

Article

Estimation of Near-Surface High Spatiotemporal Resolution Ozone Concentration in China Using Himawari-8 AOD

Yixuan Wang ^{1,2}, Chongshui Gong ^{1,2,*} , Li Dong ³  and Yue Huang ⁴

- ¹ Key Laboratory of Arid Climatic Change and Reducing Disaster of Gansu Province, Institute of Arid Meteorology, China Meteorological Administration, Lanzhou 730030, China; wangyx21@lzu.edu.cn
² Key Laboratory of Arid Climatic Change and Disaster Reduction, Institute of Arid Meteorology, China Meteorological Administration, Lanzhou 730030, China
³ Tianshui Meteorological Bureau, Tianshui 741000, China; dongl19@lzu.edu.cn
⁴ Key Laboratory of Environmental Pollution Control Technology of Zhejiang Province, Eco-Environmental Science Research & Design Institute of Zhejiang Province, Hangzhou 310013, China; huangy2018@lzu.edu.cn
* Correspondence: gongcs@iamcma.cn

Abstract: Near-surface ozone is a secondary pollutant, and its high concentrations pose significant risks to human and plant health. Based on an Extra Tree (ET) model, this study estimated near-surface ozone concentrations with the high spatiotemporal resolution based on Himawari-8 aerosol optical depth (AOD) data and meteorological variables from 1 January 2016 to 31 December 2020. The SHapley Additive exPlanation (SHAP) method was employed to evaluate the contribution of AOD and meteorological factors on ozone concentration. The results indicate that (1) the ET model achieves a sample-based cross-validation R^2 of 0.75–0.87 and an RMSE ($\mu\text{g}/\text{m}^3$) of 17.96–20.30. The coefficient of determination (R^2) values of the model in spring, summer, autumn, and winter are 0.81, 0.80, 0.87, and 0.75, respectively. (2) Higher temperature and boundary layer heights were found to positively contribute to ozone concentration, whereas higher relative humidity exerted a negative influence. (3) From 11:00 to 15:00 (Beijing time, UTC+08:00), ozone concentration increases gradually, with the highest occurring in the summer, followed by spring. This study has obtained high spatial and temporal resolution ozone concentration data, offering valuable insights for the development of fine-scale ozone pollution prevention and control strategies.



Academic Editor: Biao Zhang

Received: 3 December 2024

Revised: 17 January 2025

Accepted: 29 January 2025

Published: 4 February 2025

Citation: Wang, Y.; Gong, C.; Dong, L.; Huang, Y. Estimation of Near-Surface High Spatiotemporal Resolution Ozone Concentration in China Using Himawari-8 AOD. *Remote Sens.* **2025**, *17*, 528. <https://doi.org/10.3390/rs17030528>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Extra Tree model; near-surface ozone concentration; meteorological factors; Himawari-8 AOD

1. Introduction

Near-surface ozone (O_3) represents a significant trace gas, significantly impacting air quality due to its pronounced oxidative potential, thereby posing substantive deleterious effects on human health, animal growth, and vegetation viability [1–4]. O_3 , as an oxidant, accelerates the deterioration of stone surfaces, affecting historical buildings and monuments [5]. Tropospheric ozone, a key component of photochemical smog, forms through complex photochemical reactions driven by nitrogen oxides (NO_x) from combustion and volatile organic compounds (VOCs) from anthropogenic and biological sources [6–9]. Following the implementation of air pollution prevention and control measures in Chinese regions, a gradual reduction in fine particulate matter (PM) pollution has been observed; however, ozone pollution has exhibited a concurrent increasing trend [10–12]. In China, ground monitoring stations effectively monitor ozone pollution, but the density of stations is very low in sparsely populated areas [13,14]. The uneven distribution of ground-based

monitoring stations complicates the accurate assessment of ozone's spatial distribution. Consequently, achieving high spatiotemporal resolution in ozone concentration measurements has become imperative for the fine-scale investigation of ozone pollution.

Satellite observational data capture pollutant information, and the utilization of machine learning techniques to process these data has arisen as an innovative approach to ozone estimation. The Ozone-Monitoring Instrument (OMI) and TROPOspheric Monitoring Instrument (TROPOMI) polar orbit satellite datasets have been extensively employed for ozone estimation. It has been established that machine learning models utilizing polar orbit satellite data exhibit strong predictive capabilities [13,15–17]. Liu et al. [18] designed an Extreme Gradient Boosting model to forecast daily maximum 8 h average (MDA8) ozone in China (2013–2017) with OMI satellite data. The model yielded a coefficient of determination (R^2) of 0.78 for sample-based cross-validation. These high-precision models have introduced innovative techniques for ozone estimation.

In comparison to polar-orbiting satellites, geostationary satellites substantially enhance temporal and spatial resolution, thereby facilitating more timely and precise surveillance of hazardous meteorological phenomena and atmospheric pollutants [14,19]. Himawari-8, a third-generation geostationary meteorological satellite, carries Advanced Himawari Imager (AHI), which delivers Earth observation data at approximately 2 km spatial resolution and 10 min temporal resolution [20–22].

A robust correlation exists between aerosol optical depth (AOD) and ground-level air pollutants [23]. Based on AOD and meteorological factors, Wang and Sun [24] introduced gaseous pollutants as input factors and constructed a particulate matter with a diameter of 2.5 μm or smaller ($\text{PM}_{2.5}$) estimation model for the Beijing–Tianjin–Hebei region using deep neural networks. The cross-validation R^2 reached 0.87. There is a complex synergistic relationship between $\text{PM}_{2.5}$ and near-surface ozone [25–28]. Heterogeneous chemical reactions are an important way in which $\text{PM}_{2.5}$ and O_3 interact [29,30]. Under low-temperature conditions, $\text{PM}_{2.5}$ effectively suppresses O_3 formation by reducing the photolysis rate, while in high-temperature environments, the high concentration of O_3 and active photochemical activity promote the formation of secondary particles [31]. Therefore, utilizing Himawari-8 AOD data to infer near-surface ozone concentration enriches our understanding of ozone formation and provides targeted guidance for pollution control and urban planning.

Meteorological factors exert a substantial influence on ozone concentration [32–35]. Temperature and relative humidity impact the photochemical reactions responsible for ozone production, while boundary layer height (BLH) affects the aggregation and diffusion of pollutants [2,36,37]. By employing machine learning models, the inversion of near-surface ozone concentrations using meteorological factors facilitates the quantification of meteorological contributions to ozone concentration variations [38–40]. Li et al. [41] used a machine learning model to fill in the missing O_3 on Hainan Island, China, based on meteorological factors and other geographical variables. The results revealed that the Random Forest (RF) model exhibited the best performance among all filling methods, achieving an R^2 of 0.87.

The Extra Tree (ET) model is an ensemble model with fast training speed, high efficiency, and the ability to handle large-scale data [42]. In this study, a high spatiotemporal resolution AOD- O_3 model was constructed using Himawari-8 AOD data, geographic information, meteorological factors, and spatiotemporal information as input variables based on the ET model. The model's performance was then cross-validated. SHapley Additive exPlanation (SHAP) analysis is a method used to explain the output of machine learning models; therefore, using SHAP analysis, the impact of AOD and meteorological variables on ozone concentration was quantified. The model was employed to estimate

the distribution of ozone in China, resulting in the spatiotemporal distribution of ozone concentration with a time resolution of hours and a spatial resolution of 0.05° . Finally, the variation characteristics of ozone concentration at different times were analyzed based on the model's estimated ozone concentration results.

2. Introduction of the Data and Methods

2.1. Study Datasets

2.1.1. O₃ Site Observation

The hourly average values of O₃, which were calibrated and quality-controlled in accordance with the Chinese national standard GB 3095-2012 and HJ 818-2018, were provided by the China National Environmental Monitoring Center (CNEMC) [12,43] (<https://www.aqistudy.cn/historydata/>, accessed on 20 September 2022).

2.1.2. The Himawari-8 AOD Dataset

This study utilized Himawari-8 L3 500 nm AOD product; the temporal resolution of the product is 10 min, and the spatial resolution is 0.05° (<http://www.eorc.jaxa.jp/ptree/index.html>, accessed on 18 June 2022). In order to ensure compatibility with other datasets, only the data corresponding to full hours were selected. Yang et al. [44] compared the Himawari-8 AOD product with AERONET AOD product and found that the two products were similar ($R^2 = 0.72$), while the Himawari-8 AOD values exhibited a slight tendency to be lower, the overall agreement was noteworthy. Gao et al. [45] and Song et al. [23] also discovered a similar phenomenon.

2.1.3. Meteorological and Geographic Data

ERA5 Land, a global reanalysis dataset (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview>, accessed on 20 September 2022), is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) as part of the 5th generation of its reanalysis efforts [46]. ERA5 data include 37 layers of atmosphere from the surface to an altitude of 80 km. The dataset exhibits a temporal resolution of hourly intervals and spatial resolutions of either $0.25^\circ \times 0.25^\circ$ or $0.1^\circ \times 0.1^\circ$. The key meteorological variables incorporated in this study comprise the U and V components of 10 m wind (U10, V10), boundary layer height (BLH), 2 m temperature (TM), relative humidity (RH), and surface pressure (SP). The land use/cover change (LUCC) data were derived from the MODIS LUCC product (MCD12C1 V061, with a spatial resolution of $0.5^\circ \times 0.5^\circ$ and a temporal resolution of 1 year, <https://doi.org/10.5067/MODIS/MCD12C1.061>, accessed on 17 June 2022), which were used to describe land surface types and land use. TM and RH are known to influence the photochemical reactions that generate ozone [2], while BLH, U10, and V10 are recognized to have a significant impact on ozone diffusion [47,48]. SP is typically associated with weather systems, and different weather systems are known to affect the generation and removal of ozone [49]. LUCC is linked to the release of ozone precursors, such as VOCs [6]. Therefore, the above variables were selected as input variables for the model. A detailed overview of the datasets utilized in this study is presented in Table S1.

2.2. Research Methods

2.2.1. Data Integration

The spatial resolution of meteorological variables and auxiliary data was adjusted to $0.05^\circ \times 0.05^\circ$ using bilinear interpolation to ensure alignment with the AOD data. Hourly ozone measurements from ground stations were then matched with the AOD data. For

grids containing multiple stations, the ozone concentrations were averaged to represent the grid's value.

2.2.2. Ten-Fold Cross-Validation

This method can fully utilize data and obtain more reliable evaluation results through ten-fold cross-verifications while avoiding overfitting issues by dividing the dataset into ten parts and taking turns using nine parts as training data and one part as testing data as experiments are conducted. The corresponding accuracy will be obtained for each experiment, and the average accuracy of the ten experiments will be used as an estimate of the algorithm's accuracy [18,50–52]. This study used the ten-fold cross-validation method to assess the model's performance. The model is evaluated using several metrics, including mean absolute error (MAE), where a value of 0 signifies a perfect model and larger errors lead to higher values; root mean square error (RMSE), which equals 0 when predictions are perfect and increases with larger errors; and the coefficient of determination (R^2), with higher values indicating better model performance. The calculation formulas for these indicators are as follows:

$$R^2 = \frac{\sum_{i=1}^n (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (S_i - O_i)^2}{n}} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^n |S_i - O_i|}{n} \quad (3)$$

where S_i , O_i , $\bar{S}$, and $\bar{O}$ represent the predicted value, the true value, the average of the prediction, and the average of the true value, respectively.

2.2.3. The Extra Tree Model

The Extreme Randomized Tree (ET) model is an ensemble learning algorithm that builds multiple decision trees by randomly selecting a subset of features from the original dataset [42]. At each node, these features are used for partitioning, with a randomly chosen threshold for each feature. The partitioning process continues until leaf nodes are created, which occurs either when the sample size becomes too small or when a preset tree depth is reached. By introducing randomness in both feature selection and partition thresholds, the ET model enhances generalization and robustness [53,54]. Furthermore, ET provides advantages in interpretability [55], as it can output feature importance scores, addressing the interpretability challenges often associated with neural networks [56].

2.2.4. Comparative Models

Random forests (RFs) are a combination of the bagging algorithm and decision trees, which is a more advanced form of parallel ensemble learning [53]. The model creates multiple decision trees by taking various samples from the data. Within each decision tree, nodes are split into sub-nodes using randomly chosen optimal features until all samples in a node belong to the same class. Ultimately, the decision trees are aggregated to form a random forest.

Gradient boosting decision tree (GBDT) is an ensemble learning algorithm of the Boosting type, which combines the predictions of multiple decision trees to obtain the result [57]. The core idea of GBDT is to iteratively train decision trees to minimize the loss function, thereby improving model performance [58]. In each iteration, GBDT calculates the negative gradient of the loss function and uses it as the residual to train the next decision

tree, continuing until the preset number of iterations is reached or model performance stops improving.

Light Gradient Boosting Machine (LightGBM) is an efficient gradient-boosting decision tree framework that accelerates tree construction and reduces memory usage through histogram-based algorithms. It directly handles categorical features, avoiding the complexity of encoding. By employing gradient-based one-side sampling and a leaf-wise growth strategy, LightGBM improves training efficiency and model accuracy [59]. Additionally, LightGBM supports data and feature parallelism, making it well-suited for large-scale datasets and enabling efficient operation in distributed environments.

The RF, GBDT, LightGBM, and ET models were trained using the processed training data, and ten-fold cross-validation was applied to evaluate the performance of each model. The mean R^2 and RMSE values from cross-validation for each model are shown in Table 1. The ET model achieved the highest R^2 and lowest RMSE, with a mean R^2 of 0.852 and an RMSE of $19.29 \mu\text{g}/\text{m}^3$. The RF model followed with an R^2 of 0.826, while the GBDT model performed the worst, with an R^2 of only 0.640. Therefore, this study selected the ET model, which demonstrated the best-fitting performance, to estimate the near-surface ozone concentration. The ET model's specific parameter settings are as follows: $n_estimators = 600$, $n_jobs = -1$, with all other parameters set to default values. Figure 1 illustrates a flowchart outlining the principal methods. The detailed steps for estimating ozone concentration using the ET model are shown in Figure S1.

Table 1. Performance comparison of various models.

Model	R^2	RMSE ($\mu\text{g}/\text{m}^3$)
Random Forest (RF)	0.826	20.62
Gradient Boosting Decision Tree (GBDT)	0.640	29.71
Light Gradient Boosting Machine (LightGBM)	0.738	25.06
Extra Tree (ET)	0.852	19.29

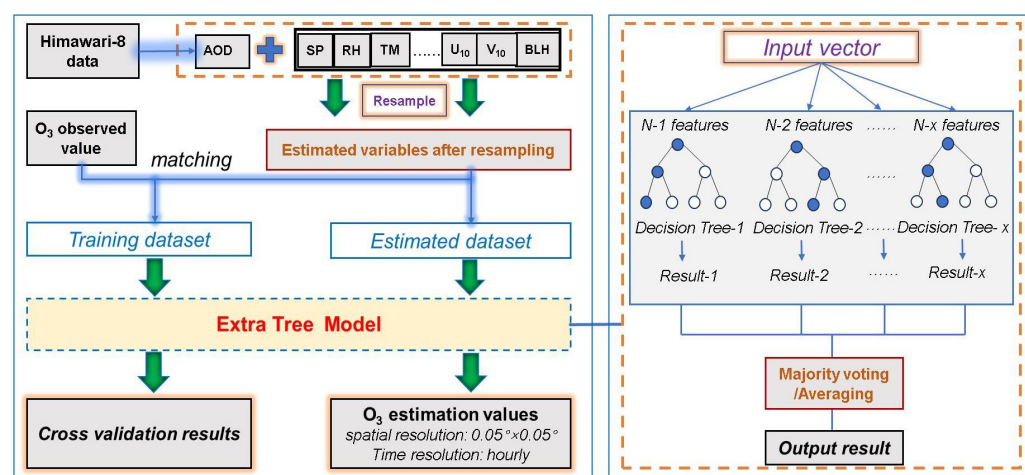


Figure 1. Flowchart for estimating near-surface ozone concentration using the Extra Tree model. The left panel illustrates the input data, methods, and results, with AOD (aerosol optical depth), SP (surface pressure), RH (relative humidity), TM (2 m temperature), U10 and V10 (10 m wind speed), and BLH (boundary layer height). The right panel illustrates the structure of the Extra Tree model.

2.2.5. SHapley Additive ExPlanation (SHAP)

SHAP is a method for addressing model interpretability [60–62]. SHAP is founded on the Shapley value, a concept from cooperative game theory introduced by economist Lloyd Shapley. In game theory, a game denotes a scenario where multiple agents aim to

optimize their individual outcomes. This approach quantifies the importance of each agent by evaluating their contribution.

2.3. Study Period and Area

The study period was from 1 January 2016 to 31 December 2020, and all the times were expressed in Beijing time (UTC+08:00). In this study, winter encompasses the months of December to February, spring spans March to May, summer covers June to August, and autumn includes September to November. Figure 2 illustrates the locations of O₃ environmental monitoring stations, with 1624 sites selected for the entire region. The six boxes delineate six study areas, detailed information for which is provided in Table S2.

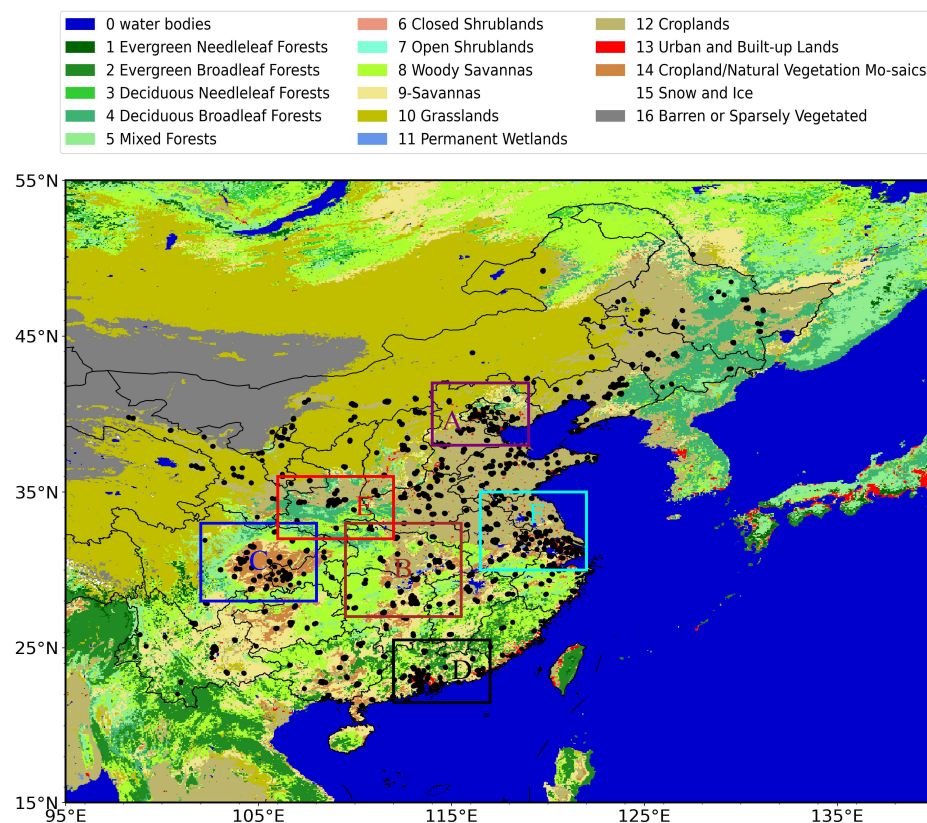


Figure 2. Land cover type and area division, the small black dots represent O₃ monitoring stations. The six boxes represent six large urban agglomerations within the country: (A) Beijing–Tianjin–Hebei region (BTH, purple box), (B) Central China (CC, brown box), (C) Cheng-Yu (CY, blue box), (D) Pearl River Delta (PRD, black box), (E) Guanzhong plain (GZP, red box), (F) Yangtze River Delta (YRD, cyan box).

3. Results

3.1. Model Cross-Validation Results

3.1.1. Temporal Scale Validation Results

Since ozone pollution mainly occurs during the daytime, with concentrations typically peaking in the afternoon due to sunlight-driven photochemical reactions [63], the validation was carried out between 10:00 and 15:00. This six-hour window was selected to assess the model's performance during the period of highest ozone activity. The cross-validation results are shown in Figure 3. From 10:00 to 14:00, the fitting performance of the ET model increased gradually, with a cross-validation R² of 0.75–0.87 and a fitting slope of 0.77–0.85. The RMSE (μg/m³) ranged from 17.96 to 20.30, and the MAE (μg/m³) ranged from 12.11 to 14.48. Between 14:00 and 15:00, the model's fitting effect remained basically

unchanged, with the cross-validation R^2 stable at 0.87, indicating that the AOD- O_3 model can estimate hourly ozone concentration well. As time goes by, from 10:00 to 15:00, a rise in temperatures because of solar radiation gradually increases; additionally, there is an increase in the concentration of ozone precursors due to human activities. This, in turn, results in a rise in ozone levels through photochemical reactions. Overall, the hourly cross-validation results of the AOD- O_3 model showed an R^2 value above 0.75, and the R^2 values were all above 0.80 from 11:00 to 15:00, indicating that the AOD- O_3 model performed well.

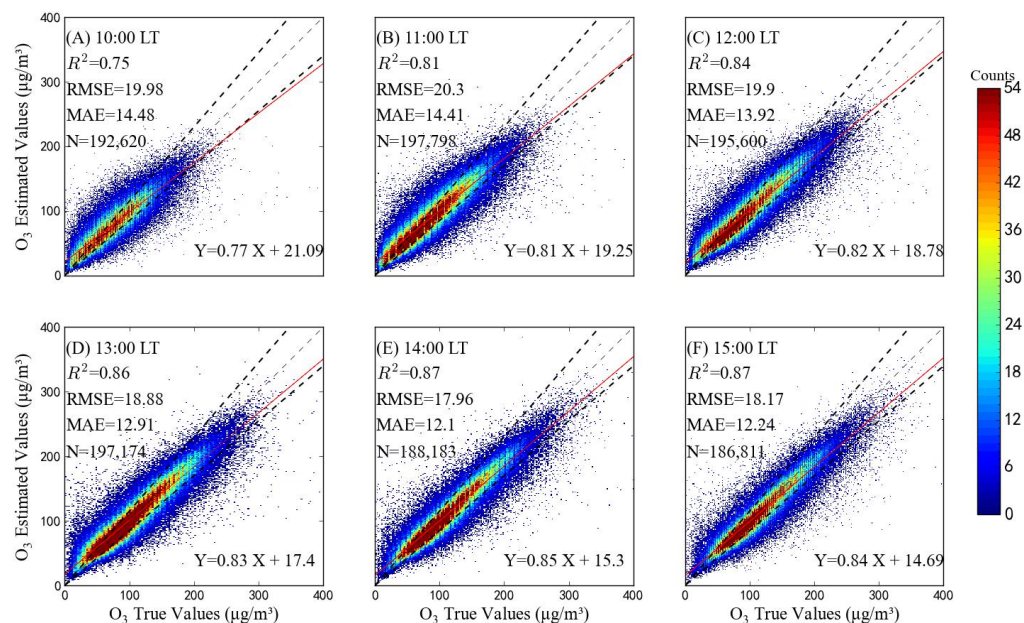


Figure 3. Hourly cross-validation results of AOD- O_3 model at grid points (time is in Beijing time, UTC+08:00). The two black lines represent the expected error; the gray line denotes the 1:1 line; the red line indicates the linear regression fit; the fitting equation is located at the bottom right corner of each panel, where Y represents the O_3 true values; and X represents O_3 estimated values. R^2 is the coefficient of determination, RMSE is the root mean square error, and MAE is the mean absolute error. N represents the size of the dataset included.

The model's performance varies across different seasons, as shown in Figure 4, and across different years, as shown in Figure S2. The highest R^2 value occurred in 2018, reaching 0.89, while the lowest R^2 value was observed in 2016, at 0.76. From a seasonal perspective, the model performed relatively well in autumn, with a cross-validation R^2 of 0.87, an RMSE of $18.29 \mu\text{g}/\text{m}^3$, and an MAE of $12.91 \mu\text{g}/\text{m}^3$. The model also performed relatively well in the other three seasons, with cross-validation R^2 values of 0.81, 0.80, and 0.75 for spring, summer, and winter, respectively. The RMSE is $20.46 \mu\text{g}/\text{m}^3$, $25.33 \mu\text{g}/\text{m}^3$, $15.71 \mu\text{g}/\text{m}^3$, and the MAE is $14.35 \mu\text{g}/\text{m}^3$, $18.15 \mu\text{g}/\text{m}^3$, and $11.06 \mu\text{g}/\text{m}^3$, respectively. The seasonal estimation accuracy of the ET models is influenced by multiple factors, such as surface emission sources, meteorological conditions, and the number of matching samples. In winter, lower temperatures and reduced solar radiation intensity decrease the rate of photochemical reactions that generate O_3 . Consequently, the ozone concentration is lower, and the estimation effect of the model based on Himawari-8 AOD data is lower in winter compared to other seasons. In summer, although temperature and solar radiation are favorable for photochemical reactions to generate ozone [64], the number of samples matched by the model is significantly smaller than in other seasons, preventing the model from showing better performance. In autumn, the model performs the best, mainly because the meteorological conditions are favorable for the generation of ozone; precipitation and

relative humidity in autumn are lower than that in summer, which is impactful in the generation and accumulation of ozone.

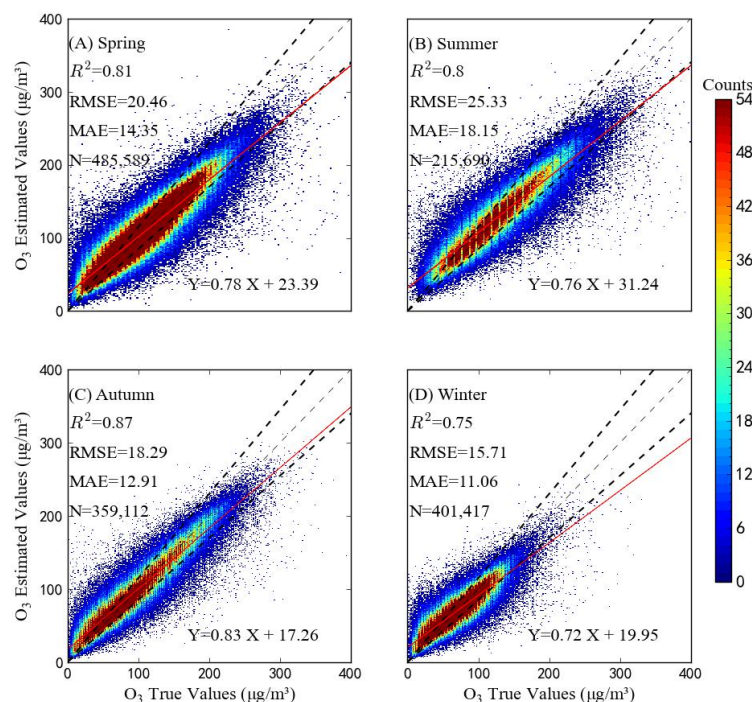


Figure 4. Similar to Figure 3, but depicting seasonal cross-validation results of the AOD-O₃ model based on grid points.

3.1.2. Spatial Scale Verification Results

The model shows good performance in each region, as shown in Figure 5A, with mean R^2 values above 0.8. However, there are also spatial differences in the model's performance. The variations of R^2 , RMSE, and MAE for hourly averages are shown in Figure S3. Regions with higher R^2 (A) and lower RMSE (B) and MAE (C) are generally distributed in economically developed regions, such as BTH, YRD, and CC. In these regions, there are more ozone precursors emitted by human activities, leading to higher ozone concentrations generated by photochemical reactions. In addition, previous studies have shown that the regional transport of ozone precursors (NO_x and VOC_s) also has a significant impact on ozone in areas such as the BTH and YRD [63,65]. In the northwest region, the annual average ozone concentration is lower (Figure 5D), and the model's performance is suboptimal due to sparse stations and pronounced meteorological influences like wind speed, characterized by low R^2 , high RMSE, and MAE.

We calculated the cross-validation results for each of the six regions divided in Figure 2. As shown in Figure 6, the model performs best in the BTH, with an R^2 of 0.91, RMSE of 18.30 µg/m³, and MAE of 12.10 µg/m³. The CC follows with an R^2 of 0.87. In the PRD and GZP, R^2 is basically the same, with an average of 0.85, while RMSE and MAE in GZP are slightly lower than those in PRD. The R^2 value in CY is slightly lower, with an average of 0.81, predominantly due to the complex terrain and sophisticated meteorological factor variations. Overall, the model demonstrates strong performance across various regions, with a mean R^2 of 0.81 or above, RMSE < 22.74 µg/m³, and MAE < 16.65 µg/m³. This indicates that AOD-O₃ has universal applicability for estimating ozone concentrations in diverse regions.

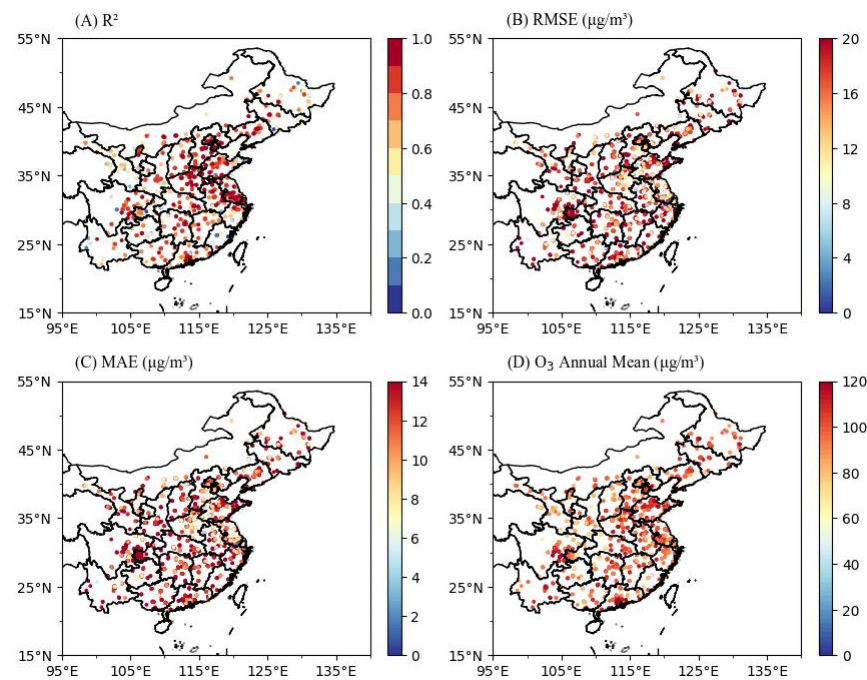


Figure 5. Spatial distribution of AOD- O_3 model evaluation indexes (A–C), with the values shown representing hourly averages. R^2 is the coefficient of determination, RMSE is the root mean square error, and MAE is the mean absolute error. Spatial distribution of annual average O_3 concentrations (D).

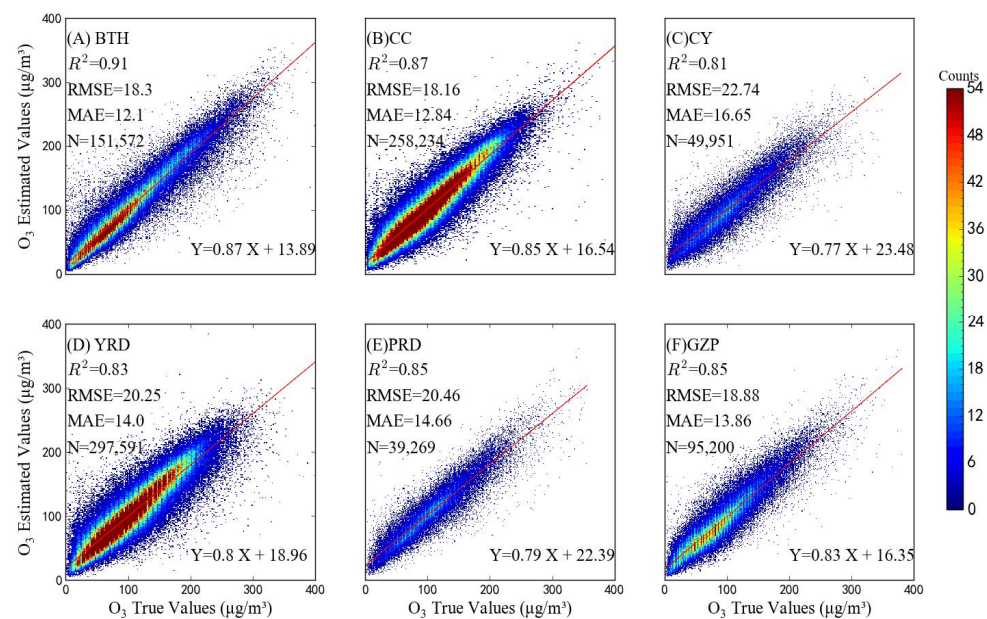


Figure 6. Hourly cross-validation outcomes of the AOD- O_3 model across six typical regions, (A) Beijing–Tianjin–Hebei region (BTH), (B) Central China (CC), (C) Cheng-Yu area (CY), (D) Pearl River Delta (PRD), (E) Guanzhong plain (GZP), (F) Yangtze River Delta (YRD). The red line indicates the linear regression fit; the fitting equation is located at the bottom right corner of each panel, where Y represents the O_3 true values; and X represents O_3 estimated values. R^2 is the coefficient of determination, RMSE is the root mean square error, and MAE is the mean absolute error. N represents the size of the dataset included.

3.2. SHAP Analysis

From the SHAP value analysis (Figure 7), it is noticeable that temperature, relative humidity, and BLH are impactful to the model, while other meteorological factors have minimal impact. This predominance is primarily due to these three elements' crucial roles

in ozone generation and dissipation. An increase in temperature and BLH alters their influence on the model from negative to positive, with higher temperatures yielding greater positive contributions. Conversely, rising relative humidity shifts its impact from positive to negative, most notably during summer (Figure 7B). NO_x and VOCs produce ozone via photochemical reactions, with temperature influencing the rate of these reactions. An increase in temperature accelerates photochemical processes, enhancing ozone generation and thereby elevating ozone concentrations [32]. The influence of BLH on O₃ concentration is complex; elevated BLH enhances ozone diffusion, reducing near-surface O₃ concentration. However, the O₃ generated by photochemical reactions can increase the concentration of ozone [66]. Humidity influences photochemical reactions, either directly via water vapor participating in chemical processes or indirectly by affecting meteorological conditions like temperature and precipitation, thereby modulating ozone generation. Elevated humidity increases atmospheric water vapor, potentially altering photochemical reaction rates. Excessive relative humidity can induce precipitation, causing ozone to settle and reducing its concentration. Furthermore, high humidity can elevate aerosol concentrations, which may absorb ultraviolet radiation and diminish ozone production [67]. The summer temporal variable contributes positively to the model (Figure 7B), reflecting significant temporal changes in summer ozone levels. Summer surface pressure exhibits a positive correlation with ozone, contributing positively to the model (Figure 7B), whereas its impact is less pronounced in autumn (Figure 7C) and winter (Figure 7D). V10 contributes positively to the model in spring (Figure 7A) and summer (Figure 7B), while U10 shows insignificant contributions. Notably, AOD positively influences the model, particularly in summer (Figure 7B), consistent with previous findings on its strong correlation with PM_{2.5} [68–70]; however, the relationship between PM_{2.5} and O₃ is highly complex, so using AOD is also a feasible approach to analyze the relationship between PM_{2.5} and ozone.

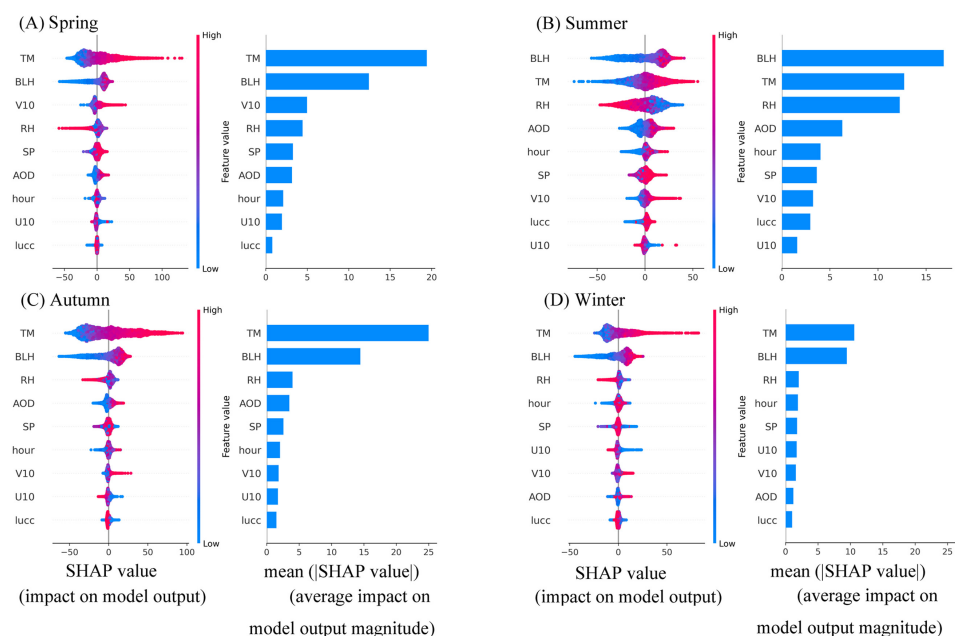


Figure 7. The SHAP importance scores for spring, summer, autumn, and winter, based on the AOD-O₃ model, with panels (A–D) corresponding to each season. The left side of each panel shows the SHAP values, where values closer to red indicate larger variable values, while values closer to blue indicate smaller ones. The x-axis displays the positive and negative contributions of the variables. The right side of each panel shows the mean absolute SHAP values, representing the magnitude of the variable contributions. (BLH: boundary layer height, TM: 2 m temperature, RH: relative humidity, U10: 10 m u component of wind, V10: 10 m v component of wind, SP: surface pressure, AOD: aerosol optical depth, lucc: land use/cover change).

4. Changes in Surface Ozone

4.1. Hourly Variation

The variation characteristics of hourly mean ozone concentration during the study period are shown in Figure 8. Starting from 11:00, ozone rapidly increases in various regions, especially in areas such as BTH, CY, GZP, and the YRD. As the solar altitude angle increases, solar radiation intensifies, leading to a gradual rise in temperature. This enhances the photochemical reactions that produce ozone, resulting in a rapid increase in ozone concentration [71–73]. After 13:00, the rate of ozone increase slows down, but due to the accumulation of ozone concentration in the earlier stage, the total near-surface ozone concentration at 14:00 and 15:00 remains relatively high. Overall, the ozone concentration increase rate in the six study areas selected in this paper is relatively high, and the accumulated total ozone is also considerable. This is mainly due to the high population density and developed economy in these regions, which produce a large amount of ozone precursors that contribute to ozone formation [74,75].

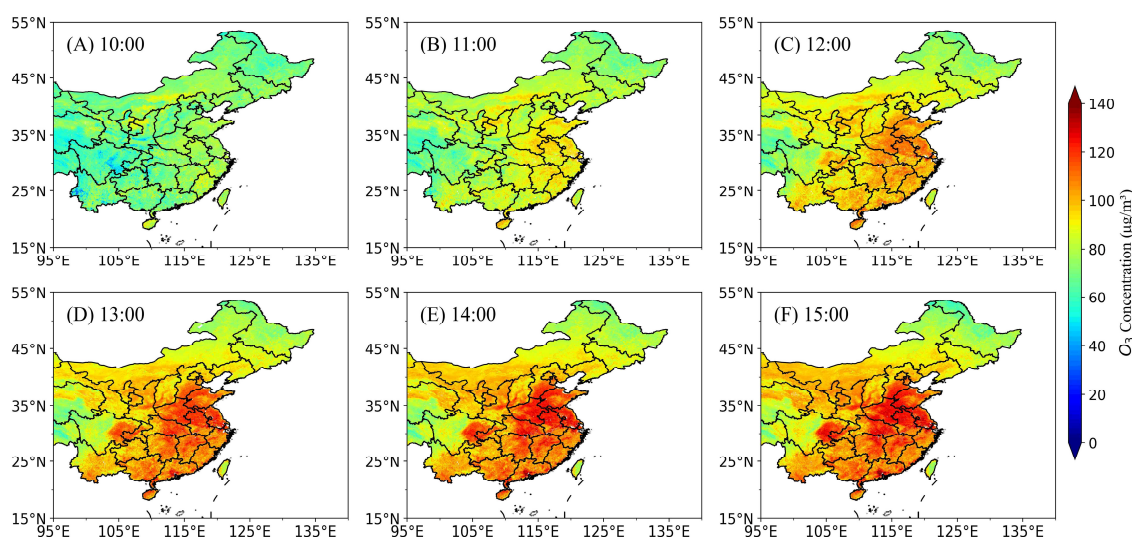


Figure 8. Spatial distribution of hourly mean O₃ concentrations from 10:00 to 15:00 (Beijing time, UTC+08:00), with data corresponding to each hour from 10:00 to 15:00 (A–F).

4.2. Seasonal Variation

The characteristics of mean ozone concentration changes in different seasons are shown in Figure 9. Ozone pollution is most severe in summer, with heavily polluted areas mainly concentrated in the BTH, CY, YRD, and GZP (Figure 9B). It is worth noting that ozone pollution in PRD is not very severe in summer (Figure S4B). This may be linked to the subtropical high pressure affecting summer precipitation across the country. Influenced by the warm and humid airflow around the subtropical high pressure, there is a significant increase in summer precipitation, with the rain belt also shifting according to the varying positions of the subtropical high pressure [76–78]. In addition, due to the control of VOCs on ozone generation in most regions of the country [7,34], reducing the decrease in NO_x in summer is also beneficial for ozone generation. Ozone pollution in spring ranks second to summer, primarily concentrated in the BTH, YRD, and CC regions (Figure 9A). Autumn ozone pollution is predominantly observed in CC, YRD, and PRD (Figure 9C); ozone concentrations in CY are higher than in the other regions, particularly in autumn 2019, when the average concentration in CY exceeded 120 µg/m³ (Figure S4C). Winter exhibits the lowest ozone concentration with minimal pollution nationwide (Figure 9D); the ozone concentration in PRD is the highest, ranging from 70 to 90 µg/m³, while the average ozone concentration in the other regions is below 75 µg/m³ (Figure S4D).

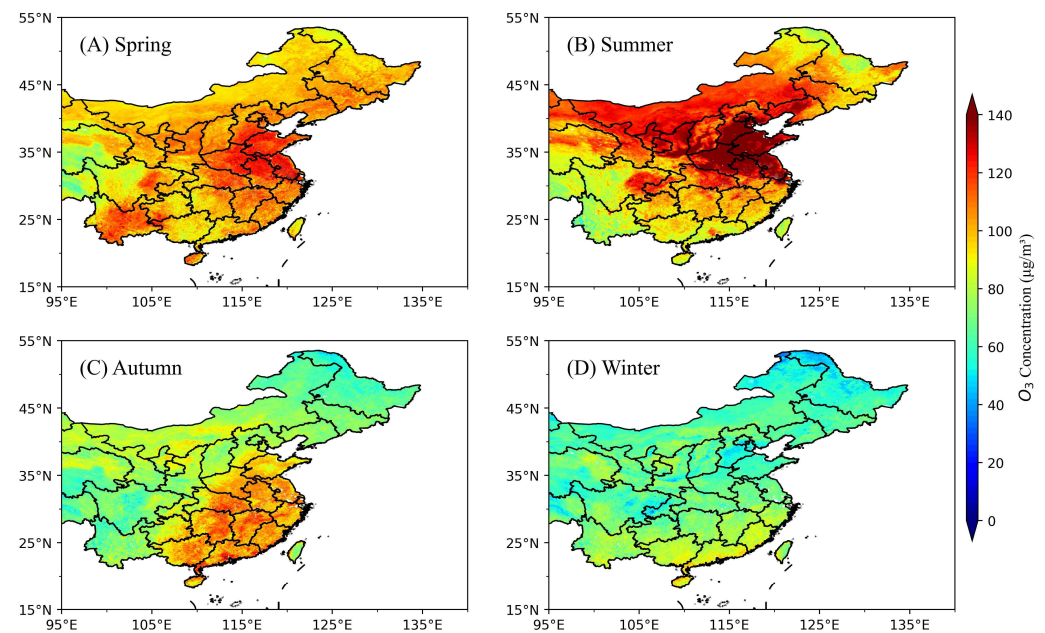


Figure 9. Spatial distribution of seasonal mean O₃ concentrations for spring (A), summer (B), autumn (C), and winter (D).

4.3. Annual Variation

To examine the annual variation characteristics of ozone concentration across regions, the annual difference was computed by subtracting the mean ozone concentration of the previous year from that of the actual year during the study period, as depicted in Figure 10. In comparison to 2016, the annual average ozone concentration increased in most Chinese regions during 2017, with an average increment of approximately 10 µg/m³. The ozone concentration in CY, CC, and YRD regions increased in 2018 compared to 2017, and in 2019, it also increased in PRD and CC relative to 2018. In 2020, an overall decrease in ozone concentration was observed across regions, except for a slight increase in GZP, encompassing BTH and PRD (Figures 10 and S5). The monthly average changes in ozone concentration across six regions are illustrated in Figure S6. In the monthly variations, the peak ozone concentration in the BTH region is the highest, while the trough concentration is relatively low. As a result, the annual average ozone concentration in BTH is lower than that in the YRD and PRD. In contrast, the PRD region has a lower peak ozone concentration but the highest trough concentration, leading to a higher annual average ozone concentration compared to the other five regions (Figure S6). Ozone generation is predominantly controlled by volatile organic compounds (VOCs) in most majority of Chinese regions [7,34]. Compared with 2016, nitrogen oxide emissions decreased in 2017, whereas volatile organic compounds (VOCs) increased rather than decreased [79]. Furthermore, the higher annual average temperature in China during 2017 created favorable meteorological conditions for ozone generation, contributing to an increase in ozone concentration. Compared to 2017, 2018 experienced more high-temperature days nationwide, particularly with extremely high temperatures in the northeastern and central-eastern regions [80]. However, the annual average precipitation across the country was also greater than in 2017 [81]. Under the influence of high relative humidity, ozone concentrations decreased in certain areas. In 2019, the southern region experienced sustained high temperatures during summer, coupled with reduced precipitation compared to the same period in previous years, creating favorable conditions for ozone generation [82]. The northeastern and northwestern regions experienced higher annual precipitation, which was unfavorable for ozone generation. In 2020, the Chinese government intensified remediation efforts for VOCs. Compared

to previous years, the YRD, CC, BTH, and other regions had increased precipitation [83], resulting in a decrease in ozone concentration in these areas.

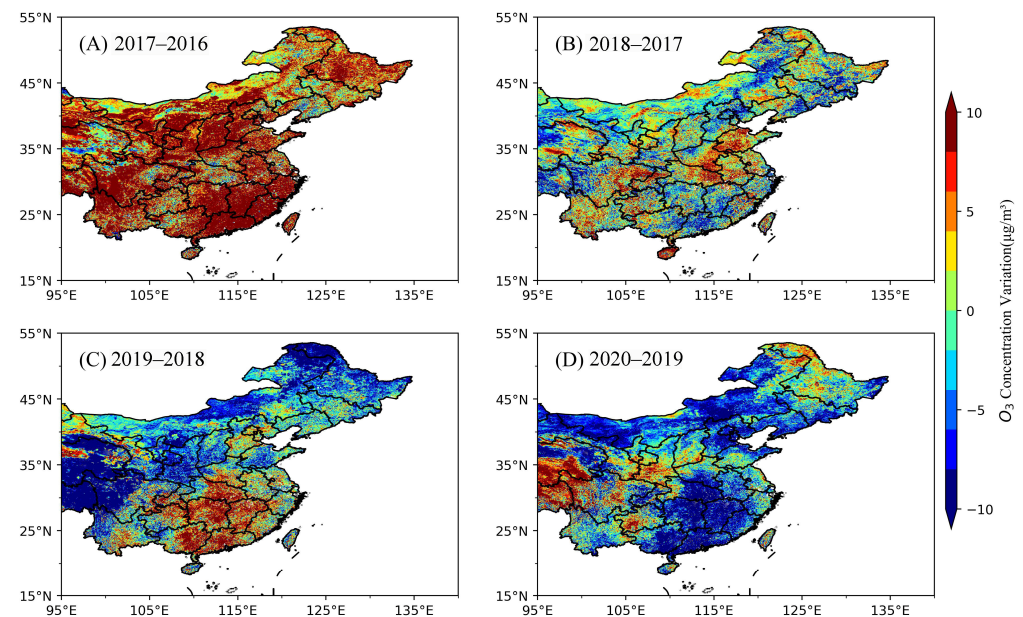


Figure 10. Spatial distribution of annual changes in ozone concentration compared to the previous year: (A) 2017 vs. 2016, (B) 2018 vs. 2017, (C) 2019 vs. 2018, and (D) 2020 vs. 2019.

5. Discussion and Conclusions

This study used an Extra Tree model to obtain high spatiotemporal resolution near-surface ozone data in China, demonstrating the spatiotemporal distribution of ozone. In our research, we revealed that the performance of the model is closely related to the spatiotemporal variation of ozone. This indicates that when constructing estimation models, it is essential to consider the impact of various factors in different regions on ozone concentration and select appropriate inversion variables based on local conditions to obtain more accurate ozone estimation data. Our research also highlights that spring ozone pollution ranks second to summer, necessitating the development of targeted pollution prevention measures. In addition, AOD has a positive contribution to the AOD-O₃ model, especially in summer. This is an issue that needs to be further investigated in future research.

The spatial resolution of the ozone estimation data are primarily constrained by the resolution of the input AOD data, which is 0.05°. To match this resolution, the lower-resolution ERA5 meteorological data (with spatial resolutions of 0.1° or 0.25°) were downsampled using interpolation techniques. As a result, the final spatial resolution of the ozone estimation data in this study is 0.05°. Temporally, the ozone estimation data align with the hourly resolution of the input meteorological data. The resolution of ozone concentration estimation is limited by the resolution of the input data. Higher-resolution observational data can improve the resolution of ozone concentration estimation.

The main conclusions of this study are as follows:

- (1) The Extra Tree model demonstrates effective capability in obtaining hourly near-surface ozone concentrations. Based on sample-based cross-validation, the model achieves an R^2 range of 0.82–0.91, with RMSE ($\mu\text{g}/\text{m}^3$) values spanning from 8.79 to 14.72. Seasonal performance evaluation reveals R^2 values of 0.87, 0.76, 0.9, and 0.88 for spring, summer, autumn, and winter, respectively.
- (2) The model's spatial performance varies across regions, with suboptimal results observed at certain sites in northwestern China. However, it shows satisfactory perfor-

mance in the Beijing–Tianjin–Hebei region, Sichuan Basin, Guanzhong Plain, Central China, Yangtze River Delta, and Pearl River Delta regions, where R^2 values exceed 0.89.

- (3) Temperature and boundary layer height positively influence ozone concentration, whereas relative humidity exerts a negative effect. Across three seasons (excluding summer), temperature is identified as the most significant factor influencing ozone concentrations in the ET model. In summer, boundary layer height becomes the dominant contributing factor.
- (4) Ozone concentrations exhibit notable spatial and temporal variations. Between 11:00 and 15:00 (Beijing time, UTC+08:00), concentrations increase progressively, with the highest levels occurring in summer, followed by spring. In autumn, high ozone concentrations are predominantly distributed in southeastern China, while winter records the lowest average ozone concentrations.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs17030528/s1>, Table S1. Detailed information of data used in the study. Table S2. Detailed information of six regions in the study. Figure S1. The six steps for estimating near-surface ozone concentration using the Extra Tree model. Figure S2. Yearly cross-validation results of AOD- O_3 model at grid points from 2016 to 2020 (time is in Beijing time). Figure S3. Hourly mean variations of the coefficient of determination (R^2). Figure S4. Seasonal average ozone concentration from 2016 to 2020 in six study areas across four seasons (A–D), the specific description of the regions is the same as in Figure S3. Figure S5. Annual average ozone concentration in six study regions from 2016 to 2020, the specific description of the regions is the same as in Figure S3. Figure S6. Monthly mean ozone concentration in six study regions from 2016 to 2020, the specific description of the regions is the same as in Figure S3.

Author Contributions: Conceptualization, C.G.; methodology, Y.W.; software, Y.W.; validation, C.G.; formal analysis, Y.W.; investigation, C.G.; resources, Y.W.; data curation, L.D., Y.H. and Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, L.D., Y.H. and Y.W.; visualization, Y.W.; supervision, C.G.; project administration, C.G.; funding acquisition, C.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (42375085) and the Natural Science Foundation of Gansu (22JR5RA749, 23JRRA1329).

Data Availability Statement: The O_3 observation data were obtained from <https://www.aqistudy.cn/historydata/>, (accessed on 20 September 2022). The Himawari-8 AOD data provided by the Japan Meteorological Agency, download from: <http://www.eorc.jaxa.jp/ptree/index.html>, (accessed on 18 June 2022). The ERA-5 data are available from <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview>, (accessed on 20 September 2022). The authors would like to thank the China National Environmental Monitoring Center and the European Centre for Medium-Range Weather Forecasts for their datasets.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Lu, X.; Zhang, L.; Wang, X.L.; Gao, M.; Li, K.; Zhang, Y.Z.; Yue, X.; Zhang, Y.H. Rapid Increases in Warm-Season Surface Ozone and Resulting Health Impact in China Since 2013. *Environ. Sci. Technol. Lett.* **2020**, *7*, 240–247. [CrossRef]
2. Ghahremanloo, M.; Choi, Y.; Lops, Y. Deep learning mapping of surface MDA8 ozone: The impact of predictor variables on ozone levels over the contiguous United States. *Environ. Pollut.* **2023**, *326*, 121508. [CrossRef] [PubMed]
3. He, J.; Wang, Y.; Hao, J.; Shen, L.; Wang, L. Variations of surface O_3 in August at a rural site near Shanghai: Influences from the West Pacific subtropical high and anthropogenic emissions. *Environ. Sci. Pollut. Res. Int.* **2012**, *19*, 4016–4029. [CrossRef] [PubMed]

4. Li, K.; Jacob, D.J.; Liao, H.; Shen, L.; Zhang, Q.; Bates, K.H. Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. *Proc. Natl. Acad. Sci. USA* **2018**, *116*, 422–427. [[CrossRef](#)] [[PubMed](#)]
5. Yan, Y.; Wang, Y. A Review of Atmospheric Deterioration and Sustainable Conservation of Calcareous Stone in Historical Buildings and Monuments. *Sustainability* **2024**, *16*, 10751. [[CrossRef](#)]
6. Li, L.; Cao, J.; Hao, Y. Spatial and species-specific responses of biogenic volatile organic compound (BVOC) emissions to elevated ozone from 2014–2020 in China. *Sci. Total Environ.* **2023**, *868*, 161636. [[CrossRef](#)] [[PubMed](#)]
7. Wang, T.; Xue, L.; Brimblecombe, P.; Lam, Y.F.; Li, L.; Zhang, L. Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. *Sci. Total Environ.* **2017**, *575*, 1582–1596. [[CrossRef](#)]
8. Yan, D.; Jin, Z.; Zhou, Y.; Li, M.; Zhang, Z.; Wang, T.; Zhuang, B.; Li, S.; Xie, M. Anthropogenically and meteorologically modulated summertime ozone trends and their health implications since China’s clean air actions. *Environ. Pollut.* **2024**, *343*, 123234. [[CrossRef](#)]
9. Zhong, H.; Zhen, L.; Yao, Q.; Xiao, Y.; Liu, J.; Chen, B.; Xu, W. Understanding the spatial and seasonal variation of the ground-level ozone in Southeast China with an interpretable machine learning and multi-source remote sensing. *Sci. Total Environ.* **2024**, *917*, 170570. [[CrossRef](#)]
10. He, Z.; Liu, P.; Zhao, X.; He, X.; Liu, J.; Mu, Y. Responses of surface O₃ and PM_{2.5} trends to changes of anthropogenic emissions in summer over Beijing during 2014–2019: A study based on multiple linear regression and WRF-Chem. *Sci. Total Environ.* **2021**, *807*, 150792. [[CrossRef](#)]
11. 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 (O₃) pollution from 2013 to 2020 across China. *Remote Sens. Environ.* **2021**, *270*, 112775. [[CrossRef](#)]
12. Dong, L.; Chen, B.; Huang, Y.; Song, Z.; Yang, T. Analysis on the Characteristics of Air Pollution in China during the COVID-19 Outbreak. *Atmosphere* **2021**, *12*, 205. [[CrossRef](#)]
13. Kang, Y.; Choi, H.; Im, J.; Park, S.; Shin, M.; Song, C.K.; Kim, S. Estimation of surface-level NO₂ and O₃ concentrations using TROPOMI data and machine learning over East Asia. *Environ. Pollut.* **2021**, *288*, 117711. [[CrossRef](#)] [[PubMed](#)]
14. Song, Z.; Chen, B.; Zhang, P.; Guan, X.; Wang, X.; Ge, J.; Hu, X.; Zhang, X.; Wang, Y. High temporal and spatial resolution PM_{2.5} dataset acquisition and pollution assessment based on FY-4A TOAR data and deep forest model in China. *Atmos. Res.* **2022**, *274*, 106199. [[CrossRef](#)]
15. Zhao, F.; Liu, C.; Cai, Z.; Liu, X.; Bak, J.; Kim, J.; Hu, Q.; Xia, C.; Zhang, C.; Sun, Y.; et al. Ozone profile retrievals from TROPOMI: Implication for the variation of tropospheric ozone during the outbreak of COVID-19 in China. *Sci. Total Environ.* **2021**, *764*, 142886. [[CrossRef](#)]
16. Hu, L.; Jacob, D.J.; Liu, X.; Zhang, Y.; Zhang, L.; Kim, P.S.; Sulprizio, M.P.; Yantosca, R.M. Global budget of tropospheric ozone: Evaluating recent model advances with satellite (OMI), aircraft (IAGOS), and ozonesonde observations. *Atmos. Environ.* **2017**, *167*, 323–334. [[CrossRef](#)]
17. Wang, W.; Liu, X.; Bi, J.; Liu, Y. A machine learning model to estimate ground-level ozone concentrations in California using TROPOMI data and high-resolution meteorology. *Environ. Int.* **2022**, *158*, 106917. [[CrossRef](#)] [[PubMed](#)]
18. Liu, R.; Ma, Z.; Liu, Y.; Shao, Y.; Zhao, W.; Bi, J. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach. *Environ. Int.* **2020**, *142*, 105823. [[CrossRef](#)]
19. Tan, Y.; Wang, Q.; Zhang, Z. Near-real-time estimation of global horizontal irradiance from Himawari-8 satellite data. *Renew. Energy* **2023**, *215*, 118994. [[CrossRef](#)]
20. Wang, Y.; Yuan, Q.; Zhu, L.; Zhang, L. Spatiotemporal estimation of hourly 2-km ground-level ozone over China based on Himawari-8 using a self-adaptive geospatially local model. *Geosci. Front.* **2022**, *13*, 101286. [[CrossRef](#)]
21. Yin, J.; Mao, F.; Zang, L.; Chen, J.; Lu, X.; Hong, J. Retrieving PM_{2.5} with high spatio-temporal coverage by TOA reflectance of Himawari-8. *Atmos. Pollut. Res.* **2021**, *12*, 14–20. [[CrossRef](#)]
22. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; Ohno, T.; et al. An Introduction to Himawari-8/9—Japan’s New-Generation Geostationary Meteorological Satellites. *J. Meteorol. Soc. Jpn. Ser. II* **2016**, *94*, 151–183. [[CrossRef](#)]
23. Song, Z.; Chen, B.; Huang, J. Combining Himawari-8 AOD and deep forest model to obtain city-level distribution of PM_{2.5} in China. *Environ. Pollut.* **2022**, *297*, 118826. [[CrossRef](#)] [[PubMed](#)]
24. Wang, X.; Sun, W. Meteorological parameters and gaseous pollutant concentrations as predictors of daily continuous PM_{2.5} concentrations using deep neural network in Beijing–Tianjin–Hebei, China. *Atmos. Environ.* **2019**, *211*, 128–137. [[CrossRef](#)]
25. Chen, C.; Gao, B.; Xu, M.; Liu, S.; Zhu, D.; Yang, J.; Chen, Z. The spatiotemporal variation of PM_{2.5}–O₃ association and its influencing factors across China: Dynamic Simil-Hu lines. *Sci. Total Environ.* **2023**, *880*, 163346. [[CrossRef](#)] [[PubMed](#)]
26. Shao, M.; Lv, S.; Wei, Y.; Zhu, J. The various synergistic impacts of precursor emission reduction on PM_{2.5} and O₃ in a typical industrial city with complex distributions of emissions. *Sci. Total Environ.* **2024**, *940*, 173497. [[CrossRef](#)] [[PubMed](#)]

27. Wang, S.; Ren, Y.; Xia, B. PM_{2.5} and O₃ concentration estimation based on interpretable machine learning. *Atmos. Pollut. Res.* **2023**, *14*, 101866. [[CrossRef](#)]
28. Liu, S.; Fu, Z. Identifying nonlinear interactions between PM_{2.5} and O₃ on small time scales with visibility graph approaches. *Atmos. Pollut. Res.* **2023**, *14*, 101883. [[CrossRef](#)]
29. Stadtler, S.; Simpson, D.; Schröder, S.; Taraborrelli, D.; Bott, A.; Schultz, M. Ozone impacts of gas–aerosol uptake in global chemistry transport models. *Atmos. Chem. Phys.* **2018**, *18*, 3147–3171. [[CrossRef](#)]
30. Lou, S.; Liao, H.; Zhu, B. Impacts of aerosols on surface-layer ozone concentrations in China through heterogeneous reactions and changes in photolysis rates. *Atmos. Environ.* **2014**, *85*, 123–138. [[CrossRef](#)]
31. Zhu, J.; Chen, L.; Liao, H.; Dang, R. Correlations between PM_{2.5} and Ozone over China and Associated Underlying Reasons. *Atmosphere* **2019**, *10*, 352. [[CrossRef](#)]
32. Chen, Z.; Li, R.; Chen, D.; Zhuang, Y.; Gao, B.; Yang, L.; Li, M. Understanding the causal influence of major meteorological factors on ground ozone concentrations across China. *J. Clean. Prod.* **2020**, *242*, 118498. [[CrossRef](#)]
33. He, Y.; Wang, H.; Wang, H.; Xu, X.; Li, Y.; Fan, S. Meteorology and topographic influences on nocturnal ozone increase during the summertime over Shaoguan, China. *Atmos. Environ.* **2021**, *256*, 118459. [[CrossRef](#)]
34. Li, K.; Chen, L.; Ying, F.; White, S.J.; Jang, C.; Wu, X.; Gao, X.; Hong, S.; Shen, J.; Azzi, M.; et al. Meteorological and chemical impacts on ozone formation: A case study in Hangzhou, China. *Atmos. Res.* **2017**, *196*, 40–52. [[CrossRef](#)]
35. Ma, S.; Shao, M.; Zhang, Y.; Dai, Q.; Xie, M. Sensitivity of PM_{2.5} and O₃ pollution episodes to meteorological factors over the North China Plain. *Sci. Total Environ.* **2021**, *792*, 148474. [[CrossRef](#)]
36. Hu, J.; Zhao, T.; Liu, J.; Cao, L.; Wang, C.; Li, Y.; Shi, C.; Tan, C.; Sun, X.; Shu, Z.; et al. Exploring the ozone pollution over the western Sichuan Basin, Southwest China: The impact of diurnal change in mountain-plains solenoid. *Sci. Total Environ.* **2022**, *839*, 156264. [[CrossRef](#)]
37. Shu, X.; Xia, Z.; Ying, Q.; Fu, Y.; Qiao, X.; Tang, Y. Investigating the causes of O₃ pollution in the western rim of sichuan basin, southwestern China. *Atmos. Pollut. Res.* **2023**, *14*, 101803. [[CrossRef](#)]
38. Zheng, H.; Kong, S.; He, Y.; Song, C.; Cheng, Y.; Yao, L.; Chen, N.; Zhu, B. Enhanced ozone pollution in the summer of 2022 in China: The roles of meteorology and emission variations. *Atmos. Environ.* **2023**, *301*, 119701. [[CrossRef](#)]
39. Tang, J.-H.; Pan, S.-R.; Li, L.; Chan, P.-W. A machine learning-based method for identifying the meteorological field potentially inducing ozone pollution. *Atmos. Environ.* **2023**, *312*, 120047. [[CrossRef](#)]
40. Do, K.; Yeganeh, A.K.; Gao, Z.; Ivey, C.E. Performance of machine learning for ozone modeling in Southern California during the COVID-19 shutdown. *Environ. Sci. Atmos.* **2024**, *4*, 488–500. [[CrossRef](#)]
41. Li, R.; Cui, L.; Hongbo, F.; Li, J.; Zhao, Y.; Chen, J. Satellite-based estimation of full-coverage ozone (O₃) concentration and health effect assessment across Hainan Island. *J. Clean. Prod.* **2020**, *244*, 118773. [[CrossRef](#)]
42. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
43. Deng, C.; Tian, S.; Li, Z.; Li, K. Spatiotemporal characteristics of PM_{2.5} and ozone concentrations in Chinese urban clusters. *Chemosphere* **2022**, *295*, 133813. [[CrossRef](#)] [[PubMed](#)]
44. Yang, X.; Zhao, C.; Luo, N.; Zhao, W.; Shi, W.; Yan, X. Evaluation and Comparison of Himawari-8 L2 V1.0, V2.1 and MODIS C6.1 aerosol products over Asia and the oceania regions. *Atmos. Environ.* **2020**, *220*, 117068. [[CrossRef](#)]
45. Gao, L.; Chen, L.; Li, C.; Li, J.; Che, H.; Zhang, Y. Evaluation and possible uncertainty source analysis of JAXA Himawari-8 aerosol optical depth product over China. *Atmos. Res.* **2021**, *248*, 105248. [[CrossRef](#)]
46. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
47. Mao, J.; Wang, L.; Lu, C.; Liu, J.; Li, M.; Tang, G.; Ji, D.; Zhang, N.; Wang, Y. Meteorological mechanism for a large-scale persistent severe ozone pollution event over eastern China in 2017. *J. Environ. Sci.* **2020**, *92*, 187–199. [[CrossRef](#)] [[PubMed](#)]
48. Shen, L.; Liu, J.; Zhao, T.; Xu, X.; Han, H.; Wang, H.; Shu, Z. Atmospheric transport drives regional interactions of ozone pollution in China. *Sci. Total Environ.* **2022**, *830*, 154634. [[CrossRef](#)]
49. Dong, Y.; Li, J.; Guo, J.; Jiang, Z.; Chu, Y.; Chang, L.; Yang, Y.; Liao, H. The impact of synoptic patterns on summertime ozone pollution in the North China Plain. *Sci. Total Environ.* **2020**, *735*, 139559. [[CrossRef](#)]
50. Song, Z.; Chen, B.; Huang, Y.; Dong, L.; Yang, T. Estimation of PM_{2.5} concentration in China using linear hybrid machine learning model. *Atmos. Meas. Tech.* **2021**, *14*, 5333–5347. [[CrossRef](#)]
51. Chen, B.; Song, Z.; Pan, F.; Huang, Y. Obtaining vertical distribution of PM_{2.5} from CALIOP data and machine learning algorithms. *Sci. Total Environ.* **2022**, *805*, 150338. [[CrossRef](#)]
52. Di, Q.; Rowland, S.; Koutrakis, P.; Schwartz, J. A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. *J. Air Waste Manag. Assoc.* **2017**, *67*, 39–52. [[CrossRef](#)] [[PubMed](#)]
53. Ahmad, M.W.; Reynolds, J.; Rezgui, Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *J. Clean. Prod.* **2018**, *203*, 810–821. [[CrossRef](#)]

54. Jafari, S.; Byun, Y.-C. Efficient state of charge estimation in electric vehicles batteries based on the extra tree regressor: A data-driven approach. *Heliyon* **2024**, *10*, e25949. [[CrossRef](#)] [[PubMed](#)]
55. Liu, Z.; Qi, Z.; Ni, X.; Dong, M.; Ma, M.; Xue, W.; Zhang, Q.; Wang, J. How to apply O₃ and PM_{2.5} collaborative control to practical management in China: A study based on meta-analysis and machine learning. *Sci. Total Environ.* **2021**, *772*, 145392. [[CrossRef](#)] [[PubMed](#)]
56. Yan, X.; Zang, Z.; Jiang, Y.; Shi, W.; Guo, Y.; Li, D.; Zhao, C.; Husi, L. A Spatial-Temporal Interpretable Deep Learning Model for improving interpretability and predictive accuracy of satellite-based PM_{2.5}. *Environ. Pollut.* **2021**, *273*, 116459. [[CrossRef](#)] [[PubMed](#)]
57. Jiang, J.; Cui, B.; Zhang, C.; Fu, F. DimBoost: Boosting Gradient Boosting Decision Tree to Higher Dimensions. In Proceedings of the 2018 International Conference on Management of Data, Houston, TX, USA, 10–15 June 2018; pp. 1363–1376.
58. Wang, J.; Li, P.; Ran, R.; Che, Y.; Zhou, Y. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree. *Appl. Sci.* **2018**, *8*, 689. [[CrossRef](#)]
59. Cai, W.; Wei, R.; Xu, L.; Ding, X. A method for modelling greenhouse temperature using gradient boost decision tree. *Inf. Process. Agric.* **2022**, *9*, 343–354. [[CrossRef](#)]
60. Zhang, L.; Wang, L.; Ji, D.; Xia, Z.; Nan, P.; Zhang, J.; Li, K.; Qi, B.; Du, R.; Sun, Y.; et al. Explainable ensemble machine learning revealing the effect of meteorology and sources on ozone formation in megacity Hangzhou, China. *Sci. Total Environ.* **2024**, *922*, 171295. [[CrossRef](#)] [[PubMed](#)]
61. Tseng, K.-Y.; Hsieh, Y.-T.; Lin, H.-C. Machine learning prediction on wetland succession and the impact of artificial structures from a decade of field data. *Sci. Total Environ.* **2024**, *937*, 173426. [[CrossRef](#)]
62. Yang, C.; Guan, X.; Xu, Q.; Xing, W.; Chen, X.; Chen, J.; Jia, P. How can SHAP (SHapley Additive exPlanations) interpretations improve deep learning based urban cellular automata model? *Comput. Environ. Urban Syst.* **2024**, *111*, 102133. [[CrossRef](#)]
63. Cheng, Y.; Peng, Y.; Cao, L.-M.; Huang, X.-F.; He, L.-Y. Identifying the geospatial relationship of surface ozone pollution in China: Implications for key pollution control regions. *Sci. Total Environ.* **2024**, *930*, 172763. [[CrossRef](#)] [[PubMed](#)]
64. Chen, B.; Wang, Y.; Huang, J.; Zhao, L.; Chen, R.; Song, Z.; Hu, J. Estimation of near-surface ozone concentration and analysis of main weather situation in China based on machine learning model and Himawari-8 TOAR data. *Sci. Total Environ.* **2023**, *864*, 160928. [[CrossRef](#)] [[PubMed](#)]
65. Yao, Y.; Wang, W.; Ma, K.; Tan, H.; Zhang, Y.; Fang, F.; He, C. Transmission paths and source areas of near-surface ozone pollution in the Yangtze River delta region, China from 2015 to 2021. *J. Environ. Manag.* **2023**, *330*, 117105. [[CrossRef](#)] [[PubMed](#)]
66. Zhang, C.; Jiang, Z.; Liu, M.; Dong, Y.; Li, J. Relationship between summer time near-surface ozone concentration and planetary boundary layer height in Beijing. *Atmos. Res.* **2023**, *293*, 106892. [[CrossRef](#)]
67. Ma, P.; Quan, J.; Jia, X.; Liao, Z.; Wang, Q.; Cheng, Z.; Dou, Y.; Su, J.; Pan, Y. Effects of ozone and relative humidity in secondary inorganic aerosol formation during haze events in Beijing, China. *Atmos. Res.* **2021**, *264*, 105855. [[CrossRef](#)]
68. 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. [[CrossRef](#)]
69. Stafoggia, M.; Bellander, T.; Bucci, S.; Davoli, M.; de Hoogh, K.; de' Donato, F.; Gariazzo, C.; Lyapustin, A.; Michelozzi, P.; Renzi, M.; et al. Estimation of daily PM₁₀ and PM_{2.5} concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environ. Int.* **2019**, *124*, 170–179. [[CrossRef](#)] [[PubMed](#)]
70. Rosanka, S.; Flesch, M.M.; Chiu, Y.T.T.; Carlton, A.G. Chemically specific sampling bias: The ratio of PM_{2.5} to surface AOD on average and peak days in the U.S. *Environ. Sci. Atmos.* **2024**, *4*, 547–556. [[CrossRef](#)]
71. Sadeghi, B.; Ghahremanloo, M.; Mousavinezhad, S.; Lops, Y.; Pouyaei, A.; Choi, Y. Contributions of meteorology to ozone variations: Application of deep learning and the Kolmogorov-Zurbenko filter. *Environ. Pollut.* **2022**, *310*, 119863. [[CrossRef](#)] [[PubMed](#)]
72. Nelson, D.; Choi, Y.; Sadeghi, B.; Yeganeh, A.K.; Ghahremanloo, M.; Park, J. A comprehensive approach combining positive matrix factorization modeling, meteorology, and machine learning for source apportionment of surface ozone precursors: Underlying factors contributing to ozone formation in Houston, Texas. *Environ. Pollut.* **2023**, *334*, 122223. [[CrossRef](#)]
73. Alduwais, A.K.; Dasari, H.P.; Karumuri, R.K.; Gandham, H.; Koteswararao, V.; Saharwardi, M.S.; Ashok, K.; Hoteit, I. Spatiotemporal variability of surface ozone and associated meteorological conditions over the Arabian Peninsula. *Atmos. Pollut. Res.* **2024**, *15*, 102210. [[CrossRef](#)]
74. Chen, X.; Situ, S.; Zhang, Q.; Wang, X.; Sha, C.; Zhou, L.; Wu, L.; Wu, L.; Ye, L.; Li, C. The synergetic control of NO₂ and O₃ concentrations in a manufacturing city of southern China. *Atmos. Environ.* **2019**, *201*, 402–416. [[CrossRef](#)]
75. Flynn, C.M.; Pickering, K.E.; Crawford, J.H.; Lamsal, L.; Krotkov, N.; Herman, J.; Weinheimer, A.; Chen, G.; Liu, X.; Szykman, J.; et al. Relationship between column-density and surface mixing ratio: Statistical analysis of O₃ and NO₂ data from the July 2011 Maryland DISCOVER-AQ mission. *Atmos. Environ.* **2014**, *92*, 429–441. [[CrossRef](#)]
76. Li, M.; Yang, Y.; Wang, H.; Wang, P.; Liao, H. Unique impacts of strong and westward-extended western Pacific subtropical high on ozone pollution over eastern China. *Environ. Pollut.* **2024**, *358*, 124515. [[CrossRef](#)] [[PubMed](#)]

77. Zhao, Z.; Wang, Y. Influence of the West Pacific subtropical high on surface ozone daily variability in summertime over eastern China. *Atmos. Environ.* **2017**, *170*, 197–204. [[CrossRef](#)]
78. Liu, C.; He, C.; Wang, Y.; He, G.; Liu, N.; Miao, S.; Wang, H.; Lu, X.; Fan, S. Characteristics and mechanism of a persistent ozone pollution event in Pearl River Delta induced by typhoon and subtropical high. *Atmos. Environ.* **2023**, *310*, 119964. [[CrossRef](#)]
79. Chen, N.; Yang, Y.; Wang, D.; You, J.; Gao, Y.; Zhang, L.; Zeng, Z.; Hu, B. Changing ozone sensitivity in Fujian Province, China, during 2012–2021: Importance of controlling VOC emissions. *Environ. Pollut.* **2024**, *359*, 124757. [[CrossRef](#)]
80. Ma, K.; Gong, H.; Wang, L. Attribution of the concurrent extreme heatwaves in Northern Europe and Northeast Asia in July 2018. *Atmos. Res.* **2024**, *307*, 107506. [[CrossRef](#)]
81. Si, P.; Li, Q.-X.; Chen, X.-Y.; Wang, M.; Luo, C.-J. Construction of daily precipitation series and the observational characteristics of extreme precipitation in Tianjin, China during 1888–2022. *Adv. Clim. Change Res.* **2024**, *15*, 52–61. [[CrossRef](#)]
82. Gong, H.; Zhang, M.; Xiang, X.; Liu, H. 1 km monthly precipitation and temperatures dataset for China from 1952 to 2019 based on new baseline climatology surfaces. *Sci. Total Environ.* **2024**, *906*, 167613. [[CrossRef](#)] [[PubMed](#)]
83. Mun, T.; Park, H.; Cha, D.-H.; Song, C.-K.; Min, S.-K.; Son, S.-W. Did recent sea surface temperature warming reinforce the extreme East Asian summer monsoon precipitation in 2020? *Weather Clim. Extrem.* **2024**, *44*, 100682. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.