

Article

Surface Ozone-Induced Yield Losses and Economic Costs of Winter Wheat Across China: Spatiotemporal Estimates Based on Reconstructed M7 and AOT40 Exposures (2015–2022)

Hui Zheng ^{1,2,3} , Rong Li ^{1,3}, Xian Su ^{1,3}, Yudi Li ^{1,3} and Junling Jin ^{1,2,3,*}

¹ Guangxi Key Laboratory of Environmental Processes and Remediation in Ecologically Fragile Regions, Guangxi Normal University, 15 Yucai Road, Guilin 541004, China; zhenghui@vip.henu.edu.cn (H.Z.); lirong@stu.gxnu.edu.cn (R.L.); suxian@stu.gxnu.edu.cn (X.S.); liyudi@stu.gxnu.edu.cn (Y.L.)

² University Engineering Research Center of Green Remediation and Low Carbon Development for Lijiang River Basin, 15 Yucai Road, Guilin 541004, China

³ College of Environment and Resources, Guangxi Normal University, 15 Yucai Road, Guilin 541004, China

* Correspondence: junling.jin@gxnu.edu.cn

Abstract

Surface ozone (O₃) adversely affects winter wheat production. However, due to the lack of spatially continuous ozone exposure datasets directly linked to crop yield responses, the assessment of ozone-induced crop losses remains uncertain. Winter wheat, a major food crop in China, is highly sensitive to ozone. Its primary cultivation region—the North China Plain (NCP)—overlaps with severe ozone pollution areas, creating significant risk asuperposition. To address this, we employed a spatiotemporal LightGBM (ST-LightGBM) model with ground observations, atmospheric reanalysis, and satellite retrievals to reconstruct a 10 km resolution dataset of daily 7 h average (M7) and daytime accumulated ozone exposure index (AOT40) across China from 2015 to 2022. The model performed robustly; SHAP analysis showed that downward solar radiation and 2 m temperature explained over 42% of ozone variability. From 2015 to 2022, despite pollution controls and the COVID-19 pandemic, surface ozone showed a phased upward trend (0.83 μg m⁻³ yr⁻¹), with O₃ concentrations increasing across 79.22% of the study area and the NCP most affected. Applying M7 and AOT40, the average annual relative yield loss of winter wheat ranged from 3.44% to 7.20% (M7) and from 26.50% to 36.54% (AOT40), corresponding to economic losses of 2.21–4.44 billion USD and 16.16–26.08 billion USD. Both metrics showed consistent trends, but AOT40 losses were substantially higher, highlighting the need to incorporate cumulative high-concentration exposure into agricultural risk assessments. These findings demonstrate that O₃ pollution poses a substantial threat to winter wheat production in China. The reconstructed exposure metrics provide a quantitative basis for identifying high-risk agricultural regions and inform targeted mitigation strategies to safeguard food security under intensifying O₃ pollution.



Academic Editor: Gianni Bellocchi

Received: 23 July 2026

Revised: 28 August 2026

Accepted: 29 August 2026

Published: 31 August 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: ozone pollution; ST-LightGBM; SHAP; winter wheat; relative yield loss

1. Introduction

Air pollution has become a critical global environmental issue since the onset of the industrial age. It not only impedes regional economic development, but also exerts detrimental effects on the ecological environment and human health [1,2]. Since 2013, China has established air pollution monitoring stations to obtain real-time pollution levels

and implement strict air pollution control policies to significantly reduce domestic air pollution levels. Owing to the implementation of air pollution prevention and control policies, particulate matter concentration in China has decreased [3,4]. However, this progress has been accompanied by a concurrent increase in O₃ concentration [5–7]. High levels of O₃ are associated with an increased risk to human health [8,9]. Moreover, O₃ pollution significantly reduces the yield of key food crops. Research indicates that O₃ exposure results in yield reductions of 33% for wheat, 23% for rice, and 9% for corn [10]. Consequently, O₃ pollution has become a core issue for pollution prevention and control. Given the current O₃ concentration levels in China, acquiring accurate spatiotemporal distribution data on O₃ concentrations is crucial.

There are two primary methods for obtaining O₃ data: site and remote sensing monitoring. The number of air pollution monitoring stations in China has increased. Air pollution monitoring stations provide continuous time-series data on O₃ concentrations at specific locations [7]. However, most stations are located in the eastern and central regions, with sparse distribution in the west, and predominantly in urban areas. This uneven distribution makes it challenging to obtain accurate and comprehensive spatiotemporal data on O₃ concentration. The development of satellite atmospheric detection technology has led to the launch of numerous atmospheric O₃ monitoring satellite payloads that provide global, all-weather, full-coverage atmospheric O₃ column concentration data [11,12]. These advantages help address the limitations of ground-monitoring stations in air pollution research. However, there are few satellite O₃ products with high spatial resolution and long time series, and the existing satellite remote sensing data often contain many missing values. In addition, O₃ differs from other pollutants in that it mainly exists in the stratosphere, with only small amounts present in the troposphere or at the surface. Therefore, the total O₃ column concentration data and tropospheric O₃ concentration data provided by satellite inversion cannot accurately reflect the spatial and temporal distributions of surface O₃.

In recent years, there have been different main methods for estimating surface O₃ worldwide, one of which is the chemical transport model. Currently used mature models include CMAQ [13,14], WRF-Chem [15,16], and GEOS-Chem [17,18]. Many researchers have used the above models to simulate and predict the sources, causes, transport processes, and spatiotemporal distribution of different pollutants, such as particulate matter and O₃ in various regions. Although widely used, they have limitations, such as high professional knowledge requirements, complex applications, and high equipment requirements. In addition, chemical transport models rely mainly on emission inventory data, which are highly uncertain and difficult to obtain [19,20]. Another approach is the statistical model, which attempts to obtain a continuous surface O₃ concentration distribution with a higher spatiotemporal resolution based on O₃ pollution and its potential factors using different regression methods, such as the generalized additive model [21], land use regression model [22,23], geographically weighted regression [24–26], and multivariate linear regression [27]. Empirical statistical models are relatively simple and easy to implement but are highly sensitive to extreme values. However, this method has certain disadvantages when simulating O₃ concentrations with nonlinear variation characteristics. It is easily affected by collinear variables, and it is difficult to accurately simulate the O₃ concentrations. In addition, as the sample size increased, the computational efficiency of these models decreased, and the spatial resolution and accuracy of the model results made it difficult to meet the application requirements. Therefore, accurately estimating and predicting O₃ remains a challenging task. In the research progress in recent years, with the emergence of machine-learning algorithms, these problems have gradually been solved. Machine learning uses different methods to mine valuable information from big data and accurately understand the complex relationships among the interactions of various variables. This

effect is superior to that of most traditional models. Many researchers have also applied machine learning to air pollutant estimation, including random forest [27–29], extreme gradient boosting trees [30,31], a self-adaptive geospatially local scheme based on categorical boosting [32], space–time extremely randomized trees [5], light gradient boosting machine models [3,18,33], and deep learning models [34,35].

A light gradient boosting machine (LightGBM) is a distributed gradient boosting framework that builds on a decision tree algorithm. It utilizes weak classifiers such as decision trees to construct an ensemble and achieve an optimal model [36]. The model has the advantages of fast training speed, small memory usage, and high accuracy. It performs well when processing high-dimensional data. At present, the model has been applied in the field of meteorology, and PM_{2.5} prediction research based on the LightGBM model has been established, proving the feasibility and efficiency of the model in predicting air pollutant concentrations [37].

High O₃ pollution occurs from April to September, which is the growing season of China's major grain crops. O₃ pollution poses a serious threat to agricultural production and food security [38,39]. Currently, because of the absence of remote sensing data for O₃-exposure phytotoxicity indices specifically linked to crop yield losses, researchers often use MDA8 (maximum daily eight-hour average), which is usually used to assess human health risks, alongside O₃ data as a surrogate to assess the impact of O₃ on crop yield reductions [18,40]. For example, Wang et al. [41] evaluated O₃-induced yield losses of winter wheat in the North China Plain using the WRF-Chem model with AOT40 and POD₁₂ metrics, reporting an average annual relative yield loss of 10.9% based on AOT40 and 14.6% based on POD₁₂ over 2015–2018. While their study provided valuable regional-scale insights, it was limited to a single region (the North China Plain) and a relatively short study period (2015–2018). Moreover, the WRF-Chem model used in their study relies on emission inventories with inherent uncertainties and requires extensive computational resources. These limitations highlight the need for a more comprehensive, national-scale assessment with extended temporal coverage and improved modeling accuracy. To address these challenges, this study aims to provide a robust scientific basis for assessing the impact of O₃ pollution on winter wheat production. Specifically, we developed a spatiotemporal machine learning framework to reconstruct high-resolution surface O₃ exposure metrics (M7 and AOT40) across China from 2015 to 2022, and then quantitatively evaluated the yield losses and economic costs of winter wheat attributable to O₃ exposure using established crop-specific response functions. The model's inputs mainly included meteorological data, Ozone Monitoring Instrument (OMI)/Aura tropospheric NO₂, OMI/Aura total O₃ column concentrations, and other auxiliary data. Owing to cloud cover and abnormalities, the OMI data contain numerous missing values. We employed the Singular Spectrum Analysis (SSA) method to interpolate missing values. Furthermore, to ensure the robustness of our estimates, we utilized the SHapley Additive exPlanation (SHAP) method to dissect the individual contributions of various features to ozone retrieval. After comprehensive verification, we obtained a high-quality surface M7 O₃ dataset with 10 km resolution coverage across China from 2015 to 2022. In addition, we reconstructed the AOT40 index using the same method to facilitate comparison in crop yield loss assessments. This dataset can scientifically and effectively predict and analyze surface O₃, provide basic data for ecological environment and agricultural production research, it is of great significance for the protection of air pollution and governance of ecosystems.

2. Materials

2.1. Surface O₃ Observation Data

The surface O₃ hourly concentration observation station data were obtained from the China National Environmental Quality Monitoring Center. The data include hourly surface O₃ concentrations ($\mu\text{g m}^{-3}$) collected from 1483 monitoring stations in 2015 and 1735 stations in 2022. We selected M7 (average O₃ concentration for seven hours from 9:00 to 16:00) as the concentration index to represent the surface O₃ concentration during the day. Each station must have at least 75% of the required data samples per day, meaning that O₃ concentration measurements must be taken at least six times daily between 9:00 and 16:00. Figure 1 shows the spatial distribution of the O₃ observation stations in mainland China during the study period. Air pollution monitoring stations are mainly distributed in the eastern coastal areas and more developed cities, and there are relatively few stations in the central and western regions.

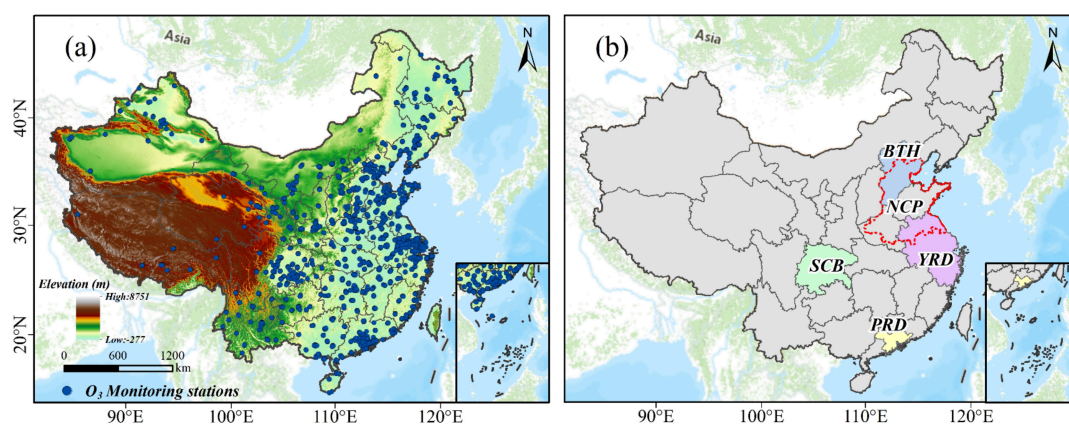


Figure 1. (a) Study area and spatial distribution of the ground O₃ monitoring stations, where the background is the surface elevation (m). The solid black lines represent China's regional and provincial boundaries. (b) Five regions of interest in this study: the North China Plain (NCP), the Beijing–Tianjin–Hebei (BTH) region, the Yangtze River Delta (YRD), the Sichuan Basin (SCB), and the Pearl River Delta (PRD). For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

2.2. Meteorological Data

Meteorological conditions are important factors affecting the generation, accumulation, and diffusion of O₃. Adding meteorological elements to the O₃ prediction model can improve prediction accuracy. The meteorological data used in this study were obtained from ERA5 and ERA5-Land Reanalysis datasets with a temporal resolution of 1 h. We selected various meteorological parameters, including the ground 2 m temperature (TEM), 10 m eastward wind component (WU), 10 m northward wind component (WV), boundary layer height (BLH), total precipitation (PRE), evaporation (ET), relative humidity (RH), downward solar radiation (DSR), and surface pressure (SP). These factors affect the formation of O₃ and are widely used to estimate O₃ concentrations. For example, TEM and DSR significantly affect the photochemical reaction of O₃ generation [42,43]. RH, SP, and ET indirectly affect O₃ generation by affecting the photolysis rate via atmospheric water vapor content [44–47]. Changes in the BLH have a complex nonlinear effect on O₃ generation, which affects the distribution and concentration of O₃ [48,49]. The WU and WV affect the propagation direction and speed of the precursors and O₃ [50,51]. The PRE itself has a flushing effect on O₃ and its precursors, thereby reducing the concentrations of O₃ and other pollutants in the atmosphere [52,53].

2.3. OMI/Aura Data Products

OMI is a near-ultraviolet/visible CCD spectrometer on NASA's Earth Observing System (EOS) Aura satellite. The spectral measurement region is 264–504 nm, with a spectral resolution of 0.42 nm–0.63 nm, which can achieve daily global coverage [54]. The Aura satellite operates in a sun-synchronous orbit at an altitude of 705 km, repeating accurately for 16 days, with the ascending node at $13:45 \pm 0:15$ local time. The OMI NO₂ column concentration data used in this study are the third version of the OMI L3 (OMNO₂d) daily average product released by NASA. These data were obtained by quality control screening based on the secondary NO₂ product, removing the “row anomaly” pixels. Only pixels with a cloud fraction of less than 30% and surface reflectivity less than 30% were selected [55]. The spatial resolution of the NO₂ column concentration data product was $0.25^\circ \times 0.25^\circ$, and the total uncertainty did not exceed 20%. Compared with the previous version, the calculations of the air mass factor (AMF) and vertical column concentration (VCD) were improved. The improved DOAS algorithm was used, which optimizes the visible light spectrum channel of the OMI and performs spectral fitting of NO₂ in the range of 402–465 nm, greatly reducing the problem of overestimation of the NO₂ oblique column concentration, making the result closer to the NO₂ oblique column concentration obtained by independent observation [56]. We used the L3_V003 OMDOAO₃e product generated using the OMI DOAS algorithm from 2015 to 2022 [57]. The OMI total O₃ column concentration product was extracted from OMI L1b level data using differential absorption spectroscopy (DOAS). The proposed algorithm consists of three steps. First, the reference differential absorption spectrum of O₃ was fitted to the measured Earth's radiation and solar irradiance spectra using the DOAS method to obtain the oblique column density. Second, the oblique column density was converted to vertical column density using the AMF. Finally, the cloud effects were corrected. Among them, the best-fitting window was selected as 331.1 nm–336.1 nm [58,59] to reduce the sensitivity to atmospheric temperature.

2.4. Auxiliary Data

Studies have shown that factors such as land cover (LUC), population (Pop), digital elevation model (DEM), and normalized difference vegetation index (NDVI) can influence the generation of O₃ [60,61]. The land cover data used were the MODIS global 0.05° product MCD12C1, and the population data came from the 1 km resolution population data provided by WorldPop. Terrain data were obtained from the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) with a spatial resolution of 30 m. NDVI data were obtained from the MODIS MOD13A3 product with a spatial resolution of 1 km and a temporal resolution of 1 month. It should be noted that the current land use data were only updated to 2021. Because the interannual changes in land use were small, we used the 2021 land use data as a substitute for the 2022 data in the model. Population data were updated to 2020, and the 2020 data were used as substitutes for the population data for 2021 and 2022. Table S1 summarizes the relevant information for all the data used in this study, and Figure S1 shows the histograms and statistics of the M7 and other explanatory variables used to build the model, covering the 2015–2022 time range.

2.5. Winter Wheat Statistics

Studies have shown that winter wheat is highly sensitive to O₃, causing yield losses of up to 33% [10]. Taking winter wheat as an example, we used the M7 O₃ dataset estimated above to assess winter wheat yield and economic losses in recent years. Yield data for the provinces in China's winter wheat-growing areas were obtained from the China Agricultural Yearbook 2016–2022 (<http://www.stats.gov.cn/>, accessed on 15 February 2025). Each year's statistical yearbook covers data from the previous year, and the China Agricultural

Yearbook 2023 has not yet been published; therefore, we only obtained winter wheat yield data from 2015 to 2021. The minimum purchase price of winter wheat was based on public information available on the website of the National Development and Reform Commission of the People's Republic of China (<https://www.ndrc.gov.cn/>, accessed on 15 March 2025). For ease of comparison and discussion, the purchase price was converted into U.S. dollars, and the historical annual average exchange rate was used for the exchange rate calculations (<https://chl.cn/>, accessed on 15 May 2025).

2.6. Data Preprocessing

Prior to ST-LightGBM model training, all predictor variables were unified to a consistent 10 km spatial resolution across mainland China. For aligning multi-source datasets, continuous meteorological fields (e.g., 2 m temperature, wind speed, boundary layer height) were resampled using bilinear interpolation to preserve smooth spatial gradients. Categorical land surface datasets, such as the MODIS land cover product (MCD12C1), were resampled using the nearest neighbor method to maintain the integrity of discrete classification schemes. Gridded demographic datasets, including WorldPop population density, were aggregated via area-weighted spatial averaging to ensure that total population counts remained consistent within each 10 km pixel. All layers were subsequently reprojected to the WGS84 coordinate system and temporally matched to the 2015–2022 study period. After preliminary processing and integration of all data, the original dataset still had problems, such as missing data and outliers; therefore, it was necessary to filter and screen the original dataset. Before creating the model, the correlation between the M7 O₃ values obtained from the ground monitoring stations and each variable was explored (Table S2). Additionally, we employed the variance inflation factor (VIF) method to assess collinearity among independent variables, ensuring the avoidance of redundant information. After the variables were screened for collinearity, the VIF values of all the independent variables did not exceed 10, indicating that there was no strong collinearity problem for any independent variable (Table S3). Therefore, 15 independent variables, including RH, ET, DSR, BLH, TEM, PRE, WU, WV, SP, NDVI, DEM, Pop, LUC, OMI NO₂, and OMI O₃, were included in the training model.

3. Methods

3.1. Imputation for OMI NO₂/O₃ Columns

There are many missing values in the OMI data products owing to cloud cover and row anomalies. We found that the missing rate of OMI NO₂/O₃ products from 2015 to 2022 reached more than 50%, which resulted in a large amount of useful information missing from OMI data and unreliable output results during model training. Addressing these missing data issues directly influences the reliability and authenticity of research findings. To address the serious problem of missing OMI NO₂/O₃ products, we adopted a strategy of interpolating missing data values based on the singular spectrum analysis (SSA) method of time series by Yi and Sneeuw [62]. The basic principle of the SSA method is to extract effective information from existing data series, reconstruct the time series according to the time correlation, establish a suitable data interpolation model, and fill in the data gaps in the time series using features obtained from the original data [63,64].

SSA fills the data gaps in the time series based on the characteristics of the available data samples, which are inferred by integrating the coefficients obtained from the available observational data. The embedding window length and the number of principal components, two key parameters for the SSA algorithm, must be determined to reconstruct the time series, so it is necessary to choose the optimal window length and number of components. The steps of SSA data interpolation are as follows.

- (1) Time series decomposition. Perform SSA decomposition on the time series of OMI data products and select an appropriate window length L based on the characteristics of the time series.
- (2) Construct the trajectory matrix. Using the selected window length L , the one-dimensional time series $X = [x_1, x_2, \dots, x_N]$ at a certain grid point is converted into its trajectory matrix Y .

$$Y = \begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ x_2 & x_3 & \cdots & x_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{bmatrix} \quad (1)$$

Here, the number of columns $M = N + 1 - L$, and N is the length of the time series. In order to separate the periodic part, N is generally taken as the least common multiple of the sequence period. The window length L is generally not more than half of the time series length ($L \leq N/2$).

- (3) Singular value decomposition (SVD). A lag-covariance matrix YY' is constructed using the trajectory matrix Y . Let $\lambda_1, \lambda_2, \dots, \lambda_m$ denote the eigenvalues of YY' , and let the corresponding eigenvectors be $U_1, U_2, \dots, U_m$. Then, define $k = \text{rank}(Y)$ as follows:

$$Y = Y_1 + Y_2 + \cdots + Y_k \quad (2)$$

- (4) Diagonal averaging. The purpose of this step is to reconstruct the elementary matrix, Y_k , obtained by decomposition, into a new sequence with a time series of length N . Let Z denote the reconstructed time series as follows:

$$Z_i = \begin{cases} \frac{1}{i} \sum_{m=1}^i z_{m,i-m+1}^* & 1 \leq i < L^* \\ \frac{1}{L^*} \sum_{m=1}^{L^*} z_{m,i-m+1}^* & L^* \leq i < M^* \\ \frac{1}{N-i+1} \sum_{m=i-M^*+1}^{N-M^*+1} z_{m,i-m+1}^* & M^* \leq i < N \end{cases} \quad (3)$$

Here, Z is the sequence of pixels with missing values interpolated by the SSA method. SSA interpolation is performed on the pixels with missing data in the study area to obtain the interpolation results of the OMI data products in the study area.

$$OMI_{ij}^* = SSA(OMI_{ij}) \quad (4)$$

3.2. ST-LightGBM Model

3.2.1. LightGBM Model

LightGBM, an algorithm framework proposed by Microsoft in 2017, is an enhancement of GBDT. It maintains accuracy without increasing the model complexity [36]. LightGBM is optimized from three perspectives: the number of features, the number of feature splitting points, and the number of samples. In terms of the number of features, the Exclusive Feature Bundling (EFB) algorithm was used to bundle two mutually exclusive features together to form a new feature, thereby reducing the number of features and improving the training speed while ensuring the original feature information; in terms of the number of feature splitting points, the histogram-based algorithm was used to reduce the number of feature splitting points and more efficiently find the best feature splitting point; in terms of the number of samples, based on the Gradient-based One-Side Sampling (GOSS) algorithm, from the perspective of reducing samples, most of the samples with smaller weights in

the training process were eliminated, and only the information gain was calculated for the remaining sample data. The purpose was to retain samples that help calculate the information gain with maximum efficiency and improve the model training speed.

In addition to the Histogram, GOSS, and EFB algorithms, LightGBM also proposes a layer-wise leaf node growth method that differs from XGBoost, namely, a leaf-wise decision tree generation algorithm with depth restrictions. It is more accurate and efficient, and it can save unnecessary computational overhead. Simultaneously, a depth restriction mechanism was added to prevent excessive growth of a certain node, which can prevent overfitting to a certain extent while ensuring accuracy.

To ensure model stability and reproducibility, we performed a grid search to determine the optimal hyperparameters for the ST-LightGBM model. The hyperparameter search space included the learning rate, maximum depth, and number of leaves. The learning rate was tested in the range of 0.10 to 0.13 with a step size of 0.01; the maximum depth was varied between 10 and 12 with a step size of 1; and the number of leaves was determined based on the maximum depth, with an adaptive step size. Model performance was evaluated using the R^2 metric, and the combination yielding the best results was selected. The final optimized hyperparameters were as follows: learning rate = 0.13; maximum depth = 12; number of leaves = 2940. Other parameters were kept at their LightGBM defaults. Model performance was subsequently evaluated using a rigorous 10-fold cross-validation framework.

3.2.2. Model Extension

O_3 concentrations exhibit significant temporal and spatial heterogeneity, indicating that concentrations in a specific area can be influenced by nearby observations and the previous day's levels, with these influences varying over time and space. To improve the accuracy of the O_3 prediction, we introduced spatiotemporal feature information into the LightGBM framework, thereby developing the ST-LightGBM model for surface O_3 estimation. For a given point (pixel), the spatial feature information can be represented by the geographical distance between its longitude and latitude and the four points on the boundary of the study area and the center point (73.2° E–135.3° E, 18.1° N–53.8° N) [65,66], which is calculated using the Haversine method (Equations (5)–(7)) [67]. Temporal features were represented by the day of the year (DOY) to distinguish the different days of the year for each data record during the model training process (Equation (8)).

$$\text{haversine}(\alpha) = \sin^2\left(\frac{\alpha}{2}\right) = \frac{1 - \cos \alpha}{2} \quad (5)$$

$$\text{Space}_w = \text{haversine}(\alpha_2 - \alpha_1) + \cos(\alpha_1)\cos(\alpha_2)\text{haversine}(\alpha_2 - \alpha_1) \quad (6)$$

$$\text{Space} = 2 \times r \times \arcsin\sqrt{\text{Space}_w} \quad (7)$$

$$\text{Time} = \cos\left(2\pi \frac{\text{DOY}}{\text{Year}}\right) \quad (8)$$

Here, Space_w represents the distance from a pixel to the boundary and center of the grid; α and β represent the longitude and latitude of a point, respectively; Space and Time represent the spatial and temporal characteristics; r represents the radius of the Earth (≈ 6371.393 km); DOY (day of the year) represents the day of the year; Year represents the total number of days in the year.

To avoid information redundancy, we first calculated the normalized importance (%) of each feature for O_3 estimation during model training (Figure S2). Variables with higher scores contribute more to the overall performance of the model, whereas variables with lower scores have less impact on the model and may generate redundant information.

Variables with low importance scores were screened to improve the model efficiency and avoid overfitting owing to redundant input variables. All meteorological factors had a significant impact on the estimation of O_3 , particularly TEM and DSR, which are relatively important features, accounting for approximately 17% and 11.3%. Overall, all features after data preprocessing and screening had a non-negligible impact on O_3 estimation (importance > 1%). Thus, we included all 15 selected variables in the ST-LightGBM for the final model construction.

3.2.3. Model Validation

This study adopted the widely used 10-fold cross-validation (10-CV) method to evaluate the overall performance of the model in predicting surface O_3 concentrations. We used a sample-based 10-CV method to assess the prediction accuracy. The dataset was randomly divided into 10 parts: nine parts were used for training and one part for validation, which was repeated 10 times to ensure that all samples were validated. In addition, a site-based 10-CV method was used to test the spatial prediction ability of the model in areas without ground monitoring points. This method involves dividing the data based on ground monitoring stations. Initially, all sample data were grouped according to their respective stations. Subsequently, nine groups of sample data were randomly selected as the training set, and the remaining group was used as the validation set. The training and validation samples represented sample data obtained at different locations, which enabled us to evaluate the prediction accuracy of the model in areas where ground monitoring data could not be obtained. This study used a linear regression equation (e.g., slope and intercept), coefficient of determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE), and mean relative error (MRE) as statistical indicators to evaluate model performance.

Although LightGBM has a higher prediction accuracy, faster calculation speed, and stronger prediction ability, its low interpretability seriously limits its application, and it is generally considered a black-box model. The feature importance value of the LightGBM model can only indicate which environmental variables are important but cannot measure the relationship between each feature and the prediction result. To solve the problem of low model interpretability, SHAP explains the model results [68–70]. Because SHAP attribution analysis uses the Shapley value in game theory as an explanation metric, it has the advantages of strong global and local interpretability of variables, fair distribution of variable contributions, and excellent visualization effects, which compensate for the poor interpretability of black box models [71,72]. Therefore, the SHAP framework is commonly used to explain the results of machine learning models. In this study, we used SHAP to explain the influence of environmental variables on O_3 prediction based on ST-LightGBM and the model prediction results. Figure 2 shows a flowchart of the M7 O_3 data processing in our study, which mainly includes data collection and preprocessing, spatiotemporal matching, and surface M7 O_3 dataset estimation based on the ST-LightGBM model.

3.3. Calculation Methods of Yield and Economic Reductions

The relative yield loss (RYL) refers to the percentage reduction in crop yield in an O_3 -contaminated environment compared with that in an O_3 -free environment. The M7 index, established by the United States Environmental Protection Agency (US EPA), quantifies the relationship between the yield of major crops and O_3 concentration. This index is widely used to assess the O_3 exposure in winter wheat [73–75]. The AOT40 ozone metric is the standard for the protection of vegetation in Europe and is also widely used in the United States and Asia as a reliable indicator for assessing crop yield losses [39,76,77]. This metric is particularly sensitive to the impacts of extremely high O_3 concentrations

and is considered to provide more accurate evaluations under conditions of elevated O₃ levels [78,79].

$$AOT40 = \sum_{i=1}^n ([O_3]_i - 0.04), \quad \text{for } [O_3]_i \geq 0.04 \text{ ppm} \quad (9)$$

Here, [O₃]_i is the hourly O₃ concentration, and n is the total number of hours within the province-specific O₃ exposure window. Following the internationally recognized crop-calendar definition [78–80], this window is defined as the three months immediately preceding the local harvest date for each province. This period corresponds to the grain-filling stage, during which winter wheat is most sensitive to O₃-induced yield reduction [10,81]. The specific exposure months for each province are detailed in Table S10.

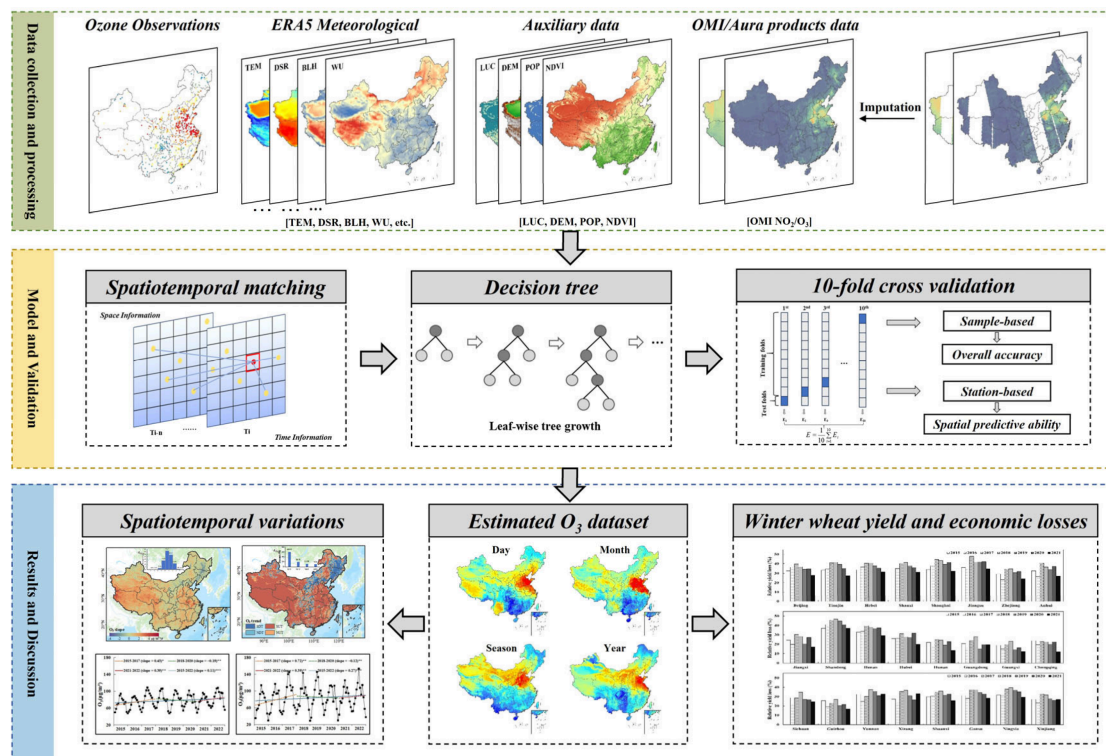


Figure 2. Flowchart of the mapping tree process of the M7 O₃ dataset in our study.

The winter wheat assessment response function based on the M7 and AOT40 ozone metrics can be used to obtain relative yield (RY) and then calculate relative yield loss (RYL) [10,39,73,74].

$$RY = \exp\left(\left(-\frac{M7}{137}\right)^{234}\right) / \exp\left(\left(-\frac{25}{137}\right)^{234}\right) \quad (10)$$

$$RY = \{1 - 0.02 \times [AOT40 + 3.24 / (1 + 0.12 \times AOT40)]\} / 0.95 \quad (11)$$

$$RYL = 1 - RY \quad (12)$$

Based on the method adopted by Sinha et al. [82], the crop yield and economic loss were calculated using Equations (11) and (12), respectively.

$$CPL = RYL \times CP / (1 - RYL) \quad (13)$$

$$ECL = CPL \times CPP \quad (14)$$

Here, CPL (crop production loss) is the loss of winter wheat yield, CP (crop production) is the actual yield of winter wheat, CPP (crop purchase price) is the minimum purchase price

of winter wheat released annually by the Chinese government (<https://www.ndrc.gov.cn/>, accessed on 15 March 2025), and ECL (economic cost loss) is economic loss, which is the loss of property due to O₃ exposure. For the convenience of comparison and discussion, the purchase price has been converted into US dollars, and the exchange rate calculation uses the historical annual average exchange rate.

4. Results

4.1. Imputation Results for OMI NO₂/O₃ Columns

We employed the Singular Spectrum Analysis (SSA) method to interpolate the missing OMI NO₂/O₃ data and validated its accuracy using a self-supervised learning approach. The results showed that the interpolated data were highly consistent with the satellite observation data. The average R² and RMSE of the interpolated OMI NO₂ data are 0.83 and 4.21×10^{14} molecules cm⁻², and the average R² and RMSE of the interpolated OMI O₃ data are 0.73 and 28.58 DU (Figures S3 and S4). Figures S5 and S6 show four examples of the original missing values and interpolated daily OMI NO₂/O₃, namely 14 April 2016, 26 July 2018, 12 October 2020, and 7 January 2022. These examples demonstrate that the interpolated OMI NO₂/O₃ covered the entire study area with minimal deviation compared to the original data, indicating stable data changes and effective interpolation.

4.2. Impact of Predictors on O₃ Predictions

Figure S2 shows the importance ranking of each characteristic variable generated by the ST-LightGBM model, indicating that TEM and DSR are the most important predictors of surface O₃ inversion. Figure 3 shows the SHAP summary and the relative contributions of the feature variables to the model output. Similar to the important features generated by the model itself, DSR and TEM made the most significant contributions (Figure 3b), with a SHAP value of approximately 46.62%, followed by RH and DEM, accounting for 10.81% and 6.08%, respectively. The SHAP method also provides an assessment of the direction in which the characteristic variable affects the O₃ inversion (Figure 3a). The color of each point represents the value of the characteristic variable. The larger the characteristic variable value, the larger the SHAP value, indicating that the positive correlation between O₃ and the characteristic variable is more obvious; the larger the characteristic variable value, the smaller the SHAP value, indicating that the negative correlation between O₃ and the characteristic variable is more obvious. For example, DSR and TEM made the greatest contributions to O₃ prediction, and their impact directions were similar. The contributions of DSR and TEM to the surface O₃ inversion increased with increasing intensity and were positively correlated. This is mainly because enhanced radiation and increased temperature promote the production of oxygen atoms and oxidants, thereby enhancing the photochemical reaction process. RH is closely related to the dry deposition of O₃ and has a scavenging effect on O₃. This further demonstrates that meteorological conditions with high temperatures, strong radiation, and low humidity are more likely to cause O₃ pollution. The SHAP method provides a basis for the selection of characteristic variables and helps us gain a deeper understanding of the physical interpretability of the ST-LightGBM.

4.3. Accuracy Assessment

4.3.1. Overall Model Performance

First, we evaluated the overall performance of the model by applying the sample-based 10-CV method to all sample data from 2015 to 2022 (N = 3,795,495) and to annual sample data (Figure 4). The M7 O₃ estimates of the total sample data across China are highly consistent with the measurements at ground monitoring stations (CV-

$R^2 = 0.91$), with a slope and y-intercept of 0.9 and $7.99 \mu\text{g m}^{-3}$, respectively, and an average RMSE, MAE, and MRE of $11.48 \mu\text{g m}^{-3}$, $8.06 \mu\text{g m}^{-3}$, and 14.63%, respectively (Figure 4a). The model also performed very well in terms of annual CV results (Figure 4b–i). The overall accuracy of M7 O_3 estimates since 2017 ($\text{CV-R}^2 = 0.93\text{--}0.95$, $\text{RMSE} = 8.19\text{--}10.92 \mu\text{g m}^{-3}$, $\text{MAE} = 5.74\text{--}7.74 \mu\text{g m}^{-3}$, $\text{MRE} = 8.94\text{--}13.09\%$) is better than that of 2015–2016 ($\text{CV-R}^2 = 0.87\text{--}0.89$, $\text{RMSE} = 13.08\text{--}14.59 \mu\text{g m}^{-3}$, $\text{MAE} = 9.1\text{--}9.99 \mu\text{g m}^{-3}$, $\text{MRE} = 18.67\text{--}22.44\%$). Statistical indicators revealed a gradual improvement in the model fitting accuracy from 2015 to 2022, with R^2 increasing annually and RMSE, MAE, and MRE decreasing. This improvement can be attributed to the increasing number of monitoring sites and sample sizes over the years, enriching the model training data and enhancing prediction accuracy. Overall, our model performed robustly across the study period and individual years.

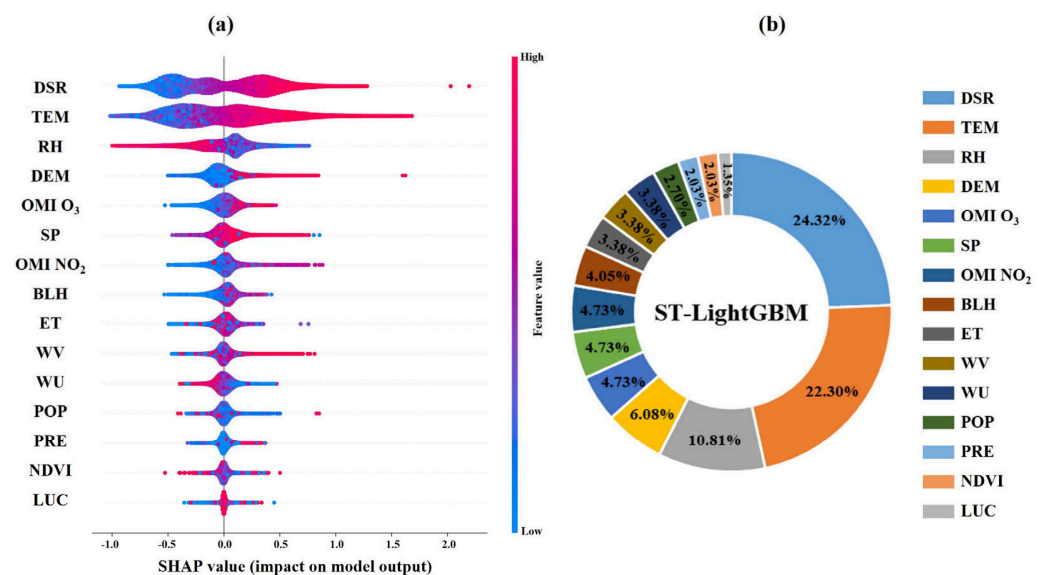


Figure 3. SHAP waterfall diagrams of the ST–LightGBM model (a). The color represents the value of the feature from low to high. The abscissa denotes each feature’s contribution to the output charts of mean absolute SHAP values for the ST-LightGBM model; it represents the relative contributions of the features to the model output (b).

Cross-validation results for the model inverting the AOT40 ozone metric are shown in Table S4. The estimated values also showed high consistency with ground monitoring data ($\text{CV-R}^2 = 0.84$), with average RMSE and MAE of 71.36 ppb and 47.18 ppb, respectively. Overall accuracy results also showed a similar trend to M7, with performance improving after 2017 ($\text{CV-R}^2 = 0.87\text{--}0.92$) compared to 2015–2016 ($\text{CV-R}^2 = 0.72\text{--}0.78$). With the increase in monitoring data and sample size, fitting accuracy improved annually, while uncertainty gradually decreased. Overall, the model’s predictive performance for the AOT40 ozone metric remains robust and can well support subsequent crop yield assessment research. Second, we tested the performance of the model in typical regions of China (Table 1). Our model performed well in different regions, with $\text{CV-R}^2 = 0.91\text{--}0.95$, $\text{RMSE} = 10.89\text{--}11.86 \mu\text{g m}^{-3}$, $\text{MAE} = 7.3\text{--}8.24 \mu\text{g m}^{-3}$, and $\text{MRE} = 12.55\text{--}19.10\%$. Overall, the uncertainties and differences in the model were small, and the model was relatively stable. The model performed best in the BTH region, with a CV-R^2 of 0.95. The model performance in the NCP, YRD, PRD, and SCB regions was slightly worse than that in the BTH region. This variation can be attributed to differences in the monitoring site density and the variability of independent variables across regions, influencing the robustness of the model. In typical areas (Table S5), the model also showed good fitting accuracy in retrieving the AOT40 ozone metric ($\text{CV-R}^2 = 0.83\text{--}0.91$).

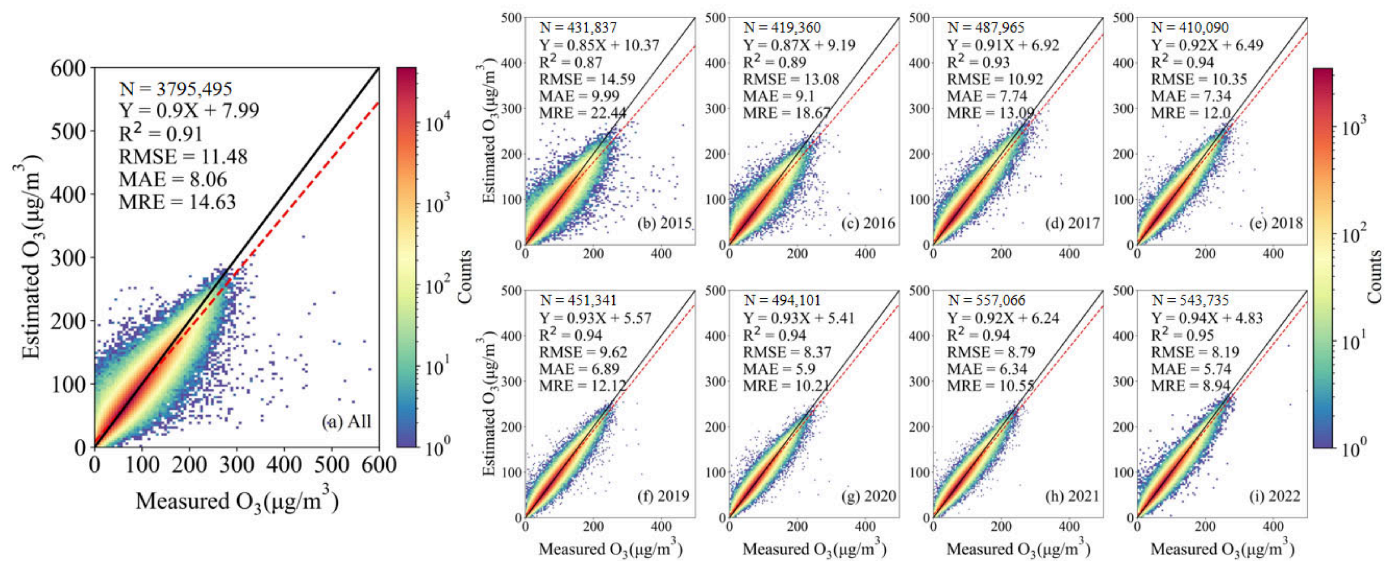


Figure 4. Sample-based cross-validation results of the M7 O₃ estimates from 2015 to 2022. (a) Overall accuracy. (b–i) Annual accuracy. The totals in the legend on the right represent the total number of data samples in each cell. The solid line represents the 1:1 line, and the dashed line represents the best-fit linear regression line.

Table 1. Statistics of cross-validation results of the estimated M7 O₃ values (μg m^{−3}) in China, North China Plain, Beijing–Tianjin–Hebei, Yangtze River Delta, Pearl River Delta, and Sichuan Basin from 2015 to 2022.

Region	Sample Size	Overall Accuracy				Spatial Predictive Ability			
	N	R ²	RMSE	MAE	MRE	R ²	RMSE	MAE	MRE
China	3,795,495	0.91	11.48	8.06	14.63	0.82	16.59	11.98	22.61
NCP	685,503	0.93	11.7	7.95	12.55	0.87	16.68	11.72	19.09
BTH	213,829	0.95	10.97	7.3	12.74	0.9	15.06	10.22	18.2
YRD	580,145	0.92	11.86	8.24	13.82	0.81	18.14	12.97	23.0
PRD	142,909	0.92	11.17	7.9	13.8	0.85	15.42	11.27	20.04
SCB	260,224	0.91	10.89	7.74	19.1	0.81	15.95	11.72	31.64

At the scale of a single station (Figure 5), owing to differences in the number of observation records and useful data samples from 2015 to 2022, the sample sizes of the different stations were different. A few stations had fewer than 10 samples, which did not meet the conditions of the 10-CV method, and were deleted. Except for a few stations established in the later stages of the study, most stations had a sufficient number of data samples, and more than 78% of the stations had at least five years of data (i.e., N > 1825).

Regarding model accuracy, approximately 81.64% of the stations had CV-R² values exceeding 0.78 (Figure 5a), with higher fitting accuracy observed for stations in the central region. The model performance exhibited significant spatial heterogeneity across different regions. The best performance was observed in central and eastern China, whereas several stations in western China exhibited poorer performance. This discrepancy may be attributed to the sparse distribution of stations in the western region, which affects the ability of the model to capture the relationship between O₃ and the independent variables. In addition, poor natural conditions in these regions, particularly in southern Xinjiang, lead to large uncertainties in air pollution estimates [5]. Overall, the model performed well at the station scale, with average CV-R², RMSE, MAE, and MRE values of 0.83, 17.32 μg m^{−3}, 11.64 μg m^{−3}, and 23.55%, respectively.

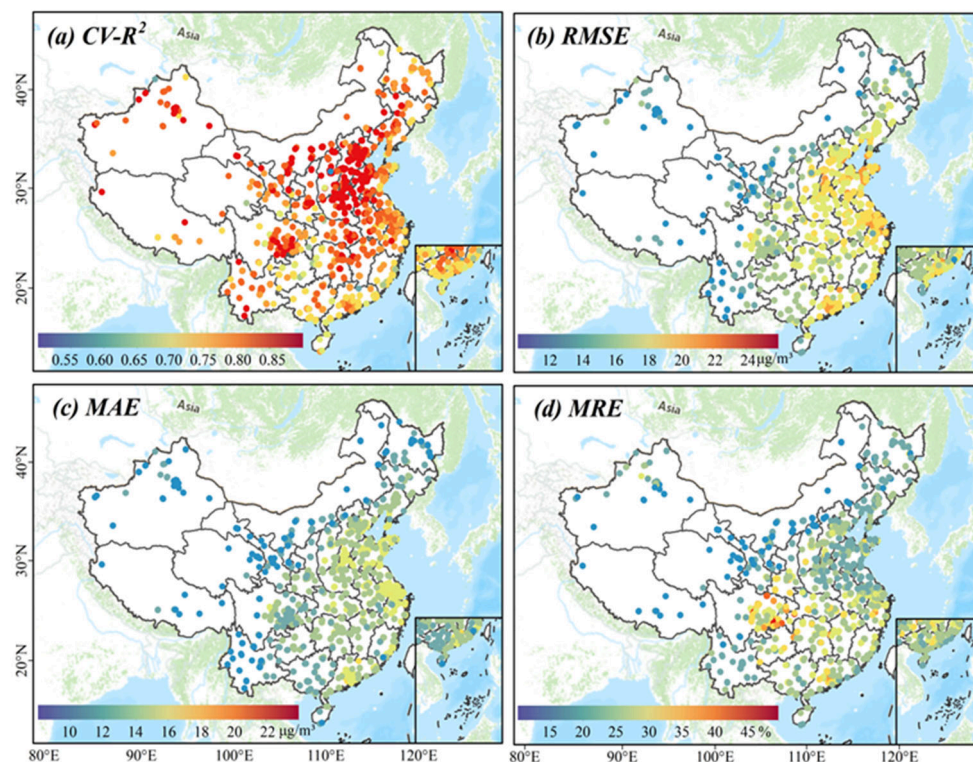


Figure 5. Individual-site-scale sample-based cross-validation results of the M7 O₃ estimates ($\mu\text{g m}^{-3}$) from 2013 to 2020 in China.

4.3.2. Spatial Predictive Capability

To further verify the robustness and spatial generalizability of our model, we employed the site-based 10-CV method to evaluate the predictive ability of the model across different spatial scales. Across the entire region (China), the model-predicted O₃ estimates were well correlated with ground site observations ($\text{CV-R}^2 = 0.82$), with a slope and y-intercept of 0.81 and $14.22 \mu\text{g m}^{-3}$, respectively, and an average RMSE, MAE, and MRE of $16.59 \mu\text{g m}^{-3}$, $11.98 \mu\text{g m}^{-3}$, and 22.61%, respectively (Figure S7a). The prediction accuracy of the model was slightly lower than that of the sample-based test results (Figure 4a), indicating good spatial prediction capability. Similarly, as the sample size increased over time, the spatial prediction ability of the model improved (Figure S7).

Regional differences exist in the model's estimation of surface O₃. The BTH region ($\text{CV-R}^2 = 0.9$, $\text{RMSE} = 15.06 \mu\text{g m}^{-3}$), NCP region ($\text{CV-R}^2 = 0.87$, $\text{RMSE} = 16.68 \mu\text{g m}^{-3}$) and PRD region ($\text{CV-R}^2 = 0.85$, $\text{RMSE} = 15.42 \mu\text{g m}^{-3}$) show better prediction accuracy for surface O₃. In contrast, the YRD and SCB regions (both $\text{CV-R}^2 = 0.81$, $\text{RMSE} = 18.14$ and $15.95 \mu\text{g m}^{-3}$, respectively) exhibit slightly lower prediction accuracy (Table 1). Different climatic environments, production, and living conditions in various regions lead to strong uncertainties in the performance of the model. Additionally, areas with numerous monitoring stations, which experience significant variability in O₃ concentration levels and sample sizes, have a relatively high estimation accuracy owing to the richness of the sample information.

4.3.3. Temporal-Scale Validation

Figure 6 presents the time series changes in the evaluation indicators used to assess the overall performance of the model in estimating daily surface M7 O₃ concentration from 2015 to 2022 and to test the model's generalization ability over time. The daily sample size was substantial, ranging from 7486 to 11,635 data points, with an average of 10,654 data points, and remained consistent over time. This is a unique advantage of the dataset used in

this study, that is, the full coverage of the Chinese region. Although the performance varies with the season, the amplitude of the change is moderate throughout the year, with $CV-R^2$ ranging from 0.66 to 0.9 (average = 0.8), and the model accuracy reaches above 0.78 on approximately 75.41% of the days in a year (Figure 6a). The RMSE and MAE values for O_3 concentration estimation fluctuated seasonally, being lower in spring and winter and higher in summer and autumn (Figure 6b,c). Conversely, MRE showed the opposite seasonal trend (Figure 6d). The reason for the larger absolute uncertainty error of O_3 in summer is that a longer daytime cycle, higher temperature, greater solar radiation intensity, and stronger photochemical reactions in summer lead to larger variations in the O_3 concentration values in summer. Using the mean value of each variable in the model may not accurately reflect the high concentration levels of O_3 at noon in summer, and the ground observations may deviate significantly from the model prediction results because of the peak driving of the noon O_3 concentration. The opposite seasonal phenomenon of relative uncertainty was due to the lower O_3 concentration in winter, which resulted in a larger relative uncertainty in winter. Overall, the average RMSE, MAE, and MRE values of our model over time are $13.92 \mu\text{g m}^{-3}$, $9.74 \mu\text{g m}^{-3}$, and 18.5%, respectively.

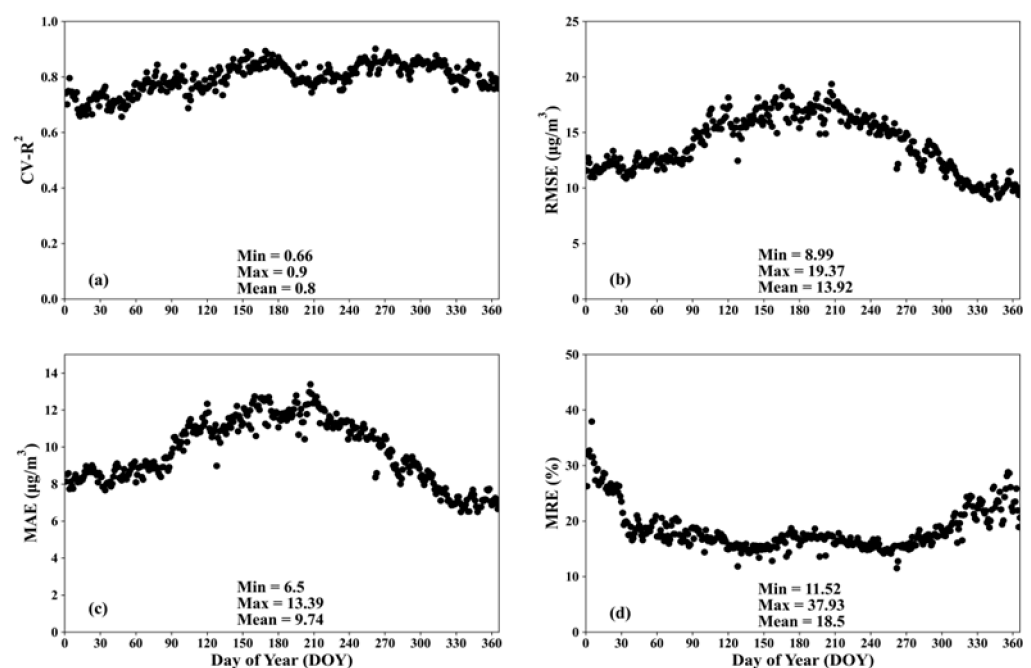


Figure 6. Daily variation time series of verification results of the M7 O_3 estimates from 2015 to 2022 in China. The minimum, maximum, and average values are given in each panel. (a) $CV-R^2$, (b) RMSE, (c) MAE, and (d) MRE. The minimum, maximum, and average values are given in each panel.

4.4. M7 O_3 Dataset Generation

4.4.1. Spatiotemporal Distributions of M7 O_3

We built the ST-LightGBM model by combining ground monitoring station data, meteorological data, and auxiliary data, such as land use. After validating the accuracy and generalization capability of the model, we produced a high-quality daily M7 O_3 dataset for China from 2015 to 2022. Figure 7 shows eight typical examples of the daily spatial distribution of M7 O_3 : 30 May and 6 November 2016; 7 June and 1 November 2018; 8 June and 5 November 2020; and 2 June and 1 November 2022. The O_3 concentrations in adjacent periods had similar spatial distribution characteristics, and the O_3 pollution conditions varied from region to region. The difference in O_3 concentrations between the north and south on different days was mainly due to the difference in sunshine intensity between southern and northern China during different seasons. Overall, the daily O_3

data estimated by the model were highly consistent with the spatial distribution pattern of ground monitoring, indicating that our model performed well and that the estimated full-coverage M7 O₃ dataset was of high quality.

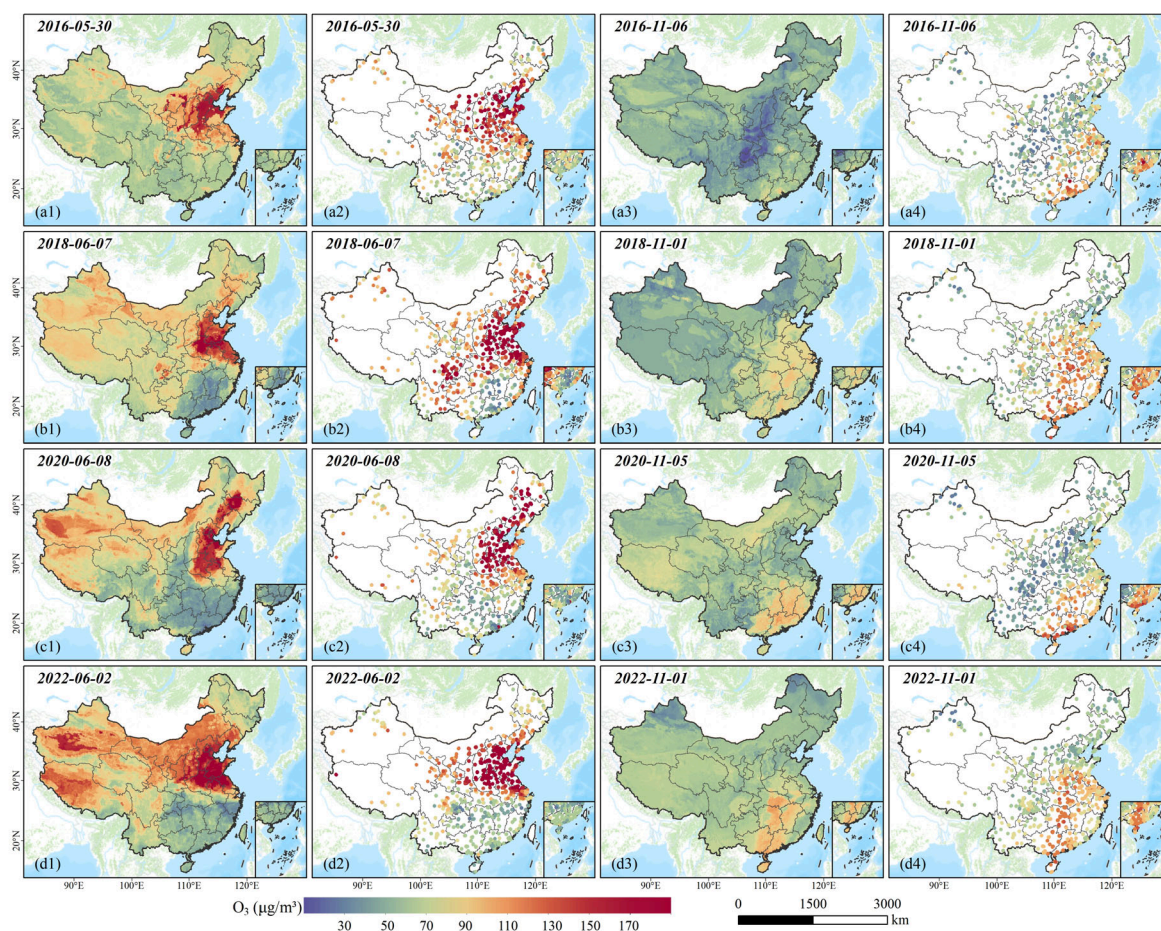


Figure 7. The typical daily M7 O₃ maps in China. (a1–d1,a3–d3) are derived from the ST-LightGBM model. (a2–d2,a4–d4) are ground-based M7 O₃ maps. 30 May 2016 (a1,a2); 6 November 2016 (a3,a4); 7 June 2018 (b1,b2); 1 November 2018 (b3,b4); 8 June 2020 (c1,c2); 5 November 2020 (c3,c4); 2 June 2022 (d1,d2); 1 November 2022 (d3,d4).

Figure 8 shows the spatial distribution of the annual average M7 O₃ concentration in China from 2015 to 2022. The annual average O₃ concentrations were $71.64 \pm 8.35 \mu\text{g m}^{-3}$, $73.75 \pm 8.61 \mu\text{g m}^{-3}$, $79.80 \pm 7.58 \mu\text{g m}^{-3}$, $80.33 \pm 7.88 \mu\text{g m}^{-3}$, $77.41 \pm 7.93 \mu\text{g m}^{-3}$, $75.03 \pm 7.41 \mu\text{g m}^{-3}$, $77.80 \pm 7.78 \mu\text{g m}^{-3}$, and $81.39 \pm 8.13 \mu\text{g m}^{-3}$. From 2015 to 2022, the overall O₃ concentration in China exhibited a phased upward trend, with a pattern similar to that observed in typical regions (Table S4), from the sporadic distribution of high-value areas in 2015 to generally high O₃ concentrations in various regions of the country by 2022. O₃ pollution has the characteristics of urban agglomeration pollution, and the scope of high-pollution areas and the intensity of pollution are increasing each year. In 2018, O₃ pollution was relatively serious, reaching the peak of O₃ concentration levels in recent years, which was mainly due to the rapid decline of PM_{2.5} caused by the implementation of the Clean Air Action Plan policy. The reduction in nitrate in PM_{2.5} released more nitrogen oxides (NO_x), breaking the balance of O₃ consumption by NO while enhancing volatile organic compound (VOC) oxidation and promoting O₃ generation. The reduction in aerosols led to enhanced ultraviolet radiation, accelerating the photolysis reaction of NO₂, which caused a significant increase in O₃ concentration. Subsequently, due to the Chinese government's launch of the "Blue Sky Defense War" (2018–2020) environmental policy and the continued impact of

COVID-19 in China, emission reduction measures and epidemic blockade measures led to a reduction in precursor emissions. The sharp decline in NO_x in the short term caused urban areas to shift from a VOC-limited regime to an NO_x-limited regime, inhibiting the rate of O₃ generation, and the overall O₃ concentration level in China and most typical areas of China decreased. However, the spatial distribution of O₃ concentrations in the atmosphere in various places did not change significantly during the COVID-19 period. This was due to control measures, reduced traffic emissions, and decreased nitrogen oxide emissions, which promoted O₃ accumulation through the “titration effect”.

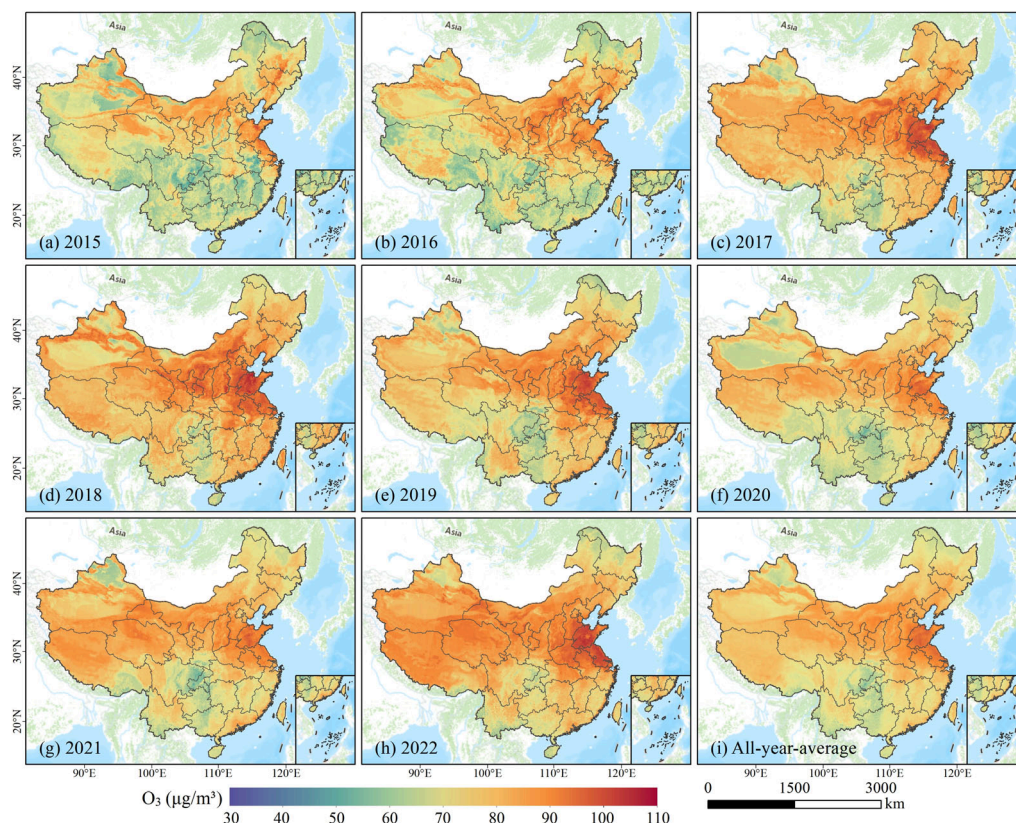


Figure 8. Spatial distribution of the annual average M7 O₃ concentration from 2015 to 2022, simulated based on the ST-LightGBM model.

From a spatial perspective, China’s O₃ distribution shows a similar pattern each year. The O₃ concentration in the east was higher than that in the west, and in the north, it was higher than that in the south. In 2022, the difference in the spatial distribution of O₃ between northern and southern China narrowed. The north and south showed large differences in different periods, mainly owing to differences in sunlight duration and O₃ chemical generation in different seasons. Among them, areas with high O₃ concentrations are concentrated in northwest, north, central, and east China, and the eastern coastal areas have higher O₃ concentrations than the inland areas. The main reason for this is that these areas are densely populated, heavy industries are well developed, and urbanization levels are high. These regions exhibit high levels of economic development in China. The O₃ concentration in northeast China is relatively low, mainly because the temperature in the northeast is low and solar radiation is weak, which inhibits the reaction of VOC and NO_x to generate O₃. In addition, the Sichuan and Tarim Basins also have low O₃ concentrations. This is mainly due to weak solar radiation, which is related to human activities and climate.

The four seasons are defined as spring (March–May), summer (June–August), autumn (September–November), and winter (December–February of the following year). Figure 9 shows the spatial distribution of the seasonal average concentration of M7 O₃ in China

from 2015 to 2022. The spatial distribution of O_3 concentrations varied significantly by season. The average O_3 concentration in summer was the highest, at $93.88 \pm 18.52 \mu g m^{-3}$. Pollution was most serious in summer, especially in BTH ($121.81 \pm 13.10 \mu g m^{-3}$) and NCP ($121.02 \pm 12.61 \mu g m^{-3}$). This was followed by spring ($89.88 \pm 9.33 \mu g m^{-3}$), autumn ($76.56 \pm 9.33 \mu g m^{-3}$), and winter ($57.58 \pm 10.27 \mu g m^{-3}$). Winter O_3 concentrations were significantly lower than those in other seasons (Table S5). The spatial distribution pattern of O_3 in autumn was similar to that in winter but was generally higher than that in winter, especially in parts of the southeast and northwest. The summer O_3 concentration pattern is similar to the interannual O_3 distribution. The NCP region is the most severely polluted area in the country, with the highest proportion of polluted days. Most of the northern parts of northeast, south, and southwest China belong to the low-value O_3 concentration area. The pollution range in summer was wider than that in the other seasons. In spring and autumn, as the O_3 concentration decreased, the national O_3 pollution range was significantly reduced; however, the NCP region remained the most serious O_3 high pollution area in the country. This dataset effectively reflects and describes the spatiotemporal distribution and changes in O_3 pollution at local and city scales.

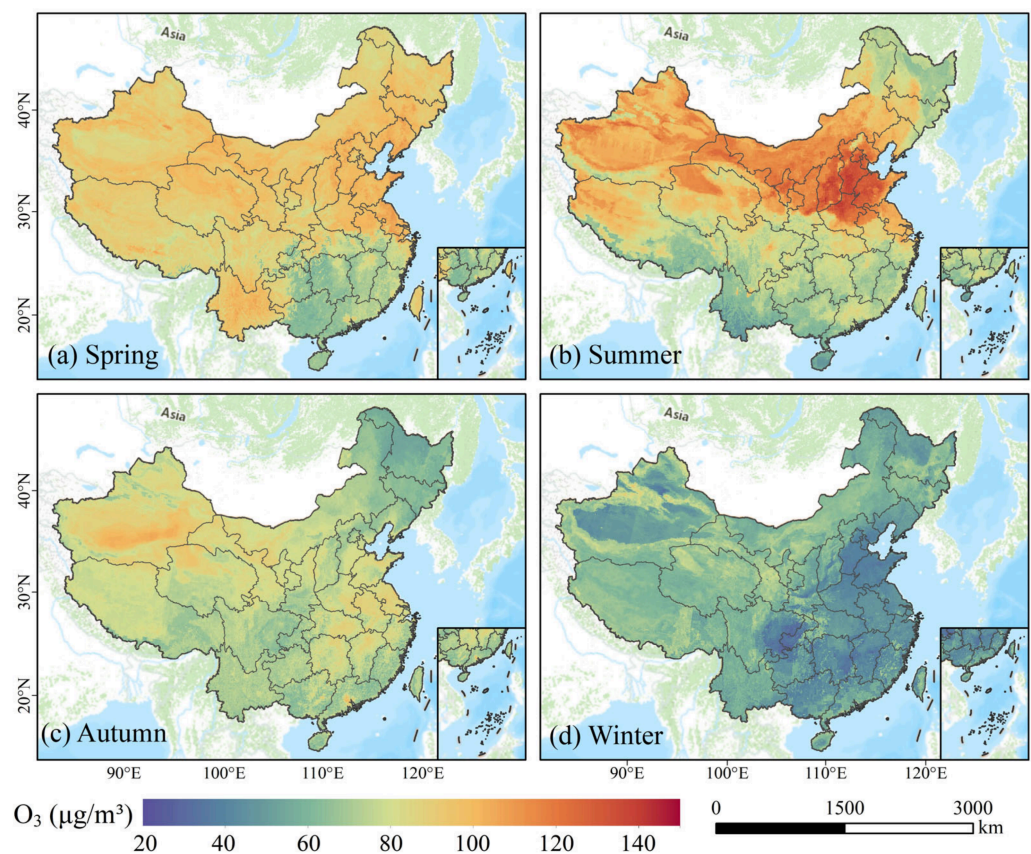


Figure 9. Spatial distribution of the multiyear seasonal average M7 O_3 concentration in China from 2015 to 2022.

4.4.2. Quantitative Analysis of Spatiotemporal Variation

Over the past eight years, O_3 concentrations have varied from the national to regional scales. The distribution of the multiyear average M7 O_3 concentrations indicated that high O_3 values were mainly concentrated in the North China Plain, whereas the southwest had relatively low O_3 concentrations (Figure 8i). The spatial variation characteristics of annual M7 O_3 were analyzed using linear slope calculations and the Mann–Kendall (M-K) trend test. Overall, O_3 pollution significantly increased in most parts of the country from 2015 to 2022, with regional variations. The annual O_3 concentration change rate was

-3.61 – $5.07 \mu\text{g m}^{-3}/\text{yr}$, with an average change rate of $0.83 \mu\text{g m}^{-3}/\text{yr}$. Approximately 79.21% of the regions exhibited an increasing trend in O_3 concentration (Figure 10a). The significance of the O_3 concentration change trend (Figure 10b) shows that the proportions of regions with a significant increase, no significant increase, significant decrease, and no significant decrease were 60.52%, 18.75%, 11.38%, and 9.36%, respectively. Regions such as Xinjiang and Tibet also face O_3 pollution problems. This shows that the problem of O_3 pollution should not be underestimated, and comprehensive measures should be taken to prevent and control it, just as haze is taken seriously.

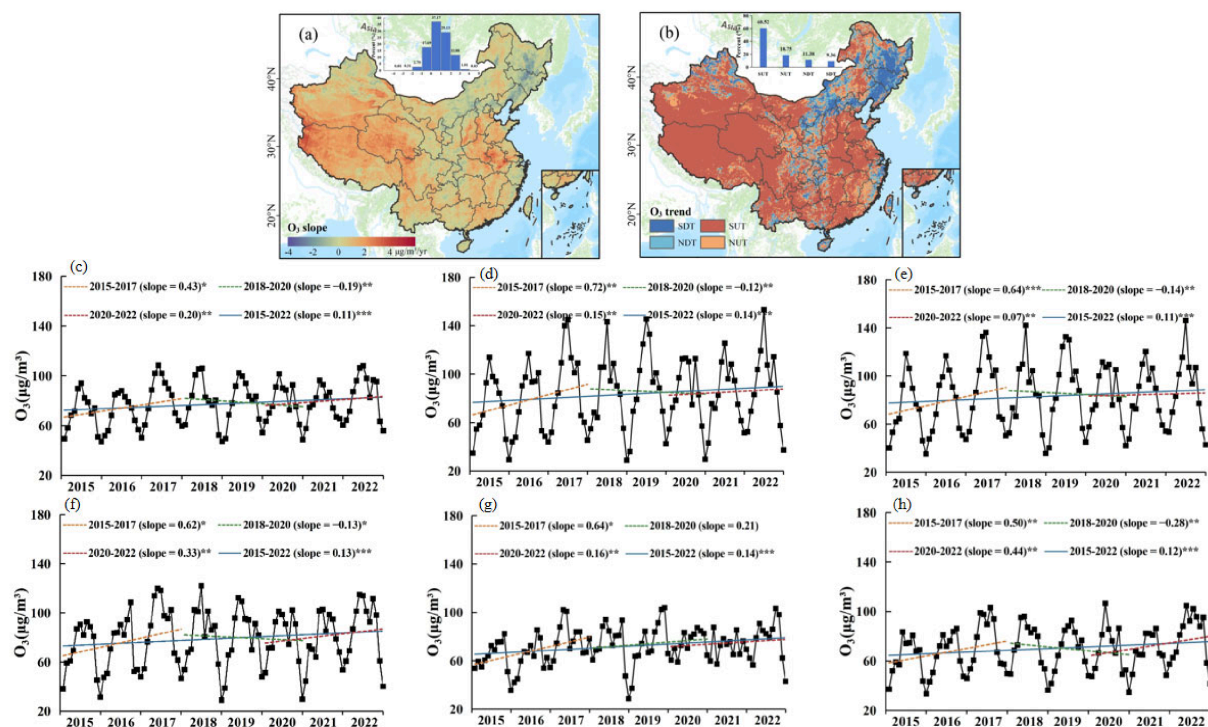


Figure 10. M7 O_3 linear trend spatial distribution and monthly average M7 O_3 concentration time series from 2015 to 2022. (a) Spatial distribution of the linear slope of annual M7 O_3 concentrations; (b) Spatial distribution of the trend significance levels, where SUT, NUT, SDT, and NDT indicate significant/non-significant uptrend and significant/non-significant downtrend, respectively; (c–h) Monthly average M7 O_3 concentration time series for China and five typical regions (NCP, BTH, YRD, PRD, and SCB). The different colored lines represent trend lines for different periods, where *, **, and *** indicate significance at the 90%, 95%, and 99% confidence levels, respectively.

The monthly average concentration time series of M7 O_3 in China and the five typical regions were calculated based on monthly M7 O_3 data (Figure 10c–h). The change in O_3 concentration exhibited an annual periodicity, with a trend of initially increasing and then decreasing each year. The NCP and BTH regions experienced a single annual peak, whereas the southern PRD and SCB regions occasionally displayed two peaks in certain years. Overall, O_3 concentrations across most of the country showed an upward trend (slope = 0.11, $p < 0.01$), with the strongest increase observed in the NCP and PRD regions (slope = 0.14, $p < 0.01$). The trend in M7 O_3 concentration values varied across different periods and regions. During the early study period (2015–2017), when the Chinese government implemented the Clean Air Action Plan (2013–2017), O_3 concentrations in China and its five typical regions showed a significant growth trend, with a growth rate of 0.43–0.72. O_3 pollution in various urban agglomerations is relatively serious, and the atmospheric environment has deteriorated, with O_3 pollution in the NCP region being the most serious. The main reason for the rapid increase in O_3 in the NCP region was the

rapid reduction in $PM_{2.5}$ due to environmental policies. $PM_{2.5}$ decreased by approximately 40%, slowing down the aerosol particles that reactively absorbed HO_2 radicals, thereby stimulating the production of O_3 [83].

Subsequently, during the BlueSky Defense Plan (2018–2020), except for the PRD region [84], the situation in all regions was the opposite, and the O_3 concentration decreased. The O_3 concentration in the SCB region decreased the fastest (slope = -0.28). The reason for this situation is that the implementation of environmental policies set targets for the reduction in O_3 precursors, NO_x , and VOCs, controlling the emission of pollutants such as NO_x and reducing the concentration of precursors required for the O_3 reaction, leading to a decrease in O_3 concentration. During the Clean Air Action Plan, China's anthropogenic nitrogen oxide emissions decreased by 20%, which also reduced pollutant emissions. However, owing to the sharp decrease in particulate matter concentrations, the change in O_3 concentrations showed an upward trend. Although the change in O_3 concentrations in China and various urban agglomerations showed a downward trend between 2018 and 2020, the overall O_3 level was higher than that at the beginning of the study, and O_3 pollution was still relatively serious.

In the later period of the study (2020–2022), during the COVID-19 lockdown, O_3 concentrations in China and typical regions increased unexpectedly. This was mainly due to the COVID-19 outbreak in 2020, when the country was in lockdown. Sensitivity to O_3 generation changed before and after the lockdown, and most of the changes were in VOC-limited regimes. The suspension of industrial activities and traffic volume caused NO_x emissions to drop rapidly to extremely low levels, and the titration consumption of O_3 rapidly weakened, leading to an increase in O_3 concentrations during the lockdown [85,86]. In addition, the COVID-19 lockdown occurred in winter, which was an active period for the transport of stratospheric O_3 to the troposphere. Studies have shown that during short periods of stratospheric intrusion, the proportion of O_3 from the stratosphere in the surface O_3 reached 40–45%. The O_3 input from the stratosphere increased the ground O_3 concentration, contributing an average of 5–10% to the increase in O_3 during the lockdown period [87]. After the lockdown was lifted, some areas may have resumed work and production, and industrial activities and traffic volume may have increased again, resulting in increased emissions of some pollutants, which may have caused the O_3 concentrations to rise again. The momentum of O_3 pollution was not effectively curbed. It is imperative to establish clearer O_3 reduction targets, adjust regional prevention and control measures according to pollution characteristics, and develop coordinated strategic policies that focus on O_3 governance.

The spatial patterns of O_3 exposure described above have direct implications for crop production. Notably, the North China Plain, which serves as China's primary winter wheat cultivation area, also exhibits the most severe O_3 pollution. This spatial overlap between high O_3 exposure and intensive agricultural production constitutes a significant risk superposition. Building on this, the following sections quantify the resulting yield losses and economic costs, explicitly establishing the link between O_3 exposure patterns and tangible agricultural outcomes.

4.5. Winter Wheat Yield Reduction and Economic Loss

4.5.1. Assessment of Winter Wheat Yield Relative Yield Loss

We used O_3 data to assess the RYL of winter wheat due to O_3 exposure across various provinces in China, based on the M7 metric. The average RYL during the winter wheat growing seasons from 2015 to 2022 was 5.61%, 5.10%, 7.12%, 7.20%, 4.62%, 4.67%, 3.44% and 4.90%, respectively (Figure S10). The trend in winter wheat RYL paralleled that of surface O_3 , with provinces experiencing higher RYL also showing more severe O_3 pollution. During

our study period (Figure 11), the highest estimated RYL of winter wheat was in the Jiangsu Province (5.48–11.31%), followed by Shandong (6.03–10.93%), Shanghai (4.88–10.19%), Tianjin (3.84–9.18%), and Shanxi (4.68–9.12%) Provinces. Jiangsu and Shandong Provinces are the major winter wheat-growing areas in China. The ranges in parentheses represent the minimum and maximum provincial RYLs during 2015–2022.

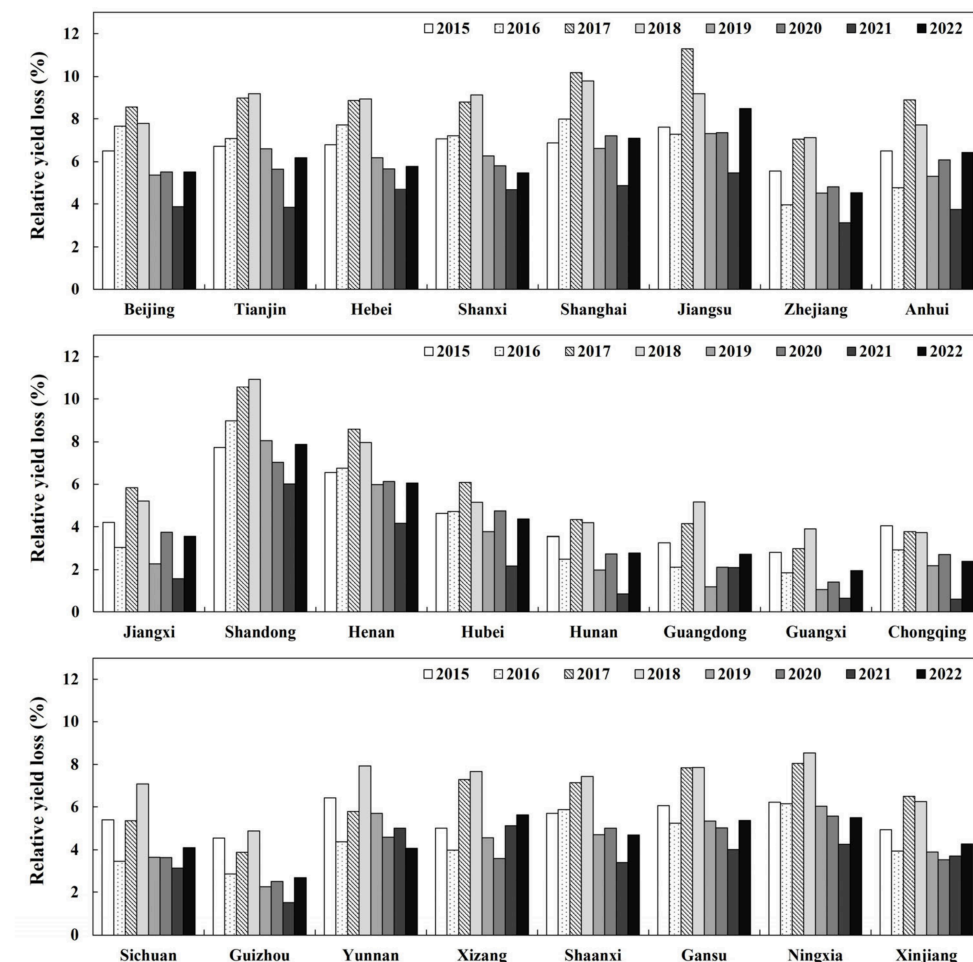


Figure 11. Relative yield losses of winter wheat based on the M7 metric in various provinces in China from 2015 to 2022.

We further assessed the RYL of winter wheat across Chinese provinces based on the AOT40 ozone metric. The average RYL during the winter wheat growing seasons from 2015 to 2022 were 26.50%, 27.11%, 30.99%, 32.78%, 28.36%, 32.08%, 27.14%, and 36.54%, respectively (Figure S10). Similar to the results derived from the M7 metric, the temporal trend of winter wheat RYL closely followed that of surface O_3 levels, with higher RYL occurring in provinces experiencing more severe O_3 pollution (Figure 12). Shandong Province exhibited the highest estimated RYL (40.14–60.87%), followed by Henan (28.52–58.88%), Shanxi (37.25–55.23%), Hebei (38.60–54.96%), and Jiangsu (23.23–54.72%). It should be noted that these provinces are the main wheat production regions of China. Overall, the AOT40-based assessment confirms the substantial impact of O_3 exposure on winter wheat yield and reinforces the significant threat posed to national food security.

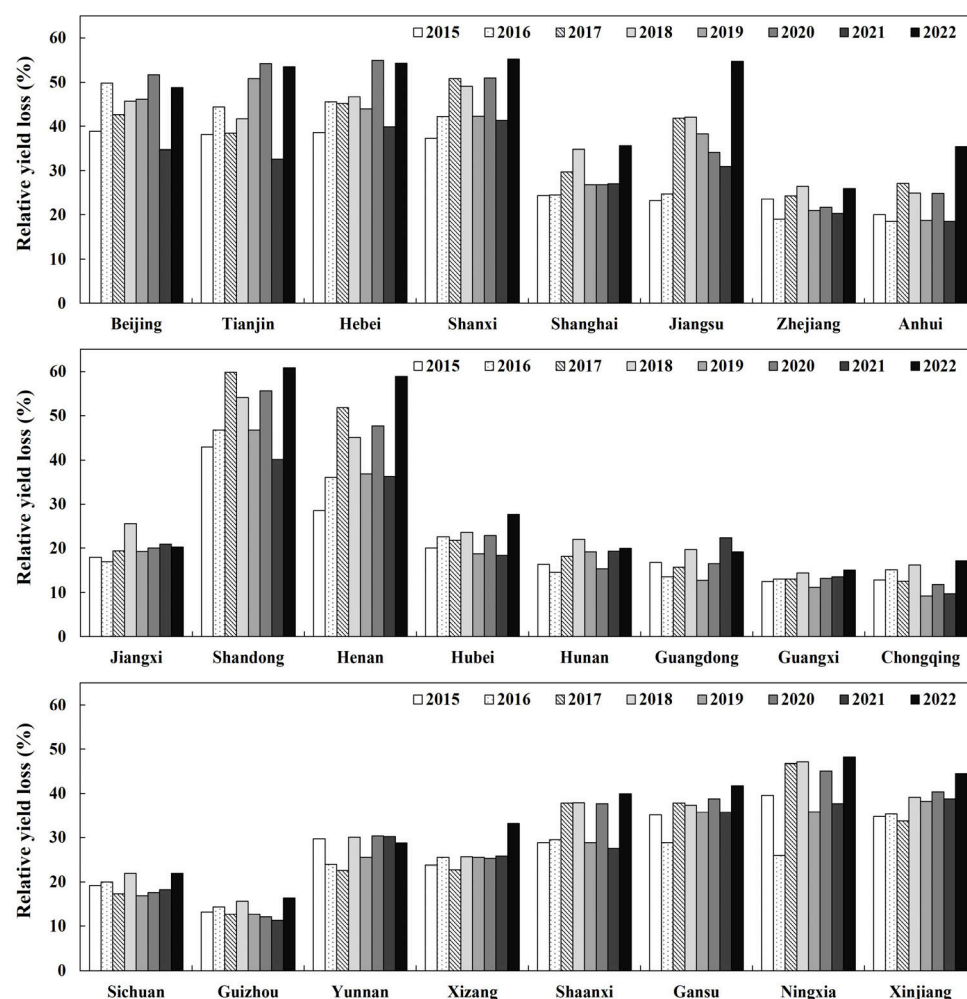


Figure 12. Relative yield losses of winter wheat based on the AOT40 metric in various provinces in China from 2015 to 2022.

4.5.2. Crop Production Loss (CPL) and Associated Economic Loss (EL)

Based on the provincial RYL estimates derived from both the M7 and AOT40 metrics, we further calculated the crop production and economic losses of winter wheat caused by O_3 exposure in China during 2015–2022. When estimated using the M7 metric (Table 2), the national winter wheat crop production losses were 8.89 million metric tons, 8.87 million metric tons, 12.71 million metric tons, 11.79 million metric tons, 8.52 million metric tons, 8.42 million metric tons, 6.30 million metric tons, and 9.16 million metric tons, respectively, accounting for 7.18%, 7.56%, 9.94%, 9.43%, 6.67%, 6.53%, 4.78% and 6.91% of the actual total winter wheat production in China that year. As the concentration of O_3 increases, the loss of winter wheat production becomes increasingly serious. During 2015–2022, the economic cost of China's winter wheat production loss is estimated to be 3.37 billion US\$, 3.15 billion US\$, 4.44 billion US\$, 4.10 billion US\$, 2.77 billion US\$, 2.73 billion US\$, 2.21 billion US\$ and 3.13 billion US\$, respectively. Among the provinces, Henan Province had the highest yield loss (1.65–3.48 million metric tons), followed by Shandong Province (1.69–3.03 million metric tons) and Hebei Province (0.72–1.46 million metric tons). The province with the highest economic cost loss is also Henan Province (0.58–1.22 billion US\$), followed by Shandong Province (0.59–1.05 billion US\$) and Hebei Province (0.25–0.51 billion US\$). The yield loss depends not only on the RYL, but also on the regional crop-planting productivity. A few provinces have high RYL, but their winter wheat planting area is small, and the yield is much lower than that of other provinces, resulting in low yield losses. In contrast,

provinces with low RYL but in the main winter wheat-producing areas may have higher yield losses.

Table 2. Impact of O₃ pollution on winter wheat yield losses and direct economic losses in various provinces in China based on the M7 metric.

Region	CPL (10 ⁴ metric tons)								ECL (10 ⁷ US\$)							
	2015	2016	2017	2018	2019	2020	2021	2022	2015	2016	2017	2018	2019	2020	2021	2022
Henan	245.83	251.66	347.77	312.20	238.35	245.70	165.26	246.06	93.15	89.41	121.56	108.5	77.39	79.79	57.89	84.14
Shandong	196.86	230.93	295.15	303.35	223.65	194.58	169.07	226.07	74.60	82.05	103.17	105.4	72.62	63.19	59.23	77.31
Hebei	104.25	119.47	145.87	141.98	95.97	86.01	72.10	89.56	39.50	42.45	50.99	49.35	31.16	27.93	25.26	30.63
Anhui	98.21	69.55	160.64	134.32	93.00	108.41	66.13	118.49	37.22	24.71	56.15	46.69	30.20	35.21	23.17	40.52
Jiangsu	96.70	87.84	165.13	130.39	103.75	105.84	77.85	126.67	36.64	31.21	57.72	45.32	33.69	34.37	27.27	43.32
Shaanxi	27.65	27.76	31.21	32.27	18.89	21.72	14.90	21.18	10.48	9.86	10.91	11.22	6.13	7.05	5.22	7.24
Sichuan	24.28	14.76	14.05	18.64	9.24	9.22	7.88	10.58	9.20	5.25	4.91	6.48	3.00	2.99	2.76	3.62
Xinjiang	21.90	18.15	27.02	24.55	15.78	15.03	18.53	21.06	8.30	6.45	9.44	8.53	5.12	4.88	6.49	7.20
Hubei	20.43	21.18	27.69	22.25	15.32	19.93	8.83	18.52	7.74	7.53	9.68	7.73	4.97	6.47	3.09	6.33
Shanxi	20.59	21.25	22.40	22.95	15.13	14.58	11.94	14.18	7.80	7.55	7.83	7.98	4.91	4.73	4.18	4.85
Gansu	11.12	9.41	14.83	14.66	9.83	9.06	7.55	10.40	4.21	3.34	5.18	5.10	3.19	2.94	2.65	3.56
Yunnan	6.21	4.08	4.53	6.41	4.34	3.36	3.25	2.52	2.35	1.45	1.58	2.23	1.41	1.09	1.14	0.86
Tianjin	3.90	4.18	5.49	5.04	3.65	3.29	2.52	3.92	1.48	1.48	1.92	1.75	1.19	1.07	0.88	1.34
Guizhou	2.94	1.75	1.66	1.70	0.76	0.86	0.51	0.80	1.11	0.62	0.58	0.59	0.25	0.28	0.18	0.27
Zhejiang	2.07	1.05	3.18	2.74	1.53	2.07	1.55	2.62	0.78	0.37	1.11	0.95	0.50	0.67	0.54	0.90
Shanghai	1.47	1.05	1.16	1.41	0.41	0.41	0.38	0.82	0.56	0.37	0.40	0.49	0.13	0.13	0.13	0.28
Chongqing	0.96	0.59	0.38	0.32	0.15	0.17	0.04	0.15	0.37	0.21	0.13	0.11	0.05	0.06	0.01	0.05
Ningxia	0.99	0.75	0.81	1.16	0.74	0.69	0.33	0.30	0.37	0.27	0.28	0.40	0.24	0.22	0.12	0.10
Xizang	0.88	0.76	1.23	1.20	0.62	0.46	0.70	0.73	0.33	0.27	0.43	0.42	0.20	0.15	0.25	0.25
Beijing	0.77	0.70	0.58	0.44	0.25	0.26	0.27	0.54	0.29	0.25	0.20	0.15	0.08	0.09	0.10	0.19
Hunan	0.35	0.15	0.44	0.35	0.15	0.22	0.07	0.22	0.13	0.05	0.15	0.12	0.05	0.07	0.02	0.07
Jiangxi	0.11	0.08	0.19	0.18	0.07	0.13	0.05	0.11	0.04	0.03	0.07	0.06	0.02	0.04	0.02	0.04
Guangxi	0.03	0.02	0.02	0.02	0.01	0.01	0.004	0.02	0.010	0.007	0.005	0.007	0.002	0.003	0.001	0.005
Guangdong	0.01	0.01	0.01	0.01	0.002	0.002	0.002	0.003	0.004	0.002	0.002	0.004	0.001	0.001	0.001	0.001
China	888.5	887.1	1271.4	1178.5	851.6	842.0	629.7	915.5	336.7	315.2	444.41	409.6	276.5	273.5	220.6	313.9

Based on the AOT40 metric (Table S9), the estimated winter wheat crop production losses in China from 2015 to 2022 were 44.61, 45.48, 57.44, 60.97, 50.58, 60.86, 49.07, and 76.26 million metric tons, respectively. The corresponding economic losses were estimated at 16.90, 16.16, 20.08, 21.19, 16.42, 19.76, 17.19, and 26.08 billion US\$. Among provinces, Shandong, Henan, and Hebei experienced the largest crop production and economic losses, consistent with their status as major winter wheat-producing regions. The AOT40-based estimates generally showed higher losses than those derived from the M7 metric, reflecting the stronger sensitivity of AOT40 to high O₃ exposure. Taken together, the results based on both M7 and AOT40 metrics consistently demonstrate that O₃ pollution has caused severe reductions in winter wheat crop production and considerable economic losses, underscoring the significant threat posed to China's food security.

5. Discussion

5.1. Uncertainty and Error Analysis

Based on the linear regression slope and y-intercept between the ground-measured M7 O₃ concentrations and the model-estimated values, our model appeared to underestimate the O₃ concentration in high-value areas. As the values measured at the observation station increased, the underestimation of the model became more evident. This indicates that the predictive effect of our model may be smaller when O₃ pollution is more severe. This may be because the number of samples with high O₃ pollution levels was small and the information obtained during model training was low, which hindered the training ability of the model. Despite this underestimation, the overall performance of the proposed model was satisfactory. At different time scales, the model's prediction slope changed by less than 0.1, and the prediction accuracy CV-R² changed by less than 0.08, reflecting the robustness and stability of the model. This shows that our model is more sensitive to changes in the independent variables and can more accurately obtain information to identify O₃ pollution.

Our model exhibits varying accuracy and spatial prediction capabilities at the regional scale, which are closely related to the density and spatial distribution of ground monitoring stations, as well as the meteorological, natural, and population conditions in each region.

Generally, a higher density of regional stations correlates with better model performance and reduced prediction uncertainty. This is because the larger the number of samples, the more information about the O₃ changes that can be covered, providing richer data for model training. In general, with an increase in the number of monitoring stations and samples, the overall accuracy and spatial prediction ability of the ST-LightGBM model prediction gradually improved, and the estimated uncertainty gradually decreased. The overall prediction accuracy of the model in different regions was higher than 0.91, and the spatial prediction accuracy was higher than 0.81, which shows that our model also has good spatial prediction ability in places without monitoring stations. This offers a practical framework for integrating satellite remote sensing with in situ observations to improve regional air pollution monitoring, though further validation and ground-based measurements remain essential for ensuring accuracy and robustness.

5.2. Comparison with Related Studies

Experimental studies using open-top chamber and free-air concentration enrichment (FACE) systems have established a well-characterized chain of physiological responses linking ambient O₃ exposure to reduced wheat grain yield [10,88–91]. Ozone enters leaves through stomata and generates reactive oxygen species (ROS) in the apoplast; when ROS accumulation exceeds the detoxification threshold, it triggers programmed cell death and suppresses photosynthetic capacity. At the leaf level, elevated O₃ reduces net photosynthetic rate, stomatal conductance, and transpiration rate in flag leaves, indicating impairment of both stomatal and non-stomatal photosynthetic components. At the canopy level, ozone accelerates leaf senescence and reduces green leaf area index, shortening the effective period of carbon assimilation. The most severe impacts occur during grain filling, when flag leaf photosynthesis supplies the bulk of assimilates for grain development. Ozone-induced senescence during this period reduces assimilate supply and kernel growth rate, ultimately lowering thousand-kernel weight rather than kernel number. These experimentally established mechanisms provide the physiological foundation for interpreting the spatial O₃–yield associations observed in our national-scale assessment.

These physiological mechanisms also help explain the divergence between M7 and AOT40 estimates observed in this study. AOT40 accumulates O₃ concentrations exceeding 40 ppb, thereby capturing the acute high-concentration episodes. In contrast, M7 averages daytime O₃ concentrations across the entire growing season, diluting the impact of these critical peak episodes. The substantially higher losses estimated by AOT40 (26.50–36.54%) compared to M7 (3.44–7.20%) are therefore consistent with the expectation that episodic high-O₃ events during grain filling impose disproportionate damage through accelerated senescence and photosynthetic collapse [10,91].

The credibility of crop yield loss estimates depends fundamentally on the quality of the underlying O₃ exposure data. To ensure that our reconstructed M7 and AOT40 datasets are sufficiently robust for agricultural risk assessment, we evaluated model performance through multiple comparisons. We compared the fitting results of the ST-LightGBM model constructed in this study with the model validation results from recent related studies in China using the cross-validation method (Table S6). Previous studies using traditional statistical or chemical models exhibited varying biases: some underestimated O₃ concentrations, while others overestimated them. Chen et al. [92] used RF to predict O₃ concentrations in China at a resolution of 6 km from 2014 to 2019. The slope of the model-fitting linear regression equation is $Y = 1.01X - 1.07$, and the predicted result is higher than the true value. In addition, the predictions of our model were better than those of related studies that used chemical model simulations. Lin et al. [77] and Liu et al. [93] used the WRF-CMAQ model to simulate O₃ concentrations in China

during 2014 and 2015, respectively. The prediction accuracy of the model constructed in our study was much higher than that of previous studies. In the study of surface O₃ concentration estimation with the same spatial resolution of 10 km, the model we constructed has improved the prediction accuracy accordingly. Our model's prediction uncertainty is smaller compared with studies of surface O₃ concentration estimation using models such as RF (CV-R² = 0.69, RMSE = 26.00 µg m⁻³; [28]), XGBoost (CV-R² = 0.78, RMSE = 21.47 µg m⁻³; [94]), data fusion (CV-R² = 0.70, RMSE = 26.20 µg m⁻³; [95]), STET (CV-R² = 0.87, RMSE = 17.10 µg m⁻³; [5]), and 3D CNN (CV-R² = 0.88, RMSE = 15.65 µg m⁻³; [34]). These comparisons demonstrate that the ST-LightGBM framework achieves robust estimation with only 15 routinely available predictors derived from reanalysis and satellite products, without relying on computationally expensive chemical transport models or emission inventories that often entail substantial uncertainties. This feature is particularly critical for agricultural regions where ground monitoring stations are sparsely distributed. The reconstructed M7 and AOT40 datasets therefore provide a reliable and updatable foundation for ozone risk assessment of winter wheat on the national scale. In relevant studies on modeling surface O₃ concentration simulations, most selected the MDA8 concentration indicator. MDA8 is typically used to assess human health risks because long-term exposure to high levels of O₃ is harmful to the human respiratory system and health [8,96]. However, O₃ pollution can also cause significant harm to crops, posing a threat to national food security, human lives, and property safety [74,97,98]. To address the lack of O₃-related data in agriculture, this study selected the M7 concentration index to simulate the surface O₃ concentration. In agriculture, high concentrations of O₃ over a short period can directly damage plants, affecting their growth and yield. M7 focuses more on the impact of O₃ exposure over a shorter period of time on plants and the environment, and is widely used to assess the impact of O₃ pollution on crops [73,99].

At the same time, many studies have used other O₃ indicators to evaluate the impact of O₃ exposure on winter wheat, such as AOT40 (accumulated value of O₃ concentration with hourly concentration exceeding 40 parts-per-billion (ppb)), SUM06 (accumulated value of O₃ with concentration greater than 60 ppb during the growth period), and W126 (weighted O₃ accumulation value) [38,100]. However, these indicators rarely consider situations where the O₃ concentration is lower than 40 ppb. Studies have shown that O₃ levels below 40 ppb can damage crops [101–103]. Therefore, the use of the M7 indicator can explain crop yield loss caused by O₃ concentrations lower than 40 ppb. The PODY—phytotoxic O₃ dose, the accumulation of stomatal O₃ uptake flux per unit area exceeding the critical value Y nmol m⁻²·s⁻¹ throughout the growing season—indicator can also explain the impact of lower O₃ concentrations [80], but this indicator has to simulate the stomatal conductance of crops throughout their entire growth period, which involves more parameters and environmental conditions, and the calculation process is more complicated. In addition to M7, we also applied the AOT40 metric to assess winter wheat yield losses and associated economic costs. This dual metric approach enables a comprehensive evaluation of O₃ impacts, facilitating comparison across studies and providing robust evidence of the threat posed by O₃ pollution to wheat production in China.

Beyond these methodological advancements, the spatial pattern of yield losses derived from our datasets reveals a clear biological reality. Provinces with higher O₃ concentrations, particularly those in the North China Plain (NCP) and adjacent regions, consistently exhibited more severe relative yield losses. For instance, Shandong and Henan—both major winter wheat producers—ranked among the highest in both O₃ concentration and yield loss. This spatial consistency is consistent with the experimentally established physiological mechanism detailed above: O₃-induced photosynthetic impairment during grain filling,

leading to reduced thousand-kernel weight rather than kernel number [91]. Given that winter wheat is one of the most O₃-sensitive staple crops, the observed yield reductions are not merely statistical correlations but are biologically interpretable as the macro-scale manifestation of O₃-induced physiological damage. The agreement between O₃ hotspots and yield loss reinforces the plausibility of a causal link between O₃ exposure and agricultural damage, highlighting the vulnerability of China's wheat production to O₃ pollution and the need for targeted mitigation in high-risk regions, such as the NCP.

5.3. Implications for Food Security

Beyond the spatial correspondence between ozone hotspots and yield losses, the magnitude of these losses itself reflects the potential threat to food security. The annual production losses estimated in this study, based on the M7 metric, range from 6.30 to 12.71 million tons, accounting for 4.78% to 9.94% of China's actual winter wheat production in the corresponding years. This is equivalent to the annual wheat production of a medium-sized province, such as Anhui or Jiangsu. At the provincial level, Henan, Shandong, and Hebei are consistently the most severely affected, with annual losses of 1.65–3.48, 1.69–3.03, and 0.72–1.46 million tons, respectively. These three provinces together account for a substantial proportion of the national total losses, implying that mitigation efforts targeted at this region could yield significant national benefits. Moreover, the AOT₄₀-based estimates, ranging from 44.61 to 76.26 million tons, further underscore the severity of ozone pollution as a threat to wheat production. However, the ozone threat is not limited to its direct yield losses; the impacts of O₃ on crops are modulated by meteorological conditions, and future climate warming may further amplify this risk.

Furthermore, the ozone risk identified in this study does not operate in isolation. Recent evidence indicates that elevated vapor pressure deficit and temperature can intensify ozone exposure through enhanced photochemical formation and altered stomatal conductance, while drought may paradoxically buffer ozone uptake by inducing stomatal closure, thereby reducing O₃ flux into leaves [104]. This is particularly concerning for the North China Plain, where spring and early summer are characterized by frequent concurrent heat and drought events. These conditions not only directly stress wheat during grain filling but also promote photochemical O₃ formation. The compound stress of O₃, heat, and drought represents a systemic risk that may exceed the sum of individual stressors. Given that VPD is a key limitation on stomatal conductance, future climate warming could further complicate the O₃ risk landscape by altering the physiological trade-off between stomatal opening, which enables CO₂ uptake but also increases O₃ influx, and water conservation. This would ultimately exacerbate the threat to China's food security. Addressing these systemic risks requires integrated strategies that account for both air quality management and agricultural adaptation, particularly in China's major wheat-producing regions.

5.4. Limitations and Further Work

The ST-LightGBM model was used to reliably simulate the Chinese surface M7 O₃ concentration dataset. The overall model performance was good; however, this study has certain limitations. The main limitations are as follows. (1) From the perspective of the spatial distribution of ground air pollution monitoring stations, the number of stations in the western region is small, and the distribution is sparse, which limits the training performance of the model. Consequently, the fitting accuracy in the eastern region is higher, whereas that in the western region may be lower. In the future, as the number of air quality monitoring stations increases, this problem will be solved, and the prediction accuracy of the model will be further improved. Another issue that must be noted is that we currently use real-time O₃ observation data from monitoring stations for incorporation into the

model to ensure the accuracy and stability of the model estimation. The key issue to be solved in future work is to obtain historical observations before the establishment of the monitoring station, while ensuring the accuracy of the prediction. (2) The spatial resolution of the M7 O₃ concentration data reconstructed in this study was 10 km. If higher-resolution model training input variable data can be obtained, a more refined O₃ concentration dataset can be simulated and reconstructed. For downscaling research, we are working diligently to obtain higher-resolution meteorological and remote sensing data and incorporate them into our research to obtain a higher-resolution O₃ concentration dataset to more accurately reveal the spatiotemporal variation of O₃ in a small area.

In addition to the methodological limitations noted above, several agronomic factors warrant consideration when interpreting our estimates. First, the dose–response functions applied in this study (M7 and AOT40) were derived from field experiments with specific wheat cultivars; however, different wheat cultivars exhibit varying sensitivity to O₃ [91]. Our national-scale assessment did not account for cultivar-specific responses or the spatial distribution of different cultivars, which may introduce uncertainty in province-level yield loss estimates. Second, the dose–response relationships used here do not explicitly incorporate the moderating effects of soil properties, irrigation, or nitrogen management—factors that have been shown to alter crop sensitivity to O₃ stress [104]. Future research integrating spatially explicit soil and management data could refine these estimates. Third, while our analysis focused on O₃ as the primary stressor, crops in the field are simultaneously exposed to multiple stresses, including drought, heat, and pest pressure. The interactions between O₃ and these co-occurring stressors may amplify or mitigate yield losses in ways that a single-stressor framework cannot capture [102]. Future assessments should therefore move toward multi-stressor approaches to better represent the actual field conditions facing wheat production in China. Acknowledging these agronomic uncertainties does not undermine our core findings but highlights directions for further refinement.

6. Conclusions

Surface ozone (O₃) is a primary pollutant adversely affecting the environment, human health, and crop production. To address the scarcity and uneven distribution of ground monitoring stations, as well as the lack of agriculturally relevant O₃ simulation data, this study reconstructed a 10 km, spatially continuous M7 O₃ dataset across China from 2015 to 2022 using the ST-LightGBM framework. The model demonstrated robust performance, with an overall CV-R² of 0.91 and a spatial CV-R² of 0.82, and its accuracy improved annually. Spatiotemporal analysis revealed a phased upward trend in surface O₃ pollution (0.83 µg m^{−3} yr^{−1}), with concentrations increasing across 79.22% of the study area; the North China Plain emerged as the primary hotspot, exhibiting the highest concentrations, particularly in summer. SHAP analysis further indicated that downward solar radiation and temperature were the dominant drivers of O₃ variability (contribution rate > 42%), providing interpretability for the model and identifying key drivers for the concentration reconstruction. Concurrently, a complementary AOT40 dataset at the same spatial resolution was generated (CV-R² = 0.84), with model performance improving consistently over time.

Based on the M7 and AOT40 metrics, we evaluated the impacts of surface O₃ exposure on winter wheat yields and associated economic losses in China. The results showed that from 2015 to 2022, the average annual relative yield loss of winter wheat ranged from 3.44% to 7.20% (M7) and from 26.50% to 36.54% (AOT40), corresponding to economic losses of USD 2.21–4.44 billion and USD 16.16–26.08 billion, respectively. Crucially, the spatial overlap between high-surface O₃ regions and major winter wheat cultivation areas—particularly in the North China Plain—means that O₃ pollution sys-

tematically threatens the core of China's wheat production system. The stark divergence between the M7 and AOT40 estimates underscores the necessity of incorporating cumulative high-concentration exposure metrics into agricultural risk assessments. In summary, the reconstructed M7 and AOT40 exposure datasets, combined with crop-specific response functions, provide a quantitative framework for identifying high-risk agricultural regions and highlight the urgency of integrating agricultural considerations into air quality management policies to safeguard food security under intensifying surface O₃ pollution.

To translate these findings into actionable strategies, we propose the following recommendations. For breeding and cultivar selection, breeding programs should prioritize ozone-tolerant winter wheat varieties in high-loss provinces, including Henan, Shandong, and Hebei, where the spatial overlap of high production and high O₃ exposure creates the greatest national risk. For agronomic practices, adjustments such as shifting sowing dates to avoid peak O₃ periods during grain filling and optimizing irrigation and nitrogen management to enhance antioxidant capacity can help buffer crops against O₃ stress. For policy measures, ozone forecasts should be incorporated into the agricultural meteorological early warning system during the grain-filling stage, and temporary precursor emission reductions should be considered for the North China Plain during critical high-ozone episodes. Integrating these agricultural considerations into air quality policy will be essential for protecting China's wheat production from the growing threat of surface O₃ pollution.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/agronomy16171671/s1>. Table S1: Summary of the data sources used in this study; Table S2: Correlation analysis between all variables; Table S3: Collinearity analysis of independent variables; Table S4: Sample-based cross-validation results of the AOT40 ozone metric from 2015 to 2022; Table S5: Statistics of cross-validation results of the estimated AOT40 ozone metric in typical regions; Table S6: Annual mean M7 O₃ concentrations from 2015 to 2022; Table S7: Seasonal mean M7 O₃ concentrations from 2015 to 2022; Table S8: Comparative statistics of model performance between the ST-LightGBM model and other O₃ studies at the Chinese scale; Table S9: National AOT40-based winter wheat production and economic losses in China (2015–2022); Table S10: Provincial growing season and O₃ exposure window for winter wheat in China; Figure S1: Histograms and statistics of variables used for modeling; Figure S2: Sorted normalized importance of each feature in the near-surface M7 O₃ estimation; Figure S3: Accuracy verification results of imputation of OMI NO₂ using the SSA method; Figure S4: Accuracy verification results of imputation of OMI O₃ using the SSA method; Figure S5: Original and imputed OMI NO₂ on four selected days; Figure S6: Original and imputed OMI O₃ on four selected days; Figure S7: Out-of-station ten-fold cross-validation results of M7 O₃ estimates from 2015 to 2022; Figure S8: Verification results of monthly average M7 O₃ estimates in China from 2015 to 2022; Figure S9: Density scatter plot of out-of-sample cross-validation results for daily M7 O₃ estimates by season; Figure S10: Relative yield loss based on M7 and AOT40 for winter wheat in China.

Author Contributions: H.Z.: Writing—Original Draft, Software, Methodology, Investigation, Conceptualization, Data Curation. R.L.: Writing—Review and Editing, Software, Data Curation, Methodology. X.S.: Methodology, Investigation, Software, Data Curation. Y.L.: Writing—Review and Editing, Methodology, Investigation, Conceptualization. J.J.: Writing—Review and Editing, Supervision, Methodology, Funding Acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (42561064), the Guangxi Natural Science Foundation (2026GXNSFAA00641000), the Guangxi Science and Technology Program (AD23026142), and the Innovation Project of Guangxi Graduate Education (YCSW2026183).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The surface O₃ hourly concentration observation station data were provided by the Chinese National Environmental Monitoring Center (<http://www.cnemc.cn/en/>, accessed on 15 February 2025). The ERA5 reanalysis is available at <https://cds.climate.copernicus.eu/>, accessed on 16 February 2025. The SRTM DEM data are available at <https://gdex.ucar.edu/>, accessed on 15 March 2025. The WorldPop data are available at <https://hub.worldpop.org/>, accessed on 15 March 2025. Tropospheric NO₂ (OMNO₂d) and total column O₃ (OMDOAO₃e) data are available at <https://disc.gsfc.nasa.gov/>, accessed on 15 March 2025. MODIS series products, including NDVI (MOD13A3) and Land Cover Type (MCD12C1) products, are available at <https://www.earthdata.nasa.gov/>, accessed on 15 March 2025.

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

References

- Chen, Y.; Shen, H.; Shen, G.; Ma, J.; Cheng, Y.; Russell, A.G.; Zhao, S.; Hakami, A.; Tao, S. Substantial Differences in Source Contributions to Carbon Emissions and Health Damage Necessitate Balanced Synergistic Control Plans in China. *Nat. Commun.* **2024**, *15*, 5880. [CrossRef] [PubMed]
- Wang, J.; Qian, J.; Chen, J.; Li, S.; Yao, M.; Du, Q.; Yang, N.; Zhang, T.; Yin, F.; Deng, Y.; et al. High-Resolution Full-Coverage Ozone (O₃) Estimates Using a Data-Driven Spatial Random Forest Model in Beijing-Tianjin-Hebei Region, China. *J. Hazard. Mater.* **2024**, *480*, 136047. [CrossRef] [PubMed]
- Yu, X.; Xi, M.; Wu, L.; Zheng, H. Spatiotemporal Weighted for Improving the Satellite-Based High-Resolution Ground PM_{2.5} Estimation Using the Light Gradient Boosting Machine. *Remote Sens.* **2023**, *15*, 4104. [CrossRef]
- Zheng, H.; Wu, D.; Wang, S.; Li, X.; Jin, L.N.; Zhao, B.; Li, S.; Sun, Y.; Dong, Z.; Wu, Q.; et al. Control of Toxicity of Fine Particulate Matter Emissions in China. *Nature* **2025**, *643*, 404–411. [CrossRef] [PubMed]
- 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.* **2022**, *270*, 112775. [CrossRef]
- Vazquez Santiago, J.; Hata, H.; Martinez-Noriega, E.J.; Inoue, K. Ozone Trends and Their Sensitivity in Global Megacities under the Warming Climate. *Nat. Commun.* **2024**, *15*, 10236. [CrossRef] [PubMed]
- Wang, P.; Chen, T.; Zhang, S.; Wang, F. High-Resolution Ground-Level O₃ Estimation in China (2019–2024) via Self-Attention Networks: Unveiling Spatiotemporal Dynamics and Key Drivers. *Atmos. Res.* **2026**, *330*, 108633. [CrossRef]
- Kasdagli, M.-I.; Orellano, P.; Pérez Velasco, R.; Samoli, E. Long-Term Exposure to Nitrogen Dioxide and Ozone and Mortality: Update of the WHO Air Quality Guidelines Systematic Review and Meta-Analysis. *Int. J. Public Health* **2024**, *69*, 1607676. [CrossRef] [PubMed]
- Wang, Y.; Yang, Y.; Yuan, Q.; Li, T.; Zhou, Y.; Zong, L.; Wang, M.; Xie, Z.; Ho, H.C.; Gao, M.; et al. Substantially Underestimated Global Health Risks of Current Ozone Pollution. *Nat. Commun.* **2025**, *16*, 102. [CrossRef] [PubMed]
- Feng, Z.; Xu, Y.; Kobayashi, K.; Dai, L.; Zhang, T.; Agathokleous, E.; Calatayud, V.; Paoletti, E.; Mukherjee, A.; Agrawal, M.; et al. Ozone Pollution Threatens the Production of Major Staple Crops in East Asia. *Nat. Food* **2022**, *3*, 47–56. [CrossRef] [PubMed]
- Liu, X.; Bhartia, P.K.; Chance, K.; Spurr, R.J.D.; Kurosu, T.P. Ozone Profile Retrievals from the Ozone Monitoring Instrument. *Atmos. Chem. Phys.* **2010**, *10*, 2521–2537. [CrossRef]
- Xu, J.; Zhang, Z.; Rao, L.; Wang, Y.; Letu, H.; Shi, C.; Tana, G.; Wang, W.; Zhu, S.; Liu, S.; et al. Remote Sensing of Tropospheric Ozone from Space: Progress and Challenges. *J. Remote Sens.* **2024**, *4*, 0178. [CrossRef]
- She, Y.; Li, J.; Lyu, X.; Guo, H.; Qin, M.; Xie, X.; Gong, K.; Ye, F.; Mao, J.; Huang, L.; et al. Current Status of Model Predictions of Volatile Organic Compounds and Impacts on Surface Ozone Predictions during Summer in China. *Atmos. Chem. Phys.* **2024**, *24*, 219–233. [CrossRef]
- Hong, Y.; Zhu, Y.; Huang, Y.; Liu, Y.; Xiong, C.; Fan, Q. The Impact of Sea Spray Aerosol on Photochemical Ozone Formation over Eastern China: Heterogeneous Reaction of Chlorine Particles and Radiative Effect. *Atmos. Chem. Phys.* **2025**, *25*, 11847–11866. [CrossRef]
- Wang, C.; Zhang, X.; Wang, J.; Pan, J.; Gao, J. Simulation of Ozone Pollution Events in Sichuan Basin under the Alternating Effects of Continental High Pressure and Subtropical High Pressure Circulation. *Atmos. Res.* **2025**, *322*, 108121. [CrossRef]
- Li, Y.; Li, C.; Li, Y.; Wang, T.; Li, M.; Qu, Y.; Wu, H.; Xie, M.; Wang, Y. Decadal Evolution of Aerosol-Mediated Ozone Responses in Eastern China under Clean Air Actions and Carbon Neutrality Policies. *Atmos. Chem. Phys.* **2026**, *26*, 1301–1319. [CrossRef]
- Li, H.; Yang, Y.; Jin, J.; Wang, H.; Li, K.; Wang, P.; Liao, H. Climate-Driven Deterioration of Future Ozone Pollution in Asia Predicted by Machine Learning with Multi-Source Data. *Atmos. Chem. Phys.* **2023**, *23*, 1131–1145. [CrossRef]

18. Mao, J.; Tai, A.P.K.; Yung, D.H.Y.; Yuan, T.; Chau, K.T.; Feng, Z. Multidecadal Ozone Trends in China and Implications for Human Health and Crop Yields: A Hybrid Approach Combining a Chemical Transport Model and Machine Learning. *Atmos. Chem. Phys.* **2024**, *24*, 345–366. [\[CrossRef\]](#)
19. Lawal, A.S.; Skipper, T.N.; Ivey, C.E.; Goldberg, D.L.; Kaiser, J.; Russell, A.G. Potential Errors in CMAQ NO:NO₂ Ratios and Upper Tropospheric NO₂ Impacting the Interpretation of TROPOMI Retrievals. *ACS EST Air* **2025**, *2*, 998–1008. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Ren, Y.; Oxford, C.R.; Zhang, D.; Liu, X.; Zhu, H.; Dillner, A.M.; White, W.H.; Chakrabarty, R.K.; Hasheminassab, S.; Diner, D.J.; et al. Black Carbon Emissions Generally Underestimated in the Global South as Revealed by Globally Distributed Measurements. *Nat. Commun.* **2025**, *16*, 7010. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Zhang, X.; Liu, C.; Liu, Y.; Yuan, X. An Analysis of the Impacts of Meteorological Factors on Ozone Concentration Using Generalized Additive Model in Tianjin, China. *J. Atmos. Sol.-Terr. Phys.* **2025**, *277*, 106669. [\[CrossRef\]](#)
22. Zhou, T.; Fang, S.; Jin, L.; Li, X.; Song, X.; Wang, Y.; Zhou, X.; Bai, Y.; Ma, X. Development of Land Use Regression Models to Characterise Spatial Patterns of Particulate Matter and Ozone in Urban Areas of Lanzhou. *Urban Clim.* **2024**, *55*, 101879. [\[CrossRef\]](#)
23. Shi, T.; Yuan, X.; Li, C.; Li, F. Synergistic and Trade-Off Influences of Combined PM_{2.5}-O₃ Pollution in the Shenyang Metropolitan Area, China: A Comparative Land Use Regression Analysis. *Sustainability* **2025**, *17*, 8046. [\[CrossRef\]](#)
24. Cheng, S.; Zhang, G.; Yang, X.; Lei, B. A Multiscale Geographically Weighted Regression Kriging Method for Spatial Downscaling of Satellite-Based Ozone Datasets. *Front. Environ. Sci.* **2024**, *11*, 1267752. [\[CrossRef\]](#)
25. Wu, Y.; Tao, T.; Wang, K.; Shi, D.; Zhang, X.; Wang, Q. What Drives the Spatiotemporal Characteristics and Evolution of Near-Surface Ozone Across Multiple Scales? Implications for Sustainable Air Quality Management in Coastal Southeast China. *Sustainability* **2026**, *18*, 6842. [\[CrossRef\]](#)
26. Zhang, X.Y.; Zhao, L.M.; Cheng, M.M.; Chen, D.M. Estimating Ground-Level Ozone Concentrations in Eastern China Using Satellite-Based Precursors. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4754–4763. [\[CrossRef\]](#)
27. Wang, X.; Kang, Y.; Duan, W.; Xv, R.; Liu, X.; Wei, W.; Cheng, S. Differential Response of Various Pollutant-Meteorology Factors on O₃ in Key Regions of China: Based on Multiple Methods and Datasets. *Atmos. Pollut. Res.* **2024**, *15*, 102086. [\[CrossRef\]](#)
28. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Tong, L.; Gu, Z.; Zhu, X.; Huang, C.; Hu, B.; Shi, Y.; Meng, Y.; Zheng, J.; He, M.; He, J.; et al. Coastal Ozone Dynamics and Formation Regime in Eastern China: Integrating Trend Decomposition and Machine Learning Techniques. *J. Environ. Sci.* **2025**, *155*, 597–612. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Du, C.; Pei, J.; Feng, Z. Unraveling the Complex Interactions between Ozone Pollution and Agricultural Productivity in China's Main Winter Wheat Region Using an Interpretable Machine Learning Framework. *Sci. Total Environ.* **2024**, *954*, 176293. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Liu, Z.; Lu, Z.; Zhu, W.; Yuan, J.; Cao, Z.; Cao, T.; Liu, S.; Xu, Y.; Zhang, X. Comparison of Machine Learning Methods for Predicting Ground-Level Ozone Pollution in Beijing. *Front. Environ. Sci.* **2025**, *13*, 1561794. [\[CrossRef\]](#)
32. 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. [\[CrossRef\]](#)
33. Wei, J.; Li, Z.; Pinker, R.T.; Wang, J.; Sun, L.; Xue, W.; Li, R.; Cribb, M. Himawari-8-Derived Diurnal Variations in Ground-Level PM_{2.5} Pollution across China Using the Fast Space-Time Light Gradient Boosting Machine (LightGBM). *Atmos. Chem. Phys.* **2021**, *21*, 7863–7880. [\[CrossRef\]](#)
34. Mu, X.; Wang, S.; Jiang, P.; Wang, B.; Wu, Y.; Zhu, L. Full-Coverage Spatiotemporal Estimation of Surface Ozone over China Based on a High-Efficiency Deep Learning Model. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103284. [\[CrossRef\]](#)
35. Li, Q.; Li, J.; Wang, Z.; Liu, B.; Wang, W.; Wang, Z. Development of a City-Level Surface Ozone Forecasting System Using Deep Learning Techniques and Air Quality Model: Application in Eastern China. *Atmos. Environ.* **2024**, *339*, 120865. [\[CrossRef\]](#)
36. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems 30*; Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R., Eds.; Curran Associates, Inc.: Long Beach, CA, USA, 2017; pp. 3146–3154.
37. Jiang, Z.; Li, J.; Liu, G.; Zhang, C. Impact of the Indian Ocean Dipole Mode on Planetary Boundary Layer Ozone in China. *Geophys. Res. Lett.* **2024**, *51*, e2024GL110108. [\[CrossRef\]](#)
38. Feng, Z.; De Marco, A.; Anav, A.; Gualtieri, M.; Sicard, P.; Tian, H.; Fornasier, F.; Tao, F.; Guo, A.; Paoletti, E. Economic Losses Due to Ozone Impacts on Human Health, Forest Productivity and Crop Yield across China. *Environ. Int.* **2019**, *131*, 104966. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Li, S.; Gao, Y.; Zhang, J.; Hong, C.; Zhang, S.; Chen, D.; Wild, O.; Feng, Z.; Xu, Y.; Guo, X.; et al. Mitigating Climate Change and Ozone Pollution Will Improve Chinese Food Security. *One Earth* **2025**, *8*, 101166. [\[CrossRef\]](#)

40. Lim, C.C.; Hayes, R.B.; Ahn, J.; Shao, Y.; Silverman, D.T.; Jones, R.R.; Garcia, C.; Bell, M.L.; Thurston, G.D. Long-Term Exposure to Ozone and Cause-Specific Mortality Risk in the United States. *Am. J. Respir. Crit. Care Med.* **2019**, *200*, 1022–1031. [[CrossRef](#)] [[PubMed](#)]
41. Wang, F.; Wang, T.; Xia, H.; Song, H.; Zhou, S.; Zhang, T. Estimation of Surface Ozone Effects on Winter Wheat Yield across the North China Plain. *Agronomy* **2024**, *14*, 2326. [[CrossRef](#)]
42. Lee, Y.C.; Shindell, D.T.; Faluvegi, G.; Wenig, M.; Lam, Y.F.; Ning, Z.; Hao, S.; Lai, C.S. Increase of Ozone Concentrations, Its Temperature Sensitivity and the Precursor Factor in South China. *Tellus B Chem. Phys. Meteorol.* **2014**, *66*, 23455. [[CrossRef](#)]
43. Li, K.; Jacob, D.J.; Shen, L.; Lu, X.; De Smedt, I.; Liao, H. Increases in Surface Ozone Pollution in China from 2013 to 2019: Anthropogenic and Meteorological Influences. *Atmos. Chem. Phys.* **2020**, *20*, 11423–11433. [[CrossRef](#)]
44. Meleux, F.; Solmon, F.; Giorgi, F. Increase in Summer European Ozone Amounts Due to Climate Change. *Atmos. Environ.* **2007**, *41*, 7577–7587. [[CrossRef](#)]
45. Shen, L.; Mickley, L.J.; Tai, A.P.K. Influence of Synoptic Patterns on Surface Ozone Variability over the Eastern United States from 1980 to 2012. *Atmos. Chem. Phys.* **2015**, *15*, 10925–10938. [[CrossRef](#)]
46. He, J.; Gong, S.; Yu, Y.; Yu, L.; Wu, L.; Mao, H.; Song, C.; Zhao, S.; Liu, H.; Li, X.; et al. Air Pollution Characteristics and Their Relation to Meteorological Conditions during 2014–2015 in Major Chinese Cities. *Environ. Pollut.* **2017**, *223*, 484–496. [[CrossRef](#)] [[PubMed](#)]
47. Li, M.; Yu, S.; Chen, X.; Li, Z.; Zhang, Y.; Wang, L.; Liu, W.; Li, P.; Lichtfouse, E.; Rosenfeld, D.; et al. Large Scale Control of Surface Ozone by Relative Humidity Observed during Warm Seasons in China. *Environ. Chem. Lett.* **2021**, *19*, 3981–3989. [[CrossRef](#)]
48. 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](#)]
49. Liu, Y.; Tang, G. Contradictory Response of Ozone and Particulate Matter Concentrations to Boundary Layer Meteorology. *Environ. Pollut.* **2024**, *343*, 123209. [[CrossRef](#)]
50. Lei, Y.; Wu, K.; Zhang, X.; Kang, P.; Du, Y.; Yang, F.; Fan, J.; Hou, J. Role of Meteorology-Driven Regional Transport on O₃ Pollution over the Chengdu Plain, Southwestern China. *Atmos. Res.* **2023**, *285*, 106619. [[CrossRef](#)]
51. Hu, F.; Xie, P.; Xu, J.; Li, A.; Lv, Y.; Zhang, Z.; Zheng, J.; Tian, X. Impacts of Synoptic Weather Patterns on Hefei's Ozone in Warm Season and Analysis of Transport Pathways during Extreme Pollution Events. *J. Environ. Sci.* **2025**, *156*, 371–384. [[CrossRef](#)] [[PubMed](#)]
52. Zhang, Y.; Liu, J.; Croft, H.; Staebler, R.; Wang, M.S.; Luo, X.; He, L.; Gonsamo, A.; Chen, J.M. Rainfall-Induced Changes in Vertical O₃ and SO₂ within and above a Boreal-Temperate Forest. *Agric. For. Meteorol.* **2025**, *373*, 110748. [[CrossRef](#)]
53. Zhang, L.; Zhu, Y.; Cui, C.; Meng, H.; Zhong, X.; Li, Y.; Huang, L.; Shen, H.; Xue, L. Nucleation-Mode Particles Burst during Convective Rainfall Linked to Ozone Intrusions at a Coastal Site in Eastern China. *Atmos. Res.* **2026**, *342*, 109150. [[CrossRef](#)]
54. Ahmad, S.P.; Levelt, P.F.; Bhartia, P.K.; Hilsenrath, E.; Leppelmeier, G.W.; Johnson, J.E. *Atmospheric Products from the Ozone Monitoring Instrument (OMI)*; Barnes, W.L., Ed.; SPEI: San Diego, CA, USA, 2003; p. 619.
55. Lamsal, L.N.; Krotkov, N.A.; Vasilkov, A.; Marchenko, S.; Qin, W.; Yang, E.-S.; Fasnacht, Z.; Joiner, J.; Choi, S.; Haffner, D.; et al. Ozone Monitoring Instrument (OMI) Aura Nitrogen Dioxide Standard Product Version 4.0 with Improved Surface and Cloud Treatments. *Atmos. Meas. Tech.* **2021**, *14*, 455–479. [[CrossRef](#)]
56. Celarier, E.A.; Brinksma, E.J.; Gleason, J.F.; Veefkind, J.P.; Cede, A.; Herman, J.R.; Ionov, D.; Goutail, F.; Pommereau, J.-P.; Lambert, J.-C.; et al. Validation of Ozone Monitoring Instrument Nitrogen Dioxide Columns. *J. Geophys. Res. Atmos.* **2008**, *113*, 2007JD008908. [[CrossRef](#)]
57. Pawan, K.B. *OMI/Aura Ozone (O₃) Total Column Daily L2 Global Gridded 0.25 Degree x 0.25 Degree V3*; NASA Goddard Earth Sciences Data and Information Services Center (GES DISC): Greenbelt, MD, USA, 2012.
58. Veefkind, J.P.; De Haan, J.F.; Brinksma, E.J.; Kroon, M.; Levelt, P.F. Total Ozone from the Ozone Monitoring Instrument (OMI) Using the DOAS Technique. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1239–1244. [[CrossRef](#)]
59. Kroon, M.; Veefkind, J.P.; Sneep, M.; McPeters, R.D.; Bhartia, P.K.; Levelt, P.F. Comparing OMI-TOMS and OMI-DOAS Total Ozone Column Data. *J. Geophys. Res. Atmos.* **2008**, *113*, 2007JD008798. [[CrossRef](#)]
60. Wang, Y.; Wild, O.; Chen, X.; Wu, Q.; Gao, M.; Chen, H.; Qi, Y.; Wang, Z. Health Impacts of Long-Term Ozone Exposure in China over 2013–2017. *Environ. Int.* **2020**, *144*, 106030. [[CrossRef](#)] [[PubMed](#)]
61. Guo, B.; Wu, H.; Pei, L.; Zhu, X.; Zhang, D.; Wang, Y.; Luo, P. Study on the Spatiotemporal Dynamic of Ground-Level Ozone Concentrations on Multiple Scales across China during the Blue Sky Protection Campaign. *Environ. Int.* **2022**, *170*, 107606. [[CrossRef](#)] [[PubMed](#)]
62. Yi, S.; Sneeuw, N. Filling the Data Gaps Within GRACE Missions Using Singular Spectrum Analysis. *J. Geophys. Res. Solid Earth* **2021**, *126*, e2020JB021227. [[CrossRef](#)]
63. Kondrashov, D.; Ghil, M. Spatio-Temporal Filling of Missing Points in Geophysical Data Sets. *Nonlinear Process. Geophys.* **2006**, *13*, 151–159. [[CrossRef](#)]

64. Golyandina, N.; Osipov, E. The “Caterpillar”-SSA Method for Analysis of Time Series with Missing Values. *J. Stat. Plan. Inference* **2007**, *137*, 2642–2653. [\[CrossRef\]](#)
65. Baez-Villanueva, O.M.; Zambrano-Bigiarini, M.; Beck, H.E.; McNamara, I.; Ribbe, L.; Nauditt, A.; Birkel, C.; Ver-bist, K.; Giraldo-Osorio, J.D.; Xuan Thinh, N. RF-MEP: A Novel Random Forest Method for Merging Gridded Precipitation Products and Ground-Based Measurements. *Remote Sens. Environ.* **2020**, *239*, 111606. [\[CrossRef\]](#)
66. Behrens, T.; Schmidt, K.; Viscarra Rossel, R.A.; Gries, P.; Scholten, T.; MacMillan, R.A. Spatial Modelling with Euclidean Distance Fields and Machine Learning. *Eur. J. Soil Sci.* **2018**, *69*, 757–770. [\[CrossRef\]](#)
67. Wei, J.; Li, Z.; Lyapustin, A.; Sun, L.; Peng, Y.; Xue, W.; Su, T.; Cribb, M. Reconstructing 1 Km Resolution High Quality PM_{2.5} Data Records from 2000 to 2018 in China: Spatiotemporal Variations and Policy Implications. *Remote Sens. Environ.* **2021**, *252*, 112136. [\[CrossRef\]](#)
68. Cheng, F.; Li, Z.; Yang, Z.; Li, R.; Wang, D.; Jia, A.; Li, K.; Zhao, B.; Wang, S.; Yin, D.; et al. First Retrieval of 24 Hourly 1 Km Resolution Gapless Surface Ozone (O₃) from Space in China Using Artificial Intelligence: Diurnal Variations and Implications for Air Quality and Phytotoxicity. *Remote Sens. Environ.* **2025**, *316*, 114482. [\[CrossRef\]](#)
69. Jiang, Y.; Shi, S.; Li, X.; Xu, C.; Kan, H.; Hu, B.; Meng, X. A 10 km Daily Level Ultraviolet Radiation Predicting Dataset Based on Machine Learning Models in China from 2005 to 2020. *Earth Syst. Sci. Data* **2024**, *16*, 4655–4672. [\[CrossRef\]](#)
70. Rodríguez-Pérez, R.; Bajorath, J. Interpretation of Machine Learning Models Using Shapley Values: Application to Compound Potency and Multi Target Activity Predictions. *J. Comput. Aided Mol. Des.* **2020**, *34*, 1013–1026. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Temenos, A.; Temenos, N.; Kaselimi, M.; Doulamis, A.; Doulamis, N. Interpretable Deep Learning Framework for Land Use and Land Cover Classification in Remote Sensing Using SHAP. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [\[CrossRef\]](#)
72. Wei, J.; Wang, Z.; Li, Z.; Li, Z.; Pang, S.; Xi, X.; Cribb, M.; Sun, L. Global Aerosol Retrieval over Land from Land-sat Imagery Integrating Transformer and Google Earth Engine. *Remote Sens. Environ.* **2024**, *315*, 114404. [\[CrossRef\]](#)
73. Wang, X.; Mauzerall, D.L. Characterizing Distributions of Surface Ozone and Its Impact on Winter wheat production in China, Japan and South Korea: 1990 and 2020. *Atmos. Environ.* **2004**, *38*, 4383–4402. [\[CrossRef\]](#)
74. Avnery, S.; Mauzerall, D.L.; Liu, J.; Horowitz, L.W. Global Crop Yield Reductions Due to Surface Ozone Exposure: 1. Year 2000 Crop Production Losses and Economic Damage. *Atmos. Environ.* **2011**, *45*, 2284–2296. [\[CrossRef\]](#)
75. Zhu, Z.; Sun, X.; Zhao, F.; Meixner, F.X. Ozone Concentrations, Flux and Potential Effect on Yield during Wheat Growth in the Northwest-Shandong Plain of China. *J. Environ. Sci.* **2015**, *34*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Lefohn, A.S.; Malley, C.S.; Smith, L.; Wells, B.; Hazucha, M.; Simon, H.; Naik, V.; Mills, G.; Schultz, M.G.; Paoletti, E.; et al. Tropospheric Ozone Assessment Report: Global Ozone Metrics for Climate Change, Human Health, and Crop/Ecosystem Research. *Elem. Sci. Anthr.* **2018**, *6*, 27. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Lin, Y.; Jiang, F.; Zhao, J.; Zhu, G.; He, X.; Ma, X.; Li, S.; Sabel, C.E.; Wang, H. Impacts of O₃ on Premature Mortality and Crop Yield Loss across China. *Atmos. Environ.* **2018**, *194*, 41–47. [\[CrossRef\]](#)
78. Lu, X.; Zhang, L.; Wang, X.; Gao, M.; Li, K.; Zhang, Y.; Yue, X.; Zhang, Y. Rapid Increases in Warm-Season Surface Ozone and Resulting Health Impact in China Since 2013. *Environ. Sci. Technol. Lett.* **2020**, *7*, 240–247. [\[CrossRef\]](#)
79. Li, D.; Shindell, D.; Ding, D.; Lu, X.; Zhang, L.; Zhang, Y. Surface Ozone Impacts on Major Crop Production in China from 2010 to 2017. *Atmos. Chem. Phys.* **2022**, *22*, 2625–2638. [\[CrossRef\]](#)
80. Sicard, P.; De Marco, A.; Dalstein-Richier, L.; Tagliaferro, F.; Renou, C.; Paoletti, E. An Epidemiological Assessment of Stomatal Ozone Flux-Based Critical Levels for Visible Ozone Injury in Southern European Forests. *Sci. Total Environ.* **2016**, *541*, 729–741. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Heck, W.W.; Cure, W.W.; Rawlings, J.O.; Zaragoza, L.J.; Heagle, A.S.; Heggstad, H.E.; Kohut, R.J.; Kress, L.W.; Temple, P.J. Assessing Impacts of Ozone on Agricultural Crops: II. Crop Yield Functions and Alternative Exposure Statistics. *J. Air Pollut. Control Assoc.* **1984**, *34*, 810–817. [\[CrossRef\]](#)
82. Sinha, B.; Singh Sangwan, K.; Maurya, Y.; Kumar, V.; Sarkar, C.; Chandra, B.P.; Sinha, V. Assessment of Crop Yield Losses in Punjab and Haryana Using 2 Years of Continuous in Situ Ozone Measurements. *Atmos. Chem. Phys.* **2015**, *15*, 9555–9576. [\[CrossRef\]](#)
83. 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* **2019**, *116*, 422–427. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Cao, T.; Wang, H.; Chen, X.; Li, L.; Lu, X.; Lu, K.; Fan, S. Rapid Increase in Spring Ozone in the Pearl River Delta, China during 2013–2022. *npj Clim. Atmos. Sci.* **2024**, *7*, 309. [\[CrossRef\]](#)
85. Wang, H.; Huang, C.; Tao, W.; Gao, Y.; Wang, S.; Jing, S.; Wang, W.; Yan, R.; Wang, Q.; An, J.; et al. Seasonality and Reduced Nitric Oxide Titration Dominated Ozone Increase during COVID-19 Lockdown in Eastern China. *npj Clim. Atmos. Sci.* **2022**, *5*, 24. [\[CrossRef\]](#)
86. Zhang, K.; Liu, Z.; Zhang, X.; Li, Q.; Jensen, A.; Tan, W.; Huang, L.; Wang, Y.; De Gouw, J.; Li, L. Insights into the Significant Increase in Ozone during COVID-19 in a Typical Urban City of China. *Atmos. Chem. Phys.* **2022**, *22*, 4853–4866. [\[CrossRef\]](#)

87. Chen, Z.; Liu, J.; Cheng, X.; Yang, M.; Shu, L. Stratospheric Influences on Surface Ozone Increase during the COVID-19 Lockdown over Northern China. *npj Clim. Atmos. Sci.* **2023**, *6*, 76. [\[CrossRef\]](#)
88. Feng, Z.; Kobayashi, K.; Ainsworth, E.A. Impact of Elevated Ozone Concentration on Growth, Physiology, and Yield of Wheat (*Triticum Aestivum* L.): A Meta-analysis. *Glob. Change Biol.* **2008**, *14*, 2696–2708. [\[CrossRef\]](#)
89. Zhu, X.; Feng, Z.; Sun, T.; Liu, X.; Tang, H.; Zhu, J.; Guo, W.; Kobayashi, K. Effects of Elevated Ozone Concentration on Yield of Four Chinese Cultivars of Winter Wheat under Fully Open-Air Field Conditions: Wheat Yield Loss to Ozone in a Chinese Field. *Glob. Change Biol.* **2011**, *17*, 2697–2706. [\[CrossRef\]](#)
90. Xu, Y.; Tang, J.; Xia, J.; Ma, Y.; Shang, B.; Liu, B.; Kobayashi, K.; Feng, Z.; Agathokleous, E. Joint Ozone Pollution and Climate Warming Reduce Yield but Enhance Grain Protein Content in a Resistant Wheat Variety. *Glob. Change Biol.* **2025**, *31*, e70351. [\[CrossRef\]](#) [\[PubMed\]](#)
91. Ma, Y.; Xu, Y.; Hu, X.; Agathokleous, E.; Kobayashi, K.; Yin, X.; Cao, R.; Feng, Z. Response of Source Capacity to Elevated Ozone during Grain Filling Determines Ozone Sensitivity of Wheat Cultivars. *Field Crops Res.* **2026**, *338*, 110296. [\[CrossRef\]](#)
92. Chen, G.; Chen, J.; Dong, G.; Yang, B.; Liu, Y.; Lu, T.; Yu, P.; Guo, Y.; Li, S. Improving Satellite-Based Estimation of Surface Ozone across China during 2008–2019 Using Iterative Random Forest Model and High-Resolution Grid Meteorological Data. *Sustain. Cities Soc.* **2021**, *69*, 102807. [\[CrossRef\]](#)
93. Liu, H.; Liu, S.; Xue, B.; Lv, Z.; Meng, Z.; Yang, X.; Xue, T.; Yu, Q.; He, K. Ground-Level Ozone Pollution and Its Health Impacts in China. *Atmos. Environ.* **2018**, *173*, 223–230. [\[CrossRef\]](#)
94. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Xue, T.; Zheng, Y.; Geng, G.; Xiao, Q.; Meng, X.; Wang, M.; Li, X.; Wu, N.; Zhang, Q.; Zhu, T. Estimating Spatiotemporal Variation in Ambient Ozone Exposure during 2013–2017 Using a Data-Fusion Model. *Environ. Sci. Technol.* **2020**, *54*, 14877–14888. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Zhang, X.; Yan, B.; Zhou, Y.; Osei, F.; Li, Y.; Zhao, H.; Cheng, C.; Stein, A. Short-Term Health Impacts Related to Ozone in China before and after Implementation of Policy Measures: A Systematic Review and Meta-Analysis. *Sci. Total Environ.* **2022**, *847*, 157588. [\[CrossRef\]](#) [\[PubMed\]](#)
97. Wang, X.; Zhang, Q.; Zheng, F.; Zheng, Q.; Yao, F.; Chen, Z.; Zhang, W.; Hou, P.; Feng, Z.; Song, W.; et al. Effects of Elevated O₃ Concentration on Winter Wheat and Rice Yields in the Yangtze River Delta, China. *Environ. Pollut.* **2012**, *171*, 118–125. [\[CrossRef\]](#) [\[PubMed\]](#)
98. Yi, F.; Feng, J.; Wang, Y.; Jiang, F. Influence of Surface Ozone on Crop Yield of Maize in China. *J. Integr. Agric.* **2020**, *19*, 578–589. [\[CrossRef\]](#)
99. Chuwah, C.; Van Noije, T.; Van Vuuren, D.P.; Stehfest, E.; Hazeleger, W. Global Impacts of Surface Ozone Changes on Crop Yields and Land Use. *Atmos. Environ.* **2015**, *106*, 11–23. [\[CrossRef\]](#)
100. Bodor, K.; Bodor, Z. The Impact of Surface Ozone on Agricultural Yields in the Ciuc Basin. *Sci. Rep.* **2026**, *16*, 9434. [\[CrossRef\]](#) [\[PubMed\]](#)
101. Ashmore, M.R. Assessing the Future Global Impacts of Ozone on Vegetation. *Plant Cell Environ.* **2005**, *28*, 949–964. [\[CrossRef\]](#)
102. Mills, G.; Hayes, F.; Simpson, D.; Emberson, L.; Norris, D.; Harmens, H.; Büker, P. Evidence of Widespread Effects of Ozone on Crops and (Semi-)Natural Vegetation in Europe (1990–2006) in Relation to AOT40- and Flux-Based Risk Maps: Ozone Effects on Vegetation in Europe. *Glob. Change Biol.* **2011**, *17*, 592–613. [\[CrossRef\]](#)
103. Hollaway, M.J.; Arnold, S.R.; Challinor, A.J.; Emberson, L.D. Intercontinental Trans-Boundary Contributions to Ozone-Induced Crop Yield Losses in the Northern Hemisphere. *Biogeosciences* **2012**, *9*, 271–292. [\[CrossRef\]](#)
104. Broberg, M.C.; Hayes, F.; Harmens, H.; Uddling, J.; Mills, G.; Pleijel, H. Effects of Ozone, Drought and Heat Stress on Wheat Yield and Grain Quality. *Agric. Ecosyst. Environ.* **2023**, *352*, 108505. [\[CrossRef\]](#)

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.