Review

A review of land-use regression models to assess spatial variation of outdoor air pollution

Gerard Hoeka,*, Rob Beelen a, Kees de Hooghb, Danielle Vienneaub, John Gulliverc, Paul Fischerd, David Briggsb

a Institute for Risk Assessment Sciences (IRAS), PO Box 80178, 3508 TD Utrecht, The Netherlands
b Imperial College, Department Epidemiology Public Health Norfolk Place, London, W2 1PG, UK
c University of the West of Scotland, Paisley, UK
d National Institute of Public Health and the Environment, P.O. Box 1, 3720 BA Bilthoven, The Netherlands

Article info

Article history:
Received 27 February 2008
Received in revised form 23 May 2008
Accepted 29 May 2008

Keywords:
Land use regression
Spatial variation
\(\text{NO}_2\)
Particulate matter
Air pollution

A B S T R A C T

Studies on the health effects of long-term average exposure to outdoor air pollution have played an important role in recent health impact assessments. Exposure assessment for epidemiological studies of long-term exposure to ambient air pollution remains a difficult challenge because of substantial small-scale spatial variation. Current approaches for assessing intra-urban air pollution contrasts include the use of exposure indicator variables, interpolation methods, dispersion models and land-use regression (LUR) models. LUR models have been increasingly used in the past few years. This paper provides a critical review of the different components of LUR models.

We identified 25 land-use regression studies. Land-use regression combines monitoring of air pollution at typically 20–100 locations, spread over the study area, and development of stochastic models using predictor variables usually obtained through geographic information systems (GIS). Monitoring is usually temporally limited: one to four surveys of typically one or two weeks duration. Significant predictor variables include various traffic representations, population density, land use, physical geography (e.g. altitude) and climate.

Land-use regression methods have generally been applied successfully to model annual mean concentrations of \(\text{NO}_2\), \(\text{NO}_x\), \(\text{PM}_{2.5}\), the soot content of \(\text{PM}_{2.5}\) and VOCs in different settings, including European and North-American cities. The performance of the method in urban areas is typically better or equivalent to geo-statistical methods, such as kriging, and dispersion models.

Further developments of the land-use regression method include more focus on developing models that can be transferred to other areas, inclusion of additional predictor variables such as wind direction or emission data and further exploration of focalsum methods. Models that include a spatial and a temporal component are of interest for (e.g. birth cohort) studies that need exposure variables on a finer temporal scale. There is a strong need for validation of LUR models with personal exposure monitoring.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

A large number of epidemiological studies have shown that current day outdoor air pollution is associated with significant adverse effects on public health (Brunekeef and
Pollutants of health concern at current day concentration levels in developed countries include particulate matter (PM), nitrogen dioxide ($\text{NO}_2$) and ozone (Brunekreef and Holgate, 2002). Time series studies have found that day-to-day changes in PM concentrations, in particular, are related to changes in hospital admissions and mortality (Katsouyanni et al., 2001; Samet et al., 2000). The relative risks in the time series studies are generally small: for example, in a large European study by Katsouyanni et al. (2001) mortality increased by 0.5% with an increase of 10 $\mu$g/m$^3$ of the 24-h average concentration of PM$_{2.5}$. In 1993 a prospective cohort study in six US cities documented an association between long-term average exposure to outdoor air pollution and reduced survival, after careful control for other individual risk factors such as smoking (Dockery et al., 1993). Mortality rates in the most polluted city were 26% higher than in the least polluted city; the difference in annual average PM$_{2.5}$ concentration between these cities was 19 $\mu$g/m$^3$. Several other studies have subsequently found associations between mortality from cardiovascular and respiratory diseases and long-term average exposure to air pollution (Pope and Dockery, 2006). In general, such long-term air pollution exposure studies have played an important role in recent health impact assessments and in the debate about pollution exposure studies have played an important role in recent health impact assessments and in the debate about new air quality guidelines for Europe (Kunzli et al., 2000).

Exposure assessment for epidemiological studies of long-term exposure to ambient air pollution remains a difficult challenge. The first cohort studies published in the mid-1990s have compared mortality rates between cities, with exposure characterized by the average concentration measured at a central site within each city (Dockery et al., 1993; Pope et al., 1995). In the past decade, various studies have documented significant variation of outdoor air pollution at a small scale within urban areas for important pollutants such as NO$_2$ and black smoke (e.g. Fischer et al., 2000; Kingham et al., 2000; Lebret et al., 2000; Monn, 2001; Jerrett et al., 2005; Zhu et al., 2002). In some settings the within-city spatial contrast may be as large as the between-city contrast. There is evidence from epidemiological studies that within-city contrasts of particulate matter air pollution are associated with larger contrasts than between-city (Miller et al., 2007). Epidemiological studies therefore need to take these contrasts into account. Monitoring alone will generally not be feasible, as the study population of epidemiological studies generally comprises several hundreds to thousands of subjects, living or working at different places. An additional complication for monitoring is that only long-term (i.e. annual) average concentrations are useful for the epidemiological study, so that multiple daily or weekly samples have to be collected.

Current approaches that have been developed to meet the challenge of assessing intra-urban air pollution contrasts have recently been reviewed (Briggs, 2005; Jerrett et al., 2005). Approaches include the use of exposure indicator variables (e.g. traffic intensity at the residential address or distance to a major road), interpolation methods (e.g. kriging, inverse distance weighing), conventional dispersion models and land-use regression models. Application of the land-use regression approach for air pollution mapping was introduced in the SAVIAH (Small Area Variations In Air quality and Health) study (Briggs et al., 1997). Land-use regression combines monitoring of air pollution at a small number of locations and development of stochastic models using predictor variables usually obtained through geographic information systems (GIS). The model is then applied to a large number of unsampled locations in the study area. The technique was initially termed regression mapping (Briggs et al., 1997). Regression mapping is probably more descriptive of the methodology as the predictor variables are not only representative of land use. Other variables such as altitude and meteorology, for example, are often included in the models. As most researchers currently refer to the method as land use regression (LUR), however, we will also use this term. There are some earlier examples of the method in environmental science (Briggs et al., 1997). In 1985 interpolation of sulfate deposition data from the USA was supplemented with a drift term using geographical coordinates (Bilonick, 1985).

After the successful pioneering work in SAVIAH, LUR methods have been increasingly used in epidemiological studies in the past decade (Briggs, 2005). Developments in GIS have contributed to the popularity of LUR methods. Initially the approach was mainly adopted in Europe, but in the past few years several applications in North America have been published (e.g. Gilbert et al., 2005; Ross et al., 2006, 2007). While most studies have developed models that explain spatial air pollution contrasts satisfactorily, the predictive models differ substantially between the studies. Although this may be due to true differences between locations, we believe that differences in the application of the approach and selection of variables also play an important role.

The goal of this paper is therefore to review the various elements of the approach by discussing studies applying LUR methods. After listing the studies identified through a systematic review, we structure the review according to the main components of LUR: monitoring data, geographic predictors and model development and validation. We will compare the validity of the LUR models to alternative quantitative approaches especially dispersion modelling, and conclude with a discussion of limitations and new developments. A short review of LUR models has been published before (Ryan and LeMasters, 2007). The review identified six studies by a search through June 2006 and had a substantially narrower scope than the current manuscript.

1.1. Literature search

We performed a systematic literature search in Pubmed and Science Direct to trace studies using land-use regression approaches. The final search was conducted on January 15 2008. We used the search terms “land use regression”, “GIS air pollution”, “regression mapping” and “air pollution stochastic”. This was supplemented by papers included in the reference lists of the traced papers and papers that were already known to us based upon previous exposure assessment and epidemiological studies. We only included papers in the English, German and Dutch language.
2. Identified studies

We identified 25 land-use regression studies. Table 1 lists some key characteristics of the design of the studies we identified. Tables 2–5 outline the performance of, and predictor variables for, the final LUR models. Most applications have been limited to nitrogen dioxide (NO\textsubscript{2}), largely because of the ease of monitoring of this pollutant (Table 2). Fewer studies have developed models for NO or NO\textsubscript{x} (Table 3), particulate matter (PM\textsubscript{2.5}) or the elemental carbon content of particulate matter (Table 4) and VOCs including benzene and toluene (Table 5).

The SAVIAH study was the first to use land use regression to model small scale variations in air pollution (Briggs et al., 1997). The aim of the study was to generate individual-level indicators of long-term average exposure to ambient air pollution to assess the risks of respiratory disease of children. As the study population involved several thousand children, monitoring ambient air pollution at the home addresses was not feasible. Instead, in each of the three cities included in the study (Amsterdam, Huddersfield and Prague), a purpose-designed monitoring network of 80 monitoring sites was established to ensure a sufficiently dense network of monitoring stations. In none of the cities did a sufficiently dense routine monitoring network exist. At each site, the NO\textsubscript{2} concentration was measured for 14 days in each season with passive samplers compiled in a GIS. The variables were then calculated for various buffers and linear regression was used to develop a model that explained the largest fraction of observed variability in annual average NO\textsubscript{2} concentration. The model was constrained by requiring that all regression coefficients have the a priori defined sign (e.g. positive for traffic and negative for altitude). The final prediction models explained between 61% and 72% of the observed variability in concentrations between sites (Briggs et al., 1997).

Slightly different models were derived for each city due to differences between the cities, as well as differences in data availability (Table 2). Altitude, for example, was not a predictor in Amsterdam due to the flat terrain in this city; traffic intensity was a predictor in the Huddersfield and Prague models, but in Amsterdam was replaced by length of road by type, due to differences in data availability. Application of the models to validation sites not used in model development resulted in similar $R^2$ values, demonstrating the robustness of the models. In all three cities, LUR performed substantially better than spatial interpolation methods such as kriging, TIN-contouring and trend surface analysis (Briggs et al., 2000). In urban areas, spatial variability is characterized more by local sources such as major roads than as a smoothly varying concentration field, as assumed in spatial interpolation. In Huddersfield the regression model predicted measured concentrations at validation sites better than the CALINE-3 dispersion model (Briggs et al., 2000).

The method has subsequently been applied in a variety of settings, including Europe and more recently North America. Most studies were performed in a large urban area, sometimes including the surrounding smaller communities (Table 1). Three studies have applied the method to entire countries, specifically the Netherlands and the UK (Stedman et al., 1997; Hoek et al., 2001a,b; Beelen et al., 2007) while the APMOSPHERE project modelled concentrations on a 1 × 1 km scale for the EU-15 (Briggs et al., 2005).

Of the 25 identified studies, 12 studies mentioned that the final model was used in a specific identified epidemiological study (Aguilera et al., 2008; Beelen et al., 2007; Brauer et al., 2003; Briggs et al., 1997; Gilbert et al., 2005; Hoek et al., 2001a,b; Jerrett et al., 2007; Morgenstern et al., 2007; Ryan et al., 2007; Smith et al., 2006; Hochadel et al., 2006; Wheeler et al., 2008). Most other studies mentioned epidemiological studies as a rationale for modelling.

3. Monitoring data

Studies differ in the monitoring data that are used to develop land use regression models. Important aspects are the use of routine versus purpose-designed networks, monitored pollutant, the number and distribution of monitoring sites and temporal resolution.

3.1. Routine versus purpose-designed monitoring

A limited number of studies have made use of air pollution monitoring data from routine networks (Stedman et al., 1997; Hoek et al., 2001a,b; Beelen et al., 2007; Briggs et al., 2005; Moore et al., 2007; Ross et al., 2007; Briggs et al., in press). Most studies, however, have undertaken monitoring specifically for the purpose of model development as routine networks in most urban areas are not dense enough to enable meaningful modelling (Table 6) of small-scale variability of outdoor air pollution. A further advantage of purpose-designed monitoring is the control the investigators have over the type of sites (e.g. traffic, background) they wish to include in model development. Disadvantages of purpose-designed monitoring include the additional cost (discussed below) and the limited temporal coverage of the measurements. In the studies to date, most purpose-designed monitoring campaigns consisted of between one and four 7–14 days sampling campaigns, whereas routine monitoring is typically continuous, especially for the gaseous components. When routine monitoring data are used, however, careful attention must be paid to the site type as routine monitors are often designed to monitor compliance with regulatory standards rather than human exposures. As a result, routine networks are often focused at potential hotspots such as heavily trafficked street locations or industrial areas, and may consequently give biased estimates of pollution levels in areas where people live. Siting of monitors may differ substantially between countries: for example, in a paper from Canada routine network monitors were seen to be preferentially placed away from hotspots (Marshall et al., 2008).

In purpose-designed studies, NO\textsubscript{2}, NO, NO\textsubscript{x} and VOCs are generally measured with passive samplers, whereas PM is...
<table>
<thead>
<tr>
<th>Reference</th>
<th>Study area</th>
<th>Pollutants</th>
<th>No. of monitoring sites</th>
<th>Selection of sites</th>
<th>Sampling period</th>
<th>Predictor variables collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Briggs et al. (1997)</td>
<td>Amsterdam (NL)</td>
<td>NO(_2)</td>
<td>80</td>
<td>Informal*</td>
<td>Four 2-week periods between June 1993 and June 1994</td>
<td>Traffic, land cover, altitude, sampling height</td>
</tr>
<tr>
<td></td>
<td>Huddersfield (UK)</td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td>Urban and suburban land cover, NO(_x) emission from major road vehicle sources</td>
</tr>
<tr>
<td></td>
<td>Prague (CZ)</td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steedman et al. (1997)</td>
<td>UK</td>
<td>NO(_2) and NO(_x)</td>
<td>37</td>
<td>All national monitoring network sites</td>
<td>Annual average 1994</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sheffield (UK)</td>
<td></td>
<td>28</td>
<td></td>
<td>Five 2-week 1997–1998</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Northampton (UK)</td>
<td></td>
<td>35</td>
<td></td>
<td>Six 2-week 1997–1998</td>
<td></td>
</tr>
<tr>
<td>Carr et al. (2002)</td>
<td>Munich, Germany</td>
<td>NO(_2), soot, VOC</td>
<td>34</td>
<td>Eighteen high traffic sites and 16 schools</td>
<td>Average December 1996–February 1998</td>
<td>Manually counted traffic intensity and percentage traffic jam</td>
</tr>
<tr>
<td>Brauer et al. (2003), Hoek et al. (2001b)</td>
<td>Netherlands</td>
<td>PM(_2.5), soot, NO(_2)</td>
<td>40</td>
<td>Informal*</td>
<td>Four 2-week periods between February 1999 and July 2000</td>
<td>Population and household density, traffic (traffic intensity, length of high traffic roads, distance to roads) and region</td>
</tr>
<tr>
<td></td>
<td>Munich, Germany</td>
<td></td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stockholm, Sweden</td>
<td></td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gilbert et al. (2005)</td>
<td>Montreal, Canada</td>
<td>NO(_2)</td>
<td>67</td>
<td>Population-weighted location-allocation model</td>
<td>One 2-week period May 2003</td>
<td>Distance and traffic counts nearest highway, area of open space and population density</td>
</tr>
<tr>
<td>Kanaroglou et al. (2005)</td>
<td>Toronto, Canada</td>
<td>NO(_2)</td>
<td>100</td>
<td>Population-weighted location-allocation model</td>
<td>One 2-week period September 2002</td>
<td>Distance to the nearest expressway traffic in buffers; park, open, recreational, or water body land in buffer; household density</td>
</tr>
<tr>
<td>Gonzalez et al. (2005)</td>
<td>El Paso, Texas, USA</td>
<td>NO(_2)</td>
<td>20 Schools</td>
<td>Informal*</td>
<td>One 7-day period, February 11–18 1999</td>
<td>Distance to US-Mexico border, distance highway and altitude</td>
</tr>
<tr>
<td>Smith et al. (2006)</td>
<td>El Paso, Texas, USA</td>
<td>NO(_2), VOC</td>
<td>Twenty-two schools</td>
<td>Informal*</td>
<td>Two 7-day periods, November–December 1999</td>
<td>Distance to US-Mexico border, distance highway and major roads, traffic intensity, altitude, population density, distance to petroleum facility</td>
</tr>
<tr>
<td>Briggs et al. (2005)</td>
<td>European Union (15 countries)</td>
<td>NO(<em>2), PM(</em>{10}) and SO(_2)</td>
<td>98–459</td>
<td>All background National Monitoring Network sites from AIRBASE</td>
<td>Annual average 2001</td>
<td>Altitude, distance to sea, climate, traffic, % non-residential urban land cover, % high density residential land cover, topography, % agriculture land cover</td>
</tr>
<tr>
<td>Ross et al. (2006)</td>
<td>San Diego County, California, US</td>
<td>NO(_2)</td>
<td>39</td>
<td>Informal*</td>
<td>One 13–16 days period in October 2003</td>
<td>Traffic intensity, length of road, distance to Pacific coast</td>
</tr>
<tr>
<td>Hochadel et al. (2006)</td>
<td>North Rhine-Westphalia, Germany</td>
<td>NO(_2), PM(_2) absorbance, PM(_2.5)</td>
<td>40</td>
<td>Informal*</td>
<td>Four 2-week periods in March 2002–March 2003</td>
<td>Traffic intensity, number of buildings, distance to major roads</td>
</tr>
<tr>
<td>Sahsuvaroglu et al. (2006)</td>
<td>Hamilton, Canada</td>
<td>NO(_2)</td>
<td>107</td>
<td>Population-weighted location-allocation model</td>
<td>One 2-week period in October 2002</td>
<td>Land use, traffic, population density, physical geography, meteorology</td>
</tr>
<tr>
<td>Beelen et al. (2007), Hoek et al., 2001a</td>
<td>Netherlands</td>
<td>NO(_2), NO, and BS</td>
<td>16–36</td>
<td>All background National Monitoring Network sites</td>
<td>Average 1976–1996</td>
<td>Region, number of inhabitants in different buffer sizes, land use, traffic</td>
</tr>
<tr>
<td>Ross et al. (2007)</td>
<td>New York City (all 28 and 9 urbanized counties)</td>
<td>PM(_2.5)</td>
<td>36–62</td>
<td>All routine Monitoring Network sites</td>
<td>3-Year averages 1999–2001 and average winter 2000</td>
<td>Traffic, population, land use and total county primary PM(_2.5) emissions</td>
</tr>
<tr>
<td>Ryan et al. (2007)</td>
<td>Cincinnati, Ohio, US</td>
<td>Elemental carbon</td>
<td>24</td>
<td>Informal*</td>
<td>December 2001–December 2004 (total 609 samples)</td>
<td>Altitude, number of trucks, length of bus routes in buffers, land use</td>
</tr>
</tbody>
</table>
typically measured with active samplers. Passive samplers that have been used to monitor NO2 include the Palmes tube (Briggs et al., 1997; Stedman et al., 1997; Brauer et al., 2003; Lewne et al., 2004; Ross et al., 2006) and the Ogawa badge (Gilbert et al., 2005; Kanaroglou et al., 2005; Sahsuvaroglu et al., 2006; Jerrett et al., 2007; Madsen et al., 2007; Henderson et al., 2007; Aguilera et al., 2008). Some studies that used the Ogawa badge also measured nitrogen oxides in the form of NO: this should represent primary emissions from combustion sources such as motorized traffic better than NO2, which has a significant secondary component (Madsen et al., 2007; Henderson et al., 2007; Aguilera et al., 2008).

Co-location of passive samplers at a few sites with continuous NOX monitors has generally shown good agreement, but it remains important to include co-location in each new study. The two studies that involved specific PM monitoring used Harvard impactors which are low volume active samplers (Brauer et al., 2003; Hochadel et al., 2006). Elemental carbon was measured with a variety of approaches, including the conventional black smoke (BS) method (Brauer et al., 2003; Ryan et al., 2007) and thermal techniques (Carr et al., 2002; Ryan et al., 2007). Several studies have documented the very high correlation of these metrics (Cyrys et al., 2003).

Because of their more reliable limitations, passive samplers have been deployed in duplicate at each site in some studies (Briggs et al., 1997; Ross et al., 2006; Jerrett et al., 2007). The low cost compared to active sampling allows duplicate sampling. The main advantage of duplicate sampling lies in detection of erroneous samples, and offering some information on measurement uncertainties, rather than increasing precision. Precision of a single measurement of NO2 determined from duplicate samples is typically 5–10%, which is acceptable.

Costs of passive sampling for NOX (n = 40 sites, four surveys) are in the order of 10–12,000 Euro (Table 7). Costs of the same survey for active PM sampling are higher, up to 30,000 Euro, assuming that equipment is available (Table 7). Actual costs are dependent on the setting (e.g. ease of selecting monitoring locations) and salary rates. In addition, application of LUR involves costs of collecting and calculating the GIS variables or stochastic modelling, together with software or data licences. Nevertheless the overall costs of exposure assessment using LUR approaches are modest given typical budgets for large scale epidemiological studies.

3.2. Number and distribution of monitoring sites

There is no rigorous methodology to determine the required number of monitoring locations given a certain study objective and setting. Published studies have included between 20 and 100 sites, with the lower range representing those studies that modelled PM using routine monitoring data. Probably 40–80 sites is a reasonable number to choose for site-specific monitoring, but the size of the population and city should be taken into account to determine the actual number. Madsen et al. (2007) reported that models developed from a random selection of 40 sites in the Oslo urban area were indistinguishable from those developed using the full set of 80 monitoring sites. Oslo is a predominantly non-industrial
Table 2: Performance of land-use regression models for NO₂

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study area</th>
<th>Predictor variables in final model</th>
<th>( R^2 ) of model</th>
<th>( R^2 ) validation</th>
<th>RMSE validation ( (\mu g , m^{-3}) )</th>
<th>Measured concentrations ( (\mu g , m^{-3}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Briggs et al. (1997)</td>
<td>Amsterdam</td>
<td>Length major roads 50, 200, 350 m + Distance major road + built up land, 100 m</td>
<td>0.62</td>
<td>0.79</td>
<td>4.5(^a)</td>
<td>44 (12–73)</td>
</tr>
<tr>
<td></td>
<td>Huddersfield</td>
<td>Traffic volume 300 m + land cover factor, 300 m + altitude + sampling height</td>
<td>0.61</td>
<td>0.82</td>
<td>3.7</td>
<td>30 (10–88)</td>
</tr>
<tr>
<td></td>
<td>Prague</td>
<td>Traffic volume, 60 m + traffic volume 60–120 m + land cover factor 60 m + altitude</td>
<td>0.72</td>
<td>0.87</td>
<td>4.7</td>
<td>32 (6–83)</td>
</tr>
<tr>
<td>Briggs et al. (2000)</td>
<td>Huddersfield</td>
<td>Traffic density in two circular buffers (0–40 m and 40–300 m) weighted to produce a weighted traffic volume factor within a 300 m buffer + high density housing in a circular buffer (0–300 m) + altitude (log)</td>
<td>0.51</td>
<td>0.76</td>
<td>9.8(^a)</td>
<td>50 (SD 16)</td>
</tr>
<tr>
<td></td>
<td>Sheffield</td>
<td></td>
<td>0.61</td>
<td>0.73</td>
<td>7.7</td>
<td>46 (SD 15)</td>
</tr>
<tr>
<td></td>
<td>Northampton</td>
<td></td>
<td>0.60</td>
<td>0.58</td>
<td>5.5</td>
<td>27 (SD 9)</td>
</tr>
<tr>
<td>Stedman et al. (1997)</td>
<td>UK</td>
<td>Rural NO₂ + urban land cover 25 km² + urban land cover 100 + NOₓ emission major road vehicles 4 km²</td>
<td>0.97(^b)</td>
<td>0.36</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Carr et al. (2002)</td>
<td>Munich, Germany</td>
<td>Traffic intensity 50 m + traffic intensity 50–300 m + traffic jam 50 m + traffic jam 50–300 m</td>
<td>0.77</td>
<td>NA</td>
<td>NA</td>
<td>50 (25–85)</td>
</tr>
<tr>
<td>Hoek et al. (2001b)</td>
<td>Netherlands</td>
<td>High traffic roads 250 m + address density 300 m + region + medium traffic roads 1000 m + address density 4000 m</td>
<td>0.85</td>
<td>–</td>
<td>4.7</td>
<td>29 (12–51)</td>
</tr>
<tr>
<td></td>
<td>Munich</td>
<td>Traffic intensity 50 m + traffic intensity 50–250 m + address density 300 m + address density 300 m + address density 300–5000 m</td>
<td>0.62</td>
<td>–</td>
<td>6.1</td>
<td>29 (16–51)</td>
</tr>
<tr>
<td>Gonzales et al. (2005)</td>
<td>El Paso</td>
<td>Distance to US–Mexico border + distance highway + elevation</td>
<td>0.81</td>
<td>NR</td>
<td>NR</td>
<td>~40 (22–75)</td>
</tr>
<tr>
<td>Gilber et al. (2005)</td>
<td>Montreal</td>
<td>Distance highway + traffic count nearest highway + length highways 100 m + length major roads 100 m + length minor roads 500 m + open space 100 m + population density 2000 m</td>
<td>0.55</td>
<td>0.52</td>
<td>–</td>
<td>23 ± 6 (10–42)(^c)</td>
</tr>
<tr>
<td>Smith et al. (2006)</td>
<td>El Paso</td>
<td>Elevation + traffic intensity 1000 m + population density + distance to border + distance petroleum facility</td>
<td>&gt;-0.90(^d)</td>
<td>–</td>
<td>~5.6</td>
<td>~44 (22–74) (^e)</td>
</tr>
<tr>
<td>Briggs et al. (2005)</td>
<td>EU</td>
<td>Altitude + distance to sea + climate factor + traffic + % non-residential urban land cover + % high density residential land cover + topeX + % agriculture land cover</td>
<td>–</td>
<td>0.61</td>
<td>6.7</td>
<td>10 ± 6</td>
</tr>
<tr>
<td>Hochadel et al. (2006)</td>
<td>North Rhine-Westphalia</td>
<td>Heavy vehicle traffic 250–10,000 m + total traffic 250 m + maximum traffic 100 m + building density 1500 × 1500 m</td>
<td>0.90</td>
<td>–</td>
<td>2.7</td>
<td>26 (19–46) (^f)</td>
</tr>
<tr>
<td>Ross et al. (2006)</td>
<td>San Diego County</td>
<td>Traffic density, 40–300 m + traffic density, 300–1000 m + road length, 40 m + distance Pacific coast</td>
<td>0.79</td>
<td>NR</td>
<td>NR</td>
<td>30 ± 9 (16–57)(^e)</td>
</tr>
<tr>
<td>Sahsuvaroglu et al. (2006)</td>
<td>Hamilton</td>
<td>Traffic density 300 m buffer + industrial land use, 200 m + open space land use, 500 m + &lt;1500 m downwind highway 403 + within 50 m highways + distance Lake Ontario + downtown industrial core, 1000 m</td>
<td>0.76</td>
<td>NR</td>
<td>NR</td>
<td>29 ± 8 (16–56)(^e)</td>
</tr>
<tr>
<td>Beelen et al. (2007)</td>
<td>Netherlands</td>
<td>Region + population 5000 m + urban land cover + city center + traffic intensity 100 m</td>
<td>0.84</td>
<td>NR</td>
<td>NR</td>
<td>–</td>
</tr>
<tr>
<td>Morgenstern et al. (2007)</td>
<td>Munich</td>
<td>Length state roads, 1000–2500 m + land cover factor, 100–250 m + land cover class + household density, 2500–5000 m + length rural roads, 250–500 m</td>
<td>0.51</td>
<td>–</td>
<td>9.51</td>
<td>29 (16–51)</td>
</tr>
<tr>
<td>Rosenlund et al. (2008)</td>
<td>Rome</td>
<td>Traffic zone + distance to busy roads + size census block + inverse population density + altitude</td>
<td>0.69</td>
<td>0.75</td>
<td>–</td>
<td>47 ± 10 (24–73)</td>
</tr>
</tbody>
</table>
Madsen et al. (2007) Oslo Altitude

Jerrett et al. (2007) Toronto Expressway, 200 m

Henderson et al. (2007) Vancouver Length expressway, 100 m

Aguilera et al. (2008) Sabadell Altitude

Wheeler et al. (2008) Windsor Distance to Ambassador bridge

Measured concentrations are mean ± standard deviation, with minimum and maximum in parentheses. RMSE, root mean squared error; NA, not available; NR, tabulated statistics not reported, but validation procedures performed resulting in typically small prediction errors.

Values are standard errors of the estimate.

Assuming 1 ppb = 2 lmg⁻¹.

Random subsets of 40 locations were drawn from the 80 sites (40 training and 40 validation samples).

Assumed 1 mg⁻¹.

Median (min–max).

city of about 500,000 inhabitants located near the sea and with significant altitude differences. In Toronto, Canada, Jerrett et al. (2007) observed that models developed from random selections of 65 sites were very similar to a model developed for all 94 monitoring sites. Toronto is a city of 2.6 million people located on the shore of Lake Ontario and a study area of 633 km². In London, LUR models derived for 75% of the PM10 sites were very similar (Briggs et al., in press). Whether these experiences are valid for other cities, in other types of environment, is unclear.

Fewer sites are most likely needed to transfer a model developed elsewhere to the study area of interest, as shown by Briggs and co-workers in the SAVIAH study. In this example, the model developed for Huddersfield was successfully applied in Huddersfield the following year and to three other UK cities, after recalibration with 10–11 sites in each city (Briggs et al., 2000). Local recalibration was necessary to take account of differences in meteorology, topography and vehicle fleet composition, and year-to-year changes in background concentrations (see further Section 5).

There are several ways in which to distribute monitoring sites over the study area once the total number of measurement sites has been fixed. Most studies have used informal methods to maximize the contrast in variables hypothesized to be potentially important predictors, by taking account of the distribution of locations to which the model will be applied. For example, in the TRAPCA study (Brauer et al., 2003) – designed to assess exposure for a birth cohort in three regions of the Netherlands – a total of 40 monitoring sites was available. It was decided to allocate 28 to urban and regional background locations and 12 to traffic locations. This decision was based upon the observation that, although only 5–10% of the population lived near major roads, those subjects were likely to experience substantially higher air pollution. It was thus decided to over-represent traffic sites in the monitoring campaign.

Kanaroglou et al. (2005) have developed a systematic methodology for selecting monitoring sites which uses the anticipated spatial variation in air pollution, as well as the distribution of addresses over the study area, to assign monitoring locations. The network density is increased in locations where concentration variability is higher and more people live. The method specifies a continuous demand (for monitoring) surface over the area. An algorithm from the general family of location-allocation problems is then used to select the optimal locations from a fixed number of monitoring sites. The demand surface incorporates an initial concentration surface, determined from, e.g. monitoring data in a wider area than the study area. The demand surface is then adapted by incorporating weights that reflect for example population density.

Another important issue is the micro-environment of monitoring sites, especially for traffic locations. If the purpose of the study is to assess exposure of people at the residential, school or work address, monitoring should take place near the façade of the homes rather than at the kerbside. Most prediction models are relatively crude (Tables 2–4) and unable to take account of small differences in distance, especially for urban roads (Brauer et al., 2003; Ross et al., 2007), though with high quality geographic data spatial resolutions of 20 m or so
are possible (Briggs et al., 1997, 2000). This may be less of a problem for exposures related to major freeways because of the generally larger distances that are affected and the more open terrain. Studies of freeway exposures have included continuous functions of distance to the freeway in their models (Gilbert et al., 2005; Kanaroglou et al., 2005).

3.3. Temporal aspects

In the SAVIAH study, four monitoring periods of 14 days were conducted spread over the four seasons. Subsequent studies that undertook purpose-designed monitoring have made between one and four repeats of 7–14 days. Monitoring is thus temporally limited and the calculated average concentrations do not necessarily agree with the annual average due to the possibility that atypical weather conditions occur during the survey period. It should be noted, however, that the original SAVIAH monitoring scheme covered 56 days of the year, close to the number of days (60) covered in the once every sixth day PM_{10} monitoring scheme in the USA, considered sufficient to establish an annual mean with good precision for regulatory purposes. More importantly, several studies have indicated that the spatial contrast between sampling sites is stable, provided that measurements are conducted simultaneously. In the SAVIAH study, the correlations between the four 14-day NO\textsubscript{2} sampling surveys ranged from 0.63 to 0.98 (Lebret et al., 2000). Correlations differed somewhat between cities: in Prague all correlations were above 0.92, whereas in Poznan the correlations were between 0.63 and 0.81. The authors also showed that between 63% and 84% of the total variability in NO\textsubscript{2} concentration was due to between-site variability (Lebret et al., 2000). In Oslo, the correlation between two 1-week average concentrations was above 0.91 for NO, NO\textsubscript{2} and NO\textsubscript{x} (Madsen et al., 2007). In Hamilton, Canada the correlation between NO\textsubscript{2} concentrations measured at 30 sites in October 2002 and May 2004 was 0.76 (Sahsuvaroglu et al., 2006). Strong support for the stability of the spatial NO\textsubscript{2} pattern was provided by the observation that the predicted NO\textsubscript{2} concentrations in Amsterdam and Huddersfield correlated very well with measurements made the following year (Briggs et al., 1997). NO\textsubscript{2} pollution surfaces in Toronto, Canada based upon measurements in September 2002 and spring 2004, respectively, were essentially the same (Finkelstein and Jerrett, 2007).

Henderson et al. (2007) have suggested a methodology to select two 14-day monitoring periods. Using data from 15 routine monitoring sites for a 5-year period, they calculated all 14-day running means for all years and compared the average of the periods separated by 26 weeks. The average of two periods closest to the actual annual mean, and not explained by extreme values, was used to select the sampling periods. They observed the February 19–March 4 and August 24–September 2 periods resulted in average NO\textsubscript{2} that were within 15% of the actual annual mean for 70 out of 75 cases (Henderson et al., 2007). There is obviously no guarantee that, during the actual campaigns, this will apply, as weather conditions are unpredictable. Because of seasonal variations in air pollution concentrations, and the potential for individual sampling campaigns to coincide with sustained periods of abnormal weather (e.g. under blocking anticyclones), we therefore recommend a minimum of at least two and preferably four campaigns to be performed. In choosing periods for survey, it is also important to avoid events that might affect air pollution conditions, such as major festivals (e.g. 'Bonfire night' in the UK, which may last for at least a week) or religious holidays.

Studies that have measured PM at a large number of locations could not perform simultaneous measurements because of insufficient equipment. In these studies, therefore, corrections were applied using the measured concentrations at a continuous monitoring location (Hoek et al., 2002b). Application of this adjustment improved the precision of annual average PM_{2.5} concentrations,

### Table 3

Performance of land-use regression models for NO and NO\textsubscript{x}.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study area</th>
<th>Predictor variables in final model</th>
<th>R\textsuperscript{2} of model</th>
<th>R\textsuperscript{2} validation</th>
<th>RMSE model (validation)</th>
<th>Measured concentrations (\mu g m\textsuperscript{-3})</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beelen et al. (2007)</td>
<td>Netherlands</td>
<td>Region + population 5000 m + urban land cover + traffic intensity 100 m</td>
<td>0.44</td>
<td>NR</td>
<td>NR</td>
<td>–</td>
</tr>
<tr>
<td>Madsen et al. (2007)</td>
<td>Oslo</td>
<td>Altitude + length high traffic roads, 100 m + length medium traffic roads, 250 m + length small roads, 1000 m</td>
<td>0.66</td>
<td>0.66–0.70</td>
<td>–</td>
<td>22 (3–124)</td>
</tr>
<tr>
<td>Henderson et al. (2007)</td>
<td>Vancouver</td>
<td>Length expressway, 100 m + length expressway, 1000 m + length major roads, 200 m + population density, 2500 m + altitude + X + Y-coordinate</td>
<td>0.62</td>
<td>0.49</td>
<td>–</td>
<td>62 ± 43</td>
</tr>
<tr>
<td>NO\textsubscript{x}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stedman et al. (1997)</td>
<td>UK</td>
<td>Rural NO\textsubscript{x} + urban land cover 25 km\textsuperscript{2} + NO\textsubscript{x} emission major road vehicles 4 km\textsuperscript{2}</td>
<td>0.96</td>
<td>NA</td>
<td>NA</td>
<td>–</td>
</tr>
<tr>
<td>Madsen et al. (2007)</td>
<td>Oslo</td>
<td>Altitude + length high traffic roads, 100 m + length medium traffic roads, 250 m + length small roads, 1000 m</td>
<td>0.73</td>
<td>0.68–0.78</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Aguilera et al. (2008)</td>
<td>Sabadell</td>
<td>Altitude + land cover factor, 500 m + road type</td>
<td>0.77</td>
<td>0.67–0.75</td>
<td>2.4</td>
<td>60 ± 30 (27–161)</td>
</tr>
</tbody>
</table>

Measured concentrations are mean ± standard deviation, with minimum and maximum in parentheses. NA, not available; NR, tabulated statistics not reported, but validation procedures performed resulting in typically small prediction errors.
expressed as the standard error of the mean, from 2.0–3.5 μg m\(^{-3}\) to 0.7–0.9 μg m\(^{-3}\) (range for the three study areas: Munich, Stockholm and the Netherlands). It relies, however, on the assumptions that the air pollutant pattern across the study area is broadly stable over time (i.e. more polluted areas remain more polluted), and that the continuous monitoring site is representative of the temporal series. Sites used to adjust for temporal variations should therefore be selected with care, and should not be subjected to local sources which might change over time (e.g. roads or major industrial emission sources).

4. Predictor variables

Most studies have typically assessed a large set of potential predictor variables to model monitored concentrations.
Frequently used predictor data include: traffic variables, population or address density, land use, altitude and topography, meteorology and location (Table 6). As an example, a study by Henderson et al. (2007) included 55 potential predictors and a study by Moore et al. (2007) examined 140 predictors. In the final models, typically only a small number of predictors are included. Definitions of the variables differ substantially between studies (Tables 1–5). This is related to data availability, unique features of the study area (e.g., distance to a lake), and probably also a result of arbitrary decisions by the investigators.

Little attention has been given in the identified papers to the potential problems associated with geographic data sets. Problems may include accessibility, completeness and...
precision. Data may also not be available for the period of interest. This is especially an issue for retrospective exposure assessment. For a general discussion, see for example Briggs and Elliot (1995), Vine et al. (1997), Nuckols et al. (2004) and Briggs (2005). Traffic intensity data, especially those for municipal roads, are often problematic as they are not always accessible for the investigators, despite the fact that most traffic data have been obtained with public funds. In many cities, traffic counts are only available for a small number of streets, and mainly on major roads (Beelen et al., 2007; Madsen et al., 2007). Assumptions about the traffic intensity of other roads therefore must be made. Significant difficulties may also arise in attributing counts to the road network, because counts have to be extrapolated along and between road links. Traffic models can be used to assign traffic counts to other roads, and in some cities modelled data are available from transport and highways authorities. In the absence of traffic data, or to avoid the problems of acquiring them, several LUR studies have successfully explored the use of the length of specific road types without traffic intensity data (Brauer et al., 2003; Madsen et al., 2007; Henderson et al., 2007). In Vancouver, the R² of the models based on traffic intensity did not differ from the models using only road length of highways and major roads for the evaluated pollutants (Henderson et al., 2007).

Predictor variables in LUR models are usually computed for circular zones around each monitoring site, using buffer functions in the GIS. The selection of buffer size is crucial in determining the performance of the model, and the spatial resolution of the estimates. Ideally, buffer sizes should be selected to take account of known dispersion patterns. Various monitoring studies have shown that the impact of a major road on concentrations of traffic-related air pollutants declines exponentially with distance to the road (Roorda-Knape et al., 1998). Beyond about 100 m from a major urban road, or 500 m from a major freeway, variability is limited. In inner-city areas, however, buildings may cause marked departures from this simple distance-decay pattern. In street canyons, for example, marked accumulation of air pollution may occur, especially against the windward side of buildings, with the result that concentrations may differ substantially from one side of the road to another. At the same time, NO₂ formation in street canyons may be limited by the availability of free oxygen, so that more of the pollution remains in NO form. There is also evidence to suggest that air pollution concentrations fall virtually to background levels behind a row of uninterrupted buildings (Bloemen et al., 1993). Especially in the compact European urban areas, much of the variation in traffic-related air pollution is therefore extremely local. Use of buffer sizes of more than 100–200 m for traffic intensity may thus be misleading, for they would incorporate sources too far removed to have a significant effect.

The large spatial variability that occurs within tens of meters from major roads is potentially challenging given the geographic precision of monitoring sites, address and road data (Briggs et al., 1997; Hochadel et al., 2006). In a series of papers from Florida, it was reported that street geocoding resulted in substantial errors of the geographical position of homes and schools. This resulted in misclassification and bias of potential traffic exposures of children at home and at school (Zandbergen, 2007; Zandbergen and Green, 2007). In recent years, however, the spatial resolution of digital geographic data has improved considerably. In the Netherlands, for example, it is claimed that digital data sets of road lines are within 10 m of their true location for ca. 95% of road sections (Beelen et al., 2007). A sample of the Address Coordinates Netherlands showed that 93.5% of the addresses were located at the centroid of the correct building (Beelen et al., 2007). In a study in New York City, on the other hand, it was observed that the coordinates of older routine monitoring sites were not geographically precise (Ross et al., 2007). Uncertainties also occur in the traffic data, due to the limited representativeness of traffic counts and difficulties in extrapolating these across the network. In a Dutch study, it was found that direct observations of traffic by a technician at the monitoring sites improved the prediction models for soot and PM₂.₅ over the GIS variables (Brauer et al., 2003). In studies with a moderate population size (several hundreds), therefore, it may be useful to include direct traffic observations at the home address. This is especially attractive if home visits have been planned to collect specific additional data for the epidemiological study of interest (e.g. collect house dust in a birth cohort study).

A further complication is the difficulty in obtaining data on height above ground for addresses. Geographical coordinates define the position using X- and Y-coordinates but do not necessarily have the height attached. In high-rise apartment buildings this may be an important issue, as several monitoring studies have suggested important differences in air pollution related to height (Vakeva et al., 1999; Janhall et al., 2003; Jo and Ky, 2002). In an elegant study in a major city in Korea, VOC concentrations were seen to be up to 70% higher outside low-floor (1st or 2nd) apartments compared to high-floor (10th–15th floor) apartments in the same building. Buildings were located within 30–100 m of a major road (Jo and Ky, 2002). Vertical

### Table 7

Approximate costs of land use regression monitoring campaigns with 40 sites and four weekly samples per site

<table>
<thead>
<tr>
<th>Pollutants</th>
<th>No. samplers</th>
<th>Investment (Euro)</th>
<th>Personnel (person days)</th>
<th>Personnel (Euro)</th>
<th>Consumables (Euro)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM₁₀ + PM₂.₅</td>
<td>12</td>
<td>60,000</td>
<td>110</td>
<td>26,000</td>
<td>7200–9600</td>
</tr>
<tr>
<td>NOₓ</td>
<td>40–80</td>
<td>3200–6400</td>
<td>30–40</td>
<td>3000–4000</td>
<td>1500–3000</td>
</tr>
</tbody>
</table>

**Notes:**
- Unit prices of 5000 Euro for PM equipment consisting of pump and impactors used in the TRAPCA study (Hoek et al., 2002a; Brauer et al., 2003) and 80 Euro for the Ogawa badge. For PM monitoring the TRAPCA design was assumed with measurements at 10 sites simultaneously.
- PM: 64 days field work, 25 days of preparation (site selection, equipment) and 20 days sample analysis and data management. NOₓ: 16 days field work, 10 days preparation and 5 days analysis (preparation and analysis costs double if duplicates used at all sites). Tariff 30 Euro h⁻¹ for a technician.
- PM: 320 filters (PM₁₀ and PM₂.₅) at 5 Euro per filter plus QA/QC; NOₓ: 320 filters (NO₂ and NOₓ) at 3 Euro per piece. If duplicates are used for each site, consumable costs are double.
profiles were more pronounced during inversion and night-time hours. This suggests that applying LUR models in study areas where a large fraction of the population lives in high-rise apartment buildings may seriously misclassify exposure. Misclassification is especially an issue for buildings near major roads. There is very limited vertical gradient at urban background locations.

5. Model development and validation

Most studies use standard linear regression techniques to develop prediction models. Forward, backward or best-subsets automatic selection methods are often applied to develop a parsimonious model from a large set of predictor variables that maximizes the percentage explained variability ($R^2$). Following the approach used in the SAVIAH study, a priori definition of a required sign of regression slopes for specific variables (e.g. positive for traffic intensity) is used by some investigators in order to increase the applicability of the model beyond the monitoring sites (Brauer et al., 2002; Henderson et al., 2007).

Four studies have used a structured approach in which different predictors were used for different spatial scales (Stedman et al., 1997; Hoek et al., 2001a; Briggs et al., 2005; Beelen et al., 2007). These were studies applying LUR models to larger geographical areas than most studies which model a large metropolitan area. In the Stedman study, a distinction was made between the rural and urban components. The rural scale was assessed by interpolating NO$_2$ concentrations from rural monitoring sites. An urban component was added by regressing measured concentrations at urban sites against land use and emissions from motorized traffic. In the other three studies the spatial variation was assumed to comprise regional, urban and local components for which a different set of monitoring locations and predictor variables was used. A model was first developed for the regional and next for the urban sites into a training data set for model development and a smaller group of sites for model validation (Briggs et al., 2005). This approach requires less intensive computer processing, but may be disadvantaged by the a priori division of sites (e.g. concentrations measured at the training and validation sites may differ). A combination of these approaches may provide a reasonable compromise. For example, a form of grouped jackknife analysis may be used, in which the monitoring sites could be divided into several equal groups, and the analysis repeated $n$ times using a different group for validation such that all sites are used in model development and validation (Briggs et al., in press). Other approaches include comparison with databases that have not been used in model development, such as data from routine monitoring stations (e.g. Stedman et al., 1997; Henderson et al., 2007). Also, comparisons with measurements at the same sites in different time periods have been considered as validation. While it is very important to document temporal stability of the typically short monitoring data, the temporal comparison does not address the question of whether the model is able to predict spatial variability well.

5.1. Transferability

As mentioned above, little attention has been given to transferability of models to other locations. We believe that further exploration in this area would be beneficial for the wider application of the land use regression method. The
first systematic attempt was the study by Briggs et al. (2000) who successfully applied the SAViaH model to other UK cities after calibration with a small number of monitoring sites. Jerrett et al. (2005) observed that application of the Amsterdam regression model reported by Briggs et al. (2000) did not explain spatial variation in the Hamilton area. The correlation of predicted NO2 concentrations with government monitoring data was poor. This suggests that transfer of models will likely be limited by the availability of equivalent variables in other areas than the area where the model was developed, and models may not be transferred to areas that have a different structure in terms of land use. Jerrett et al. (2007) discuss differences in structure between European and North–American cities.

As land use regression models are empirical models – in contrast to dispersion models which are derived from physical principles – limitations in transferability may be inevitable to some extent. When possible, it is likely to be preferable to transfer models from nearby areas that have similar predictor variables, since this will not only increase the efficiency of applying LUR models, and therefore reduce the cost of monitoring campaigns, but will also help to improve consistency between different studies. The potential of transferring models to other locations will be increased by the use of centralized (e.g. European) instead of local databases and somewhat less focus on maximizing the prediction \( R^2 \). Authors may report a best model (highest \( R^2 \)) with local predictors and a less predictive model that has more potential to be transferred.

6. Performance of models

Land-use regression methods have been applied to develop maps of NO2, NOx, PM2.5, the elemental carbon or soot content of PM2.5 and VOCs. For NO2, the percentage explained variation from the prediction model is typically about 60–70%. This is often achieved with only a few predictors, usually including variables representing traffic load, population density, altitude and land use in various representations (e.g. industrial or urban). Differences in prediction \( R^2 \) between studies may be related to the original variability in the measured concentrations, quality of the predictor variables, the modelling approach and the complexity of the city, for example, in terms of differences in topography and emission sources (Briggs, 2007). In five studies, the \( R^2 \) from the prediction model and the cross-validation were very similar, documenting the robustness of the model (Briggs et al., 1997, 2000; Gilbert et al., 2005; Madsen et al., 2007; Henderson et al., 2007). In the Stedman study, the cross-validation \( R^2 \) was much smaller than the \( R^2 \) of the prediction model, possibly related to the use of a validation data set derived from sites using a different sampling methodology. Nevertheless, it illustrates the necessity of applying validation methods. Further support for the usefulness of LUR models is that the root mean squared error, in all studies that reported it, was low compared to the range in measured NO2 concentrations.

The few studies that evaluated NO found similar \( R^2 \) values to those for NO2 in the same city (Tables 2 and 3). Henderson et al. (2007) demonstrated that the NO map developed for Vancouver showed more pronounced small-scale spatial contrast than the NO2 map, suggesting the importance of monitoring a primary pollutant.

Most studies were able to develop good predictive models for soot, and even for PM2.5 that is characterized by a high regional background concentration with only limited impact of local sources (Table 4). The exception was the study in North–Rhine Westphalia where the \( R^2 \) of the PM2.5 model was only 17% compared to 82% for the absorbance of the PM2.5 filters (Hochadel et al., 2006). Limited variability of the measured PM2.5 concentrations in the small study area, further characterized by a lack of substantial altitude differences, is one likely explanation.

Three studies have suggested that VOCs can be modelled with LUR techniques, incorporating traffic variables and industrial sources. Monitoring of VOCs could be of interest in specific industrial settings, in assessing gas stations and in studies of specific health outcomes (e.g. leukemia). Because of the more demanding nature of monitoring compared to, for example, NOx, we see little merit in using VOCs as another indicator of traffic-related air pollution.

Several studies have compared LUR to commonly used methods such as geostatistics and dispersion modelling. In general LUR has been shown to outperform geostatistical methods. In SAVIAH, LUR predicted NO2 concentrations in each of three European cities substantially better than geostatistical methods (Briggs et al., 2000). The percentage explained variability (\( R^2 \)) of measured NO2 concentrations at validation sites was between 0.79 and 0.87 for LUR models, between 0.09 and 0.56 for TIN-contouring, 0.34 and 0.44 for kriging and between 0.27 and 0.48 for trend surface analysis in the three cities. In a study by Ross et al. (2007) in New York City, land use regression slightly outperformed kriging in predicting PM2.5 concentrations at validation sites. The RMSE at prediction sites was 1.15 and 1.00 µg m\(^{-3}\) for LUR in the ‘28-county’ and ‘more urbanized 9-county’ models versus 1.30 and 1.47 µg m\(^{-3}\) for kriging in these counties. The difference between the two studies is likely due to the much smoother spatial patterns for PM2.5, which has a strong regional background component. In London, the spatial variability of PM10 across 52 monitoring sites was poorly predicted by kriging and nearest monitor site (validation \( R^2 = 0.01 \) and 0.05), while an LUR model explained 47% of the measured variability (Briggs et al., in press).

Rather few comparisons have been made between LUR methods and dispersion modelling. Those that have been done suggest that LUR models perform as well as dispersion models in explaining spatial variation. In Huddersfield, UK, for example, land-use regression explained 82% of the variability of the NO2 concentrations at validation sites whereas the CALINE-3 model explained 63% of the NO2 variability (Briggs et al., 2000). In Munich, a stochastic model explained slightly more variability in NO2 concentrations at the monitoring sites (\( R^2 = 0.62 \)) than a dispersion model (\( R^2 = 0.46 \)), although the modelled concentrations using the two approaches were highly correlated (Cyrys et al., 2005). In London, LUR models for PM10 across 52 monitoring sites (assessed using the grouped jackknife approach) gave \( R^2 \) of 0.47, compared with 0.28 for the ADMS-Urban dispersion model (Briggs et al., in press). In the Greater Vancouver area, predictions from LUR models
correlated better with concentrations from regulatory monitoring sites than the dispersion model CMAQ (NO:  
\[ R = 0.75 \] versus \[ R = 0.46 \] and NO\(_2\): \[ R = 0.76 \] and \[ R = 0.71 \], respectively) (Marshall et al., 2008). The main difference between the LUR-model, the dispersion and the evaluated interpolation models was in the spatial scale that was described. LUR-models were better in describing hotspots whereas the other methods provided a smoother concentration surface (Marshall et al., 2008).

7. Limitations

While land use regression methods have, in many cases, been applied successfully to model spatial variation of ambient air pollution, there are several limitations of the method. First, LUR models have a limited ability to separate the impact of some priority pollutants. In the TRAPCA study, the correlation of modelled NO\(_2\), PM\(_{2.5}\) and soot was above 0.95 in the three study areas (Munich, Stockholm and three regions in the Netherlands), so that the potentially independent health effects of these pollutants could not be evaluated (Brauer et al., 2002). Measured concentrations were also highly correlated (\( R^2 \) of average NO\(_2\) and PM\(_{2.5}\) between 0.64 and 0.80 and between 0.76 and 0.86 for NO\(_2\) and soot concentrations) (Lewne et al., 2004) but the relatively crude models increased the correlation of the modelled concentrations even further. In Munich, the correlation between NO\(_2\) and PM\(_{2.5}\) was 0.84 for measured concentrations and close to unity for modelled concentrations (Cyrys et al., 2005). Minor correlations were found in a small study area in Germany between PM\(_{2.5}\) and NO\(_2\) (\( r = 0.41 \)) and PM\(_{2.5}\) and soot (\( r = 0.52 \)) (Hochadel et al., 2006). The correlation between soot and NO\(_2\) was 0.93. Insofar as these correlations represent real associations between the different pollutants (rather than artifacts of the modelling procedures), the same problems will affect other methods of exposure assessment.

Second, LUR models may be unable to represent the extremely local variations in concentration (over distances of tens of meters) that may occur near sources such as major roads. This limitation, however, is mainly related to the precision of input data, and thus would also apply to dispersion models.

A third limitation of current LUR models is the often short temporal coverage of purpose-designed monitoring campaigns which does not allow the precise calculation of absolute concentrations that can be compared with air quality guidelines. Subject to the constraints mentioned earlier, this issue may be addressed by scaling the modelled concentrations using a continuous monitor in the study area. This is mainly an issue when maps are developed for regulatory purposes, much less so for application in epidemiological studies.

Fourth, dispersion models are superior when the interest is in a specific source-related component of the total concentration (e.g. traffic-derived particulates), since this can be separately modelled; in LUR models, in contrast, calibration against monitoring sites means that the model is usually designed to predict only the total concentration.

A further limitation is that, although land use regression models provide individual estimates of exposure at, for example, the residential address, the estimate refers only to the ambient concentration. The predictor variables do not take account of factors (e.g. air exchange rate) related to infiltration of outdoor air in the home where people spend a large fraction of their time. Time activity patterns are also ignored: for example, the fraction of time actually spent at home is likely to be an important determinant. A more detailed discussion can be found in Briggs (2005), though this problem affects all methods of exposure assessment except personal monitoring or biomonitoring. To our knowledge, only one study has evaluated the contrast in personal exposure related to land use regression models (Nethery et al., 2007). In a group of 62 pregnant women, personal exposure of, especially, NO was moderately correlated with ambient NO assessed by land use regression. A few studies have further suggested that some of the traffic variables used in land use regression models are associated with small but significant contrasts in personal exposure of soot (Wichmann et al., 2005; Van Roosbroeck et al., 2006, 2007). Validation of the developed pollution maps with actual personal monitoring is therefore another important research need.

A final limitation worth noting is that some of the predictors used for developing air pollution exposures with LUR could introduce confounding when applied in epidemiological studies (Moore et al., 2007). LUR models including population density, for example, may be problematic as population density may also be associated with other adverse risk factors such as low socio-economic status or poor housing stock, which could influence the disease of interest (e.g. asthma prevalence). One solution to this potential problem is the inclusion of area-level confounders that are more closely related to the disease of interest (e.g. percentage of low-income families in a neighborhood) than the variable used in predicting air pollution (number of addresses in a 300 m buffer).

8. New developments

This section briefly outlines some innovations of the LUR methodology including expanding the scope of the predictor variables, new GIS approaches, and spatio-temporal models.

Rosenlund et al. (2008) evaluated the value of adding actual emission data to surrogate variables such as traffic intensity and population density to predict NO\(_2\) concentrations in the city of Rome. No improvement of the land use regression model was obtained after adding emissions of PM, NO\(_x\), CO and benzene available at the census block level. In addition to assessing emissions from 164 census blocks, a continuous surface obtained through kriging was evaluated, but this did not improve the NO\(_2\) prediction either. The limited spatial resolution of the emission data might have contributed to the lack of prediction. It is not clear whether these findings would apply in other cities. An early study by Stedman found that traffic emissions were highly predictive of UK-wide NO\(_2\) concentrations, but the authors did not evaluate whether the same predictive power could have been obtained with traffic intensity only. In a study in New York City, primary PM\(_{2.5}\) emissions available at the county level did not enter the final
prediction model for PM$_{2.5}$ (Ross et al., 2007). The study area consisted of 28 counties; hence data were available at a rather crude spatial scale.

Arain et al. (2007) assessed the use of wind fields to improve the prediction of air pollution in the Toronto–Hamilton urban airshed. Wind direction is an important variable that determines the impact of sources on receptor points, though it has rarely been incorporated into land-use regression models. Wind direction fields were constructed from 38 weather stations in the area and the constructed wind direction was found to be a significant predictor of NO$_2$ concentration. The $R^2$ of the prediction increased from 0.65 without wind to 0.69 when wind direction was added to the model. It is not obvious whether the work required to construct the wind fields was worth the effort given this modest increase in prediction. As the authors appropriately point out, wind fields in urban areas should be very local in order to be useful.

Differences in dispersion conditions related to, for example, annual average wind speed may be important. In the simple CAR-2 model used in the Netherlands for modelling road-traffic emissions, the country is divided into different regions which are assigned different dispersion factors to reflect differences in wind speed related to distance to the sea (Eerens et al., 1993). In LUR models, this could be taken into account empirically by allowing the relationship between traffic intensity, or other source indicators, and air pollution concentrations to differ between different regions of the study area (interaction). This may be important when study areas are relatively large.

While modelled annual average concentrations are sufficient for most applications to epidemiological studies, birth cohort studies require more detailed temporal resolution. In these studies, it is common to express exposure as the average concentration per trimester of a specific pregnancy (Slama et al., 2007). The required exposure thus needs to contain a spatial and temporal component. To date this has not been systematically evaluated. One simple option is to develop LUR models using annual average concentrations as is commonly done and then use continuous routine monitoring data to adjust for the temporal component. This approach makes the assumption that the spatial pattern is constant in time. A more sophisticated approach was taken in a study on short-term air pollution variations (Maynard et al., 2007). In that study, a model was built that incorporated both time-varying and time-invariant factors in one model. Specifically, the black carbon (BC) concentration measured on a specific day at a specific location was modelled as a function of traffic within 100 m of the site (time-invariant), BC measured daily at a central continuous monitoring site, day of the week, weather and longitude and latitude. Non-parametric regression was used to allow for non-linear relationships between air pollution and predictor variables. The model performed substantially better than the central site measurement alone: cross-validation $R^2$ was 0.36 compared to 0.09 for central site BC. Another approach is taken by a group of Canadian researchers who developed a source-area land use regression (SA-LUR) model, by combining concepts of a box-type dispersion model and land use regression to provide more detailed temporal resolution (Su et al., 2008). Different scenarios with varying degrees of complexity of input data (wind direction, mixing height) were evaluated. The SA-LUR model was not superior to LUR in predicting long-term average concentrations, but did provide the flexibility for developing shorter-term exposures (Su et al., 2008). A key problem in the method was the ability to obtain sufficiently spatially detailed meteorological data.

Another innovation is the application of focalsum methods to construct weighted average prediction variables with a moving window approach. In this type of analysis, a window is 'moved' across a grid to derive a new value for the central cell, which is some function of the other cells covered by the window. In the context of air pollution modelling, this functionality can be used to smooth the spread of highly resolved data on pollution sources (e.g. emissions) to generate a map of concentrations. The fundamental principle behind the approach is that proximity to the source is one of the main determinants of pollutant concentrations. To date there exist few examples of moving window analysis for air pollution modelling. Collins (1998) was perhaps the first to explore the approach in a comparison of three GIS-based approaches in which traffic-related pollution was modelled in Huddersfield, UK. A distance-weighted kernel was used to model the near-source traffic-related component, which was subsequently added to a map of background concentrations derived from kriging. Validated against data from eight independent monitoring sites, the moving window approach ($R^2 = 0.67$) performed better than kriging alone ($R^2 = 0.44$), and was comparable to dispersion modelling ($R^2 = 0.63$). The best model, however, was derived using land use regression ($R^2 = 0.82$). Research by Loibl and Orthofer (2001) to produce a 250 m grid of NO$_2$ concentrations for Austria involved the use of weighted moving windows to apply existing dispersion profiles to point, line and area source emissions. Model validation indicated that 54% (72 sites) and 76% (101 sites) of the measurements at monitoring sites, respectively, were within ±15 and ±25 µg m$^{-3}$ of the model predictions. In general, however, their model tended to under-predict, especially at monitoring stations with high NO$_x$ contribution from traffic.

In more recent work, the moving window approach was further explored as an alternative method for modelling NO$_2$ at the 1 km level, across the EU-15 (Vienneau, 2006). The method estimated concentrations by calibrating the distance-weighted sum of emissions in concentric windows around each monitoring site to the monitored concentrations. The major advantage is that the models can be run within a raster GIS environment, which allows for rapid computation of large data sets. The models were developed using monitoring data from 714 background sites for 2001, and validated by comparing predictions with measured concentrations for a separate set of 228 reserved background sites ($R^2 = 0.60$).

9. Conclusions

Land-use regression methods have generally been applied successfully to model annual mean concentrations of NO$_2$, NO$_x$, PM$_{2.5}$, the soot content of PM$_{2.5}$ and VOCs. The method has been applied in different settings, including
European and North-American, non-industrial and industrial cities. The performance of the method in urban areas is typically better or equivalent to geo-statistical methods such as kriging and conventional dispersion models. Compared to dispersion models, the land use regression method requires less detailed input data at the expense of the need to obtain monitoring data for a sufficiently large number (40–80) of sites.

Land-use regression methods can benefit from a more systematic selection and description of monitoring locations and monitoring periods. More attention to the precision of geographic data is also important. A model strategy that incorporates greater knowledge of the factors related to spatial variation and focuses less on maximizing the percentage explained variability would probably result in models than can more readily be transferred to other areas. Where purpose-designed monitoring is included, the cost of monitoring could probably be reduced if models were transferable. Promising new developments include the use of additional predictor variables such as wind direction data or emission data and the use of the raster GIS environment – for example, to apply focalsum methods. Models that include both a spatial and a temporal component are also of interest for studies that need exposure variables on a more detailed scale, but it remains to be seen whether these LUR models can outperform dispersion models for shorter averaging periods. Finally, an area of importance for epidemiological research is the need for validation of LUR models with personal monitoring.

Acknowledgments

The work is supported by grant RGI-137 from the Dutch program Ruimte voor Geoinformatie supported by the Ministry of Housing, Spatial Planning and the Environment.

References


Bilenick, 1985. The space–time distribution of sulfate deposition in the European and North-American, non-industrial and industrial

areas. Where purpose-designed monitoring is included, the number (40–80) of sites.


