Environment International 92–93 (2016) 202–209

Contents lists available at ScienceDirect

Environment International

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

Back-extrapolated and year-specific NO_2 land use regression models for Great Britain - Do they yield different exposure assessment?

John Gulliver [a,*], Kees de Hoogh [b,c], Gerard Hoek [d], Danielle Vienneau [b,c], Daniela Fecht [a], Anna Hansell [a]

[a] UK Small Area Health Statistics Unit (SAHSU), MRC-PHE Centre for Environment & Health, Imperial College London, Norfolk Place, W2 1PG London, UK
[b] Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland
[c] University of Basel, Petersplatz 1, 4003 Basel, Switzerland
[d] Institute of Risk Assessment Sciences, University of Utrecht, Yalelaan 2, 3584 CM Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 14 January 2016
Received in revised form 21 March 2016
Accepted 29 March 2016
Available online 22 April 2016

Keywords:
Air pollution modelling
Back-extrapolation
Exposure assessment
GIS
Land use regression
Nitrogen dioxide

ABSTRACT

Robust methods to estimate historic population air pollution exposures are important tools for epidemiological studies evaluating long-term health effects. We developed land use regression (LUR) models for NO_2 exposure in Great Britain for 1991 and explored whether the choice of year-specific or back-extrapolated LUR yields 1) similar LUR variables and model performance, and 2) similar national and regional address-level and small-area concentrations. We constructed two LUR models for 1991using NO_2 concentrations from the diffusion tube monitoring network, one using 75% of all available measurement sites (that over-represent industrial areas), and the other using 75% of a subset of sites proportionate to population by region to study the effects of monitoring site selection bias. We compared, using the remaining (hold-out) 25% of monitoring sites, the performance of the two 1991 models with back-extrapolation of a previously published 2009 model, developed using NO_2 concentrations from automatic chemiluminescence monitoring sites and predictor variables from 2006/2007. The 2009 model was back-extrapolated to 1991 using the same predictors (1990 & 1995) used to develop 1991 models. The 1991 models included industrial land use variables, not present for 2009. The hold-out performance of 1991 models (mean-squared-error-based-R^2: 0.62–0.64) was up to 8% higher and ~1 μg/m^3 lower in root mean squared error than the back-extrapolated 2009 model, with best performance from the subset of sites representing population exposures. Year-specific and back-extrapolated exposures for residential addresses ($n = 1.338,399$) and small areas ($n = 10.518$) were very highly linearly correlated for Great Britain ($r > 0.83$). This study suggests that year-specific model for 1991 and back-extrapolation of the 2009 LUR yield similar exposure assessment.

1. Introduction

Land use regression (LUR) modelling has been widely used to estimate exposures for a range of air pollution metrics (Adam-Poupart et al., 2014; Aguilera et al., 2008; Briggs et al., 2000; de Hoogh et al., 2014; Montagne et al., 2015; Yang et al., 2015) over different spatial and temporal scales (Hystad et al., 2011; Liu et al., 2012; Tang et al., 2013; Vienneau et al., 2013; Wang et al., 2014). One of the potential uses of LUR is historical exposure assessment to meet the needs of cohort studies with recruitment in the past and life-course epidemiology. Limiting factors in the application of LUR to earlier years are the availability of historical air pollution measurements (i.e. required to train and evaluate models) that adequately represent population exposures and, to a lesser extent, data on historical patterns of land use, transportation and industrial activity. Transferring current or recent LUR models to earlier years (i.e. back-extrapolation) has been shown to have potential in estimating air pollution exposures for historical periods (Chen et al., 2010; Cesaroni et al., 2012; Eeftens et al., 2011; Gulliver et al., 2013; Levy et al., 2015). The main premise for back-extrapolation is that air pollution sources are similar and the spatial structure of concentration surfaces remains stable over time.

NO_2 is one of the main pollutants of health concern, which can be viewed as a proxy for transport-related exposure (HEI, 2010). We previously evaluated back-extrapolation of LUR models with a spatial resolution of 200 m for Great Britain (GB) for the year 2009 to predict concentrations of NO_2 in 1991 at 451 diffusion tube monitoring sites (Gulliver et al., 2013). The 2009 models were developed using predictor variable data from that period (2006/2007). Back-extrapolation of the 2009 models was performed using predictor variables from nearest years (1990 & 1995) as possible to the target year (1991). We showed that back-extrapolation for the whole of GB was valid for up to

* Corresponding author at: UK Small Area Health Statistics Unit, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, St Mary's campus, Norfolk Place, London W2 1PG, UK.
E-mail address: j.gulliver@imperial.ac.uk (J. Gulliver).

http://dx.doi.org/10.1016/j.envint.2016.03.037
0160-4120/

18 years earlier (mean-squared-error-based-R^2 (MSE-R^2) ~0.55), but models performed poorly in some regions, especially the regions Midlands and North of England (North): MSE-R^2: 0.01–0.24.

In this paper, we use the 451 NO_2 monitoring sites in 1991 to produce GB LUR models for 1991 for comparison with back-extrapolated models. We used historical predictor variables from the nearest years (1990 & 1995), both to develop new models for 1991 and apply the back-extrapolated 2009 model. We addressed two specific questions to inform ongoing and future epidemiological studies:

1) How different are NO_2 LUR models developed specifically for 1991 in terms of variables and performance compared to an available model for 2009 back-extrapolated to 1991?
2) Does the choice of a year specific (1991) or back-extrapolated (2009) model yield different national and regional address-level and small-area exposure assessments, given known regional differences in air pollution levels and number of monitoring sites?

2. Materials and methods

2.1. Monitored concentrations of NO_2

We acquired data from 451 diffusion tube sites for 1991 and data from 187 routine monitoring sites that form part of the Automatic Urban and Rural Network (AURN) in the UK for 2009 as detailed in Gulliver et al. (2013). We chose 1991 for developing year-specific LUR models as that year had a relatively large number of monitoring sites where all diffusion tubes had been analysed at the same laboratory (Bush et al., 2001). Average measured concentrations were broadly similar in 1991 (39.33 μg/m^3) and 2009 (36.69 μg/m^3) but are not directly comparable due to different locations of monitoring sites and monitoring methods.

2.2. Development of variables

Data on monitored NO_2 concentrations for 1991 diffusion tube sites, site name, site type, and geographic coordinates (six-figure British National Grid), were integrated into the geographic information system ArcGIS version 10 (ESRI, Redlands, CA). Characteristics of 1991 sites differ compared to the 2009 Automated Urban and Rural Network (AURN). Firstly, in 1991 geographic coordinates were recorded with a ground precision of 100 m whereas for 2009 sites have a precision of 1 m. We therefore restricted the 1991 models to a minimum buffer distance of 200 m and mapping of concentrations for both the 1991 and back-extrapolated 2009 models to a 200 m grid. Secondly, the classification of site types differs between the two years (for 1991 there are 26 classes of site type related to land cover classes and emissions sources but no information on proximity to road sources whereas for 2009, site type is classified as: industrial, roadside, urban centre, urban background, suburban, or rural).

Table S1 in Supplementary material summarises the data and variables used in LUR models. Data on land cover, road geography and altitude were integrated into ArcGIS. Land cover variables were derived from a combination of CORINE land cover for Europe (CLC) and the Land Cover Map of Great Britain (LCM) as in Gulliver et al. (2013). Both CLC, comprising 44 categories of land cover on a 100 m grid, and LCM, comprising 25 categories of land cover on a 25 m grid are available for the year 1990. LCM contains only two classes of urban land (i.e. urban and suburban). Industry, ports, airports, construction sites, green urban areas and sport and leisure facilities from CLC were, therefore, substituted into the 25 m LCM where they were intersected with 'urban land'. This enhanced land cover data set was then aggregated into seven groups for use in modelling: high density urban, low density urban, industry and commercial, urban green areas, agriculture, semi-natural land, and water.

Data on road geography came from the Ordnance Survey via the agreement for UK educational establishments (www.digimap.co.uk). Meridian-1 from 1995 was applied to 1991 models. Roads are separated into four classes: motorways, A-roads (major trunk roads, dual carriageways and arterial roads), B-roads (roads with significant traffic flows but not in 'A' class), and minor roads (urban side-roads and country lanes). Data on altitude were obtained from the PANORAMA™ digital terrain model from Ordnance Survey via Digimap.

Each dataset was, in turn, intersected with a common 25 m grid covering the whole of GB and summed within each grid cell. "Buffering", using the ArcGIS FOCALSUM tool with the circle option (subsequently referred to as buffers), was used to sum the contribution (i.e., area or length) of each land cover and road variable around each monitoring site. For land cover, buffers were created for 0.2, 0.3, 0.4, 0.5, 0.75, 1, 2, 3, 5, 10 and 20 km. For road length 0.2, 0.3, 0.4, 0.5, 0.75 and 1 km buffers were used. Both zero-centred and ring buffers were constructed. Altitude (and log-transformed altitude) was obtained for each monitoring site using a point-in-grid function. Coordinates of each monitoring site were recorded for applying either second or third order trend surfaces.

Monitoring sites were pooled into model 'training' and 'evaluation' (i.e. hold-out validation (HOV)) groups, using a 75:25 split. To control for potential geographical bias in this selection process, monitoring sites were stratified by five regions (South East, South West & Wales, Midlands, North, and Scotland) and site type. The balance of site types by region was checked for the training and evaluation groups and the LUR variables attributed to the monitoring sites were then imported into SPSS (version 20) for development of the regression models.

2.3. Modelling strategy for year-specific (1991) models

Historic monitoring networks were designed to target areas of industrial pollution in order to control air quality at source, hence 49% (221) of the 451 NO_2 monitoring sites in 1991 were located in the North region, with a concentration of heavy industry but with only 26% of the population of Great Britain in 1991. In order to study the potential effects of monitoring site selection on exposure estimates, we developed two NO_2 LUR models for 1991: 1) using all 451 available monitoring sites (model 1991A), and 2) using a reduced set of 186 monitoring sites (model 1991B) proportionate to the population in each region, relative to the region with the highest population (South East). Models were developed using a supervised forward approach that we, and others (Beelen et al., 2010; Gulliver et al., 2013; Vienneau et al., 2013), have previously used. In this study we included a rule that each of the main sources of NO_2 variability must be represented in the model (i.e. road traffic and at least one class of urban land) based on information on national source emissions contributions (Carslaw et al., 2011).

Following the supervised forward approach, LUR models were developed such that: 1) each variable in the final model had a significant correlation with the monitored concentration ($p < 0.05$), 2) the direction of effect met predetermined expectations, 3) the direction of effect of predictors already in the model did not change as subsequent predictors were added, and 4) variables adding < 1% to the explained variation in monitored concentrations were excluded.

2.4. Back-extrapolation

We used the best performing 2009 model on HOV samples of monitoring sites as described in Gulliver et al. (2013) to back-extrapolate NO_2 concentrations to 1991 for comparison with year-specific models. We applied the 2009 model using the relevant predictor variables from the same data sets (i.e. 1990 & 1995) used to develop year-specific models. Using annual average NO_2 concentrations from all five available concurrent background monitoring sites for 1991 and 2009 from the AURN, we calculated the average difference in concentrations of NO_2 between 1991 and 2009. We then added the resulting average difference of 8.2 μg/m^3 to all predictions from the 2009 model (i.e. to back-extrapolate for the 451 monitoring sites to 1991); a method previously described in Gulliver et al. (2013).

2.5. Model evaluation

Both year-specific and back-extrapolated model errors were checked for non-normality in SPSS and spatial auto-correlation using Moran's I in ArcGIS. Models were summarised by their regression equation, p-values (Sig.), adjusted coefficient of variation (Adj. R^2) and standard error of the estimate (SEE). The two 1991 models were evaluated against the HOV samples in terms of R^2, MSE-R^2, RMSE, mean bias, variance of the measurements and predictions, regression fit line (constant, slope) and confidence intervals. Using the same measures we also compared the performance of the back-extrapolated 2009 model to predict NO_2 concentrations for the 1991 model HOV samples.

2.6. Exposure assessment

NO_2 concentrations were firstly mapped to a 200 m grid across GB using each of the 1991 models and the back-extrapolated 2009 model. Postcode point locations (ground-precision of 1 m) with attached population numbers of all ca.1.34 million residential postcodes (2001) in the UK were intersected with each NO_2 grid surface. A postcode location is the nearest address location to the geometric centroid of all address locations associated with each postcode (i.e. on average 19 dwellings per postcode). Thus, all residential postcodes for the UK represent a 1.3 million sample of ca. 29 million address locations. To assess small-area exposures we aggregated NO_2 concentrations from postcodes to UK Census Wards (n = 10.518). Wards on average have an area of 20.2km^2 and are generally smaller in urban areas (mean = 4.7 km^2) and larger in rural areas (mean = 31.9 km^2). NO_2 values were population weighted to Ward level to obtain NO_2 exposures estimates for all Wards in GB. The same postcodes (2001) and Ward boundaries (1991) were used for all population-weighted exposures.

2.7. Statistical analysis

We compared exposures at postcode and Ward level from 1991 models and the back-extrapolated 2009 model in terms of correlation. We used Spearman's rho. Correlation, subsequently referred to as 'r', because we are interested in the relative ranking of exposures across the population appropriate in an epidemiological context. For the purpose of this assessment we defined correlations as very low (≥0.0–0.2), low (>0.2–0.4), moderate (>0.4–0.6), high (>0.6–0.8), and very high (>0.8–1.0). To study the linearity and slope of the relationships between the models we used locally-weighted scatterplot smoothing (LOWESS). We also mapped and quantified the percentage and absolute differences in population-weighted exposures between the LUR models at Ward level across GB.

3. Results

3.1. Model derivation

The two LUR models for 1991 and the 2009 model are summarised in Table 1. All three LUR models include information on low-density urban land within a 20 km buffer, major roads within 0.2 km (1991 models) and 0.3 km (2009 model) buffers and high-density urban land in different sized buffers. The 1991 model using all 451 sites (1991 A) includes information on length of minor roads; these are not included in the 1991 model developed using 186 sites proportionate to population (1991B) or the 2009 model. Information on industrial land is included in models 1991 A and 1991B in 4 km and 2 km buffers, respectively, but is not included in the 2009 model. Information on semi-natural land is included in models 1991B and 2009 in 2 km and 0.2 km buffers, respectively, but is not included in model 1991A. Coefficients for the predictors with the same buffer size were similar across models. Overall, we observed a higher total R^2 (R^2 = 0.71) for the 1991B model than the 1991A model (R^2 = 0.64) but both had similar values of SEE (1991A: SEE = 7.89 μg/m^3; 1991B: SEE = 8.46 μg/m^3). In all models, the variance inflation factor (VIF) indicated low levels of multi-collinearity (i.e. VIF < 3); model 1991A (Moran's I = −0.08; p = 0.88) and 1991B (Moran's I = 0.11; p = 0.12) showed a low level of residual clustering.

3.2. Model evaluation

Overall, we observed for both 1991 models good agreement between measured and predicted concentrations (Supplementary material, Fig. S1). Table 2 shows performance statistics from both 1991 models and the back-extrapolated 2009 model for the HOV samples. For both 1991 models values of R^2 are reduced by about 10% compared to model derivation (Table 1). Both 1991 models also showed better performance than the back-extrapolated 2009 model. Model 1991B had the best performance statistics, explaining 8% more of the variability in measured NO_2 concentrations (MSE-R^2 = 0.62) than the back-extrapolated 2009 model (MSE-R^2 = 0.54) on the equivalent HOV sample (Table 2). All models tend to slightly under-predict (mean bias) measured concentrations of NO_2. The 2009 model does, however,

Table 1
Summary of 1991 and 2009 NO_2 LUR (μg/m^3) models.

Model[a]	N	Variable	Buffer (km)	Constant	β	Incremental adjusted R^2	SEE	Sig.	VIF
1991A	341			17.93					
		Low density urban land	20		5.070×10^{-4}	0.477	9.56	0.000	1.73
		High density urban land	3		7.321×10^{-3}	0.571	8.66	0.000	2.38
		Length of major roads	0.2		1.012×10^{-2}	0.603	8.32	0.000	1.07
		Length of minor roads	1		2.080×10^{-4}	0.626	8.09	0.000	1.50
		Industrial land	4		1.123×10^{-2}	0.644	7.89	0.000	1.05
1991B	144			26.63					
		Low density urban land	20		4.930×10^{-4}	0.551	10.54	0.000	1.61
		High density urban land	1		5.122×10^{-2}	0.665	9.10	0.000	2.12
		Length of major roads	0.2		1.051×10^{-2}	0.686	8.81	0.000	1.20
		Industrial land	5		8.468×10^{-3}	0.701	8.60	0.010	1.09
		Semi-natural land	2		-1.092×10^{-2}	0.711	8.46	0.017	1.85
2009	140			18.23					
		Low density urban land	20		4.820×10^{-4}	0.367	13.78	0.000	1.14
		High density urban land	0.2		8.780×10^{-2}	0.501	12.22	0.001	1.44
		Length of major roads	0.3		7.777×10^{-3}	0.561	11.47	0.000	1.29
		Semi-natural land	0.2		−1.297	0.588	11.11	0.002	1.39

[a] 1991A: using all available monitoring stations; 1991B: using monitoring stations selected proportionate to population by region relative to the region with the highest population (South East); 2009: 2009 model back-extrapolated to 1991 (Gulliver et al., 2013).

Table 2
Performance statistics from model evaluation analysis.

Model	N	R^2	MSE-R^2	RMSE	Constant	β	Mean bias	Var_O[d]	Var_P[e]	95% CI (lower, upper)
1991A	110	0.56	0.55	10.35	1.08	−0.93	1.85	236.9	115.2	0.89, 1.26
1991B	42	0.62	0.62	11.27	1.08	−1.52	1.88	332.7	177.0	0.81, 1.35
2009A[a]	110	0.52	0.51	10.74	1.07	−1.57	0.97	236.9	107.1	0.87, 1.27
2009B[b]	42	0.56	0.54	12.38	1.03	1.88	3.30	332.7	174.9	0.74, 1.32
1991A–B[c]	42	0.66	0.64	10.99	1.08	−0.59	2.73	332.7	186.0	0.83, 1.33

[a] 2009 predicting NO_2 concentrations for the 1991A hold-out validation (HOV) sample.
[b] 2009 predicting NO_2 concentrations for 1991B HOV sample.
[c] 1991A applied to the 1991B HOV sample.
[d] Variance of measured NO_2 concentrations.
[e] Variance of predicted NO_2 concentration. NO_2 in μg/m^3.

provide overall a low level of mean bias in predicting 1991 monitored NO_2 concentrations, also shown by values of β and constant (i.e. regression fit line) in Table 2. Model 1991B (MSE-R^2 = 0.62; n = 42) performed better than Model 1991 A (MSE-R^2 = 0.55; n = 110) in HOV, but the results are not directly comparable due model 1991 A using a larger number of HOV sites. The lower performance of model 1991A is related to the high proportion of monitoring sites from the North, a region where we previously reported poor performance of back-extrapolation (Gulliver et al., 2013). When we assessed the performance of model 1991A on the same HOV sites used for evaluating model 1991B (i.e. with HOV sites selected proportionate to population), the performance of model 1991A (MSE-R^2 = 0.64; RMSE = 10.99 μg/m^3) was similar to model 1991B (MSE-R^2 = 0.62; RMSE = 11.27 μg/m^3) and 10% higher than back-extrapolation.

3.3. Exposure assessment

Fig. 1 shows concentration surfaces for models 1991A, 1991B and back-extrapolation of the 2009 model. Overall, spatial patterns of NO_2 concentrations across GB were similar between the three models. For Ward-level population-weighted exposures, we saw a very high level of correlation (r = 0.93; p = 0.00) between the two 1991 models (see Supplementary material, Fig. S2). We, therefore, only used model 1991B in further comparisons with the back-extrapolated 2009 model as they are based on monitoring sites selected proportionate to the population by region.

Table 3 shows Spearman's correlations between model 1991B and the back-extrapolated 2009 model for ca. 1.3 million postcode locations for GB and each of the five regions. Correlations are overall very high (r = 0.83) and high to very high for the five regions (0.73–0.89). Fig. 2 shows scatterplots of Ward-level population-weighted exposures, for GB and each of the five regions, comparing model 1991B against the back-extrapolated 2009 model with fitted LOWESS lines and correlations. We observed very high correlations (r = 0.85) of NO_2 exposures between the year-specific and back-extrapolated models for GB and each region, varying from r = 0.80 in Wales & South West to r = 0.93 in Scotland. LOWESS lines in Fig. 2 generally show linear relationships in exposure estimates between the two models, with the exception of a curvilinear relationship for Scotland which has the highest correlation of all regions.

Mapped absolute and percentage differences in exposures between model 1991B and back-extrapolation are shown in Fig. 3 and are summarised in Table 4. Full descriptive statistics of exposures from both models are shown in Table S2. We only saw small changes in exposures when using back-extrapolation rather than the year-specific 1991 model. As Table 4 shows, 90% and 62% of the population had changes in NO_2 exposures ≤20% and ≤10%, respectively. Absolute differences between 1991B and 2009 models (Table 4) for the majority of the population (82%) were also relatively small (<5 μg/m^3). As shown in Fig. 3, the

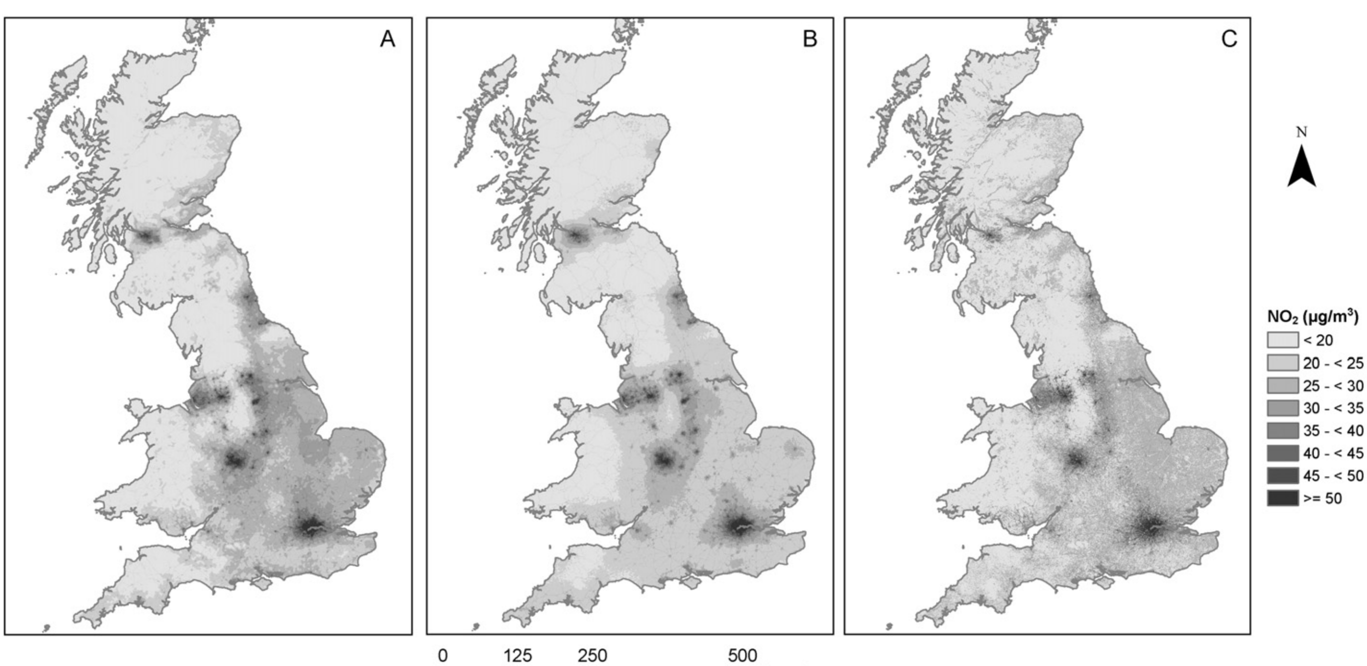

Fig. 1. Modelled NO_2 concentration surfaces (on a 200 m grid) from LUR for A) model 1991A, B) model 1991B and C) back-extrapolated 2009 model.

Table 3
Correlations (Spearman's rho. (r); two-tailed; $p < 0.01$) between NO_2 exposures from model 1991B and the back-extrapolated 2009 model for postcodes in GB and the five regions.

	Great Britain	Scotland	North	Midlands	Wales & South West	South East
N	1,338,399	128,931	335,478	202,679	214,725	456,586
% total[a]	100	9.6	25.1	15.1	16.0	34.1
r	0.83	0.81	0.75	0.85	0.73	0.89

[a] Percentage values for regions sum to 99.9 due to rounding.

largest differences in exposures between the two models are generally in rural areas which are more marked in terms of percentage differences. In subtracting estimates for model 2009 from model 1991B by Ward, back-extrapolation produced higher concentrations of NO_2 than model 1991B in some rural areas (i.e. negative values in Fig. 3), especially in the far south of England, southern, and eastern Scotland, and lower NO_2 concentrations in other rural areas, especially in central and eastern England and the west of Scotland (i.e. positive values in Fig. 3).

4. Discussion

4.1. Main findings

We developed and evaluated two LUR models for 1991: one model using all available NO_2 measurements sites, and the other with a reduced set of sites proportioned by population in each region. We found both models performed well when evaluated on the same HOV sample of NO_2 measurements with marginally better performance from the model (1991B) developed on the reduced set of sites for exposure assessment in this study. Models for 1991 performed slightly better than the back-extrapolated 2009 model but the 1991 models were given an advantage as the evaluation was done on a subset of the diffusion tube sites used to develop 1991 models. Our national and regional comparison of exposures for residential addresses and small areas from the 1991B model with those from back-extrapolation of the 2009 model generally showed a linear relationship with very high level of correlation. Absolute differences between the two models were relatively small across most of the population with larger differences generally confined to rural areas. To our knowledge this is the first study to compare exposures from back-extrapolation with those from a year-specific model at the national scale.

4.2. Performance of models

Taking the performance statistics for the 1991 models as the "gold standard" for 1991, the back-extrapolated 2009 model explained up to 8% less of the variability in monitored concentrations of NO_2 and increased the RMSE by up to ~1 μg/m^3. We consider the level of performance of the back-extrapolated 2009 model to be acceptable and within the range of performance statistics used in other epidemiological studies (Beelen et al., 2013; Carey et al., 2013).

We used monitoring sites with coordinates accurate to 100 m for the development of the 1991 models. We did not therefore use distance measures (e.g. distance to nearest major road) or use buffers <200 m to create variables. Although coordinates of 2009 monitoring sites were accurate to 1 m we restricted the development of the 2009 model in the same way as for 1991 models as the focus of its use was back-extrapolation (Gulliver et al., 2013). As a result, we have underestimated NO_2 concentrations at some monitoring sites, especially

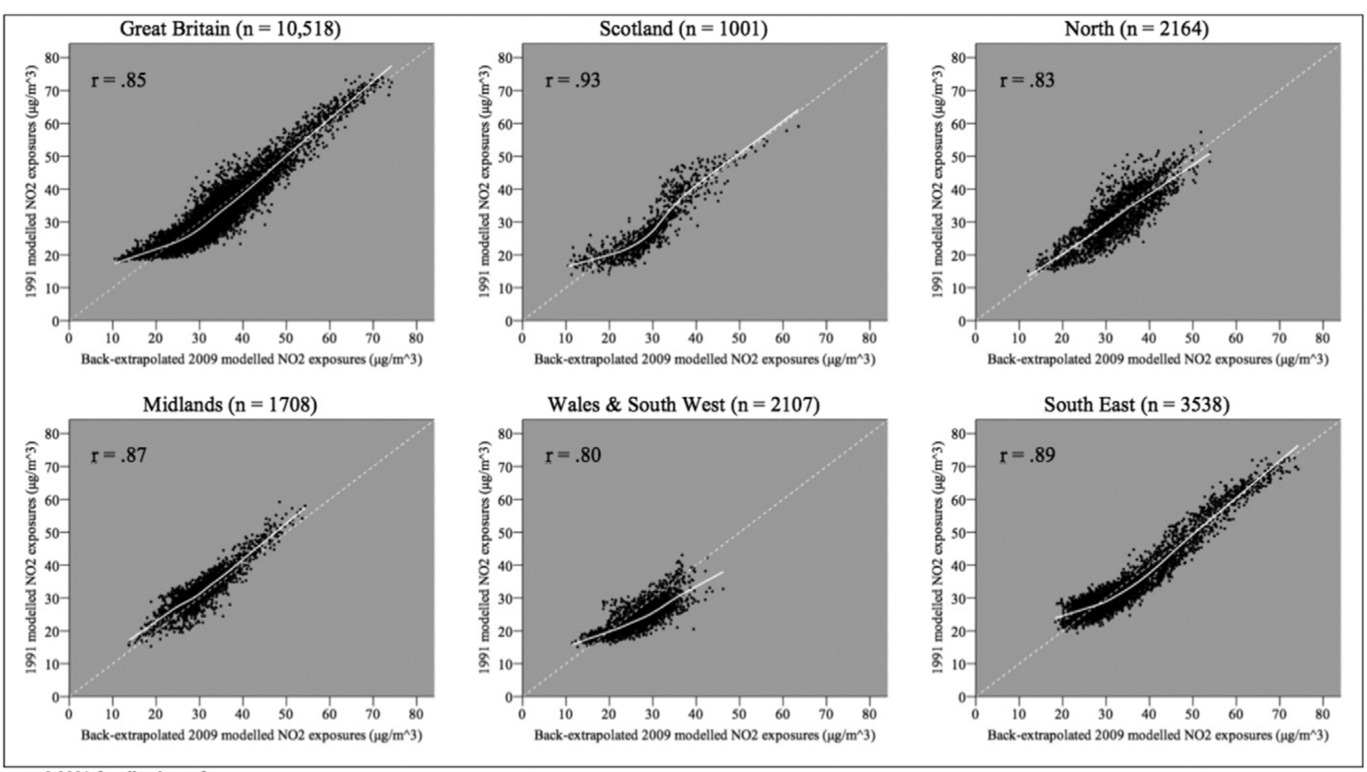

Fig. 2. Comparison of Ward-level population-weighted NO_2 exposures for GB and regions between model 1991B and the back-extrapolated 2009 model (dashed line is 1:1; solid line is LOWESS; $p < 0.0001$ for all values of r[a]).

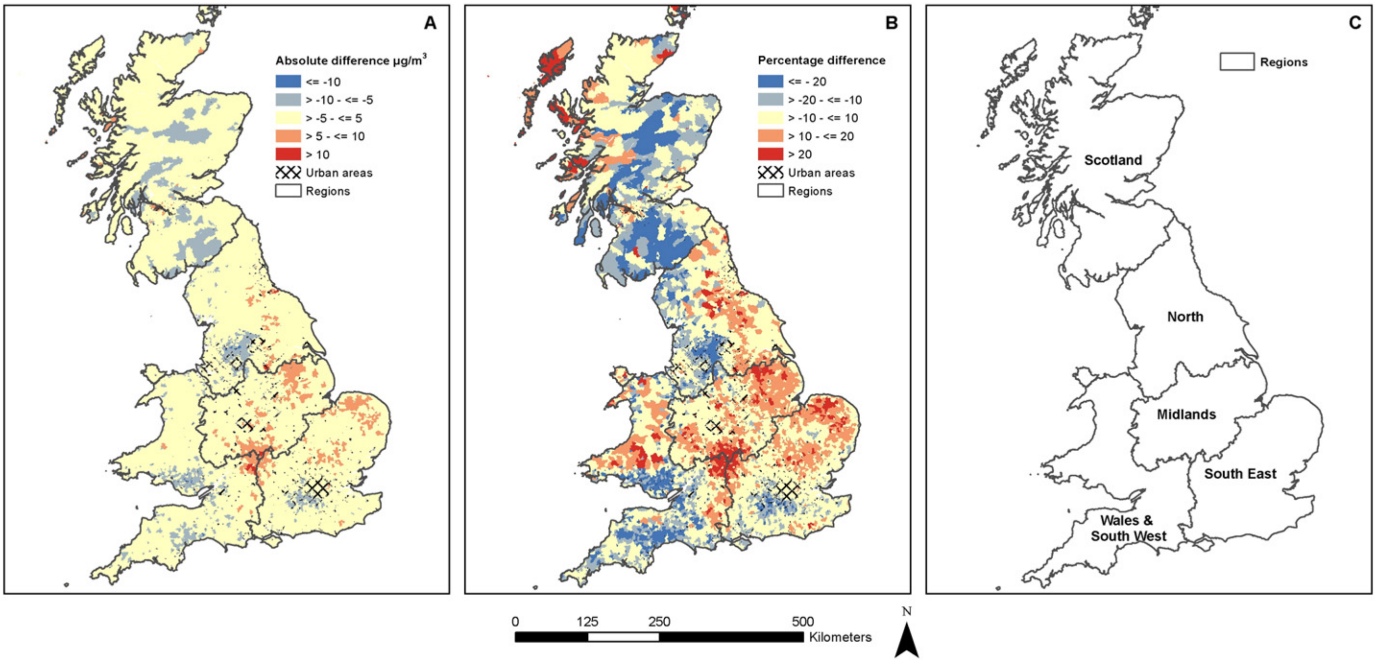

Fig. 3. Absolute (A) and percentage differences (B) in Ward-level population-weighted NO_2 exposures between model 1991B and the back-extrapolated 2009 model [negative values represent Wards where back-extrapolation produces higher exposures of NO_2 than model 1991B and vice versa]. Regions shown in C.

those close to a major source of NO_2 (e.g. major road). Indeed, all models under-predict the variance of measured NO_2 concentrations (Table 2). Nonetheless, both the 1991 models and back-extrapolated 2009 model provided good performance in HOV. Levels of NO_2 concentrations were predicted well by the back-extrapolated 2009 model compared with the 1991 models, which is remarkable given that there were only five monitoring sites available to produce data to perform back-extrapolation.

We used land cover data from 1990 and data on the road geography from 1995 to create variables for developing 1991 models and in back-extrapolation of the 2009 model to predict 1991 NO_2 concentrations. Some studies may not have access to land use data to create variables for earlier years. We have not made an assessment of the impacts of, alternatively, using recent land use data in back-extrapolation of LUR, but, as reported in an earlier study (Gulliver et al., 2013), land use patterns at the national scale are broadly similar over the time period of interest (18 years). We expect that using recent data on land use in back-extrapolation of LUR could have a large impact on local NO_2 estimates (i.e. due urban expansion and road building in the intervening years).

Weather patterns can have important impact on annual air pollution exposures, thus back-extrapolations may be affected by differences in weather between years considered. There was limited monitoring data on NO_2 concentrations other than in 1991 prior to the late 1990s so we were not able to make comparisons for other years.

We used a supervised forward approach in LUR model development whereas others have used the ADDRESS (A Distance Decay REgression Selection Strategy) (Su et al., 2009) and DSA (Deletion/Substitution/Addition) methods (Beckerman et al., 2013). We used a 25% HOV whereas others (Beelen et al., 2013; Meng et al., 2015) have tested model performance with leave-one-out cross validation (LOOCV). We used the rule of $p < 0.05$ for variable inclusion whereas others have used a less stringent $p < 0.1$ (Beelen et al., 2013). The modelling results and performance could differ if other methods of development and/or evaluation were employed. Basagaña et al. (2012), however, found that LUR models developed using the DSA and supervised methods for Girona, Spain, did not differ in HOV performance. LOOCV is popular with studies that have a very limited number of measurement sites but has been shown to overestimate the predictive power of models compared to HOV (Basagaña et al., 2012; Wang et al., 2012).

Table 4
The proportion of the population in categories of absolute and percentage differences in Ward-level population-weighted NO_2 exposures between model 1991B and the back-extrapolated 2009 model.

Absolute difference	% of the population	Percent difference	% of the population
≤−10	0.3	≤−20	9.1
>−10–≤−5	12.0	>−20–≤−10	17.5
>−5–≤5	82.1	>−10–≤10	62.0
>5–≤10	5.1	>10–≤20	10.1
>10	0.5	>20	1.3

4.3. Population exposure assessment

Some of the differences in exposures from the models are due to differences in input data, which reflect real changes in land use over the intervening years. Some differences are due to the inclusion of different variables, but not due to differences in magnitude of the constant (intercept) in the models. The constants for the 1991B and the 2009 models are 26.63 μg/m^3 and 26.23 μg/m^3, respectively, once 8.2 μg/m^3 has been added to the 2009 model for back-extrapolation (i.e. 18.23 μg/m^3 + 8.2 μg/m^3). Predicted exposures will be lower than the level of the model constants in areas where semi-natural land is present (i.e. included in both models 1991B and 2009). The largest exposure differences tended to be in rural areas with relatively low population, leading to a non-linear relationship in Scotland, a region with a relatively high proportion of rural land. These differences can be explained by the different buffer sizes and related coefficients for semi-natural land.

It is commonplace for LUR models to contain different variables reflecting the differences between study locations/periods in sources of air pollution and the effects of the built environment on dispersion patterns. In the European Study of Cohorts and Air Pollution Effects (ESCAPE) (Beelen et al., 2013), separate NO_2 models were developed for 36 locations with wide variation in the included set of variables and variations in buffer sizes where the same variables were used in different locations. Most models, however, included one or more variables on traffic (e.g. traffic intensity in a circular buffer; distance to nearest main road). The variables in our NO_2 models for 1991 and 2009 are similar. Both of the 1991 models and back-extrapolated 2009 model include information on low density urban land, high density urban land and length of major roads, reflecting the importance of urban areas and main roads on the spatial variability of NO_2 concentrations over a period extending back 18 years. Both 1991 models contain a variable on industrial land that is not included in the 2009 model. This can be related to a ~ 95% reduction in industrial emissions of nitrous oxide between 1990 and 2010 (Department of Energy and Climate Change, 2014). The low number of 2009 monitoring sites close to industrial land (e.g. 61% of monitoring sites have no industrial land within a 1 km buffer), however, probably explains the lack of a variable on industrial land in the 2009 model in addition to the global level of reduction in industrial emissions. Semi-natural land is included in 1991B and 2009 models but not in 1991A based on all sites. This can be explained by the high proportion of sites used to develop model 1991 A from urban and industrialised areas of the regions Midlands and North.

4.4. Back-extrapolation

Back-extrapolation of national models has been shown to perform well for The Netherlands 7 years earlier (Eeftens et al., 2011), Great Britain up to 18 years earlier (Gulliver et al., 2013) and Israel up to 35 years earlier (Levy et al., 2015), though in the latter study performance was assessed against historic emissions of NO_X rather than measured concentrations. Studies with a smaller geographical extent will often have a lower number of measurement sites to assess the performance of back-extrapolation. Back-extrapolation has, however, generally performed well at the urban scale. LUR model predictions for 2007 NO_2 monitoring stations in Rome (Cesaroni et al., 2012), for example, had strong linear correlation ($r = 0.83$) with measured NO_2 concentrations at the same monitoring locations ($n = 67$) 12 years earlier. Back-extrapolation of 2010 NO ($R^2 = 0.52$) and NO_2 ($R^2 = 0.63$) LUR models to 73 concurrent measurement sites for 2003 in Vancouver provided reasonable but reduced performance ($R^2 = 0.50$–0.55 for NO; $R^2 = 0.44$–0.49 for NO_2) (Wang et al., 2013). In terms of exposure assessment, Chen et al. (2010) found for Montreal, Canada, high correlation between NO_2 concentrations from 5000 random locations from a LUR surface for 2006 and from the same locations back-extrapolated to 1985 ($r = 0.70$) and 1996 ($r = 0.90$). Molnar et al. (2015) using dispersion modelling instead of LUR for 6563 cohort participant address locations in Gothenburg, Sweden, found that back-extrapolated (2009) and year-specific NO_X estimates were highly correlated 5–7 years earlier ($R^2 = 0.98$) and more weakly, but not poorly correlated 12 years earlier in 1997 ($R^2 = 0.68$) and 34 years earlier in 1975 ($R^2 = 0.60$).

Spatial-temporal LUR models, based on "mobile monitoring", where fixed-site measurements are taken at different times (days, seasons), theoretically could be back-extrapolated if temporal variables (e.g. temperature, relative humidity, air pressure) are also available for earlier years. The mobile monitoring approach has been used for routinely measured pollutants such as NO_2 and O_3 (Cavellin et al., 2016) and, especially, to develop LUR models for novel metrics such as ultra-fine particles (Hoek et al., 2011; Hankey and Marshall, 2015; Montagne et al., 2015; Weichenthal et al., 2016) and black carbon (Hankey and Marshall, 2015; Montagne et al., 2015) which are not commonly measured. Back-extrapolation of LUR models for novel metrics will be restricted, however, as they were seldom measured in the past at more than a few sites in most countries, if at all.

4.5. Application to epidemiologic studies

We developed 1991 NO_2 LUR models that performed well against HOV samples. The NO_2 surfaces that we produced can directly be used to provide exposure assessment for the early 1990s for epidemiological studies assessing long-term and life-time air pollution exposures. Previous studies have demonstrated that back-extrapolation is valid for exposure assessment based on evaluations against a limited number of measurement sites, and subsequently back-extrapolation has been applied in epidemiological studies (Slama et al., 2007; Pedersen et al., 2013). To supplement this evidence, this study suggests that back-extrapolation of NO_2 over an 18-year period in a country with only slowly evolving changes in lifestyle and land use yielded exposure estimates that were only slightly different than year-specific models, particularly for urban areas that contain most of the population. This study overall suggests that year-specific models for 1991 and back-extrapolation of the 2009 LUR yield similar exposure assessment. Gains from constructing year-specific LUR models (if historic monitoring and air pollution data are available) may therefore be small relative to efforts to source historical input information and derive new models.

Acknowledgements

The research was supported by funding from the European Community's Seventh Framework Program EXPOsOMICS study under grant agreement number FP7 308610. We thank Tony Bush at AEA Technology (UK) for providing data on monitored concentrations of nitrogen dioxide from the UK diffusion tube network. The work of the UK Small Area Health Statistics Unit is funded by Public Health England as part of the MRC-PHE Centre for Environment and Health, funded also by the UK Medical Research Council.

Appendix A. Supplementary data

Supporting material can be found in the electronic version of this paper including information on sources of data used in the study, further information on the development of the LUR models, and exposure distributions. Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.envint.2016.03.037.

References

Adam-Poupart, A., Brand, A., Fournier, M., Jerrett, M., Smargiassi, A., 2014. Spatiotemporal modeling of ozone levels in Quebec (Canada): a comparison of kriging, land-use regression (LUR), and combined Bayesian maximum entropy-LUR approaches. Environ. Health Perspect. 122 (9), 970–976.

Aguilera, I., Sunyer, J., Fernandez-Patier, R., Hoek, G., Aguirre-Alfaro, A., Meliefste, K., Bomboi-Mingarro, M.T., Nieuwenhuijsen, M.J., Herce-Garraleta, D., Brunekreef, B., 2008. Estimation of outdoor NO(x), NO(2), and BTEX exposure in a cohort of pregnant women using land use regression modeling. Environ. Sci. Technol. 42 (3), 815–821.

Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., et al., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. Atmos. Environ. 54, 634–642.

Beckerman, B.S., Jerrett, M., Martin, R.V., van Donkelaar, A., Ross, Z., Burnett, R.T., 2013. Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in California. Atmos. Environ. 77, 172–177.

Beelen, R., Voogt, M., Duyzer, J., Zanveld, P., Hoek, G., 2010. Comparison of the performances of land use regression modelling and dispersion modelling in estimating small-scale variations in long-term air pollution concentrations in a Dutch urban area. Atmos. Environ. 44 (36), 4614–4621.

Beelen, R., Hoek, G., Vienneau, D., et al., 2013. Development of NO2 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.

Briggs, D.J., de Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S., Smallbone, K., 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. Sci. Total Environ. 253 (1–3), 151–167.

Bush, T., Smith, S., Stevenson, K., Moorcroft, S., 2001. Validation of nitrogen dioxide diffusion tube methodology in the UK. Atmos. Environ. 35 (2), 289–296.

Carey, I.M., Atkinson, R.W., Kent, A.J., van Staa, T., Cook, D.G., Anderson, H.R., 2013. Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. Am. J. Respir. Crit. Care Med. 187 (11), 1226–1233.

Carslaw, D.C., Beevers, S.D., Westmoreland, E., Williams, M.L., Tate, J.E., Murrells, T., et al., **2011**. Trends in NO_X and NO_2 Emissions and Ambient Measurements in the UK. Department of Environment, Food and Rural Affairs.

Cavellin, L.D., Weichenthal, S., Tack, R., Ragettli, M.S., Smargiassi, A., Hatzopoulou, M., 2016. Investigating the use of portable air pollution sensors to capture the spatial variability of traffic-related air pollution. Environ. Sci. Technol. 50 (1), 313–320.

Chen, H., Goldberg, M.S., Crouse, D.L., Burnett, R.T., Jerrett, M., Villeneuve, P.J., et al., 2010. Back-extrapolation of estimates of exposure from current land-used regression models. Atmos. Environ. 44, 4346–4354.

Cesaroni, G., Porta, D., Badaloni, C., Stafoggia, M., Eeftens, M., Meliefste, K., Forastiere, F., 2012. Nitrogen dioxide levels estimated from land use regression models several years apart and association with mortality in a large cohort study. Environ. Health 11, 48.

de Hoogh, K., Korek, M., Vienneau, D., Keuken, M., Kukkonen, J., Nieuwenhuijsen, M.J., Badaloni, C., Beelen, R., Bolignano, A., Cesaroni, G., Pradas, M.C., Cyrys, J., Douros, J., Eeftens, M., Forastiere, F., Forsberg, B., Fuks, K., Gehring, U., Gryparis, A., Gulliver, J., Hansell, A.L., Hoffmann, B., Johansson, C., Jonkers, S., Kangas, L., Katsouyanni, K., Kunzli, N., Lanki, T., Memmesheimer, M., Moussiopoulos, N., Modig, L., Pershagen, G., Probst-Hensch, N., Schindler, C., Schikowski, T., Sugiri, D., Teixido, O., Tsai, M.Y., Yli-Tuomi, T., Brunekreef, B., Hoek, G., Bellander, T., 2014. Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. Environ. Int. 73, 382–392.

Department of Energy and Climate Change. 2014 UK Greenhouse has emissions-provision figures. 26th March 2015. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/416810/2014_stats_release.pdf (accessed: 27th November 2015).

Eeftens, M., Beelen, R., Fischer, P., Brunekreef, B., Meliefste, K., Hoek, G., 2011. Stability of measured and modelled spatial contrasts in NO(2) over time. Occup. Environ. Med. 68 (10), 765–770.

Gulliver, J., de Hoogh, K., Hansell, A., Vienneau, D., 2013. Development and back-extrapolation of NO2 land use regression models for historic exposure assessment in Great Britain. Environ. Sci. Technol. 47 (14), 7804–7811.

Hankey, S., Marshall, J.D., 2015. Land use regression models of on-road particulate air pollution (particle number, black carbon, PM2.5, particle size) using mobile monitoring. Environ. Sci. Technol. 49 (15), 9194–9202.

Hoek, G., Beelen, R., Kos, G., Dijkema, M., van der Zee, S.C., Fischer, P.H., Brunekreef, B., 2011. Land use regression model for ultrafine particles in Amsterdam. Environ. Sci. Technol. 45 (2), 622–628.

Health Effects Institute, 2010. A critical review of the literature on emissions, exposure and health effects. HEI special report 17.17. Traffic-Relat. Air Pollut.: Boston 2010http://pubs.healtheffects.org/view.php?id=334.

Hystad, P., Setton, E., Cervantes, A., Poplawski, K., Deschenes, S., Brauer, M., van Donkelaar, A., Lamsal, L., Martin, R., Jerrett, M., Demers, P., 2011. Creating national air pollution models for population exposure assessment in Canada. Environ. Health Perspect. 119 (8), 1123–1129.

Levy, I., Levin, N., Yuval, Schwartz, J.D., Kark, J.D., 2015. Back-extrapolating a land use regression model for estimating past exposures to traffic-related air pollution. Environ. Sci. Technol. 49 (6), 3603–3610.

Liu, L.J.S., Tsai, M.Y., Keidel, D., Gemperli, A., Ineichen, A., Hazenkamp-von Arx, M., et al., 2012. Long-term exposure models for traffic related NO2 across geographically diverse areas over separate years. Atmos. Environ 46, 460–471.

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

Molnar, P., Stockfelt, L., Barregard, L., Sallsten, G., 2015. Residential NOx exposure in a 35-year cohort study. Changes of exposure, and comparison with back extrapolation for historical exposure assessment. Atmos. Environ. 115, 62–69.

Montagne, D.R., Hoek, G., Klompmaker, J.O., Wang, M., Meliefste, K., Brunekreef, B., 2015. Land use regression models for ultrafine particles and black carbon based on short-term monitoring predict past spatial variation. Environ. Sci. Technol. 49 (14), 8712–8720.

Pedersen, M., Giorgis-Allemand, L., Bernard, C., Aguilera, I., Andersen, A.M., Ballester, F., Beelen, R.M., Chatzi, L., Cirach, M., Danileviciute, A., Dedele, A., Eijsden, M., Estarlich, M., Fernandez-Somoano, A., Fernandez, M.F., Forastiere, F., Gehring, U., Grazuleviciene, R., Gruzieva, O., Heude, B., Hoek, G., de Hoogh, K., van den Hooven, E.H., Haberg, S.E., Jaddoe, V.W., Klumper, C., Korek, M., Kramer, U., Lerchundi, A., Lepeule, J., Nafstad, P., Nystad, W., Patelarou, E., Porta, D., Postma, D., Raaschou-Nielsen, O., Rudnai, P., Sunyer, J., Stephanou, E., Sorensen, M., Thiering, E., Tuffnell, D., Varro, M.J., Vrijkotte, T.G., Wijga, A., Wilhelm, M., Wright, J., Nieuwenhuijsen, M.J., Pershagen, G., Brunekreef, B., Kogevinas, M., Slama, R., 2013. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). Lancet Respir. Med. 1 (9), 695–704.

Slama, R., Morgenstern, V., Cyrys, J., Zutavern, A., Herbarth, O., Wichmann, H.E., Heinrich, J., Group, L.S., 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 (9), 1283–1292.

Su, J.G., Jerrett, M., Beckerman, B., Wilhelm, M., Ghosh, J.K., Ritz, B., 2009. Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. Environ. Res. 109 (6), 657–670.

Tang, R., Blangiardo, M., Gulliver, J., 2013. Using building heights and street configuration to enhance intraurban PM10, NO(X), and NO2 land use regression models. Environ. Sci. Technol. 47 (20), 11643–11650.

Vienneau, D., de Hoogh, K., Bechle, M.J., Beelen, R., van Donkelaar, A., Martin, R.V., Millet, D.B., Hoek, G., Marshall, J.D., 2013. Western European land use regression incorporating satellite- and ground-based measurements of NO2 and PM10. Environ. Sci. Technol. 47 (23), 13555–13564.

Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., Brunekreef, B., 2012. Systematic evaluation of land use regression models for NO(2). Environ. Sci. Technol. 46 (8), 4481–4489.

Wang, M., Beelen, R., Bellander, T., Birk, M., Cesaroni, G., Cirach, M., Cyrys, J., de Hoogh, K., Declercq, C., Dimakopoulou, K., Eeftens, M., Eriksen, K.T., Forastiere, F., Galassi, C., Grivas, G., Heinrich, J., Hoffmann, B., Ineichen, A., Korek, M., Lanki, T., Lindley, S., Modig, L., Molter, A., Nafstad, P., Nieuwenhuijsen, M.J., Nystad, W., Olsson, D., Raaschou-Nielsen, O., Ragettli, M., Ranzi, A., Stempfelet, M., Sugiri, D., Tsai, M.Y., Udvardy, O., Varro, M.J., Vienneau, D., Weinmayr, G., Wolf, K., Yli-Tuomi, T., Hoek, G., Brunekreef, B., 2014. Performance of multi-city land use regression models for nitrogen dioxide and fine particles. Environ. Health Perspect. 122 (8), 843–849.

Wang, R., Henderson, S.B., Sbihi, H., Allen, R.W., Brauer, M., 2013. Temporal stability of land use regression models for traffic-related air pollution. Atmos. Environ. 64, 312–319.

Weichenthal, S., Van Ryswyk, K., Goldstein, A., Shekarrizfard, M., Hatzopoulou, M., 2016. Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: a land use regression model. Environ. Pollut. 208 (Pt A), 241–248.

Yang, A., Hoek, G., Montagne, D., Leseman, D.L., Hellack, B., Kuhlbusch, T.A., Cassee, F.R., Brunekreef, B., Janssen, N.A., 2015. Agreement of central site measurements and land use regression modeled oxidative potential of PM2.5 with personal exposure. Environ. Res. 140, 397–404.