RESEARCH ARTICLE

Elucidate long-term changes of ozone in Shanghai based on an integrated machine learning method

Jin Xue[1,2], Fangting Wang[1,2], Kun Zhang[1,2], Hehe Zhai[1,2], Dan Jin[3], Yusen Duan (✉)[3], Elly Yaluk[1,2], Yangjun Wang[1,2], Ling Huang[1,2], Yuewu Li[3], Thomas Lei[4], Qingyan Fu[3], Joshua S. Fu[5], Li Li (✉)[1,2]

1 School of Environmental and Chemical Engineering, Shanghai University, Shanghai 200444, China
2 Key Laboratory of Organic Compound Pollution Control Engineering (MOE), Shanghai University, Shanghai 200444, China
3 Shanghai Environmental Monitoring Center, Shanghai 200235, China
4 Institute of Science and Environment, University of Saint Joseph, Macao 999078, China
5 Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA

HIGHLIGHTS

• A novel integrated machine learning method to analyze O_3 changes is proposed.
• Various factors affecting long-term changes of O_3 in Shanghai are quantified.
• Meteorological, photochemical, and regional background O_3 are well separated.

GRAPHIC ABSTRACT

ABSTRACT

Surface ozone (O_3) is influenced by regional background and local photochemical formation under favorable meteorological conditions. Understanding the contribution of these factors to changes in O_3 is crucial to address the issue of O_3 pollution. In this study, we propose a novel integrated method that combines random forest, principal component analysis, and Shapley additive explanations to distinguish observed O_3 into meteorologically affected ozone ($O_{3\ MET}$), chemically formed from local emissions ($O_{3\ LC}$), and regional background ozone ($O_{3\ RBG}$). Applied to three typical stations in Shanghai during the warm season from 2013 to 2021, the results indicate that $O_{3\ RBG}$ in Shanghai was 48.8 ± 0.3 ppb, accounting for 79.6%–89.4% at different sites, with an overall declining trend of 0.018 ppb/yr. $O_{3\ LC}$ at urban and regional sites ranged from 5.9–9.0 ppb and 8.9–14.6 ppb, respectively, which were significantly higher than the contributions of 2.5–7.4 ppb at an upwind background site. $O_{3\ MET}$ can be categorized into those affecting O_3 photochemical generation and those changing O_3 dispersion conditions, with absolute contributions to O_3 ranging from 13.4–19.0 ppb and 13.1–13.7 ppb, respectively. We found that the O_3 rebound in 2017, compared to 2013, was primarily influenced by unfavorable O_3 dispersion conditions and unbalanced emission reductions; while the O_3 decline in 2021, compared to 2017, was primarily influenced by overall favorable meteorological conditions and further emissions reduction. These findings highlight the challenge of understanding O_3 change due to meteorology and regional background, emphasizing the need for systematic interpretation of the different components of O_3.

✉ Corresponding authors
duanys@sheemc.cn (Y. Duan); lily@shu.edu.cn (L. Li)

1 Introduction

Surface ozone (O_3) exerts significant impacts on air quality, climate, human health and ecosystems, displaying

deleterious effects (Lelieveld et al., 2015). As a secondary air pollutant, O_3 is formed via photochemical reactions among precursors under the presence of sunlight. Hence, meteorological variations play a vital role in the short- and long-term changes in O_3 (Wang et al., 2017). A quantitative analysis of the effects of meteorology and emissions is necessary to comprehend the fundamental relationship between O_3 and meteorology. Notably, when the effects of meteorological fluctuations are removed, variations in O_3 concentrations are governed by local photochemical reactions and background O_3. The former from local emissions, whereas the latter is primarily formed from out-of-region emissions. Therefore, determining the influence of meteorology, emissions, and their specific factors on O_3 is fundamental to comprehend O_3 trends and evaluate the impact of emission control strategies.

Various methods are employed to estimate the contributions of meteorology and emissions to O_3, including chemical transport models (CTMs), statistical methods, and machine learning (ML). CTMs involve manipulating emission inventories and meteorological fields to characterize the impact of emissions and meteorology on air quality, as demonstrated by previous studies (Li et al., 2019; Xu et al., 2019; Liu and Wang, 2020a, 2020b; Dang et al., 2021). However, CTMs methods have two main drawbacks. First, the computational resource requirements of numerical models, as well as the difficulty of ensuring the accuracy of emission inventories and meteorological fields, can pose challenges in terms of implementation and result in significant uncertainties. Second, determining the contribution of emissions and meteorology to O_3 requires extensive simulation scenarios that model the differences in O_3 concentrations resulting from interannual changes in emission inventories and meteorological fields across different years. This process can be arduous and time-consuming. Furthermore, this approach only allows for the overall contribution of meteorological conditions to O_3, hindering our understanding of the impact of individual meteorological factors on O_3. In contrast, statistical methods establish relationships between meteorological or emission data and air quality observations using multiple linear regression, empirical orthogonal functions, or Kolmogorov–Zurbenko (KZ) filtering (Camalier et al., 2007; Henneman et al., 2015; Yang et al., 2019; Li et al., 2020; Gao et al., 2021; Hu et al., 2021). These methods usually aim to fit or strip single influencing factors using mathematical equations. However, due to the nonlinearity between O_3 and its precursors, accurately capturing these nonlinear relationships can be difficult using statistics. Recently, the rapid rise of big data technology has led to the widespread application of ML methods for separating the effects of meteorology and emissions on air quality (Grange et al., 2018; Grange and Carslaw, 2019; Vu et al., 2019; Qu et al., 2020; Wang et al., 2020; Zhang et al., 2020; Dai et al., 2021; Lovrić et al., 2021; Shi et al., 2021; González-Pardo et al., 2022; Hou et al., 2022; Tang et al., 2022; Wu et al., 2022; Zhou et al., 2022; Ding et al., 2023). For example, Grange et al. (2018) employed a random forest model to examine the trend of emissions on PM_{10} in Switzerland from 1997 to 2016. Vu et al. (2019) evaluated the impact of a 5-year clean air action plan using the meteorological standardization technique. Qu et al. (2020) assessed the effectiveness of clean air policies on $PM_{2.5}$ reduction in China's "2+26" region using reinforcement tree models to remove meteorological effects. These ML methods have been demonstrated to effectively separate the effects of weather and emissions on air quality.

However, most of the existing research focuses on the overall impact of emissions or meteorology on O_3, with less emphasis on systematically quantifying the impact of specific factors of emissions, meteorology, and background on O_3. Recently, Lundberg et al. (2020) proposed a SHapley Additive exPlanation (SHAP) approach based on game-theoretic approach to explain ML models, which overcomes the black-box nature of ML methods and makes it easier to understand the attribution process of ML. The SHAP approach assigns predictions to each feature value, providing a clearer understanding of the degree of influence on the research target and quantifying the influence of each meteorological factor on O_3. In addition to meteorological factors, it is also important to differentiate O_3 to local formation and regional background considering local and regional scale pollution control. Based on the characteristics of regional background O_3, the dimensionality reduction property of principal component analysis (PCA) method can effectively isolate the regional background O_3 from the observed data (Suciu et al., 2017).

Here, we proposed a novel integrated method (IM) to quantify meteorological, emissions, and regional background contributions to changes of O_3 by integrating random forest (RF), PCA and SHAP. We applied the method to investigate changes of O_3 in Shanghai during the warm season (April–September) from year 2013 to 2021. During this period, we effectively stripped off the meteorological and chemical contribution to O_3, assigned the meteorological contribution to each meteorological factor, and decomposed the chemically generated O_3 into local emission and regional background. Eventually, we systematically quantified the influence of each specific factor on O_3. Our results demonstrate the reliability and provide strong scientific support for O_3 control.

2 Methodology

2.1 Data

For this study, three air quality monitoring stations located in Shanghai, China were selected. These stations

have continuous hourly observational data but diverse features. Pudong Huinan (PDHN, 31°3′N, 121°47′E), is an up-wind background site located in the eastern area of Shanghai and adjacent to the ocean, mainly influenced by regional transport. Putuo (PT, 31°14′N, 121°24′E) is an urban-center site strongly influenced by urban anthropogenic emissions such as vehicle exhaust, industries, and residential emissions, etc. Dianshan Lake (DSL, 31°5′N, 120°58′E) is a regional site located at the joint region of Jiangsu, Zhejiang provinces and Shanghai where O_3 accumulation occurs frequently. The reasons for selecting these three sites from multiple observational sites are detailed in Text S1, Figs. S1 and S2. Hourly observational data of air pollutants (NO_x and O_3) from 2013 to 2021 were collected from these three sites. O_3 concentration was monitored by an O_3 analyzer (Model 49i, Thermo Scientific, USA), and NO_x concentration was monitored by an NO_x analyzer (Model 42i, Thermo Scientific, USA). Outlier removal and data interpolation padding were performed on these data without changing the original data sample to meet the requirements of machine learning model training (Text S2).

Hourly reanalysis meteorological data for the years 2013–2021 were downloaded from the European Centre for Medium-Range Weather Forecasts (ECMWF). These data include 10 m u-component of wind (U10), 10 m v-component of wind (V10), 2 m dewpoint temperature (DT), 2 m air temperature (T), surface pressure (SP), surface net solar radiation (SSR), and total precipitation (TP) from ERA5-land with a spatial resolution of 0.1° × 0.1°. In addition, boundary layer height (BLH) and total cloud cover (TCC) from ERA5 were downloaded with a spatial resolution of 0.25° × 0.25°. To match the weather around the three sites above, all these meteorological variables were extracted from the grids covering these three stations in the gridded ERA5/ERA5-land data set. A comprehensive comparison between the ERA5/ERA5-land reanalysis data with surface observations (using temperature, dew point temperature, wind speed and, wind direction, at Pudong Airport and Hongqiao Airport) to verify the data reliability is presented and documented in detail in Text S3, Figs. S3 and S4.

The 72 h backward air mass trajectories arriving at an altitude of 100 m above ground level (a.g.l.) at PT, DSL and PDHN were calculated using the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT 5.0.0) model (Stein et al., 2015). The HYSPLIT was driven by reanalysis meteorological data from the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR). We conducted trajectory clustering on the results generated by HYSPLIT, as depicted in Fig S5.

2.2 Trend analysis

The Mann-Kendall (MK) trend analysis is a useful tool for detecting monotonic trends in time series data, which is widely employed in atmospheric sciences research (Hirsch et al., 1982; Pathakoti et al., 2021; Zhang et al., 2022). Compared to ordinary least squares (OLS), it is particularly appropriate when the data distribution is ambiguous or when the sample size is small. This is because the MK trend analysis does not make the assumption that error terms in the regression model are identically and independently distributed, which can lead to biased results. As a result, the MK trend analysis provides a more robust and reliable method for trend analysis in these scenarios. In this study, we employed a seasonal MK test method to determine the trend of O_3 while mitigating the influence of seasonal changes. A detailed explanation of the method can be found in the study of Hirsch et al. (1982). When the trend is statistically significant ($p < 0.05$), a positive S-value indicates that the ozone is increasing (decreasing), and the magnitude of $|S|$ is used to indicate the rate of increase (decrease) in the trend.

Furthermore, we also utilized the Theil-Sen estimator to examine the rate of O_3 increase/decrease. This method was proposed by Sen as a robust nonparametric statistical method for trend calculation (Sen, 1968). The method is impervious to measurement errors and outlier data, and it is suitable for long-term trend analysis. A detailed description of this method can be found in Zhang et al. (2022).

2.3 Random Forest (RF)-Based meteorological normalization

The RF model is a widely used ML model known for its high accuracy and stability. In this study, we employed RF to predict O_3 using time and meteorological variables. Specifically, the time variables, including Unix time, day of year, day of week, and hour, where used to represent the long-term, annual, weekly, and daily patterns of pollutant emissions, respectively. The meteorological variables used in the RF model included wind speed (WS, calculated from U10 and V10), wind direction (WD, calculated from U10 and V10), T, relative humidity (RH, calculated from T and DT, Alduchov and Eskridge, 1996), SSR, SP, TP, BLH, TCC, length of each air mass trajectory (Length) and cluster of each air mass trajectory (Cluster), to comprehensively consider the impact of meteorology on O_3. The RF model was implemented in a Python environment, and underwent hyper-parameter selection to obtain the optimal model performance. The details of specific steps are described in Text S4, Figs. S6, and S7. The resulting RF model achieved an R^2 value ranging between 0.83 and 0.85, which is acceptable compared to previous studies (Vu et al., 2019; Huang et al., 2022; Li et al., 2022; Lu et al., 2022), and the other model performance characteristics are presented in Table S1.

The meteorological normalization method used in this study is similar to those utilized in previous studies (Grange et al., 2018; Wang et al., 2020; Zhang et al., 2020; Shi et al., 2021). This method predicts O_3 trends under historical average meteorological conditions. Specifically, we predict ozone by randomly sampling meteorological parameters from the historical meteorological data set without changing Unix time (the long-term pattern of pollutant emissions) to create a new data set for O_3 prediction. By making 1000 predictions in this manner, the predicted 1000 O_3 time series are averaged, resulting in a weather-normalized O_3 variation driven only by emissions, i.e., the changes in O_3 concentrations independent of meteorological effects.

2.4 Principal component analysis (PCA)-based regional background O_3 estimation

PCA is a widely utilized method for dimensionality reduction in atmospheric environmental studies (Wren et al., 2020; Yang et al., 2020; Austin et al., 2021). Its principle lies in reducing high-dimensional data to a lower-dimensional space while retaining maximum information. This procedure facilitates the extraction of the main feature components of data, and it enables obtaining principal components that have the most significant impact on the data, thus enabling further data analysis and other subsequent operations (Joliffe and Morgan, 1992; Zhu et al., 2021).

In this study, PCA was run and two components with eigenvalues greater than 1 were retained. Based on the relationship between component loadings and variables, we identified them as regional background ozone and others. PCA transforms data linearly, therefore, it is possible to calculate O_{3_RBG} from the linear relationship in PCA. We referred to previous studies' calculation of background concentration (Langford et al., 2009; Berlin et al., 2013; Suciu et al., 2017; Wang et al., 2022a) and input daytime (10:00–17:00) data, including O_3, NO_x, WS, WD and T, into the PCA model, to determine the corresponding components of O_{3_RBG} according to the scoring coefficient matrix. The detailed calculation steps are presented in Text S5. Using the results of PCA, we calculated the background concentration of O_3 by utilizing the average and standard deviation of O_3 observations, the score, and the variance contributions of corresponding components (see Tables S2 and S3).

2.5 SHAP approach-based causal attribution

An explanatory model based on cooperative game theory, namely the SHAP approach (Lundberg and Lee, 2017; Song et al., 2022), is employed to attribute contributions from specific meteorological parameters to O_3. This approach performs an algorithm on the ML model and the input features, resulting SHAP values that quantify the impacts of each feature on the individual prediction results. The prediction results of the ML algorithm are the sum of the attributed values of each input feature (SHAP value). Therefore, using a ML model for predicting ozone (e.g., the RF model) enables calculation of the contribution of the input data (e.g., RH at 10:00 on July 1, 2013) to the predicted O_3. The core algorithm is as follows (Eqs. (1) and (2)):

$$f(x) = \Phi_0(f) + \sum_{i=1}^{M} \Phi_i(f,x), \qquad (1)$$

where f is the original model, $\phi_0(f)$ is the baseline value of the model, which is the average of the model prediction results, and M is the number of model features. $\phi_i(f,x)$ is the SHAP value of feature i, and it is the contribution of feature i to the prediction result $f(x)$. When $\phi_i(f,x) > 0$ ($\phi_i(f,x) < 0$), it indicates that feature i will increase (decrease) $f(x)$, and the absolute value of $\phi_i(f,x)$ represents the importance of the feature i.

$$\Phi_i(f,x) = \sum_{R \in \mathcal{R}} \frac{1}{M} [f_x(P_i^R \cup i) - f_x(P_i^R)], \qquad (2)$$

where $\mathcal{R}$ is the set of all feature orderings, and P_i^R is the set of all features that come before feature i in ordering $\mathcal{R}$.

The SHAP value of each meteorological factor, including T, RH, SSR, WS, WD, BLH, TCC, SP, TP, Cluster, and Length, was determined based on the trained RF model and the SHAP method. The contribution of each meteorological component to O_3 was further analyzed.

2.6 Integrated method

We propose an integrated method for identifying the contributions of meteorological, chemical / emissions and regional background factors to changes in O_3, which includes three main steps. The overall framework of this IM is presented in Fig. 1.

Step 1 (meteorology affected O_3). In this step, the meteorological normalization model is utilized to generate the time series of O_3 after meteorological normalization (denoted as "O_{3_CHEM}"). Then, the meteorologically normalized O_3 is subtracted from the observed O_3 to obtain the meteorological contribution to O_3 (denoted as "O_{3_MET}"). Next, the meteorological factors and the previously obtained meteorological contribution are used as the independent and dependent variables, respectively, in an RF model for training. Finally, utilizing the SHAP method, the SHAP values of each meteorological factor are calculated based on the trained RF model and the meteorological factors as input data. This step yields both the total meteorological contribution and the contribution of specific meteorological parameters.

Step 2 (regional background O_3). The typical PCA
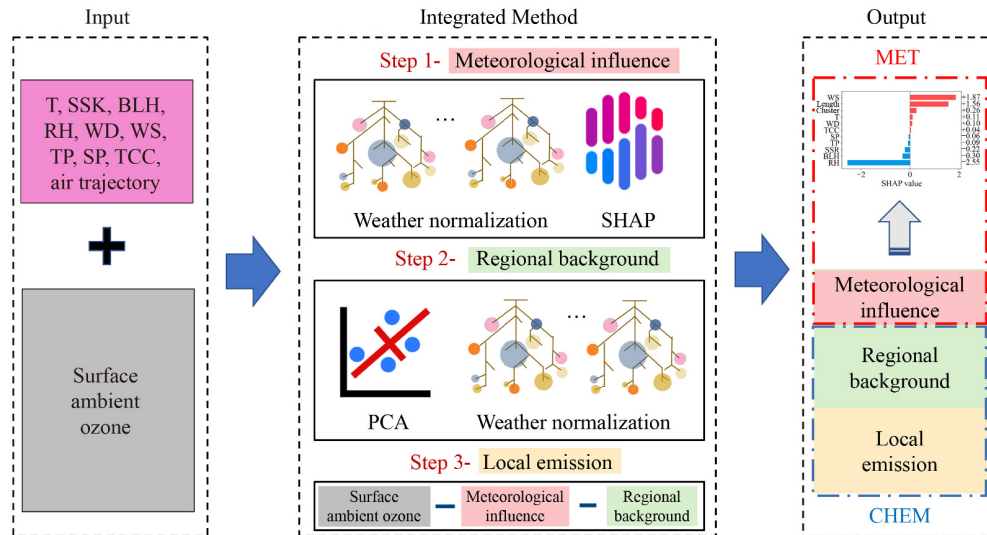
Fig. 1   Framework of the Integrated Method.

approach for determining O_3 is influenced by meteorology. To eliminate the meteorological influence and obtain the final meteorologically unaffected regional background O_3 (denoted as "O_{3_RBG}"), we first compute the regional background O_3 using the PCA model and then remove the meteorological impact using the meteorological normalization model from Step 1.

Step 3 (local formation O_3). Surface O_3 consists of the meteorological contribution, regional background O_3, and O_3 generated from local emissions via photochemical reactions. In this step, we subtract the meteorological contribution ("O_{3_MET}") and regional background ozone ("O_{3_RBG}") from the observed ozone to obtain the ozone generated from local emissions (denoted as "O_{3_LC}").

It should be noted that the relationship between O_3 and its precursors is nonlinear and separating local, meteorological, and regional background factors from the perspective of precursor emissions would be complicated. However, our approach bypasses the complexity of the nonlinear relationship between O_3 and its precursors by separating directly from the observed O_3, making the problem easier to resolve. Ultimately, this integrated method provides a systematic and quantitative separation of the observed O_3 and clarifies the specific effects of meteorological factors, regional background, and local emissions on O_3.

3   Results and discussion

3.1   Trend of O_3 pollution

The long-term trends of O_3 and nitrogen dioxide (NO_2) at each site in Shanghai are presented in Fig. 2. The concentration of NO_2 displayed a distinct pattern, with concentrations in descending order at PT, DSL, and PDHN. From trend analysis, both PT and DSL showed a decreasing trend in NO_2, while PDHN did not show a clear trend. The decrease in NO_2 at PT and DSL can be attributed to the implementation of NO_x emission control strategies in Shanghai and the Yangtze River Delta (YRD) region, respectively; while the low NO_2 concentration at PDHN, which is located near the ocean and has a sparse population, may be due to the background. Although significant differences in NO_2 concentrations were observed among the three stations, it was difficult to observe a clear trend in O_3 concentrations over the long-term. Analyzing the diurnal variations of O_3 at each site can provide a more reflective assessment of the O_3 formation regime at the three stations. PT exhibited low daytime O_3 concentration, significant diurnal variation, and sustained high NO_2 concentration, indicating strong local generation of O_3 and influence by NO titration. DSL had the highest O_3 concentration during the day among the three sites, but NO_2 was not prominent, and the diurnal variation of both NO_2 and ozone was not significant, indicating that local photochemistry had a small contribution to ozone formation, and NO_x is relatively less limiting to O_3. The daytime NO_2 concentration at PDHN was much lower than the other two sites, the O_3 concentration was comparable to that at PT, indicating that the regional transport had a more significant impact on O_3 around PDHN.

O_3 is primarily dominated by meteorology and photochemistry, therefore, meteorological conditions are crucial for O_3 formation. We selected meteorological conditions for the warm season from April to September at 10–17 for analysis, as shown in Figs. 3 and 4. From the wind rose diagram, although the dominant wind during the warm season of PT, DSL and PDHN is south-east (SE), the wind aggregation and wind speed gradually
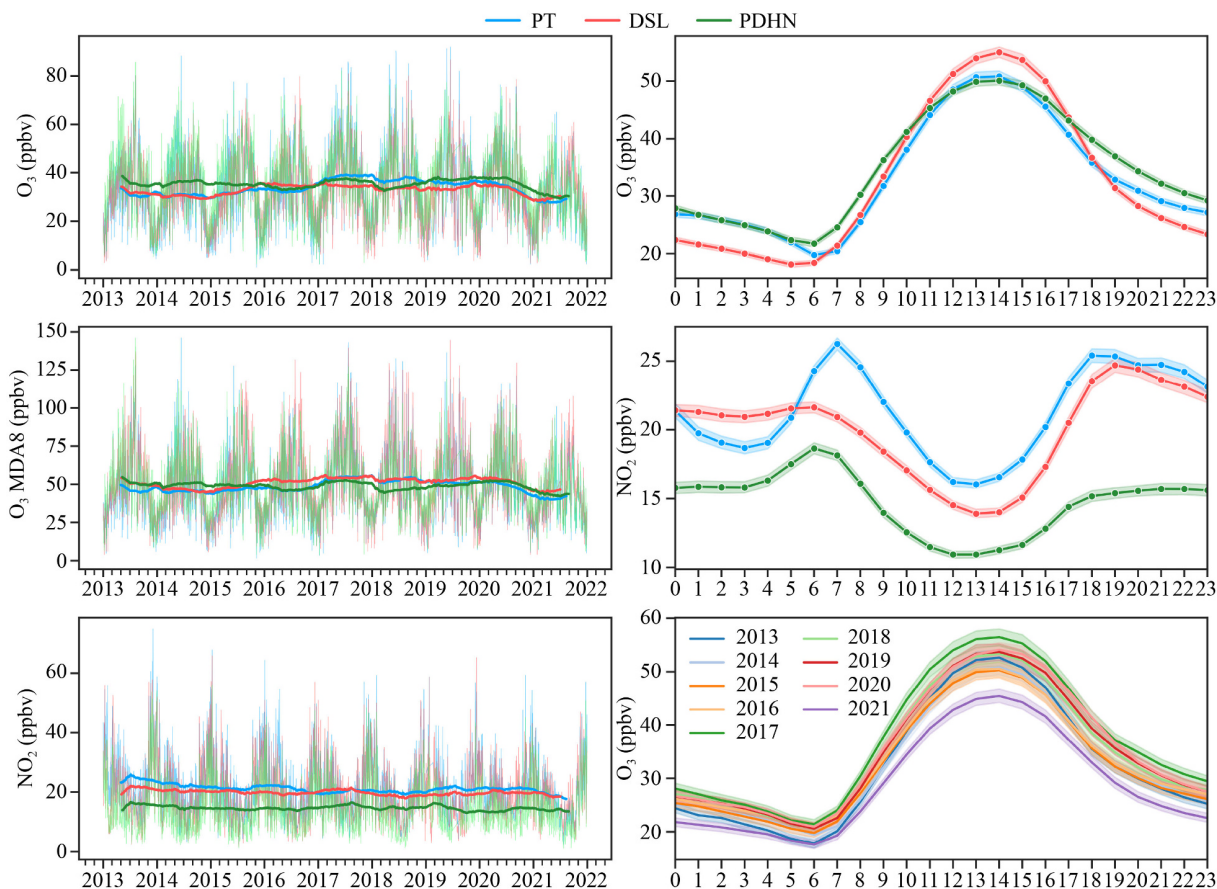
Fig. 2 Time series of O_3, maximum daily 8 hour average ozone (MDA8 O_3) and NO_2 at PT (blue), DSL (red), PDHN (green). In the left figure, the thick line is the 360-day sliding average; the thin line is the daily average. In the right figure; the shaded area is the 95% confidence interval of corresponding data.

decrease from ocean to inland (Fig. 3). Therefore, PDHN is more susceptible to one-way wind, while the wind direction at DSL is more scattered. RH, T and SSR are considered the most important meteorological conditions that determine the O_3 photochemical processes. The analysis shows that RH at PDHN is significantly higher than the other two sites; T is significantly lower than the other two sites; and SSR is between PT and DSL, which may be related to the proximity of the PDHN site to the ocean. For the two inland sites, PT and DSL, the differences in RH and T are small, but DSL's SSR is significantly higher than PT. Regarding other parameters, PDHN's TP and TCC are significantly smaller than PT and DSL, and the difference in SP among the three sites is small. The BLH of DSL, PT, and PDHN decreases in turn. Overall, the dominant wind direction of PT is southeast, solar radiation is relatively small, and other meteorological conditions are mainly between the two stations. The wind direction of the DSL station is relatively scattered, with low RH, high T and high SSR, indicating that it is more favorable for ozone generation. The meteorological parameters at PDHN are significantly influenced by the ocean, leading to distinct variations in

comparison to the other two stations. This influence is evident in the concentrated wind direction, high RH, and low temperature at PDHN. These factors contribute to the station's unique atmospheric conditions, setting it apart from the others.

3.2 Meteorological impact

The meteorological contributions to O_3 at various sites between 2013 and 2021 have been determined through the integrated method, as presented in Fig. 5, Text S6 and Figs. S8–S15. Meteorological conditions were found to have a positive impact on O_3 formation in 2013 (0.78–3.55 ppb) and 2017 (3.95–9.13 ppb), but unfavorable for O_3 formation in 2015 (−2.13 to −0.85 ppb), 2016 (−3.81 to −0.15 ppb) and 2021(−8.86 to −4.41 ppb) at PT, DSL and PDHN stations, as depicted in Fig. 5. The remaining years displayed distinct features at each site. Although O_{3_MET} displayed similar trends at the three locations in certain years, it varied significantly across most years due to their geographical locations, as illustrated in Fig. 5(a). Generally, the weather system affects O_3 in two ways: by altering T, RH, SSR, TCC and
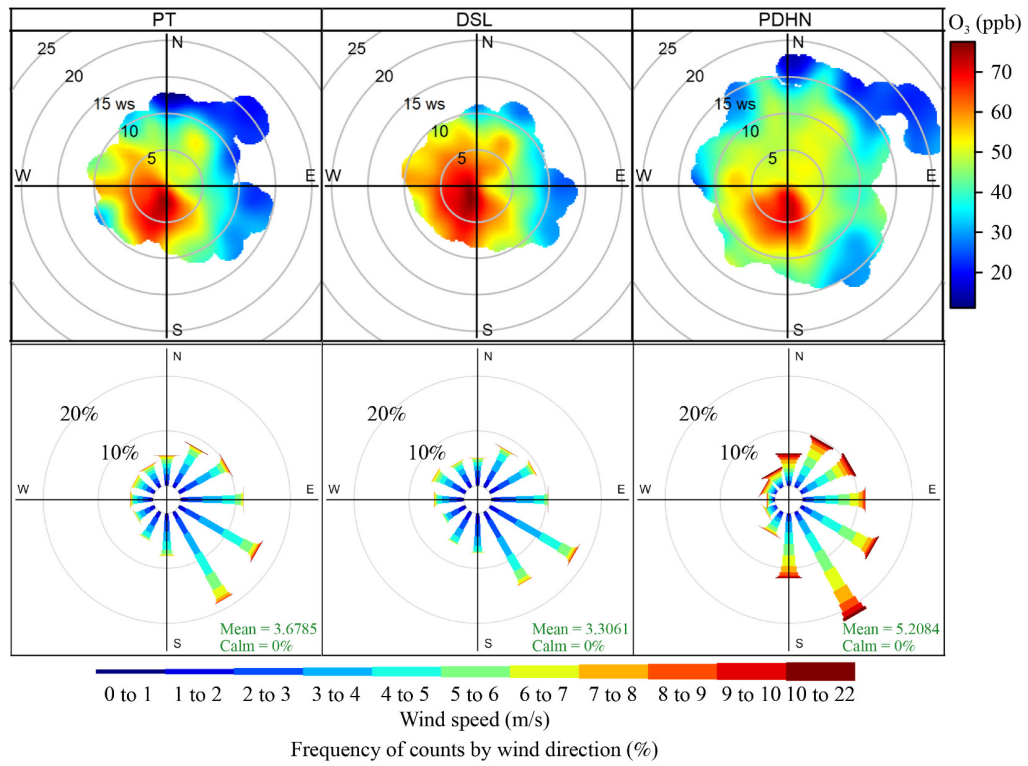
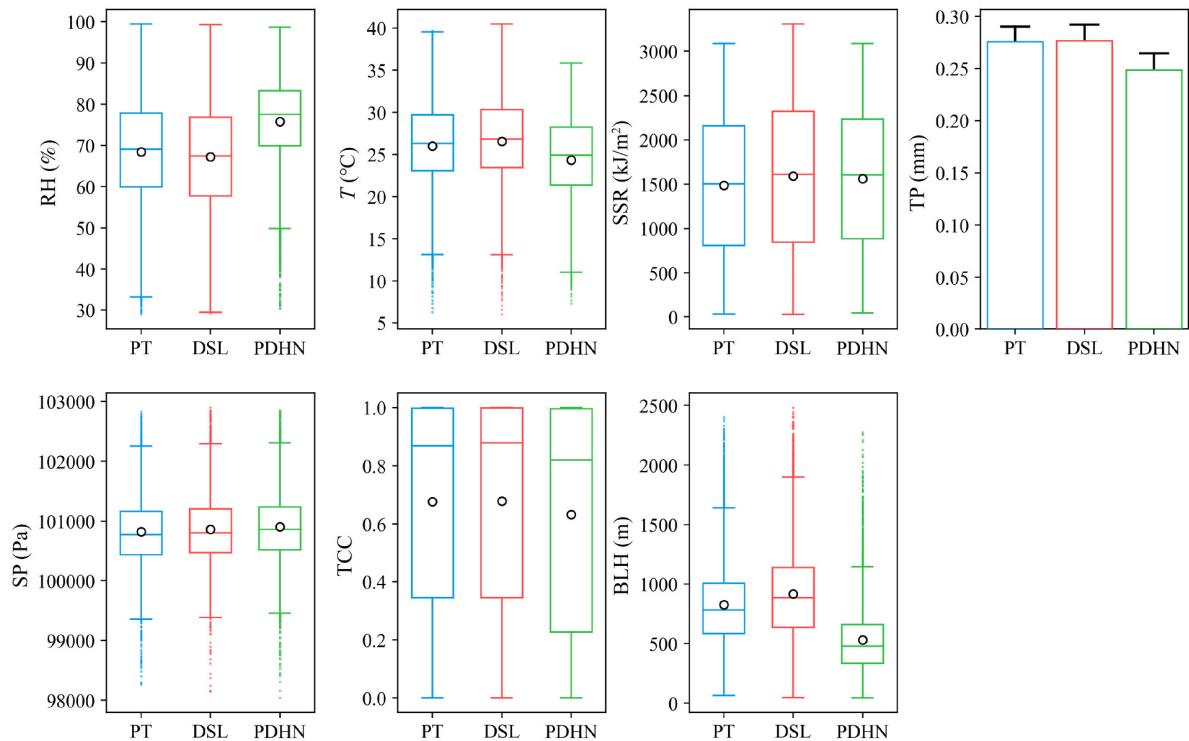**Fig. 3**　Bivariate polar plot and wind rose at PT, DSL and PDHN.

Fig. 4　Comparisons of the major meteorological factors among the three observational sites.

TP to influence O_3 photochemical generation (OPG), and by changing air mass trajectories, wind fields, BLH and SP to affect O_3 dispersion conditions (ODC). The cumulative absolute contributions to ozone ranged from 13.44–19.03 ppb and 13.18–13.66 ppb, respectively, as detailed in Table S4.
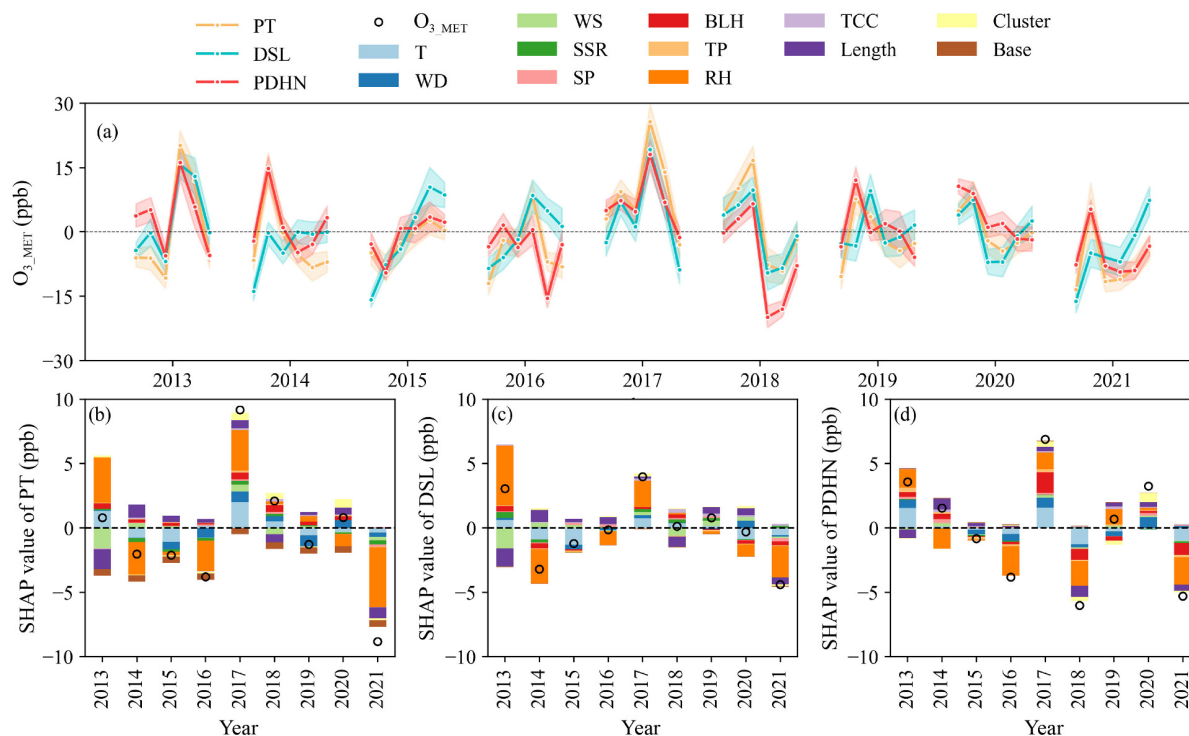
Fig. 5   Meteorological contributions to O_3 at PT, DSL and PDHN. The bands beside the lines are the 95% confidence intervals in panel (a). Base is the base value of the SHAP approach, i.e. the average of the SHAP value in panels (b)–(d).

The overall trend of meteorological impact on O_3 is primarily through influencing OPG. For example, all PT, DSL and PDHN sites indicated favorable meteorological influences on O_3 photochemical formation in 2013 and 2017, as opposed to the unfavorable impact in 2021. Both RH and T contributed significantly to meteorological ozone, as illustrated in Figs. 5(b)–(d). In comparison, RH and T had a greater impact on O_3 levels at PT and DSL stations compared to PDHN stations. This discrepancy is likely due to the location of the PDHN station near the ocean, where RH and T are relatively stable, and anthropogenic activity levels are lower. In contrast, the higher levels of anthropogenic activities around PT and DSL stations, combined with their inland location, made them more susceptible to changes in RH and T that could affect the rate of ozone generation from NO_x and volatile organic compounds (VOCs) emissions. Overall, our findings suggest that the impact of RH and T on O_3 levels is more significant at stations with high levels of anthropogenic activities, particularly in inland areas. Furthermore, this method effectively reflects the O_3 formation characteristics at different stations.

The impact of ODC at PT, DSL and PDHN varied significantly due to the different geographical locations of the stations. At the PT station, the cumulative absolute contribution of WD, Length, WS, and Cluster could reach 11.60 ppb (Table S4), signifying a clear influence of horizontal transport on O_3. The concentration of O_3 at PT was affected by the arrival of inland air mass, clean air

mass from the ocean, and changes in wind direction (Fig. S13). At the DSL site, WS had an average absolute contribution of 4.85 ppb, which was significantly higher than the PT and PDHN stations (3.10 ppb and 0.69 ppb, respectively). The reason for this was that the DSL site was located at the conjunction area of the YRD region, where O_3 and its precursors from the three regions tended to converge around the site. As a result, the dilution effect and the diffusion effect of wind speed caused significant changes in O_3 (Fig. S13). The cumulative average absolute contribution of all meteorological parameters characterizing O_3 transport at the PDHN site was 12.51 ppb, with BLH accounting for 43.0%. This indicated that the effects of transport, especially vertical transport, had a significant impact on the local variability of O_3 at PDHN. As an up-wind background station, it has little anthropogenic emissions, resulting less *in situ* chemical reactions producing O_3. However, the low concentration of NO_x made it difficult to titrate off the O_3 transported from outside the station. Figure S14 showed a clear difference between the dependence plot of BLH at PDHN and other stations. The SHAP value of BLH at PDHN station decreased with the increase of T. This was probably because the increase of T strengthened the convective intensity within the troposphere, causing the BLH to increase and further strengthening the O_3 diffusion ability at PDHN station. This was consistent with the study of Tang et al. (2021) and a more detailed explanation could be found in Text S6.

The results indicated that factors such as RH, T, and SSR, which affect OPG, determined the trend of the O_{3_MET}, whereas the influence of ODC on O_3 fully reflected the respective characteristics of PT, DSL, and PDHN sites. The meteorological impacts analyzed in this study were comparable to those reported by Lin et al. (2021). Meanwhile, Liu and Wang (2020a) revealed that weather conditions reduced ozone generation during 2014–2016, which was consistent with the meteorological findings at PT and DSL sites. Furthermore, Li et al. (2021b) investigated the daytime O_3 pollution process in Shanghai and found that the contribution of downward O_3 transport in the lower free troposphere at offshore sites could reach 49.1%. This indicated that BLH had a significant influence on ozone in PDHN, which was consistent with the results obtained in this study.

3.3   Regional background contribution

The concentration of O_{3_RBG} was estimated to be 48.8 ± 0.3 ppb, with a declining trend of −0.018 ppb/yr (p = 0.02), as presented in Fig. 6. This trend could be attributed to the stringent pollution control measures implemented in China between 2013 and 2020, which

primarily targeted the reduction of NO_x and VOCs, the precursors of O_3. The reduction rate of NO_x was higher than that of VOCs, which could be disadvantageous for O_3 mitigation in urban areas but beneficial in rural and regional areas (Li et al., 2019). Thus, the minor decline in O_{3_RBG} could be explained by the decrease in O_3 in rural regions and the complexity of the regional background O_3 components. The proportion of O_{3_RBG} in O_3 varied among the PT, DSL, and PDHN sites, with values of 86.4% (84.0% to 90.5%), 79.6% (76.7% to 85.4%), and 89.4% (86.6% to 96.0%), respectively. The contribution of O_{3_RBG} at PT station increased and then decreased, indicating a change in the local O_3 concentration in the opposite direction. The O_{3_RBG} at DSL was lower than that at PT station, which could be attributed to favorable photochemical conditions at the DSL station. The contribution of O_{3_RBG} at PDHN station had the highest percentage among the three stations, which was reasonable considering its regional location.

Our findings are overall consistent with previous studies. For example, Sahu et al. (2021) used the CMAQv5.2 model and found that the policy relevant background concentration of O_3 in the Eastern China region accounted for 79% of the total O_3, primarily
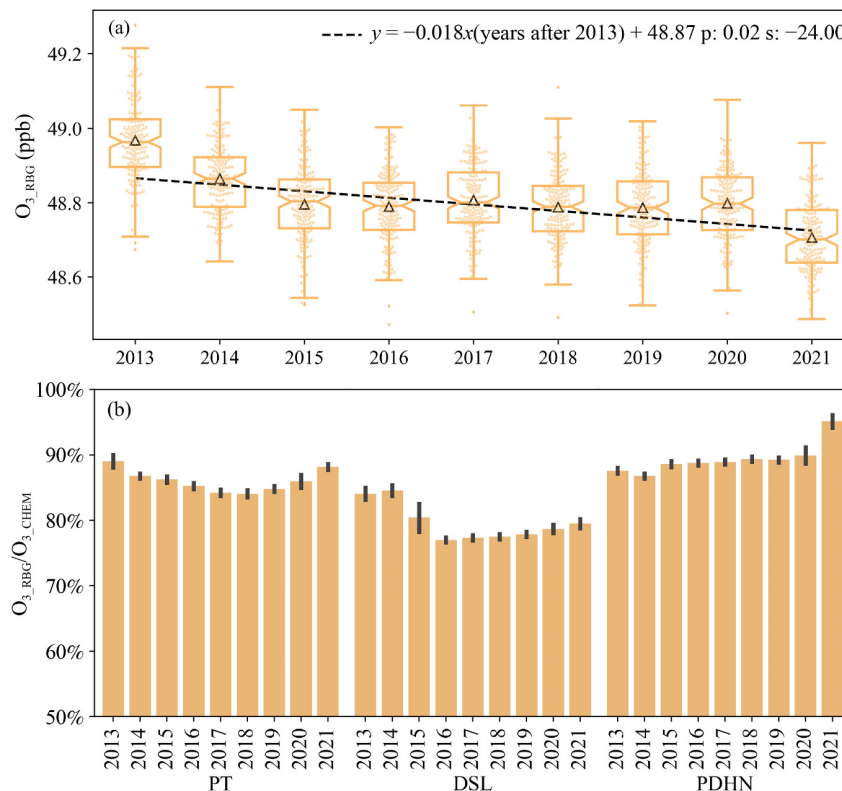

Fig. 6 Contributions of O_{3_RBG} at PT, DSL and PDHN. In panel (a), the upper and lower boundaries of the box represent the 25th percentile and 75th percentile, and the notches represent the confidence interval around the median. The line and hollow triangles inside boxes represent the average and median values of corresponding data, respectively. The error lines above and below the box represent the 25th percentile value minus 1.5 * IQR (interquartile range) and the 75th percentile value plus 1.5 * IQR (interquartile range), respectively. In panel (b), the error line indicates the 95% confidence interval of the corresponding data.

originating from South-east Asia and regions outside of Asia. Lu et al. (2019) conducted a study using the GEOS-Chem model and found that the average background concentration of O_3 for March to October in the YRD region was about 41 ppb, with biogenic volatile organic compounds (BVOCs) being the primary contributors from natural sources. Ni et al. (2018) used the GEOS-Chem model to investigate the contribution of domestic and foreign O_3 to China during the spring season and found that the background concentration of O_3 in the Shanghai region was approximately 40–45 ppb. However, it is worth noting that the regional background O_3 obtained from the aforementioned model studies was derived by simulating the shutdown of anthropogenic emissions solely within China. This approach failed to account for the contribution of O_3 pollution transport from other regions to Eastern China and the YRD region. As a result, our findings indicate a higher regional background concentration than those reported in these studies. In addition to modeling approaches, some statistical methods have also been used to calculate the regional background concentration of O_3. Chen et al. (2022) established a relationship between surface O_3 and temperature and obtained a regional background concentration of O_3 in the YRD from 2013 to 2019 of 53 ± 13 ppb. Suciu et al. (2017) used the same PCA method as our study to analyze the Houston-Galveston-Brazoria (TX) region for regional background O_3, resulting in a value of 46.72 ± 2.08 ppb. Similarly, Wang et al. (2022b) used a similar PCA method to calculate regional background O_3 concentrations of 51.0–52.1 ppb in Shandong, China. Overall, our results regarding the regional background O_3 in Shanghai are reasonable.

3.4 Local emission contribution

Figure 7 illustrates the variation of O_{3_LC} at PT, DSL, and PDHN. O_{3_LC} shows significant differences among these stations. Specifically, DSL has the highest local chemical contribution, ranging from 8.9–14.6 ppb (15.4%–23.0%), followed by PT and PDHN stations with concentrations of 5.9–9.0 ppb (10.7%–15.6%) and 2.5–7.4 ppb (4.6%–13.2%), respectively. Interestingly, O_{3_LC} began to increase at the urban PT site at a rate of 0.6 ppb/yr since 2013, reaching 9.2 ppb in 2018. In contrast, NO_2 decreased overall at a rate of 0.91 ppb/yr at PT during 2013–2018 (see Fig. S16). O_{3_LC} then decreased at a rate of 0.8 ppb/yr to 6.5 ppb during 2018–2021, while NO_2 continued to decrease at a rate of about 0.47 ppb/yr. This pattern of an increase followed by a decline in O_{3_LC}, while NO_2 levels are generally decreasing, may reveal the changes of photochemical regime related to O_3 sensitivity. It is worth mentioning that China implemented the Air Pollution Prevention and Control Action Plan (APPCAP) during 2013–2017, which mainly targeted primary pollutants including NO_x and SO_2. Subsequently during 2018–2021, China implemented the Blue Sky Protection Campaign (BSPC), which continued the previous measures and strengthened VOCs control. O_3 around the urban PT area continued to increase between 2013 and 2018, possibly related to Shanghai being under the VOC-limited regime. However, the decrease in O_3 after 2018 may be due to the continuous reduction of NO_x making Shanghai shift from VOC-limited regime to a transition regime or NO_x-limited regime. Similar findings were also reported by previous studies (Xu et al., 2019; Mousavinezhad et al., 2021; Wang et al., 2022b). Notably, reducing VOCs mitigates
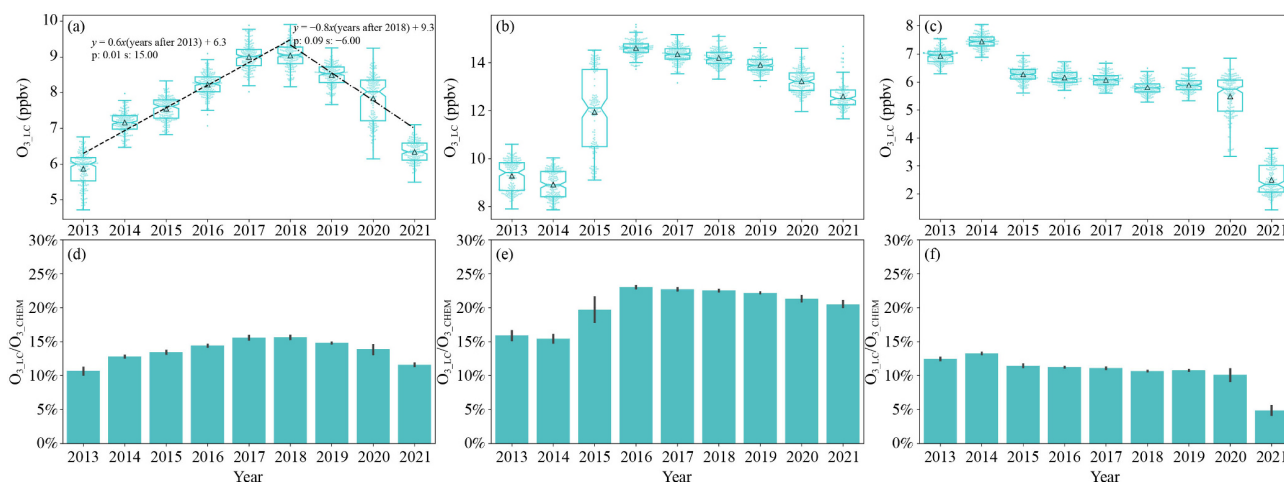

Fig. 7 Contributions of O_{3_LC} at (a) PT, (b) DSL and (c) PDHN. In panels (a–c), the upper and lower boundaries of the box represent the 25th percentile and 75th percentile, and the notches represent the confidence interval around the median. The line and hollow triangles inside boxes represent the average and median values of corresponding data, respectively. The error lines above and below the box represent the 25th percentile value minus 1.5 * IQR (interquartile range) and the 75th percentile value plus 1.5 * IQR (interquartile range), respectively. In panels (d–f), the error line indicates the 95% confidence interval of the corresponding data.

O_3, irrespective of the O_3 formation regime. Regarding DSL, O_{3_LC} similarly rises and subsequently falls, but the rising trend is not monotonic compared to that at PT. During this period, NO_2 decreased with a fluctuating speed of approximately 0.16 ppb/yr at DSL, which was much lower than the decrease rate observed at the urban station PT. This difference may be due to DSL being located at the border between Shanghai and other cities, which is influenced by the pollution reduction measures of both Shanghai and the YRD region. Hence, it is inconsistent with urban Shanghai. As for PDHN, O_{3_LC} exhibits a fluctuating trend ranging between 5.5–6.9 ppb, dropping to only 2.5 ppb in 2021 due to a large decline in NO_x (Fig. S16). Looking at the period from 2013 to 2021, NO_2 decreased overall at a rate of 0.38 ppb/yr at PDHN, which is also lower than the decline rate observed at the urban site PT. This phenomenon may be attributed to PDHN's location in an area of Shanghai with relatively low human activity, as illustrated by the significantly lower local NO_2 concentrations compared to the other two stations (Fig. S16), and this could be the reason for PDHN's continuing stability of O_3.

3.5 Investigation into O_3 changes in typical years

We quantified the numerous contributing elements of the observed O_3 after unifying the data based on the integrated method, shown in Fig. 8. Based on the findings of the integrated method, we selected the years 2013, 2017, and 2021 as representative years to investigate changes in ozone-influencing variables at three representative locations in Shanghai.

During the implementation of the APPCAP policy between 2013 and 2017, the concentrations of daytime O_3 at the PT, DSL, and PDHN stations exhibited a rebound of 11.3 ppb (20.3%), 5.8 ppb (9.5%), and 2.2 ppb (3.6%), respectively, in 2017 compared to 2013. This trend suggests that the O_3 rebound was more significant in urban areas than in regional or upwind areas. The changes

in O_{3_LC} and O_{3_MET} were the primary causes of the observed O_3 rebound, as shown in Fig. 9. Apart from a slight decrease in O_{3_LC} at PDHN, there was a substantial rebound in O_{3_LC} of approximately 3.1 ppb and 5.1 ppb at PT and DSL, respectively, accounting for 27.4% and 87.9% of the total O_3 rebound. This finding clearly indicates that the air pollution control measures implemented under the APPCAP policy were unfavorable for O_3 mitigation in urban areas. Actually, since ozone formation around PT and DSL usually falls into the VOC-limited regime (Li et al., 2021a), the emission reduction measures caused a quick drop in ambient NO_x (see Fig. S16), leading to a weakened titration effect and hence the rebound of local ozone. Moreover, under high VOCs conditions, carbonyl compounds produced by VOCs oxidation, especially formaldehyde, can accelerate O_3 production and VOCs oxidation (Li et al., 2021a). In addition, the decrease in $PM_{2.5}$ slows down the uptake of aerosol particles on hydroperoxyl radicals, thereby accelerating O_3 generation (Li et al., 2019). These factors may have contributed to the increase in O_3 levels. However, at the PDHN station, which is an upwind background station limited by NO_x, the reduction in NO_x was beneficial for O_3 mitigation, explaining the drop in O_{3_LC}.

For O_{3_MET}, all the three stations exhibited rebound during the APPCAP period, although the magnitude of the rebound differed among the three sites due to the differences in location. Notably, the increase in meteorological O_3 at all three sites was attributed to meteorological factors characterizing ODC, such as WS, Length, WD, Cluster, BLH and SP. We compare the variations in these factors between 2013 and 2017 (Tables S5–S7, Figs. S17 and S18) and found that the air mass trajectories from the ocean or primarily passing through the ocean (e.g., C7 and C11 at PT, C8 and C11 at DSL, C3 and C11 at PDHN) decreased significantly in 2017 compared to 2013, whereas air masses from inland, particularly south and west of Shanghai (e.g., C5 at PT,
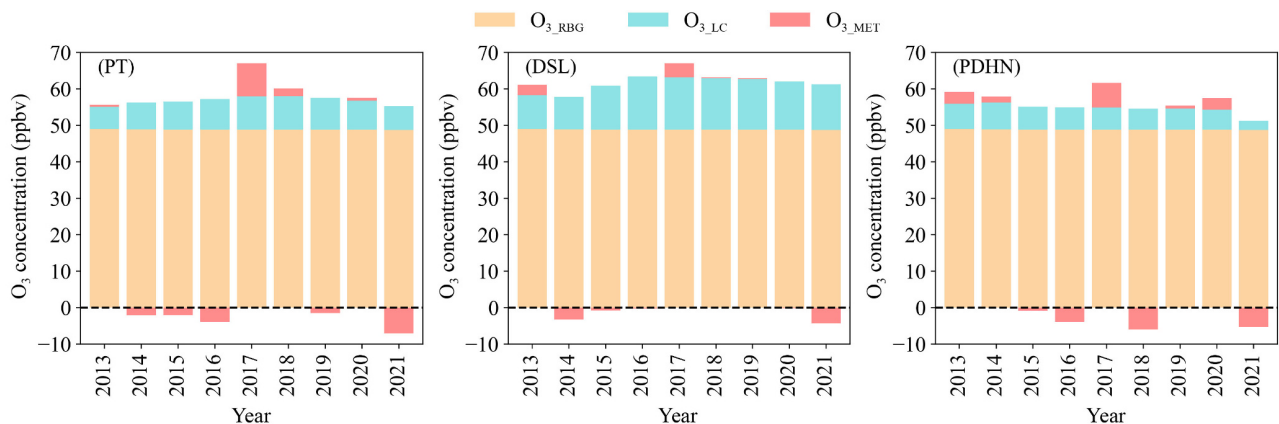

Fig. 8   Quantification of influencing factors on O_3 changes during the warm season from 2013 to 2021 in Shanghai.
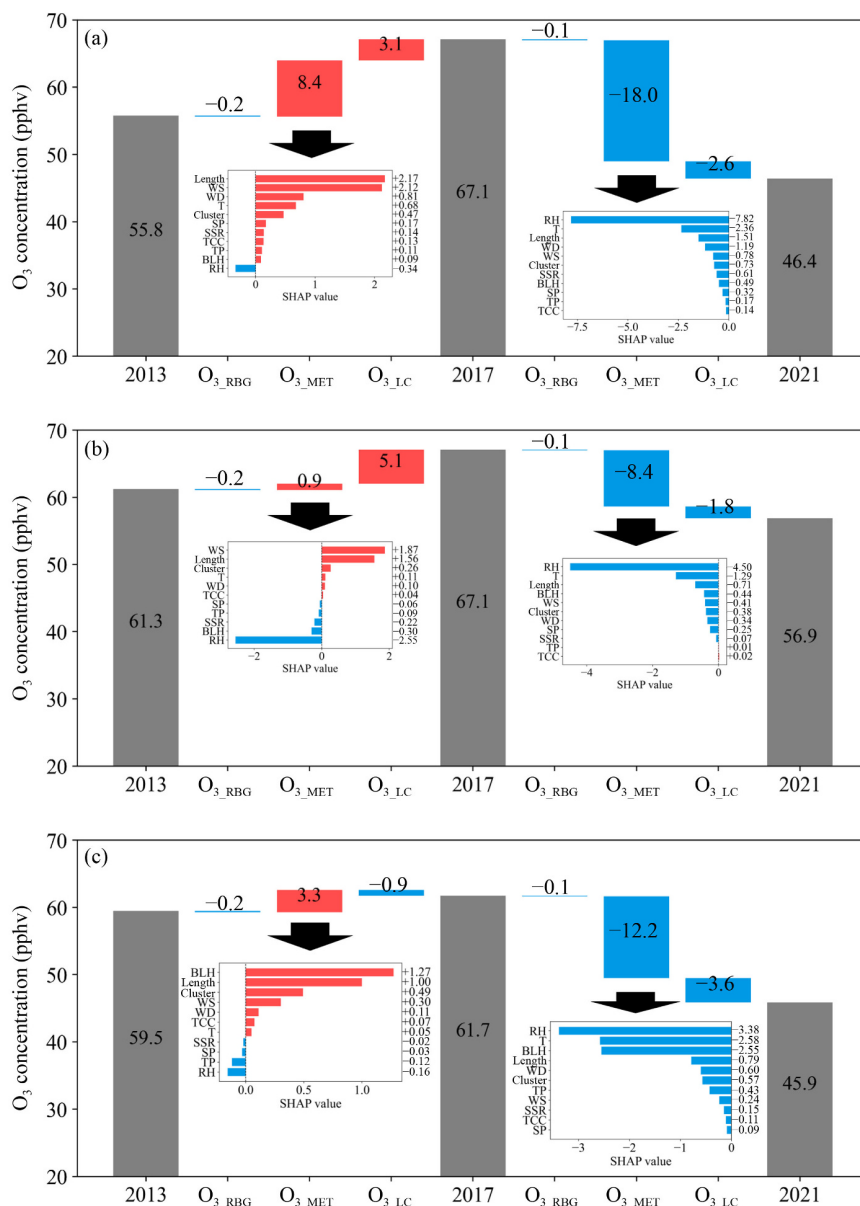
Fig. 9 Variation of influencing factors on O_3 for typical years at PT (a), DSL (b) and PDHN (c).

C4 at DSL, C8 and C10 at PDHN), increased. Meanwhile, the length of most trajectories decreased significantly in 2017 (Tables S5–S7). The south-west direction of Shanghai is the hinterland of the Yangtze River Delta with a high level of anthropogenic activities. The increase in the number of trajectories from the south-west indicates a possible increase in local O_3 transport to Shanghai, while the decrease in the number of air masses passing oceans indicates a decrease in the mitigation effect of clean air masses on O_3 pollution. In addition, the length of air mass trajectories reflects the transport rate of air masses (Hou et al., 2022), and the decrease in the length of air mass trajectories indicates a slowdown in the transport rate of air masses. This may lead to more O_3 being transported to Shanghai from surrounding areas.

When the wind rose diagram and wind speed were analyzed (Figs. S17 and S18), the wind direction and wind speed in 2017 exhibited significant differences compared to those in 2013. From the bivariate polar plot (Fig. 3), it can be seen that O_3 pollution is heavier when south-west winds dominate and the lowest O_3 concentration occurs when north-east winds occur. According to the comparison of the wind rose diagram, the number of north-easterly winds was significantly reduced in 2017 compared to 2013, and the wind speed decreased at all three stations. The comparison results indicate that in 2017, compared to 2013, the frequency of north-east winds that can alleviate O_3 pollution decreased, and wind speed significantly decreased, leading to overall favorable conditions for ozone accumulation in the 2017

wind field. These results indicate that the unfavorable ODC was the primary cause of the unfavorable O_{3_MET} in 2017 compared to 2013. In fact, according to Liu and Wang (2020a), the increase in O_3 in 2017 was caused by a change in wind field, which is consistent with our findings.

After the implementation of the BSPC during 2018–2020, there was a significant reduction in daytime O_3 at PT, DSL, and PDHN by 20.7 ppb (30.8%), 10.2 ppb (15.2%), and 15.8 ppb (25.6%), respectively, compared to 2017 (Fig. 9). Although the effects of this policy on O_{3_RBG} were moderate, there was a decrease in O_{3_LC} at PT (2.6 ppb), DSL (1.8 ppb), and PDHN (3.6 ppb), accounting for 12.6%, 17.6%, and 22.8% of the overall decline, respectively. This suggests a shift in the O_3 formation regime (Xu et al., 2019). Furthermore, O_{3_MET} decreased by 18 ppb, 8.4 ppb, and 12.2 ppb at PT, DSL, and PDHN, respectively, accounting for 87%, 82.4%, and 77.2% of the overall decline (Fig. 9). Therefore, it can be concluded that the significant reduction in O_3 in 2021 was mainly driven by favorable weather conditions for O_3 mitigation, but emission reduction also played an important role in controlling the O_3 situation, particularly at the regional level. Further investigation into the contribution of meteorological factors revealed that all types of meteorological factors favored O_3 mitigation, including decreasing SSR, decreasing T, increasing TP, increasing RH, and WS (Fig. S18). Based on the comparison of meteorological parameters, the overall decrease in O_{3_MET} may have originated from a decrease in solar radiation or an increase in precipitation.

The implementation of the APPCAP and the BSPC has resulted in a significant reduction of O_3, particularly those that are photochemically formed. Furthermore, it may have led to a shift in the O_3 formation regime in Shanghai. Therefore, if these precursor emission reduction policies are continued, they could eventually alleviate the O_3 pollution problem. Weather also plays a significant role in O_3 pollution during these periods. For example, unfavorable ODC in 2017 led to an increased O_3 pollution condition, while a decrease in solar radiation or an increase in precipitation in 2021 amplified the effect of pollution control efforts. During both periods of emission reduction policies, the reduction of O_{3_RBG} is sustained because rural and suburban areas are generally in the NO_x-limited regime, and NO_x reduction can lead to a reduction in suburban O_3, resulting in regional O_3 mitigation (Li et al., 2021a). The persistent decrease in O_{3_RBG} indicates that regional O_3 precursor emission controls have some effect on O_3 reduction. However, based on the changes in O_{3_LC}, it is clear that stringent emission reduction of regional pollutants combined with precise characterization of O_3 formation regime is more effective for O_3 reduction.

Our results were compared with previous studies. For example, Wang et al. (2019) found that anthropogenic emissions contributed positively to O_3 during the summers of 2014 and 2015 in Shanghai, compared to 2013, while meteorology and natural sources had negative contributions. Similarly, Yin et al. (2021) used eXtreme Gradient Boosting (XGBoost) model and GEOS-Chem and found that, compared to the summer of 2019, emissions caused a decrease in O_3 while meteorology caused an increase in O_3 in Shanghai during the summer of 2020, which is similar to the results of our study. Li et al. (2020) found that meteorological factors caused an increase in O_3 at a rate of 0.2 ppb/yr and emissions caused an increase in O_3 at a rate of 1.5 ppb/yr for the YRD region between 2013 and 2017 using multilinear regression (MLR) model method and GEOS-Chem, which is similar to the results of the regional station DSL. Chen et al. (2020) used MLR model and found that, in the YRD region, emissions caused O_3 to rise by 11.2 μg/m³ in 2018 compared to 2014, which is comparable to the DSL station. Meanwhile, meteorology caused O_3 to rise by 1.3 μg/m³, which is lower than the results from the DSL station. In addition, the meteorological impact, regional background O_3, and local formed O_3 were also compared with previous research in previous sections. These comparative results can well support the reliability of our research results.

4　Conclusions

The formation of O_3 is influenced by a combination of factors including meteorology, background, and local emissions, while its relationship with precursors is nonlinear. Therefore, isolating the effects of these influential factors is crucial for assessing the O_3 pollution issue. In this study, we proposed an integrated method that combines RF-based meteorological normalization technology, PCA-based regional background O_3 estimation, and SHAP-based causal attribution to separate observed O_3 into meteorological, local formed, and regional background contributions. Our investigation of daytime O_3 during the warm season between 2013 and 2021 in Shanghai revealed that meteorological factors that directly affects OPG include temperature, humidity, and solar radiation, etc., which play a significant role in O_3 variations, with a cumulative absolute contribution of 13.44–19.03 ppb, and their effects on O_3 have generally comparable yearly patterns. The overall effect of the ODC on O_3 is 13.18–13.66 ppb, with individual characteristics among different sites. The regional background O_3 is 48.8 ± 0.3 ppb, accounting for 79.6%–89.4% of the observed O_3 at different sites and shows an overall declining trend of 0.018 ppb/yr. Local formed O_3 in urban and regional sites increased and then decreased with a concentration range between 5.9–9.0 ppb and 8.9–14.6 ppb, respectively, while local formed O_3 in the upwind background site showed fluctuating variations with

concentrations ranging between 2.5–7.4 ppb. The O_3 rebound in 2017 compared to 2013 was primarily influenced by the unfavorable ODC and unbalanced emission reductions. In contrast, the O_3 decline in 2021 compared to 2017 was primarily influenced by overall unfavorable meteorological conditions for O_3 formation and further emissions reduction.

In summary, our method effectively captures the characteristics of O_3 variability and quantifies the effects of various factors on O_3 concentration. It should be noted that the intrinsic links between meteorological and chemical processes have been shown to accelerate chemical reaction rates and influence the abundance of VOCs in the troposphere, ultimately impacting surface O_3 concentrations. However, our research employs an integrated machine learning method to analyze the relationship between observed ozone and variables such as meteorology and emissions, prior to assigning the observed ozone to each influencing factor. This approach avoids potential impacts that may arise from the ozone formation process and enables the systematic quantification the impact of each factor, yielding relatively accurate results. These findings provide a more comprehensive and systematic understanding of the contribution of each factor to O_3 formation and identify which factors have a greater impact on O_3 concentrations. This information can aid in the development of more targeted and scientifically sound prevention and control strategies for local O_3 formation mechanisms, as well as proposing emergency plan based on weather forecasts. Although our method was applied to only three sites in Shanghai, these sites exhibited different geographically induced pollution characteristics, and the long-term time series provided good coverage of all types of O_3 pollution patterns. Therefore, the methods outlined in this work can be extended to other regions.

Research Station for the Yangtze River Delta at Dianshan Lake (SEED) (China) for supporting the study.

References

Alduchov O A, Eskridge R E (1996). Improved Magnus form approximation of saturation vapor pressure. Journal of Applied Meteorology, 35(4): 601–609

Austin E, Xiang J, Gould T R, Shirai J H, Yun S, Yost M G, Larson T V, Seto E (2021). Distinct ultrafine particle profiles associated with aircraft and roadway traffic. Environmental Science & Technology, 55(5): 2847–2858

Berlin S R, Langford A O, Estes M, Dong M, Parrish D D (2013). Magnitude, decadal changes, and impact of regional background ozone transported into the greater Houston, Texas, area. Environmental Science & Technology, 47(24): 13985–13992

Camalier L, Cox W, Dolwick P (2007). The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. Atmospheric Environment, 41(33): 7127–7137

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. Science of the Total Environment, 744: 140837

Chen W, Guenther A B, Shao M, Yuan B, Jia S, Mao J, Yan F, Krishnan P, Wang X (2022). Assessment of background ozone concentrations in China and implications for using region-specific volatile organic compounds emission abatement to mitigate air pollution. Environmental Pollution, 305: 119254

Dai Q, Hou L, Liu B, Zhang Y, Song C, Shi Z, Hopke P K, Feng Y (2021). Spring festival and COVID-19 lockdown: disentangling PM sources in major Chinese cities. Geophysical Research Letters, 48(11): e2021GL093403

Dang R, Liao H, Fu Y (2021). Quantifying the anthropogenic and meteorological influences on summertime surface ozone in China over 2012–2017. Science of the Total Environment, 754: 142394

Ding J, Dai Q, Fan W, Lu M, Zhang Y, Han S, Feng Y (2023). Impacts of meteorology and precursor emission change on O$_3$ variation in Tianjin, China from 2015 to 2021. Journal of Environmental Sciences-China, 126: 506–516

Gao D, Xie M, Liu J, Wang T, Ma C, Bai H, Chen X, Li M, Zhuang B, Li S (2021). Ozone variability induced by synoptic weather patterns in warm seasons of 2014–2018 over the Yangtze River Delta region, China. Atmospheric Chemistry and Physics, 21(8): 5847–5864

González-Pardo J, Ceballos-Santos S, Manzanas R, Santibáñez M, Fernández-Olmo I (2022). Estimating changes in air pollutant levels due to COVID-19 lockdown measures based on a business-as-usual prediction scenario using data mining models: a case-study for urban traffic sites in Spain. Science of the Total Environment, 823: 153786

Grange S K, Carslaw D C (2019). Using meteorological normalisation to detect interventions in air quality time series. Science of the Total

Environment, 653: 578–588

Grange S K, Carslaw D C, Lewis A C, Boleti E, Hueglin C (2018). Random forest meteorological normalisation models for Swiss PM_{10} trend analysis. Atmospheric Chemistry and Physics, 18(9): 6223–6239

Henneman L R F, Holmes H A, Mulholland J A, Russell A G (2015). Meteorological detrending of primary and secondary pollutant concentrations: method application and evaluation using long-term (2000–2012) data in Atlanta. Atmospheric Environment, 119: 201–210

Hirsch R M, Slack J R, Smith R A (1982). Techniques of trend analysis for monthly water quality data. Water Resources Research, 18(1): 107–121

Hou L, Dai Q, Song C, Liu B, Guo F, Dai T, Li L, Liu B, Bi X, Zhang Y, et al. (2022). Revealing drivers of haze pollution by explainable machine learning. Environmental Science & Technology Letters, 9(2): 112–119

Hu C, Kang P, Jaffe D A, Li C, Zhang X, Wu K, Zhou M (2021). Understanding the impact of meteorology on ozone in 334 cities of China. Atmospheric Environment, 248: 118221

Huang Y, Chen J, Duan Q, Feng Y, Luo R, Wang W, Liu F, Bi S, Lee J (2022). A fast antibiotic detection method for simplified pretreatment through spectra-based machine learning. Frontiers of Environmental Science & Engineering, 16(3): 38

Joliffe I, Morgan B (1992). Principal component analysis and exploratory factor analysis. Statistical Methods in Medical Research, 1(1): 69–95

Langford A O, Senff C J, Banta R M, Hardesty R M, Alvarez R J II, Sandberg S P, Darby L S (2009). Regional and local background ozone in Houston during Texas air quality study 2006. Journal of Geophysical Research, 114(D7): D00F12

Lelieveld J, Evans J S, Fnais M, Giannadaki D, Pozzer A (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature, 525(7569): 367–371

Li C, Zhu Q, Jin X, Cohen R C (2022). Elucidating contributions of anthropogenic volatile organic compounds and particulate matter to ozone trends over China. Environmental Science & Technology, 56(18): 12906–12916

Li K, Jacob D J, Liao H, Qiu Y, Shen L, Zhai S, Bates K H, Sulprizio M P, Song S, Lu X, et al. (2021a). Ozone pollution in the North China Plain spreading into the late-winter haze season. Proceedings of the National Academy of Sciences of the United States of America, 118(10): e2015797118

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. Proceedings of the National Academy of Sciences of the United States of America, 116(2): 422–427

Li K, Jacob D J, Shen L, Lu X, De Smedt I, Liao H (2020). Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences. Atmospheric Chemistry and Physics, 20(19): 11423–11433

Li X B, Fan G, Lou S, Yuan B, Wang X, Shao M (2021b). Transport and boundary layer interaction contribution to extremely high surface ozone levels in eastern China. Environmental Pollution, 268: 115804

Lin C, Lau A K H, Fung J C H, Song Y, Li Y, Tao M, Lu X, Ma J, Lao

X Q (2021). Removing the effects of meteorological factors on changes in nitrogen dioxide and ozone concentrations in China from 2013 to 2020. Science of the Total Environment, 793: 148575

Liu Y, Wang T (2020a). Worsening urban ozone pollution in China from 2013 to 2017 – Part 1: The complex and varying roles of meteorology. Atmospheric Chemistry and Physics, 20(11): 6305–6321

Liu Y, Wang T (2020b). Worsening urban ozone pollution in China from 2013 to 2017 – Part 2: The effects of emission changes and implications for multi-pollutant control. Atmospheric Chemistry and Physics, 20(11): 6323–6337

Lovrić M, Pavlović K, Vuković M, Grange S K, Haberl M, Kern R (2021). Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. Environmental Pollution, 274: 115900

Lu W, Huo W, Gulina H, Pan C (2022). Development of machine learning multi-city model for municipal solid waste generation prediction. Frontiers of Environmental Science & Engineering, 16(9): 119

Lu X, Zhang L, Chen Y, Zhou M, Zheng B, Li K, Liu Y, Lin J, Fu T M, Zhang Q (2019). Exploring 2016–2017 surface ozone pollution over China: source contributions and meteorological influences. Atmospheric Chemistry and Physics, 19(12): 8339–8361

Lundberg S M, Erion G, Chen H, Degrave A, Prutkin J M, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S I (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1): 56–67

Lundberg S M, Lee S I (2017). A Unified Approach to Interpreting Model Predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA. Red Hook, NY: Curran Associates Inc, 4768–4777

Mousavinezhad S, Choi Y, Pouyaei A, Ghahremanloo M, Nelson D L (2021). A comprehensive investigation of surface ozone pollution in China, 2015–2019: Separating the contributions from meteorology and precursor emissions. Atmospheric Research, 257: 105599

Ni R, Lin J, Yan Y, Lin W (2018). Foreign and domestic contributions to springtime ozone over China. Atmospheric Chemistry and Physics, 18(15): 11447–11469

Pathakoti M, Santhoshi T, Aarathi M, Mahalakshmi D V, Kanchana A L, Srinivasulu J, Shekhar R S S, Soni V K, Sai S M V R, Raja P (2021). Assessment of spatio-temporal climatological trends of ozone over the Indian region using machine learning. Spatial Statistics, 43: 100513

Qu L, Liu S, Ma L, Zhang Z, Du J, Zhou Y, Meng F (2020). Evaluating the meteorological normalized $PM_{2.5}$ trend (2014–2019) in the "2+26" region of China using an ensemble learning technique. Environmental Pollution, 266: 115346

Sahu S K, Liu S, Liu S, Ding D, Xing J (2021). Ozone pollution in China: Background and transboundary contributions to ozone concentration & related health effects across the country. Science of the Total Environment, 761: 144131

Sen P K (1968). Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association, 63(324): 1379–1389

Shi Z, Song C, Liu B, Lu G, Xu J, Van Vu T, Elliott R J R, Li W, Bloss

W J, Harrison R M (2021). Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. Science Advances, 7(3): eabd6696

Song C, Becagli S, Beddows D C S, Brean J, Browse J, Dai Q, Dall'osto M, Ferracci V, Harrison R M, Harris N, et al. (2022). Understanding sources and drivers of size-resolved aerosol in the high Arctic Islands of Svalbard using a receptor model coupled with machine learning. Environmental Science & Technology, 56(16): 11189–11198

Stein A F, Draxler R R, Rolph G D, Stunder B J B, Cohen M D, Ngan F (2015). NOAA's HYSPLIT atmospheric transport and dispersion modeling system. Bulletin of the American Meteorological Society, 96(12): 2059–2077

Suciu L G, Griffin R J, Masiello C A (2017). Regional background O_3 and NO_x in the Houston–Galveston–Brazoria (TX) region: a decadal-scale perspective. Atmospheric Chemistry and Physics, 17(11): 6565–6581

Tang G, Liu Y, Huang X, Wang Y, Hu B, Zhang Y, Song T, Li X, Wu S, Li Q, et al. (2021). Aggravated ozone pollution in the strong free convection boundary layer. Science of the Total Environment, 788: 147740

Tang M X, Huang X F, Sun T L, Cheng Y, Luo Y, Chen Z, Lin X Y, Cao L M, Zhai Y H, He L Y (2022). Decisive role of ozone formation control in winter $PM_{2.5}$ mitigation in Shenzhen, China. Environmental Pollution, 301: 119027

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. Atmospheric Chemistry and Physics, 19(17): 11303–11314

Wang F T, Zhang K, Xue J, Huang L, Wang Y J, Chen H, Wang S Y, Fu J S, Li L (2022a). Understanding regional background ozone by multiple methods: a case study in the Shandong region, China, 2018–2020. Journal of Geophysical Research. Atmospheres, 127(22): e2022JD036809

Wang P, Guo H, Hu J, Kota S H, Ying Q, Zhang H (2019). Responses of $PM_{2.5}$ and O_3 concentrations to changes of meteorology and emissions in China. Science of the Total Environment, 662: 297–306

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. Science of the Total Environment, 575: 1582–1596

Wang W, Parrish D D, Wang S, Bao F, Ni R, Li X, Yang S, Wang H, Cheng Y, Su H (2022b). Long-term trend of ozone pollution in China during 2014–2020: distinct seasonal and spatial characteristics and ozone sensitivity. Atmospheric Chemistry and Physics, 22(13): 8935–8949

Wang Y, Wen Y, Wang Y, Zhang S, Zhang K M, Zheng H, Xing J, Wu Y, Hao J (2020). Four-month changes in air quality during and after the COVID-19 lockdown in six megacities in China. Environmental

Science & Technology Letters, 7(11): 802–808

Wren S N, Mihele C M, Lu G, Jiang Z, Wen D, Hayden K, Mittermeier R L, Staebler R M, Cober S G, Brook J R (2020). Improving insights on air pollutant mixtures and their origins by enhancing local monitoring in an area of intensive resource development. Environmental Science & Technology, 54(23): 14936–14945

Wu Q, Li T, Zhang S, Fu J, Seyler B C, Zhou Z, Deng X, Wang B, Zhan Y (2022). Evaluation of NO_x emissions before, during, and after the COVID-19 lockdowns in China: A comparison of meteorological normalization methods. Atmospheric Environment, 278: 119083

Xu J, Tie X, Gao W, Lin Y, Fu Q (2019). Measurement and model analyses of the ozone variation during 2006 to 2015 and its response to emission change in megacity Shanghai, China. Atmospheric Chemistry and Physics, 19(14): 9017–9035

Yang L, Luo H, Yuan Z, Zheng J, Huang Z, Li C, Lin X, Louie P K K, Chen D, Bian Y (2019). Quantitative impacts of meteorology and precursor emission changes on the long-term trend of ambient ozone over the Pearl River Delta, China, and implications for ozone control strategy. Atmospheric Chemistry and Physics, 19(20): 12901–12916

Yang P, Yang H, Sardans J, Tong C, Zhao G, Peñuelas J, Li L, Zhang Y, Tan L, Chun K P, et al. (2020). Large spatial variations in diffusive CH_4 fluxes from a subtropical coastal reservoir affected by sewage discharge in Southeast China. Environmental Science & Technology, 54(22): 14192–14203

Yin H, Lu X, Sun Y, Li K, Gao M, Zheng B, Liu C (2021). Unprecedented decline in summertime surface ozone over eastern China in 2020 comparably attributable to anthropogenic emission reductions and meteorology. Environmental Research Letters, 16(12): 124069

Zhang K, Liu Z, Zhang X, Li Q, Jensen A, Tan W, Huang L, Wang Y, De Gouw J, Li L (2022). Insights into the significant increase of ozone during COVID-19 in a typical urban city of China. Atmospheric Chemistry and Physics, 22(7): 4853–4866

Zhang Y, Vu T V, Sun J, He J, Shen X, Lin W, Zhang X, Zhong J, Gao W, Wang Y, et al. (2020). Significant changes in chemistry of fine particles in wintertime Beijing from 2007 to 2017: impact of clean air actions. Environmental Science & Technology, 54(3): 1344–1352

Zhou W, Lei L, Du A, Zhang Z, Li Y, Yang Y, Tang G, Chen C, Xu W, Sun J, et al. (2022). Unexpected increases of severe haze pollution during the post COVID-19 period: effects of emissions, meteorology, and secondary production. Journal of Geophysical Research: Atmospheres, 127(3): e2021JD035710

Zhu Q, Gu A, Li D, Zhang T, Xiang L, He M (2021). Online recognition of drainage type based on UV-vis spectra and derivative neural network algorithm. Frontiers of Environmental Science & Engineering, 15(6): 136