Environmental Science & Technology

ARTICLE

pubs.acs.org/est

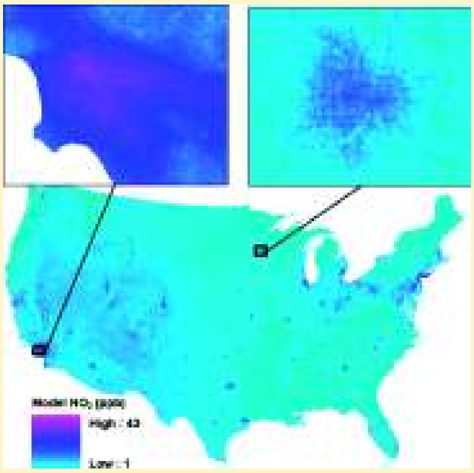
National Satellite-Based Land-Use Regression: NO_2 in the United States

Eric V. Novotny, Matthew J. Bechle, Dylan B. Millet, and Julian D. Marshall*

University of Minnesota, Twin Cities, Minneapolis, Minnesota 55455, United States

S *Supporting Information*

ABSTRACT: Land-use regression models (LUR) estimate outdoor air pollution at high spatial resolution. Previous LURs have generally focused on individual cities. Here, we present an LUR for year-2006 ground-level NO_2 concentrations throughout the contiguous United States. Our approach employs ground- and satellite-based NO_2 measurements, and geographic characteristics such as population density, land-use (based on satellite data), and distance to major and minor roads. The results provide reliable estimates of ambient NO_2 air pollution as measured by the U.S. EPA (R^2 = 0.78; bias = 22%) at a spatial resolution (∼30 m) that is capable of capturing within-urban and near-roadway gradients in NO_2. We explore several aspects of temporal (time-of-day; day-of-week; season) and spatial (urban versus rural; U.S. region) variability in the model. Results are robust to spatial autocorrelation, to selection of an alternative input data set, and to minor perturbations in input data (using 90% of the data to predict the remaining 10%). The modeled population-weighted (unweighted) mean outdoor concentration in the United States is 10.7 (4.8) ppb. Our approach could be implemented in other areas of the world given sufficient road network and pollutant monitoring data. To facilitate future use and evaluation of the results, concentration estimates for the ∼8 million U.S. Census blocks in the contiguous United States are publicly available via the Supporting Information.



1. INTRODUCTION

Predicting spatial and temporal variability in outdoor air pollution is an important research goal for environmental health and air quality engineering and provides key information for disciplines such as environmental economics, public policy, and environmental justice. For example, national epidemiology studies[1−3] have been critically important for advancing environmental regulation in the United States and globally; future studies would benefit from a national-scale model with spatial precision sufficient to resolve gradients in pollutant exposure (<100 m). Here, we present such a model for outdoor NO_2 pollution in the United States.

Land-use regression (LUR) is an empirical-statistical technique that uses in situ concentration measurements and information about surrounding land-uses to estimate concentrations for nonmeasurement locations. Details of LUR, including comparisons to alternative approaches, are published elsewhere.[4−7] LUR has been applied throughout North America and Europe,[8−10] typically for one city at a time. More recently, a small number of LUR models have been presented with broader spatial coverage.[11−15] For example, Beelen et al.[11] developed a LUR for Europe at a 1 km grid. (In contrast, continental-scale dispersion models typically use grid sizes of 12 km or larger.[16,17]) Hart et al.[13] developed a national-scale LUR for a cohort dispersed throughout the contiguous United States. While LUR is a powerful tool able to provide fine-scale estimates of pollution concentrations, its implementation requires an extensive monitoring network, and transferability among urban areas is unclear.[4,6,18,19]

Satellite measurements of tropospheric nitrogen dioxide (NO_2) column abundance (units: molec cm^{-2}) from the Ozone Monitoring Instrument (OMI), aboard the Aura satellite, offer powerful new information on urban air pollution for constraining LUR models.[20,21] OMI provides near global daily coverage, passing overhead at approximately 13:30 local time,[22−24] providing data quality that is consistent across cities, regions, and countries, and devoid of the methodological differences and political biases sometimes observed for in situ measurements.[25,26] However, OMI's spatial resolution (13 × 24 km^2 at nadir), while the best of present space-borne NO_2 sensors, cannot directly capture near-source (e.g., near-roadway) concentration variability.

Here, we present an LUR combining ground- and satellite-based measurements of air pollution (NO_2 columns from OMI) and land-use. Through this combination, we seek to achieve both the spatial resolution of urban-scale LUR and the broad coverage provided by space-borne measurements. As shown below,

Received: October 22, 2010
Revised: March 2, 2011
Accepted: March 8, 2011
Published: April 26, 2011

dx.doi.org/10.1021/es103578x | *Environ. Sci. Technol.* 2011, 45, 4407–4414

Table 1. Independent Variables

parameter	units	spatial resolution	buffer[a] or point estimate
impervious surface	%	30 m (United States only[32]); 1000 m (global[29])	buffer
tree canopy	%	30 m (United States only[33]); 500 m (global[30])	buffer
population	no.	Census block (United States only[34]); 1 km (global[31])	buffer
major road length[35]	km	NA	buffer
minor road length[35]	km	NA	buffer
total road length[35]	km	NA	buffer
elevation[36]	km	90 m	point
distance to coast	km	NA	point
OMI NO_2[25,26]	ppb	13 × 24 km^2 at nadir	point

[a] Buffers employed (m): 100; 200; 300; 400; 500; 600; 700; 800; 1000; 1200; 1500; 1800; 2000; 2500; 3000; 3500; 4000; 5000; 6000; 7000, 8000; 10000.

the results offer universal spatial coverage for the contiguous United States, at sufficient spatial precision (~30 m) to capture near-roadway pollution impacts (~200−500 m).[27] We aim to advance LUR in three main ways: (1) by incorporating satellite-based estimates of air pollution concentrations, (2) by enabling quantitative exploration of diel and other temporal variability in spatial patterns, and (3) by constructing and making publicly available a national-scale model with spatial precision capable of predicting within-urban variability. The resolution of our national-scale model (~30 m) is consistent with prior urban-scale LUR.

2. METHODS

As described next, we used multivariate linear regression to predict atmospheric NO_2 surface concentrations based on data from ambient monitors, GIS land-use characteristics, and NO_2 measurements from the OMI satellite sensor.

2.1. Input Data. *2.1.1. Monitor Data.* We obtained hourly NO_2 concentrations from the United States Environmental Protection Agency's 423 ambient monitors in the contiguous United States (AQS Datamart, year-2006; www.epa.gov/ttn/airs/aqsdatamart). While these NO_2 measurements have a known bias due to interference from other oxidized nitrogen species,[25,26,28] values reported by the EPA are used extensively in research and practice (e.g., in determining regulatory compliance and by the health community in epidemiological studies). For this reason, we use available EPA data "as is", without any interference corrections.

Using EPA's reliability criterion, stations missing more than 25% of the hourly values were excluded. Our primary model considers annual-mean concentrations at each monitor. To explore temporal variability, we also consider means by season, hour-of-day, and day-type (weekday, weekend) for a total of 126 temporally explicit models (4 seasonal, 24 hourly annual, 96 hourly seasonal, and 2 day-type [weekend/weekday] models). To explore spatial variability, we grouped stations by region (four quadrants of the United States: East Coast, Midwest, Mountain West, West Coast), by three EPA-designated location types (urban, suburban, rural), and by distance to major roadway; see Figure S1 in the Supporting Information.

2.1.2. Land-Use. Table 1 lists land-use characteristics serving as independent variables in the model. For several variables (impervious surface, tree canopy, population density), we compare two available data sources. Our main model employs globally available data ("global" data; 500−1000 m spatial resolution).[29−31] As a sensitivity analysis, we employ finer-resolution data (as fine as 30 m) covering only the contiguous United States ("United States only" data).[32−34] Both models employ year-2009 United States Census (Tiger) road networks,[35] which we divide into major roads (motorway, primary, secondary, trunk), minor roads (residential, track, tertiary), and total (major plus minor roads). Our distance-to-coast measure excludes the Great Lakes. Independent variables were analyzed either as buffer or as point estimates. A buffer estimate is an average value over a specified length scale (i.e., circular radius) from a location, while a point estimate is the specific value at that location.

2.1.3. NO_2 Column Measurements from the OMI Satellite Sensor. We employ OMI tropospheric NO_2 column data from the DOMINO product (version 1.0.2, collection 3; available at http://www.temis.nl). Following Lamsal et al.[25,26] and Bechle et al.,[37] we derived estimates of ground-level NO_2 concentrations using tropospheric NO_2 column measurements from OMI and local surface-to-column ratios from a global three-dimensional atmospheric model (GEOS-Chem; www.geos-chem.org[38]). Specifically, we generate year-2006 daily estimates of surface NO_2 concentration using 3 h average surface-to-column ratios for 12:00−15:00 local time, corresponding with Aura overpass. This approach has been shown to give surface NO_2 concentrations in urban areas that are well-correlated with corrected in situ observations, and with <30% bias that does not vary strongly throughout the year.[25,26] We calculated seasonal mean NO_2 abundance in the case of the seasonal LUR, and annual means for all other cases. Satellite-based NO_2 estimates reflect contributions from all sources, including emissions from sources (e.g., industry, airports, harbors) not explicitly included in our land-use data set.

2.2. LUR Model Formation. The LUR incorporates six land-use characteristics calculated for 22 radial buffer lengths (between 100 m and 10 km; see Table 1) and three point-derived land-use and satellite-based values for a total of 135 candidate independent variables. We tested independent variables from global and United States data sets (impervious surfaces, tree cover, population; Table 1) to determine if the higher resolution United States only data set improved model performance.

We employed the stepwise multivariate regression technique of Su et al.[39] The independent variable most correlated with the dependent variable is added to the model first. The parameter most correlated with model residuals is then selected as the next independent variable, and the process repeats. Each independent variable stays in the model if the significance level is less than 0.05 and the variance inflation factor (VIF, a check for multicollinearity) with parameters already in the model is less than 5. Following Henderson et al.,[40] when an independent variable at a

Table 2. Model Performance for the Four Annual-Mean NO_2 LURs

	R^2	adj. R^2	mean absolute error (ppb)	mean bias (%)	mean absolute bias (%)	p
global with OMI NO_2	0.78	0.77	2.4	22	39	<0.001
United States with OMI NO_2	0.77	0.76	2.4	24	41	<0.001
global without OMI NO_2	0.66	0.66	2.9	28	46	<0.001
United States without OMI NO_2	0.68	0.67	2.9	28	47	<0.001

Table 3. Annual Mean Model Using Global Data Sets and OMI NO_2[a]

parameter	units	β	std. err.	$p > \|t\|$	partial R^2	IQR	β*IQR	VIF
intercept	ppb	4.5	0.49	<0.01				
impervious (6000 m)	%	0.12	0.01	<0.01	0.55	35	8.6	1.9
annual OMI NO_2	ppb	1.0	0.06	<0.01	0.70	3.3	3.2	1.5
tree canopy (1000 m)	%	−0.09	0.02	<0.01	0.74	8.4	−0.91	1.2
major roads (800 m)	km	0.20	0.08	0.01	0.76	3.2	0.97	2.8
minor roads (100 m)	km	3.7	1.1	<0.01	0.76	0.27	0.82	1.2
elevation	km	2.0	0.47	<0.01	0.77	0.27	0.31	1.5
distance to coast	km	-1.5×10^{-3}	3.9×10^{-4}	<0.01	0.77	630	−0.86	1.6
major roads (200 m)	km	1.3	0.67	<0.05	0.78	0.19	0.25	2.4

[a] Parameters are listed in the order in which they were added to the model. Distance in parentheses is the buffer radius. Parameters without a buffer distance were taken at the station locations. IQR is the inter-quartile range for the given parameter at EPA monitor sites. β*IQR is the β coefficient multiplied by the IQR. VIF (variance inflation factor) is a check for multicollinearity.

specific buffer length enters the model, other buffer lengths of that variable are still allowed as possible candidates, subject to the VIF and significance level requirements. We stop when the next added independent variable would not be statistically significant or would fail the VIF multicollinearity check. Separate NO_2 models were created using this approach for each dependent variable: annual-mean concentration (main model) plus the 126 temporal summaries mentioned above.

2.3. Model Evaluation. We quantified model performance on the basis of correlation, error, and bias between predictions and observations, and evaluated spatial autocorrelation among the model residuals. We tested model robustness using Monte Carlo random sampling as follows. We conducted model-building using 90% of the monitoring data and tested the model's ability to predict the remaining 10%. We then calculated R^2, error, and bias for all 500 iterations. We further evaluated model robustness using spatial characteristics, temporally specific models, and by comparing United States only versus global data. Finally, we developed models with and without the OMI data to quantify the information value of the satellite-derived NO_2 estimates for LUR.

3. RESULTS

3.1. Monitor Data. Of the 423 stations in year-2006, 369 met the reliability criterion and were therefore employed in the model. Figure S2 in the Supporting Information summarizes monitor data used here. The median (mean) annual mean NO_2 concentration from these data was 11.4 (12.0) ppb. Median concentrations (ppb) at monitoring stations were higher for winter (14.0) and fall (12.0) than for spring (9.8) and summer (8.3). Median concentrations at monitoring stations were higher during weekdays than weekends (12.0 versus 9.3) and higher in urban and suburban areas (15.2 and 13.1, respectively) than in rural areas (6.2).

3.2. Annual Models. Table 2 shows model performance for the four annual-mean LUR models (United States only versus global data, and including versus excluding OMI-derived NO_2 concentrations). Model performance is nearly identical for the global versus US-only data sets, supporting the robustness of the overall LUR. This finding also suggests that the global data could be used to develop high-fidelity LURs for other world regions, where higher resolution inputs may not be available. In contrast, the OMI-based surface NO_2 estimates provide a significant improvement (~0.1 increase in R^2) in model performance, demonstrating the utility of satellite-derived air pollution data for continental-scale models.

The model with the highest R^2 in Table 2 (global, with OMI NO_2) is shown in Table 3, with variables listed in the order they were accepted into the model. Corresponding information for the remaining three Table 2 models is provided in the Supporting Information (Table S1). Signs and magnitudes for coefficients are listed in Table 3. Roads have a positive coefficient (more roads = higher NO_2 concentrations), reflecting NO_x emissions from motor vehicles. Tree canopy has a negative coefficient (more tree-cover = lower NO_2 concentrations), likely reflecting reduced combustion sources for locations with greater tree cover.[41] The satellite-based estimate of NO_2 concentration is the second variable selected by the model (after impervious surfaces, 6 km buffer), and with a coefficient (β) of 1.0, indicating the importance and direct utility of OMI estimates for the LUR presented here. The first four parameters in Table 3 contribute 97% of the model predictive power (partial R^2 = 0.76). We employ the model in Table 3 for the analyses that follow.

3.3. Model Evaluation. Mean bias, absolute bias, error, and absolute error for the core model (Table 3; equations S1−S4) are 22%, 39%, 0.0 ppb, and 2.4 ppb, respectively.

Results from the Monte Carlo simulations indicate that the model is robust and not sensitive to the population of EPA monitors used to develop the LUR. Test models built using 90% of the data (model-building) perform nearly as well at predicting the remaining 10% of the data (model-testing) as at predicting the 90% used in their construction: among the 500 iterations, median R^2 values are 0.78 for the model-building data and 0.76 for the model-testing data (Figure S3). Other metrics (error, bias)

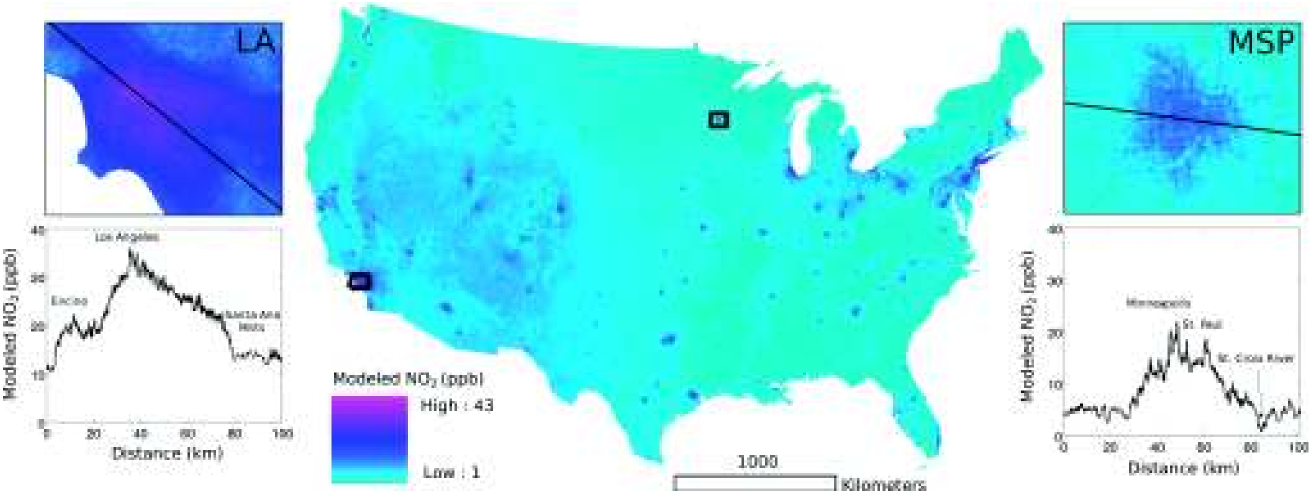


Figure 1. Modeled ground-level year-2006 annual-average NO_2 concentrations for all U.S. Census Blocks in the contiguous United States. Zoom-in panels show concentrations in Los Angeles (LA) and Minneapolis/St. Paul (MSP), employing a 30 m grid for display purposes. Line plots provide concentration estimates along the cross sections shown in the zoom-in plots for LA and MSP.

are given in Table S2 in the Supporting Information. For 86% of the 500 models, the first four parameters selected for the model were, in order, impervious surfaces (buffer distance between 3000 and 8000 m), satellite NO_2, tree canopy (buffer distances between 800 and 3000 m), and major roads (buffer distances between 200 and 800 m), that is, the same variables and inclusion order as in the core model (Table 3).

We tested spatial autocorrelation of the model residuals by calculating Moran's I, a local indicator of spatial association (LISA), at the 95% confidence interval[42] using ArcGIS. Through this process, 66 monitor outliers were suggested for removal. We created an LUR model using this new data set and compared it to the existing model. Predicted concentrations are nearly identical between the core and test models (see Figure S4; R^2 = 0.97), suggesting that spatial autocorrelation does not have a major impact on our model results.

We also investigated the extent to which monitor locations span the (independent) variable space, an important issue for any LUR. For example, monitors should measure near and away from roadways, and in more and less populated locations. As shown in Table S3, station locations range from 1 to 41 000 m from a major road. Of the EPA monitors used here, 24 are within 50 m of a major road, 31 are 50−100 m from a major road, and 51 are 100−200 m from a major road. Similar findings hold for other independent variables: the EPA monitors generally span the variable space. Among monitoring locations, impervious surface within 6000 m ranges from 0 to 74%, length of major roads within 800 m ranges from 0 to 22 km, and percent tree canopy within 1000 m ranges from 0 to 77%. Monitors are roughly evenly divided between urban (34%), suburban (39%), and rural (27%) locations.

3.4. Application of the Model. To provide visual representation of the LUR, we applied the core model (Table 2) to the centroid locations for all ~8 million Census blocks in the contiguous United States (blocks are the smallest area enumerated by the United States Census; average population per block: ~35 people) and to 30 m grids for metropolitan Minneapolis/St. Paul, MN and Los Angeles, CA (Figure 1). For visual depiction on these maps, modeled values between 0 and 1 ppb were rounded to 1 ppb. Highest concentrations are near roadways and in more populated areas, with peak (median) concentrations of 22 (4.9) ppb in Minneapolis and 43 (20) ppb in Los Angeles.

The predicted mean population-weighted (unweighted) outdoor concentration for the contiguous United States is 10.7 (4.8) ppb. Figure 1 also shows NO_2 concentration estimates along urban cross sections for the same two cities. Concentrations exhibit urban-to-rural and near-roadway concentration gradients,[43−46] with simulated concentration spikes near major highways ~200−500 m wide, broadly consistent with available estimates.[27,47]

3.5. Temporal Variability. Seasonal and day-type (weekends, weekdays) models, given in the Supporting Information (Table S4), exhibit only minor differences in performance as compared to the core model. For example, R^2 values are 0.74 versus 0.76 for summer and winter models, respectively, and 0.78 versus 0.75 for weekday and weekend models, respectively.

Diel annual models (24 models: one per hour-of-day) and hourly seasonal models (96 models: one per hour-of-day for each season) showed modest variability in model performance. Model R^2 values are higher around midday than at night (Figure 2A). Diel patterns in R^2 values depend on the inclusion of OMI NO_2 satellite data in the LUR (Figure 2B); including the OMI data provides the greatest performance enhancement during mid-day, around the time of the Aura satellite overpass.

3.6. Spatial Variability. The core model performs moderately better for urban and suburban areas (R^2 = 0.72 and 0.68, respectively) than for rural areas (0.60). This finding supports the use of our LUR for exposure assessment and environmental epidemiology, because most people (~80%[48]) in the United States live in urban or suburban areas. On the other hand, it also indicates that the model is less reliable for predicting NO_2 levels in low-population areas. NO_2 sources (e.g., motor vehicles, electricity generation) may be more directly correlated with land-use for urban than for rural areas. Results by geographic region show that model performance is highest for the East Coast domain (R^2 = 0.83) and is similar among the remaining three regions ($R^2 \approx 0.75$). The East Coast region is the most populated of the four areas, with ~37% of the United States population.[48] Comparing model performance based on station distance from a major roadway yielded the following R^2 values: 0.74 (comparing only to stations 0−200 m from major road; n = 106), 0.71

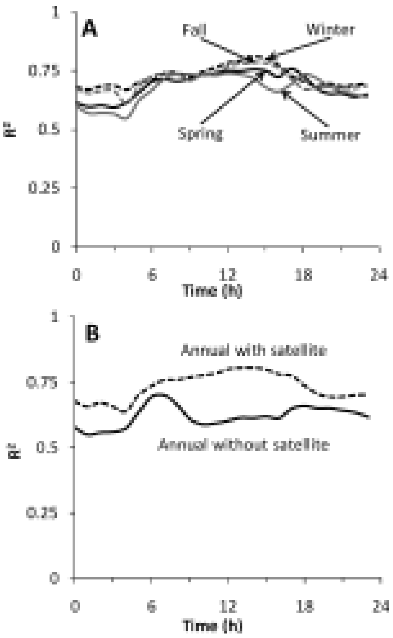


Figure 2. (A) Model performance for the 96 hourly seasonal models. (B) Model performance for hourly models with and without satellite-derived estimates of NO_2 concentrations.

(200−400 m; n = 70), 0.64 (400−600 m; n = 55), 0.75 (600−800 m; n = 29), 0.68 (800−1000 m; n = 20), and 0.74 (>1000 m; n = 89). Additional details, including modeled versus observed NO_2 by location type and region, and location-specific models (urban, suburban, rural) are in the Supporting Information (Figure S5; Table S5).

We calculate the length scale of spatial variability, defined as the distance where the mean concentration difference between points is equal to one-half the overall spatial standard deviation of all concentrations, as 2.4 km for Minneapolis and 4.5 km for Los Angeles. These values are slightly larger than the corresponding value (1.0 km) for Vancouver, Canada, based on a prior NO_2 LUR.[5,40] We examined spatial and temporal variability for one urban area (Minneapolis/St Paul; Figure 3). Concentrations are highest near major roadways and close to the city center in mornings (6:00−7:00) and at night (20:00−22:00) (Figure 3A,C). Relative to on-roadway, annual mean concentrations are 35% lower 1 km from a roadway (Figure 3B) and 50% lower 2.5 km from a major roadway (not shown). A recent literature review for multiple cities[47] reported that NO_2 concentrations 550 m from a major roadway are on average 42% lower than on-roadway; Figure 3B suggests, for Minneapolis, a 28% decline at 550 m. Concentration differences are larger for urban gradients than for near-roadway gradients (i.e., 55−70% differences in Figure 3C versus 10−35% differences in Figure 3A). Concentrations decrease sharply at close distances from major roadways (within ∼30 m) and the city center (within ∼5 km) and more gradually at larger distances.

4. DISCUSSION

We developed and implemented an LUR for the contiguous United States using fixed-site ambient monitoring station data, land-use data (including satellite-derived land-use data), and satellite-derived ground-level NO_2 concentrations. The model exhibits good predictive power [R^2 = 0.78, absolute error = 2.4 ppb], at similar spatial resolution as prior urban-scale LUR models. Our model improves on prior LUR via temporal precision (hour-of-day, day-of-week) and input data (use of satellite-derived estimates of ground-level air pollution).

Several limitations of our study merit highlighting. First, similar to prior LUR, we focus mainly on population- and traffic-related variables; we do not investigate sources such as industry, airports, or harbors. OMI-based estimates of surface NO_2 provide background concentration (incorporating emissions from all sources), but prior research[49] indicates that traffic-focused LURs perform poorly in areas with significant nontraffic sources (e.g., industrial hot-spots). Second, because traffic data are not available over the entire United States, our model is derived from road length rather than traffic volumes. Third, as is common with empirical models, our model performance evaluations employ the data used to derive the model. The 90%/10% Monte Carlo approach aims to mitigate this limitation, but we believe that future research could usefully compare our model results against independent concentration measurements or against available urban LUR models. Fourth, because no existing LUR explores time-of-day variability in concentrations (most urban-scale LUR models evaluate annual-mean concentrations, or in some cases monthly concentrations), we are unable to corroborate this aspect of our model. Further research into temporal aspects of spatial patterns in ambient concentrations is needed. Finally, our model only predicts outdoor concentrations; indoor sources such as natural-gas stoves can also contribute to total air pollutant exposure.[50]

Despite these limitations, our model captures 78% of the variability in EPA-measured ambient NO_2 throughout the contiguous United States. An urban-only model performed slightly better (R^2 = 0.80; Table S5 in the Supporting Information). We found that increased spatial precision in the land-use input data did not improve model performance, whereas including OMI NO_2 data did yield a significant improvement. The model coefficient for OMI NO_2 (Table 3) is 1.0, emphasizing the direct utility of satellite data for questions considered here; its status as the second variable selected in the model further underscores its utility. The regression approach accounts implicitly for many potential biases in the independent variables. For example, a bias in the OMI-derived NO_2 estimates would simply result in a different regression coefficient for this parameter; model predictions and model performance would remain unchanged.

Independent variables employed here (except the road network) are available globally. Thus, our approach could readily be applied to other countries or regions given sufficient ambient monitoring and road network data. Prior research demonstrates the critical need to have a sufficient number of monitoring stations to obtain a robust LUR.[4] For example, Hystad et al. demonstrate that Canadian ambient monitors are too sparse to derive a straightforward national-scale LUR.[15] For several pollutants, they combine regional concentration estimates, derived from LUR and satellite data, with literature-derived assumptions about near-roadway trends to predict concentrations throughout Canada.

We hope that our model results will be useful for epidemiological studies and other investigations seeking national-scale, spatially precise estimates of air pollution. To facilitate such research, concentration estimates for all Census blocks in the contiguous United States (Figure 1) are publicly available via the Supporting Information.

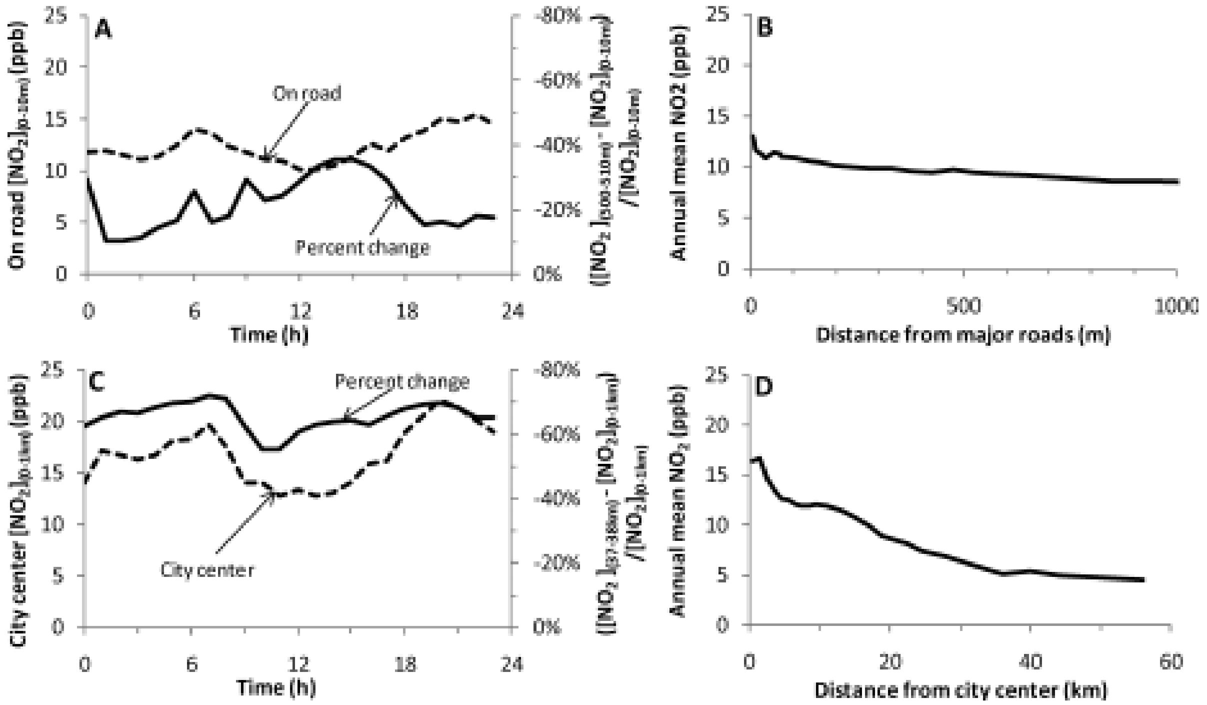


Figure 3. Spatiotemporal variability in predicted ambient NO_2 concentrations, based on annual-mean and diel LUR models for an example city (Minneapolis/St. Paul metropolitan area): distance from major roadways (upper plots) and city center (lower plots). (A) Concentrations on-road (0−10 m from a major roadway; dashed line) and percent change away from roadway (500−510 m from a major roadway) relative to on-roadway (solid line), as a function of hour-of-day. (B) Annual mean concentration versus distance from a major roadway. (C) Concentrations at city center (0−1 km from Minneapolis City Hall; dashed line) and percent change away from city center (37−38 km from city center) relative to at city center (solid line). (D) Mean annual concentration versus distance from city center.

■ ASSOCIATED CONTENT

S Supporting Information. Figure S1: EPA monitoring stations locations. Figure S2: Summary of EPA NO_2 data. Table S1: Annual LUR model results. Figure S3: Monte Carlo results. Table S2: Error and bias for Monte Carlo simulations. Figure S4: Spatial autocorrelation comparison. Table S3: Model parameter ranges. Table S4: Model statistics for seasonal models. Figure S5: Model results by location type. Table S5: LUR models by location type. Three data files provide LUR-derived NO_2 concentration estimates (ppb): one file ("Read me") describes the data, another file ("Preview") illustrates the semicolon-separated format for the database by providing data for the first 100 Census Blocks in the database, and the last file ("NO2_ByCensusBlock"; file size: 810 MB) provides estimates for all Census blocks in the contiguous United States. All files can be downloaded here: http://personal.ce.umn.edu/~marshall/data.php. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (612) 625-2397; fax: (612) 626-7750; e-mail: julian@umn.edu.

■ ACKNOWLEDGMENT

Partial funding for this work was provided by the NSF-Division of Chemical, Bioengineering, Environmental, and Transport Systems (CBET) (grant 0853467) and the University of Minnesota. We acknowledge the free use of tropospheric NO_2 column data from the OMI sensor via www.temis.nl. We thank Michael Brauer, Perry Hystad, and anonymous reviewers for helpful comments on this work.

■ REFERENCES

(1) Laden, F.; Schwartz, J.; Speizer, F. E.; Dockery, D. W. Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities Study. *Am. J. Respir. Crit. Care Med.* **2006**, *173*, 667–672.

(2) Pope, C. A.; Burnett, R. T.; Thun, M. J. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc.* **2002**, *287*, 1132–1141.

(3) Allen, R. W.; Criqui, M. H.; Diez Roux, A. V.; Allison, M.; Shea, S.; Detrano, R.; Sheppard, L.; Wong, N. D.; Stukovsky, K. H.; Kaufman, J. D. Fine particulate matter air pollution, proximity to traffic, and aortic atherosclerosis. *Epidemiology* **2009**, *20*, 254–264.

(4) Johnson, M.; Isakov, V.; Touma, J. S.; Mukerjee, S.; Özkaynak, H. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmos. Environ.* **2010**, *44*, 3660–3668.

(5) Marshall, J. D.; Nethery, E.; Brauer, M. Within-urban variability in ambient air pollution: Comparison of estimation methods. *Atmos. Environ.* **2008**, *42*, 1359–1369.

(6) Hoek, G.; Beelen, R.; Dehoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561–7578.

(7) 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*, 185–204.

(8) Su, J. G.; Jerrett, M.; Beckerman, B.; Wilhelm, M.; Ghosh, J. K.; Ritz, B. Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. *Environ. Res.* **2009**, *109*, 657–670.

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

(10) Vienneau, D.; Hoogh, K. de; Beelen, R.; Fischer, P.; Hoek, G.; Briggs, D. Comparison of land-use regression models between Great Britain and the Netherlands. *Atmos. Environ.* **2010**, *44*, 688–696.

(11) Beelen, R.; Hoek, G.; Pebesma, E.; Vienneau, D.; Hoogh, K. de; Briggs, D. J. Mapping of background air pollution at a fine spatial scale across the European Union. *Sci. Total Environ.* **2009**, *407*, 1852–1867.

(12) Gulliver, J.; Morris, C.; Lee, K.; Vienneau, D.; Briggs, D.; Hansell, A. Land use regression modeling to estimate historic (1962–1991) concentrations of black smoke and sulfur dioxide for Great Britain. *Environ. Sci. Technol.* **2011**, *45*, 3526–3532.

(13) Hart, J. E.; Yanosky, J. D.; Puett, R. C.; Ryan, L.; Dockery, D. W.; Smith, T. J.; Garshick, E.; Laden, F. Spatial modeling of PM10 and NO_2 in the continental United States, 1985–2000. *Environ. Health Perspect.* **2009**, *117*, 1690–1696.

(14) Yanosky, J. D.; Paciorek, C. J.; Suh, H. H. Predicting chronic fine and coarse particulate exposures using spatiotemporal models for the Northeastern and Midwestern United States. *Environ. Health Perspect.* **2009**, *117*, 522–529.

(15) Hystad, P.; Setton, E.; Cervantes, A.; Poplawski, K.; Deschenes, S.; Brauer, M.; van Donkelaar, A.; Lamsal, L.; Martin, R.; Jerret, M.; Demers, P. Creating national air pollution models for population exposure assessment in Canada. *Environ. Health Perspect.* **2011**, in press.

(16) Cook, R.; Phillips, S.; Houyoux, M.; Dolwick, P.; Mason, R.; Yanca, C.; Zawacki, M.; Davidson, K.; Michaels, H.; Harvey, C.; Somers, J.; Luecken, D. Air quality impacts of increased use of ethanol under the United States' Energy Independence and Security Act. *Atmos. Environ.* **2011**, in press.

(17) Zhang, Y.; Vijayaraghavan, K.; Wen, X. Y.; Snell, H. E.; Jacobson, M. Z. Probing into regional ozone and particulate matter pollution in the United States: 1. A 1 year CMAQ simulation and evaluation using surface and satellite data. *J. Geophys. Res.* **2009**, *114*, DOI: 10.1029/2009JD11898.

(18) Briggs, D. J.; Collins, S.; Elliott, P.; Fischer, P.; Kingham, S.; Lebret, E.; Pryl, K.; Reeuwijk, H. V. A. N.; Smallbone, K. *Int J. Geogr. Inf. Sci.* **1997**, *11*, 699–718.

(19) Poplawski, K.; Gould, T.; Setton, E.; Allen, R.; Su, J. Intercity transferability of land use regression models for estimating ambient concentrations of nitrogen dioxide. *J. Exposure Sci. Environ. Epidemiol.* **2009**, *19*, 107–117.

(20) Schoeberl, M. R.; Douglass, R.; Hilsenrath, E.; Bhartia, P. K.; Beer, R.; Waters, J. W.; Gunson, M. R.; Froidevaux, L.; Gille, J. C.; Barnett, J. J.; Levelt, P. F.; DeCola, P. Overview of the EOS aura mission. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1066–1074.

(21) Levelt, P. F.; Den Oord, G. H. J.; van; Dobber, M. R.; Malkki, A.; Stammes, P.; Lundell, J. O. V.; Saari, H. The ozone monitoring instrument. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1093–1101.

(22) Donkelaar, A. V.; Martin, R. V.; Brauer, M.; Kahn, R.; Verduzco, C.; Villeneuve, P. J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, 847.

(23) Martin, R. Satellite remote sensing of surface air quality. *Atmos. Environ.* **2008**, *42*, 7823–7843.

(24) Richter, A.; Burrows, J. P.; Nüss, H.; Granier, C.; Niemeier, U. Increase in tropospheric nitrogen dioxide over China observed from space. *Nature* **2005**, *437*, 129–132.

(25) Lamsal, L. N.; Martin, R. V.; Donkelaar, A. V.; Steinbacher, M.; Bucsela, E.; Dunlea, E. J.; Pinto, J. P. Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. *J. Geophys. Res.* **2008**, *113*, DOI: 10.1029/2007JD009235.

(26) Lamsal, L. N.; Martin, R. V.; Donkelaar, A. V.; Celarier, E. A.; Bucsela, E. J.; Boersma, K. F.; Dirksen, R.; Luo, C.; Wang, Y. Indirect validation of tropospheric nitrogen dioxide retrieved from the OMI satellite instrument: Insight into the seasonal variation of nitrogen oxides at northern midlatitudes. *J. Geophys. Res.* **2010**, *115*, DOI: 10.1029/2009JD013351.

(27) Zhou, Y.; Levy, J. I. Factors influencing the spatial extent of mobile source air pollution impacts: A meta-analysis. *BMC Public Health* **2007**, *7*, 89.

(28) Dunlea, E.; Herndon, S.; Nelson, D.; Volkamer, R.; San Martini, F.; Sheehy, P.; Zahniser, M.; Shorter, J.; Wormhoudt, J.; Lamb, B. Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment. *Atmos. Chem. Phys.* **2007**, *7*, 2691–2704.

(29) Elvidge, C. D.; Tuttle, B. T.; Sutton, P. C.; Howard, A. T.; Milesi, C. Global distribution and density of constructed impervious surfaces. *Sensors* **2007**, *7*, 1962–1979.

(30) Defries, R. S.; Hanson, M. C.; Townshend, J. R. G.; Janetos, A. C.; Loveland, T. R. A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biol.* **2000**, *6*, 247–254.

(31) Center for International Earth Science Information Network (CIESIN); Centro Internacional De Agricultura Tropical Gridded Population of the World Version 3 (GPWv3): Population Density Grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University, 2005.

(32) Yang, L.; Huang, C.; Homer, C. G.; Wylie, B. K.; Coan, M. J. An approach for mapping large-area impervious surfaces: Synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. *Can. J. Remote Sens.* **2003**, *29*, 230–240.

(33) Huang, C.; Yang, L.; Wylie, B.; Homer, C. *Third International Conference on Geospatial Information in agriculture and Forestry*; Denver, CO, 2001; p 10.

(34) United States Census Bureau. 2000 Census Block Statistics, 2001.

(35) United States Census Bureau. 2009 Tiger/Line Shapefiles. www.census.gov/geo/www/tiger.

(36) Jarvis, A.; Reuter, H. I.; Nelson, E. G., 2008; http://srtm.csi.cgiar.org.

(37) Bechle, M. J.; Millet, D. B.; Marshall, J. D. Effects of income and urban form on urban NO_2: Global evidence from satellites. *Environ. Sci. Technol.* **2011**, in press.

(38) Bey, I.; Jacob, J.; Yantosca, R. M.; Logan, A.; Field, B. D.; Fiore, A. M.; Li, Q.; Liu, H. Y.; Mickley, J.; Schultz, M. G. Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *J. Geophys. Res.* **2001**, *106*, 20073–23095.

(39) Su, J. G.; Jerrett, M.; Beckerman, B. A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. *Sci. Total Environ.* **2009**, *407*, 3890–3898.

(40) Henderson, S. B.; Beckerman, B.; Jerrett, M.; Brauer, M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ. Sci. Technol.* **2007**, *41*, 2422–2428.

(41) Nowak, D.; Crane, D.; Stevens, J. Air pollution removal by urban trees and shrubs in the United States. *Urban For. Urban Greening* **2006**, *4*, 115–123.

(42) Anselin, L. Local indicators of spatial association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115.

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

(44) Janssen, N. A. H.; van Vliet, P. H. N.; Aarts, F.; Harssema, H.; Brunekreef, B. Assessment of exposure to traffic related air pollution of children attending schools near motorways. *Atmos. Environ.* **2001**, *35*, 3875–3884.

(45) Rijnders, E.; Janssen, N. A. H.; van Vliet, P. H. N.; Brunekreef, B. Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density. *Environ. Health Perspect.* **2001**, *109*, 411–417.

(46) Singer, B. C.; Hodgson, A. T.; Hotchi, T.; Kim, J. J. Passive measurement of nitrogen oxides to assess traffic-related pollutant exposure for the East Bay Children's Respiratory Health Study. *Atmos. Environ.* **2004**, *38*, 393–403.

(47) Karner, A. A.; Douglas, S. E.; Niemeier, D. A. Near-roadway air quality: Synthesizing the findings from real-world data. *Environ. Sci. Technol.* **2010**, *44*, 5334–5344.

(48) U.S. Census Bureau, http://factfinder.census.gov/home/saff/main.html?_lang=enmat=US-1, 2000.

(49) Beelen, R.; Voogt, M.; Duyzer, J.; Zandveld, P.; Hoek, G. Comparison of the performances of land use regression modelling and dispersion modelling in estimating small-scale variations in long-term air pollution concentrations in a Dutch urban area. *Atmos. Environ.* **2010**, *44*, 4614–4621.

(50) Spengler, J. D.; Duffy, C. P.; Letz, R.; Tibbitts, T. W.; Ferris, B. G. Nitrogen dioxide inside and outside 137 homes and implications for ambient air quality standards and health effects research. *Environ. Sci. Technol.* **1983**, *17*, 164–168.

Supporting Information

National satellite-based land use regression: NO_2 in the United States

Eric V. Novotny, Matthew J. Bechle, Dylan B. Millet, Julian D. Marshall

Number of pages: 8

Number of Figures: 5

Number of Tables: 5

U.S. Census Block NO_2 Data

Three data files provide LUR-derived NO_2 concentration estimates (ppb): one file ("Read me") describes the data, another file ("Preview") illustrates the semicolon-separated format for the database by providing data for the first 100 Census Blocks in the database, and the last file ("NO2_ByCensusBlock"; file size: 810 MB) provides estimates for all Census blocks in the contiguous United States. All files can be downloaded here http://personal.ce.umn.edu/~marshall/data.php

Equations

Equations for mean error (ME), absolute error (AE), mean bias (MB) and absolute bias (AB):

$$ME = \frac{1}{N}\sum_{i=1}^{N}\left(C_m - C_o\right) \tag{S1}$$

$$AE = \frac{1}{N}\sum_{i=1}^{N}\left|C_m - C_o\right| \tag{S2}$$

$$MB = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{C_m - C_o}{C_o}\right) \tag{S3}$$

$$AB = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\left|C_m - C_o\right|}{C_o}\right) \tag{S4}$$

where C_m is the modeled average concentration for station i, C_o is the average observed concentration for station i, and N is the number of monitoring stations.

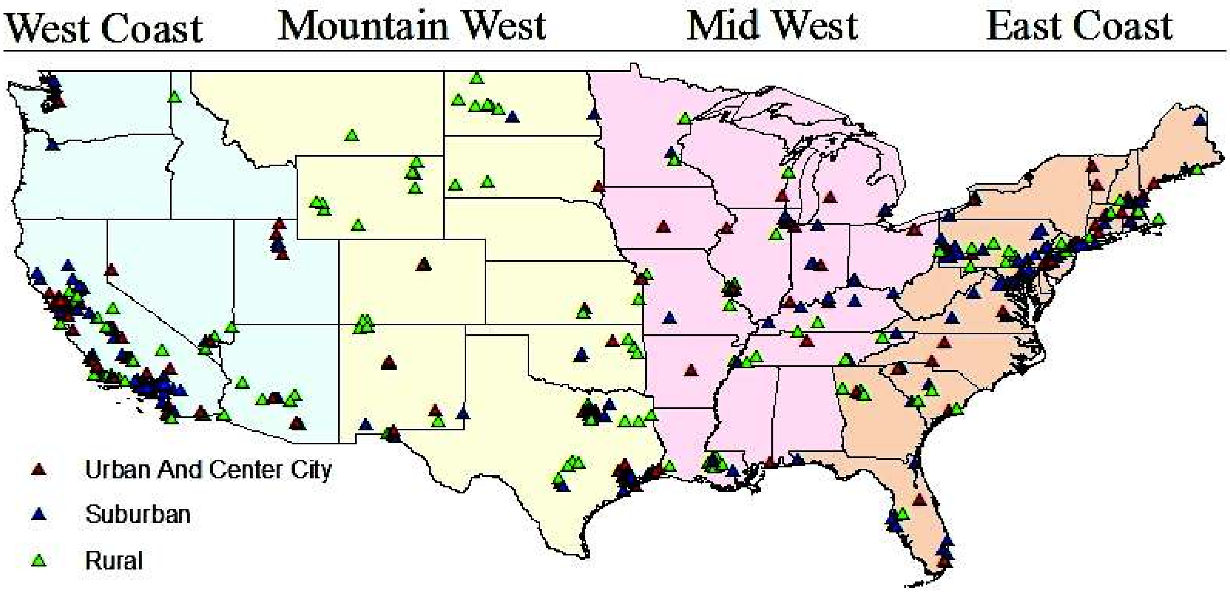


Figure S1. Station locations by region and type.

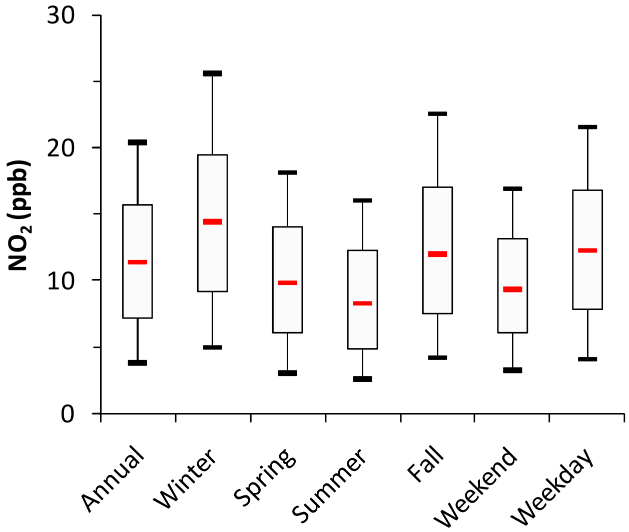


Figure S2. Box plots of year-2006 average NO_2 concentration among the EPA monitors. Inter-quartile ranges are given by the blue boxes; red lines indicate median values. Red lines show median values. Box is the IQR. Black lines outside box are 10th and 90th percentile.

Table S1. Stepwise multiple linear regression analysis for US dataset with OMI NO_2, global dataset without OMI NO_2 and US dataset without OMI NO_2. Parameters are listed in the order in which they were added to the model.

Parameter	**Unit**	***β***	**Std. Err.**	***p*>|t|**	**R^2**	**IQR**	***β* * IQR**	**VIF**
US Dataset with OMI NO_2								
Intercept	--	3.94	0.47	<0.01				
Impervious (7000m)	%	0.12	0.01	<0.01	0.58	31.2	8.48	2.5
OMI NO_2	ppb	0.92	0.07	<0.01	0.70	3.3	2.89	1.6
Tree canopy (600m)	%	-0.47	0.01	<0.01	0.72	15.1	-1.09	1.2
Major roads (700m)	km	0.30	0.07	<0.01	0.74	2.60	0.85	1.4
Impervious (100m)	%	0.03	8.83E-03	<0.01	0.75	53.7	1.91	2.4
Elevation	km	2.36	0.47	<0.01	0.76	0.27	0.42	1.5
Distance to coast	km	-1.17E-03	3.95E-04	<0.01	0.77	630	-0.73	1.5
Minor roads (100m)	km	2.53	1.14	0.03	0.77	0.27	0.68	1.3
Global Dataset without OMI NO_2								
Intercept		7.2	0.54	<0.01				
Impervious (6000m)	%	0.12	0.02	<0.01	0.55	35.1	8.61	3.8
Major roads (800m)	km	0.23	0.07	<0.01	0.58	3.20	1.17	1.4
Population (10000m)	#	7.54E-04	1.69E-04	<0.01	0.61	1100	1.05	1.8
Tree canopy (1800m)	%	-0.09	0.02	<0.01	0.63	11.1	-0.85	1.2
Distance to coast	km	-2.2E-03	7.73E-04	<0.01	0.64	630	-1.01	1.5
Elevation	km	1.81	0.57	<0.01	0.65	0.27	0.5	1.5
Major roads (10000m)	km	6.16E-03	2.17E-03	<0.01	0.66	270	1.66	4.3
US dataset without OMI NO_2								
Intercept		5.70	0.51	<0.01				
Impervious (7000m)	%	0.13	0.02	<0.01	0.58	31.2	8.48	4.7
Population (700m)	#	4.26E-04	1.15E-04	<0.01	0.61	2000	1.21	1.7
Major roads (300m)	km	0.76	0.27	<0.01	0.63	0.53	0.59	1.2
Tree canopy (500m)	%	-0.04	0.02	<0.01	0.64	14.7	-0.77	1.2
Distance to coast	km	-1.78E-03	4.68E-04	<0.01	0.65	630	-0.72	1.6
Elevation	km	2.24	0.55	<0.01	0.66	0.27	0.60	1.5
Impervious (100m)	%	0.03	9.87E-03	<0.01	0.67	53.7	1.38	2.2
Major roads (10000m)	km	7.35E-03	1.99E-03	<0.01	0.68	270	1.98	3.8

Distance in () is the buffer radius, parameters without a buffer distance were taken at the station locations. IQR is the inter-quartile range for the given parameter, *β* * IQR is the *β* coefficient multiplied by the IQR, and VIF is the variance inflation factor to check for multicollinearity.

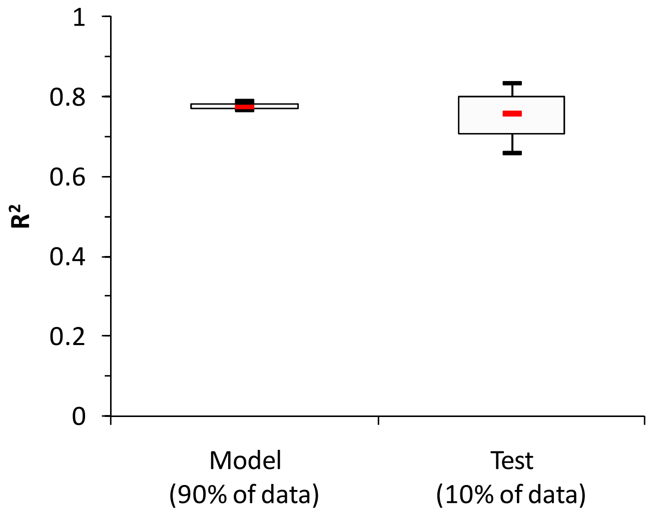


Figure S3. Box plot showing R^2 values between observed and modeled data, for the 90% of data used to create the model (model-building data) and for the remaining 10% (model-testing data) for 500 Monte Carlo simulations. Red lines show median values. Box is the IQR. Black lines outside box are 10th and 90th percentile.

Table S2. Error and bias between the measured values and the model-building and model-testing datasets for the 500 Monte Carlo simulations.

	Model-building	**Model-testing**
Mean error (ppb)	0	0.08
Mean absolute error (ppb)	2.4	2.55
Mean bias (%)	23	25
Mean absolute bias (%)	40	42

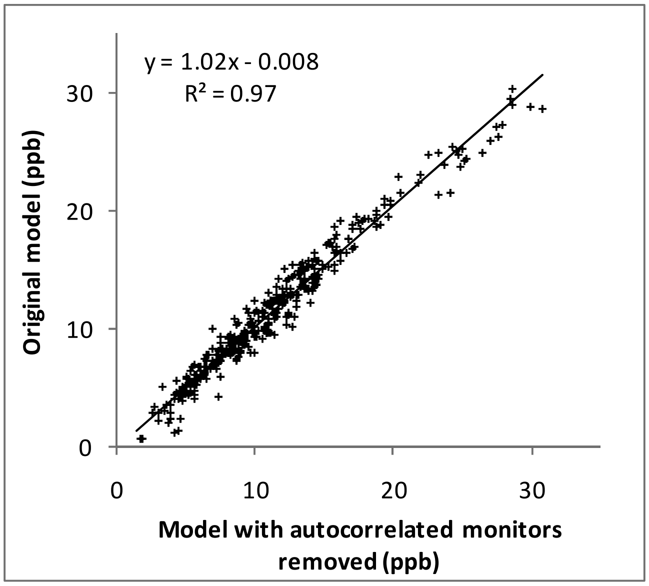


Figure S4. Comparison between the core model (Table 3 in main text) and the autocorrelation-corrected model (same as the core model, but omits 66 stations where the residuals of the models have a statistically significant spatial autocorrelation at the 95% level). We tested spatial autocorrelation of the model residuals by calculating Moran's I using ArcGIS, more information on this topic can be found here:

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Cluster_and_Outlier_Analysis:_Anselin_Local_Moran%27s_I_(Spatial_Statistics)

Table S3: Range of values for station parameters and independent variables.

	Units	Min	Max	Median (IQR)
Station Parameters				
Distance to major road	km	0.001	40.9	0.44 (0.18 - 1.04)
Annual measured NO_2	ppb	0.3	34.2	11.4 (7.2 - 15.6)
Latitude		25.73	48.64	37.14 (33.55 - 40.61)
Longitude		-124.18	-68.03	-95.08 (-115.34 to -81.16)
Independent Variables				
Impervious (6000m)	%	0	74	22.7 (5.8 - 40.9)
OMI NO_2	ppb	0.2	17.5	2.9 (1.5 - 4.8)
Tree canopy (1000m)	%	0	77	5.4 (2.4 - 10.8)
Major roads (800m)	km	0	22.4	1.56 (0 - 3.18)
Minor roads (100m)	km	0	.77	0.16 (0 - 0.27)
Elevation	km	0	2.36	0.15 (0.03 - 0.30)
Distance to coast	km	0	2,100	156 (29.1 - 651)
Major roads (200m)	km	0	2.84	0 (0 - 0.19)

Table S4. Model results for seasonal regression analysis

	R^2	**Adj. R^2**	**N**	**SSE**	**SSR**	**DFR**	**F**	***p***
Fall	0.74	0.73	358	4966	14030	6	169	<0.001
Spring	0.74	0.73	366	3737	10578	8	129	<0.001
Summer	0.76	0.75	385	3473	10736	5	238	<0.001
Winter	0.76	0.76	345	5047	16206	8	138	<0.001
Weekday	0.78	0.77	361	4010	13973	7	180	<0.001
Weekend	0.75	0.74	363	2856	8373	7	152	<0.001

N is the number of stations used in the analysis, SSE is the sum of squared error, SSR is the sum of squared residuals, DFR is the degrees of freedom, F is the F ratio and P is the significance level of the F ratio.

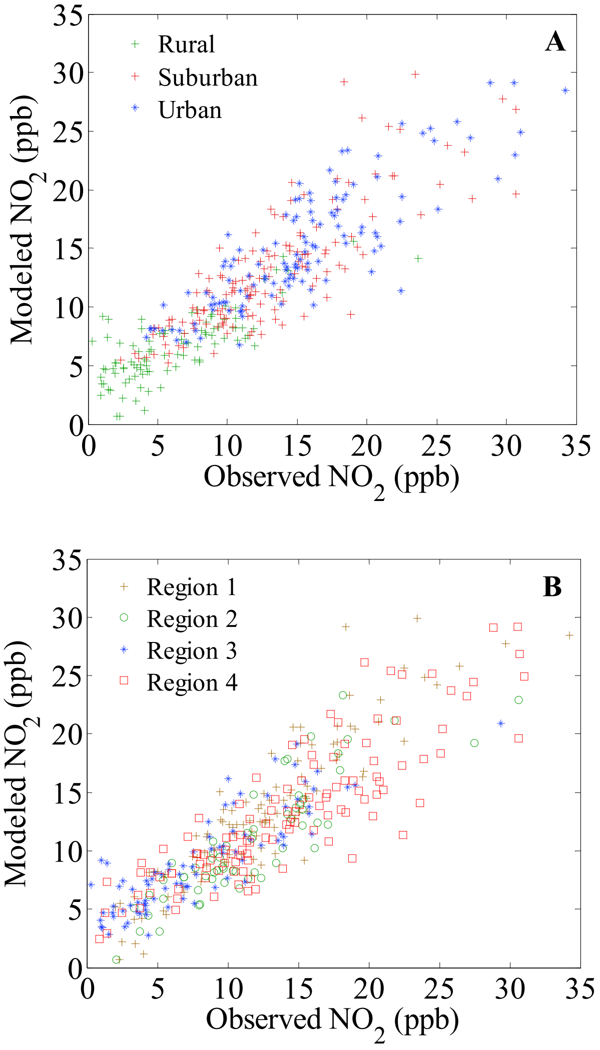


Figure S5. Modeled vs observed results for the annual average model with satellite measurements. Panel A shows the values divided into rural, urban and suburban categories and panel B is divided by regions (Figure S1).

Table S5. Stepwise multiple linear regression for urban, suburban and rural areas. Parameters are listed in the order in which they were added to the model.

					Parti		***β* ***	
Parameter	**Unit**	***β***	**std. err.**	***p*>|t|**	**al R^2**	**IQR**	**IQR**	**VIF**
Model: Urban								
Intercept	--	5.75	1.15	<0.01				
Annual OMI NO_2	ppb	1.31	0.09	<0.01	0.57	3.8	5.48	1.3
Impervious (1800m)	%	0.11	0.02	<0.01	0.68	23.8	2.63	2.4
Elevation	km	3.75	0.77	<0.01	0.74	0.22	0.89	1.2
Major roads (800m)	km	0.19	6.84E-02	0.01	0.76	3.56	0.74	1.2
Tree canopy (6000m)	%	-0.10	0.03	<0.01	0.77	8.77	-0.79	1.3
Minor roads (3000m)	km	-1.59E-02	5.07E-03	<0.01	0.79	80.0	-1.08	2.0
Minor roads (100m)	km	3.94	1.52	0.01	0.80	0.26	1.04	1.1
Model: Suburban								
Intercept		5.55	0.80	<0.01				
Annual OMI NO_2	ppb	0.82	0.09	<0.01	0.49	3.37	3.82	1.4
Impervious (800m)	%	0.05	0.02	0.01	0.60	26.8	3.34	1.7
Major roads (200m)	km	3.50	0.81	<0.01	0.65	0.17	0.64	1.1
Tree canopy (8000m)	%	-0.07	0.02	<0.01	0.68	16.0	-1.46	1.1
Elevation	km	2.96	0.91	<0.01	0.70	0.25	0.61	1.1
Major roads (10000m)	km	7.97E-03	2.36E-03	<0.01	0.72	221	1.76	1.8
Model: Rural								
Intercept		3.17	0.46	<0.01				
Impervious (200m)	%	0.10	0.04	0.03	0.45	8.00	2.71	1.7
Annual OMI NO_2	ppb	1.01	0.22	<0.01	0.52	1.61	1.37	1.9
Canopy (1000m)	%	-0.06	0.02	<0.01	0.60	22.2	-1.59	1.2
Major roads (400m)	km	1.76	0.45	<0.01	0.64	0.14	0.21	1.1
Population (10000m)	#	4.80	1.45	<0.01	0.68	0.13	0.61	2.1

Distance in () is the buffer radius; parameters without a buffer distance were taken at the station locations. IQR is the inter-quartile range for the given parameter, *β* * IQR is the *β* coefficient multiplied by the IQR. VIF is the variance inflation factor to check for multicollinearity.