Atmospheric Environment 72 (2013) 10–23

Contents lists available at SciVerse ScienceDirect

Atmospheric Environment

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

Development of NO_2 and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project

Rob Beelen [a,*], Gerard Hoek [a], Danielle Vienneau [b], Marloes Eeftens [a], Konstantina Dimakopoulou [c], Xanthi Pedeli [c], Ming-Yi Tsai [d,e,f], Nino Künzli [d,e], Tamara Schikowski [d], Alessandro Marcon [g], Kirsten T. Eriksen [h], Ole Raaschou-Nielsen [h], Euripides Stephanou [i], Evridiki Patelarou [i], Timo Lanki [j], Tarja Yli-Tuomi [j], Christophe Declercq [k], Grégoire Falq [k], Morgane Stempfelet [k], Matthias Birk [l], Josef Cyrys [l,m], Stephanie von Klot [l], Gizella Nádor [n], Mihály János Varró [n], Audrius Dėdelė [o], Regina Gražulevičienė [o], Anna Mölter [p], Sarah Lindley [q], Christian Madsen [r], Giulia Cesaroni [s], Andrea Ranzi [t], Chiara Badaloni [s], Barbara Hoffmann [u], Michael Nonnemacher [v], Ursula Krämer [u], Thomas Kuhlbusch [w], Marta Cirach [x,y,z], Audrey de Nazelle [x,y,z], Mark Nieuwenhuijsen [x,y,z], Tom Bellander [aa], Michal Korek [aa], David Olsson [bb], Magnus Strömgren [cc], Evi Dons [dd,ee], Michael Jerrett [ff], Paul Fischer [gg], Meng Wang [a], Bert Brunekreef [a,hh], Kees de Hoogh [b]

[a] Institute for Risk Assessment Sciences (IRAS), Utrecht University, P.O. Box 80178, 3508 TD Utrecht, The Netherlands
[b] MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom
[c] Department of Hygiene, Epidemiology & Medical Statistics, Medical School, National and Kapodistrian University of Athens, Medical School, Athens, Greece
[d] Swiss Tropical & Public Health Institute, Basel, Switzerland
[e] University of Basel, Basel, Switzerland
[f] Department of Environmental & Occupational Health Sciences, University of Washington, Seattle, USA
[g] Unit of Epidemiology & Medical Statistics, Department of Public Health and Community Medicine, University of Verona, Italy
[h] Danish Cancer Society, Copenhagen, Denmark
[i] Environmental Chemical Processes Laboratory, University of Crete, Heraklion, Greece
[j] Department of Environmental Health, National Institute for Health and Welfare, Kuopio, Finland
[k] French Institute for Public Health Surveillance, Saint-Maurice, France
[l] Helmholtz Zentrum München, German Research Center for Environmental Health, Institutes of Epidemiology I and II, Neuherberg, Germany
[m] University of Augsburg, Environmental Science Center, Augsburg, Germany
[n] Department of Environmental Epidemiology, National Institute of Environmental Health, Budapest, Hungary
[o] Vytautas Magnus University, Kaunas, Lithuania
[p] Centre for Occupational and Environmental Health, The University of Manchester, Manchester, England
[q] School of Environment and Development (Geography), The University of Manchester, Manchester, England
[r] Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway
[s] Epidemiology Department, Lazio Regional Health Service, Rome, Italy
[t] Regional Reference Centre on Environment and Health, ARPA Emilia Romagna, Modena, Italy
[u] IUF Leibniz Research Institute for Environmental Medicine, University of Düsseldorf, Düsseldorf, Germany
[v] Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany
[w] Air Quality & Sustainable Nanotechnology, Institute for Energy and Environmental Technology (IUTA) e.V., Duisburg, Germany
[x] Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
[y] IMIM (Hospital del Mar Research Institute), Barcelona, Spain
[z] CIBER Epidemiología y Salud Pública (CIBERESP), Spain
[aa] Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
[bb] Division of Occupational and Environmental Medicine, Department of Public Health and Clinical Medicine, Umeå University, Sweden
[cc] Department of Geography and Economic History, Umeå University, Sweden

Abbreviations: CORINE, COordination and INformation on the Environmental programme; CV, Leave-one-out cross-validation; ESCAPE, European Study of Cohorts for Air Pollution Effects; GIS, Geographic Information Systems; GPS, Global Positioning System; LUR, Land Use Regression; mvh, Motor vehicles; NO_2, Nitrogen dioxide ($\mu g\ m^{-3}$); NO_x, Nitrogen oxides ($\mu g\ m^{-3}$); R^2, Explained variance; RMSE, Root Mean Squared Error.

* Corresponding author. Tel.: +31 30 253 20 59; fax: +31 30 253 94 99.

1352-2310/$ – see front matter
http://dx.doi.org/10.1016/j.atmosenv.2013.02.037

[dd] Environmental Risk and Health Unit, VITO-MRG (Flemish Institute for Technological Research), Mol, Belgium
[ee] Hasselt University, Diepenbeek, Belgium
[ff] School of Public Health, University of California, Berkeley, USA
[gg] Centre for Environmental Health, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
[hh] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

H I G H L I G H T S

► LUR models were developed in 36 study areas in Europe using a standardized approach.
► NO_2 models explained a large fraction of concentration variability (median R^2 82%).
► Local traffic intensity data were important predictors for LUR model development.

A R T I C L E I N F O

Article history:
Received 20 September 2012
Received in revised form
10 January 2013
Accepted 19 February 2013

Keywords:
ESCAPE
Air pollution
NO_2
NO_x
Land Use Regression (LUR) model

A B S T R A C T

Estimating within-city variability in air pollution concentrations is important. Land use regression (LUR) models are able to explain such small-scale within-city variations. Transparency in LUR model development methods is important to facilitate comparison of methods between different studies. We therefore developed LUR models in a standardized way in 36 study areas in Europe for the ESCAPE (European Study of Cohorts for Air Pollution Effects) project.

Nitrogen dioxide (NO_2) and nitrogen oxides (NO_x) were measured with Ogawa passive samplers at 40 or 80 sites in each of the 36 study areas. The spatial variation in each area was explained by LUR modelling. Centrally and locally available Geographic Information System (GIS) variables were used as potential predictors. A leave-one out cross-validation procedure was used to evaluate the model performance.

There was substantial contrast in annual average NO_2 and NO_x concentrations within the study areas. The model explained variances (R^2) of the LUR models ranged from 55% to 92% (median 82%) for NO_2 and from 49% to 91% (median 78%) for NO_x. For most areas the cross-validation R^2 was less than 10% lower than the model R^2. Small-scale traffic and population/household density were the most common predictors. The magnitude of the explained variance depended on the contrast in measured concentrations as well as availability of GIS predictors, especially traffic intensity data were important. In an additional evaluation, models in which local traffic intensity was not offered had 10% lower R^2 compared to models in the same areas in which these variables were offered.

Within the ESCAPE project it was possible to develop LUR models that explained a large fraction of the spatial variance in measured annual average NO_2 and NO_x concentrations. These LUR models are being used to estimate outdoor concentrations at the home addresses of participants in over 30 cohort studies.

1. Introduction

Many epidemiological studies have suggested that traffic-related air pollution contributes to health effects associated with long-term exposure to air pollution. Current estimates of the European health impact of air pollution are large (Brunekreef and Holgate, 2002). These estimates are, however, primarily based on exposure response relationships established in studies in North America (especially Pope et al., 2002). There is therefore an urgent need to perform long-term air pollution exposure and health effect studies in Europe. The European Study of Cohorts for Air Pollution Effects (ESCAPE) project was designed to help to fill this gap.

Recent epidemiological research demonstrated the importance of accounting for within-city variability in estimating air pollution concentrations (Jerrett et al., 2005; Hoek et al., 2008; Beelen et al., 2008; Brauer et al., 2003). Several methods can explain such small-scale within-city variations such as geostatistical interpolation, dispersion models, and Land Use Regression (LUR) models. Geostatistical interpolation of monitored concentrations is problematic whenever networks are not dense enough, and therefore fail to capture variability of concentrations over short distances. Dispersion models depend on detailed and spatially resolved input data if they are to capture small-scale spatial variations in air pollutants adequately. LUR modelling uses multiple linear regression to analyse associations between measured pollutant concentrations at a number of monitoring sites and predictor variables such as traffic, land use and topography. LUR models have been shown to be a cost-effective method to explain the spatial variation in air pollution in a number of studies (Hoek et al., 2008; Marshall et al., 2008).

Within the ESCAPE project LUR models for 36 study areas have been developed to estimate outdoor pollutant concentrations at the home addresses of participants in a large number of cohort studies conducted all over Europe. In this paper, we describe the standardized approach we used to develop these models for nitrogen dioxide (NO_2) and nitrogen oxides (NO_x), and we discuss the performance of the models in terms of explained variance and cross-validation. We also discuss some of the methodological issues occurring in LUR model development. LUR models for particulate matter have been published elsewhere (Eeftens et al., 2012a).

2. Methods

Annual average NO_2 and NO_x concentrations from an intensive monitoring campaign and predictor variables were used to develop LUR models in each ESCAPE study area. A standardized approach described in the ESCAPE exposure manual (available on www.escapeproject.eu) was used to develop LUR models in all areas. Predictor variables were derived from Europe-wide and local Geographic Information System (GIS) databases. European-wide GIS data were centrally obtained to facilitate consistency between study areas. In addition, local GIS data were collected in each study area as not all potential predictor variables were centrally available.

Local GIS data might also be more accurate than the central GIS data. Common criteria were followed by each center for the collection of local GIS data. Both the central and local GIS predictor variables were included in exposure model development. Models were optimised locally with no attempt to force a common model to all study areas. This decision was based on the diversity of study areas and differences in available GIS predictor variables.

2.1. Air pollution measurement data

The ESCAPE air pollution measurements have been described previously (Eeftens et al., 2012b; Cyrys et al., 2012). Briefly, measurements of both particulate matter ($PM_{2.5}$, $PM_{2.5}$ absorbance and PM_{10}) and nitrogen oxides (NO_2 and NO_x) were conducted in 20 European study areas at 20–40 sites for PM and at 40–80 sites for NO_x per area. In an additional 16 areas only NO_x measurements were conducted. So in total, NO_x was measured in 36 areas (Fig. 1). The Barcelona measurement campaign ($N = 40$ NO_x sites) was also part of the Catalunya measurement campaign ($N = 80$ NO_x sites). A detailed description of each study area is given in the online supplement of Cyrys et al. (2012). Study areas varied significantly in size. Most study areas consisted of a major city and surrounding area, but large areas were studied in e.g. Catalunya and the Netherlands.

Measurements were conducted between October 2008 and April 2011. Sites were selected following a common study manual (available on www.escapeproject.eu) to represent the anticipated spatial variation of air pollution in the included cohort studies. Based on the ESCAPE manual, each local group made a site selection proposal with a detailed characterization of the sites including pictures of the study area and sites. The proposal was evaluated by the ESCAPE exposure working group to harmonize the site selection across the groups. Sites were classified as regional background, urban background and street sites. In each area street sites were over represented as we anticipated more variation between street sites than between regional or urban background sites. Sites representing the full range of local traffic intensities were selected. Other sources were also considered during site selection, e.g. industries or major ports. Physical geography characteristics such as altitude were also considered.

At each site measurements were conducted during three 14-day periods representing the warm, cold and one intermediate season. Ogawa badges were used for NO_x measurements (Van Roosbroeck et al., 2006). For each site an annual average concentration was calculated, correcting for temporal variation using measurements obtained from a centrally located reference site or from a routine monitoring site (areas with only NO_x measurements) which was operated year-round (Cyrys et al., 2012).

The geographical coordinates of each site were recorded with Global Positioning System (GPS) readings at each site visit. Because unacceptable variation of repeated GPS readings of more than 10 m was found in some sites and because of the large spatial variability of traffic-related air pollutants that occurs within tens of metres from major roads, the GPS readings were used as a proxy for the

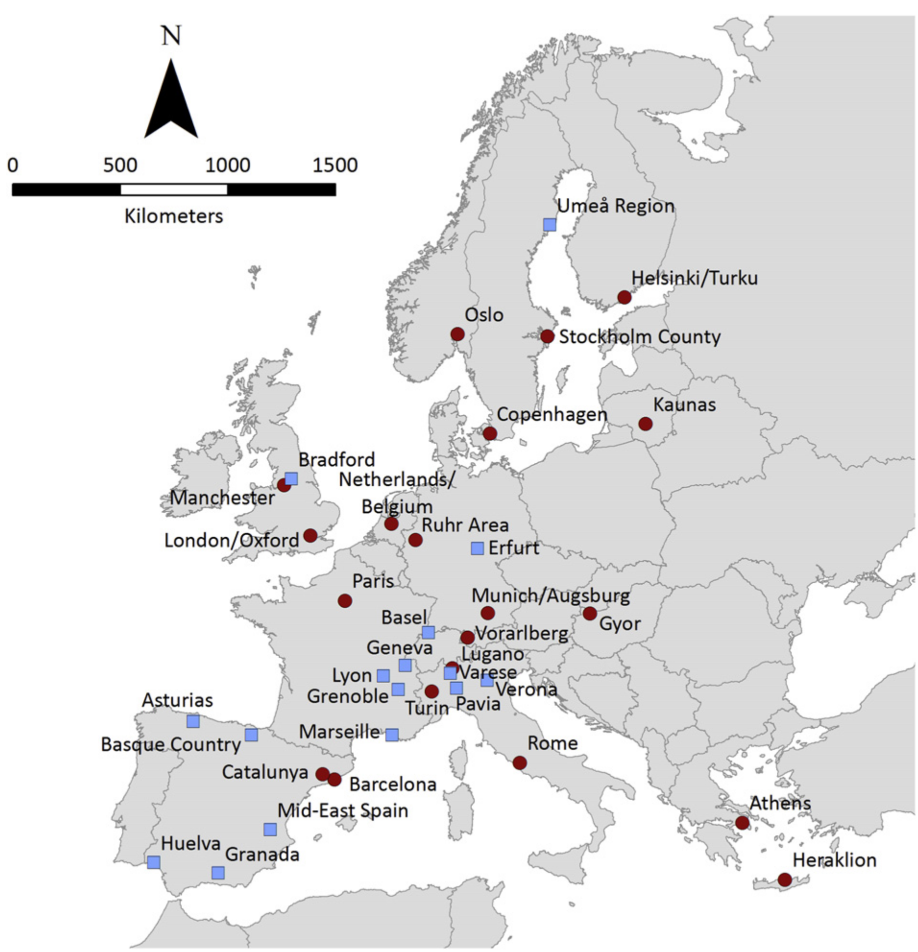

Fig. 1. ESCAPE study areas. Circles mark the study areas where both PM and NO_x were measured. Squares indicate the study areas where only NO_x was measured.

correct coordinates. Final coordinates of the monitoring locations were therefore extracted from the same, accurate digital maps which were also used for extracting GIS predictors.

2.2. Predictor data for LUR models

GIS analyses were conducted to derive the values for the predictor variables for the coordinates of the monitoring sites. Table 1 shows the potential predictor variables, and the a priori choices we made in the design for LUR model development including the buffer sizes, transformations of the variable and the a priori defined direction of effect. The buffer sizes were selected to take account of known dispersion patterns. Both small-scale and larger-scale buffer sizes were used for the traffic variables indicating two scales of influence: near source and urban background levels representing larger-area traffic density (Su et al., 2009).

2.2.1. Central GIS data

The following central datasets were available for all study areas:

1. Digital road network

High resolution road data were obtained from the Eurostreets version 3.1 digital road network (1:10,000 resolution) which is based on the TeleAtlas MultiNet TM for the year 2008. Attributes include name of street, route number, speed class, length and road classification (0: Motorways; 1: Roads belonging to 'Main road' major importance; 2: Other major roads; 3: Secondary roads; 4: Local connecting roads; 5: Local roads of high importance; 6: Local roads; 7: Local roads of minor importance; 8: Others). In the absence of traffic intensity data, several LUR studies have successfully explored the use of the length of specific road types without traffic intensity (Hoek et al., 2008). All roads of class 0, 1 and 2 were classed as major roads (and classes 3 and 4 based on local knowledge and decision). This dataset was used to calculate the total length of all roads, and of major roads in buffers of 25, 50, 100, 300, 500 and 1000 m. We also calculated the distance to the nearest road and the nearest major road.

2. Land use data

The centrally available CORINE (COordination and INformation on the Environmental programme, initiated by the European Commission) land use data have been shown to be predictors in land use regression models (Vienneau et al., 2010; Beelen et al., 2009). CORINE land cover 2000 data were available from the European Environment Agency as a 1:100,000 vector database (except Norway and Switzerland) (EEA, 2005). It comprises 44 land cover classes and has a spatial minimum mapping resolution of 25 ha. The 44 categories were regrouped in 6 classes (High density residential land; Low density residential land; Industry; Port; Urban green; and Semi-natural and forested areas) following a recent paper on LUR modelling in the UK and The Netherlands (Vienneau et al., 2010). The surface area (in m^2) of each land-use was calculated in buffers of 100, 300, 500, 1000 and 5000 m.

3. Population density data

Population data modelled at a 100 m grid were available from the INTARESE project (http://www.integrated-assessment.eu/resource_centre/eu_agesex_stratified_population_100_metre_grid). Following the approach of Briggs et al. (2007) in which census data were spatially disaggregated on the basis of land cover data, the 2001 population density available from the EEA (http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2) (Gallego, 2010) was enhanced and converted to a grid of population numbers. Because the data are based on CORINE land cover, the population data excludes Switzerland and Norway. Total population was calculated in buffers of 100, 300, 500, 1000 and 5000 m around each site.

4. Altitude

Digital elevation data (SRTM 90 m) were obtained through the Shuttle Radar Topographic Mission, and available globally from CGIAR-CSI GeoPortal (http://srtm.csi.cgiar.org/). The data have a resolution of 90 m at the equator.

2.2.2. Local GIS data

Local GIS data were collected in each study area (if available). Online supplement 1 describes in detail the available local GIS data for each area.

1. Local digital road network with linked traffic intensity data

Because the central road network only contained a road classification, local road networks with linked traffic intensity data were collected. If available, traffic intensity data were collected for different traffic types (light-duty and heavy-duty). The accuracy for the local road network should be at least 10 m compared to the central road network.

The local road network was used to collect total traffic load in vehicles*metres, calculated as the length of a road segment*the traffic intensity on that road segment. This was done for buffers of 25, 50, 100, 300, 500 and 1000 m for both heavy traffic and total traffic. Similar calculations were done for major roads only (where major was defined as a road with over 5000 motor vehicles (mvh)/24 h). We also calculated the traffic intensity on the nearest road and the nearest major road, and the distance to these roads.

2. Land use data

CORINE may not incorporate specific land use for some study areas. Local land use data, with accuracy of at least 100 m compared to central data, were therefore collected. The surface area (in m^2) of each land-use was calculated in buffers of 100, 300, 500, 1000 and 5000 m.

3. Population density data

Local population density data were collected because the centrally available population density data were modelled. If available, household density data were also collected. The accuracy should be at least 100 m. Both the number of inhabitants and households were calculated in buffers of 100, 300, 500, 1000 and 5000 m.

4. Altitude

Local altitude data were collected in areas where local data were better than the central altitude data. The required resolution for altitude data was 100 m.

5. Study area specific local data

This may include information about wood smoke, distance to sea/lake and distance to major air pollution sources. Region indicators or coordinate variables were in some areas used to model large-scale spatial trends that could not be explained by the other predictors which had a maximum buffer of 5000 m.

Table 1
Predictor variables with predefined variable names, units, defined buffer sizes, transformations of the predictor variables, a priori defined direction of effect, and indicator whether data are central or local data, divided in background and traffic variables.

GIS dataset	Predictor variable	Name variable[a]	Unit	Buffer size (radius of buffer in metre)	Transformation	Direction of effect	Central or local data
Background variables							
CORINE	High density residential land[b]	HDRES	m^2	100, 300, 500, 1000, 5000	NA	+	Central
CORINE	Low density residential land[b]	LDRES	m^2	100, 300, 500, 1000, 5000	NA	+	Central
CORINE	Sum of high density and low density residential land[b]	HLDRES	m^2	100, 300, 500, 1000, 5000	NA	+	Central
CORINE	Industry[b]	INDUSTRY	m^2	100, 300, 500, 1000, 5000	NA	+	Central
CORINE	Port[b]	PORT	m^2	100, 300, 500, 1000, 5000	NA	+	Central
CORINE	Urban green[b,c]	URBGREEN	m^2	100, 300, 500, 1000, 5000	NA	−	Central
CORINE	Semi-natural and forested areas[b,d]	NATURAL	m^2	100, 300, 500, 1000, 5000	NA	−	Central
CORINE	Sum of urban green and semi-natural and forested areas[b]	GREEN	m^2	100, 300, 500, 1000, 5000	NA	−	Central
Local land use	Similar as CORINE variables[b]	Similar as CORINE variables	m^2	100, 300, 500, 1000, 5000	NA	Following CORINE	Local
Local land use	Area[b]/number of buildings	BUILDINGS	m^2/N(umber)	100, 300, 500, 1000, 5000	NA	+	Local
Local land use	Area of water[b]	WATER	m^2	100, 300, 500, 1000, 5000	NA	−	Local
Population density	Number of inhabitants	POP	N(umber)	100, 300, 500, 1000, 5000	NA	+	Central/ Local
Household density	Number of households	HHOLD	N(umber)	100, 300, 500, 1000, 5000	NA	+	Local
Altitude	Altitude	SQRALT	m	NA	square root	−	Local
–	Regional estimate[c]	REGIONALESTIMATE	NA	NA	Local decision	NA	Local
–	Coordinate variables[c]	XCOORD, YCOORD or other combinations	m	NA	Local decision	NA	Local
–	Area indicator[c]	AREA	NA	NA	NA	NA	Local
–	Distance to sea	DISTINVSEA DISTINVSEA2	m	NA	Inverse distance and inverse distance squared	−	
Traffic variables							
Central road network	Road length of all roads in a buffer	ROADLENGTH	m	25, 50, 100, 300, 500, 1000	NA	+	Central
Central road network	Road length of major roads in a buffer[d]	MAJORROADLENGTH	m	25, 50, 100, 300, 500, 1000	NA	+	Central
Central road network	Distance to the nearest road	DISTINVNEARC1 DISTINVNEARC2	m^{-1}, m^{-2}	NA	Inverse distance and inverse distance squared	+	Central
Central road network	Distance to the nearest major road[d]	DISTINVMAJORC1 DISTINVMAJORC2	m^{-1}, m^{-2}	NA	Inverse distance and inverse distance squared	+	Central
Local road network	Traffic intensity on nearest road[e]	TRAFNEAR	Veh. day^{-1}	NA	NA	+	Local
Local road network	Distance to the nearest road	DISTINVNEAR1 DISTINVNEAR2	m^{-1}, m^{-2}	NA	Inverse distance and inverse distance squared	+	Local
Local road network	Product of traffic intensity on nearest road[e] (INT) and inverse of distance to the nearest road (INVDIST) and distance squared (INVDIST2)	INTINVDIST INTINVDIST2	Veh. day^{-1} m^{-1} Veh. day^{-1} m^{-2}	NA	NA	+	Local
Local road network	Traffic intensity on nearest major road[e,f]	TRAFMAJOR	Veh. day^{-1}	NA	NA	+	Local
Local road network	Distance to the nearest major road[f]	DISTINVMAJOR1 DISTINVMAJOR2	m^{-1}, m^{-2}	NA	Inverse distance and inverse distance squared	+	Local
Local road network	Product of traffic intensity on nearest major road[e] (INTMAJOR) and inverse of distance to the nearest major road (INVDIST) and distance squared (INVDIST2)[f]	INTMAJORINVDIST INTMAJORINVDIST2	Veh. day^{-1} m^{-1} Veh. day^{-1} m^{-2}	NA	NA	+	Local
Local road network	Total traffic load of major roads in a buffer (sum of (traffic intensity*length of all segments))[f]	TRAFMAJORLOAD	Veh. day^{-1} m	25, 50, 100, 300, 500, 1000	NA	+	Local
Local road network	Total traffic load of all roads in a buffer (sum of (traffic intensity*length of all segments))	TRAFLOAD	Veh. day^{-1} m	25, 50, 100, 300, 500, 1000	NA	+	Local
Local road network	Heavy-duty traffic intensity on nearest road[e]	HEAVYTRAFNEAR	Veh. day^{-1}	NA	NA	+	Local

Local road network	Product of heavy-duty traffic intensity on nearest road[e] (HEAVYINT) and inverse of distance to the nearest road (INVDIST) and distance squared (INVDIST2)	HEAVYINTINVDIST HEAVYINTINVDIST2	Veh. day^{-1} m^{-1} Veh. day^{-1} m^{-2}	NA	NA	+	Local
Local road network	Heavy-duty traffic intensity on nearest major road[e,f]	HEAVYTRAFMAJOR	Veh. day^{-1}	NA	NA	+	Local
Local road network	Total heavy-duty traffic load of major roads in a buffer (sum of (heavy-duty traffic intensity*length of all segments))[f]	HEAVYTRAFMAJORLOAD	Veh. day^{-1} m	25, 50, 100, 300, 500, 1000	NA	+	Local
Local road network	Total heavy-duty traffic load of all roads in a buffer (sum of (heavy-duty traffic intensity*length of all segments))	HEAVYTRAFLOAD	Veh. day^{-1} m	25, 50, 100, 300, 500, 1000	NA	+	Local

NA is not applicable.

[a] Variable name: Combining name and buffer size, for example for HDRES: HDRES_100, HDRES_300, HDRES_500, HDRES_1000, HDRES_5000.

[b] Area of that land use in the buffer (m^2). Area for BUILDINGS refers to the footprint of all buildings in a buffer.

[c] Variables were only offered if a model has been developed to test if the model with more explicit variables could be improved with these variables (describing slow trends in background). For the models in the Netherlands and Belgium the regional estimate was a regional background concentration estimate for each site location, based on inverse distance weighted interpolation of regional background sites.

[d] Definition of major road for central road network: classes 0, 1, and 2 (+classes 3 and 4 based on local knowledge and decision).

[e] Traffic intensities are traffic intensities per 24 h.

[f] Definition of major road for local road network: road with traffic intensity >5000 mvh/24 h.

2.3. LUR model development

The average NO_2 and NO_x concentrations and the GIS based values for potential predictor variables were used to develop LUR models following a standardized approach described in the ESCAPE exposure manual (available on www.escapeproject.eu). A workshop was organized for all persons developing models to further standardize model development procedures. LUR models were developed locally and then centrally evaluated by the ESCAPE exposure working group for final approval.

Regression models that maximize the adjusted percentage explained variance (R^2) were developed to explain these concentrations, using a supervised forward stepwise procedure with a priori defined predictor variables (Table 1). First, univariate regression analyses were conducted for all predictor variables. The model with the highest adjusted R^2 and a slope of the pre-specified direction was regarded as the start model. Second, to this start model all remaining variables, regardless how high the R^2 was in univariate analyses, were added consecutively and the effect on the adjusted R^2 recorded. The predictor variable with the highest additional increase in adjusted R^2 was maintained if three criteria were fulfilled: 1) the absolute increase in adjusted R^2 was more than 1%, 2) the coefficient conformed to the predefined direction of effect, and 3) the direction of effect for predictors already in the model did not change. When a variable was included, other buffer sizes of the same variable continued to be offered to the model, both smaller and larger buffers. When a variable enters a model with different buffer sizes these are overlapping. To make the model more intuitively interpretable, in the final model such a variable was rewritten using concentric adjacent rings (or doughnut-shaped buffers), e.g. traffic in a 50 m buffer and traffic between 50 and 500 m (Von Klot, 2011).

The addition of variables was repeated until there were no variables that added more than 1% to the adjusted R^2 of the previous model. Finally, variables with p-values larger than 0.1 were sequentially removed from the model as a final step.

Diagnostic tests were applied to the final models: multicollinearity between included variables (Variance Inflation Factors or VIF), influential observations (Cook's D), and heteroscedasticity, normality and spatial autocorrelation (Moran's I) of residuals to assess the independence assumption. In case of a high VIF value, specified as >3, for predictor variables included in the final model, the variable with highest VIF value was excluded from the model and the decrease in predictive power of the model was evaluated. We assessed a high Cook's D value (i.e. > 1), indicating an influential observation, because it could be caused by an (extreme) high or low concentration of one of the site(s) or by an included predictor variable with extreme values or many zero values. In such case, the developed model was applied to all sites minus the site with the high Cook's D value and the changes in model structure (parameter estimates and p-values of included variables, and model R^2) were evaluated. If the high Cook's D value was caused by one of the included predictor variables (indicated by a large change in parameter estimate for that variable without the site), a new LUR model was developed using all sites but without offering that specific predictor variable to the model. We used this procedure because we aimed at having a stable model that was not influenced by one or more individual sites, which might limit external validity. In addition, this was done because the default procedure was not to exclude data from individual monitoring sites from model development. Data from individual sites were only excluded if the site was an influential observation in the analyses (based on Cook's D analyses), when the developed model changed substantially (with respect to model R^2 and parameter estimates and p-values of included variables) after excluding such a site, and if the site was not representative for locations where study participants live (e.g. very close to a freeway), i.e. an insight that the site should not have been selected. This was decided based on consensus between the local group and the ESCAPE exposure working group. For the models for Stockholm County, Bradford, Manchester, Erfurt, Geneva and Basque Country one site was excluded. A detailed description of each of the excluded sites can be found in Online supplement 2. The number of excluded sites was very small compared to the total number of monitored sites (N $\sim$ 1400 NO_x sites).

Model performance was evaluated using a leave-one-out cross-validation (CV) method (Brauer et al., 2003). The final model was fitted to $N - 1$ sites, where the structure of the model remains constant but parameter estimates can change, and the predicted concentrations were compared with the actual measured concentrations at the left-out site. This procedure was repeated N times and the over all level of fit (R^2) between the predicted and measured concentrations, across all sites, was then calculated as a measure of model performance.

We finally developed models for urban background concentrations including the regional and urban background sites and background predictors (Table 1). These models were used in ESCAPE together with traffic intensity variables as alternative exposure models. See online supplement 3 for more details about background models.

GIS analyses were conducted with ArcGIS. Statistical analyses were conducted using the statistical package available in the local study area (SAS, SPSS, STATA or R).

3. Results

3.1. Model input data

Descriptive statistics of the air pollution measurements have been reported previously (Cyrys et al., 2012). Tables 2 and 3 describe the mean and range in concentrations for NO_2 and NO_x. Substantial spatial variations were found which were larger for NO_x than for NO_2. Within-area contrasts were largest for Catalunya, Barcelona and London-Oxford and smallest for Kaunas, Gyor and Bradford. The number of selected traffic sites differed per area, with a range of 5 traffic sites in Umeå Region and 28 traffic sites in Barcelona.

Local road network data with linked traffic intensity data were available for most areas, except Heraklion, Asturias, and Basque Country. Other local GIS data such as population or household density were available in 28 of the 36 areas. Some buffer variables had for a large number of sites in each area a value of zero, for example the smaller buffers for industry, port, urban green and semi-natural and forested areas, as well as the 25 m buffers for the traffic variables. These variables could induce influential observations. The ESCAPE exposure working group recommended that a variable should have at least 5 sites with a non-zero value, otherwise it could be a priori excluded from model development by the local team. Within the CORINE database a differentiation between low density and high density residential land and between urban green and semi-natural and forested areas was available. Based on previous experience, local knowledge and after evaluation by the ESCAPE exposure working group for standardization, it was decided to sum both into new combined variables in some areas: residential land variable and green space (see Tables 2 and3).

3.2. Model results

The final LUR models are presented in Tables 2 and 3 for NO_2 and NO_x respectively. Online supplement 4 describes the distribution of the included predictor variables for the NO_2 and NO_x models.

Model R^2 ranged from 55% to 92% (median 82%) for NO_2 (Table 2, Fig. 2) and from 49% to 91% (median 78%) for NO_x (Table 3). R^2 values for NO_2 and NO_x models were very similar within a study area consistent with the high correlation between the measured concentrations of both pollutants in all study areas (>0.9) (Cyrys et al., 2012). In some areas (almost) the same predictor variables were included in both models (e.g. Umeå Region, Copenhagen, Kaunas, Ruhr Area, Erfurt, Lyon, Varese, Asturias, Basque Country, Catalunya, Mid-East Spain, Granada, Huelva, Heraklion).

Highest model R^2 were found in Pavia, Mid-East Spain and Lyon for NO_2 and in Stockholm, Bradford and London-Oxford for NO_x. Lowest model R^2 were found in Huelva and Basque Country. For the Basque Country study area neither local traffic nor local land use data were available. For the Huelva study area only limited local traffic data were available, and in addition the model was developed on only 24 sites as models for the Andalucia study area were separately developed for Huelva and Granada. These low model performance results for these areas illustrate the importance of availability of accurate local traffic data in addition to the centrally available GIS data. In an additional evaluation of the value of traffic intensity data, LUR models were centrally developed for each area with and without offering local traffic intensity data. LUR models with local traffic intensity data had R^2 values which were on average 10% higher than models without local traffic intensity data. Online supplement 5 shows the R^2 values for both models for each study area.

Fig. 2 shows the distribution of NO_2 model R^2 values for all areas, and for the regions North, West-Central and South. It shows that on average model R^2 values are highest in North and lowest in South, but the range in R^2 values is considerable within a region.

The number of included predictor variables ranged from 2 to 7 (average number $\sim$4) for both NO_2 and NO_x. All the NO_2 and NO_x models included one or more traffic variables such as traffic intensity on nearby roads, distance to nearby roads and traffic intensity in a small buffer around the site. Fig. 3 shows the frequency of categories of predictor variables in NO_2 LUR models over all study areas. In the 36 NO_2 LUR models 137 different predictor variables were included, of which 86 were traffic variables ($\sim$63%). The category of traffic intensity within 100 m variables was the most common traffic variable ($N = 22$). In addition, several models included traffic variables with a 1000 m buffer, such as the total traffic load on all roads or major roads or the length of all roads. Most models further included population or household density. In the models covering the Netherlands and Belgium a regional estimate was included to model (large-scale) spatial trends that could not be explained by the other potential predictor variables (that had a maximum buffer size of 5000 m) as the study area was $\sim$200 $\times$ 300 km. The regional estimate was based on inverse distance weighted interpolation of regional background sites concentrations. For the models for Mid-East Spain, where a combined model was developed for Albacete and Asturias, an indicator variable for region was included.

For most areas the CV R^2 was less than 10% lower than the model R^2. For a few areas the difference was larger than 10% (Tables 2 and 3). For example, in Munich the NO_2 model R^2 and CV R^2 were 86% and 67%, respectively. This was largely due to two sites as without these two sites the cross-validation R^2 increased to 76% illustrating that one or a few sites can influence the cross-validation R^2 values. Because the sites were representative sites they were not excluded from model development.

The spatial autocorrelation in the residuals evaluated with the Moran's I value was generally small and non-significant (Tables 2 and 3).

Table 2

Description of developed LUR models for NO_2, including descriptive statistics of the concentrations. Order of study areas is from North to South.

Study area	LUR model[a]	R^2 of model	R^2 cross validation	RMSE (cross validation) ($\mu g\ m^{-3}$)	Number of sites[b]	Moran's I (p-value)	Measured concentration ($\mu g\ m^{-3}$)[c]
Oslo, Norway	2.43 + 6.52E-2*MAJORROADLENGTH_50 + 2.85E-4*TRAFNEAR + 7.30E-7*HEAVYTRAFMAJORLOAD_500 + 9.34E-2*POP_100 + 2.03E-4*POP_100_5000	76%	64%	7.9	39	−0.03 (0.48)	23.2 [6.7 − 59.8]
Umeå Region, Sweden	3.51 + 0.26*MAJORROADLENGTH_25 + 2.22E-7*TRAFLOAD_500	87%	83%	2.8	40	−0.08 (0.43)	9.3 [1.5 − 35.8]
Stockholm County, Sweden	5.82 + 1.08E-3*ROADLENGTH_500 + 4.01E-4*TRAFNEAR + 5.39E-3* POP_100	82%	78%	3.5	39	−0.23 (0.00)	15.5 [2.1 − 33.0]
Helsinki and Turku, Finland	7.61 + 1.18E-5*TRAFLOAD_25 + 3.43E-8*TRAFLOAD_25_1000 + 0.04*ROADLENGTH_25 + 1.24E-3*ROADLENGTH_25_300 − 9.18E-5*URBGREEN_500[d]	83%	75%	3.4	40	0.07 (0.19)	18.9 [6.1 − 40.8]
Copenhagen, Denmark	10.64 + 3.15E-7*TRAFLOAD_300 + 2.87E-7*PORT_5000 + 488.80*DISTINVMAJOR2 − 3.00E-6*GREEN_1000[d]	88%	83%	3.9	41	−0.06(0.57)	17.8 [6.3 − 50.1]
Kaunas, Lithuania	8.18 + 76.64*DISTINVMAJOR1 + 3.95E-4*POP_1000 + 2.28E-3* HEAVYTRAFNEAR	72%	63%	3.6	40	−0.03 (0.25)	16.7 [8.3 − 36.2]
Bradford, UK	16.52 + 7.81E-5*BUILDINGS_300 + 5.86E-6*TRAFLOAD_25 + 3.20E-8*TRAFLOAD_25_1000 − 1.09E-4*NATURAL_300[d] + 7.43E-4*HEAVYTRAFMAJOR	83%	80%	2.4	40	−0.07 (0.20)	25.3 [16.7 − 36.7]
Manchester, UK	20.40 + 1.87E-2*HEAVYINTINVDIST + 3.39E-7*INDUSTRY_5000 + 1.20E-5*HDRES_500 − 1.89E-6*NATURAL_1000 + 1.61E-3*MAJORROADLENGTH_300 + 4.37E-2*ROADLENGTH_25	83%	75%	2.6	39	−0.006 (0.50)	25.6 [17.0 − 37.1]
London, Oxford, UK	8.51 + 7.30E-6*TRAFMAJORLOAD_50 + 1.10E-3*ROADLENGTH_500 + 2.00E-7*HLDRES_5000	89%	87%	6.6	40	−0.009 (0.71)	37.9 [7.3 − 102.7]
Netherlands and Belgium	−7.80 + 1.18*REGIONALESTIMATE + 2.30E-5*POP_5000 + 2.46E-6*TRAFLOAD_50 + 1.06E-4*ROADLENGTH_1000 + 9.84E-5*HEAVYTRAFLOAD_25 + 12.19*DISTINVNEARC1 + 4.47E-7*HEAVYTRAFLOAD_25_500	86%	81%	5.1	80	−0.143 (0.09)	30.9 [12.8 − 61.5]
Ruhr Area, Germany	19.66 + 3.48E-7*INDUSTRY_5000 + 2.24E-2*POP_100 + 4.10E-6*PORT_5000 + 1.31E-6*TRAFLOAD_100	89%	84%	4.3	40	−0.18 (0.08)	33.2 [20.2 − 58.4]
Erfurt, Germany	11.11 + 2.44E-3*INTMAJORINVDIST + 9.47E-7*TRAFLOAD_100 + 3.81E-8*TRAFLOAD_100_1000	89%	87%	2.1	39	0.01 (0.16)	18.6 [11.0 − 33.4]
Munich-Augsburg, Germany	7.43 + 1.98E-6*TRAFLOAD_50 + 1.35E-3*INTMAJORINVDIST + 2.37E-2*ROADLENGTH_50 + 1.47E-5*POP_5000 + 4.15E-2*MAJORROADLENGTH_50[d] + 9.85*HLDRES_500[d]	86%	67%	5.5	40	−0.04 (0.86)	26.8 [13.6 − 53.5]
Vorarlberg, Austria	8.60 + 4.16E-3*ROADLENGTH_300 + 1.02E-5*TRAFLOAD_25 − 3.56E-7*NATURAL_5000 + 2.44E-5*INDUSTRY_1000	74%	66%	4.2	40	0.03 (0.29)	22.6 [7.4 − 39.7]
Paris, France	25.86 + 4.85E-7*TRAFLOAD_300 + 0.26*MAJORROADLENGTH_25 − 5.60E-7*NATURAL_5000	77%	67%	11.6	40	−0.05 (0.71)	39.8 [6.9 − 96.8]
Grenoble, France	8.99 + 2.87E-6*HEAVYTRAFLOAD_300 + 5.18E-7*HEAVYTRAFLOAD_300_1000 + 95.96*DISTINVMAJOR1 + 1.39E-3*ROADLENGTH_300	83%	78%	4.8	40	−0.08 (0.82)	27.2 [5.5 − 53.2]
Lyon, France	6.64 + 5.41E-2*ROADLENGTH_50 + 3.04E-4*TRAFMAJOR + 2.85E-6*HDRES_5000 + 0.16*MAJORROADLENGTH_25	90%	72%	8.7	40	−0.06 (0.80)	35.0 [7.3 − 88.0]
Marseille, France	7.09 + 0.19*MAJORROADLENGTH_25 + 3.32E-4*ROADLENGTH_1000 + 83.31*DISTINVMAJOR1	59%	46%	10.7	40	−0.06 (0.87)	36.1 [10.0 − 92.8]
Gyor, Hungary	7.36 + 7.83E-8*TRAFLOAD_1000 + 5.66E-3*INDUSTRY_5000 + 1.25E-2*MAJORROADLENGTH_100	75%	69%	3.3	40	−0.04 (0.84)	16.7 [9.2 − 37.4]
Basel, Switzerland	54.35 + 1.23E-2*INTMAJORINVDIST + 1.88E-5*HLDRES_500[d] + 5.33E-7*HLDRES_500_5000[d] − 3.96*SQRALT	67%	58%	4.8	40	−0.05 (0.45)	31.0 [16.0 − 47.8]
Geneva, Switzerland	21.32 + 1.14E-5*TRAFLOAD_25 + 3.42E-7*TRAFLOAD_25_300 + 2.35E-2*BUILDINGS_100 − 4.44E-7*NATURAL_5000[d] + 2.51E-8*TRAFMAJORLOAD_1000	87%	81%	3.7	40	0.002 (0.25)	29.3 [16.1 − 51.3]
Lugano, Switzerland	39.67 + 5.45E-6*TRAFLOAD_50 + 4.95E-8*TRAFLOAD_50_1000 + 4.31E-4*TRAFMAJOR − 1.50*SQRALT + 4.33E-3*INTINVDIST	87%	82%	3.5	42	−0.04 (0.51)	28.6 [12.2 − 59.1]
Turin, Italy	19.95 + 7.79E-8*TRAFLOAD_1000 + 3.53E-2*MAJORROADLENGTH_100 + 5.63E-5*HLDRES_300 − 4.46E-7*NATURAL_5000	78%	70%	7.7	40	−0.08 (0.10)	53.3 [15.6 − 83.7]
Varese, Italy	179.70 − 4.05E-5*NATURAL_1000 + 269.00*DISTINVMAJORC2 − 6.45*SQRALT	72%	61%	10.6	20	−0.10 (0.18)	36.5 [17.8 − 77.7]
Pavia, Italy	−16.98 + 0.13*MAJORROADLENGTH_50 + 3.52E-2*ROADLENGTH_50 + 3.72E-3*HHOLD_300	92%	87%	3.3	20	−0.05 (0.99)	25.9 [15.7 − 53.4]
Verona, Italy	−31.23 + 3.64E-5*TRAFLOAD_100 + 4.37E-6*HLDRES_5000 + 1.07E-3*MAJORROADLENGTH_1000	64%	55%	10.8	40	−0.05 (0.33)	41.6 [16.3 − 100.1]

(*continued on next page*)

Table 2 (*continued*)

Study area	LUR model[a]	R^2 of model	R^2 cross validation	RMSE (cross validation) ($\mu g\ m^{-3}$)	Number of sites[b]	Moran's I (*p*-value)	Measured concentration ($\mu g\ m^{-3}$)[c]
Rome, Italy	12.88 + 6.00E-3*POP_100 + 3.28E-4*ROADLENGTH_1000 + 107.78*DISTINVNEAR2 + 1.14E-6*INDUSTRY_5000 − 1.01E-5*URBGREEN_1000 + 4.12E-6*TRAFLOAD_50 + 1.51E-2*MAJORROADLENGTH_100	87%	76%	6.8	40	0.034 (0.17)	42.6 [13.6 − 72.6]
Asturias, Spain	8.13 + 2.72E-4*HHOLD_300 + 0.28*MAJORROADLENGTH_25 − 4519.10*DISTINVSEA1	75%	69%	9.2	40	−0.11 (0.09)	32.0 [8.6 − 76.4]
Basque Country: Bilbao-San Sebastian, Spain	6.66 + 3.24E-4*ROADLENGTH_1000 + 0.13*MAJORROADLENGTH_25 + 1.47E-6*LDRES_5000	58%	50%	6.5	39	−0.07 (0.72)	25.6 [6.8 − 74.3]
Barcelona, Spain	3.16 + 6.26E-3*INTINVDIST1 + 1.18E-4*HDRES_300 + 992.09*DISTINVMAJOR2 + 3.51E-4*ROADLENGTH_1000	75%	68%	11.6	40	−0.03 (0.98)	57.7 [13.8 − 109.0]
Catalunya, Spain	13.52 + 6.09E-4*ROADLENGTH_1000 + 9.97E-6*TRAFMAJORLOAD_25 − 5.54E-7*NATURAL_5000	71%	69%	12.0	80	−0.01 (0.95)	47.8 [12.2 − 109.0]
Mid-East Spain: Albacete-Valencia, Spain	−16.00 + 3.15E-5*HDRES_500 + 5.66E-7*HHOLD_5000 + 1271.43*DISTINVMAJOR2 + 9.07*AREA	90%	87%	5.2	38	0.03 (0.37)	26.1 [1.9 − 75.5]
Granada, Spain	6.70 + 4.89E-5*HHOLD_1000 + 36.97*DISTINVMAJORC1	82%	77%	11.2	14	−0.03 (0.81)	34.8 [12.4 − 89.0]
Huelva, Spain	3.84 + 5.51E-3*TRAFMAJOR + 5.22E-4*HDRES_100 − 1.51E9*DISTINVSEA2	55%	31%	7.0	24	−0.14 (0.10)	21.9 [8.4 − 43.4]
Athens, Greece	15.46 + 2.34E-6*TRAFMAJORLOAD_25 + 0.012*ROADLENGTH_100 + 0.001*HHOLD_500 + 6.24E-5*INDUSTRY_300 − 2.66E-5*NATURAL_1000	70%	55%	8.1	40	−0.09 (0.10)	36.5 [13.3 − 71.0]
Heraklion, Crete	7.39 + 35.14*DISTINVMAJORC1 + 1.62E-5*PORT_1000 + 1.65E-2*ROADLENGTH_50	62%	49%	5.0	40	0.007 (0.84)	15.3 [5.3 − 34.3]

[a] See for detailed explanation of the variable names in Table 1. Variables with _X (e.g. POP_100) are buffers with _X indicating the radius of the buffer in metres. The following predictors were derived for all sites: the surface area (m^2) of high density residential land (HDRES_X), low density residential land (LDRES_X), all residential land (HLDRES_X), industry (INDUSTRY_X), port (PORT_X), urban green space (URBGREEN_X), natural land (NATURAL_X), urban green and natural land combined (GREEN_X), water (WATER_X), the number (N) or surface area (m^2) of buildings (BUILDINGS_X), population (N) (POP_X) or number of households (HHOLD_X), the square root of altitude (SQRALT), a regional background concentration estimate ($\mu g\ m^{-3}$) (REGIONAL ESTIMATE), *X*-coordinate (XCOORD), *Y*-coordinate (YCOORD), area indicator (AREA), inverse/inverse distance squared to sea (DISTINVSEA, DISTINVSEA2), total length (m) of all road and all major road segments (ROADLENGTH_X, MAJORROADLENGTH_X), inverse distance (m^{-1}) and inverse distance squared (m^{-2}) to the nearest road of the central road network (DISTINVNEARC1, DISTINVNEARC2) and the nearest major road in the central road network (DISTINVMAJORC1, DISTINVMAJORC2), traffic intensity on the nearest road (TRAFNEAR) and nearest major road (TRAFMAJOR), heavy traffic intensity on the nearest (HEAVYTRAFNEAR) and the nearest major road (HEAVYTRAFMAJOR), inverse distance (m^{-1}) and inverse distance squared (m^{-2}) to the nearest road of the local road network (DISTINVNEAR1, DISTINVNEAR2) and the nearest major road in the local road network (DISTINVMAJOR1, DISTINVMAJOR2), the product of inverse/inverse distance squared to the nearest road and the traffic intensity on this road (vehicles $day^{-1}\ m^{-1}$/vehicles $day^{-1}\ m^{-2}$) (INTINVDIST, INTINVDIST2), equivalent for major roads (INTMAJORINVDIST, INTMAJORINVDIST2), and for heavy traffic (HEAVYINTINVDIST, HEAVYINTINVDIST2), the sum of (total intensity $\times$ length of all road segments) within a buffer (vehicles day^{-1} m) for all roads (TRAFLOAD_X), for major roads (TRAFMAJORLOAD_X), for heavy traffic (HEAVYTRAFLOAD_X) and heavy traffic on major roads (HEAVYTRAFMAJORLOAD_X). See Online supplement 4 for description of distributions of included variables.

[b] Number of sites that have been used for model development.

[c] Mean [min − max].

[d] Local data.

Table 3
Description of developed LUR models for NO_x, including descriptive statistics of the concentrations. Order of study areas is from North to South.

Study area	LUR model[a]	R^2 of model	R^2 cross validation	RMSE (cross validation) (μg m^{-3})	Number of sites[b]	Moran's I (*p*-value)	Measured concentration (μg m^{-3})[c]
Oslo, Norway	9.23 + 2.02E-5*HEAVYTRAFMAJORLOAD_100 + 0.24*MAJORROADLENGTH_50 + 0.14*POP_100 + 3.42E-4*POP_100_5000	76%	67%	15.3	39	−0.03 (0.18)	47.1 [14.2 − 116.8]
Umeå Region, Sweden	4.79 + 0.89*MAJORROADLENGTH_25 + 4.96E-7*TRAFLOAD_500	87%	82%	7.9	40	−0.13 (0.12)	18.9 [2.3 − 95.9]
Stockholm County, Sweden	11.09 + 1.77E-2*POP_100 + 2.92E-4* HDRES_100 + 1.32E-3*TRAFNEAR	83%	79%	8.2	39	−0.11 (0.20)	28.7 [2.5 − 78.7]
Helsinki and Turku, Finland	12.56 + 3.46E-5*TRAFLOAD_25 + 4.92E-8*TRAFLOAD_25_1000 + 1.70E-2*ROADLENGTH_100 − 5.58E-5*URBGREEN_1000[d] + 2.54E-3*HHOLD_300	85%	74%	7.8	40	0.09 (0.14)	30.6 [8.6 − 94.7]
Copenhagen, Denmark	13.88 + 6.18E-7*TRAFLOAD_300 + 1864.86*DISTINVMAJOR2 − 7.42E-6*GREEN_1000[d] + 3.82E-7*PORT_5000	83%	73%	11.5	41	−0.05 (0.72)	30.3 [9.0 − 123.6]
Kaunas, Lithuania	9.51 + 7.59E-4*POP_1000 + 225.99*DISTINVMAJOR1 + 5.07E-3*HEAVYTRAFNEAR	63%	52%	11.2	40	−0.06 (0.82)	28.8 [11.3 − 89.6]
Bradford, UK	19.76 + 1.68E-5*TRAFLOAD_25 + 1.90E-6*TRAFLOAD_25_100 + 2.74E-4*BUILDINGS_300 − 2.48E-3*NATURAL_100[d] + 1.92E-4*TRAFMAJOR	90%	88%	4.9	40	−0.11 (0.02)	43.7 [22.4 − 79.9]
Manchester, UK	−1.47E2 + 1.86E2*DISTINVMAJOR1 + 3.36E-4*ROADLENGTH_1000 + 5.38E-7*TRAFLOAD_100 + 4.38E-4*YCOORD	83%	78%	5.6	39	0.02 (0.16)	44.9 [27.4 − 74.3]
London, Oxford, UK	5.12 + 2.29E-5*TRAFLOAD_50 + 7.17E-3*ROADLENGTH_300 + 2.60E-7*HLDRES_5000	91%	88%	16.2	40	−0.009 (0.78)	69.3 [18.8 − 257.4]
Netherlands and Belgium	3.25 + 0.74*REGIONALESTIMATE + 4.22E-6*TRAFLOAD_50 + 6.36E-4*POP_1000 + 2.39E-6*HEAVYTRAFLOAD_500 + 71.65*DISTINVMAJOR1 + 0.21*MAJORROADLENGTH_25	87%	82%	11.2	80	−0.16 (0.06)	51.8 [17.5 − 130.8]
Ruhr Area, Germany	23.88 + 8.54E-7*INDUSTRY_5000 + 6.35E-2*POP_100 + 1.23E-5*PORT_5000 + 8.06E-6*TRAFLOAD_50	88%	81%	13.6	40	−0.03 (0.95)	60.0 [26.9 − 135.7]
Erfurt, Germany	16.35 + 5.81E-3*INTMAJORINVDIST + 1.67E-6*TRAFLOAD_100 + 5.31E-8*TRAFLOAD_100_1000	87%	84%	4.3	39	−0.02 (0.95)	28.8 [15.6 − 61.8]
Munich-Augsburg, Germany	13.34 + 3.90E-6*TRAFLOAD_50 + 8.97E-2*MAJORROADLENGTH_50[d] + 3.81E-3*INTMAJORINVDIST + 2.46E-8*TRAFLOAD_1000 + 5.11E-2*ROADLENGTH_50 + 19.50*HLDRES_1000[d]	88%	76%	9.4	40	−0.05 (0.65)	46.9 [23.7 − 95.2]
Vorarlberg, Austria	35.54 + 0.47*MAJORROADLENGTH_25 + 0.12*POP_100 − 6.58E-7*NATURAL_5000	60%	51%	12.7	40	−0.04 (0.72)	42.7 [14.6 − 91.4]
Paris, France	34.48 + 0.89*MAJORROADLENGTH_25 + 1.17E-6*TRAFMAJORLOAD_300	75%	67%	31.6	40	−0.06 (0.77)	80.3 [12.7 − 248.3]
Grenoble, France	8.10 + 9.80E-7*HEAVYTRAFMAJORLOAD_1000 + 9.97E-3*HEAVYTRAFNEAR + 3.72E-3*ROADLENGTH_300 + 156.66*DISTINVMAJOR1	82%	74%	11.2	40	−0.07 (0.76)	48.2 [6.5 − 116.2]
Lyon, France	−1.33 + 7.82E-4*TRAFMAJOR + 6.11E-6*HDRES_5000 + 0.43*MAJORROADLENGTH_25 + 0.11*ROADLENGTH_50	75%	65%	22.5	40	−0.08 (0.88)	61.7 [6.5 − 199.2]
Marseille, France	8.18 + 0.52*MAJORROADLENGTH_25 + 259.49*DISTINVMAJOR1 + 1.92E-3*ROADLENGTH_500	53%	39%	31.6	40	−0.04 (0.65)	70.1 [11.9 − 266.1]
Gyor, Hungary	6.95 + 9.59E-2*MAJORROADLENGTH_50 + 0.61*INDUSTRY_500 + 5.59E-3*ROADLENGTH_300 − 0.90*URBGREEN_300	74%	64%	10.4	40	−0.11 (0.39)	32.5 [13.4 − 86.3]
Basel, Switzerland	−80.41 + 2.75E-2*INTMAJORINVDIST + 5.25E-5*HLDRES_500[d] + 1.84E-6*HLDRES_500_5000[d]	61%	52%	12.0	40	−0.04 (0.71)	53.1 [21.6 − 95.7]
Geneva, Switzerland	−23.40 + 3.16E-5*TRAFLOAD_25 + 7.61E-7*TRAFLOAD_25_300 + 0.16*ROADLENGTH_25 + 0.82*BUILDINGS_100 + 1.34E-2* HLDRES_5000[d]	81%	73%	9.1	40	0.003 (0.23)	55.9 [22.1 − 108.6]
Lugano, Switzerland	75.6 6 + 2.82E-5*TRAFLOAD_25 + 9.11E-8*TRAFLOAD_25_1000 + 9.96E-4*TRAFMAJOR − 3.22*SQRALT + 2.24E-4*HEAVYTRAFLOAD_100	87%	82%	7.4	42	−0.05 (0.28)	47.8 [21.2 − 116.4]
Turin, Italy	32.86 + 1.60E-5*TRAFLOAD_50 + 1.50E-4*HLDRES_300 + 7.99E-8*TRAFLOAD_50_1000	78%	72%	17.0	40	−0.05 (0.36)	101.2 [22.8 − 101.2]
Varese, Italy	398.00 − 9.64E-3*NATURAL_1000 + 816.10*DISTINVMAJORC2 − 14.73*SQRALT	74%	52%	29.8	20	−0.12 (0.09)	70.8 [29.3 − 190.5]
Pavia, Italy	25.87 + 0.32*MAJORROADLENGTH_50 + 7.28E-2*ROADLENGTH_50	88%	80%	9.6	20	−0.08 (0.46)	50.9 [29.5 − 117.9]
Verona, Italy	57.65 + 8.61E-5*TRAFLOAD_100 + 1.72E-2*TRAFNEAR	64%	54%	32.3	40	−0.02 (0.91)	91.8 [33.1 − 284.4]
Rome, Italy	82.21 + 4.28E-4*LDRES_100 + 2.97E-2*POP_100 + 4.43E-6*TRAFLOAD_100 − 5.92*SQRALT − 8.43E-5*NATURAL_1000 − 3.01E-6*NATURAL_1000_5000 + 0.28*ROADLENGTH_25	80%	69%	15.1	40	0.006 (0.48)	72.7 [25.1 − 146.0]
Asturias, Spain	5.58 + 7.16E-4*HHOLD_300 + 0.81*MAJORROADLENGTH_25 − 10837.82*DISTINVSEA1	69%	62%	28.3	40	−0.11 (0.13)	69.7 [16.2 − 220.3]
Basque Country: Bilbao-San Sebastian, Spain	14.56 + 6.26E-4*ROADLENGTH_1000 + 0.26*MAJORROADLENGTH_25 + 2.60E-6*LDRES_5000	49%	39%	14.8	39	−0.08 (0.66)	52.7 [14.3 − 224.9]
Barcelona, Spain	32.85 + 2.55E-4*HDRES_300 + 2815.14*DISTINVMAJOR2 + 3.87E-5*TRAFLOAD_25	73%	65%	27.7	40	0.02 (0.26)	101.3 [21.0 − 236.4]
Catalunya, Spain	22.56 + 0.10*ROADLENGTH_1000 + 3.06E-5*TRAFMAJORLOAD_25 − 9.76E-7*NATURAL_5000	69%	66%	26.8	80	−0.03 (0.72)	85.5 [18.2 − 236.4]
Mid-East Spain: Albacete-Valencia, Spain	−32.39 + 5.40E-5*HDRES_500 + 9.83E-7*HHOLD_5000 + 2988.61*DISTINVMAJOR2 + 16.12*AREA	88%	84%	11.0	38	0.02 (0.46)	42.7 [0.6 − 148.6]

(*continued on next page*)

Table 3 (*continued*)

Study area	LUR model[a]	R^2 of model	R^2 cross validation	RMSE (cross validation) (μg m^{-3})	Number of sites[b]	Moran's I (*p*-value)	Measured concentration (μg m^{-3})[c]
Granada, Spain	1.41 + 1.00E-4*HHOLD_1000 + 84.69*DISTINVMAJORC1	78%	62%	30.3	14	−0.29 (0.29)	59.7 [20.6 − 195.9]
Huelva, Spain	6.84 + 9.12E-3*TRAFMAJOR + 8.53E-4*HDRES_100 − 2.73E9*DISTINVSEA2	56%	31%	11.5	24	−0.15 (0.08)	33.8 [13.3 − 71.3]
Athens, Greece	26.65 + 1.07E-5*TRAFMAJORLOAD_25 + 0.001*POP_500 − 2.09*SQRALT + 0.04*ROADLENGTH_100 + 0.001*INDUSTRY_100	67%	46%	30.1	40	−0.15 (0.005)	77.9 [21.4 − 230.1]
Heraklion, Crete	10.57 + 56.13*DISTINVMAJORC1 + 6.42E-5*PORT_500 + 2.48E-2*ROADLENGTH_50	62%	39%	8.1	40	−0.04 (0.96)	21.5 [8.6 − 52.8]

[a] See for detailed explanation of the variable names in Table 1. Variables with _X (e.g. POP_100) are buffers with _X indicating the radius of the buffer in meters. The following predictors were derived for all sites: the surface area (m^2) of high density residential land (HDRES_X), low density residential land (LDRES_X), all residential land (HLDRES_X), industry (INDUSTRY_X), port (PORT_X), urban green space (URBGREEN_X), natural land (NATURAL_X), urban green and natural land combined (GREEN_X), water (WATER_X), the number (N) or surface area (m^2) of buildings (BUILDINGS_X), population (N) (POP_X) or number of households (HHOLD_X), the square root of altitude (SQRALT), a regional background concentration estimate (μg m^{-3}) (REGIONAL ESTIMATE), *X*-coordinate (XCOORD), *Y*-coordinate (YCOORD), area indicator (AREA), inverse/inverse distance squared to sea (DISTINVSEA, DISTINVSEA2), total length (m) of all road and all major road segments (ROADLENGTH_X, MAJORROADLENGTH_X), inverse distance (m^{-1}) and inverse distance squared (m^{-2}) to the nearest road of the central road network (DISTINVNEARC1, DISTINVNEARC2) and the nearest major road in the central road network (DISTINVMAJORC1, DISTINVMAJORC2), traffic intensity on the nearest road (TRAFNEAR) and nearest major road (TRAFMAJOR), heavy traffic intensity on the nearest (HEAVYTRAFNEAR) and the nearest major road (HEAVYTRAFMAJOR), inverse distance (m^{-1}) and inverse distance squared (m^{-2}) to the nearest road of the local road network (DISTINVNEAR1, DISTINVNEAR2) and the nearest major road in the local road network (DISTINVMAJOR1, DISTINVMAJOR2), the product of inverse/inverse distance squared to the nearest road and the traffic intensity on this road (vehicles day^{-1} m^{-1}/vehicles day^{-1} m^{-2}) (INTINVDIST, INTINVDIST2), equivalent for major roads (INTMAJORINVDIST, INTMAJORINVDIST2), and for heavy traffic (HEAVYINTINVDIST, HEAVYINTINVDIST2), the sum of (total intensity × length of all road segments) within a buffer (vehicles day^{-1} m) for all roads (TRAFLOAD_X), for major roads (TRAFMAJORLOAD_X), for heavy traffic (HEAVYTRAFLOAD_X) and heavy traffic on major roads (HEAVYTRAFMAJORLOAD_X). See Online supplement 4 for description of distributions of included variables.

[b] Number of sites that have been used for model development.

[c] Mean [min − max].

[d] Local data.

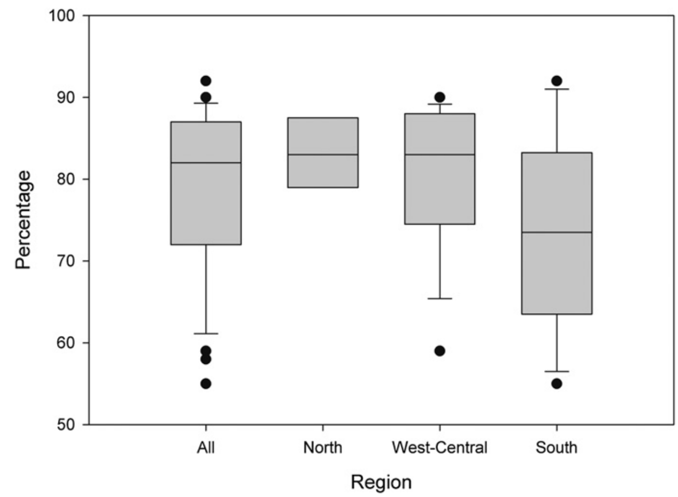

Fig. 2. Distribution of model R^2 values for NO_2 LUR models in all areas, region North, region West-Central and region South. (Region North: Oslo, Umeå Region, Stockholm County, Helsinki and Turku, and Copenhagen. Region West-Central: Kaunas, Bradford, Manchester, London and Oxford, Netherlands and Belgium, Ruhr area, Erfurt, Munich-Augsburg, Vorarlberg, Paris, Gyor, Basel, Geneva, and Lugano. Region South: Grenoble, Lyon, Marseille, Turin, Varese, Pavia, Verona, Rome, Asturias, Basque Country, Barcelona, Catalunya, Mid-East Spain, Granada, Huelva, Athens, and Heraklion.)

4. Discussion

The models developed for all sites explained a large fraction of the measured NO_2 and NO_x concentrations. This is consistent with the range in model R^2 in a recent LUR review paper (NO_2: 0.51−0.97 and NO_x: 0.73−0.96) (Hoek et al., 2008).

4.1. Comparison of models between different study areas

Differences in prediction R^2 between study areas may be related to the variability in the measured concentrations, the size

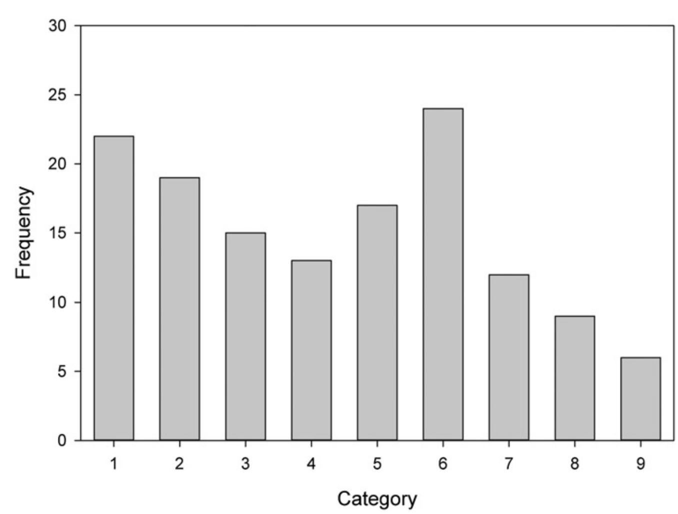

Fig. 3. Frequency of categories of included predictor variables in NO_2 LUR models over all study areas. Traffic variables: 1 = Traffic intensity =< 100 m (All traffic intensity variables with buffer size = <100 m, including traffic intensity on nearest and nearest major road); 2 = Road length = <100 m; 3 = Traffic intensity > 100 m; 4 = Road length > 100 m; 5 = Distance to traffic/road (All variables with distance to a road or traffic, including variables with product of traffic intensity and distance); and Background variables: 6 = Population/buildings/residential land/household density; 7 = Natural land/green space; 8 = Industry/port; and 9 = Other.

and complexity of the study area, the completeness and quality of predictor variables and quality of geo-coding. Substantial variation in measured concentrations was found for both NO_2 and NO_x in all areas. The on-average lower R^2 in Southern-Europe could be due to any of these factors, except concentration variability. An additional analysis showed that LUR models with local traffic intensity data had R^2 values which were on average 10% higher than models without local traffic intensity data. In several of the South-European study areas, traffic intensity data were missing or available on a limited scale only. Probably the more compact structure of cities in the South, resulting in high concentrations in canyon-like streets with moderate traffic intensity has contributed to the lower R^2 as no areas had data on street configuration available.

Models in the 36 study areas differed in the variables that were included in the models and/or the buffer size or type of road. This resulted from the decision to estimate the best prediction model for each individual study area. There are however also important similarities between the models. Fig. 3 showed that traffic variables were more often included than the background predictor variables in the NO_2 LUR models. All models contained a traffic variable. Traffic intensity of buffers smaller than 100 m was the most common traffic variable. All models contained a traffic variable describing small-scale variation, in line with studies that showed that NO_2 concentrations have a large decline in the first 100–200 m near high traffic intensity roads (Rijnders et al., 2001; Gilbert et al., 2003). In addition, several models included traffic variables with a 1000 m buffer, such as the total traffic load on all roads or major roads or the length of all roads, indicating a larger scale of influence, i.e. representing larger-area traffic density (Su et al., 2009). Population/address density was another common variable, consistent with well documented urban–rural differences related to a variety of sources including traffic and home heating (Brauer et al., 2003). Even though industry is an important emission source of NO_2 few models contained industry variables. This is mainly a result of the non-specificity of the CORINE industry variables which does not distinguish type of industry and the selection of monitoring locations.

Direct comparison of values for regression coefficients of similar predictor variables in LUR models of different areas is difficult as the coefficient depends on the other included predictors. In addition, GIS predictor data can also differ between study areas with regards to accuracy and completeness. For example, CORINE data are European wide data but are based on national data, so can have differences between countries.

4.2. Modelling experiences

Because within the ESCAPE project LUR models have been developed in so many areas using the standardized approach several lessons have been learned. We used a standardized and common site selection procedure in all study areas aiming at representing the anticipated spatial variation of air pollution in the included cohort studies, purposely over representing street sites (Cyrys et al., 2012). If the range in measured concentrations is small, then this complicates the development of good LUR models. It is therefore important that the locations of the monitoring sites represent a large diversity of potential sources of air pollution variability such as traffic intensity, population density, industry, etc. It is also important to include the whole range of each characteristic when selecting sites, and not include only the busiest road for example, because this occasionally introduced an influential observation during model development. Predictor variables with a small variation (e.g. many zero values) also were found to be problematic.

The aim was to standardize model development in all study areas. It was however not possible due to differences in availability to standardize the used predictor variables in all areas. This illustrates the need to make European-wide GIS data available and have EU-wide standards for collecting and standardizing these data. Central predictor variables were available for almost all areas, but local predictors differed per study area. Models often included a mixture of central and local predictor variables.

There is no universal standard method for developing LUR models. Within the ESCAPE project we did not use an automatic variable selection method but used a supervised forward stepwise procedure. The procedure included a priori defined predictor variables and a priori defined directions of effect for all predictor variables which were based on basic physical principles. The advantage of using a priori defined directions of effect is that a model could be applied to other study areas and also limits risk of overfitting. Basagaña et al. (2012) recently compared the performance of the ESCAPE procedure, a procedure based upon maximizing the cross-validation R^2 and a deletion–substitution algorithm in hold-out validation. The first two procedures had higher R^2 values than the deletion–substitution algorithm but the differences were very small (difference between first two procedures always <5%, difference with deletion–substitution algorithm with two other procedures <15% when $n => 40$).

4.3. Previous LUR studies

In several of the ESCAPE study areas LUR models using purpose designed monitoring campaigns have been developed in previous studies for cohort exposure assessment (Table 4). The percentage of explained variance of the ESCAPE LUR models is in most areas similar (or higher) compared to these previous LUR models.

The strength of the ESCAPE project is that a common exposure assessment approach was used in all study areas. Two previous multicenter studies also used a common exposure assessment approach: the SAVIAH project (Briggs et al., 1997) and the TRAPCA project (Hoek et al., 2001; Hochadel et al., 2006; Morgenstern et al., 2007). Within the SAVIAH project, in three cities (Amsterdam, Huddersfield and Prague) NO_2 concentrations were measured at 80 sites. The final LUR models explained between 61% and 72% of the observed variability in concentrations between sites (Briggs et al., 1997). In the TRAPCA project NO_2 concentrations were measured at 40 sites in the Netherlands, Munich, Stockholm and Ruhr Area. The LUR models predicted 51–90% of the variation in measured concentrations (Hoek et al., 2001; Hochadel et al., 2006; Morgenstern et al., 2007).

Some of the ESCAPE LUR models have higher percentages of explained variance in concentrations than in previous studies. The main reason is probably that within the ESCAPE study areas more and better GIS predictor variables were available compared with the SAVIAH and TRAPCA project. In recent years the spatial resolution of GIS data has improved considerably and traffic intensity data (linked to digital road networks) become increasingly available, while in previous studies proxies such as road length were used. Missing or inaccurate traffic variables, especially, limit the ability to account for the very local scale concentration differences between monitoring sites.

4.4. Limitations

Model performance was evaluated using a leave-one-out cross-validation method because the total number of sites was too small to have a separate test validation dataset of sufficient size. The R^2 values from the cross-validations were in general only slightly lower (i.e. <10%) than the R^2 values from the LUR models,

Table 4
Description of previously developed NO_x LUR models in ESCAPE study areas (using purpose designed monitoring campaigns).

Area	Pollutant	Number of sites	Year	Measurements	Model R^2	Reference
Oslo, Norway	NO_2	80	2005	Ogawa badges	77%	Madsen et al., 2007
Oslo, Norway	NO_x	80	2005	Ogawa badges	73%	Madsen et al., 2007
Oslo, Norway	NO_2	69	2008	Ogawa badges	74%	Madsen et al., 2011
Oslo, Norway	NO_x	69	2008	Ogawa badges	69%	Madsen et al., 2011
Stockholm, Sweden	NO_2	40	1999–2000	Palmes tubes	85%	Hoek et al., 2001
The Netherlands	NO_2	40	1999–2000	Palmes tubes	85%	Hoek et al., 2001
The Netherlands	NO_2	35	2007	Ogawa badges	86%	Eeftens et al., 2011
Munich, Germany	NO_2	34	1996–1998	Palmes tubes	77%	Carr et al., 2002
Munich, Germany[a]	NO_2	40	1999–2000	Palmes tubes	62%	Hoek et al., 2001
Munich, Germany[b]	NO_2	40	1999–2000	Palmes tubes	51%	Morgenstern et al., 2007
Ruhr Area, Germany	NO_2	40	2002–2003	Palmes tubes	90%	Hochadel et al., 2006
Rome, Italy	NO_2	68	1995–1996	Palmes tubes	69%	Rosenlund et al., 2008
Sabadell, Spain	NO_2	57	2005–2006	Ogawa and Radiello badges	68%	Aguilera et al., 2008
Sabadell, Spain	NO_x	57	2005–2006	Ogawa and Radiello badges	69%	Aguilera et al., 2008
Asturias, Spain	NO_2	67	2006–2007	Radiello badges	52%	Fernández-Somoano et al., 2011
Valencia, Spain	NO_2	93	2004–2005	Radiello badges	81%	Iñiguez et al., 2009

[a] Model restricted to Munich city (size study area ∼310 km^2).
[b] Model covering whole Munich metropolitan area (size study area ∼1200 km^2).

documenting the robustness of the models. Further support for the usefulness of these LUR models to the ESCAPE cohorts is that the root mean squared error was low compared to the range in measured concentrations. A recent study in the Netherlands evaluated the performance of NO_2 land use regression models using internal cross-validation and validation against an independent external dataset and investigated the impact of increasing numbers of training sites (Wang et al., 2012). The R^2 values of LUR models were lower in external dataset validation compared with leave-one out cross-validation, especially if the model was developed for a small number of sites ($N \sim 20$). Another recent similar exercise in Spain showed similar results and that the model R^2 and leave-one-out cross validation R^2 can be inflated for smaller sample sizes, particularly in cases where the number of potential predictor variables is large (Basagaña et al., 2012). Bayak (2004) suggested as a rule of thumb to have at least 10 observations per predictor variable in the model, implying four predictors per model. Most models had 4 or less predictors, but a few models contained more variables and are likely over-specified. Wang et al. (2012) recently showed that hold-out validation did not improve when these large models were constrained to contain fewer variables. An evaluation of the possibility to develop models for combined study areas resulting in a larger number of monitoring sites is ongoing.

Site characteristics such as street configuration, traffic speed etc (Brauer et al., 2003) might improve LUR models. Such data are, however, not usually available. In some areas (Kaunas, Bradford, Manchester, Netherlands and Belgium, Grenoble and Lugano) information about the vehicle fleet, differentiating light-duty and heavy-duty traffic, were available and were included in the final models.

5. Conclusion

Within the ESCAPE project it was possible to develop LUR models using a standardized approach that explained a large fraction of the spatial variance in measured annual average NO_2 and NO_x concentrations. Results showed that it is especially important to have accurate local traffic intensity data as predictor variables available and evaluate influential observations in LUR model development. These LUR models are being used to estimate outdoor concentrations at the home addresses of participants in over 30 cohort studies included in the ESCAPE project.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007–2011) under grant agreement number: 211250. The funding source had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the manuscript for publication.

We thank all those who are responsible for data management, model development and supervision in all study areas and especially: Sally Liu (passed away in June 2011), Kees Meliefste, Marieke Oldenwening, Marjan Tewis, Geert de Vrieze, Christophe Ampe, Erwan Gremaud, Patricia Lozano, Alexandre Armengaud, Sabrina Pontet, Francesco Forastiere, Eleonora Zirro, Claudia Galassi, Maria Morataya, Jesus Martinez Moratalla, Angelica Blanco, Vanessa Hernández, Nerea Muniozguren, Paco Farruquito, Felix Payo, Rocio Capelo, Jose Antonio Maldonado Perez, Ana Esplugues, Marisa Estarlich, Ferran Ballester, Mario Murcia, Sabrina Llop, Amparo Cases, Carmen Iñíguez, Daniel Armando Torres, Ferran Ballester, Carmen Freire, Marieta Fernández, Sara María Álvarez Avellón, Ana Souto García, María Felicitas López Cima, Ana Fernández Somoano, Patricia González Arriaga, Avelino Menéndez Crispín, Beatriz Lastra Díaz, M. Cristina Arias Díaz, Ana Fernández Somoano, Adonina Tardón García Mikel Ayerdi, Maria Dolores Martínez López di Castillo, Enara Maíz González, Aitana Lertxundi Manterola, Aitana Lertxundi Manterola, Laura Bouso, Anna Ripoll, Glòria Carrasco, Meritxell Portella, Jaume Matamala, Sònia Alvarez, Jordi Sunyer, Judith Garcia, Michelle Mendez, Xavier Basagaña, Mercè Medina, Anna Schembari, David Martinez, David Rojas, Kyriaki Papantoniou, Irene Huerta and Pablo Cerralbo.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.atmosenv.2013.02.037.

References

Aguilera, A., Sunyer, J., Fernandez-Patier, R., Hoek, G., Aguirre-Alfaro, A., Meliefste, K., et al., 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, 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.
Bayak, M.A., 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosom. Med. 66, 411–421.
Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., De Hoogh, K., Briggs, D., 2009. Mapping of background air pollution at a fine spatial scale across the European Union. Sci. Total. Environ. 407, 1852–1867.
Beelen, R., Hoek, G., van den Brandt, P.A., Goldbohm, R.A., Fischer, P., Schouten, L.J., et al., 2008. Long-term effects of traffic-related air pollution on mortality in a Dutch cohort. Environ. Health Perspect. 116, 196–202.
Brauer, M., Hoek, G., van Vliet, P., Meliefste, K., Fischer, P., Gehring, U., et al., 2003. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. Epidemiology 14, 228–239.
Briggs, D., Collins, S., Elliot, P., Fischer, P., Kingham, S., Lebret, E., et al., 1997. Mapping urban air pollution using GIS: a regression-based approach. Int. J. Geogr. Inf. Sci. 11, 699–718.
Briggs, D., Gulliver, J., Fecht, D., Vienneau, D., 2007. Dasymetric modelling of small-area population distribution using land cover and light emissions data. Remote. Sens. Environ. 108, 451–466.
Brunekreef, B., Holgate, S.T., 2002. Air pollution and health. Lancet 360, 1233–1242.
Carr, D., von Ehrenstein, O., Weiland, S., Wagner, C., Wellie, O., Nicolai, T., et al., 2002. Modeling annual benzene, toluene, NO_2, and soot concentrations on the basis of road traffic characteristics. Environ. Res. 90, 111–118.
Cyrys, J., Eeftens, M., Heinrich, J., Ampe, C., Armengaud, A., Beelen, R., et al., 2012. Variation of NO_2 and NO_x concentrations between and within 36 European study areas: results of the ESCAPE project. Atmos. Environ. 62, 374–390.
EEA, 2005. CORINE Land Cover 2000 (CLC2000) 100 m – Version 8/2005 Version 2. Copenhagen.
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, 765–770.
Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., et al., 2012a. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM_{coarse} in 20 European study areas; results of the ESCAPE project. Environ. Sci. Technol. 46, 11195–11205.
Eeftens, M., Tsai, M.-Y., Ampe, C., Anwander, B., Beelen, R., Bellander, T., et al., 2012b. Variation of $PM_{2.5}$, PM_{10}, $PM_{2.5}$ absorbance and PM_{coarse} concentrations between and within 20 European study areas – results of the ESCAPE project. Atmos. Environ. 62, 303–317.
Fernández-Somoano, A., Estarlich, M., Ballester, F., Fernández-Patier, R., Aguirre-Alfaro, A., Herce-Garraleta, M.D., et al., 2011. Outdoor NO_2 and benzene exposure in the INMA (Environment and Childhood) Asturias cohort (Spain). Atmos. Environ. 45, 5240–5246.
Gallego, F.J., 2010. A population density grid of the European Union. Popul. Environ. 31, 460–473.
Gilbert, N.L., Woodhouse, S., Stieb, D.M., Brook, J.R., 2003. Ambient nitrogen dioxide and distance from a major highway. Sci. Total. Environ. 312, 43–46.
Hochadel, M., Heinrich, J., Gehring, U., Morgenstern, V., Kuhlbusch, T., Link, E., et al., 2006. Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information. Atmos. Environ. 40, 542–553.
Hoek, G., Meliefste, K., Brauer, M., van Vliet, P., Brunekreef, B., Fischer, P., et al., 2001. Risk Assessment of Exposure to Traffic-related Air Pollution for the Development of Inhalant Allergy, Asthma and Other Chronic Respiratory Conditions in Children (TRAPCA). Final Report 2001.
Hoek, G., Beelen, R., De Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., et al., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42, 7561–7578.
Iñiguez, C., Ballester, F., Estarlich, M., Llop, S., Fernandez-Patier, R., Aguirre-Alfaro, A., et al., 2009. Estimation of personal NO_2 exposure in a cohort of pregnant women. Sci. Total. Environ. 407, 6093–6099.
Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., et al., 2005. A review and evaluation of intraurban air pollution exposure models. J. Expo. Anal. Environ. Epidemiol. 15, 185–204.
Madsen, C., Carlsen, K.C.L., Hoek, G., Oftedal, B., Nafstad, P., Meliefste, K., et al., 2007. Modeling the intra-urban variability of outdoor traffic pollution in Oslo, Norway-A GA2LEN project. Atmos. Environ. 41, 7500–7511.
Madsen, C., Gehring, U., Håberg, S.E., Nafstad, P., Meliefste, K., Nystad, W., et al., 2011. Comparison of land-use regression models for predicting spatial NO_x contrasts over a three year period in Oslo, Norway. Atmos. Environ. 45, 3576–3583.
Marshall, J.D., Nethery, E., Brauer, M., 2008. Within-urban variability in ambient air pollution: comparison of estimation methods. Atmos. Environ. 42, 1359–1369.
Morgenstern, V., Zutavern, A., Cyrys, J., Brockow, I., Gehring, U., Koletzko, S., et al., 2007. Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. Occup. Environ. Med. 64, 8–16.
Pope 3rd, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., et al., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287, 1132–1141.
Rijnders, E., Janssen, N.A., van Vliet, P.H., Brunekreef, B., 2001. Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density. Environ. Health Perspect. 109 (Suppl. 3), 411–417.
Rosenlund, M., Forastiere, F., Stafoggia, M., Porta, D., Perucci, M., Ranzi, A., et al., 2008. Comparison of regression models with land-use and emissions data to predict the spatial distribution of traffic-related air pollution in Rome. J. Expo. Anal. Environ. Epidemiol. 18, 192–199.
Su, J., Jerrett, M., Beckerman, B., 2009. A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. Sci. Total. Environ. 407, 3890–3898.
Van Roosbroeck, S., Wichmann, J., Janssen, N.A., Hoek, G., van Wijnen, J.H., Lebret, E., et al., 2006. Long-term personal exposure to traffic-related air pollution among school children, a validation study. Sci. Total. Environ. 368, 565–573.
Vienneau, D., de Hoogh, K., Beelen, R., Fischer, P., Hoek, G., Briggs, D., 2010. Comparison of land-use regression models between Great Britain and the Netherlands. Atmos. Environ. 44, 688–696.
Von Klot, S., 2011. Equivalence of using nested buffers and concentric adjacent rings as predictors in land use regression models. Atmos. Environ. 45, 4108–4110.
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, 4481–4489.