Supporting Information

Text S1 Site selection

Three typical air quality monitoring stations (PDHN, PT, and DSL) are selected in this study, representing rural Shanghai, urban Shanghai, and the YRD region, respectively. In fact, during the period from 2013 to 2021, there were ten national monitoring stations, with PDHN excluded. However, due to the lack of an observation site in eastern Shanghai, we included the long-operating municipal monitoring station, PDHN, which was upgraded to a national monitoring station in 2020.

To clarify the representativeness of the three sites, we conducted a hierarchical clustering analysis which groups sites with similar pollutant data and reveals correlations among them, as shown in Fig. S1. We found that NO2 has higher consistency at sites near downtown (Fig. S1(a)). Similarly, O3 also shows higher consistency at these sites, but the correlation is higher between sites in downtown and those near downtown, such as PDCS, due to regional characteristics of O3 pollution (Fig. S1(b)). This suggests that sites around PT or in downtown and near downtown with high consistency can represent the pollution characteristics of urban O3 well. We also analyzed the daily and monthly variations of O3 and NO2 from April to September in 2021 to obtain detailed pollution characteristics of each site and differences in pollution characteristics among the sites, as shown in Fig. S2. The PT site and sites located in downtown showed consistent changes, especially with NO2 showing a late first peak and an early second peak due to the traffic rush hour. However, the other sites, such as DSL and PDHN showed poor correlation with other sites, mainly due to their unique geographical locations. The daily and monthly variations of DSL and PDHN were different from most sites, with NO2 showing an early first peak and a late second peak, also caused by the commuting rush hour. Furthermore, in the warm season, NO2 and O3 levels in DSL were higher than most sites, while those in PDHN were lower than most sites, which also reinforces the typicality of these two sites as upwind and suburban downwind sites in the city, respectively. DSL is located in the western of Shanghai, and the wind rose results indicate that it is downwind in Shanghai and a typical regional site (Fig. S1(c)). PDHN is situated on the easternmost side of Shanghai and is adjacent to the ocean. According to the wind rose diagram, this site is upwind of the dominant wind direction in summer in Shanghai. These characteristics suggest that PDHN is ideal for serving as a background and upwind station.

In summary, downtown and surrounding sites are suitable for representing urban pollution characteristics, with the PT site being located in the center of downtown and covering our study period, making it the most suitable urban site. The PDHN and DSL sites, respectively located on the east and west sides of the city, are the most representative upwind and suburban downwind sites among all sites.

Text S2 Air quality data processing

Quality assurance and control measures are implemented to ensure the integrity of the air quality data, which includes outlier elimination and missing value filling (Garces and Sbarbaro, 2011). Negative values are first removed from the data set. For the days with less than 6 h of missing values, we employ the linear interpolation method to fill in the gaps (Junninen et al., 2004). These steps result in a data set with a reduced number of missing values and relatively fewer outliers.

Text S3 Description of ERA5 data and its comparison with surface observations

We have selected three stations, PT, DSL, and PDHN, with completely different geographical features, and the meteorological inputs used should also be in line with the characteristics of each station. The ERA5/ERA5-Land data are used as meteorological data input. Data from ERA5 reanalysis data are coarser than surface measurement data, here are reasons to use these data sets. First, ERA5/ERA5-Land data are the fourth-generation reanalysis data, with ERA5-Land having a spatial resolution of 0.1° × 0.1° (about 9 km) and covering all time periods from 1950 to present, while ERA5 has a spatial resolution of 0.25° × 0.25° (about 31 km) and covers all time periods from 1959 to present. Both have a temporal resolution of 1 h and are the highest-resolution data sets in terms of spatial resolution among all reanalysis data. In addition, the 4D-Var assimilation method used in ERA5/ERA5-Land uses observational results from over 200 satellite instruments and conventional data, making it highly accurate and reliable (Hersbach et al., 2020; Muñoz-Sabater et al., 2021). Therefore, it has been widely used in atmospheric environmental research in recent years (Ramon et al., 2019; Fan et al., 2021; Hu et al., 2021; Kong et al., 2021; Liu et al., 2021; Parida et al., 2021; Shi et al., 2021; Wang et al., 2021b; Purwanto et al., 2022; Storey & Price, 2022; Wang et al., 2022b; Xu et al., 2022; Casallas et al., 2023; Chong et al., 2023).

To evaluate the data quality, we compared the surface observation data from Pudong airport (PDAP) and Hongqiao airport (HQAP) with the corresponding grid data in ERA5, as shown in Figs. S3 and S4. It can be observed that the T and DT observed at both stations are highly correlated with ERA5 data, with R values exceeding 0.98, which indicates that ERA5 data can accurately reflect the T and RH (determined by DT and T) in the actual atmosphere. The correlation between ERA5 and ground observation data for WS and WD decreased due to the data precision at the airports, but overall acceptable. Overall, the ERA/ERA-land data set can provide sufficient and precise meteorological data support to this study.

Text S4 RF model and selection of hyper-parameters

The RF model was developed using the “RandomForestRegressor” function of the “scikit-learn” library in Python. In the RF model, the most important parameters are “n_estimators” (the number of trees in the forest), “max_features” (the number of features to consider when looking for the best split) and “min_samples_leaf” (the minimum number of samples required to be at a leaf node). These parameters play a decisive role in the model's performance. To improve the stability and

accuracy of the model, we used “GridSearchCV” function in the “scikit-learn” library to select the best parameters. The detailed parameter selection process is presented in Figs. S6 and S7.

Text S5 Regional background ozone estimation

To eliminate the possible effects of different dimensional data, we performed zero-mean normalization on the O_3, NO_x, WD, WS, and T data and input them into PCA to downscale the data and retain principal components (PC) with eigenvalues greater than 1. Table S2 provides a detailed description of the PCs obtained. PC1 and PC2 can explain most of the original data, as they account for almost 40% and 25% of the variance, respectively (Suciu et al., 2017; Wang et al., 2022a). Table S3 shows the matrix of component score coefficients obtained from PCA, where the relationship between each variable and PC1 and PC2 is different. By examining the relationship between each variable and the components, and the different relationships between sites, we can assign physical significance to the components of each site separately. Regarding the correlation between the two components and each variable at each site, the size distribution of the correlation between PC1 and PC2 and each variable is almost the same. Therefore, we only need to clarify the relationship between each variable and the components to interpret the components. First, O_3, NO_x and T are strongly correlated with PC1, with PC1 being positively correlated with O_3 and T, and negatively correlated with NO_x. This indicates that for PC1, O_3 is negatively correlated to NO_x, but consistent with T. This characteristic of PC1 is consistent with the production process of O_3 in atmospheric photochemical process. Therefore, PC1 represents locally generated O_3 (Suciu et al., 2017). NO_x and WS are strongly correlated with PC2 and poorly correlated with O_3 and T. In the atmospheric pollution process, atmospheric transport is often entrained with NO_x, and the correlation between PC2 and the variables is similar to this pattern, so we can interpret it as a component representing regional background O_3 (Suciu et al., 2017).

Text S6 Description of the meteorological impact

Using the integrated method, we calculated the meteorological contributions to O_3 at PT, DSL, and PDHN stations from 2013 to 2021 (Figs. S8–S15). While the impacts of meteorological factors on O_3 concentrations were comparable across the three stations, there were notable differences in the composition of these factors, which reflect the characteristic features of urban, regional, and background sites for PT, DSL, and PDHN, respectively.

RH is the meteorological parameter with the highest impact on O_3 at PT, with an average absolute contribution of 10.77 ppb (Figs. S8(a) and S8(d)). Positive contributions from RH were typically observed in April–May, while negative contributions were observed in June–August (Fig. S9). However, positive contribution from RH were also observed in June–August in 2013 and 2017, likely due to the shorter rainy season in these years. Additionally, T (with an average absolute contribution of 4.58 ppb) was the second most significant meteorological factor influencing O_3 concentration at the PT site. The contribution of T fluctuated in a “^” pattern, which is related to the rise and fall in temperature during warm season. Similarly, the influence of temperature on O_3 concentration was evident in the rainy season, such as the shorter rainy season in 2013 and 2017, which resulted in increased temperature and a considerable increase in its contribution to O_3 (Fig.

S12). It is worth noting that WD, Length, WS, and Cluster were also significant meteorological parameters contributing to O_3 concentration at PT (Figs. S8(a) and S8(d)). On average, the absolute contribution of these factors cumulatively reached 11.60 ppb, indicating a substantial impact of horizontal transport on O_3 concentration at the PT station. Among the meteorological factors characterizing horizontal transport, WD and Cluster contributed more at the PT site than at the other two sites (Fig. S8), which could be related to the location of the PT site between the inland and the ocean. The convergence of inland air pollution and clean air from the sea at this location means that changes in air mass trajectory or wind direction can cause significant changes in O_3 concentrations (Fig. S13).

At the DSL station, two meteorological factors, RH and T, are the most important variables affecting O_3, with average absolute contributions of 10.37 ppb and 5.10 ppb, respectively (Figs. S8(b) and S8(e)). These factors display similar trends since the energy given to geographically close regions is consistent across larger weather systems. Compared to the PT station, the DSL station shows different meteorological characteristics that distinguish horizontal transport, as evidenced by the higher average absolute contribution of WS (4.85 ppb) at DSL than at PT (3.10 ppb). Additionally, the impact of Cluster and WD on O_3 at the DSL station (0.92 ppb, 2.31 ppb) is less than at the PT station (1.62 ppb, 3.75 ppb), suggesting that wind direction and air mass path have little effect on local O_3 changes (Figs. S10 and S13). Due to its location at the intersection of Jiangsu, Zhejiang, and Shanghai, the O_3 from these regions tends to converge near the DSL station, rendering wind direction or air mass from either direction less impactful on local O_3. Nonetheless, at a location where O_3 tends to accumulate (Fig. 3), such as the DSL, the dilution effect and diffusion effect of wind speed cause significant changes in O_3. These characteristics are evident from the dependence diagram in Fig. S13.

At the PDHN station, RH continued to be the most significant climatic element contributing an average absolute value of 7.65 ppb, followed by T with an average absolute value of 2.87 ppb (Figs. S8(c) and S8(f)). From a temporal perspective, the effect of RH and T on O_3 is comparable to that of the PT and DSL stations, but with lower magnitudes (Fig. S11). Since PDHN is close to the ocean, RH remained at a high level, and T remained at a low level, resulting their impact on O_3 being less than that of the other two stations (Fig. 4). Notably, all the meteorological factors characterizing O_3 transport at the PDHN site had a total average absolute contribution of 12.51 ppb, accounting for 47.0% of the total average absolute contribution (Fig. S8). Among these factors, BLH had a significant impact on O_3, with an average absolute contribution of 5.38 ppb, which was not only significantly higher at PDHN than PT and DSL but also the second most important factor affecting O_3. This finding suggests that the effects of transport, especially vertical transport, have a significant impact on the local variability of O_3 at PDHN. Moreover, being a background station near the sea, PDHN experiences less anthropogenic emissions, leading to less in situ chemical reactions that produce O_3. In contrast, the low concentration of NO_x makes it difficult to titrate off the O_3 transported from outside the station. Figure S14 shows a clear difference between the dependence plot of BLH at PDHN and other stations. The SHAP value of BLH at the PDHN station decreases with the increase of T, likely due to the strengthening of convective intensity within the troposphere, which causes the BLH to increase, further enhancing the O_3 diffusion ability at the PDHN station. Hence, BLH becomes one of the major meteorological factors affecting O_3 at the PDHN station, consistent with the findings of Tang et al. (2021).

We classified days with average O_3 concentrations exceeding 160 $\mu g/m^3$ between 10:00 and 17:00 as pollution days, and examined the influence of meteorological parameters on O_3 pollution during these periods (Fig. S15). We compared the effects of meteorological parameters on O_3 at the three stations and found that RH, T, and SSR exhibited similar characteristics: as RH increased

(decreased), its SHAP value decreased (increased), while T and SSR showed an increase (decrease) in their SHAP values with an increase (decrease) in T or SSR. Contributions of WS, WD, Cluster, and Length to O_3 were relatively consistent throughout the day, and differences in their contributions were primarily due to their geographical locations, as previous discussed. With respect to BLH, its contribution to O_3 was similar at PT and DSL stations, with the highest SHAP value during the highest BLH hours (12:00–14:00). At PDHN station, however, its SHAP value was much higher than at the other two stations, and was highest during lowest BLH hours in the afternoon (17:00), indicating a distinct pattern of influence on O_3 compared to the other two stations. This difference in the effect of BLH on O_3 in urban and suburban areas has been confirmed in multiple previous studies (Li et al., 2021; Tang et al., 2021; Wang et al., 2021a; Liu et al., 2022). For example, Tang et al. (2021) analyzed the effect of different convective boundary layers on O_3, and found that strong convection within the troposphere led to elevated T and BLH, intensifying the physical and chemical circulation of the atmosphere within the boundary layer, thereby increasing O_3 pollution. We believe that this finding reasonably explains the difference in BLH contribution to O_3 at the three stations, in conjunction with Fig. S14. In areas with high anthropogenic emissions such as PT and DSL, enhanced atmospheric convection triggers an elevation of T and BLH, as well as a decrease in RH, which further increases the physical and chemical cycling within the boundary layer, leading to higher O_3 concentrations. In contrast, in areas with low anthropogenic emissions such as PDHN, enhanced atmospheric convection does increase the BLH, but hardly affects the chemical circulation within the BLH, and only increases the physical diffusion of O_3 within the boundary layer, leading to a decrease in O_3 concentration. However, while Tang et al.'s (Tang et al., 2021) findings largely explain our results, data on atmospheric convection are not currently available, and there have been no observational studies on the effect of convection on O_3 in Shanghai, which warrants further investigation.

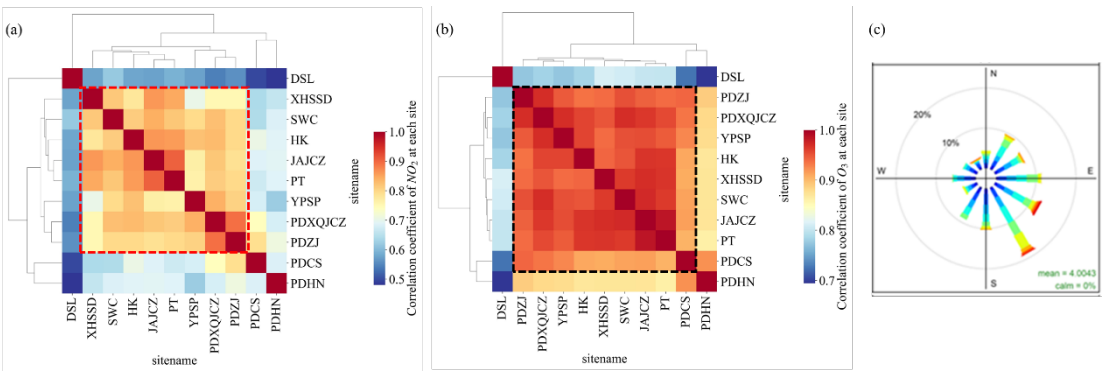

Fig. S1 Site selection diagram. Panels (a) and (b) present hierarchical clustering heatmaps of NO_2 and O_3 concentrations across 11 national monitoring stations. The color of each block in the matrix indicates the correlation between the values of two sites, with warmer colors representing higher correlation and cooler colors representing lower correlation. The dendrogram situated above and to the left of the matrix describes the clustering relationship between the sites. Clusters that are connected by shallower branches correspond to a large group of sites with similar data features, while clusters that are connected by deeper branches correspond to a more homogeneous group of sites with similar data characteristics. Panel (c) displays the wind rose diagram of the PT site from April to September, spanning the years 2013 to 2022.

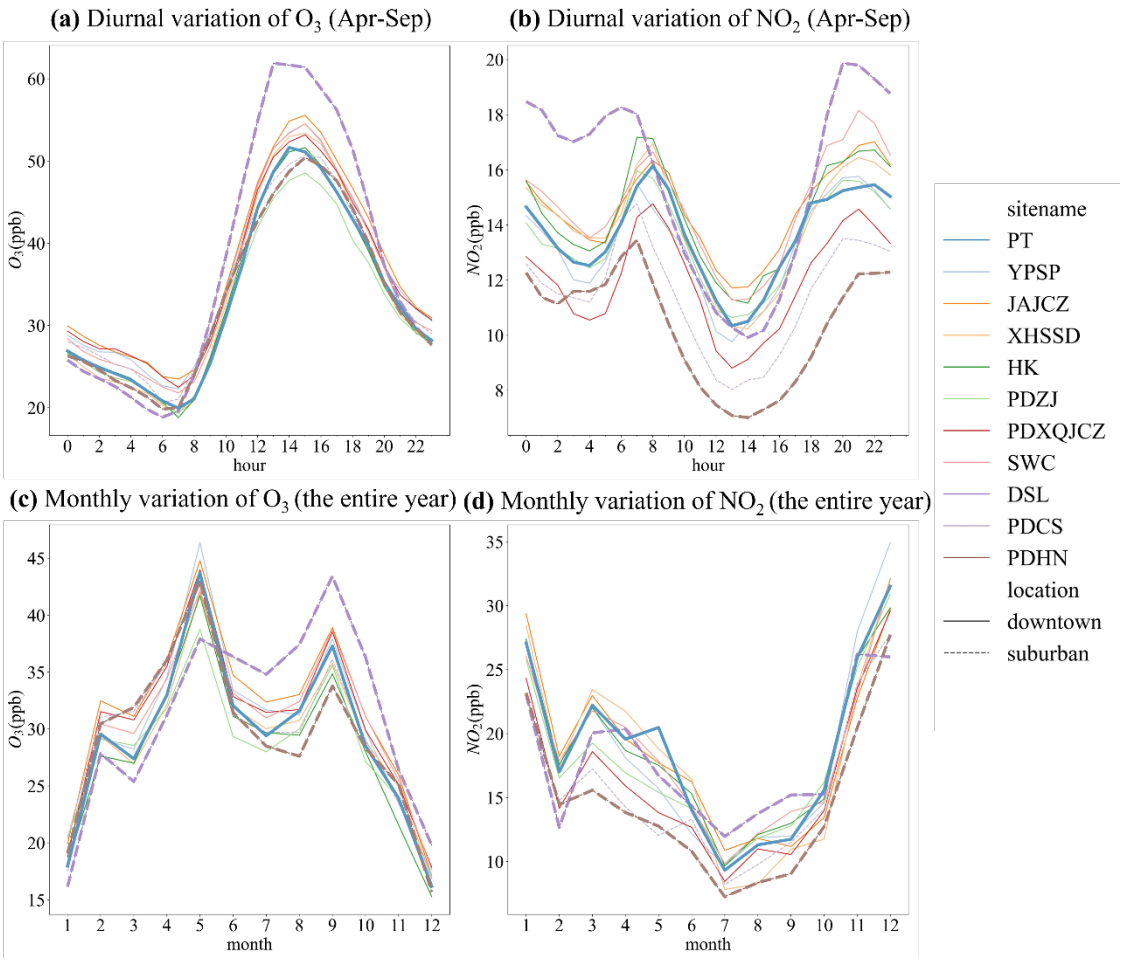

Fig. S2 Diurnal (a, b) and monthly (c, d) variations of O_3 and NO_2 at 11 national air quality monitoring stations in Shanghai. The solid line represents downtown stations, while the dotted line represents suburban stations. The data used for describing the diurnal variations corresponds to the period from April to September, while the data used for describing the monthly variations encompass the entire year.

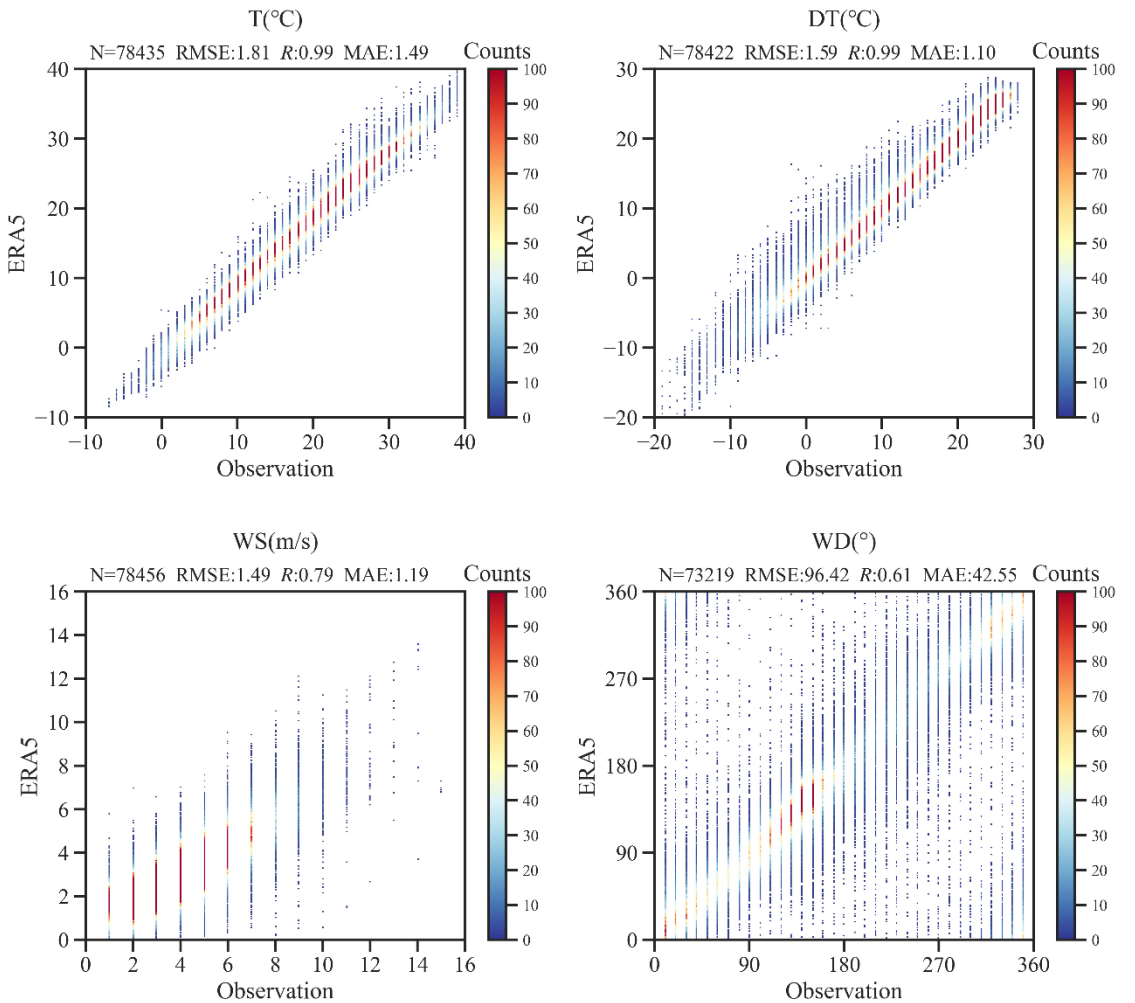

Fig. S3 Comparison of T, DT, WS and WD between ERA5 data and surface observation data in HQAP.

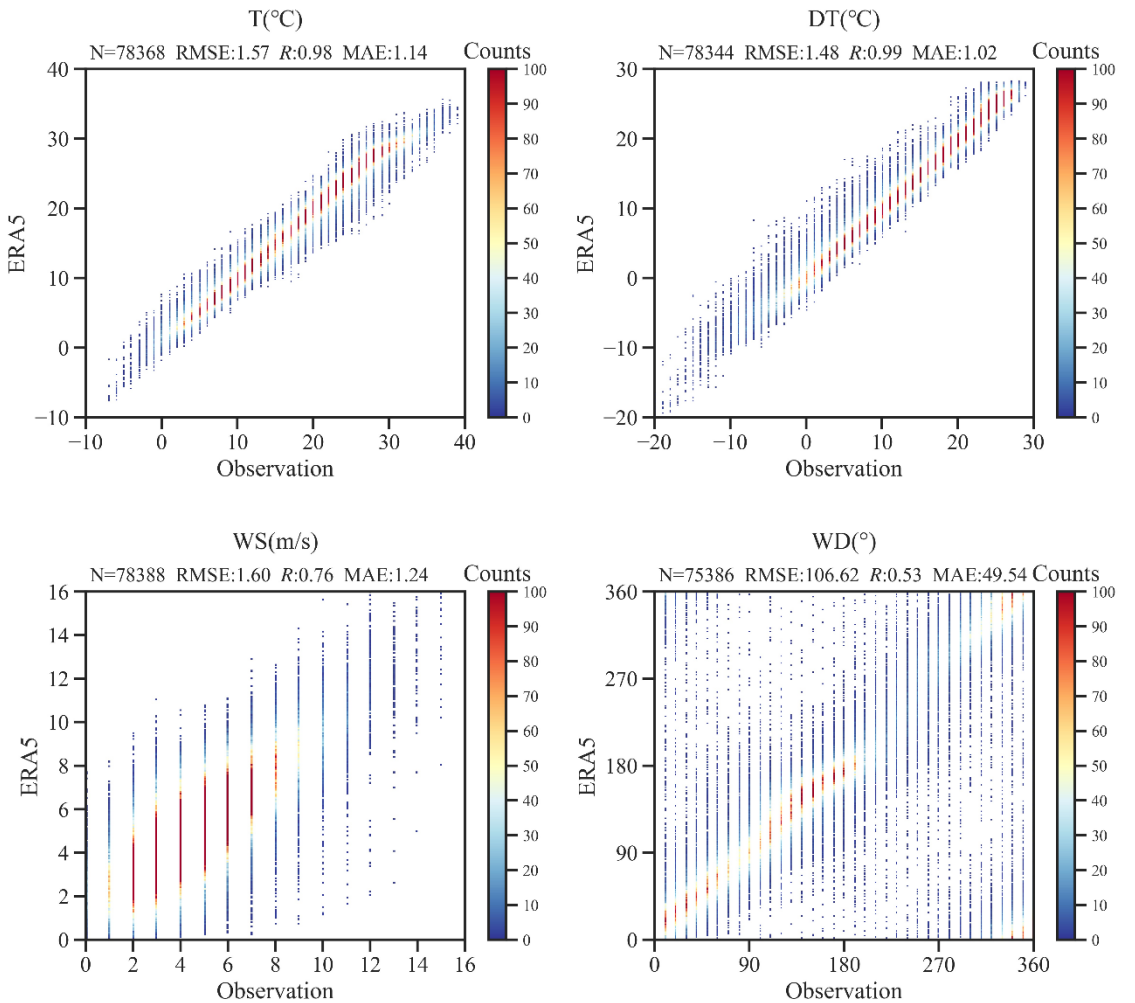

Fig. S4 Comparison of T, DT, WS and WD between ERA5 data and surface observation data in PDAP.

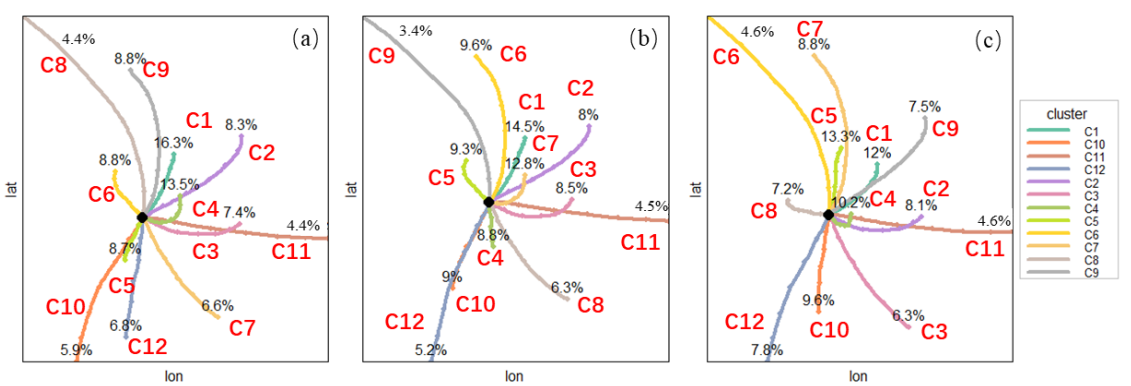

Fig. S5 Trajectory Clustering at the height of 100 m at PT (a), DSL (b) and PDHN (c) during 2013–2021.

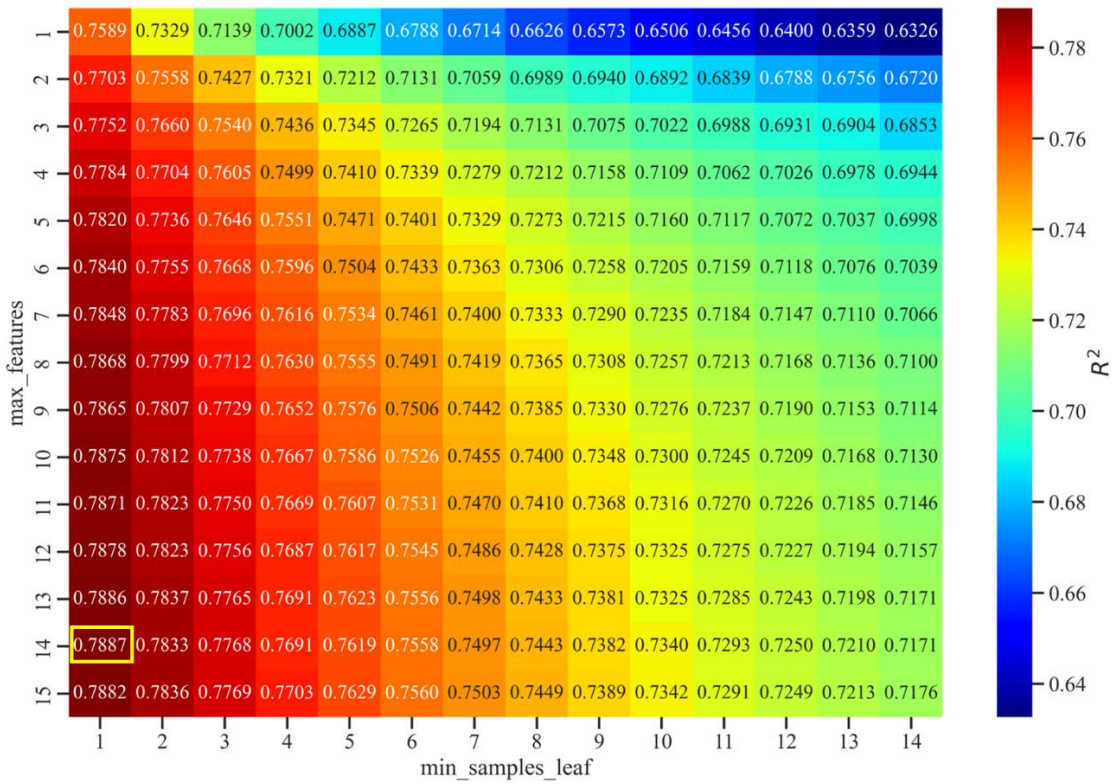

Fig. S6 Selection of hyper-parameters for RF model. R^2 is the squared correlation coefficient. According to the following figure, it can be found that max_features = 14, min_samples_leaf = 1, and n_estimators = 1950 are the parameters that the RF model performs relatively well.

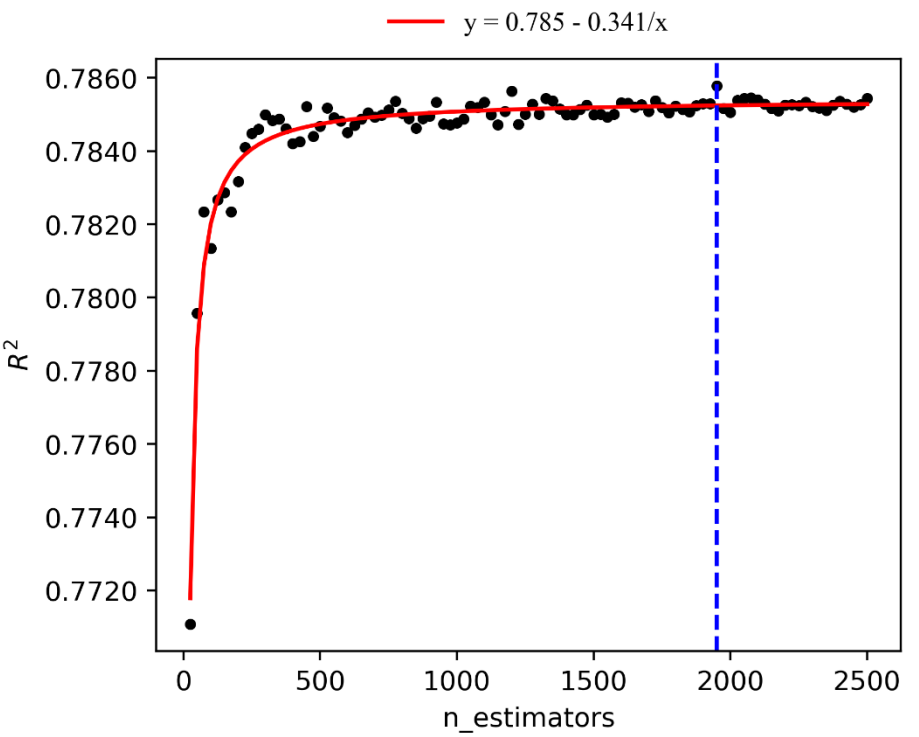

Fig. S7 Selection of hyper-parameters for RF model. R^2 is the squared correlation coefficient.

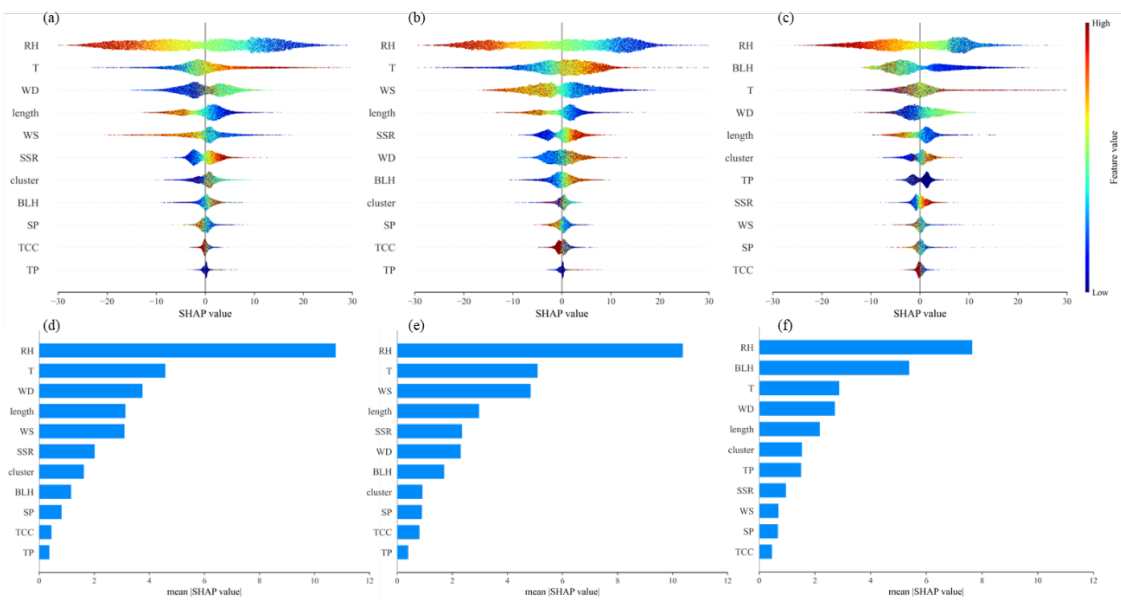

Fig. S8 Impact of meteorological factors on O_3 derived from the RF model and the SHAP approach. The color of the point indicates the relative magnitude of the feature value, and the position of the *x*-axis indicates the feature's influence on the predicted value (O_3) in the figures (a–c). The height of the bars represents the average of the absolute values of the SHAP values, i.e., the total absolute influence of the feature on the predicted values (O_3) in the figures (d–f).

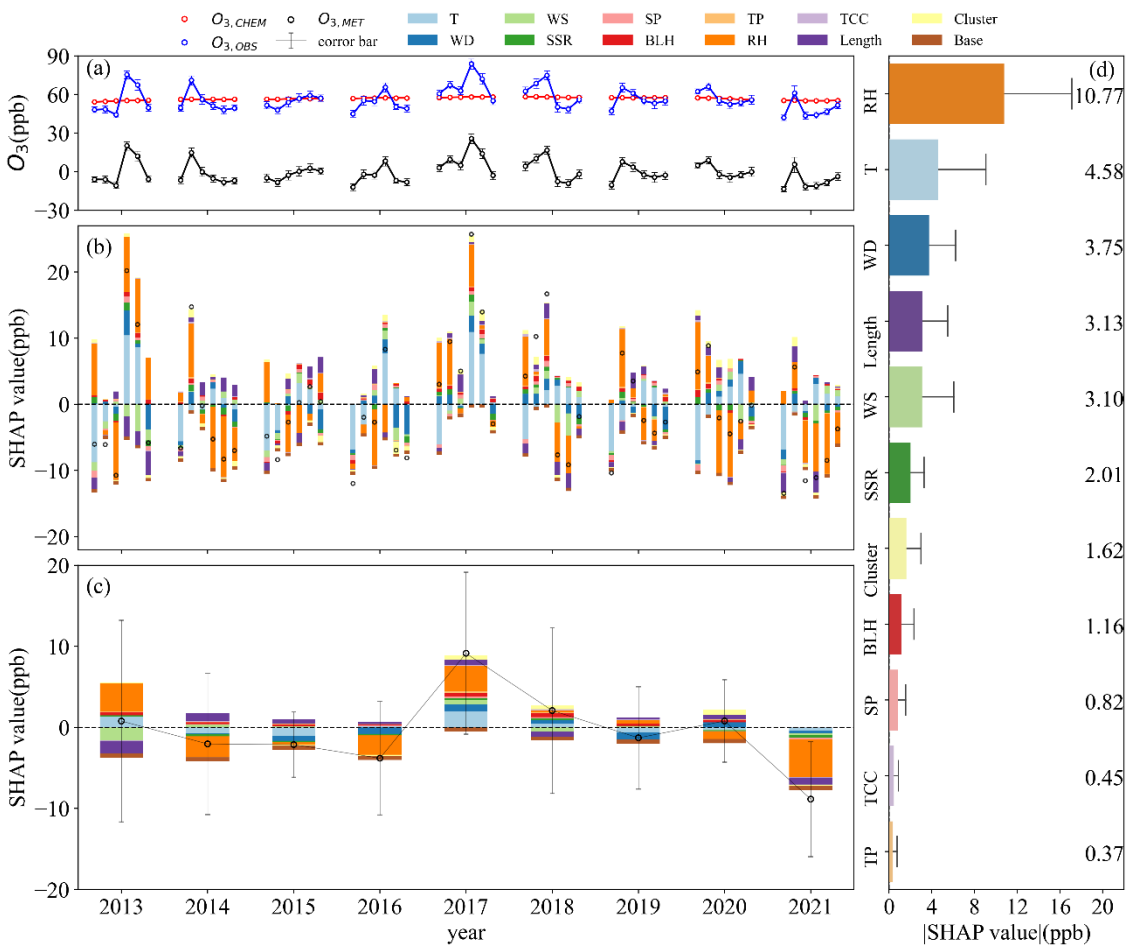

Fig. S9 Meteorological contributions to O_3 at PT station from April to September in the years 2013–2021. (a) temporal variations of O_{3_CHEM}, O_{3_OBS} and O_{3_MET}; (b) temporal variations of the SHAP values for each meteorological factor; (c) annual average SHAP values for each meteorological factor; (d) absolute mean SHAP values for each meteorological factor. The error bar in panel (a) represents the 95% confidence intervals of corresponding data, and the error bar in panels (c) and (d) represents the standard deviation of corresponding data. The base is the base value of the SHAP approach, i.e., the average of the O_{3_MET}. The same expression is used in Figs. S10 and S11.

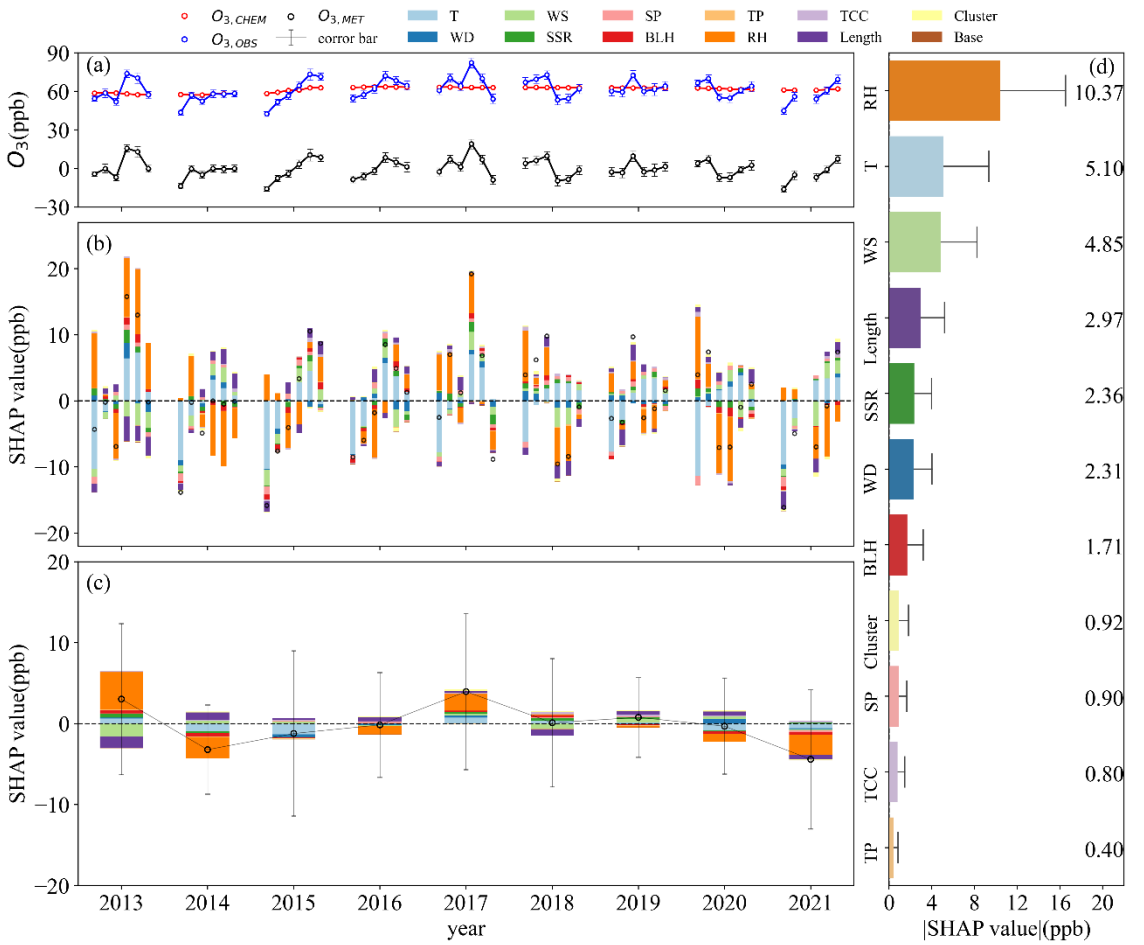

Fig. S10 Meteorological contributions to O_3 at DSL station from April to September in the years 2013–2021. (a) temporal variations of O_{3_CHEM}, O_{3_OBS} and O_{3_MET}; (b) temporal variations of the SHAP values for each meteorological factor; (c) annual average SHAP values for each meteorological factor;(d) absolute mean SHAP values for each meteorological factor.

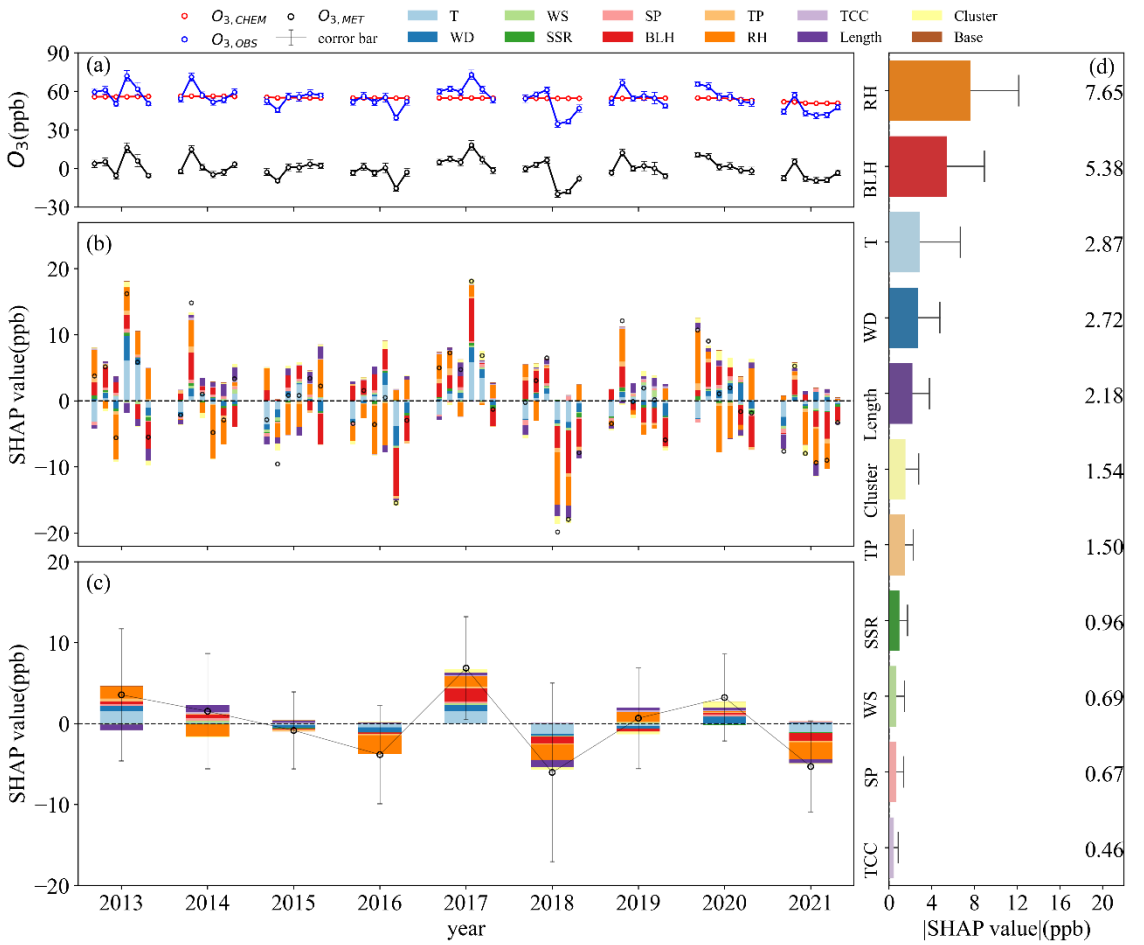

Fig. S11 Meteorological contributions to O_3 at PDHN station from April to September in the years 2013–2021. (a) temporal variations of O_{3_CHEM}, O_{3_OBS} and O_{3_MET}; (b) temporal variations of the SHAP values for each meteorological factor; (c) annual mean SHAP values for each meteorological factor; (d) absolute mean SHAP values for each meteorological factor.

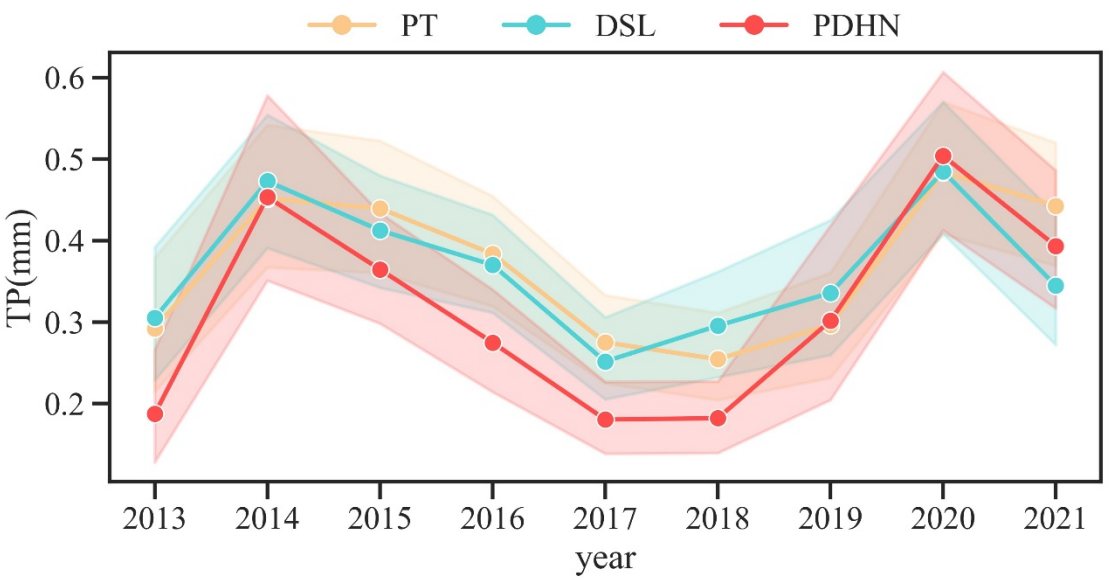

Fig. S12 Comparison of annual average precipitation in rainy seasons at PT, DSL and PDHN from 2013 to 2021.

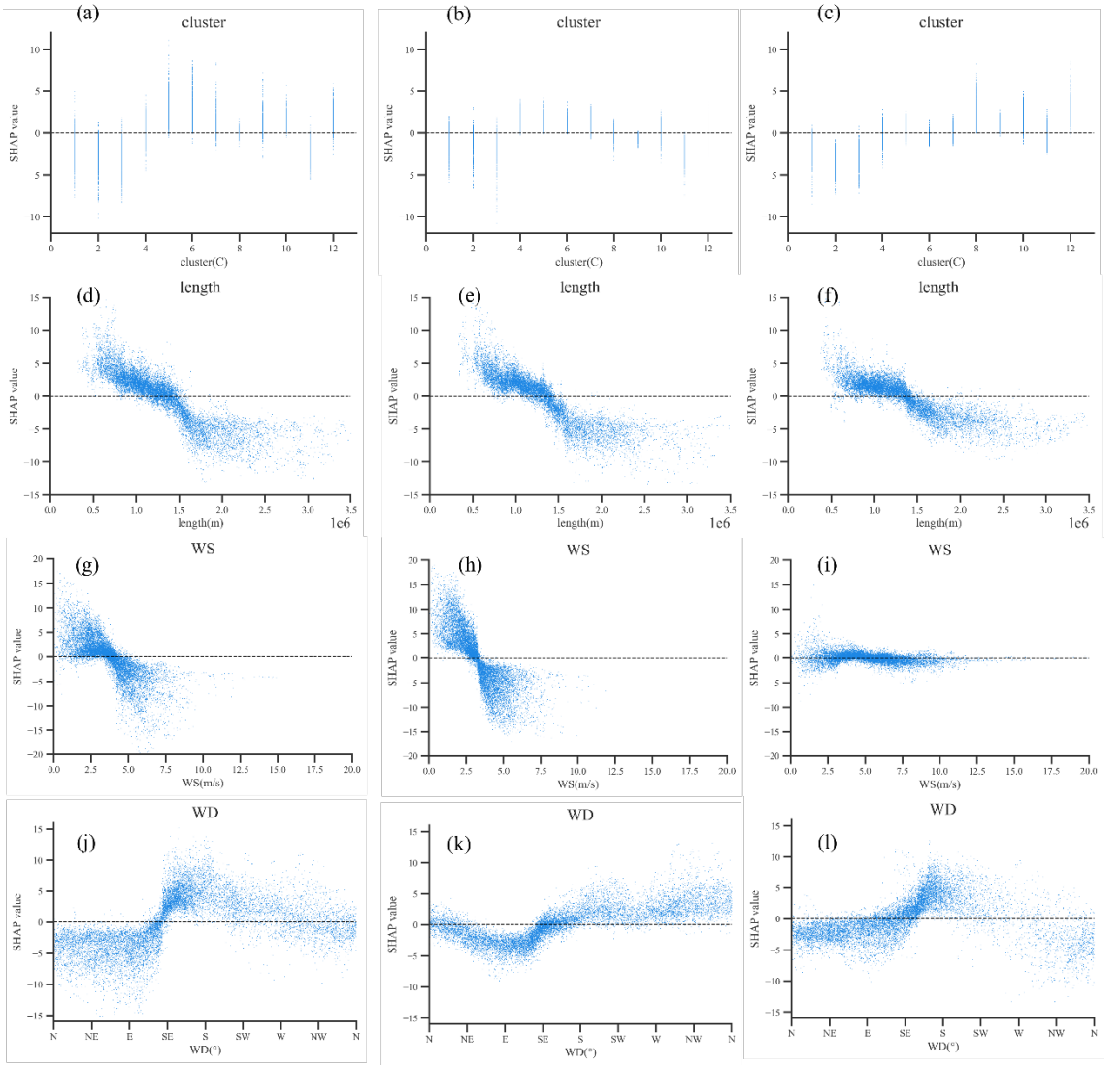

Fig. S13 Influence of horizontal transport at PT (a, d, g, and j), DSL (b, e, h, and k), and PDHN (c, f, i, and l).

Note: 1) It is obvious that the trajectory cluster and wind direction strongly influence the PT station, particularly the north-eastward and south-westward wind directions, which show significant negative and positive contributions, respectively. In contrast, the station's north-eastward and north-westward directions are the Yangtze River Delta hinterland and the East China Sea, respectively. 2) The influence of trajectory and wind direction on the DSL station is substantially lower than on the PT station, however, the importance of wind speed on O_3 around DSL grows dramatically, showing that the diffusion effect has a major influence on DSL O_3.

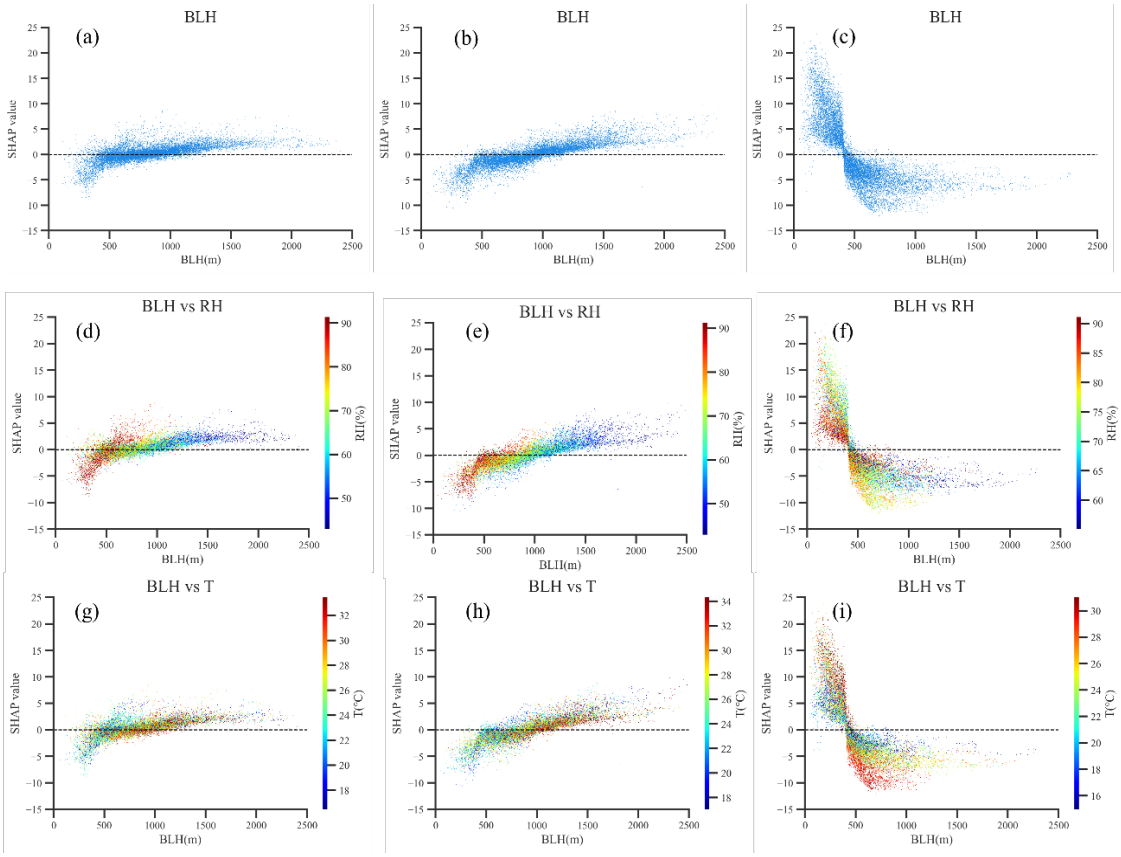

Fig. S14 Influence of BLH at PT (a, d, and g), DSL (b, e, and h), and PDHN (c, f, and i).

Note: The effect of BLH on O_3 at the PDHN station is noticeably different from that at other stations. According to the bivariate dependence plots of RH and T on BLH at the PDHN station, a decrease in RH (increase in T) may cause an increase in O_3 at the low BLH. In contrast, the opposite is true at the low BLH, indicating that a change in BLH may cause a shift in the O_3 formation regime.

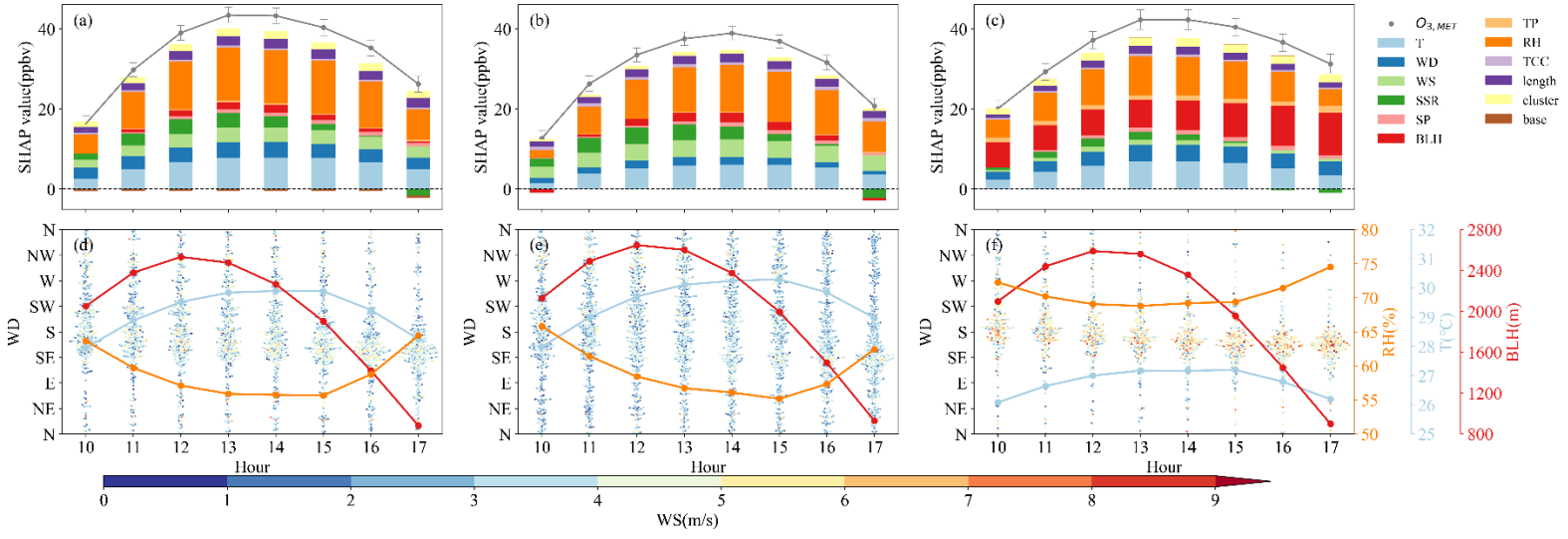

Fig. S15 Diurnal meteorological influence on O_3 at PT (a and d), DSL (b and e) and PDHN (c and f) in polluted day. The error bar in panels a–c represent the 95% confidence intervals of corresponding data. In panels d–f, the color of the scatter represents the wind speed, the position of the scatter in the *y*-axis direction represents the wind direction, and the number of scatters at a position in the *y*-axis direction represents the number of times this wind direction occurs. Base is the base value of the SHAP approach, i.e. the average of the O_{3_MET}.

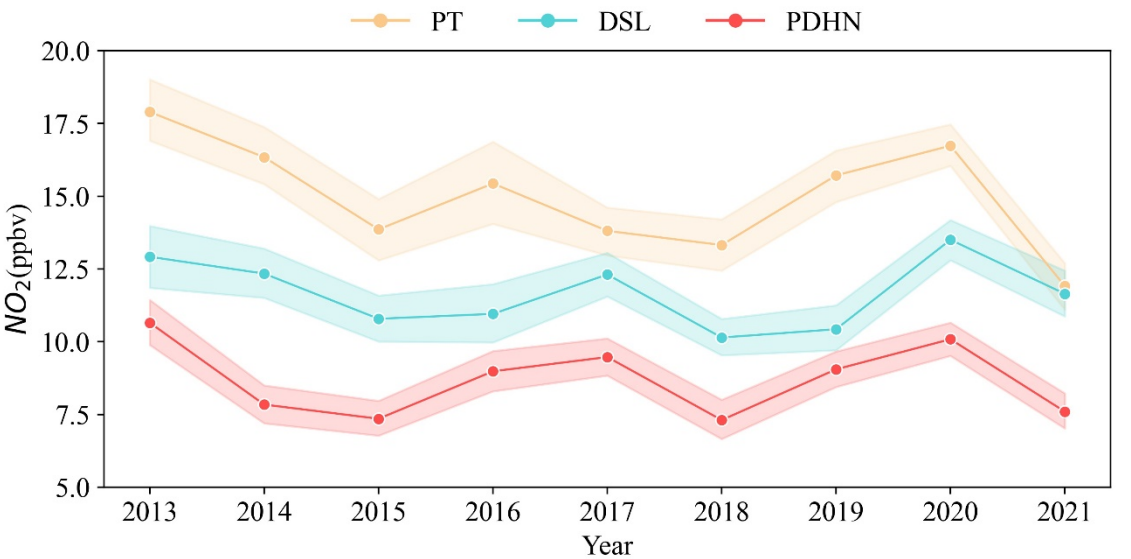

Fig. S16 Annual variation of NO_2 at PT, DSL, and PDHN stations. The shaded area is the standard deviation of corresponding data.

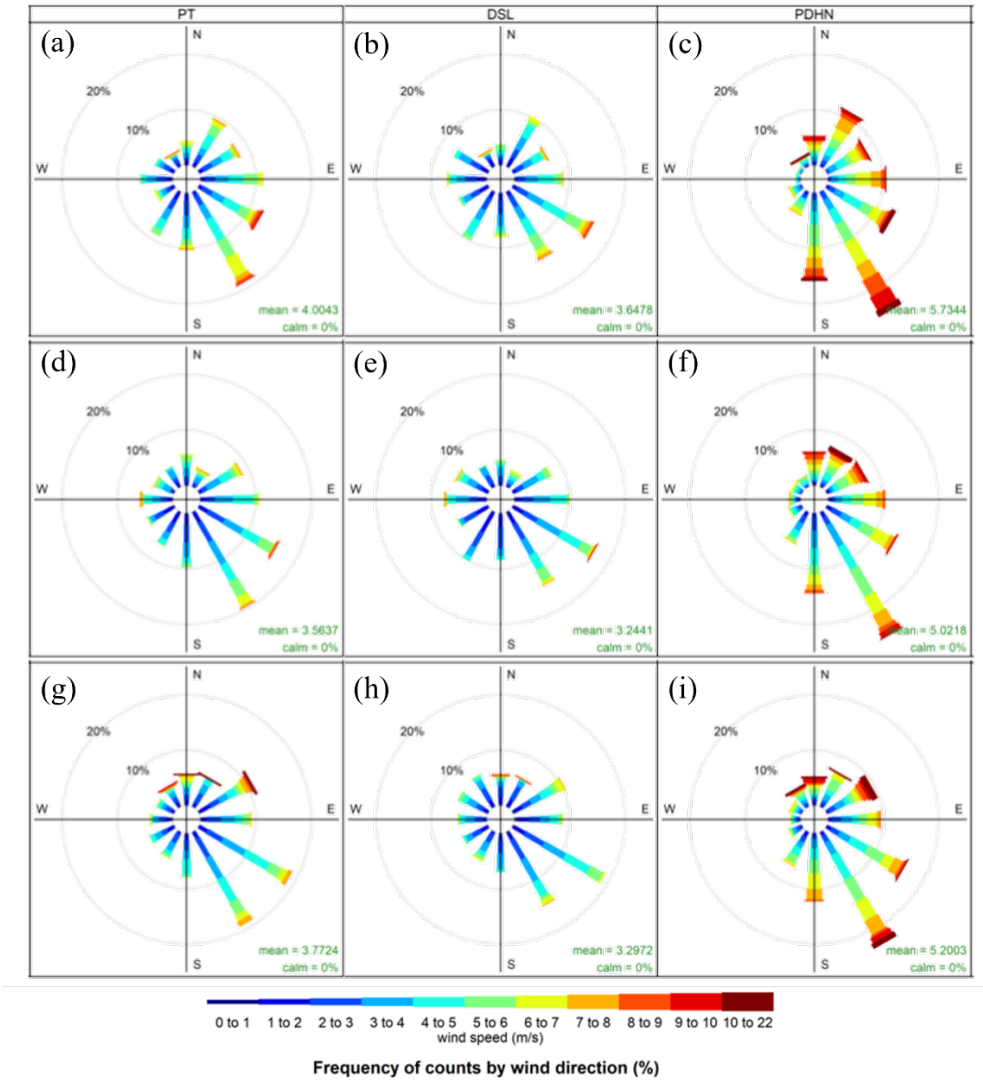

Fig. S17 Wind rose chart for 2013 (a–c), 2017 (d–f) and 2021 (g–i) at PT, DSL and PDHN.

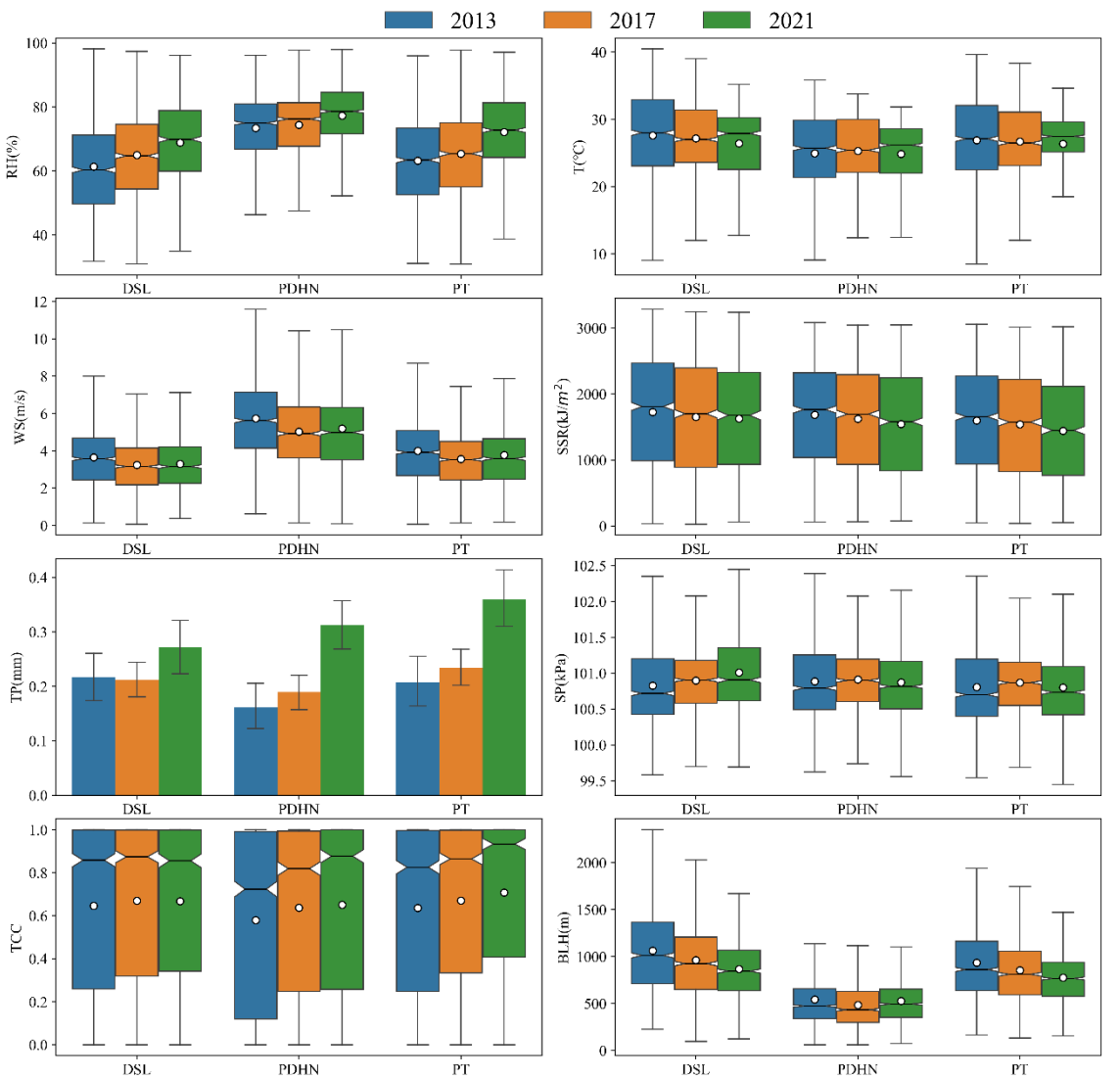

Fig. S18 Comparison of meteorological parameters for 2013, 2017, and 2021 at PT, DSL, and PDHN.

Table S1 Model performance evaluation. FAC2 is fraction of predictions within a factor of two, IOA is index of agreement, MB is mean bias, MGE is mean gross error, MNGE is mean normalized gross error, and NMB is normalized mean bias, A detailed presentation of the model performance metrics can be found in Emery et al. (2017).

Site	RMSE	R^2	MAE	IOA	MB	MGE	MNGE	NMB	FAC2
PT	10.61	0.84	7.78	0.95	−0.42	7.78	0.15	−0.007	0.98
DSL	10.82	0.83	7.85	0.95	−0.37	7.85	0.14	−0.006	0.99
PDHN	8.99	0.85	6.25	0.95	−0.26	6.25	0.12	−0.005	0.99

Table S2 Results of PCA Analysis.

Site	PC	Eigenvalue	Variance contribution (%)	Cumulative variance (%)
PT	PC1	1.980	39.596	39.596
	PC2	1.237	24.747	64.342
DSL	PC1	2.013	40.254	40.254
	PC2	1.193	23.855	64.110
PDHN	PC1	1.931	38.624	38.624
	PC2	1.189	23.787	62.411

Table S3 The loadings or correlations of the components with variables at each site.

Site	PC1					PC2				
	O_3	NO_x	T	WS	WD	O_3	NO_x	T	WS	WD
PT	0.401	−0.409	0.395	0.095	−0.107	0.294	0.309	0.258	−0.705	0.250
DSL	0.422	−0.384	0.407	0.047	−0.060	0.238	0.337	0.200	−0.731	0.306
PDHN	0.361	−0.427	0.380	0.032	−0.245	0.335	0.285	0.223	−0.732	0.248

Table S4 |SHAP value| at PT, DSL and PDHN.

No.	PT		DSL		PDHN	
	factors	|SHAP value| (ppb)	factors	|SHAP value| (ppb)	factors	|SHAP value| (ppb)
1	RH	10.77	RH	10.37	RH	7.65
2	T	4.58	T	5.10	BLH	5.38
3	WD	3.75	WS	4.85	T	2.87
4	Length	3.13	Length	2.97	WD	2.72
5	WS	3.10	SSR	2.36	Length	2.18
6	SSR	2.01	WD	2.31	Cluster	1.54
7	Cluster	1.62	BLH	1.71	TP	1.50

8	BLH	1.16	Cluster	0.92	SSR	0.96
9	SP	0.82	SP	0.90	WS	0.69
10	TCC	0.45	TCC	0.80	SP	0.67
11	TP	0.37	TP	0.40	TCC	0.46

Notes: a) The sum of |SHAP value| of all meteorological factors characterizing the photochemical ozone generation, including RH, T, SSR, TCC, and TP, determines the extent of OPG's impact on ozone concentration, which are 18.18, 19.03, and 13.44 in PT, DSL, and PDHN respectively. b) The sum of |SHAP value| of all meteorological factors characterizing the ozone dispersion conditions, including WD, Length, WS, Cluster, BLH, and SP, determines the extent of ODC's impact on ozone concentration, which are 13.58, 13.66, and 13.18 in PT, DSL, and PDHN respectively. c) The sum of |SHAP value| of all meteorological factors, determines the extent of the impact of all meteorological factors on ozone concentration, which are 31.76, 32.69, and 26.62 in PT, DSL, and PDHN respectively.

Table S5 Comparison of trajectory clustering for 2013, 2017, and 2021 at PT.

Cluster	Frequency					Average length (km)				
	2013	2017	2021	2017–2013	2021–2017	2013	2017	2021	2017–2013	2021–2017
C1	309	425	273	37.5%	−35.8%	1121.9	1109.4	1141.9	−1.1%	2.9%
C2	172	143	195	−16.9%	36.4%	1565.5	1517.4	1496.6	−3.1%	−1.4%
C3	182	187	269	2.7%	43.9%	1275.6	1329.3	1315.5	4.2%	−1.0%
C4	268	307	290	14.6%	−5.5%	1152.6	1142.0	998.4	−0.9%	−12.6%
C5	155	396	290	155.5%	−26.8%	955.7	817.6	943.0	−14.4%	15.3%
C6	221	219	179	−0.9%	−18.3%	1183.5	1018.1	1072.8	−14.0%	5.4%
C7	264	153	280	−42.0%	83.0%	1707.8	1475.2	1423.8	−13.6%	−3.5%
C8	171	92	123	−46.2%	33.7%	2398.1	2851.5	2838.9	18.9%	−0.4%
C9	271	276	154	1.8%	−44.2%	1857.0	1674.1	1773.7	−9.8%	5.9%
C10	314	152	121	−51.6%	−20.4%	1815.4	1587.5	1416.8	−12.6%	−10.8%
C11	135	75	320	−44.4%	326.7%	1796.2	1815.5	1936.7	1.1%	6.7%
C12	229	266	197	16.2%	−25.9%	1498.6	1420.5	1284.0	−5.2%	−9.6%

Table S6 Comparison of trajectory clustering for 2013, 2017, and 2021 at DSL.

Cluster	Frequency					Average length(km)				
	2013	2017	2021	2017–2013	2021–2017	2013	2017	2021	2017–2013	2021–2017
C1	312	357	229	14.4%	−35.9%	1156.1	1084.7	1139.4	−6.2%	5.0%
C2	168	125	206	−25.6%	64.8%	1554.5	1526.8	1471.7	−1.8%	−3.6%
C3	213	216	249	1.4%	15.3%	1287.7	1267.7	1280.7	−1.6%	1.0%
C4	143	326	326	128.0%	0.0%	946.3	835.6	924.6	−11.7%	10.7%
C5	212	251	192	18.4%	−23.5%	1139.0	972.6	1081.3	−14.6%	11.2%
C6	311	308	183	−1.0%	−40.6%	1889.8	1703.3	1865.4	−9.9%	9.5%

C7	224	301	317	34.4%	5.3%	1049.7	1172.9	968.4	11.7%	−17.4%
C8	279	143	226	−48.7%	58.0%	1680.2	1505.2	1471.0	−10.4%	−2.3%
C9	125	68	97	−45.6%	42.6%	2374.5	2821.4	2901.5	18.8%	2.8%
C10	260	355	282	36.5%	−20.6%	1227.9	1113.9	1134.5	−9.3%	1.9%
C11	138	81	309	−41.3%	281.5%	1786.8	1752.9	1917.6	−1.9%	9.4%
C12	306	160	75	−47.7%	−53.1%	1982.0	1808.1	1582.1	−8.8%	−12.5%

Table S7 Comparison of trajectory clustering for 2013, 2017, and 2021 at PDHN.

Cluster	Frequency					Average length(km)				
	2013	2017	2021	2017–2013	2021–2017	2013	2017	2021	2017–2013	2021–2017
C1	249	257	164	3.2%	−36.2%	1104.6	1197.6	1243.9	8.4%	3.9%
C2	184	201	272	9.2%	35.3%	1268.5	1325.5	1308.7	4.5%	−1.3%
C3	279	153	273	−45.2%	78.4%	1786.2	1550.4	1534.0	−13.2%	−1.1%
C4	202	271	266	34.2%	−1.8%	1126.0	1037.1	908.7	−7.9%	−12.4%
C5	270	332	305	23.0%	−8.1%	1125.3	1081.0	1115.8	−3.9%	3.2%
C6	171	101	112	−40.9%	10.9%	2389.1	2813.8	2796.9	17.8%	−0.6%
C7	286	244	153	−14.7%	−37.3%	1935.6	1747.0	1898.5	−9.7%	8.7%
C8	123	258	201	109.8%	−22.1%	1140.3	906.0	1038.7	−20.6%	14.6%
C9	156	158	158	1.3%	0.0%	1645.4	1554.0	1608.2	−5.6%	3.5%
C10	257	433	302	68.5%	−30.3%	1241.8	1134.6	1163.1	−8.6%	2.5%
C11	160	81	332	−49.4%	309.9%	1816.6	1817.2	1961.8	0.0%	8.0%
C12	354	202	153	−42.9%	−24.3%	1805.6	1558.5	1459.3	−13.7%	−6.4%

References

Casallas A, Castillo-Camacho M P, Sanchez E R, González Y, Celis N, Mendez-Espinosa J F, Belalcazar L C, Ferro C (2023). Surface, satellite ozone variations in Northern South America during low anthropogenic emission conditions: a machine learning approach. Air Quality, Atmosphere & Health, 16(4): 745–764

Chong H, Lee S, Cho Y, Kim J, Koo J H, Pyo Kim Y, Kim Y, Woo J H, Hyun Ahn D (2023). Assessment of air quality in Korea from satellite observations. Environment International, 171: 107708

Emery C, Liu Z, Russell A G, Odman M T, Yarwood G, Kumar N (2017). Recommendations on statistics and benchmarks to assess photochemical model performance. Journal of the Air & Waste Management Association, 67(5): 582–598

Fan H, Wang Y, Zhao C, Yang Y, Yang X, Sun Y, Jiang S (2021). The role of primary emission and transboundary transport in the air quality changes during and after the COVID-19 lockdown in China. Geophysical Research Letters, 48(7): e2020GL091065

Garces H, Sbarbaro D (2011). Outliers detection in environmental monitoring databases. Engineering Applications of Artificial Intelligence, 24(2): 341–349

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan R J, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J N (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730): 1999–2049

Hu W, Zhao T, Bai Y, Kong S, Xiong J, Sun X, Yang Q, Gu Y, Lu H (2021). Importance of regional PM2.5 transport and precipitation washout in heavy air pollution in the Twain-Hu Basin over Central China: Observational analysis and WRF-Chem simulation. Science of the Total Environment, 758: 143710

Junninen H, Niska H, Tuppurainen K, Ruuskanen J, Kolehmainen M (2004). Methods for imputation of missing values in air quality data sets. Atmospheric Environment, 38(18): 2895–2907

Kong D, Ning G, Wang S, Cong J, Luo M, Ni X, Ma M (2021). Clustering diurnal cycles of day-to-day temperature change to understand their impacts on air quality forecasting in mountain-basin areas. Atmospheric Chemistry and Physics, 21(19): 14493–14505

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

Liu H, Han X, Tang G, Zhang J, Xia X A, Zhang M, Meng L (2022). Model analysis of vertical exchange of boundary layer ozone and its impact on surface air quality over the North China Plain. Science of the Total Environment, 821: 153436

Liu N, Zou B, Li S, Zhang H, Qin K (2021). Prediction of PM2.5 concentrations at unsampled points using multiscale geographically and temporally weighted regression. Environmental Pollution, 284: 117116

Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, Boussetta S, Choulga M, Harrigan S, Hersbach H, Martens B, Miralles D G, Piles M, Rodríguez-Fernández N J, Zsoter E, Buontempo C, Thépaut J N (2021). ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth System Science Data, 13(9): 4349–4383

Parida B R, Bar S, Roberts G, Mandal S P, Pandey A C, Kumar M, Dash J (2021). Improvement in air quality and its impact on land surface temperature in major urban areas across India during the first lockdown of the pandemic. Environmental Research, 199: 111280

Purwanto P, Astuti I S, Rohman F, Utomo K S B, Aldianto Y E (2022). Assessment of the dynamics of urban surface temperatures and air pollution related to COVID-19 in a densely populated City environment in East Java. Ecological Informatics, 71: 101809

Ramon J, Lledó L, Torralba V, Soret A, Doblas-Reyes F J (2019). What global reanalysis best represents near-surface winds? Quarterly Journal of the Royal Meteorological Society, 145(724): 3236–3251

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

Storey M A, Price O F (2022). Statistical modelling of air quality impacts from individual forest fires in New South Wales, Australia. Natural Hazards and Earth System Sciences, 22(12): 4039–4062

Suciu L G, Griffin R J, Masiello C A (2017). Regional background O3 and NOx 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, Kang Y, Zhu Z, Wang M, Wang Y, Li T, Li X, Wang Y (2021). Aggravated ozone pollution in the strong free convection boundary layer. Science of the Total Environment, 788: 147740

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 M, Tang G, Liu Y, Ma M, Yu M, Hu B, Zhang Y, Wang Y, Wang Y (2021a). The difference in the boundary layer height between urban and suburban areas in Beijing and its implications for air pollution. Atmospheric Environment, 260: 118552

Wang N, Xu J, Pei C, Tang R, Zhou D, Chen Y, Li M, Deng X, Deng T, Huang X, Ding A (2021b). Air quality during COVID-19 lockdown in the Yangtze River Delta and the Pearl River Delta: two different responsive mechanisms to emission reductions in China. Environmental Science & Technology, 55(9): 5721–5730 doi:10.1021/acs.est.0c08383

Wang S, Li S, Xing J, Yang J, Dong J, Qin Y, Sahu S K (2022b). Evaluation of the influence of El Niño–Southern Oscillation on air quality in southern China from long-term historical observations. Frontiers of Environmental Science & Engineering, 16(2): 26

Xu Z, Liu Z, Lu L, Liao W, Yang C, Duan Z, Zhou Q, He W, Zhang E, Li N, Ju K (2022). Assessing the causal effects of long-term exposure to PM2.5 during pregnancy on cognitive function in the adolescence: Evidence from a nationwide cohort in China. Environmental Pollution, 293: 118560