

## Supporting Information for

# Hourly PM<sub>2.5</sub> Estimates from a Geostationary Satellite Based on an Ensemble Learning Algorithm and Their Spatiotemporal Patterns over Central East China

Jianjun Liu <sup>1</sup>, Fuzhong Weng <sup>2\*</sup>, Zhanqing Li <sup>3,4</sup>, and Maureen C. Cribb <sup>3</sup>

<sup>1</sup>Laboratory of Environmental Model and Data Optima (EMDO), Laurel, MD 20707, USA

<sup>2</sup>State Key Laboratory of Severe Weather, Beijing 100081, China

<sup>3</sup>Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA

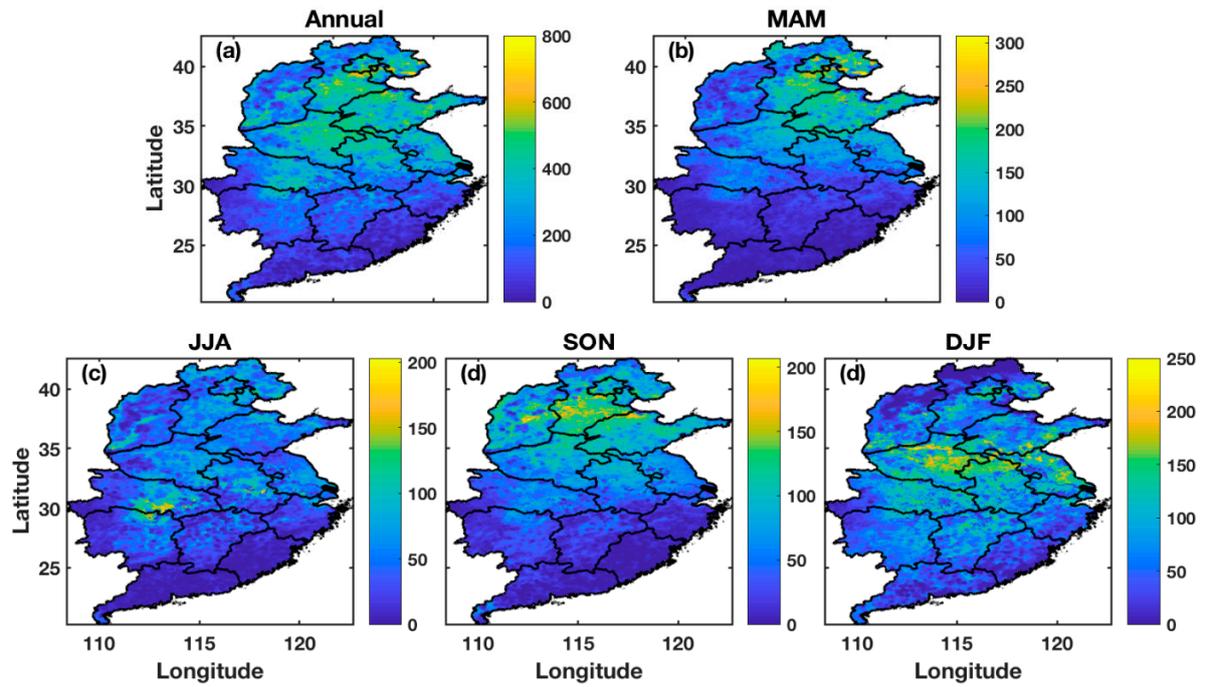
<sup>4</sup>Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742, USA

\* Correspondence: [wengfz@cma.gov.cn](mailto:wengfz@cma.gov.cn)

### Contents of this file

Figures S1 to S13

Table S1



**Figure S1.** The number of (a) annual and (b-e) seasonal AHI Level-3 AODs with the highest confidence level over central East China from 1 January 2016 to 31 December 2016. The seasons are defined by groups of months: spring (March-April-May, or MAM), summer (June-July-August, or JJA), autumn (September-October-November, or SON), and winter (December-January-February, or DJF).

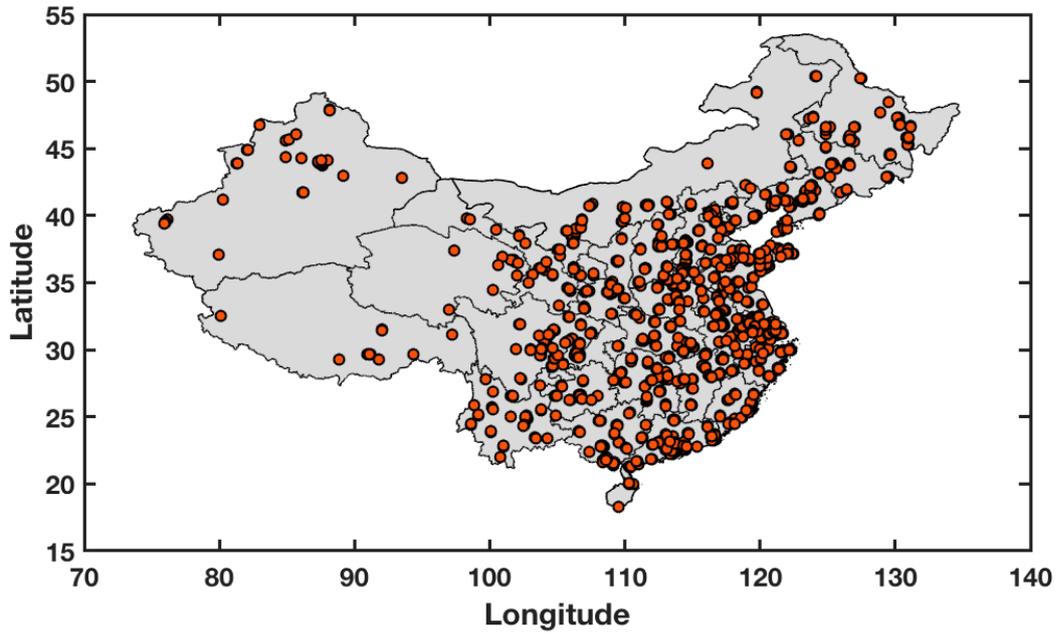


Figure S2. Spatial distribution of PM<sub>2.5</sub> monitoring sites in mainland China used in this study.

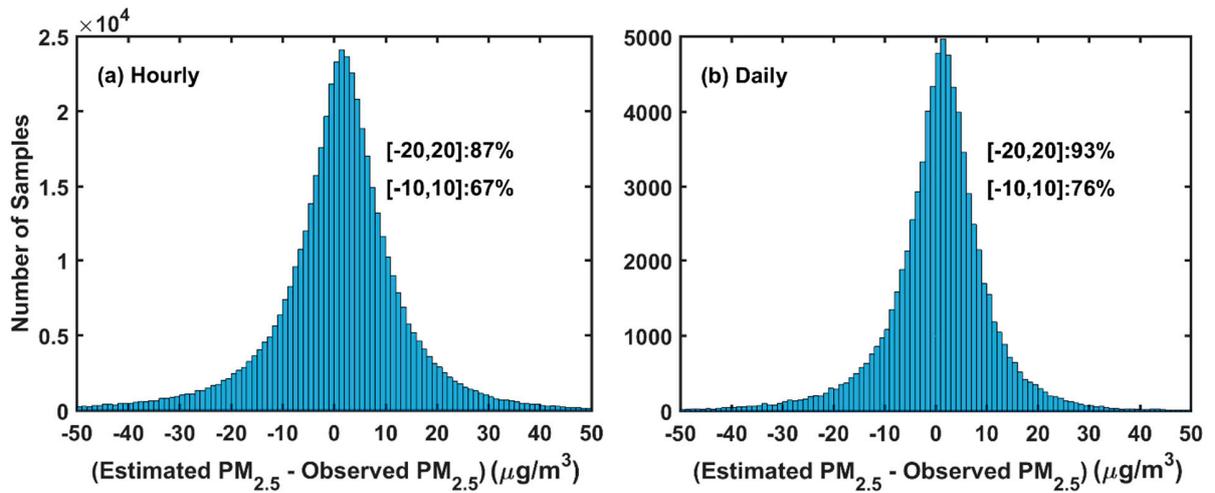
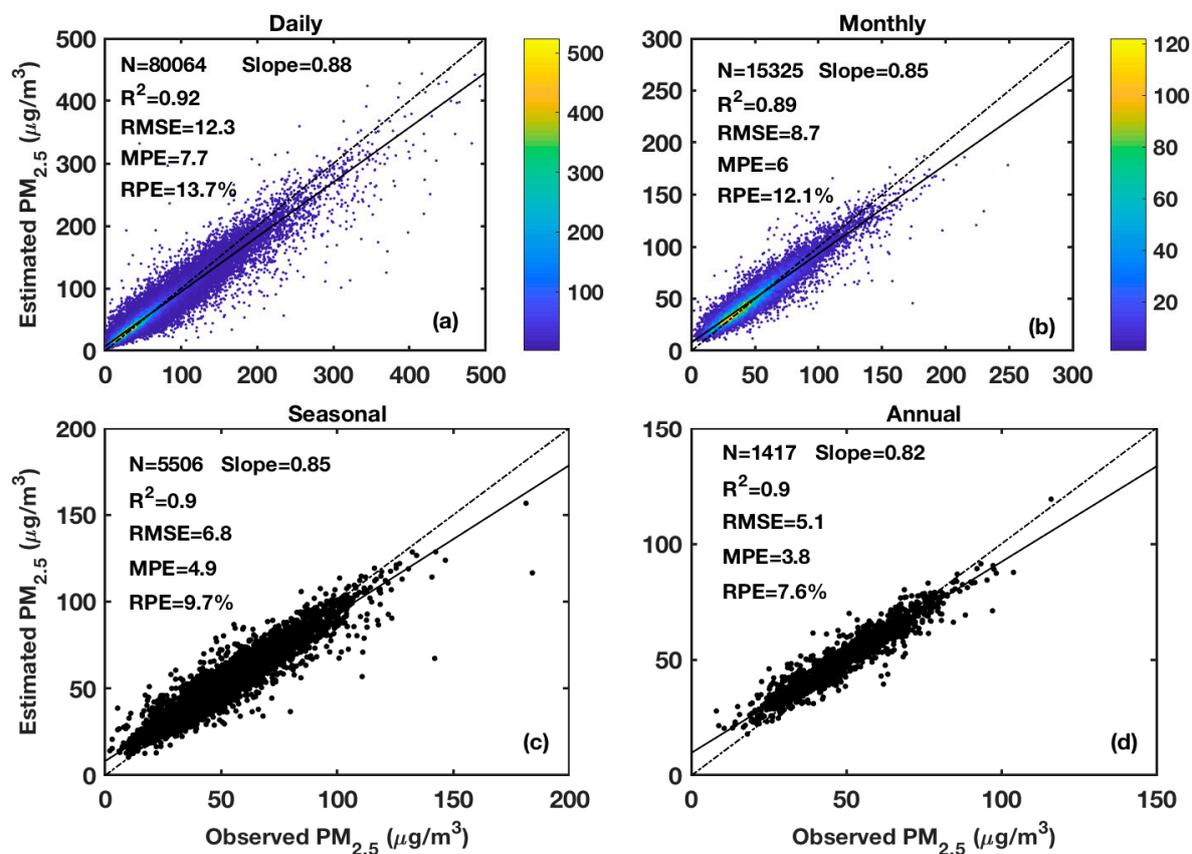
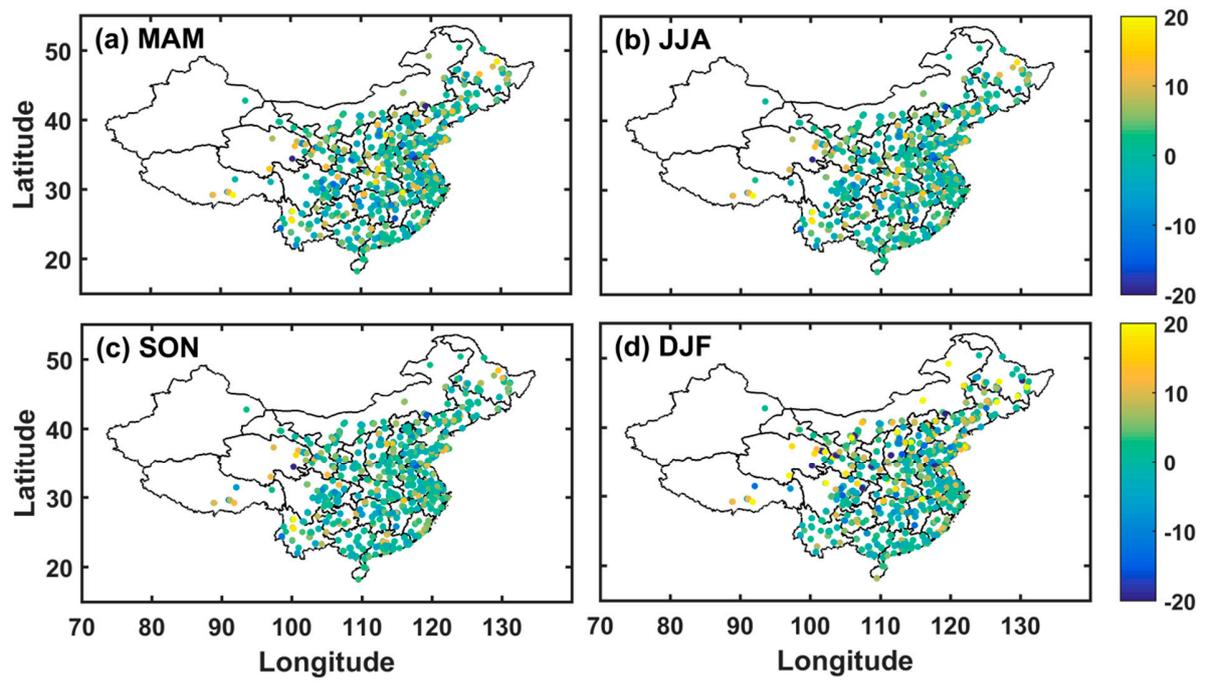


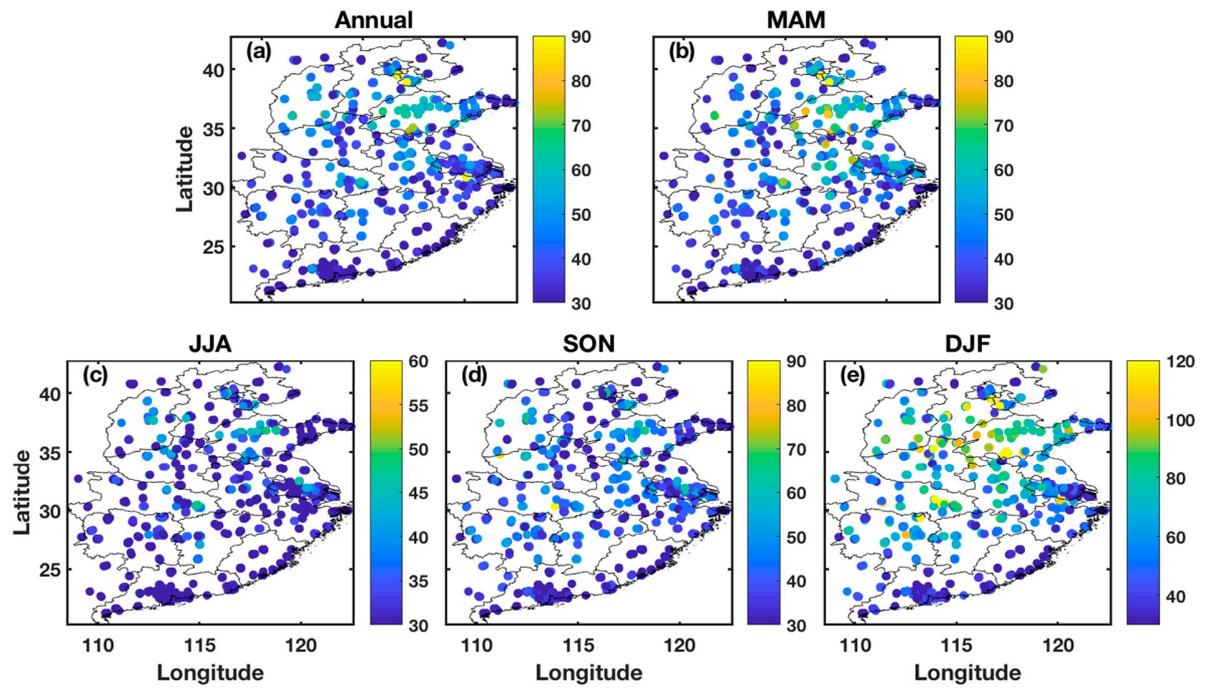
Figure S3. Histograms of the biases of model cross-validation-estimated PM<sub>2.5</sub> concentrations at (a) hourly and (b) daily levels. Each panel shows the percentage of samples falling within two ranges of values (in square brackets).



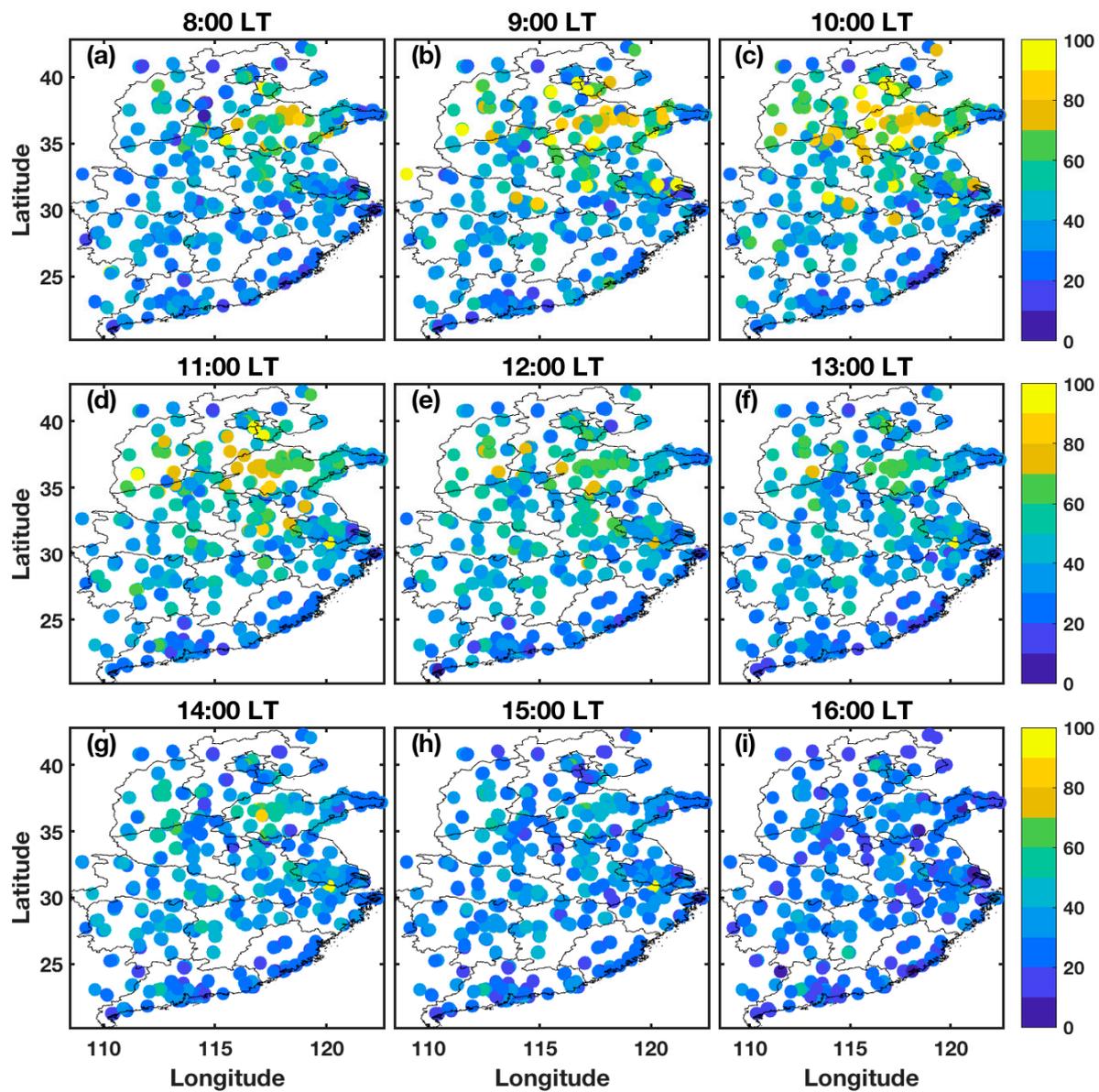
**Figure S4.** Scatter plots of the cross-validation of estimated  $PM_{2.5}$  concentrations by comparing surface-measured  $PM_{2.5}$  concentrations at (a) daily, (b) monthly, (c) seasonal, and (d) annual levels. The dashed lines are 1:1 lines. N: number of samples;  $R^2$ : coefficient of determination; RMSE: root-mean-square error ( $\mu g m^{-3}$ ); MPE: mean prediction error ( $\mu g m^{-3}$ ); RPE: relative prediction error (%).



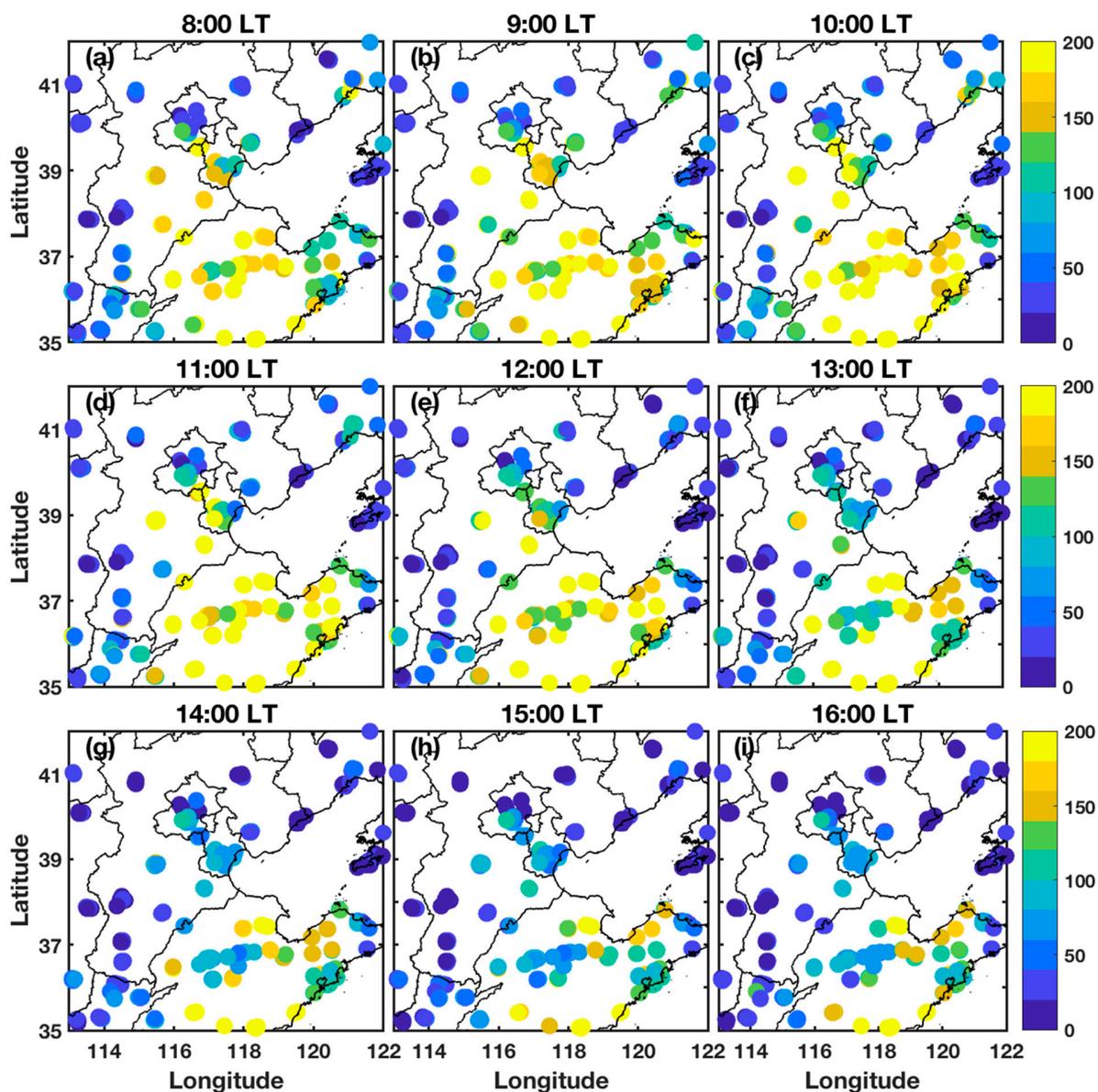
**Figure S5.** Differences between model-estimated and surface-measured seasonal mean PM<sub>2.5</sub> concentrations at each site in different seasons: (a) March, April, and May (MAM), (b) June, July, and August (JJA), (c) September, October, and November (SON), and (d) December, January, and February (DJF). Units are  $\mu\text{g m}^{-3}$ .



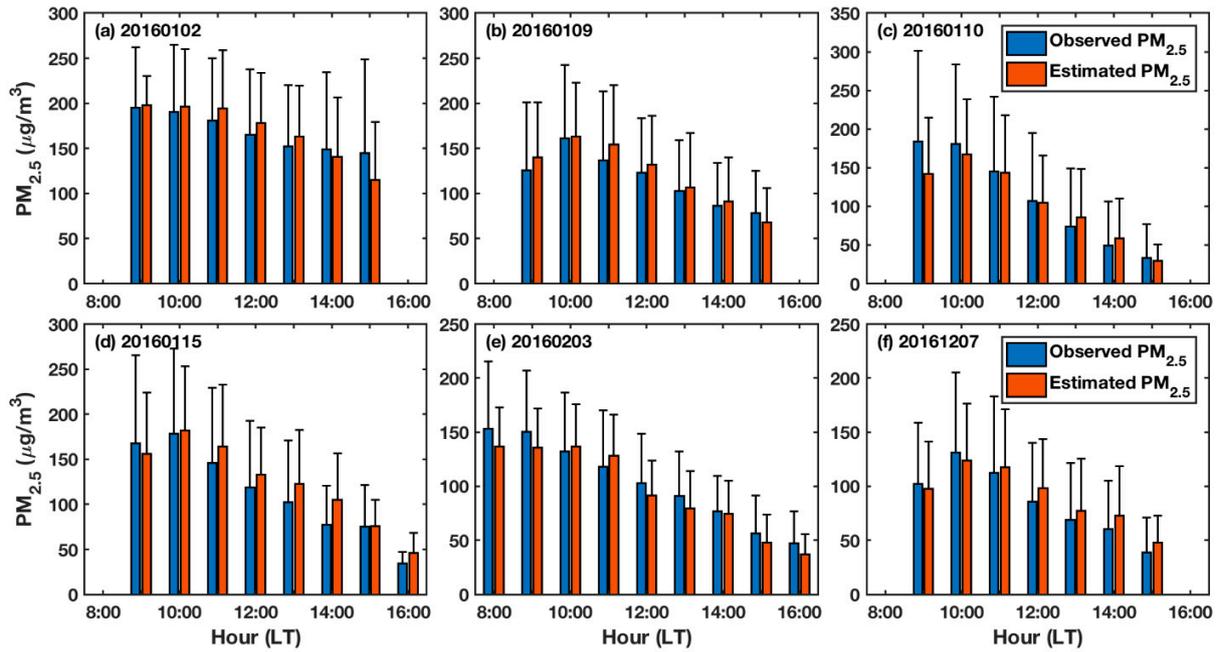
**Figure S6.** Surface-observed PM<sub>2.5</sub> concentrations (μg m<sup>-3</sup>) over central East China: (a) for the whole year of 2016, (b) March, April, and May (MAM), (c) June, July, and August (JJA), (d) September, October, and November (SON), and (e) December, January, and February (DJF).



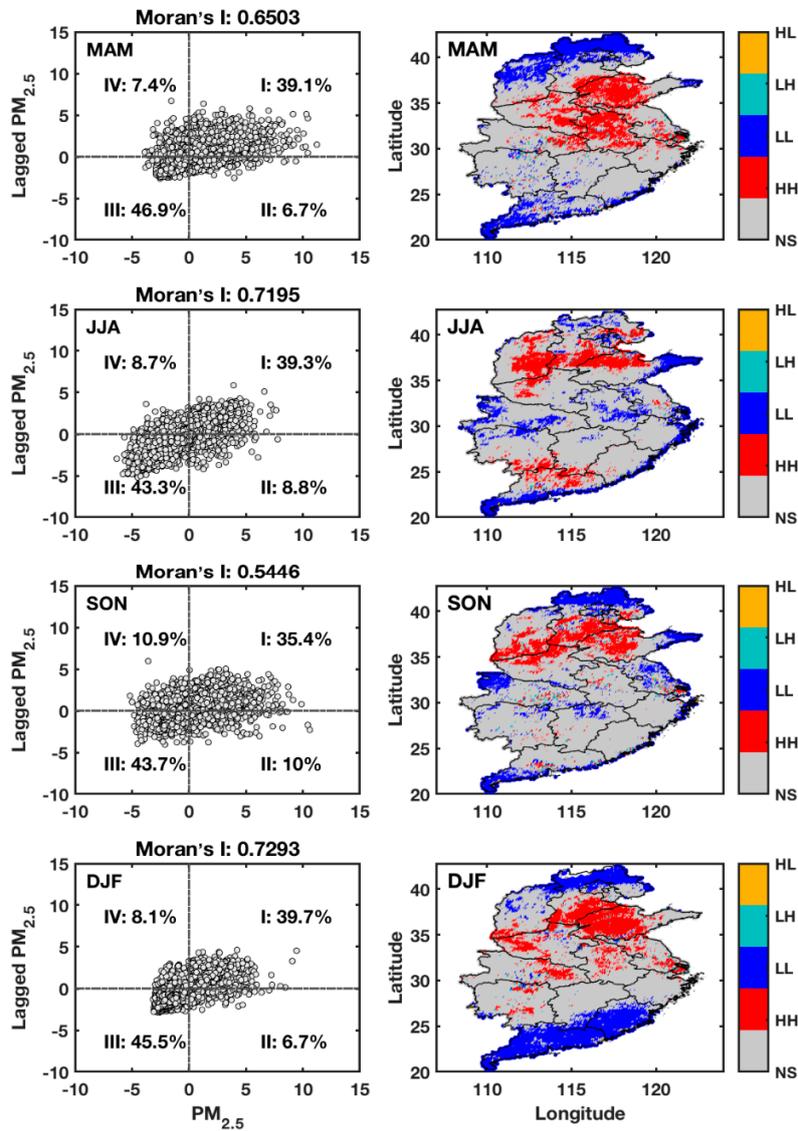
**Figure S7.** Spatial distributions of mean surface-measured PM<sub>2.5</sub> concentrations (µg m<sup>-3</sup>) over central East China for (a-i) different hours of the day (0800–1600 local time, or LT).



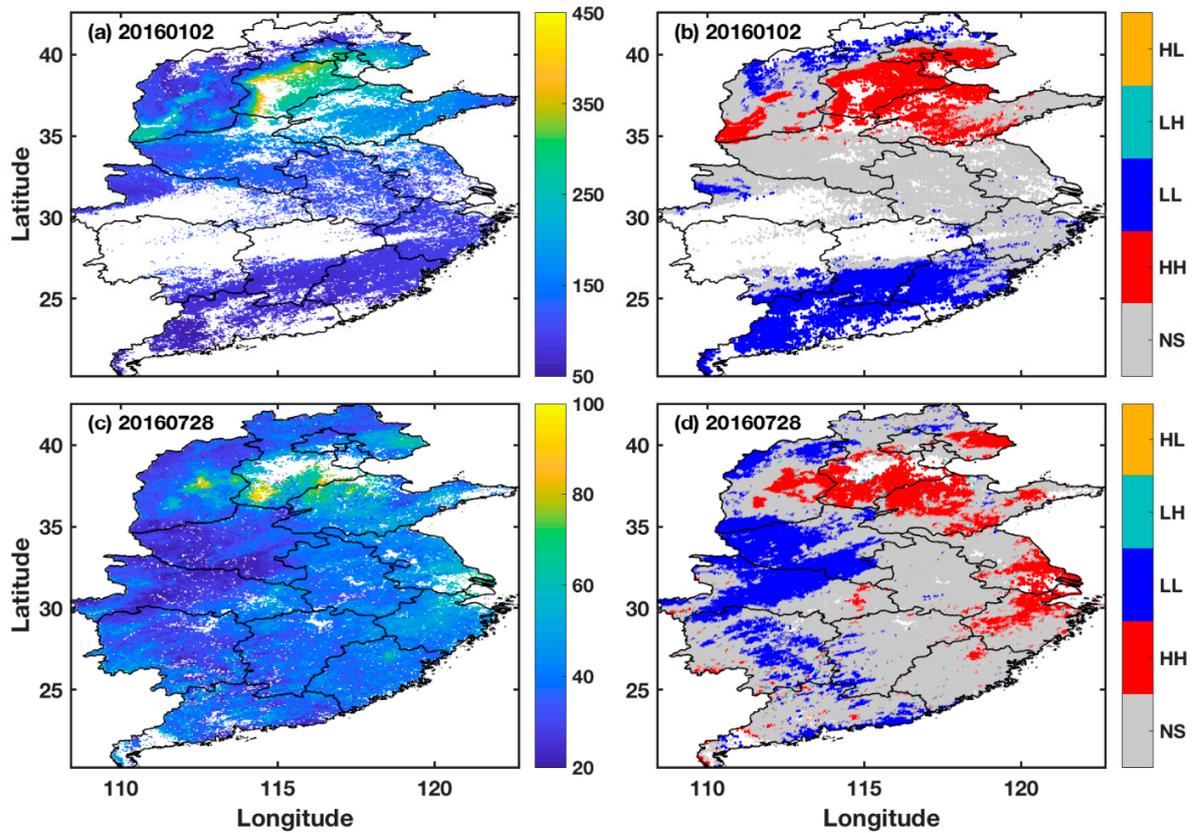
**Figure S8.** Spatial distributions of surface-measured hourly PM<sub>2.5</sub> concentrations (µg m<sup>-3</sup>) for a high pollution episode that occurred on 14 January 2019 over the North China Plain for (a-i) different hours of the day (8:00-16:00 LT). LT: local time.



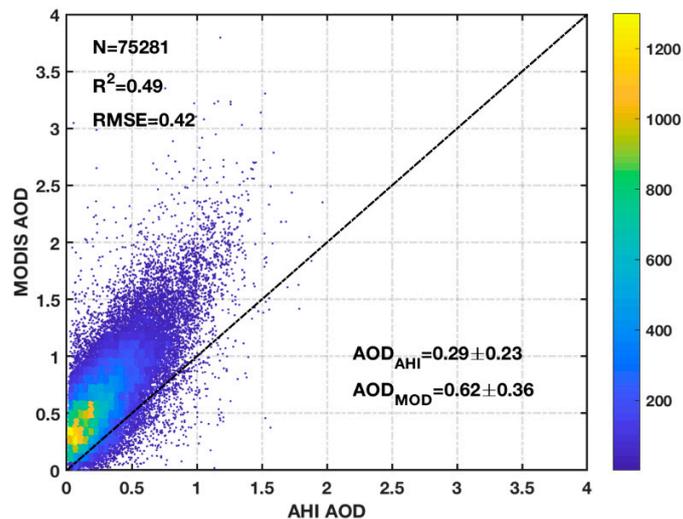
**Figure S9.** Diurnal cycles of mean model-estimated (red bars) and surface-observed (blue bars) PM<sub>2.5</sub> concentrations with standard deviations for (a-f) several high PM<sub>2.5</sub> episodes that occurred over the North China Plain (35-42°N, 113-122°E; LT: local time). The dates are in the YYYYMMDD format where YYYY = year, MM = month, and DD = day.



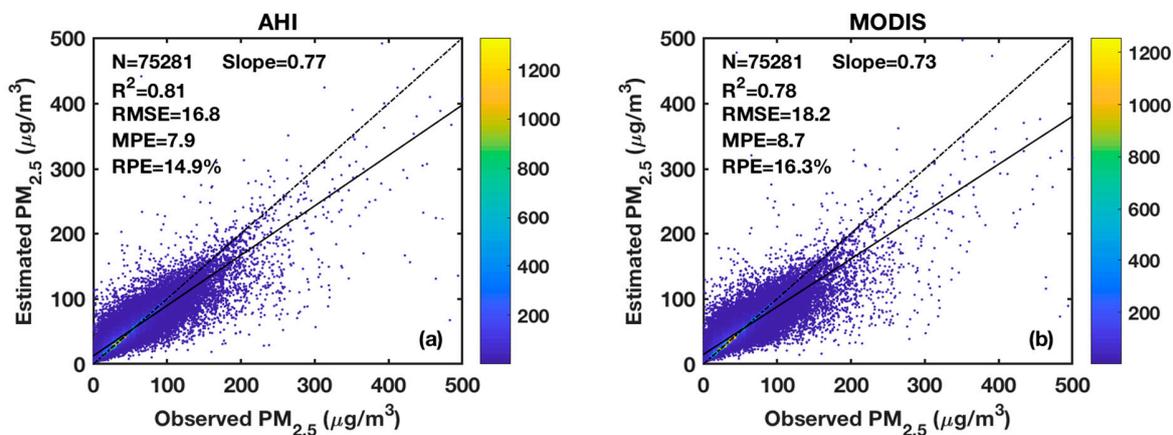
**Figure S10.** Left panels: Scatter plots of the global Moran Index for the four seasons (from top to bottom, March-April-May (MAM), June-July-August (JJA), September-October-November (SON), and December-January-February (DF)). Right panels: Spatial agglomeration diagrams of seasonal model-estimated  $PM_{2.5}$  concentrations over central East China for the four seasons. The numbers in the left panels are the percentages of samples with the aggregation patterns of (going clockwise from the upper right) I, II, III, and IV. The spatial agglomeration diagrams pass the significance test at a significance level of 0.05. The legends on the right give the spatial agglomeration category: high-low (HL), low-high (LH), low-low (LL), high-high (HH), and no significance (NS).



**Figure S11.** Spatial distributions of PM<sub>2.5</sub> concentrations (a, c) and their relevant spatial agglomeration characteristics (b, d) for a high pollution episode that occurred on 2 January 2016 over the North China Plain (a, b) and a relatively low pollution episode that occurred on 28 July 2016 (c, d). The spatial agglomeration diagrams pass the significance test at a significance level of 0.05. The legends in (b) and (d) give the spatial agglomeration category: high-low (HL), low-high (LH), low-low (LL), high-high (HH), and no significance (NS).



**Figure S12.** Scatter plot of the AHI-retrieved AOD as function of MODIS-retrieved AOD at 500 nm over all PM<sub>2.5</sub> sites in 2016. The dashed line is the 1:1 line. N: number of samples; R<sup>2</sup>: coefficient of determination; RMSE: root-mean-square error ( $\mu\text{g m}^{-3}$ ).



**Figure S13.** Scatter plots of cross-validation of the RF model of (a) AHI AOD and (b) MODIS AOD. The dashed lines are 1:1 lines. N: number of samples; R<sup>2</sup>: coefficient of determination; RMSE: root-mean-square error ( $\mu\text{g m}^{-3}$ ); MPE: mean prediction error ( $\mu\text{g m}^{-3}$ ); RPE: relative prediction error.

**Table S1.** Summary of estimates of PM<sub>2.5</sub> concentrations from satellite AODs based on statistical models at regional and national scales in China. NA stands for “not available”.

Reference	Study Region*	Study Period	Source of AOD**	Model***	R <sup>2</sup> , RMSE (µg m <sup>-3</sup> ), and Slope (sample-based CV)	Major Inputs (except AOD)****
Wang et al., 2010 [68]	Beijing	July 2007- Oct. 2008	MODIS	LR	0.47, NA, NA	Observed: RH and PBLH
Xie et al., 2015 [25]	Beijing	Mar. 2013- Apr. 2014	MODIS	LR LME	0.45, 32.40, 0.99 0.79, 18.89, 1.02	None
Li et al., 2015 [45]	Beijing	Mar. 2013- Feb. 2014	MODIS	LME	0.80, 16.04, 0.81	None
Guo et al., 2017 [64]	Beijing	Apr. 2013 - Mar. 2015	MODIS	MLR GWR TWR GTWR	0.18, 42.83, NA 027, 40.4, NA 048, 34.03, NA 0.58, 30.81, 0.59	Simulated: RH, PS, T, WD, WS, PBLH NDVI, population data, elevation data
Wu et al., 2016 [36]	BTH	2014	VIIRS	Two-stage model	0.72, 19.29, 0.88	Measured: T, RH, RF, WS, WD Simulated: PBLH, RH_PBLH Retrieved: NDVI, NO <sub>2</sub>
Lv et al., 2017 [28]	BTH	2014	MODIS	Bayesian-based statistical downscaler	0.58, 24.59, NA	Simulated: RH, T, PBLH, PS, WS Elevation data
He and Huang, 2018 [59]	BTH	2013– 2015	MODIS	GWR GTWR iGTWR	0.6, 41.41, 0.74 0.84, 27.84, 0.90	RH, WS, T, PS, PBLH, elevation, population, NDVI, water

Ma et al., 2016b [66]	YRD	2013	MODIS	NLME	0.82, 29.96, 0.84 10 km: 0.73 3 km: 0.67, 15.82, 0.67	bodies, artificial surfaces Simulated: WS, RH_PBLH forest cover
Xiao et al., 2017 [70]	YRD	2013–2014	MAIAC	Two-stage model	2013: 0.81, 25.0, 0.99 2014: 0.73, 18.0, 1.00	Simulated: PBLH, T_PBLH, RH_PBLH, WS_PBLH, SF, RH, T, RF NDVI, elevation data, population
Zheng et al., 2016 [46]	BTH, YRD, PRD	2013	MODIS	LME	BTH: 0.77, 23.07, 0.78 YRD: 0.80, 17.89, 0.80 PRD: 0.80, 12.47, 0.80	Simulated: T, RH, WS, WD, PBLH Retrieved: NO <sub>2</sub>
Bai et al., 2016 [60]	Jiangsu, Shandong, Henan, and Anhui	Nov. 2014–Feb. 2015	MODIS	OLS GWR TWR GTWR	0.41, 45.91, 0.38 0.60, 37.90, 0.57 0.68, 33.59, 0.63 0.87, 21.77, 0.89	Simulated: PBLH, RH, WS, T
Chen et al., 2018 [62]	Guangzhou and surrounding region	2014–2015	MODIS	log-linear non-log-linear non-linear exposure-lag-response	0.67, NA, 0.99 0.76, NA, 0.99 0.81, NA, 0.99	Observed: RF, WS, WD, PS, T, VP, RH, sunshine duration
Tao et al., 2013 [67]	North China	July 2007–Oct. 2008	MODIS	LR	0.62, NA, NA	Simulated: RH and PBLH
Lv et al., 2016 [29]	North China	2014	MODIS	Bayesian hierarchical	0.68, 21.40, 0.69	Simulated: PS, T, RH, PBLH, land cover, elevation data

Ma et al., 2014 [15]	National scale	2000–2013	MODIS and MISR	GWR	0.64, 32.98, 0.67	Simulated: PBLH, T, WS, RH_PBLH, PS NDVI, population data
Fang et al., 2016 [27]	National scale	June 2013–May 2014	MODIS	TSAM	0.80, 22.75, 0.79	Observed: WS, RF, RH, T, PS roads, built-up area, forest area, grass area, and water area, population and elevation data, industrial and mining enterprises, and restaurants
Ma et al., 2016a [5]	National scale	Jan. 2004–June 2014	MODIS	LME LME+GAM	0.78, 27.99, 0.77 0.79, 27.42, 0.79	Simulated: PBLH, WS, RH_PBLH, PS, CP urban and forest cover
Guo et al., 2016 [63]	National scale	Jan. 2013–Oct. 2013	PARASOL	LR QR PR LOGR	0.41, 22.68, NA 0.40, 24.21, NA 0.38, 23.91, NA 0.32, 21.76, NA	None
Li et al., 2016 [65]	National scale	Feb. 2013–Dec. 2014	MODIS	CLR MLR SEM GWR BPNN GRNN	0.24, 31.51, NA 0.28, 30.52, NA 0.30, 30.80, NA 0.37, 28.68, NA 0.48, 25.96, NA 0.67, 20.93, 0.62	Simulated: RH, T, WS, P, PBLH

You et al., 2016b [71]	National scale	2014	MODIS	GWR	0.79, 18.6, 0.83	Measured: WS, T, Vis, RH
You et al., 2016a [23]	National scale	2014	MODIS MISR	GWR	MODIS: 0.79, 20.85, 0.82 MISR: 0.85, 24.86, 0.87	Measured: WS, T, Vis, RH
Li et al., 2017 [17]	National scale	2015	MODIS	DBN	0.88, 13.03, 0.88	Simulated: RH, T, WS, PS, PBLH; NDVI; population and road data
Yu et al., 2017 [6]	National scale	2013	MODIS	Gaussian process GWR LME	Gaussian: 0.81, 21.87, 0.73 GWR: 0.74, 25.71, 0.72 LME: 0.48, 30.38, 0.48	None
Zhan et al., 2017 [31]	National scale	2014		GW-GBM	0.76, 23.0, 0.77	Measured: T, PS, evaporation, RH, RF, sunshine duration, WS
Chen et al., 2018 [61]	National scale	2014–2016	MODIS	DLNM+GAM GAM Random forest	0.51, 30.29, 0.98 0.55, 29.13, 0.98 0.83, 18.08, 1.07	Observed: T, R, PS, WS, NDVI, urban cover, elevation
Xiao et al., 2018 [69]	National scale except less regions with PM <sub>2.5</sub> sites	Nov. 2015–Feb. 2016	MODIS	BEM+GWR	0.88, 11.39, NA	Measured: precipitation, T, RH, PS, WS grass, water, urban, city, forest, roads, elevation data, and population data
Wei et al., 2019 [72]	National scale	2016	MODIS	STRF	0.85, 15.57, 0.82	Simulated: precipitation, T,

---

PBLH, RH, PS,  
WS, WD, ET,  
NDVI, DEM,  
LUC, NTL

---

\*BTH: Beijing-Tianjin-Hebei; YRD: Yangtze River Delta region; PRD: Pearl River Delta region.

\*\*MODIS: Moderate Resolution Imaging Spectroradiometer; MISR: Multi-angle Imaging Spectro Radiometer; MAIAC: Multi-Angle Implementation of Atmospheric Correction algorithm; PARASOL: Polarization & Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar.

\*\*\*LME: linear mixed-effects model; GAM: generalized additive model; GWR: geographically weighted regression model; LR: linear regression; OLS: ordinary least squares; TWR: temporally weighted regression model; QR: quadratic regression model; PR: power regression mode; LOGR: logarithmic regression model; GTWR: geographically and temporally weighted regression model; CLR: corrected linear regression; MLR: multiple linear regression; SEM: semi-empirical model; BPNN: back-propagation neural network; DBN: deep belief network; GRNN: generalized regression neural network model; DLNM: distributed lag non-linear model; BEM: Bayesian maximum entropy; iGTWR: improved geographically and temporally weighted regression model; NLME: nested linear mixed-effects model; GW-GBM: geographically weighted gradient boosting machine; STRF: space-time random forest model; TSAM: timely structure adaptive modeling.

\*\*\*\*T: temperature; RH: relative humidity; RF: rainfall; WS: wind speed; WD: wind direction; PBLH: planet boundary layer height; RH\_PBLH: average RH at the PBLH; T\_PBLH: average T at the PBLH; WS\_PBLH: average WS at the PBLH; NDVI: normalized difference vegetation index; PS: surface pressure; Vis: visibility; VP: vapor pressure; SF: surface incident shortwave flux; CP: cumulative precipitation; ET: evaporation; LUC: land-use cover; DEM: digital elevation model; NTL: nighttime lights.

## References

60. Bai, Y.; Wu, L.; Qin, K.; Zhang, Y.; Shen, Y.; Zhou, Y. A geographically and temporally weighted regression model for ground-level PM<sub>2.5</sub> estimations from satellite-derived 500-m resolution AOD. *Remote Sens.* **2016**, *8*, 262.
61. Chen, G.; Li, S.; Knibbs, L.D.; Hamm, N.A.S.; Cao, W.; Li, T.; Guo, J.; Ren, H.; Abramson, M.J.; Guo, Y. A machine learning method to estimate PM<sub>2.5</sub> concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* **2018**, *636*, 52–60.
62. Chen, Z.Y.; Zhang, T.H.; Zhang, R.; Zhu, Z.M.; Ou, C.Q.; Guo, Y. Estimating PM<sub>2.5</sub> concentrations based on non-linear exposure-lag-response associations with aerosol optical depth and meteorological measures. *Atmos. Environ.* **2018**, *173*, 30–37.
63. Guo, H.; Cheng, T.; Gu, X.; Chen, H.; Wang, Y.; Zheng, F.; Xiang, K. Comparison of four ground-level PM<sub>2.5</sub> estimation models using PARASOL aerosol optical depth data from China. *Int. J. Environ. Res. Public Health* **2016**, *13*, 180, doi:10.3390/ijerph13020180.
64. Guo, Y.; Tang, Q.; Gong, D.Y.; Zhang, Z. Estimating ground-level PM<sub>2.5</sub> concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sens. Environ.* **2017**, *198*, 140–149.
65. Li, T.; Shen, H.; Zeng, C.; Yuan, Q.; Zhang, L. Point-surface fusion of station measurements and satellite observations for mapping PM<sub>2.5</sub> distribution in China: Methods and assessment. *Atmos. Environ.* **2016**, *152*, 477–489.
66. Ma, Z.; Liu, Y.; Zhao, Q.; Liu, M.; Zhou, Y.; Bi, J. Satellite-derived high resolution PM<sub>2.5</sub> concentrations in Yangtze River Delta Region of China using improved linear mixed effects model. *Atmos. Environ.* **2016**, *133*, 156–164.

67. Tao, J.; Zhang, M.; Chen, L.; Wang, Z.; Su, L.; Ge, C.; Han, X.; Zou, M. A method to estimate concentrations of surface-level particulate matter using satellite-based aerosol optical thickness. *Sci. China Earth Sci.* **2013**, *56*, 1422–1433.
68. Wang, Z.; Chen, L.; Tao, J.; Zhang, Y.; Su, L. Satellite-based estimation of regional particulate matter (PM) in Beijing using vertical-and-RH correcting method. *Remote Sens. Environ.* **2010**, *114*, 50–63.
69. Xiao, L.; Lang, Y.; Christakos, G. High-resolution spatiotemporal mapping of PM<sub>2.5</sub> concentrations at mainland China using a combined BME-GWR technique. *Atmos. Environ.* **2018**, *173*, 295–305.
70. Xiao, Q.; Wang, Y.; Chang, H.H.; Meng, X.; Geng, G.; Lyapustin, A.; Liu, Y. Full-coverage high-resolution daily PM<sub>2.5</sub> estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ.* **2017**, *199*, 437–446.
71. You, W.; Zang, Z.; Zhang, L.; Li, Y.; Pan, X.; Wang, W. National-scale estimates of ground-level PM<sub>2.5</sub> concentration in China using geographically weighted regression based on 3-km resolution MODIS AOD. *Remote Sens.* **2016**, *8*, 184, doi:10.3390/rs8030184.
72. Wei, J.; Huang, W.; Li, Z.; Xue, W.; Peng, Y.; Sun, L.; Cribb, M. Estimating 1-km-resolution PM<sub>2.5</sub> concentrations across China using the space-time random forest approach. *Remote Sens. Environ.* **2019**, *231*, doi:10.1016/j.rse.2019.111221.