Application of Land Use Regression to Estimate Long-Term Concentrations of Traffic-Related Nitrogen Oxides and Fine Particulate Matter

ARTICLE *in* ENVIRONMENTAL SCIENCE AND TECHNOLOGY · MAY 2007

Impact Factor: 5.33 · DOI: 10.1021/es0606780 · Source: PubMed

4 AUTHORS, INCLUDING:

Sarah B Henderson
BC Centre for Disease Control

Bernardo Beckerman
University of California, Berkeley

Michael Brauer
University of British Columbia - Vancouver

Application of Land Use Regression to Estimate Long-Term Concentrations of Traffic-Related Nitrogen Oxides and Fine Particulate Matter

SARAH B. HENDERSON,[†]
BERNARDO BECKERMAN,[‡,§,||]
MICHAEL JERRETT,[‡,§,||] AND
MICHAEL BRAUER*,[†]

School of Occupational and Environmental Hygiene, The University of British Columbia, Vancouver, British Columbia, Canada, Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, California School of Geography and Earth Sciences, McMaster University, Hamilton, Ontario, Canada, and Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California

Land use regression (LUR) is a promising technique for predicting ambient air pollutant concentrations at high spatial resolution. We expand on previous work by modeling oxides of nitrogen and fine particulate matter in Vancouver, Canada, using two measures of traffic. Systematic review of historical data identified optimal sampling periods for NO and NO_2. Integrated 14-day mean concentrations were measured with passive samplers at 116 sites in the spring and fall of 2003. Study estimates for annual mean NO and NO_2 ranged from 5.4−98.7 and 4.8−28.0 ppb, respectively. Regulatory measurements ranged from 4.8−29.7 and 9.0−24.1 ppb and exhibited less spatial variability. Measurements of particle mass concentration ($PM_{2.5}$) and light absorbance (ABS) were made at a subset of 25 sites during another campaign. Fifty-five variables describing each sampling site were generated in a Geographic Information System (GIS) and linear regression models for NO, NO_2, $PM_{2.5}$, and ABS were built with the most predictive covariates. Adjusted R^2 values ranged from 0.39 to 0.62 and were similar across traffic metrics. Resulting maps show the distribution of NO to be more heterogeneous than that of NO_2, supporting the usefulness of this approach for assessing spatial patterns of traffic-related pollution.

Introduction

Several epidemiological studies have assessed the association between health impacts and simple measures of proximity to traffic (*1−7*) based on observations of steep pollutant concentration gradients as a function of distance to roads (*8−11*). To better capture this intraurban variability in pollutant concentrations more sophisticated exposure assessment methods are needed (*12*, *13*). Dispersion models that simulate pollution fate and transport can be useful, but are often infeasible at high spatial resolution throughout large areas. Interpolation of concentrations measured by regulatory air quality monitors can identify regional patterns, but networks are typically too sparse to reflect localized variations in concentration. Land use regression (LUR) is a promising alternative to these conventional approaches. By establishing a statistical relationship between land use characteristics and pollutant measurements, LUR can be used to predict concentrations at any spatiotemporal resolution within the framework of a Geographic Information System (GIS). Here we use LUR to model spatial variability in long-term average concentrations of nitrogen oxides (NO and NO_2), fine particulate matter ($PM_{2.5}$), and its light absorbance (ABS).

No standard method for conducting LUR exists, but descriptions of the general approach can be found elsewhere (*12*, *14−19*). In brief, a pollutant is sampled at multiple sites, and variables describing each site in terms of location, surrounding land use, population density, and traffic patterns are generated in GIS. Linear regression is used to model concentrations of the pollutant as a function of the most predictive variables, and GIS is used to render the models as maps that can estimate exposure across the domain. We expand on previous LUR research by outlining a sequential framework for model development and evaluation, assessing model performance for different types of pollutants, and comparing sets of models built using different traffic indices.

Most LUR analyses have used ad hoc procedures to determine sampling periods and sampler locations; two factors that may impact study results. We address these limitations by (a) presenting a novel approach for sampling period selection based on historical data and (b) applying a location-allocation approach to identify sampler locations (*20*).

In addition, most previous LUR studies have modeled nitrogen dioxide (NO_2) because it can be inexpensively measured with passive samplers. However, NO_2 is a secondary pollutant formed mainly from nitric oxide (NO), the predominant species in vehicle exhaust. Therefore we expect concentrations of NO_2 to be more spatially homogeneous than those of NO, which decrease rapidly with increasing distance from roadways. To assess whether LUR is sensitive to this distinction we built models for NO and NO_2 and generated prediction surfaces for both species. Vehicles are also an important source of fine particulate matter, and LUR models of particle constituents may help to differentiate between the impacts of vehicle types. For example, diesel combustion results in higher concentrations of light-absorbing (elemental) carbon relative to gasoline combustion (*21*). Heavy-duty vehicles are an important source of diesel emissions, and we measured the absorbance coefficient (ABS) of particulate matter at a subset of sites to assess specific associations with measures of truck traffic.

Finally, previous applications of LUR have used readily available geographic data without rigorous assessment of its utility. Differences in traffic monitoring programs between urban areas result in different types of traffic indicators. Where cars and trucks are systematically enumerated, variables reflecting traffic intensity can be generated. Where traffic counts do not exist, road classifications can be used as a surrogate. To assess whether equally predictive surfaces can be obtained from these different data types we developed all of our models using two sets of traffic variables. The first relies on traffic volume estimates from the widely used EMME/2 (INRO Consultants, Montreal, Canada) transportation model, which simulates traffic flow based on vehicle count data, sociodemographic patterns, and travel-demand surveys (*22*). The second uses the presence of highways and major roads as a proxy for traffic intensity.

* Corresponding author phone: 604-822-9585; fax: 604-822-9588; e-mail: brauer@interchange.ubc.ca.
[†] The University of British Columbia.
[‡] Currently located at University of California.
[§] McMaster University.
[||] University of Southern California.

Materials and Methods

Sampling Location Determination. Our estimation domain covered a 2200 km^2 area of the Greater Vancouver Regional District (GVRD) in British Columbia, Canada. Five years of regulatory NO_2 data from ten air quality monitoring sites were used as input for a location-allocation model (*20*). Briefly, this involves a two-step algorithm that (1) builds a demand surface and (2) solves a constrained spatial optimization problem to determine locations for a pre-specified number of samplers. The demand surface uses regulatory air quality data, land use coverage, and population density to estimate how concentrations of a pollutant are distributed in the study domain. Sites are then selected so that they capture the complete range of values while being placed at maximum distance from one another. We used this technique to identify 100 sites that would optimize the variability in measured NO_2 concentrations, and 16 more sites were added to ensure that areas of particular interest were well-characterized. A random subset of 25 sites from the location-allocation group was selected for particle sampling, though four substitutions were made to accommodate the equipment. All sites are mapped in the Supporting Information.

Sampling and Analysis for Oxides of Nitrogen. Because nitrogen oxide concentrations in the GVRD follow a seasonal cycle (*23*) we analyzed 5 years of NO_2 data from 15 regulatory monitoring stations to identify optimal sampling periods. Starting on January 1 of each year we calculated running two-week averages for the entire year, took the means of diametric values (i.e., those separated by 26 weeks), and compared results to the annual mean at each station. In 70 out of 75 cases the combined means for Feb 19 to Mar 4 and Aug 20 to Sep 2 were within 15% of the annual value. Two other periods produced slightly better results but captured the annual mean as a product of extremes in NO_2 concentration whereas the Feb/Mar and Aug/Sep periods could each provide an independent estimate. Accordingly, our field campaigns ran from Feb 24 through Mar 14 and Sep 8 through Sep 26, 2003.

During both campaigns we measured the integrated 14-day mean concentrations of NO and NO_2 with Ogawa passive samplers at 116 locations. Spring and fall results were averaged to estimate the annual mean for each site. All samplers were placed on lamp posts, utility poles, or street signs at a height of 2.5 meters. In addition we collocated a total of 27 samplers with regulatory chemiluminescence monitors, and collected 38 duplicates and 26 field blanks. All samples were extracted into 6 mL of deionized water within two weeks of collection and refrigerated until analysis by ion chromatography. Aqueous concentrations were converted to airborne concentrations according to the Ogawa protocol (*24*).

Particulate Matter Sampling and Analysis. Fine particles ($PM_{2.5}$) were collected on PTFE filters (Teflo, Pall Corp. East Hills, NY) using Harvard Impactors (Air Diagnostics and Engineering, Harrison, ME). The field campaign ran from March 5 through May 8, 2003. Programmable, battery-powered pumps (SKC Inc., model 224-PCXR8, Eighty Four, PA) with solar chargers were used to collect a 24-hour sample over 7 days at a flow rate of 4 L/min. Five sampling units were rotated between the 25 sites and one was collocated with a Tapered Element Oscillating Microbalance (TEOM, Thermo Electron Corp, East Greenbush, NY) at a central GVRD station (T18). Eight blanks were deployed. Mass concentrations were calculated as described elsewhere (*25*). Filter reflectance was measured with a Smokestain Reflectometer (Diffusion Systems Ltd., model 43, Harwell, UK) and converted to absorption coefficients (ABS) according to standard methods (*26*, *27*). Strong correlations (R^2 0.7−0.8) between ABS and traditional measurements of elemental carbon have previously been demonstrated (*28*−*30*). Because measurements were not concurrent, $PM_{2.5}$ and ABS values were adjusted to reflect temporal trends observed at station T18 using a ratio method (*1*). The ratio of the TEOM-measured mean for each week of sampling to the mean for the entire sampling campaign was used to obtain weekly adjustment factors.

Spatial Variable Generation. We generated 50 variables in 4 categories and 10 subcategories to characterize the street network, traffic intensity, land use, and population density at different radii around each sampling site (Table 1). All variables in each category were derived from a single spatial dataset in vector format. Input files for the *Road Length*, *Land Use*, and *Population Density* categories were taken from the 2001 census package prepared by DMTI Spatial (Markham, Ontario) and distributed by Statistics Canada. The input file for variables in the *Vehicle Density* category was generated by the GVRD transit authority's EMME/2 model of morning rush-hour traffic volume.

Table 1 describes how input files were manipulated in ArcView 3.2 (ESRI, Redlands, CA) and processed with the Spatial Analyst extension to produce variable layers in grid format. Values of the grid cells underlying our sampling locations were recorded for each variable layer to be used in the regression analyses. Five additional variables describing the geographic location of each site in terms of its elevation (ELEV), longitude (X), latitude (Y), distance to the nearest highway (DIST), and distance from the seashore (SHOR) were included, for a total of 55 potential predictors.

Model Building. Variables in the *Road Length* and *Vehicle Density* categories (Table 1) were treated as mutually exclusive traffic metrics and independently combined with the remaining 31 variables to build 2 models for each of NO, NO_2, $PM_{2.5}$, and ABS in S-Plus (Insightful Corporation, Seattle, WA). To ensure interpretability of model parameters and to reduce the potential for collinearity between traffic-related covariates we made the *a priori* assumption that all *Road Length* and *Vehicle Density* variables should have positive regression coefficients. The following model-building algorithm was used: (1) Rank all variables by the absolute strength of their correlation with the measured pollutant. (2) Identify the highest-ranking variable in each sub-category. (3) Eliminate other variables in each sub-category that are correlated (Pearson's $r \geq 0.6$) with the most highly ranked variable. (4) Enter all remaining variables into a stepwise linear regression. (5) Remove from the available pool any variables that have (a) insignificant *t*-statistics ($\alpha = 0.05$) and/or (b) coefficients that are inconsistent with *a priori* assumptions. 6) Repeat steps 4 and 5 to convergence and remove any variable that contributes less than 1% to the R^2 value for a parsimonious final model.

Regression Mapping. The regression equation $y = \beta_0 + \beta_1 x_1 + ... + \beta_i x_i$ is rendered as a map by multiplying all cells in the contributing variable layers ($x_{1,...,i}$) against their associated coefficients ($\beta_{1,...,i}$) and summing the resulting grids with the constant intercept β_0. We used algebraic features of Spatial Analyst to transform our 10 regression equations into maps that estimate the annual concentrations of NO, NO_2, $PM_{2.5}$, and ABS across the study domain. Where the regression equations produced negative estimates, grid cell values were set to zero. Where estimates exceeded the maximum measured concentrations by more than 20%, grid cell values were truncated.

Model Evaluation. To quantitatively evaluate models we compared estimates for NO and NO_2 to the 2003 annual averages measured by chemiluminescence at 16 air quality monitoring sites in the GVRD. Estimates for $PM_{2.5}$ were compared to the 2004 annual averages measured by TEOM at 8 sites. To test how well the models predicted extremes in traffic-related pollution we used temporally adjusted NO,

TABLE 1. Description of Spatial Variable Categories, Their Inputs, and Summary of the GIS Methods Used to Generate Variables in Each

category (*N* variables)	description	variable sub-categories	buffer radii (m)	input file source & type	methods
road length (12)	total length (in km) of two road types	RD1 (highways), RD2 (major roads)	100, 200, 300, 500, 750 & 1000	Federal street network file (DMTI Spatial), line format	(1) convert RD1 and RD2 line files into raster files with 5 m pixels (2) use *Neighborhood Statistics*[a] to sum the number of road pixels in each search radius; divide by 200 to estimate the length in km.
vehicle density (12)	density (in vehicles/ hectare) of two vehicle types during morning rush hour	AD (automobiles), TD (trucks)	100, 200, 300, 500, 750 & 1000	output from the EMME/2 traffic flow model (TransLink), line format	(1) convert line file output from traffic model to points spaced at 1 m; assign each point the automobiles/m and trucks/m values of the line segment from which it was derived (2) use *Calculate Density*[a] (simple) to estimate the density values within each search radius
land use (20)	total area (in hectares) of 5 land use types	RES (residential), COM (commercial), GOV (governmental), IND (industrial), OPN (open)	300, 400, 500 & 750	Federal land use classification file (DMTI Spatial), polygon format	(1) convert land use polygons into 5 raster files for RES, COM, GOV, IND, and OPN with 5 m pixels (2) use *Neighborhood Statistics*[a] to sum the area of each land use type within the search radii
population density (6)	density (in persons/ hectare) of the population	POP (persons)	750, 1000, 1250, 1500, 2000 & 2500	dissemination area data from the 2001 census (DMTI Spatial), polygon format	(1) convert census polygons to centroids; assign each centroid the population count of the polygon from which it was derived (2) use *Calculate Density*[a] (kernel) to estimate the values within each search radius

[a] Features of the Spatial Analyst extension to ESRI's ArcView 3.2 GIS.

NO_2, $PM_{2.5}$, and ABS data collected with a mobile air monitoring unit (MAMU) at 8 sites as part of the 2002 pilot for this study (*31*). The predictive error for all models was estimated with leave-one-out (LOO) cross-validation where each model is repeatedly parametrized on $N - 1$ data points and then used to predict the excluded measurement. The mean difference between predicted and measured values estimates the model error. To assess the equivalency of *Road Length* and *Vehicle Density* models we compared their estimates at 58,000 points for each pollutant. These points represent the spatial distribution of the GVRD population (DMTI Spatial, Markham, Ontario). To qualitatively assess the predictive capacity of the models we conducted sensitivity analyses to examine the impact of (1) using alternate methods to generate the *Road Length* and *Vehicle Density* variables; (2) combining variables from both traffic metrics; and (3) subsetting the data input data sets.

Results and Discussion

NO_x Measurements. Valid samples were collected at 107 and 112 of the 116 sites during spring and fall sampling campaigns, respectively. Mean concentrations in the spring were 38.6 (SD = 24.2) and 19.6 (SD = 4.9) parts per billion (ppb) for NO and NO_2, respectively. These were significantly higher than the fall concentrations of 24.2 (SD = 15.9) and 12.9 (SD = 4.1) ppb. However, annual means measured by 16 monitors in the GVRD regulatory network show a strong 1:1 relationship with the campaign-specific means. Slopes are 1.03 (R^2 = 0.96) and 0.89 (R^2 = 0.98) for NO and NO_2, respectively. Results from the two campaigns were averaged for the 105 sites with complete data. Spring and fall measurements alone were used for two and seven sites, respectively. At two locations no measurements were successful. Missing samples were the result of sampler vandalism (N = 12) and a laboratory processing error (N = 1).

Annual estimates for NO were log-normally distributed with an arithmetic mean of 31.3 (SD = 21.6) ppb while NO_2 followed a normal distribution with a mean of 16.2 (SD = 5.6) ppb. The spatial autocorrelation between annual estimates was weak, with Moran's I (*32*) values (analogous to Pearson's *r*) of 0.05 and 0.15 for NO and NO_2, respectively. Considerably less variability was recorded by 16 regulatory monitors during the same time periods with the mean NO and NO_2 concentrations measuring 14.6 (SD = 8.3) and 15.8 (SD = 4.3) ppb, respectively. Comparison of 14 springtime measurements made by collocated chemiluminescence and Ogawa samplers produced slopes of 0.88 (R^2 = 0.93) and 1.09 (R^2 = 0.85) for NO and NO_2, respectively. The range of concentrations measured during the fall campaign was approximately half of that measured during the spring campaign for both species. As such, the relationships between collocated measurements were somewhat weaker with slopes of 0.87 (R^2 = 0.59) and 1.20 (R^2 = 0.73) for NO and NO_2, respectively. The overall slopes for the 38 duplicates were 1.07 (R^2 = 0.94) for NO and 0.93 (R^2 = 0.87) for NO_2. These results suggest no systematic under- or over-measurement by the Ogawa samplers.

$PM_{2.5}$ Measurements. Six concurrent mass concentration measurements made by the Harvard Impactor and TEOM collocated at GVRD station T18 had a slope of 0.24 (R^2 = 0.19), suggesting considerable discordance between the methods. Temporal adjustment factors ranged from 0.62 to 1.45. Adjusted $PM_{2.5}$ concentrations at the 25 sites ranged from 0.92 to 8.91 micrograms per cubic meter of air ($\mu g/m^3$) and were normally distributed with a mean of 4.08 (SD = 1.98) $\mu g/m^3$. Adjusted ABS coefficients ranged from 0.24 to $2.36 \times 10^{-5}\ m^{-1}$ and were normally distributed with a mean of 0.84 (SD = 0.47). All measurements were spatially independent as measured by Moran's I.

TABLE 2. Linear Regression Models Built to Predict Spatial Variability in Ambient Concentrations of Traffic-Related NO, NO_2, $PM_{2.5}$, and Light Absorption (ABS) by Fine Particulate Matter

response (*M*)	traffic metric	equation ([a]variable key given below)	A R^2	B R^2 for RL vs VD[b] (slope)	C R^2 for GVRD[c] validation	D R^2 for MAMU[d] validation	E LOO[e] error estimate (SD)
logNO (114)	road length	74.4 + **RD1.100** × 1.65 + **RD1.1000** × 0.037 + **RD2.100** × 2.19 + **POP.2500** × 0.007 − **ELEV** × 0.003 − **X** × 0.089 − **Y** × 0.123	0.62	0.50 (0.70)	0.49	0.36	+1.65 (12.7)
	vehicle density	116.0 + **AD.100** × 0.001 + **TD.1000** × 0.132 − **ELEV** × 0.002 − **X** × 0.129 − **Y** × 0.196	0.57		0.65	0.45	+1.90 (13.1)
NO_2 (114)	road length	42.6 + **RD1.100** × 10.5 + **RD1.1000** × 0.275 + **RD2.200** × 4.24 + **POP.2500** × 0.074 + **COM.750** × 0.116 − **ELEV** × 0.020− **X** × 0.591	0.56	0.76 (0.83)	0.69	0.44	0.00 (2.75)
	vehicle density	41.1 + **AD.100** × 0.002 + **TD.200** × 0.161 + **TD.1000** × 0.603 + **COM.750** × 0.116 + **POP.2500** × 0.068 − **ELEV** × 0.017	0.60		0.79	0.31	0.00 (2.61)
$PM_{2.5}$ (25)	road length	0.036 − **ELEV** × 0.019 + **COM.300** × 2.58 + **RES.750** × 0.035 + **IND.300** × 0.319	0.52	0.79 (0.61)	0.09	0.14	0.00 (1.50)
	vehicle density	1.01 + **AD.100** × 0.002 − **ELEV** × 0.018 + **COM.300** × 2.88 + RES.750 × **0.025**	0.52		0.07	0.00	0.00 (1.53)
ABS (25)	road length	0.837 + **RD1.1000** × 0.049 + **RD2.100** × 1.45 − **DIST** × 0.056 − **OPN.500** × 0.020	0.39	0.21 (0.48)		0.10	0.00 (0.37)
	vehicle density	0.509 + **TD.1000** × 0.144	0.41			0.23	0.00 (0.36)

[a] Left-hand side of period denotes variable type and right-hand side denotes buffer size so that **RD1.100** = length of highways within 100 m buffer. Variable types are: **RD1** = length of highways; **RD2** = length of major roads; **DIST** = distance to highway; **AD** = automobile density; **TD** = truck density; **POP** = population density; **ELEV** = elevation; **X** = longitude; **Y** = latitude; **COM** = commercial area; **RES** = residential area; **IND** = industrial area; and **OPN** = open area. [b] Comparison of *Road Length* (RL) and *Vehicle Density* (VD) estimates taken at 58,000 points representing the population distribution within the domain. [c] NO_x surfaces compared to chemiluminescence measurements at 16 sites in 2003; $PM_{2.5}$ surface compared to TEOM measurements at 8 stations in 2004. [d] Comparison to adjusted chemiluminescence and filtered particulate measurements made at 8 sites by the Mobile Air Monitoring Unit (MAMU) in 2002. [e] Leave-one-out cross-validation.

Models. Model equations are reported in Table 2. There was no evidence of spatial autocorrelation between model residuals, suggesting that simple linear regression was appropriate for these data. Both NO models are driven by traffic impact variables with 100 and 1000 m buffers while both NO_2 models include traffic variables with a 200 m buffer. The influence of these variables may be attributable to the secondary nature of NO_2, a hypothesis that is consistent with the inclusion of COM.750 (commercially zoned area in a 750 m buffer) in both NO_2 models. The $PM_{2.5}$ models are both driven by COM.300. No *Road Length* variable was identified as being significantly predictive for $PM_{2.5}$, and the effect of AD.100 (automobile density in a 100 m buffer) in the *Vehicle Density* model is weak. Conversely, the *Vehicle Density* model for ABS relies entirely on truck density, which suggests that LUR is useful for discriminating between pollutants associated with different sources. We attempted to capture the influence of diesel traffic in the *Road Length* models by including designated truck routes, but overlap with the RD1 and RD2 variables produced instability in the parameters.

Although distance to the nearest major road has been used in many epidemiological studies as a proxy for residential exposure to traffic-related pollution, this variable (DIST) is only included in the *Road Length* model of ABS. Correlations (Pearson's *r*) between DIST and measured NO, NO_2, $PM_{2.5}$, and ABS were −0.11, −0.16, 0.20, and −0.32, respectively. Other studies using saturation sampling have produced similar correlations between simple distance measures and pollutant concentrations (*33*). Where the resources to conduct LUR do not exist, a variable like RD2.100 (r = 0.35, 0.53, 0.16, and 0.34 for NO, NO_2, $PM_{2.5}$, and ABS, respectively) might be more informative than traditional measures of distance to traffic.

Model Evaluation. The R^2 values in column A of Table 2 are consistent across models built with different traffic variables, and both metrics produce similar residual plots (not shown). To further test their comparability we generated scatter and quantile−quantile (QQ) plots (not shown) of *Road Length* versus *Vehicle Density* estimates at 58,000 points for all four pollutants. Column B of Table 2 reports the scatter plot R^2 and slope values, with *Road Length* models consistently producing higher estimates than those built with *Vehicle Density* variables. The QQ plots showed strong 1:1 linearity between quantiles for NO, NO_2, and $PM_{2.5}$, meaning that estimates from both metrics follow the same distribution. This was not true for the ABS estimates. We attempted to explain more variability for each pollutant by combining both traffic metrics in a single model, but only marginal increases in the R^2 values were obtained. All these results suggest that models built with *Road Length* and *Vehicle Density* metrics are equally able to explain small-scale variability in pollutant concentrations. This finding confirms that valuable LUR models can be developed in the absence of traffic count data, which are unreliable or nonexistent in many areas. For example, Briggs et al. used road type data for their Amsterdam model, but could not speculate on whether better results were achievable with traffic density information (*14*).

The predictive value of both traffic metrics might be improved by resolution of inaccuracies inherent to the variable generation processes. First, the *Vehicle Density* variables were derived from output of a simulation model, the errors in which have not been well described. When we compared rush hour predictions from EMME/2 to rush hour and 24-hour traffic count data at 20 sites the R^2 values were 0.72 and 0.43, respectively (*31*). Second, the vector-to-raster operations required for *Road Length* and *Vehicle Density* variable generation (described in Table 1) produce errors that are inversely proportional to the buffer radii in magnitude. To address this problem we used vector operations to generate "gold standard" values for both variable types (computationally possible for 114 points, but not for the entire domain), compared them to raster-derived values, and assessed how the inaccuracies affected our final models. The R^2 values for NO models were increased by 0.02 and 0.06, respectively, for *Road Length* and *Vehicle Density* variables.

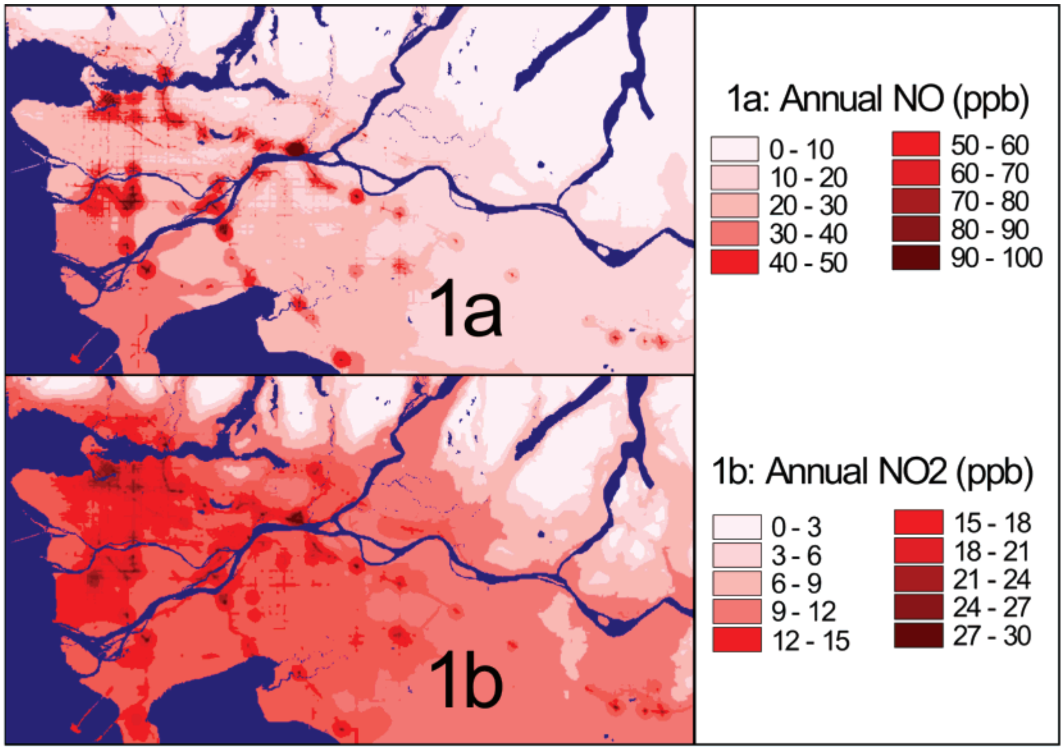

FIGURE 1. Maps estimating annual mean concentrations of NO (a) and NO_2 (b) over the study domain rendered from models built with *Vehicle Density* measures of traffic impact. Note how NO_2 concentrations are more homogeneous that those of NO.

Marginal increases for the NO_2 models were achieved, and no differences were seen for $PM_{2.5}$ or ABS.

In general, validation exercises show that *Vehicle Density* models perform better than *Road Length* models when used to predict pollutant concentrations. Column C of Table 2 shows R^2 values for measurements made at GVRD regulatory monitoring stations compared to model estimates. We have already shown how the regulatory sites fail to reflect intraurban variability, so these data cannot be used to evaluate model performance at high concentration areas. Column D shows a similar comparison for measurements made in 2002 at five high-traffic and three low-traffic sites. All slopes were less than one in both cases, with models consistently under-predicting measured values. Several other studies have used regulatory monitors or a reserved subset of the data to evaluate model performance with R^2 values ranging from 0.43 (N = 30; *19*) to 0.87 (N = 10; *14*).

Mean error estimates calculated with leave-one-out cross-validation are reported for all models in column E of Table 2. Overestimation by the logNO models results from log transformation. Mean errors for the unexponentiated values were zero with standard deviations of ~10% of the sample mean (not shown in table). Mean errors for the NO_2, $PM_{2.5}$, and ABS models are also zero, but their standard deviations are ~15%, ~40%, and ~45% of sample means, respectively. Higher variability in the $PM_{2.5}$ and ABS estimates may reflect the inability of smaller samples to produce robust models.

The question of how many sites should be sampled for development of LUR models has never been formally addressed, so this result warrants further discussion. Removing an influential outlier from our fine particulate data resulted in models with R^2 values of 0.68 and 0.79 for $PM_{2.5}$ and ABS, respectively. Despite the increase in R^2 values, these N = 24 models performed more poorly than their N = 25 counterparts in all validation exercises. To investigate further, *Road Length* and *Vehicle Density* models of NO and NO_2 using the N = 25 subset resulted in little improvement in R^2 values, with the standard deviations of the LOO error estimates remaining <15% of the sample means in both cases. We conclude that greater than 25 sites are required to produce high-quality LUR models for particulate matter and its absorbance, but fewer sites may be adequate for other pollutants.

Existing LUR models for NO_2 report R^2 values ranging from 0.54 (*15*) to 0.89 (*19*) with sample sizes ranging from 18 to 101 (Table S1 in the Supporting Information). Although our models are comparable (by R^2) to those fitted elsewhere, only the results for a study in Montreal reported a lower R^2 value (*15*). Like Montreal, Vancouver is surrounded by a complex network of waterways which may not be adequately captured by the variable set. For example, our distance to shoreline variable (SHOR) was weakly correlated with all pollutants (r = −0.02, −0.20, −0.03, and −0.20 for NO, NO_2, $PM_{2.5}$, and ABS, respectively). In addition, it remains to be seen whether sites chosen by time- and resource-intensive location-allocation algorithms produce more robust models than those chosen by less rigorous methods.

Regression Maps. Figure 1a and b show maps predicting annual concentrations of NO and NO_2, respectively, in the GVRD based on *Vehicle Density* LUR model variables. While the street network remains discernible in both figures, the NO_2 concentrations are more homogeneous across the district while the NO concentrations are high near to freeways and major roads, but drop more rapidly (within 100 m) to background levels. This demonstrates the ability of LUR to distinguish between primary and secondary pollutants. For NO and NO_2 the *Road Length* and *Traffic Density* variables produced similar exposure surfaces (available in the Supporting Information). Less than 2% of populated grid cells were truncated to 120% of the maximum measured values in all cases.

Our main motivation for LUR is to produce tools that can generate reliable estimates of ambient air pollutant concentrations for epidemiological and risk assessment applications. To illustrate this, we use population data associated with the 58,000 spatial distribution points to examine estimated exposures for ~1.93 million residents. Based on the *Traffic Density* surfaces the mean exposure estimates were 25.5 (SD = 11.1) and 15.2 (SD = 3.0) ppb for NO and NO_2, respectively. Values for the *Road Length* surfaces were almost identical. In the case of NO the mean expected exposure was 75% higher than that recorded by the monitoring network. When divided by exposure quartiles we estimate that 29, 34, 33, and 26% of Vancouver residents fall into the highest categories for NO, NO_2, $PM_{2.5}$, and ABS concentrations, respectively.

Through systematic site selection, sampling period identification, and model evaluation we have demonstrated some important characteristics of LUR. First, the method is sensitive to differences in spatial distribution between NO, NO_2, and light-absorbing (elemental) carbon (*34*), which improves our confidence in its overall credibility. Second, models of equally predictive value can be generated from different traffic metrics, confirming that LUR can be applied in varied contexts. We have also more clearly defined methodological questions that remain to be addressed regarding required sample sizes and sampling site identification. Within the geographically and climatologically complex setting of the GVRD we found the performance of LUR to be relatively consistent with results from other regions.

Acknowledgments

We acknowledge the following: Tracy Kirkham, Yang Chang, Owen Thomas, Tim Ma, and David Thomson for technical assistance with sampling and analysis; Fred Prystarz, Michiyo McGaughy, and Al Percival at the Greater Vancouver Regional District for logistical support; Clark Lim at Translink for access to EMME/2 data; numerous engineers within GVRD municipalities for their cooperation; Bob Nash at BC Hydro for access to utility poles; peer reviewers for their incisive comments; and financial support from the GVRD, Environment Canada, the BC Ministry of Environment, the Michael Smith Foundation for Health Research, and the Center for Health and Environment Research at The University of British Columbia.

Supporting Information Available

Map of sampling sites; maps of concentrations of NO_x and fine particulate matter; table of comparison of land use regression in different locations. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Hoek, G.; Meliefste, K.; Cyrys, J.; Lewne, M.; Bellander, T.; Brauer, M.; Fischer, P.; Gehring, U.; Heinrich, J.; van Vliet, P.; Brunekreef, B. Spatial variability of fine particle concentrations in three European areas. *Atmos. Environ.* **2002**, *36* (25), 4077−4088.

(2) Wilhelm, M.; Ritz, B. Local variations in CO and particulate air pollution and adverse birth outcomes in Los Angeles County, California, U.S.A. *Environ. Health Perspect.* **2005**, *113* (9), 1212−1221.

(3) Ponce, N.; Hoggatt, K.; Wilhelm, M.; Ritz, B. Preterm birth: the interaction of traffic-related air pollution with economic hardship in Los Angeles neighborhoods. *Am. J. Epidemiol.* **2005**, *162* (2), 140−148.

(4) Gordian, M.; Haneuse, S.; Wakefield, J. An investigation of the association between traffic exposure and the diagnosis of asthma in children. *J. Exposure Anal. Environ. Epidemiol.* **2006**, *16* (1), 49−55.

(5) Finkelstein, M. M.; Jerrett, M.; Sears, M. R. Traffic air pollution and mortality rate advancement periods. *Am. J. Epidemiol.* **2004**, *160* (2), 173−7.

(6) Heinrich, J.; Schwarze, P.; Stilianakis, N.; Momas, I.; Medina, S.; Totlandsdal, A.; von Bree, L.; Kuna-Dibbert, B.; Krzyzanowski, M. Studies on health effects of transport-related air pollution. In *Health Effects of Transport-Related Air Pollution*; Krzyzanowski, M., Kuna-Dibbert, B., Schneider, J., Eds.; World Health Organization Press: Geneva, 2005.

(7) Gauderman, W.; Avol, E.; Lurmann, F.; Kuenzli, N.; Gilliland, F.; Peters, J.; McConnell, R. Childhood asthma and exposure to traffic and nitrogen dioxide. *Epidemiology* **2005**, *16* (6), 737−743.

(8) Zhu, Y.; Hinds, W.; Kim, S.; Sioutas, C. Concentration and size distribution of ultrafine particles near a major highway. *J. Air Waste Manage. Assoc.* **2002**, *52*, 1032−1042.

(9) Zhu, Y.; Hinds, W.; Kim, S.; Shen, S.; Sioutas, C. Study on ultrafine particles and other vehicular pollutants near a busy highway. *Atmos. Environ.* **2002**, *36*, 4375−4383.

(10) Zhang, K.; Wexler, A.; Zhu, Y.; Hinds, W.; Sioutas, C. Evolution of particle number distribution near roadways. Part II: The 'road-to-ambient' process. *Atmos. Environ.* **2004**, *38*, 6655−6665.

(11) Gilbert, N.; Woodhouse, S.; Stieb, D.; Brook, J. Ambient nitrogen dioxide and distance from a major highway. *Sci. Total Environ.* **2003**, *312*, 43−46.

(12) 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. *J. Exposure Anal. Environ. Epidemiol.* **2005**, *15* (2), 185−204.

(13) Van Atten, C.; Brauer, M.; Funk, T.; Gilbert, N.; Graham, L.; Kaden, D.; Miller, P.; Wheeler, A.; White, R. Assessing population exposure to motor vehicle exhaust. *Rev. Environ. Health.* **2005**, *20* (3), 195−214.

(14) Briggs, D. J.; Collins, S.; Elliot, P.; Fischer, P.; Kingham, S.; Lebret, E.; Pryl, K.; van Reeuwijk, H.; Smallbone, K.; van der Veen, A. Mapping urban air pollution using GIS: a regression-based approach. *Int. J. Geogr. Inf. Sci.* **1997**, *11* (7), 699−718.

(15) Gilbert, N. L.; Goldberg, M. S.; Beckerman, B.; Brook, J. R.; Jerrett, M. Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. *J. Air Waste Manage. Assoc.* **2005**, *55*, 1059−1063.

(16) Gonzales, M.; Qualls, C.; Hudgens, E.; Neas, L. Characterization of a spatial gradient of nitrogen dioxide across a United States-Mexico border city during winter. *Sci. Total Environ.* **2005**, *337* (1−3), 163−173.

(17) Hoek, G.; Meliefste, K.; Brauer, M.; van Vliet, P.; Brunekreef, B.; Fischer, P. *Risk assessment of exposure to traffic-related air pollution for the development of inhalant allergy, asthma and other chronic respiratory conditions in children (TRAPCA), Final Report*; IRAS, Utrecht University: Utrecht, The Netherlands, 2001.

(18) Ross, Z.; English, P. B.; Scalf, R.; Gunier, R.; Smorodinsky, S.; Wall, S.; Jerrett, M. Nitrogen dioxide prediction in Southern California using land use regression modeling: Potential for environmental health analyses. *J. Exposure Anal. Environ. Epidemiol.* **2006**, *16* (2), 106−114.

(19) Hochadel, M.; Heinrich, J.; Gehring, U.; Morgenstern, V.; Kuhlbusch, T.; Link, E.; Wichmann, H.-E.; Kramer, U. Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information. *Atmos. Environ.* **2006**, *40* (3), 542−553.

(20) Kanaroglou, P. S.; Jerrett, M.; Morrison, J.; Beckerman, B.; Arain, M. A.; Gilbert, N. L.; Brook, J. R. Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmos. Environ.* **2005**, *39* (13), 2399−2409.

(21) Gray, A.; Cass, G. Source contributions to atmospheric fine carbon particle concentrations. *Atmos. Environ.* **1998**, *32* (22), 3805−3825.

(22) Yu, L.; Yue, P.; Teng, H. Comparative study of EMME/2 and QRS II for modeling a small community. *Transport. Res. Rec.* **2003**, *1858*, 103−111.

(23) Brauer, M.; Brumm, J.; Ebelt, S. *Evaluation of ambient air pollution in the Lower Mainland of British Columbia: Public health impacts, spatial variability, and temporal patterns*; Report to the Administrative Council of Lower Mainland Medical Health Officers: Vancouver, BC, 2000; 56 pp. http://www.cher.ubc.ca/PDFs/LMAirQualityReport.pdf (accessed February 12, 2007).

(24) *NO, NO_2, NO_x and SO_2 Sampling Protocol using the Ogawa Sampler*, 4th ed.; Ogawa and Company, USA, Inc.: Pompano Beach, FL. 1998.

(25) Ebelt, S.; Fisher, T.; Petkau, A.; Vedal, S.; Brauer, M. Exposure of chronic obstructive pulmonary disease (COPD) patients to particles: relationship between personal exposure and ambient air concentrations. *J. Air Waste Manage. Assoc.* **2000**, *50*, 1081−1094.

(26) Determination of absorption coefficient using reflectometric method. In *Exposure and Risk Assessment for Fine and Ultrafine Particles in Ambient Air*; KTL National Public Health Institute: Kuopio, Finland; 1998. http://www.ktl.fi/ultra/adobe/out/sop-abs.pdf (accessed February 12, 2007).

(27) *ISO 9835:1993; (E): Ambient air - Determination of a black smoke index*, 15th ed.; American National Standards Institute: Washington, DC, 1993.

(28) Noullett, M.; Jackson, P.; Brauer, M. Winter measurements of children's personal exposure and ambient fine particle mass, sulphate and light-absorbing components in a northern community. *Atmos. Environ.* **2006**, *40*, 1971−1990.

(29) Rich, K. Air pollution and patients with implanted cardiac defibrillators: An epidemiological analysis and assessment of exposure. MSc, University of British Columbia, Vancouver, 2003.

(30) Cyrys, J.; Heinrich, J.; Hoek, G.; Meliefste, K.; Lewné, M.; Gehring, U.; Bellander, T.; Fischer, T.; van Vliet, P.; Brauer, M.; Wichmann, E.; Brunekreef, B. Comparison between different traffic-related particle indicators: Elemental carbon (EC), PM2.5 mass, and absorbance. *J. Exposure Anal. Environ. Epidemiol.* **2003**, *13* (2), 134−143.

(31) Brauer, M.; Henderson, S. B. *Diesel exhaust particles and related air pollution from traffic sources in the Lower Mainland*; Report submitted to Health Canada: Vancouver, BC, 2003; 23 pp. http://www.cher.ubc.ca/PDFs/diesel02.pdf (accessed February 12, 2007).

(32) Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17−23.

(33) Brauer, M.; Hoek, G.; Van Vliet, P.; Meliefste, K.; Fischer, P.; Gehring, U.; Heinrich, J.; Cyrys, J.; Bellander, T.; Lewne, M.; Brunekreef, B. Prediction of long term average particulate air pollution concentrations by traffic indicators for epidemiological studies. *Epidemiology* **2003**, *14*, 228−239.

(34) Yang, C. Y.; Yu, S. T.; Chang, C. C. Respiratory symptoms in primary schoolchildren living near a freeway in Taiwan. *J. Toxicol. Environ. Health A* **2002**, *65* (10), 747−55.

Received for review March 21, 2006. Revised manuscript received November 3, 2006. Accepted January 19, 2007.

ES0606780