Published in final edited form as:

Atmos Environ. 2010 December 1; 44(39): 5156–5164. doi:10.1016/j.atmosenv.2010.08.058.

Modeling effects of traffic and landscape characteristics on ambient nitrogen dioxide levels in Connecticut

Katherine J. Skene, Janneane F. Gent, Lisa A. McKay, Kathleen Belanger, Brian P. Leaderer, and Theodore R. Holford

Center for Perinatal, Pediatric and Environmental Epidemiology Department of Epidemiology and Public Health Yale University School of Medicine New Haven, CT USA

Abstract

An integrated exposure model was developed that estimates nitrogen dioxide (NO₂) concentration at residences using geographic information systems (GIS) and variables derived within residential buffers representing traffic volume and landscape characteristics including land use, population density and elevation. Multiple measurements of NO₂ taken outside of 985 residences in Connecticut were used to develop the model. A second set of 120 outdoor NO₂ measurements as well as cross-validation were used to validate the model. The model suggests that approximately 67% of the variation in NO₂ levels can be explained by: traffic and land use primarily within 2 km of a residence; population density; elevation; and time of year. Potential benefits of this model for health effects research include improved spatial estimations of traffic-related pollutant exposure and reduced need for extensive pollutant measurements. The model, which could be calibrated and applied in areas other than Connecticut, has importance as a tool for exposure estimation in epidemiological studies of traffic-related air pollution.

Keywords

traffic-related air pollution; exposure modeling; outdoor nitrogen dioxide; traffic volume; land use

1. Introduction

Outdoors, traffic is the primary source of nitrogen oxides (NO_x) that convert to nitrogen dioxide (NO₂). Traffic is believed to be responsible for at least half of NO_x emissions in general and accounts for a higher proportion in urban areas (HEI, 2010). NO₂ is also a precursor for several other air pollutants including nitric acid and ozone (WHO, 2003). Of particular relevance when assessing exposure to NO₂ is the observation that levels can vary significantly over small distances (Jerrett et al., 2005; WHO, 2003).

© 2010 Elsevier Ltd. All rights reserved

Corresponding author: Janneane F. Gent, PhD Yale Center for Perinatal, Pediatric and Environmental Epidemiology One Church Street, 6th Floor New Haven, CT 06510 USA Tel: 203-764-9375 Fax: 203-764-9378 janneane.gent@yale.edu.

Mailing address for all authors: Yale Center for Perinatal, Pediatric and Environmental Epidemiology One Church Street, 6th Floor New Haven, CT 06510 USA katherine.skene@yale.edu janneane.gent@yale.edu lisa.mckay@yale.edu kathleen.belanger@yale.edu brian.leaderer@yale.edu theodore.holford@yale.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Residential exposure to traffic or to specific traffic-related pollutants has been assessed using a variety of measurements including traffic proximity (self-reported (Duhme et al., 1996) or measured distance (Brunekreef et al., 1997)), traffic volume (Lin et al., 2002; Livingstone et al., 1996), traffic-related pollutant levels measured at central site monitors (Gent et al., 2009), and estimates from complex chemical transport models (Byun and Schere, 2006; Jerrett et al., 2005). Each method has limitations. Measures using distance (self-reported or measured), or traffic counts, for example, can result in exposure misclassification by failing to capture variability in traffic volume on road networks in the vicinity of a residence. Traffic volume is considered in some models that assume a pollutant dispersion function that is Gaussian (English et al., 1999; Pearson et al., 2000) or logarithmic (Pleijel et al., 2004). The most complicated chemical/dispersion models, e.g., CALINE4 (Bensen, 1989; US EPA, 2003), require detailed data inputs that are generally unavailable: pollution sources and emissions figures, fleet composition and fuel types, topographical information, atmospheric and meteorological data. The applicability of CALINE4 to locations and time periods other than California in the 1970s (where the model was developed and when the vehicle fleet had different pollution characteristics than the 2000s) is questionable.

Holford et al. (2010) developed a framework for estimating integrated exposure to traffic using publicly available data to estimate traffic volume in buffers surrounding a residence (Holford et al., 2010). No prior assumptions were made as to the pollutant dispersion functions, which were explored with step, polynomial, and spline models. Using a step function to predict NO₂ measured outside of 120 homes in Connecticut, Holford et al. (2010) present models of NO₂ as isotropic functions of traffic volume (i.e., independent of direction with respect to a residence) ($R^2 = 0.510$) and anisotropic functions (traffic volume in buffers divided into four quadrants north, east, south, west of a residence) ($R^2 = 0.513$). No additional covariates were considered. The model presented here extends these results by using many more NO₂ samples taken over a larger area and longer time span and by including landscape characteristics (land use, elevation, population density) and season of sampling. We used NO₂ measurements at 985 residences throughout the state of Connecticut combined with other readily available, geographically-based information to develop a model to predict residential outdoor NO₂ concentration.

2. Variables used for model development

Geographic data used to create variables for model development (e.g., residential locations of nitrogen dioxide [NO₂] measurements, traffic counts on interstate and numbered highway road segments, land use, population density and elevation) were processed with ESRI's ArcGIS 9.3.1 (ESRI, 2006), using Universal Transverse Mercator (UTM) Zone 18 North projection. Locations of NO₂ samples used for model development were determined by taking Global Positioning System (GPS) measurements at study subjects' homes. Trained research assistants using GPS devices (Garmin® models *eTrex Legend Cx* and *Nuvi 760*) were instructed to stand at the front door (or nearby in the case of heavy ground cover at the front door, i.e., porch or trees), wait for an indication of a minimum of three satellites, then record the latitude and longitude. Addresses of NO₂ sample sites used in model validation were geocoded to point locations using the ESRI StreetsUSA database (ESRI, 2003) to obtain latitude and longitude coordinates. Hawth's Tools extension for ArcGIS was used to derive land use covariates (Beyer, 2004).

2.1. Residential nitrogen dioxide [NO₂] measurements

NO₂ levels were sampled outside of 985 residences of families participating in a year-long, follow-up study of air pollution and childhood asthma (Leaderer). The study took place in Connecticut, began in April, 2006, and concluded in August, 2009. NO₂ samples, up to four at each residence (one per season) were taken using Palmes tubes (Palmes et al., 1976) placed

outside the back door of the home and left in place for one month. Sampler analysis resulted in a mean daily NO₂ concentration for the monitoring period. A total of 3140 samples were available for use in model development: 508 (52%) contributed 4 samples, 262 (27%) contributed 3 and 22% contributed 2 (N=107) or 1 (N=108). Similarly collected NO₂ data from Connecticut residents participating in a study conducted in 1994 were used to validate the model and assess its accuracy (Beckett et al., 2006; Pettigrew et al., 2004; Triche et al., 2002; Triche et al., 2005; Triche et al., 2006). The 120 NO₂ samples used here for model validation were used previously for our earlier model development (Holford et al., 2010). Sample collection date for each sample was defined as the mid-point of the monitoring period. This date was used in analyses involving season of sampling (see section 3, below).

2.2. Traffic measurements

In an approach described by Holford et al. (2010) and using traffic data for 2006 provided by the Connecticut Department of Transportation (CT DOT), a line file of interstates and numbered highways was divided into approximately 50-meter segments, each of which had an associated average daily traffic (ADT) count (CT DOT, 2006). The midpoint of each segment was defined and a measure of vehicle-distance traveled on that segment was created by assuming that the contribution of this point source to NO₂ concentration was proportional to the product of segment length and ADT. Residential buffers were defined in 500 m steps within 1 km of each residence and in 1 km steps beyond 1 km (i.e., buffer 1 = 0 to 500 m, buffer 2 = 500 m to 1 km, buffer 3 = 1 km to 2 km, buffer 4 = 2 km to 3 km, ..., buffer 11 = 9 km to 10 km). Traffic variables used in model development were calculated by summing the contributions of all point sources within a buffer and dividing by 10 000 (in order to produce similar orders of magnitude for all model inputs). This created a measure of vehicle-kilometers (unit = 10 000 vehicle-kilometers) traveled within the range of distances that defined each buffer. For example, a value of 4.5 in a buffer would indicate an average of 45 000 vehicle-kilometers traveled per day within that buffer. Fig. 1 illustrates location of the ADT counts on road segments within a circular buffer surrounding a residence.

2.3. Other covariates

Land use variables—Land use data for Connecticut were obtained from the University of Connecticut's Center for Land Use Education and Research (CLEAR) (CLEAR, 2006). The most recent data available at the time of analysis showed land use in the year 2006, stored as a raster file with 30-meter pixels and classified into 12 categories of land use: developed, deciduous forest, coniferous forest, turf and grass, other grasses, agricultural field, water, non-forested wetland, forested wetland, tidal wetland, barren, utility rights-of-way (CLEAR, 2006).

Since all pixels in the file are of equal size, the proportion of a particular category of land use within a buffer is equal to the count of pixels of that land use category divided by the total pixels within a buffer. Land use measures were calculated using the 'Thematic Raster Summary (by polygon)' tool in the Hawth's Tools ArcGIS extension (Beyer, 2004). The sum of pixels of each category of land use within each buffer was multiplied by 900 (area in m² of a 30 m pixel) / 10 000 (area in acres of a hectare) to give the area in hectares of each category of land use within a buffer. Fig. 1 illustrates "pixels" within a residential buffer where different colors represent different land use categories.

Population density—Population density (per square mile) for each residence was obtained from a polygon layer of Connecticut census tracts from 2005 in ArcGIS (ESRI, 2005) and assigned to each residence as the mid-year census tract population divided by the area of the census tract in square miles.

Elevation—Elevation, in meters, was assigned to each residence according to its raster cell from the US Geological Survey (USGS) 'National Map' (Gesch, 2007; Gesch et al., 2002).

2.4. Prevailing wind

For the final model, residential buffers were divided into quadrants to reflect the cumulative prevailing wind during the course of the study (illustrated by the orange lines in Fig. 1). Meteorological data, including wind speed and direction, were obtained for four weather stations in Connecticut (Sikorsky Memorial Airport in Bridgeport, Municipal Airport in Danbury, Tweed Airport in New Haven and Bradley International Airport in Hartford)(NRCC, 2010). Using wind direction data from all four stations, a cumulative wind rose plot for the study period (April 2006–August 2009) was created by determining the distribution (as a percent of total study time) of prevailing wind from each of 16 directions (Lakes Environmental Software, 2008). Data for wind speed were also incorporated into the wind rose plot.

3. Model development

We developed a model for estimating exposure to residential NO₂, by integrating over highways and other landscape features in buffers surrounding a residence using publicly available data. Traffic-related NO₂ resulting from point (x,y) on a highway is proportional to traffic volume (ADT(x,y) at that point) and dispersion ($f[d(x,y)]$), which depends on distance ($d(x,y)$) to the residence. Traffic elements affecting level of NO₂ at a residence (illustrated in Fig. 1) are cumulated using a line integral,

$$\int_C \text{ADT}(x,y) \bullet f[d(x,y)] dx dy$$

where C represents all highways within a buffer (Holford et al., 2010). We modeled dispersion of NO₂ levels using a step-function, and also included as regressors integrated land use within concentric buffers, elevation, population density and season of sampling using linear regression (PROC REG in SAS (SAS Institute Inc., 2004)). Normal QQ plots were used to examine the distribution of residuals. No transformation of either the dependent or independent variables had any material effect on goodness of fit (R²). Initial models contained variables for traffic volume and land use representing each of the buffers surrounding a residence, population density for the census tract containing the residence and elevation at the residence. Isotropic models (independent of direction) and anisotropic models (dependent on direction, i.e., dependent on effects of prevailing wind) were considered. For anisotropic models, covariates for traffic were calculated for each residential buffer and directional quadrant combination.

Variable selection for the final model followed a rule that no more distant buffer could be included without the inclusion of all buffers of the same variable type (traffic volume or land use category) closer to the residence. For example, traffic in the 3 to 4 km buffer could not be included in the model without also including the covariates for traffic for all buffers within 3 km (0 to 500 m, 500 m to 1 km, 1 km to 2 km, and 2 km to 3 km). Similarly, 'developed' land use in the 1 km to 2 km buffer could not be included in the model without also including the covariates for 'developed' land for all buffers within 1 km (0 to 500 m, and 500 m to 1 km). The number of covariates was reduced by working from outer buffers inward, comparing nested models with F-tests, then retaining all variables contributing significantly (at the 5% level) to model fit. The use of multiple covariates for traffic volume and land use created the possibility of multicollinearity. Elimination of non-significant covariates from the furthest point from the residence and working inward established an inferential hierarchy in which near points more plausibly affected the response than those farther away.

To account for the effect of seasonality on level of NO₂, a cubic bspline function of date specifying six equally spaced knots was fit to the sampled NO₂ (PROC TRANSREG, SAS (SAS Institute Inc., 2004)) and was included as the seasonality covariate in one version of the final model. The bspline function is dependent on the specific time period of the NO₂ sampling study, thus, in order to adapt the model for predicting NO₂ levels at times outside of this specific time period, a more generalizable trigonometric (cosine and sine) and linear function of date was used to account for the effect of season in a second version of the final model.

By design, NO₂ samples were collected at each residence four times, once in each season over the course of the year-long study, and over three-fourths of the residences contributed samples from 3 or 4 seasons. Prior to fitting the final model, we examined correlations among residuals for measurements made at different time points and found low, but significant correlations (ranging from 0.13 to 0.24). Therefore, the final model was fit using repeated measures regression (PROC GENMOD with GEE [generalized estimating equations] (SAS Institute Inc., 2004)) specifying an “exchangeable” correlation structure.

To assess the validity of the model as a predictive tool, output from the (generalizable) final model was compared to 120 NO₂ measurements from a study conducted in Connecticut in 1994. Residential addresses were geocoded and land use and traffic covariates were developed using the methods described above. Parameter estimates used in the validation model were those obtained from the NO₂ samples collected from 2006–2009 with the exception of the intercept, which was adjusted to reflect the difference between mean NO₂ levels measured in each study. Differences between observed and predicted NO₂ values were examined. In addition, we performed cross-validation using a strategy of fitting the model excluding NO₂ observations for each residence in turn, then using the model to obtain predicted values for excluded observations.

In order to illustrate NO₂ levels predicted by the final model, we selected an area of the state with a range of traffic density and land use characteristics. Prediction maps of NO₂ concentrations for two dates (February 1 and August 1, 2009) were created for a 616 km² area around Hartford, Connecticut. A raster was created using 300 m × 300 m pixels then the latitude and longitude coordinates were calculated for the centroid of each pixel. Each of these 8030 points were treated as “residences” and covariates necessary for the final model (i.e., traffic volume for each buffer/quadrant combination, land use for each buffer, and for each point population density, elevation and date of prediction) were developed using the methods described above. The final model including covariates for these locations was used to obtain predicted NO₂ values. Predicted values were reassigned to the original raster grid which was resampled to a resolution of 30 m and displayed using a color ramp to indicate NO₂ concentration.

4. Results

Summary statistics for all covariates used to develop the step-function, dispersion model of NO₂ concentration in the state of Connecticut are shown in Table 1. The geographic distribution of average daily traffic (ADT) counts is displayed in Fig. 2. The unadjusted, isotropic dispersion function relating NO₂ to traffic volume in concentric buffers out to 10 km from a residence (unadjusted R² = 0.30) is shown in Fig. 3.

An examination of the wind rose representing the composite wind speed and direction for the study period reveals prevailing north/south winds (Fig. 4). Therefore, an anisotropic model was created by dividing buffers into four quadrants (north, east, south and west of the residence as indicated by the labeled quadrants in the compass inset in each panel of Fig. 5) and calculating traffic volume for each buffer/quadrant combination.

Prior to inclusion in the final model, categories of land use were collapsed from 12 to 3 according to their effect on NO₂ level in a model that included only land use variables: increased NO₂ levels were significantly associated with only the “developed” land category; decreased NO₂ levels were significantly associated with “forest/grass” categories (including deciduous forest, coniferous forest, turf and grass and other grasses); and no significant associations were found between NO₂ level and any other category (including agricultural field, water, non-forested wetland, forested wetland, tidal wetland, barren, utility rights-of-way). Geographic distributions of land use (“developed,” “forest/grass,” “other”), population density and elevation are displayed in Figs. 6, 7 and 8, respectively. Land use variables for “developed” and “forest/grass” land were calculated for each residential buffer prior to inclusion in the model. Model covariates for elevation and population density were calculated once for each residence.

Initial analyses using traffic volume variables calculated out to 10 km from the residence showed no significant improvement in model fit compared to using variables within 6 km, therefore covariates for the most distant residential buffers were dropped from further consideration. After variable selection, following the rules described in section 3, above, the final model incorporated significant traffic buffer/quadrant combinations, significant buffers for “developed” and “forest/grass” land as well as population density, elevation and variables representing season (Table 2).

Fig. 9 displays NO₂ concentrations measured over the course of the study as well as NO₂ as predicted by two different functions of date: a third-degree bspline with six knots (red line) and a trigonometric and linear function of date (green line), which were used to represent seasonality in model development and model prediction, respectively. The overall R² (adjusted) for the final model was 0.6728 using the bspline function of date to represent seasonality (Table 2, Model 1) and 0.6430 using the trigonometric and linear function of date (Table 2, Model 2). The relationship between traffic volume in the final adjusted model (Table 2, Model 1) and NO₂ level is shown for each of the four quadrants surrounding a residence (Fig. 5).

Data used for validating the model included 120 NO₂ samples taken outside of Connecticut residences in 1994, and ranged in value from 4.39 to 33.10 ppb (mean [SD] of 13.75 [5.35]). Using the trigonometric and linear function of date to adjust for seasonality, in addition to covariates from the final model including traffic, land use, population density, elevation, and adjusting the intercept to reflect the 3.75 ppb difference in means between the two sample sets, the correlation between observed and predicted NO₂ was 0.68. Results of the additional cross-validation analysis using the leave-one-residence-out strategy produced a RMSE of 2.40 compared to 2.38 for the final model (Model 2, Table 2) fitted using all observations. The similarity of RMSEs between the cross-validation sample and the final model indicates that the model performs well in estimating exposure given relevant data on traffic, land use and time.

A variogram was constructed (not shown) and revealed no evidence of spatial correlation for these data. There was, however, evidence of correlation among the repeated temporal measurements from each residence and this was incorporated into our final models (Models 1 and 2, Table 2).

The final model (with the trigonometric and linear function of date for seasonality) was used to predict NO₂ levels in a 616 square km area around Hartford, Connecticut at two times in a year (Fig. 10). Orange areas indicate the highest predicted concentrations of NO₂ and blue areas the lowest. Predicted values for February 1 ranged from 5.9 to 19.8 ppb (mean [SD] 10.4

[2.3]) (Fig. 10, left panel), and for August 1 ranged from 0.5 to 14.4 ppb (mean [SD] 5.0 [2.3]) (Fig. 10, right panel).

5. Discussion

The model described here demonstrates significant relationships between traffic volume, land use, population density, elevation, season and NO₂ concentration over a large geographic area. The final model incorporates methods used by Holford et al. (2010) and explains approximately two-thirds of observed variation in NO₂ measurements (Table 2). This is an improvement over the isotropic step function or anisotropic cubic spline models reported by Holford et al. (2010) which were unadjusted for other covariates and were able to explain half of the variability. This in turn was an improvement on use of a Gaussian dispersion model that weighted the effect of average daily traffic (ADT) as a function of distance from residence (Holford et al., 2010) which was able to explain less than 10% of the variability in NO₂ level. In the model we present here, inclusion of publicly available variables (namely, landscape characteristics including land use, population density and elevation) has resulted in a significant improvement in predicting NO₂ concentration at a residence.

5.1. Traffic volume: isotropic model (unadjusted) vs. anisotropic model (adjusted)

In the isotropic model unadjusted for any other covariates, the largest contribution to NO₂ comes from traffic volume within 500 m of a residence, with significant (though small) contributions to NO₂ from traffic volume as far as 6 km distant (Fig. 3). This is contrary to previous reports suggesting that NO₂ is significantly elevated only close to roadways (Carr et al., 2002; Jerrett et al., 2007). Many previous models of NO₂ have assumed a Gaussian dispersion function that precludes finding effects at greater distances, whereas the step function used here makes no such assumptions.

When an anisotropic model is considered, the impact of prevailing winds is easy to see. The wind rose summarizing prevailing winds during the three-year study (Fig. 4) suggests that the strongest effects should be observed for traffic volume in the northern, southern and western quadrants, with little effect accruing from traffic east of a residence, and this is exactly what we see in the model. Fig. 5 illustrates the traffic portion of the final adjusted model and shows a similarly-sized effect of traffic volume on NO₂ levels for the quadrants north, south and west of a residence and a much smaller contribution to NO₂ levels from traffic volume east of a residence (Fig. 5, Table 2). The strongest effect on NO₂ level at a residence is from traffic volume within 4 km north, 4 km south and 2 km west of a residence (Fig. 5 and Table 2).

5.2. Land use variables in traffic model

Our results suggest that developed land within 500 m of a residence is associated with increased levels of NO₂, while forested and grass land within 2 km is associated with decreased NO₂ (Table 2). Only interstates and numbered highways are included in state-collected traffic data, thus it is possible that a portion of the developed land category accounts for contributions to NO₂ concentration coming from roadways not included in the state count. Developed land may also contribute to NO₂ levels by accounting for the presence of other NO₂ sources (e.g., non-traffic combustion processes related to industry, power generation or domestic heating). One limitation of our land use data source (CLEAR, 2006) was that there was only one category of “developed” land available. Additional categories such as “commercial,” “residential,” “industrial” which may be available from other land use data sources would be interesting to explore in future analyses. There is a roughly inverse relationship between the land use categories of developed and forest/grass since together these categories account for the majority of land use in any given buffer (Fig. 6). Thus, the negative effect on NO₂ seen for forest/grass land may be an artifact of the absence of developed land. There is evidence, however, that

certain types of land use have a true negative effect on NO₂ levels, i.e., trees and plants may actually reduce levels of NO₂ by a process of absorption (Atkins and Lee, 1995; Gesler et al., 2002; Hanson et al., 1989; Hargreaves et al., 1992), or by creating a physical barrier to NO₂ dispersion from a roadway.

5.3. Other covariates: population density, elevation

Interestingly, population density had a small, but significant positive effect on estimated NO₂, while elevation had a small, but significant negative effect. In Connecticut, the highest population density (Fig. 7) and the highest percentage of developed land (Fig. 6) are along the shoreline (i.e., at sea level) and in the Connecticut River valley (Fig. 8). While these three covariates (population density, land use, elevation) are certainly related, they can also serve as proxies for unmeasured factors. For example, population density can represent factors associated with urban areas that may impact NO₂ levels (e.g., volume of diesel truck traffic or emissions from fuel combustion for home heating). Elevation may represent factors of topology that impact meteorology that in turn modify levels of NO₂.

5.4. Adjustment for season of NO₂ sampling

The final model was developed using NO₂ measurements from a recent 3.5 year period (2006 – 2009). The bspline function developed to account for seasonality was tied to this time period and thus did not seem appropriate to use to account for seasonality at the time the validation samples were collected over 12 years ago. Therefore, the more generalizable cosine, sine and linear function of date was used to represent seasonality. In addition, an adjustment was made to the intercept to account for the steady decrease in annual levels of NO₂. In the state of Connecticut, the annual NO₂ concentration has fallen from a mean (SD) of 24.6 (11.8) ppb in 1994 to 15.3 (8.5) ppb in 2006 (CT DEP, 2006; US EPA, 2010). The difference between the two sample sets was not as dramatic (13.75 (5.35) ppb in 1994 compared to 10.00 (3.99) ppb for the years of the study 2006–2009), but was significant (t-test, $p < 0.0001$). Therefore, the intercept in the validation model was adjusted by 3.75 ppb. The difference in means between the two studies suggests that model calibration would be wise before using it to predict NO₂ levels in different time periods.

5.5. Final model

Fig. 10 shows an NO₂ surface predicted by the final model for an area around Hartford, Connecticut. Areas with highest estimated values for NO₂ correspond to areas with the busiest roadways, most developed land, highest population density and lowest elevation. Predicted NO₂ levels are shown for a winter day (February 1, Fig. 10, left panel), and a summer day (August 1, Fig. 10, right panel). As would be expected from what was observed with the measured values, NO₂ levels are predicted to be higher in winter than summer. This is reflected in the February map by the greater intensity and larger size of orange areas adjacent to the interstates passing through the center of Hartford, compared to the August map with its lower intensity and smaller area shaded orange in Hartford center and increased blue intensity in areas away from the city center.

One of the challenges to creating and implementing our model has to do with the temporal congruence, or lack thereof, of the covariate data sources. This was not a major issue for model development: NO₂ measurements were from 2006 – 2009, traffic data from 2006, land use data from 2006, and population density data from 2005. Temporal mismatches were greater for the validation analysis where NO₂ measurements were from 1994, which may explain some of the residual variance in the model. Although changes in land use and elevation may be negligible over this time period, better fits might have been obtained by using traffic and population density data from closer to the time of NO₂ measurement.

The model we present is relatively simple, in that it considers only the main-effects of traffic and landscape characteristics. Although this enables straightforward interpretation of the covariates, it may fail to capture certain complexities that could be described using interaction terms. For instance, one might imagine that land use on the line between a road segment and the residence location might interact with the contribution of the segment, and this could be captured by a more complex model. As we have shown, there are clearly anisotropic effects of traffic due to meteorology and topography. We adjusted for the composite effect of prevailing wind over the period of the study. The temporal variability in the modifying effects of wind direction and wind speed could be captured and incorporated into the model in greater detail by calculating wind roses for shorter time periods and for more than four directional sectors.

5.6. Conclusion

The model presented here provides a method to estimate exposure to NO₂, often used as a proxy for traffic-related pollution, without expensive monitoring efforts. This model is based on readily accessible information and permits a cost-effective assessment of residential exposures using a dispersion function for NO₂ around roadways that also considers additional variables including land use, elevation, population density and season. The model, which could be calibrated and applied in areas other than Connecticut, has importance as a tool for exposure estimation in epidemiological studies of traffic-related air pollution.

Acknowledgments

This study was funded by the National Institutes of Health grants ES05410, ES011013, and ES017416.

References

- Atkins D, Lee D. Spatial and temporal variation of rural nitrogen dioxide concentrations across the United Kingdom. *Atmospheric Environment* 1995;29:223–239.
- Beckett WS, Gent JF, Naeher LP, Belanger K, Triche EW, Bracken MB, Leaderer BP. Peak expiratory flow rate variability is not affected by home combustion sources in a group of nonsmoking women. *Archives of Environmental and Occupational Health* 2006;61(4):176–182. [PubMed: 17867572]
- Bensen, P. Caline4. California Department of Transportation; 1989. A Dispersion Model for Predicting Air Pollution Concentrations Near Roadways.
- Beyer, HL. Hawth's Analysis Tools for ArcGIS. 2004 [accessed 10 April 2009]. <http://www.spatial ecology.com/htools>
- Brunekreef B, Janssen NA, de Hartog J, Harssema H, Knape M, van Vliet P. Air pollution from truck traffic and lung function in children living near motorways. *Epidemiology* 1997;8(3):298–303. [PubMed: 9115026]
- Byun DW, Schere KL. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Review* 2006;59:51–77.
- Carr D, Ehrenstein O, Weiland S, Wagner C, Wellie O, Nicolai T, von Mutius E. Modeling annual benzene, toluene, NO₂ and soot concentrations on the basis of road traffic characteristics. *Environmental Research* 2002;A90:111–118. [PubMed: 12483801]
- CLEAR. Center for Land Use Education and Research, University of Connecticut; Storrs, CT: 2006 [accessed 10 April 2009]. <http://clear.uconn.edu/projects/landscape/download.asp>
- CT DEP. Air Quality Trends - NO₂. Connecticut Department of Environmental Protection; Hartford, CT: 2006 [accessed 1 January]. http://www.ct.gov/Dep/cwp/view.asp?a=2684&q=322060&depNav_GID=1744
- CT DOT. Traffic monitoring volume information - Traffic count data. Connecticut Department of Transportation; Hartford, CT: 2006 [accessed June 2009]. <http://www.ct.gov/dot/cwp/view.asp?a=3532&q=330402>

- Duhme H, Weiland SK, Keil U, Kraemer B, Schmid M, Stender M, Chambless L. The association between self-reported symptoms of asthma and allergic rhinitis and self-reported traffic density on street of residence in adolescents. *Epidemiology* 1996;7(6):578–582. [PubMed: 8899382]
- English P, Neutra R, Scalf R, Sullivan M, Waller L, Zhu L. Examining associations between childhood asthma and traffic flow using a geographic information system. *Environmental Health Perspectives* 1999;107(9):761–767. [PubMed: 10464078]
- ESRI. StreetMapUSA. Environmental Systems Research Institute; Redlands, CA: 2003 [accessed 10 April 2009]. <http://www.esri.com/data/data-maps/data-and-maps-dvd.html>
- ESRI. US Census Tracts, 2005 Edition. ESRI; Redlands, CA: 2005 [accessed 10 April 2009]. <http://www.esri.com/data/data-maps/data-and-maps-dvd.html>
- ESRI. ArcGIS. Environmental Systems Research Institute; Redlands, CA: 2006 [accessed 10 April 2009]. http://www.esri.com/products/index.html#arcgis_panel
- Gent JF, Koutrakis P, Belanger K, Triche E, Holford TR, Bracken MB, Leaderer BP. Symptoms and medication use in children with asthma and traffic-related sources of fine particle pollution. *Environmental Health Perspectives* 2009;117(7):1168–74. [PubMed: 19654929]
- Gesch, D. The National Elevation Dataset. In: Maune, D., editor. *Digital Elevation Model Technologies and Applications: The DEM Users Manual*. American Society for Photogrammetry and Remote Sensing; Bethesda, MD: 2007. p. 99–118.
- Gesch D, Oimoen M, Greenlee S, Nelson C, Steuck M, Tyler D. The National Elevation Dataset: Photogrammetric Engineering and Remote Sensing 2002;68(1):5–11.
- Gesler A, Rienks M, Reenenberg H. Stomatal Uptake and Cuticular Adsorption Contribute to Dry Deposition of NH₃ and NO₂ to Needles of Adult Spruce (*Picea abies*) trees. *New Phytologist* 2002;156(2):179–194.
- Hanson P, Rott K, Taylor GE Jr, Gunderson C, Lindberg S, Ross-Todd B. NO₂ deposition to elements representative of a forest landscape. *Atmospheric Environment* 1989;23:1783–1794.
- Hargreaves KJ, Fowler D, Storeton-West RL, Duyzer JH. The exchange of nitric oxide, nitrogen dioxide and ozone between pasture and the atmosphere. *Environmental Pollution* 1992;75(1):53–9. [PubMed: 15092049]
- HEI. Special Report 17. Health Effects Institute; Boston, MA: 2010 [accessed 1 January 2010]. Traffic-related air pollution: A critical review of the literature on emissions, exposure, and health effects. <http://pubs.healtheffects.org/getfile.php?u=552>
- Holford TR, Ebisu K, McKay LA, Gent JF, Triche EW, Bracken MB, Leaderer BP. Integrated exposure modeling: A model using GIS and GLM. *Statistics in Medicine* 2010;29(1):116–129. [PubMed: 19823976]
- Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, Morrison J, Giovis C. A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology* 2005;15(2):185–204. [PubMed: 15292906]
- Jerrett M, Arain MA, Kanaroglou P, Beckerman B, Crouse D, Gilbert NL, Brook JR, Finkelstein N, Finkelstein MM. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *Journal of Toxicology and Environmental Health Part A* 2007;70(3–4):200–12. [PubMed: 17365582]
- Lakes Environmental Software. WRPLOT View - Wind Rose Plots for Meteorological Data. 2008 [accessed 1 January 2010]. Freeware <http://www.weblakes.com/products/wrplot/index.html>
- Leaderer, BP. Indoor and Outdoor NO₂ and Asthma in Children. National Institute of Environmental Health Sciences Grant (NIEHS) ES05410-18; 2007 [accessed 1 January 2010]. http://projectreporter.nih.gov/project_info_description.cfm?aid=7629628&icde=2576392
- Lin S, Munsie JP, Hwang SA, Fitzgerald E, Cayo MR. Childhood asthma hospitalization and residential exposure to state route traffic. *Environmental Research* 2002;88(2):73–81. [PubMed: 11908931]
- Livingstone AE, Shaddick G, Grundy C, Elliott P. Do people living near inner city main roads have more asthma needing treatment? Case control study. *British Medical Journal* 1996;312(7032):676–677. [PubMed: 8597735]
- NRCC. Northeast Regional Climate Center. Cornell University; Ithaca, NY: 2010 [accessed 1 January 2010]. <http://www.nrcc.cornell.edu/index.html>
- Palmes ED, Gunnison AF, DiMattio J, Tomczyk C. Personal sampler for nitrogen dioxide. *American Industrial Hygiene Association Journal* 1976;37(10):570–577. [PubMed: 983946]

- Pearson RL, Wachtel H, Ebi KL. Distance-weighted traffic density in proximity to a home is a risk factor for leukemia and other childhood cancers. *Journal of Air & Waste Management Association* 2000;50(2):175–80.
- Pettigrew M, Gent J, Belanger K, Triche E, Bracken MB, Leaderer BP. Infant otitis media and the use of secondary heating sources. *Epidemiology* 2004;15:13–20. [PubMed: 14712142]
- Pleijel H, Pihl Karlsson G, Binsell Gerdin E. On the logarithmic relationship between NO₂ concentration and the distance from a highroad. *The Science of the Total Environment* 2004;332(1–3):261–4. [PubMed: 15336908]
- SAS Institute Inc.. SAS OnlineDoc 9.1.3. Fourth Edition. SAS Institute Inc.; Cary, NC: 2004.
- Triche E, Belanger K, Beckett W, Bracken MB, Holford TR, Gent J, Jankun T, McSharry J.-e. Leaderer B. Infant respiratory symptoms associated with indoor heating sources. *American Journal of Respiratory and Critical Care Medicine* 2002;166:1105–1111. [PubMed: 12379555]
- Triche EW, Belanger K, Bracken MB, Beckett WS, Holford TR, Gent JF, McSharry J.-e. Leaderer BP. Indoor heating sources and respiratory symptoms in nonsmoking women. *Epidemiology* 2005;16(3):377–384. [PubMed: 15824555]
- Triche EW, Gent JF, Holford TR, Belanger K, Bracken MB, Beckett WS, Naeher L, McSharry JE, Leaderer BP. Low-level ozone exposure and respiratory symptoms in infants. *Environmental Health Perspectives* 2006;114(6):911–916. [PubMed: 16759994]
- US EPA. User's Guide to MOBILE6.1 and MOBILE6.2: Mobile Source Emission Factor Model. US Environmental Protection Agency; 2003 [accessed 1 January 2010].
http://www.weblakes.com/products/mobile/resources/docs/MOBILE6_model_user_guide.pdf
- US EPA. Technology Transfer Network (TTN), Air Quality System (AQS) - Nitrogen dioxide. US Environmental Protection Agency; 2010 [accessed 1 January 2010].
<http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdta.htm>
- WHO. Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide. World Health Organization; Bonn, Germany: 2003 [accessed 1 January 2010].
<http://www.euro.who.int/document/e79097.pdf>

Figure 1.

Traffic volume and land use components of model for estimating residential NO_2 levels calculated within buffer of radius “D” (dotted black line) surrounding residence. Within this buffer, traffic volume component uses state-provided average daily traffic (ADT) counts on curve segment “ C_i ” of roadways (thick black lines) subdivided into 50-m roadway sub-segments “ s_{ij} .” Distance from residence to roadway segment is “d.” Land use categories, shown as color-coded “pixels,” enter model as area in hectares of each category within a buffer. Also shown are axes (orange lines) dividing the buffer into four directional quadrants (i.e., north, east, south and west of residence).

Figure 2. Average daily traffic (ADT) counts on interstates and numbered highways in Connecticut in 2006 (CT DOT, 2006). Residences where NO₂ samples were taken are indicated (black dots).

Figure 3. Step function relating isotropic NO₂ dispersion to traffic volume in 11 concentric buffers around a residence. Model is unadjusted for any other covariates. (Connecticut 2006–2009)

Figure 4.
Wind rose representing prevailing wind direction and speed in Connecticut, 2006–2009 (Lakes Environmental Software, 2008; NRCC, 2010).

Figure 5. Traffic volume portion of model relating NO₂ dispersion to residential covariates (Model 1, Table 2). Results are shown for each of four directional quadrants. (Connecticut 2006–2009)

Figure 6.
Land use in Connecticut, 2006, derived from satellite imagery (CLEAR, 2006).

Figure 7.
Population density (per square mile), Connecticut census tracts (2005) (ESRI, 2005).

Figure 8.
Elevation (in meters) in Connecticut (Gesch, 2007; Gesch et al., 2002).

Figure 9. Seasonality of NO₂ levels (measured NO₂, black dots) represented as a function of time of year by a cubic bspline (with 6 equally spaced knots) function of measurement date (red line) or a trigonometric (cosine and sine) and linear function of date (green line).

Figure 10.

Predicted NO₂ levels (indicated by color ramp from blue (low) to bright orange (high)) for the Hartford, Connecticut area for a winter day (February 1, 2009, left panel) and a summer day (August 1, right panel). Interstates 84 (east/west) and 91 (north/south) (parallel black lines) and other numbered highways (solid black lines) are shown. Town borders are indicated (dotted black lines). The Connecticut River widens as it flows south of Hartford and is indicated in white.

Table 1

Summary statistics for model covariates.

Variable	Observations N	Mean (SD)	Range
NO ₂ (ppb) ^a	3140	10.0 (4.0)	0.1 – 27.5
Traffic (in buffers [km]) (10,000 vehicle-km) ^b	985		
0 – 0.5		1.8 (3.4)	0.0 – 24.2
0.5 – 1.0		6.4 (8.0)	0.0 – 41.6
1.0 – 2.0		26.4 (20.8)	0.0 – 134.1
2.0 – 3.0		39.8 (25.9)	1.0 – 167.1
3.0 – 4.0		50.0 (29.3)	3.9 – 187.0
4.0 – 5.0		56.8 (28.5)	3.8 – 185.2
5.0 – 6.0		63.1 (27.5)	7.2 – 199.5
Land use (in buffers [km]) (hectares) ^c	985		
Developed			
0 – 0.5		43.8 (17.2)	3.6 – 76.2
0.5 – 1.0		125.9 (51.3)	13.6 – 245.4
1.0 – 2.0		470.4 (176.5)	63.1 – 852.7
Forest/grass			
0 – 0.5		21.7 (15.2)	0.0 – 68.8
0.5 – 1.0		72.6 (45.6)	2.0 – 222.3
1.0 – 2.0		314.2 (167.2)	19.4 – 812.7
Other			
0 – 0.5		13.0 (10.5)	1.9 – 60.7
0.5 – 1.0		37.1 (36.6)	0.0 – 208.0
1.0 – 2.0		157.8 (121.1)	0.0 – 784.7
Population density (/mi ²) ^d	985	5268.7 (5238.5)	206.3 – 31742.9
Elevation (m) ^e	985	57.2 (52.2)	0.0 – 318.0

^aNO₂ sampled outside of 985 homes in Connecticut between April 2006 – June 2009.^bAverage daily volume (in 10 000 vehicle-kilometer traveled) on numbered highways and roads in Connecticut in 2006 (CT DOT, 2006). For example, mean (SD) traffic in the closest buffer to the residence (0 – 0.5 km) is 1.8 (3.4) × 10⁴ veh-hm (i.e., 18 000(34 000) veh-km), and ranges from “0”, i.e., no numbered highways within the buffer and thus no traffic density values available, to 24.2 × 10 veh-km.^cLand use (in hectares) within 0.5 km of residences in Connecticut in 2006 (CLEAR, 2006). Of 12 original categories, “Developed” includes only the developed land category; “Forest/grass” includes deciduous forest, coniferous forest, turf and grass and other grasses); and “Other” includes agricultural field, water, non-forested wetland, forested wetland, tidal wetland, barren, utility rights-of-way.^dPopulation density in census tracts from 2005 US census update (ESRI, 2005).^eElevation from 2005 (Gesch, 2007; Gesch et al., 2002).

Table 2Final model^a of NO₂ concentration adjusted for variables shown. (Connecticut 2006 – 2009)

Parameter	Model 1		Model 2	
	Estimate	95% CI	Estimate	95% CI
Intercept	4.86	3.32, 6.39	8.74	7.94, 9.54
Traffic vol (quadrant/buffer (km))				
North				
0 – 0.5	0.18	0.050, 0.30	0.18	0.047, 0.31
0.5 – 1.0	0.059	0.013, 0.11	0.055	0.0079, 0.10
1.0 – 4.0	0.010	0.0052, 0.016	0.011	0.0059, 0.016
4.0 – 5.0	–0.0050	–0.014, 0.0038	–0.0033	–0.013, 0.0059
5.0 – 6.0	0.010	0.0024, 0.017	0.001	0.002, 0.018
East 0 – 6.0	0.0037	0.00010, 0.0074	0.004	0.0003, 0.0076
South				
0 – 0.5	0.22	0.13, 0.32	0.23	0.14, 0.32
0.5 – 2.0	0.033	0.020, 0.046	0.033	0.020, 0.046
2.0 – 4.0	0.012	0.006, 0.018	0.013	0.007, 0.019
West				
0 – 0.5	0.19	0.045, 0.34	0.20	0.049, 0.36
0.5 – 2.0	0.034	0.019, 0.048	0.033	0.019, 0.048
Land use (category/buffer (km))				
Developed 0 – 0.5	0.033	0.022, 0.043	0.030	0.020, 0.041
Undeveloped				
0 – 0.5	–0.023	–0.035, –0.011	–0.022	–0.034, –0.0093
0.5 – 2.0	–0.0031	–0.0041, –0.0021	–0.0029	–0.0039, –0.0019
Population density	0.00008	0.00003, 0.00012	0.00008	0.00004, 0.00012
Elevation	–0.0028	–0.0052, –0.0005	–0.0035	–0.0059, –0.0011
R ²	0.6748		0.6452	
R ² , adjusted	0.6728		0.6430	
Correlation (exchangeable)	0.1984		0.1630	

^aModel 1: NO₂ (in ppb) = Traffic volume variables + Land use variables + Population density + Elevation + Seasonality, where seasonality is represented by bspline function of date of NO₂ measurement. Model 2: Identical to Model 1 with the exception of Seasonality which is represented by a trigonometric (cosine, sine) and linear function of date.