



# EPA Public Access

Author manuscript

*Environ Sci Technol.* Author manuscript; available in PMC 2020 March 11.

About author manuscripts

Submit a manuscript

Published in final edited form as:

*Environ Sci Technol.* 2020 February 04; 54(3): 1372–1384. doi:10.1021/acs.est.9b03358.

## Assessing NO<sub>2</sub> Concentration and Model Uncertainty with High Spatiotemporal Resolution across the Contiguous United States Using Ensemble Model Averaging

Qian Di<sup>a,b,\*</sup>, Heresh Amini<sup>b</sup>, Lihua Shi<sup>b,c</sup>, Itai Kloog<sup>d</sup>, Rachel Silvern<sup>e</sup>, James Kelly<sup>f</sup>, M. Benjamin Sabath<sup>g</sup>, Christine Choirat<sup>g</sup>, Petros Koutrakis<sup>b</sup>, Alexei Lyapustin<sup>h</sup>, Yujie Wang<sup>i</sup>, Loretta J. Mickley<sup>j</sup>, Joel Schwartz<sup>b</sup>

<sup>a</sup>Research Center for Public Health, Tsinghua University, Beijing, China, 100084

<sup>b</sup>Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States, 02215

<sup>c</sup>Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States, 30322

<sup>d</sup>Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Beer Sheva, Israel, P.O.Box 653

<sup>e</sup>Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, United States, 02138

<sup>f</sup>U.S. Environmental Protection Agency, Office of Air Quality Planning & Standards, Research Triangle Park, North Carolina, United States, 27711

<sup>g</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States, 02115

<sup>h</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland, United States, 20771

<sup>i</sup>University of Maryland, Baltimore County, Baltimore, Maryland, United States, 21250

<sup>j</sup>John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge Massachusetts, United States, 02138

### Abstract

NO<sub>2</sub> is a combustion byproduct that has been associated with multiple adverse health outcomes.

To assess NO<sub>2</sub> level with high accuracy, we propose an ensemble model to integrate multiple machine learning algorithms, including neural network, random forest, and gradient boosting, with a variety of predictor variables, including chemical transport models. This NO<sub>2</sub> model covers the

\*Corresponding author: Qian Di (qiandi@mail.harvard.edu).

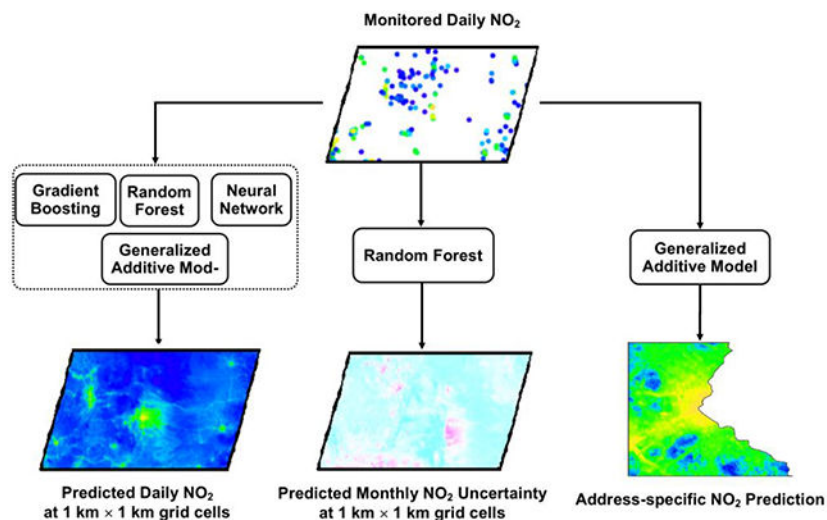
#### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website:

Detailed list of meteorological variables, details of land-cover variables used in the analysis, details of dealing with missing values, parameters tuned for base learners, model performance from Step 1 and Step2, contribution of predictor variables, model performance comparison between the ensemble model and base learners, flowchart of ensemble modeling, linearity between monitored NO<sub>2</sub> and predicted NO<sub>2</sub> at the annual level, downscaled NO<sub>2</sub> levels in the great Boston area, and long-term averages of PM<sub>2.5</sub>, NO<sub>2</sub> and ozone

entire contiguous U.S. with daily predictions on 1-km-level grid cells from 2000 to 2016. The ensemble produced a cross-validated  $R^2$  of 0.788 overall, a spatial  $R^2$  of 0.844, and a temporal  $R^2$  of 0.729. The relationship between daily monitored and predicted  $\text{NO}_2$  is almost linear. We also estimated the associated monthly uncertainty level for the predictions and address-specific  $\text{NO}_2$  levels. This  $\text{NO}_2$  estimation has a very high spatiotemporal resolution and allows the examination of health effects of  $\text{NO}_2$  in unmonitored areas. We found the highest  $\text{NO}_2$  levels along highways and in cities. We also observed that nationwide  $\text{NO}_2$  levels declined in early years and stagnated after 2007, in contrast to the trend at monitoring sites in urban areas, where the decline continued. Our research indicates that integrating different predictor variables and fitting algorithms can achieve an improved air pollution modeling framework.

## Graphical Abstract



## Keywords

$\text{NO}_2$ ; Ensemble Model; Machine Learning; Neural Network; Gradient Boosting; Random Forest

## 1. Introduction

$\text{NO}_2$ , or nitrogen dioxide, is a gaseous air pollutant, which can affect the respiratory system<sup>1</sup> by increasing susceptibility to respiratory infections<sup>2</sup>, exacerbating asthma symptoms<sup>3</sup>, and decreasing pulmonary function<sup>4</sup>. In addition to respiratory symptoms, evidence is mounting on the association of  $\text{NO}_2$  with low birth-weights<sup>5</sup>, cardiovascular diseases<sup>6</sup>, hospital admission, and mortality. For some health outcomes, the evidence is moderate<sup>7</sup>. Besides its direct health impacts,  $\text{NO}_2$  can mediate the formation of secondary organic aerosol from biogenic (e.g., terpenes) and anthropogenic (e.g., aromatics from vehicle exhaust) sources via reactions with organic gases and by influencing oxidant abundance<sup>8–10</sup>. It similarly drives reactions that produce the surface pollutant ozone.

$\text{NO}_2$  is an oxidative gas which reacts with other chemicals in the atmosphere. Mobile emissions are the major source of  $\text{NO}_2$  in the United States<sup>11</sup>, although power plants and

other large fossil fuel combustors are also important, resulting in local hotspots of NO<sub>2</sub>. This results in a heterogenous distribution of NO<sub>2</sub>. NO<sub>2</sub> modeling, therefore, needs to capture small-scale variations, which can be challenging.

NO<sub>2</sub> concentrations also vary considerably from day to day due to its short lifetime. Chemical sources and sinks, the height of the planetary boundary layer, wind speed, and wind direction all influence concentrations in any location, with substantial variations from day to day, even given similar emissions. Hence accurate modeling must also capture these temporal patterns.

Many existing NO<sub>2</sub> models were based on land-cover regression. Land-cover terms are proxies for traffic emissions and related to NO<sub>2</sub> concentrations indirectly. A typical NO<sub>2</sub> model is based on a land-cover regression with such quantities as road length, population density, tree canopy coverage, impervious surface, elevation, distance to coast<sup>12</sup>, traffic flow<sup>13</sup>, traffic intensity<sup>14</sup>, land-cover type<sup>13, 15, 16</sup>, road types<sup>17</sup>, building density<sup>18</sup>, and urban density<sup>19</sup>, as predictor variables.

Some of these land-cover regressions incorporated satellite measurements as well. NO<sub>2</sub> column density from OMI (Ozone Monitoring Instrument) have been widely used in NO<sub>2</sub> modeling<sup>15, 17, 20–25</sup>, for its relatively fine spatial resolution (13 km × 24 km) and continuous operation since 2004. SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric CHartography) and GOME-2A (Global Ozone Monitoring Experiment-2A) also provided NO<sub>2</sub> column density, but were used less often<sup>24, 26–28</sup>, since GOME-2A and SCIAMACHY have much coarser resolutions (80 km × 40 km and 60 km × 30 km respectively) and daily coverage for GOME-2A was not available after 2012 due to a change in viewing configuration. SCIAMACHY stopped data collection after 2012. NO<sub>2</sub> retrievals from satellite measurement are column concentrations. To obtain surface-level NO<sub>2</sub>, existing studies used scaling factors (i.e., the vertical distribution of NO<sub>2</sub>) from chemical transport models to derive the relationships between satellite retrievals and surface-level NO<sub>2</sub> concentrations<sup>21, 29</sup>. Chemical transport models can also directly simulate surface-level NO<sub>2</sub><sup>15,30,31</sup>, in addition to providing scaling factors.

In terms of fitting algorithms, most previous studies have used simple regression with some variable selection process, or more advanced regression methods, such as geographically weighted regression<sup>32</sup>. Most recently, several studies estimated NO<sub>2</sub> concentration using machine learning approaches. Gardner et al. used multilayer neural networks to model hourly NO<sub>2</sub> in Central London, which outperformed a regression model<sup>33</sup>. Kukkonen et al. also found that a neural network outperformed a regression model when estimating NO<sub>2</sub> levels in central Helsinki, Finland<sup>34</sup>. Yeganeh et al. employed an adaptive neuro-fuzzy inference system, a kind of artificial neural network, to estimate monthly mean NO<sub>2</sub> levels in a selected area in Australia, with model performance superior to that of a multiple regression model<sup>35</sup>. Other machine learning algorithms were also utilized to model NO<sub>2</sub> in other regions, including Hong Kong, where a support vector machine predicted hourly NO<sub>2</sub><sup>36</sup>, and urban Hungary, where a forecast model used a neural network and a support vector machine<sup>37</sup>.

After reviewing existing NO<sub>2</sub> models, we found two major areas for improvement. **First**, no existing study achieved high spatial resolution, high temporal resolution, and large spatial coverage at the same time. NO<sub>2</sub> models with fine spatial resolution or/and temporal resolution were often constrained to a small study area, usually at the city level<sup>13, 19, 31, 33, 38–43</sup>, while studies extending over a larger area had either relative low temporal resolution<sup>16, 18, 22, 30</sup> (e.g., national models only available at the annual level) or low spatial resolution<sup>23</sup>. **Second**, most existing studies relied on a single model and a single fitting algorithm to estimate NO<sub>2</sub>, even though recent studies suggest that a hybrid model is better at integrating monitoring data, land-cover regression, remote sensing data, and dispersion data<sup>44</sup> and could potentially improve model performance<sup>23</sup>.

Therefore, in this study we integrated multiple types of predictor variables and multiple types of machine learners into an ensemble model to estimate NO<sub>2</sub> with high spatial resolution (1 km × 1 km), high temporal resolution (daily), and large spatial coverage (the contiguous United States) from 2000 to 2016. We further added a land cover regression with meteorology to estimate within-grid variation. The ensemble model integrated neural network, random forest, and gradient boosting algorithms into a unified framework based on a generalized additive model for ensemble averaging. For predictor variables, we used satellite-based NO<sub>2</sub> measurements, an extensive number of land-cover variables, meteorological variables, simulation results from multiple chemical transport models, and some predictor variables not used by previous studies. We validated our model using 10-fold cross-validation and predicted daily NO<sub>2</sub> levels for every 1 km × 1 km grid cell in the entire contiguous United States from 2000 to 2016. We also quantified the uncertainty level by estimating the monthly standard deviation of the difference between daily monitored NO<sub>2</sub> and predicted NO<sub>2</sub>, for the same 1 km × 1 km grid cells. This high resolution daily NO<sub>2</sub> estimation, along with predicted uncertainty, will help epidemiologists to better assess both long-term and short-term exposures for studies of large cohorts with residents in locations far from or without monitors.

## 2. Data

### 2.1. Study Area and NO<sub>2</sub> Measurements

Our study area is the contiguous United States, including 48 states and Washington, D.C. The contiguous United States has several NO<sub>2</sub> monitoring networks included in the Air Quality System (AQS) from the Environmental Protection Agency (EPA), encompassing 912 monitoring sites. Monitoring sites are not evenly distributed in the study area, with more monitoring sites in populous regions and urban areas. Mountainous regions and some remote border areas have almost no monitoring sites. We extracted or calculated 1-hour daily maximum NO<sub>2</sub> concentrations, the NO<sub>2</sub> metric used for EPA regulation, from these monitoring sites. We used the term “daily NO<sub>2</sub>” to stand for 1-hour daily maximum NO<sub>2</sub> in this paper, unless specified otherwise. The study period is from January 1<sup>st</sup>, 2000 to December 31<sup>st</sup>, 2016, a total of 6,210 days. Not all monitoring sites were operating during the entire study period. Missing data within monitoring sites were excluded from the follow-up model training process.

Like other air pollutants, the distribution of NO<sub>2</sub> demonstrates some degree of spatial and temporal autocorrelation. NO<sub>2</sub> measurements from nearby monitoring sites are more correlated than those from sites far apart; NO<sub>2</sub> measurements from neighboring days are more correlated than measurements distant in time. Using autocorrelation can improve model fit, and we incorporated spatially and temporally lagged NO<sub>2</sub> measurements. Spatially lagged terms were calculated as inverse distance weighted NO<sub>2</sub> measurements at other locations, as well as their one-day, three-day and five-day lagged moving average values.

## 2.2. Meteorological Data

Reanalysis data sets rely on data sourced from land-surface monitors, ship, aircraft, satellite radiosondes, pibals, and other sources. The National Oceanic and Atmospheric Administration (NOAA) assimilates these data sets into a data assimilation system and provides gridded atmospheric fields. Compared with meteorological measurements from monitoring sites, reanalysis data provide almost continuous spatial and temporal coverage, often with no or few missing values. We used daily values of 16 meteorological variables (Section 1, Supporting Information), with spatial resolution approximately 32 km.

## 2.3. NO<sub>2</sub> Column Density and Chemical Transport Model Simulations

We used NO<sub>2</sub> column density from the OMI instrument aboard the Aura satellite. The OMI NO<sub>2</sub> data product is available every day at 13 km × 24 km grid cells. OMI NO<sub>2</sub> retrievals are column measurements. To relate OMI NO<sub>2</sub> retrievals to surface-level NO<sub>2</sub> level, we used chemical transport models to simulate scaling factors.

A chemical transport model (CTM) simulates the chemistry, transport, and deposition of air pollutants in discrete three-dimension grid cells, based on surface-level emission inventories and meteorological fields. The models capture the relevant atmospheric photochemical reactions, including the secondary formation of air pollutants. We used the vertical distribution of NO<sub>2</sub> from two different CTMs – the global-scale GEOS-Chem (<http://acmg.seas.harvard.edu/geos/>) and the regional-scale Community Multi-scale Air Quality Model (CMAQ, <https://www.epa.gov/cmaq>) – and calculated scaling factors as the percentage of surface-level NO<sub>2</sub> contributing to the total NO<sub>2</sub> column density. We then related the satellite-retrieved NO<sub>2</sub> column to surface-level NO<sub>2</sub>, as in previous existing studies<sup>29, 45, 46</sup>. In addition, we used surface-level NO<sub>2</sub> estimates from the CTMs as a predictor variable in NO<sub>2</sub> modeling. Details of both CTMs can be found elsewhere<sup>47–49</sup>. The spatial resolution of GEOS-Chem output was 0.5° × 0.625°; the spatial resolution of CMAQ output was 12 km for all years, except 36-km resolution for the Western U.S. in the early years. Neither GEOS-Chem nor CMAQ was calibrated or tuned with NO<sub>2</sub> monitoring data.

## 2.4. Land-cover Variables

A large percentage of surface NO<sub>2</sub> concentrations stems from local traffic emissions, which are sensitive to land-cover patterns<sup>50</sup> and can be approximated by land-cover terms. Hence, land-use variables are among the most important predictor variables in NO<sub>2</sub> modeling. Land-cover variables have been used in nationwide NO<sub>2</sub> models<sup>12, 51</sup>, as well as some regional or neighborhood models<sup>38, 41</sup>. Following previous studies<sup>52</sup>, this study included

seven categories of land-cover variables, including land-cover terms (water bodies, developed areas, barren land, forest, shrubland, herbaceous land, planted/cultivated land, wetlands, impervious surface, and tree canopy), truck traffic (truck traffic volume, truck route density, and shortest distance to truck route), road density (road density for primary roads, secondary roads, and all roads, respectively), restaurant density, elevation (minimum elevation, maximum elevation, mean elevation, median elevation, standard deviation of elevation, and breakline emphasis), normalized difference vegetation index, nighttime light, with details listed in the Supporting Information (Section 2).

We also prepared selected land-cover variables at three resolutions: 100 m  $\times$  100 m, 1 km  $\times$  1 km and 10 km  $\times$  10 km. OMI column NO<sub>2</sub> has a resolution of 13 km  $\times$  24 km; the horizontal resolution of GEOS-Chem, CMAQ, and reanalysis data sets are at similar levels, or even coarser than OMI. The kilometer-level variables capture local emissions, especially from traffic, and emissions from neighboring areas, and the 10-km variables capture more of the overall pattern of urban emissions. We incorporated 1-km- and 10-km-level land-cover variables to fit the three machine learning models. The 100-m-level land-cover variables were used to fit the local models of address-specific deviations from the 1 km grid cell.

## 2.5. Other Ancillary Variables

The retrieval algorithm of satellite-based NO<sub>2</sub> is affected by aerosol, surface reflectance<sup>53</sup>/ surface albedo, and cloud contamination<sup>54</sup>, although the agreement of satellite-based NO<sub>2</sub> with in situ measurements is usually good<sup>55</sup>. To correct possible errors in the NO<sub>2</sub> retrieval, we further added the following variables to our model. (1) Variables related to aerosol concentration and aerosol type, including simulated elemental carbon, organic carbon, sulfate, nitrate, aerosol mass from both GEOS-Chem and CMAQ; sulfate aerosol, hydrophilic black carbon, hydrophobic black carbon, hydrophilic organic carbon, and hydrophobic organic carbon from MERRA-2<sup>56</sup>; and absorbing aerosol index in the ultraviolet and visible ranges (OMAERUVd, OMAEROe) from OMI<sup>57, 58</sup>. (2) Cloud coverage, including cloud area fractions at low, medium, and high altitudes from the NCEP/NCAR reanalysis data set<sup>59</sup>. (3) Surface albedo from the NCEP/NCAR reanalysis data set<sup>59</sup> and surface reflectance from MODIS (MOD09A1)<sup>60</sup>.

OMI retrievals have many missing values. We also acquired NO<sub>2</sub> column simulations from Copernicus Atmosphere Monitoring Service (CAMS), another reanalysis data set<sup>61</sup>. The CAMS reanalysis data for NO<sub>2</sub> rely on observations from multiple satellites, without observations from NO<sub>2</sub> monitoring sites, combined with state-of-the-art computer models. CAMS NO<sub>2</sub> columns have a spatial resolution of 0.125°  $\times$  0.125°, similar to that of OMI NO<sub>2</sub> retrievals and with no missing values, providing additional information where OMI NO<sub>2</sub> retrievals are missing.

## 3. Methods

### 3.1. Overview

Our NO<sub>2</sub> model was based on an ensemble model that took estimates from three independent machine learning algorithms. We first fit neural network, random forest, and

gradient boosting algorithms with all input predictor variables and monitored NO<sub>2</sub> as the dependent variable. Then, a generalized additive geographically weighted model combined the NO<sub>2</sub> estimates from the three algorithms and produced a final NO<sub>2</sub> estimation. NO<sub>2</sub> concentrations demonstrate some degree of temporal and spatial autocorrelation. To leverage this autocorrelation, we used the above NO<sub>2</sub> estimates, calculated their spatially and temporally lagged values, used them as additional input predictor variables in refitting the three machine learning algorithms and ensemble model again (Figure S1). In this two-step modeling framework, each step combines a neural network, random forest, gradient boosting, and a generalized additive model into an ensemble model.

We applied ten-fold cross-validation in choosing the model hyperparameters to avoid overfitting. We also used 10-fold cross-validation to evaluate the final model performance. We randomly divided all monitoring sites into 10 splits. We trained the model with 90% of the monitoring sites and predicted NO<sub>2</sub> at the remaining 10% monitoring sites; then we repeated the process for other 9 splits. We aggregated cross-validated NO<sub>2</sub> predictions from 10 splits together, compared with corresponding NO<sub>2</sub> monitoring values, and calculated total R<sup>2</sup>, temporal R<sup>2</sup>, spatial R<sup>2</sup>, root mean square error (RMSE), and other metrics for model performance. The definition of total R<sup>2</sup>, temporal R<sup>2</sup>, spatial R<sup>2</sup> and RMSE are based on previous literature<sup>62</sup>. It is worth mentioning that spatial R<sup>2</sup> is calculated by regressing annual-averaged monitored NO<sub>2</sub> against the predicted value, so spatial R<sup>2</sup> evaluates model performance of long-term averages.

### 3.2. Three Machine Learning Algorithms

Previous studies have used neural network, random forest<sup>63</sup>, and other machine learning algorithms to estimate surface-level NO<sub>2</sub><sup>17, 23, 33, 34</sup>. In these studies, land-cover variables, satellite measurements and other predictors were input variables of the machine learning algorithm; monitored NO<sub>2</sub> was the dependent variable. We used neural network, random forest, and gradient boosting algorithms to estimate monitored NO<sub>2</sub> separately, with all predictors as input variables. Hyperparameters of the machine learning algorithms, such as the number of hidden layer and the number of neurons for a neural network and learning rate for gradient boosting, were determined by a grid search process and imbedded cross-validated process (Table S1). To improve efficiency, we standardized all input variables by  $x_{standard} = \frac{x - mean(x)}{sd(x)}$  and took the logarithm of the monitored NO<sub>2</sub>. We also used imputation to fill in missing values of predictor variables before model training and model prediction (Section 3, Supporting Information).

### 3.3. Ensemble Model

To blend NO<sub>2</sub> estimations from the three machine learning algorithms, we used a generalized additive model with penalized spline on both location and NO<sub>2</sub> estimation to account for geographic weights:

$$\widehat{NO_2} = f_1(\text{Location}_i, \widehat{NO_2, nni_j}) + f_2(\text{Location}_i, \widehat{NO_2, rfi_j}) + f_3(\text{Location}_i, \widehat{NO_2, gbi_j})$$

where  $f_1$  denotes a thin plate spline for an interaction between location  $i$  and the  $\text{NO}_2$  estimation from the neural network at location  $i$  and on day  $j$  ( $\widehat{\text{NO}}_{2, \text{nn}ij}$ );  $f_2$  and  $\widehat{\text{NO}}_{2, \text{rf}ij}$  and  $f_3$  and  $\widehat{\text{NO}}_{2, \text{gb}ij}$  stand for similar quantities, but from random forest and gradient boosting at location  $i$  and on day  $j$ , respectively. By employing this generalized additive model, we allowed the contribution of each algorithm to the final  $\text{NO}_2$  estimate to potentially depend on the  $\text{NO}_2$  concentration (i.e., non-linear response) and vary in different locations (geographically weighted regression).

To fit the local address deviations from a grid cell level, we took the daily residuals at each monitor and modeled these as a function of local land cover within 100 m and meteorology, using a random forest. Downscaling predictors included NLCD land-cover, truck traffic, traffic volume, elevation, and road density. We also included air temperature, humidity, wind speed, and planetary boundary layer height.

### 3.4. Model Prediction

We predicted daily  $\text{NO}_2$  at  $1 \text{ km} \times 1 \text{ km}$  grid cells in the study area with the trained model. In total there are over 11 million grid cells in the entire study area. The trained model here included trained neural networks, random forests, gradient boosting models, and generalized additive models in both steps. Model prediction repeated the same process as model training: obtain  $\text{NO}_2$  prediction from three learning algorithms, put them into the ensemble model and calculated  $\text{NO}_2$  estimation, calculate spatially and temporally averaged  $\text{NO}_2$  estimation, and use these averages as additional predictors and repeat above process again (Figure S1).

The address-specific exposure can be used to assign better exposure in studies where addresses or geocodes are available. To illustrate this while avoiding confidentiality issues, we estimated the final  $\text{NO}_2$  estimation on a 100-m grid in the greater Boston metropolitan area. We calculated the residual of the  $\text{NO}_2$  model (monitored  $\text{NO}_2$  minus predicted  $\text{NO}_2$ ) and used downscaling predictor variables to estimate the residual in a random forest. After training the random forest model, we prepared those downscaling variables and predicted residuals in each  $100 \text{ m} \times 100 \text{ m}$  grid cell.

### 3.5. Uncertainty Estimation

We also estimated the uncertainty in the  $\text{NO}_2$  predictions. We used the following generalized additive model to estimate the monthly uncertainty of  $\text{NO}_2$  estimation:

$$\begin{aligned} sd(\Delta \text{NO}_{2ij}) = & f_1(\text{Location}_i) + f_2(\text{Location}_i, \widehat{\text{NO}}_{2ij}) + f_3(\text{elevation}) + f_3(\text{elevation s.d.}) \\ & + f_4(\text{truck traffic}) + f_5(\text{traffic volume}) + f_6(\text{humidity}) + f_7(\text{tree canopy}) + f_8(\text{NDVI}) + f_9(\text{urban}) + \text{Year} \\ & + e_{ij} \end{aligned}$$

where  $sd(\Delta \text{NO}_{2ij})$  is the standard deviation of the difference between monitored daily  $\text{NO}_2$  and estimated daily  $\text{NO}_2$  at location  $i$  and month  $j$ ;  $f_1$  is a penalized spline for location  $i$ ;  $f_2$  is a thin plate spline for an interaction between location and monthly averaged predicted  $\text{NO}_2$  at location  $i$  and month  $j$ ;  $f_3 \sim f_9$  are splines on elevation, standard deviation of elevation,



truck traffic, traffic volume, humidity, tree canopy, NDVI, and urban areas, respectively. The error term is  $e_{ij}$ .

#### 4. Results

The mean cross-validated  $R^2$  was 0.79 for daily  $\text{NO}_2$ . The two-step modeling framework indeed improved model performance, with total  $R^2$  improved from 0.77 in Step 1 to 0.79 in Step 2 (Table S2). The spatial  $R^2$ , which we defined as the  $R^2$  between annual averaged monitored  $\text{NO}_2$  and estimated  $\text{NO}_2$ , varied between 0.78 to 0.86 by year, with a mean spatial  $R^2$  of 0.84, indicating a good model performance at the annual level (Table 1). The average RMSE was 7.15 ppb overall (4.51 ppb spatially and 5.57 ppb temporally). The ensemble model outperformed the three base learners ( $R^2$ , neural network: 0.763, random forest: 0.787, and gradient boosting: 0.752). Temporally, model performance remained stable, but less satisfying in early and most recent years. Among the three machine learners, the random forest outperformed the neural network and gradient boosting. Overall, ensemble averaging further improved model performance compared to the best single learner, although only modestly. Figure 1 presents the maps of uncertainty level, with better model performance in California (except the south) and the Northeastern United States. Performance was worse in mountainous regions, such as Rocky and Appalachian Mountains, where site monitors are sparse.

Although the ensemble model only had a limited impact on daily  $R^2$ , it improved the linearity of the relationship between monitored  $\text{NO}_2$  and predicted  $\text{NO}_2$ . The neural network underestimated  $\text{NO}_2$  at high concentrations, while the random forest overestimated at the high end. The overestimation at the high end was even more serious for gradient boosting. The ensemble model showed good linearity until 150 ppb, an extremely high daily concentration seldom seen in the contiguous United States (Figure 2). At the annual level, the linearity between monitored and predicted  $\text{NO}_2$  was even better, with linearity at concentrations up to 55 ppb, a very high annual average that only 0.2% monitoring data reached (Figure S2). Both Figure 2 and Figure S2 indicated that our model estimated  $\text{NO}_2$  accurately at common pollution levels in the contiguous United States, with slight underestimation at extremely high (and also rare) concentration levels.

The distribution of  $\text{NO}_2$  exhibits clear spatial clustering, with high concentration clustering around urban areas, especially major cities, and along highways. From the 2000 national maps, we can clearly identify several  $\text{NO}_2$  hotspots, such as Seattle, Los Angeles, Phoenix, Salt Lake City, Denver, Albuquerque, Chicago, Indianapolis, Louisville, New York, and Philadelphia (Figure 3). This clustering pattern of  $\text{NO}_2$  is clearer in the downscaled prediction of the greater Boston metropolitan area, using a 100-m resolution grid to illustrate the address specific model (Figure S3). We can clearly identify the central urban area with generally high concentrations, but lower concentrations in rural areas, over greenspaces and waterbodies.

$\text{NO}_2$  concentrations fell substantially in the U.S. during the study period, with annual level in 2016 about 50% of the 2000 concentrations, but the decline stagnated after 2007. The nationwide  $\text{NO}_2$  level in 2016 was almost identical (100.08%) to that of 2007 (Figure 4). By

constraining only to predictions at monitoring sites, we observed a different pattern, with long-term decline and a steady decrease after 2007, consistent with the trend reported in a previous GEOS-Chem model study<sup>49</sup>. The average predicted NO<sub>2</sub> level at the monitoring sites in 2016 was only 71.62% of the 2007 level (Figure 4).

We also reported the relative importance of different variables from the three machine learning algorithms (Table S3). Specific approaches to assess variable importance were mentioned in the footnote of Table S3. Spatially lagged NO<sub>2</sub> and its 1-day-lagged values were both important predictor variables. Multiple land-cover variables, such as impervious surface, developed land, road density, traffic volume of truck route also ranked as important predictors. The explanatory power of CMAQ-simulated NO<sub>2</sub>, and elemental carbon, which derives from similar sources as NO<sub>2</sub> was also high. The standard deviation of elevation, maximum elevation, nighttime light, and traffic volume of trucks -- variables seldom used in previous studies -- also demonstrated important explanatory power.

## 5. Discussion

In this paper, we present an ensemble model to incorporate neural network, random forest, and gradient boosting to estimate daily NO<sub>2</sub> across the contiguous United States. Performance of the ensemble model was excellent, with cross-validated mean R<sup>2</sup> of 0.79, mean spatial R<sup>2</sup> of 0.84, RMSE of 7.15 ppb, and spatial RMSE of 4.51 ppb. Our model used various types of predictors (satellite remote sensing, chemical transport models, multiple land-cover terms) that are not often combined in such models, as well as ensembled results from them using three different machine learning algorithms. We predicted daily NO<sub>2</sub> at every 1 km × 1 km grid cell in the contiguous United States, which should be useful for epidemiology and health impact assessment that require small area estimates (e.g., over census tracts or ZIP Codes). The ability to predict well outside of major urban areas is an important feature of this model, with good performance in rural areas as well. A key addition is the modeling of the standard deviation of exposure error for each month of each year in each grid cell. This will enable researchers to incorporate the measurement errors in epidemiological studies<sup>64</sup>.

This study exhibits several advantages over existing studies. **First**, our modeling framework incorporated multiple machine learning algorithms, and assembles them in an innovative way. These complementary machine learning algorithms improved model performance, especially at high concentrations. In contrast to many ensemble methods, which give fixed weights to each machine learner, our approach lets the weights vary spatially and by NO<sub>2</sub> concentration. This modeling framework, with several independent algorithms estimating NO<sub>2</sub> individually and a generalized additive model combining them, can be extended to additional fitting algorithms and is applicable to modeling other air pollutants. For example, several existing studies on NO<sub>2</sub> modeling used a support vector machine, which could become another base learner in future ensemble models. **Second**, this study achieved high spatiotemporal resolution, with 1-km-level and potentially address-specific predictions available every day. Most existing studies estimated NO<sub>2</sub> at the annual level, which would not be appropriate for pregnancy outcomes or acute effects. In addition, previous studies exhibited a tradeoff between resolution and study area. NO<sub>2</sub> models with large spatial

coverage (e.g., nationwide models) generally had to compromise either on spatial/temporal resolution or both. Our study, using multiple land-cover variables as spatial predictors and meteorological variables and CTM simulations as temporal predictors, achieved fine spatial and temporal resolutions for the entire contiguous United States. **Third**, we developed a sophisticated model to fill in the missing values. Unlike previous studies estimating annual NO<sub>2</sub> and simply excluding missing values, daily estimation of NO<sub>2</sub> requires filling in missing values before training the model. Moreover, annual average estimates can be biased if the data are not missing at random, a situation our method avoids. Some studies used values from the nearest locations to fill in missing values<sup>65</sup>, but we have argued in a related study on PM<sub>2.5</sub> modeling that this strategy can be problematic, especially when the number of missing values is large and missing values are spatially clustered. While our method of filling in missing values requires separate prediction models for each variable with missing data, which is computationally intensive, growing computational capacity makes the process less formidable. We leveraged computational power from the Harvard Odyssey Supercomputer.

We have additionally used the standard deviation of elevation, maximum elevation, restaurant density, nighttime light, and truck route as predictor variables, which, to the best of our knowledge, have seldom been used in previous studies. Results on variable importance indicate that these variables are important for NO<sub>2</sub> estimation. Truck exhaust is the largest source of NO<sub>x</sub> in the United States<sup>11</sup>, and responsible for a large portion of NO<sub>x</sub> emission in other countries as well<sup>66</sup>. Indeed, NO<sub>x</sub> emission from trucks is many times higher than normal passenger cars<sup>67</sup>. Thus, it is reasonable to separate truck emissions from generic traffic emissions. Elevation is a predictor variable widely used in NO<sub>2</sub> estimation. Our study suggests that elevation variation, instead of elevation itself, is more important. This is consistent with common sense: topography, as well as stable tropospheric structure in the winter due to temperature inversion, affects dispersion of air pollutants. For a similar reason, breakline emphasis of elevation was an important variable, which again demonstrates that elevation variation matters in air pollution modeling. Nighttime light corresponds to the level of urbanization, energy consumption, and overall economic activity<sup>68–70</sup>, and thus is related to pollution emission. Previous studies have used nighttime light in PM<sub>2.5</sub> modeling<sup>71</sup>. Nighttime light is available globally over multiple years and could serve as an important predictor variable for NO<sub>2</sub> modeling in other countries. In contrast, other variables we used here are not always available. Cooking is a major source of air pollution, especially in cities<sup>72</sup>. Thus, restaurants are an important source, creating local hotspots of NO<sub>2</sub>. Incorporating restaurants as a predictor variable can improve model performance at finer scales, especially in cities.

The three machine learning algorithms gave the highest weights to different predictor variables. Spatially lagged terms of monitored NO<sub>2</sub> (i.e. nearby monitors) play important roles in all three algorithms. Gradient boosting predominantly depends on these lagged terms; random forest relies primarily on these terms plus additional land-cover variables and CTM simulations, while neural network relies primarily on land-cover variables. The relative importance of land-cover variables also varies. Our results indicate that the contribution or importance of predictor variables depends on the fitting algorithm. Similarly,

for the ensemble model, the contribution of three individual machine learning algorithms varied by concentrations and location. Based on these results, we conclude that the model performance of different fitting algorithms and the contribution of different predictor variables are context-based. In other words, it is difficult to foresee which variable(s) are most informative and which fitting algorithm is most appropriate to an air pollution model without actually running the model. Answers to both questions depend on the research topic, time period, and study area. Some previous studies compared the performance of machine learning algorithms with a statistical model<sup>34, 73</sup>, or compared the performance of different model specifications<sup>30, 45</sup>. Our study suggests that it would be more useful to propose a framework integrating multiple predictor variables and estimations from different fitting algorithms, as we did in this study. We also conclude that the specific structure of ensemble model depends on the practical interest. For our study, the ensemble model aggregated daily NO<sub>2</sub> estimation to improve model performance at daily level, thus total R<sup>2</sup> improved (Table 1); but model performance at annual level may not be optimized at the same time, thus spatial R<sup>2</sup> of the ensemble model decreased slightly compared with random forest (Table S4). To optimize spatial R<sup>2</sup>, another ensemble model is required to aggregate NO<sub>2</sub> estimations at annual level.

We found that satellite-derived NO<sub>2</sub> column measurements are not as important as other predictor variables, contributing to less than 1% of the prediction in the neural network, random forest, and gradient boosting methods. This result contrasts with PM<sub>2.5</sub> modeling, where satellite-derived aerosol optical depth (AOD) is an important predictor variable. The reasons are multiple-fold: first, NO<sub>2</sub> column measurement from OMI is much coarser (13 km × 24 km) compared with AOD from MODIS (the finest resolution of MODIS AOD is 1 km × 1 km). Coarse satellite-based NO<sub>2</sub> measurements average out heterogeneous NO<sub>2</sub> levels within each cell<sup>74</sup>. This is especially an issue when modeling NO<sub>2</sub>, an air pollutant primarily coming from local traffic emission, and fine-scale measurement is essential. Second, the sensitivity of OMI and any satellite-based measurement of NO<sub>2</sub> increases with altitude, such that the measurement is least sensitive at the surface due to scattering of radiation at the surface and through the atmosphere<sup>75</sup>. Third, CTM outputs already contributed to the temporal variations and the satellite-derived NO<sub>2</sub> was less important as a result.

In the long term, the spatial distribution of surface NO<sub>2</sub> contrasts with that of PM<sub>2.5</sub> and ozone (Figure S4). High NO<sub>2</sub> levels cluster along highways and cities. Traffic emissions are also an important source of primary PM<sub>2.5</sub>, as well emitting precursor gases that form PM<sub>2.5</sub> in the atmosphere. However, PM<sub>2.5</sub> has a longer atmospheric lifetime than NO<sub>2</sub> and can be transported further; it also has more widespread sources of importance, such as the biosphere or aqueous phase production in clouds. Thus, for example, the entire Southeastern United States experiences high PM<sub>2.5</sub> concentration in the summer, while NO<sub>2</sub> is more locally enhanced. The spatial distribution of ozone also exhibits different patterns from NO<sub>2</sub>, with high concentrations occurring over rural areas surrounding or downwind of urban areas, and in mountainous regions. The distinct patterns of NO<sub>2</sub>, PM<sub>2.5</sub>, and ozone at the national level suggest that a nationwide environmental epidemiological study could separate and identify the adverse health effect of each pollutant.

In terms of temporal trend, we found a discrepancy between the nationwide average trend and the average trend across the monitoring sites. We observed a steadily decreasing trend of averaged NO<sub>2</sub> level at monitoring sites, but at the national level (i.e., averaged NO<sub>2</sub> level for every 1-km grid in the contiguous U.S.) NO<sub>2</sub> declined from 2000 to 2007 and stagnated after 2007. For example, from 2007 to 2008, the site-averaged NO<sub>2</sub> level dropped from 21.9 ppb to 21.1 ppb at monitoring sites with about 3.5% decrease, but our ensemble model predicted that the nationwide averaged NO<sub>2</sub> rebounded from 8.1 ppb to 9.7 ppb over the same period, with a nearly 20% increase.

The steady 2000–2016 decrease of NO<sub>2</sub> concentrations predicted at the monitoring sites is consistent with observations and with the National Emission Inventory (NEI) of the U.S. Environmental Protection Agency (EPA), underscoring the success of clean air regulations (Silvern et al., 2019). However, the discrepancy between the predicted site-based trend and the nationwide trend suggests a different pattern of NO<sub>2</sub> pollution in less urban areas where there is scant monitor coverage. Whether this is due to a lower rate of replacement of more polluting vehicles, increased wood combustion, increased influence of background NO<sub>x</sub>, or widespread reduction in anthropogenic VOC emission in urban areas, deserves further attention, particularly as rural NO<sub>x</sub> pollution may impact production of secondary organic particles and ozone.

Our model has some limitations. There are still differences between predicted and observed NO<sub>2</sub> values, and we only model outdoor concentrations, and not personal exposure. However, a recent study pointed out that ambient exposure has an advantage over personal exposure in epidemiology studies in that it is much less correlated with individual level confounders<sup>76</sup>. The model also depends on the existing monitoring network, and so was unable to take advantage of local intensive monitoring campaigns, which are often used in land cover regressions. On the other hand, the model covers many years, whereas land cover regression suffers from challenges in representing the influence of changing emissions over time using largely static land-cover terms. Despite these limitations, the daily NO<sub>2</sub> concentrations at high spatial resolution provided by our ensemble model promise to improve estimates of both long- and short-term exposures for epidemiological studies of large cohorts of U.S. residents, even those living far from monitors.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgement

This publication was made possible by U.S. EPA grant numbers RD-834798, RD-835872, and 83587201; HEI grant 4953-RFA14-3/16-4. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the U.S. Environmental Protection Agency (EPA). The views expressed in this article are those of the authors and do not necessarily represent the views or policies 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. Research described in this article was also conducted under contract to the Health Effects Institute (HEI), an organization jointly funded by the U.S. EPA (Assistance Award No. CR-83467701) and certain motor vehicle and engine manufacturers. The contents of this article do not necessarily reflect the views of HEI, or its sponsors, nor do they necessarily reflect the views and policies of the EPA or motor vehicle and engine manufacturers. The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

## References

1. Kagawa J, Evaluation of biological significance of nitrogen oxides exposure. *The Tokai journal of experimental and clinical medicine* 1985, 10, (4), 348–353. [PubMed: 3836516]
2. Chauhan A; Krishna M; Frew A; Holgate S, Exposure to nitrogen dioxide (NO<sub>2</sub>) and respiratory disease risk. *Reviews on environmental health* 1998, 13, (1–2), 73–90. [PubMed: 9718623]
3. Weinmayr G; Romeo E; De Sario M; Weiland SK; Forastiere F, Short-term effects of PM<sub>10</sub> and NO<sub>2</sub> on respiratory health among children with asthma or asthma-like symptoms: a systematic review and meta-analysis. *Environmental health perspectives* 2009, 118, (4), 449–457. [PubMed: 20064785]
4. Speizer FE; Ferris B Jr; Bishop YM; Spengler J, Respiratory disease rates and pulmonary function in children associated with NO<sub>2</sub> exposure. *American Review of Respiratory Disease* 1980, 121, (1), 3–10. [PubMed: 7352711]
5. Brauer M; Lencar C; Tamburic L; Koehoorn M; Demers P; Karr C, A cohort study of traffic-related air pollution impacts on birth outcomes. *Environmental health perspectives* 2008, 116, (5), 680. [PubMed: 18470315]
6. Chiusolo M; Cadum E; Stafoggia M; Galassi C; Berti G; Faustini A; Bisanti L; Vigotti MA; Dessì MP; Cernigliaro A, Short-term effects of nitrogen dioxide on mortality and susceptibility factors in 10 Italian cities: the EpiAir study. *Environmental health perspectives* 2011, 119, (9), 1233. [PubMed: 21586369]
7. Latza U; Gerdes S; Baur X, Effects of nitrogen dioxide on human health: systematic review of experimental and epidemiological studies conducted between 2002 and 2006. *International journal of hygiene and environmental health* 2009, 212, (3), 271–287. [PubMed: 18771952]
8. Zhao Y; Saleh R; Saliba G; Presto AA; Gordon TD; Drozd GT; Goldstein AH; Donahue NM; Robinson AL, Reducing secondary organic aerosol formation from gasoline vehicle exhaust. *Proceedings of the National Academy of Sciences* 2017, 114, 6984–6989.
9. Xu L; Guo H; Boyd CM; Klein M; Bougiatioti A; Cerully KM; Hite JR; Isaacman-VanWertz G; Kreisberg NM; Knote C; Olson K; Koss A; Goldstein AH; Hering SV; de Gouw J; Baumann K; Lee S-H; Nenes A; Weber RJ; Ng NL, Effects of anthropogenic emissions on aerosol formation from isoprene and monoterpenes in the southeastern United States. *Proceedings of the National Academy of Sciences* 2015, 112, 37–42.
10. Pye HOT; D'Ambro EL; Lee BH; Schobesberger S; Takeuchi M; Zhao Y; Lopez-Hilfiker F; Liu J; Shilling JE; Xing J; Mathur R; Middlebrook AM; Liao J; Welti A; Graus M; Warneke C; de Gouw JA; Holloway JS; Ryerson TB; Pollack IB; Thornton JA, Anthropogenic enhancements to production of highly oxygenated molecules from autoxidation. *Proceedings of the National Academy of Sciences* 2019, 116, 6641–6646.
11. Preble C; Harley R; Kirchstetter T In N<sub>2</sub>O and NO<sub>2</sub> Emissions from Heavy-Duty Diesel Trucks with Advanced Emission Controls, AGU Fall Meeting Abstracts, 2014; 2014.
12. Novotny EV; Bechle MJ; Millet DB; Marshall JD, National satellite-based land-use regression: NO<sub>2</sub> in the United States. *Environmental science & technology* 2011, 45, (10), 4407–4414. [PubMed: 21520942]
13. Kim Y; Guldman J-M, Land-use regression panel models of NO<sub>2</sub> concentrations in Seoul, Korea. *Atmospheric Environment* 2015, 107, 364–373.
14. Kashima S; Yorifuji T; Sawada N; Nakaya T; Eboshida A, Comparison of land use regression models for NO<sub>2</sub> based on routine and campaign monitoring data from an urban area of Japan. *Science of The Total Environment* 2018, 631, 1029–1037. [PubMed: 29727929]
15. De Hoogh K; Chen J; Gulliver J; Hoffmann B; Hertel O; Ketzel M; Bauwelinck M; van Donkelaar A; Hvidtfeldt UA; Katsouyanni K, Spatial PM<sub>2.5</sub>, NO<sub>2</sub>, O<sub>3</sub> and BC models for Western Europe—Evaluation of spatiotemporal stability. *Environment international* 2018, 120, 81–92. [PubMed: 30075373]
16. Kim S-Y; Song I, National-scale exposure prediction for long-term concentrations of particulate matter and nitrogen dioxide in South Korea. *Environmental pollution* 2017, 226, 21–29. [PubMed: 28399503]

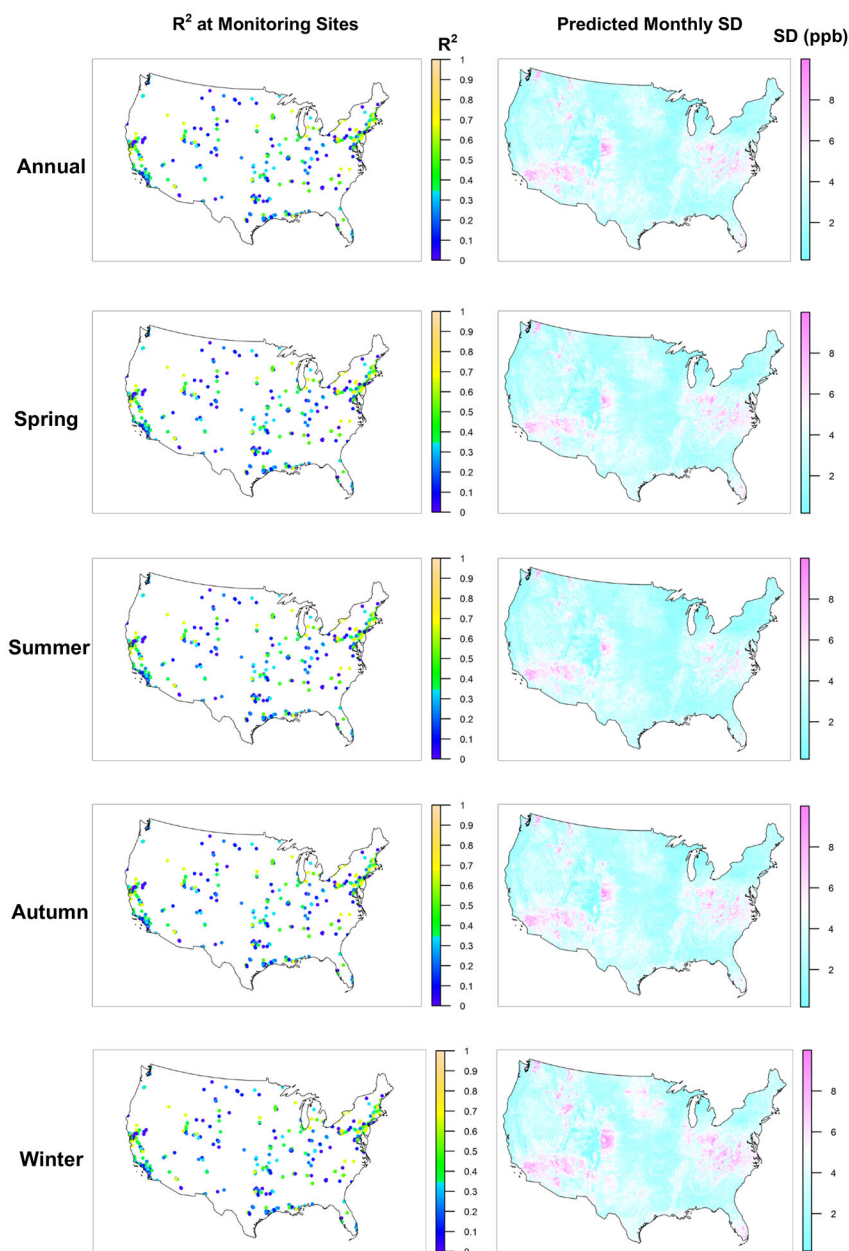
17. Araki S; Shima M; Yamamoto K, Spatiotemporal land use random forest model for estimating metropolitan NO<sub>2</sub> exposure in Japan. *Science of The Total Environment* 2018, 634, 1269–1277. [PubMed: 29710628]
18. Eeftens M; Meier R; Schindler C; Aguilera I; Phuleria H; Ineichen A; Davey M; Ducret-Stich R; Keidel D; Probst-Hensch N, Development of land use regression models for nitrogen dioxide, ultrafine particles, lung deposited surface area, and four other markers of particulate matter pollution in the Swiss SAPALDIA regions. *Environmental Health* 2016, 15, (1), 53. [PubMed: 27089921]
19. Gulliver J; de Hoogh K; Hoek G; Vienneau D; Fecht D; Hansell A, Back-extrapolated and year-specific NO<sub>2</sub> land use regression models for Great Britain-Do they yield different exposure assessment? *Environment international* 2016, 92, 202–209. [PubMed: 27107225]
20. Daneshvar MRM; Abadi NH, Spatial and temporal variation of nitrogen dioxide measurement in the Middle East within 2005–2014. *Modeling Earth Systems and Environment* 2017, 3, (1), 20.
21. Lamsal L; Martin R; Parrish DD; Krotkov NA, Scaling relationship for NO<sub>2</sub> pollution and urban population size: a satellite perspective. *Environmental science & technology* 2013, 47, (14), 7855–7861. [PubMed: 23763377]
22. Xu H; Bechle MJ; Wang M; Szpiro AA; Vedal S; Bai Y; Marshall JD, National PM<sub>2.5</sub> and NO<sub>2</sub> exposure models for China based on land use regression, satellite measurements, and universal kriging. *Science of The Total Environment* 2019, 655, 423–433. [PubMed: 30472644]
23. Zhan Y; Luo Y; Deng X; Zhang K; Zhang M; Grieneisen ML; Di B, Satellite-Based Estimates of Daily NO<sub>2</sub> Exposure in China Using Hybrid Random Forest and Spatiotemporal Kriging Model. *Environmental science & technology* 2018, 52, (7), 4180–4189. [PubMed: 29544242]
24. Bechle MJ; Millet DB; Marshall JD, Remote sensing of exposure to NO<sub>2</sub>: Satellite versus ground-based measurement in a large urban area. *Atmospheric Environment* 2013, 69, 345–353.
25. Lee HJ; Koutrakis P, Daily ambient NO<sub>2</sub> concentration predictions using satellite ozone monitoring instrument NO<sub>2</sub> data and land use regression. *Environmental science & technology* 2014, 48, (4), 2305–2311. [PubMed: 24437539]
26. Boersma K; Jacob DJ; Trainic M; Rudich Y; DeSmedt I; Dirksen R; Eskes H, Validation of urban NO<sub>2</sub> concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities. *Atmospheric Chemistry and Physics* 2009, 9, (12), 3867–3879.
27. Richter A; Burrows JP; Nüß H; Granier C; Niemeier U, Increase in tropospheric nitrogen dioxide over China observed from space. *Nature* 2005, 437, (7055), 129. [PubMed: 16136141]
28. Anand JS; Monks PS, Estimating daily surface NO<sub>2</sub> concentrations from satellite data—a case study over Hong Kong using land use regression models. *Atmospheric Chemistry and Physics* 2017, 17, (13), 8211–8230.
29. Lamsal L; Martin R; Van Donkelaar A; Steinbacher M; Celarier E; Bucsela E; Dunlea E; Pinto J, Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. *Journal of Geophysical Research: Atmospheres* 2008, 113, (D16).
30. de Hoogh K; Gulliver J; van Donkelaar A; Martin RV; Marshall JD; Bechle MJ; Cesaroni G; Pradas MC; Dedele A; Eeftens M, Development of West-European PM<sub>2.5</sub> and NO<sub>2</sub> land use regression models incorporating satellite-derived and chemical transport modelling data. *Environmental research* 2016, 151, 1–10. [PubMed: 27447442]
31. Hanigan IC; Williamson GJ; Knibbs LD; Horsley J; Rolfe MI; Cope M; Barnett AG; Cowie CT; Heyworth JS; Serre ML, Blending multiple nitrogen dioxide data sources for neighborhood estimates of long-term exposure for health research. *Environmental science & technology* 2017, 51, (21), 12473–12480. [PubMed: 28948787]
32. Song W; Jia H; Li Z; Tang D; Wang C, Detecting urban land-use configuration effects on NO<sub>2</sub> and NO variations using geographically weighted land use regression. *Atmospheric Environment* 2019, 197, 166–176.
33. Gardner M; Dorling S, Neural network modelling and prediction of hourly NO<sub>x</sub> and NO<sub>2</sub> concentrations in urban air in London. *Atmospheric Environment* 1999, 33, (5), 709–719.
34. Kukkonen J; Partanen L; Karppinen A; Ruuskanen J; Junninen H; Kolehmainen M; Niska H; Dorling S; Chatterton T; Foxall R, Extensive evaluation of neural network models for the

- prediction of NO<sub>2</sub> and PM<sub>10</sub> concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment* 2003, 37, (32), 4539–4550.
35. Yeganeh B; Hewson MG; Clifford S; Tavassoli A; Knibbs LD; Morawska L, Estimating the spatiotemporal variation of NO<sub>2</sub> concentration using an adaptive neuro-fuzzy inference system. *Environmental Modelling & Software* 2018, 100, 222–235.
  36. Lu W; Wang W; Leung AY; Lo S-M; Yuen RK; Xu Z; Fan H In Air pollutant parameter forecasting using support vector machines, *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on, 2002; IEEE: 2002; pp 630–635.*
  37. Juhos I; Makra L; Tóth B, Forecasting of traffic origin NO and NO<sub>2</sub> concentrations by Support Vector Machines and neural networks using Principal Component Analysis. *Simulation Modelling Practice and Theory* 2008, 16, (9), 1488–1502.
  38. Mavko ME; Tang B; George LA, A sub-neighborhood scale land use regression model for predicting NO<sub>2</sub>. *Science of the Total Environment* 2008, 398, (1–3), 68–75. [PubMed: 18436280]
  39. Huang Y-K; Luvsan M-E; Gombojav E; Ochir C; Bulgan J; Chan C-C, Land use patterns and SO<sub>2</sub> and NO<sub>2</sub> pollution in Ulaanbaatar, Mongolia. *Environmental research* 2013, 124, 1–6. [PubMed: 23522614]
  40. Liu C; Henderson BH; Wang D; Yang X; Peng Z. r., A land use regression application into assessing spatial variation of intra-urban fine particulate matter (PM<sub>2.5</sub>) and nitrogen dioxide (NO<sub>2</sub>) concentrations in City of Shanghai, China. *Science of The Total Environment* 2016, 565, 607–615. [PubMed: 27203521]
  41. Johnson M; MacNeill M; Grgicak-Mannion A; Nethery E; Xu X; Dales R; Rasmussen P; Wheeler A, Development of temporally refined land-use regression models predicting daily household-level air pollution in a panel study of lung function among asthmatic children. *Journal of Exposure Science and Environmental Epidemiology* 2013, 23, (3), 259. [PubMed: 23532094]
  42. Liu W; Li X; Chen Z; Zeng G; León T; Liang J; Huang G; Gao Z; Jiao S; He X, Land use regression models coupled with meteorology to model spatial and temporal variability of NO<sub>2</sub> and PM<sub>10</sub> in Changsha, China. *Atmospheric Environment* 2015, 116, 272–280.
  43. Rahman MM; Yeganeh B; Clifford S; Knibbs LD; Morawska L, Development of a land use regression model for daily NO<sub>2</sub> and NO<sub>x</sub> concentrations in the Brisbane metropolitan area, Australia. *Environmental Modelling & Software* 2017, 95, 168–179.
  44. He B; Heal M; Reis S, Land-use regression modelling of intra-urban air pollution variation in China: current status and future needs. *Atmosphere* 2018, 9, (4), 134.
  45. Vienneau D; De Hoogh K; Bechle MJ; Beelen R; Van Donkelaar A; Martin RV; Millet DB; Hoek G; Marshall JD, Western European land use regression incorporating satellite-and ground-based measurements of NO<sub>2</sub> and PM<sub>10</sub>. *Environmental science & technology* 2013, 47, (23), 13555–13564. [PubMed: 24156783]
  46. Geddes JA; Martin RV; Boys BL; van Donkelaar A, Long-term trends worldwide in ambient NO<sub>2</sub> concentrations inferred from satellite observations. *Environmental health perspectives* 2015, 124, (3), 281–289. [PubMed: 26241114]
  47. Fisher JA; Jacob DJ; Travis KR; Kim PS; Marais EA; Chan Miller C; Yu K; Zhu L; Yantosca RM; Sulprizio MP, Organic nitrate chemistry and its implications for nitrogen budgets in an isoprene- and monoterpene-rich atmosphere: constraints from aircraft (SEAC 4 RS) and ground-based (SOAS) observations in the Southeast US. *Atmospheric chemistry and physics* 2016, 16, (9), 5969–5991. [PubMed: 29681921]
  48. Kelly JT; Jang CJ; Timin B; Gantt B; Reff A; Zhu Y; Long S; Hanna A, A system for developing and projecting PM<sub>2.5</sub> spatial fields to correspond to just meeting National Ambient Air Quality Standards. *Atmospheric Environment* 2018.
  49. Silvern RF; Jacob DJ; Mickley LJ; Sulprizio MP; Travis KR; Marais EA; Cohen RC; Laughner JL; Choi S; Joiner J; Lamsal LN, Using satellite observations of tropospheric NO<sub>2</sub> columns to infer long-term trends in US NO<sub>x</sub> emissions: the importance of accounting for the free tropospheric NO<sub>2</sub> background. *Atmos. Chem. Phys. Discuss* 2019, 2019, 1–26.
  50. Zheng S; Zhou X; Singh RP; Wu Y; Ye Y; Wu C, The spatiotemporal distribution of air pollutants and their relationship with land-use patterns in Hangzhou city, China. *Atmosphere* 2017, 8, (6), 110.



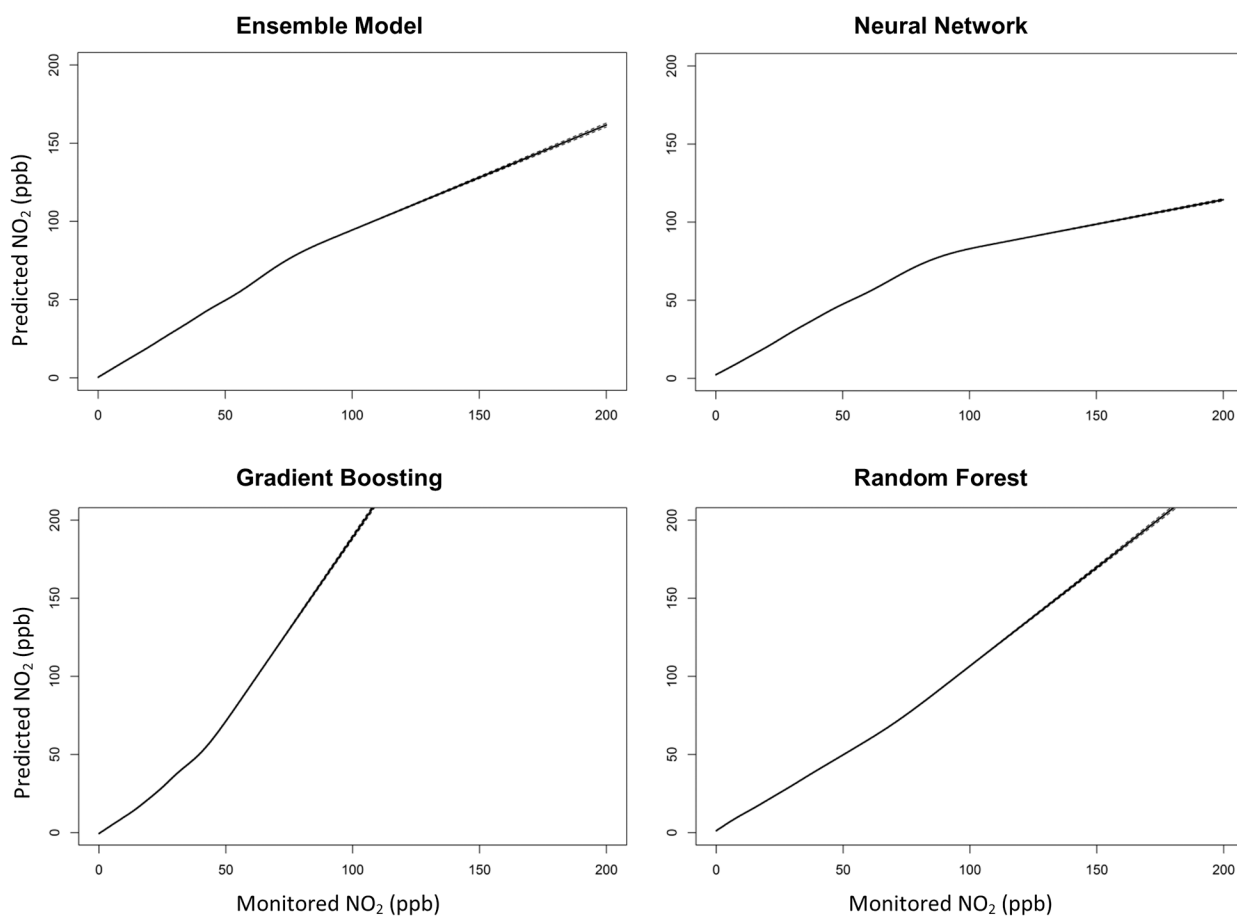
51. Knibbs LD; Hewson MG; Bechle MJ; Marshall JD; Barnett AG, A national satellite-based land-use regression model for air pollution exposure assessment in Australia. *Environmental research* 2014, 135, 204–211. [PubMed: 25282278]
52. Di Q; Amini H; Shi L; Kloog I; Silvern R; Kelly J; Sabath MB; Choirat C; Koutrakis P; Lyapustin A, An ensemble-based model of PM<sub>2.5</sub> concentration across the contiguous United States with high spatiotemporal resolution. *Environment international* 2019, 130, 104909. [PubMed: 31272018]
53. Lin J-T; Martin R; Boersma K; Sneep M; Stammes P; Spurr R; Wang P; Van Roozendael M; Clémer K; Irie H, Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide. *Atmospheric Chemistry and Physics* 2014, 14, (3), 1441–1461.
54. Boersma K; Eskes H; Brinksma E, Error analysis for tropospheric NO<sub>2</sub> retrieval from space. *Journal of Geophysical Research: Atmospheres* 2004, 109, (D4).
55. Bucselá E; Perring A; Cohen R; Boersma K; Celarier E; Gleason J; Wenig M; Bertram T; Wooldridge P; Dirksen R, Comparison of tropospheric NO<sub>2</sub> from in situ aircraft measurements with near-real-time and standard product data from OMI. *Journal of Geophysical Research: Atmospheres* 2008, 113, (D16).
56. Modeling G, MERRA-2 inst3\_3d\_aer\_Nv: 3d,3-Hourly,Instantaneous,Model-Level,Assimilation,Aerosol Mixing Ratio V5.12.4. In NASA Goddard Earth Sciences Data and Information Services Center: 2015.
57. Herman J; Bhartia P; Torres O; Hsu C; Sefort C; Celarier E, Global distribution of UV-absorbing aerosols from Nimbus 7/TOMS data. *J. Geophys. Res* 1997, 102, (16), 911–16.
58. Torres O; Bhartia P; Herman J; Ahmad Z; Gleason J, Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation: Theoretical basis. *Journal of Geophysical Research: Atmospheres* (1984–2012) 1998, 103, (D14), 17099–17110.
59. Kalnay E; Kanamitsu M; Kistler R; Collins W; Deaven D; Gandin L; Iredell M; Saha S; White G; Woollen J; Zhu Y; Leetmaa A; Reynolds R; Chelliah M; Ebisuzaki W; Higgins W; Janowiak J; Mo KC; Ropelewski C; Wang J; Jenne R; Joseph D, The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* 1996, 77, 437–471.
60. Vermote E, MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006 In DAAC, N. E. L. P., Ed. 2015.
61. Inness A; Ades M; Agusti-Panareda A; Barré J; Benedictow A; Blechschmidt A-M; Dominguez JJ; Engelen R; Eskes H; Flemming J; Huijnen V; Jones L; Kipling Z; Massart S; Parrington M; Peuch V-H; Razinger M; Remy S; Schulz M; Suttie M, The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics Discussions* 2018, 1–55.
62. Kloog I; Koutrakis P; Coull BA; Lee HJ; Schwartz J, Assessing temporally and spatially resolved PM<sub>2.5</sub> exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment* 2011, 45, 6267–6275.
63. Zhu Y; Zhan Y; Wang B; Li Z; Qin Y; Zhang K, Spatiotemporally mapping of the relationship between NO<sub>2</sub> pollution and urbanization for a megacity in Southwest China during 2005–2016. *Chemosphere* 2018.
64. Spiegelman D, Evaluating Public Health Interventions: 4. The Nurses' Health Study and Methods for Eliminating Bias Attributable to Measurement Error and Misclassification. *American Journal of Public Health* 2016, 106, (9), 1563–1566. [PubMed: 27509282]
65. Kharol S; Martin R; Philip S; Boys B; Lamsal L; Jerrett M; Brauer M; Crouse D; McLinden C; Burnett R, Assessment of the magnitude and recent trends in satellite-derived ground-level nitrogen dioxide over North America. *Atmospheric Environment* 2015, 118, 236–245.
66. Velders GJ; Geilenkirchen GP; de Lange R, Higher than expected NO<sub>x</sub> emission from trucks may affect attainability of NO<sub>2</sub> limit values in the Netherlands. *Atmospheric environment* 2011, 45, (18), 3025–3033.
67. Soltic P; Weilenmann M, NO<sub>2</sub>/NO emissions of gasoline passenger cars and light-duty trucks with Euro-2 emission standard. *Atmospheric Environment* 2003, 37, (37), 5207–5216.
68. Doll CN; Muller J-P; Morley JG, Mapping regional economic activity from nighttime light satellite imagery. *Ecological Economics* 2006, 57, (1), 75–92.

69. Shi K; Yu B; Hu Y; Huang C; Chen Y; Huang Y; Chen Z; Wu J, Modeling and mapping total freight traffic in China using NPP-VIIRS nighttime light composite data. *GIScience & Remote Sensing* 2015, 52, (3), 274–289.
70. Zhang Q; Seto KC, Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sensing of Environment* 2011, 115, (9), 2320–2329.
71. Wang J; Aegerter C; Xu X; Szykman JJ, Potential application of VIIRS Day/Night Band for monitoring nighttime surface PM<sub>2.5</sub> air quality from space. *Atmospheric Environment* 2016, 124, 55–63.
72. Zhao Y; Zhao B, Emissions of air pollutants from Chinese cooking: A literature review. *Building Simulation* 2018, 11, (5), 977–995.
73. Meng X; Chen L; Cai J; Zou B; Wu C-F; Fu Q; Zhang Y; Liu Y; Kan H, A land use regression model for estimating the NO<sub>2</sub> concentration in Shanghai, China. *Environmental research* 2015, 137, 308–315. [PubMed: 25601733]
74. Celarier E; Brinksma E; Gleason J; Veefkind J; Cede A; Herman J; Ionov D; Goutail F; Pommereau JP; Lambert JC, Validation of Ozone Monitoring Instrument nitrogen dioxide columns. *Journal of Geophysical Research: Atmospheres* 2008, 113, (D15).
75. Martin RV, An improved retrieval of tropospheric nitrogen dioxide from GOME. *Journal of Geophysical Research* 2002, 107, (D20).
76. Weiskopf MG; Webster TF, Trade-offs of Personal Versus More Proxy Exposure Measures in Environmental Epidemiology. *Epidemiology (Cambridge, Mass.)* 2017, 28, (5), 635–643.



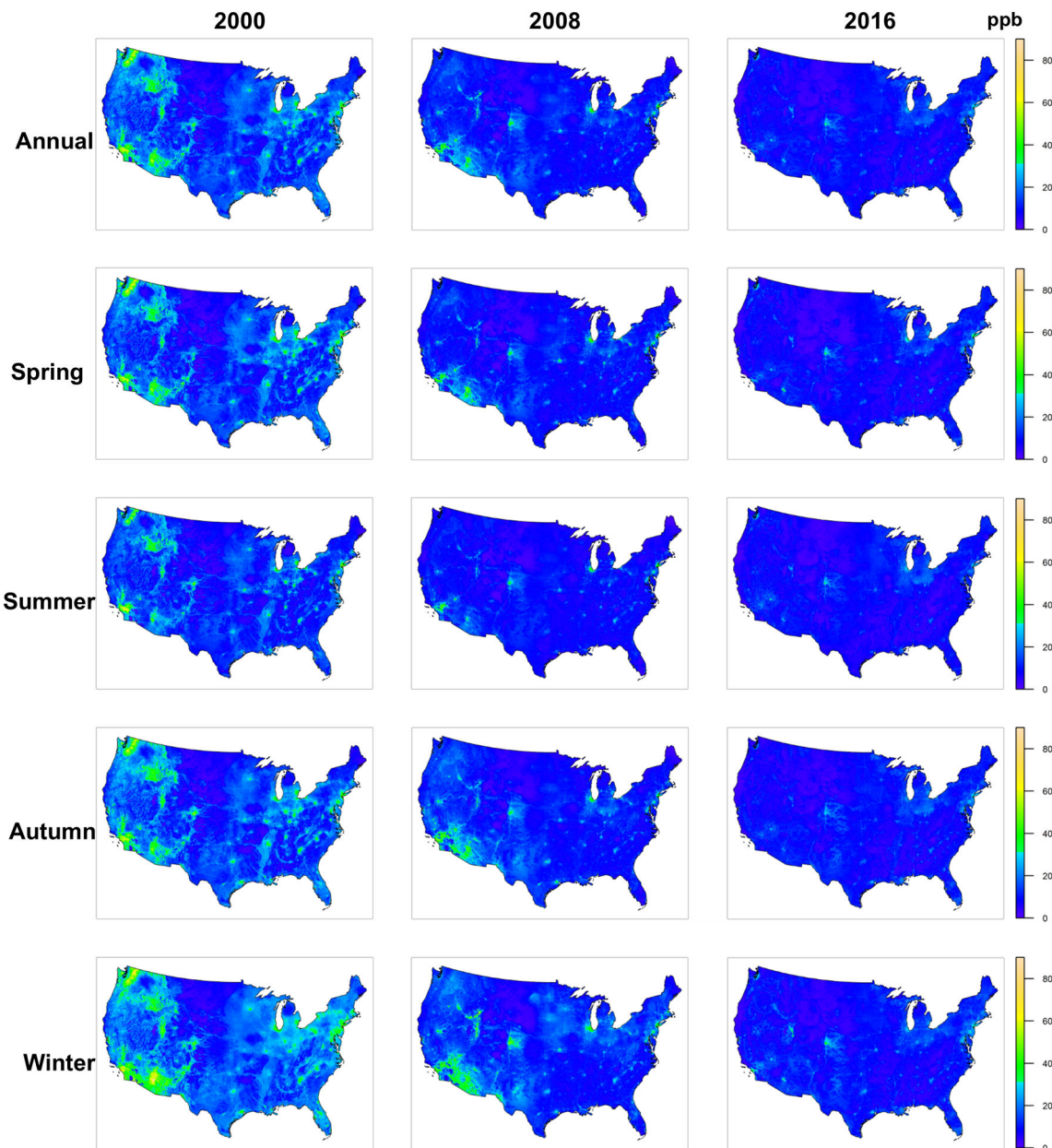
**Figure 1. Cross-validated  $R^2$  at Monitoring Sites and Predicted Uncertainty.**

Left column shows the cross-validated  $R^2$  at each monitoring site ; at right are the monthly mean standard deviations (SD) of the differences between daily monitored  $\text{NO}_2$  and daily predicted  $\text{NO}_2$ , averaged over each  $1 \text{ km} \times 1 \text{ km}$  grid cell for the entire study period. Spring is March to May; summer is June to August; autumn is September to November; winter is December, January, and February.



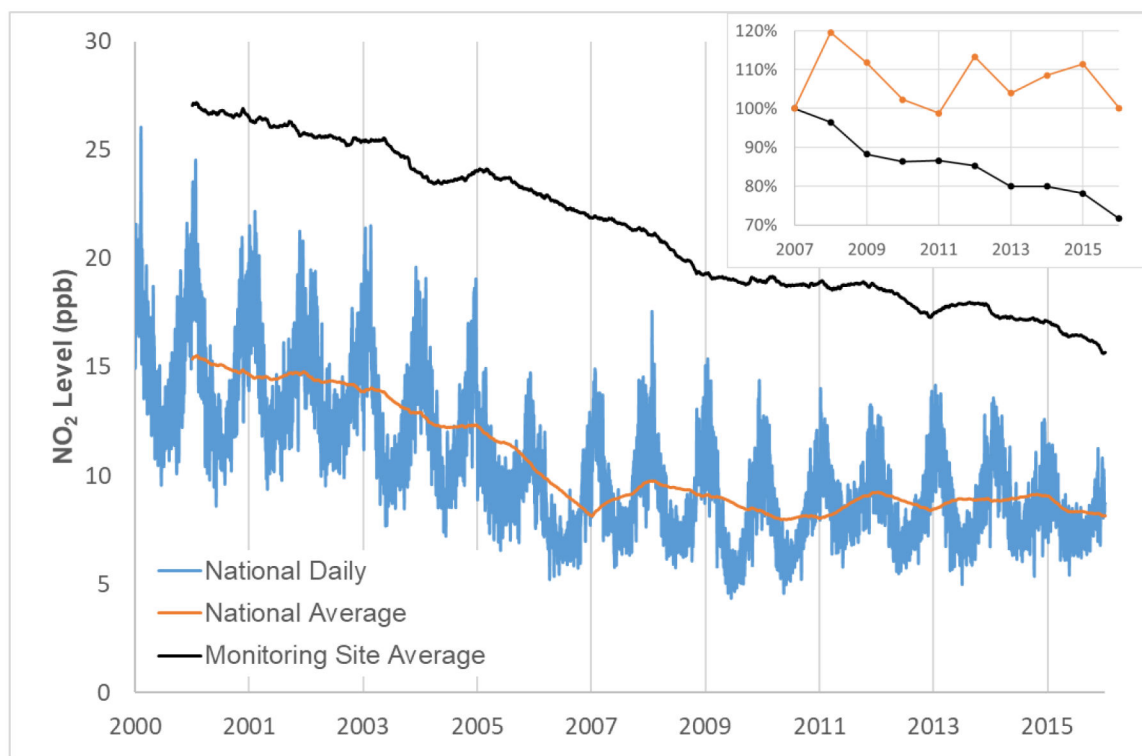
**Figure 2. Linearity between Monitored NO<sub>2</sub> and Predicted NO<sub>2</sub>**

We compared monitored NO<sub>2</sub> and predicted NO<sub>2</sub> from the ensemble model and three machine learners, respectively, with a spline on monitored NO<sub>2</sub> in a generalized additive model. Dashed lines stand for 95% confidence intervals. The 95% confidence intervals are very narrow here because of the large sample size.



**Figure 3. Spatial Distribution of Predicted NO<sub>2</sub>**

The panels show daily NO<sub>2</sub> estimate for 1 km × 1 km grid cells, averaged annually and for four seasons. Here “daily NO<sub>2</sub>” means 1-hour daily maximum NO<sub>2</sub>. Rows show the four seasons, defined in Figure 1.



**Figure 4. Nationwide NO<sub>2</sub> Trend over the Study Period**

We calculated the daily NO<sub>2</sub> for all 1 km × 1 km grid cells in the contiguous U.S., and plotted the daily average over the entire study period (blue line), as well as the one-year moving-average (orange line). For comparison, we also plotted the one-year moving average of NO<sub>2</sub> level at just the monitoring sites (black line). To visualize the relative changes after 2007, we show the timeseries of the annual averaged changes relative to the 2007 NO<sub>2</sub> levels (upper right figure). “Daily NO<sub>2</sub>” means 1-hour daily maximum NO<sub>2</sub>.

**Table 1**

## Cross-validated Model Performance

Year	Ensemble Model						Neural Network	Random Forest	Gradient Boosting
	R <sup>2</sup>	MSE (ppb)	Spatial R <sup>2</sup>	Temporal R <sup>2</sup>	Bias (ppb)	slope	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>
2000	0.692	10.175	0.804	0.602	1.330	0.962	0.668	0.693	0.677
2001	0.762	8.440	0.827	0.709	0.705	0.984	0.721	0.760	0.741
2002	0.780	7.872	0.824	0.734	0.464	0.993	0.745	0.774	0.751
2003	0.801	7.289	0.845	0.751	0.317	0.995	0.789	0.799	0.770
2004	0.782	7.249	0.833	0.734	0.374	0.985	0.754	0.781	0.755
2005	0.767	7.443	0.816	0.730	0.683	0.971	0.748	0.764	0.737
2006	0.771	7.305	0.820	0.735	0.610	0.979	0.750	0.769	0.738
2007	0.782	6.997	0.840	0.730	0.488	0.982	0.759	0.778	0.747
2008	0.785	6.964	0.799	0.764	0.323	0.984	0.744	0.787	0.753
2009	0.804	6.267	0.859	0.764	-0.157	1.000	0.775	0.803	0.765
2010	0.789	6.377	0.829	0.763	0.065	0.993	0.769	0.786	0.749
2011	0.797	6.284	0.846	0.755	-0.090	0.998	0.777	0.798	0.756
2012	0.777	6.263	0.832	0.738	0.029	0.996	0.754	0.772	0.738
2013	0.792	5.999	0.835	0.762	-0.165	1.000	0.755	0.796	0.761
2014	0.787	6.113	0.819	0.761	-0.031	0.997	0.767	0.785	0.756
2015	0.779	6.227	0.817	0.755	-0.059	1.001	0.749	0.775	0.742
2016	0.749	6.459	0.780	0.724	0.334	0.968	0.733	0.749	0.722
<b>Total</b>	0.788	7.146	0.844	0.729	0.233	0.990	0.763	0.787	0.752

Note: the definition of spatial and temporal R<sup>2</sup>'s were based on a previous study<sup>62</sup>. For bias and slope, we regressed daily predicted NO<sub>2</sub> at monitors against daily monitored NO<sub>2</sub> in a linear regression model to obtain slope and bias (the intercept).