

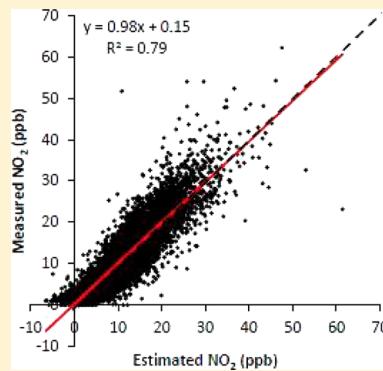
## Daily Ambient NO<sub>2</sub> Concentration Predictions Using Satellite Ozone Monitoring Instrument NO<sub>2</sub> Data and Land Use Regression

Hyung Joo Lee\* and Petros Koutrakis

Exposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard School of Public Health, 401 Park Drive, Landmark Center West Room 417, Boston, Massachusetts 02215, United States

### S Supporting Information

**ABSTRACT:** Although ground measurements have contributed to revealing the association between ambient air pollution and health effects in epidemiological studies, exposure measurement errors are likely to be caused because of the sparse spatial distribution of ground monitors. In this study, we estimate daily ground NO<sub>2</sub> concentrations in the New England region, U.S., for the period 2005–2010 using satellite remote sensing data in combination with land use regression. To estimate ground-level NO<sub>2</sub> concentrations, we constructed a mixed effects model by taking advantage of spatial and temporal variability in satellite Ozone Monitoring Instrument (OMI) tropospheric column NO<sub>2</sub> densities. Using fine-scale land use parameters, we derived NO<sub>2</sub> concentrations at point locations, which can be further used for subject-specific exposure estimates in epidemiological studies. A mixed effects model showed a reasonably high predictive power for daily NO<sub>2</sub> concentrations (cross-validation  $R^2 = 0.79$ ). We observed that the model performed similarly in each season, year, and state. The spatial patterns of model estimates reflected emission source areas (such as high populated/traffic areas) in the study region and revealed the seasonal characteristics of NO<sub>2</sub>. This study suggests that a combination of satellite remote sensing and land use regression can be useful for both spatially and temporally resolved exposure assessments of NO<sub>2</sub>.



### INTRODUCTION

Numerous studies have suggested that nitrogen dioxide (NO<sub>2</sub>) exposures are associated with adverse respiratory and cardiovascular health effects.<sup>1–4</sup> In epidemiological studies, NO<sub>2</sub> exposures are assessed by measuring ambient NO<sub>2</sub> concentrations at central monitoring sites or estimating the concentrations using various modeling techniques. Among the modeling procedures, land use regression (LUR) has been widely used to assess fine-scale long-term exposures to ambient air pollutants, including but not limited to NO<sub>2</sub> in areas between ground monitors.<sup>5,6</sup> More recently, satellite remote sensing data have been increasingly applied to estimate ambient pollutant concentrations by taking advantage of their expanded spatial monitoring coverage.<sup>7–10</sup>

Because of the limited number of central monitors, assessing NO<sub>2</sub> exposures using ground measurements is likely to introduce measurement errors. These sparsely distributed monitors may not capture the spatial gradients of the NO<sub>2</sub> concentration in regions without monitors. While the LUR models generate spatially resolved exposure estimates, they can only provide long-term average exposure estimates and are, thus, unable to indicate temporal variability in NO<sub>2</sub> concentrations. Although the spatial resolution of satellite data is coarser than that of land use data,<sup>11,12</sup> satellite remote sensing can provide both spatial and temporal variability in ambient NO<sub>2</sub> levels, complementing ground measurements and

LUR modeling and, thus, potentially improving exposure assessment for epidemiological studies.

In an effort to estimate daily satellite-based air pollutant concentrations, Lee et al. developed a mixed effects model to enhance the predictability of satellite moderate resolution imaging spectroradiometer (MODIS) aerosol optical depth (AOD) for ground PM<sub>2.5</sub> concentrations.<sup>8</sup> Subsequently, various studies have used a mixed effects model and provided model estimates for use with health effect studies.<sup>13–17</sup>

In this study, we estimated daily ambient NO<sub>2</sub> concentrations using a mixed effects model in combination with satellite Ozone Monitoring Instrument (OMI) NO<sub>2</sub> and LUR in the New England region, U.S., for the period 2005–2010. These estimated NO<sub>2</sub> concentrations were compared to measured NO<sub>2</sub> concentrations to evaluate model performance. Also, the spatial patterns of estimated NO<sub>2</sub> concentrations were investigated to examine if the concentration patterns reflected emission source areas and seasonal pollutant characteristics in the study region.

Received: October 30, 2013

Revised: January 6, 2014

Accepted: January 17, 2014

Published: January 17, 2014



## MATERIALS AND METHODS

**Ground-Level NO<sub>2</sub> Data.** Hourly NO<sub>2</sub> concentrations [in units of parts per billion (ppb)] were measured at 22 United States Environmental Protection Agency (U.S. EPA) monitoring sites located in the states of Massachusetts, Connecticut, and Rhode Island for the period 2005–2010. Among those 22 monitoring sites, there were 14, 5, and 3 spatial sites in Massachusetts, Connecticut, and Rhode Island, respectively. Although the chemiluminescence NO<sub>2</sub> measurements performed by the U.S. EPA might be biased because of non-NO<sub>x</sub> reactive nitrogen species (NO<sub>x</sub>),<sup>18</sup> these reported concentrations are widely used in regulatory processes and research associated with NO<sub>2</sub>. Hourly concentrations were averaged to calculate daily (24 h) NO<sub>2</sub> concentrations when at least 75% of hourly measurements were available for a site on a given day. This data completeness was expected to reduce the bias, potentially caused by missing hourly data with a diurnal variation, in daily average NO<sub>2</sub> concentrations. We used daily NO<sub>2</sub> concentrations because these are typically used for epidemiological studies.

**Satellite OMI NO<sub>2</sub>.** The standard product of tropospheric OMI NO<sub>2</sub> vertical column density data was obtained from the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) Aura satellite in Massachusetts, Connecticut, and Rhode Island from 2005 to 2010. The Aura satellite, launched in July 2004, operates at an altitude of 705 km and crosses the equator at approximately 1:45 pm local time. OMI aboard the Aura satellite provides column NO<sub>2</sub> information (in units of molecules/cm<sup>2</sup>) at a nominal resolution of 13 × 24 km (at nadir) on a global scale every day. The satellite data retrieval algorithm first obtains total slant column NO<sub>2</sub> densities and then subtracts stratospheric NO<sub>2</sub> from the total column NO<sub>2</sub>, which produces tropospheric slant NO<sub>2</sub> densities. Finally, tropospheric vertical column NO<sub>2</sub> densities are derived by applying conversion factors, called air mass factors (AMFs), to tropospheric slant NO<sub>2</sub> densities. Because of retrieval uncertainty in each of these steps, negative tropospheric vertical column NO<sub>2</sub> values are occasionally generated. To avoid biasing the distribution, these negative NO<sub>2</sub> values were included in our analysis. The estimated retrieval uncertainty with cloud fraction less than 0.5 is reported to be approximately 30% in urban areas.<sup>19</sup> In our study, tropospheric OMI NO<sub>2</sub> data with cloud fraction less than 0.3 were used, resulting in 55.5% of total OMI NO<sub>2</sub> data remaining for the further analysis. Pixels with OMI row anomaly<sup>20</sup> were excluded, and the median OMI pixel size used in the modeling procedure was 534 km<sup>2</sup>. More details about the OMI sensor and NO<sub>2</sub> data retrievals can be found in the studies by Levelt et al., Bucsela et al., and Celarier et al.<sup>21–23</sup>

**Land Use and Meteorology Parameters.** We considered spatial predictors of population density (U.S. Census), distance to major highways (class 1 roads, ESRI), percent developed area (U.S. Geological Survey, 30 m resolution), NO<sub>x</sub> source emissions (U.S. EPA), and elevation (National Elevation Data set, 1 arc second resolution). To extract land use information, we created buffers with various size radii (i.e., 125, 250, 500, 1000, 2000, and 5000 m) around each NO<sub>2</sub> monitoring site. Among these buffer radii, one of them, showing the highest correlation with the measured NO<sub>2</sub> concentrations across all of the monitoring sites, was chosen for each land use variable. It is noted that point estimates such as elevation and distance to major highways were independent of the buffer size radii. To

minimize confounding by collinearity (correlation coefficient  $r \approx 1$ ) between the different continuous predictor variables, we first estimated correlations between these variables. For all pairs of variables with a correlation coefficient larger than 0.60, only the variable that showed the stronger correlation with the measured NO<sub>2</sub> concentrations was selected for the modeling procedure.

Previous studies have reported that traffic-related pollution strongly affected nearby areas within several hundreds of meters from major roads.<sup>24</sup> To reflect the disproportionately decreasing impact of traffic emissions by distance from highways,<sup>25,26</sup> we categorized the distance to major highways into three groups (distance < 0.6 km, 0.6 ≤ distance < 3 km, and distance ≥ 3 km) with approximately the same number of observations in each group. This was to reflect traffic pollution as efficiently as possible and, at the same time, keep numerical stability (i.e., separating observations fairly equally into each group) as a categorical variable. In addition, the variable of elevation included comparatively high values (>300 m) for two monitoring sites. To minimize the potential bias of elevation on the NO<sub>2</sub> concentration, we recoded elevation as a categorical variable (elevation < 6 m, 6 ≤ elevation < 15 m, 15 ≤ elevation < 79 m, and elevation ≥ 79 m). Each of these four subgroups included a similar number of observations. Finally, we included the following spatial predictor variables for the modeling procedure: (1) population density (5 km), (2) percent developed area (1 km), (3) categorical distance to major highways, and (4) categorical elevation.

In addition, we included meteorological parameters (temperature and wind speed, National Oceanic and Atmospheric Administration), observed at the nearest weather station from each NO<sub>2</sub> monitoring site. There were 15 weather stations operated in the study region, and we selected weather stations with at least 95% completeness of data. When the stations had missing data, we used data from the second closest station. The average distance between weather stations and NO<sub>2</sub> monitoring sites was 23 km.

**Statistical Model.** A mixed effects model was constructed to calibrate tropospheric NO<sub>2</sub> vertical column densities for ground-level NO<sub>2</sub> concentrations by years and season. Years were split into 2005–2007 and 2008–2010. Seasons were categorized as warm (April 15–October 14) and cold (October 15–April 14). This led us to develop four different mixed effects models for the study period (i.e., cold for 2005–2007, warm for 2005–2007, cold for 2008–2010, and warm for 2008–2010). In addition to satellite NO<sub>2</sub>, the mixed effects model included land use variables and meteorological parameters (temperature and wind speed) as follows:

$$\begin{aligned} \text{NO}_{2ij} = & (\alpha + u_j) + (\beta_1 + v_j)\text{OMI NO}_{2ij} + \beta_2\text{Temp}_{ij} \\ & + \beta_3\text{WS}_{ij} + \sum_{m=1}^4 \beta_{1m}X_{im} + \varepsilon_{ij} (u_j, v_j) \sim N[(00), \Sigma] \end{aligned} \quad (1)$$

where NO<sub>2ij</sub> is the NO<sub>2</sub> concentration measured at a monitoring site  $i$  on day  $j$ , OMI NO<sub>2ij</sub> is the OMI tropospheric NO<sub>2</sub> vertical column density in the grid cell corresponding to site  $i$  on day  $j$ ,  $\alpha$  and  $u_j$  are the fixed and random intercepts, respectively,  $\beta_1$  and  $v_j$  are the fixed and random slopes of OMI NO<sub>2</sub>, respectively, Temp<sub>ij</sub> and WS<sub>ij</sub> are the temperature and wind speed observed at the nearest weather station from site  $i$  on day  $j$ , respectively,  $\beta_2$  and  $\beta_3$  are the fixed slopes of temperature and wind speed, respectively,  $X_{im}$  is  $m$ th spatial

predictor [among population density (5 km), percent developed area (1 km), distance to major highways (categorical), and elevation (point, categorical)] at site  $i$ ,  $\varepsilon_{ij} \sim N(0, \sigma^2)$  is the error term at site  $i$  on day  $j$ , and  $\Sigma$  is the variance–covariance matrix for the day-specific random effects. The values of all of the variables used in the modeling procedure are described in Table S1 of the Supporting Information.

The fixed intercept and slope of OMI NO<sub>2</sub> indicated the average relationship of OMI NO<sub>2</sub> on the ground NO<sub>2</sub> concentration over the study period. Also, the random intercept and slope of OMI NO<sub>2</sub> represented the day-specific relationships, which might be attributed to daily varying vertical profile of NO<sub>2</sub>, diurnal profile of NO<sub>2</sub>, or other parameters.

The land use terms explained site-specific characteristics affecting NO<sub>2</sub> concentrations on a fine scale. These spatial parameters played a role as surrogates of local emission sources (e.g., traffic, home heating, and any other human activities) and reflected the vertical extent of ground-level emissions (e.g., elevation). The high resolution of geographic variables was especially useful to estimate the NO<sub>2</sub> concentration at point locations, leading to more reliable subject-specific exposure estimates. In addition to the land use terms, temperature and wind speed were expected to further help explain the temporal and spatial variability of NO<sub>2</sub> concentrations. The formation and depletion of NO<sub>2</sub> depend upon the temperature, presumably correlated with solar radiation, and, thus, atmospheric photochemical activity.<sup>27</sup> An increase in wind speed accelerates atmospheric mixing of ground-level pollutants, which decreases ambient NO<sub>2</sub> levels.

**Model Validation.** The mixed effects model was validated using a 10-fold cross-validation (CV) technique. To avoid potential overfitting, we split observations into 10 randomly selected subgroups, fitted a model with 9 subgroups (calibration group), and then estimated NO<sub>2</sub> for a remaining subgroup (test group). The CV model estimates are less biased in comparison to model fit estimates. The performance of the mixed effects model was assessed by comparing the estimated NO<sub>2</sub> concentrations to the measured ones for both original and cross-validated models. The coefficient of determination ( $R^2$ ), bias, and absolute (ppb) and relative (%) root-mean-square error (RMSE) between the measured and estimated NO<sub>2</sub> concentrations were used to determine whether the model estimates were accurate enough to be reliably applied to epidemiological studies. The RMSE was calculated as the square root of the mean of the squared errors, and % RMSE was estimated as  $[100 \times (\text{RMSE}/\text{measured mean NO}_2 \text{ concentration})]$ . The measured mean NO<sub>2</sub> concentration values were based on all of the site measurements in each season, year, or state.

## RESULTS AND DISCUSSION

**Descriptive Statistics.** Table 1 shows summary statistics of NO<sub>2</sub> concentrations measured at ground-level monitors and tropospheric column NO<sub>2</sub> densities retrieved from the satellite Aura OMI. The average NO<sub>2</sub> concentration measured at 22 U.S. EPA monitoring sites for the period 2005–2010 was 11.6 ppb [standard deviation (SD) = 8.1 ppb]. On average, the NO<sub>2</sub> concentration during the cold season (14.7 ppb, SD = 8.8 ppb) was higher than that during the warm season (9.1 ppb, SD = 6.4 ppb). These NO<sub>2</sub> concentrations also presented a downward trend through the years. The average tropospheric OMI NO<sub>2</sub> in the grid cells corresponding to the ground

**Table 1. Summary Statistics (Mean  $\pm$  SD) of Ground NO<sub>2</sub> Measurements and Satellite OMI NO<sub>2</sub> Densities for the Period 2005–2010**

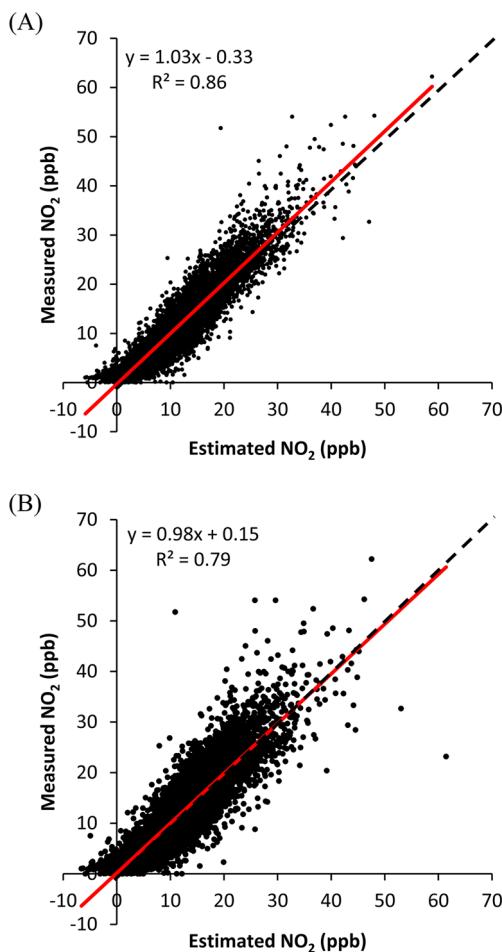
	NO <sub>2</sub> concentration (ppb)	OMI NO <sub>2</sub> ( $\times 10^{15}$ , molecules/cm <sup>2</sup> )
overall	11.6 $\pm$ 8.1	4.9 $\pm$ 5.0
season		
cold	14.7 $\pm$ 8.8	7.5 $\pm$ 6.5
warm	9.1 $\pm$ 6.4	3.4 $\pm$ 2.9
year		
2005	13.5 $\pm$ 9.0	5.7 $\pm$ 5.1
2006	12.2 $\pm$ 8.3	5.7 $\pm$ 6.0
2007	12.1 $\pm$ 8.0	4.9 $\pm$ 5.0
2008	11.6 $\pm$ 7.9	4.9 $\pm$ 4.8
2009	10.5 $\pm$ 7.6	4.2 $\pm$ 4.1
2010	9.9 $\pm$ 7.0	3.6 $\pm$ 3.7

monitoring sites was  $4.9 \times 10^{15}$  (SD =  $5.0 \times 10^{15}$ ) molecules/cm<sup>2</sup> for the same study period. The OMI NO<sub>2</sub> also varied by season, showing higher NO<sub>2</sub> during the cold season ( $7.5 \times 10^{15}$  molecules/cm<sup>2</sup>, SD =  $6.5 \times 10^{15}$ ) than the warm season ( $3.4 \times 10^{15}$  molecules/cm<sup>2</sup>, SD =  $2.9 \times 10^{15}$ ). The satellite OMI NO<sub>2</sub> densities showed a downward trend through the years as well.

**NO<sub>2</sub> Prediction.** The mixed effects model generated daily relationships between ground NO<sub>2</sub> measurements and satellite tropospheric OMI NO<sub>2</sub> for 1313 days from 2005 to 2010 (60% of the 6 year period). The number of day-specific relationships was fairly consistent over the years (mean = 219, and range = 180–250). There were more daily relationships during the warm season (736) compared to the cold season (577). The total number of observations used in the model (i.e., pairs between the measured NO<sub>2</sub> and all of the predictor variables across all monitoring sites for 2005–2010) was 5359 and 3280 during the warm and cold seasons, respectively.

Figure 1 presents the comparison between the measured and estimated NO<sub>2</sub> concentrations across all monitoring sites for 2005–2010. The mixed effects model explained 86% of variability in daily NO<sub>2</sub> concentrations ( $R^2 = 0.86$ ), showing a good agreement between the measured and estimated NO<sub>2</sub> concentrations {slope = 1.03 [standard error (SE) = 0.004] and intercept = -0.33 (SE = 0.06)}. Also, the CV mixed effects model presented similar predictive power [CV  $R^2 = 0.79$ , slope = 0.98 (SE = 0.005), and intercept = 0.15 (SE = 0.07)]. Both models demonstrated reasonably high predictability for NO<sub>2</sub> concentrations, confirming further that model overfitting was not substantial. As a sensitivity analysis, we found that 4-fold CV (i.e., 25% of observations as a subgroup; 4 subgroups) resulted in a similar predictive power of NO<sub>2</sub> concentrations (CV  $R^2 = 0.79$ ).

The comparison between our mixed effects model (fixed and random effects of OMI NO<sub>2</sub>) and a multivariate regression model (fixed effect of OMI NO<sub>2</sub>) showed a higher predictive power of NO<sub>2</sub> concentrations from the mixed effects model [CV  $R^2 = 0.79$ , and RMSE = 3.59 ppb (31.4%)] relative to the multivariate regression model [CV  $R^2 = 0.69$ , and RMSE = 4.39 ppb (38.4%)]. This indicates that the random effect of OMI NO<sub>2</sub> was useful to explain the variability in daily NO<sub>2</sub> concentrations. Also, the model performance with land use parameters and meteorology (i.e., no satellite data included) [CV  $R^2 = 0.62$ , and RMSE = 4.83 ppb (42.2%)] was lower than the NO<sub>2</sub> predictability from the mixed effects model with satellite OMI NO<sub>2</sub> data, demonstrating that OMI NO<sub>2</sub> data



**Figure 1.** Comparison between the measured and estimated  $\text{NO}_2$  concentrations: (A) mixed effects model and (B) CV mixed effects model. The solid and dashed lines represent the regression and 1:1 lines, respectively.

enabled us to estimate more reliable daily  $\text{NO}_2$  concentrations. Furthermore, our years- and season-specific mixed effects models slightly outperformed a combined mixed effects model (i.e., a single model for the entire study period) [ $\text{CV } R^2 = 0.77$ , and RMSE = 3.76 ppb (32.9%)].

Table 2A shows  $\text{NO}_2$  predictability obtained by the mixed effects model for warm and cold seasons. The mixed effects model for each season well-explained the variability in  $\text{NO}_2$  concentrations, with a high  $R^2$  value and good agreement between the measured and estimated concentrations (slope

close to unity and intercept close to zero). The predictive power for  $\text{NO}_2$  concentrations was slightly higher during the cold season ( $\text{CV } R^2 = 0.80$ ) than during the warm season ( $\text{CV } R^2 = 0.75$ ). The absolute RMSE was smaller during the warm season (3.34 ppb) compared to the cold season (3.97 ppb). However, because of the higher average  $\text{NO}_2$  concentration during the cold season, the relative RMSE during the cold season (27.6%) was smaller than that during the warm season (34.7%).

Although the  $\text{NO}_2$  predictability was not substantially different between two seasons, it is worth discussing potential factors influencing the seasonally varying predictability. First, the effectiveness of land use variables to explain  $\text{NO}_2$  variability varies by season because there are different types of emission sources in each season. As indicated by Novotny et al., the predictability is based only on land use variables included in the model.<sup>10</sup> Moreover, any land use variables that are more closely related to source emissions only for one of the seasons can disproportionately affect the model predictability in each season. For example, population density can reflect the spatial distribution of home heating more strongly during the cold season (as shown by Table S2 of the Supporting Information), while other human activities related to population density may be independent of season. Second, higher temperature and solar radiation during the warm season can shorten the lifetime of atmospheric  $\text{NO}_2$  because of the higher photolysis rate of  $\text{NO}_2$  and subsequent  $\text{O}_3$  formation.<sup>28</sup> Third,  $\text{NO}_2$  can be enhanced because of lightning in the upper troposphere, particularly during the summer,<sup>29</sup> which makes the contribution of  $\text{NO}_2$  in the planetary boundary layer to the total or tropospheric column  $\text{NO}_2$  smaller during the season. Further, higher emissions of biogenic volatile organic compounds (VOCs) during the warm season may affect the  $\text{NO}-\text{NO}_2-\text{O}_3$  reactions and, thus, may affect the seasonal predictability of  $\text{NO}_2$  concentrations.<sup>27,30</sup>

Table 2B represents annual predictability of  $\text{NO}_2$  concentrations using the mixed effects model. Note that these annual results were based on combined years- and season-specific mixed effects model estimates (four separate models, eq 1). We found reasonably high predictive power of the mixed effects model for  $\text{NO}_2$  concentrations throughout the years. The variability in  $\text{NO}_2$  concentrations explained by the CV model ( $\text{CV } R^2$ ) ranged from 0.75 in 2010 to 0.81 in 2006, and there was a good agreement between the measured and estimated  $\text{NO}_2$  concentrations in each year. In addition, the absolute (3.22–4.16 ppb) and relative (29.2–35.4%) RMSE values were fairly consistent over the years. The decreasing  $\text{NO}_2$

**Table 2. Model Performance of CV Mixed Effects Model for Ground-Level  $\text{NO}_2$  Concentrations by (A) Season and (B) Year<sup>a</sup>**

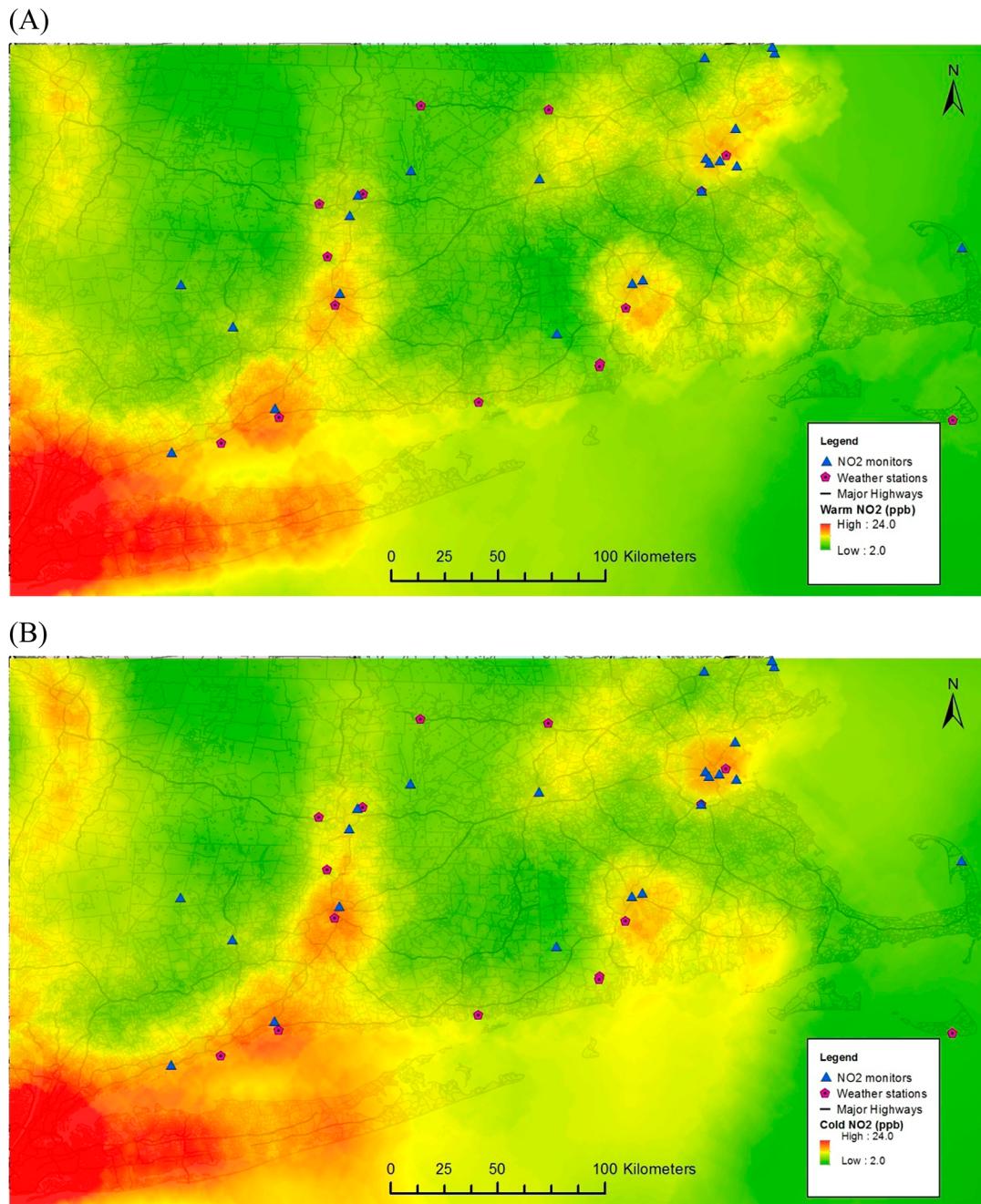
	N	$\text{NO}_2$ measured	$\text{NO}_2$ estimated	bias	RMSE (%)	$R^2$	slope (SE)	intercept (SE)
<b>(A) Season</b>								
cold	3280	14.39	14.43	0.04	3.97 (27.6)	0.80	0.98 (0.01)	0.22 (0.14)
warm	5359	9.63	9.65	0.02	3.34 (34.7)	0.75	0.98 (0.01)	0.13 (0.09)
<b>(B) Year</b>								
2005	1630	13.35	13.20	0.14	4.16 (31.2)	0.78	1.02 (0.01)	-0.12 (0.21)
2006	1519	12.22	12.31	0.09	3.73 (30.5)	0.81	0.96 (0.01)	0.42 (0.18)
2007	1747	11.43	11.59	0.16	3.34 (29.2)	0.80	0.96 (0.01)	0.32 (0.16)
2008	1474	11.01	10.95	0.06	3.47 (31.5)	0.79	1.01 (0.01)	0.01 (0.17)
2009	1082	10.10	10.18	0.08	3.22 (31.9)	0.79	0.99 (0.02)	-0.03 (0.19)
2010	1187	9.56	9.63	0.07	3.38 (35.4)	0.75	0.95 (0.02)	0.38 (0.18)

<sup>a</sup>The measured and estimated  $\text{NO}_2$  concentrations, bias, and RMSE are in units of ppb.

Table 3. State-Specific Model Performance of CV Mixed Effects Model<sup>a</sup>

	N	NO <sub>2</sub> measured	NO <sub>2</sub> estimated	bias	RMSE (%)	R <sup>2</sup>	slope (SE)	intercept (SE)
state								
Massachusetts	5734	11.11	11.00	0.10	3.37 (30.3)	0.82	0.98 (0.01)	0.34 (0.08)
Connecticut	2076	12.95	12.85	0.10	4.38 (33.8)	0.70	0.99 (0.01)	0.19 (0.21)
Rhode Island	829	9.91	11.18	1.27	2.79 (28.1)	0.87	1.00 (0.01)	-1.25 (0.17)

<sup>a</sup>The measured and estimated NO<sub>2</sub> concentrations, bias, and RMSE are in units of ppb.

Figure 2. Spatial patterns of estimated NO<sub>2</sub> concentrations (ppb) by season: (A) warm season and (B) cold season.

concentration trends for the study period imply that the emission mitigation strategies and economic recession were likely to have changed source emission patterns (i.e., the number of sources and the emission rates), which may affect the model performance differently in each year.<sup>31,32</sup> The model estimates from the years- and season-specific models (eq 1)

reflected such potential changes in emission patterns, as shown by Table S2 of the Supporting Information.

Table 3 shows a state-specific comparison between the measured and estimated NO<sub>2</sub> concentrations. Overall, the model performance for three states was reasonably high (CV R<sup>2</sup> = 0.70–0.87, and % RMSE = 28.1–33.8%). In comparison to

the other two states, the model predictive power was slightly lower in Connecticut. This may be attributed to a stronger impact of transported pollution from metropolitan New York City on NO<sub>2</sub> concentrations observed in Connecticut because of close geographic proximity. While the land use variables are expected to capture local source emissions, the OMI NO<sub>2</sub> reflects a combination of daily local and transported pollution. The OMI NO<sub>2</sub> data represent area-averaged NO<sub>2</sub> levels (13 × 24 km at nadir), therefore, likely underestimating the NO<sub>2</sub> concentrations at spatial sites strongly affected by transported pollution. When a monitoring site is influenced by strong local emissions, NO<sub>2</sub> concentrations also tend to be underestimated at the site. Because the model was less likely to be robust at some locations particularly with comparatively high and low values for predictor variables (such as high elevation and near roadways), the extent of model robustness for the locations may also explain the state-specific predictive power of NO<sub>2</sub>. Despite the differences in relative amounts of local and transported contributions to NO<sub>2</sub> concentrations and model robustness in three states, each state has similar types of local combustion sources and meteorology, resulting in fairly comparable model predictability.

The model prediction errors may be caused by satellite NO<sub>2</sub> retrieval errors, ground NO<sub>2</sub> measurement errors, and temporal inconsistency between daily 24 h average NO<sub>2</sub> measurements and tropospheric OMI NO<sub>2</sub> column densities retrieved at around 1:45 pm local time. To examine the effect of temporal difference on NO<sub>2</sub> predictability, we first calculated correlation coefficients between 24 and 2 h (1–3 pm) average NO<sub>2</sub> concentrations and OMI NO<sub>2</sub>. The correlation coefficients were similar but slightly higher for the comparison between 24 h average NO<sub>2</sub> concentrations and tropospheric OMI NO<sub>2</sub> densities ( $r = 0.50$  versus 0.44). Furthermore, this led to higher predictive power of the mixed effects model using daily average NO<sub>2</sub> concentrations (CV  $R^2 = 0.79$  versus 0.63). This may be attributed to the fact that the relatively short (2 h) observation period reflects rather transient phenomena. Hence, averaging longer temporal window for NO<sub>2</sub> measurements can be more closely related to OMI NO<sub>2</sub> considering the lifetime and transport of NO<sub>2</sub> and the spatial resolution of OMI NO<sub>2</sub> data. In addition to the relationship between measured NO<sub>2</sub> concentrations and OMI NO<sub>2</sub>, the higher predictability using daily average NO<sub>2</sub> concentrations may be because the daily averages are more representative of the land use parameters, which are temporally constant.

**Spatial Patterns of NO<sub>2</sub>.** Figure 2 illustrates the season-specific spatial patterns of NO<sub>2</sub> concentrations estimated by the mixed effects model. Both spatial patterns showed higher NO<sub>2</sub> concentrations in urban areas, reflecting more anthropogenic combustion sources compared to rural areas. The large difference in NO<sub>2</sub> concentrations between urban and rural areas suggests that ground-level NO<sub>2</sub> measurements at central monitoring sites mostly in urban areas are unlikely to be representative of NO<sub>2</sub> levels in rural areas while overestimating subjects' exposures to NO<sub>2</sub> in the areas. With regard to emission sources, high NO<sub>x</sub> line (e.g., highways) and point (latitude and longitude from National Emissions Inventory) (e.g., power plants) source areas in the region were pronounced in both concentration maps,<sup>33</sup> and the spatial maps also showed NO<sub>2</sub> transport from metropolitan New York City. The difference between two seasonal concentration patterns needs to be noted. The urban increment of NO<sub>2</sub> was larger during the cold season than during the warm season, likely because of NO<sub>2</sub>

emissions from space heating and lower boundary layer height during the cold season.

The spatial precision of model estimates depends upon the selected predictor variables (i.e., buffer size and category) in the modeling procedure. Considering the buffer size and category in the mixed effects models, our model estimates may be less spatially detailed in comparison to conventional LUR for a single urban area. Figure S1 of the Supporting Information presents the local scale variability of estimated NO<sub>2</sub> concentrations using axes,<sup>34</sup> which helps understand the spatial precision of NO<sub>2</sub> in the Boston urban area. Also, the spatial patterns of estimated NO<sub>2</sub> concentrations (Figure 2) are more likely to be representative of days without clouds, although we did not find a significant difference in the measured NO<sub>2</sub> concentrations between days excluded because of OMI cloud fraction and those included in the mixed effects models in each season.

In conclusion, our mixed effects model enhanced the predictive power of daily ground-level NO<sub>2</sub> concentrations using satellite remote sensing data in combination with fine-scale LUR. The model provided location-specific NO<sub>2</sub> concentrations in both urban and rural areas. Improved daily NO<sub>2</sub> exposure estimates will be important for epidemiological studies investigating acute and chronic health effects associated with ambient NO<sub>2</sub> concentrations because they can reduce exposure measurement errors. Furthermore, these estimated NO<sub>2</sub> concentrations can be used to investigate the atmospheric photochemistry of nitrogen oxides in the context of climate change. While our statistical model showed reasonably high predictability of ground NO<sub>2</sub> concentrations in the New England region, the modeling technique needs to be applied to other regions to be more generalized. In the near future, when the spatial resolution of OMI NO<sub>2</sub> becomes higher, the satellite-based exposure modeling is expected to further enhance the understanding of subjects' exposures to NO<sub>2</sub>.

## ASSOCIATED CONTENT

### Supporting Information

Descriptive statistics of dependent and independent variables (pairs between the measured NO<sub>2</sub> and all of the predictor variables across all monitoring sites for the period 2005–2010) used in the mixed effects model (Table S1), coefficients of four separate mixed effects models (fixed effects of predictor variables) (Table S2), and estimated NO<sub>2</sub> concentrations (ppb) for the period 2005–2010 (Figure S1). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Telephone: +1-617-384-8749. Fax: +1-617-384-8833. E-mail: hlee@hsph.harvard.edu.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank the Harvard Clean Air Research Center. This publication was made possible by U.S. EPA Grant RD 83479801. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the U.S. EPA. Further, the U.S. EPA does not endorse the purchase of any commercial products or services mentioned in the publication.

## ■ REFERENCES

- (1) Chiusolo, M.; Cadum, E.; Stafoggia, M.; Galassi, C.; Berti, G.; Faustini, A.; Bisanti, L.; Vigotti, M. A.; Dessi, M. P.; Cerniglio, A.; Mallone, S.; Pacelli, B.; Minerba, S.; Simonato, L.; Forastiere, F. Short-term effects of nitrogen dioxide on mortality and susceptibility factors in 10 Italian cities: The EpiAir study. *Environ. Health Perspect.* **2011**, *119*, 1233–1238.
- (2) Esplugues, A.; Ballester, F.; Estarlich, M.; Llop, S.; Fuentes-Leonarte, V.; Mantilla, E.; Vioque, J.; Iniguez, C. Outdoor, but not indoor, nitrogen dioxide exposure is associated with persistent cough during the first year of life. *Sci. Total Environ.* **2011**, *409*, 4667–4673.
- (3) Gauderman, W. J.; Avol, E.; Lurmann, F.; Kuenzli, N.; Gilliland, F.; Peters, J.; McConnell, R. Childhood asthma and exposure to traffic and nitrogen dioxide. *Epidemiology* **2005**, *16*, 737–743.
- (4) Samoli, E.; Aga, E.; Touloumi, G.; Nislotis, K.; Forsberg, B.; Lefranc, A.; Pekkanen, J.; Wojtyniak, B.; Schindler, C.; Nicu, E.; Brunstein, R.; Fikfak, M. D.; Schwartz, J.; Katsouyanni, K. Short-term effects of nitrogen dioxide on mortality: An analysis within the APHEA project. *Eur. Respir. J.* **2006**, *27*, 1129–1137.
- (5) Ross, Z.; English, P. B.; Scalf, R.; Gunier, R.; Smorodinsky, S.; Wall, S.; Jerrett, M. Nitrogen dioxide prediction in southern California using land use regression modeling: Potential for environmental health analyses. *J. Exposure Sci. Environ. Epidemiol.* **2006**, *16*, 106–114.
- (6) Wheeler, A. J.; Smith-Doiron, M.; Xu, X.; Gilbert, N. L.; Brook, J. R. Intra-urban variability of air pollution in Windsor, Ontario—Measurement and modeling for human exposure assessment. *Environ. Res.* **2008**, *106*, 7–16.
- (7) Lamsal, L. N.; Martin, R. V.; van Donkelaar, A.; Steinbacher, M.; Celarier, E. A.; Bucsel, 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*, No. D16308.
- (8) Lee, H. J.; Liu, Y.; Coull, B. A.; Schwartz, J.; Koutrakis, P. A novel calibration approach of MODIS AOD data to predict PM<sub>2.5</sub> concentrations. *Atmos. Chem. Phys.* **2011**, *11*, 7991–8002.
- (9) Liu, Y.; Paciorek, C. J.; Koutrakis, P. Estimating regional spatial and temporal variability of PM<sub>2.5</sub> concentrations using satellite data, meteorology, and land use information. *Environ. Health Perspect.* **2009**, *117*, 886–892.
- (10) Novotny, E. V.; Bechle, M. J.; Millet, D. B.; Marshall, J. D. National satellite-based land-use regression: NO<sub>2</sub> in the United States. *Environ. Sci. Technol.* **2011**, *45*, 4407–4414.
- (11) Martin, R. V. Satellite remote sensing of surface air quality. *Atmos. Environ.* **2008**, *42*, 7823–7843.
- (12) Valin, L. C.; Russell, A. R.; Hudman, R. C.; Cohen, R. C. Effects of model resolution on the interpretation of satellite NO<sub>2</sub> observations. *Atmos. Chem. Phys.* **2011**, *11*, 11647–11655.
- (13) Chudnovsky, A. A.; Lee, H. J.; Kostinski, A.; Kotlov, T.; Koutrakis, P. Prediction of daily fine particulate matter concentrations using aerosol optical depth retrievals from the Geostationary Operational Environmental Satellite (GOES). *J. Air Waste Manage. Assoc.* **2012**, *62*, 1022–1031.
- (14) Hyder, A.; Ebisu, K.; Lee, H. J.; Koutrakis, P.; Belanger, K.; Bell, M. L. Using satellite- and monitor-based data to assess the relationship between PM<sub>2.5</sub> exposure and birth outcomes in Connecticut and Massachusetts (2000–2006). *Epidemiology* **2014**, *25*, 58–67.
- (15) Kloog, I.; Koutrakis, P.; Coull, B. A.; Lee, H. J.; Schwartz, J. Assessing temporally and spatially resolved PM<sub>2.5</sub> exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ.* **2011**, *45*, 6267–6275.
- (16) Lee, H. J.; Coull, B. A.; Bell, M. L.; Koutrakis, P. Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM<sub>2.5</sub> concentrations. *Environ. Res.* **2012**, *118*, 8–15.
- (17) Yap, X. Q.; Hashim, M. A robust calibration approach for PM<sub>10</sub> prediction from MODIS aerosol optical depth. *Atmos. Chem. Phys.* **2013**, *13*, 3517–3526.
- (18) Dunlea, E. J.; Herndon, S. C.; Nelson, D. D.; Volkamer, R. M.; San Martini, F.; Sheehy, P. M.; Zahniser, M. S.; Shorter, J. H.; Wormhoudt, J. C.; Lamb, B. K.; Allwine, E. J.; Gaffney, J. S.; Marley, N. A.; Grutter, M.; Marquez, C.; Blanco, S.; Cardenas, B.; Retama, A.; Ramos Villegas, C. R.; Kolb, C. E.; Molina, L. T.; Molina, M. J. Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment. *Atmos. Chem. Phys.* **2007**, *7*, 2691–2704.
- (19) Lamsal, L. N.; Martin, R. V.; Parrish, D. D.; Krotkov, N. A. Scaling relationship for NO<sub>2</sub> pollution and urban population size: A satellite perspective. *Environ. Sci. Technol.* **2013**, *47*, 7855–7861.
- (20) National Aeronautics and Space Administration (NASA). *OMI Row Anomalies*; NASA: Washington, D.C., 2013; <http://disc.sci.gsfc.nasa.gov/Aura/data holdings/OMI> (accessed Dec 20, 2013).
- (21) Levelt, P. F.; van den Oord, G. H. J.; Dobber, M. R.; Malkki, A.; Visser, H.; de Vries, J.; Stammes, P.; Lundell, J. O. V.; Saari, H. The Ozone Monitoring Instrument. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1093–1101.
- (22) Bucsel, E. J.; Celarier, E. A.; Wenig, M. O.; Gleason, J. F.; Veefkind, J. P.; Boersma, K. F.; Brinksma, E. J. Algorithm for NO<sub>2</sub> vertical column retrieval from the ozone monitoring instrument. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1245–1258.
- (23) Celarier, E. A.; Brinksma, E. J.; Gleason, J. F.; Veefkind, J. P.; Cede, A.; Herman, J. R.; Ionov, D.; Goutail, F.; Pommereau, J. P.; Lambert, J. C.; van Roozendael, M.; Pinardi, G.; Wittrock, F.; Schonhardt, A.; Richter, A.; Ibrahim, O. W.; Wagner, T.; Bojkov, B.; Mount, G.; Spinei, E.; Chen, C. M.; Pongetti, T. J.; Sander, S. P.; Bucsel, E. J.; Wenig, M. O.; Swart, D. P. J.; Volten, H.; Kroon, M.; Levelt, P. F. Validation of ozone monitoring instrument nitrogen dioxide columns. *J. Geophys. Res.: Atmos.* **2008**, *113*, No. D15S15.
- (24) Hoek, G.; Beelen, R.; de Hoogh, 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.
- (25) Roorda-Knape, M. C.; Janssen, N. A. H.; de Hartog, J. J.; van Vliet, P. H. N.; Harssema, H.; Brunekreef, B. Air pollution from traffic in city districts near major motorways. *Atmos. Environ.* **1998**, *32*, 1921–1930.
- (26) Zhu, Y. F.; Hinds, W. C.; Kim, S.; Shen, S.; Sioutas, C. Study of ultrafine particles near a major highway with heavy-duty diesel traffic. *Atmos. Environ.* **2002**, *36*, 4323–4335.
- (27) Seinfeld, J. H.; Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*; John Wiley and Sons: Hoboken, NJ, 2006.
- (28) Trebs, I.; Bohn, B.; Ammann, C.; Rummel, U.; Blumthaler, M.; Konigstedt, R.; Meixner, F. X.; Fan, S.; Andreae, M. O. Relationship between the NO<sub>2</sub> photolysis frequency and the solar global irradiance. *Atmos. Meas. Tech.* **2009**, *2*, 725–739.
- (29) Sioris, C. E.; McLinden, C. A.; Martin, R. V.; Sauvage, B.; Haley, C. S.; Lloyd, N. D.; Llewellyn, E. J.; Bernath, P. F.; Boone, C. D.; Brohede, S.; McElroy, C. T. Vertical profiles of lightning-produced NO<sub>2</sub> enhancements in the upper troposphere observed by OSIRIS. *Atmos. Chem. Phys.* **2007**, *7*, 4281–4294.
- (30) Steiner, A. L.; Cohen, R. C.; Harley, R. A.; Tonse, S.; Millet, D. B.; Schade, G. W.; Goldstein, A. H. VOC reactivity in central California: Comparing an air quality model to ground-based measurements. *Atmos. Chem. Phys.* **2008**, *8*, 351–368.
- (31) Castellanos, P.; Boersma, K. F. Reductions in nitrogen oxides over Europe driven by environmental policy and economic recession. *Sci. Rep.* **2012**, *2*, 1–7.
- (32) Russell, A. R.; Valin, L. C.; Cohen, R. C. Trends in OMI NO<sub>2</sub> observations over the United States: Effects of emission control technology and the economic recession. *Atmos. Chem. Phys.* **2012**, *12*, 12197–12209.
- (33) United States Environmental Protection Agency (U.S. EPA). *National Emissions Inventory*; U.S. EPA: Washington, D.C., 2013; <http://www.epa.gov/ttn/chief/eiinformation.html> (accessed Aug 19, 2013).
- (34) Vienneau, D.; de Hoogh, K.; Bechle, M. J.; Beelen, R.; van Donkelaar, A.; Martin, R. V.; Millet, D. B.; Hoek, G.; Marshall, J. D. Western European land use regression incorporating satellite- and ground-based measurements of NO<sub>2</sub> and PM<sub>10</sub>. *Environ. Sci. Technol.* **2013**, *47*, 13555–13564.