Supplementary Information for

Improving air quality assessment using physics-inspired deep graph learning

Lianfa Li *et al.*

*Corresponding author. Email: lilf@lreis.ac.cn; wangjf@lreis.ac.cn

This PDF file includes:

Notes 1 to 4
Algorithms 1 to 2
Figures 1 to 20
Tables 1 to 11
References 1 to 8

Supplementary Notes

Supplementary Note 1: Dataset

We had 56 spatial and temporal input variables including 2 spatial coordinates (latitude and longitude), elevation, day of year, ground aerosol coefficient, Aura Ozone Monitoring Instrument (OMI) NO_2, planetary boundary layer height (PBLH), 4 high-resolution meteorological variables (air temperature, air pressure, relative humidity and wind speed), normalized difference vegetation index (NDVI), an indicator variable of workday (0: not a workday; 1 workday), roadway length within a 10 km buffer around each monitoring site, 26 component pollutant variables and 10 reanalysis vertical meteorological variables (three heights: 2m, 10m, 50m; used to capture vertical variations). We obtained these component pollutant and reanalysis meteorological variables from MERRA2-GMI, which is a simulation for the atmospheric composition community driven by MERRA-2 meteorology and coupled to the Global Modeling Initiative's (GMI) stratosphere-troposphere chemical mechanism run at ~50 km horizontal resolution and output onto the native MERRA-2 grid (0.625° longitude x 0.5° latitude) [1]. The ground aerosol coefficient [2] was obtained through the conversion based on the imputed Multi-Angle Implementation of Atmospheric Correction (MAIAC) aerosol optical depth (AOD), PBLH and relative humidity. MAIAC AOD, OMI-NO_2 and 26 component pollutant variables contain important information about emission, chemical transformation and deposition of air pollutants. The meteorological variables are the driving factors for advection and diffusion of air pollutants. For the full list of variables, please refer to Supplementary Table 1.

Supplementary Note 2: Full residual deep network with the residual term of continuity equation

Based on Eq. 4, we have:

$$e_1 = \frac{1}{N}\ell_{RSE}\left(C, f_{\theta_{\mathbf{W},\mathbf{b}}}(\mathbf{x})\right) \tag{1}$$

$$e_2 \approx e_2' = \frac{1}{M}\sum_{i=1}^{M}\left(\begin{array}{l}\frac{\partial C_i'}{\partial d_i} + v_{l_{x_i}}\frac{\partial C_i'}{\partial l_{x_i}} + v_{l_{y_i}}\frac{\partial C_i'}{\partial l_{y_i}} - \\ p_{l_{x_i}}\left(\frac{\partial^2 C_i'}{\partial^2 l_{x_i}} + \left(\frac{\partial C_i'}{\partial l_x}\right)^2\right) - p_{l_{y_i}}\left(\frac{\partial^2 C_i'}{\partial^2 l_y} + \left(\frac{\partial C_i'}{\partial l_y}\right)^2\right) - \rho C'_i\end{array}\right)^2 \tag{2}$$

$$e_3 = \Omega\left(\theta_{\mathbf{W},\mathbf{b}}\right) \tag{3}$$

where C is the observed concentration of air pollutant, $\mathbf{x}$ is the input vector, N is the number of samples for e_1, M is the number of samples (including N supervised collocated samples) for e_2, d_i is the temporal index of day for sample i, l_{x_i} is the x coordinate for sample i, l_{y_i} is the y coordinate for sample i, $f_{\theta_{\mathbf{W},\mathbf{b}}}(\mathbf{x})$ is the output of the model, ℓ_{RSE} is the mean square error (MSE) loss function, $\theta_{\mathbf{W},\mathbf{b}}$ is the set of parameters (weight matrix and bias) and $\Omega\left(\theta_{\mathbf{W},\mathbf{b}}\right)$ is the

normalization of $\theta_{\mathbf{w},\mathbf{b}}$. Since the dependent variable (C) is log-transformed into $C' = \log(C)$ and so the continuity equation (Eq. 1) changes correspondingly and thus e_2 can be approximated using e_2'. We minimize the loss function, so e_2 will be close to 0 and we can make the predictions to maintain the continuity equation as much as possible.

Here, the time variable of day (d) and the horizontal directions (x and y) were used in the residual term, e_2. The extraction of e_2 is shown in Algorithm 2.

Supplementary Note 3: Neighborhood heatmap and SHAP values

The output from the final graph convolution layer represents the spatial agglomeration of neighborhoods in the study area, and the linear activation was used for the output of the last graph convolution. For a daily grid of 1x1 km^2 over the study area, the outputs of the final graph convolution layers for all the pixels that represent the spread feature of air pollution at the longest spatial scale were merged into a heatmap, which provides a visualization of the contours of the neighborhood.

As a game-theoretic approach, SHAP (Shapley Additive exPlanations) [3] measures the contribution of each feature to the prediction taking into account the interaction with other variables. It is calculated by comparing what a model predicts across all possible combinations of the other covariates with and without the particular feature under analysis [4]. With the national mean concentration of an air pollutant as the background value and 1000 selected observations as background samples, the SHAP value was estimated for each covariate of every prediction. Then the statistics of SHAP values were reported to evaluate spatiotemporal distribution of influential factors. We estimated the contribution of each covariate to $PM_{2.5}$, especially those related to emissions.

Supplementary Note 4: AQI output

The air quality sub-index for each pollutant was calculated [5]:

$$\mathrm{IAQI_p} = \frac{\mathrm{IAQI_{H_i}} - \mathrm{IAQI_{L_o}}}{\mathrm{BP_{H_i}} - \mathrm{BP_{L_o}}}\left(C_p - BP_{\mathrm{L_o}}\right) + IAQI_{L_o} \tag{4}$$

where $\mathrm{IAQI_p}$ is the sub-index of air quality index for the air pollutant, p, C_p is the measured or estimated concentration for, p, $\mathrm{BP_{L_o}}$ and $\mathrm{BP_{H_i}}$ are respectively the low and high values closest to the limit of C_p, $\mathrm{IAQI_{H_i}}$ is the sub-index of AQI corresponding to $\mathrm{BP_{H_i}}$, and $IAQI_{L_o}$ is the sub-index of AQI corresponding to $\mathrm{BP_{L_o}}$.

Then, the AQI was calculated:

$$\mathrm{AQI} = \max\left\{\mathrm{IAQI_1}, \mathrm{IAQI_2}, \mathrm{IAQI_3}, .., \mathrm{IAQI_n}\right\} \tag{5}$$

where IAQI is a sub-index of AQI for each air pollutant, n is the number of air pollutants. The primary air pollutant is defined as having the maximum sub-index of AQI.

In addition, we calculated the national averages of air pollutant concentrations and AQI weighted by the 1x1 km^2 population density. The population-weighted concentration or AQI was defined as:

$$\hat{M}_p = \frac{\sum_{i=1}^{N} \tilde{p}_i \tilde{c}_i}{\sum_{i=1}^{N} \tilde{p}_i} \tag{6}$$

where $\hat{M}_p$ is the population-weighted air pollutant concentration, AQI sub-index or AQI, i is the cell index, $\tilde{p}_i$ is the estimated population density for cell i, and $\tilde{c}_i$ is the estimated pollutant concentration, AQI sub-index or AQI for cell i. We got the population density data from https://www.worldpop.org.

Supplementary Algorithms

Supplementary Algorithm 1: MGraphAir Minibatch Forward Algorithm

Input : Set of minibatch sample indices: $\mathcal{B}$;
Input features: $\{\mathbf{x}_b, \forall b \in \mathcal{B}\}$;
Depth for convolutions: $\mathcal{K}$;
Thresholds for diffusion, velocity and ratio ;
Weight matrices of reciprocal distances: $\mathbf{W}_d^k, \forall k \in \{1, \cdots, \mathcal{K}\}$

Output : Multilevel convolution outputs: $\mathcal{O}$

Function : k-NN nearest function: $\mathcal{N}_k, \forall k \in \{1, \cdots, \mathcal{K}\}$

Parameter: Diffusion coefficient: $\mathbf{p}^k, \forall k \in \{1, \cdots, \mathcal{K}\}$;
Velocity: $\mathbf{v}^k, \forall k \in \{1, \cdots, \mathcal{K}\}$;
Ratio for emissions, deposition etc.: $\rho^k, \forall k \in \{1, \cdots, \mathcal{K}\}$;
Weight matrices for total summary: $\mathbf{W}_c^k, \forall k \in \{1, \cdots, \mathcal{K}\}$;
Weight matrices for neighbor aggregation: $\mathbf{W}_u^k, \forall k \in \{1, \cdots, \mathcal{K}\}$;
Weight matrices of reciprocal distances: $\mathbf{W}_d^k, \forall k \in \{1, \cdots, \mathcal{K}\}$

$\mathcal{B}^{\mathcal{K}} \leftarrow \mathcal{B}$;
for $k = \mathcal{K} \cdots 1$ **do**
  $\mathcal{B}^{k-1} \leftarrow \mathcal{B}^k$;
  for $i \in \mathcal{B}^k$ **do**
    $\mathcal{B}^{k-1} \leftarrow \mathcal{B}^{k-1} \cup \mathcal{N}_k(i)$;
  end for
end for
$\mathbf{h}_b^0 \leftarrow \mathbf{x}_b, \forall b \in \mathcal{B}^0$;
Initialize learnable parameters: $\mathbf{p}^k, \mathbf{v}^k, \rho^k, \mathbf{W}_c^k, \mathbf{W}_u^k, \mathbf{W}_d^k$;
Set thresholds for diffusion, velocity and ratio ;
for $k = 1 \cdots \mathcal{K}$ **do**
  for $i \in \mathcal{B}^k$ **do**
    $\mathbf{h}_{\mathcal{N}(i)}^k \leftarrow \mathbf{p}^k \cdot \sum_{j \in \mathcal{N}_k(i)} (\mathbf{h}_j^{k-1} - \mathbf{h}_i^{k-1}) - \mathbf{v}^k \cdot \sum_{j \in \mathcal{N}_k(i)} w_{d_{ij}}^{k-1} (\mathbf{h}_j^{k-1} - \mathbf{h}_i^{k-1})$;
    $\mathbf{h}_{\mathcal{N}(i)}^k \leftarrow \mathbf{h}_{\mathcal{N}(i)}^k + \rho^k \cdot \mathbf{h}_i^{k-1}$;
    $\mathbf{h}_i^k \leftarrow \sigma(\mathbf{W}_c^k \cdot \mathbf{h}_i^{k-1} - \mathbf{W}_u^k \cdot \mathbf{h}_{\mathcal{N}(i)}^k)$;
    $\mathbf{h}_i^k \leftarrow \mathbf{BN}(\mathbf{h}_i^k)$;
  end for
  if $k = 1$ **then**
    $\mathcal{O} = \mathbf{h}_i^k, \forall i \in \mathcal{B}^{\mathcal{K}}$;
  else
    $\mathcal{O} \leftarrow \mathcal{O} \cup \mathbf{h}_i^k, \forall i \in \mathcal{B}^{\mathcal{K}}$;
  end if
end for
Return $\mathcal{O}$

Supplementary Algorithm 2: Continuity Equation Residual Forward Algorithm

Input : Module of main MGraph: $\mathcal{M}$;
Input features: $\{\mathbf{x}\}$;
Adjacent topology matrix: $\mathcal{N}_{\mathbf{x}}$;
List of the node numbers for all layers: $\mathcal{L}$;
List of decoding layers: $\mathcal{L}_d$;
Time column index (day): d ;
Thresholds for velocity, diffusion and ratio

Output : Prediction and residual outputs: $\mathcal{O}$

Function : Partial derivative by AD: $grad$

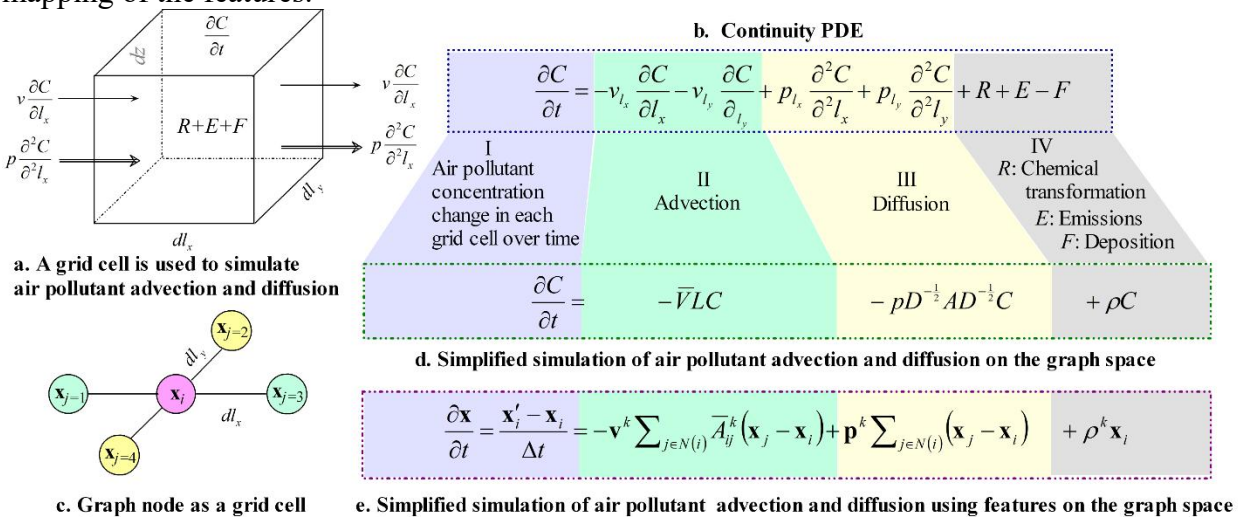
Parameter: Hidden layers: $\mathcal{H}_l, \forall l \in \{1, \cdots, \mathcal{L}+2\}$;
Velocity: v_{l_x}, v_{l_y}; Diffusion: p_{l_x}, p_{l_y}; Ratio: ρ

Initialize the parameters ;
Set the thresholds for $v_{l_x}, v_{l_y}, p_{l_x}, p_{l_y}, \rho$;
$\tilde{\mathbf{c}} \leftarrow \mathcal{M}(\mathbf{x}, \mathcal{N}_{\mathbf{x}})$;
$\partial\tilde{\mathbf{c}}/\partial l_x \leftarrow grad(\tilde{\mathbf{c}}, l_x)$;
$\partial^2\tilde{\mathbf{c}}/\partial^2 l_x \leftarrow grad(\partial\tilde{\mathbf{c}}/\partial l_x, l_x)$;
$\partial\tilde{\mathbf{c}}/\partial l_y \leftarrow grad(\tilde{\mathbf{c}}, l_y)$;
$\partial^2\tilde{\mathbf{c}}/\partial^2 l_y \leftarrow grad(\partial\tilde{\mathbf{c}}/\partial l_y, l_y)$;
$\partial\tilde{\mathbf{c}}/\partial d \leftarrow grad(\tilde{\mathbf{c}}, d)$;
$S.push(\mathbf{x})$;
$\mathbf{x} \leftarrow \mathbf{BN}(\sigma(\mathcal{H}_1(\mathbf{x})))$;
for $i = 2 \cdots \text{len}(\mathcal{L})$ **do**
  if $i \leq (\mathcal{L}_d - 1)$ **then**
    $S.push(\mathbf{x})$;
  $\mathbf{x} \leftarrow \mathbf{BN}(\sigma(\mathcal{H}_i(\mathbf{x})))$;
  if $i \geq (\mathcal{L}_d)$ **then**
    $\mathbf{x} \leftarrow \mathbf{x} + S.pop()$;
end for
$\mathbf{x} \leftarrow \mathbf{BN}(\sigma(\mathcal{H}_{\mathcal{L}+1}(\mathbf{x})))$;
$\mathbf{x} \leftarrow \mathbf{x} + S.pop()$;
$\zeta \leftarrow \mathcal{H}_{\mathcal{L}+2}(\mathbf{x})$;
$v_{l_x} \leftarrow \zeta[1], v_{l_y} \leftarrow \zeta[2], p_{l_x} \leftarrow \zeta[3], p_{l_y} \leftarrow \zeta[4], \rho \leftarrow \zeta[5]$;
$\mathcal{R} = \frac{\partial\tilde{\mathbf{c}}}{\partial d} + \frac{\partial\tilde{\mathbf{c}}}{\partial l_x} v_{l_x} + \frac{\partial\tilde{\mathbf{c}}}{\partial l_y} v_{l_y} - \rho\tilde{\mathbf{c}}$;
$\mathcal{R} = \mathcal{R} - (\frac{\partial^2\tilde{\mathbf{c}}}{\partial^2 l_x} + (\frac{\partial\tilde{\mathbf{c}}}{\partial l_x})^2)p_{l_x} - (\frac{\partial^2\tilde{\mathbf{c}}}{\partial^2 l_y} + (\frac{\partial\tilde{\mathbf{c}}}{\partial l_y})^2)p_{l_y}$;
$\mathcal{O} \leftarrow (\tilde{\mathbf{c}}, \mathcal{R})$;
Return $\mathcal{O}$

Supplementary Figures

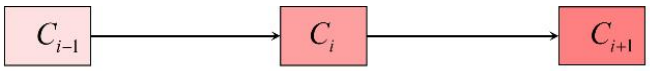
Supplementary Figure 1

Mapping of advection, diffusion and the others (chemical transformation, emissions and deposition). a, using the continuity PDE in a grid cell to simulate. b, c and d, the graph space. e, mapping of the features.



Supplementary Figure 2

Traffic simulation of air pollutants in the discrete graph space. a, diffusion was simulated using PDE based on thermodynamic laws. b, the graph convolution process was used to simulate the horizontal advection and diffusion, and the input contains the meteorological driving force for the pollutant movement and other covariates to account for the other changes. c, messaging and aggregating from neighbors in the local graph space were used to represent local multiscale spread of air pollutants.

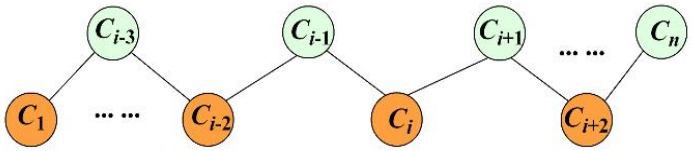


$$\frac{\partial C_i}{\partial t} = -\bar{V}\cdot\frac{(C_{i+1}-C_{i-1})}{2} + p\left((C_{i+1}-C_i)-(C_i-C_{i-1})\right) = -\bar{V}\cdot\frac{(C_{i+1}-C_{i-1})}{2} + p\sum_j (C_j - C_i) \qquad p=k_x+m$$

$$\Longrightarrow \quad \frac{\partial C_x}{\partial t} = -\bar{V}_x\frac{\partial C_x}{\partial x} + p\frac{\partial^2 C_x}{\partial^2 x}$$

p: Diffusion coefficient that is the sum of eddy diffusion coefficient (k_x) and molecular diffusion coefficient (m)

a. Simplified simulation of the diffuson of air pollutant in the *x* direction



Simulation of second-order derivatives: Use Laplacian matrix

$$-p\sum_j A_{ij}(C_i - C_j) = -p\left(D(C_i)C_i - \sum_j A_{ij}C_j\right) \quad \Rightarrow \quad p\frac{\partial^2 C}{\partial^2 x} = -p(D-A)C = -pLC$$

A: Ajacency matrix; D: Degree matrix; L: Laplacian matrix (L=D-A)

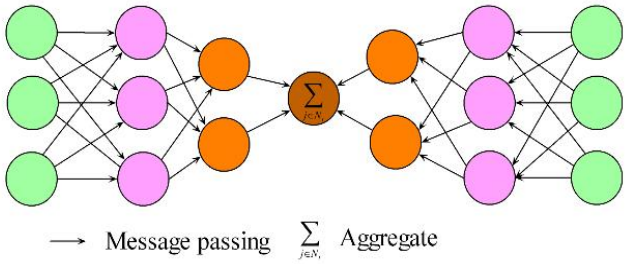
Corresponding Laplacian convolutional operator: $D^{-1/2}AD^{-1/2}H^{(l)}W^{(l)}$

Simulation of first-order derivatives: Decompose the differential equation along the linked directions (nodes)

$$-\frac{\partial \bar{V}C_i}{\partial x} = -\sum_j \bar{V}_{ij}(C_j - C_i) \quad \Rightarrow \quad -\frac{\partial \bar{V}C}{\partial x} = -\bar{V}LC$$

$\bar{A}$: Learnable speed matrix based on the Ajacency matrix A

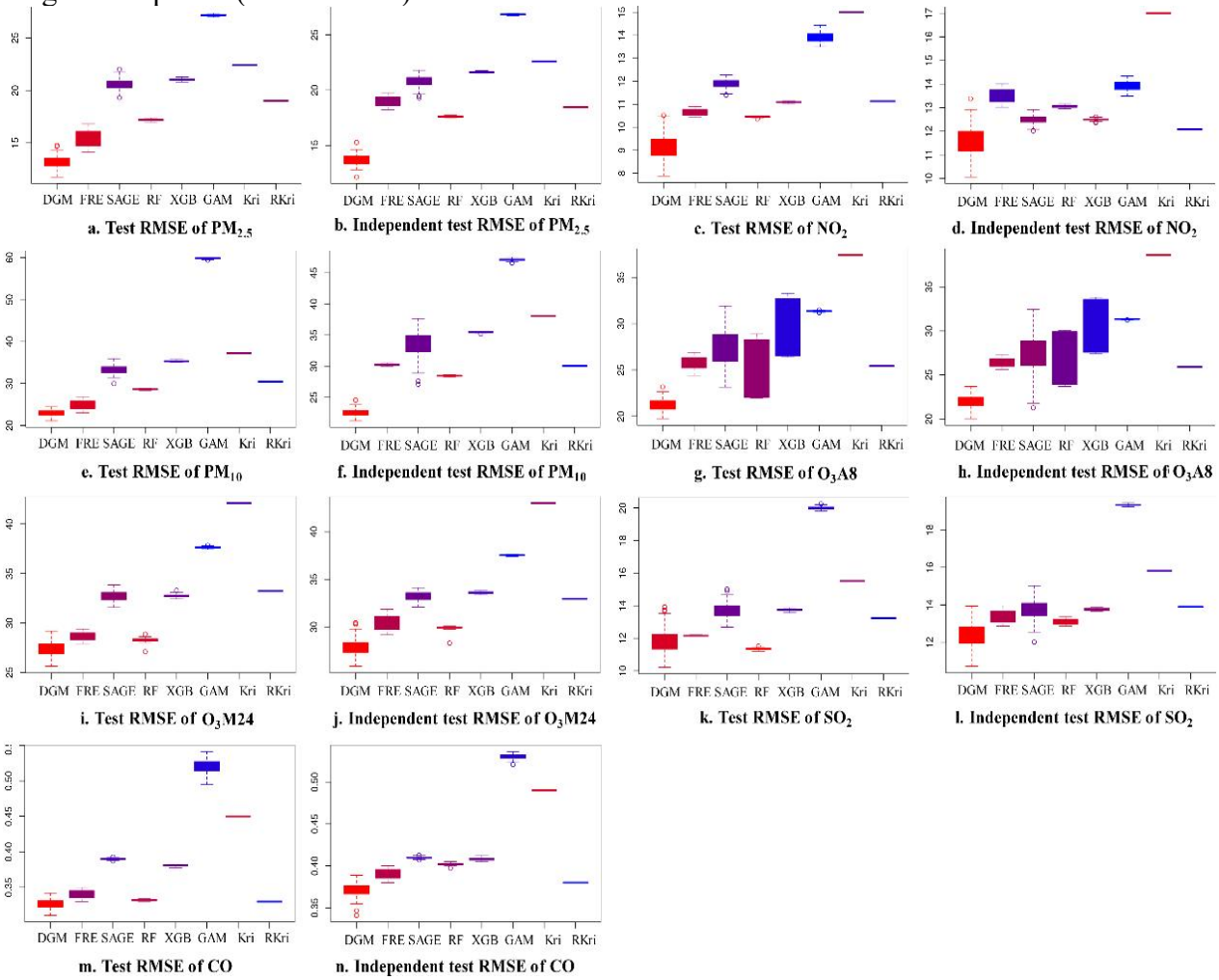
b. Simulation of air pollutant convection on the graph space



c. Multiscale message passing and aggregating on the local graph space

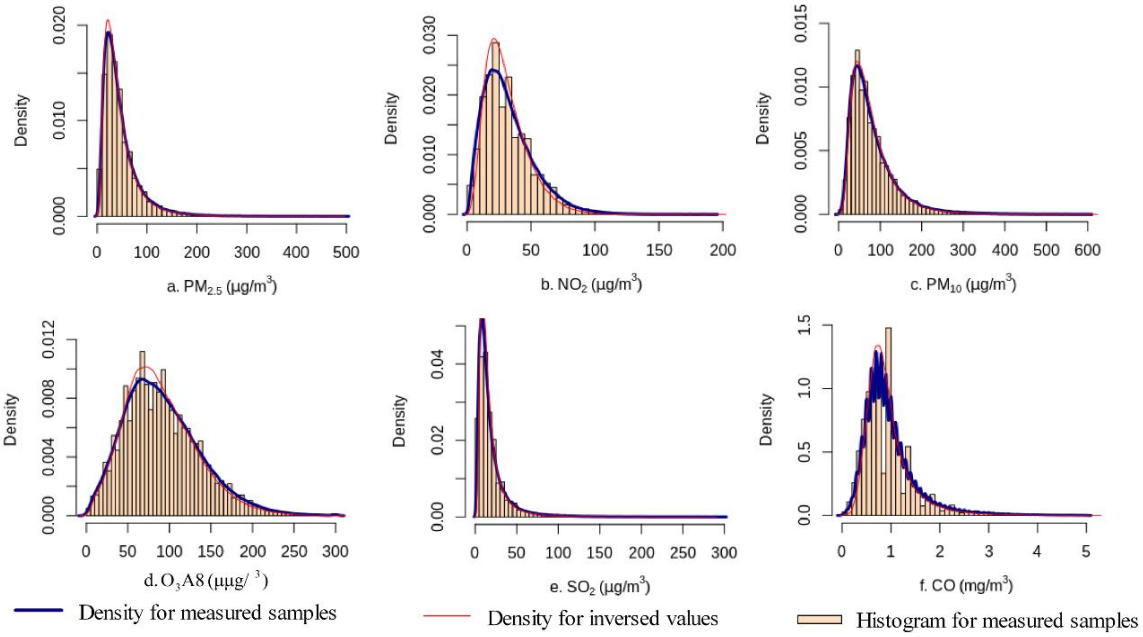
Supplementary Figure 3

Site-based test RMSE boxplot of seven air pollutants by eight methods. a, c, e, g, i, k and m show testing RMSE and b, d, f, h, j, l and n show site-based independent testing RMSE. Method: Deep graph modeling (DGM) proposed; Full residual encoder-decoder (FRE); GraphSAGE (SAGE); Random forest (RF); XGBoost (XGB); Kriging (Kri); Regression kriging (RKri); Generalized additive model (GAM). The whiskers (solid horizontal lines) represent the minimum and maximum RMSE, excluding outliers. The box bounds the interquartile range (IQR) from the 25th to 75th percentiles, with the 50th percentile (median) as the bold line inside. The whiskers (dashed vertical lines) extend 1.5 times the IQR from the box, indicating the normal RMSE range. Data points (small circles) outside the whiskers are RMSE outliers.



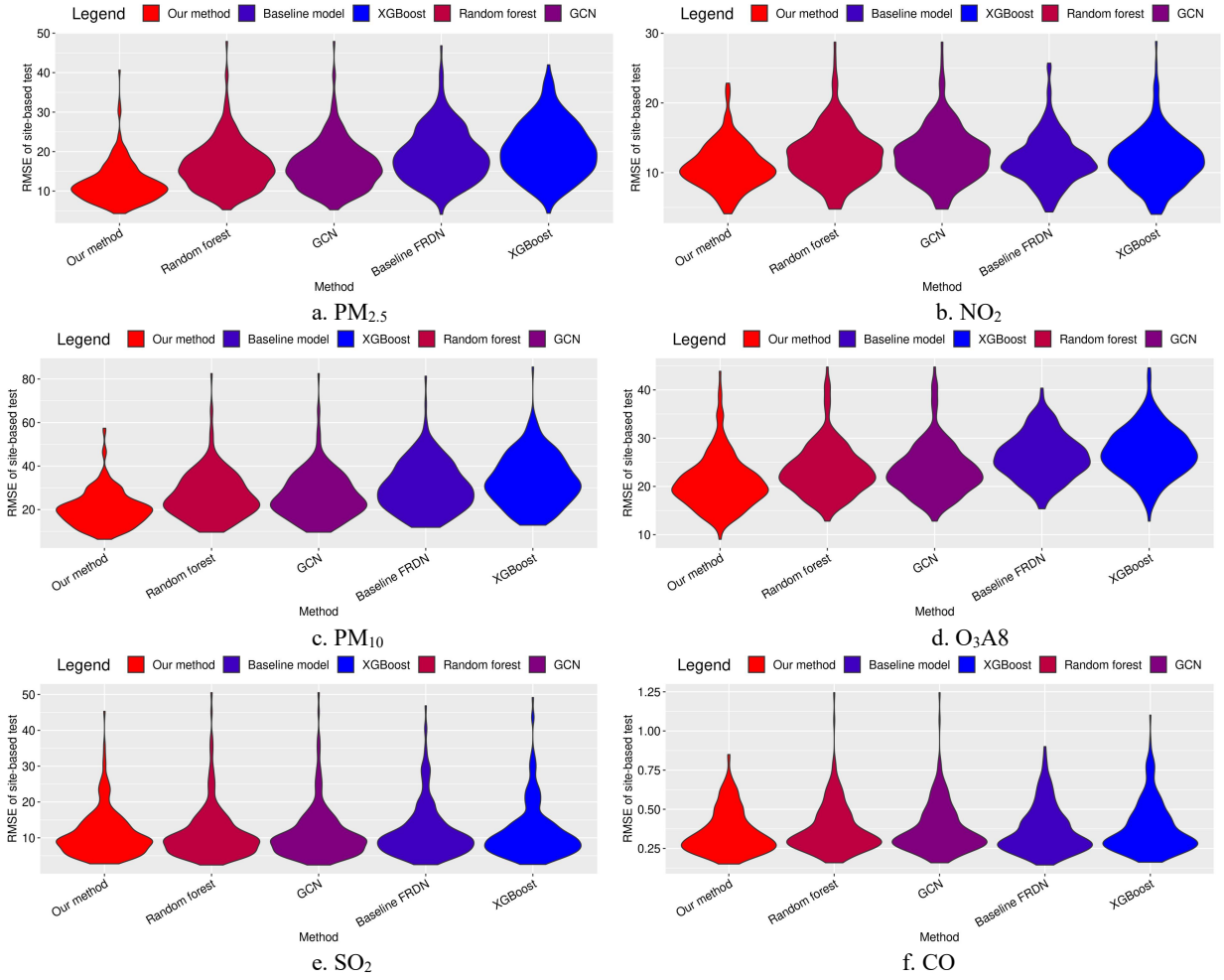
Supplementary Figure 4

Histogram and density plots for the measured values vs. inversed samples in the site-based independent tests of six air pollutants.



Supplementary Figure 5

Comparison of RMSE violin plots between our DGM method and various machine learning approaches, illustrating the predicted versus observed time series for all sites in the site-based independent test.



a. $PM_{2.5}$

b. NO_2

c. PM_{10}

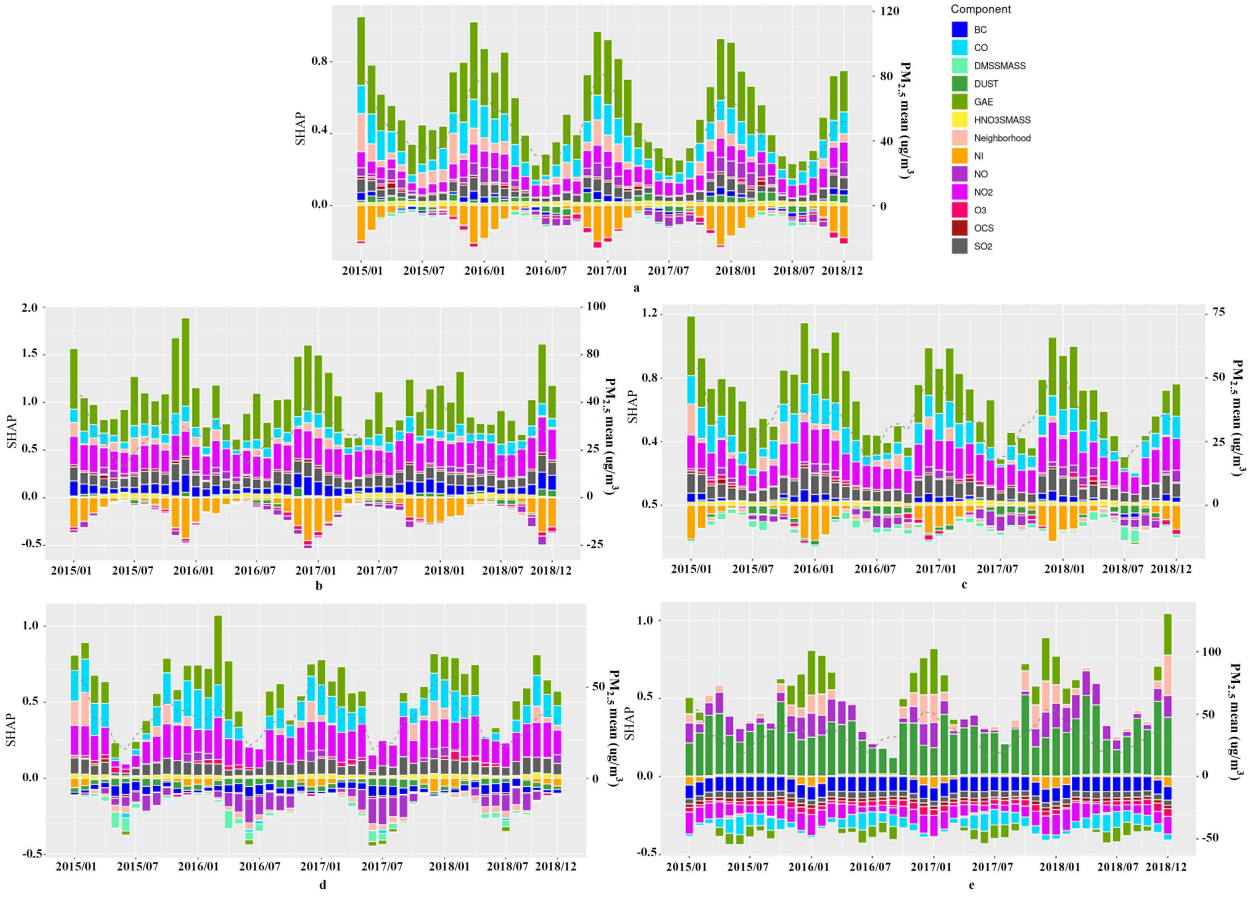
d. O_3A8

e. SO_2

f. CO

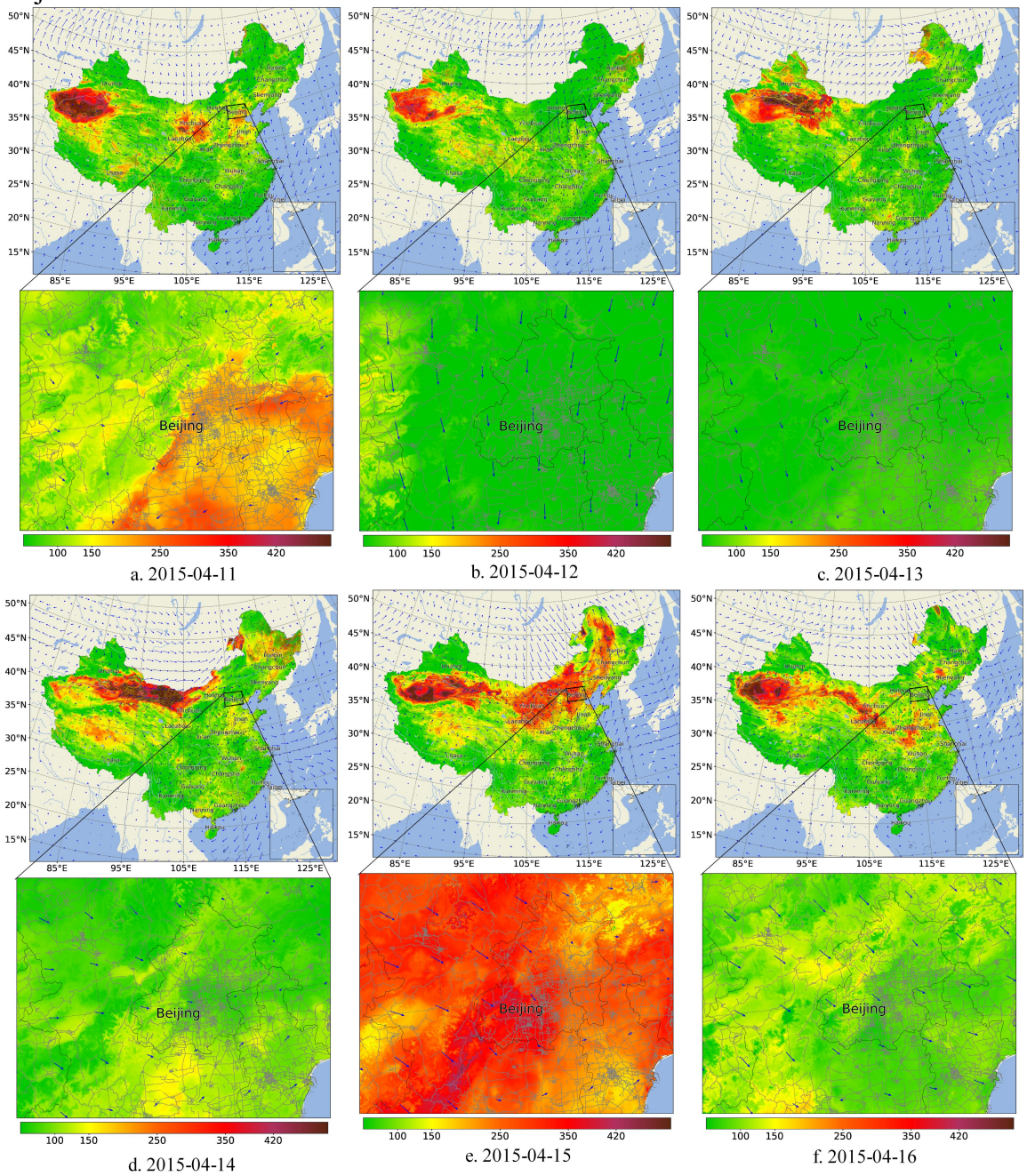
Supplementary Figure 6

Contributions (SHAP) of the emission-related components to the $PM_{2.5}$ predictions for four representative regions of China. a, China Mainland. b, the Jing-Jin-Ji metropolitan region. c, Yangtze River Delta. d, Pearl River Delta. e, Xinjiang Zhuang Autonomous Region. BC: Black carbon mass mixing ratio; DMSSMASS: DMS surface mass concentration; DUST: Dust surface mass concentration; GAE: Ground aerosol extinction coefficient; HNO3SMASS: Nitric acid surface mass concentration; Neighborhood: Neighborhood feature extracted using GC; NI: Nitrate Mass Mixing Ratio; NO: Nitric oxide; NO_2: Nitrogen dioxide; O_3: Ozone; OCS: Organic Carbon Surface Mass Concentration; SO_2: SO_2 Surface Mass Concentration.

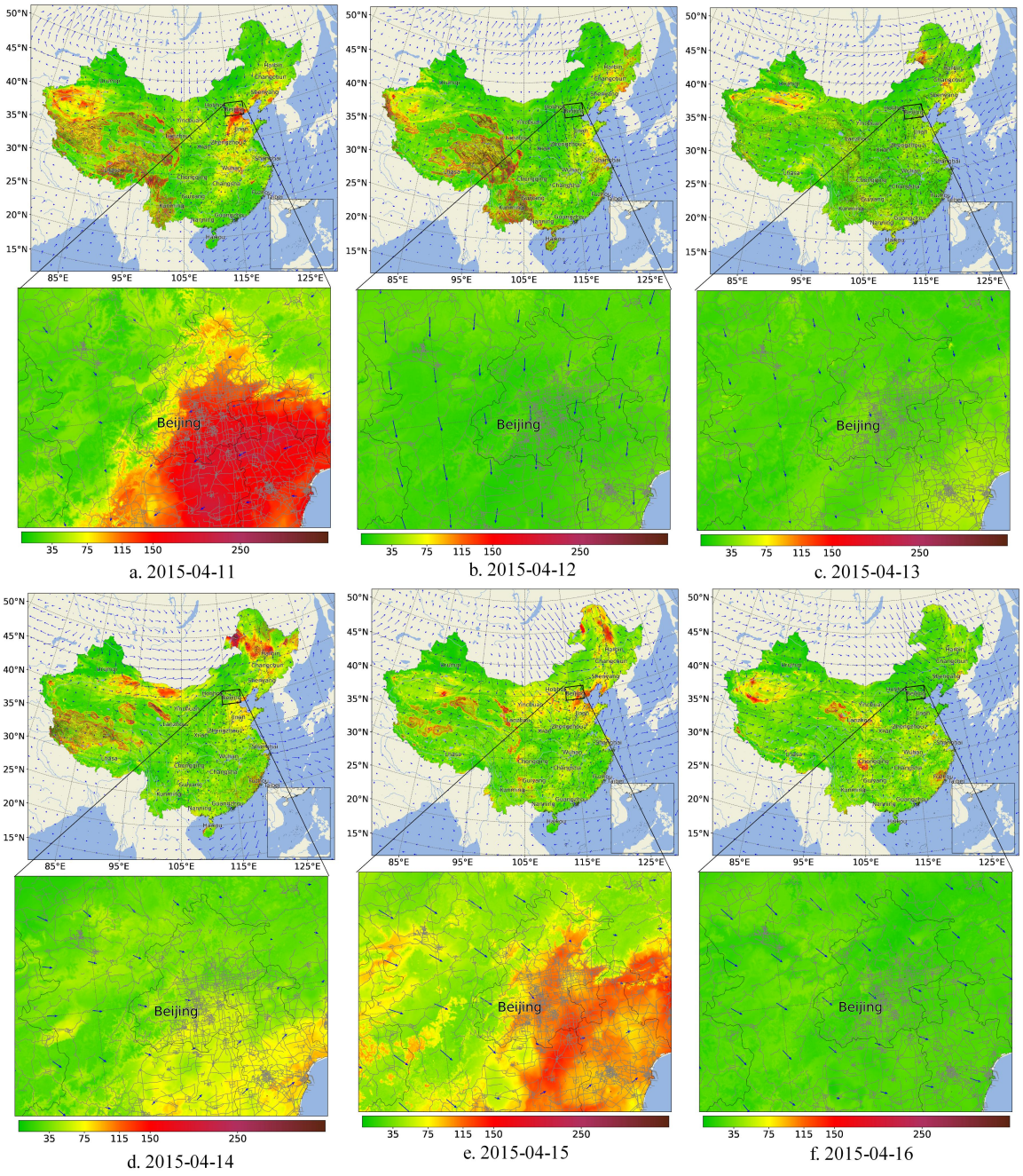


Supplementary Figure 7

Spatiotemporal evolution of PM_{10} for the sandstorm event in North China of April 15, 2015. a-f, national (the upper part) and local (the lower part of the Jingjinji metropolitan region) spatial distributions of estimated PM_{10} concentrations from April 11 to 16, 2015. The blue arrow is mean ground wind vector (m/s), the color gradient lines in each upper image show the contours of the graphical convolution heatmap, and the gray lines in each below enlarged image show major traffic roads.



Supplementary Figure 8
Spatiotemporal evolution of $PM_{2.5}$ for the sandstorm event in North China of April 15, 2015. a-f, national (the upper part) and local (the lower part of the Jingjinji metropolitan region) spatial distributions of estimated $PM_{2.5}$ concentrations from April 11 to 16, 2015. The blue arrow is mean ground wind vector (m/s), the color gradient lines in each upper image show the contours of the graphical convolution heatmap, and the gray lines in each below enlarged image show major traffic roads.

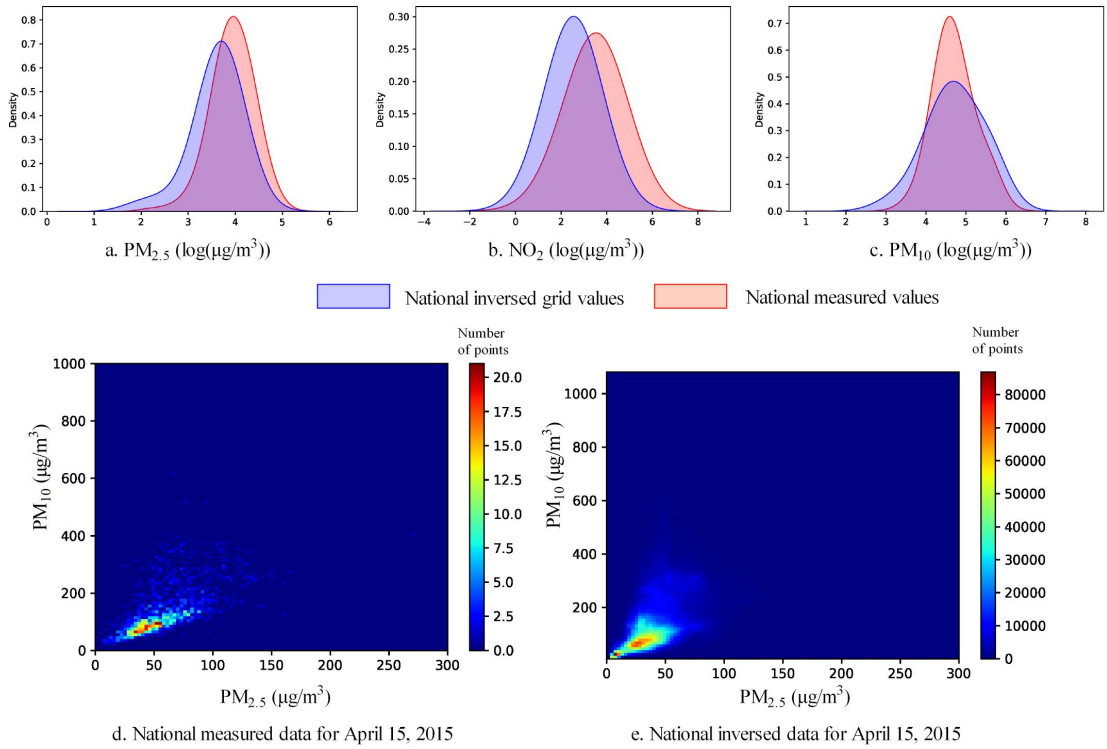


a. 2015-04-11 b. 2015-04-12 c. 2015-04-13

d. 2015-04-14 e. 2015-04-15 f. 2015-04-16

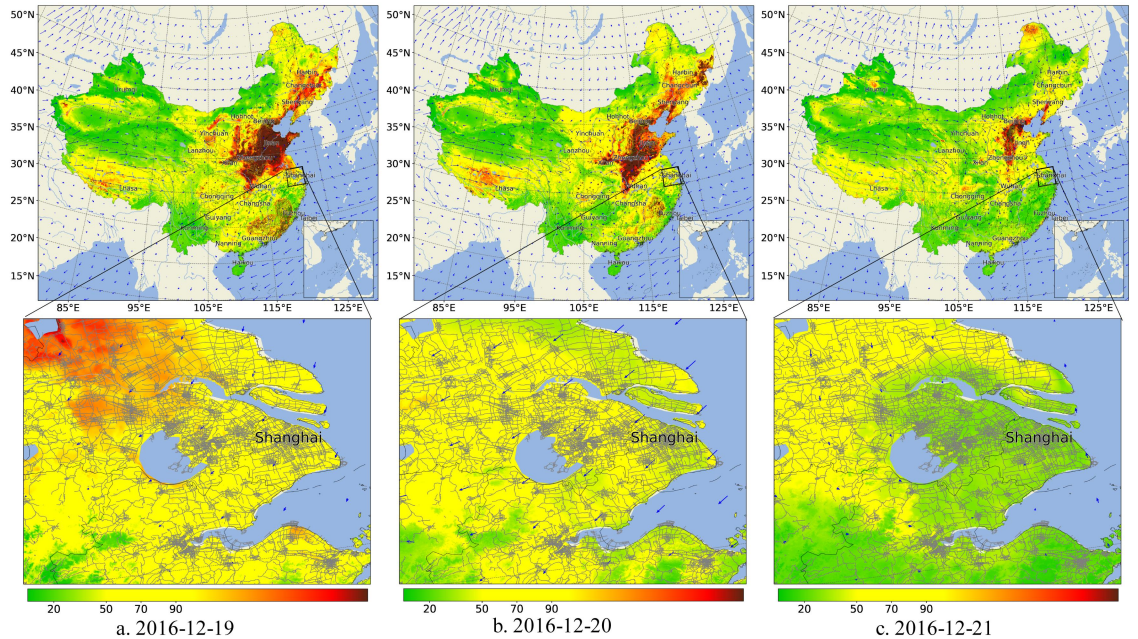
Supplementary Figure 9

Comparison between estimated and measured PM_{10}, $PM_{2.5}$ and NO_2 for mainland China on April 15, 2015. a, b and c, log-transformed probability density of $PM_{2.5}$ (a), NO_2 (b) and PM_{10} (c). d and e, 2-d histograms of $PM_{2.5}$ and PM_{10} for national measured data (d) and national inversed data (e).



Supplementary Figure 10

Spatiotemporal evolution of $PM_{2.5}$ for the haze event in East China of December 23, 2016. a-f, national (the upper part) and local (the lower part of the Yangtze River Delta) spatial distributions of estimated $PM_{2.5}$ concentrations from December 19 to 24, 2016. The blue arrow is mean ground wind vector (m/s), the color gradient lines in each upper image show the contours of the graphical convolution heatmap, and the gray lines in each below enlarged image show major traffic roads.

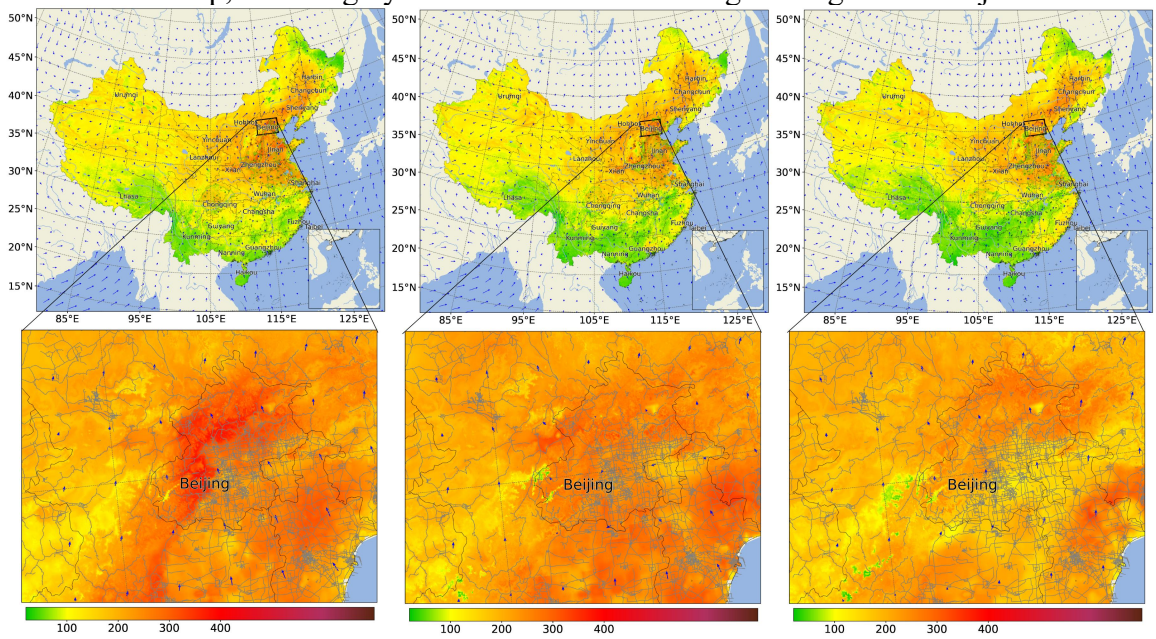


a. 2016-12-19 b. 2016-12-20 c. 2016-12-21

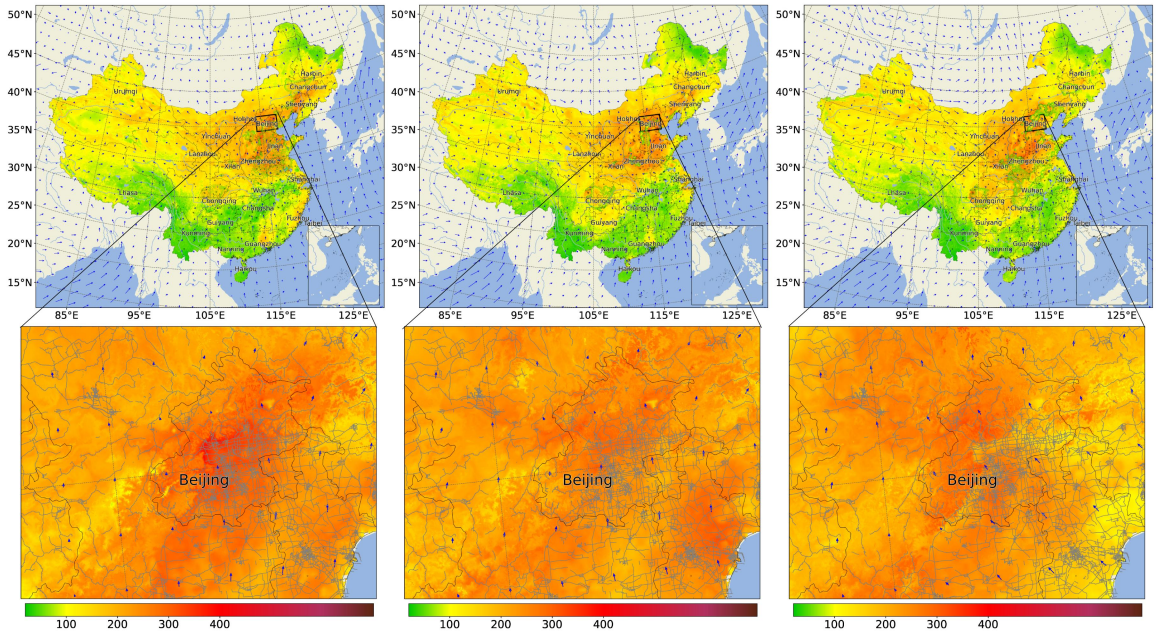
d. 2016-12-22 e. 2016-12-23 f. 2016-12-24

Supplementary Figure 11

Spatiotemporal evolution of O_3 for the ozone event of July 1, 2017. a-f, national (the upper part) and local (the lower part of Beijing) spatial distributions of the estimated 8-hour O_3 concentration average from June 27 to July 2, 2017. The blue arrow is mean ground wind vector (m/s), the color gradient lines in each upper image show the contours of the graphical convolution heatmap, and the gray lines in each below enlarged image show major traffic roads.



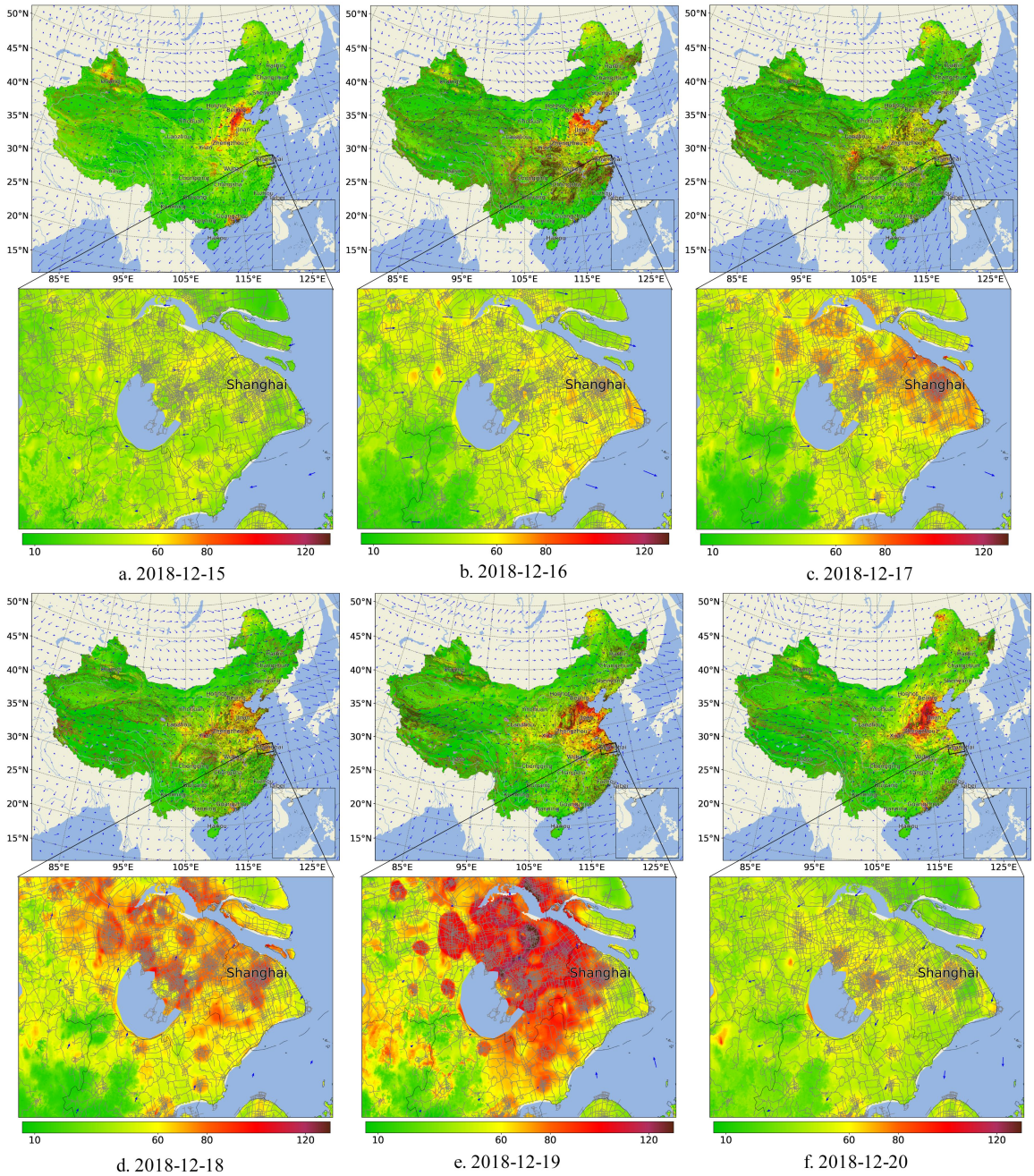
a. 2017-06-27 b. 2017-06-28 c. 2017-06-29



d. 2017-06-30 e. 2017-07-01 f. 2017-07-02

Supplementary Figure 12

Spatiotemporal evolution of NO_2 for the haze event in Shanghai of December 19, 2018. a-f, national (the upper part) and local (the lower part of the Shanghai) spatial distributions of estimated NO_2 concentrations from December 15 to 20, 2018. The blue arrow is mean ground wind vector (m/s), the color gradient lines in each upper image show the contours of the graphical convolution heatmap, and the gray lines in each below enlarged image show major traffic roads.

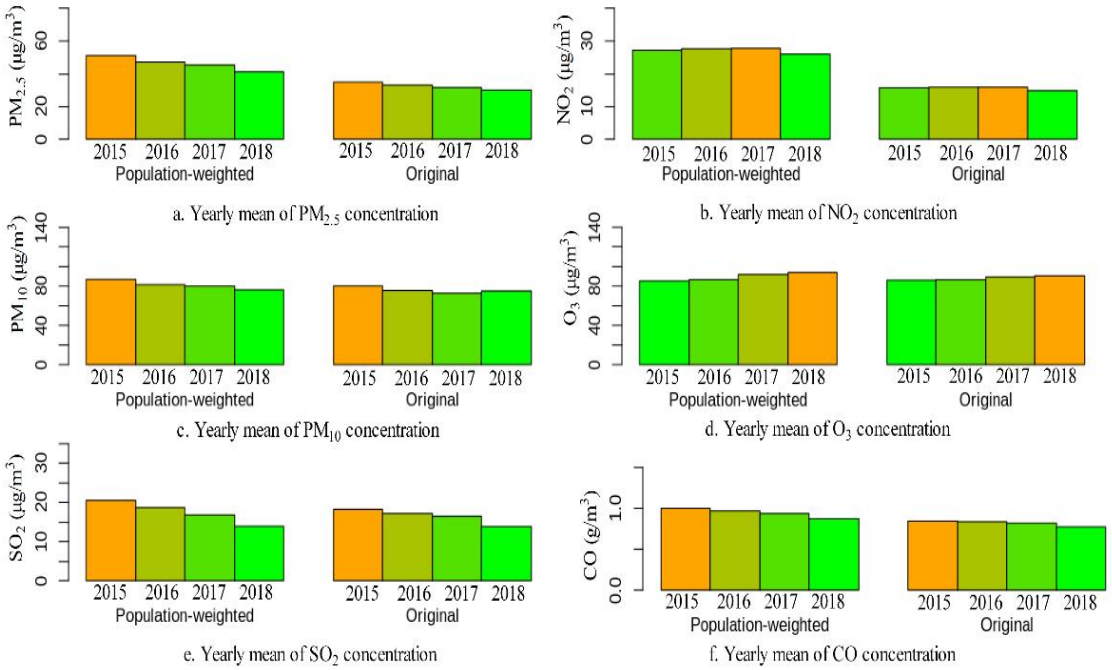


a. 2018-12-15 b. 2018-12-16 c. 2018-12-17

d. 2018-12-18 e. 2018-12-19 f. 2018-12-20

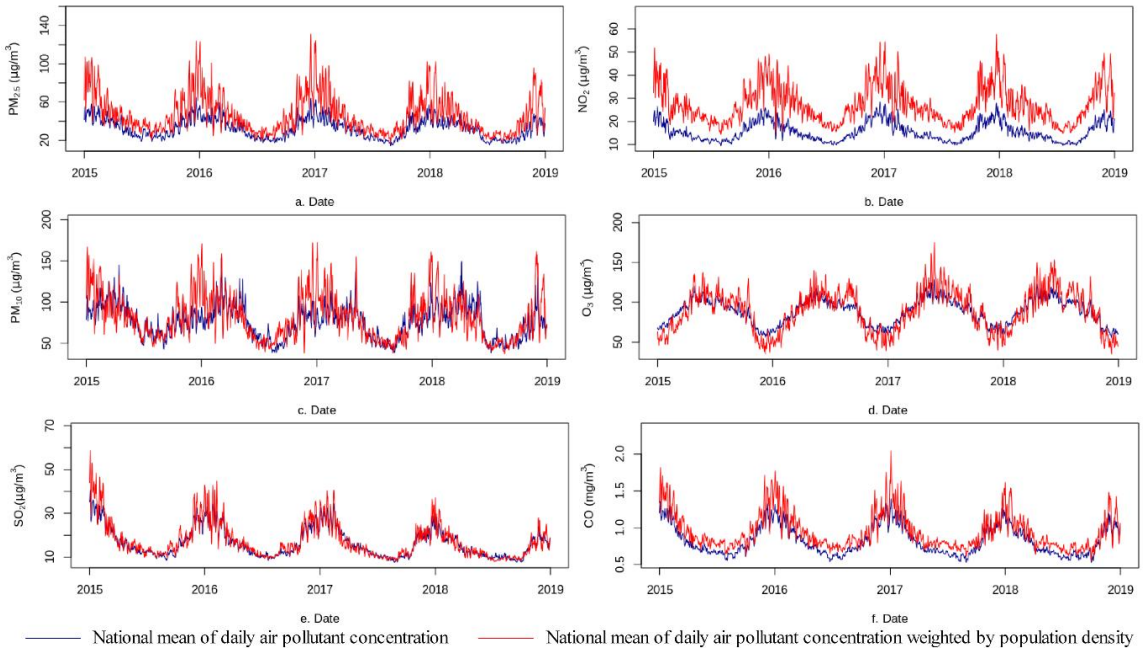
Supplementary Figure 13

Yearly means of air pollutant concentrations.



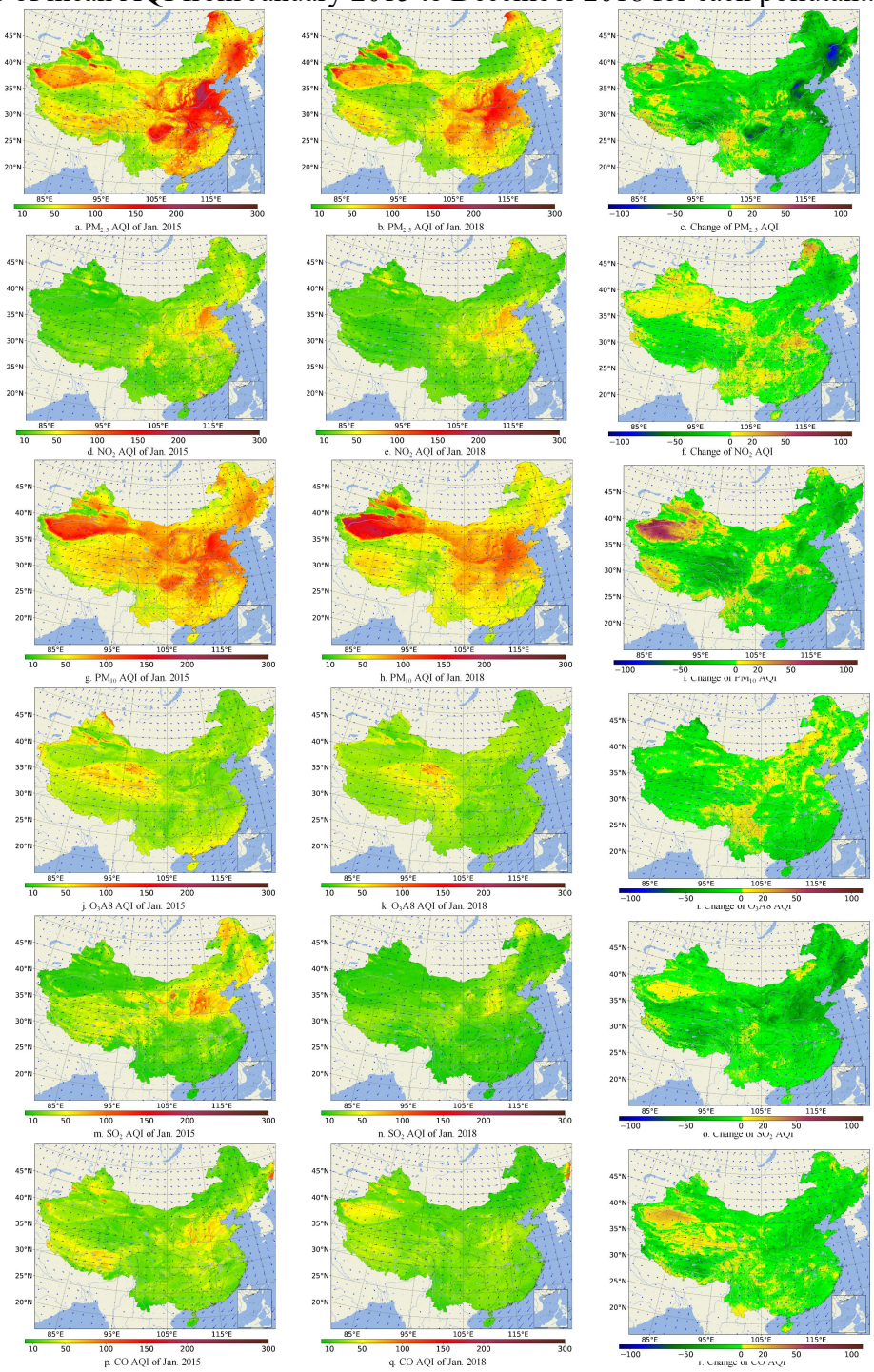
Supplementary Figure 14

Daily national estimated averages and daily national population-weighted averages of predicted concentrations across mainland China for air pollutants.



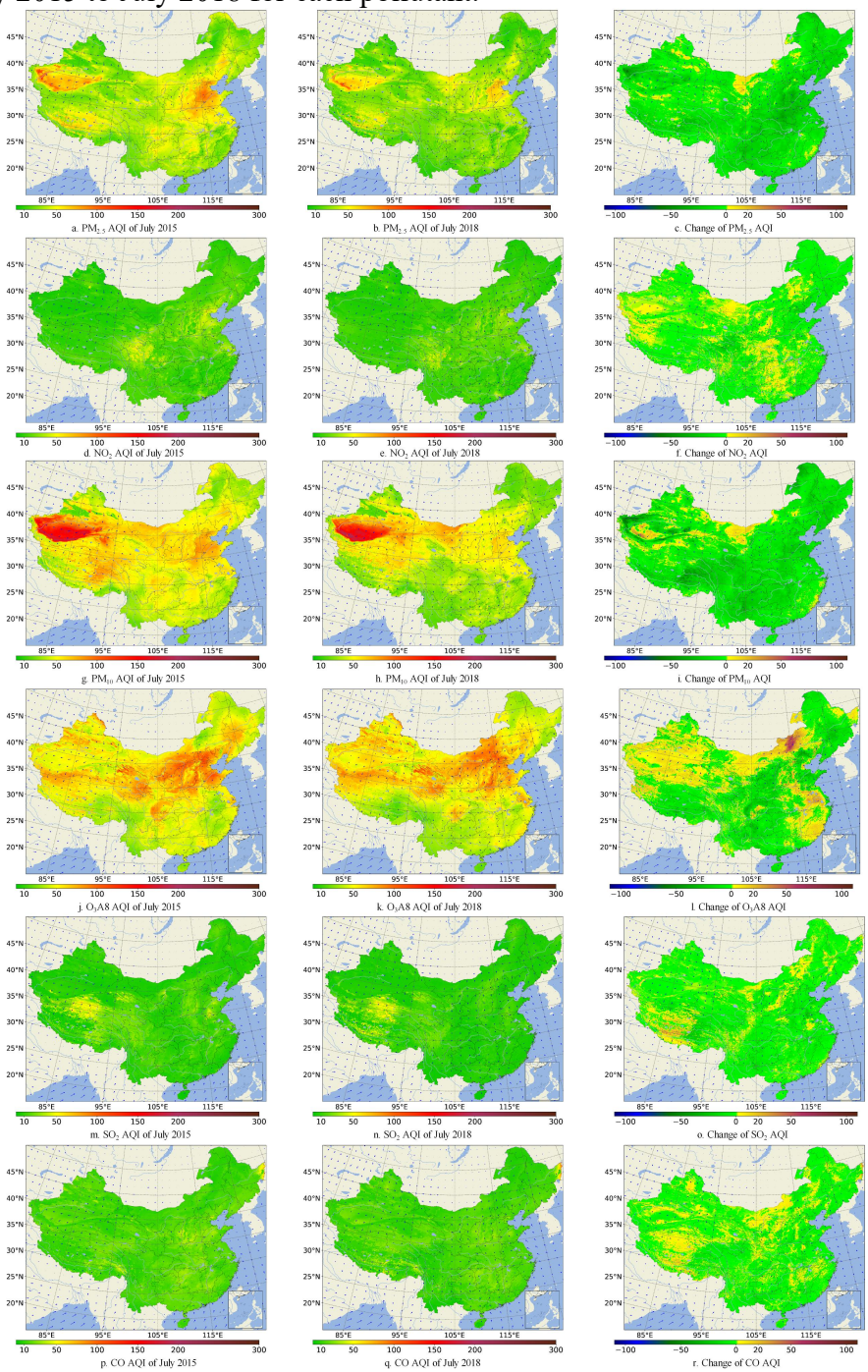
Supplementary Figure 15

Monthly mean AQI and its change for each air pollutant from January 2015 to December 2018 in mainland China. a, b and c: $PM_{2.5}$. d, e and f: NO_2. g, h and i: PM_{10}. j, k and l: O_3A8, m, n and o: SO_2. p, q and r: CO. a, d, g, j, m and p, monthly mean AQI of January 2015 for each pollutant. b, e, h, k, n and q, monthly mean AQI of December 2018 for each pollutant. c, f, i, l, o and r, change of mean AQI from January 2015 to December 2018 for each pollutant.



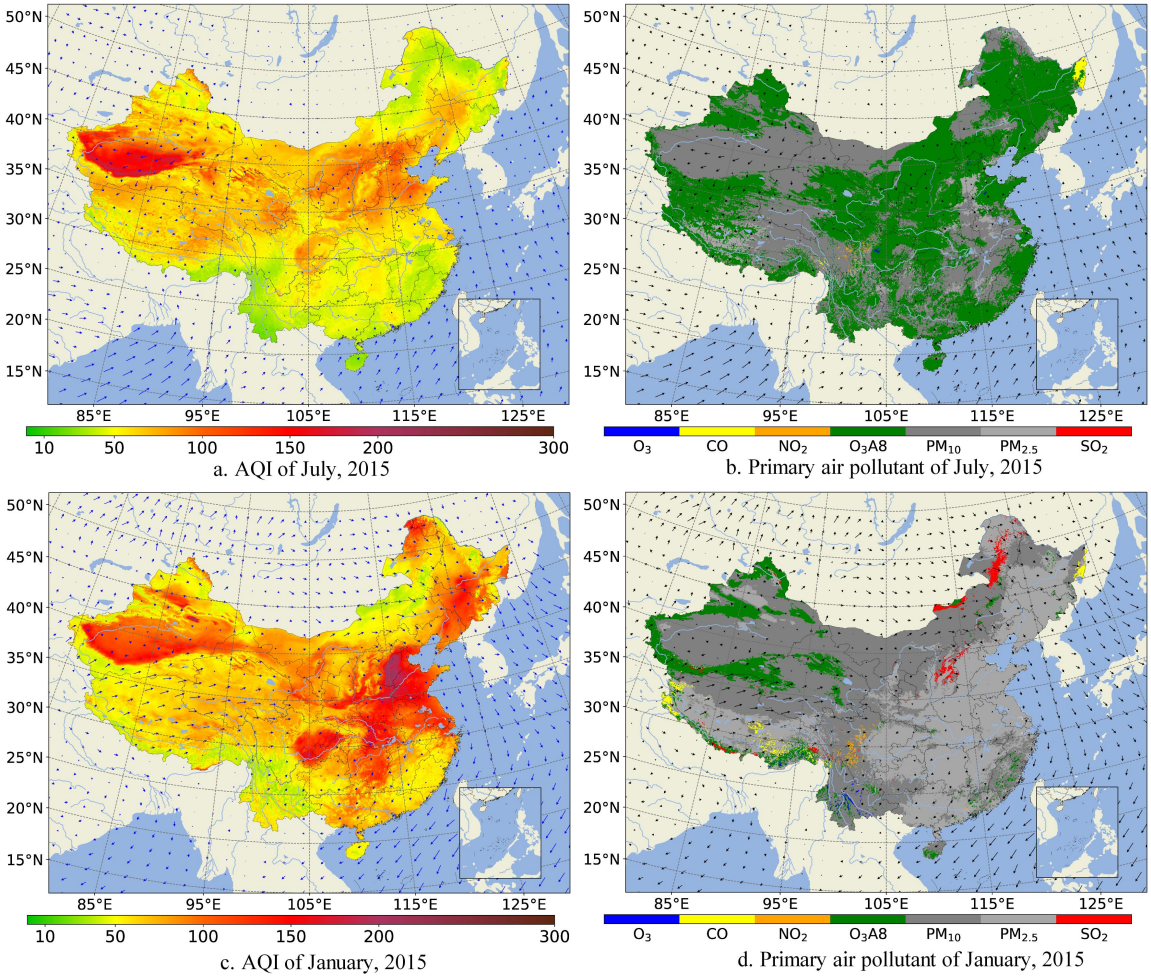
Supplementary Figure 16

Monthly mean AQI and its change for each air pollutant from July 2015 to July 2018 in mainland China. a, b and c, $PM_{2.5}$. d, e and f, NO_2. g, h and I, PM_{10}. j, k and l, O_3A8. m, n and o, SO_2. p, q and r, CO. a, d, g, j, m and p, monthly mean AQI of July 2015 for each pollutant. b, e, h, k, n and q: Monthly mean AQI of July 2018 for each pollutant. c, f, i, l, o and r, change of mean AQI from July 2015 to July 2018 for each pollutant.



Supplementary Figure 17

Summer and winter AQI and primary air pollutant for 2015. a, mean AQI of July 2015. b, primary air pollutant for mean AQI of July 2015. c, mean AQI of January 2015. d, primary air pollutant for mean AQI of January 2015.



a. AQI of July, 2015

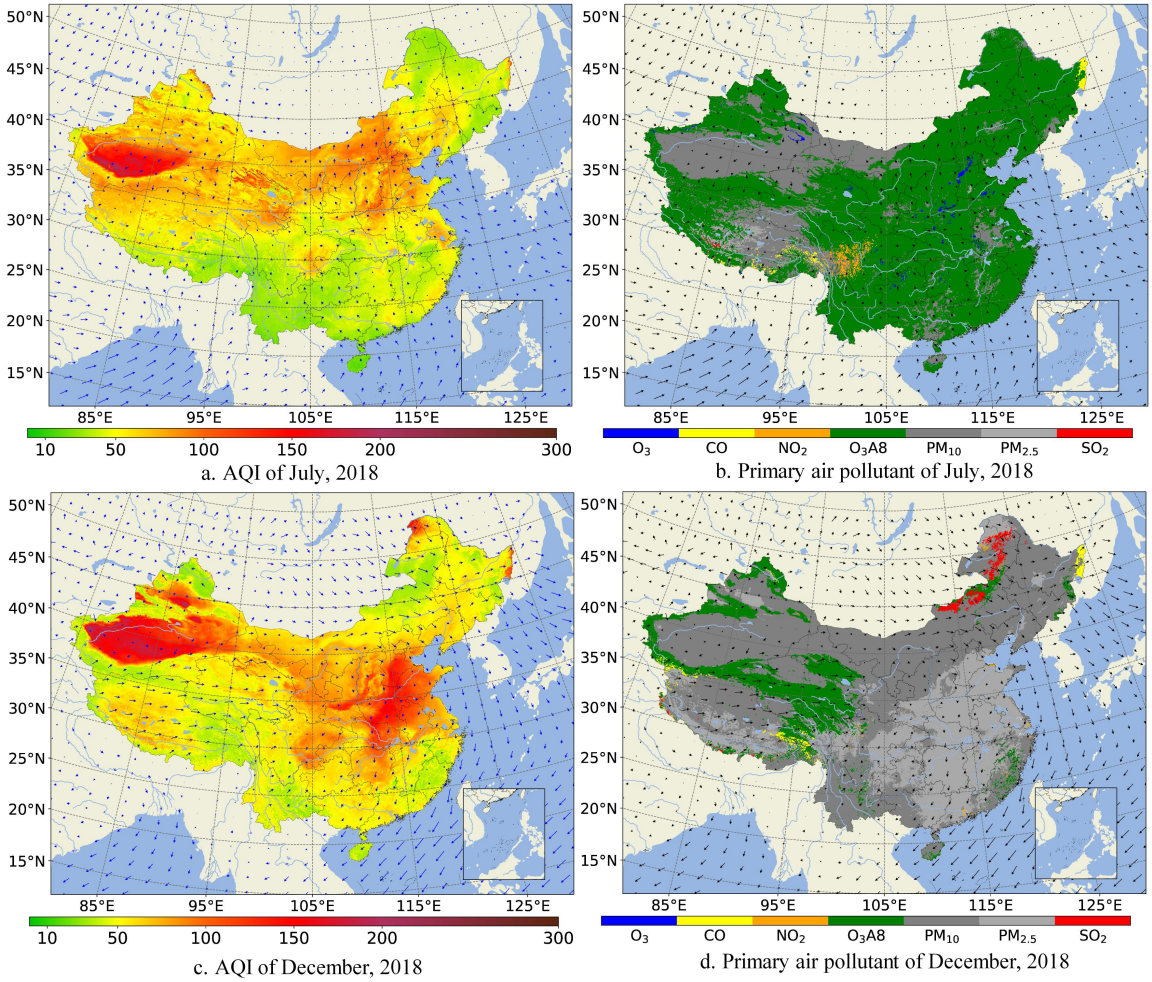
b. Primary air pollutant of July, 2015

c. AQI of January, 2015

d. Primary air pollutant of January, 2015

Supplementary Figure 18

Summer and winter AQI and primary air pollutant for 2018. a, mean AQI of July 2018. b, primary air pollutant for mean AQI of July 2018. c, mean AQI of December 2018. d, primary air pollutant for mean AQI of December 2018.



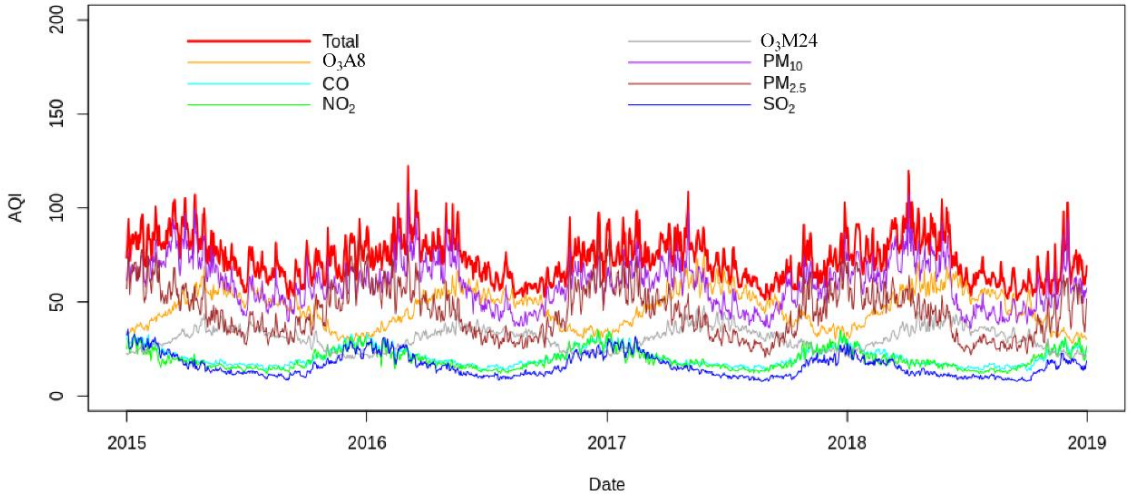
a. AQI of July, 2018

b. Primary air pollutant of July, 2018

c. AQI of December, 2018

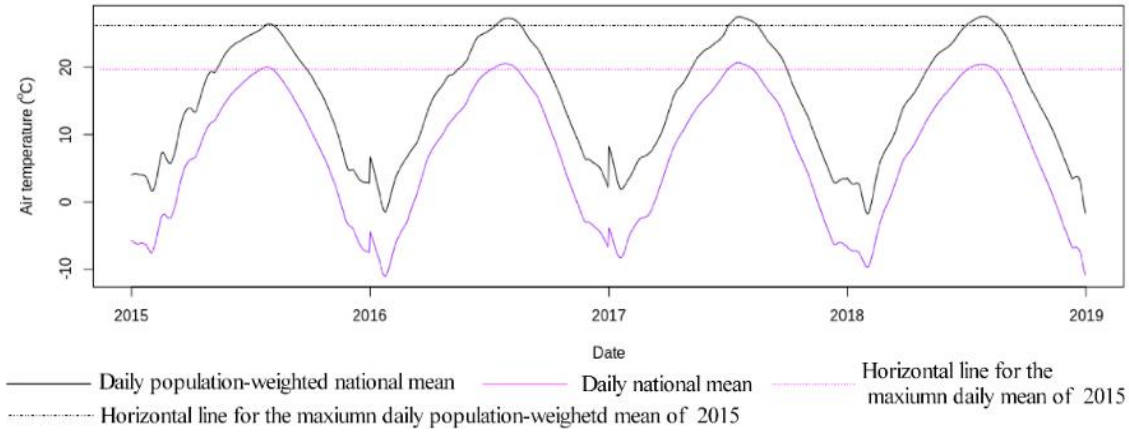
d. Primary air pollutant of December, 2018

Supplementary Figure 19
Daily national AQI for mainland China.



Supplementary Figure 20

Summer and winter AQI and primary air pollutant for 2015. a, mean AQI of July 2015. b, primary air pollutant for mean.



Supplementary Tables

Supplementary Table 1

Covariates and data sources.

Class	Variable	Unit	Description and source
Coordinates	Latitude	º	To capture spatial variability.
	Longitude	º	Elevation: Space Shuttle Radar
Elevation	Elevation	meter	Topology Mission (SRTM)
Time	Day of year	-	To capture temporal variability
Proxy factors	Ground aerosol coefficient	-	Derived from MAIAC AOD by PBLH and relative humidity
	OMI-NO_2	molec/cm^2	From the Dutch-Finnish OMI sensor to capture NO_2 column
PBLH	Planetary boundary layer height (PBLH)	meter	To constrain spreading of air pollutants. From NASA GMI.
MERRA2-GMI	Carbon monoxide	mol mol-1	Bottom layer diagnostics (50-60 meters above the ground): Hourly values for a variety of surface-level values, including PM, $PM_{2.5}$, NO, and NO_2. These will likely provide the most valuable model inputs.
	Dry deposition of OX	kg m-2 s-1	
	Dust mass mixing ratio $PM_{2.5}$	kg kg-1	
	Nitrate mass mixing ratio	kg/kg	
	Nitrogen dioxide	mol mol-1	
	Ozone	mol mol-1	
	Organic carbon mass mixing ratio	kg kg-1	
	Total reconstructed $PM_{2.5}$	kg m-3	
	Sea salt mass mixing ratio $PM_{2.5}$	kg kg-1	
	Black carbon surface mass concentration	kg m-3	Aerosol diagnostics: Total aerosol extinction on an hourly basis, which may be used for model input (and estimates of surface mass concentrations for black carbon, dust, etc..).
	DMS surface mass concentration	kg m-3	
	Dust surface mass concentration - $PM_{2.5}$	kg m-3	
	Dust surface mass concentration	kg m-3	
	Nitric acid surface mass concentration	kg m-3	
	Ammonia surface mass concentration	kg m-3	
	Ammonium surface mass concentration	kg m-3	
	Nitrate surface mass concentration	kg m-3	
	Organic carbon surface mass concentration	kg m-3	
	Total reconstructed $PM_{2.5}$	kg m-3	
	SO_2 surface mass concentration	kg m-3	
	SO_4 surface mass concentration	kg m-3	
	Sea salt surface mass concentration $PM_{2.5}$	kg m-3	
	Total aerosol extinction AOT [550 nm]	-	
	Sulphur dioxide 10am local	mol mol-1	Daily satellite overpass fields: instantaneous measurements at 10AM and 2PM local time.
	Nitrogen dioxide 10am local	mol mol-1	
	Sulphur dioxide 2pm local	mol mol-1	
	Nitrogen dioxide 2pm local	mol mol-1	
	50-meter eastward wind	m s-1	Single-level diagnostics: basic meteorological information such as temperature and humidity at 2m and 10m. These may be useful model inputs
	10-meter eastward wind	m s-1	
	2-meter eastward wind	m s-1	
	50-meter northward wind	m s-1	
	10-meter northward wind	m s-1	
	2-meter northward wind	m s-1	
	10-meter specific humidity	kg kg-1	
	2-meter specific humidity	kg kg-1	
	10-meter air temperature	ºC	
	2-meter air temperature	ºC	
	Surface skin temperature	ºC	
	total tropospheric ozone	Dobsons	

Derived wind variables	Vertical stagnation of wind speed	m/s	$w_{stag} = \left(\sqrt{u_{50}^2 + v_{50}^2} - \sqrt{u_{10}^2 + v_{10}^2}\right)_6$, derived from wind speeds of single-level diagnostics.
	Wind shear/mechanical mixing	m/s	$w_{stag} = \left(\sqrt{u_{10}^2 + v_{10}^2} - \sqrt{u_2^2 + v_2^2}\right)_6$, derived from wind speeds of single-level diagnostics.
High-resolution meteorology	Air temperature	^oC	High-resolution spatiotemporal interpolation [7,8].
	Air pressure	hPa	
	Relative humidity	%	
	Wind speed	m/s	
	Precipitation	mm	
For traffic-related NO_2	Roadway length within the buffer of 10 km	meter	Extracted from the open street traffic network.
Greenness	NDVI	-	average NDVI from NASA's Aqua and Terra satellite.
An indirect indicator	Is workday?	-	As an indirect indicator of air pollutants.

Note: The variables including the MERRA2-GMI input of pollutants (shown in gray background) were not used in our sensitivity analysis.

Supplementary Table 2

Hyperparameters tested in the sensitivity analysis for the compared machine learning or statistical regression methods.

Algorithm	Description	Hyperparameters (value range; step) or (value setup)
DGM (no MERRA2 GMI)	Our physics-inspired graph hybrid model (no MERRA2 GMI used): a hybrid system of local graph convolutions, full residual connections and attention layers, with the soft constraint of mass conservation.	Number of neighbors (4-30; 2), number of graph convolutions (1-10; 1), numbers of hidden encoding layers (1-20; 1), number of attention layers (1-8; 1); learning rate (0.001-0.2; 0.01), and mini batch size (256-4096; 256)
FRE	Full residual encoder-decoder: a neural network of multilayer perceptron with full residual connections, as a general nonlinear approximator of any continuous functions.	Numbers of hidden layers (1-20; 1), number of attention layers (1-10; 1); learning rate (0.001-0.2; 0.01), and mini batch size (256-4096; 256)
SAGE	GraphSAGE: a stochastic generalization of graph convolutions, especially useful for massive, dynamic graphs that contain rich feature information. It has the same network topology as local graph convolution in our DGM method.	Number of neighbors (4-30), number of graph convolutions (1-10); learning rate (0.001-0.2), and mini batch size (256-4096)
RF	Random forest: a method of averaging multiple deep decision trees, trained on different parts of the same training set, with the goal of reducing the variance.	Number of trees (50-500; 50), criterion (mse), min samples split (1-10; 2), bootstrap (True, False), max feature ('auto')
XGB	XGBoost: an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It is run under the gradient boosting framework.	Booster (gbtree, gblinear), number of trees (50-500; 50), learning rate (0-1; 0.02), max depth (3-10; 1), min child weight (0-2; 0.3), lambda (0-1; 0.2), alpha (0-1; 0.2), tree method ('auto')
Kri[a]	Ordinary kriging: estimates a value at a point of a region for which a variogram is known, using data in the neighborhood of the estimation location. With local second-order stationarity, it implicitly evaluates the mean in a moving neighborhood.	Semivariogram fitting: model (stable, circular, spherical, exponential, Gaussian), anisotropy (No), partial sill (True), number of lags (6-20; 2); search neighborhood: max neighbors (3-16; 2), min neighbors (2-6; 1).
RKrig	Regression kriging: combines a regression model with simple kriging of the regression residuals. The experimental variogram of residuals is first modeled, and then simple kriging (SK) is applied to the residuals to give the spatial prediction of the residuals.	Regression model: multilinear regression (MLR); kriging of the residuals from the MLR: ordinary kriging that also tested its similar hyperparameters including variogram models (model and number of lags), search neighborhood (min and max neighbors)
GAM	Generalized additive model: a generalized additive model in which the linear response variable depends linearly on unknown smooth functions of some predictor variables. It was originally developed to blend properties of generalized linear models with additive models.	Smoothing basis (cr: cubic regression spline), link function (normal distribution), optimizing hyperparameters ("auto"): InitialLearnRateForPredictors, NumTreesPerPredictor, Interactions, InitialLearnRateForInteractions, and NumTreesPerInteraction.

[a] No covariates used for ordinary kriging; all the same covariates used for all the other methods.

Supplementary Table 3

Mean R^2 of different methods across seven air pollutants.

Method	Type	$PM_{2.5}$	NO_2	PM_{10}	O_3A8	O_3M24	SO_2	CO
DGM	Training	0.93	0.93	0.92	0.88	0.85	0.85	0.83
	Testing	0.88	0.78	0.86	0.81	0.76	0.73	0.70
	Ind. testing	0.87	0.72	0.85	0.78	0.73	0.63	0.58
DGM (No MERRA2-GMI pollutants)	Training	0.92	0.94	0.93	0.88	0.81	0.84	0.80
	Testing	0.88	0.77	0.86	0.80	0.71	0.73	0.69
	Ind. testing	0.87	0.68	0.85	0.78	0.70	0.61	0.55
FRE	Training	0.93	0.93	0.92	0.90	0.84	0.86	0.84
	Testing	0.81	0.70	0.81	0.73	0.70	0.68	0.64
	Ind. testing	0.72	0.59	0.71	0.70	0.66	0.57	0.54
SAGE	Training	0.68	0.67	0.68	0.67	0.63	0.57	0.57
	Testing	0.67	0.62	0.66	0.66	0.62	0.54	0.52
	Ind. testing	0.66	0.59	0.65	0.65	0.61	0.53	0.49
RF	Training	0.94	0.93	0.94	0.94	0.93	0.92	0.92
	Testing	0.78	0.71	0.76	0.74	0.72	0.74	0.67
	Ind. testing	0.77	0.56	0.75	0.70	0.68	0.58	0.50
XGB	Training	0.68	0.68	0.64	0.65	0.63	0.62	0.57
	Testing	0.67	0.67	0.64	0.65	0.62	0.62	0.56
	Ind. testing	0.66	0.59	0.61	0.62	0.59	0.54	0.41
Krig	Training	-	-	-	-	-	-	-
	Testing	0.55	0.46	0.56	0.51	0.50	0.48	0.45
	Ind. testing	0.55	0.41	0.54	0.49	0.48	0.46	0.42
RKrig	Training	-	-	-	-	-	-	-
	Testing	0.70	0.67	0.71	0.69	0.65	0.62	0.61
	Ind. testing	0.72	0.62	0.70	0.68	0.66	0.60	0.56
GAM	Training	0.53	0.53	0.42	0.55	0.49	0.34	0.37
	Testing	0.53	0.53	0.41	0.54	0.50	0.34	0.36
	Ind. testing	0.54	0.54	0.46	0.54	0.49	0.30	0.35

Method: Deep graph modeling proposed in this paper (DGM); Full residual encoder-decoder (FRE); GraphSAGE (SAGE); Random forest (RF); XGBoost (XGB); Kriging (Krig); Regression kriging (RKrig); Generalized additive model (GAM). Ind. testing: independent testing. Air pollutants: $PM_{2.5}$: Fine inhalable particles, with diameters that are generally 2.5 micrometers and smaller; NO_2: Nitrogen dioxide; PM_{10}: Inhalable particles, with diameters that are generally 10 micrometers and smaller; O_3A8: 8-hour average of O_3; O_3M24: Hourly maximum of daily O_3; SO_2: Sulfur dioxide; CO: Carbon monoxide.

Supplementary Table 4

Mean RMSE for different methods across seven air pollutants.

Method	Type	$PM_{2.5}$	NO_2	PM_{10}	O_3A8	O_3M24	SO_2	CO
Unit		$\mu g/m^3$	$\mu g/m^3$	$\mu g/m^3$	$\mu g/m^3$	$\mu g/m^3$	$\mu g/m^3$	mg/m^3
DGM	Training	9.41	6.58	16.05	15.81	20.44	8.71	0.24
	Testing	12.66	9.14	21.90	20.23	26.07	11.72	0.31
	Ind. testing	13.42	11.53	21.96	21.47	27.58	12.37	0.37
DGM (No MERRA2-GMI pollutants)	Training	9.47	6.13	16.01	15.93	22.69	8.79	0.25
	Testing	12.76	9.15	22.13	20.50	28.42	11.49	0.32
	Ind. testing	13.41	11.69	21.92	21.86	29.03	12.67	0.38
FRE	Training	9.65	5.11	16.01	14.73	21.14	8.55	0.24
	Testing	15.46	10.63	24.94	25.79	28.63	12.17	0.34
	Ind. testing	16.89	13.51	30.24	26.41	30.48	13.41	0.39
SAGE	Training	20.45	11.15	33.24	27.03	32.19	13.40	0.37
	Testing	20.58	11.89	33.36	27.48	32.73	13.69	0.39
	Ind. testing	20.78	12.49	33.51	27.43	32.18	13.77	0.41
RF	Training	8.82	5.19	14.65	12.62	14.28	5.88	0.17
	Testing	17.22	10.45	28.61	25.22	28.30	11.35	0.33
	Ind. testing	17.60	13.05	28.45	26.91	29.88	13.11	0.40
XGB	Training	20.75	10.97	34.77	29.37	32.40	13.61	0.38
	Testing	21.02	11.09	35.32	29.69	32.75	13.74	0.38
	Ind. testing	21.69	12.49	35.47	30.63	33.60	13.76	0.41
Krig	Training	-	-	-	-	-	-	-
	Testing	22.5	15.03	37.23	37.41	42.13	15.53	0.45
	Ind. testing	22.6	17.01	38.10	38.61	43.02	15.81	0.49
RKrig	Training	-	-	-	-	-	-	-
	Testing	19.01	11.13	30.41	25.42	33.23	13.23	0.40
	Ind. testing	18.46	12.08	30.03	25.92	32.97	13.89	0.37
GAM	Training	27.40	13.42	57.92	31.31	37.62	20.13	0.51
	Testing	27.18	13.46	59.77	31.36	37.63	20.00	0.52
	Ind. testing	26.83	13.64	47.08	31.33	37.52	19.33	0.53

Method: Deep graph modeling proposed in this paper (DGM); Full residual encoder-decoder (FRE); GraphSAGE (SAGE); Random forest (RF); XGBoost (XGB); Kriging (Krig); Regression kriging (RKrig); Generalized additive model (GAM). Ind. testing: independent testing. Air pollutants: $PM_{2.5}$: Fine inhalable particles, with diameters that are generally 2.5 micrometers and smaller; NO_2: Nitrogen dioxide; PM_{10}: Inhalable particles, with diameters that are generally 10 micrometers and smaller; O_3A8: 8-hour average of O_3; O_3M24: Hourly maximum of daily O_3; SO_2: Sulfur dioxide; CO: Carbon monoxide.

Supplementary Table 5

Mean RMSE for the e^2 and the total loss terms of seven air pollutants.

Loss type	$PM_{2.5}$ $\mu g/m^3$	NO_2 $\mu g/m^3$	PM_{10} $\mu g/m^3$	O_3A8 $\mu g/m^3$	O_3M24 $\mu g/m^3$	SO_2 $\mu g/m^3$	CO mg/m^3
Total loss	9.41	6.58	16.05	15.81	20.44	8.71	0.24
e_2	0.037	0.056	0.053	0.094	0.134	2.33E-4	1.62E-5

Supplementary Table 6

Increase in variance (R^2) and decrease in RMSE by our DGM in comparison with representative machine learning methods for criteria air pollutants.

Metric	Type	$PM_{2.5}$ $\mu g/m^3$	NO_2 $\mu g/m^3$	PM_{10} $\mu g/m^3$	O_3A8 $\mu g/m^3$	O_3M24 $\mu g/m^3$	SO_2 $\mu g/m^3$	CO mg/m^3
R^2	Increase range	0.10-0.33	0.10-0.31	0.10-0.39	0.08-0.29	0.05-0.25	0.03-0.33	0.02-0.23
	Mean increase	0.21	0.16	0.22	0.15	0.13	0.11	0.11
	Increase percentage	13-61%	16-76%	13-85%	11-59%	7-52%	5-110%	4-66%
	Mean increase percentage	34%	31%	38%	27%	25%	30%	27%
RMSE	Decrease range	3.47-13.41	0.55-5.48	6.49-25.12	4.45-17.14	2.30-15.44	0.74-6.96	0.00-0.16
	Mean decrease	7.28	1.93	12.73	8.14	6.66	2.34	0.06
	Decrease percentage	21-50%	5-32%	23-53%	17-44%	8-36%	6-36%	0-30%
	Mean decrease percentage	34%	13%	35%	26%	18%	15%	12%

Supplementary Table 7

Comparison of RMSE for the temporal trends: summarized monthly means of air pollutants for all sites in the site-based independent test.

Type	$PM_{2.5}$ $\mu g/m^3$	NO_2 $\mu g/m^3$	PM_{10} $\mu g/m^3$	O_3A8 $\mu g/m^3$	SO_2 $\mu g/m^3$	CO mg/m^3
DGM	0.79	1.26	1.61	1.74	1.16	0.024
FRE	3.60	2.18	4.68	5.03	2.67	0.065
SAGE	3.05	2.41	3.83	3.91	1.77	0.049
XGBoost	3.92	2.74	6.06	5.43	2.31	0.064
Random forest	3.05	2.40	3.83	3.91	1.76	0.049

Supplementary Table 8
Comparison of the point estimates, grid mean estimates and MERRA2-GMI related variables.

MERRA2-GMI/Satellite variables	Source	Unit	Cor.	Pol.	Model prediction	Cor. [a]	R^{2b}
Total reconstructed $PM_{2.5}$	Aerosol diagnostics	kg m^{-3}	0.50	$PM_{2.5}$	Point estimate	0.93	0.87
Total reconstructed $PM_{2.5}$	Bottom layer diagnostics	kg m^{-3}	0.51	$PM_{2.5}$	1x1 km^2 grid mean	0.82	0.72
NO_2	Bottom layer diagnostics	mol mol^{-1}	0.53	NO_2	Point estimate	0.85	0.72
NO_2 (average over 10am and 2pm)	Satellite overpass fields	mol mol^{-1}	0.41	NO_2	1x1 km^2 grid mean	0.72	0.51
OMI-NO_2	Aura OMI	mol cm^{-2}	0.18	PM_{10}	Point estimate	0.92	0.85
Total reconstructed PM	Bottom layer diagnostics	mol mol^{-1}	0.36	PM_{10}	1x1 km^2 grid mean	0.83	0.68
MAIAC AOD	MODIS Terra/Aqua	-	0.22	O_3A8	Point estimate	0.89	0.78
Total tropospheric O_3	Single-level diagnostics	dobsons	0.005	O_3A8	1x1 km^2 grid mean	0.75	0.57
O_3(average over 10am and 2pm)	Satellite overpass fields	mol mol^{-1}	0.35	O_3M24	Point estimate	0.85	0.73
					1x1 km^2 grid mean	0.7	0.46
SO_2 surface mass concentration	Aerosol diagnostics	kg m^{-3}	0.15	SO_2	Point estimate	0.79	0.63
SO_2 (average over 10am and 2pm)	Satellite overpass fields	mol mol^{-1}	0.17		1x1 km^2 grid mean	0.71	0.49
CO	Bottom layer diagnostics	mol mol^{-1}	0.22	CO	Point estimate	0.76	0.57

[a]Cor.: Pearson correlation with the ground observed values; [b]R^2 between estimates and observed values.

Supplementary Table 9

Seasonal performance metrics (R^2 and RMSE) for the site-based independent test.

Year	Season	$PM_{2.5}$		PM_{10}		NO_2		O_3A8		SO_2		CO	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
2015	Spring (AMJ[a])	0.76	13.35	0.77	23.44	0.53	10.73	0.53	32.83	0.36	13.81	0.36	0.40
	Summer (JAS[b])	0.73	11.73	0.73	19.12	0.51	9.84	0.59	29.51	0.28	12.47	0.37	0.35
	Autumn (OND[c])	0.85	18.97	0.83	27.91	0.63	13.24	0.72	21.50	0.60	17.10	0.60	0.44
	Winter (JFM[d])	0.85	17.40	0.83	27.93	0.63	13.52	0.51	23.81	0.63	21.33	0.48	0.56
2016	Spring (AMJ)	0.77	10.73	0.79	22.35	0.62	10.06	0.60	27.61	0.46	10.92	0.46	0.36
	Summer (JAS)	0.78	9.19	0.73	16.29	0.56	9.41	0.68	25.91	0.35	9.60	0.45	0.31
	Autumn (OND)	0.89	14.96	0.86	24.26	0.76	11.21	0.76	16.29	0.67	14.00	0.59	0.42
	Winter (JFM)	0.82	18.68	0.83	28.69	0.66	12.38	0.61	21.12	0.63	17.10	0.56	0.45
2017	Spring (AMJ)	0.84	9.24	0.85	20.71	0.67	9.47	0.82	20.75	0.51	8.06	0.43	0.30
	Summer (JAS)	0.82	7.37	0.79	14.08	0.62	8.43	0.82	20.02	0.43	7.28	0.53	0.26
	Autumn (OND)	0.90	12.54	0.87	20.34	0.79	10.63	0.82	14.18	0.60	11.01	0.63	0.32
	Winter (JFM)	0.89	16.17	0.86	24.52	0.76	11.23	0.70	18.01	0.69	14.25	0.62	0.40
2018	Spring (AMJ)	0.84	8.16	0.87	19.49	0.67	8.99	0.85	18.26	0.46	7.00	0.46	0.24
	Summer (JAS)	0.76	6.43	0.74	12.07	0.61	7.24	0.83	17.85	0.38	6.02	0.44	0.23
	Autumn (OND)	0.91	10.61	0.90	18.83	0.78	9.47	0.89	12.56	0.58	8.16	0.61	0.29
	Winter (JFM)	0.90	13.34	0.88	21.50	0.77	10.46	0.82	14.04	0.60	12.06	0.61	0.32
All seasons	2015	0.85	15.67	0.83	24.91	0.60	12.86	0.66	27.32	0.59	16.61	0.51	0.46
	2016	0.86	14.02	0.84	23.38	0.65	11.84	0.73	23.08	0.63	13.36	0.55	0.41
	2017	0.90	11.72	0.87	20.18	0.71	10.88	0.85	18.42	0.66	10.21	0.60	0.34
	2018	0.91	9.99	0.89	18.35	0.71	9.95	0.88	15.90	0.60	8.66	0.59	0.28

Note: [a]. April, May and June; [b]. July, August and September; [c]. October, November and December; [d]. January, February and March.

Supplementary Table 10

Seasonal performance metrics (Pearson's correlation and mean bias) for the site-based independent test.

Year	Season	$PM_{2.5}$		PM_{10}		NO_2		O_3A8		SO_2		CO	
		Cor.[a]	MB[b]	Cor.	MB	Cor.	MB	Cor.	MB	Cor.	MB	Cor.	MB
2015	Spring (AMJ[c])	0.87	-0.96	0.88	-1.13	0.71	-1.72	0.73	-2.86	0.59	-1.81	0.59	-0.036
	Summer (JAS[d])	0.86	-0.77	0.86	-0.35	0.70	-1.34	0.77	-2.28	0.52	-1.59	0.61	-0.017
	Autumn (OND[e])	0.93	-0.77	0.91	-1.41	0.76	-1.88	0.85	-0.98	0.77	-1.82	0.76	0.006
	Winter (JFM[f])	0.92	-0.50	0.91	-2.04	0.76	-1.68	0.71	-1.07	0.80	-2.10	0.69	-0.047
2016	Spring (AMJ)	0.88	-0.51	0.89	-0.66	0.75	-0.47	0.78	-1.71	0.67	-0.94	0.66	-0.017
	Summer (JAS)	0.88	-0.41	0.85	-0.47	0.72	-0.71	0.83	-2.91	0.60	-0.89	0.65	-0.020
	Autumn (OND)	0.94	-0.42	0.93	-1.59	0.84	-1.58	0.87	-1.12	0.82	-0.67	0.76	-0.025
	Winter (JFM)	0.91	-0.84	0.91	-1.74	0.78	-1.28	0.77	-0.10	0.80	-1.61	0.74	-0.023
2017	Spring (AMJ)	0.92	-1.00	0.92	-1.70	0.79	-0.89	0.91	-2.59	0.71	-0.46	0.64	-0.012
	Summer (JAS)	0.90	-0.66	0.89	-1.62	0.75	-0.55	0.91	-2.04	0.65	-0.06	0.72	-0.008
	Autumn (OND)	0.95	-0.79	0.93	-1.87	0.85	-1.91	0.91	-1.03	0.77	-0.25	0.78	0.00001
	Winter (JFM)	0.94	-1.16	0.93	-1.52	0.84	-1.22	0.84	-0.93	0.83	-0.35	0.78	-0.028
2018	Spring (AMJ)	0.92	-0.25	0.93	-1.67	0.79	-0.68	0.92	-1.35	0.68	-0.42	0.66	0.003
	Summer (JAS)	0.87	-0.06	0.86	-0.75	0.75	-0.31	0.91	-1.01	0.62	-0.17	0.65	0.010
	Autumn (OND)	0.95	0.07	0.95	-0.79	0.85	-0.62	0.94	-0.48	0.76	-0.49	0.77	0.002
	Winter (JFM)	0.95	-1.17	0.94	-2.25	0.84	-0.73	0.91	-0.02	0.78	-0.34	0.77	-0.004
All seasons	2015	0.92	-0.75	0.91	-1.24	0.77	-1.66	0.81	-1.79	0.77	-1.83	0.71	-0.023
	2016	0.93	-0.55	0.92	-1.13	0.81	-1.03	0.86	-1.44	0.79	-1.03	0.74	-0.021
	2017	0.95	-0.90	0.93	-1.68	0.84	-1.14	0.92	-1.66	0.81	-0.28	0.77	-0.012
	2018	0.95	-0.36	0.94	-1.38	0.84	-0.59	0.94	-0.72	0.77	-0.36	0.76	0.0072

Note: [a]. Pearson's correlation (defined in [9]), unit: $\mu g/m^3$ for $PM_{2.5}$, PM_{10}, NO_2, O_3A8, SO_2 and mg/m^3 for CO; [b]. Mean bias; April, May and June; [b]. July, August and September; [c]. October, November and December; [d]. January, February and March.

Supplementary Table 11

Yearly mean of criteria air pollutants for mainland China.

Year	National Mean	All[a]	$PM_{2.5}$	NO_2	PM_{10}	O_3A8[b]	O_3M24[c]	SO_2	CO
2015	Con. [d]	-	35	16	80	86	94	18	0.85
	Con. PD [e]	-	51	27	87	85	99	20	1
	AQI[f]	75	48	19	62	46	30	18	21
	AQI. PD[g]	84	69	34	66	47	33	19	24
2016	Con.	-	32	16	75	87	94	17	0.84
	Con. PD	-	47	28	81	87	101	19	0.97
	AQI	73	45	19	60	46	30	17	21
	AQI. PD	80	64	34	63	47	34	18	24
2017	Con.	-	31	16	72	90	97	16	0.82
	Con. PD	-	42	28	80	92	107	17	0.94
	AQI	72	43	19	58	48	31	16	20
	AQI. PD	80	62	33	62	51	37	16	23
2018	Con.	-	30	15	73	90	97	14	0.78
	Con. PD	-	42	26	76	92	108	14	0.88
	AQI	72	41	18	59	48	31	14	19
	AQI. PD	78	57	32	60	51	37	14	22

Note: [a]. AQI for all air pollutants; [b]. O_3A8: 8-hour average of O_3; [c]. O_3M24: Hourly maximum of daily O_3; [d]. Air pollutant concentration (Unit: $\mu g/m^3$ for SO_2, O_3, $PM_{2.5}$, PM_{10}. NO_2; mg/m^3 for CO); [e]. Air pollutant concentration weighted by population density; [f]. AQI; g. AQI weighted by population density.

Supplementary References

1 NASA. *MERRA-2 GMI* <https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/> (2018).

2 Li, L. F. High-Resolution Mapping of Aerosol Optical Depth and Ground Aerosol Coefficients for Mainland China. *Remote Sensing* **13**, doi:10.3390/rs13122324 (2021).

3 Lundberg, S. M. & Lee, S. I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems 30 (Nips 2017)* **30** (2017).

4 Lundberg, S. M. *Welcome to the SHAP documentation*, <https://shap.readthedocs.io/en/latest/index.html> (2018).

5 Ministry of Environmental Protection of China. *Technical Regulation on Ambient Air Quality Index*. (China Environmental Science Press, 2012).

6 Li, L. F. *et al.* Ensemble-based deep learning for estimating PM2.5 over California with multisource big data including wildfire smoke. *Environment International* **145**, doi:10.1016/j.envint.2020.106143 (2020).

7 Fang, Y. & Li, L. Estimation of high-precision high-resolution meteorological factors based on machine learning. *Journal of Geo-information Sciences (in Chinese)* **21**, 799-813 (2019).

8 Li, L. Geographically Weighted Machine Learning and Downscaling for High-Resolution Spatiotemporal Estimations of Wind Speed. *Remote Sensing* **11**, 1378 (2019).

9 Sicard, P. *et al.* High spatial resolution WRF-Chem model over Asia: Physics and chemistry evaluation. *Atmospheric Environment* **244**, 118004 (2021).