1. Constructed features in prior research

In this section, we introduce the complicated constructed features in prior research [26–29]. The features are as follows:

1. Geo-intelligent deep belief network (Geoi-DBN): For a specific grid, the spatiotemporal features are represented as:

$$S-PM_{2.5} = \frac{\sum_{i=1}^{n} ws_i PM_{2.5,i}}{\sum_{i=1}^{n} ws_i} \quad ws_i = \frac{1}{ds_i^2} \quad \text{(S1)}$$

$$T-PM_{2.5} = \frac{\sum_{j=1}^{m} wt_j PM_{2.5,j}}{\sum_{j=1}^{m} wt_j} \quad wt_j = \frac{1}{dt_j^2} \quad \text{(S2)}$$

$$DIS = \min\left(\frac{1}{ds_i}\right) \quad i = 1, 2, \ldots, n \quad \text{(S3)}$$

The ds, dt are the distances in space and time, respectively. ws_i, ws_j are the distance weighting coefficient of the i-th station close to the current station in space and the time weighting coefficient of the j-th day before. $PM_{2.5,i}$ is the ground-based $PM_{2.5}$ observation of the nearest i-th station, and $PM_{2.5,j}$ is the observation of a specific station j-th day before. S-$PM_{2.5}$ and T-$PM_{2.5}$ are the constructed features in space and time ($m = 3$, $n = 10$). The DIS is applied to reflect the uneven distributions of stations.

2. Space-time random forest (STRF): The spatial (Ps) and temporal (Pt) characteristics for a given pixel can be expressed as:

$$Ps = \frac{\sum_{w=1}^{W} \frac{1}{ds_w^2} Psw}{\sum_{w=1}^{W} \frac{1}{ds_w^2}} \quad \text{(S4)}$$

$$Pt = \frac{\sum_{l=1}^{L} \frac{1}{dt^2} Ptl}{\sum_{l=1}^{L} \frac{1}{dt_l^2}} \quad \text{(S5)}$$

The ds and dt represent the distances in space and time. The W and L are the W pixels near the current site and the L days before for the same pixel. The Psw is the observation of a station that nears (W pixels) the specific pixel (i.e., the current station), Ptl represents the observation of the current site l-th day ago.

3. Space-time Extra-Trees (STET): The two improved space-time features ($P_{S(i,j,t)}$, $P_{T(i,j,t)}$) can be formulated as:

$$h = f\left(Lon_{i,j,t}, Lat_{i,j,t}\right) = \text{haversin}(\alpha_1 - \alpha_2) + \cos(\alpha_1)\cos(\alpha_2)\text{haversin}(\beta_1 - \beta_2) \quad \text{(S6)}$$

$$\text{haversin}(\theta) = \sin^2(\theta/2) = [1 - \cos(\theta)]/2 \quad \text{(S7)}$$

$$P_{S(i,j,t)} = 2 \times r \times \text{asin}(\sqrt{h}) \quad \text{(S8)}$$

$$P_{T(i,j,t)} = \cos\left(2\pi \frac{d_{i,j,t}}{T}\right) \quad \text{(S9)}$$

For a given station (the i-th one of total stations) on the t-th day of the study period, information from the nearest station (the j-th station) on the same day is helpful. For Equation (6), α_1 and α_2 are the latitudes of two points, and β_1 and β_2 denote the longitudes of the i-th, and j-th points in space. The r represents the earth's radius (km), and d is day of year. T denotes the total days of the year used in the study.

4. Fast space-time LightGBM (STLG): An enhanced $P_{S(i,j,t)}$ can be expressed as:

$$DIS = 2 \cdot r \cdot asin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos\left(\varphi_1\right)\cos\left(\varphi_2\right)\sin^2\left(\frac{\gamma_2 - \gamma_1}{2}\right)} \right) \quad (S10)$$

$$P_{S(i,j,t)} = 2 \times r \times asin(\sqrt{DIS}) \quad (S11)$$

The features in the STLG method are based on the STET method, and this feature (Equation (9)) is also the temporal feature in the STLG method. φ and γ represent the latitude and longitude of the current station on the sphere, respectively, and r denotes the earth's mean radius ($\approx$6371 km).

2. Employed Features in this study

2.1 Raw features

1. Stage 1 step0: **X0** month   **X1** day of month   **X2** hour
   X3 day of week   **X4** day of year   **X5** week of year
   X6 season   **X7** height   **X8** latitude
   X9 longitude   **X10** boundary-layer height
   X11 2m dewpoint temperature   **X12** relative humidity
   X13 surface pressure   **X14** 2m temperature
   X15 10m v-component of wind   **X16** 100m v-component of wind
   X17 $PM_{2.5}$ of MERRA2   **X18** black carbon of $PM_{2.5}$
   X19 dust of $PM_{2.5}$ **X20** organic carbon of $PM_{2.5}$ **X21** sulfate of $PM_{2.5}$
   X22 CO   **X23** O_3 **X24** sea salt of $PM_{2.5}$
   X25 total precipitation   **X26** K index
   X27 100m u-component of wind   **X28** 10m u-component of wind
   The $PM_{2.5}$ of MERRA2 is derived from this formula:
   $$PM_{2.5} = SeaSalt + 1.6*OC + BC + 1.375*SO_4 + DUST \quad (S12)$$
   In this formula, the SeaSalt is the sea salt aerosol. The OC represents the organic carbon aerosol. The BC is the black carbon aerosol, the SO_4 denotes the sulfate aerosol, and the DUST means the dust aerosol.

2. Stage 2 step0:

Table S1. All features used in the stage2 step0.The features are the output of the model at different steps.

Features	Model	Step	Target
X1	Catboost	2	CHAP
X2	Catboost	0	Real
X3	ET	0	CHAP
X4	ET	0	Real
X5	GBDT	2	CHAP
X6	GBDT	2	Real
X7	HistGBM	2	CHAP
X8	HistGBM	2	Real
X9	LightGBM	2	CHAP
X10	LightGBM	2	Real
X11	RF	0	CHAP
X12	RF	1	Real

X13	Xgboost	1	CHAP
X14	Xgboost	2	Real

2.2 Constructed features

According to the outputs of the python codes, a necessary correction was made. The ** and the sqrt denote the x^n (the n-th power operation), and the $\sqrt{x}$ (the square root operation), respectively. The log represents the $\ln x$ (logarithmic operation). The exp means the e^x (the exponential operation). The Abs is the $|x|$ (the absolute value operation). The constructed features in the step1, step2 of the target towards the two values (before the stacking part) and the target towards the ground-based observations during the day and night (the stacking part) are presented here.

1. CHAP step1:   1/X10   1/X12   X21**3   X11**3   X11**2
   sqrt(X24)   1/X4   X18**3   X22**2   X10**3
   X19**2   X28**3   X22**3   X24**2   X20**3
   1/X0   1/X5   X27**2   X19**3   X16**3
       X23**3   log(X25)   1/X1   exp(X15)   sqrt(X25)X17**3
   X16**2   exp(X3)

2. CHAP step2:   X21/X0   X6**3*X21   X4*X22**3
   sqrt(X21)*X22   X2**3*X22**2   sqrt(X19)*X22
   sqrt(X18)*X22   X6**3*X12**2   X18*sqrt(X24)
   sqrt(X6)*sqrt(X23)   X22/X0   X15/X1
   X2**3*X23   X22**3/X12   X17**3/X4
   X18**2*X23   X22*log(X1)   log(X5)/X0
   X12*log(X12)   Abs(X16)/X16   log(X0)/X0
   X22**2*X28**2   X7**3*X22**3   X2**3*Abs(X26)
   X2**3*sqrt(X3)   X2**3*sqrt(X10)   sqrt(X12)*sqrt(X17)
   X18*X21   X4*X23   X4**2*X11
   X12**3*X26   X6**2/X1   X22**3/X4
   X12**2/X0   X17**2*X26**3   X2*sqrt(X18)
   X6**3*X18**3   sqrt(X6)/X10   X2**3*log(X12)
   sqrt(X6)*X22**3   sqrt(X19)*log(X4)   X13/X8
   X4**3*X17   X7*X18**3   X4**3*X21
   X27**2/X5   Abs(X16)/X4   exp(X3)/X1
   X22**3*26**2   X0**3*X6**3   X16**2*X21**3
   X18**2*X23**2   X16**2*sqrt(X22)   X4**3*sqrt(X24)
   sqrt(X12)*sqrt(X19)   X10**2/X4   X3**3*X28
   X21**3/X1   X22**2*X26   1/(X4*X12)X22**3*X27**3
   X15**3*X19**2   X13**2*log(X10)
   X1**3*log(X0)   X2**3*sqrt(X22)   sqrt(X3)*X16**2
   log(X2)*log(X10)   sqrt(X24)*exp(X3)   sqrt(X12)*sqrt(X18)

3. Real step1:   1/X10   1/X12   X22**3   sqrt(X24)   1/X4   X22**2
   X10**3   X21**3   X11**3   X11**2   X16**2   X18**3
   X27**3   X19**2   X24**3   X17**3   X16**3   exp(X3)

4. Real step2:   X22/X0   X15/X1   X21/X0
   X13/X8   X12**2/X0   X17**3/X4
   X22**3/X12   X18**2*X23   X2**3*X23
   log(X0)/X0   eXp(X3)/X1   Abs(X16)/X16
   sqrt(X6)/X10   sqrt(X19)*X22   X22**2*X28**2
   sqrt(X18)*X22   X7**3*X22**3   X6**3*X12**2

X18*sqrt(X23)   X2**3*X22**2   X7*sqrt(X19)
X18*sqrt(X24)   sqrt(X6)*sqrt(X23)   X18*X21
X6**3*X21   X5*X22**3   1/(X4*X12)
Abs(X16)/X4   X12*log(X12)   log(X5)/X0
X22**3*X27**3   X2**3*log(X12)   X2**3*sqrt(X3)
X16**2*sqrt(X22)   X2**3*sqrt(X10)   sqrt(X6)*X22**3
X4**2*X26   X22**3/X4   X2**2*X18
X12**3*X26   X22*log(X1)   X11**3*X17**2
X16**2*X21**3   X2**3*Abs(X11)   sqrt(X19)*log(X4)
sqrt(X12)*sqrt(X17)   X6/X1   X4*X23
X21*X26**3   X27**2/X5   X4**3*X17
X22**2*X26   X3**3*X28   X0**3*X6**3
X18**2*X23**2   X6**3*X18**3   X11**2*X22**3
sqrt(X13)/X9   X10**2*X19**3   X4**3*sqrt(X24)
sqrt(X24)*eXp(X3)   X22**4   X11*X21
X18*X23   X26/X1   X27**2/X4
X22**3/X23   X4**3*X21   X27**2/X1
X7*X18**3   X16**3*X19   X22**3/X5
X19**3/X15   X26**2/X1   X18*log(X5)
X4*log(X12)   X3**3*X20**2   X10**2*X21**3
sqrt(X21)/X23   sqrt(X7)/X10   X16**3*X17**2
X19**3*X27**2   X2**3*log(X25)   X11**2*Abs(X27)
X2**3*Abs(X16)   X11**2*sqrt(X22)   X2**3*sqrt(X22)
sqrt(X3)*X16**2   sqrt(X12)*sqrt(X18)   sqrt(X12)*sqrt(X19)
sqrt(X10)*sqrt(X19)

5. Stage 2 step1: 1/X12   1/X10   1/X3
   X3**3   X5**3   X7**3
   X4**3   log(X4) log(x5)

6. Stage 2 step2: X11**5   X3/X5   X3/X2
   X3/X11   X3/X10   X2/X4
   X10/X2   X2**3/X4   X2**2*X14
   X1*X14**2   X4**3/X1   X14**3/X10
   X11**3/X4   X13**3/X12   X1**3/X4
   X3**3/X12   X6**3/X13   X7**3/X1
   X11**3/X5   X12**3/X5   X9**3/X5
   X1**2/X13   1/(X2*X3)   X4*Abs(X8)
   X5**2*X6**3   X1**3*X10**3   sqrt(X14)/X10
   X2**3*X14**3   sqrt(X10)/X3   sqrt(X2)*X14**3
   X11/X2   X13/X11   X10/X11
   X11/X3   X12**2/X1   X9**3/X4
   Abs(X7)/X12   Abs(X1)/X2   X4/X2
   X5/X3   X2**2/X12   X11**2/X12
   X8**2/X4   Abs(X6)/X10   X9/X4
   X4/X1   X11/X4   X13**2/X1
   X12**2/X13   X11**2/X13   X11**2/ X1
   X8/X6   X4/X11   X4/X5
   X9/X12   X1**2/X6   X9**2/X6
   X12**2/X4   X5**2/ X1   X4**2/X6
   1/(X5*X13)   Abs(X7)/X4   Abs(X12)/X3
   sqrt(X14)/X12   X5**3*X14**2   X5**3*X9**2

7. Feature importance

Table S2. top 10 features of the Catboost method in stage 2, * means the useful information.

feature	SHAP (%)	feature	SHAP (%)
X2	7.14	X4	3.57
X8	5.60	X30*	3.33
X36*	5.35	X7	3.10
X13	4.53	X9	3.00
X14	3.77	X3	2.38

[1](X36: X9^3/X5   X30: X1^3/X4, ^ denotes the n-th power operation (x^n).)

Table S3. Hyperparameters of the Catboost method in Stage 2, Step 2.

depth	eta	iterations	l2_leaf_reg
16	0.059	9479	7.37

Depth means the largest depth of a concrete tree, eta (i.e., learning rate) controls the learning speed which could be used to discover the local minimum value and the best result. Iterations mean the number of trees. L2_leaf_reg represents the L2 regularization method that decreases the possibility of over-fitting.

3. Summer

The RMSE, variance, and R^2 can be formulated as:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - \hat{y}^{(i)}\right)^2} \qquad (S13)$$

$$\mathrm{var}(y) = \left(\sum_{i=1}^{m}\left(y^{(i)} - \bar{y}\right)^2\right)/m \qquad (S14)$$

$$R^2 = 1 - \frac{\sum_i \left(\hat{y}^{(i)} - y^{(i)}\right)^2}{\sum_i \left(\bar{y} - y^{(i)}\right)^2}$$

$$= 1 - \frac{\left(\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2\right)/m}{\left(\sum_{i=1}^{m}\left(y^{(i)} - \bar{y}\right)^2\right)/m} \qquad (S15)$$

$$= 1 - \frac{RMSE^2}{\mathrm{var}(y)}$$

For a specific dataset, the $y(i)$, $\hat{y}(i)$, $\bar{y}$ mean the i-th ground-based PM$_{2.5}$ observation in this dataset, the estimated PM$_{2.5}$ concentration according to the i-th observation value, and the mean value of this dataset, respectively. As for a concrete dataset, the variance is a constant. For the dataset used in this study, the PM$_{2.5}$ concentrations in summer have a lower range (0–256 µg/m^3) than in the other three seasons, and the variance in summer is also lower. The CV-R^2 is 0.67, the CV-RMSE is 12.24 µg/m^3 (var(y) ≈ 454.00 µg/m^3) (spring, autumn, winter: 1989.69 µg/m^3, 2072.82 µg/m^3, 4181.78 µg/m^3). The lower RMSE does not mean that the model performs well in the summer. On the contrary, the proposed model is worse in summer. In other words, the autogeoi-stacking model is relatively under-fitting in summer. The reason may be the

frequent weather changes, the cloud cover, nighttime gaps, and the uncertainty in the CHAP data.
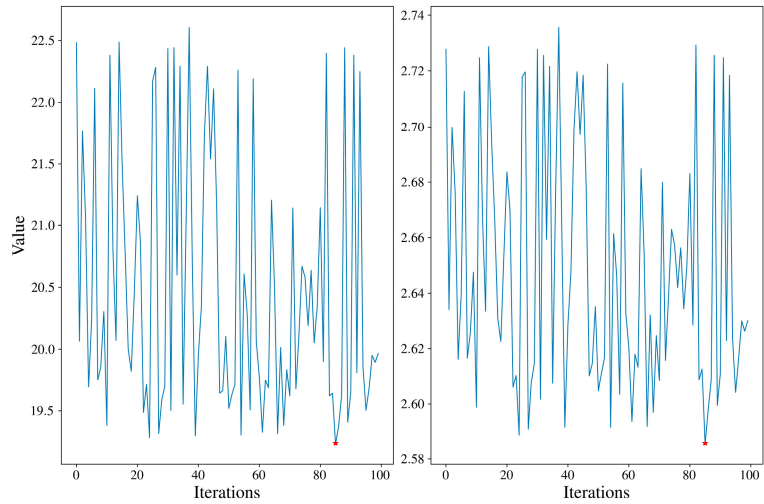

Figure S1. Two optimization objectives in the process of optimizing the hyperparameters of CBR model with the optuna library (version 2.9.1), RMSE on the left and log (MAE) on the right.
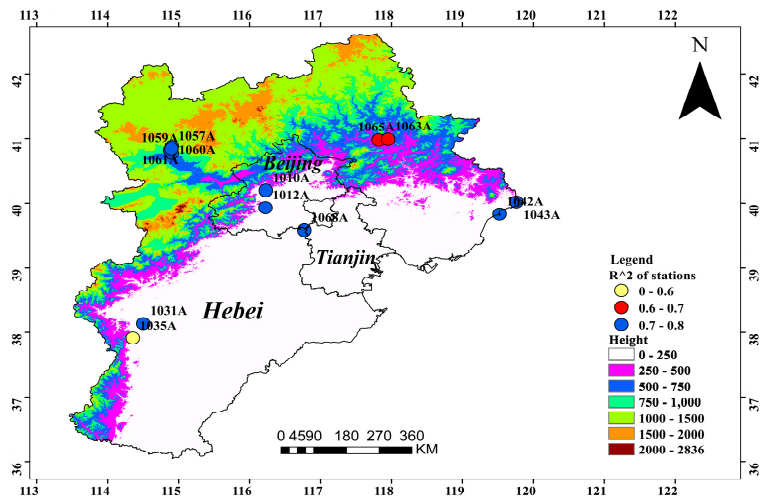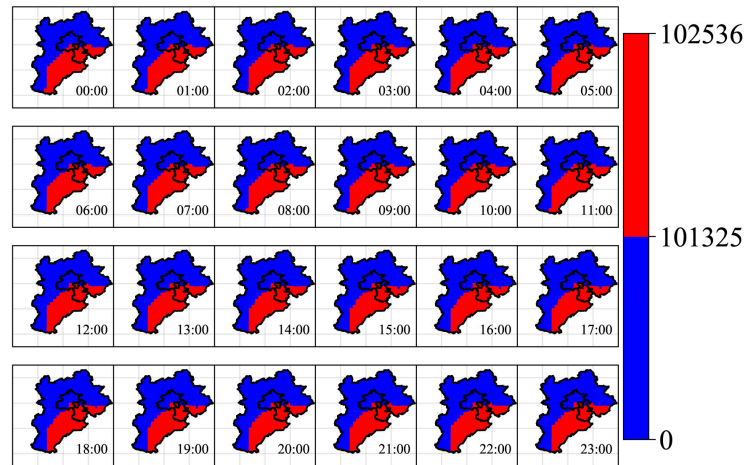

Figure S2. Distributions of the 13 stations that perform lower than other stations used in this study.
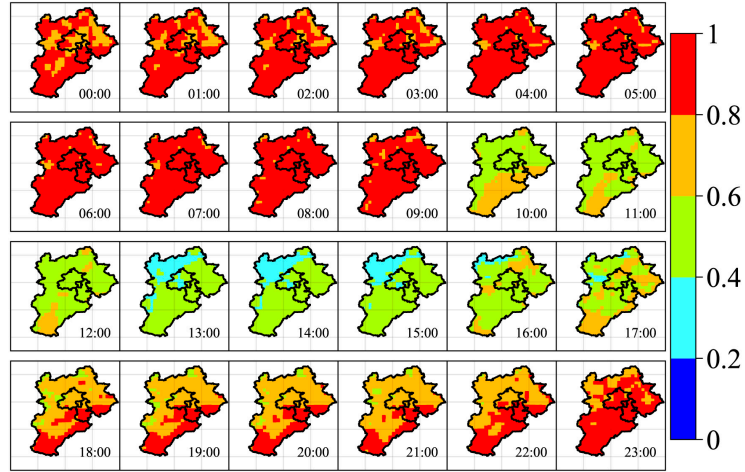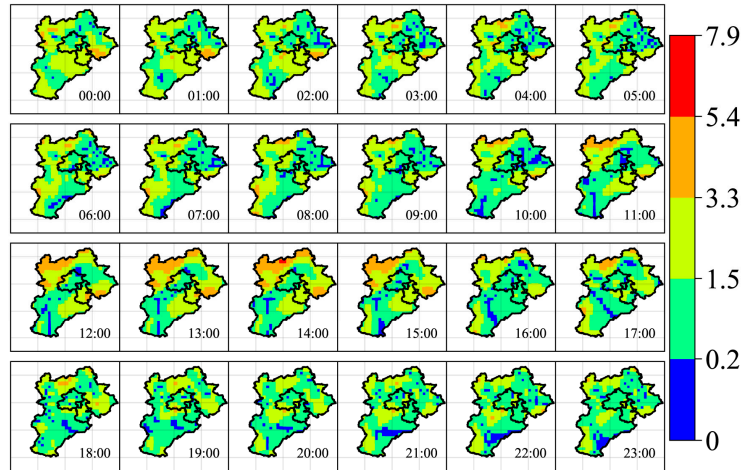
Figure S3. Distributions of the boundary layer height in the BTH region on November 12, 2018.
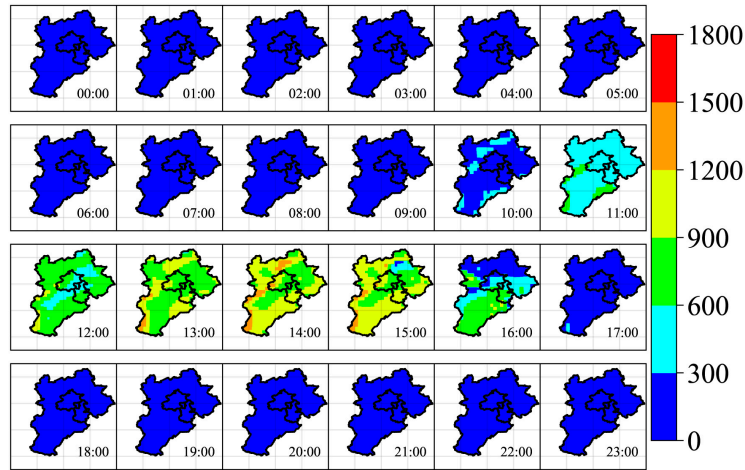

Figure S4. Distributions of the relative humidity in the BTH region on November 12, 2018.

Figure S5. Distributions of the surface pressure in the BTH region on November 12, 2018. The 101325 pa means the standard atmospheric pressure.

Figure S6. Distributions of the wind direction in the BTH region on November 12, 2018. 0 (360) means the north direction. 90 represents the east direction. 180 is the south direction and 270 indicates the west direction.
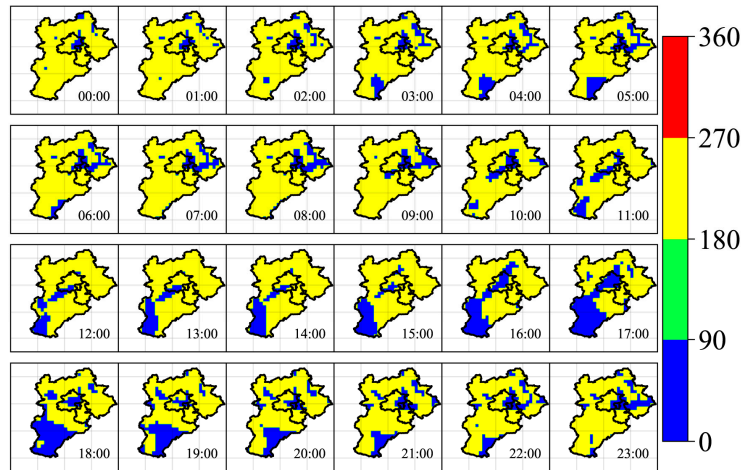
Figure S7. Distributions of the wind speed in the BTH region on November 12, 2018.

References

26   Li, T.; Shen, H.; Yuan, Q.; Zhang, X.; Zhang, L. Estimating Ground-Level PM$_{2.5}$ by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach: Deep Learning for PM$_{2.5}$ Estimation. *Geophys. Res. Lett.* **2017**, *44*, 11985–11993. https://doi.org/10.1002/2017GL075710.

27   Wei, J.; Li, Z.; Cribb, M.; Huang, W.; Xue, W.; Sun, L.; Guo, J.; Peng, Y.; Li, J.; Lyapustin, A.; et al. Improved 1 Km Resolution PM$_{2.5}$ Estimates across China Using Enhanced Space–Time Extremely Randomized Trees. *Atmos. Chem. Phys.* **2020**, *20*, 3273–3289. https://doi.org/10.5194/acp-20-3273-2020.

28   Wei, J.; Huang, W.; Li, Z.; Xue, W.; Peng, Y.; Sun, L.; Cribb, M. Estimating 1-Km-Resolution PM$_{2.5}$ Concentrations across China Using the Space-Time Random Forest Approach. *Remote Sens. Environ.* **2019**, *231*, 111221. https://doi.org/10.1016/j.rse.2019.111221.

29   Li, H.; Yang, Y.; Wang, H.; Li, B.; Wang, P.; Li, J.; Liao, H. Constructing a Spatiotemporally Coherent Long-Term PM$_{2.5}$ Concentration Dataset over China during 1980–2019 Using a Machine Learning Approach. *Sci. Total Environ.* **2021**, *765*, 144263. https://doi.org/10.1016/j.scitotenv.2020.144263.