Supporting Information

Diagnosing Ozone-NO_x-VOCs-Aerosols Sensitivity to Uncover Urban-nonurban Discrepancies in Shandong, China using Transformer-based High-resolution Air Pollution Estimations

Chenliang Tao[1], Yanbo Peng[1,2,], Qingzhu Zhang[1,*], Yuqiang Zhang[1], Bing Gong[3], Qiao Wang[1], Wenxing Wang[1]*

[1]Big Data Research Center for Ecology and Environment, Environmental Research Institute, Shandong University, Qingdao 266237, P.R. China

[2]Shandong Academy for Environmental Planning, Jinan 250101, P. R. China

[3]Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany

Keywords:

Air pollution, Deep learning, Transformer, Satellite, Urban-rural difference, Ozone Regime

Text S1 Variables Selected

Satellite data have been extensively used to derive surface air pollutant concentration[1,2]. The daily tropospheric NO_2 vertical column densities (VCDs) and O_3 total VCDs with a horizontal resolution of 5.5 × 3.5 km^2 were measured by TROPOMI. The daily AOD data and atmospheric properties with a 1 km resolution were obtained from MODIS Terra and Aqua combined multiangle implementation of atmospheric correction (MAIAC) land AOD product (MCD19A2)[3]. In addition, we used AOD estimates from the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) as the supplement of MODIS for filling extensively missing values. The meteorological reanalysis data were obtained from the fifth generation ECMWF reanalysis for the global climate and weather (ERA5) hourly products[4]. Ancillary data related to human activity and geographical information were retrieved and rasterized, including daily dynamic industrial emissions, moonlight-adjusted nighttime lights (NTL) product, population density, road density, land use data, the shuttle radar topography mission digital elevation model (DEM), the MOD13Q1 vegetation index (VI) product, and the MOD11A1 land surface temperature (LST) product. Industrial emissions amount (unit: kg) contains three categories, i.e. sulfur dioxide (SO_2), NO_x, and particulate matter (PM), collected from SDEM. Geographic covariates directly related to pollution emissions, such as industrial emission, and road density were decomposed into magnitude-related data by using Gaussian convolution kernels to account for the impact of neighboring sources (Text S2).

Text S2 Data Extension of Emission Proxies

The procedure of data extension follows from a previous study[5], geographic covariates directly related to pollution emissions like industrial emission, road density, and population density were decomposed into magnitude-related data by using Gaussian convolution kernels to account for the impact of neighboring sources. In this study, after rasterizing all spatial data to match with the tarted grid, the Gaussian convolution with the size of width   (ranging from 1.5 to 31.5 km) was used to consider the impact of nearby sources. For the Gaussian convoluted values with various at each location, the maximum value () was assigned as the characteristic magnitude of the emission proxy map for describing the influence of potential air pollution emission.

Text S3 Spatiotemporal Proxies

Taking the space-time-variant into consideration, three Euclidean spherical coordinates $(eqs\ 1-3)$ and three helix-shape trigonometric sequences $(eqs\ 4-6)$ were calculated as following:[6]

$$s_1 = \cos\left(2\pi\frac{longitude}{360}\right)\cos\left(2\pi\frac{latitude}{180}\right) \tag{1}$$

$$s_2 = \cos\left(2\pi\frac{longitude}{360}\right)\sin\left(2\pi\frac{latitude}{180}\right) \tag{2}$$

$$s_3 = \sin\left(2\pi\frac{longitude}{360}\right) \tag{3}$$

$$\text{cos\_sea} = \cos\left(2\pi\frac{month}{12}\right) \tag{4}$$

$$\text{sin\_sea} = \sin\left(2\pi\frac{month}{12}\right) \tag{5}$$

$$\text{cos\_mon} = \frac{month}{360} \tag{6}$$

Text S4 Data Fusion and Gap filling

Due to the various data sources and types, we bilinearly interpolated predictor variables to the targeted grid with 500 m resolution to harmonize with other data. The daily Ozone (O_3), fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2) concentrations were assigned to their overlay cells by spatial aggregation.

The detection of trace gases information below-cloud was prevented by the shielding of ubiquitous clouds in optical remote sensing images, causing the existence of gaps in satellite productions. We utilize the efficient machine-learning model, called Light Gradient Boosting Machine (LightGBM)[7], to fill the gaps in satellite data. LightGBM is designed to be distributed and efficient with the advantages of faster training speed and higher accuracy. Thus, it can impute a large dataset (1407 × 863 grids in the targeted resolution) with missing data in multiple variables using an iterative way. For each iteration, available daily satellite-based data are regarded as the observations, and the missing values are predicted by the LightGBM with meteorological reanalysis and geographical coordinates. The number of iterations corresponds to the number of satellite products with missing values. Here, the satellite-based production contains MOIDS AOD, TROPOMI NO_2 and O_3 column density, normalized difference vegetation index (NDVI); enhanced vegetation index (EVI), and land surface temperature (LST). Applying the model of filling missing values, the predictions of all variables are reliable, with the average coefficient of determination (R^2) values ranging from 0.87 to 0.99 in the validation set.

Text S5 The Detail of Air Transformer (AiT)

In this study, V, T, H and W are configured to 57, 8, 5 and 5, respectively, according to the number of chosen variables and the empirical range of time and spatial. The data size remains unchanged $V \times 8 \times 5 \times 5$ for the first AiT encoder blocks, while for the next 3 blocks, 2 blocks, and 1 block, the temporal dimensions and spatial window size are reduced by the convolutional embedded block, which includes convolution operation with $2 \times 2 \times 2$ filter with the stride of $2 \times 1 \times 1$, and the number of variables' channels is 64, 96 and 128, resulting in data size of $64 \times 4 \times 4 \times 4$, $96 \times 2 \times 3 \times 3$ and $128 \times 1 \times 2 \times 2$. The data dimensionality is transformed through a linear layer in decoder blocks.

We train AiT via backpropagation using an AdamW optimizer with a learning strategy of warmup, a learning rate of 0.0005 and a batch size of 256, and apply early stopping on the validation loss using patience of 300 epochs. we combat overfitting by dropout within each layer of linear and self-attention. A GeLU activation function is applied throughout the network. The loss function of mean squared error was applied to the errors for the computation of gradients in the optimization. The model is coded and trained using the Pytorch library. Before the data is fed into the model for training, it is normalized over the entire dataset. The total dataset for training and testing has 262,656 instances.

For sensitivity analysis, we first simply applied the image and video recognition Transformers for the estimation and also achieved good prediction performance (R^2 of 0.96 for O_3 in Timesformer). However, the spatial distribution of estimation exhibits severe "reticular phenomenon" (Figure S5). We briefly analyze the reasons why original Transformer-based models fell into trouble in terms of pollutant maps. Firstly, these original Transformer-based image models are purely based on pixel units for self-attention computation. Air pollution estimation often involves various features (satellite, meteorological, and emission proxies, etc.) [2,8–10], which is unlike image data with just three channels (red, green, and blue). These models overly focus on the correlation between neighboring grids and lack extraction of deep features, resulting in a discontinuous distribution of estimation for our study. Secondly, they paid attention to the full domain of pixels and there were no overlaps between samples, so only the encoder part of Transformer was used. Air quality estimation could be troubled by the overlearning of neighborhood features and extensive data duplication of adjacent samples when existing deep learning models are directly applied. Summarizing the above factors, we believe that it is necessary to build upon a tradeoff between the spatial distribution of estimations and the performance of the model.

Text S6 Multi-task Learning Strategy

It not only leverages large amounts of cross-task data but also benefits from a regularization effect that leads to more general representations to help adapt to estimating multiple pollutants simultaneously and efficiently,[11] and alleviating overfitting to a specific pollutant. As shown in the bottom right of Figure 1, the encoder and decoder blocks are shared across all predictions, while the last block is task-specific combining different estimations of $PM_{2.5}$, O_3, and NO_2. The shared blocks can take advantage of the interrelationship between different air pollutants by learning the intrinsic features of data. The task-specific blocks can capture the relevant information needed for the single task from extracted potential features of Transformer blocks.

Text S7 Method: Inferring Surface HCHO

Column-to-surface Conversion Factor

The satellite-derived surface HCHO concentrations (S_g) from Tropospheric Ozone Monitoring Instrument (TROPOMI) formaldehyde (HCHO) vertical columns density (VCD) by the simulated surface-to-column conversion factor method described in literatures[12,13]:

$$S_g = \frac{vV_M - V_M^{upper}}{V_M^{lower}} \times \frac{S_M}{V_M} \times V_g^- \tag{7}$$

where, S_g is the inferred surface level HCHO mixing ratio, S_M and V_M are the surface and tropospheric HCHO concentration, V_M^{lower} is the lower partial column, V_M^{upper} is the upper partial columns simulated by the CAM-Chem chemical transport model, V_g^- is the averaged tropospheric TROPOPMI HCHO VCD within the WRF-model, and v represents the satellite-observed sub-model-grid spatial variability calculated as:

$$v = \frac{V_g}{V_g^-} \tag{8}$$

where V_g is the tropospheric HCHO VCD in the TROPOMI grid. HCHO below the lower layer is considered to be well mixed in the vertical direction, and a large portion of HCHO (~70%) appears over the boundary layer, causing a nonhomogeneous distribution of upper partial columns. Therefore, in this study, the altitude where the HCHO partial column reaches the half maximum of its profile is regarded as the lower layer, following a previous study[12].

ECMWF Atmospheric Composition Reanalysis 4 (EAC4)

To derive the surface HCHO concentration, we used the European Centre for Medium-Range Weather Forecasts (ECMWF) Atmospheric Composition Reanalysis 4 (EAC4) at 0.75×0.75 horizontal resolution simulation with 25 vertical levels.[14] Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using a model of the atmosphere based on the laws of physics and chemistry. The monthly averaged field of EAC4 was used in our study.

Text S8 Cross Validation

The performance of our AiT model is evaluated through two cross-validation (CV) methods: out-of-sample 10-fold CV and out-of-site 10-fold CV. The out-of-sample CV, where all samples are randomly divided into 10 folds, saving one-fold for testing, is widely used for comparing measurements with the predictions of the out-of-bag sample. In addition, the generalization capability of spatial prediction at the location without monitors is evaluated by out-of-site CV, which randomly divides all sites into 10 subsets and then trains the model using four subsets and tests the model on the remaining subset.

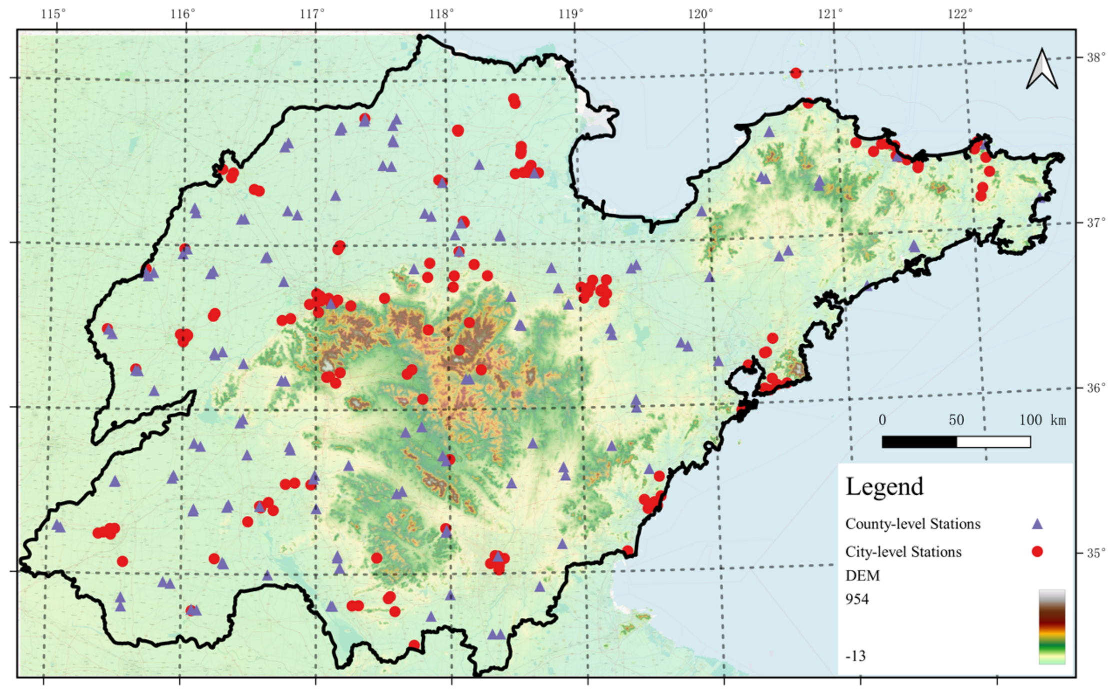

Figure S1. Map of study domain and location of monitoring stations. Purple triangles show the county-level air quality monitoring stations from SDEM, and red markers show the city-level air quality monitoring stations from CNEMC. The base map is the overlay of the © Google Maps and Digital Elevation Model (DEM) data.

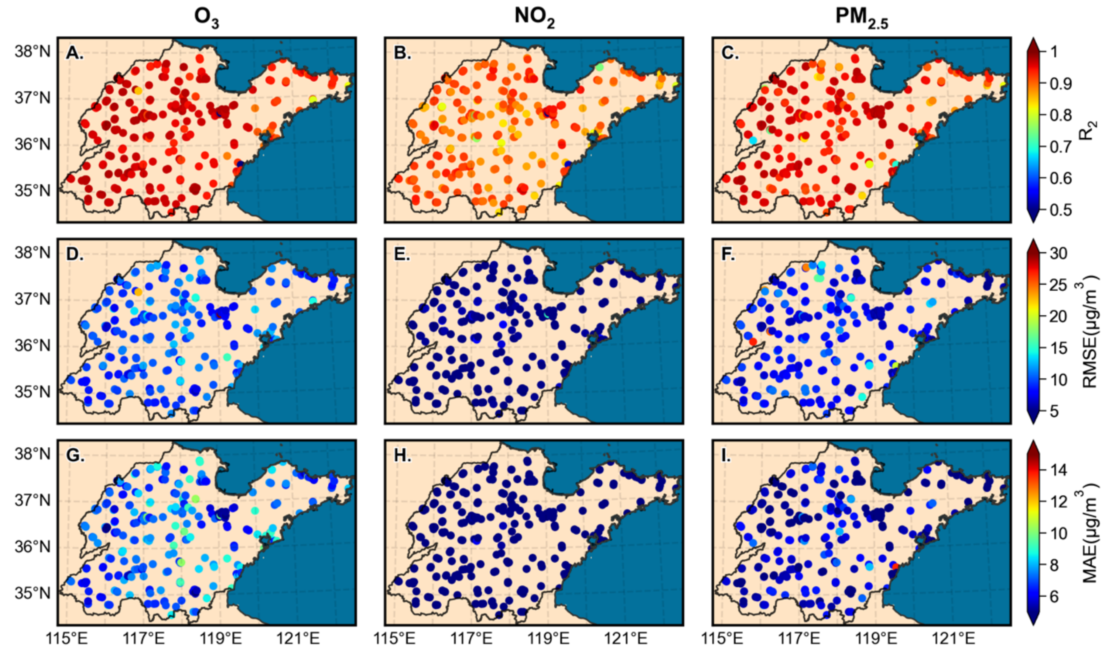

Figure S2. Out-of-sample cross-validation of daily surface O_3, NO_2 and O_3 estimates at each monitoring site.

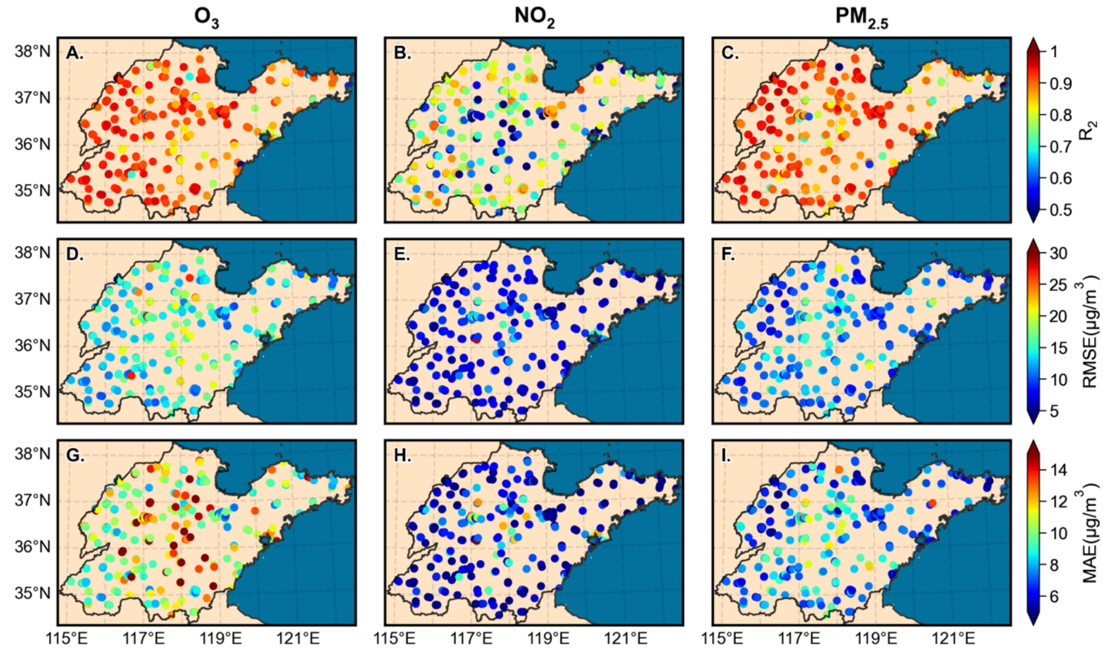

Figure S3. Out-of-site cross-validation of daily surface O_3, NO_2 and O_3 estimates at each monitoring site.

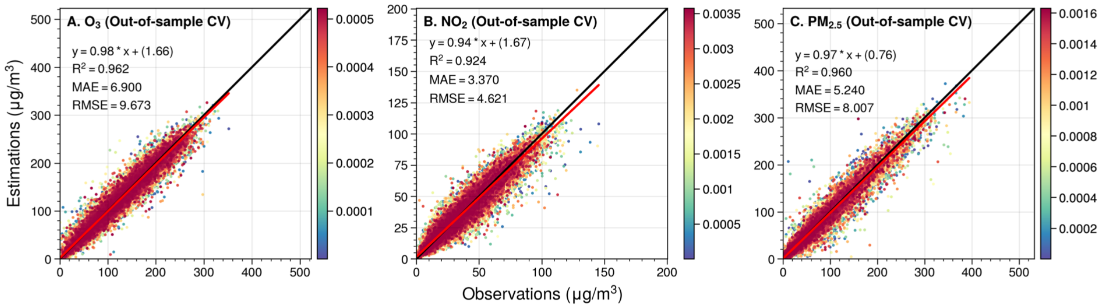

Figure S4. Out-of-sample cross-validation (A-C) of daily ground-level O_3, NO_2 and $PM_{2.5}$ concentration in the validation set based on the AiT model trained by monitoring data of CNEMC.

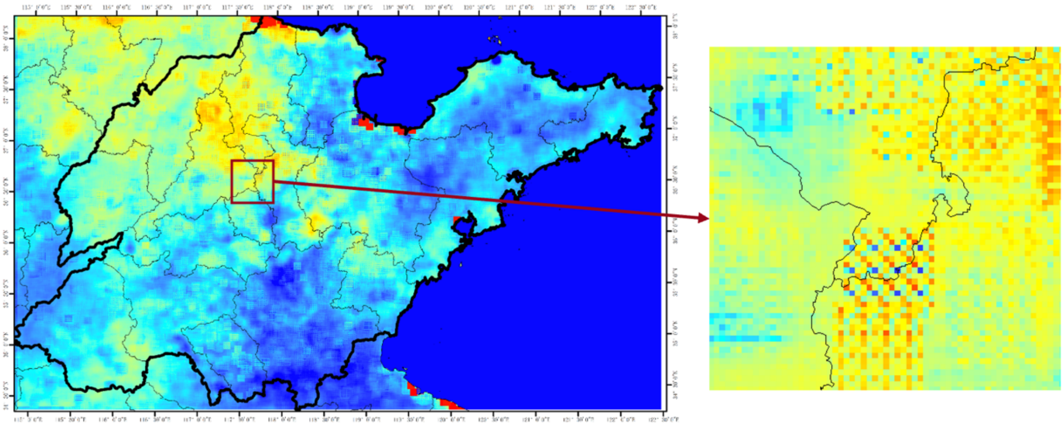

Figure S5. The estimated O_3 concentration on May 12, 2018, in Shandong, China using Timesformer (left) and also the zoomed-in map in region-scale distribution (right). The blue area represents the ocean.

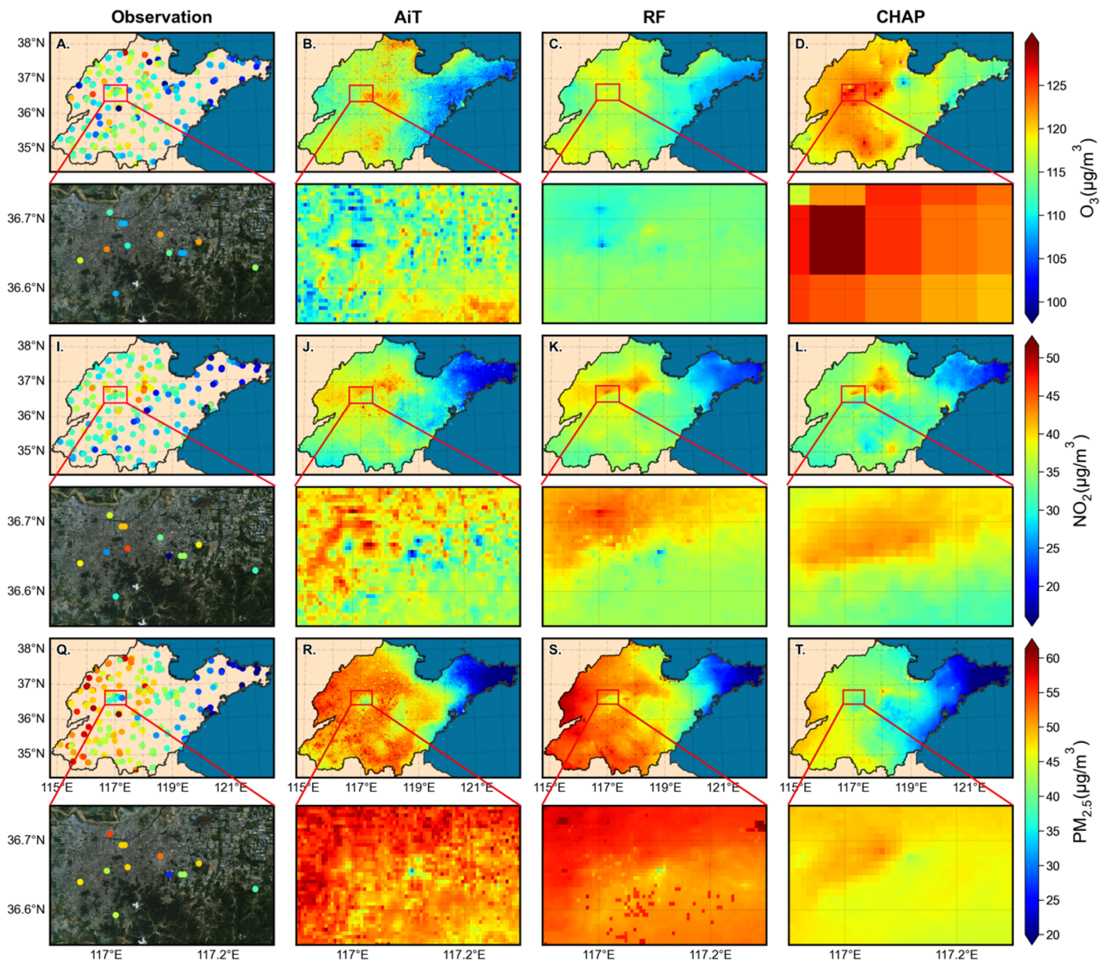

Figure S6. Spatial distribution of the annual mean (A-D) O_3, (I-L) NO_2 and (Q-T) $PM_{2.5}$ concentrations from observations, Air Transformer (AiT), Random Forest (RF) and ChinaHighAirPollutants (CHAP), respectively, in 2020. The region enclosed by the red rectangular box in (A-T) corresponds to the zoomed-in maps of satellite (© Tianditu: www.tianditu.gov.cn) and pollutant concentrations at a city scale for the capital city of Shandong Province, Jinan.

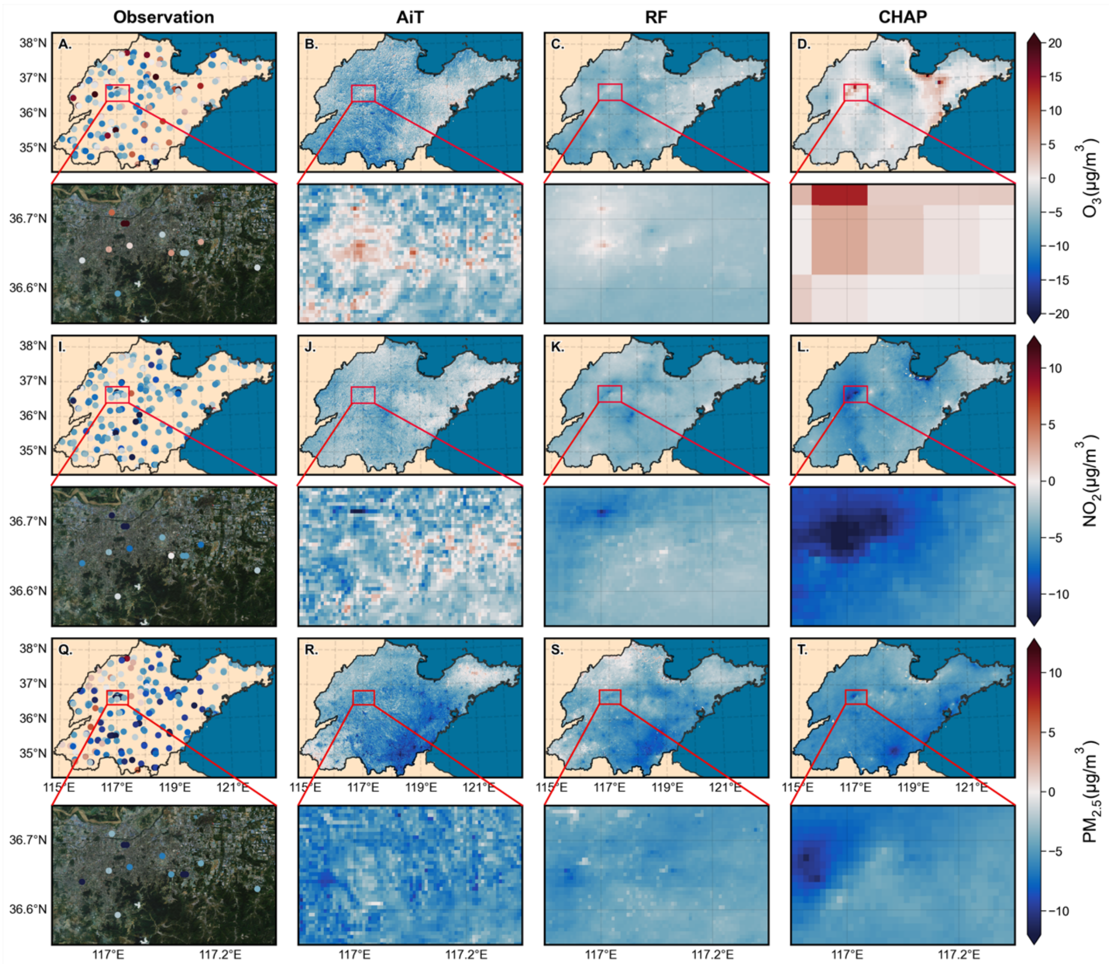

Figure S7. Spatial distribution of annual mean disparities for (A-D) O_3, (I-L) NO_2 and (Q-T) $PM_{2.5}$ concentrations from observations, Air Transformer (AiT), Random Forest (RF) and ChinaHighAirPollutants (CHAP), respectively, during 2019-2020. The region enclosed by the red rectangular box in (A-T) corresponds to the zoomed-in maps of satellite (© Tianditu: www.tianditu.gov.cn) and pollutant concentrations at a city scale for the capital city of Shandong Province, Jinan.

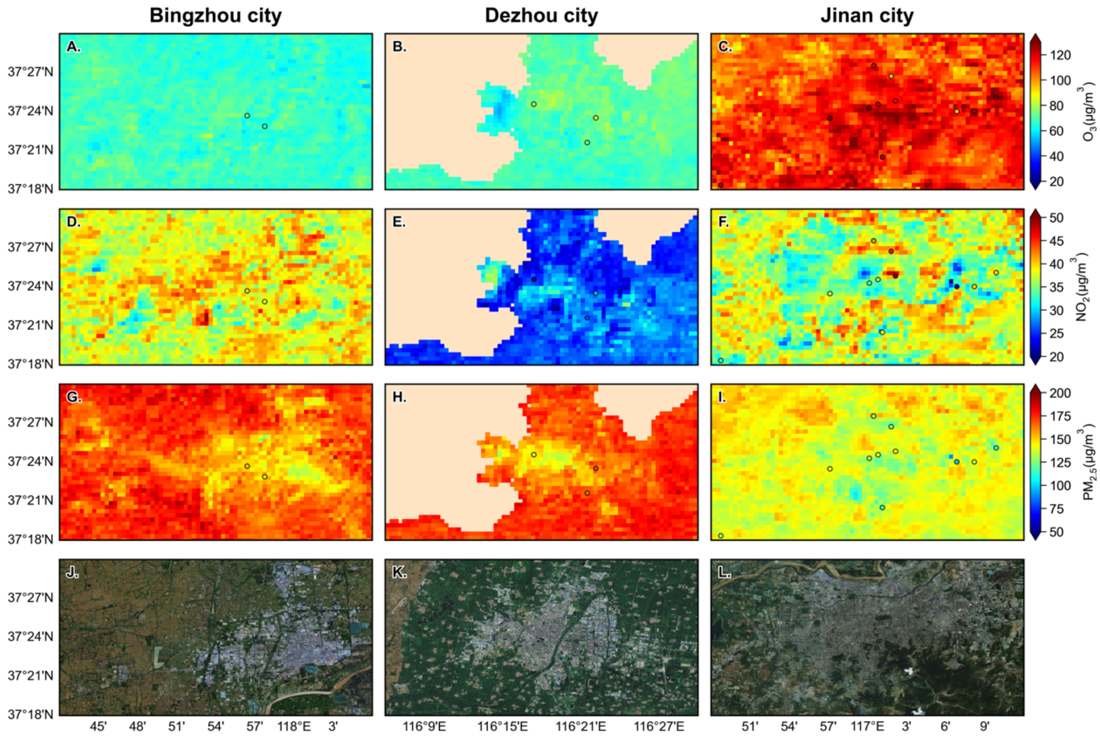

Figure S8. The spatial distribution of ground-level O_3 (A-C), NO_2 (D-F), and $PM_{2.5}$ (G-I) from AiT and monitoring stations in three cities experiencing diverse dust storm pollution on 15 March 2021 in Shandong, China. J-L represents the satellite maps of these cities (© Tianditu: www.tianditu.gov.cn).

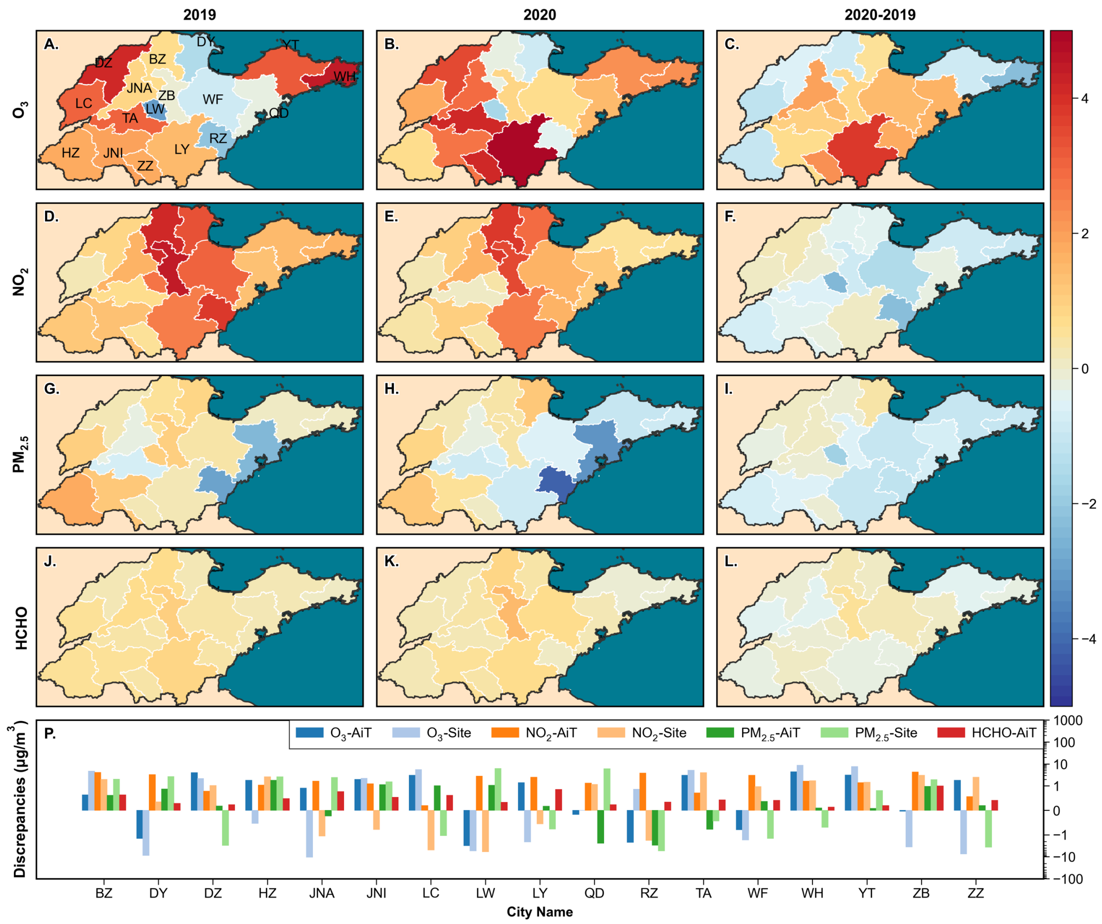

Figure S9. The urban-nonurban disparities of O_3, NO_2, $PM_{2.5}$ and HCHO calculated by AiT across cities with administrative divisions in Shandong, China during summer in 2019 (A, D, G) and 2020 (B, E, H), and the changes of differences between 2019 and 2020 (C, F, I).

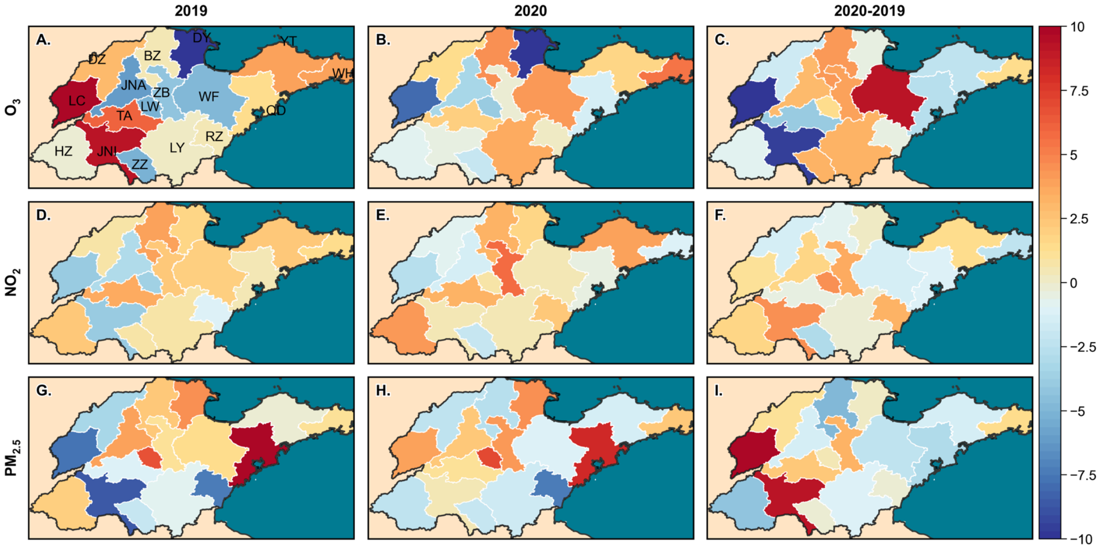

Figure S10. The urban-nonurban disparities of O_3, NO_2, and $PM_{2.5}$ were calculated by monitoring station data across cities in Shandong, China in 2019 (A, D, G) and 2020 (B, E, H), and the changes of differences between 2019 and 2020 (C, F, I).

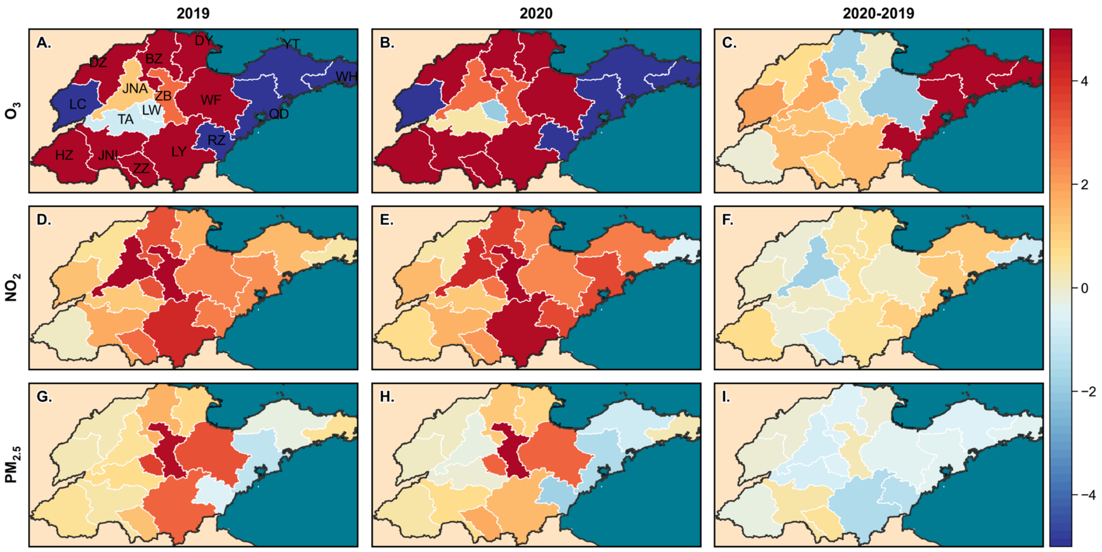

Figure S11. The urban-nonurban disparities of O_3, NO_2, and $PM_{2.5}$ calculated by CHAP across cities in Shandong, China in 2019 (A, D, G) and 2020 (B, E, H), and the changes of differences between 2019 and 2020 (C, F, I).

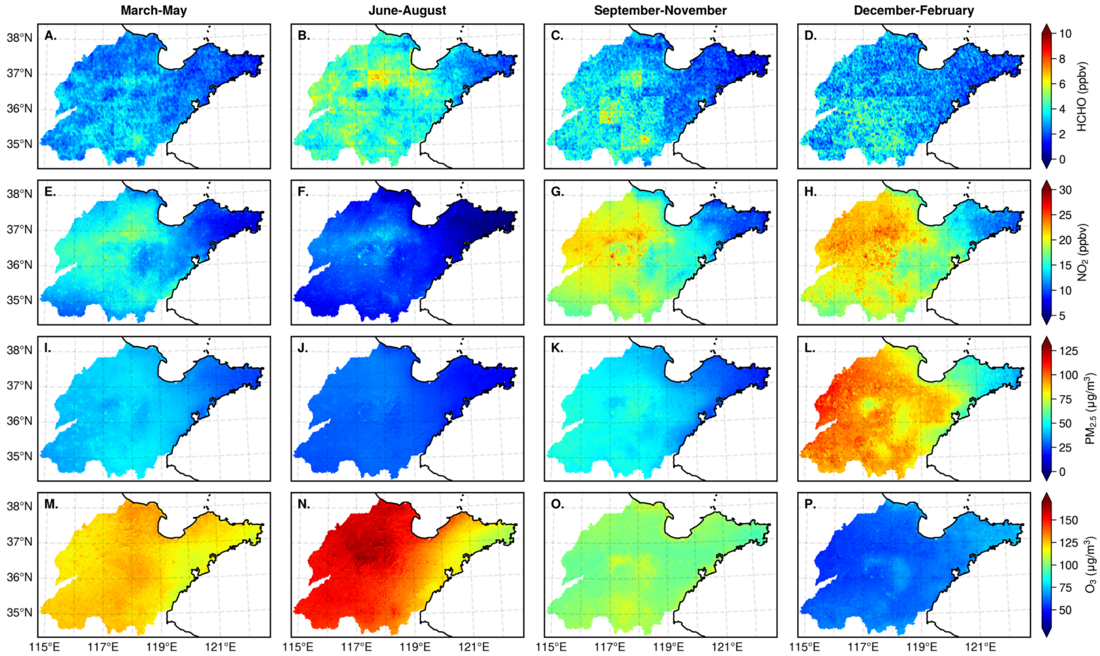

Figure S12. The seasonal changes of surface HCHO mixing ratio inferred from TROPOMI and EAC4 (A-D), and surface NO_2 (E-D), $PM_{2.5}$ (I-L) and O_3 (M-P) derived from Air Transformer across Shandong, China, in 2010 and 2020.

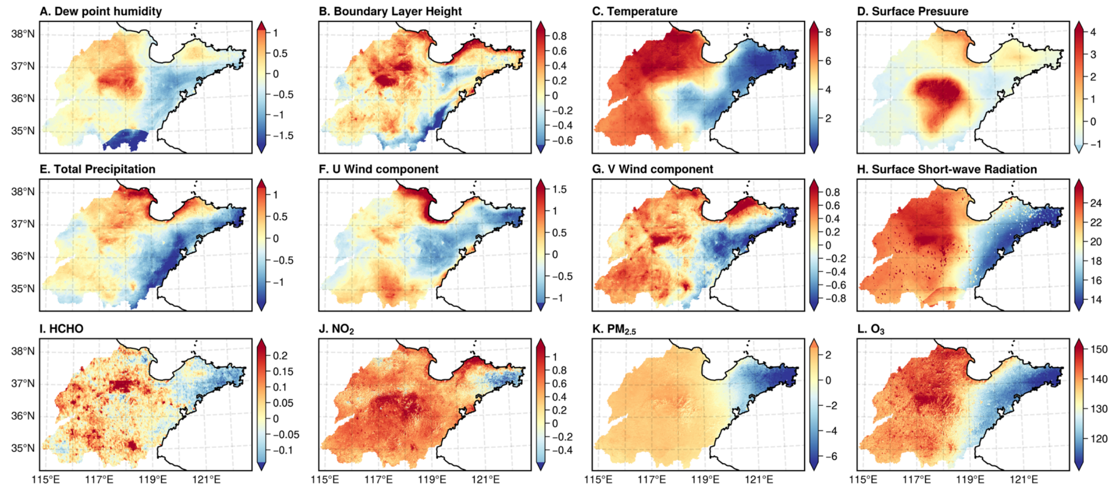

Figure S13. The geographical distribution of the averaged SHAP values for the important driving factors of O_3 production (A-K) in XGBoost model, and O_3 concentration (L) from May to October across Shandong, China in 2019 and 2020. The above color demonstrates how different variables each contribute to pushing the model output away from the base value (the average model output over the training dataset) towards the actual model output. Variables pushing the O_3 higher are shown in red, indicating they promote O_3 formation. In contrast, variables pushing the estimations lower are in blue, revealing they inhibit O_3 formation.

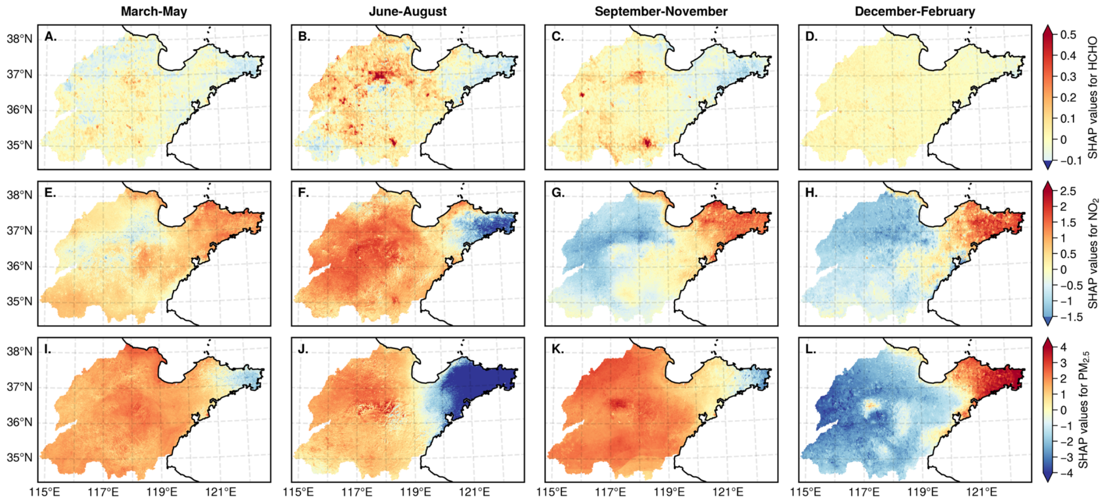

Figure S14. The seasonal changes of SHAP values in HCHO (A-D), NO_2 (E-H) and $PM_{2.5}$ (I-L) for O_3 formation across Shandong, China in 2019 and 2020.

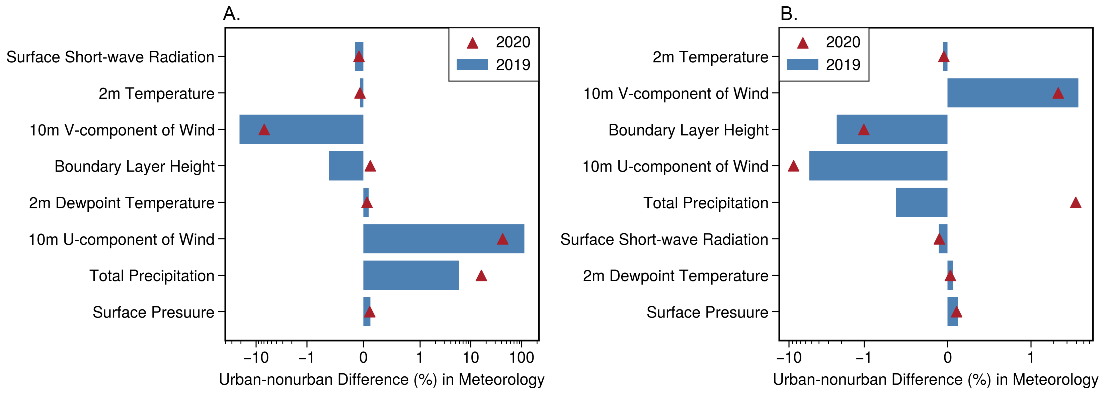

Figure S15. The changes in urban-nonurban discrepancies of meteorological conditions between 2019 and 2020 in Shandong, China during the lockdown periods (A) and summertime (B).

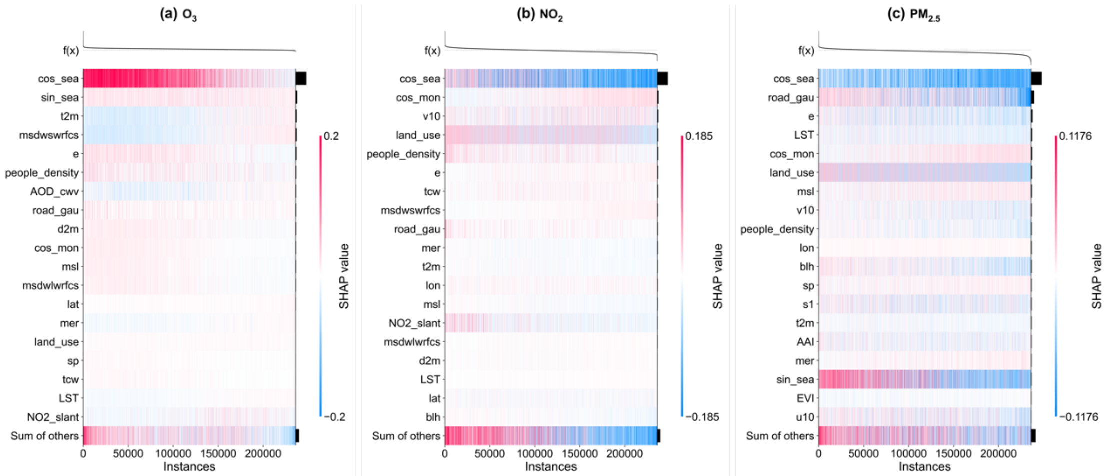

Figure S16. Contribution of each covariate to the near-surface O_3 (a), NO_2 (b), and $PM_{2.5}$ (c) concentration quantified with the Shapley Additive explanations (SHAP) method in the training dataset. The estimations of the model are shown above the heatmap matrix and the global importance of each model input is shown as a bar plot on the right side of the plot. The top fifteen variables of global importance are listed in order from top to bottom. The abbreviation of "people_density", "road_gau", and "land_use" represents the people density, road density and land use data, respectively. Another full form of the abbreviation can be found in Text S2 and Table S1.

Table S1. Summary of the dataset used in Air Transformer from multiple sources*

Data category	Data name	Spatial resolution	Temporal resolution	Data source
Ground observation	O_3、NO_2、$PM_{2.5}$ measurements	Point	Hourly	http://www.sdem.org.cn http://www.cnemc.cn
Satellite data	TROPOMI O_3, NO_2 [1]	5.5 × 3.5 km [2]	Daily	https://scihub.copernicus.eu
	MAIAC AOD [3]	1 × 1 km	Daily	https://lpdaac.usgs.gov/products/mcd19a2v006/
Meteorological fields	ERA5 [4]	0.25° × 0.25°	Hourly	https://cds.climate.copernicus.eu
Ancillary data	Industry emission	Point	Hourly	http://www.sdem.org.cn
	Land use	30 × 30 m	-	http://www.globallandcover.com
	People density	100m	-	https://hub.worldpop.org
	Road density	0.5 × 0.5 km	-	https://www.openstreetmap.org
	Digital elevation model (DEM)	0.5 × 0.5 km	-	https://www.resdc.cn
	MODIS vegetation index [5]	0.25 × 0.25 km	16-daily	https://lpdaac.usgs.gov/products/mod13q1v061/
	Nighttime lights (NTL)	0.5 × 0.5 km	Daily	https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/VNP46A2/
	Land surface temperature (LST)	1 × 1 km	Daily	https://e4ftl01.cr.usgs.gov
	MERRA-2 AOD reanalysis [6]	0.625° × 0. 5°	3-hourly	https://disc.gsfc.nasa.gov/datasets/M2I3NXGAS_5.12.4/summary
Spatial-temporal information	Euclidean spherical coordinates Temporal trend [7]	-	-	-

* The dataset covers the Shandong province of China from May 1, 2018 to July 1, 2021.

[1] TROPOMI satellite data contains: Tropospheric NO_2 column density (NO2); Total O_3 column density (O3); NO_2 slant columns density (NO2_slant); Absorbing aerosol index (AAI); cloud fraction. The Level-2 data from TROPOMI were filtered based on quality assurance values (>0.5).

[2] 7.5 × 3.5 km from 30. May 2018 to 6. August 2019.

[3] MAIAC AOD data including Aerosol Optical Depth (AOD) and column water vapor over land and clouds (AOD_cwv). The AOD was calculated by averaging the AOD at 0.47 µm and 0.55 µm. MAIAC AOD has better accuracy in the brighter areas[15] compared with AOD products generated from the Deep Blue[16] or Dark Target algorithms[17].

[4] ERA5 hourly data on single levels (reanalysis). It contains 18 variables: 10 meter U wind component (u10), 10 meter V wind component (v10), 2 meter dewpoint temperature (d2m); 2 meter temperature (t2m); Boundary layer height (blh); Evaporation (e); Total precipitation (tp); Surface pressure (sp); Boundary layer dissipation; Cloud base height; Low vegetation cover; Forecast albedo; Instantaneous large-scale surface precipitation fraction; Medium cloud cover; Mean evaporation rate (mer); Mean surface downward long-wave radiation flux, clear sky (msdwlwrfcs); Mean surface downward short-wave radiation flux, clear sky (msdwswrfcs); Mean sea level pressure (msl); Total columns ozone; Total columns water (tcw).

[5] MODIS vegetation index contains: Normalized Difference Vegetation Index (NDVI); Enhanced Vegetation Index (EVI).

[6] MERRA-2 AOD reanalysis contains: Aerosol Optical Depth Analysis, Aerosol Optical Depth Analysis Increment.

[7] Temporal trends contain: Helix-shape trigonometric month sequence; Julian day; Year; Month. One-hot encoding was used to process categorical variables.

Table S2. The performances of AiT in estimating multiple targeted pollutants as well as single targeted pollutants. All four model was trained using the same input dataset, but different targets (The targets of AiT is O_3, NO_2, $PM_{2.5}$. The target of AiT_O_3, AiT_NO_2, AiT_$PM_{2.5}$ is O_3, NO_2 and $PM_{2.5}$, respectively).

Model	AiT			AiT_O_3	AiT_NO_2	AiT_$PM_{2.5}$
	O_3	NO_2	$PM_{2.5}$	O_3	NO_2	$PM_{2.5}$
R^2	0.96	0.92	0.90	0.97	0.92	0.90
RMSE (μg/m^3)	9.96	4.72	11.99	9.27	4.75	12.57
MAE (μg/m^3)	7.06	3.48	5.38	6.35	3.46	6.14

Table S3. Comparison of model performance with previous studies.

Model	Spatial resolution	Cross-validation		Pollutant	Literature
		R^2	RMSE (μg/m³)		
RF	0.05°	0.87	13.03	O_3	Zhu et al., 2022[18]
STET	0.1°	0.87	17.1	O_3	Wei et al., 2022[19]
LSTM	0.1°	0.94	10.64	O_3	Wang et al., 2022[20]
DP	0.003°	0.94	11.29	O_3	Li et al., 2022[10]
LightGBM	0.05°	0.91	14.14	O_3	Wang et al., 2021[2]
XGBoost	0.05°	0.83	7.58	NO_2	Liu, 2021[21]
LightGBM	0.05°	0.83	6.62	NO_2	Wang et al., 2021[2]
GTWR-SK	0.025°	0.84	6.70	NO_2	Wu et al., 2021[22]
FSDN	0.01°	0.82	8.80	NO_2	Li & Wu, 2021[23]
SWDF	0.01°	0.93	4.89	NO_2	Wei et al., 2022[24]
DP	0.04°	0.88	11.27	$PM_{2.5}$	Song et al., 2022[1]
DEML	0.01°	0.87	5.38	$PM_{2.5}$	Yu et al., 2022[25]
RF	0.1°	0.83	13.9−22.1	$PM_{2.5}$	Geng et al., 2021[26]
STET	0.01°	0.89	10.33	$PM_{2.5}$	Wei et al., 2020[9]
RF	0.01°	0.88	15.73	$PM_{2.5}$	Huang et al., 2021[27]
		0.90	15.5	O_3	
RF*	0.005°	0.82	7.2	NO_2	This study
		0.92	10.72	$PM_{2.5}$	
		0.96	**10.11**	O_3	
AiT	0.005°	**0.92**	**4.82**	NO_2	This study
		0.95	**8.54**	$PM_{2.5}$	

STET: Space-time extremely randomized trees; LSTM: Long short-term memory network; DP: deep forest; semi-SILDM: tree-based ensemble deep learning model; LightGBM: Light gradient boosting machine; XGBoost: Extreme gradient boosting; GTWR-SK: Geographically and temporal weighted regression with spatiotemporal kriging; SFDN: Full residual deep networks; SWDF: Spatiotemporally weight deep forest; DEML: deep ensemble machine learning; RF: random forest; AiT: Air Transformer.

*: While training RF with variables involving neighboring grids is necessary, ML models are limited to accepting only one-dimensional data. Flattening four-dimensional data $(X \in R^{57\times8\times5\times5})$ causes a significant increase in the number of features, which results in a reduction in model performance. Thus, to ensure optimal performance, only variables in situ were employed to train RF.

Table S4. The average concentration of four pollutants across urban and non-urban areas in 2019 and 2020.

Year	Type	O_3	NO_2	$PM_{2.5}$	HCHO
2019	Nonurban	141.1	24.7	33.3	3.5
	Urban	141.1	26.3	32.6	4.2
2020	Nonurban	129.2	24.2	30.8	3.3
	Urban	130.4	25.4	29.5	4.0
Relative Changes (%)	Nonurban	-8.43	-2.02	-7.51	-5.71
	Urban	-7.58	-3.42	-9.51	-4.76

References

(1) Song, Z.; Chen, B.; Zhang, P.; Guan, X.; Wang, X.; Ge, J.; Hu, X.; Zhang, X.; Wang, Y. High Temporal and Spatial Resolution PM2.5 Dataset Acquisition and Pollution Assessment Based on FY-4A TOAR Data and Deep Forest Model in China. *Atmospheric Research* **2022**, *274*, 106199. https://doi.org/10.1016/j.atmosres.2022.106199.

(2) Wang, Y.; Yuan, Q.; Li, T.; Zhu, L.; Zhang, L. Estimating Daily Full-Coverage near Surface O_3, CO, and NO_2 Concentrations at a High Spatial Resolution over China Based on S5P-TROPOMI and GEOS-FP. *ISPRS Journal of Photogrammetry and Remote Sensing* **2021**, *175*, 311–325. https://doi.org/10.1016/j.isprsjprs.2021.03.018.

(3) Lyapustin, A.; Wang, Y. MODIS/Terra+Aqua Land Aerosol Optical Depth Daily L2G Global 1km SIN Grid V061. *NASA EOSDIS Land Processes DAAC* **2022**. https://doi.org/10.5067/MODIS/MCD19A2.061.

(4) Hersbach, H.; Bell, B.; Berrisford, G.; Horányi, A.; Muñoz Sabater, J.; Nicolas, J.; Peubey, C.; Rozum, I.; Schepers, D.; Simmons, A.; Soci, C.; Dee, D.; Thépaut, J.-N. ERA5 Hourly Data on Single Levels from 1959 to Present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)* **2023**. https://doi.org/10.24381/cds.adbb2d47.

(5) Kim, M.; Brunner, D.; Kuhlmann, G. Importance of Satellite Observations for High-Resolution Mapping of near-Surface NO2 by Machine Learning. *Remote Sensing of Environment* **2021**, *264*. https://doi.org/10.1016/j.rse.2021.112573.

(6) Sun, H.; Shin, Y. M.; Xia, M.; Ke, S.; Wan, M.; Yuan, L.; Guo, Y.; Archibald, A. T. Spatial Resolved Surface Ozone with Urban and Rural Differentiation during 1990–2019: A Space–Time Bayesian Neural Network Downscaler. *Environ. Sci. Technol.* **2022**, *56* (11), 7337–7349. https://doi.org/10.1021/acs.est.1c04797.

(7) Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems 30* **2017**, 3149-3157.

(8) Zang, Z.; Guo, Y.; Jiang, Y.; Zuo, C.; Li, D.; Shi, W.; Yan, X. Tree-Based Ensemble Deep Learning Model for Spatiotemporal Surface Ozone (O_3) Prediction and Interpretation. *International Journal of Applied Earth Observation and Geoinformation* **2021**, *103*, 102516. https://doi.org/10.1016/j.jag.2021.102516.

(9) Wei, J.; Li, Z.; Cribb, M.; Huang, W.; Xue, W.; Sun, L.; Guo, J.; Peng, Y.; Li, J.; Lyapustin, A.; Liu, L.; Wu, H.; Song, Y. Improved 1 Km Resolution $PM_{2.5}$ Estimates across China Using Enhanced Space–Time Extremely Randomized Trees. *Atmos. Chem. Phys.* **2020**, *20* (6), 3273–3289. https://doi.org/10.5194/acp-20-3273-2020.

(10) Li, M.; Yang, Q.; Yuan, Q.; Zhu, L. Estimation of High Spatial Resolution Ground-Level Ozone Concentrations Based on Landsat 8 TIR Bands with Deep Forest Model. *Chemosphere* **2022**, *301*, 134817. https://doi.org/10.1016/j.chemosphere.2022.134817.

(11) Liu, X.; He, P.; Chen, W.; Gao, J. Multi-Task Deep Neural Networks for Natural Language Understanding. arXiv May 29, 2019. http://arxiv.org/abs/1901.11504 (accessed 2022-11-22).

(12) Su, W.; Hu, Q.; Chen, Y.; Lin, J.; Zhang, C.; Liu, C. Inferring Global Surface HCHO Concentrations from Multisource Hyperspectral Satellites and Their Application to HCHO-Related Global Cancer Burden Estimation. *Environment International* **2022**, *170*, 107600. https://doi.org/10.1016/j.envint.2022.107600.

(13) Cooper, M. J.; Martin, R. V.; McLinden, C. A.; Brook, J. R. Inferring Ground-Level Nitrogen Dioxide Concentrations at Fine Spatial Resolution Applied to the TROPOMI Satellite Instrument. *Environ. Res. Lett.* **2020**, *15* (10), 104013. https://doi.org/10.1088/1748-9326/aba3a5.

(14) Inness, A.; Ades, M.; Agustí-Panareda, A.; Barré, J.; Benedictow, A.; Blechschmidt, A.-M.; Dominguez, J. J.; Engelen, R.; Eskes, H.; Flemming, J.; Huijnen, V.; Jones, L.; Kipling, Z.; Massart, S.; Parrington, M.; Peuch, V.-H.; Razinger, M.; Remy, S.; Schulz, M.; Suttie, M. The CAMS Reanalysis of Atmospheric Composition. *Atmos. Chem. Phys.* **2019**, *19* (6), 3515–3556. https://doi.org/10/ghdkrm.

(15) Lyapustin, A.; Wang, Y.; Laszlo, I.; Kahn, R.; Korkin, S.; Remer, L.; Levy, R.; Reid, J. S. Multiangle Implementation of Atmospheric Correction (MAIAC): 2. Aerosol Algorithm. *Journal of Geophysical Research: Atmospheres* **2011**, *116* (D3). https://doi.org/10.1029/2010JD014986.

(16) Hsu, N. C.; Jeong, M.-J.; Bettenhausen, C.; Sayer, A. M.; Hansell, R.; Seftor, C. S.; Huang, J.; Tsay, S.-C. Enhanced Deep Blue Aerosol Retrieval Algorithm: The Second Generation. *Journal of Geophysical Research: Atmospheres* **2013**, *118* (16), 9296–9315. https://doi.org/10.1002/jgrd.50712.

(17) Kaufman, Y. J.; Wald, A. E.; Remer, L. A.; Gao, B.-C.; Li, R.-R.; Flynn, L. The MODIS 2.1-/Spl Mu/m Channel-Correlation with Visible Reflectance for Use in Remote Sensing of Aerosol. *IEEE Transactions on Geoscience and Remote Sensing* **1997**, *35* (5), 1286–1298. https://doi.org/10.1109/36.628795.

(18) Zhu, Q.; Bi, J.; Liu, X.; Li, S.; Wang, W.; Zhao, Y.; Liu, Y. Satellite-Based Long-Term Spatiotemporal Patterns of Surface Ozone Concentrations in China: 2005–2019. *Environ Health Perspect* **2022**, *130* (2), 027004. https://doi.org/10.1289/EHP9406.

(19) Wei, J.; Li, Z.; Li, K.; Dickerson, R. R.; Pinker, R. T.; Wang, J.; Liu, X.; Sun, L.; Xue, W.; Cribb, M. Full-Coverage Mapping and Spatiotemporal Variations of Ground-Level Ozone (O_3) Pollution from 2013 to 2020 across China. *Remote Sensing of Environment* **2022**, *270*, 112775. https://doi.org/10.1016/j.rse.2021.112775.

(20) Wang, S.; Mu, X.; Jiang, P.; Huo, Y.; Zhu, L.; Zhu, Z.; Wu, Y. New Deep Learning Model to Estimate Ozone Concentrations Found Worrying Exposure Level over Eastern China. *International Journal of Environmental Research and Public Health* **2022**, *19* (12), 7186. https://doi.org/10.3390/ijerph19127186.

(21) Liu, J. Mapping High Resolution National Daily NO_2 Exposure across Mainland China Using an Ensemble Algorithm. *Environmental Pollution* **2021**, *279*, 116932. https://doi.org/10.1016/j.envpol.2021.116932.

(22) Wu, S.; Huang, B.; Wang, J.; He, L.; Wang, Z.; Yan, Z.; Lao, X.; Zhang, F.; Liu, R.; Du, Z. Spatiotemporal Mapping and Assessment of Daily Ground NO2 Concentrations in China Using High-Resolution TROPOMI Retrievals. *Environmental Pollution* **2021**, *273*, 116456. https://doi.org/10.1016/j.envpol.2021.116456.

(23) Li, L.; Wu, J. Spatiotemporal Estimation of Satellite-Borne and Ground-Level NO_2 Using Full Residual Deep Networks. *Remote Sensing of Environment* **2021**, *254*, 112257. https://doi.org/10.1016/j.rse.2020.112257.

(24) Wei, J.; Liu, S.; Li, Z.; Liu, C.; Qin, K.; Liu, X.; Pinker, R. T.; Dickerson, R. R.; Lin, J.; Boersma, K. F.; Sun, L.; Li, R.; Xue, W.; Cui, Y.; Zhang, C.; Wang, J. Ground-Level NO_2

Surveillance from Space Across China for High Resolution Using Interpretable Spatiotemporally Weighted Artificial Intelligence. *Environ. Sci. Technol.* **2022**, acs.est.2c03834. https://doi.org/10.1021/acs.est.2c03834.

(25) Yu, W.; Li, S.; Ye, T.; Xu, R.; Song, J.; Guo, Y. Deep Ensemble Machine Learning Framework for the Estimation of PM2.5 Concentrations. *Environ Health Perspect* **2022**, *130* (3), 037004. https://doi.org/10.1289/EHP9752.

(26) Geng, G.; Xiao, Q.; Liu, S.; Liu, X.; Cheng, J.; Zheng, Y.; Xue, T.; Tong, D.; Zheng, B.; Peng, Y.; Huang, X.; He, K.; Zhang, Q. Tracking Air Pollution in China: Near Real-Time $PM_{2.5}$ Retrievals from Multisource Data Fusion. *Environ. Sci. Technol.* **2021**, *55* (17), 12106–12115. https://doi.org/10.1021/acs.est.1c01863.

(27) Huang, C.; Hu, J.; Xue, T.; Xu, H.; Wang, M. High-Resolution Spatiotemporal Modeling for Ambient $PM_{2.5}$ Exposure Assessment in China from 2013 to 2019. *Environ. Sci. Technol.* **2021**, *55* (3), 2152–2162. https://doi.org/10.1021/acs.est.0c05815.