Estimated long-term outdoor air pollution concentrations in a cohort study

Rob Beelen\textsuperscript{a,*}, Gerard Hoek\textsuperscript{a}, Paul Fischer\textsuperscript{b}, Piet A van den Brandt\textsuperscript{c}, Bert Brunekreef\textsuperscript{a,d}

\textsuperscript{a}Institute for Risk Assessment Sciences, Division Environmental Epidemiology, Utrecht University, P.O. Box 80178, 3508 TD Utrecht, The Netherlands
\textsuperscript{b}Center for Environmental Health Research, National Institute for Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, The Netherlands
\textsuperscript{c}Department of Epidemiology, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
\textsuperscript{d}Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands

Received 22 June 2006; received in revised form 6 October 2006; accepted 10 October 2006

Abstract

Several recent studies associated long-term exposure to air pollution with increased mortality. An ongoing cohort study, the Netherlands Cohort Study on Diet and Cancer (NLCS), was used to study the association between long-term exposure to traffic-related air pollution and mortality. Following on a previous exposure assessment study in the NLCS, we improved the exposure assessment methods.

Long-term exposure to nitrogen dioxide (NO\textsubscript{2}), nitrogen oxide (NO), black smoke (BS), and sulphur dioxide (SO\textsubscript{2}) was estimated. Exposure at each home address (N = 21 868) was considered as a function of a regional, an urban and a local component. The regional component was estimated using inverse distance weighted interpolation of measurement data from regional background sites in a national monitoring network. Regression models with urban concentrations as dependent variables, and number of inhabitants in different buffers and land use variables, derived with a Geographic Information System (GIS), as predictor variables were used to estimate the urban component. The local component was assessed using a GIS and a digital road network with linked traffic intensities. Traffic intensity on the nearest road and on the nearest major road, and the sum of traffic intensity in a buffer of 100 m around each home address were assessed. Further, a quantitative estimate of the local component was estimated.

The regression models to estimate the urban component explained 67\%, 46\%, 49\% and 35\% of the variances of NO\textsubscript{2}, NO, BS, and SO\textsubscript{2} concentrations, respectively. Overall regression models which incorporated the regional, urban and local component explained 84\%, 44\%, 59\% and 56\% of the variability in concentrations for NO\textsubscript{2}, NO, BS and SO\textsubscript{2}, respectively.

We were able to develop an exposure assessment model using GIS methods and traffic intensities that explained a large part of the variations in outdoor air pollution concentrations.

\textcopyright{} 2006 Elsevier Ltd. All rights reserved.

Keywords: Long term; Exposure; Air pollution; Geographic information system; Traffic; Cohort

*Corresponding author. Tel.: +31 30 2539474; fax: +31 30 2539499.
E-mail address: r.beelen@iras.uu.nl (R. Beelen).
1. Introduction

We recently reported an association between cardiopulmonary mortality and long-term exposure to traffic-related air pollution in a random subgroup of 5000 of a cohort study of elderly people (NLCS study; N=120,000) (Hoek et al., 2002a, b). Other recent studies also found associations between long-term exposure to air pollution and health effects (Nafstad et al., 2004; Filleul et al., 2005; Finkelstein et al., 2004).

Two large cohort studies in the US also showed associations between long-term exposure to air pollution and mortality. However, in the Harvard Six Cities study and the American Cancer Society (ACS) study, exposure has been estimated as the average concentration within a city/metropolitan area, ignoring small-scale variations within cities (Dockery et al., 1993; Pope et al., 1995). Exposure studies have shown large variations in concentrations of traffic-related air pollutants on a small scale within cities (Fischer et al., 2000; Jerrett et al., 2005a). A study by Jerrett et al. (2005b) in a subpopulation of the ACS study suggested that the chronic health effects associated with intraurban gradients in fine particles may be even larger than previously reported associations across metropolitan areas.

In our previous study, long-term exposure to traffic-related air pollution was therefore considered as a function of the regional background, an urban background and an indicator variable for living near a major road (Hoek et al., 2001). Following on this previous study, we conducted a study on the association between long-term exposure to traffic-related air pollution and mortality in the full NLCS cohort of ~120,000 subjects (NLCS-AIR study). The purpose of this paper is to describe an improved method for the assessment of long-term outdoor air pollution concentrations in a cohort study.

2. Methods

2.1. Design

The design of the Netherlands Cohort Study on Diet and Cancer (NLCS) (van den Brandt et al., 1990) and the exposure assessment method used in the previous study have been described in detail elsewhere (Hoek et al., 2001). Briefly, the NLCS started in September 1986 when 120,852 participants from 204 municipalities, spread out over the Netherlands, were enrolled. The residential address of all participants in 1986 is known.

The NLCS was designed as a case–cohort study, i.e. mortality cases are derived from the entire cohort of ~120,000 participants, while the person years at risk are estimated from a random sample of ~5000 subjects (van den Brandt et al., 1990). The number of observations in a case–cohort analysis consists of the number of subcohort members plus the number of mortality cases (which depends on the cause of death). For all cause mortality, which has the largest number of observations, the number of observations in the case–cohort analysis is 22,337 (4,991 subcohort members and 17,346 mortality cases). The entire cohort was followed-up for mortality until December 31, 1996 for the current study.

In the previous study, the assessment of long-term exposure to black smoke and nitrogen dioxide consisted of a separate estimation of regional background, urban background and local traffic contributions at the centroids of the six-digit postal code areas of the home addresses. Inverse distance weighed interpolation of concentration data from regional background sites in the National Air Quality Monitoring Network (NAQMN) was used to estimate the regional background concentration. A regression model relating address density in a four-digit postal code area to measured concentrations at urban background sites was used to estimate the urban background concentration. Local traffic contributions were estimated using a GIS and a digital road network (Basisnetwerk—version year 1993), resulting in a variable indicating whether a subject lived close to a major road (within 50 m from a major urban road or within 100 m from a motorway) (Hoek et al., 2001).

For the current study we extended the exposure assessment with additional air pollutants and additional spatial determinants. Long-term outdoor air pollution concentrations for black smoke (BS), nitrogen dioxide (NO2), nitrogen oxide (NO), and sulphur dioxide (SO2) were estimated based on the exact residential home address in 1986. Data were collected from 1976 (start of the NAQMN) until 1996 (end of follow-up). NO2 and NO data were available for the years 1978–1996, BS data for the years 1985–1996 and SO2 data for the years
1977–1996. In 1986 the NAQMN was rearranged resulting in only limited days with valid measurements in 1986.

A limited assessment was made for fine particles measuring less than 2.5 \( \mu m \) (PM\(_{2.5}\)) and less than 10 \( \mu m \) (PM\(_{10}\)) in diameter, using NAQMN data from 1992 to 1996 and data from two specific monitoring studies in 1997/1998 (Janssen et al., 2001) and 1999/2000 (Brauer et al., 2003).

As in the previous study, long-term exposure to outdoor air pollution was considered as a function of a regional component, an additional urban component, and an additional local component from nearby (busy) roads: Exposure = C(regional) + C(urban) + C lokal.

The exact 1986 residential addresses of participants were geocoded into standard Dutch geographical coordinates (Address Coordinates Netherlands (ACN)) using a database from the year 2000. For 21,868 of the 22,337 study subjects (97.9%), the home address could be geocoded.

2.2. Regional component

The regional component was estimated using average daily concentrations measured at regional background sites in the NAQMN (Elskamp, 1989). Fig. 1 shows the locations of NO\(_2\) monitoring sites in the NAQMN in 1996. Before calculating the average concentration for a station, missing values were estimated to prevent bias in the comparison across sites. Missing values were estimated by calculating the seasonal means (summer/winter) of the daily ratios of the concentration at a station to the daily average concentration at the other sites. A seasonal mean ratio was used because differences in meteorological conditions between seasons can affect the mean ratio. These seasonal mean ratios were used to multiply the average concentration of all sites of a day with a missing value. Missing values were estimated for 0–17% of the days at a monitoring site. Correlations between unadjusted and adjusted annual averages were high (>0.9 for all pollutants), indicating that both unadjusted and adjusted data would result in similar exposure estimates.

In contrast with the previous study, we evaluated two methods to interpolate concentrations. Ordinary kriging using the GSTAT package within the R statistical software environment (Pebesma, 2004), and inverse distance weighted interpolation as in the previous study, were used.

Fig. 1. NO\(_2\) monitoring sites in the national air quality monitoring network (1996).

2.3. Urban component

The urban component was estimated using regression models, in which we evaluated address density and number of inhabitants in different buffers as potential predictor variables while in the previous study only address density in a four-digit postal code area was used as predictor variable (Hoeck et al., 2001). A possible limitation of the method in the previous study was that a subject can live near the border of a postal code area with a completely different address density. We also used a larger number of potential predictor variables compared with the previous study: number of inhabitants (GIS database available for 1995) and address density (available for 1998) in different buffers around the home address (300, 1000, and 5000 m), which were assessed using buffer calculations in which only cells within the buffer that have data values were used in determining the sum. For homes near the border of the Netherlands a (probably slight) underestimation thus occurred. Further, land-use variables that indicated whether a site was located in a city center, in a rural background location, in an industrial location, or
in a traffic location (available for 1999) were used. For each pollutant, regression models, using mixed modeling, were developed with residual concentrations as dependent variables. Residual concentrations for all regional background and urban monitoring sites in the NAQMN (Elskamp, 1989) were calculated as measured concentration minus estimated regional component concentration using cross-validation. The average residual concentrations for each 5-year period 1976–1980, 1981–1985, 1987–1991, and 1992–1996 were used as dependent variables in separate models. We used 5-year periods as these were the exposure variables of interest for the cohort study.

For each pollutant and each 5-year period separately, regression models were developed using a manual forward selection process. No restriction regarding the number of predictor variables was applied. Variables were only included when the direction of the effect was as defined a priori. First, univariate regression analyses were conducted for all possible predictor variables. The model with the highest adjusted explained variance ($R^2$) value was regarded as the start model. Second, to this start model the remaining variables were added separately and the effect on the adjusted $R^2$ was evaluated. If the effect on the adjusted $R^2$-value was greater than 1%, then the predictor variable with the highest addition to the previous regression model was added to the model. This was repeated until there were no predictor variables that added more than 1% to the adjusted $R^2$-value of the previous regression model.

To assess whether a single regression coefficient for each predictor variable could be applied for the whole period 1976–1996, interaction terms between time period and the predictor variables in the final models were tested. No significant interaction terms were found, indicating that regression coefficients did not differ for different periods. Therefore, overall coefficients for each predictor variable were estimated using mixed modeling and combining data of the different 5-year periods.

Values for the predictor variables for the coordinates of the home addresses of the study participants were obtained using GIS and were multiplied with the coefficients of the regression models to calculate an urban component concentration. This estimated urban component was added to the regional component resulting in a background concentration. Background concentrations were estimated for all separate years in the period 1976–1996 (if available), and average exposure estimates were made for the 5-year periods 1976–1980, 1981–1985, 1987–1991, and 1992–1996 (in 1986 the NAQMN was rearranged resulting in only limited days with valid measurements).

### 2.4. Local component

Traffic variables to characterize the local component were assessed using a digital road network to which average whole-week traffic intensity data (Monday–Sunday) were linked.

Compared to the previous study we used a more detailed, more complete and geographically more accurate digital road network: the National Road Database (Nationaal Wegen Bestand (NWB), version year 2001, which was not yet available for use in the previous study (Ministry of Transport, Public Works, and Water Management, 2001). Because the NWB has no road function or traffic intensity data attached, we collected traffic intensity data for national, provincial and municipal roads that were then linked to the road network. We aimed to obtain both total traffic and truck traffic intensity for the years 1986–1996 (follow-up time of the study). More details about the collection of traffic intensity data and the linkage to the digital road network is described in the Appendix A.

Not all municipalities had traffic counts for their municipal roads, and municipalities that count traffic did not have traffic intensities for all roads in their municipality. To municipal roads without traffic intensities a background traffic intensity value of 1225 mvh 24 h$^{-1}$ was assigned assuming that these roads are not major. This was done to avoid underestimation of local scale exposure for example when calculating total traffic intensity in a buffer. A value of 1225 mvh 24 h$^{-1}$ was used, because this was half of the value of 2450 mvh 24 h$^{-1}$ that was used in Environmental Traffic Maps to distinguish between roads for which air pollution assessments were and were not made (Harms, 2000).

Traffic intensities were available for different years on different roads. For some roads data were available for the years in the period 1986–1996, while for other roads data were available for isolated years. Because of time trends in traffic intensities, the comparability between roads with data for different years was limited. Therefore, we developed extrapolation procedures to estimate...
traffic intensities for all years in the period 1986–1996 for which data were not available.

Continuous traffic variables for the local component of exposure were estimated using GIS calculations:

- Total daily average traffic intensity in 1986 on the nearest (mvh 24 h\(^{-1}\)) and distance to this road. If a road with traffic intensity >10 000 mvh 24 h\(^{-1}\) was located within 30 m of the nearest road, the traffic intensity and distance to this road was used, covering situations where a major road was located close to a minor road (e.g. when roads run parallel).
- Total daily average traffic intensity in 1986 on the nearest major road (mvh 24 h\(^{-1}\)) (defined as a road with more than 10 000 mvh 24 h\(^{-1}\) and distance to this road.

Because of GIS calculation limitations, distances to nearby (major) roads were calculated with a maximum distance of 500 m.

- The sum of the total daily average traffic intensity in 1986 (mvh 24 h\(^{-1}\)) in a buffer of 100 m and in a buffer of 250 m around each home address was calculated using buffer calculations. Each buffer was divided in 10 × 10 m cells and for each cell the traffic intensity was assessed. Then, the traffic intensities of all cells within the specified buffer were added together. A buffer of 100 m was used because especially in the densely built Dutch cities, the largest contrast in traffic-related air pollutants is likely to occur over distances less than 100 m (Briggs et al., 1997).

Further, quantitative estimates of the local component concentrations were estimated using regression models. Because there was only a limited number of traffic sites in the NAQMN which were concentrated in urban areas (only one motorway site), data from monitoring sites of the TRAPCA study (Brauer et al., 2003) were used. Because the TRAPCA sites were not located close to motorways, they were used to estimate the local concentrations without taking into account the effect of motorways. To estimate the additional local concentrations caused by nearby motorways a Dutch study on air pollution near motorways was used (Janssen et al., 2001).

In the TRAPCA study, 40 monitoring sites were selected throughout the Netherlands. At each site NO\(_2\), PM\(_{2.5}\) and PM\(_{2.5}\) filter absorbance were measured for four 2-week periods distributed over the period February 1999 and July 2000 (Brauer et al., 2003). PM\(_{2.5}\) filter absorbance at TRAPCA sites was transformed into BS concentrations using the equation: BS (\(\mu g\) m\(^{-3}\)) = −3.663 + 9.897*PM\(_{2.5}\) filter absorbance (Roorda-Knape et al., 1998). Then, NO\(_2\), BS and PM\(_{10}\) background concentrations (sum of regional and urban component) were estimated for the TRAPCA sites using the methods described above. PM\(_{2.5}\) concentrations have not been measured in the NAQMN, therefore, estimated background PM\(_{10}\) concentrations at TRAPCA sites were transformed into PM\(_{2.5}\) concentrations using the formula: PM\(_{2.5}\) = 0.6739∗PM\(_{10}\)−0.1038 (Cyrys et al., 2003). The residual (traffic) NO\(_2\), BS and PM\(_{2.5}\) concentrations at each site were calculated (measured concentration minus estimated background concentration), and regression models were developed with the sum of traffic intensity, excluding the traffic intensities of motorways, in a buffer of 100 m as predictor. Because traffic intensities were not available for all municipalities, 22 sites were used to develop regression models. For addresses where the sum of traffic intensity in a 100 m buffer was higher than at the TRAPCA site with the highest buffer intensity (0.2%), the sum of traffic intensity in a 100 m buffer of this TRAPCA site was used as predictor variable, as unrealistically high local contributions were obtained for these addresses.

In the Dutch motorway study, weekly averaged measurements of NO\(_2\), PM\(_{2.5}\), and BS concentrations were conducted at 24 schools located within 400 m of a motorway in the Netherlands. Measurements were conducted 5–10 times in the period 1997–May 1998 (Janssen et al., 2001). For the 24 schools we assessed the truck traffic intensity on and the distance to the nearest motorway. Regression models were developed with the measured concentrations at these 24 sites as dependent variables and truck traffic intensity on nearest motorway and distance to this motorway in three categories (<100, 100–300, and >300 m (reference category)) as predictor variables. For addresses located further away than 500 m from a motorway, the quantitative local component concentration caused by nearby motorways was set to 0 \(\mu g\) m\(^{-3}\).

Both estimated quantitative local component concentrations were added to the estimated background concentration.

GIS analyses were conducted using ArcGis at the National Institute of Public Health and the
Environment (RIVM). Statistical analyses were performed using SAS 8.02.

3. Results

3.1. Regional component

We evaluated ordinary kriging and inverse distance weighed interpolation as methods to estimate regional background concentrations at the home addresses. However, there was only a limited number of BS sites (9 sites in the period 1992–1996), and therefore it was not possible to develop a suitable variogram for kriging. Further, the RIVM has used kriging procedures to decrease the density of the network during the rearrangement of the network in 1986. Therefore, it was less possible to estimate variograms for kriging for the periods after 1986.

For NO2 and SO2 both interpolation and kriging were possible and correlations between estimates with inverse distance weighed interpolation and kriging were all $\geq 0.92$, supporting that both methods performed similarly. To be consistent for all air pollutants, inverse distance weighed interpolation was used for all pollutants.

We studied how the prediction error depended on the distance criterion for inclusion of network sites. Prediction errors were estimated by cross-validation: the concentration at sites was estimated using data from the other network sites only and compared with the actually measured concentration. Table 1 lists the prediction errors related to interpolation (average root mean square error (RMSE) over the different years). For NO2 and NO, RMSE was lowest when distance criterion was 75 km. However, not all home addresses were located within 75 km of a monitoring site. Therefore, we used for NO2 and NO a distance criterion of 100 km (RMSE values for 75 km and 100 km were only slightly different). Distance criteria of 75 km (SO2) and 100 km (NO2, NO and BS) were used to estimate regional background concentrations using inverse distance weighed interpolation.

3.2. Urban component

In Table 2, regression models to estimate the urban component are shown for NO2, NO, BS and SO2. As a sensitivity analysis we used for NO2 a manual forward selection process, with as inclusion criterion whether the t-statistic of a variable was statistically significant, resulting in the same predictor model.

In Table 3, the $R^2$-values and RMSE of the regression models are shown together with the results of the regression models based on the previous study method (Hoek et al., 2001). The $R^2$-values are higher and the RMSE values are lower for the current study compared to the previous study.

Table 1

<table>
<thead>
<tr>
<th>Distance criterion (km)</th>
<th>NO2</th>
<th>BS</th>
<th>NO</th>
<th>SO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>4.38</td>
<td>2.14</td>
<td>5.23</td>
<td>4.08</td>
</tr>
<tr>
<td>75</td>
<td>3.89</td>
<td>1.56</td>
<td>4.67</td>
<td>4.03</td>
</tr>
<tr>
<td>100</td>
<td>3.93</td>
<td>1.40</td>
<td>4.70</td>
<td>4.04</td>
</tr>
<tr>
<td>150</td>
<td>4.12</td>
<td>1.69</td>
<td>4.79</td>
<td>4.10</td>
</tr>
<tr>
<td>200</td>
<td>4.25</td>
<td>1.80</td>
<td>4.87</td>
<td>4.14</td>
</tr>
<tr>
<td>No criterion</td>
<td>4.32</td>
<td>1.83</td>
<td>4.91</td>
<td>4.17</td>
</tr>
</tbody>
</table>

$^a$RMSE in $\mu g m^{-3}$, derived from cross-validation.

Table 2

Regression models for estimating the urban component for each pollutant based on residual concentration at urban and regional sites and GIS derived predictor variables: parameter estimates and standard errors (SE) ($N =$ number of observations on which regression model is based)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO2 ($N = 202$)</td>
<td>Intercept</td>
</tr>
<tr>
<td></td>
<td>(Number of inhabitants/1000) in a 5000 m buffer</td>
</tr>
<tr>
<td></td>
<td>Located in a non-rural area$^a$</td>
</tr>
<tr>
<td></td>
<td>Located in the center of a town$^a$</td>
</tr>
<tr>
<td>NO ($N = 203$)</td>
<td>Intercept</td>
</tr>
<tr>
<td></td>
<td>(Number of inhabitants/1000) in a 5000 m buffer</td>
</tr>
<tr>
<td></td>
<td>Located in a non-rural area$^a$</td>
</tr>
<tr>
<td>BS ($N = 23$)</td>
<td>Intercept</td>
</tr>
<tr>
<td></td>
<td>(Number of inhabitants/1000) in a 1000 m buffer</td>
</tr>
<tr>
<td>SO2 ($N = 508$)</td>
<td>Intercept</td>
</tr>
<tr>
<td></td>
<td>Located in a non-rural area$^a$</td>
</tr>
<tr>
<td></td>
<td>Located in an urban area$^a$</td>
</tr>
<tr>
<td></td>
<td>Located in an industrial area$^a$</td>
</tr>
</tbody>
</table>

$R^2$- and RMSE-values of the models are described in Table 3.

$^a$Indicator variable with $0 = \text{no}$ and $1 = \text{yes}$. 

The correlations between the average background concentrations (sum of regional and urban component) in different 5-year periods at the home address of study participants are shown in Table 4. The correlation between estimated concentrations for different periods is high. Because we added the same urban component to the regional component for each of the 5-year periods, this could result in artificially high correlations between background concentrations of different periods. However, this was not the case since correlations for annual measured concentrations of network sites were high as well (for NO\textsubscript{2}>0.91; for NO>0.85; for BS>0.95; and for SO\textsubscript{2}>0.84). In Fig. 2, the trend in background concentrations at the home addresses is shown for the components NO\textsubscript{2}, NO, BS, and SO\textsubscript{2}. SO\textsubscript{2} concentrations show a clear decrease over time, while for NO\textsubscript{2} and NO concentrations a more gradual decline is shown. BS concentrations appear to be more or less stable over time in the period 1976–1996, but showed a clear decline over a longer period of time.

### 3.3. Local component

#### 3.3.1. Completeness

For all national roads (for all years in the period 1986–1996) and all provincial roads (although not for all roads for all years in the period 1986–1996) total traffic intensity data were obtained. For municipal roads, total traffic intensity data were obtained for 121 of the in total 204 municipalities (59%). Especially for municipalities with a small number of participants, no data were available. These 121 municipalities contained 17,912 subjects, i.e. approximately 82% of the total number of subjects in the study (N=21,868). The percentage of municipal roads for which traffic intensity data were available was 14.3%. For most of these municipalities, to a large part of the municipal roads a background traffic intensity value of 1225 mvh\textsuperscript{24}h\textsuperscript{-1} was therefore assigned. For all national roads truck traffic intensity data were available. For the municipal roads and provincial roads for which total traffic intensity data were available, truck traffic intensity data were available for only 19% and 23%, respectively.

Of the in total 21,868 addresses, 17,592 (80.5%) had an assigned traffic intensity on nearest road of 1225 mvh\textsuperscript{24}h\textsuperscript{-1}. A total of 9632 addresses were located in municipalities with more than 100,000 inhabitants, and 7282 of these addresses (75.6%) had an assigned traffic intensity on nearest road of 1225 mvh\textsuperscript{24}h\textsuperscript{-1}. Of the 12,236 addresses in municipalities with less than 100,000 inhabitants, 10,310 addresses (84.3%) had an assigned traffic intensity on nearest road of 1225 mvh\textsuperscript{24}h\textsuperscript{-1}. These data show that for the larger municipalities for a higher percentage of the roads traffic intensity data were available compared with small municipalities.

#### 3.3.2. Temporal trends

In Fig. 3, the correlation between 1996 and 1986 traffic intensities on municipal, provincial and national roads is shown. The regression equations

<table>
<thead>
<tr>
<th>Component</th>
<th>NLCS-AIR method</th>
<th>Previous study method\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$ (%)</td>
<td>RMSE (\textmu{g} m\textsuperscript{-3})</td>
</tr>
<tr>
<td>NO\textsubscript{2}</td>
<td>67.3</td>
<td>3.57</td>
</tr>
<tr>
<td>NO</td>
<td>45.9</td>
<td>5.64</td>
</tr>
<tr>
<td>BS</td>
<td>49.2</td>
<td>1.22</td>
</tr>
<tr>
<td>SO\textsubscript{2}</td>
<td>34.8</td>
<td>3.23</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Address density of the four-digit postal code area used as predictor variable.

### Table 3

Comparison between performance of new exposure assessment (NLCS-AIR study) and previous study exposure assessment for estimation of the urban component ($R^2$-value and Root Mean Squared Error (RMSE)) (predictor variables for the different components in the NLCS-AIR method are described in Table 2)

<table>
<thead>
<tr>
<th>Component</th>
<th>NLCS-AIR method</th>
<th>Previous study method\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$ (%)</td>
<td>RMSE (\textmu{g} m\textsuperscript{-3})</td>
</tr>
<tr>
<td>NO\textsubscript{2}</td>
<td>67.3</td>
<td>3.57</td>
</tr>
<tr>
<td>NO</td>
<td>45.9</td>
<td>5.64</td>
</tr>
<tr>
<td>BS</td>
<td>49.2</td>
<td>1.22</td>
</tr>
<tr>
<td>SO\textsubscript{2}</td>
<td>34.8</td>
<td>3.23</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Address density of the four-digit postal code area used as predictor variable.
in Fig. 3 show that the increase in traffic intensity over time is substantially larger for national and provincial roads compared with municipal roads. Correlations between 1996 traffic intensities and traffic intensities from other years were all > 0.92 for the different road types, supporting the use of traffic intensities from one year to represent a long-term average. Because municipal data were not always available for the same year, trends in traffic intensity were estimated using roads with traffic intensity data for several years and mixed modeling procedures. The estimated trends (percentages increase per year) in traffic intensities on national, provincial and municipal roads within different traffic intensity classes are shown in Table 5. A linear trend for the whole period was assumed. These trends were used to estimate traffic intensities for roads that did not have data for all years in the period 1986–1996.

3.3.3. Distributions in cohort

In Table 6 the distribution of the traffic intensities on the nearest road is shown for addresses located at roads with traffic intensities higher than 1225 mvh 24 h⁻¹, for all addresses together and separately for addresses in municipalities with more than and less than 100,000 inhabitants. Table 6 shows that roads in municipalities with more than 100,000 inhabitants for which traffic intensity data were available had on average a higher traffic intensity compared with addresses in municipalities with less than 100,000 inhabitants.

The average distance to the nearest road was 19 m, with a minimum of 0 m and a maximum of 457 m. For only three subjects the nearest road was located further than 500 m away.

Of the 21,868 subjects 11,229 (51.3%) lived within 500 m of a major road, defined as a road with traffic intensity > 10,000 mvh 24 h⁻¹ in 1986. The average distance to the nearest major road was for these 11,229 subjects 233 m, and the minimum and maximum distance were 5 and 500 m, respectively. The percentage of participants who lived within 50 m of a major road was 4.7%, and 10.8% lived within 100 m of a major road.
Table 7 shows the distribution of the sum of total traffic intensity in 1986 in a buffer of 100 m around the home addresses of the study participants. The distributions of the traffic variables were all highly skewed. The correlation between the traffic intensity on the nearest road and the nearest major road was 0.14, while the correlations with the sum of traffic intensity in a 100 m buffer were 0.64 and 0.20 for the traffic intensity on the nearest road and the nearest major road respectively.

3.3.4. Quantitative estimates

In the regression models for quantitatively estimating the local component without taking into account air pollution effects of nearby motorways, the regression coefficient of the local component
was 1.61 (SE 0.60) \( \mu g m^{-3} \) per 100,000 mvh in a 100 m buffer for NO\(_2\), and the \( R^2 \) of the model was 26.3% (RMSE 5.70). For BS the regression coefficient was 1.84 (SE 0.38) \( \mu g m^{-3} \) per 100,000 mvh in a 100 m buffer with an \( R^2 \) of 54.5% (RMSE 3.54). For PM\(_{2.5}\), the regression coefficient was 0.50 (SE 0.23) \( \mu g m^{-3} \) per 100,000 mvh in a 100 m buffer with a \( R^2 \) of 18.7% (RMSE 2.21). This resulted in average predicted local component concentrations (minimum; maximum) of 2.3 \( \mu g m^{-3} \) (0; 14.4 \( \mu g m^{-3} \)) for NO\(_2\); 1.9 \( \mu g m^{-3} \) (0.1; 7.7 \( \mu g m^{-3} \)) for BS, and 2.2 \( \mu g m^{-3} \) (0.2; 7.0 \( \mu g m^{-3} \)) for PM\(_{2.5}\).

### 3.4. Performance of the exposure assessment method

Since other studies (Brauer et al., 2003; Jerrett et al., 2005a) have reported \( R^2 \)-values incorporating the three spatial scales in one overall regression method, we illustrate the overall performance of our exposure assessment method by developing one overall regression model for each air pollutant with average concentrations for the period 1987–1991 measured at both regional background, urban background and street monitoring sites in the NAQMN as dependent variables. For all pollutants, the regional component was described by an indicator variable (North, East, West, and South) as proxy for the interpolation estimates, which cannot be used as predictors in the model. The urban
component was modeled using the predictor variables for the different pollutants as described in Table 2. The sum of traffic intensity, excluding the traffic intensity of motorways, in a 100 m buffer around a home address, and truck traffic intensity on nearest motorway and distance to this motorway in three categories were used as predictor variables for the local component for all pollutants. However, none of the NAQMN sites in the period 1987–1991 was located within 500 m of a motorway, and therefore the contributing effect of nearest motorway could not be evaluated. The models explained 84%, 44%, 59%, and 56% of the variability in concentrations for NO2 (N = 36 sites), NO (N = 36 sites), BS (N = 16 sites) and SO2 (N = 81 sites), respectively. One monitoring site was located in a street canyon with extremely high measured concentrations for NO (248 µg/m³) and BS (63 µg/m³). Excluding this site resulted in models which explained 67% and 87% of the variance of NO and BS, respectively.

Fig. 4 shows the resulting distributions for the regional, urban, and local component for NO2 and BS for the period 1987–1991 at addresses of participants. Further the distributions of the background concentrations (sum of regional and urban component) and total concentration (sum of regional, urban and local component) are shown. The estimated NO2 and BS concentrations vary substantially over the home addresses of the study participants. The regional component contributes most to the total estimated concentration. Because only a small percentage of the study participants lived close to a major road, the effect for the local component is mostly in the highest concentrations.

4. Discussion

We described an improved method to estimate long-term outdoor air pollution concentrations in a Dutch cohort study. Using comprehensive data from existing air pollution monitoring stations, data on land use, population density, road networks and traffic density, we were able to develop exposure assessment models that explained a large part of the spatial variance in long-term averaged air pollution concentrations.

4.1. Spatial components

Ordinary kriging and inverse distance weighed interpolation procedures were evaluated as methods to estimate the regional component concentration. In general, all interpolation methods perform similarly when the data density is high and distributed relatively uniformly across the study area (Elliot et al., 2000). This similarity in performance was also shown by the high correlations (>0.92) between estimates assessed with ordinary kriging and inverse distance interpolation procedures for those pollutants with large number of observations (NO2 and SO2).

Compared to our previous study the regression models to estimate the urban component concentrations were improved by (1) using varying buffer sizes around the exact home addresses instead of the postal codes, and (2) using a larger number of potential predictor variables. The results showed that the models of the current study resulted in more
precise estimates of the urban component concentrations compared with the previous study. In the SAVIAH study (Briggs et al., 1997) and a study by Stedman et al. (1997) NO2 concentrations were also estimated using regression modeling. In both studies land cover variables were used, i.e. area of built up land and percentage of urban and suburban land cover, which are equivalent with the variables we used. PM2.5 filter absorbance, which can be transformed into BS concentrations, was predicted in the TRAPCA study (Brauer et al., 2003). In the TRAPCA regression models address density was used as predictor variable for the estimation of the urban component exposure, which is similar with our urban component predictor variable for BS. In contrast with the other air pollutants, the number of inhabitants around a site was no predictor variable for SO2 concentrations. However, whether a site was located in an industrial location was a significant predictor variable for SO2, which is consistent with the fact that SO2 is mainly emitted by industrial sources in the Netherlands.

Average background air pollution concentrations showed a decline during the measurement period. Average concentrations in different time periods were highly correlated ($r>0.83$). A recent study by Pope et al. (2002) showed also a general decline in particulate air pollution in the United States, with a high correlation of 0.78 between average concentrations in the period 1979–1983 and the period 1999–2000.

Exposure was estimated for the 1986 home address, as this is the baseline-address. Address history during the follow-up period was available for subcohort members and cases. For subcohort members, bi-annual follow-up questionnaires were used to determine residential history. The residential information of deceased subjects was available from the Central Bureau of Genealogy (CBG), which has an archive of municipal registration cards of subjects that deceased before October 1994. For persons that died after October 1994, the CBG has computer files on deaths containing however mostly incomplete address information. For most of these deceased subjects only the baseline address and the last address were available. Therefore, we did not assess total exposure for the entire period by estimating concentrations at different residential addresses. Approximately 30% of the participants moved between 1986 and end of follow-up. To take into account the effect of moving in the data-analysis, we used an indicator variable indicating whether a subject lived at the baseline address for the complete follow-up period.

Particulate matter was mainly characterized by BS. We had limited possibilities to evaluate PM10 and PM2.5 exposures. Measurements of PM10 in the NAQMN started only in 1992, while PM2.5 was not measured during the study period. PM10 measurements were thus available for only approximately half of the follow-up period. Measurements in the framework of the NAQMN of PM10 have documented that the contrast in PM10 concentrations in the Netherlands is relatively small (van der Wal et al., 2000). Several studies have shown that BS or elemental carbon is better related to traffic than PM10 and PM2.5 (Janssen et al., 2001).

Traffic intensity is frequently used as an indicator for exposure to traffic-related air pollution (Wjst et al., 1993; van Vliet et al., 1997; Venn et al., 2001; Janssen et al., 2003). One criticism of using traffic intensity as indicator of exposure is that data are often only available for only one point in time and not available for the relevant time period of interest (Nuckols et al., 2004). Obtaining retrospective data on traffic intensity is laborious. We illustrated, however, that traffic intensity data obtained for different years were highly correlated, even over periods of 10 years.

A background traffic intensity value was assigned to roads for which no traffic data were available assuming that these were minor roads. This may have resulted in some (non-quantifiable) misclassification of exposure since some roads may actually have substantial counts. In the epidemiological study this may result in an underestimation of the effect because the ‘low exposed’ group contains some ‘exposed’ subjects. Because the vast majority (90–95%) of subjects is ‘low exposed’, this impact is likely small.

The trend in traffic intensity on municipal roads showed only a slight relative increase compared with the relative increase in traffic intensity on provincial and national roads. This suggests that the busy roads outside the city (provincial and national roads) became busier, but that traffic intensities on roads within cities did not change much. This is due to the nation-wide introduction of regulations that prevent traffic in the more centrally located areas in municipalities and the increased development of employment facilities and other facilities at the borders of municipalities in the Netherlands (Harms, 2000). Municipal roads with traffic intensities $>4000$ mvh day$^{-1}$ showed a
percentage increase in traffic intensity of 1% per year, while the traffic intensity on municipal roads with traffic intensities <4000 mvh 24 h⁻¹ hardly increased. This is caused by municipal traffic regulations that prevent through traffic on minor municipal roads, with as consequence a concentration of traffic on the municipal main roads (Harms, 2000).

Several Dutch urban traffic studies showed that concentrations of traffic-related air pollutants are substantially higher near busy roads compared to roads with low traffic intensities or at urban background locations (Janssen et al., 1997; Fischer et al., 2000; Wichmann et al., 2005). We used therefore as first traffic variable the traffic intensity on the nearest road. This variable is however only based upon the characteristics of the nearest road, while a nearby busy road may also be important for local air pollution concentrations. Exposure studies near motorways showed that traffic-related air pollution concentrations depend on the traffic intensity on and the distance to motorways (Rooda-Knape et al., 1998; Janssen et al., 2001). As a second traffic variable we used therefore the traffic intensity of the nearest major road, defined as a road >10,000 mvh 24 h⁻¹, with the distance to this nearest major road. However, prediction of pollution values may be difficult in areas where there are a large number of intersecting line sources or where roads run parallel. Therefore, we used as third traffic variable the sum of traffic intensity in buffers of 100 and 250 m around each home address. These variables use all traffic intensities within the specified buffer. A buffer of 100 m was used, because especially in urban areas with specific local topographic and building conditions marked variations in pollution levels can occur over distances less than 100 m (Briggs et al., 1997). Using buffer calculations has however limitations, e.g. the assumption that the dispersion pattern is the same in all directions may be violated because of wind patterns or topographic conditions (Jerrett et al., 2005a).

The used digital road network was from the year 2001 while home addresses of 1986 were used. The road network may have changed between 1986 and 2001. However, we focused mainly on busy roads that have likely been in place between 1986 and 2001. We used traffic intensities of 1986, while, for most municipal roads, these were estimated using traffic intensities of more recent years. The mixed modeling procedures to estimate trends in traffic intensities showed that for most roads the trend was not significantly different from the averaged fixed trend. We use therefore the assumption that while absolute traffic intensities may have changed, relative differences are likely to be stable. In addition, the 1986 addresses were geocoded with a database from 2000 and therefore the coordinates and digital road network are compatible.

### 4.2. Overall performance

The overall regression models predicted a large part of the variation in air pollution concentrations with explained variances of 84%, 44%, 59%, and 56% for NO₂, NO, BS, and SO₂, respectively. This is probably a (slight) underestimation because in the overall regression models an indicator variable for region was used as proxy for the interpolation results. It was not possible to predict concentrations for locations in street canyons. One monitoring site was located in a street canyon with a moderately high traffic intensity of 7842 mvh 24 h⁻¹, but with extremely high measured concentrations for NO (248 μg m⁻³) and BS (63 μg m⁻³). Excluding this site resulted in models with substantially higher explained variances of the NO (67%) and BS (87%) concentrations.

Comparable to other studies which developed regression models to estimate exposure to air pollution using GIS derived variables, we found similar results (Briggs et al., 1997; Brauer et al., 2003; Ross et al., 2005). The SAVIAH study used three key variables (traffic volume/road type, land cover and altitude) as predictors of NO₂ concentrations in three European cities (Amsterdam (NL), Huddersfield (UK), Prague (CR)). The explained variances of the annual mean NO₂ concentrations were 79%, 82% and 87% for the three cities, respectively (Briggs et al., 1997). In the multicenter TRAPCA study, regression models using traffic-related variables were developed to estimate exposure to fine particles and filter absorbance. Regression models explained 81%, 67%, and 66% of the variability in annual average filter absorbance for the Netherlands, Munich and Stockholm, respectively (Brauer et al., 2003). These studies show that a large part of long-term averaged air pollution concentrations can be explained using a regression model with land use and traffic-related predictor variables and that it is a useful method to estimate exposure to air pollution for a large study population. However, to estimate concentrations...
using a regression-based approach both geographically relevant GIS data and concentration data from monitoring sites have to be available. These data may, however, not be available in all countries. GIS-based exposure assessment can be useful, but only when geo-referenced data are available at a relevant scale and with sufficient geographical accuracy (Nuckols et al. 2004).

5. Conclusion

An improved method for assessment of long-term outdoor air pollution concentrations taking into account small-scale variations in air pollution concentrations was illustrated. Despite the limitations and methodological problems, we were able to develop an exposure assessment model using GIS-methods and traffic intensities that explained a large part of the variations in concentrations outdoor air pollutants and which can be used in epidemiologic studies to estimate air pollution levels at specified locations.

Appendix A. Collection of and linkage of traffic intensity data to the digital road network

Traffic intensity data were collected for the different types of roads in the Netherlands: national, provincial and municipal roads. The road network in the Netherlands has a total length of more than 100,000 km. The national roads and provincial roads have a total length of approximately 3200 and 9000 km, respectively. The largest part of the total road network is thus the responsibility of the municipalities. Traffic intensity data were not readily available for all roads. Therefore, we put a major effort in obtaining as much traffic data as possible. For national roads, traffic intensity data for the period 1986–1996, both total traffic and truck traffic, were available from the Transport Research Center of the Ministry of Transport, Public Works, and Water Management. Data were available for all road sections between two junctions. Although traffic intensities are counted on all provincial roads, traffic intensity data for provincial roads were not available in a central database. Therefore, we contacted all 12 provinces in the Netherlands individually to obtain total traffic and truck traffic intensity data for the years 1986–1996. Obtaining traffic intensity data for municipal roads was the most difficult, especially for small municipalities. To obtain data on total and truck traffic densities on municipal roads for the years 1986–1996 we contacted all 204 municipalities included in the study. Another source of municipal traffic intensity data were Environmental Traffic Maps. In the Netherlands, municipalities with more than 40,000 inhabitants are obliged to develop Environmental Traffic Maps to assess exposure to air pollution and noise for all roads with relevant traffic intensities defined as 2450 motorvehicles 24 h\(^{-1}\) (Harms, 2000). For municipalities with an Environmental Traffic Map traffic data were available for all major roads.

For some provincial and municipal roads no average whole-week traffic intensities, but only average workday (Monday–Friday) or average daytime traffic intensities (traffic intensity between 7.00 AM and 7.00 PM) were available. We determined factors to convert these intensities to average whole-week traffic intensities by calculating ratios between whole-week and workday traffic intensities. For approximately 7% of the municipalities and for four of the 12 provinces only workday traffic intensities (Monday–Friday) were available. There are no standards for the conversion of average workday intensities into average whole-week traffic intensities (Monday–Sunday). However, for five municipalities and four provinces both average workday and whole-week traffic intensities were available for several years. Average ratios between whole-week and workday traffic intensities were calculated. To this end, the whole-week traffic intensities were expressed in counts per day as for the workday intensities. The overall average ratios for municipal and provincial roads were 0.94 (SD 0.03) and 0.93 (SD 0.02), respectively. The ratios did not change over time and did not depend on the size of a municipality. Further, the ratio was not related to the workday traffic intensity. Average workday traffic intensities were therefore multiplied by 0.94 and 0.93 for municipal and provincial roads, respectively, to estimate daily average whole-week traffic intensities. Further, for two municipalities only daytime traffic intensities (traffic intensity between 7.00 AM and 7.00 PM) were available. From the literature information was available about the percentage of the total daily traffic intensity that occurs during daytime hours on municipal roads in the Netherlands; 77.8% of the total daily traffic intensity occurred during daytime hours (Dassen et al., 2000) We assumed that this percentage did not change over time and that the percentage is independent of the traffic intensity on a road. For
municipal roads daytime traffic intensities were therefore multiplied by 1.29 ( = 1/0.778) to calculate an average whole-week traffic intensity.

Average whole-week traffic intensities for the years 1986–1996 were linked to the National Road Database (Nationaal Wegen Bestand (NWB), based on road name and road number. The NWB includes all streets and roads in the Netherlands that have a street name and/or road number. This means that more than 98% of the Dutch roads have been included. More than 95% of all road sections in NWB have a maximum location difference of 10 m compared with the true location (Ministry of Transport, Public Works, and Water Management, 2001).

References


Stedman, J.R., Vincent, K.J., Campbell, G.W., Goodwin, J.W.L., Downing, C.E.H., 1997. New high resolution maps of estimated background ambient NO\textsubscript{x} and NO\textsubscript{2} concentrations in the UK. Atmospheric Environment 31, 3591–3602.


