# Supplementary Material

# First retrieval of 24-hourly 1-km-resolution gapless surface ozone (O₃) from space in China using artificial intelligence: diurnal variations and implications for air quality and phytotoxicity

Fan Cheng[1], Zhanqing Li[2*], Zeyu Yang[1], Ruohan Li[3], Dongdong Wang[3], Aolin Jia[4],

Ke Li[5], Bin Zhao[6,7], Shuxiao Wang[6,7], Dejia Yin[6,7], Shengyue Li[6,7],

Wenhao Xue[8], Maureen Cribb[2], Jing Wei[2*]


1. College of Global Change and Earth System Science, Faculty of Geographical Science, Beijing Normal University, Beijing, China
2. Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
3. Department of Geographical Sciences, University of Maryland, College Park, MD, USA
4. Department of Environment Research and Innovation, Luxembourg Institute of Science and Technology (LIST), Belvaux, Luxembourg
5. Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China
6. State Key Joint Laboratory of Environmental Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China
7. State Environmental Protection Key Laboratory of Sources and Control of Air Pollution Complex, Beijing 100084, China
8. School of Economics, Qingdao University, Qingdao, China




*Correspondence:

Zhanqing Li (zhanqing@umd.edu), and Jing Wei (weijing@umd.edu)

# Table of Contents

# Appendix Texts

## Text S1: Necessity of high-resolution surface $O_3$ data

The lifetime of surface $O_3$ varies considerably with location and season, ranging from a few hours in polluted urban regions up to a few weeks in the upper troposphere, as reported in many previous studies (Griffiths et al., 2021; Monks et al., 2015; Stevenson et al., 2006; Young et al., 2013). Additionally, $O_3$ exhibits a distinct chemical mechanism, notably its strong nonlinearity with $NO_X$ (Lin et al., 2018; Wang et al., 2017). For example, $O_3$ concentrations in a city center are typically lower than those in suburban areas (Xie et al., 2016), and concentrations around urban roads are generally lower than in other parts of the city center (Mavroidis and Ilia, 2012). Moreover, the substantial presence of background $O_3$ may result in smaller actual spatial disparities (Reid et al., 2008).

In our study, we examined the scale of $O_3$ variability and found the existence of strong $O_3$ heterogeneity by analyzing surface $O_3$ differences at multiple closely located sites in typical regions across China. The Kruskal-Wallis H-test and Dunn's test methods were utilized to assess the differences among these sites. When $p$-values are less than 0.05, this indicates significant differences between data from the two sites at a 95% significance level. Here, we have identified six densely populated areas with multiple closely situated ground monitoring stations across China (Figure S8) and computed the variances in surface $O_3$ levels observed at these locations in comparison to their nearby sites within the region (Figure S9). Results show that the majority of $p$-values between sites fall below 0.05, indicating pronounced spatial gradients in hourly $O_3$ concentrations among the closest sites within the $0.1° × 0.1°$ (approximately 10 km × 10 km) grid at the 95% confidence level. This underscores the importance of producing hourly $O_3$ data at a high spatial resolution (e.g., 1 km) because it facilitates the monitoring of fine-scale spatial variations in surface $O_3$ concentrations, especially in urban areas.
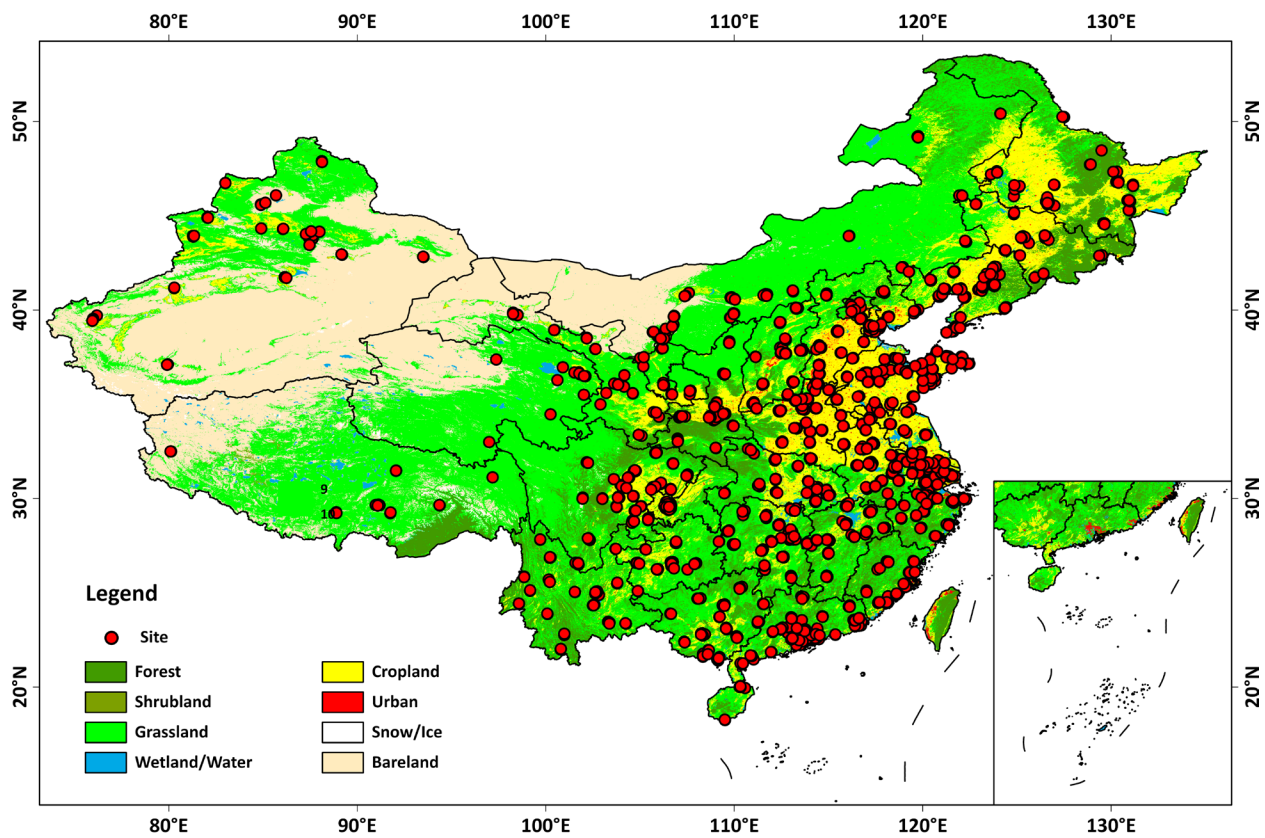
# Appendix Figures



**Figure S1.** Spatial distribution of ground-based surface O$_3$ monitoring stations in 2019 in China.
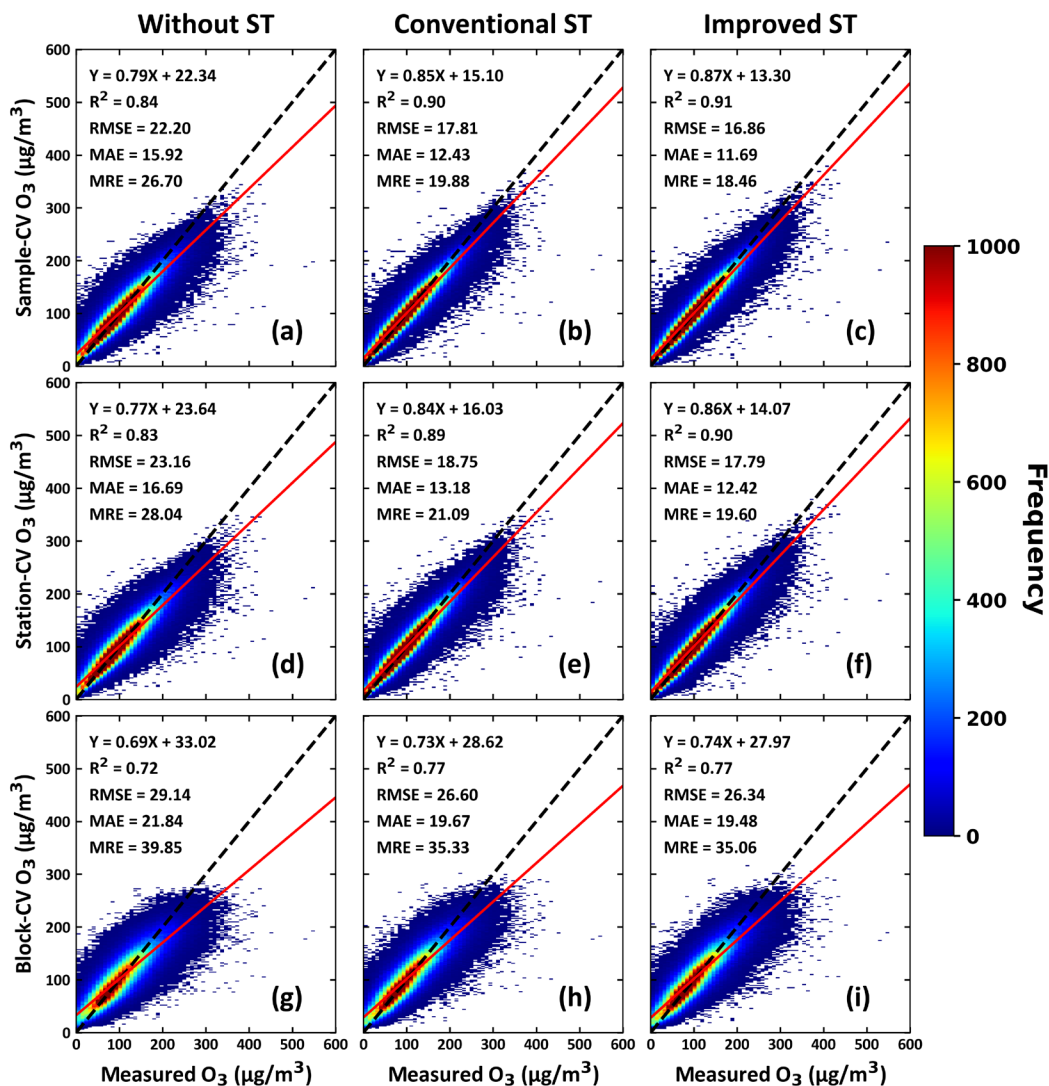
**Figure S2.** Comparisons of model performance (a, d, g) without, (b, e, h) with conventional, and (c, f, i) with improved spatial and temporal information for sample-CV (top row), station-CV (middle row), and block-CV (bottom row) results at 17:00 (Beijing Time) across China (sample size = 476,039). The linear-regression relation, coefficient of determination ($R^2$), root-mean-square error (RMSE, $\mu g/m^3$), mean absolute error (MAE, $\mu g/m^3$), and mean relative error (MRE, %) are also given.
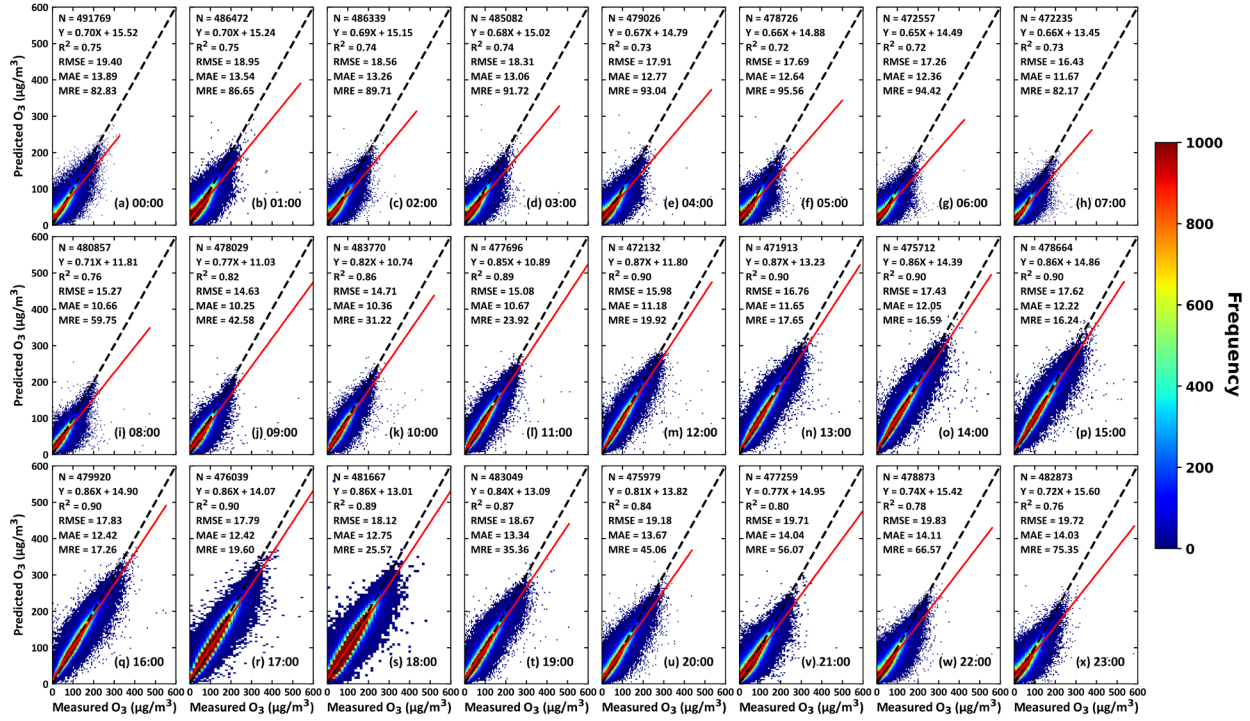
**Figure S3.** Out-of-station cross-validation results of hourly O₃ estimates (μg/m³) from 00:00 to 23:00 LT for 2019 in China using the 4D-STET model. Black dashed lines denote 1:1 lines, and red solid lines denote best-fit lines from linear regression. The sample size (N), linear-regression relation, coefficient of determination ($R^2$), root-mean-square error (RMSE, μg/m³), mean absolute error (MAE, μg/m³), and mean relative error (MRE, %) are also given.

**Figure S4.** Out-of-block cross-validation results of hourly $O_3$ estimates ($\mu g/m^3$) from 00:00 to 23:00 LT for 2019 in China using the 4D-STET model. Black dashed lines denote 1:1 lines, and red solid lines denote best-fit lines from linear regression. The sample size (N), linear-regression relation, coefficient of determination ($R^2$), root-mean-square error (RMSE, $\mu g/m^3$), mean absolute error (MAE, $\mu g/m^3$), and mean relative error (MRE, %) are also given.
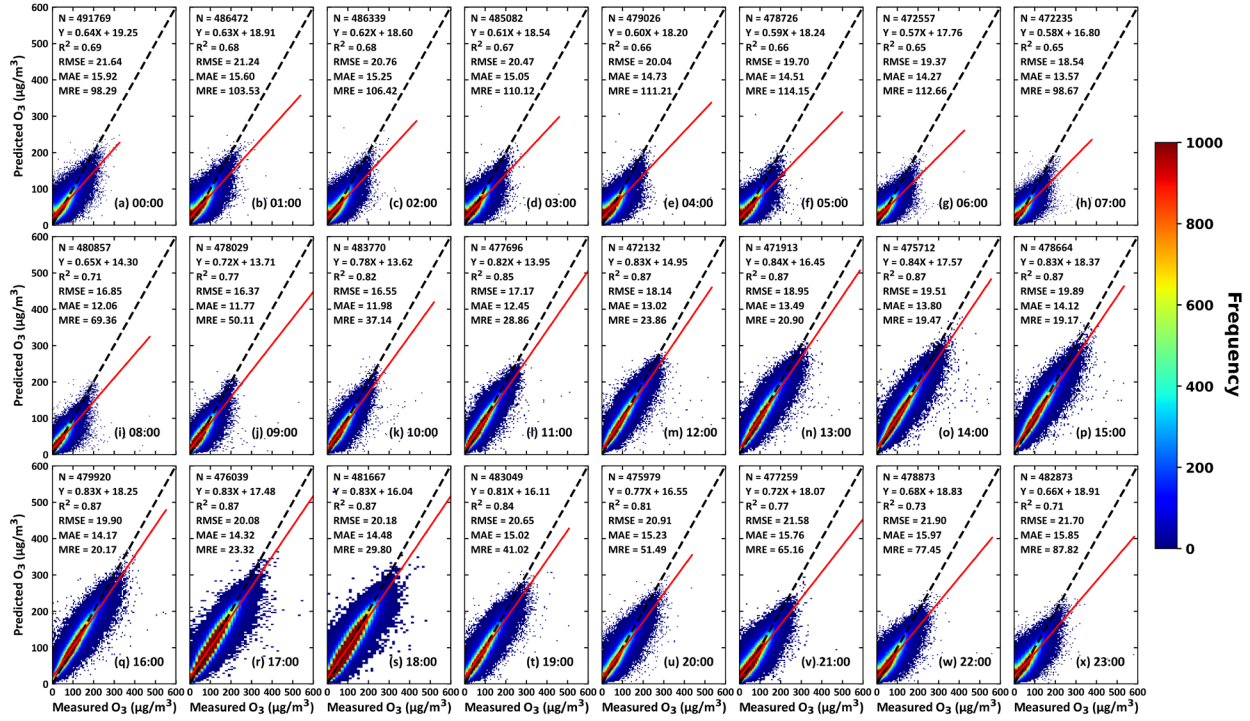
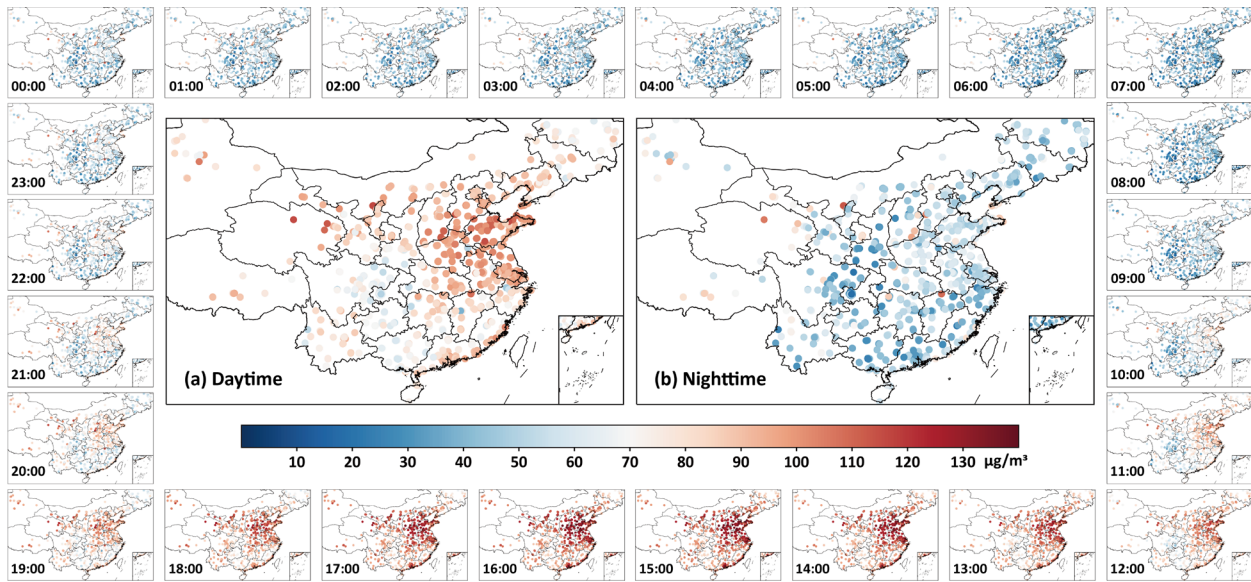**Figure S5.** Surface O₃ concentrations (μg/m³) monitored for each hour throughout the day (00:00–23:00 LT, surrounding subplots), along with average maps during the (a) daytime (08:00–20:00 LT) and (b) nighttime (20:00–08:00 LT) in 2019 across China.
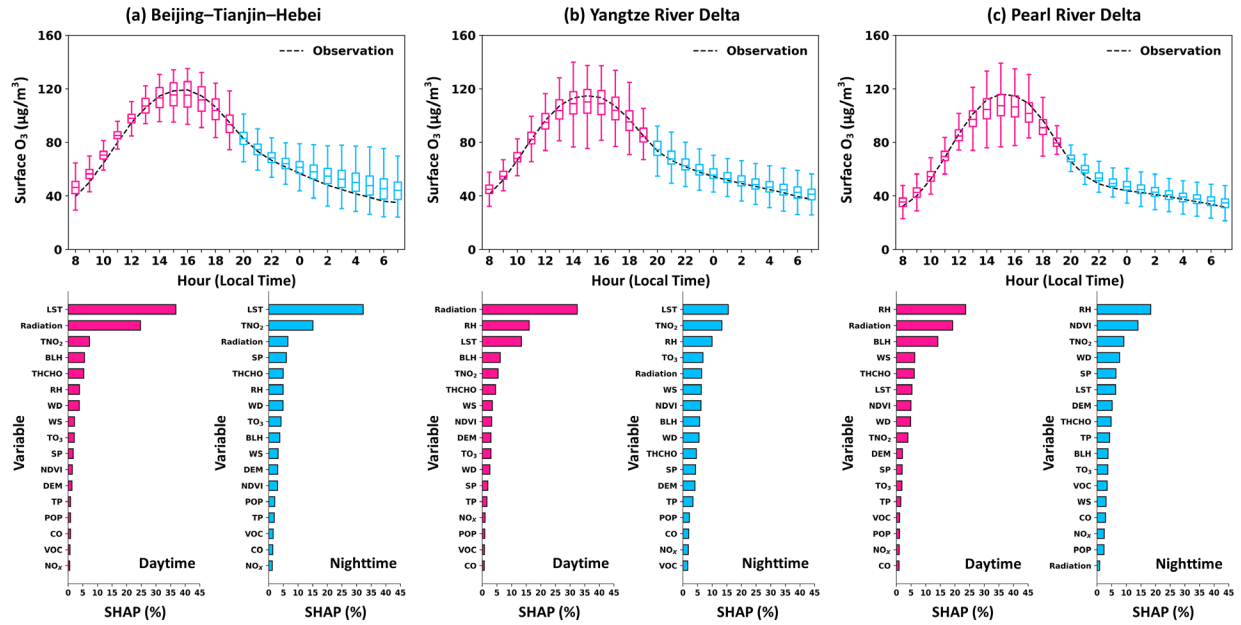
**Figure S6.** Boxplots of diurnal surface O₃ concentrations (top row; μg/m³) and the sorted SHAP importance of each variable during the daytime (08:00–20:00 LT) and nighttime (20:00–08:00 LT) in 2019 for (a) the Beijing–Tianjin–Hebei region, (b) the Yangtze River Delta region, and (c) the Pearl River Delta region. Black dashed lines represent the diurnal variations observed from ground measurements. In each box, the middle, lower, and upper horizontal black lines represent the mean bias, 25th percentile, and 75th percentile, respectively.
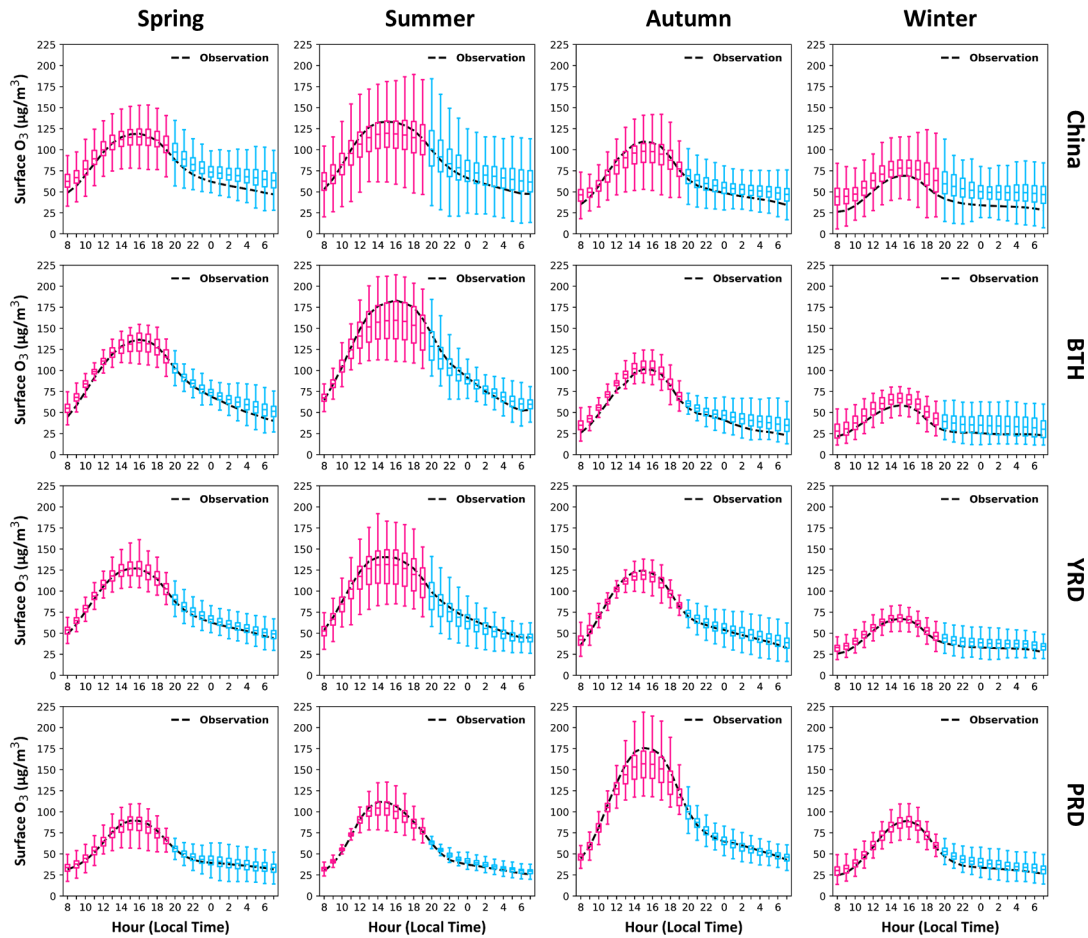
**Figure S7.** Boxplots of diurnal surface O$_3$ concentrations (μg/m³) covering the day for different seasons for China, the Beijing–Tianjin–Hebei (BTH) region, the Yangtze River Delta (YRD) region, and the Pearl River Delta (PRD) region. Black dashed lines represent the diurnal variations observed from ground measurements.

**Figure S8.** Selected typical densely populated areas with multiple closely located sites in China, where the number in brackets indicates the number of sites within the area.

**Figure S9.** Variances in surface O₃ concentrations among the sites within the six selected populated areas in China.

# Appendix Tables

**Table S1.** Summary of the main parameters of tree-based machine-learning models.

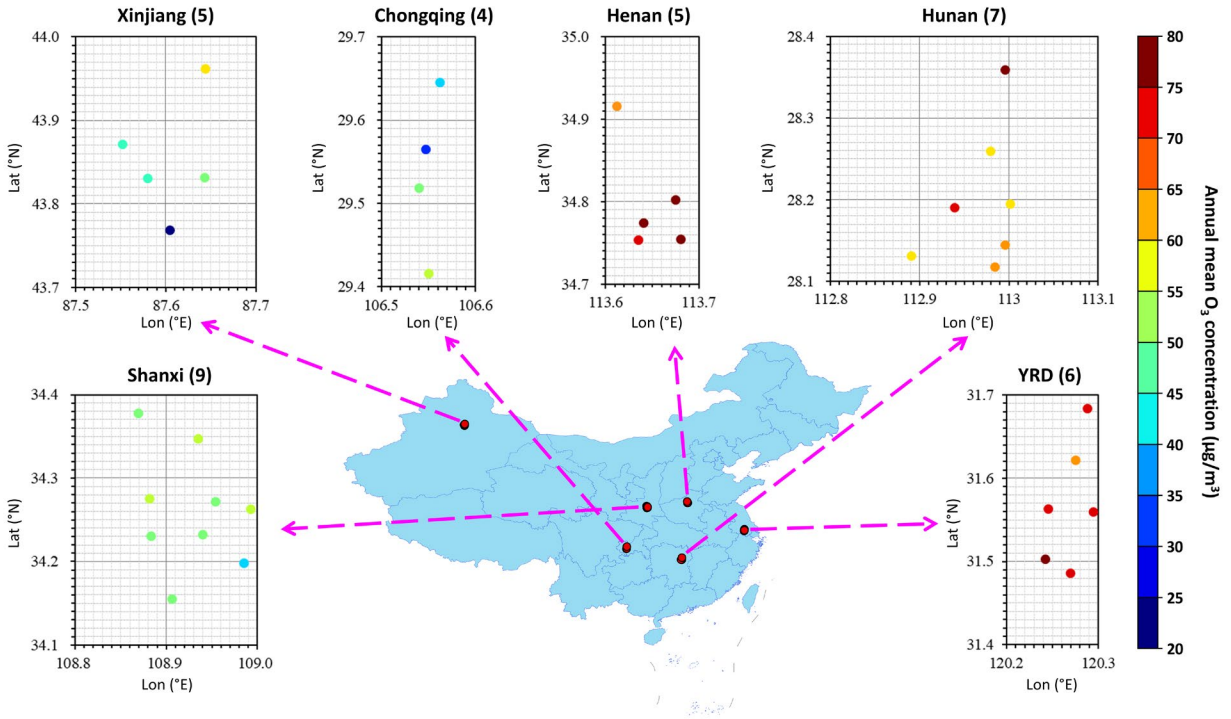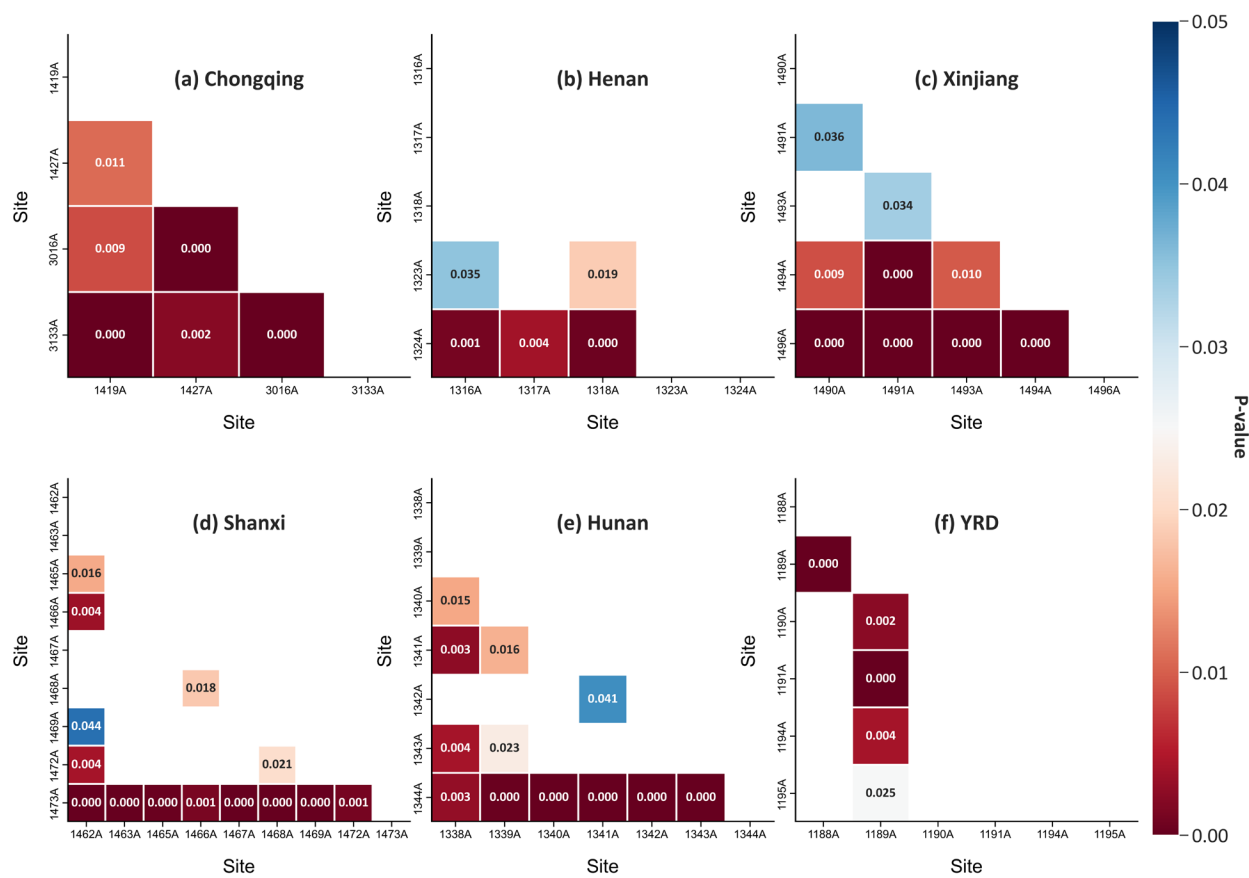| Model | Parameter | Value |
|---|---|---|
| AdaBoost | input_data | tabular data |
| | n_estimators | 200 |
| | learning_rate | 0.1 |
| | random_state | 0 |
| DT | input_data | tabular data |
| | min_samples_split | 2 |
| | random_state | 0 |
| GBDT | input_data | tabular data |
| | n_estimators | 200 |
| | alpha | 0.5 |
| | max_depth | 5 |
| | learning_rate | 0.1 |
| | min_samples_split | 2 |
| | random_state | 0 |
| | loss | 'squared_error' |
| CatBoost | input_data | tabular data |
| | iterations | 800 |
| | depth | 6 |
| | learning_rate | 0.1 |
| | random_state | 0 |
| XGBoost | input_data | tabular data |
| | objective | 'squarederror' |
| | n_estimators | 2000 |
| | max_depth | 5 |
| | subsample | 0.8 |
| | learning_rate | 0.1 |
| | colsample_bytree | 0.8 |
| | gamma | 0 |
| | reg_alpha | 0.1 |
| | reg_lambda | 0.1 |
| | random_state | 0 |
| RF | input_data | tabular data |
| | n_estimators | 200 |
| | min_samples_split | 2 |
| | random_state | 0 |
| LightGBM | input_data | tabular data |
| | objective | 'regression' |
| | num_leaves | 250 |
| | n_estimators | 2000 |
| | max_depth | 10 |
| | min_child_samples | 20 |
| | min_child_weight | 0.001 |
| | learning_rate | 0.1 |
| | bagging_fraction | 1.0 |
| | feature_fraction | 0.8 |
| | random_state | 0 |
| ET | input_data | tabular data |
| | estimators | 200 |
| | min_samples_split | 2 |
| | random_state | 0 |

| DF | input_data | tabular data |
| --- | --- | --- |
| | n_estimators | 2 |
| | n_trees | 100 |
| | max_layers | 20 |
| | criterion | 'squared_error' |
| | predictor | 'xgboost' |
| | random_state | 0 |

**Table S2.** Summary of the main parameters of deep-learning models used in this study.

| Model | Parameter | Value |
|---|---|---|
| MLP | input_data | tabular data |
| | input size | 25 |
| | hidden layers | 1 |
| | hidden size | 64 |
| | data normalization | 'StandardScaler' |
| | train epochs | 50 |
| | batch size | 64 |
| | activation function | Relu() |
| | output activation | Linear() |
| | learning rate | 0.001 |
| | optimizer | Adam() |
| | loss function | MAELoss() |
| DBN | input_data | tabular data |
| | input size | 25 |
| | RBM num | 2 |
| | hidden sizes | 30, 10 |
| | data normalization | 'StandardScaler' |
| | train epochs | 100 |
| | batch size | 64 |
| | activation function | Sigmoid() |
| | output activation | Linear() |
| | learning rate | 0.001 |
| | optimizer | Adam() |
| | loss function | MAELoss() |
| ResNet | input_data | tabular data |
| | input size | 25 |
| | Residual Block num | 2 |
| | data normalization | 'StandardScaler' |
| | train epochs | 50 |
| | batch size | 64 |
| | activation function | Relu() |
| | output activation | Linear() |
| | learning rate | 0.001 |
| | optimizer | Adam() |
| | loss function | MAELoss() |
| CNN | input_data | tabular data |
| | input channel | 1 |
| | input size | 25 |
| | kernel size | (1, 3) |
| | conv layers | 3 |
| | hidden size | 16, 32, 64 |
| | data normalization | 'StandardScaler' |
| | train epochs | 50 |
| | batch size | 64 |
| | activation function | Relu() |
| | output activation | Linear() |
| | learning rate | 0.001 |
| | optimizer | Adam() |
| | loss function | MAELoss() |
| LSTM | input_data | tabular data |
| | input size | 25 |
| | hidden layers | 3 |
| | hidden size | 90 |

| | | |
|---|---|---|
| | data normalization | 'StandardScaler' |
| | train epochs | 150 |
| | batch size | 256 |
| | output activation | Linear() |
| | learning rate | 0.001 |
| | optimizer | Adam() |
| | loss function | MAELoss() |
| ResNeXt | input_data | tabular data |
| | input size | 25 |
| | cardinality | 32 |
| | Residual Layer num | 4 |
| | Residual Block num | [3, 4, 6, 3] |
| | data normalization | 'StandardScaler' |
| | train epochs | 50 |
| | batch size | 64 |
| | activation function | Relu() |
| | output activation | Linear() |
| | learning rate | 0.001 |
| | optimizer | Adam() |
| | loss function | MAELoss() |

**Table S3.** Performance and efficiency comparisons of different 4-dimensional space-time (4D-ST) tree-based machine- and deep-learning models for estimating surface $O_3$ concentrations at 17:00 LT in China based on the station-CV approach.

| Category | Core model | $R^2$ | Slope | RMSE | MAE | Speed (s) | Memory (GB) |
|---|---|---|---|---|---|---|---|
| Tree-based machine learning | AdaBoost | 0.492 | 0.451 | 38.94 | 30.80 | 504.45 | 0.1068 |
| | DT | 0.725 | 0.855 | 30.12 | 20.22 | 12.16 | 0.0319 |
| | GBDT | 0.753 | 0.739 | 26.86 | 20.23 | 1000.68 | 0.1045 |
| | CatBoost | 0.783 | 0.769 | 25.19 | 18.96 | 21.24 | 0.1482 |
| | XGBoost | 0.815 | 0.803 | 23.27 | 17.40 | 247.53 | 0.1314 |
| | RF | 0.863 | 0.826 | 20.13 | 14.45 | 324.62 | 6.0102 |
| | LightGBM | 0.873 | 0.857 | 19.29 | 14.19 | 47.18 | 0.1495 |
| | ET | 0.875 | 0.792 | 20.36 | 14.64 | 64.81 | 10.3010 |
| | DF | 0.888 | 0.888 | 18.08 | 12.86 | 9491.03 | 11.3705 |
| Deep learning | DBN | 0.716 | 0.735 | 28.85 | 21.90 | 156.81 | 0.0624 |
| | MLP | 0.724 | 0.705 | 28.47 | 21.62 | 175.11 | 0.0930 |
| | CNN | 0.777 | 0.783 | 25.79 | 19.40 | 1282.64 | 0.1698 |
| | LSTM | 0.784 | 0.796 | 25.14 | 18.81 | 3719.45 | 0.1556 |
| | ResNet | 0.799 | 0.820 | 24.23 | 18.09 | 2628.15 | 1.0659 |
| | ResNeXt | 0.820 | 0.835 | 22.94 | 17.01 | 8144.82 | 1.1123 |

AdaBoost: Adaptive Boosting; DT: Decision Trees; GBDT: Gradient Boosting Decision Tree; CatBoost: Categorical Boosting; XGBoost: eXtreme Gradient Boosting; RF: Random Forest; LightGBM: Light Gradient Boosting Machine; ET: ExtraTrees; DF: Deep Forest; MLP: Multilayer Perceptron; DBN: Deep Belief Network; ResNet: Deep Residual Network; CNN: Convolutional Neural Network; LSTM: Long Short Term Memory; ResNeXt: ResNet Next.

**Table S4.** Performance and efficiency comparisons of different 4-dimensional space-time (4D-ST) tree-based machine- and deep-learning models for estimating surface $O_3$ concentrations at 17:00 LT in China based on the block-CV approach.

| Category | Core model | $R^2$ | Slope | RMSE | MAE | Speed (s) | Memory (GB) |
|---|---|---|---|---|---|---|---|
| Tree-based Machine Learning | DT | 0.555 | 0.749 | 39.55 | 29.06 | 12.78 | 0.0366 |
| | AdaBoost | 0.572 | 0.494 | 39.94 | 31.60 | 359.81 | 0.0106 |
| | XGBoost | 0.745 | 0.729 | 27.88 | 20.90 | 403.70 | 0.1265 |
| | GBDT | 0.747 | 0.716 | 28.55 | 21.01 | 894.83 | 0.1056 |
| | RF | 0.763 | 0.729 | 27.62 | 19.95 | 335.28 | 5.8400 |
| | LightGBM | 0.765 | 0.750 | 26.75 | 20.01 | 48.75 | 0.2034 |
| | CatBoost | 0.771 | 0.757 | 26.41 | 19.71 | 14.22 | 0.1939 |
| | ET | 0.774 | 0.737 | 26.34 | 19.48 | 67.44 | 7.4371 |
| | DF | 0.775 | 0.751 | 26.87 | 19.33 | 9914.96 | 10.6786 |
| Deep Learning | MLP | 0.714 | 0.694 | 30.33 | 22.69 | 365.61 | 0.0934 |
| | DBN | 0.731 | 0.716 | 29.50 | 21.82 | 628.19 | 1.9854 |
| | ResNet | 0.737 | 0.733 | 29.06 | 21.11 | 2453.79 | 1.0668 |
| | ResNeXt | 0.747 | 0.747 | 28.48 | 20.68 | 7165.49 | 1.2687 |
| | LSTM | 0.752 | 0.731 | 28.25 | 20.60 | 4367.56 | 0.1539 |
| | CNN | 0.754 | 0.770 | 28.12 | 20.38 | 3681.92 | 1.1686 |

AdaBoost: Adaptive Boosting; DT: Decision Trees; GBDT: Gradient Boosting Decision Tree; CatBoost: Categorical Boosting; XGBoost: eXtreme Gradient Boosting; RF: Random Forest; LightGBM: Light Gradient Boosting Machine; ET: ExtraTrees; DF: Deep Forest; MLP: Multilayer Perceptron; DBN: Deep Belief Network; ResNet: Deep Residual Network; CNN: Convolutional Neural Network; LSTM: Long Short Term Memory; ResNeXt: ResNet Next.

# Appendix References

Griffiths, P.T., Murray, L.T., Zeng, G., Shin, Y.M., Abraham, N.L., Archibald, A.T., Deushi, M., Emmons, L.K., Galbally, I.E., Hassler, B., Horowitz, L.W., Keeble, J., Liu, J., Moeini, O., Naik, V., O'Connor, F.M., Oshima, N., Tarasick, D., Tilmes, S., Turnock, S.T., Wild, O., Young, P.J., Zanis, P., 2021. Tropospheric ozone in CMIP6 simulations. Atmos. Chem. Phys. 21, 4187–4218. https://doi.org/10.5194/acp-21-4187-2021

Lin, Y., Zou, J., Yang, W., Li, C.-Q., 2018. A review of recent advances in research on $PM_{2.5}$ in China. Int. J. Environ. Res. Public Health 15, 438. https://doi.org/10.3390/ijerph15030438

Mavroidis, I., Ilia, M., 2012. Trends of $NO_x$, $NO_2$, and $O_3$ concentrations at three different types of air quality monitoring stations in Athens, Greece. Atmos. Environ. 63, 135–147. https://doi.org/10.1016/j.atmosenv.2012.09.030

Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K.S., Mills, G.E., Stevenson, D.S., Tarasova, O., Thouret, V., Von Schneidemesser, E., Sommariva, R., Wild, O., Williams, M.L., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. Atmos. Chem. Phys. 15, 8889–8973. https://doi.org/10.5194/acp-15-8889-2015

Reid, N., Yap, D., Bloxam, R., 2008. The potential role of background ozone on current and emerging air issues: An overview. Air Qual. Atmos. Health 1, 19–29. https://doi.org/10.1007/s11869-008-0005-z

Stevenson, D.S., Dentener, F.J., Schultz, M.G., Ellingsen, K., Noije, T.P.C. van, Wild, O., Zeng, G., Amann, M., Atherton, C.S., Bell, N., Bergmann, D.J., Bey, I., Butler, T., Cofala, J., Collins, W.J., Derwent, R.G., Doherty, R.M., Drevet, J., Eskes, H.J., Fiore, A.M., Gauss, M., Hauglustaine, D.A., Horowitz, L.W., Isaksen, I.S.A., Krol, M.C., Lamarque, J.-F., Lawrence, M.G., Montanaro, V., Müller, J.-F., Pitari, G., Prather, M.J., Pyle, J.A., Rast, S., Rodriguez, J.M., Sanderson, M.G., Savage, N.H., Shindell, D.T., Strahan, S.E., Sudo, K., Szopa, S., 2006. Multimodel ensemble simulations of present-day and near-future tropospheric ozone. J. Geophys. Res. Atmos. 111. https://doi.org/10.1029/2005JD006338

Wang, T., Xue, L., Brimblecombe, P., Lam, Y.F., Li, L., Zhang, L., 2017. Ozone pollution in China: a review of concentrations, meteorological influences, chemical precursors, and effects. Sci. Total Environ. 575, 1582–1596. https://doi.org/10.1016/j.scitotenv.2016.10.081

Xie, M., Zhu, K., Wang, T., Chen, P., Han, Y., Li, S., Zhuang, B., Shu, L., 2016. Temporal characterization and regional contribution to $O_3$ and $NO_x$ at an urban and a suburban site in Nanjing, China. Sci. Total Environ. 551–552, 533–545. https://doi.org/10.1016/j.scitotenv.2016.02.047

Young, P.J., Archibald, A.T., Bowman, K.W., Lamarque, J.-F., Naik, V., Stevenson, D.S., Tilmes, S., Voulgarakis, A., Wild, O., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W.J., Dalsøren, S.B., Doherty, R.M., Eyring, V., Faluvegi, G., Horowitz, L.W., Josse, B., Lee, Y.H., MacKenzie, I.A., Nagashima, T., Plummer, D.A., Righi, M., Rumbold, S.T., Skeie, R.B., Shindell, D.T., Strode, S.A., Sudo, K., Szopa, S., Zeng, G., 2013. Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP). Atmos. Chem. Phys. 13, 2063–2090. https://doi.org/10.5194/acp-13-2063-2013