Science of the Total Environment 610–611 (2018) 1439–1447

Contents lists available at ScienceDirect

Science of the Total Environment

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

Land use regression modelling estimating nitrogen oxides exposure in industrial south Durban, South Africa

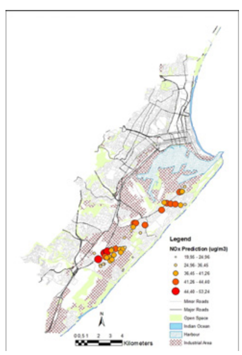
Sheena Muttoo [a,*], Lisa Ramsay [b], Bert Brunekreef [c], Rob Beelen [d], Kees Meliefste [c], Rajen N. Naidoo [a]

[a] Discipline of Occupational and Environmental Health, School of Nursing and Public Health, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa
[b] School of Agricultural, Earth and Environmental Sciences, University of Kwa-Zulu Natal, Durban, South Africa
[c] Institute for Risk Assessment Sciences, Utrecht University, The Netherlands
[d] National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

HIGHLIGHTS

- Measured NO_x levels were observed to be higher during the winter sampling campaign as compared to summer.
- Measured and modelled mean NO_x levels were highly correlated, thus illustrating the strength of the model to accurately predict exposure at un-monitored locations.
- Fewer than the recommended 40 NO_x sampling sites were sufficient for LUR model development for this study.
- This study indicates that ambient NO_x levels are strongly influenced by levels of local traffic.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 20 April 2017
Received in revised form 13 July 2017
Accepted 31 July 2017
Available online 14 September 2017

Editor: D. Barcelo

Keywords:
Air pollution
Exposure assessment
Nitrogen oxides
Land use regression modelling
South Durban

ABSTRACT

Background: The South Durban (SD) area of Durban, South Africa, has a history of air pollution issues due to the juxtaposition of low-income communities with industrial areas. This study used measurements of oxides of nitrogen (NO_x) to develop a land use regression (LUR) model to explain the spatial variation of air pollution concentrations in this area.
Methods: Ambient NO_x was measured over two two-week sampling periods at 32 sites using Ogawa badges. Following the ESCAPE approach, an annual adjusted average was calculated for these results and regressed against pre-selected geographic predictor variables in a multivariate regression model. The LUR model was then applied to predict the NO_x exposure of a sample of pregnant women living in South Durban.
Results: Measured NO_x levels ranged from 22.3–50.9 μg/m^3 with a median of 36 μg/m^3. The model developed accounts for 73% of the variance in ambient NO_x measurements using three input variables (length of minor roads within a 1000 m radius, length of major roads within a 300 m radius, and area of open space within a 1000 m radius). Model cross validation yielded a R^2 of 0.59. Subsequent participant exposure estimates indicated exposure to ambient NO_x ranged from 19.9–53.2 μg/m^3, with a mean of 39 μg/m^3.
Discussion and Conclusion: This is the first study to develop a land use regression model that predicts ambient concentrations of NO_x in a South African context. The findings of this study indicate that the participants in the South Durban are exposed to high levels of NO_x that can be attributed mainly to traffic.

* Corresponding author.
E-mail address: muttoos@ukzn.ac.za (S. Muttoo).

http://dx.doi.org/10.1016/j.scitotenv.2017.07.278

1. Introduction

Studies have shown that long term exposure to air pollution, including ambient nitrogen oxides (NOx), has been linked to a varying degree of health end points ranging from modest respiratory changes and impaired pulmonary functions to cases of mortality (Uea, 2004). Research has further shown definitive associations between air pollution exposure and elevated risk of adverse pregnancy outcomes as well as early childhood health effects (Lacasana et al., 2005; Aguilera et al., 2010; Brauer et al., 2008; Morgenstern et al., 2007; Slama et al., 2007). A key challenge for environmental epidemiology is deriving precise measures of individual exposure for the characterisation of exposure-response relationships.

To address such challenges, the development of exposure models to predict the spatial variation of air pollution and assess individual exposure has been identified as a research priority (Brunekreef and Holgate, 2002). Existing exposure assessment methods include proximity-based estimates, dispersion models, land use regression (LUR) models and interpolation methods such as kriging.

Proximity-based estimates have been associated with some health effects, but present a risk of exposure misclassification as meteorology and source characteristics are not taken into consideration (Jerrett et al., 2005). Kriging, when applied at the intra-urban scale, has been known to produce notable variation in air pollutant concentrations at very short distances (Briggs, 2005) and is more effective at a regional or national scale (Bell, 2006; Liao et al., 2006). Dispersion models could potentially incorporate both spatial and temporal variation without the need for additional air pollution monitoring but the high cost of implementation, input data demands and required expertise limit its application (Jerrett et al., 2005; Briggs et al., 1997).

LUR modelling has emerged as an effective alternative to these approaches. This model relates monitored pollutant concentrations to site-specific geographic predictors such as traffic, land use, topography and meteorology in a multivariate regression model. The parameter estimates derived from the regression equation are then used to estimate pollutant levels at unmonitored locations. The use of site-specific variables in this method has been known to effectively capture small-scale variability (Ryan and LeMasters, 2007).

The LUR method, first introduced in the SAVIAH (Small Area Variations in Air pollution and Health) Study (Briggs et al., 1997), has since been applied in a number of studies in Europe (Brauer et al., 2003; Madsen et al., 2007; Brauer et al., 2007; Rosenlund et al., 2008), North America (Brauer et al., 2007; Gilbert et al., 2005; Wheeler et al., 2008; Arain et al., 2007; Ross et al., 2005; Ross et al., 2007; Sahsuvaroglu et al., 2006) and Asia (Chen et al., 2010a; Chen et al., 2010b; Kingham et al., 2013; Lee et al., 2013) to model urban air pollution for exposure assessment, with the most recent large-scale application in the European Study of Cohorts for Air Pollution Effects (ESCAPE) (Eeftens et al., 2012). Most studies use nitrogen dioxide (NO_2) as a proxy for traffic emissions as this pollutant correlates well with traffic densities and is relatively simple and inexpensive to measure (Sahsuvaroglu et al., 2006). Frequently used predictor variables include traffic variables, population and household density, land use, altitude, topography and location (Ross et al., 2007; Sahsuvaroglu et al., 2006; Gilliland et al., 2005; Moore et al., 2007; Rosenlund et al., 2007; Bertazzon et al., 2015). The variable selection process relies on data availability, accessibility and completeness (Brauer et al., 2003; Madsen et al., 2007; Brauer et al., 2007; Hoek et al., 2008).

South Africa, as a developing country, experiences multiple challenges in mitigating environmental impacts, such as air pollution, while promoting economic development and employment. Durban is South Africa's second largest city by population and has the country's largest and busiest port. Apartheid era policy emphasis on strategic industries facilitated the growth of the paper, petroleum and chemical industries in the South Durban (SD) region. There are approximately six hundred industrial facilities in the region today, including textile, food and beverage, printing, paint, metal smelting, galvanising, fibre and automotive facilities. Further emission sources include a sewage treatment works, a large container terminal and chemical storage facility alongside the port (Guastella and Knudsen, 2007) and dense traffic (including heavy trucking) nodes (Naidoo et al., 2007).

Apartheid planning juxtaposed industrial activity with densely populated labour suburbs in the city, creating a unique air pollution exposure context. Despite the end of political apartheid, these communities remain within industrial regions today, simultaneously contesting environmental degradation while demanding local job opportunities. The South Durban Community Environmental Alliance (SDCEA) have for over two decades, taken the lead on environmental activism efforts in this region, with predominant focus on the impacts associated with the major industries (the oil refineries and paper mill) (Ramsay and Naidoo, 2012).

It is this combination of multiple polluting sources in the basin-like topography of South Durban that poses adverse public health impacts as highlighted in previous studies (Naidoo et al., 2007; Kistnasamy et al., 2008; Naidoo et al., 2013). Prevailing winds carry gaseous pollution, soot and oil spray from industries in the direction of local residences and schools (Leonard et al., 2008). Pollution stagnation in the region during surface temperature inversions in winter are well understood (Kistnasamy et al., 2008). Furthermore, the coastal context creates intermittent fumigation scenarios when emissions from elevated sources descend rapidly to accumulate at the surface.

A lack of exposure data for this region, particularly for vulnerable groups such as pregnant women, is a limitation to assessing the burden of impact on public health. In an attempt to characterise exposure and better understand the relationship between air pollution exposure and health outcomes with the context of prevailing geographical and environmental factors, the aim of this study was to develop a LUR model and derive exposure estimates for NO_x among sample of pregnant women residing in the South Durban region. The exposure estimates from the current study will be used in the Mother and Child in the Environment (MACE) birth cohort to assess the association between maternal pre-natal exposure to air pollution and health outcomes in neonates.

2. Methodology

The study area is located within the South Durban region of the eThekwini Municipality of KwaZulu-Natal, South Africa. The estimated population for this region is 595,601 (South Africa - Ethekwini Population Census, 2011) and includes the suburbs of Merebank, Wentworth, Austerville and Bluff (Fig. 1 below and Fig. S1, in the supplementary document).

2.1. Sample sites and sampling periods

Sampling sites were selected using Google Earth's® software (detailed in supplementary document). NO_x sampling locations were selected following the ESCAPE (Beelen et al., 2000) protocol, however streetlight poles were selected as opposed to building facades of homes due to restricted access to private property. Given that the distance between residential building facades and streetlight poles was minimal (<2 m), this was unlikely to a significant impact on measured concentrations. The monitoring locations were evenly distributed within the subareas (see Fig. S1 in supplementary document), with one sampler colocated at an Air Quality Monitoring Station (AQMS) for each sampling period. Samplers were distributed to represent street level concentrations as well as background concentrations.

There is no standardized methodology in the literature for determining the required number of sampling sites for model development. The use of 40 sampling sites is recommended by ESCAPE (Beelen et al., 2000). Other studies (Chen et al., 2010a; Briggs et al., 2000) however, have reported statistically significant models using fewer sites, ranging from 25 to 35. It has been noted that factors such as size of the study area, traffic intensity, level of urbanization, level of urban

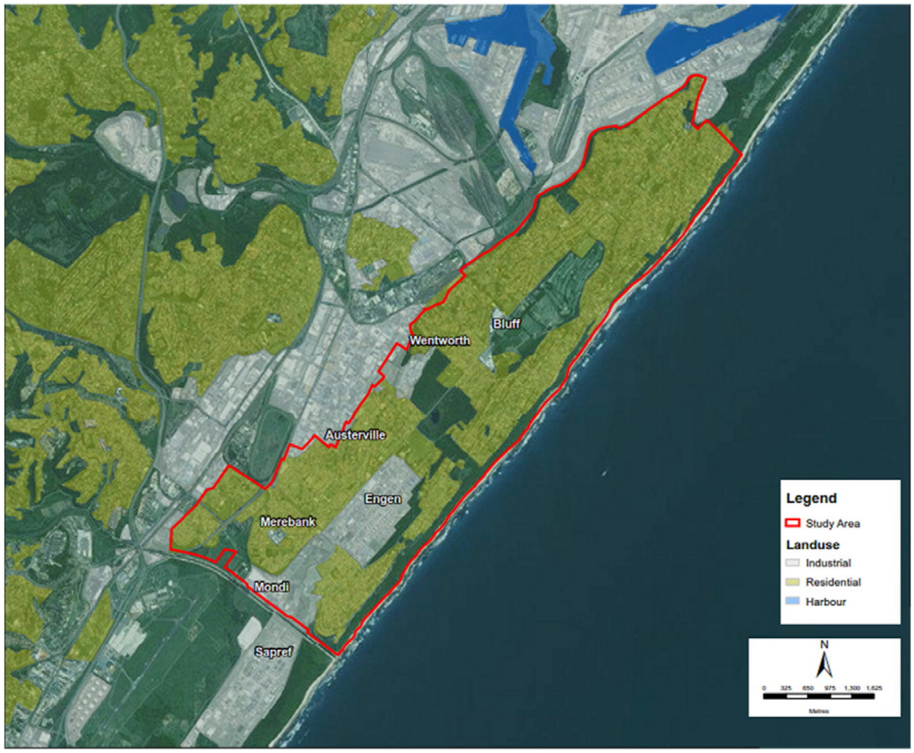


Fig. 1. The study area in South Durban (Map data©, 2016).

industrialization, topography and meteorology are important considerations when deciding on the number of samplers to be used (Kashima et al., 2009). The South Durban study domain is smaller than most documented in the literature. The explained variance (R^2) values were calculated for LUR models using an increasing number of the South Durban monitoring samplers (i.e. 10, 20, 25, 29, 30 and 31 sites). There was not a significant increase in explanatory power of the model. An increasing number of samplers is useful to capture the range of values in the various geographical variables to limit the need for truncation when the model is applied.

The South Durban region has two distinct meteorological regimes (one distinctly summer and one distinctly winter) with transitions between these in spring and autumn (Tyson et al., 2004). To capture variability across these two scenarios, two sampling periods were selected, specifically mid-summer and mid-winter.

2.2. Nitrogen oxides measurements

NO_x measurements were conducted over two two-week periods during mid-winter from 21 July to 3 August 2011 and mid-summer from 24 January to 6 February 2012. Measurements at the sites were conducted simultaneously for all sampling periods to avoid the influence of meteorological and emission variations.

Ogawa passive samplers were deployed at a height of 2.5 m using street light poles and traffic signs. At the end of sampling, the collected badges were refrigerated at 4 °C prior to its courier to Utrecht, The Netherlands. Laboratory analysis was conducted at the Institute for Risk Assessment Sciences (IRAS) laboratory facility, following the Ogawa protocol (Ogawa & Co., Inc., V3.98, USA). This procedure entailed adding a reagent to the reactive filters, upon which the developed colour was measured by 540 nm on a spectrophotometer. Concentrations of

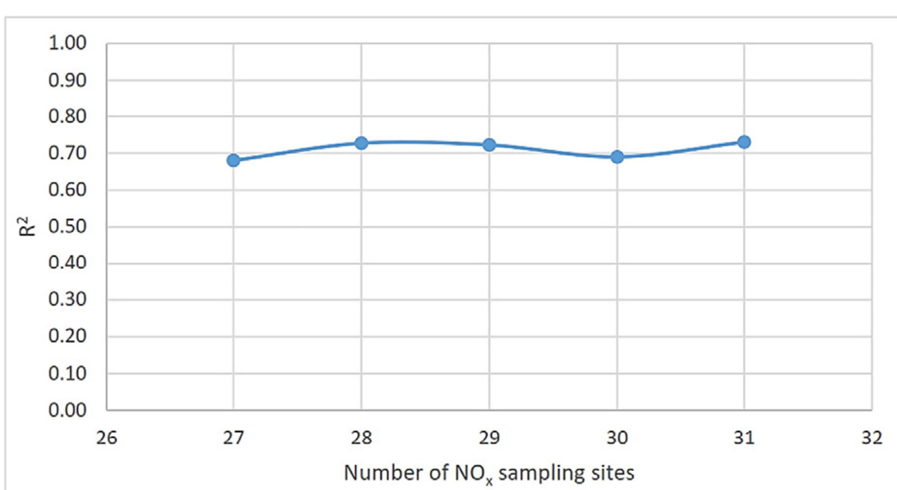


Fig. 2. Number of NO_x monitoring sites used for model development and the proportion of variability (R^2) in the 31 monitoring sites.

NO_x were calculated by means of linear regression of a five-point calibration line and corrected for temperature and relative humidity during sampling (Madsen et al., 2007).

For quality control purposes, field blanks were maintained during the deployment and collection of badges. There were three field blanks maintained for each sampling period, however in the first sampling period, one field blank was lost at collection. A limit of detection (LOD) was calculated as three times the standard deviation of the field blanks for each sampling period (Madsen et al., 2007; Van Roosbroeck, 2007; Madsen et al., 2011a). Duplicate badges (three per sampling period) were deployed at selected sites for measurement validation purposes. NO_x results at duplicate sites were averaged to provide a single estimate at these sites. Badges were co-located at the Wentworth Reservoir Air Quality Monitoring Station (AQMS) during both sampling periods and the results compared. These continuous monitoring stations are managed by the eThekwini Municipality's Air Pollution Unit using international standards of quality control and validation (EThekwini Air Quality Monitoring Network Annual Report, 2009). The AQMS data at the validation site for the period March 2011 to February 2012 was then used to adjust the monitored concentrations to an annual average of NO_x concentrations and this adjustment was applied to all samplers (Beelen and Hoek, 2010).

To determine the annual average for each sampling site, the results from the two sampling periods had to be adjusted using data from an existing AQMS station consistent with the sampling periods. As per the requirements stated in the ESCAPE (Beelen and Hoek, 2010) protocol, the station selected was required to have 75% data capture for the study period (Eeftens et al., 2012), hence the Wentworth Reservoir AQMS was selected based on its location and the high percentage of data capture. The arithmetic mean of the available measurements (i.e. both sampling periods) per site was adjusted using the difference between the sampling period and the annual average at the AQMS (see Annexure B in supplementary document), thus deriving an annual adjusted average of NO_x.

2.3. Geographic predictor variables

The selection of geographic variables was based on a desktop assessment of the local environment, and a review of variables successfully applied in previously published LUR studies (Madsen et al., 2007; Brauer et al., 2007; Sahsuvaroglu et al., 2006; Aguilera et al., 2007). The selected geographic predictors were grouped into the following broad categories: road length/type, land use, physical geography, population and housing density.

Geographical Information System (GIS) shapefiles were obtained from the eThekwini Corporate GIS Unit and GIS analyses were performed using ArcGIS ver 9.3.1, following the ESCAPE (Beelen and Hoek, 2010) procedure. These GIS shapefiles were developed by the local authority for planning purposes. Regular updates to the system are made by GIS specialists within the Corporate GIS Unit in line with rezoning's and local infrastructural and residential developments.

The selection of buffer radii, was also based on previous LUR models, taking into account local surrounding factors such as urbanization, urban industrialisation and topography. These included buffer distances in the ranges of 50, 100, 300, 500 and 1000 m for road length accounting for urban background concentrations of NO_x attributed to traffic emissions and, all remaining variables had buffer distances of 100, 300, 500, 1000 and 2000 m. (See Table 1.)

2.4. Exposure modelling

Supervised forward stepwise regression techniques were used to develop the prediction model using a previously published approach (Beelen et al., 2013). As per the ESCAPE (Eeftens et al., 2012) methodology, *a priori* definitions (i.e. direction of effect) were established. The variance inflation factor (VIF) was calculated as an indication of how much inflation of standard error was caused by collinearity between predictor variables (Madsen et al., 2007), and was considered acceptable if <5 (Beelen and Hoek, 2010). These approaches were used to address potential collinearity (Wheeler et al., 2008) and improve model interpretability (Isakov et al., 2010).

For the model development univariate regression analysis was conducted and the model with the highest adjusted explained variance (R^2) was regarded as the 'start model'. In the next step predictor variables with the highest additional increase in adjusted R^2 were maintained in the model provided they fulfilled three criteria: (Uea, 2004) the increase in adjusted R^2 was $>1\%$, (Lacasana et al., 2005) the coefficient conformed to a priori definitions (Aguilera et al., 2010) the direction of effect for predictors already included in the model did not change. The addition of variables in a supervised stepwise process was repeated until there were no remaining predictor variables that added $>1\%$ to the adjusted R^2 of the previous regression model. To evaluate the significance of the variables in the model, variables with *p*-value >0.10 were sequentially removed from the model, starting with the least significant, until all predictor variables in the 'final model' had a $p \leq 0.10$. In addition the model was tested for influential variables using the Cook's D test in which results <1 were considered acceptable.

To validate the model the leave-one-out cross validation (LOOCV) technique was used, in which the model was developed for $n - 1$ sites and the levels predicted for the excluded site and this was repeated n number of times. The mean difference between predicted and measured values estimated the model error (Brauer et al., 2007). This was determined in STATA ver.11 (Stata Corporation, USA), using the ESCAPE (Beelen and Hoek, 2010; Eeftens et al., 2012) approach.

To assess the impact of any potential seasonal influences, seasonal models were developed for each of the sampling periods (i.e. winter and summer) using actual NO_x measurements obtained during the sampling campaigns for each season.

2.5. Participant exposure assessment

The developed LUR model was then used to predict exposure at the residential addresses of a sample of study participants from the MACE cohort. For each address GPS coordinates were obtained and further data was collected on the significant geographic predictors identified in the final model. Exposure estimates for individual participants was determined by applying the model intercept and regression coefficients of the LUR model to the selected geographic predictors.

Over (or under) prediction at participant addresses may occur when using predictor variable values that are outside the range of predictor variables within the location of monitoring sites. To avoid this, the former were truncated by assigning the maximum (or minimum) value that occurred at one of the monitoring sites for that specific variable and this was performed for all values above the highest (or below the lowest) values at the monitoring sites (Beelen and Hoek, 2010).

3. Results

3.1. Nitrogen oxides measurements

Summary statistics of NO_x results for each of the seasonal sampling periods are presented in Table 2. A high correlation ($R^2 = 0.99$) was observed for the duplicate samplers during both sampling periods (Table S1 in supplementary document). For passive samplers that were co-located with an existing AQMS, differences of between 8 and 10 $\mu g/m^3$ were observed (Table S2 in supplementary document). There was an overall correlation of 61% between measurements observed during both sampling periods. (See Table 4.)

After adjustment of passive sampling results, there were valid measurements available for 31 sites (there was one and three sampler/s lost in the first and second sampling periods respectively, resulting in an

Table 1
Brief description of GIS analysis of geographic predictors and 'a priori' definitions.

Variable type	Variables (unit)	Buffers (m)	Rationale for inclusion	GIS analyses[a]	*A priori*[b]
Road length/type	Length of:		Road length is directly associated with traffic	Update of geometry in attribute table	
	major road[c] (m)	50, 100, 300, 500, 1000	patterns, thus influencing pollutant emissions	of selected feature to calculate length	+
	minor road[de*] (m)	50, 100, 300, 500, 1000	from traffic.		+
Land use	Area (m) of:		Different land uses yield different levels of	Update of geometry in attribute table	
	Industrial land use (m^2)	100, 300, 500, 1000, 2000	pollutant emission or transportation activities[5]	of selected feature to calculate area	+
	Open space land use (m^2)	100, 300, 500, 1000, 2000			−
	The harbour[e] (m^2)	1000, 2000			+
Physical geography	Elevation (m)	–	The height of polluting sources influences emission levels.	Square root calculation	−
Population & housing data	population density (km^2)	100, 300, 500, 1000, 2000	Pollutant contributions from commuting,	Area weighted calculation	+
	household density (km^2)	100, 300, 500, 1000, 2000	travelling (traffic), heating & combustion activities (indoor sources)[5]		+

[a] All GIS data were obtained from the eThekwini Corporate GIS Unit and GIS analyses were performed using ArcGIS ver 9.3.1, following the ESCAPE (Beelen and Hoek, 2010) procedure.
[b] a priori definitions based on the definitions used in the ESCAPE (Beelen and Hoek, 2010) study.
[c] Major roads – freeway, arterials and collectors (Classifications of Roads and Intersections, 2011).
[d] Minor roads – local street and cul-de-sac (Classifications of Roads and Intersections, 2011).
[e] data was only available for buffers of 1000 and 2000 m.

average of 31 sites), the adjusted annual average concentration for NO_x ranged from 22.3–50.9 μg/m^3 with a median of 36.03 μg/m^3.

3.2. Geographic predictor variables

Summary data for geographic predictor variables at the measurement sites are presented in Tables S3 of the supplementary document. The mean length of minor roads in all buffer categories (120.6–28,182.9 m) were significantly higher than major roads (0–1938.2 m) and area of industrial land use within 2000 m (3,315,988 m^2) was significantly higher than open space within the same buffer radius (1,210,818 m^2). Data availability for area of harbour was limited to buffer distances of 1000 m and 2000 m with means of 3536.8 m and 100,401.2 m respectively.

3.3. Model building and validation

A total of 33 variables were individually regressed against NO_x concentrations in the univariate analysis. In the stepwise linear regression (Table S4 in supplementary document) the start model yielded an adjusted R^2 of 0.41 and this sequentially increased to 0.73 as non-significant variables (those with beta coefficients that did not correlate with the pre-defined a priori definitions and those which had *p*-values >0.1) were removed from the model in a stepwise manner.

In the final model (Table 3), significant predictor variables identified included minor road length within 1000 m, major road length within 300 m and open space within 1000 m. VIF were considered acceptable for these variables (<5) (Beelen and Hoek, 2010). No outliers were identified in the Cook's distance test (Fig. 3). Open space within 1000 m ($r = -0.6$) was negatively associated with NO_x, while minor road length within 1000 m ($r = 0.65$) and major road length within 300 m ($r = 0.5$) were positively associated with NO_x. (See Table 3.)

Table 2
Descriptive summary of NO_x passive sampling results by sampling period.

	SD_SP1* ($N = 30$)	SD_SP2^ ($N = 28$)
Mean (μg/m^3)	54.9	13.9
Median (μg/m^3)	54.5	13.8
Standard deviation	13.7	3.3
Min (μg/m^3)	33.6	7.7
Max (μg/m^3)	84.4	21.7

SD – South Durban; SP1 – Mid-winter; SP2 – Mid-summer.
*Field blanks ($N = 2$): mean (μg/m^3) – 0.64; Standard Dev – 0.01; Limit of Detection (3Xsd) = 0.03.
^Field blanks ($N = 2$): mean (μg/m^3) – 0.98; Standard Dev – 1.4; Limit of Detection (3Xsd) = 4.2; Missing badges -2.

The model was validated using the leave one out cross validation technique, in which the final model was developed for 31 – 1 sites and predictions were made for the excluded site and this was repeated 31 times. The R^2 for the cross validation was 0.59, (Fig. 4) and the root mean square error (RMSE), was 5.16.

To further evaluate the impact of potential seasonal differences, independent models were developed for each season, with the winter model achieving an R^2 of 0.68 (Table S5 in supplementary document) and the summer model an R^2 of 0.59 (Table S6 in supplementary document). It was further noted that the winter model had the same significant predictor variables as the main model (i.e. length of minor road within 1000 m, length of major road within 300 m and area of open space within 1000 m). While length of minor road within 1000 m was a common factor among all models, household density within 2000 m and length of minor road within 50 m were relevant to the summer model. The combination of the seasonal measurements in the final model resulted in improved model performance ($R^2 = 0.73$).

3.4. Model application: participant exposure assessment

The final model was applied to 44 participant addresses. Truncation was only applied to the variable major road within 300 m. The correlation between the untruncated and truncated model was observed to be high at 0.99. (See Table 4.)

The gradient of participant NO_x exposure estimates is illustrated in Fig. 5.

4. Discussion and conclusion

The LUR model developed for the South Durban region accounts for 73% of the variation in the observed NO_x levels at 31 sites. Significant variables included length of minor road within 1000 m, major road within 300 m (positive association with NO_x) and open space within 1000 m (negative association with NO_x). With the known health impacts

Table 3
Summary of regression model predicting NO_x.

Variable	Unit	β	SE	t	P
Intercept	–	18.81385	6.72687	2.8	0.009
Length of minor road within 1000 m	m	0.0006708	0.0002085	3.22	0.003
Length of major road within 300 m	m	0.0082245	0.0025539	3.22	0.003
Area of open space within 1000 m	m^2	−9.08E−06	5.13E−06	−1.77	0.088

$R^2 = 0.73$.
Root Mean Square Error = 4.97.

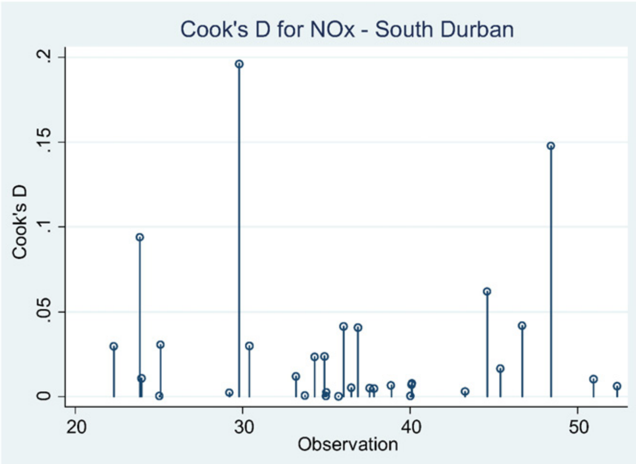


Fig. 3. Cook's D test for NO_x.

Table 4
Measured and predicted nitrogen oxides levels.

	Measured[a] NO_x (μg/m^3)			Predicted[b] NO_x (μg/m^3)		
	N	Mean	Range	N	Mean	Range
South Durban	31	36.35	22.91–50.94	44	38.97	19.94–53.24

[a] Sampling sites.
[b] Participant addresses.

associated with NO_x exposure, a broadening in focus towards the control of traffic emissions and protection of local open spaces may assist in protecting local environmental health.

The significant variables identified in this study are comparable to previous studies (Madsen et al., 2007; Aguilera et al., 2007; Madsen et al., 2011b) in which length of a major road is the common factor. It was further noted that in the ESCAPE project, LUR models developed for NO_x among 36 study areas had model explained variances (R^2) of 55% to 92% in which traffic intensity/load was among the significant predictors variables (Beelen et al., 2013). In the current study, there were challenges in attaining traffic count data as municipal records for this were not available from the relevant traffic authority. This was thus not included in model development. Road length classifications were used as a surrogate for traffic, as done in previous studies (Brauer et al., 2003; Madsen et al., 2007; Brauer et al., 2007; Hoek et al., 2008).

Other challenges faced included determining the number of samplers required. The recommended number varied across studies, and there was no standardized statistical methodology in the literature for calculating this value. It has been recommended that a minimum number of 40 sites be used (Hoek et al., 2008) to minimize the risk of exposure misclassification as a result of under sampling, however other studies have reported using fewer sampler sites while still achieving robust model development (Chen et al., 2010a). As indicated in Fig. 2, 31 sites were sufficient for this assessment with a clear plateauing of the R^2 curve. This suggests that while 40 sites might have improved the strength of the model, this would not have resulted in a substantial change in our

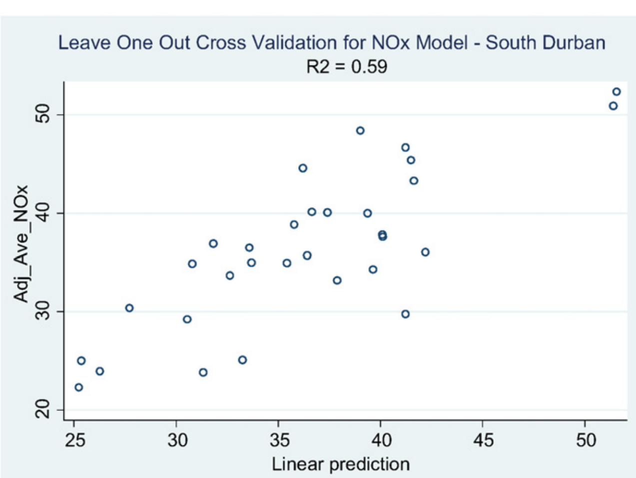


Fig. 4. Leave one out cross validation for NO_x.

estimates. This is possibly due to the smaller study area than conventionally associated with LUR, and the effective placement of the 31 monitors to cover the range of local geographical input variables.

The number of sampling campaigns to be conducted has also varied between studies in different regions (Hoek et al., 2008; Lebret et al., 2000; Sahsuvaroglu et al., 2006). It has been previously reported that geographic pollution patterns remain relatively stable throughout the year, as areas of high pollution tend to remain in the same location. Thus it is possible that as little as one-short term monitoring session could potentially adequately estimate mean annual concentrations (Lebret et al., 2000) of NO_x. In this study, two short-term monitoring campaigns (summer and winter) were conducted to capture seasonal variation.

Despite the complex industrial profile of the South Durban area, the model developed comprised just three variables (two traffic proxy variables and open space). As with previous studies (Wheeler et al., 2008; Chen et al., 2010a; Lee et al., 2014; Raux et al., 2015) NO_x levels were mostly influenced by proxy traffic markers such as length of road. Elevation, though considered an important influential factor for coastal cities, did not emerge as a significant predictor variable in this study. This is mostly due to the limited changes in elevation in the study area. Industrial land use coverage showed no significant association with NO_x levels, which could be partially attributable to the lack of distinction made between industries which emit NO_x and those that do not (Gilbert et al., 2005). Furthermore, industrial emissions generally are emitted at height, and complex dispersion factors play a role in their surface impact. The impact at ground level as a result of these emissions thus may occur at some distance from the source, rendering industrial proximity variables insufficient to explain spatial patterns in ground-level concentrations. The results, however, highlight that ambient NO_x in the region is predominantly the result of traffic emissions, which distinguish it from other pollutants such as sulphur oxides (SO_x), which are predominantly industrial emissions in the region. LUR models developed for such pollutants would require more careful assessment of how industrial variables should be included as potential predictor variables.

Although there were significant differences in measurements of NO_x between seasons in this study, the development of the seasonal models showed no significant differences in model performance (winter $R^2 = 0.68$, summer $R^2 = 0.59$ and the combined model $R^2 = 0.73$). This is the result of the high correlation (61%) between seasonal measurements for each of the measurement sites. The key meteorological difference across seasons in South Durban is the significant narrowing of the mixing layer during winter (Tyson et al., 2004; Tularam and Ramsay, 2013). This arises through radiative surface inversions with night-time cooling and non-surface subsidence inversions that are intensified during pre-frontal conditions (Preston-Whyte and Diab, 1980). This narrowing depth is a feature occurs across the south Durban region without spatial variations in the impact between sampling sites. As a result, pollutant concentrations are higher in winter relative to summer across sites but the spatial distribution of concentrations is largely unaffected. Minor differences in the spatial distribution are likely the result of changes in the wind regime between summer and winter. The prevailing winds across seasons are north-easterly and south-westerly winds, but with an increase in frequency and strengthening of the south-westerly component in winter (due to frontal systems), and an increase in the easterly component in summer (as a sea breeze) (Tyson et al., 2004; Tularam and

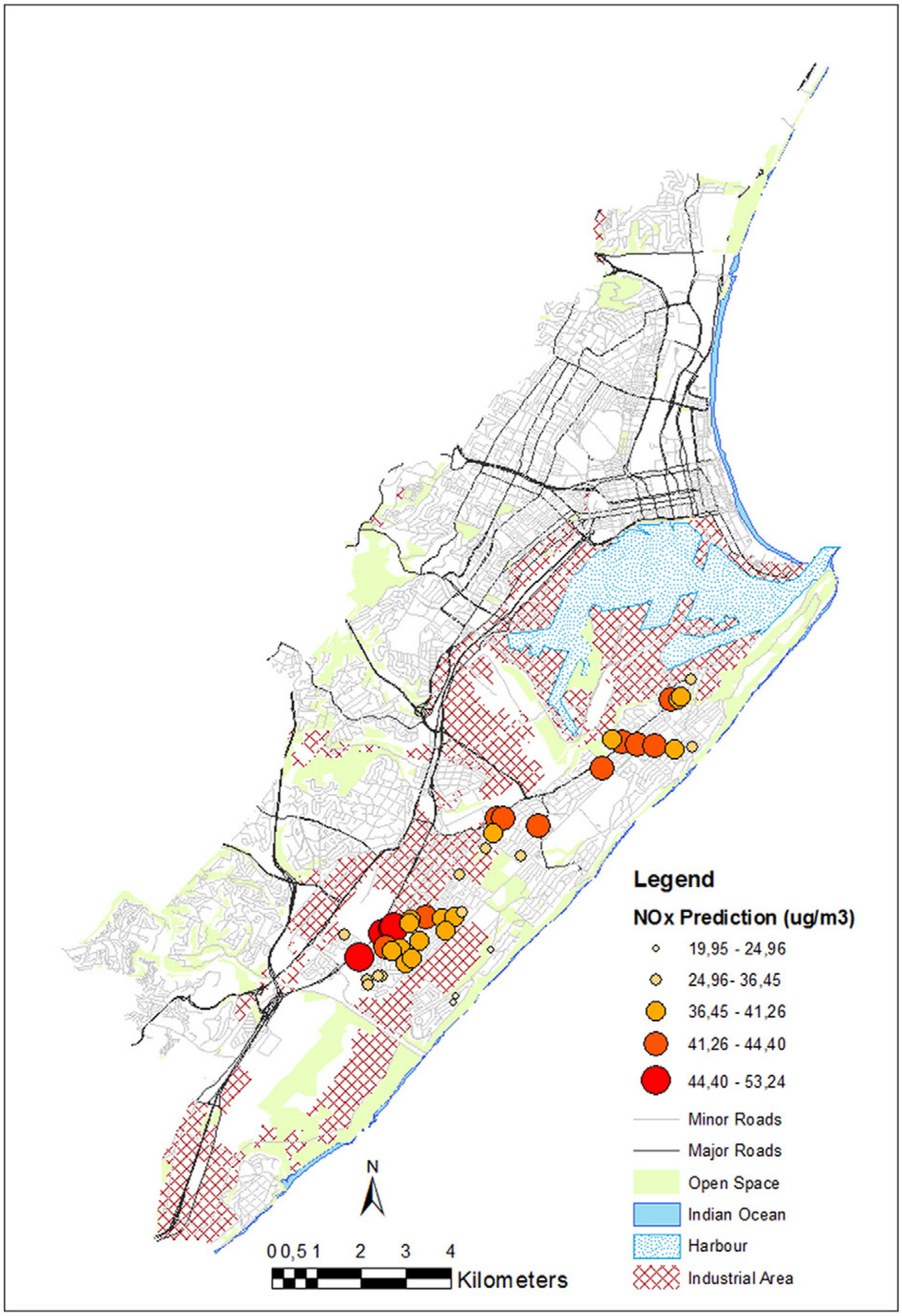


Fig. 5. NO_x estimation at participant address points in South Durban.

Ramsay, 2013). We acknowledge the limitation in this study of the absence of meteorological input variables, however the above suggests that the model would not have been significantly strengthened with their inclusion.

In a similar study conducted in Taipei City, Taiwan (Lee et al., 2013), the model developed included five variables in categories of road length as well as urban greenery and natural area accounting for 81% of the explained variance in ambient NO_x levels. This study further assessed the spatial variation within the metropolis and attributed the high concentrations observed to traffic emissions from the dense road networks likely to have been trapped within the basin like topography (Lee et al., 2013). The metropolis of Taipei with an area of 36,000 km^2 reported a routine air quality monitoring network comprising of 15 stations measuring NO_2 & NO_x in comparison to an area of approximately 100km^2 in South Durban with 4 AQMS locations. For the Taipei study 40 sampling sites were selected, relative to a population size of 23 million, compared to the 32 sites selected in our study with an approximate population of 600,000, thus suggesting an acceptable network of measurement sites. In Taipei, measurements were obtained over three seasons (classified as intermediate, cold and warm), accounting for seasonal variability. An annual average concentration was estimated from the results of the three measurements for each site and temporally adjusted using a background reference site, which was continuously measured over an entire year (Lee et al., 2013). This was similarly done in in the current study using available AQMS data, and adjusting measurements to obtain an annual average. Though seasonal differences were not reported for the Taipei study, the annual average NO_x concentrations ranged from 21 to 113 $\mu g/m^3$.

A study of a similar setting conducted in Tianjin, in Beijing, China, with an approximate population of 10 million in an area with a radius of 11,919.7 km^2 used data from 30 AQMS locations in the region, of which only 20 were used for model development and the remaining 10 used in model validation. The similarities observed between this latter study and our study was notably the composition of industries in the region and traffic influences. This study developed models for heating and non-heating seasons, in which major roads and agricultural land use were among the significant predictor variables included in the respective models. R^2 values of 0.74 (heating season) and 0.61 (non-heating season) were obtained (Chen et al., 2010a). The study reported measured NO_2 concentrations with an average of 54 μg/m^3 in the heating season and 40 μg/m^3 in the non-heating season.

The model developed for the current study is fairly comparable to previous studies of similar settings such as the Taipei and the Tianjin study. Furthermore notable challenges of LUR model development, like with other studies are based on the availability, accuracy and reliability of municipal data such as air quality monitoring stations and municipal traffic count data to account for traffic volume and density. Despite such challenges, and the identified shortcomings, this was the first study to successfully develop an LUR model in South Africa ($R^2 =$ 0.73), comparable to other larger scale applications. The results emphasize the importance of traffic (represented by the proxy variables of road coverage in this study) and open space in influencing NO_x concentrations in South Durban.

Acknowledgements

We gratefully acknowledge the resource support provided by The Institute for Risk Assessment Sciences (IRAS) in Utrecht, The Netherlands, for the provision of the Ogawa samplers, and accompanying materials used for NO_x sampling, as well as the courier and return of samplers to IRAS for laboratory analysis. Further thanks is extended to the MACE Project Manager Mrs. Kareshma Asharam, Mr. Nkosana Jafta (Senior Lecturer at the Discipline of Occupational and Environmental Health, UKZN) and all fieldworkers involved in the deployment and collection of samplers. The Air Pollution Unit of the eThekwini Municipality as well as Corporate GIS and the South African Weather Services are also acknowledged for their assistance with providing AQMS data, GIS information and meteorological data respectively. We also wish to acknowledge funding support received within MACE from the National Research Foundation (NRF – 90550 & 80861) and the Medical Research Council (MRC).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.scitotenv.2017.07.278.

References

Aguilera, I., Sunyer, J., Fernandez-Patier, R., et al., 2007. Estimation of outdoor NOx,NO2, and BTEX exposure in a cohort of pregnant women using land use regression modeling. Environ Sci Technol 42, 7.
Aguilera, I., Garcia-Esteban, R., Iniguez, C., et al., 2010. Prenatal exposure to traffic-related air pollution and ultrasound measures of fetal growth in the INMA Sabadell cohort. Environ. Health Perspect. 118, 705–711.
Arain, M.A., Blair, R., Finkelstein, N., et al., 2007. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. Atmos. Environ. 41, 3453–3464.
Beelen, R., Hoek, G., 2010. ESCAPE Exposure Assessment Manual.
Beelen, R., Hoek, G., Briggs, D., de Hoogh, K., Vienneau, D., Fischer, P., 2000. User Protocol for Land Use Regression to Model Spatial Variations of Outdoor air Pollution Concentrations. Netherlands.
Beelen, R., Hoek, G., Vienneau, D., et al., 2013. Development of NO_2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – the ESCAPE project. Atmos. Environ. 72, 10–23.
Bell, L.M., 2006. The use of ambient air quality modeling to estimate individual and population exposure for human health research: a case study of ozone in the northern Georgia region of the United States. Environ. Int. 32, 8.
Bertazzon, S., Johnson, M., Eccles, K., Kaplan, G.G., 2015. Accounting for spatial effects in land use regression for urban air pollution modeling. Spatial and Spatio-temporal Epidemiology 14–15, 9–21.
Brauer, M., Hoek, G., van Vliet, P., et al., 2003. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. Epidemiology 14, 228–239.
Brauer, M., Henderson, S.B., Beckerman, B., Jerrett, 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.
Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P., Karr, C., 2008. A cohort study of traffic-related air pollution impacts on birth outcomes. Environ. Health Perspect. 116, 680–686.
Briggs, D., 2005. The role of GIS: coping with space (and time) in air pollution exposure assessment. J Toxicol Environ Health A 68, 1243–1261.
Briggs, D.J., Collins, S., Elliott, P., et al., 1997. Mapping urban air pollution using GIS: a regression-based approach. Int. J. Geogr. Inf. Sci. 11, 699–718.
Briggs, D.J., de Hoogh, C., Gulliver, J., et al., 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. Sci. Total Environ. 253, 151–167.
Brunekreef, B., Holgate, S.T., 2002. Air pollution and health. Lancet 360, 10.
Chen, L., Bai, Z., Kong, S., et al., 2010a. A land use regression for predicting NO2 and PM10 concentrations in different seasons in Tianjin region, China. J. Environ. Sci. 22, 1364–1373.
Chen, L., Du, S., Bai, Z., et al., 2010b. Application of land use regression for estimating concentrations of major outdoor air pollutants in Jinan. China. Journal of Zhejiang University 11, 11.
Classifications of Roads and Intersections, 2011. Accessed 11 January 2012, at. http://www.ethekwini.gov.za/City_Services/development_planning_management/Land_Use_Management/Land_Use_Definitions/Part_B_General_Definitions/Pages/Classifications_Roads_Intersections.aspx.
Eeftens, M., Beelen, R., de Hoogh, K., et al., 2012. 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.
EThekwini Air Quality Monitoring Network Annual Report, 2009. Pollution Control Support Section, EThekwini Health Department. 2009.
Gilbert, N.L., Goldberg, M.S., Beckerman, B., Brook, J.R., Jerrett, M., 2005. Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. J. Air Waste Manage. Assoc. 55, 1059–1063.
Gilliland, F., Avol, E., Kinney, P., et al., 2005. Air pollution exposure assessment for epidemiologic studies of pregnant women and children: lessons learned from the centers for children's environmental health and disease prevention research. Environ. Health Perspect. 113, 1447–1454.
Guastella, L., Knudsen, S., 2007. South Durban Basin Multi-Point Plan Case Study Report. Durban, South Africa, Department of Environmental Affairs and Tourism.
Hoek, G., Beelen, R., de Hoogh, K., et al., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42, 7561–7578.
Isakov, V., Johnson, M., Touma, J.S., Mukerjee, S., Ozkaynak, H., 2010. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. Atmos. Environ. 44, 3660–3668.
Jerrett, M., Arain, A., Kanaroglou, P., et al., 2005. A review and evaluation of intraurban air pollution exposure models. J Expo Anal Env Epid 15, 185–204.
Kashima, S., Yorifuji, T., Tsuda, T., Doi, H., 2009. Application of land use regression to regulatory air quality data in Japan. Sci. Total Environ. 407, 3055–3062.
Kingham, S., Pattinson, W., Pearson, A., Longley, I., Campbell, M., Apparicio, P., 2013. The use of a land use regression model to predict NO_2 air pollution in two small areas of Auckland. Air Quality and Climate Change 47, 4.
Kistnasamy, E.J., Robins, T.G., Naidoo, R., et al., 2008. The relationship between asthma and ambient air pollutants among primary school students in Durban, South Africa. Int. J. Environ. Health Res. 2, 365–385.
Lacasana, M., Esplugues, A., Ballester, F., 2005. Exposure to ambient air pollution and prenatal and early childhood health effects. Eur. J. Epidemiol. 20, 183–199.
Lebret, E., Briggs, D., van Reeuwjik, H., et al., 2000. Small area variations in ambient NO_2 concentrations in four European areas. Atmos. Environ. 34, 9.
Lee, J.-H., Wu, C.-F., Hoek, G., et al., 2013. Land use regression models for estimating individual NOx and NO2 exposures in a metropolis with a high density of traffic roads and population. Sci. Total Environ. 472, 1163–1171.
Lee, J.-H., Wu, C.-F., Hoek, G., et al., 2014. Land use regression models for estimating individual NOx and NO2 exposures in a metropolis with a high density of traffic roads and population. Sci. Total Environ. 472, 1163–1171.
Leonard, L., Bukurura, S.H., Poonen, H., 2008. Durban Reality Tour: A Collection of Material about the 'Invisible' Side of the City.
Liao, D., Peuquet, D.J., Duan, Y., et al., 2006. GIS approaches for the estimation of residential-level ambient PM concentrations. Environ. Health Perspect. 114, 1374–1380.
Madsen, C., Carlsen, K.C.L., Hoek, G., et al., 2007. Modeling the intra-urban variability of outdoor traffic pollution in Oslo, Norway—a GA 2LEN project. Atmos. Environ. 41, 12.
Madsen, C., Gehring, U., Haberg, S.I., et al., 2011a. Comparison of land-use regression models for predicting spatial NOx contrasts over a three year period in Oslo. Norway. Atmos Environ 45, 8.
Madsen, C., Gehring, U., Håberg, S.E., et al., 2011b. Comparison of land-use regression models for predicting spatial NOx contrasts over a three year period in Oslo, Norway. Atmos. Environ. 45, 3576–3583.
Map data©, 2016. Google Earth.
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 Monitor 9, 246–252.
Morgenstern, V., Zutavern, A., Cyrys, J., et al., 2007. Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. Occup. Environ. Med. 64, 8–16.
Naidoo, R., Gqaleni, N., Batterman, S., Robins, T., 2007. South Durban Health Study. Final Project Report. University of Kwa-Zulu Natal, South Africa.
Naidoo, R.N., Robins, T.G., Batterman, S., et al., 2013. Ambient pollution and respiratory outcomes among schoolchildren in Durban, South Africa. South African Journal of Child Health 7, 127–134.

Preston-Whyte, R.A., Diab, R.D., 1980. Local weather and air pollution potential: the case of Durban. Environ. Conserv. 7, 241–244.
Ramsay, L.F., Naidoo, R., 2012. Carbon footprints, industrial, transparency and community engagement in a South Durban neighbourhood. S. Afr. Geogr. J. 94 (2), 174–190.
Raux, C., Bonnel, P., Habermann, M., Billger, M., Haeger-Eugensson, M., 2015. Toward integrated modelling of urban systems land use regression as method to model air pollution. Previous Results for Gothenburg/Sweden. Procedia Engineering. 115, pp. 21–28.
Rosenlund, M., Forastiere, F., Stafoggia, M., et al., 2007. Comparison of regression models with land-use and emissions data to predict the spatial distribution of traffic-related air pollution in Rome. J Expos Sci Environ Epidemiol 18, 192–199.
Rosenlund, M., Forastiere, F., Stafoggia, M., et al., 2008. Comparison of regression models with land-use and emissions data to predict the spatial distribution of traffic-related air pollution in Rome. J Expo Sci Environ Epidemiol 18, 192–199.
Ross, Z., English, P.B., Scalf, R., et al., 2005. Nitrogen dioxide prediction in Southern California using land use regression modeling: potential for environmental health analyses. Exposure Sci Environ Epidemiol 16, 9.
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, 15.
Ryan, P.H., LeMasters, G.K., 2007. A review of land-use regression models for characterizing intraurban air pollution exposure. Inhal. Toxicol. 19 (Suppl. 1), 127–133.
Sahsuvaroglu, T., Arain, A., Kanaroglou, P., et al., 2006. A land use regression model for predicting ambient concentrations of nitrogen dioxide in Hamilton, Ontario. Canada. Journal of Air Waste Management Association 56, 12.
Slama, R., Morgenstern, V., Cyrys, J., et al., 2007. Traffic-related atmospheric pollutants levels during pregnancy and offspring's term birth weight: a study relying on a land-use regression exposure model. Environ. Health Perspect. 115, 1283–1292.
South Africa - Ethekwini Population Census, 2011. Statistics South Africa, 2016 AfriGIS (Pty) Ltd. Google.
Tularam H, Ramsay L. Synoptic influences on air pollution events in the Durban South Basin, 2006 to 2010 [Dissertation]. http://researchspace.ukzn.ac.za/handle/10413/11065: University of Kwa-Zulu Natal; 2013.
Tyson, P.D., Preston-Whyte, R.A., Preston-Whyte, R.A., 2004. The Weather and Climate of Southern Africa. Oxford. Oxford University Press, New York.
Uea, Ackerman-Liebrich, 2004. Health aspects of air pollution. In: Rea, Anderson (Ed.), Results from the WHO Project "Systematic Review of Health Aspects of Air Pollution in Europe". Copenhagen, Denmark, World Health Organisation.
Van Roosbroeck, S., 2007. Validation of traffic-related air pollution exposure estimates for longterm studies. Utrecht. 147.
Wheeler, A.J., Smith-Doiron, M., Xu, X., Gilbert, N.L., Brook, J.R., 2008. Intra-urban variability of air pollution in Windsor, Ontario - measurement and modeling for human exposure assessment. Environ. Res. 106, 7–16.