remote sensing

MDPI

Article

High-Resolution Inversion, Driving Mechanisms, and Source Apportionment of Near-Surface Ozone in Arid Urban Clusters: A Case Study of the Tianshan North Slope Urban Agglomeration

Guangrui Pan [1,2], Yunyun Xi [1,2], Tuodi Wang [3], Liqiang Shen [1,2], Yutian Luo [1,2], Zhijun Li [1,4], Lihong Wang [1,2], Liping Xu [1,2,*], Linlin Cui [1,2], Shuliang Zhang [5], Xiangjun Lu [6] and Yongpeng Tong [7]

1 School of City and Environment, Shihezi University, Shihezi 832000, China
2 Key Laboratory of Ecological Corps of Oasis Town and Mountain Basin System, Shihezi 832000, China
3 School of Economics and Management, Xinjiang University of Technology, Hotan 848000, China
4 State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian 116000, China
5 State Key Laboratory of Climate System Prediction and Risk Management, Nanjing 210000, China
6 The Fourth Monitoring Station of Ecological Environment of Xinjiang Production and Construction Corps, Cocodala 835900, China
7 School of Science, Shihezi University, Shihezi 832000, China
* Correspondence: xlp_tea@shzu.edu.cn

Highlights

What are the main findings?

- Developed a near-surface O_3 spatial inversion model suitable for arid urban clusters, achieving high-precision daily-scale spatial reconstruction of O_3 in the Tianshan North Slope Urban Agglomeration (TNSUA).
- Revealed the spatiotemporal distribution characteristics of near-surface O_3, its transport pathways, and potential source regions in the TNSUA.
- Downward shortwave radiation and air temperature are the dominant factors controlling the spatial variation in near-surface O_3, exhibiting significant threshold effects.

What are the implications of the main findings?

- Provides high-precision spatial distribution of near-surface O_3, offering scientific support for air quality monitoring in urban clusters.
- Quantitatively elucidates the regulatory effects of meteorological conditions on near-surface O_3 formation, providing a reference for precise pollution control.
- Accurately identifies transport pathways and potential source regions, supplying a scientific basis for cross-regional coordinated emission reduction strategies in arid urban clusters.

Check for updates

Academic Editor: Manuel Antón

Received: 21 April 2026
Revised: 14 June 2026
Accepted: 1 July 2026
Published: 4 July 2026

Copyright: © 2026 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.

Abstract

Ozone (O_3), as a key secondary pollutant, exhibits pronounced spatiotemporal heterogeneity, posing significant challenges to coordinated regional air pollution control. However, systematic understanding of high-resolution O_3 spatial inversion and its driving mechanisms in arid urban agglomerations remains limited. In this study, the Tianshan North Slope Urban Agglomeration (TNSUA) was selected as the study area, and a multi-model comparative framework was established to comprehensively evaluate the O_3 inversion performance of 16 machine learning and deep learning models, including Extreme Gradient Boosting (XGBoost), Random Forest (RF), Extremely Randomized Trees (ET), and Gradient Boosting Decision Tree (GBDT). Based on the optimal model performance, high-precision daily O_3 spatial reconstruction for the year 2023 was achieved across the study region. The

contributions of individual driving factors and their nonlinear response relationships were quantitatively interpreted using Shapley Additive Explanations (SHAP). Furthermore, a backward trajectory model combined with the Weighted Potential Source Contribution Function (WPSCF) and Weighted Concentration Weighted Trajectory (WCWT) methods was employed to identify potential source regions and transport pathways of O_3. The results indicate that: (1) The XGBoost model exhibited the best performance (R^2 = 0.93, RPD > 3). The reconstructed results reveal that high O_3 concentrations in 2023 were primarily distributed in southern Urumqi, southern Changji, and southern Tacheng, with southern Urumqi identified as the most prominent hotspot. (2) The spatial variability of O_3 was predominantly driven by downward shortwave radiation (DSR) and air temperature (TEM), both of which showed significant nonlinear responses and threshold effects on O_3 formation. (3) Source apportionment analysis indicates that westerly transport serves as a major exogenous contribution pathway, with potential source regions mainly located in the surrounding areas of the northern Tianshan slope as well as Central Asia, particularly eastern Kazakhstan and northern Kyrgyzstan. This study systematically elucidates the formation mechanisms of O_3 pollution in arid urban agglomerations from three aspects—high-precision inversion, driving mechanism analysis, and cross-regional transport identification—thereby providing a scientific basis for precise air pollution control strategies.

Keywords: ozone; spatial distribution; spatial source; artificial intelligence model

1. Introduction

Air pollution has significant impacts on human health, atmospheric environmental quality, and sustainable social development and has become one of the key environmental issues of concern for governments, scientists, and the public worldwide [1–3]. As one of the six criteria air pollutants, O_3 plays a protective role in the stratosphere by absorbing ultraviolet radiation [4–6]. However, in the lower troposphere, O_3 adversely affects the human respiratory system, causing chest tightness and coughing, and in severe cases impairing lung function, thereby posing substantial risks to human health [7–10].

To prevent and control air pollution and improve atmospheric environmental quality, extensive studies have been conducted at multiple scales—global, national, regional, and urban [11–14]. These studies have addressed various aspects of air pollution, including its chemical composition, emission sources, and control strategies [15–17]. Early research primarily relied on monitoring data from air quality stations and focused on statistical analyses of pollutant composition and concentration variations [18–20]. In recent years, with the rapid advancement of big data and artificial intelligence, increasing attention has been devoted to the temporal forecasting of air pollutants and the investigation of their driving factors [21–23]. However, most existing studies have predominantly employed a single type of model, such as empirical models [24,25] or individual artificial intelligence models [26,27], lacking the systematic integration of multiple modeling approaches. Moreover, a growing body of research has recognized that air pollution is not only driven by local emissions but is also significantly influenced by cross-regional pollutant transport [28–30]. Nevertheless, current studies on the spatial origin of air pollutants often conduct source apportionment based on city centers or individual monitoring stations within the study area [31,32], which limits the ability to accurately identify pollution hotspots. In this study, we integrate O_3 spatial distribution results to precisely locate high-concentration regions at specific geographic coordinates, enabling refined source apportionment analysis for pollution hotspots.

In terms of pollutant selection, earlier studies primarily focused on particulate matter ($PM_{2.5}$ and PM_{10}) [33–35]. In recent years, however, a growing number of studies have reported that O_3 pollution has intensified, drawing substantial interest from researchers [9,36–38]. Regarding the selection of study areas, with the rapid socio-economic development of China and the acceleration of industrialization and urbanization, the country has been facing significant air pollution challenges [39–41]. At present, China has entered a critical stage in air pollution prevention and control, with overall air quality showing steady improvement nationwide [42–44]. Nevertheless, air pollution in some key regions remains a serious concern, among which the Tianshan North Slope Urban Agglomeration (TNSUA) is a notable example [45,46]. This region is located in the arid zone of northwestern China and represents a typical inland arid urban agglomeration [47]. Compared with eastern humid regions, arid and semi-arid regions are characterized by scarce precipitation, strong solar radiation, and large diurnal temperature variations, resulting in pronounced regional differences in the formation and dissipation processes of O_3 [48,49]. Previous studies have shown that O_3 pollution in arid and semi-arid urban areas is jointly influenced by meteorological conditions and anthropogenic activities, where high temperature and strong solar radiation can significantly enhance photochemical reaction rates, thereby promoting O_3 formation and accumulation [50,51]. Further studies have indicated that, unlike eastern humid regions where O_3 variations are primarily driven by anthropogenic activities, O_3 concentrations in urban agglomerations in arid regions are more sensitive to meteorological conditions, and in some cases meteorological factors even play a stronger role than anthropogenic influences [52–56]. However, research on O_3 pollution at the scale of arid urban agglomerations remains relatively limited. Therefore, a deeper understanding of the driving mechanisms of O_3 pollution in arid urban agglomerations is of great importance for improving the accurate prevention and coordinated control of regional O_3 pollution.

Based on this, the present study focuses on the TNSUA, with O_3 selected as the core research object, and constructs a multi-model integrated analytical framework from a geospatial perspective. By systematically comparing 16 machine learning (ML) and deep learning (DL) models, high-resolution spatial inversion of O_3 is conducted. The Shapley Additive Explanations (SHAP) method is employed to quantitatively identify the key driving factors of near-surface O_3 and to characterize their nonlinear and threshold response relationships. Furthermore, based on the spatial inversion results, a backward trajectory model is introduced for O_3 hotspot regions to analyze pollutant transport pathways and identify potential source areas, thereby enabling a refined attribution of high-concentration O_3 sources.

2. Materials and Methods

2.1. Overview of the Research Area

The TNSUA is located along the northern foothills of the Tianshan range, forming a 300 km-long, east–west-oriented belt of densely distributed cities, accounting for approximately 5.7% of the total area of Xinjiang. This region represents the most economically productive area in Xinjiang, with highly concentrated industrial and agricultural activities. It hosts about 83% of the region's heavy industry and 62% of its light industry, and its gross domestic product has consistently contributed more than 40% of Xinjiang's total [57,58]. However, this area also accounts for nearly half of the total coal consumption in Xinjiang. Due to low winter temperatures and a prolonged heating season, the intensive use of coal and other fossil fuels has led to severe air pollution. In recent years, with the adoption of air pollution mitigation measures, concentrations of particulate matter ($PM_{2.5}$ and PM_{10}) have shown a declining trend, whereas O_3 concentrations have exhibited an increasing

trend [57,59], drawing significant attention from policymakers and researchers. Therefore, this study selects 26 national air quality monitoring stations across seven prefectures and cities, ranging from Urumqi in the east to Bortala Mongolian Autonomous Prefecture in the west (Figure 1), to investigate the spatial distribution and source attribution of O_3.

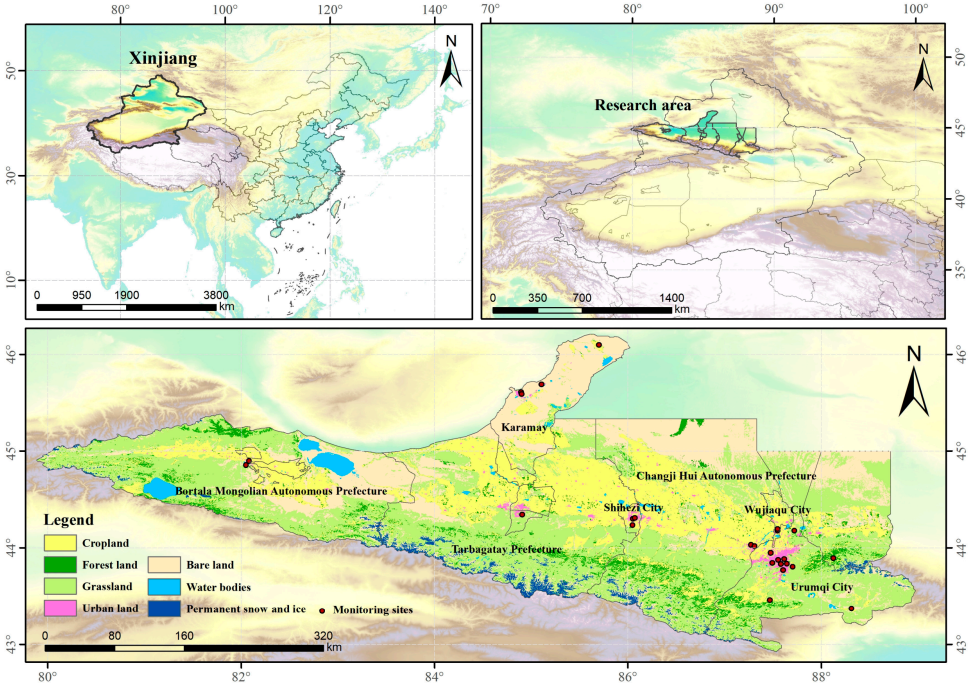


Figure 1. Research areas and monitoring sites.

2.2. Data Sources

2.2.1. Surface O_3 Observations

The O_3 observational data used in this study were obtained from 26 monitoring stations (Figure 1) within the TNSUA, provided by the China National Environmental Monitoring Center (CNEMC) (https://www.cnemc.cn/). Data processing was conducted in accordance with the Ambient Air Quality Standards of China (GB 3095–2012), with preprocessing procedures following the approaches of Fan et al. (2020) [60] and Luo et al. (2023) [57]. Specifically, the quality-control procedures for the hourly O_3 observations consisted of three steps. First, hourly O_3 concentrations falling outside the physically reasonable atmospheric range were identified as invalid and removed. Second, consecutive records with identical values over extended periods were excluded to eliminate potential artifacts caused by instrument malfunction, communication interruption, or calibration failure. Third, daily average O_3 concentrations were calculated only when more than 20 valid hourly observations were available within a day; otherwise, the corresponding daily record was discarded. After quality control, including outlier screening and missing-value removal, approximately 3.35% of the total observations were excluded, indicating generally high data quality.

2.2.2. Ancillary Data for Surface O_3 Retrievals

Downward shortwave radiation (DSR) data play an important role in ozone formation, and the DSR data used in this study are derived from NASA's Global Land Data Assimilation System (GLDAS). Meteorological factors also have significant impacts on ozone formation and transport. The meteorological datasets employed in this study are obtained from the ERA5-Land global reanalysis hourly data, including air temperature (TEM), evaporation (ET), total precipitation (TP), surface pressure (SP), relative humidity

(RH), wind direction (WD, calculated from the u and v components of wind), and wind speed (WS). Satellite remote sensing data are also considered, as total column ozone is highly correlated with surface ozone. The Total Ozone Column ($TotalO_3$) data are sourced from the TROPOMI satellite, which provides daily observations with a spatial resolution of 3.5 km × 5 km. Digital Elevation Model (DEM) data are derived from the Shuttle Radar Topography Mission (SRTM) with a spatial resolution of 30 m × 30 m. Land use/land cover change (LUCC) and Normalized Difference Vegetation Index (NDVI) data are obtained from MODIS sensors, with a spatial resolution of 500 m × 500 m. The NDVI data have a temporal resolution of 16 days. In addition, human activity variables are included: population (POP) data are sourced from LandScan™, with a spatial resolution of 1 km, and nighttime light (NL) data are obtained from NPP-VIIRS at a spatial resolution of 500 m × 500 m. Detailed parameters of all datasets are provided in Table 1. All variables were resampled to a unified spatial resolution of 300 m using bilinear interpolation.

Table 1. Summary of data sources used in artificial intelligence algorithms.

Data	Variable	Description	Unit	Temporal Resolution	Spatial Resolution	Data Source
Ground measurements	O_3	Ozone	$\mu g \cdot m^{-3}$	1 day	-	China National Environmental Monitoring Centre
Solar radiation	DSR	Downward shortwave radiation	W/m^2	3 h	0.25° × 0.25°	GLDAS
Meteorological factors	TEM	2-m air temperature	°C	1 h	0.1° × 0.1°	ERA5-Land
	ET	Evaporation	mm	1 h	0.1° × 0.1°	
	TP	Total precipitation	mm	1 h	0.1° × 0.1°	
	SP	surface pressure	hPa	1 h	0.1° × 0.1°	
	RH	Relative humidity	%	1 h	0.1° × 0.1°	
	WD	Wind direction	°	1 h	0.1° × 0.1°	
	WS	Wind speed	m/s	1 h	0.1° × 0.1°	
Satellite remote sensing products	$TotalO_3$	Total Ozone Column	mol/m^2	1 day	3.5 km × 5 km	TROPOMI Satellite
	DEM	Digital elevation model	m	-	30 m × 30 m	SRTM
	LUCC	land use-cover change	-	1 year	500 m × 500 m	MODIS products
	NDVI	Normalized Difference Vegetation Index	-	16 day	500 m × 500 m	
Human activity data	POP	Population	-	1 year	1 km × 1 km	LandScan™
	NL	Night lighting	$nW/cm^2/sr$	1 month	500 m × 500 m	NPP-VIIRS

2.2.3. Data for Spatial Source Analysis

The meteorological data used to drive the HYSPLIT backward trajectory model were obtained from the Global Data Assimilation System (GDAS) (https://ready.arl.noaa.gov/archives.php, accessed on 3 January 2025), provided by the National Centers for Environmental Prediction (NCEP), with a spatial resolution of 1° × 1°.

2.3. Methods

2.3.1. Spatial Inversion Model Development

Since O_3 is primarily generated through photochemical reactions, its formation is closely related to natural factors such as solar radiation and air temperature. Therefore,

based on observed data from monitoring stations, artificial intelligence models are employed to learn the relationships between O_3 concentrations and natural variables (e.g., solar radiation, temperature, and precipitation). This process establishes a mapping between the explanatory variables and O_3 concentrations, forming an O_3 spatial inversion model. By applying the trained model to continuous spatial datasets of the explanatory variables, continuous spatial distributions of O_3 can be generated [61,62].

Feature Selection

To ensure the accuracy of the inversion model, features that exhibit significant correlations with O_3 (the dependent variable) were selected, while redundant, irrelevant, or highly collinear variables were removed. Pearson Correlation Analysis (Pearson), Spearman Rank Correlation Analysis (Spearman), Kendall Rank Correlation Coefficient (Kendall's τ), and the Variance Inflation Factor (VIF) were employed for feature selection [63]. The corresponding calculation formulas are as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (1)$$

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (2)$$

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)} \quad (3)$$

$$VIF_i = \frac{1}{1 - R_i^2} \quad (4)$$

Here, r denotes the Pearson correlation coefficient, ranging from -1 to 1. x_i represents the value of the feature variable for the i-th sample, and y_i represents the O_3 concentration for the i-th sample. $\overline{x}$ and $\bar{y}$ denote the mean values of the feature variable and O_3 concentration, respectively, and n is the total number of samples. ρ denotes the Spearman rank correlation coefficient, ranging from -1 to 1, and d_i represents the difference in ranks between the feature variable and O_3 concentration for the i-th sample. τ represents Kendall's rank correlation coefficient, with a range of -1 to 1, where C is the number of concordant pairs and D is the number of discordant pairs. VIF_i denotes the variance inflation factor for the i-th feature variable, and $R_i{}^2$ represents the coefficient of determination obtained by regressing the i-th variable against all other variables.

Pearson, Spearman, and Kendall's τ were employed to evaluate the relationships among candidate variables. Variables exhibiting excessively high correlations with other predictors were excluded to reduce redundancy. Subsequently, VIF was used to assess multicollinearity, and variables with VIF values greater than 10 were removed. The remaining variables, characterized by moderate inter-variable correlations and acceptable multicollinearity levels, were retained for model development.

Artificial Intelligence Model Selection

The relationship between O_3 formation and natural variables is highly nonlinear. ML and DL models are well suited to capture such complexity and nonlinearity, enabling the simulation of the multi-factor coupling mechanisms involved in O_3 photochemical formation and revealing the underlying relationships between explanatory variables and O_3 concentrations. Considering the characteristics of different ML models and their applications in the field of atmospheric environmental research, 16 representative models were selected for the retrieval of O_3 concentrations in the TNSUA. Based on the retrieval

accuracy and computational efficiency of each model, the most suitable ML model for this study was identified. Detailed descriptions and applications of the 16 models are presented in Table 2.

Table 2. Artificial intelligence models selected in this study.

Model Types	Model Name	Abbreviation	Abbreviation	Reference
ML	Decision Tree	DT	Applied in research on air quality data.	[64]
	Random Forest	RF	Applied in research on CH_4 and N_2O emissions in China.	[65]
	Extremely Randomized Trees	ET	Extensively applied in research on complex non-linear environments.	[66]
	Adaptive Boosting	AdaBoost	Applied in research on the Air Quality Index (AQI) in major Indian cities.	[67]
	Gradient Boosting Decision Tree	GBDT	Applicable to the non-linear function fitting of large-scale, complex geographic environmental features.	[68]
	Extreme Gradient Boosting	XGBoost	Applied in high-precision near-surface $PM_{2.5}$ concentration retrieval and spatial mapping in major South African cities.	[69]
	Light Gradient Boosting Machine	LightGBM	Applied in research on O3 concentrations in China.	[70]
	Categorical Boosting	CatBoost	Extensively applied in research on categorical feature processing.	[69]
	Support Vector Regression	SVR	Applied in research on urban air pollutants and localized meteorological features.	[69]
DL	Multilayer Perceptron	MLP	Applied in generalized modeling and non-linear feature extraction of complex environmental monitoring data.	[71]
	Deep Neural Network	DNN	Applied in localization research of atmospheric pollution sources.	[72]
	Deep Belief Network	DBN	Applicable to research on large-scale, high-dimensional environmental sequence data.	[73]
	Convolutional Neural Network	CNN	Applied in research on urban air quality.	[74]
	Long Short-Term Memory	LSTM	Applied in research on long-term time-series forecasting of atmospheric pollutants.	[75]
	Residual Neural Network	ResNet	Extensively applied in research such as high-level feature identification of spatial remote sensing imagery.	[76]
	Residual Network	ResNeXt	Applied in the prediction research of typhoon pathways and extreme meteorological trajectories.	[77]

Model Evaluation

Model validation plays a crucial role in ensuring the accuracy and robustness of the models. In this study, a 10-fold cross-validation approach was employed to evaluate and compare model performance. The dataset was randomly divided into 10 subsets, and an iterative training process was conducted in which 9 subsets were used for model training while the remaining subset was used for testing and validation. This rotating procedure ensures that all data are involved in the validation process. For model evaluation, three metrics were selected: the Coefficient of Determination (R^2), Root Mean Squared Error

(RMSE), and the Ratio of Performance to Deviation (RPD) [78–80]. The corresponding calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

$$RPD = \frac{SD}{RMSE} = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{RMSE} \tag{7}$$

Here, y_i represents the observed O_3 value for the i-th sample, $\hat{y}_i$ denotes the model-predicted O_3 value, and $\bar{y}$ is the mean of all observed O_3 values. n is the total number of samples. SD represents the standard deviation of the observed O_3 concentrations, and $RMSE$ is the root mean squared error between the predicted and observed values.

2.3.2. Shapley Additive Explanations

Based on the trained ML model, the SHAP method was applied to investigate the effects of 14 feature variables on near-surface O_3. Using the selected optimal model, SHAP was employed to perform interpretability analysis of the model outputs, thereby quantifying the contributions of individual feature variables to near-surface O_3. SHAP is an interpretable modeling framework based on game theory and local interpretability [81]. It quantifies the marginal contribution of each input variable by evaluating how the inclusion of that variable increases or decreases the model prediction for a given sample [82]. The corresponding calculation formula is as follows:

$$\varphi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[f(S \cup \{i\}) - f(S)] \tag{8}$$

In the equation, $\varphi_i(f)$ represents the contribution of feature x_i (e.g., solar radiation) to near-surface O_3 (Y). S denotes a subset of features that excludes x_i, while N represents the set of all features. $f(S)$ is the model prediction based on the feature subset S (without x_i), and $f(S \cup \{x_i\})$ denotes the prediction after including x_i in the subset.

To facilitate the comparison of feature importance, the contribution percentage of each feature was calculated based on its mean absolute SHAP value. The contribution percentage was calculated as follows:

$$Mean(|\varphi_i|) = \frac{1}{M}\sum_{j=1}^{M}\left|\varphi_i^{(j)}\right| \tag{9}$$

$$P_i = \frac{Mean(|\varphi_i|)}{\sum_{k=1}^{N} Mean(|\varphi_i|)} \times 100\% \tag{10}$$

Here, $\varphi_i^{(j)}$ denotes the SHAP value of feature i for the j-th sample, and M represents the total number of evaluation samples. P_i signifies the relative contribution percentage of feature i, N is the total number of feature variables (N = 14), and $Mean(|\varphi_i|)$ represents the global feature importance score.

2.3.3. Spatial Source Analysis Method

Backward Trajectory Analysis

To investigate the spatial origins of O_3, the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model was employed in combination with the GDAS dataset

to perform 72-h backward trajectory simulations at an arrival height of 500 m above ground level, aiming to identify potential transport pathways of air masses reaching the study area. The MeteoInfo software (Version 3.8.7) was used for trajectory simulation and analysis, and the TrajStat plugin within MeteoInfo was applied to conduct statistical analysis and visualization of the HYSPLIT-generated trajectories [28,30,83].

Potential Source Contribution Function (PSCF) Analysis

The PSCF method was employed to identify potential pollution source regions in the study area. The calculation formula is presented in Equation (9).

$$PSCF_{ij} = \frac{m_{ij}}{n_{ij}} \tag{11}$$

In the equation, $PSCF_{(ij)}$ represents the ratio of the number of trajectory endpoints m_{ij} exceeding a specified threshold to the total number of trajectory endpoints n_{ij} within grid cell *i,j*. To reduce the influence of small n_{ij} values in some grid cells, a weighting function W_{ij} is applied to the PSCF values. The determination of this weighting function follows the approaches of Fan et al. (2019) [84] and Zhang et al. (2023) [46], with the calculation formulas provided in Equations (10) and (11), where N_{ave} represents the average n_{ij} across all grid cells.

$$WPSCF_{ij} = PSCF_{ij} \times W_{ij} \tag{12}$$

$$W_{ij} = \begin{cases} 1.00, & n_{ij} > 3n_{ave} \\ 0.70, & 3n_{ave} \geq n_{ij} > 1.5n_{ave} \\ 0.40, & 1.5n_{ave} \geq n_{ij} > n_{ave} \\ 0.17, & n_{ave} \geq n_{ij} \end{cases} \tag{13}$$

Concentration-Weighted Trajectory (CWT) Analysis

Although the PSCF method can, to some extent, identify the probability of pollution occurrence in different regions and thereby infer potential source areas, its results are strongly dependent on the selected threshold and cannot effectively reflect the severity of pollution when concentrations exceed that threshold. In contrast, the CWT method calculates concentration-weighted trajectories, which can better represent the intensity of pollution. Therefore, the CWT approach was employed to quantify the relative contributions of different potential source regions [31,32]. The corresponding calculation formulas are provided in Equations (12) and (13).

$$CWT_{ij} = \frac{1}{\sum_{l=1}^{M} \tau_{ijl}} \sum_{l=1}^{M} C_l \tau_{ijl} \tag{14}$$

$$WCWT_{ij} = CWT_{ij} \times W_{ij} \tag{15}$$

In the equation, $CWT_{(ij)}$ represents the average weighted concentration in grid cell *i,j*. *M* denotes the total number of trajectories, ${\tau_{ij}}^{l}$ is the residence time of trajectory *l* within grid cell *i,j*, and C^l represents the pollutant concentration associated with trajectory *l* when passing through grid cell *i,j*.

3. Results

3.1. Analysis of Spatial Distribution Results

3.1.1. Correlation Analysis of Characteristic Variables

The correlations and collinearity among the feature variables were analyzed. Pearson was used to assess linear relationships between variables, while Spearman and Kendall's τ

were applied to evaluate nonlinear relationships. The VIF was used to assess collinearity among the features. As shown in Figure 2, the results of the Pearson, Spearman, and Kendall's τ analyses indicate generally weak correlations among the feature variables, with no excessively high correlation coefficients observed. With respect to collinearity, all feature variables exhibited VIF values below 10, suggesting the absence of significant multicollinearity among the predictors.

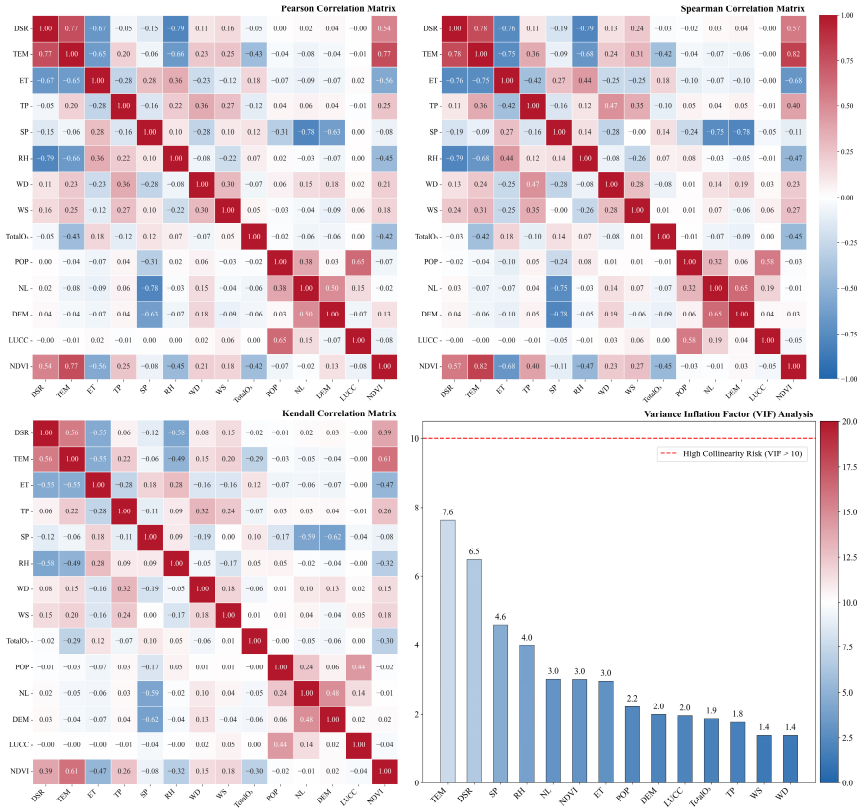


Figure 2. Correlation and collinearity analysis figure.

3.1.2. Model Comparison and Validation

Figure 3 compares the performance and efficiency of 16 artificial intelligence models in estimating O_3 concentrations across the TNSUA. All models were trained on the same dataset (sample size N = 9374) with hyperparameters optimized using Bayesian tuning. The best-performing models were XGBoost and CatBoost, both achieving R^2 values of 0.93 and RPD values above 3. Among them, XGBoost was faster, with a runtime of 2501.0 s. Models with R^2 values between 0.90 and 0.92 included LightGBM, RF, GBDT, AdaBoost, and ET, which also performed well. ET was the most efficient among these, with a runtime of 403.3 s. Models with R^2 values between 0.80 and 0.90 included SVR, DT, ResNeXt, MLP, DNN, LSTM, and ResNet, showing relatively lower performance, with RPD values below 3. The DL models in this group had particularly long runtimes, with the DNN model taking 19,223.7 s—the longest among all models. The poorest-performing models were CNN and DBN, with R^2 values of 0.59 and 0.49, respectively, indicating limited capability for estimating O_3 concentrations in this region. Moreover, DBN had a runtime of 10,481.4 s, reflecting relatively low efficiency. Considering both performance and computational efficiency, XGBoost was selected as the optimal model for estimating O_3 concentrations in the TNSUA, enabling high-resolution mapping of O_3 spatial distribution in the study area.

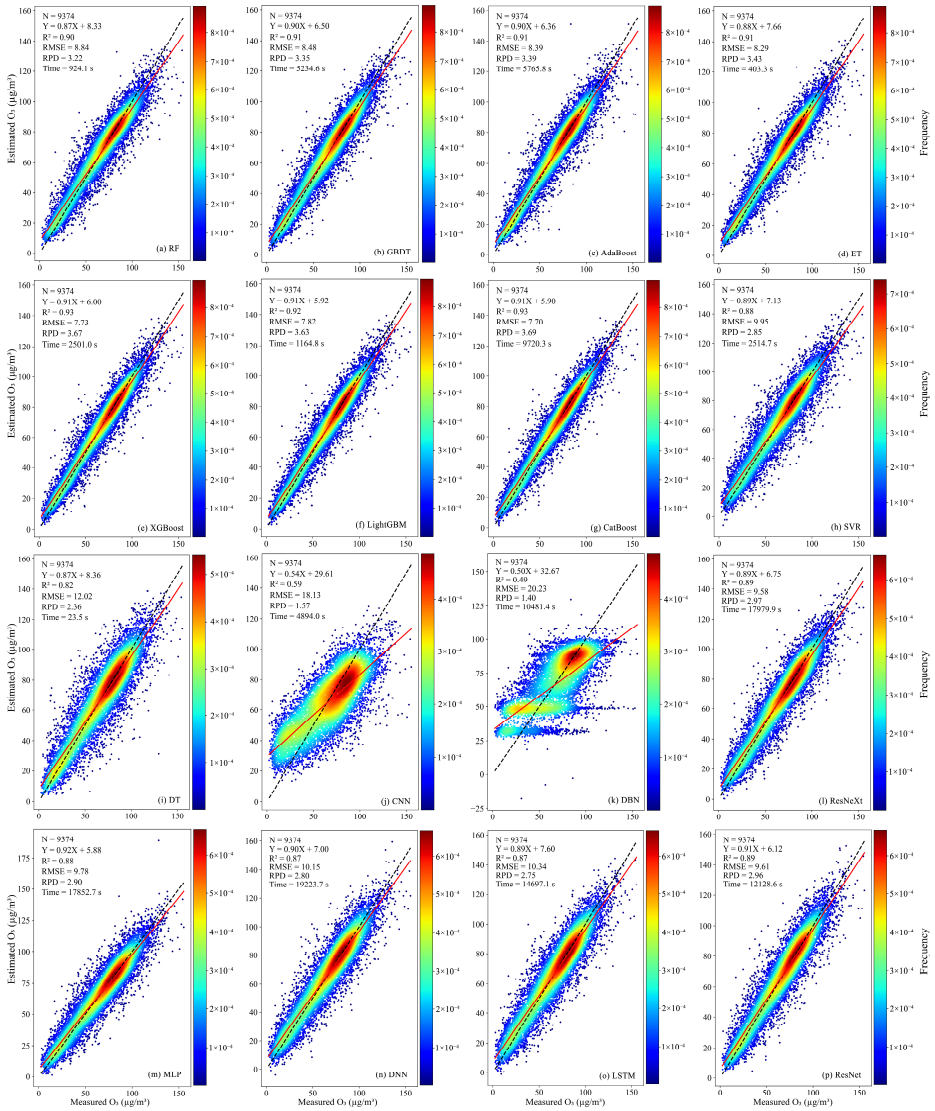


Figure 3. Model evaluation figure. The red solid line indicates the fitted regression line, and the black dashed line represents the 1:1 reference line.

3.1.3. Inversion Based on Excellent Inversion Model

Figure 4 presents the spatial distribution of O_3 concentrations estimated by the XGBoost model. According to the inversion results, the high-concentration areas of O_3 in the TNSUA are primarily located in the southern part of Urumqi, the southern part of Changji Prefecture, and the southern part of Tacheng. The reason for this phenomenon is that the southern part of Urumqi City, the southern part of Changji Prefecture, and the southern part of Tacheng Prefecture are located in the northern piedmont zone of the Tianshan Mountains. This region possesses high vegetation coverage that emits biogenic volatile organic compounds (BVOCs); moreover, because this area is situated at the southern edge of the cities, the pollutant emissions are also relatively high, thereby generating O_3 via photochemical reactions. Moreover, the high-concentration area in southern Urumqi exhibits the highest O_3 levels and the widest spatial extent. In June, the maximum O_3 concentration in southern Urumqi exceeds 120 $\mu g \cdot m^{-3}$, while the high-value areas in southern Changji and southern Tacheng are mainly in the range of 100–110 $\mu g \cdot m^{-3}$ during summer. The spatial inversion results also indicate a seasonal pattern of O_3, with higher concentrations in summer and lower in winter. The locations of high-value areas remain generally consistent across different months, showing that southern Urumqi, southern Changji, and southern Tacheng consistently have higher concentrations relative to other

regions throughout the year. This pattern is attributed to the complex interplay of multiple factors. O_3 concentrations are influenced by atmospheric processes and photochemical reactions. Higher temperatures in summer promote O_3 formation, leading to generally elevated concentrations in these regions. Vegetation also contributes by emitting BVOCs, which participate in photochemical reactions to generate O_3. The higher vegetation cover in southern Urumqi, southern Changji, and southern Tacheng further contributes to these areas being high-concentration zones for O_3.

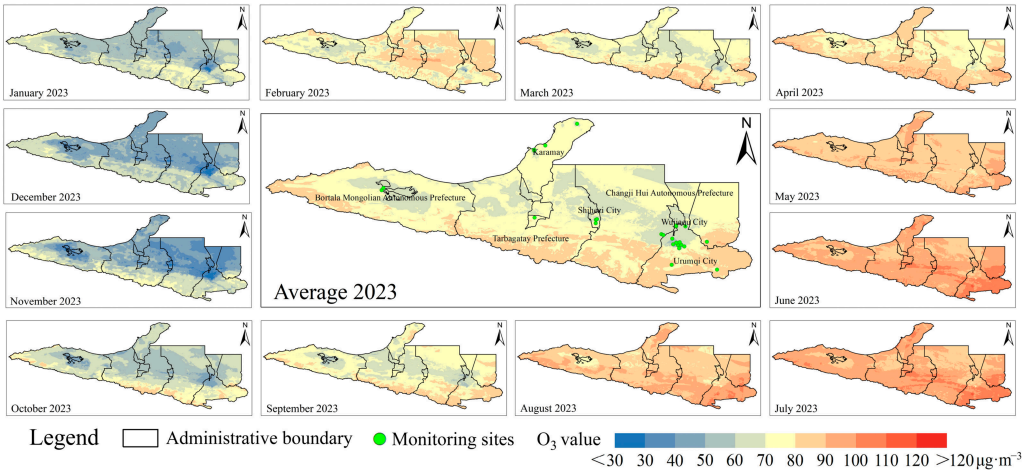


Figure 4. XGBoost-derived O_3 concentration inversion results.

3.1.4. Spatio-Temporal Feature Analysis

Figure 5 illustrates the spatiotemporal characteristics of O_3. On a daily scale, near-surface O_3 in the TNSUA in 2023 exhibited higher concentrations in summer and lower concentrations in winter, with spring concentrations generally exceeding those in autumn. On a monthly scale, the months with the highest average O_3 concentrations were June (93.98 μg·m^{-3}) and July (93.92 μg·m^{-3}), while the lowest monthly average occurred in November (51.50 μg·m^{-3}). Regarding seasonal spatial distribution, the high-concentration areas remained largely consistent across the four seasons, generally showing higher values around the periphery and lower values in the central region, with high-concentration zones primarily located in the southern part of the study area. The seasonal distribution also reflected that spring average concentrations exceeded those in autumn, with spring high-concentration areas more widely distributed. In summer, the high-concentration areas were most extensive. Regions with average O_3 concentrations of 90–100 μg·m^{-3} were mainly located in southern Urumqi, southern Changji, southern Tacheng, southern Bozhou, and Karamay. Notably, the high-concentration area in southern Urumqi reached the highest average O_3 levels, exceeding 110 μg·m^{-3}, marking it as a critical area for O_3 control in the TNSUA.

3.2. Analysis of Influencing Factors

Based on the continuous spatial distribution of near-surface O_3 derived from feature variables and the trained feature relationships, the key driving factors influencing near-surface O_3 were investigated. Using the XGBoost model combined with the SHAP algorithm, environmental variables with significant contributions to near-surface O_3 were identified. The analysis results (Figure 6) indicate that DSR and TEM are the primary drivers of near-surface O_3 in the TNSUA. DSR has the highest mean absolute SHAP value of 8.656 (20.85%), followed by TEM with a mean absolute SHAP value of 6.678 (16.08%). According to the SHAP value distributions, both DSR and TEM exhibit a strong positive correlation with O_3 concentrations (high-concentration areas correspond to positive SHAP

values), which is consistent with photochemical reaction mechanisms, where strong solar radiation and high temperatures accelerate the conversion of precursors. In contrast, WS and LUCC contributed less, with mean absolute SHAP values of 1.196 (2.88%) and 1.066 (2.57%), respectively. Their SHAP values are highly concentrated around zero, indicating that in 2023, the variability of WS and the spatial heterogeneity of LUCC had a relatively limited influence on O_3 concentrations compared with strong meteorological fluctuations such as DSR and TEM. This suggests that O_3 pollution in the region is more controlled by large-scale meteorological conditions and regional emissions rather than local microclimate or surface cover adjustments.

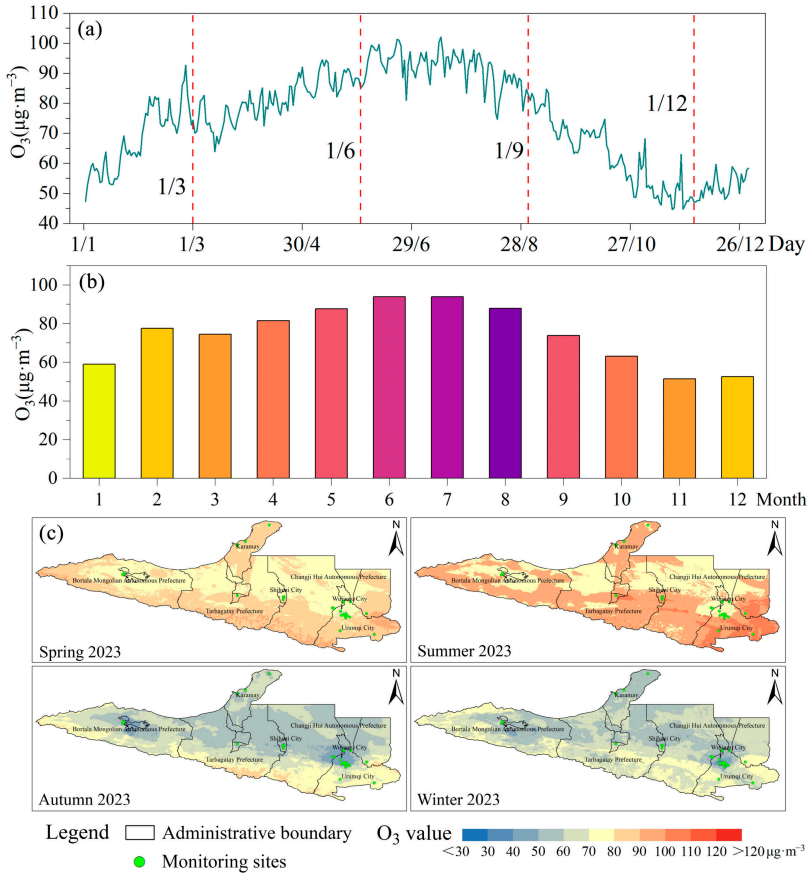


Figure 5. Spatiotemporal Characteristics Analysis Figure ((**a**): Daily mean variation in O_3, (**b**): Monthly mean variation in O_3, (**c**): Seasonal mean spatial distribution of O_3).

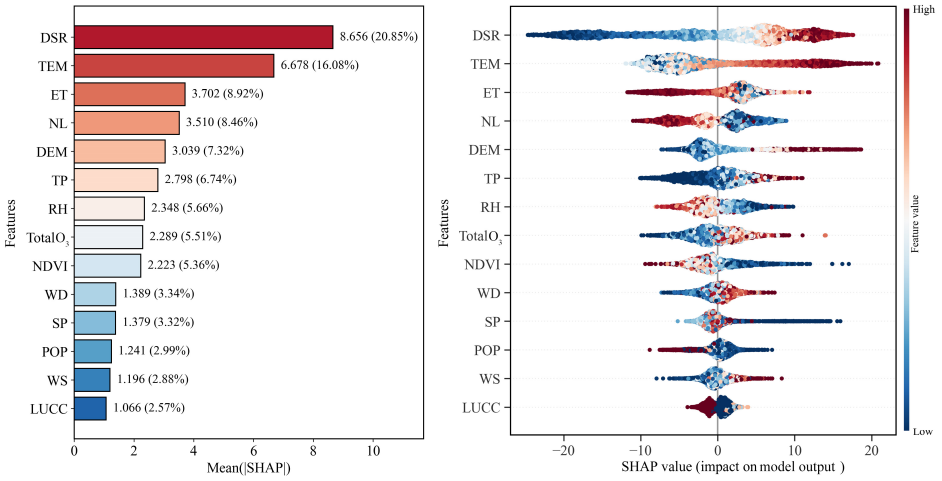


Figure 6. Feature importance and SHAP contribution analysis figure.

Based on the feature contribution analysis, feature dependence plots were generated for the most influential variables (DSR and TEM) to further investigate their threshold effects (Figure 7). Downward shortwave radiation (DSR) exhibits a significant positive contribution to near-surface O_3 concentrations. When DSR is below the threshold of 88.20 W/m^2, SHAP values are negative, indicating that low radiation conditions strongly inhibit O_3 formation. As DSR increases, SHAP values rise, and once it exceeds the median value of 155.02 W/m^2, the positive contribution sharply strengthens, highlighting DSR as a key driver of photochemical reactions. Air temperature shows a clear positive correlation with near-surface O_3 concentrations. When TEM is below 2.26 °C, SHAP values are mostly negative. As the temperature rises above 10 °C, the slope of the curve increases significantly, reflecting the strong influence of high temperatures on O_3 concentrations. This effect is likely due to the acceleration of photochemical reaction rates at higher temperatures and increased BVOC emissions from vegetation, which synergistically promote O_3 formation. Moreover, when examining the combined feature dependence plots, high-concentration samples (dark red) are concentrated in the high-value ranges of DSR and TEM (upper-right corner), where the SHAP positive contributions are maximal. This strong consistency indicates that the model accurately captures the driving mechanisms of high radiation and high temperature on near-surface O_3. The formation of near-surface O_3 in the TNSUA is strongly regulated by meteorological thresholds. For precise atmospheric management, periods when DSR and TEM exceed the thresholds (e.g., DSR > 88.20 W/m^2 and TEM > 2.26 °C) should be prioritized for control measures.

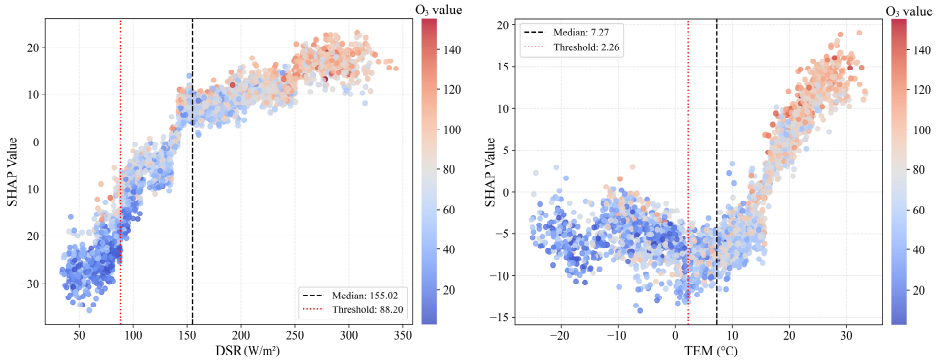


Figure 7. Dependence plots of key features.

3.3. Spatial Source Analysis

Based on the spatial distribution results of O_3, high concentrations in the TNSUA are mainly concentrated in southern Urumqi, southern Changji Prefecture, and southern Tacheng, with the highest O_3 levels observed in southern Urumqi. Therefore, the high-concentration area in southern Urumqi was selected for backward trajectory clustering analysis, potential source analysis, and CWT analysis, in order to investigate the sources and transport pathways of O_3 from a spatial perspective.

3.3.1. Results of Backward Trajectory Analysis

Figure 8 presents the clustering results of backward air mass trajectories for southern Urumqi. In spring, Cluster 1 (accounting for 41.30%) represents the dominant airflow trajectory affecting southern Urumqi, originating from eastern Kazakhstan and traveling eastward through the Ili River Valley before reaching the study area. In summer, the primary airflow trajectory (Cluster 1, accounting for 57.07%) originates from the northwest of the TNSUA, moving from northwest to southeast toward southern Urumqi. In both autumn and winter, the dominant airflow trajectories originate from Kazakhstan and move eastward to southern Urumqi. Overall, the backward trajectory analysis across all

four seasons indicates that the main airflows influencing southern Urumqi predominantly originate from the west, traveling eastward along the northern and southern foothills of the Tianshan Mountains before reaching the study area.

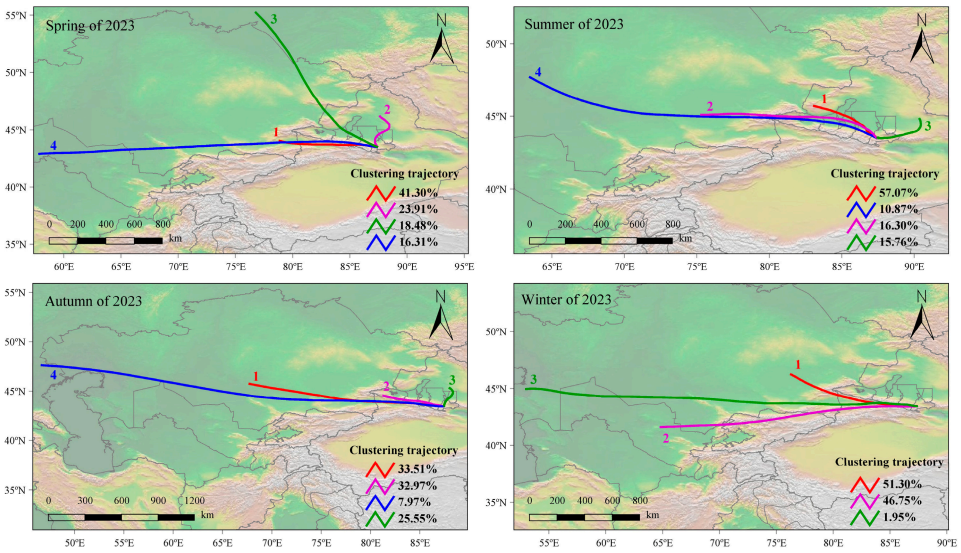


Figure 8. Backward trajectory clustering map for southern Urumqi.

3.3.2. Potential Source Analysis

Since backward trajectory clustering analysis can only identify the directional influence of air masses on pollutant transport and cannot quantify the contributions of potential source regions, the PSCF method was employed to analyze the likely source areas. The results are presented in Figure 9, where high WPSCF values indicate the presence of potential source regions. The analysis shows that in spring, the potential source regions for O_3 in southern Urumqi are mainly located in the northeastern part of the TNSUA (WPSCF > 0.9) and the Ili River Valley region (WPSCF > 0.7). In summer, the potential sources are primarily distributed around the TNSUA and eastern Kazakhstan (WPSCF > 0.9). In autumn, the potential sources are primarily located in the northeastern and southwestern edges of the urban agglomeration. In winter, the potential sources are mainly concentrated in the western and central parts of the study area. Considering all four seasons, the potential source regions of O_3 affecting southern Urumqi are generally concentrated in the surrounding areas of the TNSUA, eastern Kazakhstan, and northern Kyrgyzstan.

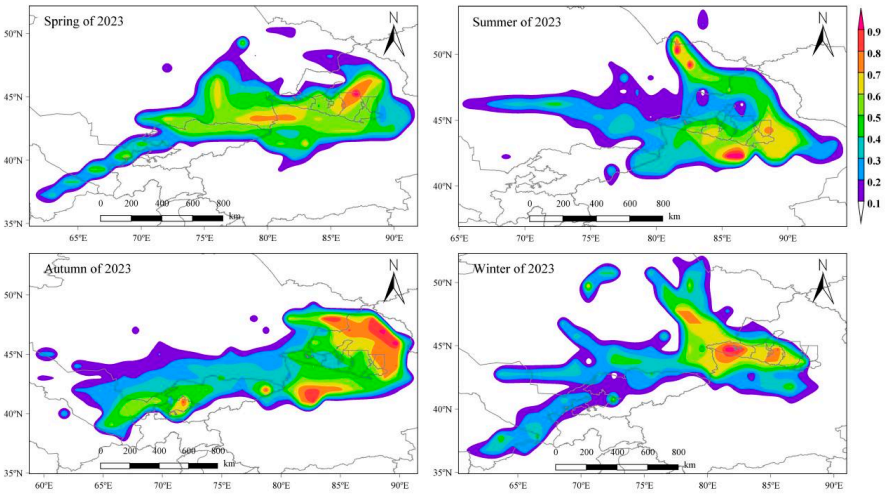


Figure 9. Seasonal PSCF distribution maps of O_3 for southern Urumqi.

3.3.3. Concentration Weight Trajectory Analysis

Although the PSCF method can analyze the contribution of potential source regions, it is constrained by threshold values and cannot assess the severity of pollution when concentrations exceed the threshold. Therefore, the CWT method was used to quantitatively reflect pollution intensity, and the results are presented in Figure 10. The analysis indicates that in spring, the high WCWT regions for O_3 in southern Urumqi are mainly distributed around the TNSUA, with the Ili River Valley in the southwest showing WCWT values greater than 90 $\mu g \cdot m^{-3}$. In summer, both the urban agglomeration and its surrounding areas are high WCWT regions for O_3 (WCWT > 100 $\mu g \cdot m^{-3}$), and eastern Kazakhstan also exhibits high WCWT values (>90 $\mu g \cdot m^{-3}$). In autumn, the high WCWT regions for O_3 are primarily located to the northeast and south of the study area, as well as near the border between Kazakhstan and Kyrgyzstan. In winter, the high WCWT regions are mainly concentrated around the TNSUA and its southern areas (WCWT > 60 $\mu g \cdot m^{-3}$), with generally lower WCWT values, reflecting the low O_3 concentrations typical of the winter season.

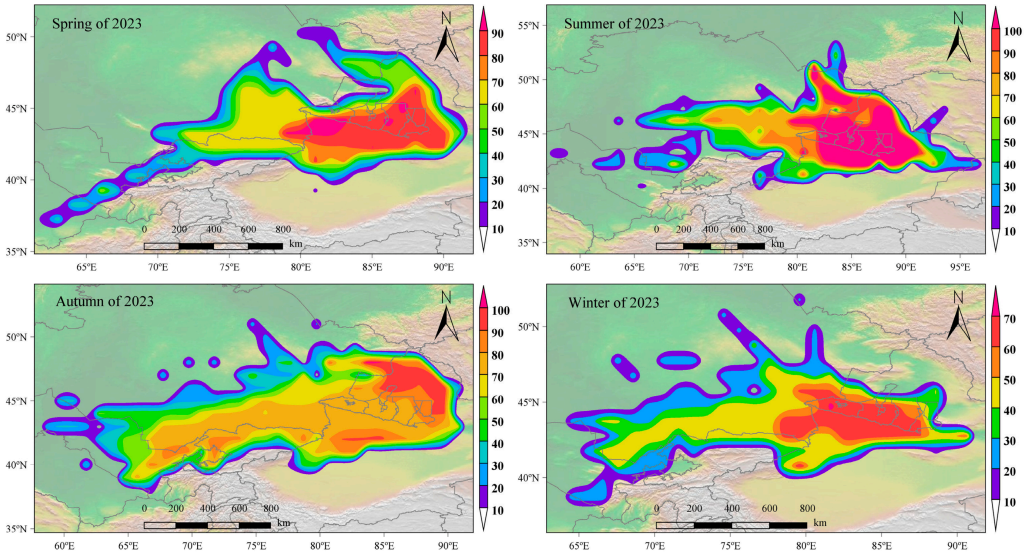


Figure 10. Seasonal CWT distribution maps of O_3 for southern Urumqi.

4. Discussion

4.1. Model Applicability and Analysis of Differences in Spatial Inversion

This study employed 16 artificial intelligence models and selected the optimal model to investigate the spatial distribution of O_3 in the TNSUA. The results were compared and discussed with previous studies focusing on near-surface O_3 distribution. Earlier studies have visualized spatial distributions based on monitoring station data [85,86], which can only provide a general characterization of spatial patterns but cannot produce continuous spatial fields. Other studies have applied classical geostatistical interpolation methods [87], such as Inverse Distance Weighted interpolation (IDW) and Kriging interpolation (Kriging). Although these methods generate continuous spatial surfaces, they are constrained by predefined mathematical formulations, resulting in limited spatial continuity [35,88]. In this study, both IDW and Kriging methods were applied to the study area and compared with artificial intelligence-based approaches, as shown in Figure 11. The interpolation results exhibit relatively poor continuity and are strongly influenced by station distribution and model assumptions. The interpolation results exhibit relatively poor continuity, whereas the XGBoost model shows much better spatial continuity. The fundamental reason for this difference lies in the distinct underlying mechanisms of the two types of approaches.

Traditional interpolation methods, such as IDW and Kriging, estimate O_3 concentrations mainly based on the spatial relationships among monitoring stations. In essence, these methods follow the 'First Law of Geography' and calculate O_3 concentrations according to spatial proximity. Therefore, their outputs are highly dependent on the distribution of monitoring stations and are prone to local over-smoothing effects, resulting in relatively poor continuity [89–91]. In contrast, the XGBoost model establishes nonlinear relationships between surface O_3 concentrations and 14 environmental variables (e.g., DSR, TEM, and NDVI), thereby overcoming the limitations of purely geometric interpolation and enabling the capture of complex nonlinear interactions associated with O_3 formation. Furthermore, conventional interpolation approaches are unable to characterize complex terrain effects and spatial heterogeneity, whereas the XGBoost model can learn the relationships between O_3 and variables such as DEM and LUCC, thereby identifying local spatial heterogeneity. In addition, O_3 concentrations are influenced by photochemical reactions and atmospheric transport processes, and their spatial distributions often deviate from the assumptions required by traditional interpolation methods. Therefore, artificial intelligence algorithms such as XGBoost are better able to capture spatial heterogeneity, leading to more spatially continuous results than those obtained using conventional interpolation methods. In studies using artificial intelligence for surface O_3 mapping, most previous research has relied on only one to three ML models [92,93], whereas this study systematically compared 16 models and identified the optimal one. The results indicate that the XGBoost model exhibited excellent performance (R^2 = 0.93, RPD > 3), which may be closely related to the unique meteorological and pollution characteristics of the TNSUA, a typical arid region. In this arid area, the photochemical processes of surface O_3 are influenced by extreme environmental conditions, including strong solar radiation and persistently high temperatures during summer, resulting in highly complex environmental conditions. Owing to its gradient boosting tree framework, XGBoost can effectively model the complex nonlinear relationships among high-dimensional environmental variables [94], while reducing bias and the risk of overfitting through iterative optimization. Therefore, it is more suitable for the spatial inversion of O_3 under the complex environmental conditions of arid regions. Compared with ensemble learning models such as RF and ET, XGBoost employs sequential tree building and gradient boosting, enabling it to more effectively identify complex interactions among environmental factors and thereby improve model generalization ability [95]. Although SVR and some DL models have advantages in handling complex pattern recognition tasks, their performance did not reach the optimal level due to limitations associated with sample size, characteristics of regional environmental variables, and model parameter sensitivity [96]. In particular, in regional air quality inversion studies with relatively limited sample sizes, DL models may fail to fully exploit their advantages because of their high parameter complexity [97]. Therefore, the superior performance of XGBoost in this study may fundamentally stem from its stable capability to model complex nonlinear relationships under limited sample conditions, as well as its strong adaptability to heterogeneous environmental variables from multiple sources. Furthermore, compared with previous studies focusing on large-scale regional O_3 distribution using artificial intelligence models [61,62], this study focuses on an urban agglomeration scale and achieves high-resolution mapping of near-surface O_3. This provides a more refined representation of spatial patterns and offers important references for air pollution prevention and control in urban agglomerations.

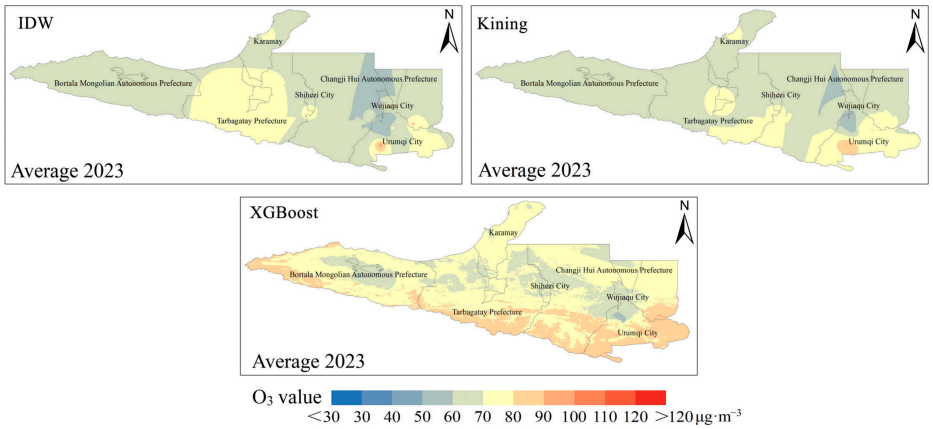


Figure 11. Comparison of interpolation results and artificial intelligence model results.

4.2. Discussion of Influencing Factors

Based on the optimal model, this study applied the SHAP algorithm to quantify the contributions of environmental variables to near-surface O_3 concentrations. During the training of the spatial distribution model, the nonlinear influence mechanisms of environmental variables on near-surface O_3 were systematically analyzed. The results indicate that meteorological factors—particularly DSR and TEM—are the dominant drivers of near-surface O_3 in the TNSUA, which is highly consistent with the physicochemical mechanisms of O_3 formation. Gagliardi et al. (2025) [98], in their study of the Basilicata Region (Southern Italy), also found that meteorological variables played a dominant role in O_3 variability throughout the study period, consistent with the findings of this study. However, their key variables included relative humidity, wind speed, and temperature, which differs from our results. In this study, relative humidity and wind speed were found to have relatively weak effects on O_3 in the study region, with solar radiation and temperature emerging as the primary drivers. This discrepancy may be attributed to regional differences: the study area of Gagliardi et al. (2025) [98] is a coastal region, where meteorological factors such as humidity have stronger influences, whereas the present study area is an inland region, where O_3 formation is more strongly controlled by large-scale meteorological conditions such as solar radiation and temperature. The identification of DSR and TEM as key drivers is consistent with previous studies investigating near-surface O_3 across China [61,62]. Furthermore, by analyzing threshold effects, this study provides deeper insights into the specific roles of DSR and TEM. The study area is a typical arid urban agglomeration. When the region exceeds the thresholds of solar radiation and temperature (i.e., DSR > 88.20 W/m^2, TEM > 2.26 °C), the positive effects of DSR and TEM on O_3 concentrations become more pronounced, leading to elevated O_3 levels. The solar radiation threshold of 88.20 W/m^2 may be associated with relatively weak solar radiation conditions, such as those occurring in the early morning or early spring. When radiation is below this threshold, the low solar irradiance limits photochemical reactions involved in O_3 formation. In contrast, when radiation exceeds this threshold, photochemical reactions are enhanced, thereby promoting O_3 production [52,99]. The temperature threshold of 2.26 °C may correspond to relatively low temperatures in early spring. When temperature exceeds this threshold, it may be associated with increased emissions of BVOCs from vegetation, and the increase in temperature further enhances the photochemical reactions involved in O_3 formation [100,101]. The identified thresholds offer valuable references for precise air pollution control in the region, suggesting that emissions should be strictly regulated during periods when these thresholds are exceeded to mitigate potential O_3 pollution peaks.

4.3. Analysis of Spatial Sources

This study conducted source analysis specifically on the high-value areas identified in the spatial distribution of O_3, enabling precise identification of the most severely polluted regions and thereby deriving conclusions about their spatial sources. Compared with prior studies—such as Guan et al. (2019) [28], where the authors used city centers as reference points, and Yin et al. (2021) [32], where the authors used monitoring stations for source tracing—this study integrates the spatial distribution results of O_3 to accurately pinpoint the exact latitude and longitude of high-concentration areas, making the spatial source analysis more precise. This provides a basis for targeted pollution control in high-value areas, aligning source analysis results with practical pollution management needs. The analysis indicates that the primary transport pathway for O_3 in southern Urumqi originates from western airflows, traveling along the northern and southern slopes of the Tianshan Mountains to the high-concentration areas. Similarly, Zhang et al. (2023) [46], in their study on aerosol sources in Urumqi, found that the aerosol optical depth (AOD) source trajectories were mainly distributed along the western slope of the Tianshan Mountains. This pattern may be attributed to the influence of the North Atlantic warm current on Xinjiang, with moisture primarily transported via western airflows [102], highlighting the significant role of western airflows in the spatial transport of O_3.

4.4. Control Strategies and Recommendations

Based on the identified spatial distribution characteristics, key driving factors, and pollution source transport pathways, the following targeted control strategies and recommendations are proposed to address the O_3 pollution problem in the TNSUA. First, in terms of spatial management, key prevention and control efforts should focus on the southern parts of Urumqi, Changji Prefecture, and the Tacheng region, allowing for the formulation of a joint prevention and control policy across these three areas. Second, since DSR and TEM are the critical driving factors influencing near-surface O_3 in this urban agglomeration, targeted regulations should be deployed against these two influencing factors; especially during the summer period, temporary emission reduction measures for O_3 precursors should be strengthened to prevent the occurrence of heavy pollution events. Third, given that western airflows constitute the primary transport pathway affecting O_3 pollution within the study region, collaborative management of potential pollution source emissions from the western direction should be tightened, while regional cooperation and information sharing with western territories should be enhanced to mitigate the impacts of spatial transport on localized O_3 pollution. Finally, as meteorological factors exert a profound influence on O_3 generation in arid urban clusters, integrating meteorological forecasting with O_3 pollution early-warning systems could substantially upgrade pollutant prevention and control. These measures may provide a scientific reference for precise O_3 pollution control, regional coordinated management, and the formulation of air quality policies in arid urban agglomerations.

4.5. Limitations and Outlook

This study, based on high-precision observational data and multi-source environmental variables, developed a near-surface O_3 spatial inversion model suitable for arid urban clusters, achieving high-accuracy daily spatial reconstruction of O_3 over the TNSUA. Using SHAP analysis, the study quantified the regulatory effects of key meteorological factors on O_3 formation, while combining HYSPLIT backward trajectories and PSCF/CWT methods to precisely identify transport pathways and potential source regions of high-concentration areas, providing scientific support for air quality monitoring and cross-regional coordinated emission reduction in urban clusters. However, certain limitations exist: the study relied

solely on data from 2023, lacking multi-year long-term monitoring, which may limit capturing long-term O_3 trends; moreover, the spatial inversion model does not fully account for complex chemical reactions and dynamic local emissions, and the potential source analysis has limited resolution in assessing local transport and terrain effects. In addition, while backward trajectory analysis combined with PSCF/WCWT methods can effectively identify potential source regions of ozone, the trajectory calculation is affected by the resolution of meteorological data (GDAS $1^\circ \times 1^\circ$), and this method mainly focuses on horizontal transport paths, without considering the influence of vertical transport processes on ozone transport. Future work could incorporate multi-year continuous observations and high-resolution emission inventories, combined with atmospheric chemistry models, to further improve the precision of O_3 formation mechanism analysis and extend the approach to a wider range of arid urban clusters, providing scientific guidance for precise air pollution control, regional coordinated management, and policy development.

5. Conclusions

(1) After comparing 16 artificial intelligence models, including DT, RF, ET, GBDT, CatBoost, XGBoost, and LightGBM, the XGBoost model showed the best performance for O_3 inversion in the TNSUA, with an R^2 of 0.93 and an RPD of 3.67, while also demonstrating relatively high computational efficiency. The second-best model was CatBoost, which also had an R^2 of 0.93, but its computational efficiency was lower. Based on the XGBoost inversion results, the high-value areas of O_3 concentration in the TNSUA are mainly concentrated in southern Urumqi, southern Changji Prefecture, and southern Tacheng region, with southern Urumqi exhibiting the highest concentrations and the most extensive high-value area, reaching a summer maximum exceeding 110 $\mu g \cdot m^{-3}$.

(2) By analyzing the contributions of the feature variables, DSR and TEM were identified as the key drivers influencing near-surface O_3 in the TNSUA. The SHAP value of DSR was 8.656 (accounting for 20.85% of the contribution), and TEM had a SHAP value of 6.678 (16.08% contribution). The threshold and median values of DSR affecting O_3 in the study area were 88.20 W/m^2 and 155.02 W/m^2, respectively, while TEM thresholds and median values were 2.26 °C and 7.27 °C. The generation of near-surface O_3 in the study area is strongly regulated by meteorological thresholds, suggesting that targeted pollution control should focus on periods when radiation and temperature exceed critical points (e.g., DSR > 155.02 W/m^2 and TEM > 7.27 °C).

(3) Based on precise spatial source analysis of the high-value areas derived from the O_3 distribution, the main transport pathway influencing the high-value O_3 regions in the TNSUA was found to be western airflows, primarily originating from northwestern Xinjiang and eastern Kazakhstan, traveling along the northern and southern slopes of the Tianshan Mountains to reach the high O_3 areas. WPSCF and WCWT results indicate that the surrounding areas of the TNSUA, eastern Kazakhstan, and northern Kyrgyzstan serve as potential source regions for O_3 pollution.

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

Funding: This research was funded by Diagnosis and evaluation project of bingtuan ecological security based on multi-source heterogeneous data (No. 2023CB008-23), Key areas of Xinjiang production and construction corps science and technology research plan project (No. 2024AB076),

Shihezi University Youth Cultivation Project (PYSK202403), Shihezi University School-level Science and Technology Planning System Project (CXFZ202310).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: Thanks to Zhou Xi of Hainan Normal University for their special guidance. During the preparation of this manuscript, the authors used ChatGPT (GPT-5) for the purposes of translation and polishing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Molina, M. Air pollution is a global problem with local solutions. *Nature* **2008**, *456*, 19. [CrossRef]
2. Li, K.; Jacob, D.J.; Liao, H.; Zhu, J.; Shah, V.; Shen, L.; Bates, K.H.; Zhang, Q.; Zhai, S. A two-pollutant strategy for improving ozone and particulate air quality in China. *Nat. Geosci.* **2019**, *12*, 906–910. [CrossRef]
3. Lin, S.; Xiao, L.; Wang, X. Does air pollution hinder technological innovation in China? A perspective of innovation value chain. *J. Clean. Prod.* **2021**, *278*, 123326. [CrossRef]
4. Williamson, C.E.; Zepp, R.G.; Lucas, R.M.; Madronich, S.; Austin, A.T.; Ballaré, C.L.; Norval, M.; Sulzberger, B.; Bais, A.F.; McKenzie, R.L. Solar ultraviolet radiation in a changing climate. *Nat. Clim. Change* **2014**, *4*, 434–441. [CrossRef]
5. Barnes, P.W.; Williamson, C.E.; Lucas, R.M.; Robinson, S.A.; Madronich, S.; Paul, N.D.; Bornman, J.F.; Bais, A.F.; Sulzberger, B.; Wilson, S.R. Ozone depletion, ultraviolet radiation, climate change and prospects for a sustainable future. *Nat. Sustain.* **2019**, *2*, 569–579. [CrossRef]
6. Stenke, A. Natural control on ozone pollution. *Nat. Clim. Change* **2020**, *10*, 101–102. [CrossRef]
7. Li, R.; Cui, L.; Hongbo, F.; Li, J.; Zhao, Y.; Chen, J. Satellite-based estimation of full-coverage ozone (O_3) concentration and health effect assessment across Hainan Island. *J. Clean. Prod.* **2020**, *244*, 118773. [CrossRef]
8. Achebak, H.; Garatachea, R.; Pay, M.T.; Jorba, O.; Guevara, M.; Pérez García-Pando, C.; Ballester, J. Geographic sources of ozone air pollution and mortality burden in Europe. *Nat. Med.* **2024**, *30*, 1732–1738. [CrossRef] [PubMed]
9. Wang, Y.; Yang, Y.; Yuan, Q.; Li, T.; Zhou, Y.; Zong, L.; Wang, M.; Xie, Z.; Ho, H.C.; Gao, M. Substantially underestimated global health risks of current ozone pollution. *Nat. Commun.* **2025**, *16*, 102. [CrossRef] [PubMed]
10. Zhu, L.; Wang, J.; Xu, Y.; Ma, N.; Song, X.; Qin, J.; Li, B.; Tsui, W.B.C.; Lv, L.; Zhang, T. Vertical Characteristics of an Ozone Pollution Episode in Hong Kong Under the Typhoon Mawar—A Case Study. *Remote Sens.* **2025**, *17*, 3904. [CrossRef]
11. Yang, W.; Wang, J.; Zhang, K.; Hao, Y. A novel air pollution forecasting, health effects, and economic cost assessment system for environmental management: From a new perspective of the district-level. *J. Clean. Prod.* **2023**, *417*, 138027. [CrossRef]
12. Rentschler, J.; Leonova, N. Global air pollution exposure and poverty. *Nat. Commun.* **2023**, *14*, 4432. [CrossRef] [PubMed]
13. Zhao, J.; Tang, Y.; Zhu, X.; Zhu, J. National environmental monitoring and local enforcement strategies. *Nat. Cities* **2025**, *2*, 58–69.
14. Musollari, S.; Pseftogkas, A.; Koukouli, M.-E.; Manders, A.; Segers, A.; Garane, K.; Balis, D. The Spatiotemporal Variability of Ozone and Nitrogen Dioxide in the Po Valley Using In Situ Measurements and Model Simulations. *Remote Sens.* **2025**, *17*, 1794. [CrossRef]
15. Daellenbach, K.R.; Uzu, G.; Jiang, J.; Cassagnes, L.-E.; Leni, Z.; Vlachou, A.; Stefenelli, G.; Canonaco, F.; Weber, S.; Segers, A. Sources of particulate-matter air pollution and its oxidative potential in Europe. *Nature* **2020**, *587*, 414–419. [CrossRef] [PubMed]
16. Ou, Y.; West, J.J.; Smith, S.J.; Nolte, C.G.; Loughlin, D.H. Air pollution control strategies directly limiting national health damages in the US. *Nat. Commun.* **2020**, *11*, 957. [CrossRef] [PubMed]
17. Ding, J.; Ren, C.; Wang, J.; Feng, Z.; Cao, S.-J. Spatial and temporal urban air pollution patterns based on limited data of monitoring stations. *J. Clean. Prod.* **2024**, *434*, 140359. [CrossRef]
18. Traversi, D.; Degan, R.; De Marco, R.; Gilli, G.; Pignata, C.; Ponzio, M.; Rava, M.; Sessarego, F.; Villani, S.; Bono, R. Mutagenic properties of PM2.5 air pollution in the Padana Plain (Italy) before and in the course of XX Winter Olympic Games of "Torino 2006". *Environ. Int.* **2008**, *34*, 966–970. [CrossRef] [PubMed]
19. Austin, E.; Coull, B.A.; Zanobetti, A.; Koutrakis, P. A framework to spatially cluster air pollution monitoring sites in US based on the PM2.5 composition. *Environ. Int.* **2013**, *59*, 244–254. [CrossRef] [PubMed]
20. Huang, R.-J.; Zhang, Y.; Bozzetti, C.; Ho, K.-F.; Cao, J.-J.; Han, Y.; Daellenbach, K.R.; Slowik, J.G.; Platt, S.M.; Canonaco, F. High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **2014**, *514*, 218–222. [CrossRef] [PubMed]
21. Liu, X.; Pu, X.; Lu, C.; Zhang, H.; Li, T.; Grieneisen, M.L.; Li, J.; Ma, N.; Yan, C.; Zhan, Y. Interpretable regional meteorological feature extraction enhances deep learning for extended 120-h $PM_{2.5}$ forecasting. *J. Clean. Prod.* **2024**, *483*, 144287. [CrossRef]

22. Zhang, Q.; Wang, Y.; Xiao, Q.; Geng, G.; Davis, S.J.; Liu, X.; Yang, J.; Liu, J.; Huang, W.; He, C. Long-range PM2.5 pollution and health impacts from the 2023 Canadian wildfires. *Nature* **2025**, *645*, 672–678. [CrossRef] [PubMed]
23. Li, Z.; Liu, Y.; Bi, J.; Hu, X. Forecasting high-resolution $PM_{2.5}$ concentrations in southeastern China by combining high-resolution satellite data and numerical simulation with machine learning. *J. Clean. Prod.* **2025**, *528*, 146759. [CrossRef]
24. Ji, W.; Wang, Y.; Zhuang, D. Spatial distribution differences in $PM_{2.5}$ concentration between heating and non-heating seasons in Beijing, China. *Environ. Pollut.* **2019**, *248*, 574–583. [CrossRef] [PubMed]
25. Shukla, K.; Kumar, P.; Mann, G.S.; Khare, M. Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi. *Sustain. Cities Soc.* **2020**, *54*, 101997. [CrossRef]
26. Tao, H.; Jawad, A.H.; Shather, A.H.; Al-Khafaji, Z.; Rashid, T.A.; Ali, M.; Al-Ansari, N.; Marhoon, H.A.; Shahid, S.; Yaseen, Z.M. Machine learning algorithms for high-resolution prediction of spatiotemporal distribution of air pollution from meteorological and soil parameters. *Environ. Int.* **2023**, *175*, 107931. [CrossRef] [PubMed]
27. Chi, Y.; Ren, Y.; Xu, C.; Zhan, Y. The spatial distribution mechanism of $PM_{2.5}$ and NO_2 on the eastern coast of China. *Environ. Pollut.* **2024**, *342*, 123122. [CrossRef] [PubMed]
28. Guan, Q.; Yang, Y.; Luo, H.; Zhao, R.; Pan, N.; Lin, J.; Yang, L. Transport pathways of PM_{10} during the spring in northwest China and its characteristics of potential dust sources. *J. Clean. Prod.* **2019**, *237*, 117746. [CrossRef]
29. Tian, Y.; Jiang, Y.; Liu, Q.; Xu, D.; Zhao, S.; He, L.; Liu, H.; Xu, H. Temporal and spatial trends in air quality in Beijing. *Landsc. Urban Plan.* **2019**, *185*, 35–43. [CrossRef]
30. Wang, F.; Li, Z.; Wang, F.; You, X.; Xia, D.; Zhang, X.; Zhou, X. Air pollution in a low-industry city in China's silk road economic belt: Characteristics and potential sources. *Front. Earth Sci.* **2021**, *9*, 527475. [CrossRef]
31. Zhao, N.; Wang, G.; Li, G.; Lang, J.; Zhang, H. Air pollution episodes during the COVID-19 outbreak in the Beijing–Tianjin–Hebei region of China: An insight into the transport pathways and source distribution. *Environ. Pollut.* **2020**, *267*, 115617. [CrossRef] [PubMed]
32. Yin, X.; Kang, S.; Rupakheti, M.; de Foy, B.; Li, P.; Yang, J.; Wu, K.; Zhang, Q.; Rupakheti, D. Influence of transboundary air pollution on air quality in southwestern China. *Geosci. Front.* **2021**, *12*, 101239. [CrossRef]
33. Yin, H.; Xu, L. Comparative study of PM_{10}/$PM_{2.5}$-bound PAHs in downtown Beijing, China: Concentrations, sources, and health risks. *J. Clean. Prod.* **2018**, *177*, 674–683. [CrossRef]
34. Chen, G.; Jin, Z.; Li, S.; Jin, X.; Tong, S.; Liu, S.; Yang, Y.; Huang, H.; Guo, Y. Early life exposure to particulate matter air pollution (PM_1, $PM_{2.5}$ and PM_{10}) and autism in Shanghai, China: A case-control study. *Environ. Int.* **2018**, *121*, 1121–1127. [CrossRef] [PubMed]
35. Ma, J.; Ding, Y.; Cheng, J.C.P.; Jiang, F.; Wan, Z. A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for $PM_{2.5}$. *J. Clean. Prod.* **2019**, *237*, 117729. [CrossRef]
36. Zhang, X.; Osei, F.; Stein, A.; Cheng, C.; Maji, K.J. Temporal and spatial evolution of short-term exposure to ozone pollution: Its health impacts in China based on a meta-analysis. *J. Clean. Prod.* **2022**, *373*, 133938. [CrossRef]
37. Qin, M.; She, Y.; Wang, M.; Wang, H.; Chang, Y.; Tan, Z.; An, J.; Huang, J.; Yuan, Z.; Lu, J. Increased urban ozone in heatwaves due to temperature-induced emissions of anthropogenic volatile organic compounds. *Nat. Geosci.* **2025**, *18*, 50–56. [CrossRef] [PubMed]
38. Kang, H.; Liu, B.; Hong, L.; Shi, J.; Lu, H.; Zhang, Y.; Guo, Z. Emission Control and Sensitivity Regime Shifts Drive the Decline in Extreme Ozone Concentration in the Sichuan Basin During 2015–2024. *Remote Sens.* **2025**, *17*, 3238. [CrossRef]
39. Zhang, L.; Liu, W.; Hou, K.; Lin, J.; Song, C.; Zhou, C.; Huang, B.; Tong, X.; Wang, J.; Rhine, W. Air pollution exposure associates with increased risk of neonatal jaundice. *Nat. Commun.* **2019**, *10*, 3741. [CrossRef] [PubMed]
40. Song, J.; Wang, B.; Yang, W.; Duan, H.; Liu, X. Extracting critical supply chains driving air pollution in China. *J. Clean. Prod.* **2020**, *276*, 124282. [CrossRef]
41. Zhang, Y.; Wang, L.; Tang, Z.; Zhang, K.; Wang, T. Spatial effects of urban expansion on air pollution and eco-efficiency: Evidence from multisource remote sensing and statistical data in China. *J. Clean. Prod.* **2022**, *367*, 132973. [CrossRef]
42. Wang, Y.; Zhao, Y.; Liu, Y.; Jiang, Y.; Zheng, B.; Xing, J.; Liu, Y.; Wang, S.; Nielsen, C.P. Sustained emission reductions have restrained the ozone pollution over China. *Nat. Geosci.* **2023**, *16*, 967–974. [CrossRef]
43. Geng, G.; Liu, Y.; Liu, Y.; Liu, S.; Cheng, J.; Yan, L.; Wu, N.; Hu, H.; Tong, D.; Zheng, B. Efficacy of China's clean air actions to tackle $PM_{2.5}$ pollution between 2013 and 2020. *Nat. Geosci.* **2024**, *17*, 987–994. [CrossRef]
44. Zhou, H.; Yue, X.; Dai, H.; Geng, G.; Yuan, W.; Chen, J.; Shen, G.; Zhang, T.; Zhu, J.; Liao, H. Recovery of ecosystem productivity in China due to the Clean Air Action plan. *Nat. Geosci.* **2024**, *17*, 1233–1239. [CrossRef]
45. Ma, W.; Ding, J.; Wang, R.; Wang, J. Drivers of $PM_{2.5}$ in the urban agglomeration on the northern slope of the Tianshan Mountains, China. *Environ. Pollut.* **2022**, *309*, 119777. [CrossRef] [PubMed]
46. Zhang, Z.; Ding, J.; Chen, X.; Wang, J. Aerosols characteristics, sources, and drive factors analysis in typical megacities, NW China. *J. Clean. Prod.* **2023**, *403*, 136879. [CrossRef]

47. Yu, T.; Abulizi, A.; Xu, Z.; Jiang, J.; Akbar, A.; Ou, B.; Xu, F. Evolution of environmental quality and its response to human disturbances of the urban agglomeration in the northern slope of the Tianshan Mountains. *Ecol. Indic.* **2023**, *153*, 110481. [CrossRef]
48. Sorooshian, A.; Arellano, A.F.; Fraser, M.P.; Herckes, P.; Betito, G.; Betterton, E.A.; Braun, R.A.; Guo, Y.; Mirrezaei, M.A.; Roychoudhury, C. Ozone in the desert Southwest of the United States: A synthesis of past work and steps ahead. *ACS EST Air* **2024**, *1*, 62–79. [CrossRef]
49. Liu, M.; Wang, X.; Wang, Y. Interactions between aerosols and surface ozone in arid and semi-arid regions of China. *Environ. Monit. Assess.* **2024**, *196*, 390. [CrossRef] [PubMed]
50. Mirrezaei, M.A.; Arellano, A.; Guo, Y.; Roychoudhury, C.; Sorooshian, A. Ozone production over arid regions: Insights into meteorological and chemical drivers. *Environ. Res. Commun.* **2024**, *6*, 051009. [CrossRef]
51. Bhuiyan, M.A.M.; Sahi, R.K.; Islam, M.R.; Mahmud, S. Machine learning techniques applied to predict tropospheric ozone in a semi-arid climate region. *Mathematics* **2021**, *9*, 2901. [CrossRef]
52. Hu, C.; Kang, P.; Jaffe, D.A.; Li, C.; Zhang, X.; Wu, K.; Zhou, M. Understanding the impact of meteorology on ozone in 334 cities of China. *Atmos. Environ.* **2021**, *248*, 118221. [CrossRef]
53. Song, X.; Tian, P.; Yu, Z.; Liu, X.; Wang, T.; Zhang, M.; Li, J.; Cao, X.; Liang, J.; Ren, Y. Probing the mechanism of prolonged ozone pollution in arid and semi-arid urban areas. *Urban Clim.* **2025**, *61*, 102399. [CrossRef]
54. Su, Y.; Yang, Y.; Zhang, Y.; Yin, J.; Mao, F. Interannual Divergence of $PM_{2.5}$ and O_3 Pollution in East-Central China Revealed by Seamless 24-Hour Retrievals (2017–2023). *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 1–12. [CrossRef]
55. Liu, H.; Liu, J.; Liu, Y.; Yi, K.; Yang, H.; Xiang, S.; Ma, J.; Tao, S. Spatiotemporal variability and driving factors of ground-level summertime ozone pollution over eastern China. *Atmos. Environ.* **2021**, *265*, 118686. [CrossRef]
56. Wang, M.; Chen, X.; Jiang, Z.; He, T.-L.; Jones, D.; Liu, J.; Shen, Y. Meteorological and anthropogenic drivers of surface ozone change in the North China Plain in 2015–2021. *Sci. Total Environ.* **2024**, *906*, 167763. [CrossRef] [PubMed]
57. Luo, Y.; Xu, L.; Li, Z.; Zhou, X.; Zhang, X.; Wang, F.; Peng, J.; Cao, C.; Chen, Z.; Yu, H. Air pollution in heavy industrial cities along the northern slope of the Tianshan Mountains, Xinjiang: Characteristics, meteorological influence, and sources. *Environ. Sci. Pollut. Res.* **2023**, *30*, 55092–55111. [CrossRef]
58. Shen, L.; Hao, J.; Cui, L.; Chen, H.; Wang, L.; Wang, Y.; Tong, Y. Assessing the Impact of Agricultural Land Consolidation on Ecological Environment Quality in Arid Areas Based on an Improved Water Benefit-Based Ecological Index. *Remote Sens.* **2025**, *17*, 2987. [CrossRef]
59. Chen, Z.; Li, Z.; Xu, L.; Zhou, X.; Zhang, X.; Wang, F.; Luo, Y. Gaseous and particulate pollution in the Wu-Chang-Shi urban agglomeration on the northern slope of Tianshan Mountains from 2017 to 2021. *Atmosphere* **2022**, *14*, 91. [CrossRef]
60. Fan, H.; Zhao, C.; Yang, Y. A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018. *Atmos. Environ.* **2020**, *220*, 117066. [CrossRef]
61. Cheng, F.; Li, Z.; Yang, Z.; Li, R.; Wang, D.; Jia, A.; Li, K.; Zhao, B.; Wang, S.; Yin, D. First retrieval of 24-hourly 1-km-resolution gapless surface ozone (O_3) from space in China using artificial intelligence: Diurnal variations and implications for air quality and phytotoxicity. *Remote Sens. Environ.* **2025**, *316*, 114482. [CrossRef]
62. 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 Sens. Environ.* **2022**, *270*, 112775. [CrossRef]
63. Abdullah, S.; Napi, N.N.; Ahmed, A.N.; Mansor, W.N.; Mansor, A.A.; Ismail, M.; Abdullah, A.M.; Ramly, Z.T. Development of Multiple Linear Regression for Particulate Matter (PM_{10}) Forecasting during Episodic Transboundary Haze Event in Malaysia. *Atmosphere* **2020**, *11*, 289. [CrossRef]
64. Babu, K.M.; Beulah, J.R. Air quality prediction based on supervised machine learning methods. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 206–212. [CrossRef]
65. Yang, C.; Du, W.J.; He, R.-L.; Hu, Y.-R.; Liu, H.; Huang, T.; Li, W.-W. Spatiotemporal patterns of methane and nitrous oxide emissions in China's inland waters identified by machine learning technique. *ACS EST Water* **2023**, *4*, 936–947. [CrossRef]
66. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
67. Difaizi, T.Z.; Camille, O.P.L.; Nadine, A.; Kalita, S.; Kumar, A.; Rakesh, N. Comparative analysis of machine learning techniques for air quality index prediction for Indian cities. *AIP Conf. Proc.* **2023**, *2023*, 050014. [CrossRef]
68. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
69. Morapedi, T.D.; Obagbuwa, I.C. Air pollution particulate matter ($PM_{2.5}$) prediction in South African cities using machine learning techniques. *Front. Artif. Intell.* **2023**, *6*, 1230087. [CrossRef] [PubMed]
70. Xiong, K.; Xie, X.; Huang, L.; Hu, J. Improved O_3 predictions in China by combining chemical transport model and multi-source data with machining learning techniques. *Atmos. Environ.* **2024**, *318*, 120269. [CrossRef]
71. Du, K.-L.; Leung, C.-S.; Mow, W.H.; Swamy, M.N.S. Perceptron: Learning, generalization, model selection, fault tolerance, and role in the deep learning era. *Mathematics* **2022**, *10*, 4730. [CrossRef]

72. Hu, J.; Cui, C.; Gao, J.; Zhang, A.; Liu, X.; Zhang, S.; Fang, Y. A Method for Locating Air Pollution Sources Combining Sensor Network and Machine Learning. *Water Air Soil Pollut.* **2026**, *237*, 421. [CrossRef]
73. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [CrossRef] [PubMed]
74. Wang, S.; McGibbon, J.; Zhang, Y. Predicting high-resolution air quality using machine learning: Integration of large eddy simulation and urban morphology data. *Environ. Pollut.* **2024**, *344*, 123371. [CrossRef] [PubMed]
75. Sun, C.; Li, J.; Sulaiman, R.; Alotaibi, B.S.; Elattar, S.; Abuhussain, M. Air quality prediction and multi-task offloading based on deep learning methods in edge computing. *J. Grid Comput.* **2023**, *21*, 32. [CrossRef]
76. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *arXiv* **2016**, arXiv:1512.03385.
77. Qiao, B.; Wang, Y.; Yao, L.; Han, D.; Wu, G. Attention mechanism fusion neural network for typhoon path prediction. *Appl. Intell.* **2025**, *55*, 244.
78. Van Roode, S.; Ruiz-Aguilar, J.J.; González-Enrique, J.; Turias, I.J. An artificial neural network ensemble approach to generate air pollution maps. *Environ. Monit. Assess.* **2019**, *191*, 727. [CrossRef] [PubMed]
79. Shen, L.; Zhao, J.; Wang, T.; Yao, C.; Hao, J.; Cheng, Z.; Wang, Y.; Zhao, X. Spatiotemporal evolution and driving mechanisms of cropland soil salinity in the irrigated area of the Manas River Basin based on random forest model. *Agric. Water Manag.* **2026**, *331*, 110437. [CrossRef]
80. Xue, Y.; Ma, J.; Ma, S.; Chen, Y.; Sun, X.; Ren, M.; Shen, L. Climate-Driven Redistribution of Early-Spring Ephemeral Plant Communities in Cold Arid Deserts: Evidence from the Gurbantunggut Desert, China. *Plants* **2026**, *15*, 1586. [CrossRef] [PubMed]
81. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
82. Chen, M.; Liu, J.; Chu, B.; Zhao, D.; Li, R.; Chen, T.; Ma, Q.; Wang, Y.; Zhang, P.; Li, H. Research on the influencing factors of $PM_{2.5}$ in China at different spatial scales based on machine learning algorithm. *Environ. Int.* **2025**, *200*, 109536. [CrossRef] [PubMed]
83. Wang, Y.Q. MeteoInfo: GIS software for meteorological data visualization and analysis. *Meteorol. Appl.* **2014**, *21*, 360–368.
84. Fan, W.; Qin, K.; Xu, J.; Yuan, L.; Li, D.; Jin, Z.; Zhang, K. Aerosol vertical distribution and sources estimation at a site of the Yangtze River Delta region of China. *Atmos. Res.* **2019**, *217*, 128–136. [CrossRef]
85. Lei, R.; Nie, D.; Zhang, S.; Yu, W.; Ge, X.; Song, N. Spatial and temporal characteristics of air pollutants and their health effects in China during 2019–2020. *J. Environ. Manag.* **2022**, *317*, 115460. [CrossRef]
86. Zhou, W.; Chen, C.; Lei, L.; Fu, P.; Sun, Y. Temporal variations and spatial distributions of gaseous and particulate air pollutants and their health risks during 2015–2019 in China. *Environ. Pollut.* **2021**, *272*, 116031. [CrossRef] [PubMed]
87. Tan, S.; Xie, D.; Ni, C.; Zhao, G.; Shao, J.; Chen, F.; Ni, J. Spatiotemporal characteristics of air pollution in Chengdu-Chongqing urban agglomeration (CCUA) in Southwest, China: 2015–2021. *J. Environ. Manag.* **2023**, *325*, 116503. [CrossRef]
88. Zhang, Z.; Zhou, Y.; Huang, X. Applicability of GIS-based spatial interpolation and simulation for estimating the soil organic carbon storage in karst regions. *Glob. Ecol. Conserv.* **2020**, *21*, e00849. [CrossRef]
89. Kravchenko, A.; Bullock, D.G. A comparative study of interpolation methods for mapping soil properties. *Agron. J.* **1999**, *91*, 393–400. [CrossRef]
90. Deligiorgi, D.; Philippopoulos, K. Spatial interpolation methodologies in urban air pollution modeling: Application for the greater area of metropolitan Athens, Greece. *Adv. Air Pollut.* **2011**, *17*, 341–362.
91. Han, S.; Kundhikanjana, W.; Towashiraporn, P.; Stratoulias, D. Interpolation-based fusion of Sentinel-5P, SRTM, and regulatory-grade ground stations data for producing spatially continuous maps of $PM_{2.5}$ concentrations nationwide over Thailand. *Atmosphere* **2022**, *13*, 161. [CrossRef]
92. Zhan, Y.; Luo, Y.; Deng, X.; Grieneisen, M.L.; Zhang, M.; Di, B. Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. *Environ. Pollut.* **2018**, *233*, 464–473. [CrossRef] [PubMed]
93. 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] [PubMed]
94. Jung, C.-R.; Chen, W.; Chen, W.-T.; Su, S.-H.; Chen, B.-T.; Chang, L.; Hwang, B.-F. A machine learning model for estimating daily maximum 8-hour average ozone concentrations using OMI and MODIS products. *Atmos. Environ.* **2024**, *331*, 120587. [CrossRef]
95. Liu, W.; Niu, Z.; Feng, X.; Zhou, W.; Liang, D.; Wang, G.; Liu, L. Determining the key meteorological factors affecting atmospheric CO2 and CH4 using machine learning algorithms at a suburban site in China. *Urban Clim.* **2025**, *59*, 102312. [CrossRef]
96. Petrić, V.; Hussain, H.; Časni, K.; Vuckovic, M.; Schopper, A.; Andrijić, Ž.U.; Kecorius, S.; Madueno, L.; Kern, R.; Lovrić, M. Ensemble machine learning, deep learning, and time series forecasting: Improving prediction accuracy for hourly concentrations of ambient air pollutants. *Aerosol Air Qual. Res.* **2024**, *24*, 230317. [CrossRef]
97. Agbehadji, I.E.; Obagbuwa, I.C. Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction. *Atmosphere* **2024**, *15*, 1352. [CrossRef]
98. Gagliardi, R.V.; Andenna, C. Exploring the Influencing Factors of Surface Ozone Variability by Explainable Machine Learning: A Case Study in the Basilicata Region (Southern Italy). *Atmosphere* **2025**, *16*, 491. [CrossRef]

99. 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]
100. Ma, Y.; Ma, B.; Jiao, H.; Zhang, Y.; Xin, J.; Yu, Z. An analysis of the effects of weather and air pollution on tropospheric ozone using a generalized additive model in Western China: Lanzhou, Gansu. *Atmos. Environ.* **2020**, *224*, 117342. [CrossRef]
101. Huang, M.; Feng, Z.; Liao, T. Shift of surface O_3-NOx-VOC sensitivity with temperature in the Guangdong-Hong Kong-Macao Greater Bay Area, South China. *Environ. Pollut.* **2025**, *372*, 125974. [CrossRef] [PubMed]
102. Shen, Z.; Zhang, Q.; Singh, V.P.; Pokhrel, Y.; Li, J.; Xu, C.-Y.; Wu, W. Drying in the low-latitude Atlantic Ocean contributed to terrestrial water storage depletion across Eurasia. *Nat. Commun.* **2022**, *13*, 1849. [CrossRef] [PubMed]

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.