Environmental Pollution 357 (2024) 124397

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Machine-learning-based corrections of CMIP6 historical surface ozone in China during 1950–2014☆

Yuanxi Tong [a], Yingying Yan [a,*], Jintai Lin [b], Shaofei Kong [a,c], Zhixuan Tong [a], Yifei Zhu [a], Yukun Yan [a], Zhan Sun [a]

[a] Department of Atmospheric Science, School of Environmental Sciences, China University of Geosciences, Wuhan, 430074, China
[b] Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, 100871, China
[c] Research Centre for Complex Air Pollution of Hubei Province, Wuhan, 430074, China

ABSTRACT

Due to a lack of long-term observations in China, reports on historical ozone concentration are severely limited. In this study, by combining observation, reanalysis and model simulation data, XGBoost machine learning algorithm is used to correct the surface ozone concentration from CMIP6 climate model, and the long-term and large-scale surface ozone concentration of China during 1950–2014 is obtained. The long-term evolutions and trends of ozone and meteorological effects on interannual ozone variations are further analyzed. The results reveal that CMIP6 historical simulations have a large underestimation in ozone concentrations and their trends. The XGB-derived ozone are closer to observations, with R^2 value of 0.66 and 0.74 for daily and monthly retrievals, respectively. Both the concentrations and exceedances of ozone in most parts of China have shown increasing trends from 1950 to 2014. The daily mean ozone concentration without climate change effects is estimated to be 117 ppb in the year 1950 averaged over China. It indicates that the increase in anthropogenic emissions of China has a significant contribution to ozone enhancement between 1950 and 2014. The higher ozone growth rates of XGB retrievals than those from the model indicate a regional surface ozone penalty due to the warming climate. The relatively significant increment in ozone are estimated in the Central and Western China. Seasonally, the ozone enhancement is largest in spring, indicating a shift in seasonal variation of ozone. Given the uncertainty in simulating historical ozone by climate model, we show that machine learning approaches can provide improved assessment of evolution in surface ozone, along with valuable information to guide future model development and formulate future ozone pollution prevention and control policies.

1. Introduction

Tropospheric ozone (O_3) poses significant health risks (Anenberg et al., 2010; Fu and Tai, 2015; Qiu et al., 2020), and also adverse effects on agricultural production (Avnery et al., 2011; Li et al., 2022a,; Liu et al., 2024) and vegetation and ecosystems (Xu et al., 2020; Mills et al., 2018; Musselman et al., 2006), from long-term exposure to high concentrations. Additionally, ozone contributes to positive radiative forcing and plays a significant role in global warming (Stevenson et al., 2013; IPCC, 2021). The generation of ozone is not only dominated by precursors of volatile organic compounds (VOCs) and nitrogen oxides (NO_x) (Yan et al., 2021; Wang et al., 2017), but also significantly affected by specific meteorological factors such as high temperatures, strong radiation, and stable atmospheric conditions (Chen et al., 2020; Elminir, 2005; Yan et al., 2018; Zhao et al., 2020). The meteorological influence becomes more pronounced in the context of decreasing anthropogenic emissions but increasing extreme weather events due to climate warming (Meehl et al., 2018), posing challenges to ozone control and prevention.

Ozone pollution in China has become increasingly serious with a growth rate of 2.0 ppb yr^{-1} for annual averages in recent years, and the burden of disease caused by ozone exposure has surpassed that of $PM_{2.5}$ (Wang et al., 2022; ,Zhang et al., 2023b). Some previous studies analyzed the changes in ozone concentration based on observations or model simulations. Chen et al. (2019) analyzed the observations from 12 urban stations of Beijing and revealed an enhancement of 0.11 ppb yr^{-1} for the annual ozone concentrations. Gao et al. (2017) also showed an increase of 67% (1.1 ppb yr^{-1}) for observed ozone at an urban site of Shanghai. ,Li et al., 2022b analyzed observation data from the Pearl River Delta (PRD) region of China during 2006–2019 and showed an increase of ozone by up to 1.02 ppb yr^{-1}. Dang et al. (2021) analyzed the maximum daily 8-h average (MDA8) ozone during 2013–2017 in North China Plain (NCP) and Yangtze River Delta (YRD) using GEOS-Chem. They found a 0.58/1.74 ppb yr^{-1} ozone concentration trend in the

☆ This paper has been recommended for acceptance by Pavlos Kassomenos.

* Corresponding author.
E-mail address: yanyingying@cug.edu.cn (Y. Yan).

https://doi.org/10.1016/j.envpol.2024.124397
Received 25 April 2024; Received in revised form 9 June 2024; Accepted 18 June 2024
Available online 19 June 2024

NCP/YRD, with 49%/84% from meteorological contributions. Through an extensive literature review, Sicard (2021) has concluded that ozone concentrations increased in most regions of East Asia with an enhancement of 0.21 ppb yr^{-1} at rural stations over the period 2000–2010 and 0.68 ppb yr^{-1} in cities between 2015 and 2014. Sicard et al. (2021) reported that the WRF-Chem model reproduced well the spatiotemporal of surface ozone concentrations across China. Tang et al. (2022) further combined the observations and WRF-Chem simulations to reveal a worsening trend (4.9 μg m^{-3} yr^{-1}) of ozone pollution in YRD during 2014–2017. However, due to the lack of long-term observations, and the large computational resources required for atmospheric chemistry models, it is difficult to achieve long-term large-scale surface ozone concentrations. Thus, the time scale of ozone variations in studies is mostly within recent 20 years, which is not enough to identify the long-term spatiotemporal evolution of ozone under the background of climate change.

A few models in CMIP6 (Climate Model Intercomparison Project Phase 6) have provided long-term simulation data of surface ozone globally since 1950 (Griffiths et al., 2021; Wang et al., 2021; Song et al., 2022). The CMIP6 is based on long transient simulations, but adds a new, more complete emission dataset, as well as the most up-to-date and complete or complex interactive models. The CMIP6 "Historical" simulations run from preindustrial times to the year 2014 (Eyring et al., 2016; Griffiths et al., 2021). However, due to oversimplified parameterizations, and inaccurate simulated meteorological field, the CMIP6 models show large biases in simulating historical surface ozone (Griffiths et al., 2021; Liu et al., 2022). Griffiths et al. (2021) compared the CMIP6 model ensemble to five remote surface ozone stations with the longest available in situ sampling record (1957–present). They showed a significant error of surface ozone concentration. In China, surface ozone observations after 2000s have been seriously underestimated by the models of CMIP6 (Wang et al., 2021, 2022). These errors may propagate within complex Earth system models, and impeding a comprehensive and accurate understanding of Earth system interactions. Overall, further improvement and validation are needed for the surface ozone data from CMIP6.

Thus, it is urgent to efficiently build continuous long-term high-precision surface ozone datasets to analyze its spatiotemporal change. There are six main approaches for inverting historical ozone data, namely statistical models, geostatistical models, machine learning algorithms, deterministic approach, chemistry-climate models, chemical reanalyses and ensemble approaches (Marco et al., 2022). Recently, artificial intelligence-based regression and clustering algorithms are gradually becoming popular methods with outstanding performance to predict future ozone changes and invert historical ozone levels (Zhang et al., 2023a,; Du et al., 2022; Zhang et al., 2022; Mao et al., 2022; Ma et al., 2021; Vu et al., 2019). The main algorithms include linear regression models, deep learning-based neural network regression models (Jia et al., 2020), support vector machines (SVM), and tree models. Linear regression models have simple principles and short computation time but are prone to overfitting (Trivedi et al., 2021). Neural network models can handle complex tasks well, but they require a long training time when faced with large training data and have high requirements for data preprocessing (Alzubaidi et al., 2021). SVM regression models can achieve more accurate multi-task classification but are difficult to handle large-scale training samples and have high computational costs (Chauhan et al., 2018). Decision tree models are simple and easy-to-use non-parametric models that are suitable for large-scale predictions, with fast computation speed, interpretable results, strong robustness, and low data requirements (Myles et al., 2004). XGBoost is an efficient machine learning algorithm first proposed by Chen et al., in 2014. Liu et al. (2020) established a nationwide MDA8 O_3 prediction model based on the XGBoost (eXtreme Gradient Boosting) algorithm, combined with the ozone observations, meteorological parameters, and land use data, achieving high prediction accuracy. Tsai (2018) applied XGBoost and used emission data, and meteorological monitoring data to construct features and to rolling forecast the hourly average ozone concentrations at four automatic monitoring stations of Xiamen City. It achieved a 90% forecast accuracy rate.

In this study, we attempt to use the XGBoost algorithm to correct the historical surface ozone concentrations of China simulated by CMIP6. Meteorological reanalysis data and CMIP6 surface ozone simulations are used as precursor factors for training and inversion to obtain improved surface ozone concentrations of China from 1950 to 2014. Based on the obtained data, we will further analyze the long-term spatiotemporal evolution of surface ozone, the meteorological effects on the interannual ozone variation, and the spatiotemporal variation of ozone exceedance events. This study can provide scientific basis for ozone prevention and control and long-term stability in the context of climate change.

2. Data and methods

2.1. Data

The meteorological variables (Table S1) from 1950 to 2022 are obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF; https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5) Reanalysis version 5 (ERA5) dataset. The dataset provides hourly estimates with a spatial resolution of 0.25° × 0.25°, and daily data are calculated by averaging the 24-hourly values. These meteorological elements were significantly correlated with the surface ozone and were introduced into machine learning models to predict ozone concentration (Ma et al., 2021; Luo et al., 2022; Cheng et al., 2023; Deng et al., 2021). For simplicity, the variable names used as input to the XGBoost model are abbreviated as Table S1.

The observed daily 8-h rolling average ozone concentration data (OBS) from 2014 to 2022 are obtained from the National Urban Air Quality Real-time Release Platform of the China Environmental Monitoring Station (https://air.cnemc.cn:18007/) and archived at https://quotsoft.net/air (Wang, 2020; Tong et al., 2024). The network was established in 2013 as part of the Clean Air Action Plan. In 2013, there were 946 monitoring sites, which increased to 2024 sites by 2022. The start date of the dataset is May 13, 2014. The original unit of ozone concentration is μg m^{-3}, which is converted to ppb.

The historical hourly surface ozone simulations during 1950–2014 are conducted by the MPI-ESM1.2-HAM (MEH) climate model (Neubauer et al., 2019; https://www.wdc-climate.de/ui/cmip6?input=CMIP6.CMIP.HAMMOZ-Consortium.MPI-ESM-1-2-HAM; 2.5° × 2.5°), which is released in 2017. And the data are downloaded from the CMIP6 (https://esgf-node.llnl.gov/search/cmip6/). The daily 8-h rolling average is calculated to correspond with the ozone observation data. The original unit of ozone concentration is mol/mol, which is also converted to ppb.

In the quality control process, we use linear interpolation of near time or space points to deal with missing values and outliers. For this study, we hypothesize that air quality monitoring station can represent the air quality in the immediate area; current emission scenarios can be considered to be approximately equivalent to those since industrial and urban development; and the boundary layer height has been reflected in the temperature, pressure and other meteorological elements.

2.2. Methods

2.2.1. Site location extraction of grid data

We use the spherical cosine formula to calculate the nearest grid point to a site:

$$d = R \cdot \arccos\left(\sin lat_s \cdot \sin lat_g + \cos lat_s \cdot \cos lat_g \cdot \cos\left(lon_g - lon_s\right)\right) \quad (1)$$

Where R is the radius of the Earth, take 6371 km. Lat_s, lat_g, lon_s, and lon_g are respectively station latitude, grid latitude, station longitude, and grid longitude (all in radians). The extracted ERA5 and MEH data of these grid points are accordingly used as the meteorological data and the simulated ozone concentration of stations.

2.2.2. XGBoost algorithm

XGBoost is an implementation of gradient boosting methods, which make predictions by ensembling multiple decision trees (Chen et al., 2014; Liu et al., 2020). Compared to traditional gradient boosting methods, XGBoost has made improvements and optimizations in several aspects. Firstly, it introduces regularization terms to prevent overfitting by controlling the model complexity. Secondly, XGBoost uses an approximate splitting algorithm to accelerate the training process. Moreover, XGBoost adopts parallel computing strategies, utilizing multi-threading or distributed computing resources to speed up the model training process. XGBoost algorithm has been widely used in the prediction and inversion of air quality (Calatayud et al., 2023; Liu et al., 2020; Tsai, 2018).

The core idea of the XGBoost algorithm is to iteratively optimize the predictive ability of the model. It uses decision trees as the basic model and continuously adds new decision trees while optimizing the weights of existing ones, so as to improve the overall prediction effect of the model. During training, XGBoost uses gradient boosting technique to minimize the loss function, gradually approaching the optimal solution. The objective function is:

$$Obj^{(t)} = \sum_{i=1}^{N} L\left(y_i, \widehat{y}_i^{(t)}\right) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

Where N is the number of samples, y_i is the true value of the *i-th* sample, $\widehat{y}_i$ is the predicted value of the *i-th* sample, L is the loss function (difference between the true value and the predicted value), $\Omega(f_k)$ is the tree complexity, and K is the number of features.

Perform Taylor second-order expansion on it:

$$Obj^{(t)} = \sum_{i=1}^{N} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^{\,2}(x_i) \right] + \Omega(f_k) + \text{Constant} \tag{3}$$

Where g_i, h_i is the first and second derivative of the loss function respectively, $f_t(x_i)$ is the structure value of tree x_i in the *t-th* iteration.

A tree is defined as:

$$f_t(x) = w_{q(x)}, w \in R^T \tag{4}$$

Where q represents the structure of the tree; T is the number of leaf nodes; w is a one-dimensional vector of length T that represents the weight of a leaf node.

The complexity of the tree is obtained by means of hyperparameters and weighting the number of leaf nodes of the tree and L2 norm of the node weight vector respectively, and is defined as:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{5}$$

By substituting equations (4) and (5) into equation (3) and ignoring the constant, the objective function can be translated as:

$$Obj^{(t)} = \gamma T + \sum_{j=1}^{T} \left[w_j G_j + \frac{1}{2} w_j^2 (\lambda + H_j) \right] \tag{6}$$

Where $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_i$.

The minimum value is obtained when $w_j = -\frac{G_j}{H_j + \lambda}$, that is, the optimal solution is:

$$Obj^{(t)*} = \gamma T - \frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} \tag{7}$$

Formula (7) is the expression of XGBoost model, and the smaller the formula is, the better the model prediction effect is.

In this study, the XGBoost model is trained with 80% datasets of the years 2014–2022 and tested with the other 20% datasets. And its performance is evaluated by comparing observations and predicted data of test datasets by coefficient of determination (R^2), the means of the root mean squared error (RMSE) and normalized root mean squared error (NRMSE).

2.2.3. Feature importance index

There are three main ways to calculate the importance of features in XGBoost (Zhang et al., 2022): weight (the number of times a feature is used to segment the data across all trees), gain (the average gain of all split features), and coverage (the average coverage of all split features). In this study, the importance of features is measured by weight. The higher the importance of the feature, the greater the influence of the feature on the target learning value.

2.2.4. Key focus regions

Here we analyze the surface ozone at all stations in China, as well as five major urban agglomerations, namely, BTH, YRD, PRD, middle Yangtze River (MYR) and Chengdu-Chongqing region (CC), as shown in Fig. 1. These regions include provincial capitals, municipalities or other megacities with representative pollution levels in recent decades.

3. Results and discussion

3.1. Feature importance and model performance evaluation

Although the XGBoost model do not make assumptions about the correlation between the input variables, from the perspective of model interpretability and model overfitting, the correlation between variables still has an impact. We select the features based on the Pearson correlation coefficient (Figs. S1a–c) for each pair of variables, as well as the feature importance scores (Fig. S1d). Spatiotemporally, the variables MEH, T2m, MT2m, and SSRD have a strong positive correlation with OBS ($p < 0.01$). In addition, variables MSL, TCC and SP have a negative correlation with OBS ($p < 0.01$). These findings are consistent with previous studies by Han et al. (2020), Jacob and Winner (2009), and Kavassalis and Murphy (2017). The correlations are more evident in Fig. S1b. Spatially, MEH, T2m, and MT2m show stronger positive correlation with OBS (0.83, 0.70, 0.68), while MSL shows stronger negative correlation with OBS (−0.70). However, the temporal correlation is relatively unrepresentative.

In terms of feature importance, we show that the rankings differ from the correlation coefficients. The top-ranking features are U10, V10, D2m, SP, and TCC. It has been reported that near-surface wind speed in China has been decreasing over the past 50 years (Ding et al., 2020; Guo et al., 2011; Chen et al., 2012). This decrease hampers the dispersion of pollutants and promotes the increase in surface ozone concentrations. This could explain why U10 and V10 are ranked high in terms of feature importance, which is similar to the analysis of Liu and Wang (2020).

Fig. 2 shows the performance evaluation results of the final model. The XGB can effectively estimate daily ozone mixing ratios at most ground-based monitoring stations. On average, for daily retrievals, the XGB model shows a reliable overall accuracy, with a high cross-validated coefficients of determination ($R^2 = 0.66$), a corresponding root-mean-square error (RMSE) of 37.89 ppb, and normalized root-mean-square error (NRMSE) of 0.06. The minor difference in the statistical meterics between the day and month ($R^2 = 0.74$; RMSE = 25.95 ppb, NRMSE = 0.08) level indicates that averaging over time reduces the errors and there is no obvious temporal overfitting in this model. These results highlight the model's capability in accurately predicting daily surface ozone levels at locations.

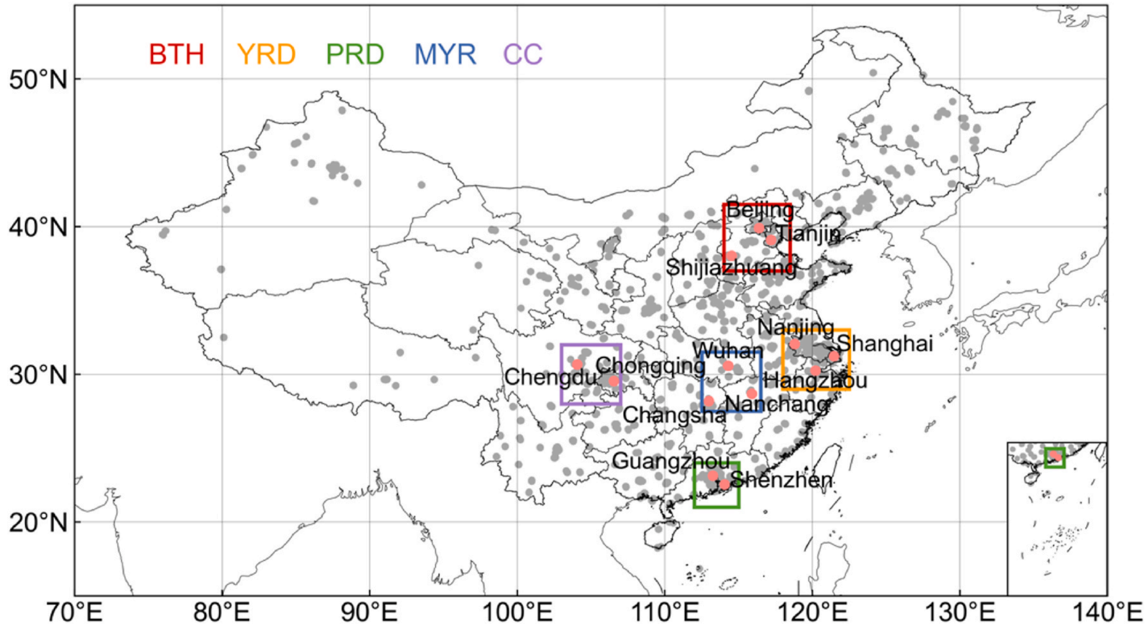

Fig. 1. The locations of monitoring stations (gray dots) and the key focus regions of BTH, YRD, PRD, MYR, CC. The red dots represent the provincial capitals, municipalities or other megacities in these regions. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

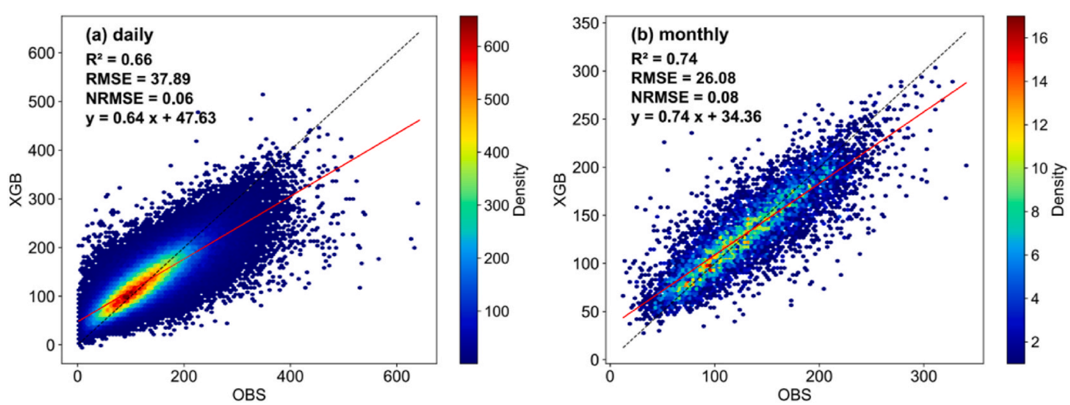

Fig. 2. Daily (a) and monthly (b) scatter density plot of the cross-validation result for the final estimator.

3.2. Spatiotemporal comparison of ozone in OBS, MEH and XGB from 2014 to 2022

Fig. 3a (3b) shows that MEH ozone simulations averaged over China (five key focus regions) significantly underestimate the measured daily (monthly) mean ozone concentrations. Spatially, average ozone mixing raitos for the years 2014–2022 at 95% of sites are underestimated by more than 60 ppb, as shown in Fig. 3c, d and 3f. Such large underestimations are effectively alleviated by the XGB model, with the biases declined to be −9.12–9.37 ppb (−6.87%–7.05%). In addition, the MEH model fails to capture the significant daily, monthly, and seasonal variations in observed ozone (Fig. 3a and b). The XGB can accurately reproduce daily, monthly, and seasonal variations in ozone, as confirmed by high correlations over China ($r = 0.90$) and over the selected regions ($r = 0.50$, 0.56, 0.53, 0.58, 0.58 for BTH, YRD, PRD, MYR, CC, respectively).

To portray national variations, Fig. 3c, d and 3e present the average ozone in observations, MEH simulations, and XGB retrievals from 2014 to 2022. The spatial distribution of observations shows a large amount of ozone in the East and Northwest China. High ozone concentrations over Northwest China may be strongly linked with high solar radiation over there. While large areas of high ozone concentration in the eastern region are probably attributed to high anthropogenic emission sources. The MEH simulations do not reproduce the severe ozone pollution over the Central and Eastern China. However, the spatial pattern of XGB-derived ozone concentrations are highly consistent with surface ozone measurements ($r = 0.90$).

3.3. Spatiotemporal changes of ozone concentration from 1950 to 2014

During the period 1950–2014, ozone concentrations show an increasing trend (Fig. 4a). The average enhancements are 0.29 ppb yr^{-1} for China, 0.29 ppb yr^{-1} for BTH, 0.30 ppb yr^{-1} for YRD, 0.08 ppb yr^{-1} for PRD, 0.14 ppb yr^{-1} for MYR, and 0.25 ppb yr^{-1} for CC, respectively. These trends are roughly consistent with those reported in other studies, such as 0.28–1.02 ppb yr^{-1} in the PRD from 2006 to 2019 (Li et al., 2022b), 0.58 ppb yr^{-1} in Hong Kong, China from 1994 to 2007 (Wang et al., 2009), and 0.51 ppb yr^{-1} in Taiwan, China from 1994 to 2012 (Chen et al., 2014). They are also comparable to the ozone trends observed in Japan from 1980 to 2005 (0.27 ppb yr^{-1}) (Nagashima et al., 2017) and in some areas of South Korea from 2001 to 2018 (0.21–0.88 ppb yr^{-1}) (Yeo and Kim, 2021), but lower than those in the North China

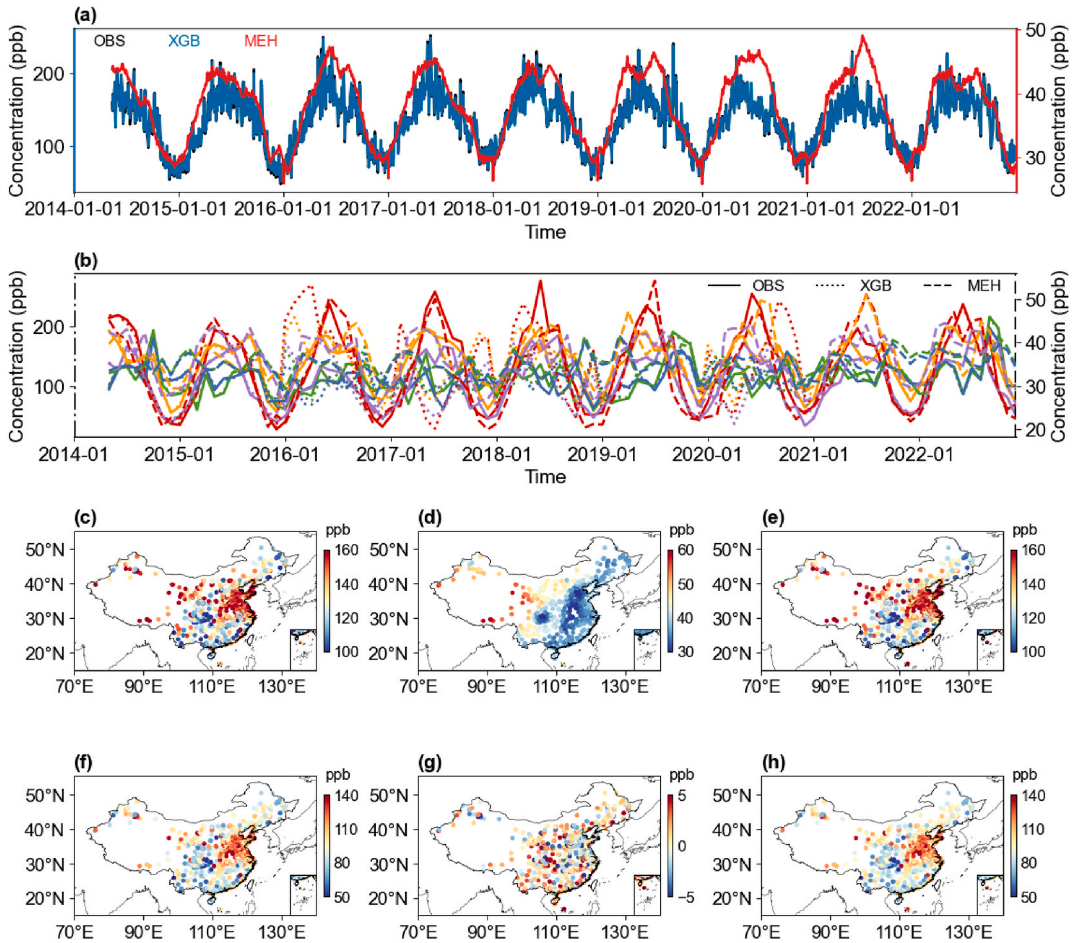

Fig. 3. The time series of daily mean ozone concentrations averaged over China (a) and monthly mean ozone averaged over five key focus regions (b) for OBS, MEH and XGB from 2014 to 2022. The ozone spatial distribution for OBS (c), MEH (d) and XGB (e) averaged over this period. Also shown are the differences between OBS and MEH (f), between XGB and OBS (g), between XGB and MEH (h).

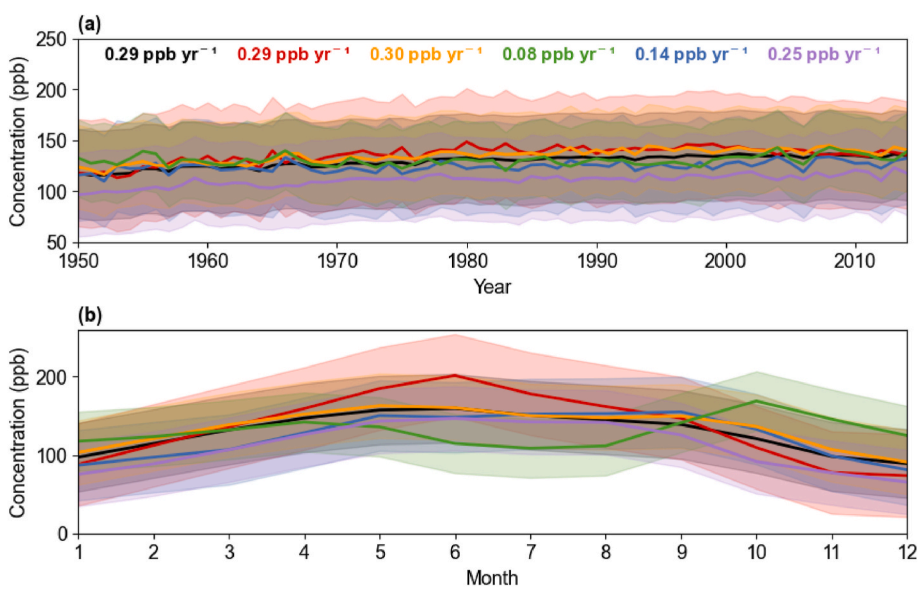

Fig. 4. The average annual (a) and monthly (b) ozone concentrations over China, BTH, YRD, PRD, MYR, CC simulated by XGBoost from 1950 to 2014, the shadow represents the standard deviation and the numbers represent the long-term trends.

Plain (1.58 ppb yr^{-1} from 2006 to 2017) and eastern China (2.3 ppb yr^{-1} from 2013 to 2017) (Cheng et al., 2019; Li et al., 2019). Through an extensive literature review, Sicard (2021) has concluded that ozone concentrations increased in most regions of East Asia with an enhancement of 0.21 ppb yr^{-1} at rural stations over the period 2000–2010 and 0.68 ppb yr^{-1} in cities between 2015 and 2014. However, the upward trends simulated by MEH are 0.17 ppb yr^{-1} (China), 0.10 ppb yr^{-1} (BTH), 0.15 ppb yr^{-1} (YRD), 0.22 ppb yr^{-1} (PRD), 0.14 ppb yr^{-1} (MYR) and 0.14 ppb yr^{-1} (CC), respectively. This indicates that the MEH model also underestimates the long-term trend of ozone concentration, except for the PRD and MYR.

The XGB ozone results show that daily mean ozone concentration without climate change effects is estimated to be 117 ppb in the year 1950 averaged over China (Fig. 4). This result is much higher than the baseline ozone concentrations of North America (<60 ppb, Emery et al., 2012), Europe (<50 ppb, Derwent et al., 2018), and East Asia (<80 ppb, Lam and Cheung, 2022). It indicates that the increase in anthropogenic emissions of China has a significant contribution to ozone enhancement between 1950 and 2014.

In addition, significant differences in the seasonal variations of ozone concentration exist due to latitude. For example, compared with the PRD, BTH is located at a higher latitude, showing a more significant change (up to 128 ppb; Fig. 4b) in ozone concentration due to the notable seasonal changes in temperature, solar radiation, and other meteorological factors. Conversely, the PRD, located at a lower latitude, shows a moderate seasonal variations (~60 ppb) in ozone concentration due to relatively stable temperature and solar radiation throughout the year. Previous studies have reached similar conclusions. Wang et al. (2022) demonstrated that the seasonal difference of ozone concentration in BTH was about 60 ppb, while ~35 ppb in YRD and ~23 ppb in

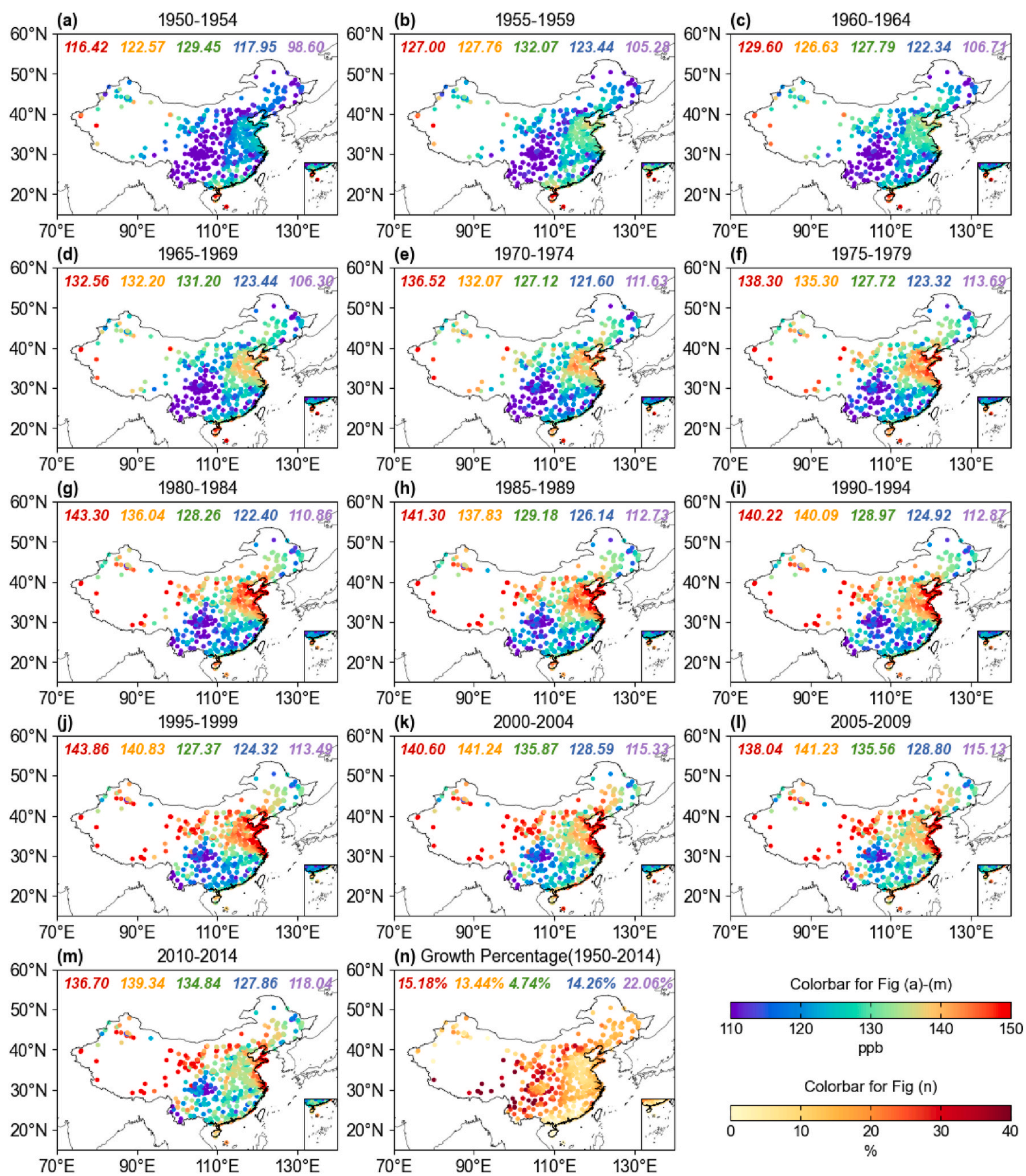

Fig. 5. The spatial distribution of the average ozone concentration of XGB retrievals in every five years from 1950 to 2014 (a–m), and the spatial distribution of ozone growth percentage from 1950 to 2014 (n), with the numbers represent the corresponding values of BTH, YRD, PRD, MYR, and CC regions.

PRD. Lu et al. (2019) showed that the seasonal difference of ozone concentration was approximately 95 ppb in Beijing (at BTH) and 70 ppb in Shanghai (at YRD). Han et al. (2020) showed that the monthly differences of daily mean ozone concentration of BTH, YRD and PRD were about 80 $\mu g\ m^{-3}$, 40 $\mu g\ m^{-3}$ and 25 $\mu g\ m^{-3}$ respectively, daily maximum 8-h average ozone concentration monthly differences were about 115 $\mu g\ m^{-3}$, 75 $\mu g\ m^{-3}$, 55 $\mu g\ m^{-3}$, respectively.

Fig. 5a–m shows the spatial distribution of ozone concentration averaged over every five years from 1950 to 2014. From 2000 to 2014, the surface ozone concentrations of BTH, PRD, YRD, MYR and other regions in Southeastern China decrease slightly due to the implementation of emission policies such as the Clean Air Actions, which reduced the emission of precursors such as VOCs (Sicard et al., 2023; Zheng et al., 2018). Moreover, the decline of annual mean site-based ozone concentrations during the years 2000–2014 is attributed to rapid decrease of winter ozone shown by XGB retrievals (Fig. 6) and MEH simulations (Fig. 7a, Fig. S2). Such winter ozone decrease is probably due to reaction with NO (i.e., ozone titration) these years (Fig. S3), during which excessive anthropogenic NO_x emissions are emitted with centralized heating in winter as well as overall rapid development.

The ozone concentration in the CC has been relatively low with an annual mean value from 99 ppb (1950–1954) to 118 ppb (2010–2014), but has increased most significantly during this period with a growth rate of 22.06% (Fig. 5n). In the regions of BTH, YRD, and the areas between them, ozone concentration increased from 1950 (BTH: 116 ppb; YRD: 123 ppb) to 1999 (BTH: 144 ppb; YRD: 141 ppb), then slightly decreased from 2000 to 2014 (BTH: 137 ppb; YRD: 139 ppb).

The ozone of MYR also experiences similar changes, but the upward and trends of ozone for the period 1950–2009 (118 ppb versus 129 ppb) are moderate, and the decline from 2010 to 2014 is less obvious. The changes of ozone concentration in the PRD are relatively even less pronounced, with an overall ozone increase of 4.74% from 1950 to 2014. From observations, Xu et al. (2008) reported a moderate decrease of -(0.56 ± 0.23) ppb yr^{-1} ($p < 0.05$) in ozone measurements (1991–2006) of a background station (Lin'an). Xu et al. (2020) applied a Mann-Kendall (M-K) test to analyze the ozone trends during 2004–2016 at eight observation sites in China. They found that there was a significant increase (2% yr^{-1}) in ozone at the background site (Shangdianzi) in North China Plain, a moderate increase at the global baseline site in western China, a significant decrease at the edge of northwest China, and almost no trend at other sites.

The seasonal variations are depicted in Fig. 6. During the period of 1950–2014, the ozone concentration increases at rates of 0.46 ppb yr^{-1}, 0.31 ppb yr^{-1}, 0.23 ppb yr^{-1}, and 0.14 ppb yr^{-1} in spring, summer, autumn, and winter, respectively. These long-term growth rates are lower than the research conducted by Chen et al. (2019) on the seasonal trends of ozone concentration from 2006 to 2016, and the results reported by Wang et al. (2022) for the period of 2014–2017. The highest growth rate in spring indicates a shift in seasonal varation of ozone, with that the high-level ozone concentration in summer gradually advances to spring. Schnell et al. (2016) have also reported that the seasonal variation of historical surface ozone in most parts of East Asia is characterized by the highest in spring. In addition, Cooper et al. (2010) have reported that in recent years (since 1980s), the ozone in Europe and the United States has increased significantly in spring, and the peak value of ozone has shifted to spring.

From the 1950s to the 2010s, due to the climate change, China has experienced a decrease in cloud cover (10.10%) and precipitation (24.12%), an enhancement in solar radiation (7.66%) and surface air temperature (0.49%), a weakened wind (28.19%), and a increased surface pressure (0.12%), as shown in Fig. 7b–h. Since the surface has received more heat and radiation, the ozone photochemical production over both urban and remote regions would be enhanced in this period, especially under the conditions of wide range extreme heat waves and strong radiation during the summertime. And the still wind could slow down the transportation. The XGB-derived ozone show a total ozone growth of 17.21 ppb, while the MEH-simulated ozone reproduce an increase of 9.52 ppb between 1950 and 2014. These results indicate a regional surface ozone penalty due to the warming climate in the absence of changes in anthropogenic polluting activities. Over polluted regions of the world, previous studies also projected a general increase of surface ozone levels in a future warmer climate, particularly during summertime (Fu and Tian, 2019; Zanis et al., 2022). Jing et al. (2017) showed that since 2008, ozone in the Midwest of the United States is more temperature-dependent than that in 1999–2007, and more frequent dry and stagnant weather also jointly enhances ozone concentration. Watson et al. (2016) found that a climate increase of 2 °C would lead to a change in summer ozone concentration of −0.1–0.8 ppb over Europe.

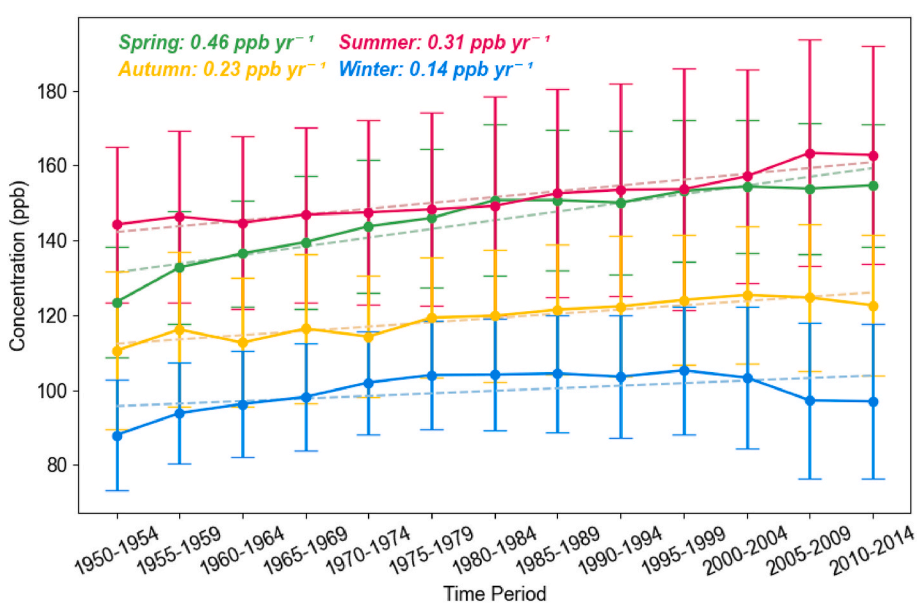

Fig. 6. Seasonal surface ozone concentration averaged over every five years from 1950 to 2014 (spring for MAM, summer for JJA, autumn for SON, winter for DJF) after bias–corrected by XGBoost, the bar represents standard error and the fitting line represents the long-term trend.

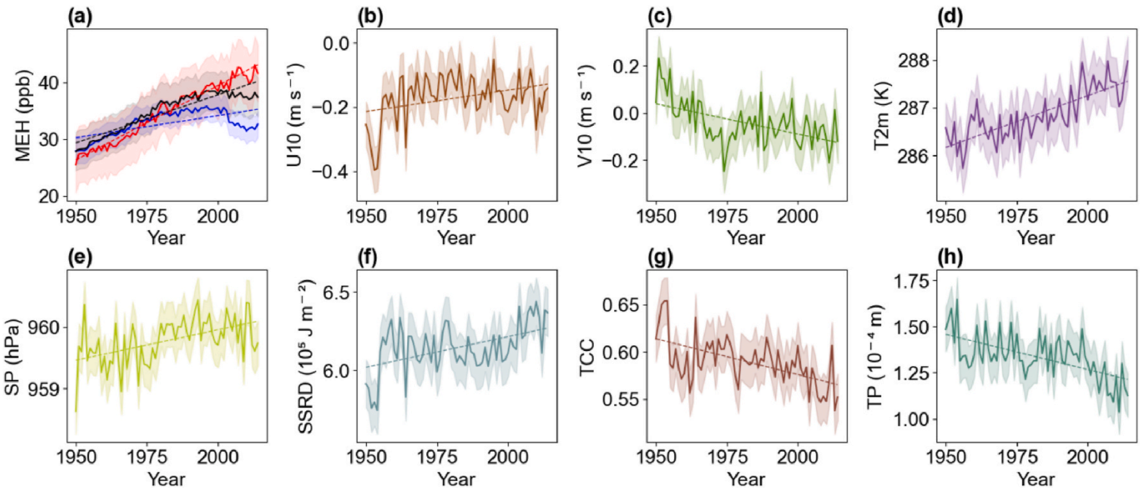

Fig. 7. Annual variation (a–h) for MEH (red represents summer (JJA), blue represents winter (DJF) and black represents the whole year), U10, V10, T2m, SP, SSRD, TCC and TP of ERA5 from 1950 to 2014 respectively. The shadow indicates the standard error and the dashed line indicates the long-term trend. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

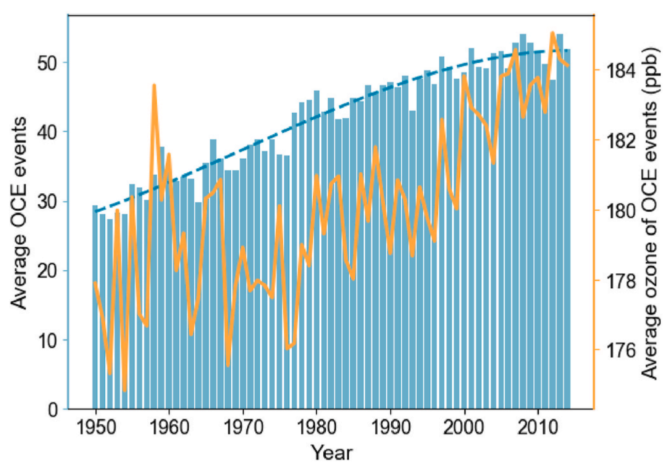

Fig. 8. The number of OCE events (blue bar) with ozone concentration exceeding the standard (daily mean ozone≥70 μg m^{-3}) per year averaged over all stations from 1950 to 2014 after bias–corrected by XGBoost. The average ozone concentration of OCE events in each year (orange line). The blue dashed line is the fitting line of OCE events. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.4. Spatiotemporal changes of ozone exceedances

Fig. 8 presents the statistical analysis of the excessive ozone levels for each year during the study period. The number of events with daily mean ozone concentration exceeding (OCE) 70 μg m^{-3} (ozone standard defined by WHO) shows a fluctuating upward trend averaged over China. The average OCE events increase from 1950s to 2010s. The maximum average OCE events of China is 54, which was occurred in 2008 and 2013. The annual mean ozone averaged over OCE events also shows a fluctuating upward pattern, with the highest value of 185.07 ppb in 2012.

Fig. 9a–m demonstrates the spatial patterns in average OCE events for a year in every five years from 1950 to 2014. From 1950 to 1979, the Eastern and Central regions generally exhibit higher OCE events compared to the Western region, especially in the case of the East China region. However, OCE events at stations in the Western region show a sharp increase, from below 20 events in the period of 1950–1954, and surged to over 260 events in the period of 2010–2014. Overall, there is a clear upward trend in OCE events of China during the years 1950–1999, particularly pronounced in regions of North and Western China. Since 2000, the East China shows a slight decline in OCE events, while the Tibetan Plateau and Western regions continue to experience relatively high OCE events. In the years 2010s, more than half of the stations have over 100 OCE events for a year, which indicates a high level of ozone pollution across the country. The regions with a larger amount of OCE events are mainly concentrated in the East China, Tibetan Plateau, and Western China, indicating that these regions suffers the most severe ozone pollution in China. South China, Southwest China and Northeast China have relatively fewer OCE events than other regions. In terms of the growth percentage from 1950 to 2014, the growth percentage of OCE events in Central and Western regions is significantly higher than that in Eastern regions (Fig. 9n).

Among the key focus regions, the BTH and YRD regions suffer relatively more severe OCE events, followed by PRD and MYR regions, with the relative lowest OCE events occurring in the CC region. In BTH and YRD, OCE events increased from 1950 (BTH: 81; YRD: 85) to 1999 (BTH: 163; YRD: 148), then declined from 1999 to 2014 (BTH: 136; YRD: 133). OCE events in the PRD and MYR regions remained relatively stable from 1950 to 1999 (PRD: 96–102; MYR: 96–97) but experienced growth from 2000 to 2014 (PRD: 117; MYR: 118). The CC region's OCE events showed a more moderate variation, consistently following a fluctuating upward trend from 1950 to 2014.

4. Conclusions

In this study, we use the XGBoost algorithm to correct surface ozone concentrations simulated by the MEH model over a long period of time and across the entire country. This correction yields results that are closer to actual observations and allows for a more accurate analysis of the spatiotemporal variations in surface ozone concentrations in China and five urban agglomerations from 1950 to 2014. We also discuss the spatiotemporal evolution of ozone exceedances and analyze their possible reasons.

The MEH model severely underestimates both the magnitude and trend of surface ozone concentrations, with approximately 95% of stations showing an underestimation of around 40%. Using the XGBoost algorithm for inversion or extrapolation is reasonable, with R^2 value of 0.66 and 0.74 for daily and monthly retrievals, respectively. Based on the XGB-derived surface ozone, concentrations in most parts of China have shown an increasing trend from 1950 to 2014, with growth rates ranging from 0.08 ppb yr^{-1} to 0.30 ppb yr^{-1}. The most significant

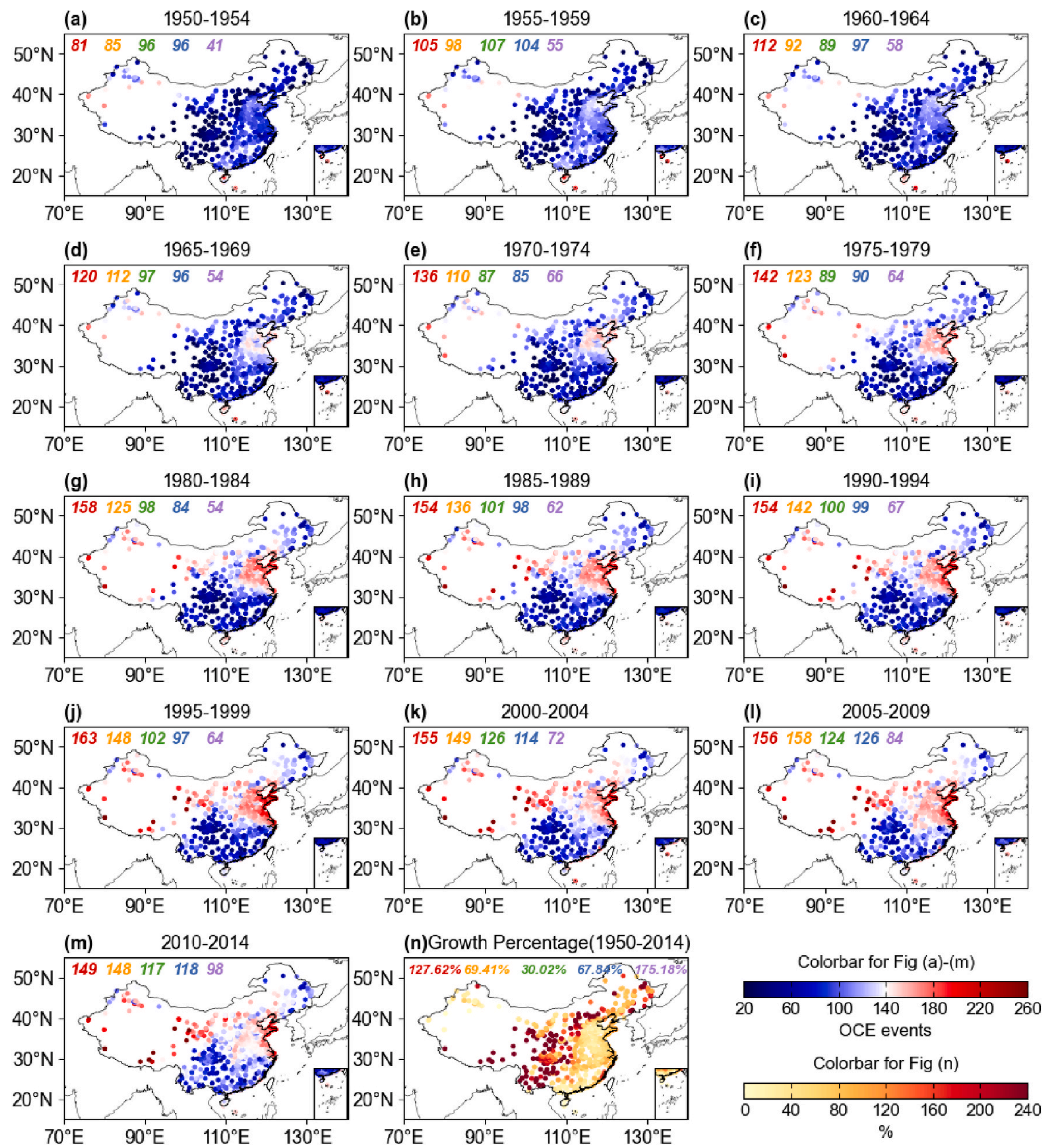

Fig. 9. Spatial distribution of the average number of OCE events per site per year over a five-year period from 1950 to 2014 (a–m), and the spatial distribution of OCE events growth percentage from 1950 to 2014 (n) after bias–corrected by XGBoost, with the numbers represent the corresponding values of BTH, YRD, PRD, MYR, and CC regions.

increment in ozone is estimated in the CC region. Seasonally, the ozone enhancement is largest in spring and smallest in winter, with increasing trends ranging from 0.14 ppb yr^{-1} to 0.46 ppb yr^{-1}. The frequency and concentration of OCE events have increased significantly over time. During the period from 1950 to 1954 and 2010 to 2014, OCE events increased by 76.89%, with a concentration increase of 6.23 ppb.

Through this work, we can understand the background concentration level of surface ozone in the period of weak human influence, and provide a certain reference for understanding the long-term change trend of surface ozone concentration and future ozone pollution prevention and control. Under the background of climate warming, the increasing frequency of weather and climate events such as extreme high temperature, heat waves and atmospheric stability will greatly reduced the efforts of anthropogenic emission reduction. The accurate prediction of meteorological conditions should work together with the accurate reduction of anthropogenic emissions to achieve long-term stability of ozone. However, in this study, due to the lack of data, we did not fully consider the specific situation of historical emissions in China, more accurate results, higher spatiotemporal resolution, and more efficient methods will require further research in the future.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

Funding

This study was financially supported by the Hubei Natural Science Foundation (Grant No. 2022CFB027). It was also supported by the State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry (Grant No. LAPC-KF-2023-07), the Key Laboratory of Atmospheric Chemistry, China Meteorological Administration (Grant No. 2023B08), and innovation and entrepreneurship training program for college students (Grant No. S202310491022).

CRediT authorship contribution statement

Yuanxi Tong: Writing – original draft, Methodology, Investigation, Data curation. **Yingying Yan:** Writing – review & editing, Methodology, Conceptualization. **Jintai Lin:** Writing – review & editing, Supervision. **Shaofei Kong:** Writing – review & editing, Supervision. **Zhixuan Tong:** Data curation. **Yifei Zhu:** Data curation. **Yukun Yan:** Data curation. **Zhan Sun:** Data curation.

Declaration of competing interest

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

Data availability

The Availability of data and materials have been written in the manuscript.

Acknowledgements

We thank the organizations and individuals for providing the research data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2024.124397.

References

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-dujaili, A., Duan, Y., Al-shamma, O., Santamaría, J., Fadhel, M.A., Al-amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data 8 (1). https://doi.org/10.1186/s40537-021-00444-8.

Anenberg, S.C., Horowitz, L.W., Tong, D.Q., Jason west, J., 2010. An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling. Environ. Health Perspect. 118 (9), 1189–1195. https://doi.org/10.1289/ehp.0901220.

Avnery, S., Mauzerall, D.L., Liu, J., Horowitz, L.W., 2011. Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage. Atmos. Environ. 45 (13), 2284–2296. https://doi.org/10.1016/j.atmosenv.2010.11.045.

Calatayud, V., Diéguez, J.J., Agathokleous, E., Sicard, P., 2023. Machine learning model to predict vehicle electrification impacts on urban air quality and related human health effects. Environ. Res. 228, 115835 https://doi.org/10.1016/j.envres.2023.115835.

Chauhan, V.K., Dahiya, K., Sharma, A., 2018. Problem formulations and solvers in linear SVM: a review. Artif. Intell. Rev. 52 (2), 803–855. https://doi.org/10.1007/s10462-018-9614-6.

Chen, L., Li, D., Pryor, S.C., 2012. Wind speed trends over China: quantifying the magnitude and assessing causality. Int. J. Climatol. 33 (11), 2579–2590. https://doi.org/10.1002/joc.3613.

Chen, L., Zhu, J., Liao, H., Yang, Y., Yue, X., 2020. Meteorological influences on $PM_{2.5}$ and O_3 trends and associated health burden since China's clean air actions. Sci. Total Environ. 744, 140837 https://doi.org/10.1016/j.scitotenv.2020.140837.

Chen, S., Chang, C., Liu, J., Chou, C.C.-., Chang, J.S., Wang, J., 2014. Recent improvement in air quality as evidenced by the island-wide monitoring network in Taiwan. Atmos. Environ. 96, 70–77. https://doi.org/10.1016/j.atmosenv.2014.06.060.

Chen, Z., Zhuang, Y., Xie, X., Chen, D., Cheng, N., Yang, L., Li, R., 2019. Understanding long-term variations of meteorological influences on ground ozone concentrations in Beijing During 2006–2016. Environ. Pollut. 245, 29–37. https://doi.org/10.1016/j.envpol.2018.10.117.

Cheng, M., Fang, F., Navon, I.M., Zheng, J., Zhu, J., Pain, C., 2023. Assessing uncertainty and heterogeneity in machine learning-based spatiotemporal ozone prediction in Beijing-Tianjin- Hebei region in China. Sci. Total Environ. 881, 163146 https://doi.org/10.1016/j.scitotenv.2023.163146.

Cheng, N., Li, R., Xu, C., Chen, Z., Chen, D., Meng, F., Cheng, B., Ma, Z., Zhuang, Y., He, B., 2019. Ground ozone variations at an urban and a rural station in Beijing from 2006 to 2017: trend, meteorological influences and formation regimes. J. Clean. Prod. 235, 11–20. https://doi.org/10.1016/j.jclepro.2019.06.204.

Cooper, O.R., Parrish, D.D., Stohl, A., Trainer, M., Nédélec, P., Thouret, V., Cammas, J.P., Oltmans, S.J., Johnson, B.J., Tarasick, D., 2010. Increasing springtime ozone mixing ratios in the free troposphere over western North America. Nature 463 (7279), 344–348. https://doi.org/10.1038/nature08708.

Dang, R., Liao, H., Fu, Y., 2021. Quantifying the anthropogenic and meteorological influences on summertime surface ozone in China over 2012–2017. Sci. Total Environ. 754, 142394 https://doi.org/10.1016/j.scitotenv.2020.142394.

Deng, T., Manders, A., Segers, A., Bai, Y., Lin, H., 2021. Temporal transfer learning for ozone prediction based on CNN-LSTM model. Proceedings of the 13th International Conference on Agents and Artificial Intelligence. https://doi.org/10.5220/0010301710051012.

Derwent, R.G., Manning, A.J., Simmonds, P.G., Gerard spain, T., O'doherty, S., 2018. Long-term trends in ozone in baseline and European regionally-polluted air at Mace Head, Ireland over a 30-year period. Atmos. Environ. 179, 279–287. https://doi.org/10.1016/j.atmosenv.2018.02.024.

Ding, Y., Li, X., Li, Q., 2020. Advances of surface wind speed changes over China under global warming. Journal of Applied Meteorological Science 31 (1), 1–12. https://doi.org/10.11898/1001-7313.20200101.

Du, J., Qiao, F., Lu, P., Yu, L., 2022. Forecasting ground-level ozone concentration levels using machine learning. Resour. Conserv. Recycl. 184, 106380 https://doi.org/10.1016/j.resconrec.2022.106380.

Elminir, H.K., 2005. Dependence of urban air pollutants on meteorology. Sci. Total Environ. 350 (1–3), 225–237. https://doi.org/10.1016/j.scitotenv.2005.01.043.

Emery, C., Jung, J., Downey, N., Johnson, J., Jimenez, M., Yarwood, G., Morris, R., 2012. Regional and global modeling estimates of policy relevant background ozone over the United States. Atmos. Environ. 47, 206–217. https://doi.org/10.1016/j.atmosenv.2011.11.012.

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J., Taylor, K.E., 2016. Overview of the coupled model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev. (GMD) 9 (5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016.

Fu, T., Tian, H., 2019. Climate change penalty to ozone air quality: review of current understandings and knowledge gaps. Current Pollution Reports 5 (3), 159–171. https://doi.org/10.1007/s40726-019-00115-6.

Fu, Y., Tai, A.P.K., 2015. Impact of climate and land cover changes on tropospheric ozone air quality and public health in East Asia between 1980 and 2010. Atmos. Chem. Phys. 15 (17), 10093–10106. https://doi.org/10.5194/acp-15-10093-2015.

Gao, W., Tie, X., Xu, J., Huang, R., Mao, X., Zhou, G., Chang, L., 2017. Long-term trend of O_3 in a mega City (Shanghai), China: characteristics, causes, and interactions with precursors. Sci. Total Environ. 603–604, 425–433. https://doi.org/10.1016/j.scitotenv.2017.06.099.

Griffiths, P.T., Murray, L.T., Zeng, G., Shin, Y.M., Luke abraham, N., Archibald, A.T., Deushi, M., Emmons, L.K., Galbally, I.E., Hassler, B., 2021. Tropospheric ozone in CMIP6 simulations. Atmos. Chem. Phys. 21 (5), 4187–4218. https://doi.org/10.5194/acp-21-4187-2021.

Guo, H., Xu, M., Hu, Q., 2011. Changes in near-surface wind speed in China: 1969–2005. Int. J. Climatol. 31 (3), 349–358. https://doi.org/10.1002/joc.2091.

Han, H., Liu, J., Shu, L., Wang, T., Yuan, H., 2020. Local and synoptic meteorological influences on daily variability in summertime surface ozone in eastern China. Atmos. Chem. Phys. 20 (1), 203–222. https://doi.org/10.5194/acp-20-203-2020.

IPCC, 2021. In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: the Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. https://doi.org/10.1017/9781009157896 (in press).

Jacob, D.J., Winner, D.A., 2009. Effect of climate change on air quality. Atmos. Environ. 43 (1), 51–63. https://doi.org/10.1016/j.atmosenv.2008.09.051.

Jia, B., Dong, R., Du, J., 2020. Ozone concentrations prediction in Lanzhou, China, using chaotic artificial neural network. Chemometr. Intell. Lab. Syst. 204, 104098 https://doi.org/10.1016/j.chemolab.2020.104098.

Jing, P., Lu, Z., Steiner, A.L., 2017. The ozone-climate penalty in the Midwestern U.S. Atmos. Environ. 170, 130–142. https://doi.org/10.1016/j.atmosenv.2017.09.038.

Kavassalis, S.C., Murphy, J.G., 2017. Understanding ozone-meteorology correlations: a role for dry deposition. Geophys. Res. Lett. 44 (6), 2922–2931. https://doi.org/10.1002/2016GL071791.

Lam, Y.F., Cheung, H.M., 2022. Investigation of policy relevant background (PRB) ozone in East Asia. Atmosphere 13 (5), 723. https://doi.org/10.3390/atmos13050723.

Li, D., Shindell, D., Ding, D., Lu, X., Zhang, L., Zhang, Y., 2022a. Surface ozone impacts on major crop production in China from 2010 to 2017. Atmos. Chem. Phys. 22 (4), 2625–2638. https://doi.org/10.5194/acp-22-2625-2022.

Li, K., Jacob, D.J., Liao, H., Shen, L., Zhang, Q., Bates, K.H., 2019. Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. Proc. Natl. Acad. Sci. USA 116 (2), 422–427. https://doi.org/10.1073/pnas.1812168116.

Li, X., Yuan, B., Parrish, D.D., Chen, D., Song, Y., Yang, S., Liu, Z., Shao, M., 2022b. Long-term trend of ozone in southern China reveals future mitigation strategy for air pollution. Atmos. Environ. 269, 118869 https://doi.org/10.1016/j.atmosenv.2021.118869.

Liu, X., Chu, B., Tang, R., et al., 2024. Air quality improvements can strengthen China's food security. Nat Food. https://doi.org/10.1038/s43016-023-00882-y.

Liu, R., Ma, Z., Liu, Y., Shao, Y., Zhao, W., Bi, J., 2020. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: a machine learning approach. Environ. Int. 142, 105823 https://doi.org/10.1016/j.envint.2020.105823.

Liu, Y., Wang, T., 2020. Worsening urban ozone pollution in China from 2013 to 2017 – Part 1: the complex and varying roles of meteorology. Atmos. Chem. Phys. 20 (11), 6305–6321. https://doi.org/10.5194/acp-20-6305-2020.

Liu, Z., Doherty, R.M., Wild, O., O'connor, F.M., Turnock, S.T., 2022. Correcting ozone biases in a global chemistry–climate model: implications for future ozone. Atmos. Chem. Phys. 22 (18), 12543–12557. https://doi.org/10.5194/acp-22-12543-2022.

Lu, X., Zhang, L., Chen, Y., Zhou, M., Zheng, B., Li, K., Liu, Y., Lin, J., Fu, T., Zhang, Q., 2019. Exploring 2016–2017 surface ozone pollution over China: source contributions and meteorological influences. Atmos. Chem. Phys. 19 (12), 8339–8361. https://doi.org/10.5194/acp-19-8339-2019.

Luo, N., Zang, Z., Yin, C., Liu, M., Jiang, Y., Zuo, C., Zhao, W., Shi, W., Yan, X., 2022. Explainable and spatial dependence deep learning model for satellite-based O_3 monitoring in China. Atmos. Environ. 290, 119370 https://doi.org/10.1016/j.atmosenv.2022.119370.

Ma, R., Ban, J., Wang, Q., Zhang, Y., Yang, Y., He, M.Z., Li, S., Shi, W., Li, T., 2021. Random forest model based fine scale spatiotemporal O_3 trends in the Beijing-Tianjin-Hebei region in China, 2010 to 2017. Environ. Pollut. 276, 116635 https://doi.org/10.1016/j.envpol.2021.116635.

Mao, W., Jiao, L., Wang, W., 2022. Long time series ozone prediction in China: a novel dynamic spatiotemporal deep learning approach. Build. Environ. 218, 109087 https://doi.org/10.1016/j.buildenv.2022.109087.

Marco, A.D., Garcia-gomez, H., Collalti, A., Khaniabadi, Y.O., Feng, Z., Proietti, C., Sicard, P., Vitale, M., Anav, A., Paoletti, E., 2022. Ozone modelling and mapping for risk assessment: an overview of different approaches for human and ecosystems health. Environ. Res. 211, 113048 https://doi.org/10.1016/j.envres.2022.113048.

Meehl, G.A., Tebaldi, C., Tilmes, S., Lamarque, J., Bates, S., Pendergrass, A., Lombardozzi, D., 2018. Future heat waves and surface ozone. Environ. Res. Lett. 13 (6), 064004 https://doi.org/10.1088/1748-9326/aabcdc.

Mills, G., Pleijel, H., Malley, C.S., Sinha, B., Cooper, O.R., Schultz, M.G., Neufeld, H.S., Simpson, D., Sharps, K., Feng, Z., 2018. Tropospheric Ozone Assessment Report: present-day tropospheric ozone distribution and trends relevant to vegetation. Elementa: Science of the Anthropocene 6. https://doi.org/10.1525/elementa.302.

Musselman, R., Lefohn, A., Massman, W., Heath, R., 2006. A critical review and analysis of the use of exposure- and flux-based ozone indices for predicting vegetation effects. Atmos. Environ. 40 (10), 1869–1888. https://doi.org/10.1016/j.atmosenv.2005.10.064.

Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D., 2004. An introduction to decision tree modeling. J. Chemometr. 18 (6), 275–285. https://doi.org/10.1002/cem.873.

Nagashima, T., Sudo, K., Akimoto, H., Kurokawa, J., Ohara, T., 2017. Long-term change in the source contribution to surface ozone over Japan. Atmos. Chem. Phys. 17 (13), 8231–8246. https://doi.org/10.5194/acp-17-8231-2017.

Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Stoll, J., Folini, D.S., Tegen, I., Wieners, K., Mauritsen, T., Stemmler, I., Barthel, S., Bey, I., Daskalakis, N., Heinold, B., Kokkola, H., Partridge, D., Rast, S., Schmidt, H., Schutgens, N., Stanelle, T., Stier, P., Watson-Parris, D., Lohmann, U., 2019. HAMMOZ-consortium MPI-ESM1.2-HAM Model Output Prepared for CMIP6 CMIP. Unknown Publisher. https://doi.org/10.22033/esgf/cmip6.1622.

Qiu, X., Wei, Y., Wang, Y., Di, Q., Sofer, T., Awad, Y.A., Schwartz, J., 2020. Inverse probability weighted distributed lag effects of short-term exposure to $PM_{2.5}$ and ozone on CVD hospitalizations in New England Medicare participants - exploring the causal effects. Environ. Res. 182, 109095 https://doi.org/10.1016/j.envres.2019.109095.

Schnell, J.L., Prather, M.J., Josse, B., Naik, V., Horowitz, L.W., Zeng, G., Shindell, D.T., Faluvegi, G., 2016. Effect of climate change on surface ozone over North America, Europe, and East Asia. Geophys. Res. Lett. 43 (7), 3509–3518. https://doi.org/10.1002/2016GL068060.

Sicard, P., 2021. Ground-level ozone over time. An observation-based global overview 19, 100226. https://doi.org/10.1016/j.coesh.2020.100226.

Sicard, P., Agathokleous, E., Anenberg, S.C., Marco, A.D., Paoletti, E., Calatayud, V., 2023. Trends in urban air pollution over the last two decades: a global perspective. Sci. Total Environ. 858, 160064 https://doi.org/10.1016/j.scitotenv.2022.160064.

Sicard, P., Crippa, P., Marco, A.D., Castruccio, S., Giani, P., Cuesta, J., Paoletti, E., Feng, Z., Anav, A., 2021. High spatial resolution WRF-Chem model over Asia: Physics and chemistry evaluation. Atmos. Environ. 244, 118004 https://doi.org/10.1016/j.atmosenv.2020.118004.

Song, A., Yan, Y., Kong, S., Ma, J., Tong, Z., 2022. Impact of synoptic climate system interaction on surface ozone in China during 1950–2014. Atmos. Environ. 279, 119126 https://doi.org/10.1016/j.atmosenv.2022.119126.

Stevenson, D.S., Young, P.J., Naik, V., Lamarque, J.-., Shindell, D.T., Voulgarakis, A., Skeie, R.B., Dalsoren, S.B., Myhre, G., Berntsen, T.K., 2013. Tropospheric ozone changes, radiative forcing and attribution to emissions in the atmospheric chemistry and climate model Intercomparison Project (ACCMIP). Atmos. Chem. Phys. 13 (6), 3063–3085. https://doi.org/10.5194/acp-13-3063-2013.

Tang, K., Zhang, H., Feng, W., Liao, H., Hu, J., Li, N., 2022. Increasing but variable trend of surface ozone in the Yangtze River Delta region of China. Front. Environ. Sci. 10 https://doi.org/10.3389/fenvs.2022.836191.

Tong, Z., Yan, Y., Kong, S., Lin, J., Chen, N., Zhu, B., Ma, J., Zhao, T., Qi, S., 2024. Distribution and formation causes of PM2.5 and O3 double high pollution events in China during 2013–20. Adv. Atmos. Sci. 41 (6), 1235–1250. https://doi.org/10.1007/s00376-023-3156-9.

Trivedi, U.B., Bhatt, M., Srivastava, P., 2021. Prevent overfitting problem in machine learning: a case focus on linear regression and logistics regression. Innovations in Information and Communication Technologies (IICT-2020), pp. 345–349. https://doi.org/10.1007/978-3-030-66218-9_40.

Tsai, W.H., 2018. Prediction of ozone concentration in the air using machine learning methods. Chinese Journal of Environmental Management (2), 78–84. https://doi.org/10.16868/j.cnki.1674-6252.2018.02.078.

Vu, T.V., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., Harrison, R.M., 2019. Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. Atmos. Chem. Phys. 19 (17), 11303–11314. https://doi.org/10.5194/acp-19-11303-2019.

Wang, T., Wei, X.L., Ding, A.J., Poon, C.N., Lam, K.S., Li, Y.S., Chan, L.Y., Anson, M., 2009. Increasing surface ozone concentrations in the background atmosphere of Southern China, 1994–2007. Atmos. Chem. Phys. 9 (16), 6217–6227. https://doi.org/10.5194/acp-9-6217-2009.

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

Wang, W., Parrish, D.D., Wang, S., Bao, F., Ni, R., Li, X., Yang, S., Wang, H., Cheng, Y., Su, H., 2022. Long-term trend of ozone pollution in China during 2014–2020: distinct seasonal and spatial characteristics and ozone sensitivity. Atmos. Chem. Phys. 22 (13), 8935–8949. https://doi.org/10.5194/acp-22-8935-2022.

Wang, X., 2020. Historical air quality data in China. https://quotsoft.net/air.

Wang, Y., Yan, Y., Duan, K., Kong, S., Lin, J., Zheng, H., Song, A., Zhang, Z., 2021. Effect of springtime thermal forcing over Tibetan Plateau on summertime ozone in Central China during the period 1950–2019. Atmos. Res. 261, 105735 https://doi.org/10.1016/j.atmosres.2021.105735.

Watson, L., Lacressonnière, G., Gauss, M., Engardt, M., Andersson, C., Josse, B., Marécal, V., Nyiri, A., Sobolowski, S., Siour, G., 2016. Impact of emissions and +2 °C climate change upon future ozone and nitrogen dioxide over Europe. Atmos. Environ. 142, 271–285. https://doi.org/10.1016/j.atmosenv.2016.07.051.

Xu, X., Lin, W., Wang, T., Yan, P., Tang, J., Meng, Z., Wang, Y., 2008. Long-term trend of surface ozone at a regional background station in eastern China 1991–2006: enhanced variability. Atmos. Chem. Phys. 8 (10), 2595–2607. https://doi.org/10.5194/acp-8-2595-2008.

Xu, X., Lin, W., Xu, W., Jin, J., Wang, Y., Zhang, G., Zhang, X., Ma, Z., Dong, Y., Ma, Q., 2020. Long-term changes of regional ozone in China: implications for human health and ecosystem impacts. Elem Sci Anth 8. https://doi.org/10.1525/elementa.409.

Yan, Y., Lin, J., He, C., 2018. Ozone trends over the United States at different times of day. Atmos. Chem. Phys. 18 (2), 1185–1202. https://doi.org/10.5194/acp-18-1185-2018.

Yan, Y., Zheng, H., Kong, S., Lin, J., Yao, L., Wu, F., Cheng, Y., Niu, Z., Zheng, S., Zeng, X., 2021. On the local anthropogenic source diversities and transboundary transport for urban agglomeration ozone mitigation. Atmos. Environ. 245, 118005 https://doi.org/10.1016/j.atmosenv.2020.118005.

Yeo, M.J., Kim, Y.P., 2021. Long-term trends of surface ozone in Korea. J. Clean. Prod. 294, 125352 https://doi.org/10.1016/j.jclepro.2020.125352.

Zanis, P., Akritidis, D., Turnock, S., Naik, V., Szopa, S., Georgoulias, A.K., Bauer, S.E., Deushi, M., Horowitz, L.W., Keeble, J., 2022. Climate change penalty and benefit on surface ozone: a global perspective based on CMIP6 earth system models. Environ. Res. Lett. 17 (2), 024014 https://doi.org/10.1088/1748-9326/ac4a34.

Zhang, A., Fu, T., Feng, X., Guo, J., Liu, C., Chen, J., Mo, J., Zhang, X., Wang, X., Wu, W., 2023a. Deep learning-based ensemble forecasts and predictability assessments for surface ozone pollution. Geophys. Res. Lett. 50 (8) https://doi.org/10.1029/2022GL102611.

Zhang, B., Zhang, Y., Jiang, X., 2022. Feature selection for global tropospheric ozone prediction based on the BO-XGBoost-RFE algorithm. Sci. Rep. 12 (1) https://doi.org/10.1038/s41598-022-13498-2.

Zhang, Q., Yin, Z., Lu, X., Gong, J., Lei, Y., Cai, B., Cai, C., Chai, Q., Chen, H., Dai, H., 2023b. Synergetic roadmap of carbon neutrality and clean air for China. Environmental Science and Ecotechnology 16, 100280. https://doi.org/10.1016/j.ese.2023.100280.

Zhao, S., Yin, D., Yu, Y., Kang, S., Qin, D., Dong, L., 2020. $PM_{2.5}$ and O_3 pollution during 2015–2019 over 367 Chinese cities: spatiotemporal variations, meteorological and topographical impacts. Environ. Pollut. 264, 114694 https://doi.org/10.1016/j.envpol.2020.114694.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., 2018. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions 18 (19), 14095–14111. https://doi.org/10.5194/acp-18-14095-2018.