Land Use Regression Approach to Model NO_2–Concentrations in a Swedish Medium-City

Mateus Habermann[1], Monica Billger[1], Marie Haeger-Eugensson[2,3]

[1]Department of Architecture, Chalmers University of Technology, Gothenburg, Sweden.

[2]Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden.

[3]COWI AB, Gothenburg, Sweden.
E-mail: mathab@gmail.com

Abstract. In order to visualize the geographical distribution of air pollution concentration realistically, we applied the Land Use Regression (LUR) model in the urban area of Gothenburg, Sweden. The concentration of NO_2 was obtained by 25 passive air samplers during 7-20 May, 2001. Explanatory variables were estimated by GIS in buffers ranging from 50 to 500 m-radii. Linear regression was calculated, and the most robust were attained to the multiple linear regression. Additionally, the LUR model was compared with a dispersion model. The final model explained 81.7% of the variance of NO_2 concentration with presence of sum of traffic within 150 m and altitude as predictor variables. Mann-Whitney Test did not exhibit significant difference between yearly concentrations of NO_2 measured by regulatory measurement sites and measurements from passive samplers, thus LUR model was extrapolated for later years and mapped. The extrapolation indicated more elevated levels of pollution for the years 2003, 2006 and 2010. The results highlight the contribution of traffic on air quality and suggest that LUR modelling may explain the variations of atmospheric pollution with good accuracy. In addition, the model puts focus on spatial and temporal variability needed to describe retrospective exposure to air pollution in studies that evaluate health effects.

Keywords: Air pollution; nitrogen dioxide; exposure modeling; geographic information system; LUR model.

1 Introduction

Exposure assessment can be used to evaluate, at various levels of detail, the degree and linkage between contaminant sources and concentration of hazards and receptors (e.g. humans) in the environment by studying different exposure pathways (e.g. air, water, and soil) and routes (e.g. inhalation, ingestion, and dermal contact) between them. As one kind of exposure assessment air pollution exposure assessment indicates human exposure to air pollutants [1].

The most exposed receptors to air pollution include either individuals whose residences, study or work places offices are located near to heavy traffic roads or individuals who remain long time on roads (bus drivers, traffic guards, street vendors etc.). Therefore the environment may influence the exposure to pollutants and thereby trigger various outcomes [2].

Although the literature has documented significant variation of outdoor air pollution at small scales within urban areas for important pollutants such as NO_2 and black smoke [3-5], many studies assessed exposure based only on the proximity to polluted source, e.g. proximity to busy traffic [6]. This approach is limited because it disregards other parameters that may influence the dispersion of pollutants such as altitude, land use, population, road type, traffic intensity, temperature and atmospheric stagnation [7]. Therefore, recent models became more refined including some of those parameters [1].

Ordinary dispersion modelling requires good databases which are updated frequently (at least every five years); however, measurements are expensive if they are conducted at numerous places [1] and when dispersion modelling (DM) does not include measurements the computational analysis can be labour intensive. In many of urban areas levels of air pollution frequently exceed environmental standards, it is rather suitable to perform specific modelling, e.g. a computational fluid dynamics (CFD) model for street canyons. The result is usually satisfactory since the emission variability within those areas is more limited than the emissions variability of a whole city. Nevertheless, in comparison to environmental standard values, total concentrations are required, meaning that areas outside street canyon areas (urban, regional and long distance shares) have to be included. This can be accomplished through the smaller scale modelling, such as Land Use Regression (LUR).

LUR modelling is based on characteristics related to the overall trends of air pollutants concentrations mainly for longer time scales. This approach adopts more stable variables characteristics of land use, traffic, demographic and geography as predictor variables and measurements of pollution samplers as dependent variable [6]. A novel approach has utilized existing data from an air dispersion model rather than measurements data [8].

Thus LUR predicts pollution concentrations based on surrounding land use and traffic characteristics within circular areas (buffers) as predictors of the measured concentrations [7]. Moreover, the enhancement of geographic information system (GIS) techniques has contributed to the dissemination of LUR method.

Hence we aim to develop a LUR model to map the geographical distribution and the level of air pollution concentrations in the urban area of Gothenburg and Mölndal, Sweden.

2 Methods

2.1 Study Area

The study was carried out in urban areas of Gothenburg and Mölndal, at the west coast of Sweden. The choice of this mid-sized urban area for this study is motivated due to their characters, for instance, the varied altitude, with flat bottom valleys surrounded with mountains. This cause limited dispersion in the valleys affecting the levels of air pollutants especially during wintertime due to temperature inversions. Furthermore, both harbour at Göta River estuary and industrial operations around the city contribute to air pollution emissions. However, the air quality in the city centre is mainly affected by the road traffic emission. Despite having fewer inhabitants, geographic conditions lead Gothenburg to exhibit levels of air pollution considerably higher than the urban area of Stockholm.

2.2 NO_2 Sampling

NO_2 is considered a good indicator of traffic-related air pollution and is easy to measure [9]. In 2001 a monitoring campaign of NO_2 named GÖTE-2001 was carried out from 7^{th} - 20^{th} May (2 weeks of spring). The measurements were taken at a height of approximately 2-2.5 m above ground. This monitoring campaign was a partnership between local government, Chalmers University of Technology and University of Gothenburg.

The measurements were done in 25 sites using passive air samplers (PAS) (figure 1). The placement of each site was determined by specific criteria: 20 PAS were distributed in cells by dividing the region into 1 x 1 km cells covering a 20 km^2 grid area. In addition to the grid, five instruments were positioned in the western and north parts of the city and also in the vicinity of the main valleys in the region Göta Älv valley, Säve valley, Mölndal valley [10].

2.3 Data for Independent Variables

The digital cartographic database on altitude, land use predominance and roads were obtained from the Lantmäteriets geodatabase (http://www.lantmateriet.se). The land use data included 9 categories by type of use (industrial, arable, forests and water) building patterns (enclosed, low, high, recreational). The road data, containing many different categories, were summarized in 04 main groups based on their width and speed: Types I and II representing local roads (until 50 km/h) and types III and IV representing expressways (> 50 km/h). Traffic information for the year 2001 was provided by Institute of Medicine of University of Gothenburg.

Meteorological data contained the annual average of wind speed and annual average of mixing height, which is a vertical stability parameter of the air column which strongly determines the dispersion of air pollutants. Demographic data contained number of inhabitants in a grid of points with a resolution of 500 m. Both data were provided by Earth Sciences Department of University of Gothenburg.

The independent variables were created on the GIS software MapInfo (Professional version 10.5; MapInfo Corporation, New York, NY, USA) in buffers of 50, 100, 150, 250 and 500 m-radii around 25 sampling locations. The variables consisted in six broad categories:

- Physical geography – altitude (m);
- Land use – shortest distance to industrial use (m), area of different land uses (km^2) estimated within buffers around each sampling location;
- Road – shortest distance to roads type IV (m), lengths of different roads (m) within buffers of different radii;
- Traffic – sum of traffic flow counts (annual average daily traffic) within buffers of different radii;
- Demographic – available in a grid of points with a resolution of 500 m. Thus number of inhabitants and population density within 500 m were estimated.
- Meteorological – annual average of wind speed and mixing height.

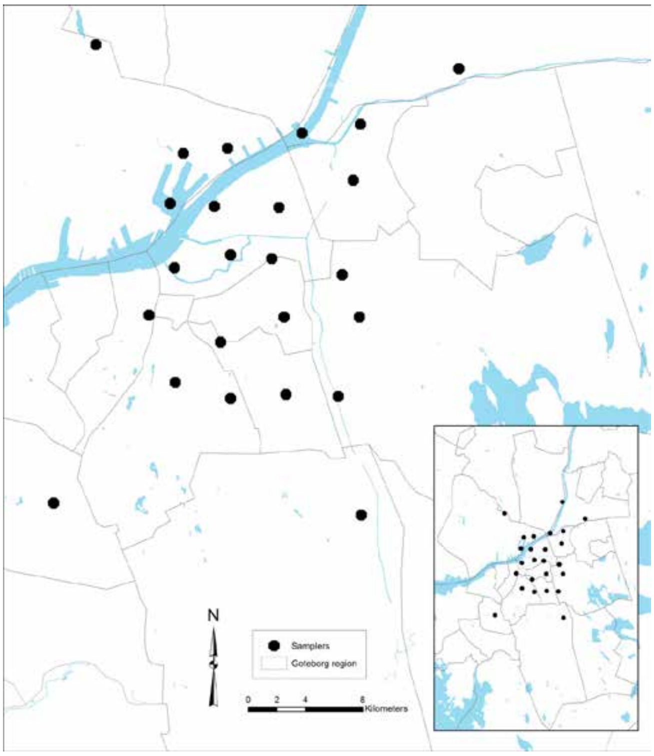

Figure 1. Location of the 25 measurement sites and regulatory monitoring stations in Gothenburg and Mölndal.

2.4 Statistical Analysis

The measured ambient NO_2 concentration was used as dependent variable and its normal distribution was evaluated with Shapiro-Wilk Test. Regression model was constructed using stepwise procedure, in which models were constrained to maximize the adjusted percentage explained variance (R^2). Variables with significant multicollinearities determined by variance inflation factors (VIF) were excluded. Residuals were examined to ensure the regression equations did not demonstrate any systematic bias or residual spatial autocorrelation. The theoretical final equation is presented below:

$$NO_2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (1)$$

NO_2 represents the value of the dependent variable. The constant is β_0. X_1 represents the explanatory variables of traffic volume, X_2 road type, X_3 land cover, X_4 altitude and X_5 demography and their respective coefficients represented by β. ε is the random error component.

The LUR-modelled results were compared with the data collected by four regulatory monitoring stations (RMS): Femman, Gårda, Järntorget and Mölndal, using Mann-Whitney Test ($\alpha = 5\%$) for the period May/2001 and yearly 2001 average. We tested Pearson correlation and quantified agreement between measured and LUR-predicted at RMS using Bland-Altman method.

We also used Bland-Altman to verify agreement between the LUR - model and the dispersion model (DM) concentrations for 2001. DM was carried out and provided by the IVL Swedish Environmental Research Institute [11].

The LUR-model was extrapolated for later years using the average of NO_2 concentration measured by the PAS (May, 2001) and RMS (yr. 2001) for the respective year (2002 – 2013) to adjust the constant of LUR formula.

$$Cy = \left(\frac{C_{2001} \times \mu M_y}{\mu S_{2001}} \right) \tag{2}$$

where C is the constant of the LUR final formula, y represents the correspondent year to be extrapolated (2002 to 2013), μS the average of NO_2 concentration at the PAS and μM is the average of NO_2 concentration at RMS.

2.5 Mapping the Model

A grid of points 7,257 lattice points with a cell resolution of 300 m was created for the studied area. Buffers were created around each point to capture the variable(s) derived from the final regression model.

A lattice interpolation using kriging at the points was applied to visualize the continuous map of LUR-modelled NO_2 concentration for the years 2001 to 2013 and DM-NO_2 for 2001, on the GIS software ArcMap 10.1 (ESRI, Redlands, CA, USA).

Statistical analysis and plots were calculated using SPSS for Windows software. (version 13.0, SPSS Inc.). Kriging and Moran I. were performed on ArcGIS (version 10.1, ESRI, Redlands, CA, USA).

3 Results

3.1 LUR Modelling

NO_2 samples at the PAS had an arithmetic mean of 22.8 µg/m^3 ranging from 11.8 to 34.3 µg/m^3 (SD = 6.1). There were 05 samples that exhibited a greater value than one standard deviation from the mean ($\geqslant$ 28.9 µg/m^3). All of these were located in proximity to expressways and industrial area. NO_2 measured at the RMS (Femman, Gårda, Järntorget and Mölndal) had an average of 30.2 µg/m^3 (SD = 5.7) for the year 2001.

Mann-Whitney Test did not exhibit a significant difference annual (RMS) and 7-20th May period (PAS) means. Thus, it was assumed that the measurements obtained by PAS in May are not different from the annual average for 2001 measured at RMS.

Table 1 exhibits the variables having significant correlation with NO_2 concentrations. The most related variables to the levels of NO_2 were altitude, deciduous forests, sum of traffic, high buildings and industrial areas. Many of the predictor variables were highly correlated within the same group e.g. sums of traffic at 50, 100, 150 m radii; area of high buildings at 50, 100, 150 and 500 m radii. Elevation and deciduous forest were highly correlated with NO_2 at 150, 250 and 500 m radii.

Industrial land use, enclosed buildings and roads type IV were associated with increasing of NO_2 concentration whereas altitude, recreational buildings, high buildings and roads type I were associated with decreasing of NO_2 levels (table 1).

When the residuals of regression analysis were investigated, the PAS named Östra Hamngatan represented an outlier (see figure 2). After excluding this PAS from the model the predictive power improved to 81.7%. A previous study that included this outlier exhibited a lower predictive power of 59.4% [12].

Table 1. Independent variables with significant correlation with concentrations of NO_2.

Variables	Buffer radius[‡]	r^*	r^2	p-value
Sum of traffic	150	0.841	0.917	0.000
Altitude	-	-0.704	0.839	0.000
Sum of traffic	100	0.664	0.815	0.000
Decidual forest	500	-0.656	0.810	0.000
Industrial	500	0.599	0.774	0.002
Sum of traffic	50	0.595	0.771	0.002
Decidual Forest	250	-0.574	0.758	0.003
High buildings	500	-0.570	0.755	0.004
Sum of traffic	250	0.543	0.737	0.006
Decidual forest	150	-0.532	0.730	0.007
Industrial	250	0.511	0.715	0.011
High buildings	250	-0.488	0.699	0.015
Decidual forest	100	-0.470	0.686	0.020
High buildings	150	-0.469	0.685	0.021
High buildings	100	-0.454	0.673	0.026
Industrial	150	0.425	0.652	0.039
High buildings	50	-0.424	0.651	0.039
Decidual Forest	50	-0.397	0.630	0.027
Low buildings	500	-0.393	0.627	0.029
Industrial	100	0.389	0.624	0.030
Industrial	50	0.376	0.613	0.035
Low buildings	250	-0.357	0.598	0.043
Sumo f traffic	500	0.357	0.597	0.043
Road type I	250	-0.351	0.592	0.046

[‡] Buffer radius refers to the distance of the circular zone around each site for which the variables were calculated.

[*] Pearson 2-tailed correlation.

Tables 2 and 3 show results from multivariate regression analysis and the variables with statistical significance and non-collinearity involved in the model. The final model explained 81.7% of the variance of NO_2 and two predictor variables were included. Sum of traffic within 150 m exhibited the strongest association with levels of NO_2 ($p < 0.001$), thus heaviest traffic areas had more elevated air pollution concentration (positive coefficient). On the other hand, elevation was related to the decrease of the concentration of NO_2 (negative coefficient) and the second strongest variable related to the dependent variable ($p < 0.001$).

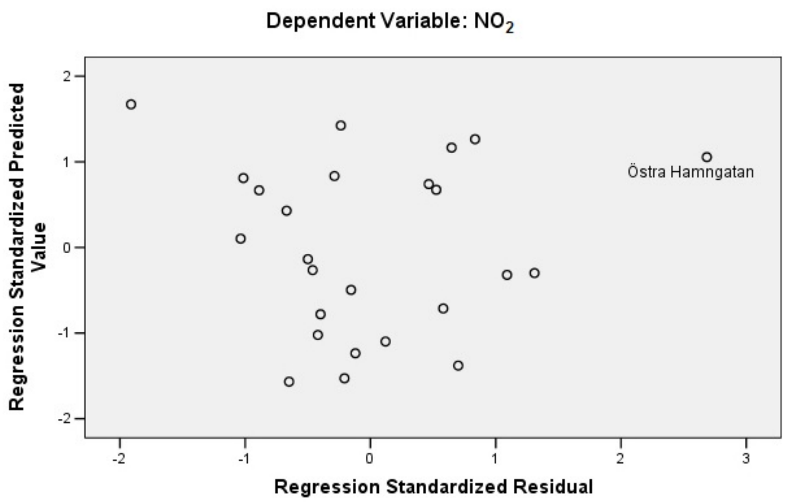

Figure 2. Box plot of regression residuals values and regression predicted values of NO_2 at the 25 PAS.

Table 2. Results of the multivariate regression model.

Model		1^a	2^b
R		0.84117	0.91266
r^2 (%)		0.70756	0.83294
Adjusted r^2 (%)		0.69427	0.81703
RMSE (µg/m³)		335.176	259.293
Change Statistics	r^2 Change (%)	0.7076	0.1254
	F Change	53.230	15.761
	df1	1	1
	df2	22	21
	Sig. F Change	<0.001	<0.001

[a] Predictors: constant, sum of traffic 150 m

[b] Predictors: constant, sum of traffic 150 m, altitude

Table 3. Description of developed LUR model.

Model		Constant	Sum of traffic[‡]	altitude
Unstandardized Coefficients	B	20.911	2.00E-05	-0.100
	Std. Error	1.220	3.08E-06	0.025
Standardized Coefficients	Beta	-	0.656	-0.400
T		17.13	6.52	-3.97
Sig.		<0.001	<0.001	0.001
Collinearity Statistics	Tolerance	-	0.785	0.785
	VIF	-	1.274	1.274

[‡] Buffer radius of 150 m

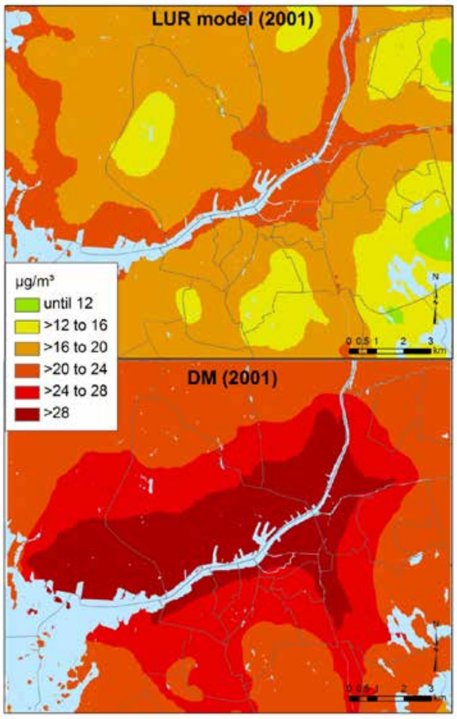

Figure 3. Interpolated surface of NO_2 concentration based on LUR and DM for 2001.

LUR residual spatial autocorrelation exhibited z-score = -1.85 (Moran's I: -0.169474; p-value: 0.064654), it means there is less than 10% likelihood that this dispersed pattern could be the result of random chance. Even though spatial autocorrelation in the NO_2 observation values was close to significance (critical z-score = 1.96) the assumption of spatial autocorrelation in the residuals was negligible.

3.2 Mapping the LUR Model

The final formula was applied at 7,257 lattice points to calculate the NO_2 concentration for each point. The average of the LUR-predicted NO_2 was 17.5 µg/m³ (SD = 3.2 µg/m³). DM was applied at 2,730 lattice points and exhibited an average of 24.5 µg/m³ (SD = 4.5µg/m³). Figure 2 compares the interpolated LUR and DM NO_2 concentration in the urban area of Gothenburg and Mölndal. Although visually both LUR and DM models display most polluted areas close to Göta River (low altitude) and the largest highways (E6 for example), LUR model exhibits lower levels of NO_2, as seen in figure 3.

3.3 Validation and Agreement of Regression Results

Validation analysis was undertaken to confirm the predictive capacity of the results. First, an attempt to compare modelled results with the data collected by the RMS for four collocated sites demonstrated correlation higher between LUR-modelled concentration and annual average (r^2 = 0.89) compared to LUR-modelled concentration and May/2001 average (r^2 = 0.57). Although well correlated, Bland-Altman analysis demonstrated significant systematic difference between annual average and LUR modelled at RMS (p=0.013).

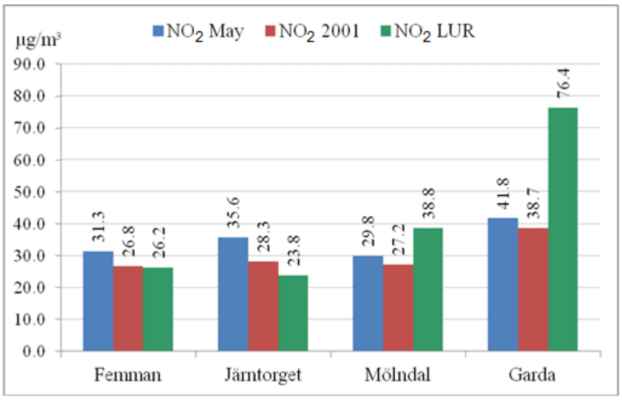

Figure 4. Estimated and measured NO_2 concentrations obtained from LUR-model for 2001 (May and 2001 averages) at the RMS.

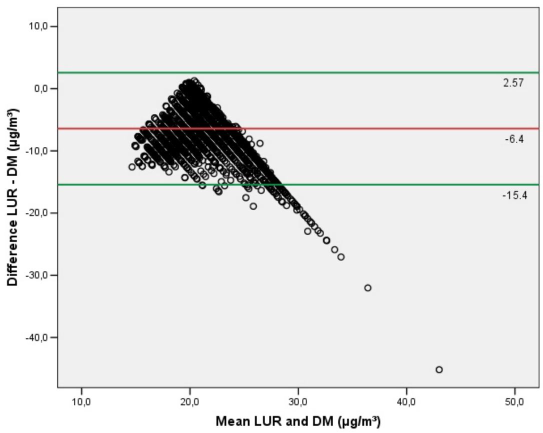

Figure 5. Bland-Altman plot of differences DM-LUR *vs.* mean DM-LUR with the representation of the limits of agreement.

The figure 4 exhibits the measured and modelled NO_2 concentration for each RMS. The LUR modelling is overestimated at Gårda station due to its proximity to a busy highway (route E6), and underestimated at Järntorget station due to lower traffic volume detected within 150 m.

For comparison between the models, Bland-Altman plot exhibited in figure 5 indicates negative bias of 6.4 µg/m³ between LUR and DM. There is a significant systematic difference between both models ($p<0.001$) and DM over-estimates NO_2 compared to LUR.

3.4 Extrapolating the LUR Model

For the years over which the LUR model was extrapolated (2002 to 2013), it was found that the mean of LUR NO_2 estimated at the RMS levels were more elevated than PAS (see table 4). NO_2 had higher measured concentrations for 2003, 2006 and 2010 ($\geqslant$ 30 µg/m³) at the RMS. Consequently, the extrapolated models also exhibited higher levels of estimated NO_2 by LUR for those years. Inversely, the years 2008 and 2009 exhibited lower levels of pollution (table 4).

Even though the LUR-modelled NO_2 concentration varies with time, the maps that have undergone kriging exhibited similar patterns of NO_2 concentration. Hotspots are related to the proximity to heavy traffic roads and low elevation (maps available in Supplemental Material - Figures S1 to S13).

Table 4. Average of NO_2 concentration at the RMS and PAS for the years 2002 to 2013.

Year	Avg. of NO_2 at the RMS§		Avg. of NO_2 at the PAS§
	Real	Estimated†	Estimated†
2002	28.6	46.6	23.9
2003	32.4	50.0	27.3
2004	32.1	49.8	27.1
2005	31.7	49.4	26.7
2006	33.1	50.7	28.0
2007	29.2	47.1	24.4
2008	25.9	44.1	21.4
2009	25.8	44.0	21.3
2010	30.4	48.3	25.6
2011	26.2	44.4	21.7
2012	29.2	47.1	24.4
2013	27.7	45.8	23.1

§ in µg/m³

† Estimated by the LUR model for its respective year.

4 Discussion

In this study concentrations of ambient NO_2 throughout the urban area of Gothenburg and Mölndal in Sweden have been modelled. The LUR model included two independent variables (altitude and traffic within 150 m radius) and it predicted almost 82% of NO_2 variability in the urban area of Gothenburg and Mölndal for 2001.

Results show that variations of NO_2 concentrations are positively correlated with increasing industrial areas, and negatively correlated with increase of forests. Geographic characteristics as altitude also contribute considerably to determine the urban air quality. Thus, by adding meteorological data the model was extrapolated to be valid also for other years based on measurements of monitoring stations.

Comparing to other LUR models [5, 6, 8, 13-18] the results also highlight the road network and its respective vehicle flow as the main indicator of air pollution, mainly in areas where cars are more concentred such as the city centre, or close to busy expressways.

Land use regression method has generally been applied successfully to model annual mean concentrations of NO_2, NO_x, $PM_{2.5}$, and VOCs. The method has been carried out in different settings, including non-industrial and industrial cities and its performance is often better or equivalent to geostatistical methods such as kriging and conventional dispersion models.

Nonetheless the method has some limitations, as highlighted by Hoek et al. [19], LUR models have a restricted capacity to separate the impact of some priority pollutants because they are collinear to each other, although the same problem affects other methods of exposure assessment.

Although LUR models provide individual estimates of ambient exposure (e.g. residential address) their predictor variables do not include infiltration of outdoor air into the houses, or only estimates concentration at rooftop levels. However, this problem may affect all methods of environmental exposure assessment due to a lack of available data, complexity and high costs of data collection.

Regarding possible differences between ambient predicted and personal exposure to air pollution Montagne et al. [20] assessed the agreement of LUR models with measured personal exposure to NO_2 in Helsinki (Finland), Utrecht (The Netherlands), and Barcelona (Spain). NO_2 LUR models significantly predicted outdoor concentrations and personal exposure in Utrecht and Helsinki. LUR-predicted and measured outdoor, indoor, and personal concentrations were highly correlated when data from the three cities were combined.

Moreover, little attention has been given to potential problems associated with datasets as accessibility, completeness and precision. Sometimes data may not be available for the period of interest. In this LUR model the meteorological, traffic and demographic data for the year 2001 were available, although there were limitations on accuracy and representativeness of the covariate data applied in this study. The demographic data were available only with a resolution of 500 m and neither land use nor traffic data were available in categories which have enabled detailed urban land uses (e.g. residential, commercial or governmental) and type of traffic (e.g. truck traffic, bus traffic or light traffic).

Usually sampling sites are chosen to capture the range in predictor variables (e.g. local roads to main roads; commercial areas, industrial areas, etc.) to be representative of the full range of NO_2 concentrations. Additionally measurements were collected from at least 3 measurement periods to capture seasonal variation of NO_2. In this case the model only relied on one rather small set (n=24) of measurements allocated on an equidistant grid method and during one time period (and extrapolated to annual average exposures), representing both spatial and temporal limitations for LUR model.

Furthermore, there were not a substantial number of PAS to internally validate the LUR model by applying them to a set aside validation dataset. However, LUR-modeled NO_2 correlated well with regulatory monitoring data.

In the LUR model for Gothenburg the predictor variables were computed for circular zones around each monitoring site ranging 50 to 500 m radii. Radii upper than 500 m were not included in our modelling since they could be overlapped due to the proximity of the samplers' sites to each other although previous studies have found influence of NO_2 from distances larger than 500 m [13, 16, 18; 21]. Although the NO_2 concentrations mostly have a local pattern, this may have constrained the selection of predictors, and may consequently have deprived the capability of our model.

Promising new developments in LUR modelling include data from air dispersion modelling [18] and also the use of additional predictor variables such as meteorological or emission data and raster GIS environment [19]. A few studies included meteorological variables such as temperature [21], humidity [14], atmospheric stability [22], wind speed [6, 15], wind direction [6, 13, 15]. However, in this study for Gothenburg and Mölndal, variables of wind speed and mixing height were included but did not modify the results.

Air pollution modelling always implies some degree of error. When two methods are compared, neither provides an unequivocally correct result. However, in comparison to air dispersion models, LUR is a less costly option to assess the intra urban variability of air pollution as it combines air pollution monitoring at a smaller number of locations and development of models using predictor variables obtained through GIS [19]. Even though the result did not identify good agreement between LUR and the dispersion model a spatial similarities among the most polluted areas could still be observed.

When LUR models represent annual average concentrations it is possible to use continuous routine monitoring data to adjust for the temporal component [19]. In our LUR model we identified considerable differences in temporal concentration of pollutants, and higher levels of air pollution for the years 2003, 2006 and 2010.

LUR modelling is definitely a good option to assess the intra-urban variability of air pollution [13] but LUR models need a sufficient network of sampling sites, properly selected to represent the spatial gradient of exposure in the study population. This condition is partially fulfilled in this study because the sampling was performed only at 24 sites. In this context, when the number of predictor variables is very high (n = 180) and the number of sites is low it is easy to end up with a final model with a high R^2 but with a poor out-of-sample predictability [23]. Consequently, the results showed higher concentrations of pollutants close to busy traffic roads and/or with low altitude the accuracy of the model decreases in others areas e.g., rural areas or islands, since samplers were sited only at the urban area.

Additionally, as personal measurements in epidemiological studies are expensive and logistically difficult, we emphasize the importance on developing indirect approaches to assessment exposure. Furthermore, our LUR model provides spatial and temporal variability needed to estimate outdoor concentrations at the home address of participants in epidemiological studies regarding air pollution in Gothenburg.

Epidemiological retrospective studies are based on several metrics to evaluate association between air pollution exposure and adverse health outcomes (e.g. pregnancy outcomes, cardiorespiratory morbidity and mortality, etc.) [1]. In future health studies, air pollution exposures can be determined and the LUR can be calculated for homes or work locations within the studies, as described in the literature [14, 24].

Despite some limitations, LUR is a fast method to carry out and to access exposure to air pollution when there is geographical covariates data availability. LUR quantifies parameters related to deterioration of air quality, e.g. high density land uses, industrial or busy traffic areas, as well as it may support policymaking for sustainable construction and development of an effective policy to decrease air pollution concentration.

Potential benefits of this model for health effects research include improved spatial estimations of atmospheric pollutant exposure and reduced need for extensive pollutant measurements.

5 Conclusion

This study adds to the literature on air pollution exposure assessment in several ways:

- The influence of different types of spatial covariates on model performance was identified;
- Pollution surfaces/maps for the study area showing the estimated distribution of ambient concentrations were created;
- The LUR model was compared to a dispersion model;
- The importance of extrapolating models to capture substantial differences in temporal distributions of pollutants between years (2002 to 2013) was emphasized. This is important for epidemiological studies based on different exposure periods.

Acknowledgements. This research was funded by the National Council for Scientific and Technological Development, Research Center and partially by the Swedish Brazilian Innovation Center and SAAB AB.

We thank the staff of Visual Arena Lindholmen Science Park and Visual Arena Research for supporting the study. Additionally, we thank Alexander Walther and Peter Molnár from Institute of Medicine of University of Gothenburg for providing us both traffic and demographic data.

References

1. Zou B, Wilson Jg, Zhan Fb, Zeng Y. Air pollution exposure assessment methods utilized in epidemiological studies. *J Environ Monit.* 2009; 11:475-490. doi: 10.1039/b813889c
2. Nuckols JR, Ward MH, Jarup l. Using geographic information systems for exposure assessment in Environm ental Epidemiology studies. *Environ Health Perspect.* 2004; 112:1007-1015. doi: 10.1289/ehp.6738
3. Zhu Y, Hinds WC, Kim S, Shen S, Sioutas C. Study of ultrafine particles near a major highway with heav y-duty diesel traffic. *Atmos Environ.* 2002; 36: 4323–4335. doi: 10.1016/S1352-2310(02)00354-0
4. World Health Organization (WHO). *WHO air quality guidelines global update 2005* – Report on a working group meeting. Bonn, Germany: 2006.
5. Jerrett M, Arain MA, Kanaroglou P, Beckerman B, Crouse D, Gilbert NL, et al. A Modeling the intraurba

n variability of ambient traffic pollution in Toronto, Canada. *J Toxicol Environ Health*, 70 (2007):200-212. doi: 10.1080/15287390600883018

6. Arain MA, Blair R, Finkelstein N, Brook JR, Sahsuvaroglu T, Beckerman B, Zhang L, Jerret M. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure stu dies. *Atmos Environ.* 2007; 41:3453-3464. doi: 10.1016/j.atmosenv.2006.11.063
7. Jerret M, Finkelstein M. Geographies of risk in studies linking chronic air pollution exposure to health outc omes. *J Toxicol Environ Health.* 2005; 68:1207-1242. doi: 10.1080/15287390590936085
8. Mölter A, Lindley S, de Vocht F, Simpson A, Agius R. Modelling air pollution for epidemiologic research--Part I: A novel approach combining land use regression and air dispersion. *Sci Total Environ.* 2010; 1; 408 (23):5862-5869. doi: 10.1016/j.scitotenv.2010.08.027
9. Rose N, Cowie C, Gillet R, Marks GB. Weighted road density: a simple way of assigning traffic-related air pollution exposure. Atmos Environ. 2009; 43:5009-5014. doi:10.1016/j.atmosenv.2009.06.049
10. Borne K, Chen D, Miao J, Achberger C, Lindgren J, Hallquist M, et al. *Data report on measurements of meteorological- and air pollution variables during the campaign GÖTE-2001.* Earth Sciences Centre, Gothen burg University. Gothenburg, 2005.
11. Haeger-Eugensson et.al. (2010). *On the increasing levels of NO_2 in some cities*: The role of primary emissi ons and shipping. Stockholm: IVL Swedish Environmental Research Institute. IVL report B- B1886. Availab le in http://www.ivl.se. Access in July/2017.
12. Habermann M, Billger M, Haeger-Eugensson M. Land use regression as method to model air pollution. Pre vious results for Gothenburg/Sweden. *Procedia Engineering.* 2015; 115:21–28. doi: 10.1016/j.proeng.2015.07.350
13. Sahsuvaroglu T, Arain A, Kanaroglou P, Finkelstein N, Newbold B. A land use regression model model fo r predicting ambient concentrations of nitrogen dioxide in Hamilton, Ontario. *J Air Waste Manag Assoc.* 20 06; 56:1059-1069. doi: 10.1080/10473289.2006.10464542
14. Iñiguez C, Ballester F, Estarlich M, Llop S, Fernandez-Patier R, Aguirre-Alfaro A, Esplugues A. Estimatio n of personal NO_2 exposure in a cohort of pregnant women. *Science of the Total Environment.* 2009; 407:60 93-6099. doi: 10.1016/j.scitotenv.2009.08.006
15. Chen L, Bai Z, Kong S, Han B, You Y, Ding X, Du S, Liu A. A land use regression for predicting NO_2 and PM10 concentrations in different seasons in Tianjin region, China. *J Environ Sci.* 2010; 22:1364–1373. doi: 10.1016/S1001-0742(09)60263-1
16. Eeftens M, Beelen R, Hoogh K, Bellander T, Cesaroni G, Cirach M, et al. Development of Land Use Reg ression Models for PM2.5, PM2.5 Absorbance, PM10 and PM coarse in 20 European Study Areas; Results of the ESCAPE Project. *Environ Sci Technol.* 2012, 46: 11195−11205. doi: 10.1021/es301948k
17. Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, et al. Development of NO_2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe e The E SCAPE project. *Atmos Environ*; 72 (2013):10-23. doi:10.1016/j.atmosenv.2013.02.037
18. Ross Z, English PB, Scalf R, Gunier R, Smorodinsky S, Wall S, et al. Nitrogen dioxide prediction in Sout hern California using land use regression modelling: potential for environmental health analyses. *J Exp Sci Environ Epidemiol.* 2006; 16:106–114. doi:10.1038/sj.jea.7500442
19. Hoek, Beelen R, Hoogh K, Vienneau D, Gulliver J, Fischer P, Briggs D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Environ.* 2008; 42:7561-7578. doi:10.1016/j. atmosenv.2008.05.057
20. Montagne D, Hoek G, Nieuwenhuijsen M, Lanki T, Pennanen A, Portella M, et al. Agreement of land use regression models with personal exposure measurements of particulate matter and nitrogen oxides air pollut ion. *Environ Sci Technol.* 2013; 47:8523-8531. doi: 10.1021/es400920a
21. Liu C, Henderson BH, Wang D, Yang X, Peng ZR. A land use regression application into assessing spatia l variation of intra-urban fine particulate matter (PM2.5) and nitrogen dioxide (NO_2) concentrations in City of Shanghai, China. *Sci Total Environ.* 2016; 15; 565:607-15. doi: 10.1016/j.scitotenv.2016.03.189
22. Li L, Wu J, Wilhelm M, Ritz B. Use of generalized additive models and cokriging of spatial residuals to improve land-use regression estimates of nitrogen oxides in Southern California. *Atmos Environ.* 2012; 55:22 0-228. doi:10.1016/j.atmosenv.2012.03.035
23. Johnson M, Isakov V, Touma JS, Mukerjee S, Özkaynak H. Evaluation of land-use regression models used

to predict air quality concentrations in an urban area. *Atmos Environ*. 2010; 44:3660-3668. doi:10.1016/j.atmosenv.2010.06.041

24. Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. A cohort study of traffic-related air pollution impacts on birth outcomes. *Environ Health Perspect*. 2008; 116:680-686. doi:10.1289/ehp.10952

Authors' Contributions

MH contributed to design, analysis, and interpretation of data and has been involved in drafting of the manuscript. MB and MHE have been involved in supervising the revising critically the manuscript and the final approval of the version to be published.

Abbreviations

μg/m^3: microgram per cubic meter
GIS: geographic information system
LUR: land use regression
RMS: regulatory monitoring stations
NO_2: nitrogen dioxide
NO_x: nitrogen oxides
PAS: passive air samplers
PM: particular matter
PM_{10}: particular matter $\leqslant$ 10 μg
$PM_{2.5}$: particular matter $\leqslant$ 2.5 μg
SD: standard deviation
VOC: Volatile organic compound

Conflict of Interest Statement

The authors declare that they do not have any competing interests.

Supporting Information

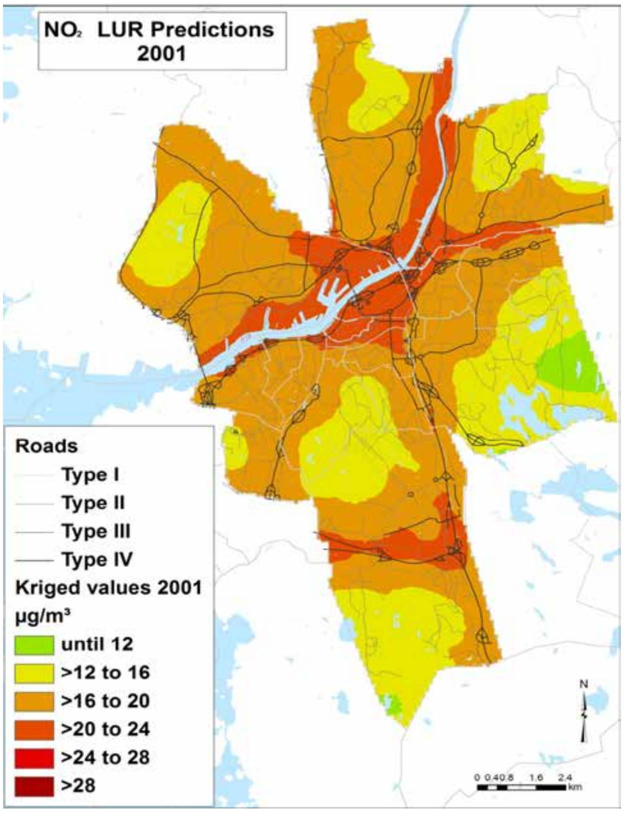

Figure S01. Kriged surface of NO_2 concentration based on LUR model for the year 2001.

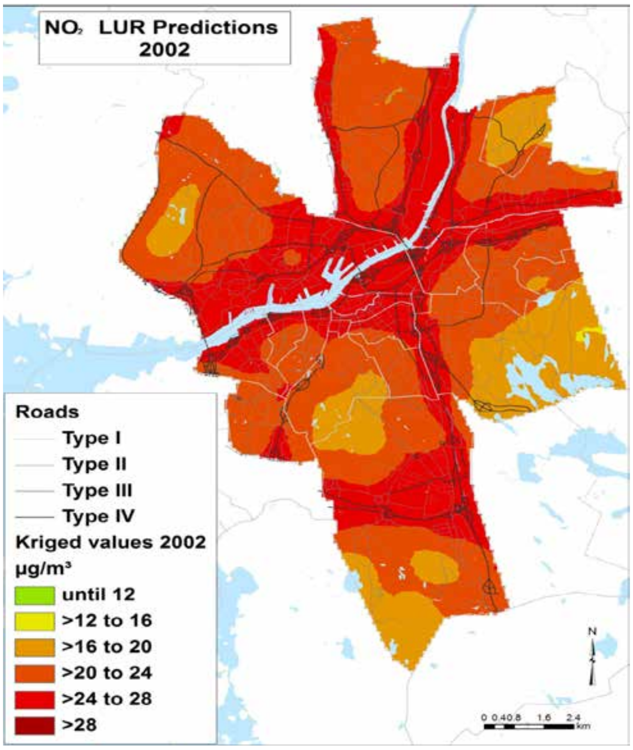

Figure S02. Kriged surface of NO_2 concentration based on LUR model for the year 2002.

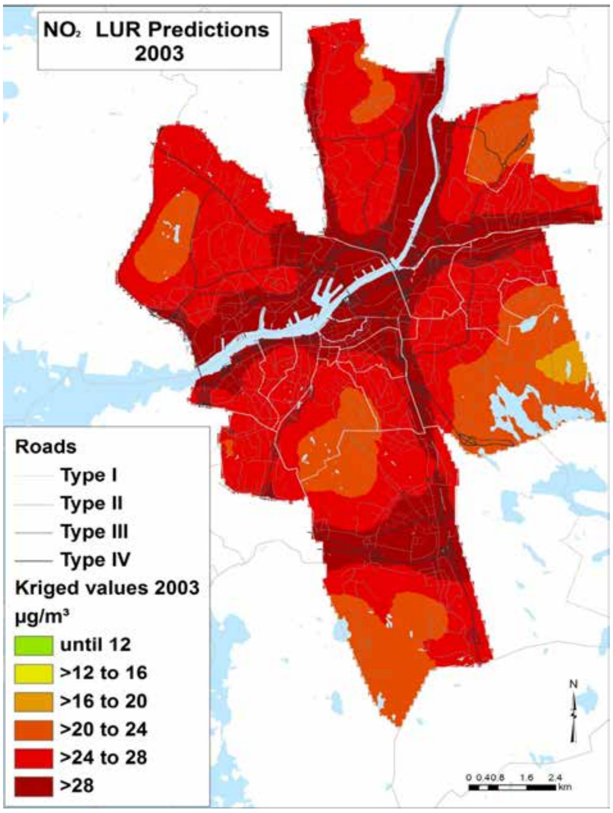

Figure S03. Kriged surface of NO_2 concentration based on LUR model for the year 2003.

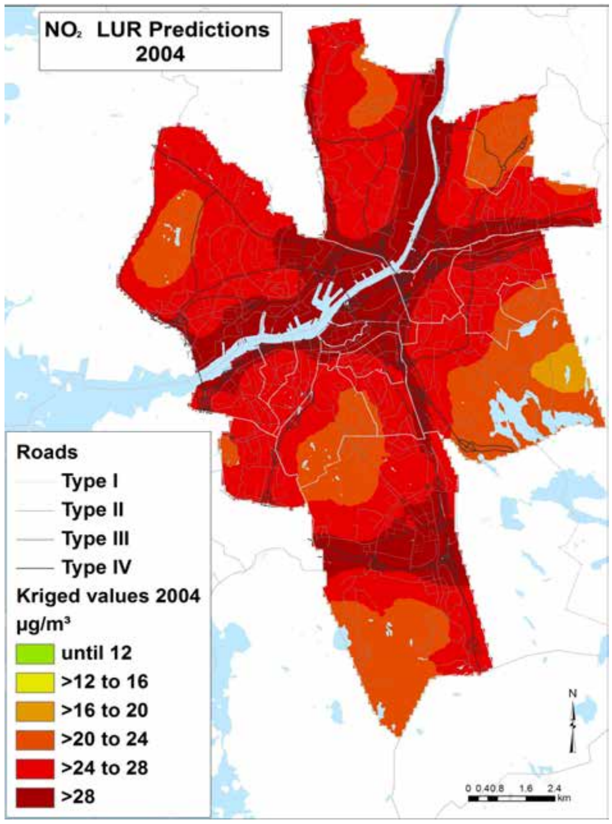

Figure S04. Kriged surface of NO_2 concentration based on LUR model for the year 2004.

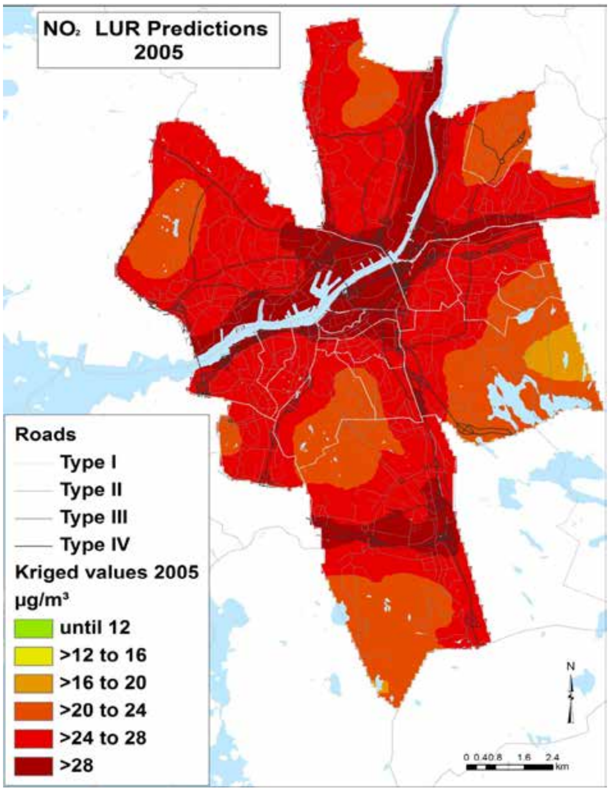

Figure S05. Kriged surface of NO_2 concentration based on LUR model for the year 2005.

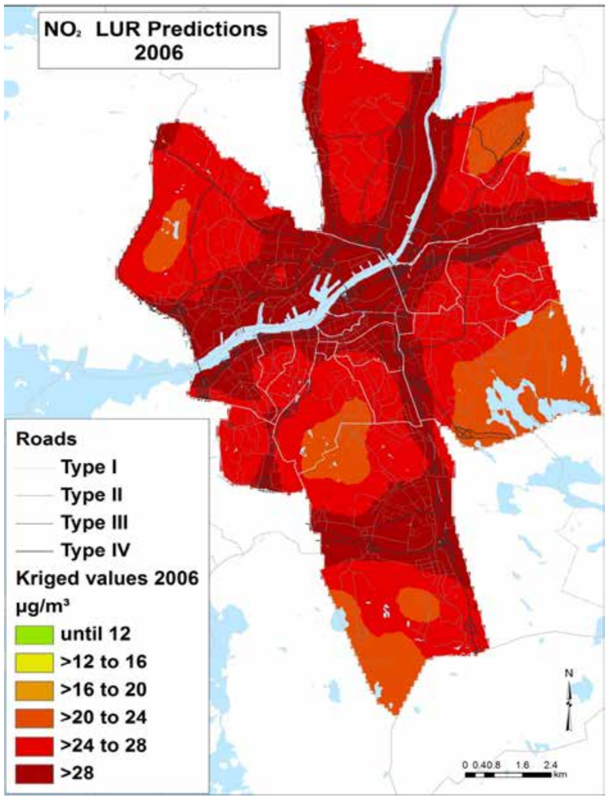

Figure S06. Kriged surface of NO_2 concentration based on LUR model for the year 2006.

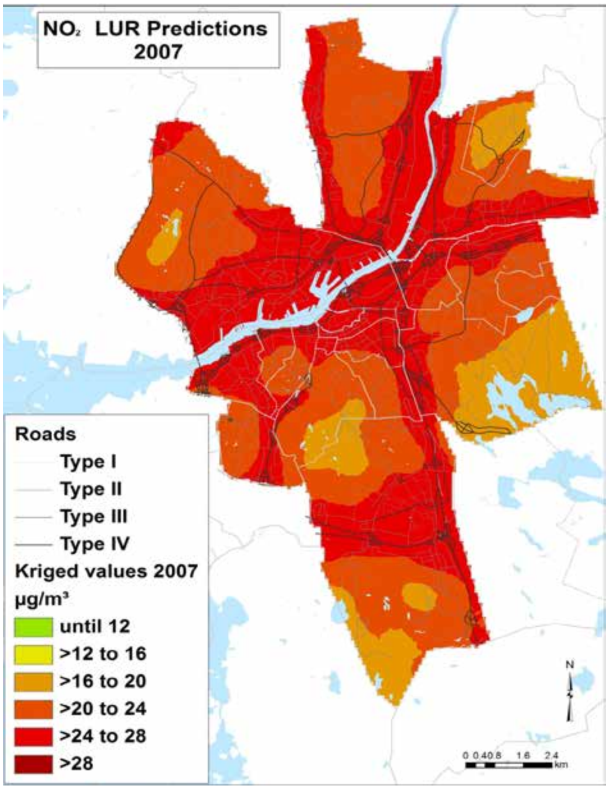

Figure S07. Kriged surface of NO_2 concentration based on LUR model for the year 2007.

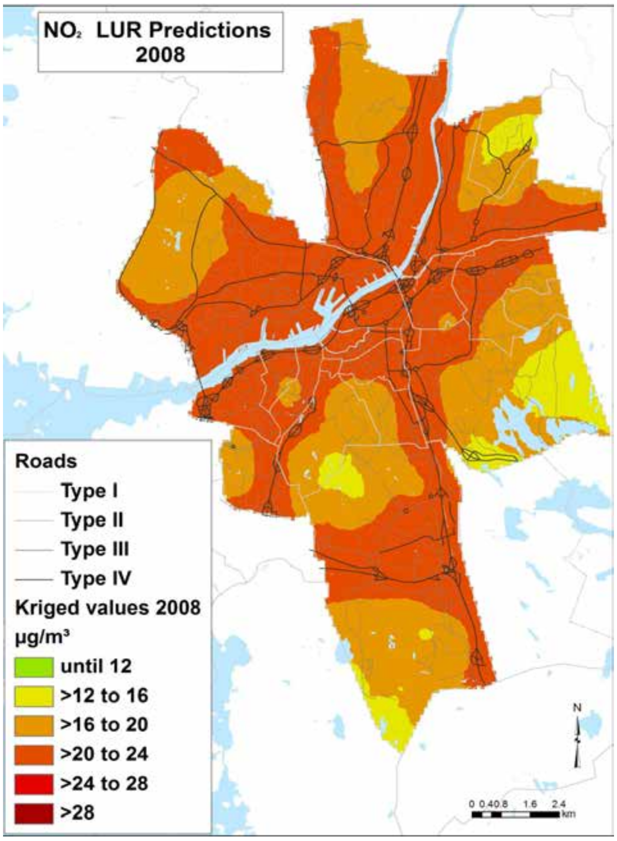

Figure S08. Kriged surface of NO_2 concentration based on LUR model for the year 2008.

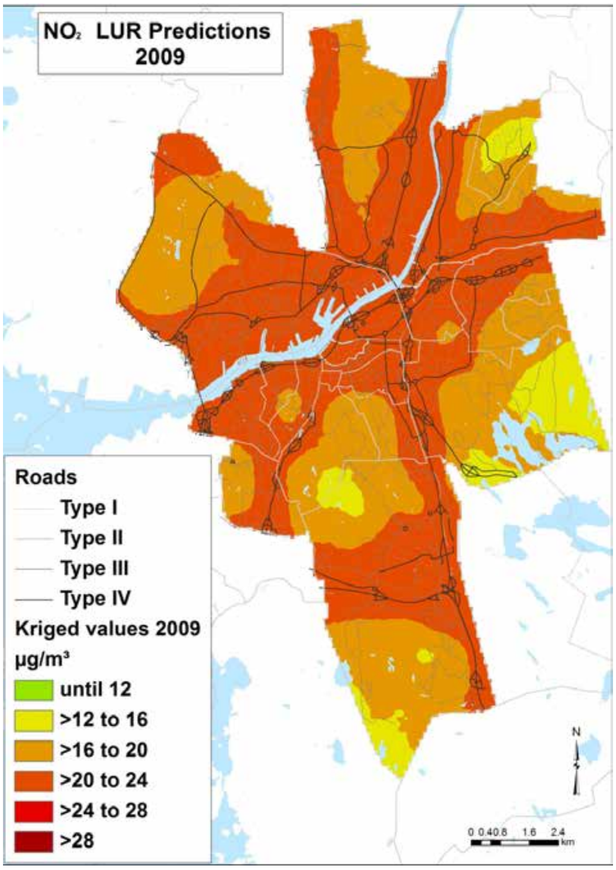

Figure S09. Kriged surface of NO_2 concentration based on LUR model for the year 2009.

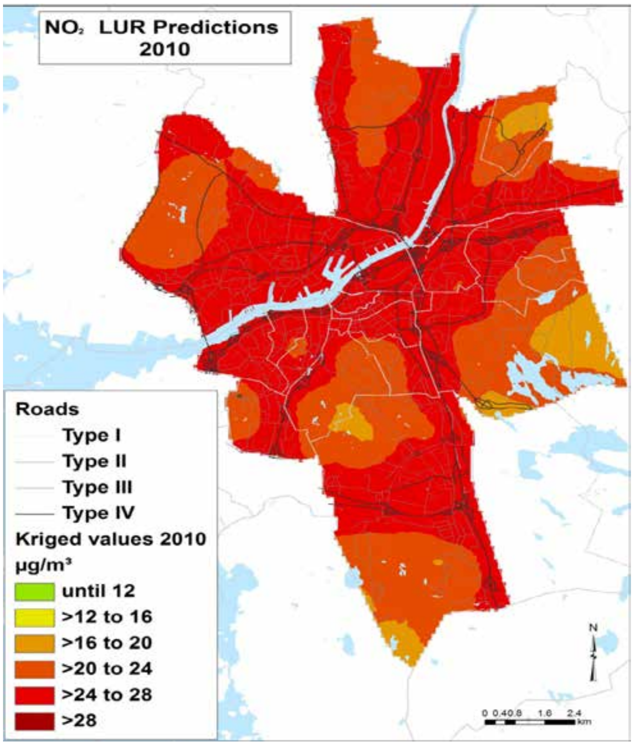

Figure S10. Kriged surface of NO_2 concentration based on LUR model for the year 2010.

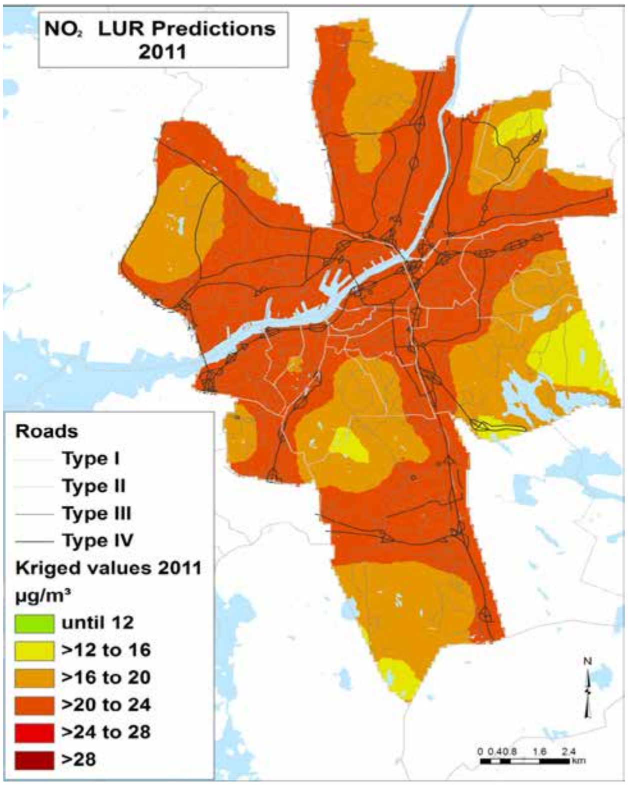

Figure S11. Kriged surface of NO_2 concentration based on LUR model for the year 2011.

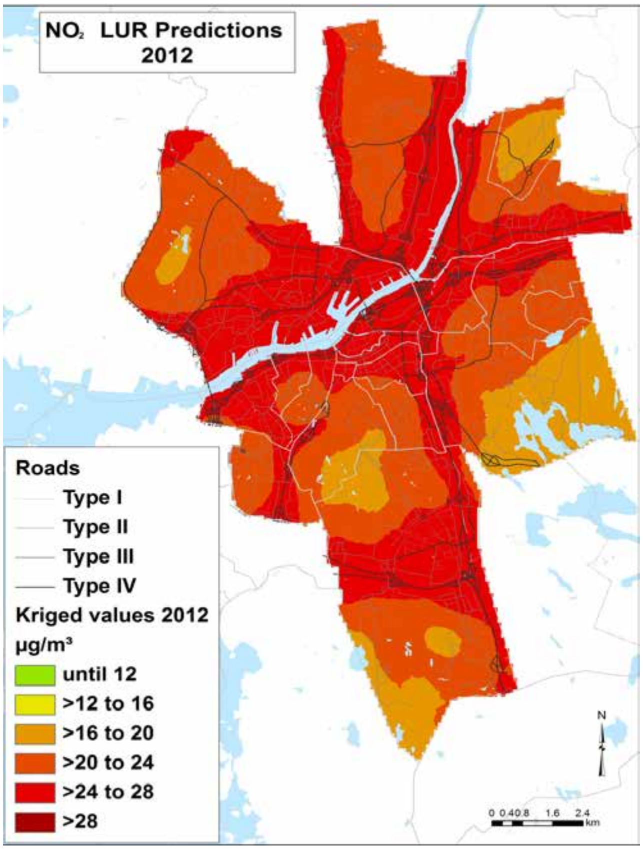

Figure S12. Kriged surface of NO_2 concentration based on LUR model for the year 2012.

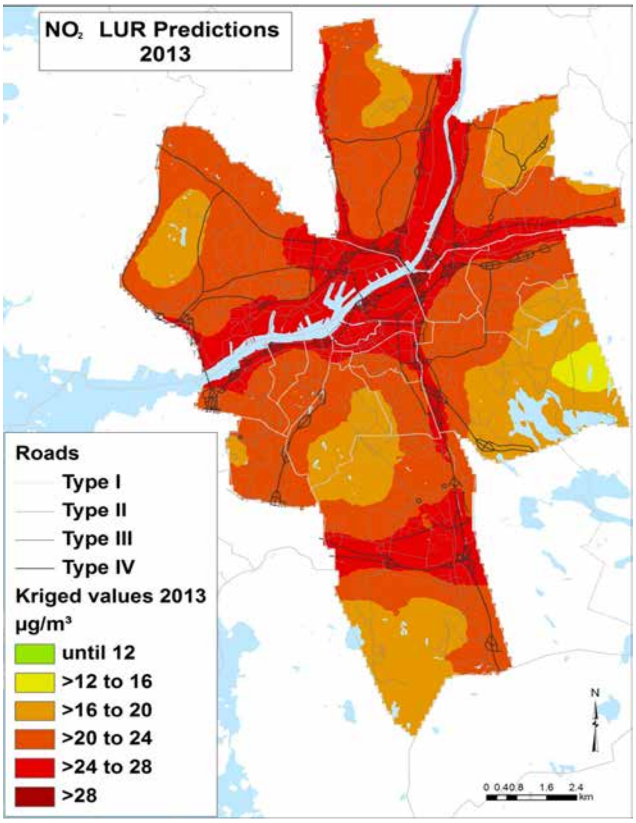

Figure S13. Kriged surface of NO_2 concentration based on LUR model for the year 2013.