Science of the Total Environment 655 (2019) 423–433

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

National $PM_{2.5}$ and NO_2 exposure models for China based on land use regression, satellite measurements, and universal kriging

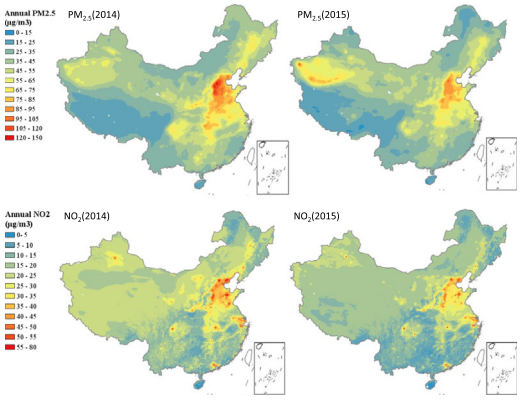
Hao Xu [a,b], Matthew J. Bechle [c], Meng Wang [d,e], Adam A. Szpiro [f], Sverre Vedal [e], Yuqi Bai [a,b,*], Julian D. Marshall [c,**]

[a] The Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China
[b] Joint Center for Global Change Studies (JCGCS), Beijing 100875, China
[c] Department of Civil & Environmental Engineering, University of Washington, Seattle, WA 98195, United States
[d] Department of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo, Buffalo, NY, United States
[e] Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA 98195, United States
[f] Department of Biostatistics, University of Washington, Seattle, WA 98195, United States

HIGHLIGHTS

- First high spatial resolution national LUR models for both NO_2 and $PM_{2.5}$ in China
- Satellite data and kriging are complementary in making predictions more accurate.
- Variable selection models perform similar or better than PLS models.
- 1 km^2 resolution prediction maps will be publicly available for future research.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 July 2018
Received in revised form 8 November 2018
Accepted 8 November 2018
Available online 12 November 2018

Editor: Lidia Morawska

Keywords:
Land use regression
Satellite data
Universal kriging
China
Air pollution

ABSTRACT

Outdoor air pollution is a major killer worldwide and the fourth largest contributor to the burden of disease in China. China is the most populous country in the world and also has the largest number of air pollution deaths per year, yet the spatial resolution of existing national air pollution estimates for China is generally relatively low. We address this knowledge gap by developing and evaluating national empirical models for China incorporating land-use regression (LUR), satellite measurements, and universal kriging (UK). Land use, traffic and meteorological variables were included for model building. We tested the resulting models in several ways, including (1) comparing models developed using forward variable selection vs. partial least squares (PLS) variable reduction, (2) comparing models developed with and without satellite measurements, and with and without UK, and (3) 10-fold cross-validation (CV), Leave-One-Province-Out CV (LOPO-CV), and Leave-One-City-Out CV (LOCO-CV). Satellite data and kriging are complementary in making predictions more accurate: kriging improved the models in well-sampled areas; satellite data substantially improved performance at locations far away from monitors. Variable-selection models performed similarly to PLS models in 10-fold CV, but better in LOPO-CV. Our best models employed forward variable selection and UK, with 10-fold CV R^2 of 0.89 (for both 2014 and

* Correspondence to: Y. Bai, The Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China.
** Correspondence to: J.D. Marshall, Department of Civil & Environmental Engineering, University of Washington, Seattle, WA 98195, United States
E-mail addresses: jdmarsh@uw.edu (J.D. Marshall), yuqibai@tsinghua.edu.cn (Y. Bai).

https://doi.org/10.1016/j.scitotenv.2018.11.125
0048-9697/

2015) for $PM_{2.5}$ and of 0.73 (year-2014) and 0.78 (year-2015) for NO_2. Population-weighted concentrations during 2014–2015 decreased for $PM_{2.5}$ (58.7 μg/m^3 to 52.3 μg/m^3) and NO_2 (29.6 μg/m^3 to 26.8 μg/m^3). We produced the first high resolution national LUR models for annual-average concentrations in China. Models were applied on 1 km grid to support future research. In 2015, >80% of the Chinese population lived in areas that exceeded the Chinese national $PM_{2.5}$ standard, 35 μg/m^3. Results here will be publicly available and may be useful for epidemiology, risk assessment, and environmental justice research.

1. Introduction

Long term exposure to air pollutants such as fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2) has been associated with many adverse health effects, including respiratory and cardiovascular diseases, and increased mortality (Pope III et al., 2011; Kaufman et al., 2016). Epidemiological research on the health effects of air pollution exposure increasingly relies on high spatial resolution air pollution predictions (Beelen et al., 2008; Cesaroni et al., 2014). Land-use regression (LUR) and other empirical modeling approaches are useful tools to improve the accuracy of air pollution exposure estimates and to explore within-urban variability of outdoor air pollutants. LUR employs ground observations and geographic covariates to build a regression model and to estimate concentrations at locations without monitoring data, typically at a city-wide scale (Hoek et al., 2008). Variables corresponding to emission sources (e.g. traffic, population density, nearby pollutant emissions) and dispersion conditions (e.g. elevation, vegetative indices, meteorology) are often included in an LUR model. More recently, LUR and other geostatistical approaches have been used to model fine scale air pollution concentrations over large areas (Novotny et al., 2011; Sampson et al., 2013; Vienneau et al., 2013; Knibbs et al., 2014; Young et al., 2016). Unlike city-wide models, national LUR models typically rely on routine monitoring data instead of purpose-designed monitoring. As a rule of thumb, typically ~40–100 monitors are necessary to build a robust LUR model at a city or regional scale (Basagaña et al., 2012; Wang et al., 2012), whereas continental scale models have typically used ~300–900 monitors (Novotny et al., 2011; Sampson et al., 2013; Knibbs et al., 2014; Young et al., 2016).

In our study, we sought to test several aspects of national-scale spatial LUR model in China and build robust models for both $PM_{2.5}$ and NO_2 for nationwide predictions. Since LUR often incorporates hundreds of variables, variable selection procedures are indispensable for conventional LUR models. Recent studies have used variable reduction approaches such as PLS regression in national LUR-like models, which could effectively choose predictors from dimension-reduced components of all the variables without variable selection procedures and avoid multicollinearity and overfitting (Sampson et al., 2013; Young et al., 2016). However, no one has evaluated whether PLS models could outperform conventional variable selection models. Inclusion of satellite-based air pollution data in LUR has been shown to improve model performance (Vienneau et al., 2013; Knibbs et al., 2014; Bechle et al., 2015). Geostatistical methods such as universal kriging (UK) have also been found to improve model performance when combined with LUR in continental scale models (Sampson et al., 2013; Young et al., 2016). However, these techniques have not been systematically evaluated in a national empirical model for China.

China is experiencing severe and widespread air pollution, along with rapid economic development and urbanization in recent years (Li and Zhang, 2014). $PM_{2.5}$ and NO_2 are two of the major air pollutants in China, which have great health impact on Chinese people. Public health studies conducted on national or regional scales have been critically important for China in advancing environmental policies to improve air quality (J. Liu et al., 2016; Zheng et al., 2017). However, many cohort studies in China still use monitoring data at the district level as the exposure metric due to lack of publicly-available high resolution data sets of air pollution concentrations (W. Liu et al., 2016; Guo et al., 2018). Owing to limitations of data access and lack of publicly-available nationwide monitoring data prior to 2012, LUR models were rarely reported in China. Most reported studies focused on city or regional scale models that relied on limited number of GIS variables (Chen et al., 2010; Meng et al., 2015; Wu et al., 2015; C. Liu et al., 2016). In recent years, satellite-data-driven national models have been emerging in China, which typically estimate the daily relations between a pollutant (e.g., $PM_{2.5}$, NO_2) and satellite-derived aerosol optical depth (AOD) (Ma et al., 2014, 2015; You et al., 2016; He and Huang, 2018) or satellite-derived NO_2 (Zhan et al., 2018). However, existing satellite-based models typically have relatively coarse spatial resolution (10 to 50 km) which may miss intra-urban variations. Additionally, missing data due to cloud cover and weather conditions may increase uncertainty of these daily satellite-based predictions. Inaccurate spatial estimation of air pollution may lead to measurement error caused by spatial misalignment in environmental epidemiology (Gryparis et al., 2008). Accurate annual models are important and necessary for estimating long-term exposure of air pollution and to investigate chronic health outcomes. Incorporating detailed point-based local indicators of air pollution such as road information and land use variables in an LUR framework could provide higher resolution predictions and better model precision. However, no such models have been developed for the entire country in China by now.

Here we developed high-quality national LUR models for China that employed open-source GIS-derived land use and meteorological variables. Satellite data were incorporated to provide additional information especially at locations where monitors are sparse. A point of interest or POI is a feature on a map (or in a geodataset) that occupies a particular point. The number of different types of POIs within a certain buffer length could be used as proxy of different emission sources. Since air pollutants such as $PM_{2.5}$ and NO_2 are typically more concentrated near their source regions and in the boundary layer, we calculated boundary-layer-height-averaged wind speed (BLHA-WS) as a potential predictor to account for diffusion and transport conditions of air pollutants.

Contributions of this paper to the literature include (1) first use of categorized POI data (e.g., gas stations, Chinese restaurants) and BLHA-WS as potential predictors in a national LUR model; (2) robust evaluation of satellite data and UK when these are incorporated in an LUR model in China, accounting for performance near and far from monitoring locations; (3) comparing model performance with forward variable selection and partial-least-squared (PLS) variable reduction methods; and (4) by focusing on long-term average concentrations, providing high spatial resolution prediction maps of $PM_{2.5}$ and NO_2 (1 × 1 km^2) in China, with evaluation of national, regional, and within-urban variations. The publicly available predictions given here will be useful in advancing environmental and health studies in China, including in epidemiology and environmental health.

2. Materials and methods

2.1. Monitoring data

Daily mean ground-level $PM_{2.5}$ and NO_2 concentrations for two years (January 1st, 2014 to December 31st, 2015) were obtained from the China Environmental Monitoring Center (CEMC, http://113.108.

142.147:20035/emcpublish/). Measurements and quality control follow regulations of Chinese Ambient Air Quality Standards (GB 3095—2012) and Ambient Air Quality Index (AQI) technology (HJ 633—2012). Stations missing >25% of daily mean measurements for each pollutant were excluded; annual averages were calculated for each remaining monitor. Finally, 902 and 893 monitors were used to build $PM_{2.5}$ and NO_2 models in 2014, while the numbers increased to 1419 and 1418 in 2015 due to the massive construction of monitoring sites in China since 2012.

2.2. Geographical predictors

We employed a combination of point, buffer, and proximity based geographic variables resulting in 292 unique covariates. Details on each covariate, including the various buffer lengths we employed, are provided in Table S1 of the Supporting Information (SI).

Road network data were extracted from OpenStreetMap data sets (available at http://download.geofabrik.de/asia/china.html), including all roads, major roads, secondary roads and railways. We calculated total length of roads (all, major, and secondary) and railways within 16 buffer lengths from 100 m to 10 km. We also calculated distance to nearest major roads, secondary roads and railways.

The percentage of land cover types for eight categories was computed within 11 sizes of moving windows (from 300 m to 30 km). Land cover type data sets were derived from the Finer Resolution Observation and Monitoring of Global Land Cover data set (FROM-GLC) with 30 m resolution (Gong et al., 2013), resampled from sinusoidal projection to Albers projection system using nearest neighbor assignment.

Five types of POIs (i.e., gas stations, heat suppliers, polluting factories, bus stops and Chinese restaurants) were extracted using Amap Application Programming Interface (API) (available at http://lbs.amap.com/api/webservice/guide/api/search/) based on categories and keywords (see Section S4 in the SI). Categorized POIs may indicate local land uses that are not well captured by other variables and have been used in city-scale LUR, but to our knowledge have not previously been employed in a national LUR model. For example, Chinese restaurants are restaurants with Chinese-style cooking, which are important source of air pollution in China (Wu et al., 2017). To capture both local and regional transport of air pollutants, we calculated POI counts using 22 buffer lengths from 100 m to 50 km.

Previous research suggests that remotely sensed fire count data could improve $PM_{2.5}$ prediction accuracy, and will have good prediction power when the buffer zone reach 50 km (Hu et al., 2014). To capture fire emissions, we used number of fire spots within 10 buffer lengths from 5 km to 100 km using Moderate-resolution Imaging Spectroradiometer (MODIS) Global Monthly Fire Location Product (MCD14ML, available at https://earthdata.nasa.gov/active-fire-data).

Other potential predictor variables included elevation (China 1 km Digital Elevation Model data based on Shuttle Radar Topography Mission, available at http://www.resdc.cn), population density calculated from Landscan 2015 population data set (Bright et al., 2016), Normalized Difference Vegetation Index (NDVI) & Enhanced Vegetation Index (EVI) (derived from MODIS MOD13A3 monthly NDVI data set, available at https://lpdaac.usgs.gov/data_access/data_pool), and coordinates (x and y coordinate in China Albers Equal Area Conic coordinate system).

2.3. Meteorological data

Boundary layer height, temperature (at 2 m), dew point temperature (at 2 m), surface pressure and wind speed (at 10 m) were extracted from the European Reanalysis (ERA) Interim reanalysis data monthly means of daily means product (Dee et al., 2011). We derived precipitation data from a 0.25° × 0.25° interpolated observational product based on 2419 monitoring stations in China (available at http://data.cma.cn/). Relative humidity (RH) and BLHA-WS were also calculated (for details, see Section S1 and S2 in the SI) (Apte et al., 2012). All meteorological data were averaged to annual means and re-sampled to 1 km grid cells using bilinear interpolation, which performs linear interpolation first in one direction, and then again in the other direction.

2.4. Satellite-based air pollution data

To reduce the influence of possible deficiency of monitoring sites and improve the modeling accuracy, we incorporated satellite-based air pollution data. Satellite measurements of air pollution are derived from observations; in general, they reflect ambient conditions and therefore contributions from all emission sources. Satellite-based estimates have been previously developed for ground-level $PM_{2.5}$ (Zheng et al., 2016; Xue et al., 2017) and NO_2 (Novotny et al., 2011; Vienneau et al., 2013; Young et al., 2016). We employed a publicly available 0.1° resolution global annual Satellite-Derived $PM_{2.5}$ product (http://fizz.phys.dal.ca/~atmos/martin/?page_id=140), which combined AOD retrievals from the NASA MODIS, Multi-angle Imaging SpectroRadiometer (MISR), and Sea-Viewing Wide Field-of-View Sensor (SeaWIFS) instruments with GEOS-Chem output to estimate ground-level $PM_{2.5}$. There are two versions of the 0.1° satellite $PM_{2.5}$ data set, one as described above and a product calibrated to ground-based measurements using geographically-weighted regression (GWR) (Van Donkelaar et al., 2016). We employed the non-GWR data set, to avoid including satellite data calibrated to measurements from monitoring sites used for constructing our models. Previous studies suggest tropospheric NO_2 column data are sufficient to track spatial patterns in ground-level NO_2 (Knibbs et al., 2014; Bechle et al., 2015). We directly employed monthly mean Ozone Monitoring Instrument (OMI) tropospheric NO_2 column data with 0.125° resolution from the Derivation of OMI tropospheric NO_2 (DOMINO) product (version 1.0.2, collection 3; available at http://www.temis.nl). A previous study has suggested that satellite NO_2 data could improve $PM_{2.5}$ model performance (Zheng et al., 2016), therefore, we also included satellite NO_2 data as a potential predictor in our $PM_{2.5}$ model. We calculated annual averages for 2014 and 2015 and then converted all of the satellite data into 1×1 km^2 grid cells using bilinear interpolation.

2.5. Statistical model building

We developed geostatistical models for mainland China for both $PM_{2.5}$ and NO_2 for the years of 2014 and 2015. We used R X64 3.4.0 'stat' and 'gstat' packages (Pebesma and Heuvelink, 2016). To determine the added value of UK and satellite data, we developed models with and without UK, and with and without satellite data.

2.5.1. Forward variable selection

At the first stage, forward variable selection approach was used to build a linear model. For each pollutant ($PM_{2.5}$, NO_2) and year, we explored four potential regression models: satellite $PM_{2.5}$ included; satellite NO_2 included; both satellite $PM_{2.5}$ and NO_2 included; and, without satellite data. We followed conventional supervised forward stepwise regression, with 290 to 292 independent variables as inputs in our variable selection procedures (Eeftens et al., 2012). Briefly, the independent variable most correlated with the dependent variable was added to the model. In subsequent steps, the remaining variable that gave the largest improvement in adjusted R^2 was added to the model if (1) the variance inflation factor (VIF, a check for multi-collinearity) of the variable was <5; (2) the p-value of the variable was <0.05; and (3) the direction of existing variables in the model did not change, in order to make sure that selected variables were significant without multi-collinearity. This procedure was repeated until the increase in adjusted-R^2 for an additional variable was <0.5% or no variable met the aforementioned criteria. We allowed multiple buffer lengths per variable (e.g. major road length) to be selected into the model as long as they follow our criteria (Henderson et al., 2007).

We also tried alternative metrics (e.g., F value, 10-fold R^2, adjusted R^2, Akaike Information Criterion (AIC)), and alternative algorithms such as backward selection, to see whether those attributes strongly influenced results from the model-building process.

Monte Carlo iterations were used to evaluate the over-fitting risk; we simulated reduced data availability via Monte Carlo sampling (500 iterations per number-of-monitors) at lower number of monitors, and then compared model fitted and cross-validated model performance. Other regression diagnostic tests included checking the normality of residuals, heteroscedasticity, and spatial autocorrelation of residuals using Moran's I.

2.5.2. Universal kriging

At the second stage, we incorporated a spatial smoothing approach (UK). Kriging will account for spatial autocorrelation in the model residuals. We leveraged a first-order polynomial function in UK as the external drift, whose independent variables were obtained from the forward selection stage. Then the spatial dependence was modeled using the exponential variogram model (for details, see Section S5 in the SI).

2.5.3. Partial least squares (PLS)

Some previous empirical models for the US were built using partial least squares (PLS) (Sampson et al., 2013; Young et al., 2016). PLS reduces the dimensions of the many predictor variables and avoids having to use a variable selection procedure. To evaluate the comparative performance of variable-selection models and PLS models, we also built PLS models with and without satellite data, and with and without UK, using all potential predictors for 2015. Selection of PLS components were based on 10-fold CV results, using the R command *selectNcomp*, which checks whether the root-mean-squared-error (RMSE) of models with fewer components are significantly larger than in the model with global minimum RMSE. The most parsimonious model not significantly worse than the model with global minimum RMSE was selected (see Fig. S6 in the SI). Details of our model building approaches are described elsewhere (Young et al., 2016). Briefly, the satellite data were used directly as a covariate in the PLS procedure and spatially varying PLS components were used in building UK models.

2.6. Cross-validation and model assessment

Our core model evaluation mainly used two types of cross-validation approaches: conventional 10-fold cross-validation (10-fold CV, as default CV method), and Leave-One-Province-Out cross-validation (LOPO-CV). These approaches divided the data into model-building and model-testing sub-sets. For conventional cross-validation, all monitoring sites were randomly divided into 10 groups. Nine groups were then used to train the model and the remaining one group to test the model. This process was repeated 10 times, until all the groups were tested, resulting in "out-of-sample" predictions at all monitoring sites.

Most monitors are in cities, and often are somewhat near to (in the same city as) other monitors. To explore model performance at locations without a nearby monitor, we performed LOPO-CV on each model, wherein we consecutively excluded monitors from one province.

To further address the city-scale performance of our national models, we additionally conducted Leave-One-City-Out cross-validation (LOCO-CV), wherein we excluded all monitors from a city during model-building, and then compared model results against (held-out) monitoring data for that city.

Statistics from 10-fold CV and LOPO-CV/LOCO-CV used to assess model performance include mean-square-error-based R^2 (assessing deviation around the 1:1 line) (Szpiro et al., 2011) and RMSE. In addition, in order to demonstrate model differences spatially, we calculated and mapped differences between the national predictions of the different models.

3. Results

3.1. Model results and comparison

For 2014 and 2015, respectively, the number of monitors that met our inclusion criteria was 893(NO_2)/902($PM_{2.5}$) and 1418(NO_2)/1419 ($PM_{2.5}$). Descriptive statistics for those input data are in Table 1. Substantial variability of both pollutants and the drop of annual mean concentrations from 2014 to 2015 can be observed in the table. The large difference in the number of monitors between 2014 and 2015 is the result of a rapid monitoring network expansion in China since 2012; there were a total of 944 monitors in 2014 and 1494 in 2015. Most monitors were located in the eastern part of China, where population density is relatively high, while the number of monitors in the western part was very sparse (see Fig. S1 in the SI). According to the detailed information of monitoring sites, most of them were set near public institutions and schools, and a few of them were set near industry or busy roads. Fig. 1 presents the R^2 of the 2015 models based on variable selection method. Table 2 shows the summary of all the models built for 2015 (2014 model results are presented in Table S8 in the SI).

3.1.1. Variable-selection models without UK

Detailed descriptions of all the linear models based on variable selection are in Tables S6 and S7 in the SI. $PM_{2.5}$ models explained 69%–76% variation in 2014 and 62%–71% in 2015. Including satellite-derived $PM_{2.5}$ and NO_2 together greatly improved 10-fold CV R^2 for $PM_{2.5}$ (e.g., 16% improvement in 2015) compared to non-satellite model. The prediction ability of satellite-derived $PM_{2.5}$ was slightly better than satellite-derived NO_2 when only using one set of satellite data. The best linear model based on variable selection ($PM_{2.5}$-4) suggests that agricultural emission source (percentage of cropland), indirect traffic/urbanization indicators (number of gas stations/bus stops, road length) and meteorological conditions (BLHA-WS, RH) are important predicting factors to $PM_{2.5}$ models. In 2015, the number of gas stations within 10 km and BLHA-WS increased 3.9% and 2.5% of explained variance respectively for our best linear model for $PM_{2.5}$ ($PM_{2.5}$-4), which were most important variables except for satellite $PM_{2.5}$.

Similarly, including NO_2 satellite data substantially improved the NO_2 model CV R^2, e.g., by 12 percentage points in 2015, whereas satellite-derived $PM_{2.5}$ was not selected into any NO_2 model (hence, it is not displayed for NO_2 models in Fig. 1). Model CV R^2's were lower for NO_2 than for $PM_{2.5}$ (0.61 [NO_2] vs. 0.76 [$PM_{2.5}$] in 2014; 0.66 [NO_2] vs. 0.71 [$PM_{2.5}$] in 2015). Key predictor variables for the NO_2 models included urbanicity (percentage of impervious surfaces, percentage of forest, number of heating suppliers), indirect traffic/urbanization indicators (number of gas stations/bus stops) and meteorological conditions (BLHA-WS). In 2015, the number of heat suppliers within 7500 m and BLHA-WS increased 5.7% and 1.9% of explained variance respectively for our best linear model for NO_2 (NO_2-2), which were also the most important variables except for satellite NO_2. Model buffer lengths were generally smaller for NO_2 than $PM_{2.5}$, consistent with $PM_{2.5}$ being a more regional pollutant than NO_2.

Results from our test to evaluate the potential for over-fitting in our year-2015 models is presented in Fig. 2 (year-2014 results are in Fig. S2 in the SI). The evaluation metric (model R^2) converges at ~ 400–450 monitoring sites, suggesting that the number of monitoring sites in our models were more than sufficient, with little risk of over-fitting.

Table 1
Descriptive statistics of monitoring data.

Pollutant	Year	N	Min	25th	Mean	Median	75th	Max	SD
$PM_{2.5}$ (μg/m^3)	2014	902	17.2	46.4	61.4	61.0	73.0	136.2	20.0
$PM_{2.5}$ (μg/m^3)	2015	1419	9.9	39.1	52.4	52.0	62.4	128.8	18.4
NO_2 (μg/m^3)	2014	893	6.5	28.8	37.8	37.7	46.6	78.2	12.8
NO_2 (μg/m^3)	2015	1418	5.8	22.4	31.8	31.4	40.1	75.9	12.4

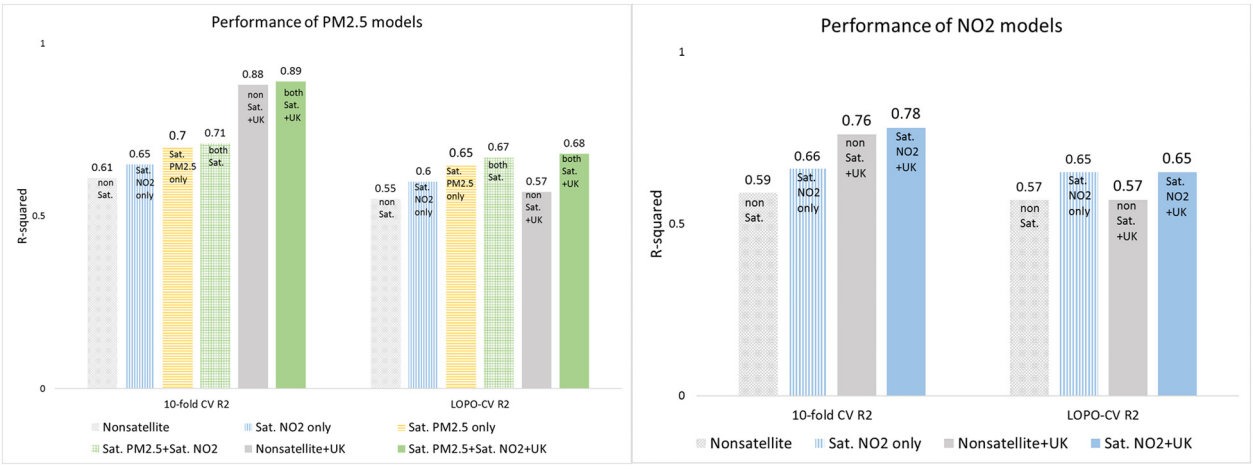


Fig. 1. Model performance for year-2015 models. (Sat. = satellite) Other statistics (e.g., RMSE; results for year-2014) are in the SI.

The difference between 10-fold CV R^2 and LOPO-CV R^2 is slightly larger in $PM_{2.5}$ models than in NO_2 models (e.g. 0.71 vs. 0.67 for $PM_{2.5}$ and 0.66 vs. 0.65 for NO_2 in 2015), suggesting that NO_2 predictors are slightly more capable of capturing spatial variance at locations far away from training samples, but differences are modest. We also classified model R^2 by rural, suburban and urban areas based upon population density using our best performing linear models ($PM_{2.5}$-4 and NO_2-2) (see Fig. S3 in the SI). The $PM_{2.5}$ linear models yielded the best predictions in urban areas (CV R^2: 0.79 in 2014, 0.73 in 2015); NO_2 linear models gave best predictions in rural areas (CV R^2: 0.64 in 2014, 0.71 in 2015), but also performed well in urban areas.

Table 2
Summary of all the models built for 2015 (VS = variable selection, PLS = partial least square, Sat. = satellite, UK = universal kriging, LOPO = Leave One Province Out, the shading rows are the final selected models).

Model	VS	PLS	Sat. PM2.5	Sat. NO2	UK	10-fold CV R^2	LOPO CV R^2	10-fold RMSE ($\mu g/m^3$)	LOPO RMSE ($\mu g/m^3$)
$PM_{2.5}$-1	✓					0.61	0.55	11.4	12.3
$PM_{2.5}$-2	✓		✓			0.70	0.65	10.2	10.9
$PM_{2.5}$-3	✓			✓		0.65	0.60	10.9	11.7
$PM_{2.5}$-4	✓		✓	✓		0.71	0.67	9.9	10.6
$PM_{2.5}$-5	✓				✓	0.88	0.57	6.3	12.1
$PM_{2.5}$-6	✓		✓	✓	✓	0.89	0.68	6.3	10.4
$PM_{2.5}$-7		✓	✓	✓		0.70	0.64	10.0	11.1
$PM_{2.5}$-8		✓	✓	✓	✓	0.89	0.66	6.1	10.8
NO_2-1	✓					0.59	0.57	7.9	8.1
NO_2-2	✓			✓		0.66	0.65	7.2	7.4
NO_2-3	✓				✓	0.76	0.57	6.1	8.1
NO_2-4	✓			✓	✓	0.78	0.65	5.9	7.4
NO_2-5		✓		✓		0.63	0.57	7.6	8.1
NO_2-6		✓		✓	✓	0.76	0.60	6.1	7.9

3.1.2. Variable-selection models with UK

Fig. 1 also presents the results of model performance of variable-selection models with addition of UK. Incorporating UK improved the 10-fold CV R^2 for $PM_{2.5}$ and NO_2; increases were 0.17 to 0.27 for non-satellite models, 0.12 to 0.18 for satellite models. The differences in 10-fold CV R^2 between kriging models with and without satellite data were small (ranging from 0.00 to 0.03), however, under LOPO-CV, kriging models were improved with the addition of satellite data (R^2 increased 0.08 to 0.11). Under 10-fold CV, the best models ($PM_{2.5}$-6 and NO_2-4) consistently included satellite data with UK.

3.1.3. Comparison with PLS models

Using 10-fold CV R^2, model performance for $PM_{2.5}$ and NO_2 was similar for PLS models with satellite data and UK (R^2: 0.89 [$PM_{2.5}$], 0.76 [NO_2] in 2015) as for conventional variable-selection models with satellite data and UK. However, using LOPO-CV, PLS models with satellite data and UK had R^2 values of 0.66 ($PM_{2.5}$) and 0.60 (NO_2) – slightly worse than similar variable-selection models with satellite data and UK (0.68 and 0.65 for $PM_{2.5}$ and NO_2, respectively). Our PLS models used from 6 to 8 components. Because PLS models performed no better than the conventional variable-selection models (see Table 1 and Table S8 in the SI), and in the case of LOPO-CV were worse, we chose the variable-selection model with satellite data and UK as our core model.

3.2. Model predictions and assessments

Fig. 3 shows our annual prediction maps for China based on our best performing models ($PM_{2.5}$-6 and NO_2-4) in 2015 (for 2014, see Fig. S11 in the SI), consisting of 9.6 million 1×1 km^2 grid cells. The most polluted areas for $PM_{2.5}$ were in the Beijing-Tianjin-Hebei urban agglomeration: predicted annual-average concentrations were above 85 $\mu g/m^3$ in 2014 and above 75 $\mu g/m^3$ in 2015. In eastern and northern China, $PM_{2.5}$ concentrations were similar (above 60 $\mu g/m^3$ in 2014 in most areas). Regions in central and western China including Hunan and Hubei provinces and the Sichuan basin also exhibited comparatively higher concentrations of $PM_{2.5}$. Also, $PM_{2.5}$ concentrations were high in the southern part of Xinjiang autonomous region where transported dust from deserts might be a major source. For NO_2, the most polluted areas were urban areas, especially the Beijing-Tianjin-Hebei urban agglomeration, Shandong province, the Yangtze River Delta and the Pearl River Delta. These regions are more economically developed and densely populated and have more industrial sources.

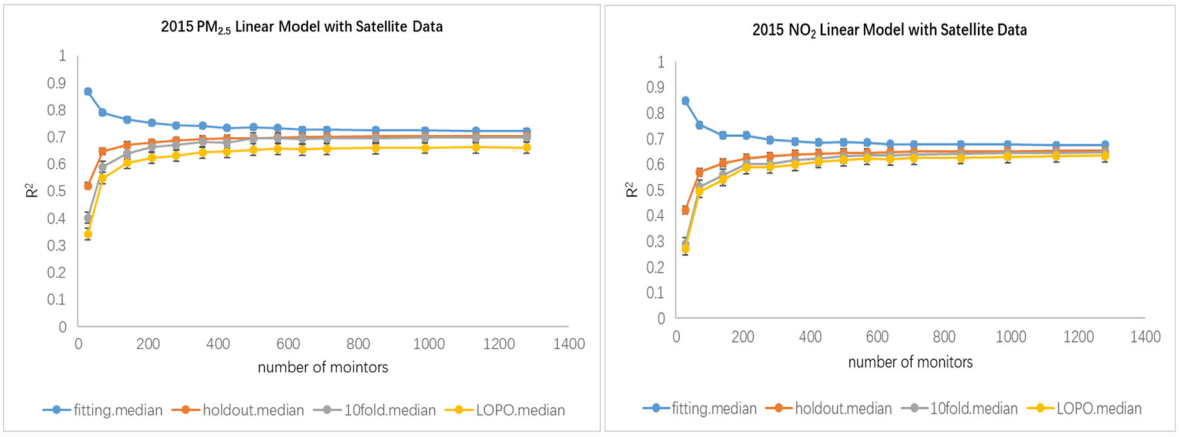


Fig. 2. Median and interquartile range R^2 for Monte Carlo random sampling for n training monitors employed in model building (left: 2015 $PM_{2.5}$ LUR model with satellite data; right: 2015 NO_2 LUR model with satellite data). *Fitting* uses N monitors to fit the model; *holdout* uses N monitors to build the model and the rest of monitors to test the model; *10-fold* used N monitors to build and test model using 10-fold CV; *LOPO* used N monitors to build and test model using LOPO-CV. (N = 28, 70, 140, 210, 280, 355, 425, 500, 570, 640, 710, 850, 990, 1135, 1280).

Maps for predictions from the various approaches (see Fig. S8 in the SI) suggest consistent patterns in spatial variation of the pollutants. However, the inclusion of satellite data had a more regional impact, and typically provided information in areas with few monitors (e.g. Xinjiang, Tibet and Northeastern China). Kriging mainly created adjustments in urban areas, where monitor density was greater.

Fig. 3 also shows year-2015 spatial predictions from the four models (with and without satellite data; with and without UK) along transects across two major cities in Northern and Southern China (Beijing and Guangzhou). For $PM_{2.5}$ models, although the addition of satellite data and kriging resulted in better model performance, the spatial concentration gradients became smoother with some potential loss of spatial variations. For predicted concentrations, differences between satellite and non-satellite models were comparatively smaller for NO_2 than for $PM_{2.5}$. Within-urban variation was greater for NO_2 predictions than $PM_{2.5}$ predictions. The patterns described here are for 2015; patterns for 2014 were similar (see Fig. S11 in the SI).

We used predictions derived from CV of our best national models ($PM_{2.5}$-6 and NO_2-4) to calculate citywide R^2 and RMSE for 10 major cities. We selected the 10 cities with the largest number of monitors. 10-Fold CV and LOCO-CV were used for model evaluation. As shown in Table 3, in 2015, 10-fold CV R^2 values of $PM_{2.5}$ model ranged from 0.02 to 0.79, with RMSE ranging from 3.9 to 6.8. Although city-scale R^2 of $PM_{2.5}$ model were relatively poor (most cities are below 0.10), the RMSE values in these cities were excellent (most below 5.0); that result suggests that the $PM_{2.5}$ model was accurately predicting (average) concentrations in each city, but that within-city spatial variability was either too low or not well captured by the model. In contrast, the NO_2 model in most cities performed reasonably well (eight cities had a 10-fold CV $R^2 > 0.50$). Table 3 also shows that the standard deviation values for $PM_{2.5}$ were much lower than the values for NO_2, further indicating that there is much more within-city variability in the NO_2 concentrations than the $PM_{2.5}$ concentrations. LOCO-CV reflected model performance when models were built excluding the monitors in the specific city; for cities with monitors, performance for our final models ($PM_{2.5}$-6 and NO_2-4) will generally be better than LOCO-CV results. Under LOCO-CV, the RMSE values were higher in Northern cities such as Beijing, Tianjin and Shenyang, where concentrations of air pollutants were also relatively high.

Fig. 4 shows cumulative exposure nationally for $PM_{2.5}$ and NO_2, based on best performing models ($PM_{2.5}$-6 and NO_2-4). For $PM_{2.5}$, >90% of people in year-2014 lived in locations that exceeded China's national standard, 35 μg/m^3 (same as WHO IT1); this number reduced to 83% in 2015. Average concentrations were above 75 μg/m^3 $PM_{2.5}$ for >20% of people, and above 40 μg/m^3 NO_2 for >20% of people.

4. Discussion

Our research developed and rigorously tested national prediction models for $PM_{2.5}$ and NO_2 in China using large open-source data sets and state-of-the-art modeling. Factors influencing air quality may vary by year; we built separate models for each year (2014; 2015) and pollutant. Our final models ($PM_{2.5}$-6 and NO_2-4) incorporated satellite data and UK and exhibited good predictive power (10-fold CV R^2: 0.89 [$PM_{2.5}$], 0.73 to 0.78 [NO_2]).

4.1. Model performance

We compared model performance of our variable-selection models with PLS models. Although PLS obviates the need for variable selection and deals with multi-collinearity, it is computationally intensive for making national predictions since all of the geographic variables need to be used for extracting the individual PLS components. In addition, not all potential variables are correlated with the dependent variables ($PM_{2.5}$ and NO_2); since PLS uses all of the variables, this aspect raises potential concern of overfitting. Further, it is not straightforward to demonstrate the contribution of each geographic variable to overall model predictions. We found that PLS (alone or combined with UK) performs similarly to, or in some cases not as well as, multivariate linear regression with variable selection (alone or combined with UK).

Performance of our final $PM_{2.5}$ models ($PM_{2.5}$-6) were comparable to those reported from previous studies in the US (10-fold CV R^2 0.89 vs. 0.88), (Sampson et al., 2013) while performance of the NO_2 models was slightly worse than that reported from the US (10-fold CV R^2 0.78 vs. 0.85) (Young et al., 2016); potential explanations include that in China relative to the US monitoring sites might be located more unevenly, or explanatory variables we employed are less relevant to pollution, or pollution may be generally less correlated with land use.

Some previous studies used geographically weighted regression (GWR) instead of LUR to account for spatially varying coefficients of impacting factors in national models (Ma et al., 2014; You et al., 2016), but they did not intensively evaluate model performance in sparsely monitored areas. We also tested GWR for $PM_{2.5}$ based on our LUR variables. However, the final CV results based on GWR (10-fold CV $R^2 = 0.80$ and 0.81 for 2014 and 2015, respectively) were no better than our method based on LUR with UK (10-fold CV $R^2 = 0.89$ for both 2014 and 2015). Under LOPO CV, the R^2 of GWR models were extremely low ($R^2 = 0.61$ and 0.01 for 2014 and 2015, respectively) due to substantial bias at regions where monitors are sparse (e.g. Xinjiang, Tibet), indicating potential difficulty in using GWR in regions with sparsely monitored areas.

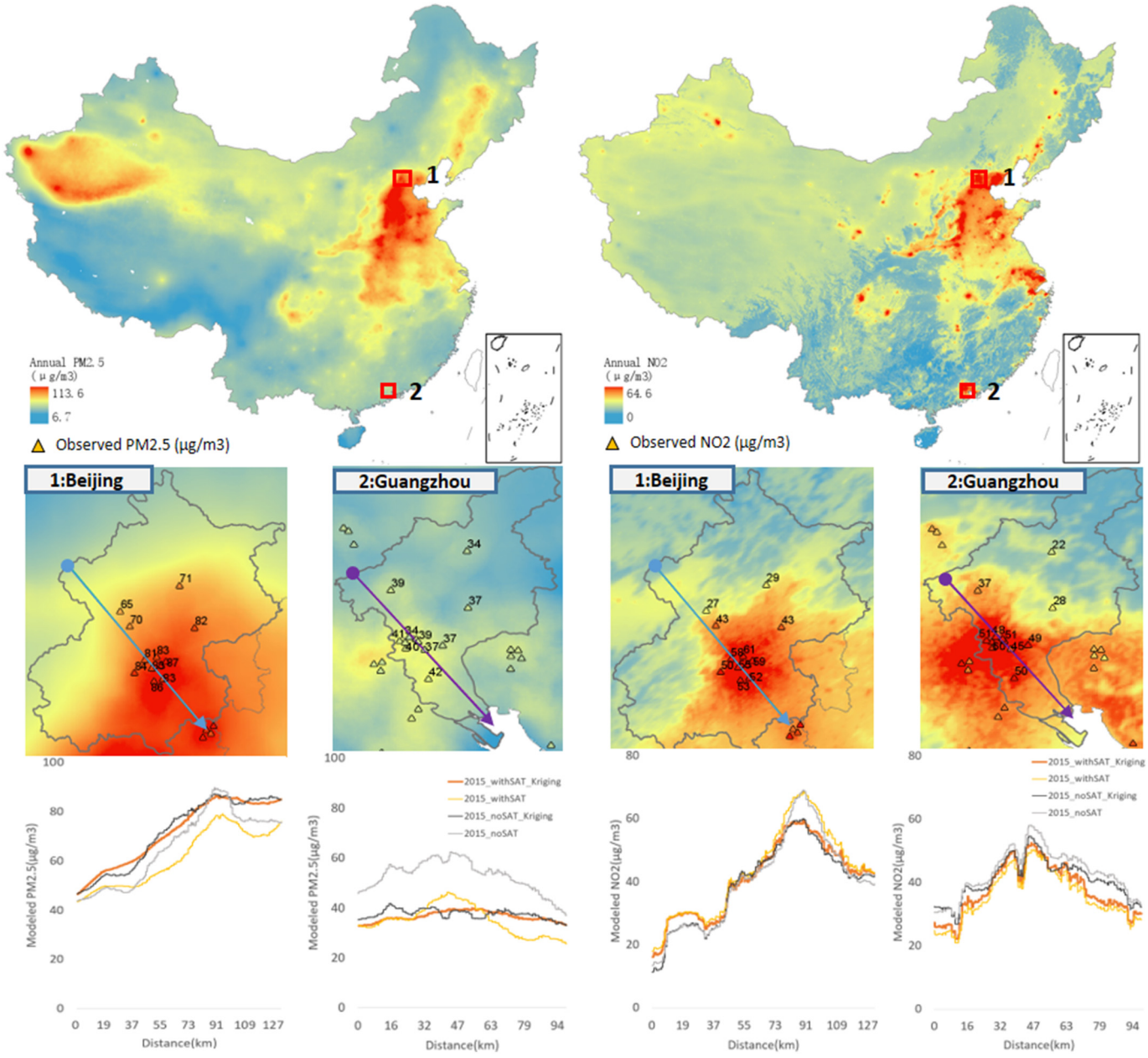


Fig. 3. National (top) and city-level (middle) predictions derived from our best models ($PM_{2.5}$-6 and NO_2-4) in 2015 and profile plots (bottom) of concentrations in two major cities in China based on four different models (left:$PM_{2.5}$ right:NO_2). Profile plots are derived from 1 × 1 km^2 estimates along the transect shown for each city. Monitor locations are indicated with triangle symbols in city-level maps along with corresponding monitor concentration.

Most existing large scale empirical models in China are daily models based on constructing relations between satellite data and ground observations (details regarding existing models are in Table S10 in the SI). Reported overall R^2's of these models are 0.62 to 0.80. Very few publications reported annual R^2. Zhan et al. used machine learning to model daily $PM_{2.5}$ in 2014 at 50 km grid cells in China, yielding an annual R^2 of 0.84 based on 10-fold CV (Zhan et al., 2017). Xue et al. estimated daily $PM_{2.5}$ in 2014 with 0.1° spatial resolution combined satellite data and CMAQ model output, yielding an annual R^2 of 0.87 based on site-based CV (Xue et al., 2017). Zhan et al. predicted 0.1° spatial resolution daily NO_2 from 2013 to 2016 using random forest and spatiotemporal kriging, yielding an annual R^2 of 0.68 (Zhan et al., 2018). Our parsimonious models with relatively low computational cost had comparable performance without requiring complex algorithms. Furthermore, our method provided very fine scale predictions at 1 km resolution while explicitly assessing the relationships between the pollutants and land use variables. Prior research has applied the LUR approach for smaller regions in China; for example, Yang et al. developed a regional LUR for the Pearl River Delta region (Yang et al., 2017).

4.2. Contribution of satellite data and UK

While satellite NO_2 could be used as a proxy of anthropogenic NO_x emissions which are highly related to energy consumption (Zhang et al., 2012), we also incorporated satellite NO_2 as a potential predictor in $PM_{2.5}$ models. When including all potential predictors, both satellite $PM_{2.5}$ and satellite NO_2 were consistently selected into $PM_{2.5}$ models, while only satellite NO_2 was consistently selected into NO_2 models. Satellite data substantially improved LOPO-CV performance of the models, suggesting that satellite data provides additional spatial information on air pollutant concentrations that land use and meteorological variables could not provide. It is worth noting that satellite NO_2 could

Table 3
City-scale performance of final national models for 2015. (LOCO = Leave One City Out, Obs. = Observations, SD = Standard Deviation).

City	N	10-Fold R^2	10-Fold RMSE (μg/m^3)	LOCO R^2	LOCO RMSE (μg/m^3)	LOCO mean (μg/m^3)	Obs. mean (μg/m^3)	LOCO SD (μg/m^3)	Obs. SD (μg/m^3)
(1) $PM_{2.5}$ model									
Chongqing	17	0.06	6.8	0.24	7.1	58.9	54.7	3.0	6.7
Beijing	12	0.79	4.1	0.78	10.0	71.3	80.3	9.3	7.4
Tianjin	11	0.18	6.0	0.21	15.0	84.8	72.4	5.2	5.2
Hangzhou	11	0.76	5.4	0.78	5.9	57.6	54.3	5.8	9.5
Shenyang	11	0.55	4.3	0.44	8.6	63.7	70.7	3.0	6.6
Guangzhou	10	0.04	3.9	0.10	4.0	38.9	38.4	4.0	3.0
Wuhan	10	0.09	4.4	0.21	6.6	74.1	69.1	4.6	4.2
Changchun	10	0.34	4.2	0.54	4.9	61.1	64.2	2.4	5.5
Changsha	10	0.04	4.7	0.17	5.0	57.4	60.4	2.1	4.6
Shanghai	9	0.02	4.8	0.30	8.6	60.4	54.4	3.6	3.8
(2) NO_2 model									
Chongqing	17	0.65	6.4	0.76	7.8	37.9	43.1	6.6	10.8
Beijing	12	0.77	6.0	0.83	8.6	54.3	48.3	14.6	11.1
Tianjin	11	0.19	4.9	0.07	14.3	55.3	42.8	7.7	2.3
Hangzhou	11	0.78	6.3	0.83	6.8	42.1	45.3	9.5	13.7
Shenyang	11	0.66	5.9	0.75	5.2	47.3	46.0	10.7	9.7
Guangzhou	10	0.79	5.2	0.83	5.5	40.5	43.7	10.2	11.3
Wuhan	10	0.77	5.4	0.85	7.5	42.3	48.1	8.1	11.4
Changchun	10	0.71	6.7	0.79	7.5	39.0	42.6	7.8	12.9
Changsha	10	0.57	3.9	0.71	4.0	38.8	36.3	4.5	6.0
Shanghai	9	0.43	4.5	0.41	7.9	51.4	45.8	7.8	4.3

compliment satellite $PM_{2.5}$ data in $PM_{2.5}$ models, serving as a proxy for polluted urban plumes, however, satellite $PM_{2.5}$ has low correlation with NO_2 concentration; satellite $PM_{2.5}$ was never selected into our NO_2 models. According to model R^2, satellite data contribute more to improving models for $PM_{2.5}$ than for NO_2, a finding consistent with results reported previously for Europe (Vienneau et al., 2013). Previous national models in the US show an increase in R^2 of 0.22 for a $PM_{2.5}$ model (Beckerman et al., 2013) and an increase in R^2 of 0.12 for an NO_2 model (Young et al., 2016) by incorporating satellite data.

As shown in Tables 2 and S8, UK made substantial improvement in 10-fold CV performance, with little distinction in performance between UK models with and without satellite data. Under extreme conditions like LOPO-CV, performance of all models was reduced owing to prediction errors in large unmonitored areas, however, UK models with satellite data performed better in LOPO-CV than UK models without satellite data. This suggests that models with UK may mask the importance of satellite data (or possibly other regional predictors) when evaluated with 10-fold CV, and highlights the importance of alternative CV evaluation such as LOPO-CV. These findings are consistent with a previous study in the US (Young et al., 2016) though the improvement from UK (0.12–0.18 increase in 10-fold CV R^2 for satellite models) was greater than for the US (0.04 increase in 10-fold CV R^2).

The overall performance of the $PM_{2.5}$ models was consistently better than the NO_2 models, perhaps because our predictors were better at explaining regional concentrations of a more regional pollutant such as $PM_{2.5}$ rather than a pollutant with more local sources, such as NO_2. Although all of our models yielded reasonable R^2 and RMSE, there was systematic underestimation for both $PM_{2.5}$ and NO_2, especially when measured concentrations were high (see scatter plots in Figs. S3 and S4 in the SI). Possible reasons for this could be some inadequacy of our predictors in national scale modeling and possible non-linear relationships between the dependent variables and the predictors. This finding is unsurprising; most models are better at detecting central tendencies than at accurately predicting extremes.

4.3. Variable selection

Because variables selected into our final models tended to be relatively stable across the different approaches (described in Section 2.5.1), we chose a relatively simpler and more conventional approach to select variables (forward variable selection based on adjusted-R^2).

Some previous LUR studies used emission data as an important predictor (Sampson et al., 2013; Knibbs et al., 2014). However, inclusion of emission data as potential predictor may not improve the model performance due to their uncertainties and possible crude spatial scale (Yang et al., 2017). Satellite data and other GIS data could also be used as proxy of emissions (Novotny et al., 2011; Yang et al., 2017). Since we lacked emission data and detailed information on pollution sources, we employed some alternative variables from open-source data sets. We

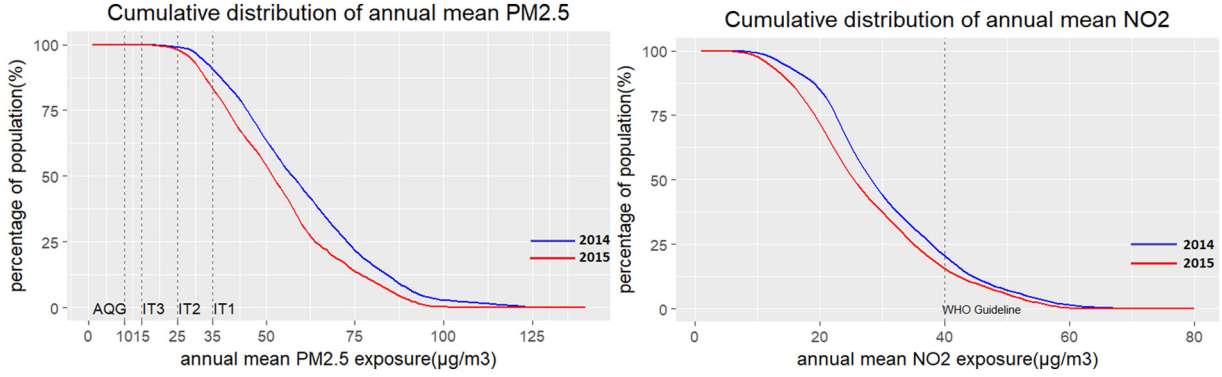


Fig. 4. Cumulative exposure for $PM_{2.5}$ and NO_2 based on the best performing models ($PM_{2.5}$-6 and NO_2-4). For context, air quality guidelines (AQG) and interim targets (IT1-3) from the World Health Organization are shown.

used number of fire spots to reflect pollution from biomass burning, and number of different types of POIs to reflect industrial source pollution, heating suppliers, urban transportation and cooking fumes. POI data was an important predictor in our final models (e.g., gas stations, bus stops, heating suppliers), suggesting that POI data from online mapping services (e.g., Amap, Google Maps) may provide information on local and regional sources that are not well captured from other nationally available data. We also incorporated percentages of eight categories of land cover types to represent different land uses (see Table S1 in the SI for details).

Prior studies have reported that meteorological factors like wind speed (Liu et al., 2009), wind direction (Arain et al., 2007), precipitation (Knibbs et al., 2014) and boundary layer height (Lee et al., 2016) are useful for predicting $PM_{2.5}$ and NO_2 concentrations. Here, we incorporated BLHA-WS to represent diffusion conditions of air pollutants and found it contributed in both $PM_{2.5}$ and NO_2 models. In our non-satellite LUR models, percentage of cropland and number of gas stations consistently entered in $PM_{2.5}$ models, while number of gas stations and percentage of impervious land entered in NO_2 models. Crop land could be a non-negligible source of $PM_{2.5}$ when generated from ammonia, acid gases and straw burning and (for primary $PM_{2.5}$) dust (Xu et al., 2016; Zhang et al., 2016). Variables with larger buffer lengths were more likely to be selected into $PM_{2.5}$ models, while buffer lengths of variables selected into NO_2 models were comparatively smaller. This may be because $PM_{2.5}$ is a more regional pollutant affected by long-range transport: variables with large buffer sizes could reflect more regional transport, while NO_2 concentrations are more likely to be affected by local pollution sources.

4.4. Cross validation

Distance between a test-set monitor and its nearest training-set neighbor for LOPO-CV ranged from 28 km to 1454 km with a mean (median) value in 2015 of 184 km (148.7 km); for 10-fold CV the same value (distance between test monitor and nearest training-set neighbor) ranged from 0.3 km to 350 km with a mean (median) of 8.9 km (3.8 km) (see also Table S4 in the SI). Fig. S13 indicates that most people live in areas <50 km from the nearest monitor. That result indicates that LOPO-CV is a more extreme (more stringent) test of model performance than would be applicable to most people in mainland China. Average model performance across the population should fall between 10-fold CV performance and LOPO-CV performance. Our findings indicate that 10-fold CV may overestimate model performance at locations far from monitors, whereas LOPO-CV likely underestimates model performance for most people. Tables 2 and S9 show that 10-fold CV performance was consistently better than LOPO-CV performance. This finding also suggests that performance of both kriging and non-kriging models is reduced in unmonitored regions. Previous studies have also implemented some approaches that non-randomly select cross-validation groups, such as spatially clustered cross-validation (Young et al., 2016) and isolated-site cross-validation (Xue et al., 2017), which tend to select the test-set to be far away from the training-set. Based on the characteristic of the monitoring sites distribution in China, we chose to use provinces or cities as fixed groups to conduct cross-validation, which is convenient for quantitatively analyzing differences in model performance of the CV methods and in evaluating model performance at the province or city scale (Bengio and Grandvalet, 2004).

4.5. Within-urban variability

We were able to assess within-urban variability in concentrations using our national predictions. Compared to $PM_{2.5}$ models, NO_2 predictions typically show more within-urban variation. Models without UK show more within-urban spatial variation for $PM_{2.5}$ than models with UK, however, this may be an artifact of local land use data serving as a proxy for explaining regional concentration variations. A similar phenomenon has been shown for $PM_{2.5}$ models with and without satellite data (Beckerman et al., 2013). For $PM_{2.5}$ models, land-use variables tended to under-predict in high concentration areas like Beijing, and over-predict in relatively low concentration areas like Guangzhou. Satellite data made less of a contribution to our NO_2 models than to our $PM_{2.5}$ models. At a city scale, $PM_{2.5}$ models had a relatively low R^2 but a reasonable RMSE in most cities, which might be due to undetectable within-urban variability, or a lack of within-urban variability altogether. NO_2 models had better R^2 than $PM_{2.5}$ in most cities. Our city scale performance is comparable to some city scale models for Beijing (R^2 0.78 vs. 0.58 for $PM_{2.5}$) (Wu et al., 2015) and Shanghai (R^2 0.70 vs. 0.61 for NO_2) (C. Liu et al., 2016), but worse in some other locations (0.25 vs. 0.73 for $PM_{2.5}$ in Tianjin) (Chen et al., 2017). Regional or city-scale models may better capture within-urban variability than national models such as ours, especially for $PM_{2.5}$ for which pollution sources can be complex and vary by regions in China.

Since spatial resolutions of existing national-scale empirical models in China were 3 km or larger, we also quantitatively computed modeled value variance within each 3×3 km^2 and 10×10 km^2 moving window based on 1×1 km^2 resolution predictions derived from our best performing models ($PM_{2.5}$-6 and NO_2-4, see Fig. S12 in the SI). NO_2 models consistently have much higher ratio than $PM_{2.5}$ models; that finding indicates that finer resolution predictions based on our models help to reveal with-urban variability for NO_2 but less so for $PM_{2.5}$.

4.6. Limitations

A critical aspect of our approach is that it relies on regulatory monitors; such monitors often are located near specific land uses (e.g. public institutions, parks, schools), but may not capture the full range of land uses. Very few monitors were set as traffic sites. The average distance to road of each monitor is larger than 500 m (see Tables S2 and S3 in the SI), this may have led to lack of spatial gradients related to major roads in our predictions.

Moreover, our approach used a single model with fixed variable parameters to predict $PM_{2.5}$ and NO_2 for the whole country, however, the relationship between land uses and concentrations may vary by region. Incorporation of satellite data and universal kriging help to partially remedy this shortcoming.

Another potential limitation is lack of detailed traffic and industrial related inputs in our study. All of our variables were extracted from open-source data sets. The eight categories of land cover types we employed were extracted from land cover product from satellite remote sensing, which may not fully represent emission sources such as industry, ports, and airports. Our road data were extracted from an open-source website, which was not officially provided by a government agency and provides no information on traffic volume. Small-scale traffic variables (buffer lengths <1000 m) were hardly selected into our models, which may be also owing to the quality of our road data.

Finally, we did not incorporate kriging in variable selection in order to lighten the computational load, which may have resulted in underestimating the benefit of kriging.

5. Conclusion

We built national LUR models for ambient annual average NO_2 and $PM_{2.5}$ concentrations in China and generated publicly available 1×1 km^2 spatial resolution national prediction maps which could be used for national-scale long-term exposure analyses. Our models leveraged information from ~900–1400 regulatory monitors, satellite-based measurements of NO_2 and $PM_{2.5}$, and 290 land use and meteorological variables. We find that Satellite data and UK could complement in making predictions more accurate. We also find that parsimonious variable selection approach provides similar or better model performance than more computationally-intensive PLS variable reduction, an important finding for fine spatial resolution national predictions. New variables

for national LUR models such as categorized POI data and BLHA-WS contribute to capturing spatial variations of $PM_{2.5}$ and NO_2 concentrations in China. Our models are capable of providing point predictions, such as at individual residential locations, which could be useful for other population-based environmental health studies in China, including in epidemiology, risk assessment, and environmental justice research. The general approach could usefully be applied to future years of data. Methodological findings here can inform future LUR research.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2018.11.125.

Acknowledgments

This article was developed in part under Assistance Agreement no. RD835873 awarded by the US Environmental Protection Agency (EPA). It has not been formally reviewed by EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the agency. EPA does not endorse any products or commercial services mentioned in this publication. This research was also supported by Tsinghua Scholarship for Overseas Graduate Studies (No. 2016143).

References

Apte, J.S., Bombrun, E., Marshall, J.D., Nazaroff, W.W., 2012. Global intraurban intake fractions for primary air pollutants from vehicles and other distributed sources. Environ. Sci. Technol. 46 (6), 3415–3423. https://doi.org/10.1021/es204021h.

Arain, M.A., Blair, R., Finkelstein, N., Brook, J.R., Sahsuvaroglu, T., Beckerman, B., Zhang, L., Jerrett, M., 2007. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. Atmos. Environ. 41 (16), 3453–3464. https://doi.org/10.1016/j.atmosenv.2006.11.063.

Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., de Nazelle, A., Nieuwenhuijsen, M., Vila, J., Künzli, N., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. Atmos. Environ. 54, 634–642. https://doi.org/10.1016/j.atmosenv.2012.01.064.

Bechle, M.J., Millet, D.B., Marshall, J.D., 2015. National spatiotemporal exposure surface for NO2: monthly scaling of a satellite-derived land-use regression, 2000–2010. Environ. Sci. Technol. 49 (20), 12297–12305. https://doi.org/10.1021/acs.est.5b02882.

Beckerman, B.S., Jerrett, M., Serre, M., Martin, R.V., Lee, S.J., Van Donkelaar, A., Ross, Z., Su, J., Burnett, R.T., 2013. A hybrid approach to estimating national scale spatiotemporal variability of $PM_{2.5}$ in the contiguous United States. Environ. Sci. Technol. 47 (13), 7233–7241. https://doi.org/10.1021/es400039u.

Beelen, R., Hoek, G., van den Brandt, P.A., Goldbohm, R.A., Fischer, P., Schouten, L.J., Armstrong, B., Brunekreef, B., 2008. Long-term exposure to traffic-related air pollution and lung cancer risk. Epidemiology, 702–710 https://doi.org/10.1097/EDE.0b013e318181b3ca.

Bengio, Y., Grandvalet, Y., 2004. No unbiased estimator of the variance of k-fold cross-validation. J. Mach. Learn. Res. 5 (Sep), 1089–1105.

Bright, E.A., Rose, A.N., Urban, M.L., 2016. Landscan 2015 High-Resolution Global Population Data Set (No. LandScan 2015 High-Resolution Global Pop Data Set; 005130MLTPL00). Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States) Available at. http://web.ornl.gov/sci/landscan/landscan_data_avail.shtml (Verified 28 August 2018).

Cesaroni, G., Forastiere, F., Stafoggia, M., Andersen, Z.J., Badaloni, C., Beelen, R., Caracciolo, B., de Faire, U., Erbel, R., Eriksen, K.T., Fratiglioni, L., 2014. Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project. BMJ 348, f7412. https://doi.org/10.1136/bmj.f7412.

Chen, L., Bai, Z., Kong, S., Han, B., You, Y., Ding, X., Du, S., Liu, A., 2010. A land use regression for predicting NO2 and PM10 concentrations in different seasons in Tianjin region, China. J. Environ. Sci. 22 (9), 1364–1373. https://doi.org/10.1016/S1001-0742(09)60263-1.

Chen, L., Shi, M., Li, S., Bai, Z., Wang, Z., 2017. Combined use of land use regression and BenMAP for estimating public health benefits of reducing $PM_{2.5}$ in Tianjin, China. Atmos. Environ. 152, 16–23. https://doi.org/10.1016/j.atmosenv.2016.12.023.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, D.P., Bechtold, P., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597. https://doi.org/10.1002/qj.828.

Eeftens, M., Tsai, M.Y., Ampe, C., Anwander, B., Beelen, R., Bellander, T., Cesaroni, G., Cirach, M., Cyrys, J., de Hoogh, K., De Nazelle, A., 2012. Spatial variation of $PM_{2.5}$, PM_{10}, $PM_{2.5}$ absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2–results of the ESCAPE project. Atmos. Environ. 62, 303–317. https://doi.org/10.1016/j.atmosenv.2012.08.038.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens. 34 (7), 2607–2654. https://doi.org/10.1080/01431161.2012.748992.

Gryparis, A., Paciorek, C.J., Zeka, A., Schwartz, J., Coull, B.A., 2008. Measurement error caused by spatial misalignment in environmental epidemiology. Biostatistics 10 (2), 258–274. https://doi.org/10.1093/biostatistics/kxn033.

Guo, T., Wang, Y., Zhang, H., Zhang, Y., Zhao, J., Wang, Q., Shen, H., Wang, Y., Xie, X., Wang, L., Xu, Z., 2018. The association between ambient $PM_{2.5}$ exposure and the risk of preterm birth in China: a retrospective cohort study. Sci. Total Environ. 633, 1453–1459. https://doi.org/10.1016/j.scitotenv.2018.03.328.

He, Q., Huang, B., 2018. Satellite-based mapping of daily high-resolution ground PM 2.5 in China via space-time regression modeling. Remote Sens. Environ. 206, 72–83. https://doi.org/10.1016/j.rse.2017.12.018.

Henderson, S.B., Beckerman, B., Jerrett, M., Brauer, M., 2007. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. Environ. Sci. Technol. 41 (7), 2422–2428. https://doi.org/10.1021/es0606780.

Hoek, G., Beelen, R., De Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42 (33), 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057.

Hu, X., Waller, L.A., Lyapustin, A., Wang, Y., Liu, Y., 2014. Improving satellite-driven $PM_{2.5}$ models with Moderate Resolution Imaging Spectroradiometer fire counts in the southeastern US. J. Geophys. Res. Atmos. 119 (19). https://doi.org/10.1002/2014JD021920.

Kaufman, J.D., Adar, S.D., Barr, R.G., Budoff, M., Burke, G.L., Curl, C.L., Daviglus, M.L., Roux, A.V.D., Gassett, A.J., Jacobs Jr., D.R., Kronmal, R., 2016. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. Lancet 388 (10045), 696–704. https://doi.org/10.1016/S0140-6736(16)00378-0.

Knibbs, L.D., Hewson, M.G., Bechle, M.J., Marshall, J.D., Barnett, A.G., 2014. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. Environ. Res. 135, 204–211. https://doi.org/10.1016/j.envres.2014.09.011.

Lee, H.J., Chatfield, R.B., Strawa, A.W., 2016. Enhancing the applicability of satellite remote sensing for $PM_{2.5}$ estimation using MODIS deep blue AOD and land use regression in California, United States. Environ. Sci. Technol. 50 (12), 6546–6555. https://doi.org/10.1021/acs.est.6b01438.

Li, M., Zhang, L., 2014. Haze in China: current and future challenges. Environ. Pollut. 189, 85–86. https://doi.org/10.1016/j.envpol.2014.02.024.

Liu, Y., Paciorek, C.J., Koutrakis, P., 2009. Estimating regional spatial and temporal variability of $PM_{2.5}$ concentrations using satellite data, meteorology, and land use information. Environ. Health Perspect. 117 (6), 886. https://doi.org/10.1289/ehp.0800123.

Liu, C., Henderson, B.H., Wang, D., Yang, X., Peng, Z.R., 2016. A land use regression application into assessing spatial variation of intra-urban fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO2) concentrations in City of Shanghai, China. Sci. Total Environ. 565, 607–615. https://doi.org/10.1016/j.scitotenv.2016.03.189.

Liu, J., Han, Y., Tang, X., Zhu, J., Zhu, T., 2016. Estimating adult mortality attributable to $PM_{2.5}$ exposure in China with assimilated $PM_{2.5}$ concentrations based on a ground monitoring network. Sci. Total Environ. 568, 1253–1262. https://doi.org/10.1016/j.scitotenv.2016.05.165.

Liu, W., Huang, C., Hu, Y., Fu, Q., Zou, Z., Sun, C., Shen, L., Wang, X., Cai, J., Pan, J., Huang, Y., 2016. Associations of gestational and early life exposures to ambient air pollution with childhood respiratory diseases in Shanghai, China: a retrospective cohort study. Environ. Int. 92, 284–293. https://doi.org/10.1016/j.envint.2016.04.019.

Ma, Z., Hu, X., Huang, L., Bi, J., Liu, Y., 2014. Estimating ground-level $PM_{2.5}$ in China using satellite remote sensing. Environ. Sci. Technol. 48 (13), 7436–7444. https://doi.org/10.1021/es5009399.

Ma, Z., Hu, X., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L., Liu, Y., 2015. Satellite-based spatiotemporal trends in $PM_{2.5}$ concentrations: China, 2004–2013. Environ. Health Perspect. 124 (2), 184–192. https://doi.org/10.1289/ehp.1409481.

Meng, X., Chen, L., Cai, J., Zou, B., Wu, C.F., Fu, Q., Zhang, Y., Liu, Y., Kan, H., 2015. A land use regression model for estimating the NO2 concentration in Shanghai, China. Environ. Res. 137, 308–315. https://doi.org/10.1016/j.envres.2015.01.003.

Novotny, E.V., Bechle, M.J., Millet, D.B., Marshall, J.D., 2011. National satellite-based land-use regression: NO2 in the United States. Environ. Sci. Technol. 45 (10), 4407–4414. https://doi.org/10.1021/es103578x.

Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. RFID J. 8 (1), 204–218.

Pope III, C.A., Burnett, R.T., Turner, M.C., Cohen, A., Krewski, D., Jerrett, M., Gapstur, S.M., Thun, M.J., 2011. Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure–response relationships. Environ. Health Perspect. 119 (11), 1616. https://doi.org/10.1289/ehp.1103639.

Sampson, P.D., Richards, M., Szpiro, A.A., Bergen, S., Sheppard, L., Larson, T.V., Kaufman, J.D., 2013. A regionalized national universal kriging model using Partial Least Squares regression for estimating annual $PM_{2.5}$ concentrations in epidemiology. Atmos. Environ. 75, 383–392. https://doi.org/10.1016/j.atmosenv.2013.04.015.

Szpiro, A.A., Paciorek, C.J., Sheppard, L., 2011. Does more accurate exposure prediction necessarily improve health effect estimates? Epidimiology 22 (5), 680. https://doi.org/10.1097/EDE.0b013e3182254cc6.

Van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., Lyapustin, A., Sayer, A.M., Winker, D.M., 2016. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. Environ. Sci. Technol. 50 (7), 3762–3772. https://doi.org/10.1021/acs.est.5b05833.

Vienneau, D., De Hoogh, K., Bechle, M.J., Beelen, R., Van Donkelaar, A., Martin, R.V., Millet, D.B., Hoek, G., Marshall, J.D., 2013. Western European land use regression incorporating satellite-and ground-based measurements of NO2 and PM10. Environ. Sci. Technol. 47 (23), 13555–13564. https://doi.org/10.1021/es403089q.

Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., Brunekreef, B., 2012. Systematic evaluation of land use regression models for NO2. Environ. Sci. Technol. 46 (8), 4481–4489. https://doi.org/10.1021/es204183v.

Wu, J., Li, J., Peng, J., Li, W., Xu, G., Dong, C., 2015. Applying land use regression model to estimate spatial variation of $PM_{2.5}$ in Beijing, China. Environ. Sci. Pollut. Res. 22 (9), 7045–7061. https://doi.org/10.1007/s11356-014-3893-5.

Wu, C.D., Chen, Y.C., Pan, W.C., Zeng, Y.T., Chen, M.J., Guo, Y.L., Lung, S.C.C., 2017. Land-use regression with long-term satellite-based greenness index and culture-specific sources to model $PM_{2.5}$ spatial-temporal variability. Environ. Pollut. 224, 148–157. https://doi.org/10.1016/j.envpol.2017.01.074.

Xu, W., Wu, Q., Liu, X., Tang, A., Dore, A.J., Heal, M.R., 2016. Characteristics of ammonia, acid gases, and $PM_{2.5}$ for three typical land-use types in the North China Plain. Environ. Sci. Pollut. Res. 23 (2), 1158–1172. https://doi.org/10.1007/s11356-015-5648-3.

Xue, T., Zheng, Y., Geng, G., Zheng, B., Jiang, X., Zhang, Q., He, K., 2017. Fusing observational, satellite remote sensing and air quality model simulated data to estimate spatiotemporal variations of $PM_{2.5}$ exposure in China. Remote Sens. 9 (3), 221. https://doi.org/10.3390/rs9030221.

Yang, X., Zheng, Y., Geng, G., Liu, H., Man, H., Lv, Z., He, K., de Hoogh, K., 2017. Development of $PM_{2.5}$ and NO2 models in a LUR framework incorporating satellite remote sensing and air quality model data in Pearl River Delta region, China. Environ. Pollut. 226, 143–153. https://doi.org/10.1016/j.envpol.2017.03.079.

You, W., Zang, Z., Zhang, L., Li, Y., Wang, W., 2016. Estimating national-scale ground-level PM25 concentration in China using geographically weighted regression based on MODIS and MISR AOD. Environ. Sci. Pollut. Res. 23 (9), 8327–8338. https://doi.org/10.1007/s11356-015-6027-9.

Young, M.T., Bechle, M.J., Sampson, P.D., Szpiro, A.A., Marshall, J.D., Sheppard, L., Kaufman, J.D., 2016. Satellite-based NO2 and model validation in a national prediction model based on universal kriging and land-use regression. Environ. Sci. Technol. 50 (7), 3686–3694. https://doi.org/10.1021/acs.est.5b05099.

Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M.L., Shen, X., Zhu, L., Zhang, M., 2017. Spatiotemporal prediction of continuous daily $PM_{2.5}$ concentrations across China using a spatially explicit machine learning algorithm. Atmos. Environ. 155, 129–139. https://doi.org/10.1016/j.atmosenv.2017.02.023.

Zhan, Y., Luo, Y., Deng, X., Zhang, K., Zhang, M., Grieneisen, M.L., Di, B., 2018. Satellite-based estimates of daily NO2 exposure in China using hybrid random forest and spatiotemporal kriging model. Environ. Sci. Technol. 52 (7), 4180–4189. https://doi.org/10.1021/acs.est.7b05669.

Zhang, Q., Geng, G., Wang, S., Richter, A., He, K., 2012. Satellite remote sensing of changes in NOx emissions over China during 1996–2010. Chin. Sci. Bull. 57 (22), 2857–2864. https://doi.org/10.1007/s11434-012-5015-4.

Zhang, L., Liu, Y., Hao, L., 2016. Contributions of open crop straw burning emissions to $PM_{2.5}$ concentrations in China. Environ. Res. Lett. 11 (1), 014014. https://doi.org/10.1088/1748-9326/11/1/014014.

Zheng, Y., Zhang, Q., Liu, Y., Geng, G., He, K., 2016. Estimating ground-level $PM_{2.5}$ concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. Atmos. Environ. 124, 232–242. https://doi.org/10.1016/j.atmosenv.2015.06.046.

Zheng, Y., Xue, T., Zhang, Q., Geng, G., Tong, D., Li, X., He, K., 2017. Air quality improvements and health benefits from China's clean air action since 2013. Environ. Res. Lett. 12 (11), 114020. https://doi.org/10.1088/1748-9326/aa8a32.