Rahman, Md Mahmudur, Yeganeh, Bijan, Clifford, Samuel, Knibbs, Luke, & Morawska, Lidia
(2017)
Development of a land use regression model for daily NO2 and NOx concentrations in the Brisbane metropolitan area, Australia.
Environmental Modelling and Software, *95*, pp. 168-179.

https://doi.org/10.1016/j.envsoft.2017.06.029

Development of a land use regression model for daily NO_2 and NO_x concentrations in the Brisbane metropolitan area, Australia

Md Mahmudur Rahman[†], Bijan Yeganeh[†], Sam Clifford[†#], Luke D. Knibbs[‡], and Lidia Morawska[†]*

[†]International Laboratory for Air Quality and Health, Institute of Health and Biomedical Innovation, Queensland University of Technology, GPO Box 2434, Brisbane QLD, 4001, Australia

[#]ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology, GPO Box 2434, Brisbane QLD, 4001, Australia

[‡]School of Public Health, The University of Queensland, Herston, QLD, 4006

Abstract

Land use regression models are an established method for estimating spatial variability in gaseous pollutant levels across urban areas. Existing LUR models have been developed to predict annual average concentrations of airborne pollutants. None of those models have been developed to predict daily average concentrations, which are useful in health studies focused on the acute impacts of air pollution. In this study, we developed LUR models to predict daily NO_2 and NO_x concentrations during 2009−2012 in the Brisbane Metropolitan Area (BMA), Australia's third-largest city. The final models explained 64% and 70% of spatial variability in NO_2 and NO_x, respectively, with leave-one-out-cross-validation R^2 of 3−49% and 2−51%. Distance to major road and industrial area were the common predictor variables for both NO_2 and NO_x, suggesting an important role for road traffic and industrial emissions. The novel modeling approach adopted here can be applied in other urban locations in epidemiological studies.

Keywords: Air pollution, Land Use Regression (LUR) model, Nitrogen dioxide (NO_2), Oxides of nitrogen (NO_x), Urban area.

*Corresponding author: Lidia Morawska (Email: l.morawska@qut.edu.au ☏ 61-7-3138 2616)

1. Introduction

Exposure to ambient air pollution is an important public health concern. A variety of diseases such as lower respiratory infection, lung cancer, chronic obstructive pulmonary disease (COPD), stroke and ischaemic heart disease are linked to ambient air pollutants (WHO, 2014). In order to accurately assess people's exposure to air pollutants in epidemiological studies, it is important to capture the spatial variability in concentrations. Although air quality monitoring networks can capture large-scale spatial variability, their sparseness means that they may not capture variability across spatial areas of interest (e.g. cities) in epidemiological studies (Hoek et al., 2008).

To overcome this limitation, Land Use Regression (LUR) has been widely used to model spatial variability in air pollutants (Hoek et al., 2008). The LUR modeling method includes air quality monitoring data from fixed points along with geographical predictor variables (e.g., land use area, road length, and population density) to predict the pollutant concentration at unmeasured locations (Hoek et al., 2008). With the rapid development of geographic information system (GIS) to calculate spatial predictors, LUR has emerged as an efficient tool for modeling human exposure to ambient air pollutants (Beelen et al., 2013). LUR models have been applied successfully to predict various gaseous (e.g., NO_2 and NO_x) and particle (e.g., $PM_{2.5}$ and PM_{10}) pollutants, incorporating a number of predictor variables such as population density, land usage, and traffic characteristics in urban areas (Hoek et al., 2008; Johnson et al., 2010). The performance of LUR models for estimating NO_2 and NO_x concentrations has been found to be better than a number of GIS-based interpolation methods, such as kriging and inverse distance weighting, as most GIS interpolation methods create a smooth concentration surface without considering underlying land use information (Lee et al., 2014; Meng et al., 2015).

LUR models have been developed to estimate long-term exposure to NO_2 and NO_x in China (Meng et al., 2015), Korea (Kim and Guldmann, 2015), Taiwan (Lee et al., 2014), and Norway (Madsen et al., 2011), among others. In the European Study of Cohorts for Air Pollution Effects (ESCAPE) project, NO_2 and NO_x LUR models were developed and applied in 36 study areas across Europe (Beelen et al., 2013). Allen et al. (2011) found that the transferability of a location-specific LUR models is often limited because of poor model performance, and that better performance has been observed in locally-calibrated LUR models. Such models for NO_x, $PM_{2.5}$ and benzene perform better when the number of monitoring stations also is increased (Johnson et al., 2010).

To date, most LUR models in different geographic locations were developed to predict annual average concentrations of NO_2 and NO_x. Therefore, LUR models capable of predicting daily average concentrations would be useful in health studies focused on acute effects of air pollution. Moreover, there have been very few investigations of within-city NO_2 and NO_x LUR models in Australia (Dirgawati et al., 2015; Rose et al., 2011). In this study, we aimed to address the knowledge gaps outlined above by developing and applying LUR models to predict the daily average ambient NO_2 and NO_x concentration across the Brisbane Metropolitan Area (BMA), in Queensland, Australia.

2. Materials and methods

Brisbane is the third-largest city in Australia, and its metropolitan area (BMA) is one of the fastest growing urban areas in Australia, with a population of 2.3 million in 2014 (ABS, 2015). The city is located on the Brisbane River, in the south-east corner of the state of Queensland (Figure 1). BMA has a subtropical climate with relatively small seasonal variation, and the monthly mean temperatures during the warmer (November – April) and cooler (May – October) months are 29.9℃ and 24℃, respectively (http://www.bom.gov.au). The major urban pollution sources in the BMA are vehicular traffic and local non-traffic sources, mostly located in the lower reaches of the Brisbane river, approximately 15-18 km NE of the Central Business District (CBD). The main local non-traffic sources of air pollutants in Brisbane include aviation, a seaport, and other local industries (e.g., oil refineries).

2.1 Air pollution monitoring data

In total, 31 sites were selected for this study across the BMA over 2009 – 2012. The sites consisted of 6 long-term and 25 short-term monitoring locations (Figure 1). The location and data availability for each site is shown in Figure 2. All short-term sites were schools located between 1.5 and 30 km from Brisbane city. Traffic emissions were the major emission source to the nearby sites, with significant variability of hourly averaged vehicle counts, ranging from 44 to 1217 vehicles/hour (Laiman et al., 2014). Further details regarding the UPTECH project and its design can be found in our previous publication (Salimi et al., 2013). Long-term regulatory monitoring data were collected from the Queensland Department of Environment and Heritage Protection (DEHP) (coded as sites R1 to R5) (http://www.ehp.qld.gov.au/). DEHP provided hourly averaged NO_2 and NO_x monitoring data. In addition, short-term NO_2 and NO_x monitoring data were collected at 20 (coded S05 to S25) and 25 (coded S01 to S25) sites, respectively, during the Ultrafine Particles from Traffic Emissions and Children's Health (UPTECH) project between October 2010 and August 2012 (https://www.qut.edu.au/research/research-projects/uptech). NO_2 data was not recorded at five sites (S01–S05) due to data logging malfunction. Measurement campaigns were conducted at one site at a time for two consecutive weeks. Also, a one-year-long ambient NO_2 and NO_x monitoring campaign was conducted at an inner city site (S26) as part of the UPTECH project. An outdoor NO_x monitoring station was located at each site. The NO_x concentration was measured by an Ecotech NO_x

analyzer (EC9841A). Zero checks were conducted using GasCal (Ecotech 1100) and a ZeroAir Generator (Ecotech 8301LC) for the NO_x analyzer. Span and zero checks were performed regularly. More details on the data collection scheme and project design have been published previously (Salimi et al., 2013). The NO_2 and NO_x measurement method at the DEHP reference sites (R1 to R5) used analogous methods (e.g. chemiluminescence approach. The sampling interval was 30 s. Negative and zero values (2%) were considered missing and removed. All data was converted to hourly averages for further analysis.

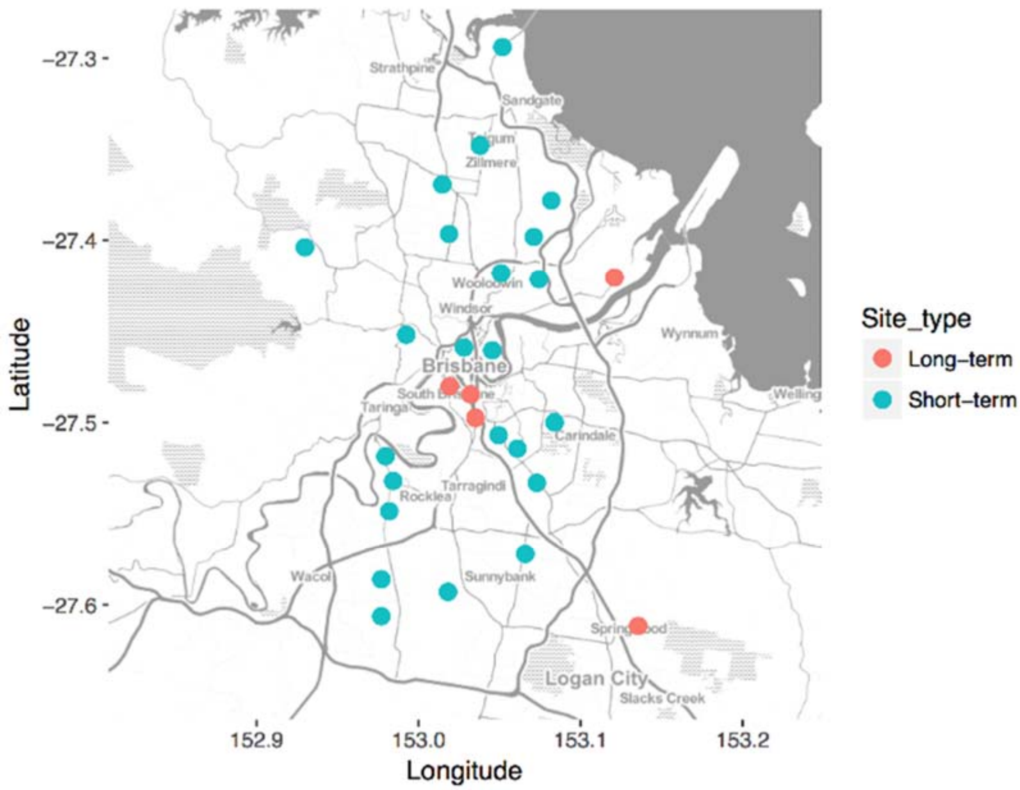


Figure 1. The map (openstreetmap.org) shows the geographical location of short-term and long-term monitoring stations across Brisbane Metropolitan Area (BMA), Australia.

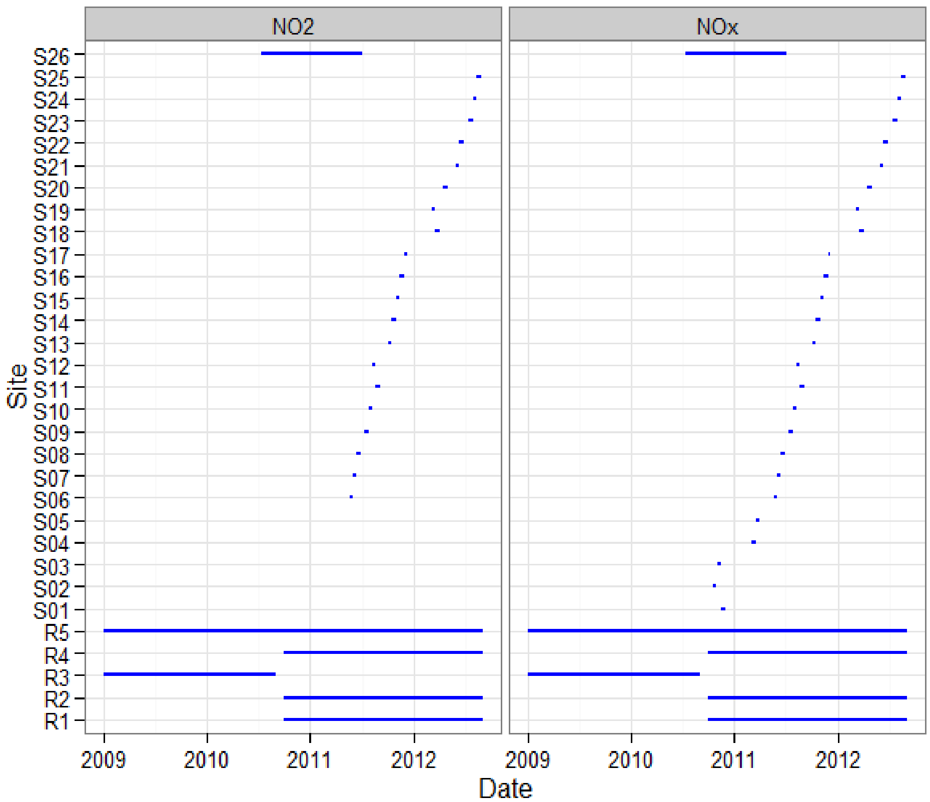


Figure 2. NO_2 and NO_x monitoring campaign period and data availability at 31 sites. NO_2 data was not available at five sites (S01–S05).

Four reference sites (R1, R2, R4, and R5) data were used to adjust temporal variability in short–term monitoring sites (S01–S25), which is a similar approach to previous studies (Beelen et al., 2013; Dirgawati et al., 2015; Hoek et al., 2002). R3 site was not considered for the adjusted average calculation, as the data did not cover the entire period when the short–term monitoring was conducted. Temporal variability in the short term sites was adjusted to eliminate uncertainties due to the shortness of the measurement campaigns, as well as to calculate adjusted the yearly average concentration of NO_2 and NO_x, which is a similar approach to the one used in ESCAPE study (Beelen et al., 2013). Adjusted hourly averaged NO_2 and NO_x concentrations at the short–term sites (S01–S25) were calculated using following equation:

$$X_{adj} = \frac{A\ (Long - term)}{B\ (Long - term)} \times S\ (short - term)$$

Where, X_{adj} is the adjusted average concentration at a given short–term monitoring site, A is the overall average concentration at the long–term sites for the entire period of short–term monitoring, B is the average concentration at the long–term monitoring sites concurrent with the two–weeks of monitoring at the short–term site (October 2010 – August 2012), and S is the hourly average concentration at a the short–term site.

2.2 Predictor variable data

In total, 139 predictor variables were selected to develop LUR model. Table 1 summarizes all predictor variables in 10 categories. Road length and land use variables were calculated in buffer rings around each monitoring site (of which there were 22 – see Table 2), and other variables were calculated at a single point.

Population density

The population density data were obtained from the Australian Bureau of Statistics (ABS) at a 1km^2 grid across BMA (Australian Bureau of Statistics., 2014). The value of each cell represents Usual Resident Population (URP) from the 2011 Census of Population and Housing.

Elevation

Elevation data (m above sea level) were obtained through the Shuttle Radar Topography Mission (SRTM) data acquired by National Aeronautics and Space Administration (NASA) in February 2000. SRTM data is distributed by the U.S. Geological Survey (USGS) and provides global ground surface topography.

Land use data

We obtained data on natural and anthropogenic land use (spatial predictors) variables that have a possible relation with measured NO_2 and NO_x concentrations. Land use variables were used to improve the predictions of NO_2 and NO_x, since their influence has been shown to be important in other recent studies (Knibbs et al., 2014; Vienneau et al., 2013). Land use data were derived from the Australian Bureau of Statistics census 2011. The land use types were classified into residential, commercial, industrial, and open space (the sum of water, parks and agricultural land) (Rose et al., 2010). We generated 22 ring buffers (annulus buffers) from 100 m to 10 km around each monitoring site using ArcGIS in order to capture sources of variability in NO_x concentrations over short and longer distances (Table 1)

Road length

Roads length and classification were derived from the Transport and Topography product from the Public Sector Mapping Agencies (PSMA), Australia (https://www.psma.com.au/). The data set positional accuracy is ±2 m in urban areas, ±10 m in rural and remote areas, and attributed 99.09% accuracy. A total, major, and minor road length in each buffer ring was calculated using ArcGIS, as summarized in (Table 1).

Table 1. List of land uses predictor variables.

Variables category (units) [a]	Spatial resolution	Point or buffer [b]	Data source
Distance to coast (km)	–	Point	ArcGIS geoprocessing tools
Distance to port (km)	–	Point	ArcGIS geoprocessing tools
Distance to airport (km)	–	Point	ArcGIS geoprocessing tools
Distance to nearest major road (km)	–	Point	ArcGIS geoprocessing tools
Distance to nearest minor road (km)	–	Point	ArcGIS geoprocessing tools
Major road (km)	–	Buffer	PSMA Australia Transport and Topography product
Minor road (km)	–	Buffer	PSMA Australia Transport and Topography product
Population density (person/km^2)	1 × 1 km	Point	Australian Bureau of Statistics
Land use by type (km^2)[c]	Mesh block [d]	Buffer	Australian Bureau of Statistics
Elevation (m)	30 m	Point	U.S. Geological Survey

[a]In total 139 variables were calculated in GIS, including 7 point variables (as shown in Table 2), 44 road length (major and minor) buffer variables (2×22), and 88 land use (industrial, commercial, residential, and open space) buffer variables (4×22)

[b] 22 Ring buffers were created with radii of 100 m, 200 m, 300 m, 400 m, 500 m, 600 m, 700 m, 800 m, 1000 m, 1200 m, 1500 m, 1800 m, 2000 m, 2500 m, 3000 m, 3500 m, 4000 m, 5000 m, 6000 m, 7000 m, 8000 m, and 10,000 m.

[c] Land use type was classified in residential, commercial, industrial, and open space.

[d] A mesh block is the smallest geographical location in Australian Statistical Geography Standard.

2.3 LUR model development

In this study, the temporal model was set as $\log\ (\mathrm{NO_2}\ \text{or}\ \mathrm{NO_x}) = \beta_0 + \mathrm{s(yday)} + \mathrm{factor(wday)} + \varepsilon$, where $\mathrm{s(yday)}$ is a smooth, periodic function of the day of the year, fit with penalised splines (Eilers and Marx, 1996), and $\mathrm{factor(wday)}$ is a categorical term for the day of the week. The temporal model does not adjust for any spatial effect, and makes use of the long-term monitoring site data to estimate the city-wide annual trend. The residuals from the temporal model are then fit with a Least Angle Regression approach (described below) with at most 50 land use variables added sequentially. Of these 50 models, the one with the lowest Bayesian information criterion (BIC) is selected, representing the minimum number of predictors to obtain the best model fit. This model is then pruned with a backwards selection method that sequentially drops covariates from the model whose 95% confidence intervals contain zero. The final spatial model therefore contains those predictors which have led to a decrease in BIC but are found to be significant at 5%. The final predictions were therefore constructed from adding the predictions from the spatial model of the residuals to the predictions from the temporal model. The Least Angle Regression (LARS), a form of forward stepwise regression (Efron et al., 2004), model started with the residuals from the temporal model; at each step in the modelling process, the angle between the current residuals and each predictor variable was calculated and a linear function of the variable with the least angle is added to the model for the next step, up to a maximum of 50 included predictor variables (of the 139 derived variables). LAR method was found as a robust stepwise regression with better performance and faster computational capability compared to the traditional stepwise regression method (Capizzi and Masarotto, 2011). As the buffers are annular in shape, the same land use variable can be included in the model at multiple buffer sizes without including the same information multiple times. As the land use variables did not vary temporally, and we assumed that the spatial and temporal variability were separable (a necessary assumption due to the study design) they were only included in the spatial residual model to account for changes over space. However, the inclusion of the smooth effect of day of the year and day of the week in the temporal model ensures that the long-term monitoring sites' data explains the annual trend in NO_2 and NO_x concentrations, which enable the model to estimate the daily concentration of the pollutants. In this way, the spatial variation in the schools' baseline NO_2 and NO_x concentrations can be identified, as the spatial effects explain the remaining variability after accounting for the annual trend. This allows prediction for every day of the year at every site, particularly at each short-term site for the 50 weeks of the year when measurements were not taken. In order to allow each measurement location to contribute equally to the estimates of the spatial effects, the data were weighted by the inverse of the number of observations at that location.

Model choice was informed by calculating the Variance Inflation Factors (VIFs) (Kutner et al., 2004) for each model and choosing the model which explained the most variability and for which all VIFs were less than 5. Rogerson (2001) concluded that the VIFs>5 means potential multicollinearity in a model. Thus the resulting model with VIFs < 5 contains the set of variables which maximize model fit while reducing multicollinearity and rejected the noisy variables from the model To assess goodness of fit among competing models, Akaike information criterion (AIC) was applied (Akaike, 1974). The lower AIC values indicate the best model fit. Therefore, the variable which has limited contribution to the model was dropped in order to choose a model that performs well with a small number of covariates. Leave-one-out cross-validation (LOOCV) was performed to assess the final model performance by fitting the model to N-1 stations and predicted the concentration at the left-out station. The procedure was repeated N times and estimated LOOCV R^2 between observed and predicted NO_2 and NO_x concentrations (Johnson et al., 2010).

3. Results

3.1 NO_2 and NO_x measurement

The overall average NO_2 and NO_x concentration at 31 monitoring station was 10 (SD 4) and 22 (SD 18) ppb, respectively. Figure 3 shows the overall average NO_2 and NO_x concentrations at each site. It should note that adjusted concentrations were applied in short-term sites (S01–S25). Figure 4 shows the average hourly NO_2 and NO_x concentration distribution over the whole study period. More than 75% of NO_2 and 60% of NO_x concentrations were below 16 and 25 ppb, respectively, as shown in Figure 4. The NO_2 and NO_x concentration range (maximum – minimum)/mean ratios were 148% and 369%, respectively, which indicates relatively high spatial variability between investigated sites. Analysis of variance (ANOVA) tests at all sites found that the average daily NO_2 and NO_x concentrations were significantly different (p<0.001).

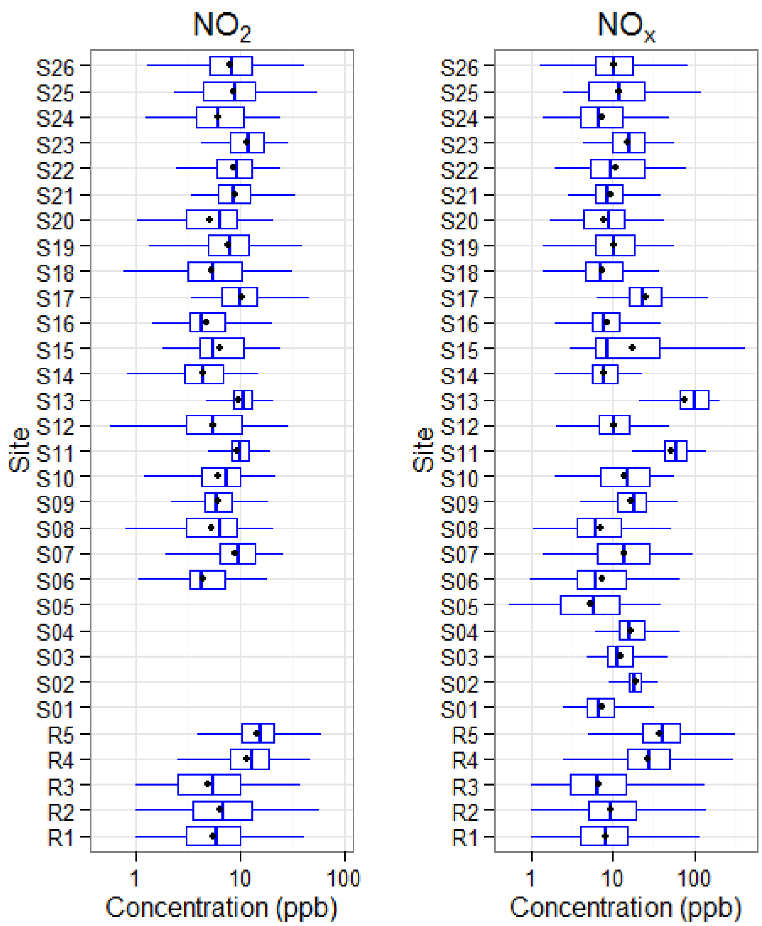


Figure 3. NO_2 and NO_x concentrations recorded at 31 sites. The solid circle symbols inside the boxplot represent average concentrations.

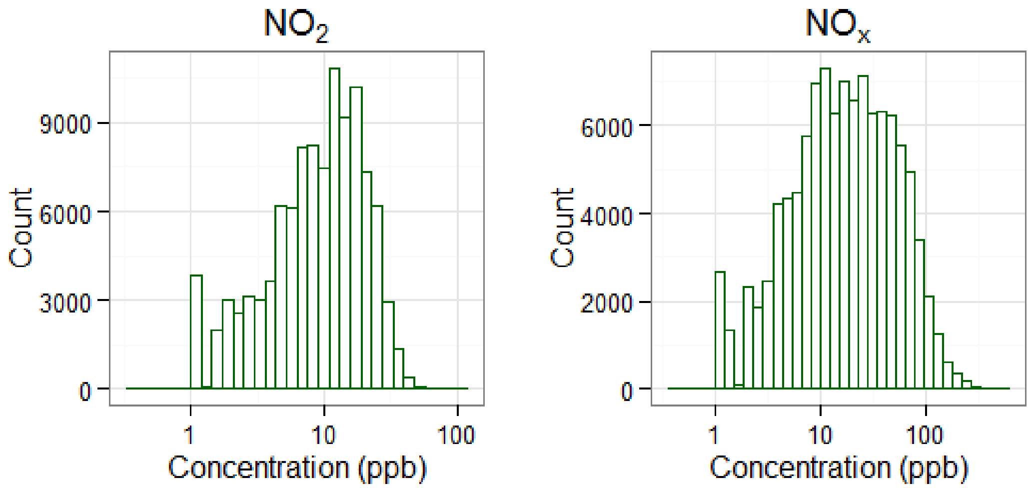


Figure 4. Hourly averaged NO_2 and NO_x concentrations over the whole study period.

3.2 Model predictor variables

Table 2 summarizes the data of all 139 predictor variables, separated into 22 buffer ring and 7 point categories at 31 monitoring stations. Commercial and residential areas were observed in all buffer radiuses, whereas, the industrial areas were not available below 300 m buffer. Industrial areas were located at a large extent beyond 1000 m buffer. The statistics of land use areas show that the most of the sites were located in the residential and commercial land use category. Commercial area density was higher (≥ 2.93 km^2) outside the 3000 m buffer. Open space areas that include water, park, and agricultural land were not present in 100 m buffer range at any of the sites. Open space areas were high (≥ 2.32 km^2) outside the 1500 m buffer radius. Both major and minor roads were present in all buffers. The average distance of the monitoring sites to the major and minor roads were 1.67 (SD 1.77) and 0.21 (SD 0.28) km, respectively. Overall average population density at the study sites was 2,103 (SD 1,179) persons per km^2, ranging from 217 and 6,216 persons per km^2. The population density was higher in the inner-city area sites, as expected, compared to urban background sites. The average distance to the port and airport to the monitoring sites were 21.8 (SD 10.4) and 13.1 (SD 9.94) km, respectively, which indicate a very limited contribution from the airport and port areas to the recorded data. Figure 5 shows the relationship between land use variables and ring buffer radius across 31 sites.

Table 2. Summary statistics of the buffer (A) and point (B) predictor variables at 31 monitoring sites.

A. Predictor variable with buffer

Order	Buffer (m)	Commercial area (Km^2)				Industrial area (Km^2)				Open space area (Km^2)				Residential area (Km^2)				Major road length (Km)				Minor road length (Km)			
		Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
1	100	0.02	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.01	0.00	0.03	0.02	0.10	0.00	0.42	0.10	0.16	0.00	0.70
2	200	0.05	0.03	0.00	0.13	0.00	0.00	0.00	0.02	0.01	0.02	0.00	0.08	0.06	0.03	0.00	0.13	0.07	0.24	0.00	1.02	0.44	0.52	0.00	2.51
3	300	0.08	0.07	0.00	0.28	0.01	0.02	0.00	0.08	0.03	0.04	0.00	0.18	0.17	0.07	0.00	0.28	0.12	0.38	0.00	1.51	0.82	0.84	0.00	3.72
4	400	0.10	0.11	0.00	0.49	0.02	0.04	0.00	0.20	0.05	0.07	0.00	0.31	0.33	0.13	0.00	0.50	0.31	0.88	0.00	3.60	1.33	1.23	0.00	5.52
5	500	0.13	0.16	0.00	0.73	0.04	0.09	0.00	0.40	0.08	0.11	0.00	0.43	0.53	0.21	0.00	0.78	0.48	1.31	0.00	5.34	1.99	1.79	0.00	8.47
6	600	0.17	0.22	0.00	1.00	0.06	0.14	0.00	0.67	0.12	0.14	0.00	0.58	0.77	0.30	0.00	1.12	0.66	1.69	0.00	6.88	2.73	2.31	0.00	11.2
7	700	0.22	0.29	0.00	1.29	0.09	0.21	0.00	1.00	0.18	0.18	0.00	0.77	1.04	0.41	0.00	1.51	0.83	2.04	0.00	8.07	3.54	3.00	0.00	14.7
8	800	0.28	0.36	0.00	1.57	0.12	0.29	0.00	1.36	0.25	0.24	0.00	1.02	1.33	0.53	0.00	1.98	1.06	2.48	0.00	9.93	4.63	3.92	0.00	18.8
9	1000	0.41	0.51	0.00	2.18	0.21	0.46	0.00	2.12	0.43	0.38	0.01	1.66	2.02	0.80	0.00	3.06	1.63	3.45	0.00	13.4	7.03	5.73	0.00	26.8
10	1200	0.58	0.69	0.00	2.88	0.33	0.69	0.00	3.02	0.69	0.52	0.07	2.36	2.80	1.13	0.00	4.34	2.50	4.73	0.00	18.4	10.2	7.79	0.00	34.1
11	1500	0.88	0.96	0.00	4.19	0.59	1.19	0.00	4.99	1.17	0.83	0.22	3.65	4.18	1.69	0.00	6.60	3.98	6.18	0.00	23.9	15.8	11.7	0.97	53.2
12	1800	1.24	1.26	0.00	5.41	0.94	1.80	0.00	7.63	1.82	1.27	0.33	5.29	5.79	2.31	0.00	9.13	5.89	8.02	0.00	30.4	22.4	16.0	2.12	75.4
13	2000	1.50	1.46	0.00	6.12	1.25	2.28	0.00	9.70	2.32	1.55	0.42	6.50	7.01	2.78	0.00	11.0	7.31	9.30	0.00	34.3	27.4	18.9	2.83	88.2
14	2500	2.20	2.02	0.00	8.44	2.23	3.66	0.00	15.8	3.91	2.41	0.82	11.0	10.5	4.10	0.00	16.5	11.3	11.9	0.00	44.0	42.3	26.4	6.49	117.6
15	3000	2.93	2.46	0.00	10.2	3.42	5.26	0.00	22.8	5.89	3.34	1.69	16.0	14.8	5.53	0.05	22.7	15.5	14.5	0.00	52.5	58.6	34.2	9.77	151.7
16	3500	3.72	3.00	0.02	12.1	4.91	6.93	0.03	29.7	8.11	4.25	2.50	20.4	20.1	7.02	0.71	31.1	20.9	17.5	0.00	63.6	78.7	43.0	15.6	190.3
17	4000	4.66	3.52	0.04	14.2	6.71	8.72	0.13	36.9	10.7	5.33	3.30	25.9	26.2	8.81	2.56	40.3	27.4	21.5	0.00	82.5	101.7	52.7	22.0	229.0
18	5000	6.90	4.54	0.27	18.5	10.8	12.5	0.24	51.8	17.2	8.04	5.64	39.5	40.5	12.58	7.42	60.0	43.3	29.4	0.00	114.6	153.3	70.1	36.6	312.9
19	6000	9.77	5.74	0.78	24.6	15.6	15.8	0.25	63.4	25.5	10.6	10.3	55.5	57.4	16.25	17.44	82.3	59.6	37.0	0.00	149.2	214.3	92.5	47.3	416.1
20	7000	13.4	7.3	1.7	31.2	21.0	18.6	0.4	76.8	36.4	14.3	16.5	80.1	76.2	20.3	25.4	107.1	78.0	46.4	0.00	192.5	283.9	116.6	60.0	518.9
21	8000	16.8	8.9	2.2	37.8	26.7	21.3	0.5	87.5	49.9	18.9	22.6	108.0	97.8	24.2	37.9	134.1	97.1	55.5	0.00	240.1	358.3	143.7	76.3	636.8
22	10000	23.7	11	4	48.1	39.0	27.8	2.17	111.9	85.0	31.0	38.2	174.3	148.9	33	68	201.8	136.4	71.2	8.11	318.7	533.2	192.4	124.4	891.0

B. Point predictor variable

Order	Predictor Variable	Mean	SD	Min	Max
1	Elevation (m)	30.65	18.41	5.00	65.00
2	Distance to Coast (Km)	16.52	10.63	3.00	45.91
3	Distance to Major Road (Km)	1.67	1.77	0.01	9.56
4	Distance to Minor Road (Km)	0.21	0.28	0.02	1.41
5	Distance to Port (Km)	21.77	10.38	6.86	49.87
6	Distance to Airport (Km)	13.05	9.94	0.18	40.00
7	Population Density (person/Km^2)	2103	1179	217	6216

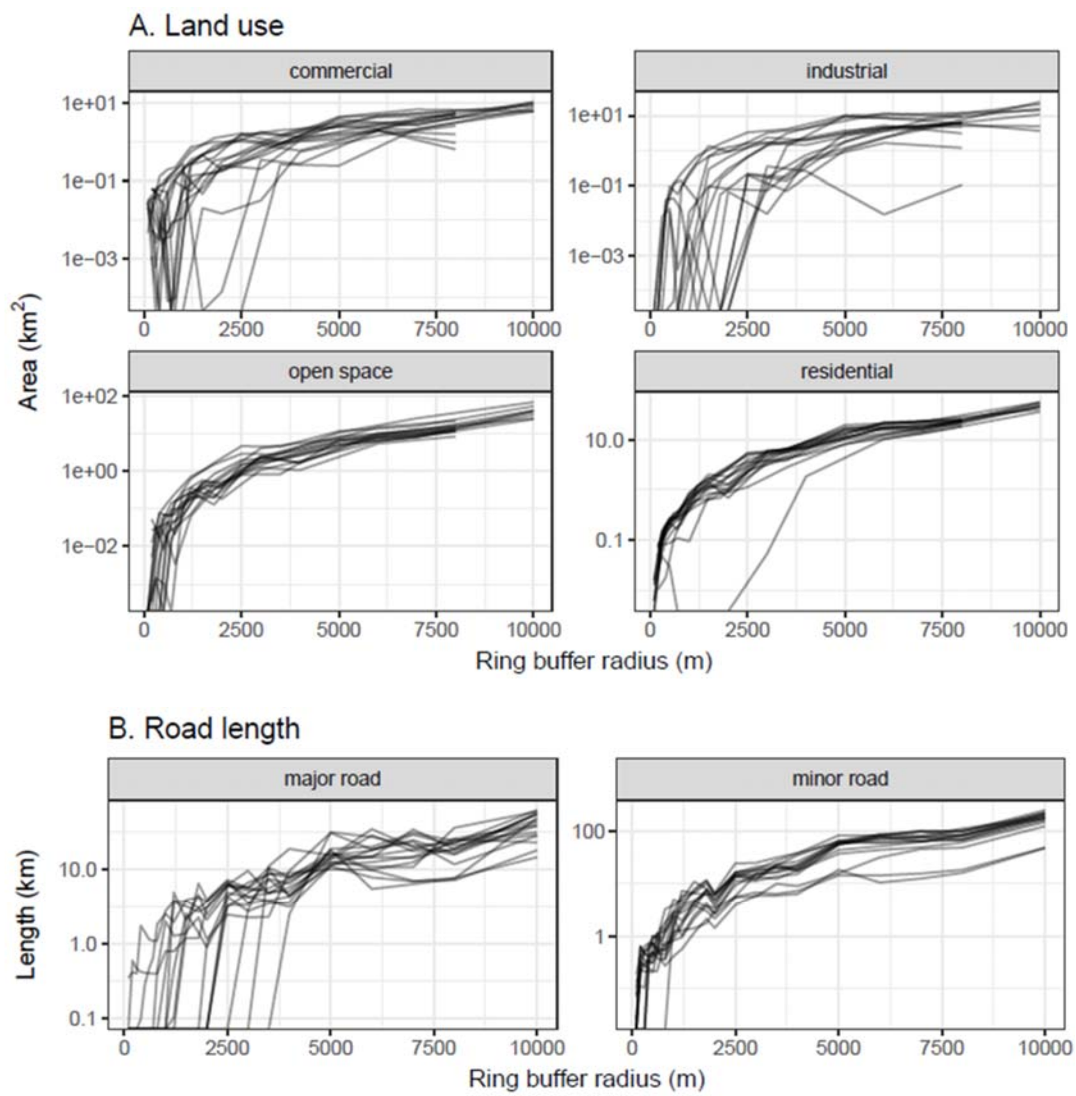


Figure 5. Relationship between land use variables and ring buffer radius across 31 sites. Each line correspondence to a site, and the y-axis is in the log scale.

3.3 LUR model result

The final NO_2 and NO_x models are shown in Table 3. The selected NO_2 model includes population density, open space area in the 0 – 100 m buffer, major road in 100 – 200 m buffer, industrial area in 400 – 500 m buffer, residential area in 100 – 200 m buffer, and distance to major road. The selected NO_x model includes open space area in 0 – 100 m and 600 – 700m buffer, industrial area in 100 – 200 m buffer, residential area 100 – 200 m buffer, distance to major road.. The fitted model for NO_2 and NO_x explained 63% and 70%, respectively, of spatial variability in daily NO_2 levels. The mean LOOCV R^2 for the fitted NO_2 and NO_x model across BMA was 23% and 27%, respectively. We conducted LOOCV for the model evaluation for each site across BMA. The LOOCV is the best-suited approach for the model evaluation in this study as it leaves out all data from an entire site when predicting the concentration at that site. In this way, we were able to evaluate the performance in the spatial locations across BMA. The overall average LOOCV R2 for the fitted NO_2 and NO_x model was 23% (range: 3-49%) and 27% (range: 2-51%), respectively. The LOOCV R^2 was less than 10% at seven short-term sites indicating that those sites have characteristics which are not common when compared to the remaining sites and therefore result in poor predictive performance. The models predict overall NO_2 and NO_x concentration on each day of a year (Figure 6). The partial effect of the modeled smooth splines ranged between -0.4 and 0.4, which are 0.67 and 1.49 in the normal scale, respectively, centered around the average value of zero (or one in the normal scale) (Figure 6). Partial effects for both NO_2 and NO_x were the highest between day 150 (May) and 240 (August) of the year. Figure 7 shows the relation between modeled and fitted daily average concentration of the pollutants. Figure 8 shows the predicted NO_2 and NO_x concentration at each day of the year at each site.

To investigate the LUR model performance without long-term data, the model was run with data from only 25 short-term sites. The results showed the adj-R^2 of 33.6% for NO_2 and 33% for NO_x, which is approximately 50% reduction of the model performance compared to the original model that included both long-term and short-term sites data.

Table 3. Parameter estimates, lower and upper bounds of their 95% confidence intervals, and variance inflation factors.

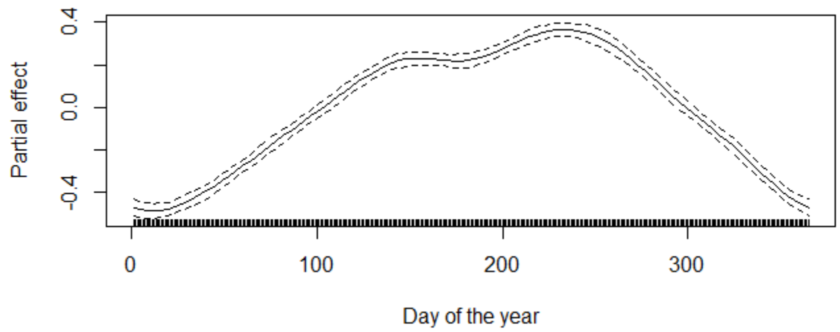
Term	Buffer	Estimate	2.5th	97.5th	R^2 Partial[2]	R^2 (adj)	R^2 LOOCV[3]
A. NO_2							
Day of the year[1]	–	–	–	–			
Day of the week[1]	–	–	–	–			
Intercept	–	-0.48	-0.51	-0.45			
Open area (km^2)	100-200	-7.20	-8.46	-5.94	43%		
Population Density (person/Km^2)	–	1.15E-04	1.04E-04	1.26E-04	40%		
Major road (km)	100-200	1.88	1.82	1.94	46%	64%	23 (3-49)%
Industrial area (km^2)	400-500	1.74	1.54	1.93	43%		
Residential area (km^2)	100-200	-1.18	-1.61	-0.76	43%		
Distance to major road (km)	–	0.01	0.01	0.02	44%		
B. NO_x							
Day of the year	–	–	–	–			
Day of the week	–	–	–	–			
Intercept		-0.57	-0.63	-0.52			
Open area (km^2)	0-100	25.25	23.54	26.95	53%		
Industrial area (km^2)	100-200	11.00	8.76	13.24	51%	70%	27 (2-51)%
Open area (km^2)	600-700	-1.79	-2.39	-1.18	51%		
Residential area (km^2)	100-200	0.65	0.07	1.24	52%		
Distance to major road (km)	–	1.95E+04	1.91E+04	2.00E+04	60%		

[1]Day of the year and day of the week term represented the temporal model (see section 2.3)

[2] Spatial LUR model for NO_2 and NO_x and estimated partial R^2

[3]Leave out all data from an entire site when predicting the concentration at that site

A. NO_2



B. NO_x

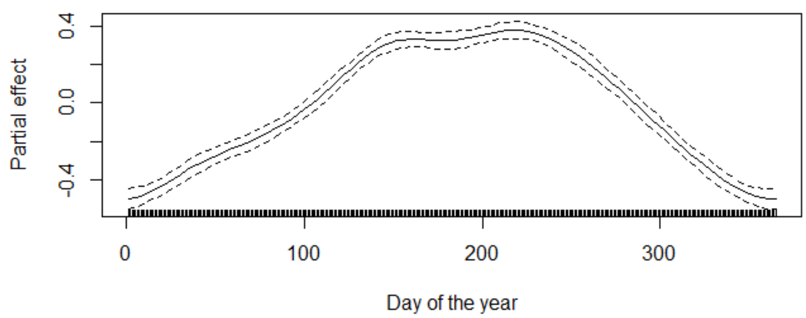
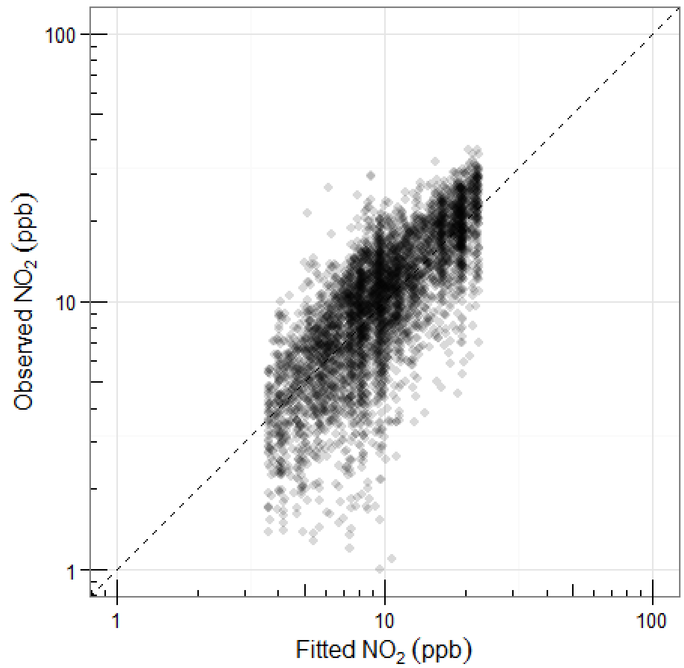


Figure 6. Smooth annual trend term (year of the day) from LUR model (solid line) and its 95% confidence interval (dotted region). Partial effects are the contribution of that explanatory variable to the fitted value.

A. Observed vs. fitted NO_2



B. Observed vs. fitted NO_x

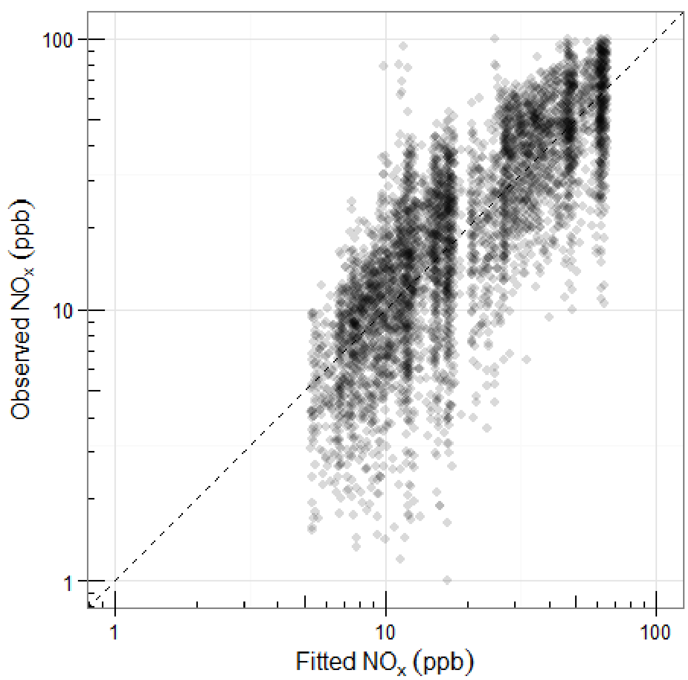
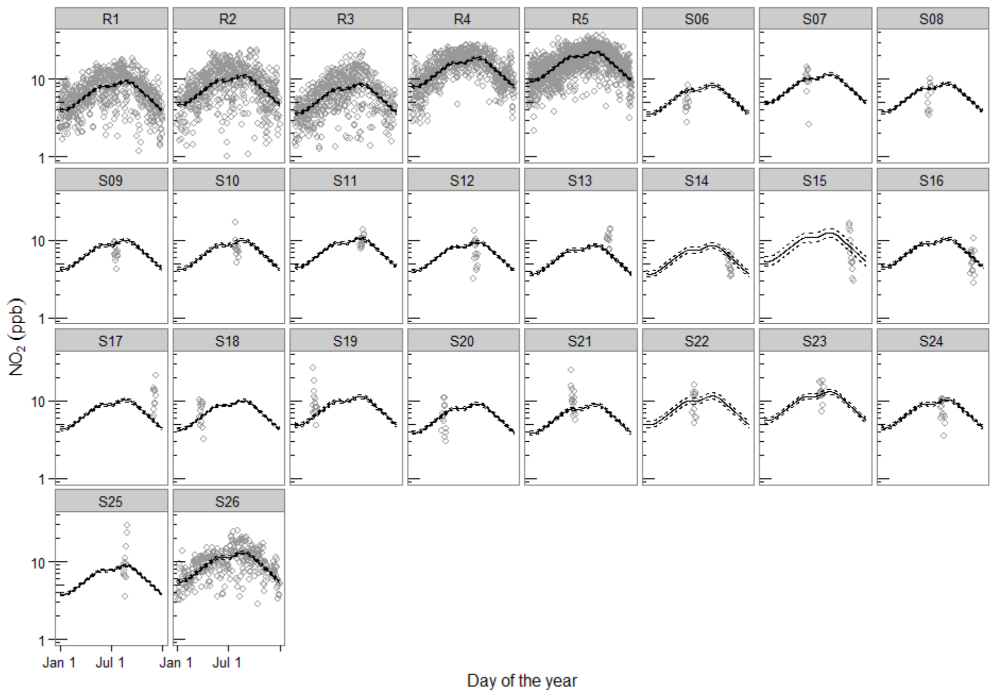


Figure 7. Observed vs. fitted values of average daily NO_2 (A) and NO_x (B) concentration.

A. NO_2



B. NO_x

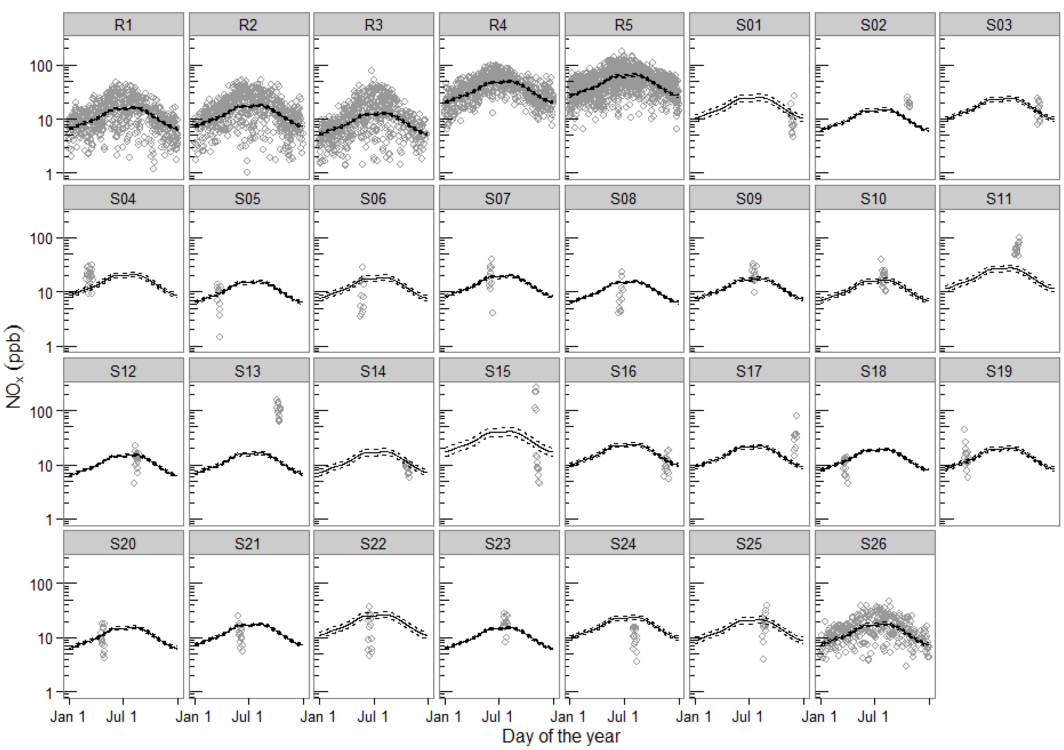


Figure 8. Observed (open circle) and predicted (solid line) NO_2 (A) and NO_x (B) concentration at each day of the year (with 95% prediction interval, dashed lines) at each location.

4. Discussion

In this study, we developed LUR models to predict average daily NO_2 and NO_x concentrations in the Brisbane Metropolitan Area, Australia from 2009 –2012. We found the best LUR models for NO_2 and NO_x explained 64% and 70% of spatial variability, respectively. This performance falls within the range of many recent studies, which varied between 53% – 98% for NO_2 and 48% – 91% for NO_x (as shown in Table 4 and Table 5.The ESCAPE study developed 36 LUR models for NO_2 and NO_x for 36 European cities (Beelen et al., 2013). The adjusted R^2 ranged from 59% (Marseille, France) to 90% (Albacete-Valencia, Spain) for NO_2, and 49% (Basque Country, Spain) to 91% (London, UK) for NO_x (Beelen et al., 2013). In addition, Hoek et al. (2008) reviewed LUR models for NO_2 and NO_x from the year 1997 to 2008, which showed significant variation between studies for NO_2 (R^2: 51% – 90%) and NO_x (R^2: 73% – 96%). It is clear from the previous studies that the LUR models performance varied significantly in different geographical locations, which is due to unique characteristics of each study site, variability in the effect of predictor variables on investigated pollutants, and the selection of an LUR modeling method. Our LUR model included a day of the year term that estimated day to day variations of NO_2 and NO_x concentrations. It should note that the day of the year variable accounts seasonality. Previous studies (Table 4 and 5) included only short-term monitoring campaign data in the model to predict the annual average concentration of NO_2 and NO_x; whereas, our model included both long-term and short-term monitoring data and predicted the daily average concentration of NO_2 and NO_x. Overall, this study developed LUR models in estimating daily NO_2 and NO_x concentrations based on LARS in which periodic function of the day of the year fit with penalised splines. The model under-predicted NO_2 and NO_x concentration at a few short-term sites, for example, the model underpredicted NO_2 concentration at S02, S11 and S13. The limited datasets at short-term sites and the variability in local pollution sources are the most likely reasons of uncertainty in the modeling performance at those sites. We note that our results, which are for daily NO_2, are not directly comparable to the majority of results in the LUR literature, which are for annual NO_2.

In a recent investigation in Perth, Australia, Dirgawati et al. (2015) applied an LUR model for NO_2 and NO_x, which explained 69% (LOOCV R^2 62%) and 75% (LOOCV R2 65%) of variation in annual mean concentration, respectively. The modeling method and study design were identical to the ESCAPE study. The major differences between our study and Dirgawati et al. (2015) study are the modeling method and data collection scheme. Our LUR model is developed (based on LARS) to predict daily NO_2 and NO_x concentration, incorporating six long-term air quality monitoring datasets, while Dirgawati et al. (2015) applied the LUR model to predict annual NO_2 and NO_x concentrations.

Lee et al. (2014) also applied an ESCAPE-based NO_2 and NO_x LUR model to high-density traffic roads and population area in an Asian city, Taipei, which explained 74% and 81% of the variation (Table 4 and Table 5). The total area in BMA is 20 times higher than Taipei (786 km^2) and has greater pollutant dispersion due to strong sea breezes; those are some factors that could explain the variability in model performance between the two studies. In a European city, Sabadell, Spain, Aguilera et al. (2008) developed and applied annual NO_x LUR model that explained a similar NO_x concentration variation (68%) to our study, although the LUR model in Spain incorporated 57 short-term monitoring sites.

In our study, industrial areas and distance to major road, were predictor variables in the final NO_2 LUR model, similar to the previous study in Perth, Australia (Dirgawati et al., 2015). Household density, which was a predictor variable in LUR model in Perth city, was not applied in our study; instead, we incorporated population density data in the model, and it was identified as a predictor variable in the final LUR model for NO_2. The ESCAPE study also found population density as a predictor variable in the final NO_x LUR model in 9 out of 36 European cities, including Norway, Sweden, Lithuania, Netherlands, Belgium, Germany, Austria, Rome, and Greece (Beelen et al., 2013). Population density is a key predictor variable in a large geographic location because it provides a prediction of anthropogenic emission source contribution in that area.

In this study, distance to major road was one of the key predictor variables in the final model. It represents the traffic-related contribution to the overall NO_2 and NO_x concentration in BMA, and as such was conspicuous in our models. Traffic related emission were identified as a predictor variable in most previous studies, which is comparable to our findings (Table 4 and Table 5). In the ESCAPE study, more than 65% of the final LUR models for both NO_2 and NO_x contained road length predictor variable in different buffer radius out of 36 LUR models (Beelen et al., 2013). While road length predictor variables were not included in a number of previous LUR models for NO_2 and NO_x (Table 4 and Table 5), traffic counts were present in those final models, which aimed to capture the influence of vehicle emissions. Our study applied road length data, instead of the traffic count data due to their unavailability at all sites. The use of traffic counts in LUR models is often not possible due to the logistics of collecting these data for all roads in a large study area (Hoek et al., 2008). Henderson et al. (2007) incorporated road lengths and traffic count data separately in the LUR models in Vancouver, Canada, and the results showed that the models differed by only 7% variability in the R^2 values, which suggests that road length is an acceptable traffic emission proxy in LUR models for nitrogen oxides.

Our LUR model predicted average daily NO_2 and NO_x concentrations, which are an important criterion in exposure assessment as part of health studies that examine short-term impacts of air pollutants. The modeled annual trend in Figure 6 showed that during May to August (the cooler part of the year), the NO_2 and NO_x concentrations were the highest in BMA. The developed LUR model predicted daily average concentration variation over the year at both long-term and short-term sites, and the daily modelled NO_2 and NO_x concentration trends at short-term sites were agreed with the long-term sites trends (Figure 8). Previously developed LUR models (Table 4 and Table 5) were applied almost entirely to predict the annual average concentration of NO_2 and NO_x. Therefore, in this study we focused on developing a framework for LUR models to estimating daily NO_2 and NO_x concentrations based on LARS, in which a periodic function of the day of the year and the day of the week was fit with penalised splines.

The major limitation of this study is a limited number of monitoring site data (31 sites) compared with other studies where authors used 57–80 sites (Aguilera et al., 2008; Madsen et al., 2011). Another limitation of this study was that at short-term sites only two weeks of monitoring data were collected, which is similar to the several recent studies (Beelen et al., 2013; Dirgawati et al., 2015). In addition, monitoring campaigns at the short-term sites were conducted only once, which means in one season. Although a limited number of monitoring sites (14 to 24) have successfully been applied in Spain and Italy to develop LUR models (Beelen et al., 2013), a recent study in Spain showed that LUR model R^2 and LOOCV-R^2 can be exaggerated when the number of measurement sites is small and there are many predictor variables (Basagaña et al., 2012; Wang et al., 2012). This should therefore be considered when interpreting our results. However, the LUR model in our study incorporated six long-term continuous monitoring site data, which may increase the robustness of the results presented here. BMA has very different weather pattern and ambient pollutant sources compared to Europe, Asia, and America, due to its sub-tropical climate, relatively small seasonal variability, good air dispersion due to the eastern sea breeze, limited point source industrial emission sources. Nevertheless, the LUR models we developed explained NO_2 and NO_x spatial variability with reasonable performance.

Table 4. Comparison of land use regression models for NO_2

Study area	Total sites	Sampling period at each site	Predictor variables in final model	R^2 of the model	Model validation	Reference
Oslo, Norway	80	Two weeks	Altitude, total length high-traffic road (100 m), total length-medium traffic road (250 m), total length small and low-traffic roads (1000 m).	77%	63– 71%	Madsen et al. (2007)
Tianjin, China	30	Daily average (2005 - 2006)	Heating seasons: Total length of major roads (2000 m), residential area (2000 m), population density, wind index, temperature	74%	-	Chen et al. (2010)
			Non-heating seasons: Total length of major roads (2000 m), residential area (2000 m), agricultural land (500 m), population density, wind index	61%	-	Chen et al. (2010)
Los Angeles and Seattle, USA	145	Two weeks	Shortest distance to coastline, population density (2250 m), traffic density (250 m), length of primary roadway with limited access to highway (100 m), commercial area (750 m), length of secondary roadway (25 m).	53 – 54%	71 – 74%	Wilton et al. (2010)
Oslo, Norway	80	Two weeks	Altitude, total length high-traffic road (100 m), total length-medium traffic road (250 m), total length small and low-traffic roads (1000 m).	74%	67%	Madsen et al. (2007)
Edmonton, Alberta and Winnipeg, Canada	50	Two 2-weeks period	Industrial (2000 m), industrial (200 m), total road length (75 m), major road length (50 m), population density (2500 m), latitude of the site	84%	77%	Allen et al. (2011)
Southern California, USA	181	Two weeks	Weighted traffic flow, weighted truck flow, atmospheric stability, local road length, minimum distance to local streets and major freeways, surface temperature, transportation land use proportion, residential and transport land use proportion, farm and open land use proportion.	58 – 68%	55 – 64%	Li et al. (2012)
36 European cities (ESCAPE study)	14 – 80	Two weeks	Predictor variables of 36 LUR models developed for 36 European cities are available in the published article.	58 – 92%	31 – 87%	Beelen et al. (2013)
Taipei, Taiwan	40	Three 2-week period	Major road length (25 m), low-density residential area (500 m), urban green area (300 - 500 m), natural area (500 m).	74%	63%	Lee et al. (2014)
Perth, Australia	43	Two weeks	Traffic intensity on nearest road, household density (100 m), Industrial area (5000 m), road length (50 m).	69%	62%	Dirgawati et al. (2015)
Seoul, Korea	37	One year	Wind frequency weighted vehicle kilometer traveled, commercial area, residential area (1000 m), Industrial area (4500 m), solar radiation, wind speed, temperature, humidity, a number of dummy variables.	95- 98%	-	Kim and Guldmann (2015)
Veneto, Italy	40	Three 2-week period	Building (5000 m), industry (1000 m), altitude, total road length (100 m), inverse distance to motorways	75%	64%	Marcon et al. (2015)
Shanghai, China	38	Annual average	Major road length (2000 m), count of industrial sources (10000 m), agricultural land (5000 m), population density	82%	75%	Meng et al. (2015)
Great Britain	140	Annual average	Low-density urban land (20,000 m), high-density urban land (200 m), length of major roads (300 m), semi-natural land (200 m)	58%	51 – 54%	Gulliver et al. (2016)
Rome, Italy	46	Two 2-weeks period	Product of traffic intensity of the nearby road and inverse distance to the major roads, total traffic count all roads (300 m), altitude, traffic intensity nearby major roads, low-density residential land (1000 m)	72%	69	Gaeta et al. (2016)
Brisbane, Australia	26	20 short-term (two weeks) and six long-term (more than a year) sites	Day of the year, day of the week, population density, open space area (0 – 100 m), major road (100 – 200 m), industrial area (400 – 500 m), residential area (100 – 200 m), distance to major road.	64%	3 – 49%	This study

Table 5. Comparison of land use regression models for NO_x

Study area	Total sites	Sampling period at each site	Predictor variables in final model	R^2 of the model	Model validation	Reference
Oslo, Norway	80	Two weeks	Altitude, total length high-traffic road (100 m), total length-medium traffic road (250 m), total length small and low-traffic roads (1000 m).	73%	68 – 78%	Madsen et al. (2007)
Sabadell, Spain	57	Two months	Altitude, land cover factor, land cover factor (500 m), road type.	77%	67 – 75%	Aguilera et al. (2008)
Newhaven, UK	25 – 285	Two months average	Traffic intensity, proximity to ports and harbors, population and housing density, proximity to roadways, proximity to industrial sources.	63 – 79%	28 – 63%	Johnson et al. (2010)
Los Angeles and Seattle, USA	145	Two weeks	Shortest distance to coastline, population density (2250 m), length of primary roadway with limited access to highway (75 m), length of secondary roadway (25 m), traffic density (250 m)	53 – 55%	45%	Wilton et al. (2010)
			Shortest distance to coastline, area of commercial and industrial land (900 m), population density (2250 m), traffic density (250 m).	55%	50%	Wilton et al. (2010)
Los Angeles, USA	150	Three 2-week period	Distance to the nearest highway, distance to coast, distance to commercial source, length of primary and secondary roads (50 m), length of primary and secondary roads (400 m), population density (3000 m), intense land use (3000 m)[2].	48 – 70%	–	Mercer et al. (2011)
Oslo, Norway	80	Two weeks	Altitude, total length high-traffic road (100 m), total length medium traffic road (250 m), small and low-traffic roads (1000 m).	69%	59 – 68%	Madsen et al. (2007)
Southern California, USA	181	Two weeks	Weighted traffic flow, weighted truck flow, atmospheric stability, minimum distance to local streets, surface temperature, transportation land use proportion, farm and open land use proportion.	64 – 66%	42 – 45%	Li et al. (2012)
36 European cities (ESCAPE study)	14 – 80	Two weeks	Predictor variables of 36 LUR models developed for 36 European cities are available in the published article.	49 – 91%	31 – 88%	Beelen et al. (2013)
Taipei, Taiwan	40	Three 2-week period	Major road length (25 m), urban green area (300 m), urban green area (300 - 500 m), major road length (50 - 500 m), major road length (25 - 50 m), natural area (500 m).	81%	75%	Lee et al. (2014)
Perth, Australia	43	Two weeks	Traffic intensity on nearest road, household density (100 m), Industrial area (5000 m), road length (50 m).	75%	65%	Dirgawati et al. (2015)
Brisbane, Australia	31	25 short-term (two weeks) and six long-term (more than a year) sites	Day of the year, day of the week, open space area (0 – 100 m and 600 – 700), industrial area (100 – 200 m), residential area (100 – 200 m), distance to major road.	70%	2 – 51%	This study

5. Conclusion

LUR models for NO_x in BMA were developed based on least angle regression method. The models had reasonable predictive ability for average daily NO_2 and NO_x concentrations. Population density, distance to major road, and land use variables in the final model captured 64% and 70%, respectively, of NO_2 and NO_x spatial variability. Our study is the first LUR study to predict daily average NO_2 and NO_x concentration in Australia. Our LUR modeling approach can be applied to other ambient gaseous and particle pollutants in BMA and elsewhere.

6. Reference

ABS, 2015. Australian Bureau of Statistics, Regional Population Growth, Australia, 2013-14. Available in- http://www.abs.gov.au/.

Aguilera, I., Sunyer, J., Fernández-Patier, R., Hoek, G., Aguirre-Alfaro, A., Meliefste, K., Bomboi-Mingarro, M.T., Nieuwenhuijsen, M.J., Herce-Garraleta, D., Brunekreef, B., 2008. Estimation of outdoor NO x, NO2, and BTEX exposure in a cohort of pregnant women using land use regression modeling. Environmental science & technology 42, 815-821.

Akaike, H., 1974. A new look at the statistical model identification. Automatic Control, IEEE Transactions on 19, 716-723.

Allen, R.W., Amram, O., Wheeler, A.J., Brauer, M., 2011. The transferability of NO and NO2 land use regression models between cities and pollutants. Atmospheric Environment 45, 369-378.

Australian Bureau of Statistics., 2014. Australian Population Grid. Australian Bureau of Statistics, Australia.

Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., de Nazelle, A., Nieuwenhuijsen, M., Vila, J., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. Atmospheric Environment 54, 634-642.

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.-Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K.T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nádor, G., Varró, M.J., Dėdelė, A., Gražulevičienė, R., Mölter, A., Lindley, S., Madsen, C., Cesaroni, G., Ranzi, A., Badaloni, C., Hoffmann, B., Nonnemacher, M., Krämer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Strömgren, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., de Hoogh, K., 2013. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. Atmospheric Environment 72, 10-23.

Capizzi, G., Masarotto, G., 2011. A least angle regression control chart for multidimensional data. Technometrics 53, 285-296.

Chen, L., Bai, Z.P., Kong, S.F., Han, B., You, Y., Ding, X.A., Du, S.Y., Liu, A.X., 2010. A land use regression for predicting NO2 and PM10 concentrations in different seasons in Tianjin region, China. Journal of Environmental Sciences 22, 1364-1373.

Dirgawati, M., Barnes, R., Wheeler, A.J., Arnold, A.-L., McCaul, K.A., Stuart, A.L., Blake, D., Hinwood, A., Yeap, B.B., Heyworth, J.S., 2015. Development of Land Use Regression models for predicting exposure to NO2 and NOx in Metropolitan Perth, Western Australia. Environmental Modelling & Software 74, 258-267.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. 407-499.

Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Statistical Science 11, 89-121.

Gaeta, A., Cattani, G., di Bucchianico, A.D.M., De Santis, A., Cesaroni, G., Badaloni, C., Ancona, C., Forastiere, F., Sozzi, R., Bolignano, A., 2016. Development of nitrogen dioxide and volatile organic compounds land use regression models to estimate air pollution exposure near an Italian airport. Atmospheric Environment 131, 254-262.

Gulliver, J., de Hoogh, K., Hoek, G., Vienneau, D., Fecht, D., Hansell, A., 2016. Back-extrapolated and year-specific NO 2 land use regression models for Great Britain-Do they yield different exposure assessment? Environment international 92, 202-209.

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. Environmental science & technology 41, 2422-2428.

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. Atmospheric Environment 42, 7561-7578.

Hoek, G., Meliefste, K., Cyrys, J., Lewne, M., Bellander, T., Brauer, M., Fischer, P., Gehring, U., Heinrich, J., van Vliet, P., Brunekreef, B., 2002. Spatial variability of fine particle concentrations in three European areas. Atmospheric Environment 36, 4077-4088.

Johnson, M., Isakov, V., Touma, J.S., Mukerjee, S., Ozkaynak, H., 2010. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. Atmospheric Environment 44, 3660-3668.

Kim, Y., Guldmann, J.M., 2015. Land-use regression panel models of NO2 concentrations in Seoul, Korea. Atmospheric Environment 107, 364-373.

Knibbs, L.D., Hewson, M.G., Bechle, M.J., Marshall, J.D., Barnett, A.G., 2014. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. Environ. Res. 135, 204-211.

Kutner, M.H., Nachtsheim, C.J., Neter, J., 2004. Applied Linear Regression Models, 4 ed. McGraw-Hill Irwin.

Laiman, R., He, C.R., Mazaheri, M., Clifford, S., Salimi, F., Crilley, L.R., Mokhtar, M.A.M., Morawska, L., 2014. Characteristics of ultrafine particle sources and deposition rates in primary school classrooms. Atmospheric Environment 94, 28-35.

Lee, J.-H., Wu, C.-F., Hoek, G., de Hoogh, K., Beelen, R., Brunekreef, B., Chan, C.-C., 2014. Land use regression models for estimating individual NOx and NO2 exposures in a metropolis with a high density of traffic roads and population. Science of The Total Environment 472, 1163-1171.

Li, L., Wu, J., Wilhelm, M., Ritz, B., 2012. Use of generalized additive models and cokriging of spatial residuals to improve land-use regression estimates of nitrogen oxides in Southern California. Atmospheric Environment 55, 220-228.

Madsen, C., Carlsen, K.C.L., Hoek, G., Oftedal, B., Nafstad, P., Meliefste, K., Jacobsen, R., Nystad, W., Carlsen, K.-H., Brunekreef, B., 2007. Modeling the intra-urban variability of outdoor traffic pollution in Oslo, Norway—A GA 2 LEN project. Atmospheric Environment 41, 7500-7511.

Madsen, C., Gehring, U., Håberg, S.E., Nafstad, P., Meliefste, K., Nystad, W., Lødrup Carlsen, K.C., Brunekreef, B., 2011. Comparison of land-use regression models for predicting spatial NOx contrasts over a three year period in Oslo, Norway. Atmospheric Environment 45, 3576-3583.

Marcon, A., de Hoogh, K., Gulliver, J., Beelen, R., Hansell, A.L., 2015. Development and transferability of a nitrogen dioxide land use regression model within the Veneto region of Italy. Atmospheric Environment 122, 696-704.

Meng, X., Chen, L., Cai, J., Zou, B., Wu, C.-F., Fu, Q., Zhang, Y., Liu, Y., Kan, H., 2015. A land use regression model for estimating the NO2 concentration in shanghai, China. Environmental Research 137, 308-315.

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). Atmospheric Environment 45, 4412-4420.

Rogerson, P., 2001. Chapter 7.1. Multiple Regression, Statistical methods for geography. Sage.

Rose, N., Cowie, C., Gillett, R., Marks, G.B., 2010. Validation of a spatiotemporal land use regression model incorporating fixed site monitors. Environ. Sci. Technol. 45, 294-299.

Rose, N., Cowie, C., Gillett, R., Marks, G.B., 2011. Validation of a Spatiotemporal Land Use Regression Model Incorporating Fixed Site Monitors. Environmental Science & Technology 45, 294-299.

Salimi, F., Mazaheri, M., Clifford, S., Crilley, L.R., Laiman, R., Morawska, L., 2013. Spatial Variation of Particle Number Concentration in School Microscale Environments and Its Impact on Exposure Assessment. Environmental Science & Technology 47, 5251-5258.

Vienneau, D., de Hoogh, K., Bechle, M.J., Beelen, R., van Donkelaar, A., Martin, R.V., Millet, D.B., Hoek, G., Marshall, J.D., 2013. Western European land use regression incorporating satellite-and ground-based measurements of NO2 and PM10. Environ. Sci. Technol. 47, 13555-13564.

Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., Brunekreef, B., 2012. Systematic evaluation of land use regression models for NO2. Environmental science & technology 46, 4481-4489.

WHO, 2014. Burden of disease from the joint effects of Household and Ambient Air Pollution for 2012, World Health Organization. Available in- http://www.who.int/phe/health_topics/outdoorair/databases/FINAL_HAP_AAP_BoD_24March2014.pdf?ua=1.

Wilton, D., Szpiro, A., Gould, T., Larson, T., 2010. Improving spatial concentration estimates for nitrogen oxides using a hybrid meteorological dispersion/land use regression model in Los Angeles, CA and Seattle, WA. Science of The Total Environment 408, 1120-1130.