

ESTIMATING DAILY NITROGEN DIOXIDE LEVEL: EXPLORING TRAFFIC EFFECTS

BY LIXUN ZHANG¹, YONGTAO GUAN², BRIAN P. LEADERER¹ AND THEODORE R. HOLFORD¹

Yale University

Data used to assess acute health effects from air pollution typically have good temporal but poor spatial resolution or the opposite. A modified longitudinal model was developed that sought to improve resolution in both domains by bringing together data from three sources to estimate daily levels of nitrogen dioxide (NO₂) at a geographic location. Monthly NO₂ measurements at 316 sites were made available by the Study of Traffic, Air quality and Respiratory health (STAR). Four US Environmental Protection Agency monitoring stations have hourly measurements of NO₂. Finally, the Connecticut Department of Transportation provides data on traffic density on major roadways, a primary contributor to NO₂ pollution. Inclusion of a traffic variable improved performance of the model, and it provides a method for estimating exposure at points that do not have direct measurements of the outcome. This approach can be used to estimate daily variation in levels of NO₂ over a region.

1. Introduction. The relationship between traffic and air pollutants such as NO₂ has been examined using many different approaches [e.g., [Maantay \(2007\)](#), [McConnell et al. \(2010\)](#)]. Proximity to traffic has frequently been used as a proxy for traffic related air pollution exposure in environmental health [[Jerrett et al. \(2005\)](#), [McConnell et al. \(2006\)](#)]. In such studies, the goal is to determine whether there is a relationship between air pollution and health outcomes. When direct measurements of specific pollutant levels are not available, proximity to roadways and traffic levels are sometimes used as proxies. In general, NO₂ levels decline with distance from a highway [[Cape et al. \(2004\)](#), [Fрати et al. \(2006\)](#), [Gilbert et al. \(2003\)](#), [Rodes and Holland \(1981\)](#)].

While data on proximity to major roads have proven to be a cost-effective approach in epidemiological studies of traffic exposure, they do not necessarily account for traffic volume. Inclusion of volume further improves the quality of traffic exposure measurement [[Rose et al. \(2009\)](#)]. For instance, [Gauvin et al. \(2001\)](#) found that including an index of traffic intensity and proximity in a model, along

Received March 2012; revised February 2013.

¹Supported in part by NIH Grants ES017416 and ES005410.

²Supported in part by NSF Grant DMS-08-45368.

Key words and phrases. Bayesian model, longitudinal model, nitrogen dioxide, EPA, air pollution.

with an indicator of gas cooker use in the home, improved the correlation between model estimates and levels of nitrogen dioxide measured from a monitor located close to a child's home or school. Other studies [e.g., Brauer et al. (2003), Carr et al. (2002), Cesaroni et al. (2008), Heinrich et al. (2005), Ryan et al. (2005), Schikowski et al. (2005), Venn et al. (2000)] also used traffic volume to improve the quality of exposure information.

One way to include traffic volume information in a model is to introduce vehicular counts within a buffer zone, which Rose et al. (2009) call weighted-road-density. The idea is to calculate the total (road length \times traffic volume) for a given circle and divide it by the area, that is, $\frac{\sum_{i=1}^n L_i V_i}{\pi r^2}$, where L_i is the length of a segment, V_i the traffic volume and r the radius of the circle. Either actual traffic counts or a road classification system can be used for V_i . The authors found that actual traffic counts were better at predicting NO₂ than a simple hierarchical classification of roads. In addition, weighted road density was found to be a better predictor than proximity to a major road.

Rose et al.'s (2009) method assumed that all roads within a circle had the same effect regardless of distance to the point of interest. Holford et al. (2010) proposed a method that made use of road density, traffic volume and distance to roads from points of interest. They were able to estimate a dispersion function for a pollutant, which improved estimates of NO₂ over those obtained using only average daily traffic (ADT: number of vehicles/day) on the closest highway, ADT on the busiest highway within a buffer and the sum for all road segments within a buffer.

The underlying framework for the methods reviewed above is land use regression which uses traffic-related variables as predictors for NO₂ [e.g., Briggs et al. (1997), Gilbert et al. (2005), Gonzales et al. (2005), Jerrett et al. (2007), Rosenlund et al. (2008), Ross et al. (2006), Wheeler et al. (2008)]. Ibarra-Berastegi et al. (2003) added a time-varying component to a model using multiple linear regression to forecast NO₂ levels up to 8 hours in advance by using current and past 15 hours meteorology along with traffic information.

Further methods for assessing intraurban exposure were reviewed by Jerrett et al. (2005): (i) statistical interpolation [Jerrett et al. (2001)], (ii) line dispersion models [Bellander et al. (2001)], (iii) integrated emission-meteorological models [Frohn, Christensen and Brandt (2002)], and (iv) hybrid models combining personal or household exposure monitoring with one of the preceding methods [Kramer et al. (2000), Zmirou et al. (2002)], or combining two or more of the preceding methods with regional monitoring [Hoek et al. (2001)]. Rose et al. (2009) broke down the alternatives into just two categories: dispersion-based models and empirical models.

As pointed out by Jerrett et al. (2005), a disadvantage of geostatistical interpolation is the limited availability of monitoring data. This approach requires a reasonably dense network of sampling sites. Government monitoring data generally come from a sparse network of stations, giving rise to systematic errors in estimates at

sites far from the monitoring stations. Increasing the number of monitoring sites can be helpful but costly, so it has not been used extensively. Researchers often have to use pollution measurements over relatively short time periods as a substitute for the comparatively long periods covered by health histories. This poses a choice between relying on a government network that provides temporal detail for a limited number of sites or on their own more detailed spatial network, which usually covers a short period of time.

To address the limitations inherent in each source of available data, Zhang (2011) applied a longitudinal model that established a relationship between data from US Environmental Protection Agency (EPA) monitoring sites with daily or finer temporal resolution and those from the Study of Traffic, Air quality and Respiratory health in children (STAR) with monthly resolution. It was assumed that the relationship at the monthly level held at the daily level, using a model in which data from EPA sites were used to estimate pollution information at study sites. This model performed well as measured by R^2 in a simple linear model that used STAR site observations as the response variable and the predictions based on EPA measurements as the predictor variable. The model showed that about 73% of the variability at the STAR sites can be explained by the predictions. This article extends and seeks to improve Zhang's (2011) method by including traffic as predictors in the model. A traffic-related variable can then be used to explain the spatial variation observed in the random intercept of the longitudinal model, thus providing a practical way for estimating the temporal/spatial distribution of NO_2 in a region.

2. Methods.

2.1. *EPA and STAR data.* STAR is an epidemiological study of childhood asthma designed to investigate whether common air contaminants are related to disease severity. Four monthly outdoor NO_2 measurements were taken for each subject, with three months separating each consecutive measurement. Observations used in this analysis were taken between April 25, 2006 and March 21, 2008. In contrast to the STAR study, the EPA monitoring sites provide hourly NO_2 measurements. Average daily NO_2 was calculated from these hourly measurements. Figure 1 shows the locations of four EPA sites in Connecticut and 316 STAR study sites used in this analysis. We selected randomly 266 STAR learning sites for model development and the remaining 50 sites were used for model validation.

Inverse distance weighting (IDW) was used to interpolate daily NO_2 values at STAR sites based on daily averages at the four EPA sites. Let $Z_{i,j}$ denote the j th NO_2 measurement at STAR site i (between days t_1 to t_2 , say), and let $V_{i,t}$ denote the IDW interpolated NO_2 value at site i on day t , for $i = 1, 2, \dots, n$, and $t = 1, 2, \dots, T$. A new variable $U_{i,j}$ can be created by taking the average of $V_{i,t}$ for site i over the same period as $Z_{i,j}$. Figure 2 plots $Z_{i,j}$ against $U_{i,j}$ for the 316 sites in Figure 1, where weights are the reciprocal of distance.

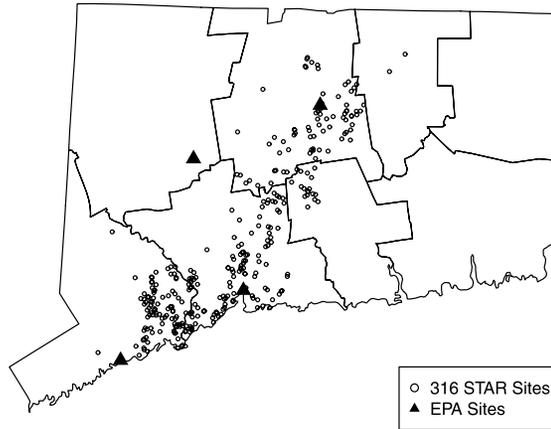


FIG. 1. Locations of 4 EPA sites which have hourly NO_2 measurements and 316 STAR sites which have monthly measurements.

2.2. *Traffic data.* The Connecticut Department of Transportation reports ADT for all state roads on a three-year cycle. The data for 2006 were used in this analysis. Figure 3 shows these road segments which have reported ADT. There are 5196 road segments, with lengths ranging from 16 meters to 12,295 meters, median of

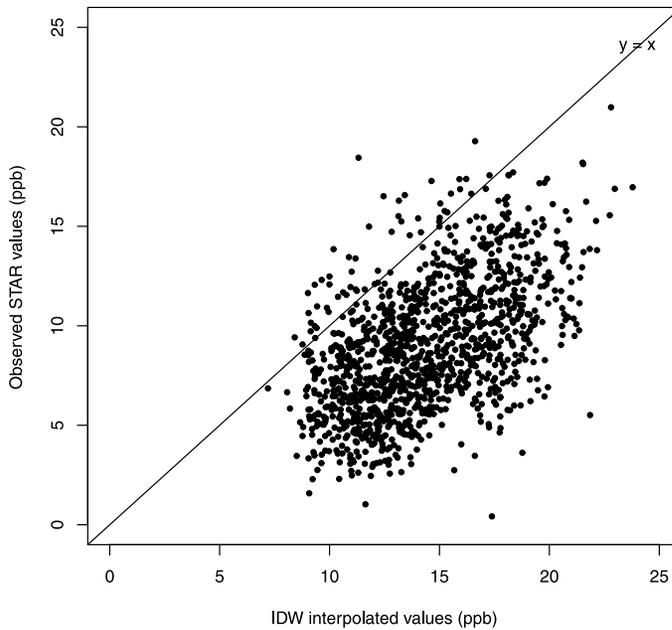


FIG. 2. Observed NO_2 values at 316 STAR sites ($Z_{i,j}$) vs average of IDW interpolated values from EPA NO_2 measurements over the same period as $Z_{i,j}$.

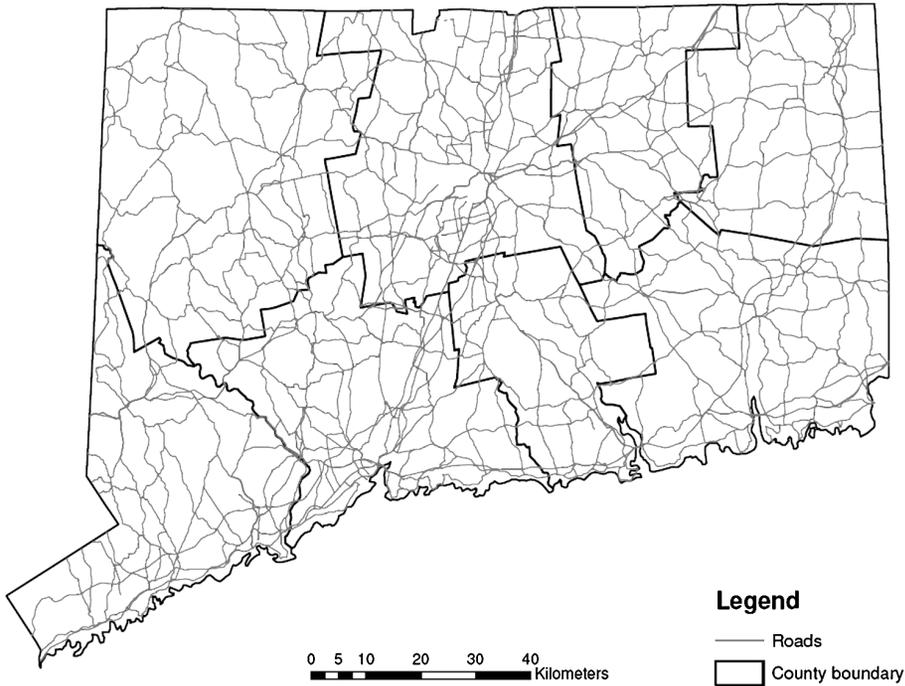


FIG. 3. Roads with traffic information in Connecticut.

740 meters and mean of 1207 meters. The range for ADT was 0 to 184,000 (mean of 22,323 and median of 11,400).

2.3. *Models.* Three models were compared in this study. First, we considered a linear model:

$$(2.1) \quad Y_i = \alpha_0 + \alpha_1 \times x_i + \sum_k \gamma_k W_{i,k} + \varepsilon_i,$$

where Y_i denotes the i th NO_2 measurement on the natural log scale, x_i is the natural log of the average IDW interpolated NO_2 for that site over the corresponding period, $W_{i,k}$ is the traffic information (ADT), and $\varepsilon_i \sim N(0, \sigma^2)$ is some random error, for $i = 1, 2, \dots, 1064$.

Second, we specified a longitudinal model with random effects for sites:

$$(2.2) \quad Y_{i,j} = \beta_0 + b_{0,i} + \beta_1 \times x_{i,j} + \sum_k \gamma_k W_{i,k} + \varepsilon_{i,j},$$

where $Y_{i,j}$ denotes the j th NO_2 measurement at STAR site i on the natural log scale, $x_{i,j}$ is the corresponding average of IDW interpolated NO_2 on the natural log scale, $W_{i,k}$ is the traffic information, $b_{0,i} \sim N(0, \sigma_b^2)$ is a random intercept for site i , and $\varepsilon_{i,j} \sim N(0, \sigma_\varepsilon^2)$ is some random error, for $i = 1, 2, \dots, 266$ and

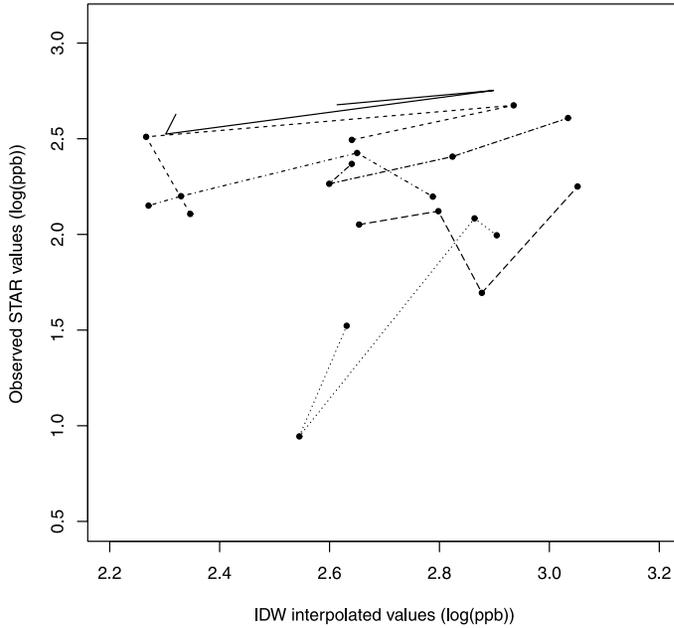


FIG. 4. Observed NO₂ levels vs averages of IDW interpolated NO₂ levels for six randomly selected STAR sites, with lines connecting values in temporal order.

$j = 1, 2, 3, 4$. The random effects $b_{0,i}$ and $\varepsilon_{i,j}$ are mutually independent. A scatter plot showing this relationship for these data is shown in Figure 4, which shows $Z_{i,j}$ (the j th NO₂ measurement at site i) against $U_{i,j}$ (average of IDW interpolated daily NO₂ values at site i over the period corresponding to $Z_{i,j}$) for six randomly selected sites, with lines connecting values for a site in temporal order.

Finally, we specified a modified longitudinal model which allowed for spatial correlation among site effects for the model in equation (2.2), that is, $\mathbf{b}_0 = (b_{0,1}, b_{0,2}, \dots, b_{0,n})^T \sim N(\mathbf{0}, \sigma_b^2 \times \Sigma(\phi))$. Elements in the covariance matrix $\Sigma(\phi)$ are given by $\exp(-\frac{d}{\phi})$, where d denotes spatial distance. The random effects \mathbf{b}_0 and $\varepsilon_{i,j}$'s are mutually independent.

We adjusted for traffic effects using the integrated exposure model proposed by Holford et al. (2010) which introduced covariates into the linear predictor in a regression model. The contribution of traffic was expressed as

$$\int z(s)\phi(s) ds,$$

where $z(s)$ denotes ADT for point s on a line representing a highway and $\phi(s)$ is a dispersion function for the pollutant generated at s . We can achieve computational efficiency with little loss in accuracy by representing this contribution numerically—taking the sum of the product of ADT, the segment length and the

unknown dispersion function which depends on distance. Holford et al. (2010) discussed alternative forms of linear dispersion functions, for example, stepped, polynomial or spline. In this example we used a step function, in which we estimated a value for the level of dispersion between specified distance intervals, D_{k-1} and D_k : $\sum_j z_{k,j} \gamma_k \delta_{k,j} = \gamma_k \sum_j z_{k,j} \delta_{k,j}$, where γ_k is the pollution effect from a unit intensity source within the interval, $z_{k,j}$ is ADT, and $\delta_{k,j}$ is length of the segment. The linear predictor related to traffic effects can now be written as

$$\int z(s)\phi(s) ds = \sum_k \left(\gamma_k \sum_j z_{k,j} \delta_{k,j} \right) = \sum_k \gamma_k W_k,$$

where $W_k = \sum_j z_{k,j} \delta_{k,j}$.

ADT is reported in highly variable lengths, and while this approach might work well for short segments, it can become problematic for long segments, for example, if the center of one road is close to a site but most of the remaining segments are relatively far away. To mitigate this problem, we divided the segments into smaller subsegments and found that 50-meter segments provided an adequate accuracy. To show this, we tested lengths such as 10-meter, 50-meter, 100-meter and up to 5000-meter and found little difference in the resulting estimates between 10 and 50 meters. For this example, we used 50-meter. Segments were divided into subsegments using a Python (<http://www.python.org/>) script which calls relevant ArcGIS [Environmental Systems Resource Institute (2010)] functions.

Values of D_k 's were predetermined by our experience with earlier analysis. Setting the values of D_k beforehand leaves the values of γ_k 's to be estimated as regression parameters. Two possible approaches for incorporating traffic effects were examined: a single-step model which sets the contribution of highway segments within 2000 meters as equal and for distances farther than 2000 meters as 0; and a multi-step model with steps at 400 meters, 800 meters, 1200 meters, 1600 meters and 2000 meters.

While models (2.1) and (2.2) were fitted using a frequentist approach, we obtained parameter estimates for the third model under the Bayesian framework.

The three models were fitted to NO_2 levels at the 266 learning sites and the results were used to estimate levels not only at these sites but at the 50 validation sites as well. By assuming that the relationship at the monthly level also holds at the daily level, we also obtained daily estimates. One predictor variable was based on daily pollution levels obtained by interpolating with IDW measurements from the four EPA sites. We also included the remaining predictors representing traffic-related effects $W_{i,k}$.

Once daily NO_2 predictions at the sites were obtained, they were averaged over the same periods as the STAR observations. Systematic departures for site estimates were evaluated using simple linear regression:

$$(2.3) \quad Z_{i,j} = \alpha_0 + \alpha_1 * P_{ij} + \varepsilon_{i,j},$$

where $Z_{i,j}$ is the j th observation at STAR site i , $P_{i,j}$ is the average of the estimated daily NO_2 values at site i over the same period as $Z_{i,j}$, and $\varepsilon_{i,j} \sim N(0, \sigma^2)$. In addition, we calculated the root mean square error (RMSE):

$$\sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^4 (Z_{i,j} - P_{i,j})^2}{4n}}$$

3. Results. Table 1 shows results from fitting the model in equation (2.1) using the single-step and multi-step dispersion models for the traffic effect. Table 2 shows results from fitting the corresponding longitudinal model in equation (2.2). In Table 1, the results from the multi-step dispersion model reveal that the effects of the first two steps (0–400 m and 400–800 m) are not significantly different from zero at the 0.05 significance level. While parameter estimates of the next three steps (800–1200 m, 1200–1600 m and 1600–2000 m) are significantly different from zero, their values are nearly the same (0.0622, 0.0675 and 0.0495). Similar observations can be made on the results from the longitudinal model in Table 2. While one might expect values to decline with distance, this could be due to the high correlation among traffic covariates for the five steps. The variance inflation factor (VIF) for each traffic variable in model (2.1) was above one and the VIFs for two of them were above three. While multi-collinearity does not greatly affect prediction severely in general, it can be difficult to diagnose the potential issue of extrapolation with multiple predictors when making a prediction at a new site. Moreover, note from Table 1 that the adjusted R^2 only improved marginally with the use of multi-step variables. For these reasons we focused on the model using the single-step traffic variable.

The single-step dispersion function was also used for the modified longitudinal model and the results are shown in Table 3. Table 4 summarizes results from a comparison of the fitted and the observed levels at the 50 validation sites using

TABLE 1
Results from fitting the linear model in (2.1) with different traffic variables

Traffic		Estimate	SE	<i>t</i> -value	<i>p</i> -value	Adjusted R^2
Single-step	α_0	-0.3728	0.1181	-3.1570	0.0016	0.3857
	α_1	0.9428	0.0447	21.0930	<0.0001	
	γ	0.1524	0.0098	15.5110	<0.0001	
Multi-step	α_0	-0.3963	0.1184	-3.3470	0.0008	0.3911
	α_1	0.9341	0.0446	20.9230	<0.0001	
	γ_1	-0.0133	0.0283	-0.4710	0.6378	
	γ_2	0.0062	0.0236	0.2630	0.7926	
	γ_3	0.0622	0.0233	2.6660	0.0078	
	γ_4	0.0675	0.0151	4.4810	<0.0001	
	γ_5	0.0495	0.0099	4.9900	<0.0001	

TABLE 2
Results from fitting the longitudinal model in (2.2) with different traffic variables

Traffic		Estimate	SE	DF	t-value	p-value
Single-step	β_0	-0.5974	0.1033	797	-5.7826	<0.0001
	β_1	1.0281	0.0389	797	26.4628	<0.0001
	γ	0.1529	0.0146	264	10.5075	<0.0001
	σ_b^2	0.0402				
	σ_Y^2	0.0619				
Multi-step	β_0	-0.6344	0.1053	797	-6.0222	<0.0001
	β_1	1.0250	0.0389	797	26.3591	<0.0001
	γ_1	-0.0117	0.0419	260	-0.2797	0.7800
	γ_2	0.0070	0.0350	260	0.1985	0.8428
	γ_3	0.0627	0.0346	260	1.8149	0.0707
	γ_4	0.0653	0.0223	260	2.9300	0.0037
	γ_5	0.0503	0.0147	260	3.4294	0.0007
	σ_b^2	0.0398				
	σ_Y^2	0.0619				

the model in equation (2.3). Also included are a comparison of results for models with and without the traffic variable. Including the traffic variable improved performance of both the linear and the longitudinal models. For instance, the predictive R^2 for model (2.3) changed from 0.2617 to 0.4375 and RMSE from 2.9527 to 2.5763 after including traffic variable in the longitudinal model. The additive bias α_0 in the longitudinal model changed from 1.0821 (p -value 0.283) to 1.2584 (p -value 0.0637).

For the modified longitudinal model that included spatial correlation, the estimated α_0 was not significantly different from zero, thus being similar to the estimates from the model without the traffic variable. However, when the traffic variable was included in this model, the predictive R^2 was 0.6106, which was

TABLE 3
Results from fitting the modified longitudinal model that includes spatial correlation in (2.2) with a single-step traffic variable

	Mean	SE	2.50%	50%	97.50%
β_0	-0.8524	0.0896	-0.9838	-0.8748	-0.6251
β_1	1.0828	0.0312	1.0068	1.0828	1.1365
γ	0.1023	0.0153	0.0725	0.1023	0.1333
σ_b^2	0.0748	0.0203	0.0419	0.0722	0.1207
σ_Y^2	0.0648	0.0033	0.0588	0.0647	0.0716
ϕ	12.3184	3.6682	6.5307	12.2449	19.5918

TABLE 4
Results from a comparison of predicted and observed values for the 50 validation sites

		Estimate	SE	<i>t</i> -value	<i>p</i> -value	Predictive R^2	RMSE	Traffic
Linear model	α_0	0.1163	1.1220	0.104	0.9180	0.2605	2.9687	N
	α_1	1.0526	0.1260	8.352	<0.0001			
	α_0	0.8978	0.7073	1.269	0.206	0.4342	2.5843	Y
	α_1	0.9468	0.0768	12.327	<0.0001			
Longitudinal model	α_0	1.0821	1.0057	1.076	0.2830	0.2617	2.9527	N
	α_1	0.9333	0.1114	8.377	<0.0001			
	α_0	1.2584	0.6748	1.865	0.0637	0.4375	2.5763	Y
	α_1	0.8998	0.0725	12.409	<0.0001			
Modified longitudinal model	α_0	0.5247	0.5539	0.947	0.3450	0.5807	2.2081	N
	α_1	0.9703	0.0586	16.560	<0.0001			
	α_0	0.6802	0.5131	1.326	0.1860	0.6106	2.1311	Y
	α_1	0.9527	0.0541	17.622	<0.0001			

slightly higher than 0.5807 for the model without traffic. Comparing RMSEs led to similar conclusions, that is, the model that included traffic had a lower RMSE compared with the model without traffic. Figure 5 shows a scatter plot of observed vs predicted NO₂ from the modified longitudinal model with traffic effects.

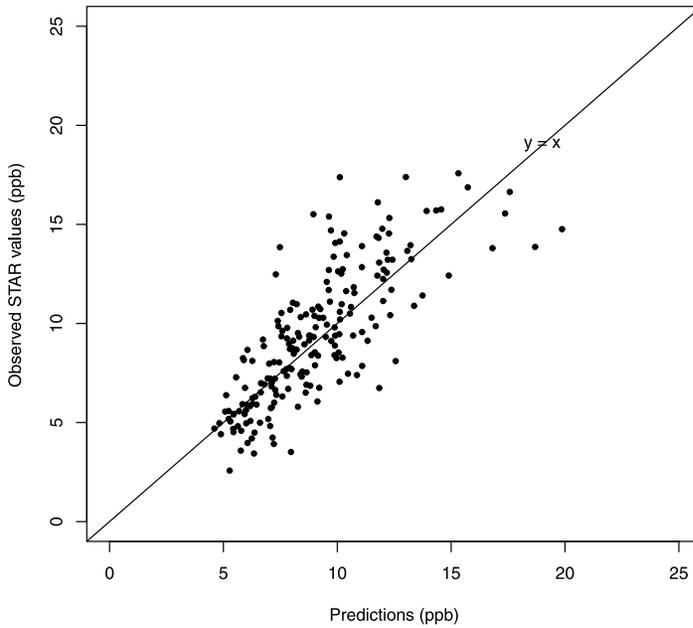


FIG. 5. *Observed vs predicted NO₂ values at 50 validation STAR sites.*

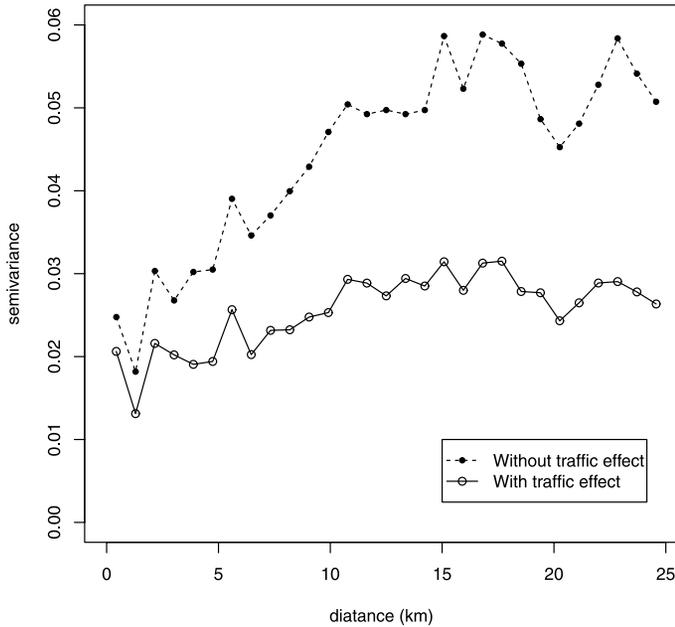


FIG. 6. Semivariograms of the random intercept in the longitudinal model, before and after including the predictor for traffic effect.

To see whether traffic effects explain the spatial correlation in the random intercepts of the longitudinal model, we compared the sample semivariograms for two versions of the longitudinal model (2.2), one with traffic and the other without (Figure 6). We can see that the semivariogram after accounting for traffic is almost flat compared with the one without traffic. This suggests that the spatial correlation in the random intercept has been partially explained by the inclusion of traffic in the model.

4. Discussion. Based on the estimated α_0 , predictive R^2 and RMSE for the 50 validation sites, we concluded that inclusion of traffic effects improved the linear, the longitudinal and the modified longitudinal models. In addition, the modified longitudinal model worked reasonably well for making predictions at random sites.

In the modified longitudinal model, no temporal correlation structure was assumed for the residual $\varepsilon_{i,j}$. An area for future research would be to develop a model that allows for both spatial and temporal correlation. Brown et al. (2001) and Romanowicz et al. (2006) demonstrated how such models could be estimated. From an application perspective, however, assuming only spatial correlation has the advantage of being less computationally demanding. One would need to weigh the benefits and costs of using a more complex model that includes a spatiotemporal correlation structure.

Another area for further research is to allow for additional predictors such as land use, population density and elevation similar to that used by Skene et al. (2010). In addition, one needs to explore whether these models can be applied to different temporal resolutions. The EPA sites record NO₂ levels on an hourly basis, so if the level of pollutant varies with time of day as a subject moves from place to place, this could have relevant health consequences.

It would also be useful to determine whether the proposed model can be applied to other pollutants generated by traffic. The US EPA monitors a variety of relevant pollutants, including carbon monoxide, ozone, particulate matter 2.5 and sulfur dioxide. Epidemiological studies have been carried out to explore the relationship between exposure to these pollutants and health [e.g., Bell and Dominici (2006), Islam et al. (2008), Son, Bell and Lee (2011)]. If this approach also performs well for these pollutants, one would be able to study the effect of daily pollution levels on health.

Finally, it would be interesting to develop alternative models for estimating the daily pollution levels at multiple sites, for example, similar to the latent spatial process used by Smith, Zhang and Field (2007). As a result, it would be no longer necessary to assume that the relationship between monthly EPA measures and STAR sites would hold at the daily level. However, implementation of such models would be computationally expensive, which could pose a significant challenge for potential users.

Acknowledgments. The authors thank an Associate Editor for the many constructive comments that greatly enhanced our paper. The authors appreciate the computing services provided by Yale University Biomedical High Performance Computing Center, which was funded by NIH Grant RR19895. The authors also thank Dr. Janneane Gent for her valuable input about the Study of Traffic, Air Quality and Respiratory Health in Children.

REFERENCES

- BELL, M. and DOMINICI, F. (2006). Analysis of threshold effects for short-term exposure to ozone and increased risk of mortality. *Epidemiology* **17** S223–S223.
- BELLANDER, T., BERGLIND, N., GUSTAVSSON, P., JONSON, T., NYBERG, F., PERSHAGEN, G. and JARUP, L. (2001). Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. *Environ. Health Perspect.* **109** 633–639.
- BRAUER, M., HOEK, G., VAN VLIET, P., MELIEFSTE, K., FISCHER, P., GEHRING, U., HEINRICH, J., CYRYS, J., BELLANDER, T., LEWNE, M. and BRUNEKREEF, B. (2003). Estimating long-term average particulate air pollution concentrations: Application of traffic indicators and geographic information systems. *Epidemiology* **14** 228–239.
- BRIIGGS, D. J., COLLINS, S., ELLIOTT, P., FISCHER, P., KINGHAM, S., LEBRET, E., PRYL, K., VANREEUWIJK, H., SMALLBONE, K. and VANDERVEEN, A. (1997). Mapping urban air pollution using GIS: A regression-based approach. *International Journal of Geographical Information Science* **11** 699–718.

- BROWN, P. E., DIGGLE, P. J., LORD, M. E. and YOUNG, P. C. (2001). Space-time calibration of radar rainfall data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **50** 221–241. MR1833274
- CAPE, J. N., TANG, Y. S., VAN DIJK, N., LOVE, L., SUTTON, M. A. and PALMER, S. C. F. (2004). Concentrations of ammonia and nitrogen dioxide at roadside verges, and their contribution to nitrogen deposition. *Environmental Pollution* **132** 469–478.
- CARR, D., VON EHRENSTEIN, O., WEILAND, S., WAGNER, C., WELLIE, O., NICOLAI, T. and VON MUTIUS, E. (2002). Modeling annual benzene, toluene, NO₂, and soot concentrations on the basis of road traffic characteristics. *Environ. Res.* **90** 111–118.
- CESARONI, G., BADALONI, C., PORTA, D., FORASTIERE, F. and PERUCCI, C. A. (2008). Comparison between various indices of exposure to traffic-related air pollution and their impact on respiratory health in adults. *Occup. Environ. Med.* **65** 683–690.
- ENVIRONMENTAL SYSTEMS RESOURCE INSTITUTE (2010). ArcMap 10.0. ESRI, Redlands, CA.
- FRATI, L., CAPRASECCA, E., SANTONI, S., GAGGI, C., GUTOVA, A., GAUDINO, S., PATI, A., ROSAMILIA, S., PIRINTOS, S. A. and LOPPI, S. (2006). Effects of NO₂ and NH₃ from road traffic on epiphytic lichens. *Environ. Pollut.* **142** 58–64.
- FROHN, L. M., CHRISTENSEN, J. H. and BRANDT, J. (2002). Development of a high-resolution nested air pollution model—The numerical approach. *J. Comput. Phys.* **179** 68–94.
- GAUVIN, S., MOULLEC, Y. L., BREMONT, F., MOMAS, I., BALDUCCI, F., CIOGNARD, F., POILVE, M. P., ZMIROU, D. and VESTA INVESTIGATORS (2001). Relationships between nitrogen dioxide personal exposure and ambient air monitoring measurements among children in three French metropolitan areas: VESTA study. *Arch. Environ. Health* **56** 336–341.
- GILBERT, N. L., WOODHOUSE, S., STIEB, D. M. and BROOK, J. R. (2003). Ambient nitrogen dioxide and distance from a major highway. *Sci. Total Environ.* **312** 43–46.
- GILBERT, N. L., GOLDBERG, M. S., BECKERMAN, B., BROOK, J. R. and JERRETT, M. (2005). Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. *Journal of the Air Waste Management Association* **55** 1059–1063.
- GONZALES, M., QUALLS, C., HUDGENS, E. and NEAS, L. (2005). Characterization of a spatial gradient of nitrogen dioxide across a United States-Mexico border city during winter. *Science of the Total Environment* **337** 163–173.
- HEINRICH, J., GEHRING, U., CYRYS, J., BRAUER, M., HOEK, G., FISCHER, P., BELLANDER, T. and BRUNEKREEF, B. (2005). Exposure to traffic related air pollutants: Self reported traffic intensity versus GIS modelled exposure. *Occup. Environ. Med.* **62** 517–523.
- HOEK, G., FISCHER, P., BRANDT, P. V. D., GOLDBOHM, S. and BRUNEKREEF, B. (2001). Estimation of long-term average exposure to outdoor air pollution for a cohort study on mortality. *J. Expo. Anal. Environ. Epidemiol.* **11** 459–469.
- HOLFORD, T. R., EBISU, K., MCKAY, L. A., GENT, J. F., TRICHE, E. W., BRACKEN, M. B. and LEADERER, B. P. (2010). Integrated exposure modeling: A model using GIS and GLM. *Stat. Med.* **29** 116–129. MR2751384
- IBARRA-BERASTEGI, G., MADARIAGA, I., AGIRRE, E. and URIA, J. (2003). Short-term forecasting of ozone and NO₂ levels using traffic data in Bilbao (Spain). *Urban Transport IX: Urban Transport and the Environment in the 21st Century* **14** 235–242, 709.
- ISLAM, T., MCCONNELL, R., GAUDERMAN, W. J., AVOL, E., PETERS, J. M. and GILLILAND, F. D. (2008). Ozone, oxidant defense genes, and risk of asthma during adolescence. *Am. J. Respir. Crit. Care Med.* **177** 388–395.
- JERRETT, M., BURNETT, R. T., KANAROGLOU, P., EYLES, J., FINKELSTEIN, N., GIOVIS, C. and BROOK, J. R. (2001). A GIS—Environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A* **33** 955–973.
- JERRETT, M., ARAIN, A., KANAROGLOU, P., BECKERMAN, B., POTOGLOU, D., SAHSU-VAROGLU, T., MORRISON, J. and GIOVIS, C. (2005). A review and evaluation of intraurban air pollution exposure models. *J. Expo. Anal. Environ. Epidemiol.* **15** 185–204.

- JERRETT, M., ARAIN, M. A., KANAROGLOU, P., BECKERMAN, B., CROUSE, D., GILBERT, N. L., BROOK, J. R., FINKELSTEIN, N. and FINKELSTEIN, M. M. (2007). Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *Journal of Toxicology and Environmental Health-Part A-Current Issues* **70** 200–212.
- KRAMER, U., KOCH, T., RANFT, U., RING, J. and BEHRENDT, H. (2000). Traffic-related air pollution is associated with atopy in children living in urban areas. *Epidemiology* **11** 64–70.
- MAANTAY, J. (2007). Asthma and air pollution in the Bronx: Methodological and data considerations in using GIS for environmental justice and health research. *Health Place* **13** 32–56.
- MCCONNELL, R., BERHANE, K., YAO, L., JERRETT, M., LURMANN, F., GILLILAND, F., KUNZLI, N., GAUDERMAN, J., AVOL, E., THOMAS, D. and PETERS, J. (2006). Traffic, susceptibility, and childhood asthma. *Environmental Health Perspectives* **114** 766–772.
- MCCONNELL, R., ISLAM, T., SHANKARDASS, K., JERRETT, M., LURMANN, F., GILLILAND, F., GAUDERMAN, J., AVOL, E., KUNZLI, N., YAO, L., PETERS, J. and BERHANE, K. (2010). Childhood incident asthma and traffic-related air pollution at home and school. *Environ. Health Perspect.* **118** 1021–1026.
- RODES, C. E. and HOLLAND, D. M. (1981). Variations of No, No₂ and O₃ concentrations downwind of a Los-Angeles freeway. *Atmospheric Environment* **15** 243–250.
- ROMANOWICZ, R., YOUNG, P., BROWN, P. and DIGGLE, P. (2006). A recursive estimation approach to the spatio-temporal analysis and modelling of air quality data. *Environmental Modelling Software* **21** 759–769.
- ROSE, N., COWIE, C., GILLETT, R. and MARKS, G. B. (2009). Weighted road density: A simple way of assigning traffic-related air pollution exposure. *Atmospheric Environment* **43** 5009–5014.
- ROSENLUND, M., FORASTIERE, F., STAFOGGIA, M., PORTA, D., PERUCCI, M., RANZI, A., NUS-SIO, F. and PERUCCI, C. A. (2008). Comparison of regression models with land-use and emissions data to predict the spatial distribution of traffic-related air pollution in Rome. *J. Expo. Sci. Environ. Epidemiol.* **18** 192–199.
- ROSS, Z., ENGLISH, P. B., SCALF, R., GUNIER, R., SMORODINSKY, S., WALL, S. and JERRETT, M. (2006). Nitrogen dioxide prediction in Southern California using land use regression modeling: Potential for environmental health analyses. *J. Expo. Sci. Environ. Epidemiol.* **16** 106–114.
- RYAN, P. H., LEMASTERS, G., BIAGINI, J., BERNSTEIN, D., GRINSHPUN, S. A., SHUKLA, R., WILSON, K., VILLAREAL, M., BURKLE, J. and LOCKEY, J. (2005). Is it traffic type, volume, or distance? Wheezing in infants living near truck and bus traffic. *Journal of Allergy and Clinical Immunology* **116** 279–284.
- SCHIKOWSKI, T., SUGIRI, D., RANFT, U., GEHRING, U., HEINRICH, J., WICHMANN, H. E. and KRAMER, U. (2005). Long-term air pollution exposure and living close to busy roads are associated with COPD in women. *Respiratory Research* **6** 152.
- SKENE, K. J., GENT, J. F., MCKAY, L. A., BELANGER, K., LEADERER, B. P. and HOLFORD, T. R. (2010). Modeling effects of traffic and landscape characteristics on ambient nitrogen dioxide levels in Connecticut. *Atmos. Environ.* (1994) **44** 5156–5164.
- SMITH, B. J., ZHANG, L. and FIELD, R. W. (2007). Iowa radon leukaemia study: A hierarchical population risk model for spatially correlated exposure measured with error. *Stat. Med.* **26** 4619–4642. MR2411891
- SON, J., BELL, M. and LEE, J. T. (2011). Survival analysis to estimate the association between long-term exposure to different sizes of airborne particulate matter and risk of infant mortality using a birth cohort in Seoul, Korea. *Epidemiology* **22** S166–S167.
- VENN, A., LEWIS, S., COOPER, M., HUBBARD, R., HILL, I., BODDY, R., BELL, M. and BRITTON, J. (2000). Local road traffic activity and the prevalence, severity, and persistence of wheeze in school children: Combined cross sectional and longitudinal study. *Occup. Environ. Med.* **57** 152–158.

WHEELER, A. J., SMITH-DOIRON, M., XU, X., GILBERT, N. L. and BROOK, J. R. (2008). Intra-urban variability of air pollution in Windsor, Ontario—Measurement and modeling for human exposure assessment. *Environ. Res.* **106** 7–16.

ZHANG, L. (2011). A Bayesian spatio-temporal model for estimating daily nitrogen dioxide levels. Ph.D. thesis, Yale Univ., New Haven, CT.

ZMIROU, D., GAUVIN, S., PIN, I., MOMAS, I., JUST, J., SAHRAOUI, F., MOULLEC, Y. L., BRÉ-MONT, F., CASSADOU, S., ALBERTINI, M., LAUVERGNE, N., CHIRON, M., LABBÉ, A. and VESTA INVESTIGATORS (2002). Five epidemiological studies on transport and asthma: Objectives, design and descriptive results. *J. Expo. Anal. Environ. Epidemiol.* **12** 186–196.

L. ZHANG
STATE STREET CORPORATION
SFC 1507
1 LINCOLN ST
BOSTON, MASSACHUSETTS 02111
USA
E-MAIL: lixun.zhang@aya.yale.edu

B. P. LEADERER
DIVISION OF ENVIRONMENTAL HEALTH SCIENCES
YALE SCHOOL OF PUBLIC HEALTH
NEW HAVEN, CONNECTICUT 06520
USA
E-MAIL: brian.leaderer@yale.edu

Y. GUAN
DEPARTMENT OF MANAGEMENT SCIENCE
UNIVERSITY OF MIAMI
CORAL GABLES, FLORIDA 33124-6544
USA
E-MAIL: yguan@bus.miami.edu

T. R. HOLFORD
DIVISION OF BIostatISTICS
YALE SCHOOL OF PUBLIC HEALTH
NEW HAVEN, CONNECTICUT 06520
USA
E-MAIL: theodore.holford@yale.edu