



Uncovering the evolution of ozone pollution in China: A spatiotemporal characteristics reconstruction from 1980 to 2021

Su Ding ^{a,b,*}, Zhiwei Wei ^{c,d,**}, Shuiling Liu ^e, Rong Zhao ^f

^a School of Environmental and Resources Science, Zhejiang A & F University, Hangzhou 311300, China

^b State Key Laboratory of Subtropical Silviculture, Zhejiang A & F University, Hangzhou 311300, China

^c Key Laboratory of Network Information System Technology (NIST), Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China

^d Guangdong Laboratory of Artificial Intelligence and Digital Economy, Shenzhen 518132, China

^e Wuhan Natural Resources Conservation and Utilization Center, Wuhan 430010, China

^f Vanke School of Public Health, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Keywords:

Ozone pollution
Long-term trend
Machine learning
Data-driven modeling

ABSTRACT

Ozone pollution emerged as a significant environmental concern in China in recent years. Given the enduring impact of anthropogenic activities on the atmosphere, understanding the long-term evolutionary patterns of ozone pollution is crucial for effective pollution mitigation strategies. However, there have been limited studies to reconstruct spatiotemporal continuity ozone concentrations in China since 1980. In this research, monthly MDA8 ozone concentrations from 1980 to 2021 were first estimated by reproducing the numerical relations between ozone pollution and the atmospheric environment based on a machine learning approach. The verification results from temporal, spatial, and sample-based cross-validation methods demonstrated the model performance in simulating MDA8 ozone concentration spatiotemporally, with correlation coefficients (R) ranging from 0.76 to 0.79, mean square error ($RMSE$) ranging from $22.72 \mu\text{g}\cdot\text{m}^{-3}$ to $24.28 \mu\text{g}\cdot\text{m}^{-3}$, and mean absolute error (MAE) ranging from $17.06 \mu\text{g}\cdot\text{m}^{-3}$ to $18.25 \mu\text{g}\cdot\text{m}^{-3}$. The evolution of ozone pollution in China can be delineated into two phases, characterized by an increasing trend since 2001, coinciding with significant rises in average annual temperature and hot weather frequency over the past four decades. The 90th percentile of ozone concentration from 2001 to 2021 increased by 1.79% compared to the period from 1980 to 2001. Notably, ozone episodes have expanded beyond summer months, partly due to global warming and frequent extreme climate events, contributing to an annual temperature increase of 0.03°C . Despite several pollutant emission restrictions since 2013, the ozone pollution in China has persisted due to the increased emissions of VOCs. Spatially, severe ozone pollution hotspots are evident in Xinjiang province, the Jing-Jin-Ji urban agglomeration, and Northeast China, with notable increases observed in Tibet and the Guangdong-Hong Kong-Macao Greater Bay Area over the past six years. Moreover, the influence factors were identified by the average reduction of node impurity calculation. The evolution of ozone pollution was inextricably linked to solar radiation and atmospheric temperature, which played pivotal positive roles in ozone formation. Whilst, there is significant heterogeneity in the effect of meteorological conditions on ozone concentrations due to the disparity between land and oceanic climates.

1. Introduction

The severity of ozone pollution in China in recent years has exceeded expectations. Both tropospheric ozone concentrations and the number of days above the pollution limit show a consistent increase with time (Wang et al., 2022). Ozone pollution poses a serious threat to human

health (Cakmak et al., 2016; Chen et al., 2023; Rich et al., 2018). At the same time, it also exerts negative impacts on climate change, ecosystem carbon sequestration capacity, and food security (Agathokleous et al., 2020; Ren, 2021; Sampedro et al., 2020). As a result, the scientific and precise promotion of ozone pollution control has been given great importance. Reconstructing spatiotemporal features of tropospheric

* Corresponding authors at: School of Environmental and Resources Science, Zhejiang A & F University, Hangzhou 311300, China.

** Corresponding author.

E-mail addresses: suding@zafu.edu.cn (S. Ding), 2011301130108@whu.edu.cn (Z. Wei).

ozone is fundamental to understanding the evolution of ozone pollution characteristics and the formulation of treatment strategies.

In-situ observations and simulation outputs from ozone pollution models are effective for discerning ozone pollution signatures. Monitoring stations offer direct access to ozone concentrations, yet solely represent local pollution. Interpolation methods for continuous spatio-temporal ozone data from monitoring data entail uncertainties due to limited spatial autocorrelation considerations. Furthermore, monitoring inventories in China only encompass ozone pollution post-2013. Modeling approaches like CMAQ, WRF-CHEM, and GEO-CHEM, along with data-driven methods, offer proficient ozone concentration estimations (Li et al., 2023; Xiong et al., 2022). Machine learning methods, a subset of data-driven modeling, have gained widespread use in ozone pollution simulations, particularly benefiting from the nonlinear correlation between ozone concentration and influencing factors (Yafouz et al., 2021; Zhang et al., 2022). Given that tropospheric ozone originates from both stratospheric inputs and photochemical pollution, it is crucial to incorporate total ozone column and meteorological conditions into the construction of data-driven models for ozone concentration simulation (Gao et al., 2021; Weng et al., 2022; Zhan et al., 2021).

With the escalating severity of ozone pollution in China, numerous scientific publications have delved into the spatiotemporal evolution of ozone concentration (Liu et al., 2020; Ma et al., 2021b; Wei et al., 2021). However, research on reconstructing surface ozone concentrations from 1980 onwards remains scarce. Ozone is a secondary pollutant produced by nitrogen oxides (NO_x) and volatile organic compounds (VOCs) under solar radiation and is affected by both anthropogenic activities and the natural environment. Its formation entails a complex non-linear relationship with driving forces. On one hand, anthropogenic pollutant emissions significantly contribute to changes in atmospheric environmental composition and oxidation (Lu et al., 2023). Unreasonable

reductions in NO_x may result in an increase in ozone concentration due to the attenuation of the NO titration effect (Li et al., 2021b). Meanwhile, a decrease in particulate matter slows down the aerosol sink of hydroperoxy (HO_2) radicals, thereby stimulating ozone production (Li et al., 2019). On the other hand, global warming is also inextricably linked to anthropogenic activity and can drive photochemical reactions. Consequently, long-term anthropogenic impacts significantly elevate ozone concentrations by altering the atmospheric environment (Ding et al., 2021; Huang et al., 2021; Liu et al., 2022). The detection of long-term features of ozone pollution is beneficial for identifying the impacts of socioeconomic activities and formulating effective control strategies. Considering the lack of ozone pollution simulations of China over the past 42 years, this research reconstructed ozone concentrations in China from 1980 to 2021 employing a random forest model based on total ozone column and meteorological reanalysis data.

2. Data and method

2.1. Data sources and processing

In this research, surface observation data and reanalysis dataset describing ozone pollution and meteorological conditions in China from 1980 to 2021 were collected for ozone concentration simulations. The detail description of data is shown in Table S1.

2.1.1. In-situ ozone measurement

Surface ozone concentrations were monitored by China National Environmental Monitoring Centre (CNEMC). A total of 2024 stations recorded in-situ ozone observations on an hourly scale between 2015 and 2021. Furthermore, 29 stations across China measuring hourly ozone concentration during 1991 to 2015 was collected from

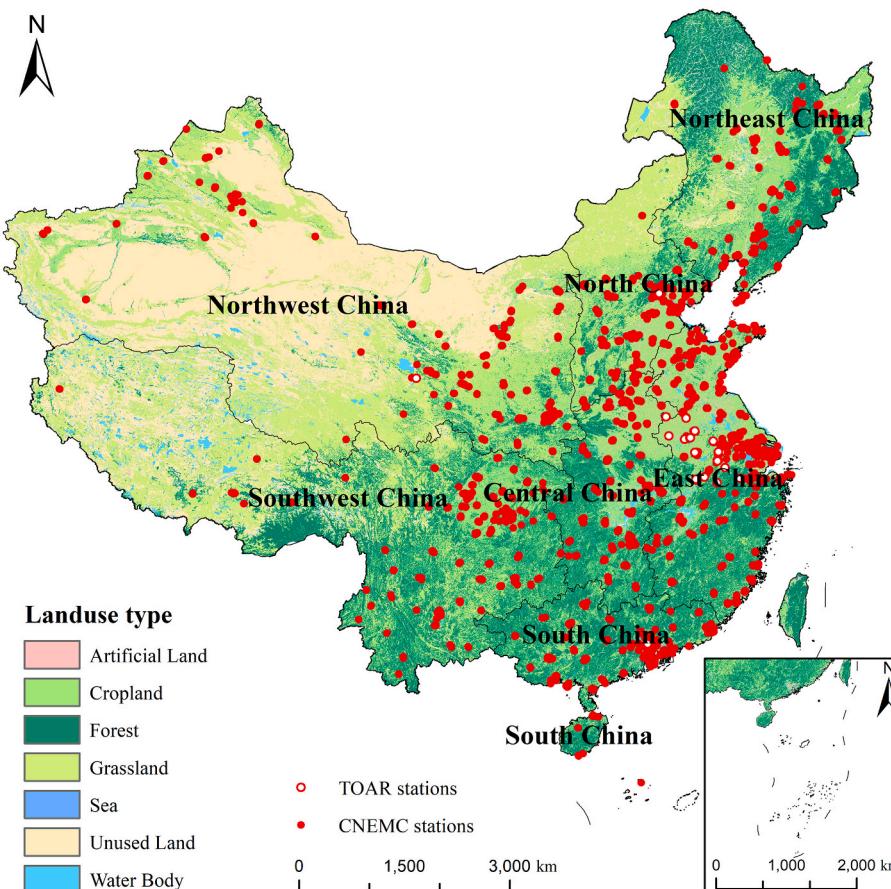


Fig. 1. The spatial distributions of atmospheric pollution monitoring stations.

Tropospheric Ozone Assessment Report (TOAR) data center (Fig. 1) (Schultz et al., 2017). Raw data preprocessing is requested for training dataset integration. Missing data removal and daily maximum 8 h average (MDA8) concentration calculation was initially undertaken. Following the criteria set by the China national ambient air quality standard (GB3095–2012), MDA8 ozone concentration is defined as the average of at least 14 valid 8-h average concentrations from 8:00 to 24:00 within a natural day. In cases where 14 valid data points are not met, statistical validity is maintained if the daily maximum 8-h average concentration exceeds the concentration threshold. Finally, a spatial grid with resolution of 0.25° (monthly dataset) and 1.875° (daily dataset) were established to ensure spatial matching. The near-ground station index value of each grid point is derived from the average of all monitoring stations within the corresponding grid.

2.1.2. Reanalysis dataset

Three meteorological reanalysis datasets were utilized for ozone concentration reconstruction. Total ozone column data were obtained from the MERRA2 reanalysis dataset (GMAO, 2015) representing ozone concentrations throughout the vertical extent of the atmosphere, encompassing the troposphere and a significant portion of the stratosphere. Notably, tropospheric ozone levels are influenced by both stratospheric ozone input and local emissions of precursor gases, such as NO_x and VOCs. Thus, while total column ozone primarily reflects stratospheric ozone levels, it indirectly mirrors tropospheric ozone variations due to the interaction between stratospheric ozone input and tropospheric processes. Additionally, various studies have underscored the significance of total column ozone in modeling ozone pollution (Liu et al., 2020; Luo et al., 2022; Wei et al., 2021). The fifth-generation ECMWF reanalysis dataset (ERA5) was employed for monthly MDA8 ozone concentration estimates from 1980 to 2021. The NCEP/DOE Reanalysis II dataset provides daily meteorological conditions and facilitates daily MDA8 ozone concentration retrieval. Previous research has demonstrated the close relationship between solar radiation, temperature, wind factors and the atmospheric physical and chemical processes contributing to ozone pollution (Gao et al., 2021; Weng et al., 2022; Zhan et al., 2021). These datasets were spatially and temporally matched to the study regions.

2.2. Ozone reconstruction and spatiotemporal analysis

Fig. 2 shows the process of ozone concentration reconstruction as well as the analysis of spatiotemporal characteristics. In this research, daily ozone concentration and meteorological dataset were used as training set for a random forest model called RF-O₃ construction. Based on the validation results, the optimal parameter scheme for the model was selected. Subsequently, the spatial dataset was fed into the RF-O₃ model and the spatiotemporal continuity dataset of ozone concentration was obtained.

2.2.1. RF-O₃ model construction and verification

Daily data combining ozone observations and meteorological variables from 1991 to 2021 served as the training dataset for the construction of ozone simulation model. In this research, random forest (RF) algorithm was employed for constructing ozone model (RF-O₃). RF algorithm refers to an ensemble machine learning algorithm consisting of multiple Class and Regression Tree (CART) (Breiman, 2001). It is equipped with the advantages of high efficiency and accuracy. It has been verified in recent scientific publications that RF method was capable of estimating air pollutants concentrations (Ding et al., 2022; Ma et al., 2021b). Moreover, the contribution of each meteorological variable to ozone pollution can be obtained from the average reduction of node impurity calculated from the random forest model.

Ten-fold cross verification method (CV) was applied to examine model performance. To validate the RF-O₃ historical estimates and assess potential spatial overfitting issues, alongside sample-based cross-

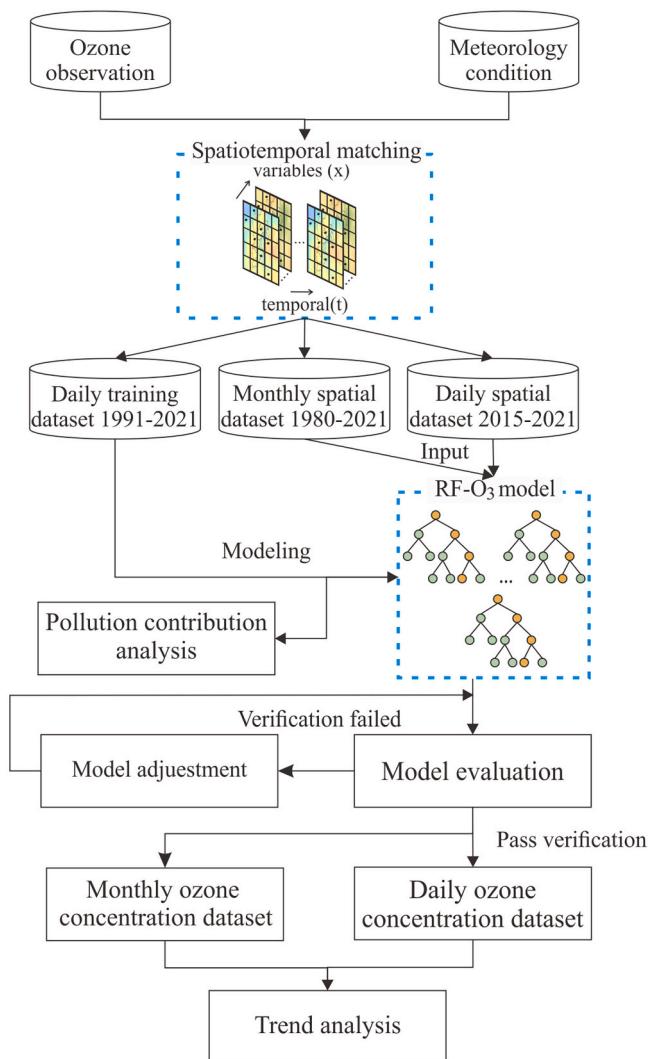


Fig. 2. Methodological flow for ozone concentration reconstruction.

validation (CV), temporal-based CV, and spatial-based CV were utilized in this study. For temporal-based CV, the dataset was divided into seven segments according to the sampling year of the data. Six segments were used as the training dataset for the RF-O₃ model and the remaining segments served as validation dataset for the accuracy of the RF-O₃ model. This process was repeated iteratively until all the data had been modeled and validated. For the spatial-based and sample-based CV, ten groups were divided according to the grid code of the study area and the number of data samples, respectively. In particular, the results of the temporal-based CV were applied for parameter scheme adjustment and the results of the remaining two CV methods were considered as the basis for RF-O₃ performance.

Correlation coefficient (R), root mean square error (RMSE) and mean absolute error (MAE) (Eqs. 1–3) were selected for RF-O₃ model performance evaluation. R denotes the similarity in variation between simulations and observations, while RMSE and MAE quantify the bias between model outputs and observations.

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

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

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

where n is the number of samples, MDA8 ozone concentration simulations and observation data are labeled as M_i and O_i , respectively. O represents mean value of observation.

2.2.2. The trend analysis of ozone pollution

Daily and monthly spatial datasets were fed into the RF-O₃ model with the best parameter scheme to obtain spatiotemporal continuity dataset of MDA8 ozone concentrations. Subsequently, Theil-Sen Median and Mann-Kendall methods were applied to analyze the spatiotemporal trends in ozone pollution. The combination of Theil-Sen median and Mann-Kendall methods has been extensively applied in previous studies (Ding et al., 2022; Munir et al., 2013). The Theil-Sen Median method is employed to capture the long-term trend analysis of ozone pollution, with calculations conducted as follows:

$$S_{O_3} = Median\left(\frac{C_j - C_i}{j - i}\right) \quad (4)$$

where MDA8 ozone concentration variation is quantitatively expressed as S_{O_3} . C_j and C_i refer to the 90th percentile value of annual MDA8 ozone concentration (the number of days exceeding pollution limits) in j and i year ($1980 \leq i \leq j \leq 2021$). The increase trend of ozone concentration is labeled as $S_{O_3} > 0$. Otherwise, $S_{O_3} < 0$ means ozone pollution improved in these years.

The significance of the trend is evaluated using the Mann-Kendall method. The Mann-Kendall statistic for the MDA8 ozone concentration dataset is calculated using Equs.5–6.

$$S = \sum_{i=1}^{n-1} \sum_{j=n+1}^n sgn(C_j - C_i) \quad (5)$$

$$sgn(X_j - X_i) = \begin{cases} +1, & \text{if } (C_j - C_i) > 0 \\ 0, & \text{if } (C_j - C_i) = 0 \\ -1, & \text{if } (C_j - C_i) < 0 \end{cases} \quad (6)$$

where sign function is symbolled as sgn . When $C_j - C_i > 0$, the value of $sgn(C_j - C_i)$ is 1. Otherwise, $C_j - C_i < 0$, $sgn(C_j - C_i) = -1$; $C_j - C_i = 0$, $sgn(C_j - C_i) = 0$.

The trend test is performed using the test statistic Z , as shown in Equs.7–8

$$Z = \begin{cases} \frac{S - 1}{\sqrt{Var(S)}}, & \text{if } S > 0 \\ 0, & \text{if } S = 0 \\ \frac{S + 1}{\sqrt{Var(S)}}, & \text{if } S < 0 \end{cases} \quad (7)$$

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{p=1}^q t_p(t_p-1)(2t_p+5)}{18} \quad (8)$$

$Var(S)$ is variance of test statistics S . t_p and n are data series and the number of the p th group points, respectively. The Z -value is an indicator to judge the significance. When the absolute value of Z is larger than 1.96, the trend of the O₃ concentration is significant at the 95% level.

3. Results and discussion

3.1. Results of RF-O₃ model verification

Fig. 3 illustrates the performance of the RF-O₃ model assessed through temporal-based CV, spatial-based CV, and sample-based CV methods. The R values for all three methods exceed 0.75. The deviation between simulated values and observations is minimal, with $RMSE$ and MAE values of about 20 $\mu\text{g}\cdot\text{m}^{-3}$, indicating consistency between simulated ozone variations and observations, comparable to the accuracy of previous models at the daily level (Ding et al., 2018; Liu et al., 2020). External verification was conducted to assess the performance of the RF-O₃ model using surface ozone observations in Hong Kong during 1990 to 2021. The correlation coefficients for daily and monthly MDA8 ozone concentrations were 0.59 and 0.52, respectively, with corresponding RMSEs and MAEs ranging from 33.43 $\mu\text{g}\cdot\text{m}^{-3}$ to 46.18 $\mu\text{g}\cdot\text{m}^{-3}$. Although most points clustered near the line of $y = x$, indicating small deviations between observations and simulations, the RF-O₃ model tended to underestimate ozone concentrations during periods of high pollution, with a regression linear slope of approximately 0.6. This is attributed to the infrequent occurrence of high ozone levels, as only 4.71% of samples in the training dataset exceeded the secondary concentration threshold (160 $\mu\text{g}\cdot\text{m}^{-3}$) outlined in the China national ambient air quality standard (GB3095–2012). Consequently, the RF algorithm, due to the limited occurrence of high ozone episodes, produces a smoothed maximum value. Additionally, concentrations beyond the range of the training dataset cannot be accurately predicted by the machine learning model (Requia et al., 2020).

3.2. The spatiotemporal characteristics of MDA8 ozone concentration during 1980 to 2021

3.2.1. Annual temporal evolution of MDA8 ozone concentration

Fig. 4 illustrates the 90th percentile of MDA8 ozone concentration over the entire year and statistical results concerning ozone concentrations exceeding both the primary (100 $\mu\text{g}\cdot\text{m}^{-3}$) and secondary concentration threshold (160 $\mu\text{g}\cdot\text{m}^{-3}$) across China throughout the study period, in accordance with the China national ambient air quality standard (GB3095–2012). Over the last 42 years, the 90th percentile of MDA8 ozone concentration have increased 0.08 $\mu\text{g}\cdot\text{m}^{-3}$ on average every year across the country. The turning points came in 2001 and ozone contamination was relatively mild, averaging approximately 115 $\mu\text{g}\cdot\text{m}^{-3}$ before that year. Across China, the frequency of high ozone episodes was about 92%. After 2001, the 90th percentile of MDA8 ozone concentrations exhibited a rapid increase, persisting at elevated levels. Throughout the years 2001 to 2021, the 90th percentile of MDA8 ozone concentration consistently surpassed the average value of 115 $\mu\text{g}\cdot\text{m}^{-3}$ recorded during the period from 1980 to 2000. Simultaneously, severe ozone pollution occurred more frequently. In 2007 and 2001, heavily ozone polluted events were most frequent, with 94.77% and 0.10% of areas exceeding the first-level and the secondary-level concentration threshold, respectively. The peak value of the 90th percentile of MDA8 ozone concentration was observed in 2006, with value of 118 $\mu\text{g}\cdot\text{m}^{-3}$. A comprehensive analysis reveals that both anthropogenic activities and the physical environment have contributed to ozone pollution over the past 42 years. Anthropogenic activities significantly contribute to the emission of VOCs and NO_x, crucial ozone precursors. Previous studies indicate a rising trend in anthropogenic VOCs emissions since 1949, with accelerated growth post-2001 (Li et al., 2021a). Similarly, NO_x emissions exhibited an increasing trend until 2011 (Itahashi et al., 2019). Concurrently, global warming rates have accelerated. Since 1997, the average temperature in China has increased by approximately 1 °C compared to pre-1997 levels. (Fig. S1).

After 2001, ozone pollution showed slight improvements in 2008, 2011, and 2020 compared to neighboring years, with the 90th percentile of annual concentrations at 117 $\mu\text{g}\cdot\text{m}^{-3}$, 117 $\mu\text{g}\cdot\text{m}^{-3}$, and 116 $\mu\text{g}\cdot\text{m}^{-3}$,

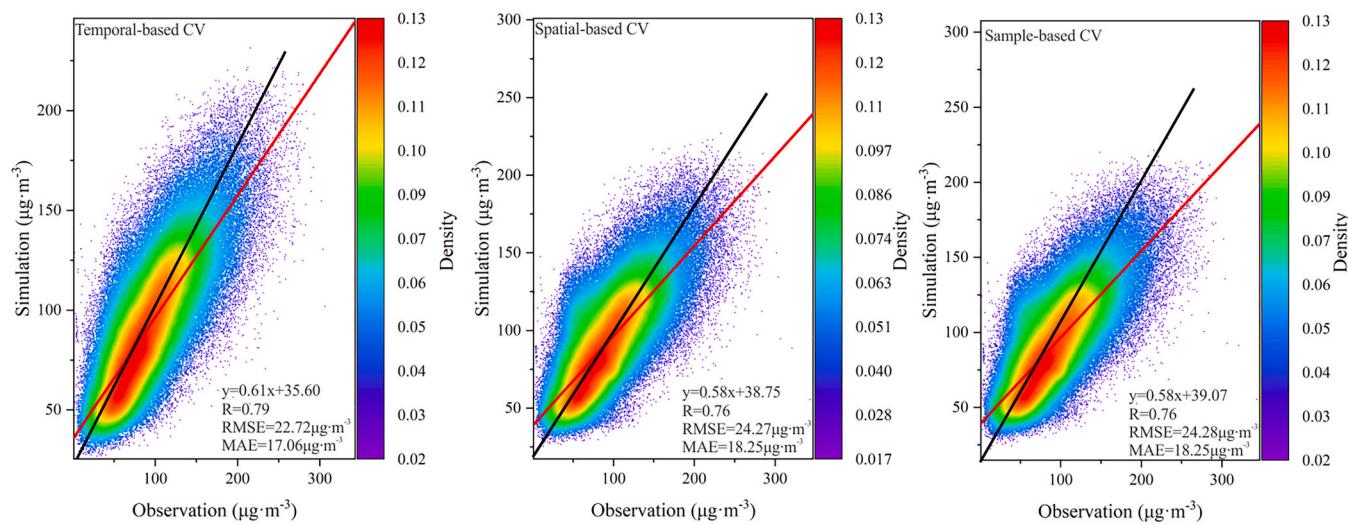


Fig. 3. RF-O₃ performance evaluation based on three CVs methods (the black lines and red lines represent 1:1 line and linear regression line, respectively). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

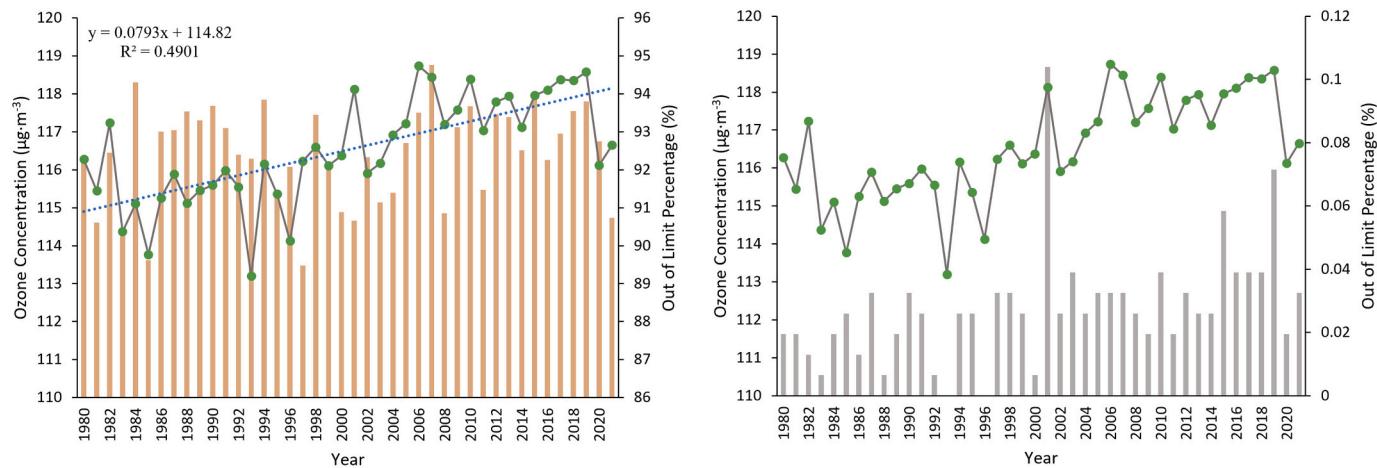


Fig. 4. Out of limit percentage (histogram) and annual average concentration (line) of monthly MDA8 ozone concentration during 1980–2021.

respectively. In 2008 and 2011, lower average temperatures in China were attributed to La Niña events (Fig. S1). However, from 2015 to 2019, consistently high levels of ozone pollution were observed, with the 90th percentile of MDA8 ozone concentration reaching 118 µg·m⁻³ and high pollution frequency of 93.28% (the first level concentration threshold) and 0.05% (the second level concentration threshold). Despite efforts under the China Clean Air Act, NO_x emissions decreased by 21% from 2013 to 2017, effective mitigation measures for non-methane volatile organic compounds (NMVOCs) emission were lacking from 2010 to 2017 (Zheng et al., 2018). Ozone pollution remains a challenge due to inadequate emission reduction strategies. In 2020, ozone concentrations experienced a significant decline due to the COVID-19 pandemic and subsequent economic slowdown. However, with the resumption of economic activities, ozone concentrations rebounded. Overall, ozone pollution persists and shows an increasing trend across China from 1980 to 2021.

3.2.2. Spatial pattern of MDA8 ozone concentration

The spatial distribution of the 90th percentile of MDA8 ozone concentration from 1980 to 2021 was also mapped in Fig. 5. The 90th percentile of MDA8 ozone concentration exceeded 100 µg·m⁻³ in most of regions, except Tibet and northern Northeast China. Sporadic occurrences of the 90th percentile of MDA8 ozone concentrations surpassing

140 µg·m⁻³ are observed in Xinjiang province, Jing-Jin-Ji, northern Central China, and Northeast China. These regions also represent a significant increasing trend, with an average annual increase of at least 0.12 µg·m⁻³ in ozone concentration. The distribution pattern of ozone pollution mirrors that of population density, owing to the close association between ozone formation and anthropogenic activity. Ozone precursors, such as NO_x and VOCs, primarily stem from anthropogenic sources (Ding et al., 2021).

It is notable that elevated ozone concentrations were observed in some regions of Xinjiang province and Northwest China. Trend analysis of the annual 90th percentile of MDA8 ozone concentrations indicates a significant deterioration of ozone pollution in these areas, characterized by an average annual increase in ozone concentration exceeding 0.20 µg·m⁻³. The extreme hot weather and intensive solar radiation in Xinjiang province provide favorable conditions for photochemical reactions. Through the analysis of temperature trend, a significant increase in temperature has been observed in Xinjiang over the last 42 years (Fig. S3). As part of national strategies such as the “Grand Western Development” and “West–East energy transmission project”, numerous large-scale chemical enterprises in Northwest China contribute significantly to air pollutants emissions (Liang et al., 2019). Furthermore, the ozone pollution was also prominent in coastal areas of China. On the one hand, the atmosphere over offshore regions is affected by land transport,

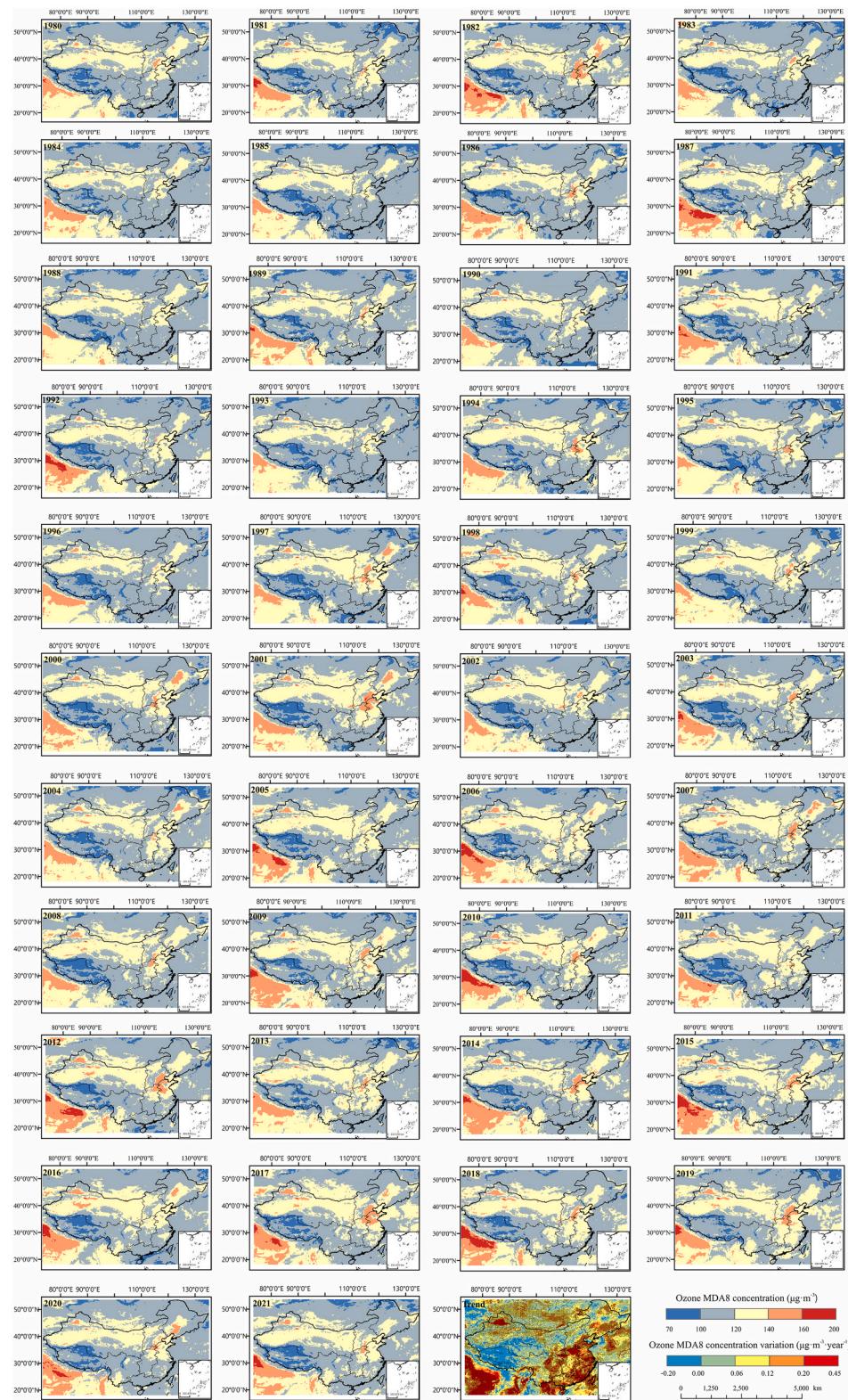


Fig. 5. The 90th percentile of the monthly MDA8 ozone concentration and trend analysis during 1980 to 2021 (black dots signify the pass of significance test in trend analysis map).

leading to high ozone concentrations also observed in this region. On the other hand, fishing and shipping activities at sea also contribute to air pollutants emissions (Merico et al., 2016).

In terms of spatial evolution, the regions with elevated ozone concentrations have expanded since 1997, extending to southern China.

Despite improvements in extremely high ozone pollution in 2005, 2011, and 2013, the national average ozone concentration remained high, with the high pollution areas continuing to expand. Remarkable declines in ozone concentrations were observed in Northeast China during 1990 to 1996, accompanied by a reduction in the area with concentrations

exceeding $120 \mu\text{g}\cdot\text{m}^{-3}$. However, the ozone pollution increase subsequently until 2020. In 2020 and 2021, the ozone concentration decreased significantly due to the COVID-19 lockdown. However, Xinjiang province and some areas in North, Central and East China remain covered by high ozone concentrations. During the COVID-19 lockdown, the observed increase in surface ozone concentrations is likely linked to the reduction in NO_x emissions from vehicle exhaust, leading to a diminished NO titration effect. Unreasonable reduction of precursor emissions may result in an increase in ozone levels (Huang et al., 2021). A balanced decrease in VOCs and NO_x is beneficial for mitigating ozone pollution (Wang et al., 2022).

3.3. Influencing factors contributions to MDA8 ozone concentrations

The node impurity average reduction of the RF- O_3 model, as depicted in Fig. 6, elucidates the contribution of each factor to ozone concentration. Downward solar radiation flux at the surface (DSWRF) notably influences surface concentration, being a pivotal factor in various atmospheric chemistry and photochemical reactions, providing energy for photochemistry. Tropospheric ozone is facilitated by photochemical reactions under solar radiation (Kou et al., 2023). Previous researches have identified significant variations between ozone concentrations and solar radiation intensities with similar trends (Ma et al., 2021a; Xia et al., 2022). Temperature (T) and relative humidity (RH) also exhibit strong correlations with ozone concentration, with temperature affecting ozone concentration positively through enhanced photochemical reactions and BVOCs emissions at higher temperatures. Downward longwave radiation flux (DLWRF), approximately twice the average solar radiation received by the ground (Kiehl et al., 1997), significantly influences meteorological and environmental conditions. DLWRF, being strongly temperature-dependent, plays a role in ozone formation.

4. Conclusion

Anthropogenic activities change the atmospheric environment by discharging the large amounts of air pollutants into the atmosphere, which is one of the main causes of ozone pollution. Understanding how ozone pollution responds to the changes in the atmospheric environment is the fundamental to the formulation of pollution treatment strategies. Besides, the impacts of anthropogenic activities on ozone pollution take a long time to become apparent. To fill the gap in the study of long-term time series ozone concentration reconstruction, the long-term evolution estimates of ozone pollution between 1980 and 2021 are estimated using the RF- O_3 model based on the machine learning algorithm. Spatiotemporal features of ozone concentration and the contributions from independent variables were also detected. It indicated:

- (1) To comprehensively analyze the performance of the RF- O_3 model from both spatial and temporal perspectives, temporal-based, spatial-based and sample-based CVs methods were employed to validate the RF- O_3 model. According to the results of the three kinds of CVs methods, the RF- O_3 model was competent for ozone concentration reconstruction with R of temporal-based CV, spatial-based CV and sample-based CV ranging from 0.76 to 0.79 and RMSE ranging from $22.72 \mu\text{g}\cdot\text{m}^{-3}$ to $24.28 \mu\text{g}\cdot\text{m}^{-3}$ and MAE ranging from $17.06 \mu\text{g}\cdot\text{m}^{-3}$ to $18.25 \mu\text{g}\cdot\text{m}^{-3}$. RF- O_3 model has greater advantage for capturing the spatial pattern of ozone pollution.
- (2) The spatiotemporal features of daily and monthly MDA8 ozone concentrations were reconstructed based on the evaluation criteria in the Chinese national ambient air quality standards. For temporal evolution, ozone pollution began to prominent since 2001 with average concentration and pollution episodes climbing to high levels. The 90th percentile of ozone concentration from 2001 to 2021 increased by 1.79% compared to the period from

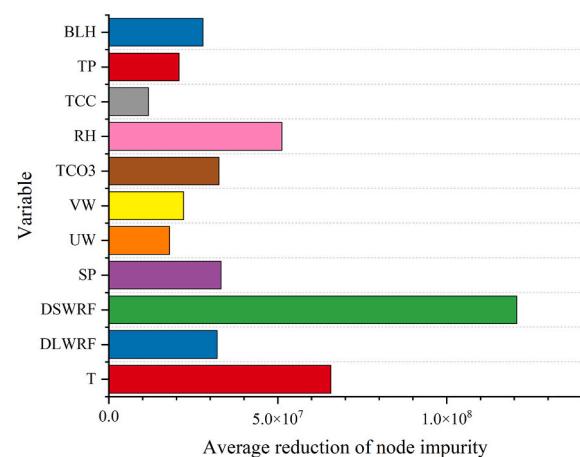


Fig. 6. Influencing factors contributions to ozone MDA8 concentration.

1980 to 2000. The most severe ozone episode was observed in 2006 with the 90th percentile of MDA8 ozone concentration of $118 \mu\text{g}\cdot\text{m}^{-3}$ and the most frequency high pollution occurred in 2007. For the spatial distribution, most regions had ozone concentrations above the first concentration threshold, with high pollution centers located in Xinjiang, Jing-Jin-Ji and Northeast China exceeding $140 \mu\text{g}\cdot\text{m}^{-3}$. In the last six years, northern China has suffered more from ozone pollution than southern China. However, ozone concentrations in Tibet and the Pearl River Delta increased between 2015 and 2021, with an increase of at least 5 days per year. Based on previous research results and climate change analyses, unscientific emission planning and increasing global warming have contributed to the intensification of ozone pollution in China.

- (3) The upward trend in ozone pollution was more pronounced during the summer months. Besides, the frequency of episodes of high ozone during the cold season has also increased in recent years. Seasonal variations in ozone concentration are mostly attributed to meteorological conditions. Climatic anomalies such as La Niña and El Niño significantly influence ozone pollution. As for meteorological factors, solar radiation contributes the most to ozone concentrations. It has been observed that the difference in climate between land and ocean of southern China result in the opposite effects of temperature, downward solar radiation flux at surface and downward longwave radiation flux at surface. Moreover, the correlation between total column ozone and surface ozone concentration varies from northern to southern China due to differences in atmospheric environment. This spatial heterogeneity was also observed in the correlation between northward wind and ozone concentration.

CRediT authorship contribution statement

Su Ding: Conceptualization, Funding acquisition, Investigation, Writing – original draft. **Zhiwei Wei:** Data curation, Visualization, Writing – review & editing. **Shuiling Liu:** Writing – review & editing. **Rong Zhao:** Writing – review & editing.

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

Data will be made available on request.

Acknowledgements

This work was financially supported by Research Development Fund of Zhejiang A & F University (2020FR083).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.atmosres.2024.107472>.

References

- Agathokleous, E., et al., 2020. Ozone affects plant, insect, and soil microbial communities: a threat to terrestrial ecosystems and biodiversity[J]. *Sci. Adv.* 6 (33), eabc1176. <https://doi.org/10.1126/sciadv.abc1176>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Cakmak, S., et al., 2016. Ozone exposure and cardiovascular-related mortality in the Canadian Census Health and Environment Cohort (CANCHEC) by spatial synoptic classification zone[J]. *Environ. Pollut.* 214, 589–599. <https://doi.org/10.1016/j.envpol.2016.04.067>.
- Chen, C., et al., 2023. Short-term exposure to ozone and cause-specific mortality risks and thresholds in China: evidence from nationally representative data, 2013–2018 [J]. *Environ. Int.* 171, 107666 <https://doi.org/10.1016/j.envint.2022.107666>.
- Ding, S., et al., 2018. An applied research of decision-tree-based statistical model in forecasting the spatial-temporal distribution of O₃[J]. *Acta Sci. Circumst.* 38 (8), 3229–3242.
- Ding, S., et al., 2021. Investigating the biophysical and socioeconomic determinants of China tropospheric O₃ pollution based on a multilevel analysis approach[J]. *Environ. Geochem. Health* 43 (8), 2835–2849. <https://doi.org/10.1007/s10653-020-00797-8>.
- Ding, S., et al., 2022. Estimates of PM_{2.5} concentrations spatiotemporal evolution across China considering aerosol components in the context of the Reform and Opening-up [J]. *J. Environ. Manag.* 322, 115983 <https://doi.org/10.1016/j.jenvman.2022.115983>.
- Gao, D., et al., 2021. Ozone variability induced by synoptic weather patterns in warm seasons of 2014–2018 over the Yangtze River Delta region, China[J]. *Atmos. Chem. Phys.* 21 (8), 5847–5864. <https://doi.org/10.5194/acp-21-5847-2021>.
- GMAO, 2015. MERRA-2 instU_2d_asm_Nx: 2d,diurnal,Instantaneous,Single-Level,Assimilation,Single-Level Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC).
- Huang, X., et al., 2021. Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China[J]. *Natl. Sci. Rev.* 8 (2), nwaa137. <https://doi.org/10.1093/nsr/nwaa137>.
- Itahashi, S., et al., 2019. Inverse estimation of NO_x emissions over China and India 2005–2016: contrasting recent trends and future perspectives[J]. *Environ. Res. Lett.* 14 (12) <https://doi.org/10.1088/1748-9326/ab4d7f>.
- Kiehl, J.T., et al., 1997. Earth's Annual Global mean Energy Budget[J]. *Bull. Am. Meteorol. Soc.* 78 (2), 197–208. [https://doi.org/10.1175/1520-0477\(1997\)078<0197:EAGMEB>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<0197:EAGMEB>2.0.CO;2).
- Kou, W., et al., 2023. High downward surface solar radiation conducive to ozone pollution more frequent under global warming[J]. *Sci. Bull.* 68 (4), 388–392. <https://doi.org/10.1016/j.scib.2023.01.022>.
- Li, K., et al., 2019. Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China[J]. *Proc. Natl. Acad. Sci. USA* 116 (2), 422–427. <https://doi.org/10.1073/pnas.1812168116>.
- Li, B., et al., 2021a. A comprehensive review on anthropogenic volatile organic compounds (VOCs) emission estimates in China: Comparison and outlook[J]. *Environ. Int.* 156 <https://doi.org/10.1016/j.envint.2021.106710>.
- Li, K., et al., 2021b. Ozone pollution in the North China Plain spreading into the late-winter haze season[J]. *Proc. Natl. Acad. Sci. USA* 118 (10). <https://doi.org/10.1073/pnas.2015797118>.
- Li, H., et al., 2023. Climate-driven deterioration of future ozone pollution in Asia predicted by machine learning with multi-source data[J]. *Atmos. Chem. Phys.* 23 (2), 1131–1145. <https://doi.org/10.5194/acp-23-1131-2023>.
- Liang, X., et al., 2019. Chemical composition and source apportionment of PM₁ and PM_{2.5} in a national coal chemical industrial base of the Golden Energy Triangle, Northwest China[J]. *Sci. Total Environ.* 659, 188–199. <https://doi.org/10.1016/j.scitotenv.2018.12.335>.
- Liu, R., et al., 2020. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: a machine learning approach[J]. *Environ. Int.* 142, 105823 <https://doi.org/10.1016/j.envint.2020.105823>.
- Liu, T., et al., 2022. Atmospheric oxidation capacity and ozone pollution mechanism in a coastal city of southeastern China: analysis of a typical photochemical episode by an observation-based model[J]. *Atmos. Chem. Phys.* 22 (3), 2173–2190. <https://doi.org/10.5194/acp-22-2173-2022>.
- Lu, K., et al., 2023. Ozone mitigations beyond the control of nitrogen oxides and volatile organic compounds[J]. *Sci. Bull.* 68 (18), 1989–1992. <https://doi.org/10.1016/j.scib.2023.07.051>.
- Luo, N., et al., 2022. Explainable and spatial dependence deep learning model for satellite-based O₃ monitoring in China[J]. *Atmos. Environ.* 290 <https://doi.org/10.1016/j.atmosenv.2022.119370>.
- Ma, M., et al., 2021a. Distinct spatiotemporal variation patterns of surface ozone in China due to diverse influential factors[J]. *J. Environ. Manag.* 288, 112368 <https://doi.org/10.1016/j.jenvman.2021.112368>.
- Ma, R., et al., 2021b. Random forest model based fine scale spatiotemporal O₃ trends in the Beijing-Tianjin-Hebei region in China, 2010 to 2017[J]. *Environ. Pollut.* 276, 116635 <https://doi.org/10.1016/j.envpol.2021.116635>.
- Merico, E., et al., 2016. Influence of in-port ships emissions to gaseous atmospheric pollutants and to particulate matter of different sizes in a Mediterranean harbour in Italy[J]. *Atmos. Environ.* 139, 1–10. <https://doi.org/10.1016/j.atmosenv.2016.05.024>.
- Munir, S., et al., 2013. Quantifying temporal trends of atmospheric pollutants in Makkah (1997–2012)[J]. *Atmos. Environ.* 77, 647–655. <https://doi.org/10.1016/j.atmosenv.2013.05.075>.
- Ren, J., 2021. Effects of O₃ pollution near formation on crop yield and economic loss[J]. *Environ. Technol. Innov.* 22 <https://doi.org/10.1016/j.eti.2021.101446>.
- Requia, W.J., et al., 2020. An Ensemble Learning Approach for estimating High Spatiotemporal Resolution of Ground-Level ozone in the Contiguous United States [J]. *Environ. Sci. Technol.* 54 (18), 11037–11047. <https://doi.org/10.1021/acs.est.0c01791>.
- Rich, D.Q., et al., 2018. Cardiovascular function and ozone exposure: the Multicenter ozone Study in older Subjects (MOSES)[J]. *Environ. Int.* 119, 193–202. <https://doi.org/10.1016/j.envint.2018.06.014>.
- Sampedro, J., et al., 2020. Future impacts of ozone driven damages on agricultural systems[J]. *Atmos. Environ.* 231 <https://doi.org/10.1016/j.atmosenv.2020.117538>.
- Schultz, M.G., et al., 2017. Tropospheric ozone Assessment Report: Database and metrics data of global surface ozone observations[J]. *Elementa: Sci. Anthropol.* 5, 58. <https://doi.org/10.1525/elementa.244>.
- Wang, W., et al., 2022. Long-term trend of ozone pollution in China during 2014–2020: distinct seasonal and spatial characteristics and ozone sensitivity[J]. *Atmos. Chem. Phys.* 22 (13), 8935–8949. <https://doi.org/10.5194/acp-22-8935-2022>.
- Wei, J., et al., 2021. Full-coverage mapping and spatiotemporal variations of ground-level ozone (O₃) pollution from 2013 to 2020 across China[J]. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2021.112775>.
- Weng, X., et al., 2022. A machine learning approach to quantify meteorological drivers of ozone pollution in China from 2015 to 2019[J]. *Atmos. Chem. Phys.* 22 (12), 8385–8402. <https://doi.org/10.5194/acp-22-8385-2022>.
- Xia, Y., et al., 2022. Concurrent hot extremes and high ultraviolet radiation in summer over the Yangtze Plain and their possible impact on surface ozone[J]. *Environ. Res. Lett.* 17 (6) <https://doi.org/10.1088/1748-9326/ac6c3c>.
- Xiong, K., et al., 2022. Improving the accuracy of O₃(3) prediction from a chemical transport model with a random forest model in the Yangtze River Delta region, China [J]. *Environ. Pollut.* 120926 <https://doi.org/10.1016/j.envpol.2022.120926>.
- Yafouz, A., et al., 2021. Ozone Concentration forecasting based on Artificial Intelligence Techniques: a Systematic Review[J]. *Water Air Soil Pollut.* 232 (2) <https://doi.org/10.1007/s11270-021-04989-5>.
- Zhan, C., et al., 2021. Surface ozone in the Yangtze River Delta, China: a Synthesis of Basic Features, Meteorological Driving Factors, and Health Impacts[J]. *J. Geophys. Res. Atmos.* 126 (6) <https://doi.org/10.1029/2020JD033600> e2020JD033600.
- Zhang, B., et al., 2022. Deep learning for air pollutant concentration prediction: a review [J]. *Atmos. Environ.* 290, 119347 <https://doi.org/10.1016/j.atmosenv.2022.119347>.
- Zheng, B., et al., 2018. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions[J]. *Atmos. Chem. Phys.* 18 (19), 14095–14111. <https://doi.org/10.5194/acp-18-14095-2018>.