Journal of Exposure Science and Environmental Epidemiology (2006) 16, 106–114
© 2006 Nature Publishing Group All rights reserved 1559-0631/06/$30.00
www.nature.com/jes

Nitrogen dioxide prediction in Southern California using land use regression modeling: potential for environmental health analyses

ZEV ROSS,[a] PAUL B. ENGLISH,[b] RUSTY SCALF,[c] ROBERT GUNIER,[b] SVETLANA SMORODINSKY,[b] STEVE WALL[b] AND MICHAEL JERRETT[d]

[a]*ZevRoss Spatial Analysis, 303 Fairmount Ave., Ithaca, New York 14850, USA*
[b]*Division of Environmental and Occupational Disease Control, California Department of Health Services, 1515 Clay St., #1700, Oakland, California 94612, USA*
[c]*Impact Assessment Inc., 1515 Clay St., #1700, Oakland, California 94612, USA*
[d]*Division of Biostatistics, Department of Preventive Medicine and Department of Geography, University of Southern California, 1540 Alcazar Street, CHP-220, Los Angeles, California 90089-9011, USA*

We modeled the intraurban distribution of nitrogen dioxide (NO_2), a marker for traffic pollution, with land use regression, a promising new exposure classification technique. We deployed diffusion tubes to measure NO_2 levels at 39 locations in the fall of 2003 in San Diego County, CA, USA. At each sample location, we constructed circular buffers in a geographic information system and captured information on roads, traffic flow, land use, population and housing. Using multiple linear regression, we were able to predict 79% of the variation in NO_2 levels with four variables: traffic density within 40–300 m of the sampling location, traffic density within 300–1000 m, length of road within 40 m and distance to the Pacific coast. Applying this model to validation samples showed that the model predicted NO_2 levels within, on average, 2.1 p.p.b for 12 training sites initially excluded from the model. Our evaluation of this land use regression model showed that this method had excellent prediction and robustness in a North American context. These models may be useful tools in evaluating health effects of long-term exposure to traffic-related pollution.
Journal of Exposure Science and Environmental Epidemiology (2006) **16,** 106–114. doi:10.1038/sj.jea.7500442; published online 27 July 2005

Keywords: *nitrogen dioxide, traffic, land use regression, GIS.*

Introduction

Leading researchers have identified the development of models for assessing air pollution exposure *within* cities as a priority for future research (Brunekreef and Holgate, 2002; National Research Council, 2002; Brauer et al., 2003). This paper reports on the first stage of a multiyear study sponsored by the National Cancer Institute to compare and evaluate different models for assigning within-community or intraurban air pollution exposures. Our goal is to model the intraurban distribution of nitrogen dioxide (NO_2), a marker for traffic pollution, with the land use regression method developed by Briggs et al. (1997). Although this method has performed well in European contexts, no studies have used this method to estimate criteria air pollutant exposures in the United States, and a need exists to understand whether the method will perform adequately in a US setting, where land use, traffic and climate patterns differ from those in Europe. Land use regressions have cost and implementation advantages over dispersion models, and they arguably model pollution concentrations more accurately than proximity measures that use distance from source to approximate exposures (Jerrett et al., 2005). In future research, we will compare assigned exposures from these land use regression models with those from proximity models, geostatistical kriging and ADMS urban dispersion models.

Recent studies have shown that within-community or intraurban exposure gradients may be associated with larger health effects than the between-community exposures used in the earlier studies (Krewski et al., 2000; Pope et al., 2002). For example, Hoek et al. (2002) reported a near doubling of cardiopulmonary mortality (relative risk = 1.95, 95% CI 1.09–3.52) for Dutch subjects living near major roads in a cohort of 5000 people, where control was available for many confounding variables. Recent cohort studies from Canada have also discerned large, significant health effects at the intraurban scale (Finkelstein et al., 2003, 2004). Although these findings may be robust, the basic exposure models used in these analyses may misclassify exposure because they treat the continuous air pollution field as a discrete entity, that is,

1. Address all correspondence to: Dr. P.B. English, Environmental Health Investigations Branch, Division of Environmental and Occupational Disease Control, California Department of Health Services, 1515 Clay St., #1700, Oakland, CA 94612, USA. Tel.: +1 510 622 4508.
Fax: +1 510 622 4505. E-mail: PEnglish@dhs.ca.gov
Received 11 October 2004; accepted 8 June 2005; published online 27 July 2005

either within or outside a specified distance from a road. With the high probability of errors in exposure estimates from these studies, questions remain about the validity of results from health effect studies that use exposure surrogates such as road buffers. This scientific uncertainty impedes efforts by policy-makers to protect public health. Studies comparing different exposure models in relation to each other and observed health effects will reduce some of the these uncertainties.

Methods

Study Site

San Diego County, a region of 2.8 million people, was chosen as the study site due to concerns about high levels of traffic-related pollutants and the availability of traffic and other spatial data. Expanding trade and truck traffic and population growth throughout the region due to the North American Free Trade Agreement has placed strain on the local airshed. Local evaluation of a traffic pollution model is a high priority in this region in order to develop maps of long-term exposure to air pollutants for use in epidemiologic studies.

Monitoring Site Selection

We used public buildings such as police stations, libraries and religious institutions as locations for NO_2 monitoring. We selected from approximately 1300 potential monitoring locations based on (1) terciles of estimated diesel particulate matter concentrations at the census tract level (United States Environmental Protection Agency (USEPA), 2001); (2) the number of potential subjects within 200 m of the location for various epidemiological studies; (3) proximity to freeways and (4) location within the western part of San Diego County, the more populated area, due to field work time constraints. We identified 61 potential monitoring sites that agreed to participate based on the above criteria. For comparison purposes, we included three locations where the San Diego Air Pollution Control District continuously monitors for NO_2 using chemiluminescence analyzers (Model 42, Thermo Environmental Instruments Inc.). We also included a background site at a field office used by the Environmental Health Investigations Branch (EHIB) of the Department of Health Services near the San Diego coast. While in the field, we found that a number of eligible sites were not suitable for NO_2 monitor placement (due, for example, to lack of safe space to place a monitor). Altogether, we monitored for NO_2 at 39 sites (Figure 1).

NO_2 Monitoring and Analysis

We sampled for NO_2 using passive diffusion tubes (Gradko International Ltd., P/N: DIF100RTU-R). NO_2 travels by diffusion from the bottom of the acrylic tubes (7 cm long $\times$ 1 cm diameter) and is collected on metal mesh screens coated with triethanolamine at the top of the tubes (Palmes et al., 1976). We determined the mass of nitrite as NO_2 from each tube using ion chromatography (Dionex Corp., 500 IC/AS9 Analytical Column). We placed duplicate tubes at each monitoring site inside a plastic shelter to protect the diffusion tubes from oversampling at high wind speeds. The NO_2 concentration used in this analysis was the average of the two tubes. We used cable ties to mount the shelters to drain pipes, fence posts, trees or plastic stakes with the diffusion tubes perpendicular to the ground. We attempted to place the tubes at breathing zone height (1.5–2 m) but the ultimate height of the sampler was determined by the availability of a safe attachment site and ranged from 1.2 to 8.2 m, with the majority between 1.5 and 3 m. The tubes were exposed for 13–16 days in early October 2003. Due to logistic constraints, not all of the samples could be deployed on the same day but were deployed over a 3-day period (October 1–3).

NO_2 concentrations vary throughout the year, with the lowest levels in June–August and the highest levels in December and January. NO_2 levels in October are slightly above the average annual levels for most recent years but are more representative of average annual concentrations than the summer and winter extremes.

Buffering

For each sample location, we constructed a series of concentric circle buffers with radii of 40, 300, 500 and 1000 m, using ArcGIS version 8.3. All geographic information system (GIS) layers were then clipped to these buffers. Although an examination of dispersion models suggests that 80–90% of the decay of pollutants occurred within 150 and 200 m, research in Toronto, a city of similar size to San Diego, found that a detectable effect could be identified at higher distances necessitating the inclusion of wider buffers (Jerrett et al., in press). In particular, they found that both traffic counts within 500 m and location within 1500 m (downwind) of an expressway were statistically significant predictors of NO_2 concentration.

Traffic Data

The roads and traffic data were developed by EHIB based on publicly and nonpublicly available data. Road lines come from the San Diego Association of Governments (SANDAG). Single freeway lines were replaced with more accurate double lines based on SANDAG's transportation planning layer. On-ramps and off-ramps were added using linework from Navigational Technologies (2002).

We obtained average daily traffic (ADT) count data for 2000 from SANDAG and the California Department of Transportation (CALTRANS). ADT is defined as the annual average number of cars traveling in both directions per weekday. These values are based on traffic counters and

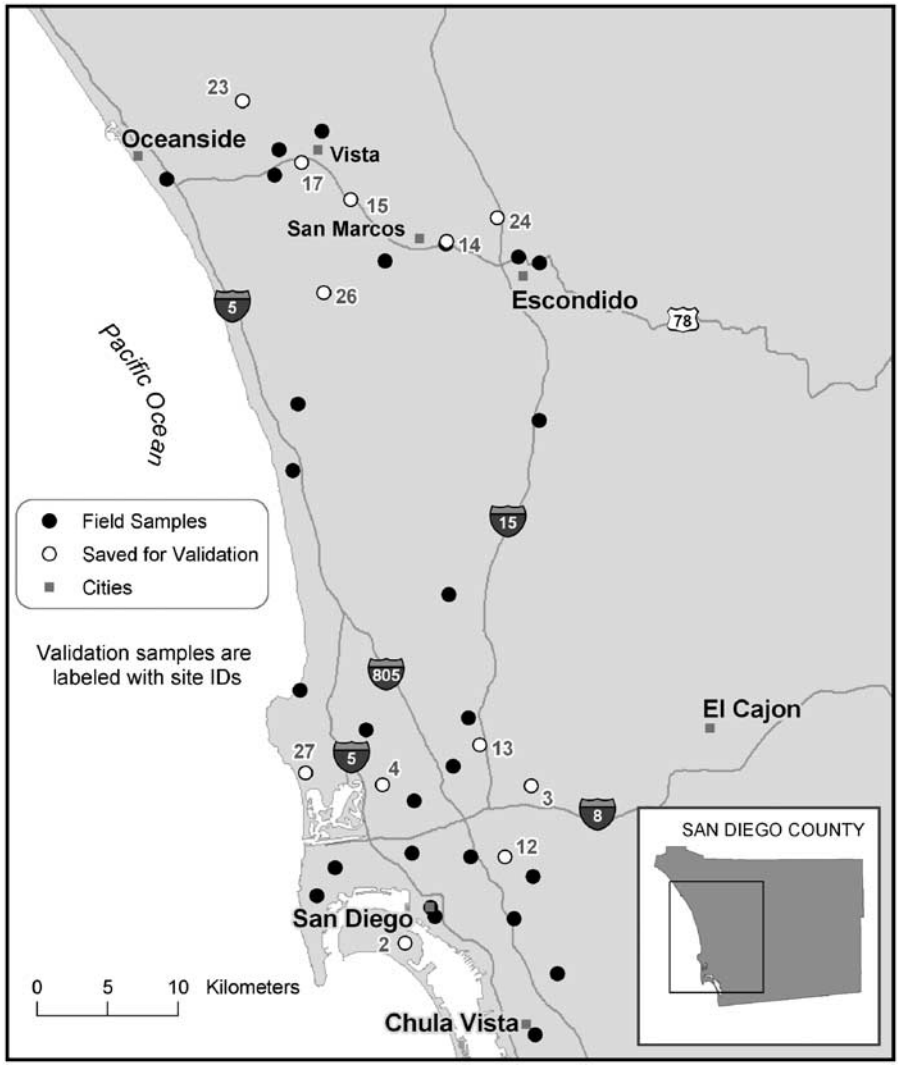

Figure 1. Sampling locations.

are collected for nearly all highways and major roads in the county. A smaller proportion of local roads was also measured. Nonattributed roads were assigned default values based on the number of lanes as well as ADT for roads nearby. ADT values range from a minimum of 400 to more than 250,000 vehicles per day. In all cases, the ADT was multiplied by the length of the road segment and divided by 24,000 representing vehicle-kilometers per hour.

Candidate predictor variables were derived from the buffered road layer and traffic data. Road- and traffic-related variables included the total length of road (km), traffic density derived from ADT $\times$ length of road (vehicle-kilometers per hour), length of interstate highways (km) and a traffic variable weighted by road class.

Land Use Data

Land use data were obtained from SANDAG. The 2003 land use layer was derived by SANDAG from aerial imagery, SanGIS landbase and other sources (SANDAG, 2004). Each polygon in the land use layer has been assigned a land use code representing the type of land use. There are approximately 30 major land use categories and many of these major land use categories have been subdivided into minor land use categories. Candidate land use-related predictor variables included total area (square meters) covered by industry, heavy industry, freeways and multifamily residential housing.

Census Data

We used the US Bureau of the Census Summary File 1 (2000) at the block group level to calculate population and housing density for buffers. Similar to the road and land use layers, the Census block group layer was clipped to the buffers discussed above. We estimated Census characteristics such as population in each buffer by computing the proportion of a block group's total area located in that buffer and multiplying the variables by that proportion. This

assumes an even distribution of these variables within a block group, but this unrealistic assumption is difficult to avoid due to limitations on publicly available population data. Candidate Census predictor variables included population, population by race, housing units and households.

Additional Variables

We also included several additional variables that showed promise in previous analyses. These include minimum distance to a major road (defined as those with ADT of at least 50,000 vehicles), altitude and sampler height (Brauer et al., 2003).

There was also some concern that proximity to major urban areas could have some residual effect on NO_2 levels. While suburban areas may have significantly lower traffic density and lower density road networks, NO_2 levels may be influenced by nearby urban areas. In order to test this effect, we included as a potential predictor minimum distance to a large urban area ($>100{,}000$ people).

Coastal winds can have a significant effect on air pollutant levels. The prevailing daytime winds in San Diego County generally drive coastal pollutants inland. We, therefore, included distance (due west) to the Pacific coast (in meters) as a potential predictor variable.

Statistical Modeling

For the initial modeling, a random sample of 12 monitoring locations stratified by quadrant (NW, NE, SW, SE) was removed for subsequent validation. Three samples were removed from each quadrant.

Initial modeling was conducted on the remaining 27 measured concentrations. Candidate predictors were evaluated using primarily exploratory data analysis and manual forward selection procedures. Stepwise and all-subsets procedures were also used to identify strong models that would not otherwise be identified using forward selection. All variables above were considered both independently of other variables in the forward selection process and in combination with other variables in both stepwise and all-subsets procedures.

Variables were included based on an assessment of their explanatory power (sums of squares), Mallow's Cp value, statistical significance and collinearity with other added variables. The best model based on 27 samples was then used to predict the 12 samples removed for validation. The variables from the best model were also included in a model using the full set of 39 samples to test the sensitivity of the coefficients.

We evaluated residual spatial autocorrelation to assess the independence assumption (which spatially autocorrelated errors violate). We used a lattice approach to evaluate residual spatial autocorrelation. Using the 39 sample points, we constructed a Thiessen polygon theme and then calculated Moran's I and Geary's C values (measures of spatial autocorrelation) using a Queen's adjacency neighborhood structure.

The sensitivity of the model parameters to the selection of samples was evaluated with a bootstrap. In each iteration, five randomly selected samples were removed, the final model was run and the coefficients were recorded. The five samples were returned to the pool and a new randomly selected group of five was removed. This procedure was repeated 20,000 times.

Finally, for visualization purposes, we used both regression and kriging to generate a continuous map of NO_2. Although, in an epidemiologic setting, regression would be used to predict NO_2 concentrations at point locations (e.g., case/control residences), a smooth, region-wide map of concentrations enables a visual assessment of model results. In order to generate a continuous surface, we applied the final regression model at 500 randomly selected sample locations and then kriged these 500 predictions.

Statistical models and variograms were calculated using S-PLUS statistical software (version 6.1, Insightful Corporation, Seattle, WA). Kriging was performed with the ArcGIS Geostatistical Analyst (version 8.3, ESRI, Redlands, CA, USA).

Results

Laboratory Results

We recovered diffusion tubes from 100% of the monitoring sites, although one tube cracked during removal from the plastic shelter and was discarded. At all other sites, we calculated the 2-week average NO_2 concentration (p.p.b) from the duplicate tubes. For the one site without a duplicate tube, the single measured concentration was used in place of an average. NO_2 concentrations at the 39 sites ranged from 8.08 to 28.74 p.p.b. The mean concentration was 14.84 p.p.b, with a median of 15.25 p.p.b and a standard deviation of 4.70 p.p.b, indicating a fairly normal distribution. There was good agreement (± 2 p.p.b) between the average NO_2 concentrations measured by the diffusion tubes and San Diego Air Pollution Control District's chemiluminescence analyzers for the 2-week monitoring period. The interclass correlation coefficient for the collocated tubes was 0.97 and the mean coefficient of variation was 3.3%, indicating good reproducibility of the diffusion tubes.

Buffering and Potential Predictors

Although all the sample points are located in the Census-defined San Diego "urbanized" area, there is significant diversity in the traffic and population patterns in sample buffers (Table 1). The samples are located in areas ranging from heavily urbanized, high-traffic areas to areas with few people or roads. Total population within the 1000 m buffers, for example, ranged from approximately 400 to 29,000.

Table 1. Summary statistics for NO_2 and selected potential predictor variables.

Variable	Mean	Median	Minimum	Maximum	First quartile	Third quartile
Nitrogen dioxide (p.p.b.)	14.84	15.25	8.08	28.74	10.78	17.47
Sampler height (m)	2.41	1.85	1.21	8.15	1.26	2.46
Elevation (m)	96	94	7	234	40	129
Industrial land use, 300 m (m^2)	11,801	—	—	133,071	—	568
Industrial land use, 1000 m (m^2)	113,806	8201	—	1,236,896	—	55,927
Length of road, 40 m (m)	71	76	—	153	—	115
Length of road, 300 m (m)	3431	3436	192	5728	2738	4115
Traffic density, 300 m (vehicle-km/h)	1061	769	38	4920	399	1148
Traffic density, 1000 m (vehicle-km/h)	11,408	8814	1831	31,202	5416	16,500
Total population, 1000 m	8437	7917	392	28,743	4569	10,854
Total housing units, 1000 m	3220	2862	133	9193	1837	4414
Distance to the coast (m)	11,079	10,758	315	23,974	7075	15,000
Distance to nearest high-traffic road (m)	1228	892	48	5947	426	1703

Note: Summary statistics are for complete buffers. In order to avoid collinearity, internal buffers have been subtracted from larger buffers for the final models.

Traffic density within the 1000 m buffers also varied substantially among sample points, ranging from approximately 1800 to 31,000 vehicle-kilometers per hour. All of the four locations with the highest traffic density within 1000 m are located within 15 km of the San Diego city center.

Statistical Modeling

The approximately normal distribution of the samples permitted us to use multiple linear regression with the standard Gaussian assumptions. Traffic density within the 300 m buffer (traffic300) was consistently the strongest predictor of NO_2 levels explaining over 50% of the variation in the 27 samples. A plot of traffic300 against NO_2 reveals a strong linear relationship with some indication of a slight curvature (not shown). We explored accounting for the curvilinear relationship with a second-order polynomial term, but found that it was insignificant.

Distance to the coast was, independent of other variables, the second strongest predictor (after removing from consideration collinear variables). Distance to the coast was also the strongest predictor after adding traffic300. Traffic300 and distance to the coast together accounted for 64% of the variation in the NO_2 concentrations.

After adding both traffic300 and distance to the coast, heavy industry within the 1000 m buffer becomes the strongest predictor. While this association is consistent with previous work (Briggs et al., 1997), just two of the 27 samples had any heavy industry within 1000 m. Furthermore, the results from a model with the heavy industry variable included reveal an extremely high Cook's distance value (7.4) for one of these two samples. When this one sample is removed, heavy industry becomes nonsignificant, suggesting an unstable relationship. Heavy industry, therefore, was excluded.

The next most important variable, after removing from consideration collinear variables and heavy industry within the 1000 m buffer, was traffic density within the 1000 m buffer. To ensure that we did not "double count" roadways, we subtracted traffic in the 300 m buffer from traffic within the 1000 m buffer producing a doughnut-shaped buffer of traffic 300–1000 m (traffic300–1000). This variable remained the next most important predictor. These three variables, traffic300, distance to the coast and traffic300–1000, accounted for 70% of the variation in the 27 samples.

The last variable to be included was the length of road in the 40 m buffer. Again, to ensure that we did not double count roads, the 40 m portion of the 300 m buffer was removed. This reduced buffer (buffer 40–300) was still the strongest NO_2 predictor when added to the model with three other variables. This model with four variables (Table 2) explained 79% of the variation. Model diagnostics suggested that the model conformed to underlying assumptions of the regression model.

The model with four predictors was then applied to the 12 validation samples and performed well. It predicted, on average, to within 2.1 p.p.b (with a maximum residual of 5.6 p.p.b) (Figure 2). The model overpredicts nine of the 12 samples, resulting in a fractional bias (2(AvgPred − AvgObs)/(AvgPred + AvgObs)) of 11.9%.

The 12 validation samples were then returned to the pool and a new regression equation was calculated using the same predictors. This model also performed well (Table 3), accounting for 77% of the variation in the 39 samples. In all cases, variance inflation factors were below 1.3, indicating acceptable levels of collinearity. Diagnostic plots reveal no major problems, although two samples show relatively higher Cook's distances (0.35 and 0.26). The model F-statistic was 27.94 on 4 and 34 degrees of freedom ($P < 0.00005$).

A bootstrap to evaluate the sensitivity of the parameters to the actual samples included in the modeling suggests that the intercept, road length (40 m), traffic density (300–1000 m) and distance to the coast variables are resistant to the sample

choice (Figure 3). Traffic density (40–300 m) appeared to be somewhat affected. In particular, the histogram for this variable shows a modest bimodal shape. Further investigation revealed that the inclusion of a single sample (located in Escondido) resulted in a 30% change in the parameter. This sample location has the highest traffic density in the 40–300 m buffer and the second highest NO_2 value. Overall, however, the parameter shift has little overall effect on predictions. Replacing the current coefficient value (0.017) with the mean value of the right tail (0.022) yields predictions that are, on average, just 3% different.

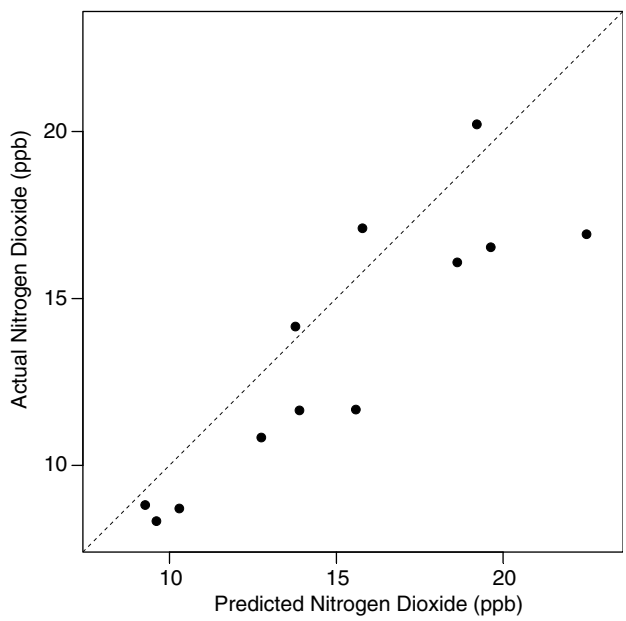

Figure 2. Predictions for 12 samples withheld for validation.

The full model accounts for a significant proportion of the autocorrelation in NO_2 concentrations. NO_2 values themselves are highly spatially autocorrelated with statistically significant Moran and Geary coefficients of 0.37 ($P<0.0005$) and 0.57 ($P<0.0005$), respectively. (For positive spatial autocorrelation, Moran ranges from 0 to 1 with 1 representing perfect autocorrelation and Geary ranges from 1 to zero with zero representing perfect autocorrelation.) Under the full model, the Moran and Geary values become 0.17 ($P=0.04$) and 0.77 ($P=0.03$), respectively. While the autocorrelation levels remain statistically significant at the 0.05 level, the modest levels observed would not be expected to have a significant effect on model results.

We were able to apply this final regression model and generate predictions for 499 of the 500 random locations (one fell just outside the boundaries of the road layer). An exponential variogram function fit the variogram based on these 499 points well and yielded a range of approximately 5 km, a nugget of 4.90 and a partial sill of 12.41. The final kriged surface is shown in Figure 4.

Discussion

We used field monitoring data in San Diego County, CA, to build and evaluate a land use regression model. Traffic within a 300 m buffer of the sampling location was the most consistent predictor of NO_2 levels, accounting for over 54% of the variation. When combining this variable with three

Table 2. Final model results based on 27 sample locations (the 12 validation samples were excluded).

	Coeff.	SE	Scaled coeff.	Scaled SE	t	P	VIF
Intercept	5.7423	1.4231	15.47	0.47	4.04	0.0006	—
Road length (40 m)	28.4658	8.7174	1.64	0.50	3.27	0.0035	1.10
Traffic volume (40–300 m)	0.0020	0.0005	2.37	0.56	4.25	0.0003	1.36
Traffic volume (300–1000 m)	0.0002	0.0001	1.51	0.50	3.04	0.0060	1.08
Distance to coast	0.0003	0.0001	2.10	0.56	3.75	0.0011	1.38

$R^2=79\%$. Variables scaled to mean = 0 and standard deviation = 1.

Table 3. Final model results after 12 validation samples are returned to the pool of samples (39 total sample locations).

	Coeff.	SE	Scaled coeff.	Scaled SE	t	P	VIF
Intercept	5.3051	1.1039	14.84	0.38	4.81	0.0000	—
Road length (40 m)	29.4083	7.0382	1.67	0.40	4.18	0.0002	1.05
Traffic volume (40–300 m)	0.0017	0.0004	1.87	0.44	4.23	0.0002	1.29
Traffic volume (300–1000 m)	0.0002	0.0001	1.51	0.41	3.72	0.0007	1.08
Distance to coast	0.0003	0.0001	2.01	0.44	4.62	0.0001	1.25

$R^2=77\%$. Variables scaled to mean = 0 and standard deviation = 1.

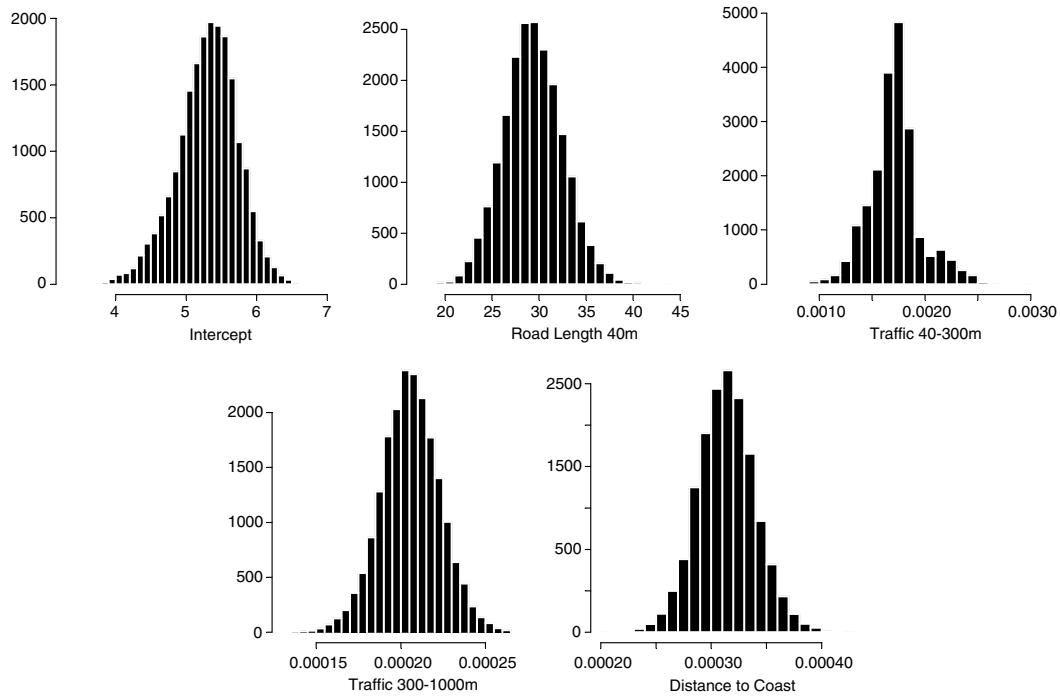

Figure 3. Histograms of parameter values after a bootstrap with 20,000 iterations.

other variables (traffic in the 300–1000 m buffer, length of road in the 40 m buffer and distance to the coast), the model explains 79% of the variation in the model.

When applying this four variable model to our validation samples, we found that the model predicted NO_2 levels very well, predicting levels on average within 2.1 p.p.b and all of the predicted estimates were within a factor of 1.5 times the range of the observed levels. Briggs et al., in applying a similar model in four urban areas of the UK, found that their model gave estimates of the mean NO_2 concentrations within 70–90% of a factor of 1.5 of the observed mean (Briggs et al., 2000). In general, residuals from the model were less than 20% of the measured values. The residuals exhibit a somewhat bimodal form, which is likely explained by the bimodal shape of the NO_2 concentrations themselves.

While most previous studies did not incorporate wind or coastal influences in their models, we found strong coastal effects. In San Diego, distance to the coast may be acting as a proxy for the background NO_2 concentration. The wind blows predominantly from the west, so locations near the coast receive relatively clean ocean air while the more inland locations receive air with emissions that occurred along the coast. We did not evaluate the effect of local wind speed or direction in the regression model due to lack of data at the monitoring sites, but this information would likely improve the overall model prediction.

This study focuses on a 2-week period in October to demonstrate the utility of land use regression for NO_2 prediction. As such, the predictions and kriged surface should only be seen as representative of fall conditions in the study area, not annual averages. It should be noted that, unlike the current study, previous studies using the same approach (e.g., Briggs et al., 1997) attempted to model annual average concentrations by taking multiple samples (2–4) at the same locations over the course of the year. Although applying this approach to the current study would result in different NO_2 concentrations, we would expect that the relative differences of sample locations would be preserved (e.g., areas of relatively high NO_2 concentrations would continue to have relatively high NO_2 concentrations) and that the same variables would continue to be predictive.

The exposure assessment methods used to estimate levels of traffic-related pollutants are a critical component of health effect studies. Unfortunately, the exposure assessment piece of these studies is often the weakest. Simple models such as proximity of households to roadways have been used in epidemiologic studies (English et al., 1999; Wilkinson et al., 1999; Hoek et al., 2002). Although NO_2 levels in households have been found to correlate with distance to roadways and traffic density (Rijinders et al., 2001), the proximity method has been criticized due to a high potential for exposure misclassification (Jerrett et al., in press). Assigning exposure based solely on proximity measures ignores other traffic, land use and topographical factors that may influence exposure. On the other hand, the most sophisticated dispersion models that integrate meteorological data and atmospheric chemistry

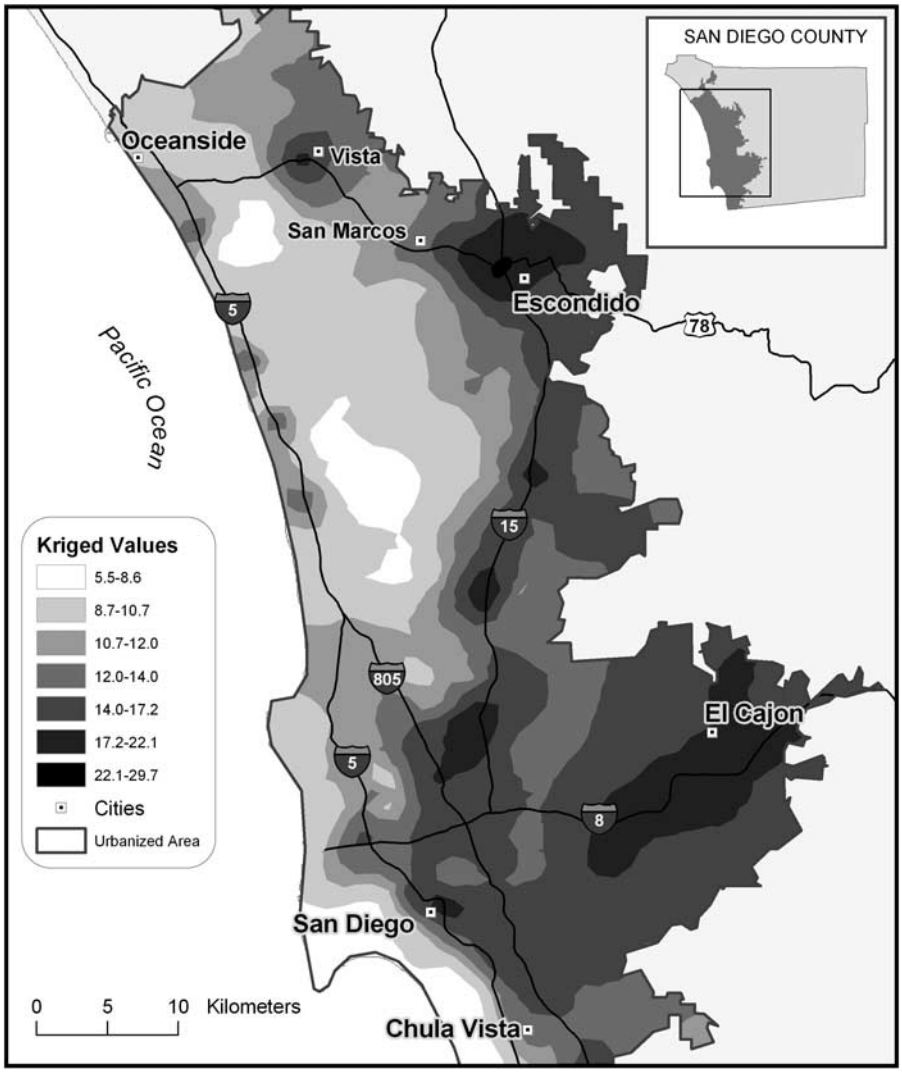

Figure 4. Kriged prediction surface (fall 2003).

require many data inputs and months of training. The land use regression approach is attractive, as it predicts well, is not nearly as data intensive as dispersion modeling but can improve upon proximity methods by incorporating a wider range of factors that influence exposure.

In summary, we were able to evaluate a land use regression method to predict levels of NO_2 from traffic sources in a California urban area using field measurements. We found, in general, that the model produced good predictions, was robust and was relatively simple in terms of data inputs required and analysis. To our knowledge, this research represents the first application of the land use regression method for a criteria pollutant in a large US city. The method performed as well or better in the US than in the other study locations. This model should be further evaluated in additional locations and with other pollutants but appears to have great promise for use in evaluating chronic exposures to traffic exhaust in epidemiologic studies.

Acknowledgements

This publication was supported by grant number 5-R21-CA094723-03 from the National Cancer Institute. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute. Dr. Jerrett's time was partly funded by grants number, 5P30 ES07048, 5P30 ES05605-14, 5P01 ES09581, and IP01 ES11627 from the National Institute of Environmental Health Sciences.

References

Brauer M., Hoek G., van Vliet P, et al. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology* 2003: 14: 228–239.

Briggs D.J., Collins S., Elliott P., et al. Mapping urban air pollution using GIS: a regression-based approach. *Int J Geogr Inform Sci* 1997: 11: 699–718.

Briggs D.J., de Hoogh C., Guiliver J., et al. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. *Sci Total Environ* 2000: 253: 151–167.

Brunekreef B., and Holgate S.T. Air pollution and health. *Lancet* 2002: 360: 1233–1242.

English P., Neutra R., Scalf R., et al. Examining associations between childhood asthma and traffic flow using a geographic information system. *Environ Health Perspect* 1999: 107: 761–767.

Finkelstein M.M., Jerrett M., DeLuca P., et al. Relation between income, air pollution and mortality: a cohort study. *Can Med Assoc J* 2003: 169: 397–402.

Finkelstein M.M., Jerrett M., and Sears M.R. Traffic air pollution and mortality rate advancement periods. *Am J Epidemiol* 2004: 160: 173–177.

Hoek G., Brunekreef B., Goldbohm S., et al. Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *Lancet* 2002: 360: 1203–1209.

Jerrett M., Arain A., and Kanaroglou P. A review and evaluation of intra-urban air pollution exposure models. *J Exposure Anal Environ Epidemiol* 2005: 15: 185–204.

Jerrett M., Arain A., Kanaroglou P., et al. Modelling the intra-urban variability of ambient traffic pollution in Toronto, Canada, in press.

Krewski D., Burnett R.T., Goldberg D.E., et al. *Reanalysis of the Harvard Six-Cities Study and the American Cancer Society of Air Pollution and Mortality, Phase II: Sensitivity Analysis*. Health Effects Institute, Cambridge, MA, 2000.

National Research Council. *Estimating Public Health Benefits of Proposed Air Pollution Regulations*. National Academy Press, Washington, DC, 2002.

Palmes E., Gunnison A., Dimattio J, et al. Personal sampling for nitrogen dioxide. *Am Ind Hyg Assoc J* 1976: 37: 570–577.

Pope C.A., Burnett R.T., Thun M.J., et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 2002: 287: 1132–1141.

Rijinders E., Janssen N., Van Vliet P., et al. Personal and outdoor concentrations in relation to degree of urbanization and traffic density. *Environ Health Perspect* 2001: 109: 411–417.

SANDAG. SANDAG GIS Metadata: Current Landuse, 2004: http://www.sandag.cog.ca.us/resources/maps_and_gis/gis_downloads/downloads/metadata/ludoc.htm. Accessed 2004.

United States Environmental Protection Agency (USEPA). National-Scale Air Toxics Assessment for 1996. Estimated Concentrations by Census Tract. Cumulative Exposure Project, 2001 Accessed.

Wilkinson P., Elliott P., Grundy C., et al. Case–control study of hospital admission with asthma in children aged 5–14 years: relation with road traffic in north west London. *Thorax* 1999: 54: 1070–1074.