Science of the Total Environment 514 (2015) 178–184

Contents lists available at ScienceDirect

Science of the Total Environment

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

LUR models for particulate matters in the Taipei metropolis with high densities of roads and strong activities of industry, commerce and construction

Jui-Huna Lee [a], Chang-Fu Wu [a], Gerard Hoek [b], Kees de Hoogh [c,d,e], Rob Beelen [b], Bert Brunekreef [b,f], Chang-Chuan Chan [a,*]

[a] Institute of Occupational Medicine and Industrial Hygiene, College of Public Health, National Taiwan University, Taipei, Taiwan
[b] Institute for Risk Assessment Sciences, Utrecht University, P.O. Box 80178, 3508 TD Utrecht, The Netherlands
[c] Swiss Tropical and Public Health Institute, Basel, Switzerland
[d] University of Basel, Basel, Switzerland
[e] MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom
[f] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

HIGHLIGHTS

- The ESCAPE LUR modeling approach can be applied to the Taipei metropolis.
- Incorporating local variables relevant to PM emissions improve model performance.
- Road area is a good surrogate for traffic intensity data in the Taipei metropolis.
- $PM_{2.5}$ and $PM_{2.5}$ absorbance models yielded 96% and 95% of explained variability.

ARTICLE INFO

Article history:
Received 30 October 2014
Received in revised form 20 January 2015
Accepted 27 January 2015
Available online 5 February 2015

Editor: Lidia Morawska

Keywords:
Land use regression
Particulate matter
GIS
Road area
Elevated highway
Long-term exposure

ABSTRACT

Traffic intensity, length of road, and proximity to roads are the most common traffic indicators in the land use regression (LUR) models for particulate matter in ESCAPE study areas in Europe. This study explored what local variables can improve the performance of LUR models in an Asian metropolis with high densities of roads and strong activities of industry, commerce and construction. By following the ESCAPE procedure, we derived LUR models of $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse} ($PM_{2.5-10}$) in Taipei. The overall annual average concentrations of $PM_{2.5}$, PM_{10}, and PM_{coarse} were 26.0 ± 5.6, 48.6 ± 5.9, and 23.3 ± 3.1 μg/m^3, respectively, and the absorption coefficient of $PM_{2.5}$ was $2.0 \pm 0.4 \times 10^{-5}$ m^{-1}. Our LUR models yielded R^2 values of 95%, 96%, 87%, and 65% for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}, respectively. $PM_{2.5}$ levels were increased by local traffic variables, industrial, construction, and residential land-use variables and decreased by rivers; while $PM_{2.5}$ absorbance levels were increased by local traffic variables, industrial, and commercial land-use variables in the models. PM_{coarse} levels were increased by elevated highways. Road area explained more variance than road length by increasing the incremental value of 27% and 6% adjusted R^2 for $PM_{2.5}$ and PM_{10} models, respectively. In the $PM_{2.5}$ absorbance model, road area and transportation facility explain 29% more variance than road length. In the PM_{coarse} model, industrial and new local variables instead of road length improved the incremental value of adjusted R^2 from 39% to 60%. We concluded that road area can better explain the spatial distribution of $PM_{2.5}$ and $PM_{2.5}$ absorbance concentrations than road length. By incorporating road area and other new local variables, the performance of each PM LUR model was improved. The results suggest that road area is a better indicator of traffic intensity rather than road length in a city with high density of road network and traffic.

Abbreviations: LUR, land use regression; ESCAPE, European Study of Cohorts for Air Pollution Effect; PM, particulate matter; SSR, Smoke Stain Reflectometer; GIS, Geographic Information System; VIF, Variance Inflation Factor; LOOCV, leave-one-out cross-validation; RMSE, root mean square error; IQR, interquartile range.

* Corresponding author at: Institute of Occupational Medicine and Industrial Hygiene, College of Public Health, National Taiwan University, 7F., No. 17, Xuzhou Rd., Zhongzheng Dist., Taipei City 100, Taiwan (R.O.C.).
E-mail address: ccchan@ntu.edu.tw (C.-C. Chan).

http://dx.doi.org/10.1016/j.scitotenv.2015.01.091
0048-9697/

1. Introduction

Land use regression (LUR) has been widely used and has rapidly become an important approach to predict long-term average pollutant concentration at an intra-urban scale. Recently, the European Study of Cohorts for Air Pollution Effects (ESCAPE) project described the development and performance of LUR models of 20 European study areas for particulate matter (PM) (Eeftens et al., 2012a). Based on the exposures estimated from the LUR models, this project illustrated an association between mortality and average annual exposure to fine particles (Beelen et al., 2014). Several studies have also developed LUR models for assessing intra-urban contrast of PM in North America (Henderson et al., 2007; Moore et al., 2007; Ross et al., 2007). The above-mentioned European and North American LUR models yielded a predictive capacity (as R^2) ranging between 35% and 94% for $PM_{2.5}$, 39% and 97% for $PM_{2.5}$ absorbance (i.e., soot), 50% and 90% for PM10, and 32% and 81% for PM_{coarse} ($PM_{10-2.5}$). To estimate the concentrations of air pollutants at any point in the area, air pollution data and predictor variables are required. The selection of variables hence plays an important role in developing LUR models. Most study areas in Europe and North America, such as the Netherlands, Munich, London, Los Angeles, and New York City, share the same characteristics. However, cities in rapidly developing countries, such as Taipei, Taiwan, usually have different land use attributes from those in developed countries. One of such variable is the traffic-related predictor of road length, which was suitably used in several LUR studies (Hoek et al., 2008) as a good indicator of traffic intensity in European and North American cities, but may not be good enough in traffic-jammed cities like Taipei. Under such circumstances, road area can be an alternative to road length as a better land use variable for traffic intensity in LUR models in Asian cities with crowded roads. To our knowledge, no research has yet focused on the role of road area as a surrogate for traffic intensity when traffic flow data was not available. Elevated highways, which have been commonly constructed to increase road capacity across the city and soothe the traffic jams in city centers in Asian cities, were seldom considered in previous LUR models. Intra-city variation in air pollution can be better described by including elevated highways in the LUR models as one previous study reported that overpass structures could affect the distribution of air pollutants (Tong et al., 2011).

The continuous construction of buildings and infrastructures to meet residents' living demands is another common feature in rapidly urbanized cities. Previous LUR models have overlooked spatial variation in particulate air pollution arising from fugitive emissions of such construction activities, which are even more important in Asian cities. The mix of residential, commercial, and industrial areas in one small city district is not uncommon in Asian cities, where zoning policies are not as strictly enforced as in most European and North American cities. The individual contribution of residential, commercial, and industrial areas to spatial variation in PM needs to be adequately considered when they are included in LUR models.

Rivers and green lands in cities, by contrast, can alleviate air pollution in nearby areas by restricting or diluting pollution emissions. River, a component of the green land variable in the ESCAPE LUR models, can also be treated as an independent variable as it can cover a larger area in some urban environments and its air pollution dilution effects can be different, compared to urban green land.

This study aimed at developing LUR models by following the ESCAPE modeling procedures to characterize the spatial distribution of $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}. We focused on the role of specific variables that are important and are associated with emissions of PM in the Taipei metropolis but which have rarely been used in European and North American LUR studies. Further, we investigated whether the performance of PM LUR models in Taipei can be equivalent to those of European cities in the ESCAPE project by incorporating specific new variables. Results of this study could provide a higher spatial resolution of estimation of exposure to PM for future health studies in Taipei, Taiwan.

2. Materials and methods

2.1. Study area

The Taipei metropolis lies north of Taiwan Island with the total size of the study area accounting for about 800 km^2 and with a population of 6 million. It falls within the Tamsui River Basin, which is considered as the heartland of Taiwan (the detailed description of current traffic and land use status is presented in Appendix A). Table 1 shows the land use characteristics of Taipei in 2010. The inhabitable area is only about half of the study domain area, while other areas are conservation areas including forest, mountain and natural areas. Road area is the largest land cover area in the inhabitable part of the study domain, followed by residential land-use, river area and industrial land-use. River area, industrial, and commercial land-use are more than a quarter of residential land-use. Increased industrialization and construction of buildings could be the main reason of the urbanization of Taipei. Since over 30% of the population of Taiwan resides in a region covering less than 7% of the total land area of Taiwan (Huang and Chan, 2014), the high pressure of population growth and urban expansion results in a massive development of primary transport zones and peri-urban areas, which have become one of the main areas of urban development in Taipei.

2.2. Measurement program

Three major pollutants were measured in the study, ambient PM_{10} mass, $PM_{2.5}$ mass and $PM_{2.5}$ absorbance. The measurement and sampling site selection were according to ESCAPE SOPs (http://www.escapeproject.eu/manuals) and have been described elsewhere (Eeftens et al., 2012b). At each site, PM samples were collected over a two-week period during three seasons: cold (January to March, 2010), warm (June to August, 2010) and one season with intermediate temperatures (October to December, 2009) under the assumption of stable spatial contrast between sampling sites. The sampling sites were chosen with a purpose of maximizing the contract in variables. We selected 9 street sites and 11 urban background sites (Fig. 1). Urban background sites were defined as sites with no more than 10,000 vehicles per day typically passing within a 50-meter radius. Street sites are urban sites in a major road with more than 10,000 vehicles per day and no nearby air pollution sources other than traffic. At one additional background reference site, PM_{10} mass, $PM_{2.5}$ mass and $PM_{2.5}$ absorbance were continuously measured over the entire year with the same instruments (October 2009 to September 2010). The discontinuous site-specific measurements were then adjusted to long-term averages over the observation period (Hoek et al., 2002) using the data from the background reference site. Such approach relies on the assumptions that the spatial pattern of air pollutant across the study area is stable over time, and the continuous sampling site is representative of the temporal series.

PM_{10} and $PM_{2.5}$ were measured by Harvard Impactors (Air Diagnostics and Engineering Inc., Naples, ME, USA). PM_{coarse} was calculated as the difference between PM_{10} and $PM_{2.5}$. The reflectance of all $PM_{2.5}$ filters was measured using a Smoke Stain Reflectometer (SSR, Model 43 (M43D), Diffusion Systems, Ltd., London, UK). The reflectance measurements correlated with actual measurements of elemental carbon in previous studies (Janssen et al., 2001; Kinney et al., 2000) and were considered markers for traffic emissions ('diesel soot') in many areas. Absorbance, as defined by ISO 9835, was calculated from the reflectance. Reflectance measurements were performed after the filters were weighed. Because most elemental carbon is associated with sub-micrometer particles, the reflectance of $PM_{2.5}$ and PM_{10} filters was shown to be nearly equal and highly correlated. Therefore, reflectance measurements were limited to $PM_{2.5}$ filters.

Table 1
Land use characteristics of Taipei in 2010.

	Area (km^2)	%	Relative to residential land-use (%)
Study domain	807	–	–
Conservation area	444	–	–
Inhabitable area	363	100	–
River	23.3	6	43
Residential land-use	54.0	15	100
Commercial land-use	14.1	4	26
Industrial land-use	16.5	5	31
Road (major road)	64.0 (10.7)	18 (3)	118 (20)
Transportation facility	8.9	2	17
	Length (km)	km per km^2	Relative to residential land-use (km/km^2)
Elevated highway	617	1.70	11.4

2.3. Geographic information datasets

Potential predictor variables were derived from Geographic Information System (GIS) datasets. Digital road networks, land use data, population and household density data, and altitude data, which are in line with the predictor data used in the ESCAPE LUR models in Europe (Eeftens et al., 2012a) with some modification to fit local geological features of Taipei, were used. A detailed description of all predictor variables, and the a priori choices made in the design for LUR model development including the buffer sizes, is shown in Table A. 1. In addition to those common factors, characteristics of land use and transportation in Taipei were considered and some predictor variables not found in the ESCAPE study were included in the model to evaluate the applicability or adaptability of the LUR methods developed in Europe to Asia. Those additional predictor variables are various traffic representations such as the length of and the nearest distance to the elevated highway, as well as the probably more applicable surrogate for traffic flow, the road area. Other important predictor variables include commercial land, construction land, transportation facility, and the area of and the nearest distance to the river. GIS analyses were conducted to calculate the predictor variables for each site, using the site coordinates and digital data sets within a GIS system (ESRI ArcGIS v.10.0). The buffer sizes are derived from dispersion patterns; for traffic variables, circular buffers with radii of 25, 50, 100, 300, 500, and 1000 m around each site were calculated. For land use and population, buffers of 100, 300, 500, 1000, and 5000 m were calculated.

The traffic representations of Taipei were obtained from a digital map of traffic networks in Taiwan, which was completed by the Institute of Traffic and Transportation in 2010. This dataset was used to calculate the total length of all roads, major roads, and elevated highway, as well as the nearest distance to those roads. Population and household data for 2010 were collected from the Household Registration Office of each district in Taipei. The topographic altitude map was obtained from the City Offices. Land use information was taken from the Land

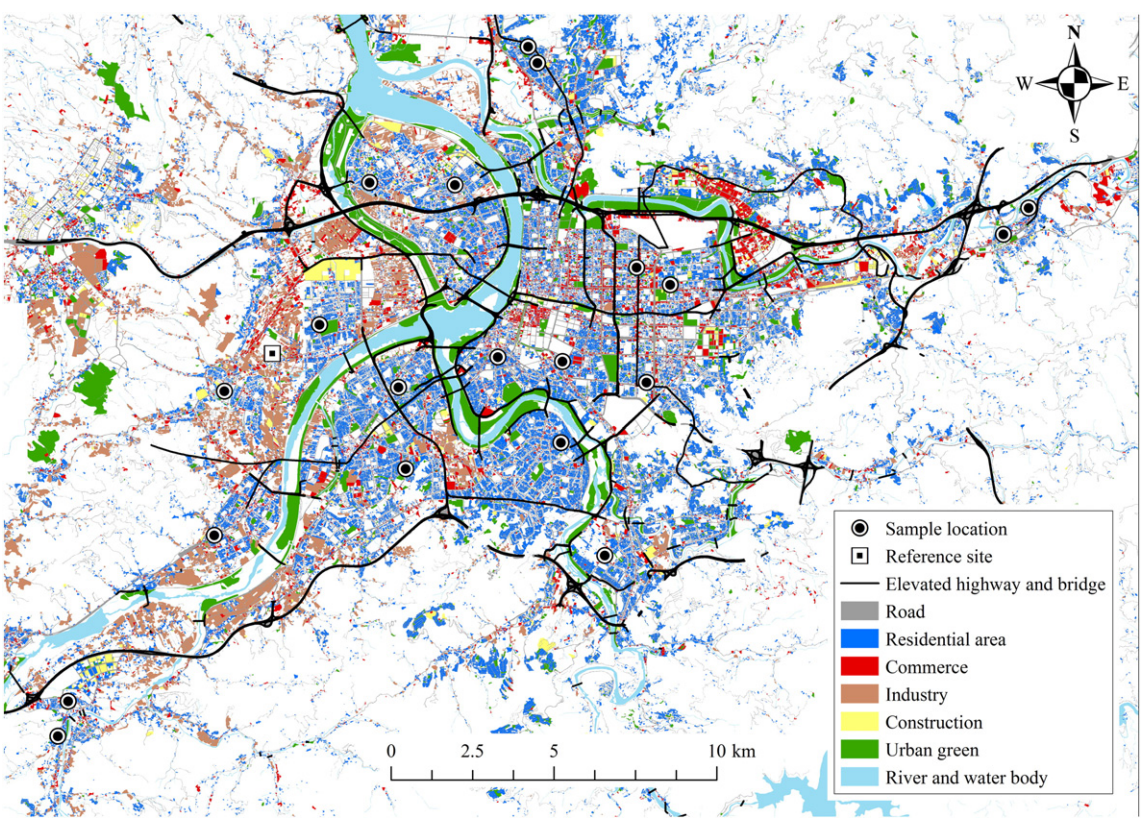

Fig. 1. Map showing the 20 sample sites, 1 background reference site for adjustment of the temporal variation, roads, elevated highways, and land use categories in Taipei.

Use Investigation of Taiwan conducted by the National Land Surveying and Mapping Center in 2007. The land use dataset includes a category of road area which was used to calculate the area of all roads and major roads (consisted of the following class of road: national highway, provincial highway, and expressway). Since traffic flow data was not available in Taipei, road area data served as a surrogate for this traffic indicator. It should be noted that the land use data from 2007 included road area data that were used as surrogates for road area conditions within the study's time of 2009–10. It should also be noted that land use changes between those years may induce bias using these variables in the development of models. The 2007 land use dataset were the most recent available data at the time.

2.4. LUR model development and evaluation

Models were developed for PM_{10}, PM_{coarse}, $PM_{2.5}$, and $PM_{2.5}$ absorbance in Taipei with the statistical software (SAS v.9.2) for regression analysis. Predictor variables were regressed against measured concentration of each pollutant. The same procedures as used in ESCAPE (available on www.escapeproject.eu) were followed for model development and validation. In the first step of the model development, univariate regression analyses were conducted for all possible predictor variables via regression of each predictor variable against monitored PM concentrations. The model with the largest adjusted explained variance (adjusted R^2) was regarded as the 'start model'. The remaining variables were then added separately to the 'start model'. The predictor variable with the largest additional increase in adjusted R^2 and the right direction of effect was maintained in the model. In this supervised stepwise process, variables were added until no predictor variables added more than 1% to the adjusted R^2 of the previous regression model, which resulted in an 'intermediate model'.

As a final step, variables with p value >0.10 were sequentially removed from the model. In the 'final model', all selected variables have a p value <0.10. Subsequently we checked the Variance Inflation Factor (VIF) and excluded those variables with VIF >3 and re-evaluated the model. The COOK's D statistic was used to detect influential observations. By excluding observations with high COOK's D (i.e., >1), we assessed the extent of changes in model coefficients. If large changes in the coefficient of a specific variable were observed, the model was re-developed without offering this variable.

We used leave-one-out cross-validation (LOOCV) to evaluate overall model performance. The R^2 and root mean square error (RMSE) were used to verify the fit for the final LUR models. In addition, Moran's I, a statistic of spatial autocorrelation, was calculated to test whether the assumption of independence for the residuals was violated.

3. Results

3.1. PM measurements

Table 2 summarizes the PM_{10}, and PM_{coarse}, $PM_{2.5}$ concentrations and the $PM_{2.5}$ absorbance measured from 2009–2010 at 9 street sites and 11 urban background sites in Taipei. The annual averages of $PM_{2.5}$, PM_{10}, and PM_{coarse} mass concentrations for the 20 sampling sites were 26.0, 48.6 and 23.3 μg/m^3, respectively, and the absorption coefficient of $PM_{2.5}$ ($PM_{2.5}$ absorbance) was 2.0×10^{-5} m^{-1}. As expected, the PM concentrations measured at street sites were higher than those at urban background sites. The ratios of average concentrations at street sites over urban background sites were 1.30 for $PM_{2.5}$ absorbance, 1.24 for $PM_{2.5}$ mass, 1.14 for PM_{10} mass and 1.11 for PM_{coarse} mass. The range/mean ratios indicated that the spatial contrasts in $PM_{2.5}$ mass (89%) and $PM_{2.5}$ absorbance (72%) were higher than those for PM_{10} mass (51%) and PM_{coarse} (54%).

3.2. Predictor variables in models

Table 3 summarizes the local predictor variables selected in the final LUR models in specific buffer zones for the 20 sampling sites. As mentioned above, Taipei has a dense road network. The mean road area of 25 m and 50 m buffer zones was about 550 m^2 and 2000 m^2, respectively. There was on average 4 km of elevated highway or bridge in a 1000 m buffer zone. The mean of the nearest distance to major road and elevated highway or bridge were less than 60 m and 200 m, respectively. The mean area of transportation facility in a small buffer zone (100 m) was more than 1500 m^2. Multiple types of land use also coexisted within a small (100 m) to median (500 m) buffer zone. The mean area of commercial use within 100 m and industrial use within 500 m was more than 900 m^2 and 12,000 m^2, respectively. A considerable area of construction land was observed with an average of more than 250 m^2 within a 100 m buffer zone.

3.3. Land use regression modeling

Table 4 presents the four models derived from using predictor variables defined in the ESCAPE project with some modification to fit local features of Taipei for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}. Regression coefficients (β), the absolute contribution (β ∗ interquartile range (IQR, the difference between the first quartile and third quartile)), significance (p-value), incremental values of adjusted R^2, and partial R^2 of each variable to predicted concentrations, and the performance of each model (in terms of R^2, root mean squared error (RMSE), and

Table 2
Comparisons of $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse} concentrations between street sites and urban background sites in Taipei, 2009–2010.

Site type (n)	$PM_{2.5}$ mass (μg/m^3)			$PM_{2.5}$ absorbance (10^{-5}/m)			PM_{10} mass (μg/m^3)			PM_{coarse} (μg/m^3)		
	Mean (SD)	Range	Range/mean (%)	Mean (SD)	Range	Range/mean (%)	Mean (SD)	Range	Range/mean (%)	Mean (SD)	Range	Range/mean (%)
All (20)	26.0 (5.6)	17.4–40.6	89%	2.0 (0.4)	1.2–2.6	72%	48.6 (5.9)	39.2–64.0	51%	23.3 (3.1)	18.6–31.3	54%
ST[a] (09)	29.1 (6.3)	22.2–40.6	63%	2.2 (0.3)	1.8–2.6	36%	52.2 (5.7)	45.5–64.0	35%	24.7 (3.7)	19.4–31.3	48%
UB[a] (11)	23.5 (3.3)	17.4–27.4	43%	1.7 (0.2)	1.2–2.2	58%	45.7 (4.2)	39.2–51.7	27%	22.2 (2.1)	18.6–25.2	30%
Ratio ST/UB	1.24[b]			1.3[b]			1.14[b]			1.11[c]		

[a] ST = street site and UB = urban background.
[b] Significant difference between the site types on p < 0.10 level.
[c] Significant on p < 0.05 level.

Table 3
Localized predictor variables for the 20 measurement sites used in the LUR models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}.

Variables	Mean (SD)	Minimum	Median	Maximum
Road area (m^2)				
25 m	547.9 (311.1)	0.0	586.1	1195.9
50 m	2021.8 (1092.8)	0.0	1708.0	4283.6
Length of elevated highway (m)				
1000 m	4052.2 (3757.0)	0.0	3689.0	14,840.9
Commerce (m^2)				
100 m	958.5 (944.8)	0.0	622.8	3097.3
1000 m	140,794.0 (81,978.9)	60,480.7	107,823.3	360,547.0
Construction (m^2)				
100 m	272.3 (868.9)	0.0	0.0	3666.0
Transportation facility (m^2)				
100 m	1592.2 (2040.8)	0.0	433.3	5574.1
500 m	25,388.9 (12,427.8)	3334.7	23,824.7	46,987.8
River (m^2)				
5000 m	3,969,048.5 (2,350,761.7)	1,609,455.2	3,384,582.1	10,242,119.0
Industry (m^2)				
500 m	12,194.4 (23,031.7)	0.0	1216.2	85,892.9
5000 m	3,241,427.6 (2,721,945.1)	222,954.5	192.8582.7	8,251,340.1
Residential area (m^2)				
1000 m	611,380.0 (165,629.4)	327,519.3	631,573.0	943,122.2
Inverse distance to the nearest major road (m^{-1})	0.017 (0.044)	8.847×10^{-4}	4.524×10^{-3}	0.206
Inverse distance squared to the nearest elevated road (m^{-2})	3.064×10^{-5} (7.137×10^{-5})	2.246×10^{-7}	2.272×10^{-6}	3.123×10^{-4}

those obtained from leave-one-out cross-validation) were included in this table.

The derived LUR models in Taipei yielded R^2 values of 95%, 96%, 87%, and 65% for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}, respectively. The final models included four to six variables. In $PM_{2.5}$ model, the first selected variable was road area in a 50 m buffer zone explaining 56% of variability in $PM_{2.5}$ concentrations. The other five variables together yielded 37% of additional explained variability. For $PM_{2.5}$ absorbance, the first three selected variables (area of transportation facilities between 100 m and 500 m buffer zone, road area in 25 m buffer zone, and area of transportation facilities in 100 m buffer zone) yielded 77% of explained variability. In the PM_{10} model, as was observed for $PM_{2.5}$, road area in a 50 m buffer zone was selected as the first variable, explaining 59% of variability. With the other three variables, an additional explained variability of 25% was obtained. For PM_{coarse}, industrial land-use in a 1000 m buffer zone coupled with the local variables (inverse distance squared to the nearest elevated road and land of commercial use in 1000 m buffer zone) yielded 59% of explained variability.

With regard to the LOOCV, the four LUR models yielded R^2 values of 91%, 92%, 74%, and 52% for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}, respectively. The changes in R^2 between model development and LOOCV R^2 were 4%, 4%, 13%, and 13% for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}, respectively, indicating that the derived LUR models were stable.

Models derived by using variables listed in ESCAPE study only are shown in Table A. 2. Consistently, length of major road was the first variable selected in all models but only yielding 29%, 42%, 53%, and 39% of explained variability for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}

Table 4
LUR model results for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse} using predictor variables based on 20 measurement sites in Taipei.

Model	Variable	β	β ∗ IQR[a]	p-Value	Adj R^2	Partial R^2	Global statistics
$PM_{2.5}$	(Intercept)	13.81		<0.001			$R^2 = 95\%$
	Area of roads (50 m)	2.23×10^{-3}	2.54	<0.001	56%	58%	RMSE = 1.51 μg/m^3
	Inverse distance to the nearest major road	72.07	0.55	<0.001	70%	16%	LOOCV $R^2 = 91\%$
	Industry (500 m)	9.89×10^{-5}	1.45	<0.001	78%	7%	LOOCV RMSE = 1.75 μg/m^3
	Construction (100 m)	2.22×10^{-3}	0.00	<0.001	83%	6%	
	Residential area (1000 m)	1.13×10^{-5}	2.44	<0.001	89%	5%	
	River (5000 m)	-5.62×10^{-7}	−1.79	0.015	93%	3%	
$PM_{2.5}$ absorbance	(Intercept)	1.07		<0.001			$R^2 = 96\%$
	Transportation (500–100 m)	5.76×10^{-6}	4.56×10^{-2}	0.023	17%	22%	RMSE = 0.09 10^{-5}/m
	Area of roads (25 m)	8.69×10^{-4}	3.23×10^{-2}	<0.001	76%	56%	LOOCV $R^2 = 92\%$
	Transportation (100 m)	3.75×10^{-5}	3.11×10^{-2}	0.007	77%	3%	LOOCV RMSE = 0.11 10^{-5}/m
	Industry (500 m)	4.40×10^{-6}	1.40×10^{-2}	0.001	84%	7%	
	Length of elevated roads (1000 m)	2.17×10^{-5}	3.78×10^{-2}	0.003	91%	5%	
	Commerce (100 m)	6.93×10^{-5}	4.09×10^{-2}	0.016	94%	3%	
PM_{10}	(Intercept)	34.41		<0.001			$R^2 = 87\%$
	Area of roads (50 m)	3.41×10^{-3}	3.89	<0.001	59%	61%	RMSE = 2.36 μg/m^3
	Industry (5000 m)	1.10×10^{-6}	5.18	<0.001	73%	15%	LOOCV $R^2 = 74\%$
	Commerce (1000 m)	2.46×10^{-5}	1.91	0.006	81%	8%	LOOCV RMSE = 3.05 μg/m^3
	Construction (100 m)	1.16×10^{-3}	0.00	0.078	84%	3%	
PM_{coarse}	(Intercept)	13.88		<0.001			$R^2 = 65\%$
	Industry (5000 m)	1.20×10^{-6}	5.66	<0.001	8%	13%	RMSE = 2.83 μg/m^3
	Inverse distance squared to the nearest elevated road	2.23×10^{4}	0.26	0.055	39%	32%	LOOCV $R^2 = 52\%$
	Commerce (1000 m)	2.97×10^{-5}	2.30	0.008	59%	20%	LOOCV RMSE = 3.15 μg/m^3

[a] β ∗ IQR is the β coefficient multiplied by the inter-quartile range for the given parameter at 20 measurement sites.

respectively in the incremental value of adjusted R^2 (data not shown). The model R^2 values were 38%, 74%, 86%, and 43% for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}, respectively, and the corresponding validation R^2 were 14%, 63%, 69%, and 25%.

4. Discussion

Our study has developed LUR models for the Taipei study area by following the same model development and validation procedures as applied in the European ESCAPE project. With extra-potential localized predictor variables, including road area, industry, commerce, construction, transportation facility, river, length of elevated highway and major road, and proximity to road, moderate to great explained variance was obtained for PM_{10}, PM_{coarse}, $PM_{2.5}$, and $PM_{2.5}$ absorbance. Explained variance of the LUR models was highest for $PM_{2.5}$ absorbance and $PM_{2.5}$, followed by PM_{10}, and was lowest for PM_{coarse}. Variables which represent local sources of traffic and industrial emissions were selected in LUR models and contributed significantly in the model performance, especially for $PM_{2.5}$ and $PM_{2.5}$ absorbance. Compared to models derived from using common variables specified in the ESCAPE study, our models gained an additional 57%, 22%, 1%, and 22% of explained variance in measured concentrations for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}, respectively.

4.1. Spatial contrast of PM concentrations

Overall contrast (total range/mean) indicated that among all measurement sites the spatial contrasts in $PM_{2.5}$ absorbance and $PM_{2.5}$ were higher than those for PM_{10} and PM_{coarse}. This relationship was also observed in terms of model performance (R^2), indicating that spatial contrasts dominated the performance of LUR models for the three size fractions of particulate matter and $PM_{2.5}$ absorbance. The contrasts between urban background and street sites were also higher for $PM_{2.5}$ absorbance and $PM_{2.5}$ than for PM_{10} and PM_{coarse} (Table 2). The ratio between street and urban background sites was greater than 1 for all the four pollutants, suggesting that traffic emissions have significant impacts on the pollution concentrations. This is also reflected in that traffic emissions have more impact in $PM_{2.5}$ absorbance and $PM_{2.5}$ than PM_{10} and PM_{coarse}.

4.2. Model performance across particle fractions

The performance of LUR models for $PM_{2.5}$ and $PM_{2.5}$ absorbance was higher than PM_{10} and PM_{coarse}. Six variables are included in the former two models. Among these variables, traffic emission is consistently the first variable selected into the model; industry emission is the second most important variable in the models; other variables differ from model to model. Road area within a buffer size of 50 m and area of transportation facilities between buffers 500 m and 100 m were the first selected variables in the model for $PM_{2.5}$ and $PM_{2.5}$ absorbance, respectively. They gave the largest incremental value of adjusted R^2 and hence represented the major sources of each air pollutant. The variables added next, i.e., inverse distance to the nearest major road for the $PM_{2.5}$ model; road area within a buffer size of 25 m, and area of transportation facilities within a buffer size of 100 m for the $PM_{2.5}$ absorbance model, were also related to traffic emissions.

The road area variable provided substantial improvement for the LUR $PM_{2.5}$ and $PM_{2.5}$ absorbance models (Tables 4 and A. 2), which can be explained by its better representativeness of traffic emissions and higher spatial variability than the road length variable. As shown in Table A. 3, the coefficient of variation (CV) for road area within 50 m buffer was larger than that for road length at all sites; the spatial variability of road area was also higher than that of road length at street sites.

Various traffic variables selected in the two models reflect that $PM_{2.5}$ and $PM_{2.5}$ absorbance were generated from different traffic emissions. Road area and proximity of major road in the $PM_{2.5}$ model represent emissions from on-road vehicles, usually private cars or motorcycles, whereas area of transportation facilities selected in the $PM_{2.5}$ absorbance model (e.g., bus terminal, station, and stop) represents emissions from buses that contributed to the soot component of $PM_{2.5}$. In Taiwan, a substantial fraction of buses and long-distance coaches use diesel. Industrial land-use, selected in both $PM_{2.5}$ and $PM_{2.5}$ absorbance models with exactly the same buffer size, represents contributions from industrial processes in Taipei, such as combustion or the movement of large quantities of goods. Their impact can be observed up to a range of 500 m.

In the $PM_{2.5}$ model, the first included variable explained a large part of the spatial variability (56%) in $PM_{2.5}$ concentrations. This was an improvement compared to the 29% of variability explained by the length of road which was the first variable selected in the model derived from using common variables defined in the ESCAPE study. The other two local variables (i.e., construction and river area) also provided a greater additional explained variability of 9%, in contrast to that of 3% by industrial land when only considering the common variables used in the ESCAPE study. For $PM_{2.5}$ absorbance, the first three local variables yielded 77% explained variability, while all the four variables in the model which were derived from variables used in the ESCAPE (two variables of length of major road with different buffer size, industrial land, and urban green) explained only 67%.

In addition to traffic and industrial emission factors, construction area along with residential land-use and river area was included in the $PM_{2.5}$ model. Some specific construction activities (e.g., combustion of fossil fuels) have been linked to $PM_{2.5}$ emissions (Ketchman and Bilec, 2013; Muleski et al., 2005). The 100-m buffer zone reflects the small spatial scale of impact from construction activities on $PM_{2.5}$ concentration. In a residential region, many household activities generate substantial air pollution. For example, cooking and incense burning were significant contributors to indoor particle levels in Taiwan (Liao et al., 2006) and consequently may also be a major source of ambient air pollution. The river area is the only variable selected that has a negative correlation with the concentrations of traffic-related air pollutants. The dilution effect of rivers is different from green spaces in Taipei because they have different distribution patterns, coverage, and extent. It only contributes to the $PM_{2.5}$ model with a large scale influence of a 5000 m buffer zone. The possible explanation is that rivers can dilute fine particulates among a large area, while most PM_{10} and PM_{coarse} have settled down within that distance.

The length of elevated roads and commercial land-use were factors for explaining the variance of $PM_{2.5}$ absorbance besides traffic and industrial emission factors. Many near-road air quality studies have characterized open highway conditions (Baldauf et al., 2013; Zhu et al., 2002; Zwack et al., 2011) and highway-building environment (Tong et al., 2011), and highlighted the need to account for the complexity of urban features in estimating near-road population exposures to traffic emissions. The observed effects of elevated highway within a 1000 m buffer radius on $PM_{2.5}$ absorbance (surrogate of soot component or black carbon) are probably because the carbonaceous species of PM is mostly associated with smaller particles from diesel emission and transported over longer distances. A similar effect of elevated highway was also found in the PM_{coarse} model. The spatial distribution of PM_{coarse} concentration was associated with the inverse distance squared to the nearest elevated highway. The importance of including elevated highway as a local variable was shown by comparing model performance with and without this variable. The LUR models improved 7% and 31% of explained variability (incremental adjusted R^2) for the concentration of $PM_{2.5}$ absorbance and PM_{coarse}, respectively by including the elevated highway variable in the models. Additionally, activities in commerce, such as emissions from moving cars and heavy goods vehicles as well as idling emissions due to traffic congestion in commercial areas, have influences on $PM_{2.5}$ absorbance.

Variables selected in PM_{10} model were similar to those in $PM_{2.5}$ model, but it performed less well. The lower spatial contrast of PM_{10} than $PM_{2.5}$ in Taipei is the possible explanation for this relationship

which was mentioned in previous section (Section 4.1). It also reflects that the linear relationship between PM_{10} concentration and major emission sources was not as high as for $PM_{2.5}$ due to its coarser size. Traffic and industrial emissions dominate the explained variance with the incremental value of adjusted R^2 of 73% which is slightly lower than that in the $PM_{2.5}$ model (78%). It is likely that combustion-related PM emission from vehicle tailpipes and factory stacks contributes more to fine particles than to coarse particles in the study area. There was no robust tendency about model performance between $PM_{2.5}$ and PM_{10} observed across European study areas (Eeftens et al., 2012a). The road area variable which was firstly selected in the PM_{10} model yielded a slightly better explained variability of 59% than that of 53% by length of major road in the model derived from common ESCAPE variables. Other local variables in the PM_{10} model (commercial land and construction area) together yielded an additional explained variability of 11%, also better than that of 4% by residential land in the model using ESCAPE variables. The performance of the PM_{coarse} model was the lowest and no traffic variables were selected, also due to the size effect. Local variables yielded a better explained variability (59%) in PM_{coarse} concentrations, in contrast to that of 39% provided by the only one variable in the model derived from ESCAPE variables. As a result, the PM_{coarse} model explained the spatial variability in measured concentrations moderately similar to the results of the ESCAPE project in 20 European study areas (Eeftens et al., 2012a). The possible reason is the lack of sufficient data representing local sources.

4.3. Limitations

In the present study, models for $PM_{2.5}$ and $PM_{2.5}$ absorbance have six predictor variables based on 20 measurement sites and both yield almost 100% R^2 values. There is a potential problem of model overfitting arising from over-specified variables with limited number of measurement sites. A rule of thumb to avoid the overfitting problem is to constrain the number of observations per predictor variable between 10 and 15 in the final model (Babyak, 2004). This is the case for most $PM_{2.5}$ and $PM_{2.5}$ absorbance models in the ESCAPE project, which usually include three to five variables in the 20 European study areas. By applying the same constraints to construct LUR models with four variables for $PM_{2.5}$ and $PM_{2.5}$ absorbance, the R^2 values of the constrained models became 87% and 88% for $PM_{2.5}$ and $PM_{2.5}$ absorbance, respectively, which were 8% lower than that of unconstrained models. The present study uses the same study design as the ESCAPE project. Limitations of our approach, such as the temporal restriction of the sampling campaign, the gap between modeled individual concentrations and personal exposure, and the temporal representativeness of traffic and land use data, have also been described and discussed in detail elsewhere (Eeftens et al., 2012a).

5. Conclusions

In conclusion, the ESCAPE LUR modeling approach can be applied to Asian cities with high densities of roads and significant industrial, commerce and construction activities to develop LUR models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}, and PM_{coarse}. This study has shown that incorporating local variables improves the performance of LUR models in Taipei. The introducing of road area data as a traffic variable has gained the greatest advancement of model performance for $PM_{2.5}$, $PM_{2.5}$ absorbance, and PM_{10} when the traffic intensity data were not available. It provides an alternative traffic indicator to represent intra-urban variation in PM for other similar Asian cities. PM LUR models developed from the present study are being used to estimate ambient concentrations, giving better performance in explaining spatial variation than other modeling methods, at the home address of participants in future cohort studies in Taipei.

Funding source

This research was supported by the project NSC97-2923-I002-001-MY4 of the National Science Council of Taiwan.

Financial disclosure

Authors have no financial relationships relevant to this article to disclose.

Conflict of interest

Authors have no conflicts of interest to disclose.

Acknowledgments

This research was supported by the project NSC97-2923-I002-001-MY4 of the National Science Council of Taiwan. The results of this research have received scientific contribution from the European Community's Seventh Framework Program (FP/2007-2011) under the following grant agreement number: 211250.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.scitotenv.2015.01.091.

References

Babyak, M.A., 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosom. Med. 66, 411–421.
Baldauf, R.W., Heist, D., Isakov, V., Perry, S., Hagler, G.S.W., Kimbrough, S., et al., 2013. Air quality variability near a highway in a complex urban environment. Atmos. Environ. 64, 169–178.
Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z.J., Weinmayr, G., Hoffmann, B., et al., 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.
Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., et al., 2012a. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM_{coarse} in 20 European study areas; results of the ESCAPE project. Environ. Sci. Technol. 46, 11195–11205.
Eeftens, M., Tsai, M.Y., Ampe, C., Anwander, B., Beelen, R., Bellander, T., et al., 2012b. Spatial variation of $PM_{2.5}$, PM_{10}, $PM_{2.5}$ absorbance and PM_{coarse} concentrations between and within 20 European study areas and the relationship with NO_2 — results of the ESCAPE project. Atmos. Environ. 62, 303–317.
Henderson, S.B., Beckerman, B., Jerrett, M., Brauer, M., 2007. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. Environ. Sci. Technol. 41, 2422–2428.
Hoek, G., Brunekreef, B., Goldbohm, S., Fischer, P., van den Brandt, P.A., 2002. Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. Lancet 360, 1203–1209.
Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., et al., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42, 7561–7578.
Huang, K.C., Chan, S.L., 2014. A study on the land-cover change indicators of Taipei metropolitan areas. IOP Conf. Ser.: Earth Environ. Sci. 18, 012105.
Janssen, N.A.H., van Vliet, P.H.N., Aarts, F., Harssema, H., Brunekreef, B., 2001. Assessment of exposure to traffic related air pollution of children attending schools near motorways. Atmos. Environ. 35, 3875–3884.
Ketchman, K., Bilec, M., 2013. Quantification of particulate matter from commercial building excavation activities using life-cycle approach. J. Constr. Eng. Manag. 139, A4013007.
Kinney, P.L., Aggarwal, M., Northridge, M.E., Janssen, N.A.H., Shepard, P., 2000. Airborne concentrations of $PM_{2.5}$ and diesel exhaust particles on Harlem sidewalks: a community-based pilot study. Environ. Health Perspect. 108, 213–218.
Liao, C.M., Chen, S.C., Chen, J.W., Liang, H.M., 2006. Contributions of Chinese-style cooking and incense burning to personal exposure and residential PM concentrations in Taiwan region. Sci. Total Environ. 358, 72–84.
Moore, D.K., Jerrett, M., Mack, W.J., Kunzli, N., 2007. A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA. J. Environ. Monit. 9, 246–252.
Muleski, G.E., Cowherd Jr., C., Kinsey, J.S., 2005. Particulate emissions from construction activities. J. Air Waste Manag. Assoc. 55, 772–783.
Ross, Z., Jerrett, M., Ito, K., Tempalski, B., Thurston, G.D., 2007. A land use regression for predicting fine particulate matter concentrations in the New York City region. Atmos. Environ. 41, 2255–2269.
Tong, Z., Wang, Y.J., Patel, M., Kinney, P., Chrillrud, S., Zhang, K.M., 2011. Modeling spatial variations of black carbon particles in an urban highway-building environment. Environ. Sci. Technol. 46, 312–319.
Zhu, Y., Hinds, W.C., Kim, S., Shen, S., Sioutas, C., 2002. Study of ultrafine particles near a major highway with heavy-duty diesel traffic. Atmos. Environ. 36, 4323–4335.
Zwack, L.M., Paciorek, C.J., Spengler, J.D., Levy, J.I., 2011. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. Environ. Health Perspect. 119, 852–859.