

Effect of monitoring network design on land use regression models for estimating residential NO₂ concentration



Hao Wu ^{a, b, *}, Stefan Reis ^{b, c}, Chun Lin ^a, Mathew R. Heal ^{a, *}

^a University of Edinburgh, School of Chemistry, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ, UK

^b Centre for Ecology & Hydrology, Bush Estate, Penicuik, Edinburgh, Midlothian, EH26 0QB, UK

^c University of Exeter Medical School, Knowledge Spa, Truro, TR1 3HD, UK

HIGHLIGHTS

- Land-use regression (LUR) models for NO₂ were evaluated using a dispersion model.
- The number of monitoring sites improved LUR model performance, but not > ~30 sites.
- Networks including sites in populated areas better estimated across residential NO₂.
- Roadside sites needed to better characterise the high end of residential NO₂.
- No specific monitoring site design estimated both overall and high NO₂ levels well.

ARTICLE INFO

Article history:

Received 23 June 2016

Received in revised form

20 October 2016

Accepted 6 November 2016

Available online 8 November 2016

Keywords:

Land-use regression model

Dispersion model

Exposure assessment

ABSTRACT

Land-use regression (LUR) models are increasingly used to estimate exposure to air pollution in urban areas. An appropriate monitoring network is an important component in the development of a robust LUR model. In this study concentrations of NO₂ were simulated by a dispersion model at ‘virtual’ monitoring sites in 54 network designs of varying numbers and types of site, using a 25 km² area in Edinburgh, UK, as an example location. Separate LUR models were developed for each network. The LUR models were then used to estimate NO₂ concentration at all residential addresses, which were evaluated against the dispersion-modelled concentration at these addresses. The improvement in predictive capability of the LUR models was insignificant above ~30 monitoring sites, although more sites tended to yield more precise LUR models. Monitoring networks containing sites located within highly populated areas better estimated NO₂ concentrations across all residential locations. LUR models constructed from networks containing more roadside sites better characterised the high end of residential NO₂ concentrations but had increased errors when considering the whole range of concentrations. No particular composition of monitoring network resulted in good estimation simultaneously across all residential NO₂ concentration and of the highest NO₂ levels. This evaluation with dispersion modelling has shown that previous LUR model validation methods may have been optimistic in their assessment of the model’s predictive performance at residential locations.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The assessment of long-term exposure to air pollution for epidemiological and health burden studies has been a challenge because of the high spatial variation of pollutant concentration in the urban environment, particularly nitrogen dioxide (NO₂) (Briggs,

2005; Jerrett et al., 2005). Over the years, land use regression (LUR) modelling has demonstrated better or equivalent performance to other geostatistical methods (Hoek et al., 2008), and therefore has become popular in health studies to estimate long-term exposure to ambient NO₂ (Beelen et al., 2014; Jerrett et al., 2009). LUR modelling is a stepwise multiple regression method that regresses the pollutant concentration at the measurement sites against the land-use variables within buffer areas around the measurement sites (Jerrett et al., 2007). The derived empirical relationship between pollutant concentration and surrounding land use is then

* Corresponding authors. University of Edinburgh, School of Chemistry, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ, UK.

E-mail addresses: h.wu@ed.ac.uk (H. Wu), m.heal@ed.ac.uk (M.R. Heal).

applied to un-sampled locations to provide a spatially-resolved seasonal or annual average pollution field.

The selection of monitoring sites to build the LUR model has been identified as one of the factors affecting the quality of the LUR model, but a rigorous method to determine the number and distribution of monitoring sites is lacking (Hoek et al., 2008). One study (Kanaroglou et al., 2005) aimed to develop a formal method to locate air quality monitors for LUR model development. However, the method has been rarely applied due to its complexity and the extensive prior knowledge required on the population and pollutant distributions. A few studies (Basagaña et al., 2012; Johnson et al., 2010; Wang et al., 2012) evaluated the effect of number of monitoring sites on LUR model performance, but the effect of the distribution of monitoring sites remains to be investigated.

Evaluation of an LUR model has always been limited to the measurements available in a monitoring campaign (Hoek et al., 2008). The ultimate goal of exposure assessment is to accurately predict the exposure of hundreds or thousands of study subjects, but evaluation of an LUR model at this level through measurements is practically impossible. However, with the use of a dispersion model it is possible to simulate a pseudo-measured concentration at every residential address, which can then be compared with an LUR-model estimated concentration to assess the validity of the latter.

The aim of this study was to evaluate a large suite of LUR models built from different monitoring network designs by using dispersion modelled concentration at each home address to assess the predictive power of the LUR models. This modelling study used as its basis the city of Edinburgh (population ~460,000) in the east of Scotland, UK (55.94° N, 3.18° W). The outcome of the evaluation is to recommend sampling strategies and to highlight how particular monitoring network designs may lead to potential exposure misclassification.

2. Method

The evaluation of the performance of a monitoring network for constructing an LUR model for estimating concentration at home addresses was carried out in four stages. An overview of the methodology is presented first, with details described in subsequent sub-sections. A schematic of the overall workflow is shown in Supporting Information (SI) Fig. S1.

The ADMS-Urban model v3.4 (CERC, 2015) was used to simulate NO₂ concentrations for each of the population home addresses (centre points of residential buildings) in a 5 km × 5 km study area in Edinburgh (Fig. 1). This area covers the commercial (city centre) and residential areas of the city and encompasses 7445 residential buildings housing a total population of 144,715. The dispersion modelled NO₂ concentration is considered to be the reference, on which the subsequent LUR model development and evaluation are based. Next, three different types of monitoring networks were designed (comprising different numbers of monitoring sites) based on household density and proximity to road. The NO₂ concentration at each monitoring site was modelled with ADMS-Urban using the same setup as the modelling of residential NO₂ concentration. The third stage was to develop a separate LUR model for each monitoring network, which was then applied to residential address to provide an LUR-model estimate. Finally, the LUR-model-estimated residential concentration was compared with the dispersion modelled residential concentration. The extent of agreement between the two indicates the performance of the LUR model and, in turn, the performance of the monitoring network from which the LUR model was constructed.

2.1. Stage 1 – dispersion modelling of residential NO₂ concentration

2.1.1. Data preparation

All the Geographic Information System (GIS) data, including buildings and road networks for the City of Edinburgh, were obtained from EDINA Digimap Ordnance Survey Service (Ordnance Survey, 2015) as ESRI Shapefiles. Annual average daily traffic (AADT) of traffic for the major roads in 2013 were downloaded from the Department for Transport (DfT, 2015). The population of each postcode area for the 2011 census was distributed to the buildings within the polygon area based on the volume of the building (building polygon area × building height). The centre of the building polygon with assigned population was used as the home address.

2.1.2. ADMS-urban setup

The model domain (12 km × 12 km) covered most of the City of Edinburgh where all the emissions of NO_x and NMVOC (non-methane volatile organic compound) were modelled (Fig. 1). Within this larger domain, a 5 km × 5 km subset was chosen to output the concentration at each home address. To allow the receptors on the edges of the inner domain to be modelled smoothly, a 1 km buffer zone was added to the 5 km × 5 km output area, within which all the major and minor roads were explicitly modelled as road sources, whereas emissions outside the buffer zone were modelled as a 1 km × 1 km gridded area source. NO_x and NMVOC emissions were downloaded from the UK National Atmospheric Emissions Inventory (NAEI) for 2012 (NAEI, 2015) with a resolution of 1 km². Road emissions were calculated by dividing the total emissions for the major or minor road subsector by the total length of the corresponding roads within each 1 km² grid. For grids in which road emissions were explicitly modelled, the road emissions were subtracted from the grid total emission. Measured meteorological data for the model, including wind speed/direction, cloud cover and temperature, were obtained from a WMO station to the west of the model domain (Gogarbank: 55.93° N, 3.35° W) (Met Office, 2012). An urban canopy file was prepared to account for the variation in the vertical profiles of wind speed and turbulence caused by the presence of buildings. Background concentrations of hourly-average NO₂, NO_x and O₃ were obtained from a rural national-network monitoring station to the south of the model domain (Bush Estate: 55.86° N, 3.21° W). For 2012, 0.8% and 23% of O₃ and NO_x measurements, respectively, were missing. These were replaced by the average concentration for that particular hour over the whole year. Monthly average concentration was calculated at each receptor in ADMS-Urban, from which the annual-average concentration was calculated as the metric in the subsequent analysis.

2.2. Stage 2 – sampling network design

Three different types of sampling networks were investigated. The aim in the selection of monitoring sites was to investigate how network selection criteria and number of sites influence the representation of the spatial variation of NO₂ at the residential home addresses. Specifically, the exposure study area was first disaggregated into 25 m × 25 m grid cells. The following GIS variables were then calculated for the centroid of each grid cell: total home addresses within a 100 m buffer (HH100) and distance to major/minor road edge (MJRDDIST/MNRDDIST). Three types of monitoring sites were then defined:

- High household density sites (HH sites): centroid of the cells with HH100 falling in the top 10% of all the 25 m grid cells;

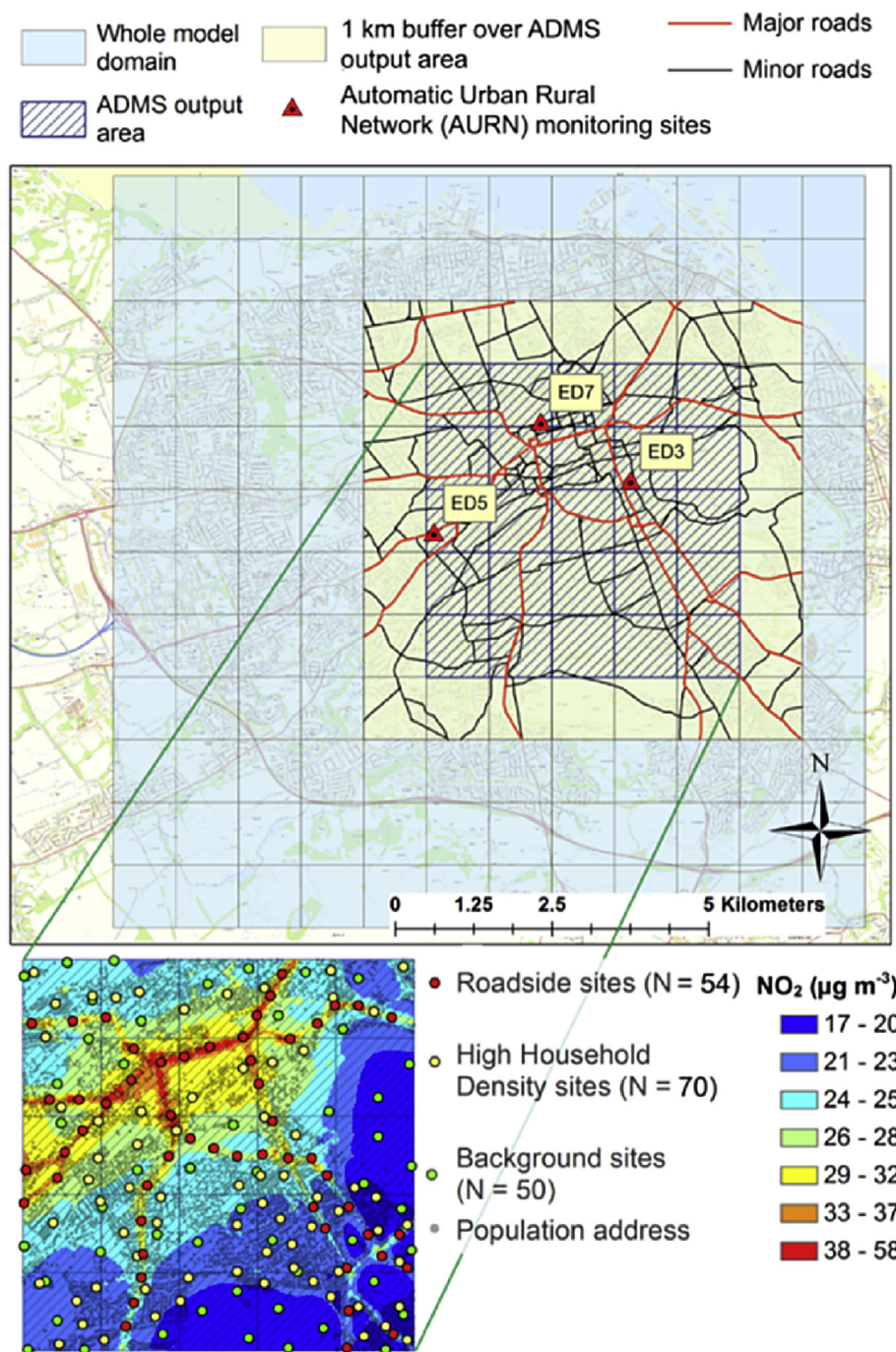


Fig. 1. The modelling domain for the city of Edinburgh. The inset map shows the locations of all potential monitoring sites. The underlying contour plot shows the monthly average dispersion model concentration of NO_2 for April 2012.

- Roadside sites: centroid of the cells with MJRDDIST between 0 and 5 m;
- Background sites: centroid of the cells with both MJRDDIST and MNRDDIST > 200 m.

A subset of each of the total set of HH and roadside sites was then randomly selected, subject to a minimum distance of 300 m between any pair of sites, to form two pools of potential HH and roadside monitoring sites to be used in the sampling networks. The locations of these sets of monitoring sites are shown in Fig. 1. The purpose of adding a minimum distance constraint to the random

selection was to ensure that potential network sites were distributed across the range of localities in the study area. A third subset of sites was randomly selected from the total set of background sites, but with a minimum distance constraint of 500 m, since the background concentrations in this modelling study are mainly determined by the gridded emissions which have a resolution of 1 km^2 rather than by the road network which has a finer resolution. Due to the minimum distance constraint added to the random selection, the number of monitoring sites of each type from which a monitoring network could subsequently be selected comprised 54, 70 and 50 for roadside, HH and background sites, respectively

(Fig. 1).

From these potential monitoring sites, the following three types of sampling networks were designed by randomly selecting different numbers of sites from each type of monitoring site.

- Household density based network (*HH* network): randomly selecting from the HH sites only;
- Proximity to road based network (*Road* network): randomly selecting equal numbers of roadside sites and background sites;
- *Mixed* network: randomly selecting equal numbers of roadside sites and HH sites.

Eleven different numbers of monitoring sites were tested for each type of network design ranging from 10 to 60 (in steps of $N = 5$). Random sampling of each number of monitoring sites was repeated 30 times to obtain a statistical distribution of a particular network configuration, resulting in 990 unique networks (3 network designs \times 11 network sizes \times 30 random replications). Table S1 summarises the configurations of all the networks examined. As a further network sensitivity test, different proportions of roadside and HH sites within the *Mixed* network were investigated to evaluate the effect of network composition in estimating residential NO_2 concentration. Table S2 summarises the different network compositions investigated.

The *HH* networks were designed in the anticipation that such networks would more accurately estimate concentrations at most residential addresses. However this sampling design might under-predict concentrations for a small fraction of population who live close to roads. The *Road* networks, being a mixture of roadside and background sites, should capture the greatest NO_2 variation in the study area; this is the network site selection design used in many monitoring campaigns (Beelen et al., 2013). The *Mixed* networks of roadside and HH sites aimed to capture similar spatial variation of NO_2 as the *Road* network, but also to represent where most the population live. This sampling design resembles the concept of a formal methodology for locating monitoring sites (Kanaroglou et al., 2005), namely locating monitors where the expected pollution spatial variability and density of the study subjects are high. Unlike the formal methodology, however, the sampling design here does not require prior knowledge of the pollutant concentration surface, therefore the application of this sampling design is less restricted.

2.3. Stage 3 – LUR modelling

2.3.1. Predictor variables

A total of 15 predictor variables were selected for model development (Table 1). These variables were chosen based on prior knowledge that they may correlate with the input emissions in ADMS-Urban and their inclusion in previous LUR models for NO_2 (Beelen et al., 2013). As shown in Fig. S2, NO_x emissions for each of the 1 km² grids in the study area are mostly dominated by road transport and combustion in commercial/residential sectors. The total road length, population counts and building plan area within a buffer radius are considered to reflect these emissions. In addition, in some areas, NO_x emissions from ‘other’ transport (most likely resulting from railways) are also significant. Therefore total railway length within a buffer was also included as a predictor variable. Since the emissions apart from major roads and some minor roads were modelled as 1 km² grid sources in ADMS-Urban, the buffer radii for the relevant predictor variables were chosen to be comparable with the resolution of input emissions, namely 0.5 and 1 km (Table 1). The rest of the predictor variables attempt to account for the increase in NO_2 concentration close to road sources (Table 1).

2.3.2. LUR model development and diagnostics

The development of the LUR models followed the method used in the ESCAPE project (Beelen et al., 2013). The method is a supervised forward stepwise procedure which aims to maximise the adjusted R^2 of the model while also ensuring that the included variables are associated with coefficients with pre-defined directions (Table 1).

First, all variables were individually regressed against the NO_2 concentrations in that monitoring network. The variable with the highest adjusted R^2 and a coefficient with pre-defined direction formed the initial model. Second, the remaining variables were successively added to the start model and the change in adjusted R^2 recorded. The variable resulting in the highest increase in adjusted R^2 was added to the model if: (i) the increase in adjusted R^2 was greater than 1%; and (ii) the coefficients of this variable and the variables already in the model conformed to the pre-defined direction. The selection process was continued until no variable fulfilled the above criteria. At the final step, variables with p -value greater than 0.1 were subsequently removed from the model starting from the variable with the highest p -value.

Diagnostic tests were performed on the final model. Multicollinearity in the variables was checked using Variance Inflation Factor (VIF). Predictors with high VIF value (>3) were excluded from the model one at a time starting with the variable with the highest VIF. Potential influential observations were investigated using Cook's D value. An influential observation (indicated by a Cook's $D > 1$) was generally caused by including a variable with extreme values or many zero values. A sensitivity test was therefore conducted on a model with an influential observation problem by fitting a new model without using the observation with Cook's $D > 1$. If the change in the coefficient for that variable was large (over 100% of the coefficient derived from using all the observations), a new LUR model was developed following the above procedure but excluding that specific variable from the outset.

For the LUR model validation, leave-one-out-cross-validation (LOOCV) was used to assess the generalisability of the LUR model. LOOCV uses the variables in the final model to develop a regression model using $N - 1$ observations (N = total number of observations in a monitoring network), which was then applied to the leave-out site. The procedure was repeated N times at which point all the predicted concentrations are compared with the observations to test the validity of the model within the dataset. Values of R^2 and Root Mean Squared Error (RMSE) calculated from LOOCV were used to assess the LUR model's capability to predict the concentrations within a monitoring network.

2.4. Stage 4 – evaluation of LUR model's capability at estimating simulated NO_2 concentrations at residential addresses

This aspect of LUR model evaluation compares the LUR modelled concentration at residential address with that modelled by ADMS-Urban. In essence this is similar to the concept of hold-out validation (HV) in a regression model validation, where the training data and testing data are completely independent. However the validation dataset is based on ADMS-Urban output and is of constant size and much larger (7445 residential addresses) than the traditional HV validations based on measurement data. In this context, the evaluation results not only reflect the performance of the LUR model but also indicate the relative effectiveness of the underlying monitoring sites used to build the LUR model. R^2 , RMSE and Mean Bias (MB) were used here to evaluate the LUR modelled concentration for all population addresses and for different concentration ranges.

All GIS calculations were conducted in the Feature Manipulation Engine (FME) (Safe Software Inc., 2015). Statistical analyses were

Table 1
Predictor variables with buffer sizes and a priori defined directions of effect on NO₂ concentration.

Predictor variables	Abbreviation	Unit	Buffer radius (m)	Direction of effect
Population counts	POP	n	500, 1000	+
Building plan area	BA	m ²	500, 1000	+
Total major and minor road length	RDLEN	m	500, 1000	+
Total railway length	RAILLEN	m	500, 1000	+
Distance to the nearest major road (inverse distance and inverse distance squared)	INVMJRDDIST, INVMJRDDIST2	m ⁻¹ , m ⁻²	NA	+
Distance to the nearest road (inverse distance and inverse distance squared)	INVDIST, INVDIST2	m ⁻¹ , m ⁻²	NA	+
Traffic volume on nearest major road	ALLTRAF	veh.day ⁻¹	NA	+
Product of traffic intensity on nearest major road and inverse distance to the nearest major road and inverse distance squared	TRAFDIST, TRAFDIST2	veh.day ⁻¹ m ⁻¹ , m ⁻²	NA	+

conducted in R software (R Core Team, 2015).

3. Results

3.1. ADMS-urban model validation

ADMS-Urban was evaluated against measurements taken by both reference chemiluminescence analyser and passive diffusion tube (PDT). Comparison between the modelled annual average concentration of 2012 and the measurement by reference analyser at three monitoring stations in the study area showed that the bias was small at urban background (ED3) and minor roadside (ED7) (Table 2). The relatively large underestimation at major roadside (ED5) could be associated with the known issue of under-reporting of NO_x from diesel vehicles (Carslaw and Rhys-Tyler, 2013).

A network of 30 PDT sampling sites within this study area were deployed weekly for 6 weeks during summer and winter periods of 2013/2014. Detail of the site locations and characteristics can be found in Table S3 and in Lin et al. (2016). Seasonal average concentration (i.e. mean of 6 weekly average NO₂ concentrations) was compared with ADMS-Urban output. Sites containing more than 1 week's missing weekly NO₂ were excluded from the model evaluation. Some PDTs were located next to bus stop or at the traffic junction, where additional emissions from buses and traffic queueing are considered to be great but not modelled in the current model setup. These sites therefore do not reflect the general predictive ability of the ADMS-Urban model and were also excluded from the evaluation. Fig. 2 shows the relationship between modelled and PDT-measured NO₂ concentrations during different seasons. Overall the model underestimated NO₂ concentrations compared to the PDT measurements. However the spatial variation in the measured NO₂ was explained very well by the model ($R^2 = 73\%$ and 77% for summer and winter, respectively) and was comparable to a previous ADMS model evaluation study (Dédélé and Miškinytė, 2014). This indicates that although there is bias between modelled and PDT-measured NO₂ concentration the spatial pattern predicted by the model is consistent with the measurements. The bias could result from both the errors in the model and the errors in the PDT measurements. Large discrepancy (55% for summer and 82% for winter) between PDT measurement

and reference analyser was observed during the deployment period at one co-location site (Table S4). This partly explains the general underestimation in the modelled NO₂ compared to the PDT measurements.

Given the good agreement between the model and real-time analyser measurements at the urban background and minor roadside monitor locations, and the very good capture of spatial pattern indicated by the dense PDT network, it can be deduced that the dispersion model here fulfils the purpose of this study; that is, to simulate a realistic pollution surface of NO₂ for the evaluation of the LUR model validity and of the monitoring sites used to build the LUR model.

3.2. Evaluation of the LUR models constructed from different monitoring networks

The distributions of NO₂ concentrations at the locations of each type of monitoring site, and at all the population addresses, are summarised in Fig. S3. Consistent with the expectations underpinning the network design principles, Fig. S3 shows that a Road network (roadside sites + background sites) is likely to cover the whole range of concentration across the modelled domain, whereas a HH network (only HH sites) matches most closely the interquartile range of residential NO₂ concentration.

Fig. 3 summarises the following statistics evaluating LUR model performance as a function of network design and size: (i) the percentage of variance explained within the data used to build the LUR model (LUR R^2); (ii) the ability of the LUR model to predict the observed concentrations at the virtual monitoring sites (LOOCV R^2 and LOOCV RMSE); and (iii) the effectiveness of the monitoring networks at predicting concentrations at all the residential addresses (Residential R^2 and Residential RMSE). Fig. 3 shows that LUR R^2 and LOOCV R^2 slightly decreased with increasing network size, while LOOCV RMSE slightly increased. In contrast, the effectiveness of the monitoring networks at predicting residential NO₂ concentration improved with increasing network size as shown by the increasing Residential R^2 and decreasing Residential RMSE (Fig. 3). The improvement in the prediction of residential concentration (Residential R^2 and RMSE) was, however, insignificant between LUR models constructed with >30 monitoring sites, as

Table 2
Dispersion model versus measured annual average NO₂ concentration in 2012. ED3, ED5 and ED7 are three real-time monitoring stations (shown in Fig. 1) located at urban background, major roadside and minor roadside, respectively.

	Modelled NO ₂ (μg m ⁻³)	Measured NO ₂ (μg m ⁻³)	Over/under-estimation by the model
ED3	24.8	24.1	2.8%
ED5	35.4	39.1	−9.5%
ED7	27.5	28.1	−2.0%

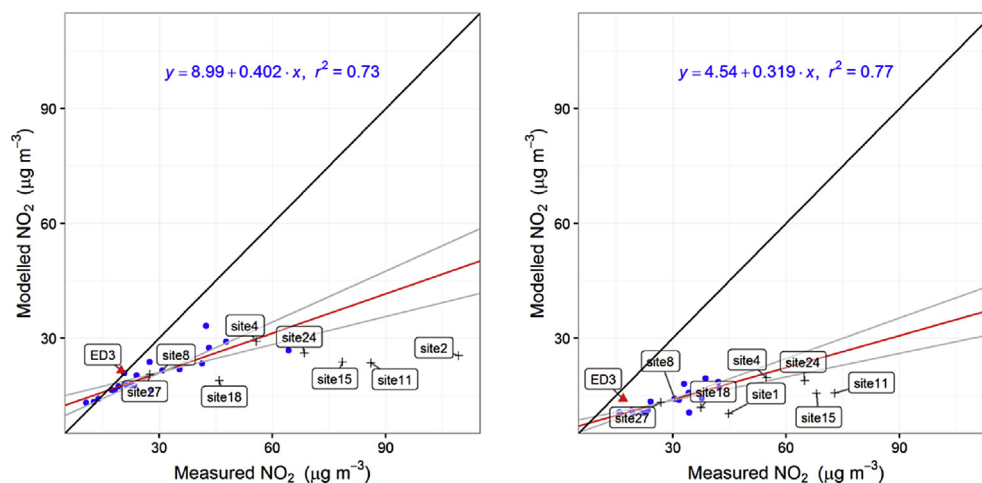


Fig. 2. Comparison of modelled and measured NO_2 concentration for summer (a) and winter (b) seasons. The cross markers denote the sites excluded from regression analysis due to special local effects as described in the text. Site 8 PDTs were co-located with reference analyser at ED3 marked by the red triangle. This site is also marked on Fig. 1.

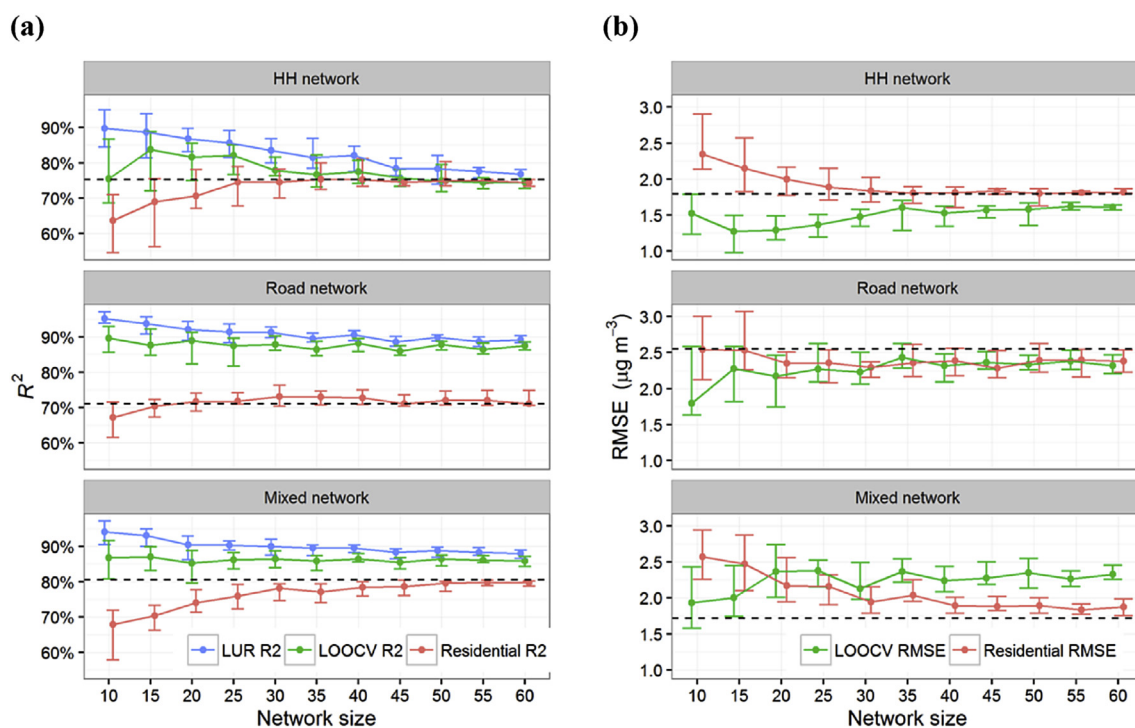


Fig. 3. Diagnostic statistics for LUR models as a function of network design and size for simulating network site concentrations (LUR R^2 and LOOCV R^2 shown in (a); LOOCV RMSE shown in (b)); and for predicting residential NO_2 concentration (Residential R^2 in (a); Residential RMSE in (b)). The points represent the median of the statistics for the 30 random repetitions of each network configuration. The whiskers extend to 25th and 75th percentiles of the statistics for the 30 random repetitions of each network configuration. The horizontal dashed lines denote the Residential R^2 and RMSE if all the potential monitoring sites (70, 104 and 124 for HH, Road and Mixed networks, respectively) are used for calculation.

indicated by the overlap of inter-quartile range of the statistic calculated from 30 random repetition. The fact that the LOOCV R^2 was significantly higher than the Residential R^2 across the network size for Road and Mixed networks (Fig. 3a) suggests that using LOOCV to evaluate the LUR model's predictive ability might be overly optimistic. The contrast between the performance of the LUR model and its ability to predict residential NO_2 concentration was especially large for the Road network design (comprising a mixture of roadside and background sites), and for the other network designs when there were only 10 or 15 monitoring sites (Fig. 3a). The

most effective type of monitoring network was the Mixed network, as indicated by the highest Residential R^2 limit and lowest Residential RMSE limit. The variability of Residential R^2 and RMSE in the 30 random repetitions of each network configuration (whiskers in Fig. 3) decreased with increasing network size, suggesting that larger number of monitoring sites better capture the actual relationship between predictor variables and NO_2 concentration, hence less between-LUR-model variabilities.

The performances of the LUR models in estimating residential concentration within three separate ranges of NO_2 concentration

are compared in Fig. 4. At the low end of NO_2 concentration ($<20 \mu\text{g m}^{-3}$), RMSE was similar between *Mixed* and *HH* networks, but both *HH* and *Road* networks significantly overestimated (MB) the overall residential NO_2 concentration. For NO_2 concentrations between 20 and $30 \mu\text{g m}^{-3}$, the *HH* networks generally underestimated the residential concentration (Fig. 4b). The most distinctive difference between the three network designs was observed at the high end of NO_2 concentration ($>30 \mu\text{g m}^{-3}$). For these NO_2 concentrations, the prediction errors (RMSE) and the extent of overall underestimation (MB) were significantly higher for the *HH* networks (Fig. 4). *Mixed* and *Road* networks performed similarly at high NO_2 concentration, although they both still, on average, underestimated (Fig. 4b). Similar to the statistics in Fig. 3, the variability of estimation errors also reduced with increasing network size. Overall, considering the results shown in Figs. 3 and 4 together, the *Mixed* networks were most effective in estimating residential NO_2 concentration when considering both all residential addresses together and subsets of addresses in different ranges of NO_2 concentrations.

Fig. 5 shows the results of the investigation of the different proportions of HH sites and roadside sites within the *Mixed* network on the LUR model predictions of residential NO_2 concentrations. There was no significant trend in the R^2 values across the different proportions of roadside sites in the *Mixed* network (Fig. 5a). The prediction error (RMSE) increased with increasing proportions of roadside sites (Fig. 5b). The variability of the RMSE resulting from networks consisting only of roadside sites was the largest, whereas the LUR models derived from networks with more HH sites were relatively more precise (Fig. 5b).

The performance of networks containing different percentages of roadside sites at different NO_2 concentration ranges are compared in Fig. 6. For NO_2 concentration $<30 \mu\text{g m}^{-3}$, the RMSE increased with increasing percentage of roadside sites (Fig. 6a) and the MB suggested overestimation for networks with high roadside

site composition (Fig. 6b). However at NO_2 concentration $>30 \mu\text{g m}^{-3}$, the RMSE decreased with increasing percentage of roadside sites (Fig. 6a) and higher roadside site composition led to reduced bias (Fig. 6b). Fig. 6a shows that LUR models constructed with only roadside sites resulted in high variabilities in the RMSE at NO_2 concentration $<30 \mu\text{g m}^{-3}$, whereas LUR models constructed with all HH sites resulted in high variabilities in the RMSE at high level of NO_2 concentrations ($>30 \mu\text{g m}^{-3}$).

4. Discussion

In most LUR studies, LOOCV and/or hold-out validation (dividing monitoring sites into two independent sets for model development and validation) have been used to validate the LUR models. LOOCV tests how well the LUR model predicts the observation within the training dataset. Hold-out validation evaluates the predictive ability of the LUR model at locations that were not used in model development. The latter evaluation is of more interest for an LUR model; in practice, however, there is always a trade-off between building a more robust LUR model using a larger training dataset and giving more power to the evaluation using a larger validation dataset. A limited number of monitoring sites in many studies makes the division of the dataset even more difficult. Evaluation of LUR models on all potential exposure subjects has been unfeasible in reality. However, this can be achieved by using a dispersion model to provide a realistic spatial field of urban ambient NO_2 concentration. Although there may be uncertainties in the dispersion-modelled concentrations, the nature of the errors should be similar at the virtual monitoring sites and at the residential addresses.

As expected, more monitoring sites yielded better estimation of residential NO_2 concentration by the LUR model (Fig. 3). For all three network types, however, the improvement in the estimation was insignificant for networks with more than ~30 monitoring

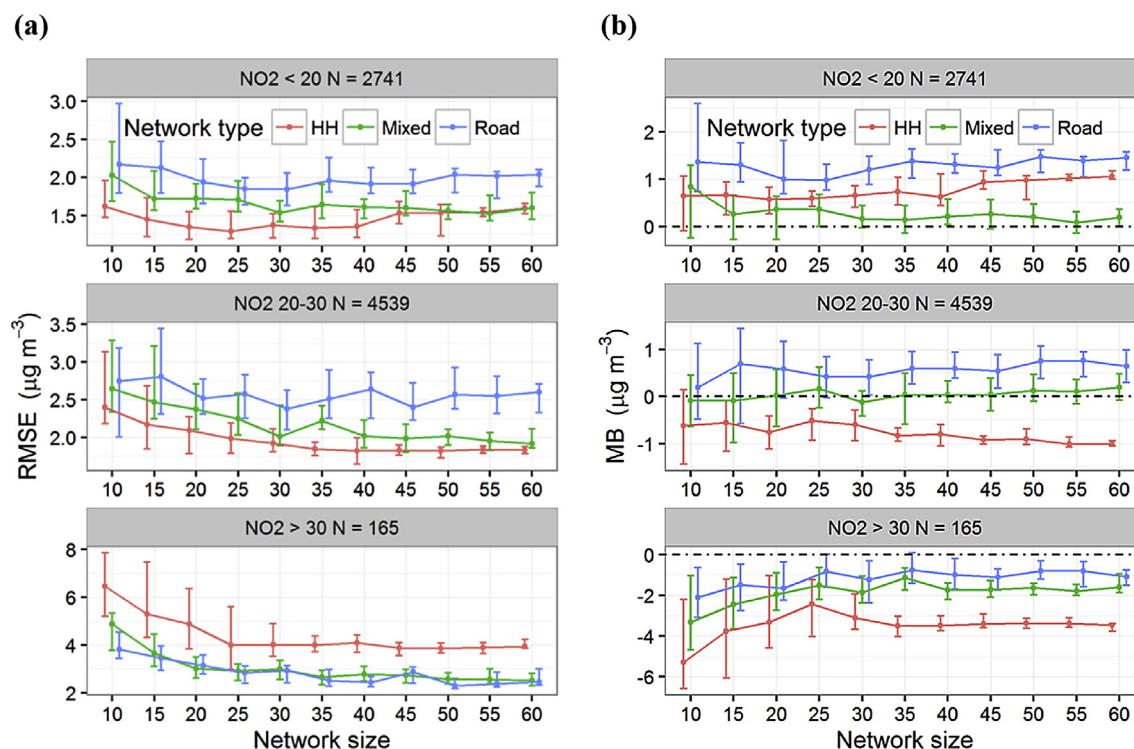


Fig. 4. Summary statistics of (a) RMSE and (b) MB in estimating residential concentration for different ranges in NO_2 concentration. The whiskers extend to 25th and 75th percentiles of the statistics for the 30 repetitions of each network configuration.

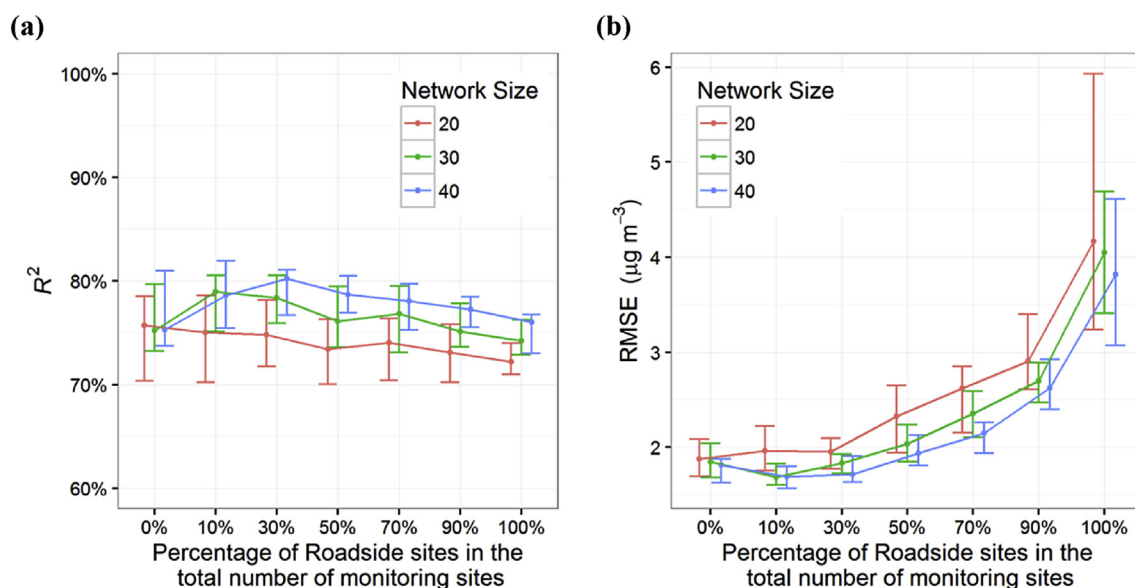


Fig. 5. Summary statistics of (a) R^2 and (b) RMSE for Mixed networks containing different proportions of roadside sites in predicting residential NO₂ concentration. The whiskers extend to 25th and 75th percentiles of the statistics for the 30 repetitions of each network configuration.

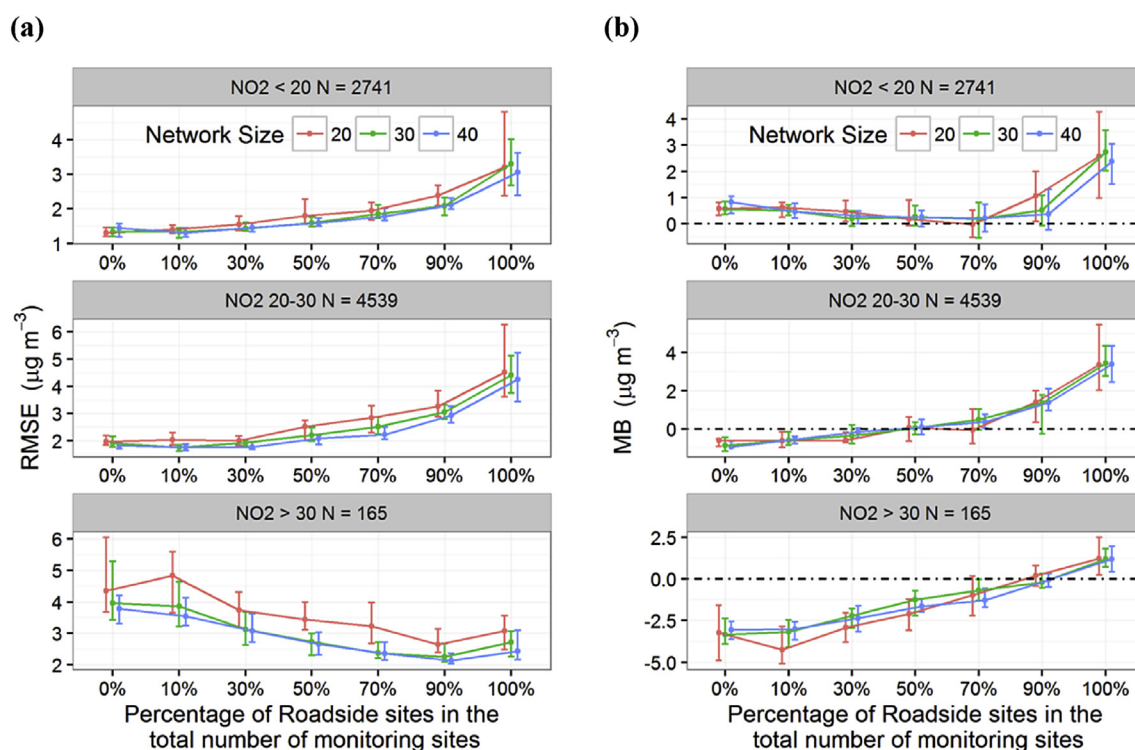


Fig. 6. Summary statistics of (a) RMSE and (b) MB in estimating residential concentration in different NO₂ concentration ranges for Mixed networks containing different proportions of roadside sites. The whiskers extend to 25th and 75th percentiles of the statistics for the 30 repetitions of each network configuration.

sites. Although the improvement was insignificant, higher number of monitoring sites increased the stability of the developed LUR models as shown by the very small inter-quartile range for the statistics at larger network sizes in Figs. 3 and 4. As the number of monitoring sites increased, the number of unique variables appearing in the LUR models decreased (Table S5), indicating that a greater number of monitoring sites was more effective at eliminating insignificant predictor variables. This is consistent with the

findings of Basagaña et al. (2012) using actual NO₂ measurements. In our work, it was found that ~30 observations are sufficient to capture the spatial variation of the residential NO₂ concentrations in a dispersion modelled pollution surface of an urban area of 25 km², but this number is expected to be larger in reality due to local effects (e.g. street canyon effect and traffic queueing) that were not modelled by ADMS-Urban and for larger areas than simulated in this study. Basagaña et al. (2012) showed that the

improvement of R^2 in hold-out validation was minor after ~60 monitoring sites in a study area of 45.7 km². In a national wide Netherlands study, LUR models constructed with over ~90 monitoring sites seemed to result in similar prediction ability. Collectively these results suggest that a minimum optimal number of monitoring exists but depends on the actual study area.

In the ESCAPE study (Beelen et al., 2013) and many other LUR studies (Aguilera et al., 2008; Madsen et al., 2007), urban background sites were selected in conjunction with roadside sites to build the LUR model. Urban background sites are usually defined with respect to the distance to road source or traffic activity within a certain buffer, irrespective of the distribution of the exposure study subjects, as was represented by the *Road* network design in this study. Fig. 3a shows that LUR models derived from such networks were generally poorer at estimating NO₂ concentration at residential addresses than LUR models derived from networks with sites selected on the basis of household density. LUR models derived from *Mixed* networks were better at estimating residential NO₂ concentration than those derived from *Road* networks (Fig. 3) and also gave comparable errors to *Road* network-derived LUR models for estimating concentrations at the high end of the distribution (Fig. 4). This observation emphasises the importance of characterising both the concentration and population distribution in the study area when designing a monitoring network.

The composition of different types of measurement sites in most monitoring networks used to construct LUR models to date has been rather arbitrary. Some researchers (Cyrys et al., 2012) followed the principle of over-representing the roadside sites with respect to the fraction of addresses close to the roads, as this captures the spatial variation of NO₂. In our work, LUR models constructed from networks containing 0–30% of roadside sites (compared with 0.2% of addresses within 10 m to the roads in the study area) showed lower estimation errors (Fig. 5b) compared to other network compositions for all three network sizes tested. When examining the estimated residential concentrations at different NO₂ levels, LUR models constructed from networks containing 0–30% of roadside sites resulted in larger errors at high NO₂ concentrations compared to networks containing higher proportions of roadside sites (Fig. 6). The results here suggest that a greater proportion of roadside sites in a monitoring network yielded LUR models that better characterised the higher end of the residential NO₂ concentration (Fig. 6) but also introduced greater prediction error considering the population as a whole (Fig. 5), and vice versa for LUR models derived from networks containing a greater proportion of HH sites. No particular network composition was simultaneously able to provide an LUR model capable of good overall estimation of the residential NO₂ concentrations and a good estimation of the higher end concentration. This illustrates the limitation of LUR models to capture the spatial contrast in residential NO₂ concentration predicted by the dispersion model.

As a common LUR model evaluation method, the LOOCV R^2 statistic was found to overestimate the LUR predictive ability, consistent with the limited number of other studies on the same topic (Basagaña et al., 2012; Johnson et al., 2010; Wang et al., 2012). Collectively, the results from these studies highlight the limited predictability of empirical NO₂ LUR models that are highly dependent on the measurement sites. Dispersion modelling, as demonstrated in this study, is a potentially useful tool to design an effective monitoring network and to better evaluate the LUR models in a way that is otherwise unfeasible in reality.

We acknowledge that the area of our domain (25 km²) is smaller than some LUR studies (Aguilera et al., 2008; Fernández-Somoano et al., 2011). This choice was mainly limited by the intensive computational requirement of the dispersion model to calculate concentration at the large number of residential addresses.

Clustering addresses with similar characteristics would reduce the calculation time and facilitate dispersion modelling over a larger area. In this study, a dispersion model provided the NO₂ concentrations for development of the LUR models and for evaluation of their predictive capabilities. Whilst accepting potential discrepancies between dispersion model and real measurements, this work shows that more comprehensive evaluation of LUR models and their underpinning monitoring networks is needed. Although we only evaluated the LUR models for NO₂, results for the effect of number and type of monitoring sites on LUR model performance should be transferable to other traffic-related air pollutants such as black carbon and ultrafine particle number, given their mutual high correlations.

5. Conclusions

Using a greater number of sites to build an LUR model improved its ability to estimate residential NO₂ concentrations. However, improvement in LUR model predictive capability was not significant beyond a certain number of monitoring sites: the predictive capability achieved using ~30 monitoring sites was similar to that achieved using 70–100 monitoring sites, but a greater number of monitoring sites tended to decrease imprecision. LUR models constructed from a network design incorporating both high household density areas and roadside sites better characterised the full range of residential concentrations and specifically those with highest concentrations. It is therefore recommended to incorporate monitoring sites representing most of the study subjects when designing of a monitoring network aimed at studying the health effects of air pollutants. The more roadside sites included in a monitoring network used to construct LUR model, the larger the RMSE for the estimation of residential NO₂ concentrations, but the lower the estimation error for high NO₂ concentrations. The fact that no particular proportion of roadside sites within the network design estimated well both the overall residential concentration and higher level of NO₂ concentrations suggested a lack of spatial contrast in LUR modelled pollution surface. A dispersion model has been shown to be a useful tool for both designing a monitoring network for LUR models and for the evaluation of the LUR models.

Acknowledgements

Hao Wu acknowledges funding from the University Of Edinburgh and the NERC Centre for Ecology & Hydrology (NERC CEH project number NEC04544).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.atmosenv.2016.11.014>.

References

- Aguilera, I., Sunyer, J., Fernández-Patier, R., Hoek, G., Aguirre-Alfaro, A., Meliefste, K., Bombai-Mingarro, M.T., Nieuwenhuijsen, M.J., Herce-Garraleta, D., Brunekreef, B., 2008. Estimation of outdoor NO_x, NO₂, and BTEX exposure in a cohort of pregnant women using land use regression modeling. *Environ. Sci. Technol.* 42, 815–821. <http://dx.doi.org/10.1021/es0715492>.
- Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., de Nazelle, A., Nieuwenhuijsen, M., Vila, J., Künzli, N., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmos. Environ.* 54, 634–642. <http://dx.doi.org/10.1016/j.atmosenv.2012.01.064>.
- Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.-Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K.T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nádor, G., Varró, M.J., Dèdèlè, A., Gražulevičienė, R., Mölter, A., Lindley, S., Madsen, C., Cesaroni, G., Ranzi, A.,

- Badaloni, C., Hoffmann, B., Nonnemacher, M., Krämer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Strömgren, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., de Hoogh, K., 2013. Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – the ESCAPE project. *Atmos. Environ.* 72, 10–23. <http://dx.doi.org/10.1016/j.atmosenv.2013.02.037>.
- Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z.J., Weinmayr, G., Hoffmann, B., Wolf, K., Samoli, E., Fischer, P., Nieuwenhuijsen, M., Vineis, P., Xun, W.W., Katsouyanni, K., Dimakopoulou, K., Oudin, A., Forsberg, B., Modig, L., Havulinna, A.S., Lanki, T., Turunen, A., Oftedal, B., Nystad, W., Nafstad, P., De Faire, U., Pedersen, N.L., Östenson, C.-G., Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K.T., Overvad, K., Ellermann, T., Eeftens, M., Peeters, P.H., Meliefste, K., Wang, M., Bueno-de-Mesquita, B., Sugiri, D., Krämer, U., Heinrich, J., de Hoogh, K., Key, T., Peters, A., Hampel, R., Concin, H., Nagel, G., Ineichen, A., Schaffner, E., Probst-Hensch, N., Künzli, N., Schindler, C., Schikowski, T., Adam, M., Phuleria, H., Vilier, A., Clavel-Chapelon, F., Declercq, C., Grioni, S., Krogh, V., Tsai, M.-Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Katsoulis, M., Trichopoulou, A., Brunekreef, B., Hoek, G., 2014. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* 383, 785–795. [http://dx.doi.org/10.1016/S0140-6736\(13\)62158-3](http://dx.doi.org/10.1016/S0140-6736(13)62158-3).
- 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. <http://dx.doi.org/10.1080/15287390590936094>.
- Carslaw, D.C., Rhys-Tyler, G., 2013. New insights from comprehensive on-road measurements of NO_x, NO₂ and NH₃ from vehicle emission remote sensing in London, UK. *Atmos. Environ.* 81, 339–347. <http://dx.doi.org/10.1016/j.atmosenv.2013.09.026>.
- CERC, 2015. ADMS User Guides [WWW Document]. <http://www.cerc.co.uk/environmental-software/user-guides.html> (accessed 3.11.16).
- Cyrys, J., Eeftens, M., Heinrich, J., Ampe, C., Armengaud, A., Beelen, R., Bellander, T., Beregszaszi, T., Birk, M., Cesaroni, G., Cirach, M., de Hoogh, K., de Nazelle, A., de Vocht, F., Declercq, C., Dedele, A., Dimakopoulou, K., Eriksen, K., Galassi, C., Grauleviciene, R., Grivas, G., Gruzdeva, O., Gustafsson, A.H., Hoffmann, B., Iakovides, M., Ineichen, A., Krämer, U., Lanki, T., Lozano, P., Madsen, C., Meliefste, K., Modig, L., Mölter, A., Mosler, G., Nieuwenhuijsen, M., Nonnemacher, M., Oldenwening, M., Peters, A., Pontet, S., Probst-Hensch, N., Quass, U., Raaschou-Nielsen, O., Ranzi, A., Sugiri, D., Stephanou, E.G., Taimisto, P., Tsai, M.-Y., Vaskövi, E., Villani, S., Wang, M., Brunekreef, B., Hoek, G., 2012. Variation of NO₂ and NO_x concentrations between and within 36 European study areas: results from the ESCAPE study. *Atmos. Environ.* 62, 374–390. <http://dx.doi.org/10.1016/j.atmosenv.2012.07.080>.
- DfT, 2015. City of Edinburgh [WWW Document]. Dep. Transp. – Traffic Counts. <http://www.dft.gov.uk/traffic-counts/cp.php?la=City+of+Edinburgh>.
- Dedele, A., Miskinyte, A., 2014. Estimation of inter-seasonal differences in NO₂ concentrations using a dispersion ADMS-Urban model and measurements. *Air Qual. Atmos. Health* 8, 123–133. <http://dx.doi.org/10.1007/s11869-014-0272-9>.
- Fernández-Somoano, A., Estarlich, M., Ballester, F., Fernández-Patier, R., Aguirre-Alfaro, A., Herce-Garraleta, M.D., Tardón, A., 2011. Outdoor NO₂ and benzene exposure in the INMA (environment and childhood) Asturias cohort (Spain). *Atmos. Environ.* 45, 5240–5246. <http://dx.doi.org/10.1016/j.atmosenv.2011.02.010>.
- Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* 42, 7561–7578. <http://dx.doi.org/10.1016/j.atmosenv.2008.05.057>.
- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. *J. Expo. Sci. Environ. Epidemiol.* 15, 185–204. <http://dx.doi.org/10.1038/sj.jea.7500388>.
- Jerrett, M., Arain, M.A., Kanaroglou, P., Beckerman, B., Crouse, D., Gilbert, N.L., Brook, J.R., Finkelstein, N., Finkelstein, M.M., 2007. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *J. Toxicol. Environ. Health A* 70, 200–212. <http://dx.doi.org/10.1080/15287390600883018>.
- Jerrett, M., Finkelstein, M.M., Brook, J.R., Arain, M.A., Kanaroglou, P., Stieb, D.M., Gilbert, N.L., Verma, D., Finkelstein, N., Chapman, K.R., Sears, M.R., 2009. A cohort study of traffic-related air pollution and mortality in Toronto, Ontario, Canada. *Environ. Health Perspect.* 117, 772–777. <http://dx.doi.org/10.1289/ehp.11533>.
- Johnson, M., Isakov, V., Touma, J.S., Mukerjee, S., Özkaynak, H., 2010. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmos. Environ.* 44, 3660–3668. <http://dx.doi.org/10.1016/j.atmosenv.2010.06.041>.
- Kanaroglou, P.S., Jerrett, M., Morrison, J., Beckerman, B., Arain, M.A., Gilbert, N.L., Brook, J.R., 2005. Establishing an air pollution monitoring network for intra-urban population exposure assessment: a location-allocation approach. *Atmos. Environ.*, 12th International Symposium, Transport and Air Pollution 12th International Symposium. *Transp. Air Pollut.* 39, 2399–2409. <http://dx.doi.org/10.1016/j.atmosenv.2004.06.049>.
- Lin, C., Feng, X., Heal, M.R., 2016. Temporal persistence of intra-urban spatial contrasts in ambient NO₂, O₃ and O_x in Edinburgh, UK. *Atmos. Pollut. Res.* 7, 734–741. <http://dx.doi.org/10.1016/j.apr.2016.03.008>.
- Madsen, C., Carlsen, K.C.L., Hoek, G., Oftedal, B., Nafstad, P., Meliefste, K., Jacobsen, R., Nystad, W., Carlsen, K.-H., Brunekreef, B., 2007. Modeling the intra-urban variability of outdoor traffic pollution in Oslo, Norway—a GA2LEN project. *Atmos. Environ.* 41, 7500–7511. <http://dx.doi.org/10.1016/j.atmosenv.2007.05.039>.
- NAEI, 2015. UK Emissions Interactive Map [WWW Document]. Natl. Atmospheric Emiss. Inventory. <http://naei.defra.gov.uk/data/gis-mapping>.
- Met Office, 2012. Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853–current).
- Ordnance Survey, 2015. OS MasterMap Topography Layer [Shape Geospatial Data].
- R Core Team, 2015. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Safe Software Inc., S.S.I., 2015. FME.
- Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., Brunekreef, B., 2012. Systematic evaluation of land use regression models for NO₂. *Environ. Sci. Technol.* 46, 4481–4489. <http://dx.doi.org/10.1021/es204183v>.