Supporting Information

Satellite-Based Estimates of Daily NO_2 Exposure in China Using Hybrid Random Forest and Spatiotemporal Kriging Model

Yu Zhan,[†,‡,§] Yuzhou Luo,[‖] Xunfei Deng,[⊥] Kaishan Zhang,[†] Minghua Zhang,[‖]
Michael L. Grieneisen,[‖] Baofeng Di[*,‡,†]

[†]Department of Environmental Science and Engineering, Sichuan University, Chengdu, Sichuan 610065, China
[‡]Institute for Disaster Management and Reconstruction, Sichuan University, Chengdu, Sichuan 610200, China
[§]Sino-German Centre for Water and Health Research, Sichuan University, Chengdu, Sichuan 610065, China
[‖]Department of Land, Air, and Water Resources, University of California, Davis, CA 95616, USA
[⊥]Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang 310021, China

Corresponding author
[*]Phone: +86 2885996656; fax +86 2885405613; e-mail: dibaofeng@scu.edu.cn.

Number of pages: 34
Number of tables: 11
Number of figures: 19

S.1 Temporal convolution

Temporal convolution with a Gaussian kernel was employed to process the original OMI-NO2 satellite retrievals.

$$O(t) = \sum_n \left[I(n) \cdot W(t-n) \right] \Big/ \sum_n W(t-n) \qquad (1)$$

where $O(t)$ is the output value on the tth day iterating through the whole study period, $I(n)$ is the original OMI-NO2 value on the nth day if available, and $W(t\text{-}n)$ is the weight of $I(n)$ in calculating $O(t)$. $W(t\text{-}n)$ is determined by the Gaussian kernel function:

$$W(t-n) = \exp\left[-\left(t-n\right)^2 \Big/ \left(2\sigma^2\right) \right] \qquad (2)$$

where σ is the standard deviation of the Gaussian function, which is set to 60 days based on the sensitivity analyses considering the input completeness and output smoothness.

S.2 Algorithm of random forests[1]

For *tree* = 1 to 500:

- Randomly draw a sample from the training data with replacement, and the sample size is the same as the training data;
- Grow a tree by starting with a single node, and then repeat the steps below until only one observation presents in each terminal node:
    - Randomly select one third of the predictors;
    - Find the split that reduces the squared error the most;

Average the predictions of all trees as the model output.

S.3 Hourly-scaling approach

The random forest-spatiotemporal kriging (RF-STK) model was used to predict the hourly NO_2 concentrations when the Aura satellite passed, which were then scaled to daily concentrations. The detailed procedures are as follows:

(1) Train the RF-STK model to simulate the NO_2 for the overpass hour of the Aura satellite;
(2) Use the RF-STK model to predict the hourly NO_2 for unmonitored areas;
(3) Calculate the scaling factors (i.e., the ratio of observed hourly to daily NO_2) for the monitoring sites;
(4) Estimate the scaling factors for the unmonitored areas by using kriging interpolation;
(5) Divide the hourly NO_2 predictions by the estimated scaling factor to get the daily NO_2 predictions for the unmonitored areas.

Table S1 Coverage rates of OMI-NO2 satellite retrievals across China (%)

Year(s)	Spring	Summer	Fall	Winter	Annual
2013	68.0	78.0	68.5	51.7	66.6
2014	70.0	77.1	61.7	52.5	65.4
2015	65.3	76.7	62.8	49.5	63.7
2016	66.2	76.5	61.6	47.5	62.5
2013-2016	67.4	77.1	63.6	50.3	64.6

Table S2 Correlation coefficient squares (R^2) between the observed ambient NO_2 concentrations and the OMI satellite-retrieved vertical column density of the tropospheric NO_2

Temporal Resolution	Level-2	Level-3	Level-3 Conv.[a]
Day	0.18	0.24	0.27
Month	0.26	0.38	0.42
Season	0.38	0.44	0.48
Year	0.45	0.48	0.48
Spatial[b]	0.53	0.53	0.52

[a] OMI retrievals processed with the temporal convolution.
[b] Multiyear averages, i.e., no temporal resolution.

Table S3 List of variable symbols and definitions

Symbol	Unit	Variable definition	Spatial[a]	Temporal[a]	Convolution[b]
OMI	molecule/cm^2	OMI-retrieved tropospheric NO_2 density	0. 25°×0.25°	Day	Temporal
DOY	-	Day of year	-	-	-
YEAR	-	Year	-	-	-
EVP	mm	Evaporation	Point	Day	-
PRE	mm	Precipitation	Point	Day	-
PRS	hPa	Atmospheric pressure	Point	Day	-
RHU	%	Relative humidity	Point	Day	-
SSD	hour	Sunshine duration	Point	Day	-
TEM	°C	Temperature	Point	Day	-
WIN	m/s	Wind speed	Point	Day	-
PBLH	km	Planetary boundary layer height	0.625°×0.5°	Day	-
ELV	m	Elevation	90m×90m	None	Spatial
NDVI	-	Normalized Difference Vegetation Index	250m×250m	8 Days	Spatial
POP	people/km^2	Population density	30"×30"	None	Spatial
LU10	%	Cultivated land area	30m×30m	None	Spatial
LU20	%	Forest area	30m×30m	None	Spatial
LU30	%	Grassland area	30m×30m	None	Spatial
LU40	%	Shrubland area	30m×30m	None	Spatial
LU50	%	Wetland area	30m×30m	None	Spatial
LU60	%	Waterbody area	30m×30m	None	Spatial
LU80	%	Artificial surface area	30m×30m	None	Spatial
LU90	%	Bareland area	30m×30m	None	Spatial
LU100	%	Permanent frozen land area	30m×30m	None	Spatial
LU255	%	Sea area	30m×30m	None	Spatial
ROAD	km/grid	Road density	Polyline	None	Spatial
eBC	Mg/grid	Emission of black carbon	0.25°×0.25°	Month	Spatial
eCO	Mg/grid	Emission of CO	0.25°×0.25°	Month	Spatial
eCO2	Mg/grid	Emission of CO_2	0.25°×0.25°	Month	Spatial
eNH3	Mg/grid	Emission of NH_3	0.25°×0.25°	Month	Spatial
eNOx	Mg/grid	Emission of NO_2 and NO	0.25°×0.25°	Month	Spatial
eOC	Mg/grid	Emission of organic carbon	0.25°×0.25°	Month	Spatial
ePM25	Mg/grid	Emission of $PM_{2.5}$	0.25°×0.25°	Month	Spatial
ePMcoar	Mg/grid	Emission of PM-coarse	0.25°×0.25°	Month	Spatial
eSO2	Mg/grid	Emission of SO_2	0.25°×0.25°	Month	Spatial
eVOC	Mg/grid	Emission of VOC	0.25°×0.25°	Month	Spatial

[a] Spatial or temporal resolution of raw data.
[b] Temporal: OMI is processed with the temporal convolution. Spatial: these variables have accompanying variables processed with the spatial convolution.

Table S4 Descriptive statistics of the meteorological variables

Variable	Number of observations	Unit	Mean	Standard deviation	Median	Interquartile range
Evaporation	1202560	mm	2.7	2	2.4	2.6
Precipitation	1140852	mm	2.7	9.5	0	0.6
Atmospheric pressure	1219978	hPa	924	108	969	120
Relative humidity	1219982	%	66	20	69	28
Sunshine duration	1219032	hour	5.9	4.1	6.8	7.7
Temperature	1220028	℃	12.5	11.9	14.4	16.9
Wind speed	1219439	m/s	2.2	1.4	1.8	1.4

Table S5 Performance of the random-forest-spatiotemporal-kriging model (RFSTK) in predicting the ambient NO_2 concentrations for China during 2013-2016

Metric[a]	Daily	Monthly	Seasonal	Annual	Spatial
R^2	0.62	0.65	0.68	0.68	0.73
RMSE	13.3	10.2	9.0	7.7	6.5
Slope	0.67	0.71	0.73	0.72	0.77
RPE	39.5%	30.1%	26.2%	22.4%	19.9%
MFB	0.066	0.044	0.037	0.029	0.027
MFE	0.32	0.25	0.22	0.19	0.17
MNB	0.27	0.19	0.18	0.07	0.06
MNE	0.48	0.36	0.34	0.21	0.19

[a] R^2: coefficient of determination; RMSE: root mean square error (μg/m^3); RPE: relative prediction error; MFB: mean fractional bias; MFE: mean fractional error; MNB: mean normalized bias; MNE: mean normalized error.

Table S6 Prediction performance for data points with or without OMI-NO2 values prior to imputation by temporal convolution[a]

OMI-NO2	R^2	Slope	RMSE	RPE	MFB	MFE	MNB	MNE
Available	0.61	0.67	13.2	40%	0.07	0.32	0.27	0.48
Missing	0.62	0.67	13.6	39%	0.06	0.32	0.26	0.47

[a] R^2: coefficient of determination; RMSE: root mean square error (μg/m^3); RPE: relative prediction error; MFB: mean fractional bias; MFE: mean fractional error; MNB: mean normalized bias; MNE: mean normalized error. These metrics are evaluated in the 10-fold cross-validation. Approximately 65% of the data had OMI-NO2 values prior to the imputation.

Table S7 Number of NO_2 monitoring sites within each cross-validation fold for each climate region

Climate Region[a]	Stratified Sampling[b]	Fold									
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Qinghai-Tibet Plateau	Y	4	4	4	4	4	3	3	3	3	3
	N	3	2	3	4	5	2	2	6	2	6
Subtropical Monsoon	Y	87	87	87	87	87	86	86	86	86	86
	N	82	85	75	77	88	86	100	85	89	98
Temperate Continental	Y	7	6	6	6	6	6	6	6	6	6
	N	9	5	7	7	7	7	5	8	2	4
Temperate Monsoon	Y	66	66	66	66	66	66	66	65	65	65
	N	68	70	79	72	63	65	57	61	67	55
Tropical Monsoon	Y	4	4	4	4	4	4	4	4	4	3
	N	4	4	2	6	3	6	2	5	5	2

[a] Please refer to Figure S10 for the map of these climate regions.

[b] Y means random sampling stratified by climate regions, with cross-validation results of R^2=0.61, Slope=0.67, RMSE=13.4, RPE=40%, MFB=0.06, MFE=0.32, MNB=0.26, MNE=0.48; and N represents non-stratified random sampling, with cross-validation results of R^2=0.62, Slope=0.67, RMSE=13.3, RPE=40%, MFB=0.07, MFE=0.32, MNB=0.27, MNE=0.48.

Table S8 Comparisons of the statistical models in predicting daily NO_2 for China during 2013-2016 with the same setting of 10-fold-cross-validation

Metric[a]	LR[b]	STK[b]	LR-STK[b]	RF_0[b]	RF[b]	RF-STK[b]	$RF\text{-}STK_h$[b]
R^2	0.38	0.60	**0.64**	0.60	0.61	0.62	0.48
Slope	0.38	0.57	0.62	0.63	0.64	**0.67**	**0.67**
RMSE	16.8	13.5	**12.9**	13.5	13.4	13.3	16.0
RPE	50%	40%	**38%**	40%	40%	40%	48%
MFB	0.11	0.10	0.08	0.09	0.08	0.07	**0.05**
MFE	0.40	0.33	**0.31**	0.33	0.32	0.32	0.36
MNB	0.40	0.34	0.29	0.30	0.29	**0.27**	0.28
MNE	0.64	0.53	0.49	0.50	0.49	**0.48**	0.53

[a] R^2: coefficient of determination; RMSE: root mean square error (μg/m^3); RPE: relative prediction error; MFB: mean fractional bias; MFE: mean fractional error; MNB: mean normalized bias; MNE: mean normalized error. Bold: the best performance of each evaluation metric. Lower values are better for each metric except R^2 and slope.

[b] LR: Linear Regression model; STK: Spatiotemporal Kriging model; LR-STK: Linear Regression-Spatiotemporal Kriging hybrid model; RF_0: Random Forest model without variable selection; RF: Random Forest model with variable selection; RF-STK: Random Forest-Spatiotemporal Kriging model; $RF\text{-}STK_h$: Random Forest-Spatiotemporal Kriging model with hourly scaling.

Table S9 Performance of the previous statistical models in predicting NO_2

Reference	Model	Study Area	Study Period	Validation	Metric
2	Satellite-based LUR; lasso	Australia	2006-2011	Fitting	R^2=0.81 (annual) R^2=0.76 (monthly)
3	Mixed effects model	New England region, United States	2005-2010	10-fold sample-based cross-validation	R^2=0.79 (daily)
4	LUR	Canada	2006	Fitting	R^2=0.73 (spatial)
5	LUR	Western Europe	2005-2007	Fitting	R^2=0.50 (spatial)
6	LUR	Netherlands	2007	Fitting	R^2=0.84 (annual)
7	LUR	Changsha, China	2010	Leave-25%-out-validation	R^2=0.67 (annual)
8	LUR	Shanghai, China	2008-2011	Leave-one-out-cross-validation	R^2=0.75 (annual)
9	LUR	Pearl River Delta, China	2013-2014	Leave-one-out-cross-validation	R^2=0.71 (annual)
10	LUR	Western Europe	2009-2010	Hold-out-validation on 20% of monitoring site	R^2=0.60 (annual)
11	LUR	Seoul, Korea	2003	Fitting	R^2=0.95~0.98 (seasonal)
12	LUR; Lasso regression	Global	2011	Bootstrap 10% cross-validation	R^2=0.53 (annual)
13	LUR; monthly scaling	Contiguous United States	2000-2010	Fitting	R^2=0.79 (spatial); R^2=0.84 (monthly)
14	LUR-UK; partial least square	Contiguous United States	1990-2012	20-fold site-based cross-validation	R^2=0.85 (annual)
15	LUR; Stepwise multivariate regression	Contiguous United States	2006	Leave-10%-out validation	R^2=0.76 (annual)

Table S10 Comparisons of the random forest and the linear regression models in predicting NO_2 for temporal extrapolation[a]

Metric[b]	Daily NO_2		Spatial NO_2[c]	
	Linear Regression	Random Forest	Linear Regression	Random Forest
R^2	0.39	**0.56**	0.61	**0.83**
Slope	0.39	**0.54**	0.62	**0.81**
RMSE	16.8	**14.3**	7.9	**5.2**
RPE	50%	**42%**	24%	**16%**
MFB	0.10	**0.08**	0.03	**0.01**
MFE	0.39	**0.33**	0.21	**0.13**
MNB	0.38	**0.29**	0.08	**0.03**
MNE	0.63	**0.50**	0.24	**0.14**

[a] The two models are trained with the data of 2014 and 2015, and then are used to make predictions for 2013 and 2016. The two models have the same set of predictor variables.
[b] R^2: coefficient of determination; RMSE: root mean square error ($\mu g/m^3$); RPE: relative prediction error; MFB: mean fractional bias; MFE: mean fractional error; MNB: mean normalized bias; MNE: mean normalized error. Bold: the best performance of each evaluation metric. Lower values are better for each metric except R^2 and slope.
[c] Two-year averages, i.e., no temporal resolution.

Table S11 Population-weighted ambient NO_2 concentrations (mean ± standard deviation; μg/m^3) and temporal trends (μg/m^3/year) during 2013-2016 for the main economic zones in China

Economic Zone[a]	Spring	Summer	Fall	Winter	Annual	Trend	Trend 95% CI	Trend *P*
Beijing-Tianjin Metro	46.8 ± 7.4	35.5 ± 6.2	54.1 ± 7.4	62.0 ± 7.1	49.5 ± 6.7	-0.61	(-1.82, 0.60)	=0.32
Pearl River Delta	41.7 ± 9.1	29.4 ± 6.8	39.3 ± 7.8	49.9 ± 9.4	40.1 ± 7.9	-1.37	(-2.19, -0.55)	<0.01
Sichuan Basin	30.4 ± 10.9	23.6 ± 9.3	29.6 ± 9.9	38.0 ± 10.9	30.4 ± 10.1	-1.10	(-1.44, -0.76)	<0.01
Yangtze River Delta	40.6 ± 8.2	28.3 ± 6.0	39.8 ± 7.2	50.2 ± 8.5	39.7 ± 7.3	-1.03	(-1.62, -0.44)	<0.01

[a] The four economic zones are located in North, East, South, and Southwest China, respectively.

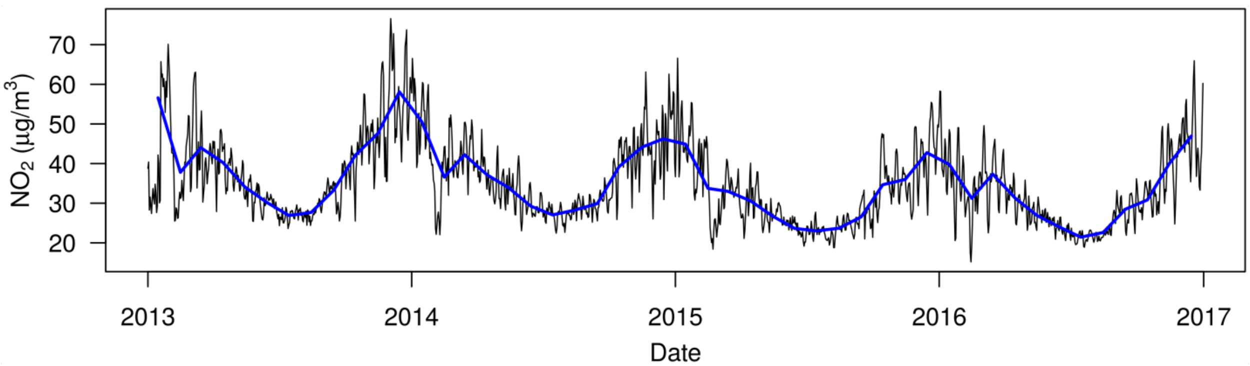

Figure S1. Average daily (black line) and monthly (blue line) NO_2 concentrations across all the monitoring sites for China during 2013-2016.

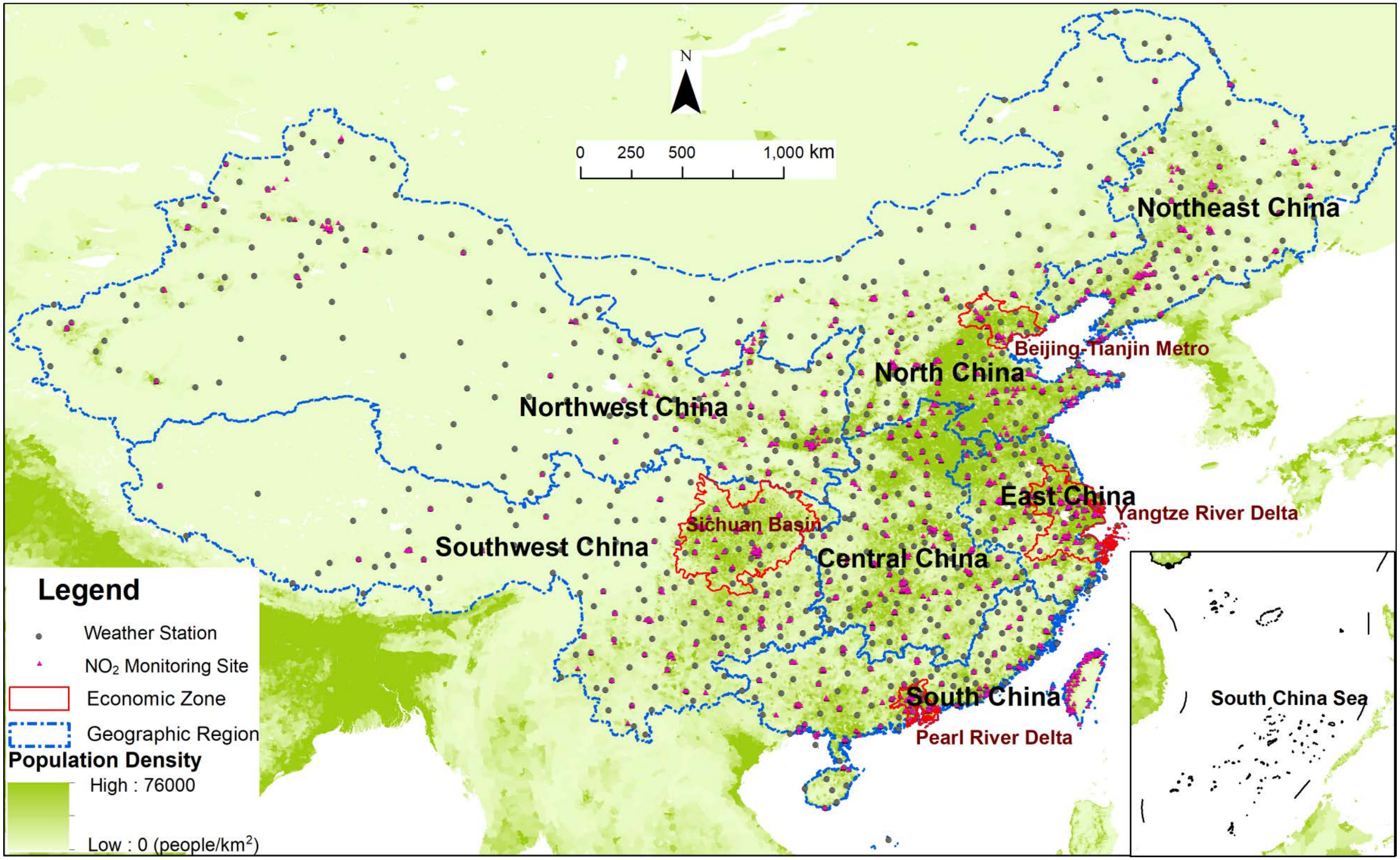

Figure S2. Spatial distribution of the NO_2 monitoring sites in China during 2013-2016. The basemap of population density data for 2015 is obtained from the Gridded Population of the World (GPWv4; 30 arc-second resolution).[16]

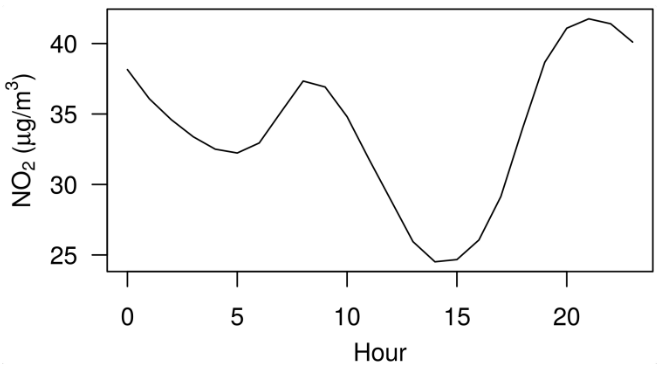

Figure S3. Average diurnal pattern in NO_2 concentrations across all the monitoring sites for China during 2013-2016. Two peaks appeared at 8am and 21pm.

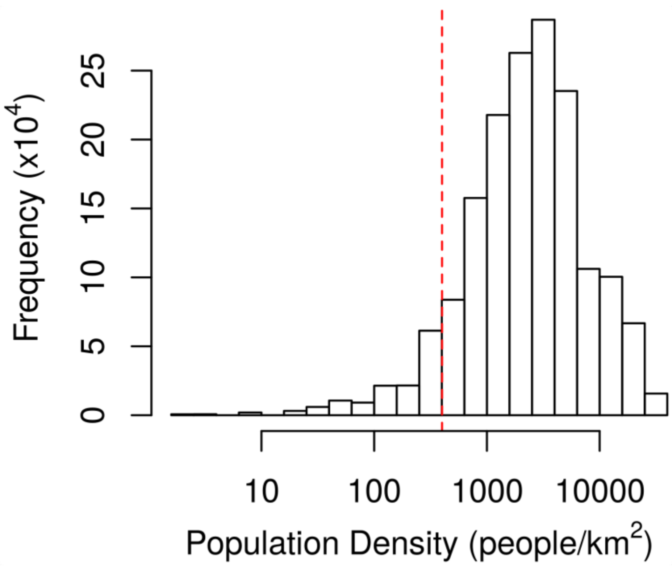

Figure S4. Frequency distributions of the numbers of NO_2 observations by population density. The population density of 400 people/km^2 is indicated by the red line.

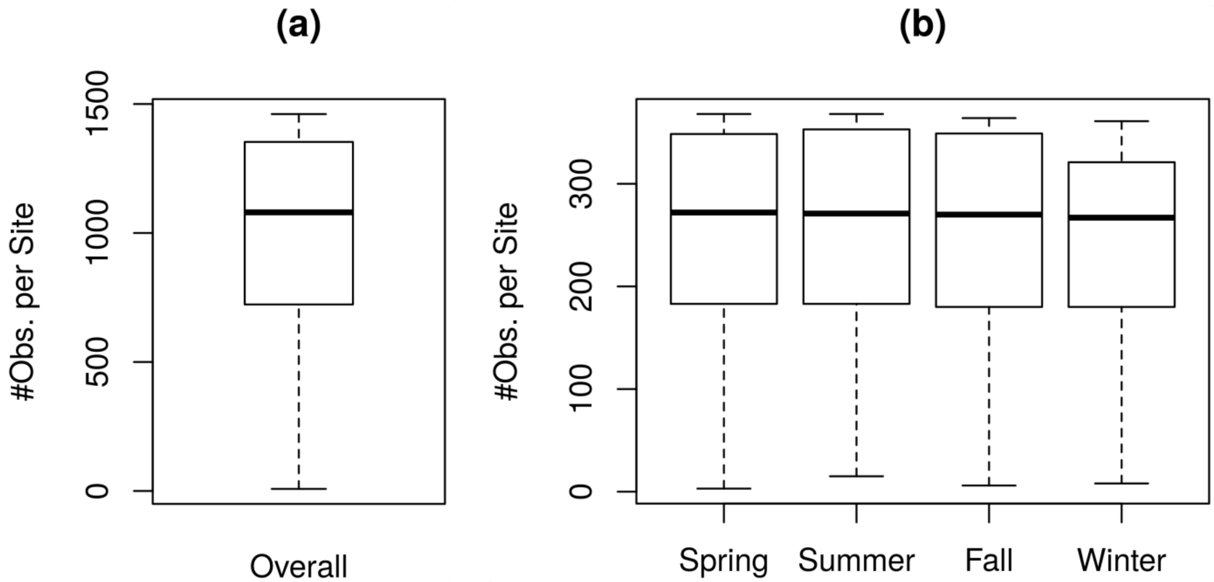

Figure S5. (a) Overall and (b) seasonal number of NO_2 observations per monitoring site during 2013-2016 for China. The overall mean±standard deviation was 1008±344. The seasonal mean±standard deviation values were 258±83, 258±83, 260±85, and 246±79 for spring, summer, fall, and winter, respectively.

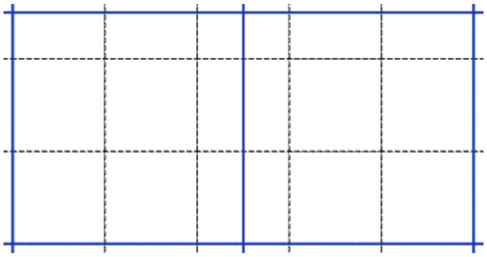

Figure S6. Overlay of the 0.25° grid (blue line) on the 0.1° grid (black dashed line).

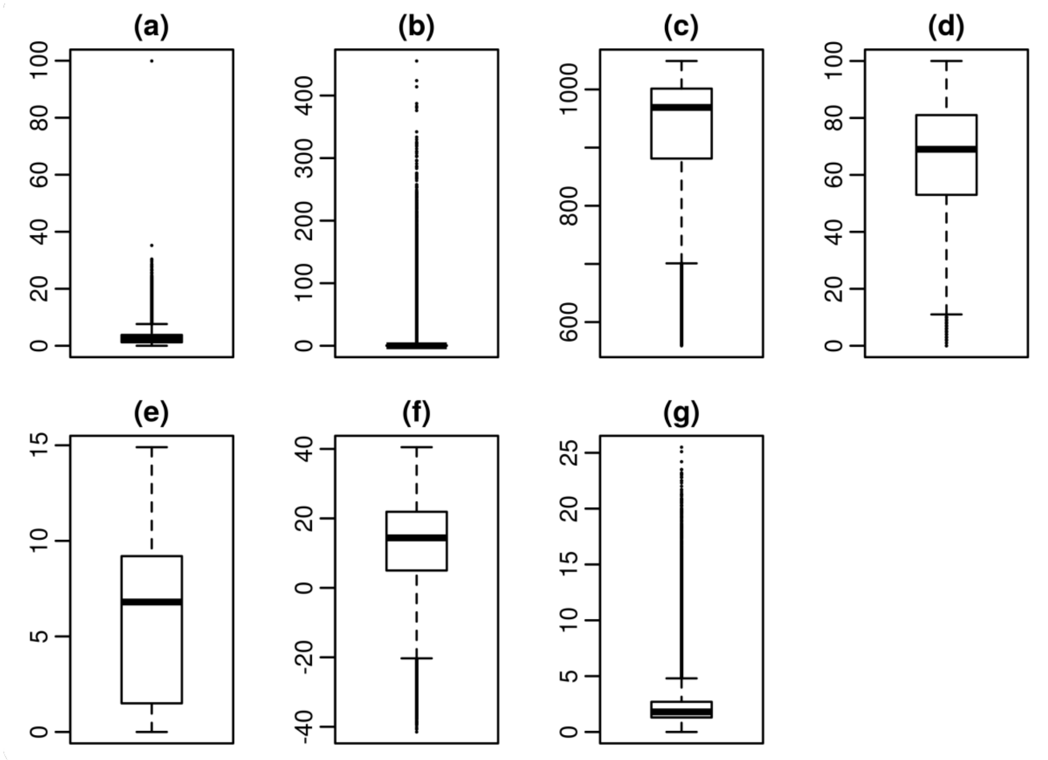

Figure S7. Boxplots of the meteorological variables: (a) evaporation (mm), (b) precipitation (mm), (c) atmospheric pressure (hPa), (d) relative humidity (%), (e) sunshine duration (hour), (f) temperature (°C), and (g) wind speed (m/s).

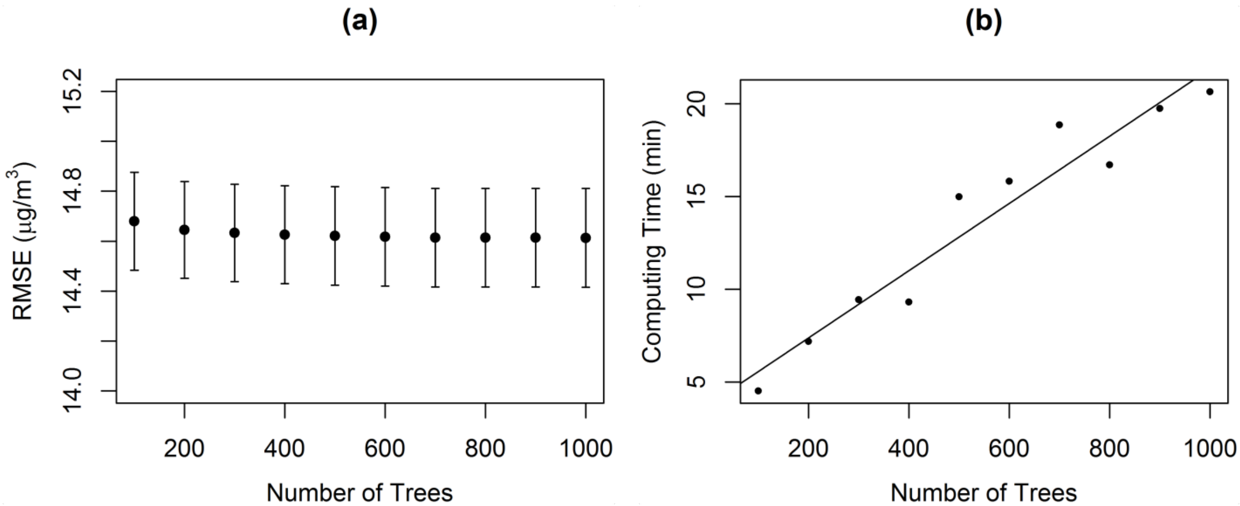

Figure S8. (a) Root mean square error (RMSE) as a function of the number of trees for the random forest model with the selected predictors. The RMSE is evaluated with 10-fold cross-validation on a random subset (100,000 samples) of the training data. The error bars represent the standard errors across the 10 folds. (b) Computing time for different numbers of trees in this evaluation, with a fitted regression line.

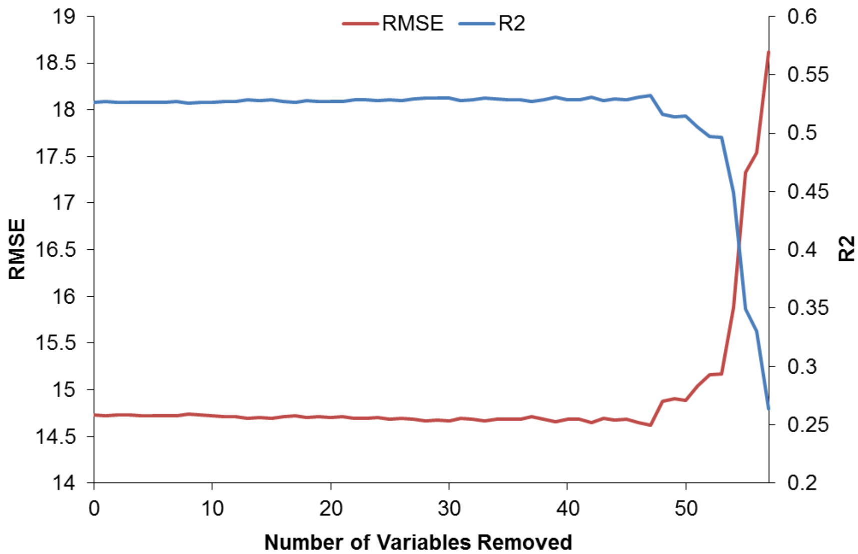

Figure S9. Evolution of cross-validation RMSE (μg/m^3) and R^2 for the random forest submodels through the variable selection process. Refer to Table S3 for the detailed descriptions of the variables.

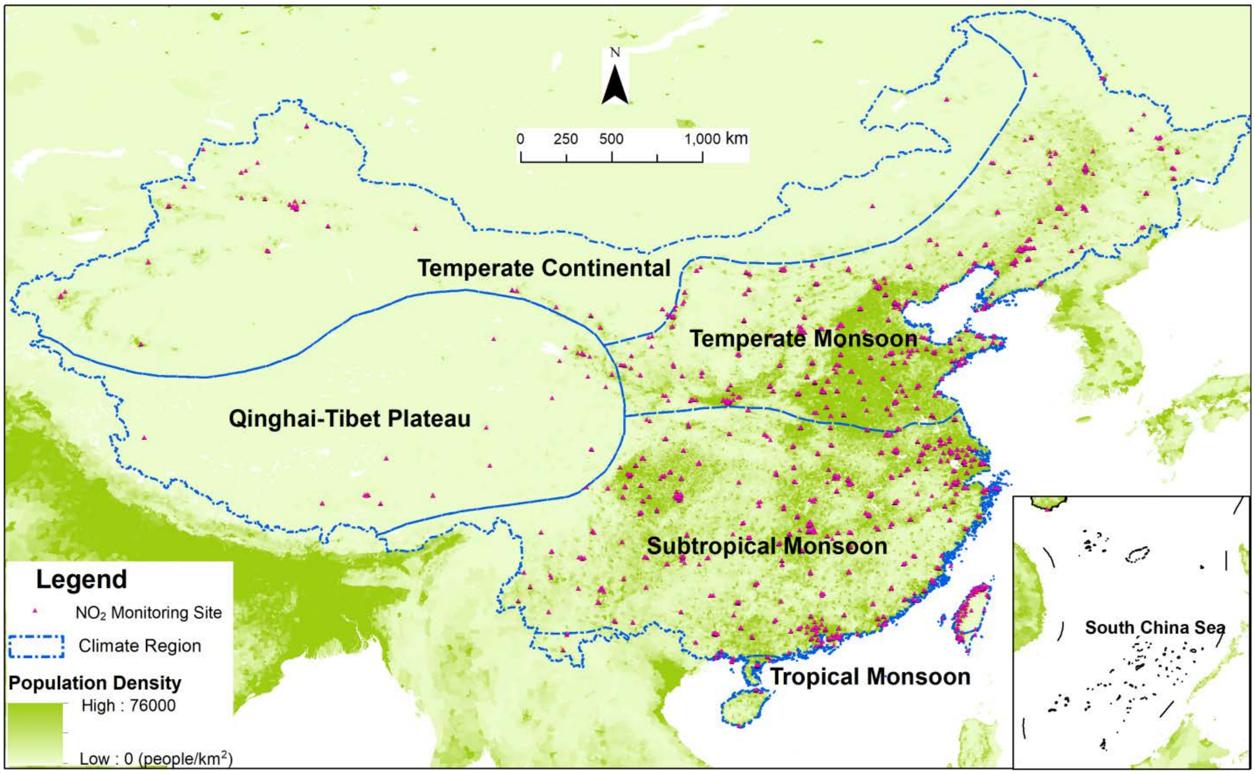

Figure S10. Major climate regions of China.

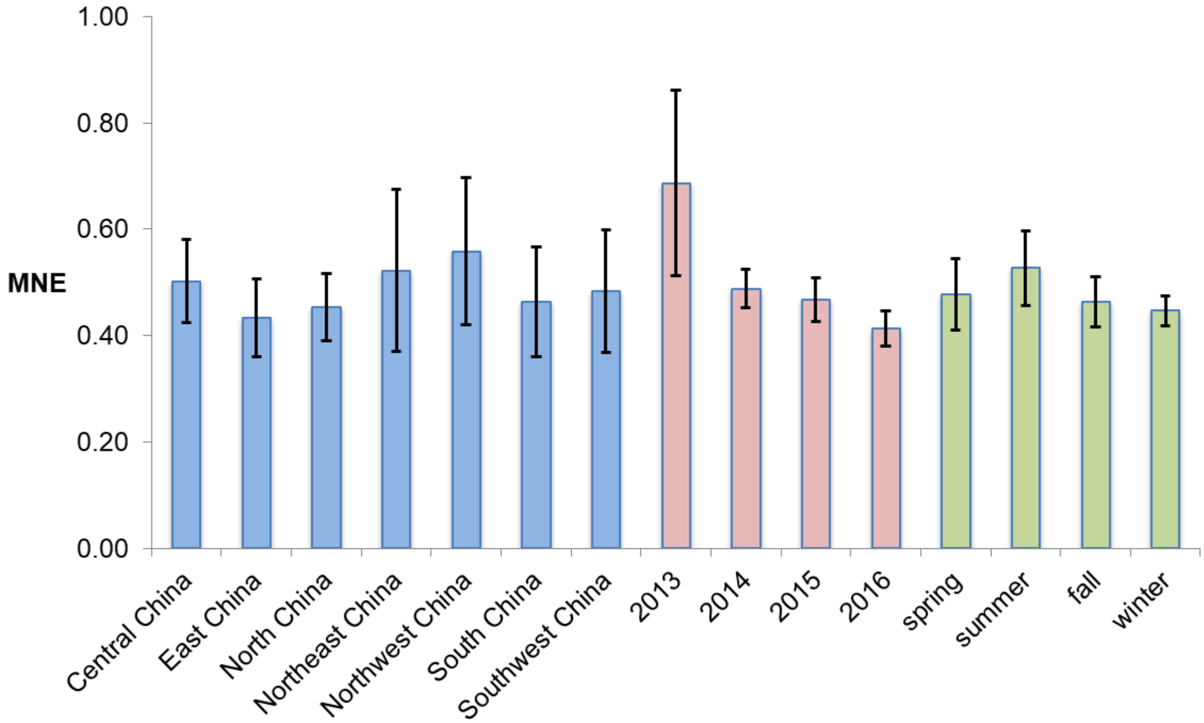

Figure S11. Performance of the random-forest-spatiotemporal-kriging (RF-STK) model in predicting daily NO_2 by regions, years, and seasons. The mean and standard deviation of the mean normalized error (MNE) over all of the 10 cross-validation folds are presented. The numbers of monitoring sites in 2013, 2014, 2015, and 2016 are 744, 1022, 1612, and 1604, respectively. The annual average numbers of monitoring sites in Central, East, North, Northeast, Northwest, South, and Southwest China are 173, 200, 256, 125, 112, 239, and 140, respectively.

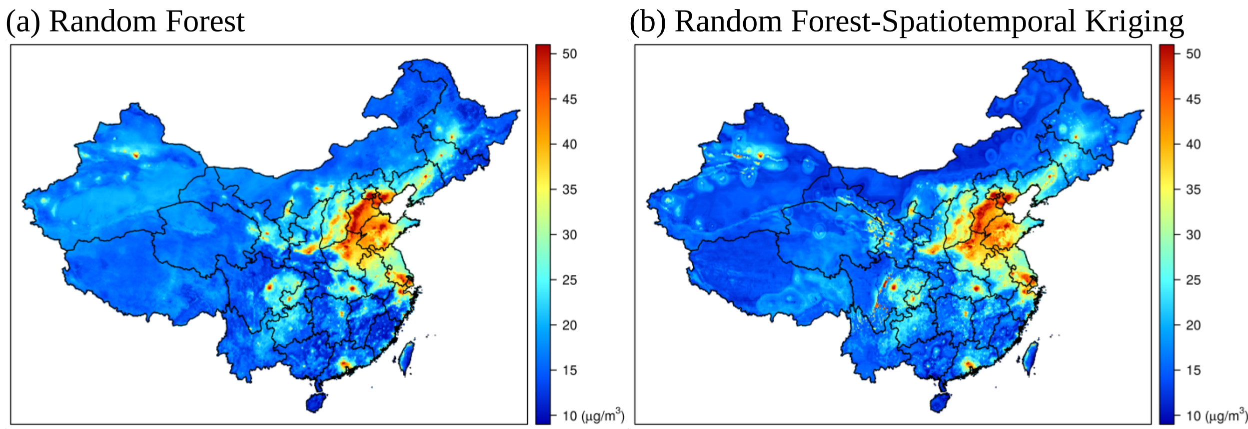

Figure S12. Average NO_2 during 2013-2016 predicted by (a) the random forest model and (b) the random-forest-spatiotemporal-kriging model.

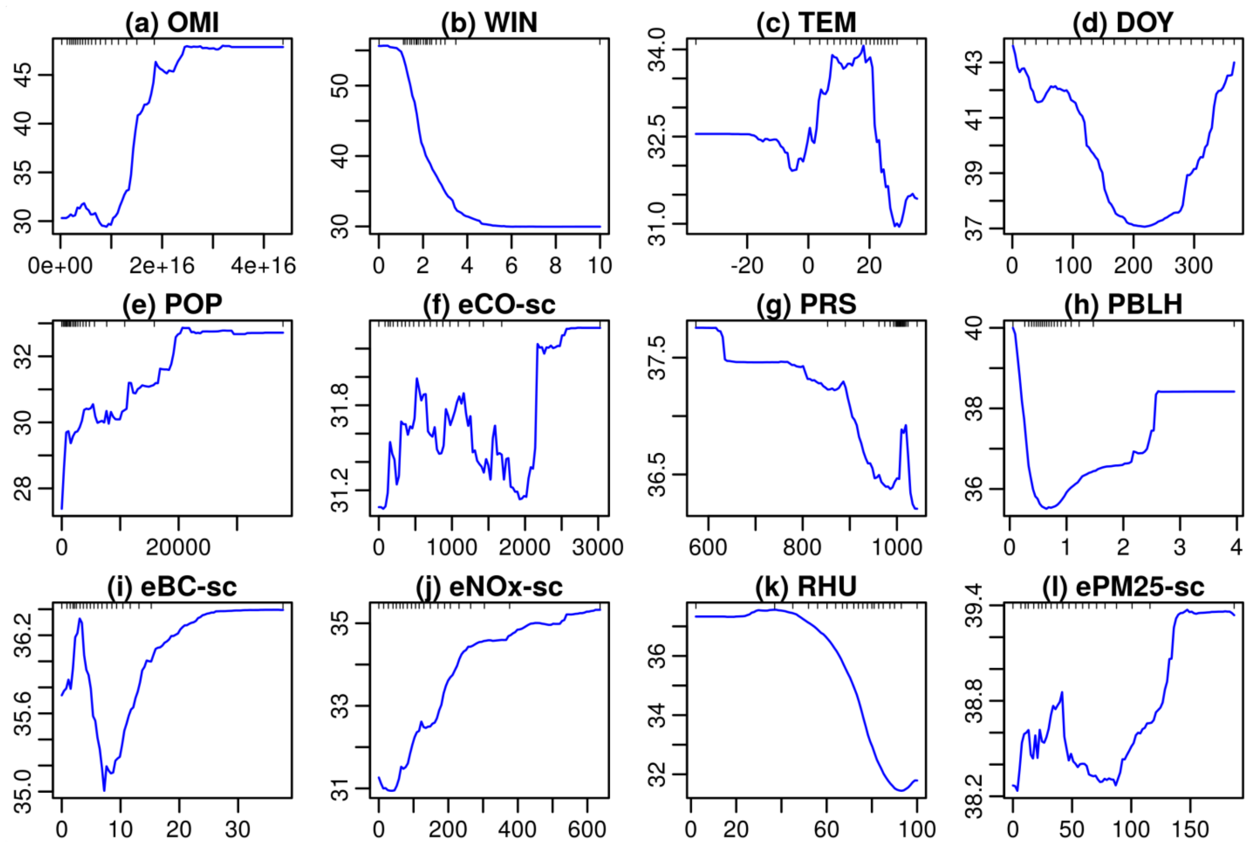

Figure S13. Partial dependence plots for the random forest model, indicating the effects of the predictor variables (a) OMI, (b) WIN, (c) TEM, (d) DOY, (e) POP, (f) eCO-sc, (g) PRS, (h) PBLH, (i) eBC-sc, (j) eNOx-sc, (k) RHU, and (l) ePM25-sc on the NO_2 predictions. The rug plot indicates the data density. Note that the partial dependence estimation tends to be unreliable at the two ends of horizontal axis due to their low data densities. Refer to Table S3 for the descriptions of the predictor variables.

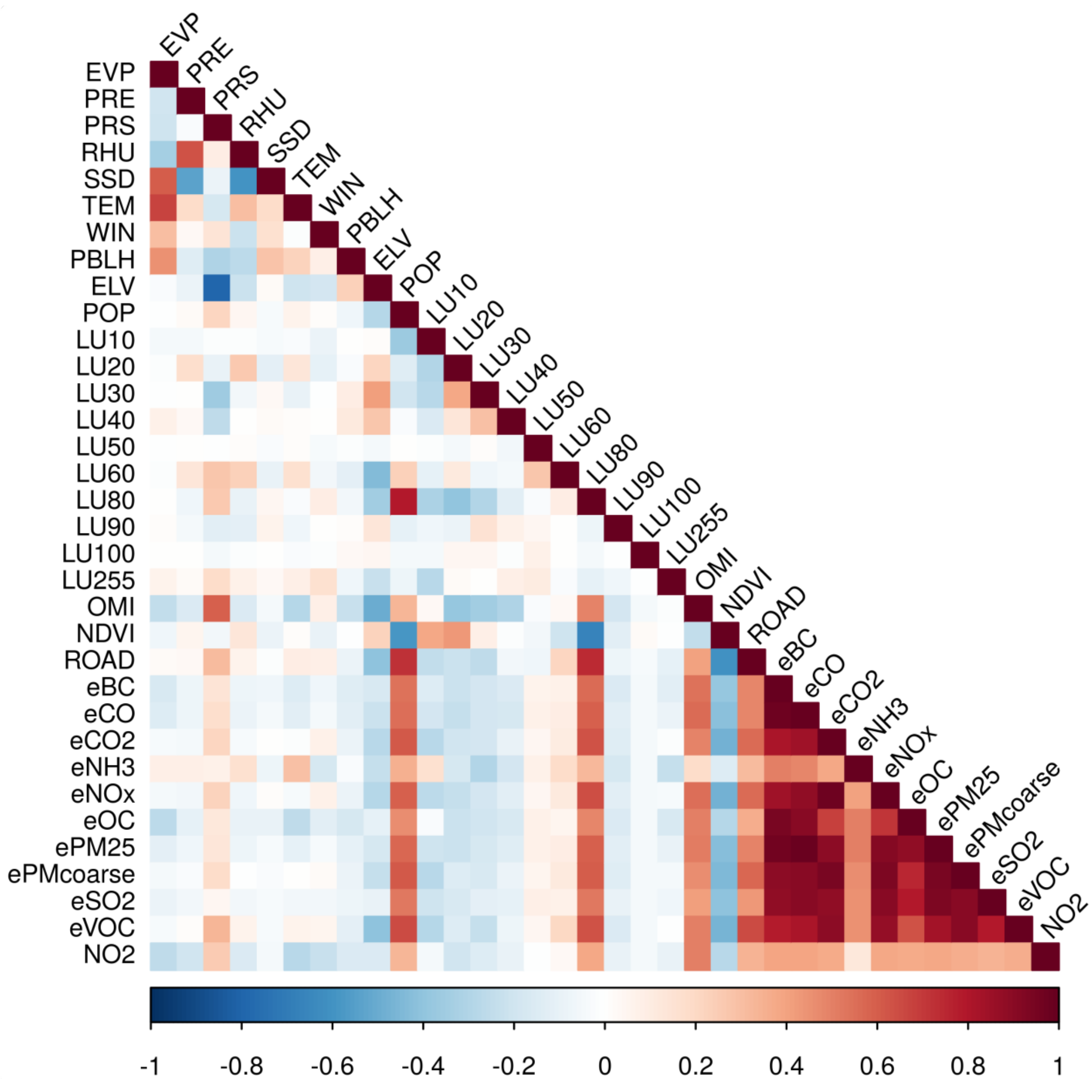

Figure S14. Correlations among the geographic factors and NO_2. The Spearman's rank correlation coefficients are used due to the prevalence of nonlinearity. Refer to Table S3 for the detailed descriptions of the variables.

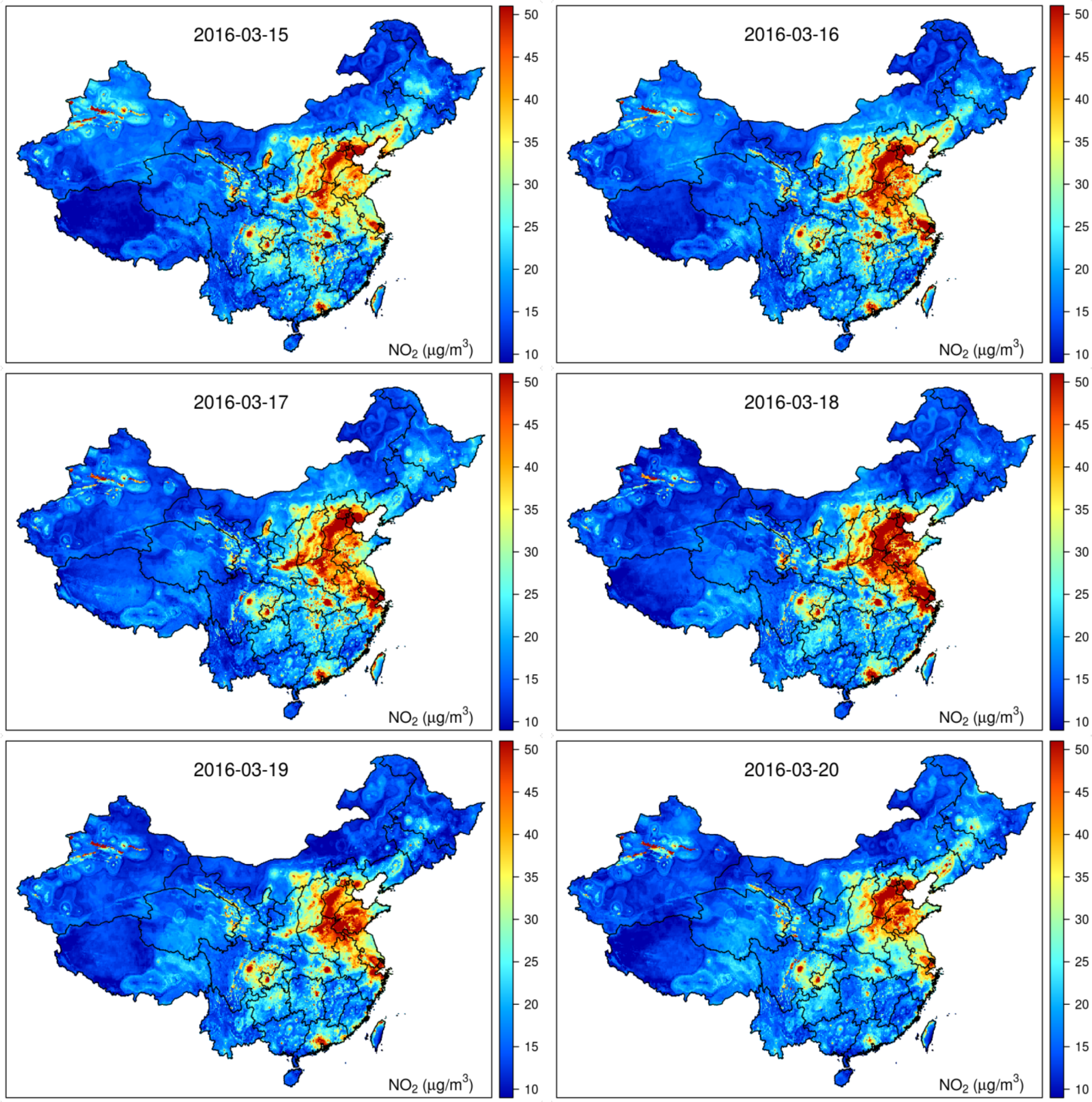

Figure S15. Daily surfaces of the ambient NO_2 concentrations across China from March 15 to March 20, 2016, which are examples of daily predictions by the random forest-spatiotemporal kriging (RF-STK) model.

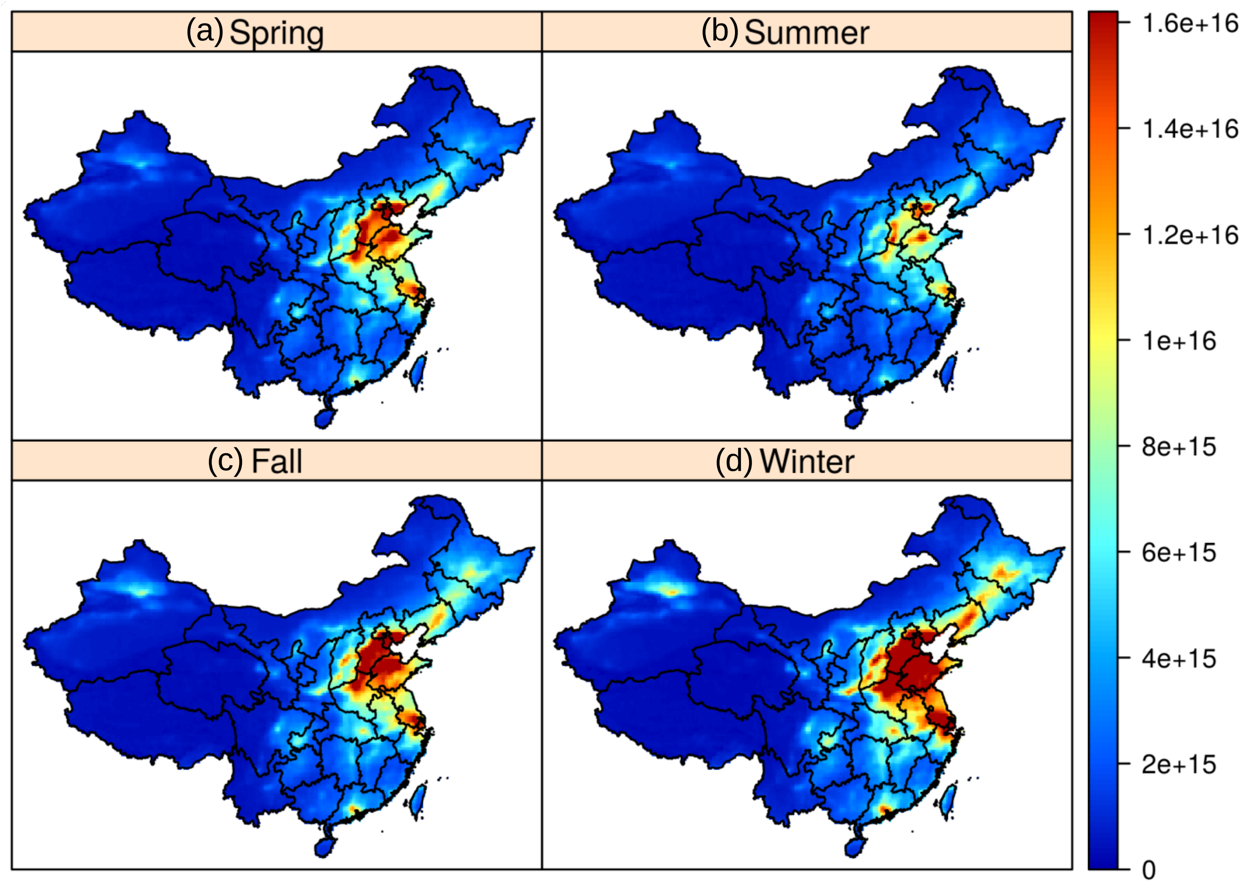

Figure S16. Vertical column density (molecules/cm^2) of tropospheric NO_2 retrieved from the OMI Level-3 product and processed by temporal convolution for (a) spring, (b) summer, (c) fall, and (d) winter during 2013-2016.

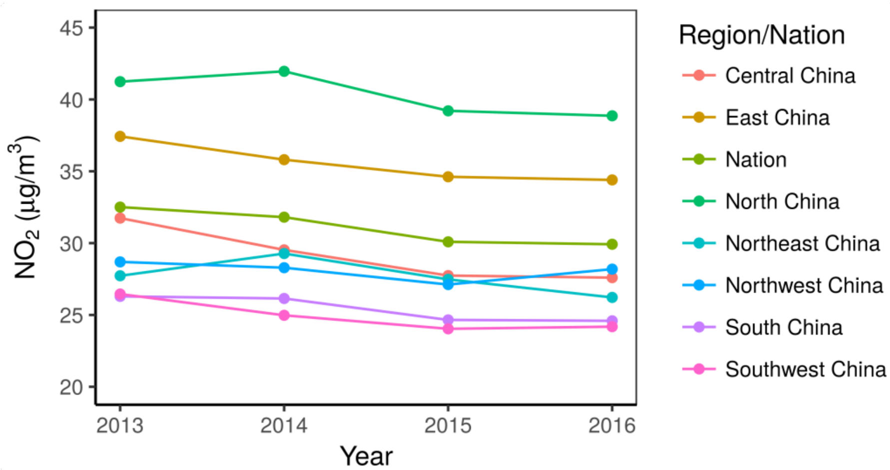

Figure S17. Population-weighted annual averages of NO_2 predictions for the major regions and the whole nation of China.

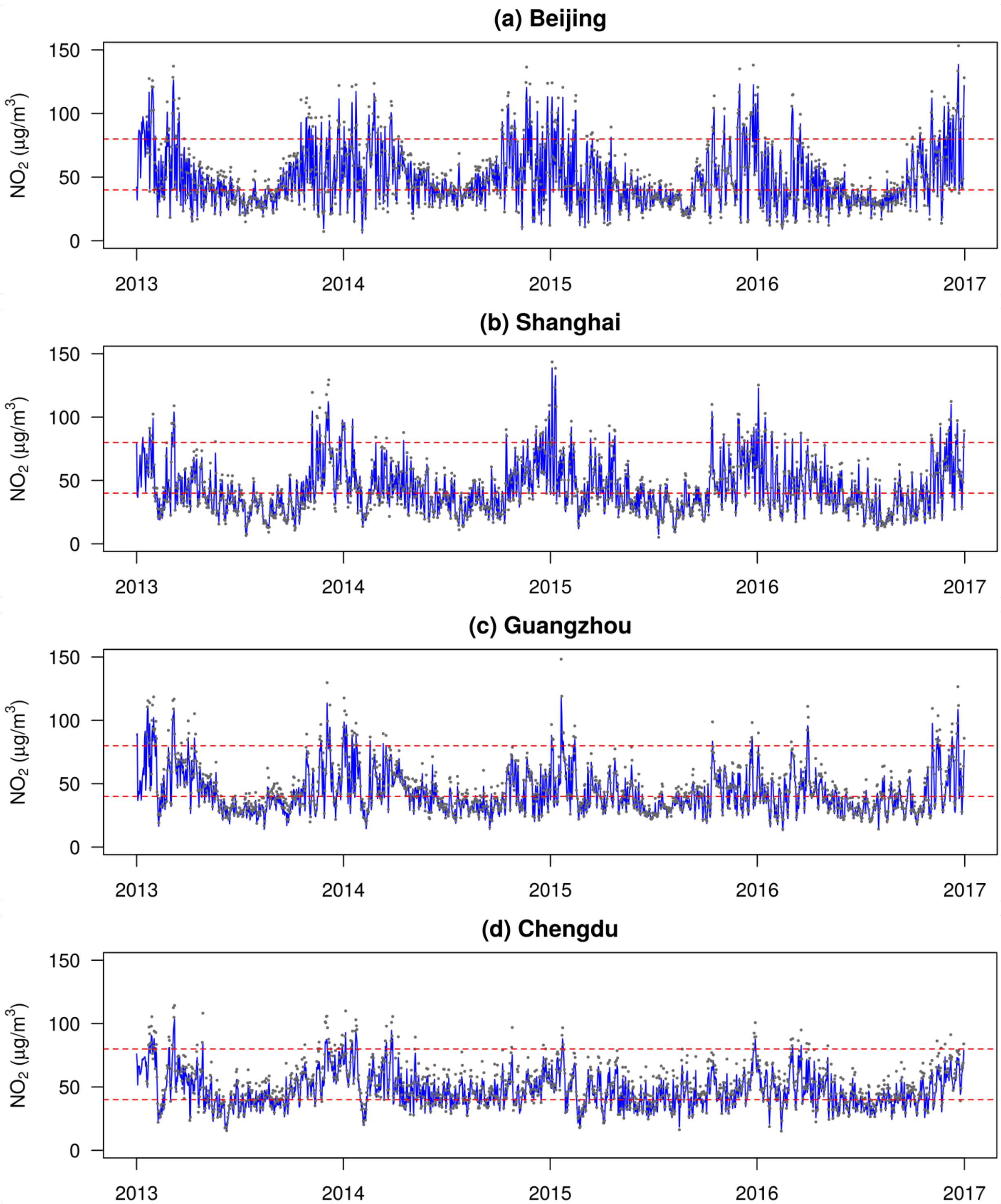

Figure S18. Observations (black dots) and predictions (blue lines) of daily NO_2 during 2013-2016 for (a) Beijing, (b) Shanghai, (c) Guangzhou, and (d) Chengdu, which are the major cities in the Beijing-Tianjin Metro, Yangtze River Delta, Pearl River Delta, the Sichuan Basin regions, respectively.

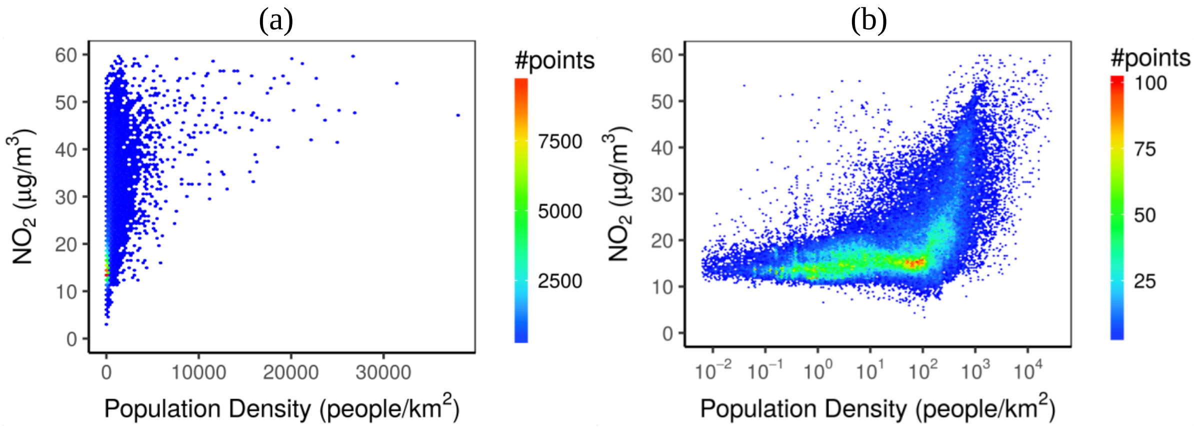

Figure S19. Scatterplots showing the relationship between the population density and the estimated average ambient NO_2 concentrations for 2013-2016 in China. The population density is at (a) original and (b) logarithm scales.

REFERENCES

1. Breiman, L., Random Forests. *Mach. Learn.* **2001,** *45* (1), 5-32.
2. Knibbs, L. D.; Hewson, M. G.; Bechle, M. J.; Marshall, J. D.; Barnett, A. G., A national satellite-based land-use regression model for air pollution exposure assessment in Australia. *Environ. Res.* **2014,** *135*, 204-211.
3. Lee, H. J.; Koutrakis, P., Daily ambient NO_2 concentration predictions using satellite ozone monitoring instrument NO_2 data and land use regression. *Environ. Sci. Technol.* **2014,** *48* (4), 2305-2311.
4. Hystad, P.; Setton, E.; Cervantes, A.; Poplawski, K.; Deschenes, S.; Brauer, M.; van Donkelaar, A.; Lamsal, L.; Martin, R.; Jerrett, M., et al., Creating national air pollution models for population exposure assessment in Canada. *Environ. Health Perspect.* **2011,** *119* (8), 1123-1129.
5. Vienneau, D.; de Hoogh, K.; Bechle, M. J.; Beelen, R.; van Donkelaar, A.; Martin, R. V.; Millet, D. B.; Hoek, G.; Marshall, J. D., Western European land use regression incorporating satellite- and ground-based measurements of NO_2 and PM_{10}. *Environ. Sci. Technol.* **2013,** *47* (23), 13555-13564.
6. Hoek, G.; Eeftens, M.; Beelen, R.; Fischer, P.; Brunekreef, B.; Boersma, K. F.; Veefkind, P., Satellite NO_2 data improve national land use regression models for ambient NO_2 in a small densely populated country. *Atmos. Environ.* **2015,** *105*, 173-180.
7. Liu, W.; Li, X.; Chen, Z.; Zeng, G.; León, T.; Liang, J.; Huang, G.; Gao, Z.; Jiao, S.; He, X., et al., Land use regression models coupled with meteorology to model spatial and temporal variability of NO_2 and PM_{10} in Changsha, China. *Atmos. Environ.* **2015,** *116*, 272-280.
8. 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_2 concentration in Shanghai, China. *Environ. Res.* **2015,** *137*, 308-315.
9. Yang, X.; Zheng, Y.; Geng, G.; Liu, H.; Man, H.; Lv, Z.; He, K.; de Hoogh, K., Development of $PM_{2.5}$ and NO_2 models in a LUR framework incorporating satellite remote sensing and air quality model data in Pearl River Delta region, China. *Environ. Pollut.* **2017,** *226*, 143-153.
10. de Hoogh, K.; Gulliver, J.; Donkelaar, A. v.; Martin, R. V.; Marshall, J. D.; Bechle, M. J.; Cesaroni, G.; Pradas, M. C.; Dedele, A.; Eeftens, M., et al., Development of West-European $PM_{2.5}$ and NO_2 land use regression models incorporating satellite-derived and chemical transport modelling data. *Environ. Res.* **2016,** *151*, 1-10.
11. Kim, Y.; Guldmann, J.-M., Land-use regression panel models of NO_2 concentrations in Seoul, Korea. *Atmos. Environ.* **2015,** *107*, 364-373.
12. Larkin, A.; Geddes, J. A.; Martin, R. V.; Xiao, Q.; Liu, Y.; Marshall, J. D.; Brauer, M.; Hystad, P., A global land use regression model for nitrogen dioxide air pollution. *Environ. Sci. Technol.* **2017,** *51* (12), 6957-6964.
13. Bechle, M. J.; Millet, D. B.; Marshall, J. D., National spatiotemporal exposure surface for NO_2: Monthly scaling of a satellite-derived land-use regression, 2000–2010. *Environ. Sci. Technol.* **2015,** *49* (20), 12297-12305.
14. Young, M. T.; Bechle, M. J.; Sampson, P. D.; Szpiro, A. A.; Marshall, J. D.; Sheppard, L.; Kaufman, J. D., Satellite-based NO_2 and model validation in a national prediction model based on universal kriging and land-use regression. *Environ. Sci. Technol.* **2016,** *50* (7), 3686-3694.
15. Novotny, E. V.; Bechle, M. J.; Millet, D. B.; Marshall, J. D., National satellite-based land-use regression: NO_2 in the United States. *Environ. Sci. Technol.* **2011,** *45* (10), 4407-4414.

16. CIESIN, Gridded Population of the World, Version 4 (GPWv4): Population count. In NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY, 2016.