when predicting health effects across the global exposure range (Burnett et al., 2014; Lelieveld et al., 2015). The exposure-response function developed by Burnett et al. considered only relative risks associated with ambient $PM_{2.5}$ levels < 29 μg/m^3 to model the shape of the relationship between $PM_{2.5}$ and ischemic heart disease (Burnett et al., 2014). By comparison, the ambient concentrations measured in our study area were similar to the $PM_{2.5}$ levels for passive smoking considered by Burnett et al. (Burnett et al., 2014).

Well-performing models to predict exposure in low-/middle-income countries is essential to fill the gap in epidemiological evidence regarding health effects of long-term exposure to outdoor air pollution at exposures higher than those typically observed in North America and Europe. We showed that coupling a LUR approach with a built-environment survey was more effective in capturing within-village spatial variability, an essential point in epidemiology, than the ordinary kriging approach. Future air pollution exposure assessment in areas with limited geographic data and/or culturally specific sources could benefit from the combination of manually-collected built-environment data and remote sensing data. The LUR models we developed will be applied locally to all APCAPS participants living in the Southeast region of Hyderabad, India, supporting the epidemiological investigation of the long-term effects of $PM_{2.5}$ and black carbon in this population.

Abbreviations

LUR	land-use regression
HIC	high-income country
LMIC	low-/middle-income country
GIS	Geographic Information System
NTLI	night-time light intensity
NDVI	Normalized Difference Vegetation Index
NRP	non-residential places
$PM_{2.5}$	particulate matter with a diameter ≤ 2.5 μm

Conflict of interest

The authors declare no competing financial interest.

Acknowledgements

The research leading to these results received funding from the European Research Council under ERC Grant Agreement no 336167 for the CHAI Project. The built-environment survey within the framework of the APCAPS study was funded by the Wellcome Trust (084674/Z). Cathryn Tonne is funded by a Ramón y Cajal fellowship (RYC-2015-17402) awarded by the Spanish Ministry of Economy and Competitiveness.

We thank Christopher Bulkley-Logston who designed and built the structure of the stations for the sampling campaign. We are grateful to all members of the study teams of the APCAPS and CHAI studies who made the research possible.

Appendix A. Supplementary data

Description of the evaluated predictors for $PM_{2.5}$ and black carbon (Table S1); sensitivity analysis of the land-use regression models for $PM_{2.5}$ and black carbon with no restriction on the included number of predictors (Table S2); sensitivity analysis of the land-use regression models for $PM_{2.5}$ and black carbon considering only satellite-derived predictors (Table S3); predictions from ordinary kriging for annual $PM_{2.5}$ and annual black carbon (Table S4); diagnostic plots of the developed land-use regression models for $PM_{2.5}$ (Fig. S1) and black carbon (Fig. S2); maps of predicted annual concentrations using ordinary kriging for 23,605 households of the study area for $PM_{2.5}$ and black carbon (Fig. S3). Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2018.03.308.

References

Alexeeff, S.E., Schwartz, J., Kloog, I., Chudnovsky, A., Koutrakis, P., Coull, B.A., 2015. Consequences of kriging and land use regression for $PM_{2.5}$ predictions in epidemiologic analyses: insights into spatial variability using high-resolution satellite data. J. Expo. Sci. Environ. Epidemiol. 25:138–144. https://doi.org/10.1038/jes.2014.40.

Balakrishnan, K., Sambandam, S., Ramaswamy, P., Ghosh, S., Venkatesan, V., Thangavel, G., Mukhopadhyay, K., Johnson, P., Paul, S., Puttaswamy, N., Dhaliwal, R.S., Shukla, D.K., 2015. Establishing integrated rural-urban cohorts to assess air pollution-related health effects in pregnant women, children and adults in Southern India: an overview of objectives, design and methods in the Tamil Nadu Air Pollution and Health Effects (TAPHE). BMJ Open 5, e008090. https://doi.org/10.1136/bmjopen-2015-008090.

Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., de Nazelle, A., Nieuwenhuijsen, M., Vila, J., Künzli, N., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. Atmos. Environ. 54:634–642. https://doi.org/10.1016/j.atmosenv.2012.01.064.

Basagaña, X., Aguilera, I., Rivera, M., Agis, D., Foraster, M., Marrugat, J., Elosua, R., Künzli, N., 2013. Measurement error in epidemiologic studies of air pollution based on land-use regression models. Am. J. Epidemiol. 178:1342–1346. https://doi.org/10.1093/aje/kwt127.

Baugh, K., Elvidge, C.D., Ghosh, T., Ziskin, D., 2010. Development of a 2009 Stable Lights Product using DMSP-OLS data. Proc. Asia-Pacific Adv. Netw. 30:114–130. https://doi.org/10.7125/APAN.30.17.

Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z.J., Weinmayr, G., Hoffmann, B., Wolf, K., Samoli, E., Fischer, P., Nieuwenhuijsen, M., Vineis, P., Xun, W.W., Katsouyanni, K., Dimakopoulou, K., Oudin, A., Forsberg, B., Modig, L., Havulinna, A.S., Lanki, T., Turunen, A., Oftedal, B., Nystad, W., Nafstad, P., De Faire, U., Pedersen, N.L., Östenson, C.-G., Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K.T., Overvad, K., Ellermann, T., Eeftens, M., Peeters, P.H., Meliefste, K., Wang, M., Bueno-de-Mesquita, B., Sugiri, D., Krämer, U., Heinrich, J., de Hoogh, K., Key, T., Peters, A., Hampel, R., Concin, H., Nagel, G., Ineichen, A., Schaffner, E., Probst-Hensch, N., Künzli, N., Schindler, C., Schikowski, T., Adam, M., Phuleria, H., Vilier, A., Clavel-Chapelon, F., Declercq, C., Grioni, S., Krogh, V., Tsai, M.-Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Katsoulis, M., Trichopoulou, A., Brunekreef, B., Hoek, G., 2014. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. Lancet 383: 785–795. https://doi.org/10.1016/S0140-6736(13)62158-3.

Both, A.F., Balakrishnan, A., Joseph, B., Marshall, J.D., 2011. Spatiotemporal aspects of real-time $PM_{2.5}$: low- and middle-income neighborhoods in Bangalore, India. Environ. Sci. Technol. 45:5629–5636. https://doi.org/10.1021/es104331w.

Brokamp, C., Jandarov, R., Rao, M.B., LeMasters, G., Ryan, P., 2017. Exposure assessment models for elemental components of particulate matter in an urban environment: a comparison of regression and random forest approaches. Atmos. Environ. 151: 1–11. https://doi.org/10.1016/j.atmosenv.2016.11.066.

Burnett, R.T., Pope, C.A., Ezzati, M., Olives, C., Lim, S.S., Mehta, S., Shin, H.H., Singh, G., Hubbell, B., Brauer, M., Anderson, H.R., Smith, K.R., Balmes, J.R., Bruce, N.G., Kan, H., Laden, F., Prüss-Ustün, A., Turner, M.C., Gapstur, S.M., Diver, W.R., Cohen, A., 2014. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. Environ. Health Perspect. 122: 397–403. https://doi.org/10.1289/ehp.1307049.

Chan, S.H., Van Hee, V.C., Bergen, S., Szpiro, A.A., DeRoo, L.A., London, S.J., Marshall, J.D., Kaufman, J.D., Sandler, D.P., 2015. Long-term air pollution exposure and blood pressure in the sister study. Environ. Health Perspect. 123:951–958. https://doi.org/10.1289/ehp.1408125.

Dirgawati, M., Heyworth, J.S., Wheeler, A.J., McCaul, K.A., Blake, D., Boeyen, J., Cope, M., Yeap, B.B., Nieuwenhuijsen, M., Brunekreef, B., Hinwood, A., 2016. Development of Land Use Regression models for particulate matter and associated components in a low air pollutant concentration airshed. Atmos. Environ. 144:69–78. https://doi.org/10.1016/j.atmosenv.2016.08.013.

Dons, E., Van Poppel, M., Int Panis, L., De Prins, S., Berghmans, P., Koppen, G., Matheeussen, C., 2014. Land use regression models as a tool for short, medium and long term exposure to traffic related air pollution. Sci. Total Environ. 476–477:378–386. https://doi.org/10.1016/j.scitotenv.2014.01.025.

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dėdelė, A., Dons, E., de Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Gražulevičienė, R., Heinrich, J., Hoffmann, B., Jerrett, M., Keidel, D., Korek, M., Lanki, T., Lindley, S., Madsen, C., Mölter, A., Nádor, G., Nieuwenhuijsen, M., Nonnemacher, M., Pedeli, X., Raaschou-Nielsen, O., Patelarou, E., Quass, U., Ranzi, A., Schindler, C., Stempfelet, M., Stephanou, E., Sugiri, D., Tsai, M.-Y., Yli-Tuomi, T., Varró, M.J., Vienneau, D., von Klot, S., Wolf, K., Brunekreef, B., Hoek, G., 2012. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PMcoarse in 20 European Study Areas; Results of the ESCAPE Project. Environ. Sci. Technol. 46: 11195–11205. https://doi.org/10.1021/es301948k.

Eeftens, M., Meier, R., Schindler, C., Aguilera, I., Phuleria, H., Ineichen, A., Davey, M., Ducret-Stich, R., Keidel, D., Probst-Hensch, N., Künzli, N., Tsai, M.-Y., 2016. Development of land use regression models for nitrogen dioxide, ultrafine particles, lung deposited surface area, and four other markers of particulate matter pollution in the Swiss SAPALDIA regions. Environ. Health 15, 53. https://doi.org/10.1186/s12940-016-0137-9.

Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R., Davis, C.W., 1997. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. Int. J. Remote Sens. 18:1373–1379. https://doi.org/10.1080/014311697218485.

Gautam, S., Yadav, A., Tsai, C.-J., Kumar, P., 2016. A review on recent progress in observations, sources, classification and regulations of $PM_{2.5}$ in Asian environments. Environ. Sci. Pollut. Res. 23:21165–21175. https://doi.org/10.1007/s11356-016-7515-2.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42:7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057.

Hoek, G., Beelen, R., Kos, G., Dijkema, M., van der Zee, S.C., Fischer, P.H., Brunekreef, B., 2011. Land use regression model for ultrafine particles in Amsterdam. Environ. Sci. Technol. 45:622–628. https://doi.org/10.1021/es1023042.

Huang, L., Zhang, C., Bi, J., 2017. Development of land use regression models for $PM_{2.5}$, SO_2, NO_2 and O_3 in Nanjing, China. Environ. Res. 158:542–552. https://doi.org/10.1016/j.envres.2017.07.010.

Janssen, N.A.H., Hoek, G., Simic-Lawson, M., Fischer, P., van Bree, L., ten Brink, H., Keuken, M., Atkinson, R.W., Anderson, H.R., Brunekreef, B., Cassee, F.R., 2011. Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM_{10} and $PM_{2.5}$. Environ. Health Perspect. 119:1691–1699. https://doi.org/10.1289/ehp.1003369.

Karagulian, F., Belis, C.A., Dora, C.F.C., Prüss-Ustün, A.M., Bonjour, S., Adair-Rohani, H., Amann, M., 2015. Contributions to cities' ambient particulate matter (PM): a systematic review of local source contributions at global level. Atmos. Environ. 120:475–483. https://doi.org/10.1016/j.atmosenv.2015.08.087.

Kaufman, J.D., Adar, S.D., Barr, R.G., Budoff, M., Burke, G.L., Curl, C.L., Daviglus, M.L., Roux, A.V.D., Gassett, A.J., Jacobs, D.R., Kronmal, R., Larson, T.V., Navas-Acien, A., Olives, C., Sampson, P.D., Sheppard, L., Siscovick, D.S., Stein, J.H., Szpiro, A.A., Watson, K.E., 2016. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. Lancet 388:696–704. https://doi.org/10.1016/S0140-6736(16)00378-0.

Kinra, S., Radha Krishna, K., Kuper, H., Rameshwar Sarma, K., Prabhakaran, P., Gupta, V., Walia, G.K., Bhogadi, S., Kulkarni, B., Kumar, A., Aggarwal, A., Gupta, R., Prabhakaran, D., Reddy, K.S., Davey Smith, G., Ben-Shlomo, Y., Ebrahim, S., 2014. Cohort profile: Andhra Pradesh Children and Parents Study (APCAPS). Int. J. Epidemiol. 43:1417–1424. https://doi.org/10.1093/ije/dyt128.

Lee, M., Brauer, M., Wong, P., Tang, R., Tsui, T.H., Choi, C., Cheng, W., Lai, P.-C., Tian, L., Thach, T.-Q., Allen, R., Barratt, B., 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. https://doi.org/10.1016/j.scitotenv.2017.03.094.

Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 525: 367–371. https://doi.org/10.1038/nature15371.

Lepeule, J., Litonjua, A.A., Coull, B., Koutrakis, P., Sparrow, D., Vokonas, P.S., Schwartz, J., 2014. Long-term effects of traffic particles on lung function decline in the elderly. Am. J. Respir. Crit. Care Med. 190, 542–548.

Liu, C., Henderson, B.H., Wang, D., Yang, X., Peng, Z., 2016. A land use regression application into assessing spatial variation of intra-urban fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2) concentrations in City of Shanghai, China. Sci. Total Environ. 565:607–615. https://doi.org/10.1016/j.scitotenv.2016.03.189.

Ma, J., Simonich, S., Tao, S., 2017. New discoveries to old problems: a virtual issue on air pollution in rapidly industrializing countries. Environ. Sci. Technol. 51: 11497–11501. https://doi.org/10.1021/acs.est.7b04885.

Mercer, L.D., Szpiro, A.A., Sheppard, L., Lindström, J., Adar, S.D., Allen, R.W., Avol, E.L., Oron, A.P., Larson, T., Liu, L.-J.S., Kaufman, J.D., 2011. Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen (NOx) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). Atmos. Environ. 45:4412–4420. https://doi.org/10.1016/j.atmosenv.2011.05.043.

Montagne, D.R., Hoek, G., Klompmaker, J.O., Wang, M., Meliefste, K., Brunekreef, B., 2015. Land use regression models for ultrafine particles and black carbon based on short-term monitoring predict past spatial variation. Environ. Sci. Technol. 49:8712–8720. https://doi.org/10.1021/es505791g.

Paliwal, U., Sharma, M., Burkhart, J.F., 2016. Monthly and spatially resolved black carbon emission inventory of India: uncertainty analysis. Atmos. Chem. Phys. 16: 12457–12476. https://doi.org/10.5194/acp-16-12457-2016.

Patton, A.P., Zamore, W., Naumova, E.N., Levy, J.I., Brugge, D., Durant, J.L., 2015. Transferability and generalizability of regression models of ultrafine particles in urban neighborhoods in the Boston area. Environ. Sci. Technol. 49:6051–6060. https://doi.org/10.1021/es5061676.

Rice, M.B., Ljungman, P.L., Wilker, E.H., Dorans, K.S., Gold, D.R., Schwartz, J., Koutrakis, P., Washko, G.R., O'Connor, G.T., Mittleman, M.A., 2015. Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the Framingham Heart Study. Am. J. Respir. Crit. Care Med. 191:656–664. https://doi.org/10.1164/rccm.201410-1875OC.

Saraswat, A., Apte, J.S., Kandlikar, M., Brauer, M., Henderson, S.B., Marshall, J.D., 2013. Spatiotemporal land use regression models of fine, ultrafine, and black carbon particulate matter in New Delhi, India. Environ. Sci. Technol. 47:12903–12911. https://doi.org/10.1021/es401489h.

Sharma, D., Kulshrestha, U.C., 2014. Spatial and temporal patterns of air pollutants in rural and urban areas of India. Environ. Pollut. 195:276–281. https://doi.org/10.1016/j.envpol.2014.08.026.

Smargiassi, A., Brand, A., Fournier, M., Tessier, F., Goudreau, S., Rousseau, J., Benjamin, M., 2012. A spatiotemporal land-use regression model of winter fine particulate levels in residential neighbourhoods. J. Expo. Sci. Environ. Epidemiol. 22:331–338. https://doi.org/10.1038/jes.2012.26.

Tonne, C., Salmon, M., Sanchez, M., Sreekanth, V., Bhogadi, S., Sambandam, S., Balakrishnan, K., Kinra, S., Marshall, J.D., 2017. Integrated assessment of exposure to $PM_{2.5}$ in South India and its relation with cardiovascular risk: design of the CHAI observational cohort study. Int. J. Hyg. Environ. Health 220:1081–1088. https://doi.org/10.1016/j.ijheh.2017.05.005.

Tyagi, A., Singh, P., 2013. Applying kriging approach on pollution data using GIS software. Int. J. Environ. Eng. Manag. 4, 185–190.

U.S. Geological survey, 2017. LANDSAT 8 OLI-TIRS 30 m pixel. GloVis: Global Visualization Viewer. (WWW Document. URL). https://earthexplorer.usgs.gov.

van Nunen, E., Vermeulen, R., Tsai, M.-Y., Probst-Hensch, N., Ineichen, A., Davey, M., Imboden, M., Ducret-Stich, R., Naccarati, A., Raffaele, D., Ranzi, A., Ivaldi, C., Galassi, C., Nieuwenhuijsen, M., Curto, A., Donaire-Gonzalez, D., Cirach, M., Chatzi, L., Kampouri, M., Vlaanderen, J., Meliefste, K., Buijtenhuijs, D., Brunekreef, B., Morley, D., Vineis, P., Gulliver, J., Hoek, G., 2017. Land use regression models for ultrafine particles in six European areas. Environ. Sci. Technol. 51:3336–3345. https://doi.org/10.1021/acs.est.6b05920.

Wang, M., Beelen, R., Bellander, T., Birk, M., Cesaroni, G., Cirach, M., Cyrys, J., de Hoogh, K., Declercq, C., Dimakopoulou, K., Eeftens, M., Eriksen, K.T., Forastiere, F., Galassi, C., Grivas, G., Heinrich, J., Hoffmann, B., Ineichen, A., Korek, M., Lanki, T., Lindley, S., Modig, L., Mölter, A., Nafstad, P., Nieuwenhuijsen, M.J., Nystad, W., Olsson, D., Raaschou-Nielsen, O., Ragettli, M.S., Ranzi, A., Stempfelet, M., Sugiri, D., Tsai, M.-Y., Udvardy, O., Varró, M.J., Vienneau, D., Weinmayr, G., Wolf, K., Yli-Tuomi, T., Hoek, G., Brunekreef, B., 2014. Performance of multi-city land use regression models for nitrogen dioxide and fine particles. Environ. Health Perspect. 122:843–850. https://doi.org/10.1289/ehp.1307271.

Wang, M., Brunekreef, B., Gehring, U., Szpiro, A., Hoek, G., Beelen, R., 2016. A new technique for evaluating land-use regression models and their impact on health effect estimates. Epidemiology 27:51–56. https://doi.org/10.1097/EDE.0000000000000404.

Watson, J.G., Chow, J.C., 2001. Estimating middle-, neighborhood-, and urban-scale contributions to elemental carbon in Mexico city with a rapid response aethalometer. J. Air Waste Manage. Assoc. 51:1522–1528. https://doi.org/10.1080/10473289.2001.10464379.

Weichenthal, S., Van Ryswyk, K., Goldstein, A., Bagg, S., Shekkarizfard, M., Hatzopoulou, M., 2016. A land use regression model for ambient ultrafine particles in Montreal, Canada: a comparison of linear regression and a machine learning approach. Environ. Res. 146:65–72. https://doi.org/10.1016/j.envres.2015.12.016.

Wolf, K., Cyrys, J., Harciníková, T., Gu, J., Kusch, T., Hampel, R., Schneider, A., Peters, A., 2017. Land use regression modeling of ultrafine particles, ozone, nitrogen oxides and markers of particulate matter pollution in Augsburg, Germany. Sci. Total Environ. 579:1531–1540. https://doi.org/10.1016/j.scitotenv.2016.11.160.

World Health Organization, 2016. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease. World Health Organization, Geneva (doi:9789241511353).

Wu, C.-F., Lin, H.-I., Ho, C.-C., Yang, T.-H., Chen, C.-C., Chan, C.-C., 2014. Modeling horizontal and vertical variation in intraurban exposure to $PM_{2.5}$ concentrations and compositions. Environ. Res. 133:96–102. https://doi.org/10.1016/j.envres.2014.04.038.

Wu, J., Li, J., Peng, J., Li, W., Xu, G., Dong, C., 2015. Applying land use regression model to estimate spatial variation of $PM_{2.5}$ in Beijing, China. Environ. Sci. Pollut. Res. 22: 7045–7061. https://doi.org/10.1007/s11356-014-3893-5.

Wu, C.-D., Chen, Y.-C., Pan, W.-C., Zeng, Y.-T., Chen, M.-J., Guo, Y.L., Lung, S.-C.C., 2017. Land-use regression with long-term satellite-based greenness index and culture-specific sources to model $PM_{2.5}$ spatial-temporal variability. Environ. Pollut. 224: 148–157. https://doi.org/10.1016/j.envpol.2017.01.074.

Young, M.T., Bechle, M.J., Sampson, P.D., Szpiro, A.A., Marshall, J.D., Sheppard, L., Kaufman, J.D., 2016. Satellite-based NO_2 and model validation in a national prediction model based on universal kriging and land-use regression. Environ. Sci. Technol. 50: 3686–3694. https://doi.org/10.1021/acs.est.5b05099.

Zhang, K., Larson, T.V., Gassett, A., Szpiro, A.A., Daviglus, M., Burke, G.L., Kaufman, J.D., Adar, S.D., 2014. Characterizing spatial patterns of airborne coarse particulate ($PM10_{-2.5}$) mass and chemical components in three cities: the multi-ethnic study of atherosclerosis. Environ. Health Perspect. 122:823–831. https://doi.org/10.1289/ehp.1307287.