Supplementary Information for

Trends of peroxyacetyl nitrate and its impact on ozone over 2018–2022 in urban atmosphere

Ziyi Lin[1,2,3], Lingling Xu[1,2*], Chen Yang[1,2,3], Gaojie Chen[1,2,3], Xiaoting Ji[1,2,3], Lingjun Li[1,2,3], Keran Zhang[1,2,4], Youwei Hong[1,2], Mengren Li[1,2], Xiaolong Fan[1,2], Baoye Hu[5], Fuwang Zhang[6], Jinsheng Chen[1,2*]

[1]Center for Excellence in Regional Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

[2]Fujian Key Laboratory of Atmospheric Ozone Pollution Prevention, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

[3]University of Chinese Academy of Sciences, Beijing 100049, China

[4]Fujian Agriculture and Forest University, Fuzhou, 350002, China

[5]Minnan Normal University, Zhangzhou 363000, China

[6]Environmental Monitoring Center of Fujian, Fuzhou 350003, China

*Correspondence to: Jinsheng Chen (jschen@iue.ac.cn); Lingling Xu (linglingxu@iue.ac.cn)

Supplementary Notes

1. Calculation of PA and the lifetime of PAN

The major chemical reactions about PA and PAN can be simply summarized as the following five reactions R1−R5.

$$VOCs + OH, O_3, NO_3 \rightarrow CH3C(O)O_2\ (PA) + product \quad \text{(R1)}$$

$$PA + NO_2 \xrightarrow{k1} PAN \quad \text{(R2)}$$

$$PAN \xrightarrow{k2} PA + NO_2 \quad \text{(R3)}$$

$$PA + NO \xrightarrow{k3} CH_3 \cdot + NO_2 + CO_2 \quad \text{(R4)}$$

$$PAN + OH \xrightarrow{k4} HCHO + NO_2 + CO \quad \text{(R5)}$$

As shown in R1, PA radicals are mainly produced from the oxidation of VOCs (OVOCs) by OH, O_3 and NO_3, especially during the daytime by OH. [1] PA radicals can react with NO_2 to form PAN (R2), which is the only way for PAN formation. The thermal decomposition of PAN will release PA radicals (R3) and it is also the main loss pathway of PAN [2]. And PA radicals react with NO (R4) as permanent loss [3]. In the steady-state condition, PA radicals can be estimated as[4]:

$$[PA]_{calculated} = \frac{\frac{d[PAN]}{dt} + k_2[PAN]}{k_1[NO_2]}$$

In the case of only considering R2 to R4 without considering other removal pathways, the lifetime $(\boldsymbol{\tau})$ of PAN could be derived as following:[5]

$$\tau(PAN) = \frac{1}{k_2}\left(1 + \frac{k_1}{k_3 \frac{NO}{NO_2}}\right)$$

Where k_1 is 9.5×10^{-12} cm³/molecule/s, the thermal decomposition coefficient rate k_2 is $2.52\times1016\ \exp^{(-13573/T)}$/s, and k_3 is 1.85×10^{-11} cm³/molecule/s [6, 7].

2. The principle of Kolmogorov-Zurbenko Adaptive filter method

The KZA filter is a kind of time series method based on an iterative moving average KZ [8]. The main parameters of KZ (m, n) are the number of applications moving

average n (iterative times) and the moving average points m. The process can be expressed as:

$$Y_i = \frac{1}{2k+1} \sum_{j=-k}^{k} X_{i+j}$$

Where X_{i+j} is the input data, Y_i is the result of applying the moving average to the input data, which then becomes the input for the next iteration of averaging, k is the half-length of m.

Based on the KZ process as mentioned above, KZA filter introduces the rate of change for dynamically adjusting the length of average window (parameter m) to avoid removing some sharp changes.[9] Furthermore, the time series can be partitioned as:

$$R(t) = W(t) + S(t) + e(t)$$

Where $R(t)$ is raw time series, $W(t)$ is short-term variation, $S(t)$ is seasonal change and $e(t)$ is long-term trend.

According to a previous study,[10] the sum of $S(t)$ and $e(t)$ can be captured by using KZ (15, 5), and $e(t)$ can be estimated by KZ (365, 3). Thus, $W(t)$ and $S(t)$ can be calculated as follows: $W(t)$ is obtained by subtracting KZ (15, 5) from $R(t)$, and $S(t)$ is obtained by subtracting and subtracting KZ (365, 3) from KZ (15, 5). In this work, the KZ filter is replaced by KZA filter and the setting is implemented like the previous study.[10]

3. Factor Selection

Based on the physicochemical meaning, 11 factors including nitric dioxide (NO_2), nitric oxide (NO), ozone (O_3), relative humidity (RH), carbon monoxide (CO), temperature (T), ultraviolet radiation (UV), 10m v-component of wind (V_{10}), 10m u-component of wind (U_{10}), surface pressure (SP), and total precipitation (TP) were considered in this study. Since NO_2 can react with PA radicals to promote PAN formation and NO can consume PA radicals, the factor NO_2 and the factor NO can be characterized as the level of PAN precursors and the potential loss capability through

NO pathway, respectively. CO is usually considered to be produced from primary emissions. The variation of CO could reflect the primary emission intensity to some extent. The factor O_3 was used as a proxy of atmospheric oxidation capacity in this work. The factors RH, T and UV influence the photochemical generation of PAN by affecting liquid phase chemistry, thermal decomposition, and photochemical generation of PAN. These three factors could be regarded as meteorological factors that can affect the atmospheric chemical reactions. As for SP, TP, U_{10} and V_{10}, they affected PAN levels through physical processes. The factor SP reflects the diffusion condition, the factor TP shows the ability of physical removal, and factor U_{10} as well as factor V_{10} represent the transport of PAN. In addition, all factors were tested for collinearity using Variance Inflation Factor (VIF) method, and the VIF values were less than 1.

4. Machine learning.

4.1 Detailed information about RF model and XGBoost model.

(1) Random Forest (RF) Model

The RF algorithm applies the bootstrap aggregating (bagging) ensemble approach which is based on decision trees learner and introduces random selection of predictors in the decision trees training process.[11] This method improves predictive performance and reduces the overfitting by creating many individual decisions trees for predictions and averaging all the predictions for the final prediction. The main advantage of RF is that it could be easier to investigate and interpret compared with other ML techniques like artificial neural networks.[12] In this work, RF regression model was conducted to explore the complex nonlinear relationship between PAN and its influencing factors by using “scikit-learn” library[13] in python environment.

(2) XGBoost model

The XGBoost algorithm, a kind of tree ensemble model like RF, optimizes the gradient boosted decision tress (GBDP) algorithm to deal with classification or

regression problem.[14] Its advantages include handling missing value, consuming less computing time, and overcoming overfitting. Unlike RF model, XGBoost obtains more accurate results by accumulating and progressively optimizing the prediction impact of each decision tree. XGBoost regression model also conducted to investigate PAN by using "xgboost" library (https://github.com/dmlc/xgboost/tree/master) in python environment in this work.

5. Model Setting

The input of model are hourly data during observation in UV>0 condition. When the model was being adjusted, 80% of the data was used as the training set, and 20% of the data was used as the test set. The hyperparameters were tuned using grid search and cross-validation method. Specifically, for a single hyperparameter, grid search was employed to determine its optimal value range. For combinations of hyperparameters, the entire training set was split into five folds. A grid search was then conducted over the pre-adjusted combinations of hyperparameters by training on four folds and predicting on the fifth fold, following a cross-validation procedure. The hyperparameters of the final model were determined according to the R^2 function score. To retain as many time periods of data as possible in the characteristic analysis of PAN, all available data was used as input for the pre-trained machine learning model. For main hyperparameters of RF model, the number of decision trees was set as 300, max depth was 8. For key hyperparameters of XGBoost model, the number of trees was 150, learning rate was 0.07, max depth was 7 and the regularization parameter lambda was set as 1.38. When segregating the train and test sets, the RF model achieved an R-squared value of 0.68, while the XGBoost model achieved an R-squared value of 0.81. When using the entire dataset for modeling, the performance of both models is illustrated in **Supplementary Fig. 11**.

6. Shapley Additive Explanation theory

In this study, the target of SHAP method is to explain the variation of prediction by computing the contribution of each feature to the variation of prediction. Based on the coalitional game theory,[17] SHAP value can fairly allocate the total concentration of anomalous values across different features. In the SHAP theory, each sample x_i containing M features produces a predicted value $g(x_i)$, the feature weighted linear model as:

$$g(x_i) = \phi(x)_{base} + \sum_{j=1}^{M} \phi(x_{i,j})$$

Where $\phi(x)_{base}$, the base value, represents the expected value of the model output over the whole data set. $\phi(x_{i,j})$ is the SHAP value of the feature j in sample x_i showing its impact on the prediction.

In this work, if $\phi(x_{i,j})$>0, it means that the feature j of sample x_i has a positive impact on PAN relative to the base value. While conversely, it exerts a negative impact.

Additionally, the SHAP value method is capable of assessing not only the primary effects of each feature but also the interactions among them. As for variable a and variable b, the SHAP interaction values among them is defined as:

$$\phi_{a,b}(x_{i,j}) = \sum_{S \subseteq K \setminus \{a,b\}} \frac{|S|!\,(K - |S| - 2)!}{2(K-1)!} \nabla_{ab}(x_{i,j}, S)$$

When $a \neq b$, and

$$\nabla_{ab}(x_{i,j}, S) = f_x(S \cup \{a,b\}) - f_x(S \cup \{a\}) - f_x(S \cup \{b\}) + f_x(S)$$

Where $S \subseteq \{0,1\}^K$ and K is the set of all input factors. $|S|$ is the number of non-zero entries in S, The SHAP interaction value between *a* and *b* is split equally for each variable, thus $\phi_{a,b}(x_{i,j}) = \phi_{b,a}(x_{i,j})$. The SHAP interaction value can be interpreted as the difference between the SHAP values of variable a when variable b is present and the SHAP values for variable a when b is absent.[18]

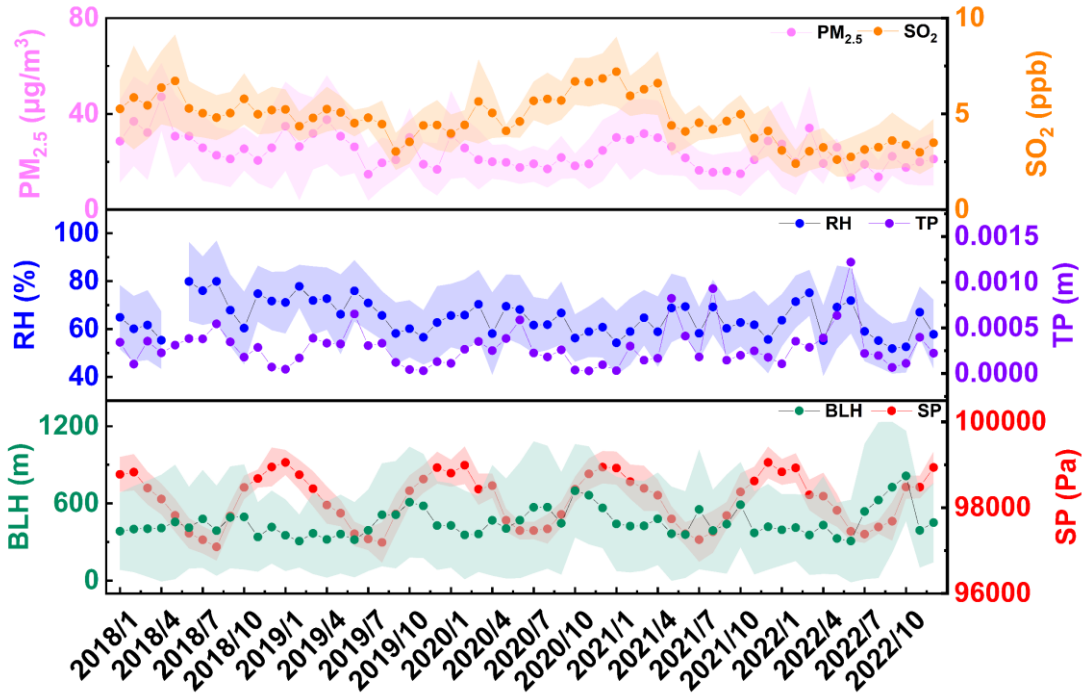
7. Detailed setting of F0AM model

In this study, a chemical box model was developed by the Framework for 0-D Atmospheric Modeling (F0AM) to investigate the formation of PAN and its impact on

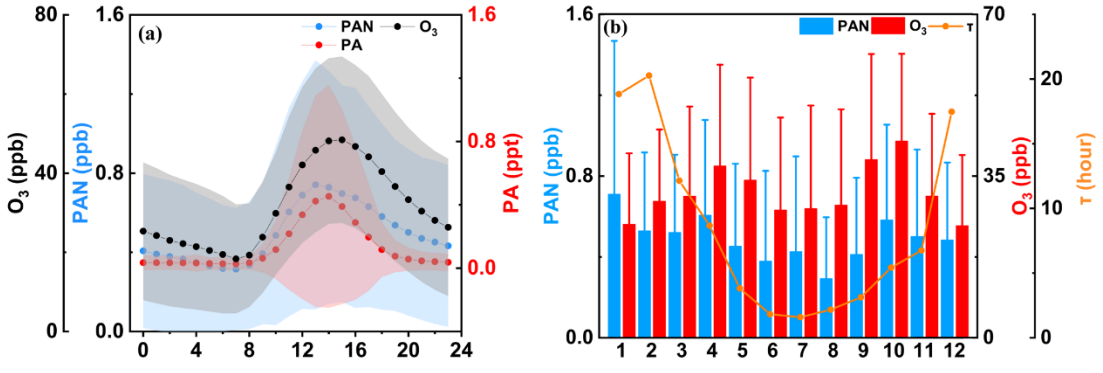
O_3. Measured hourly interval data of trace gases (VOCs, PAN, O_3, NO, NO_2, CO, SO_2), meteorological variable (T, RH, P, and photolysis frequencies) and reanalysis data (BLH) were used as constraints in the box model. It should be noted that the VOCs constraints in this model were only considered if they were MCM species and had valid values. And the descriptive statistics of the measured VOCs are shown as **Supplementary Table 3**. The photolysis rate constants parameterized by the model were corrected using the observed and simulated photolysis rate of NO_2. To better simulate the concentration of intermediate species, 3-day spin-up was set in all model simulation. Physical loss including dilution and dry deposition was also considered in the model, the basic dilution rate was 6×10^{-5} s^{-1} adjusting by the BLH, and the dry deposition velocity for constrained species was set as Liu et al.[19] The PAN-related mechanism herein consists of reactions R2, R3 and R5 presented in **Supplementary Note 1**. Additionally, the heterogeneous HONO mechanism presented in **Supplementary Table S4** was also incorporated to better simulate the photochemistry process.

8. Theil-Sen trend test method

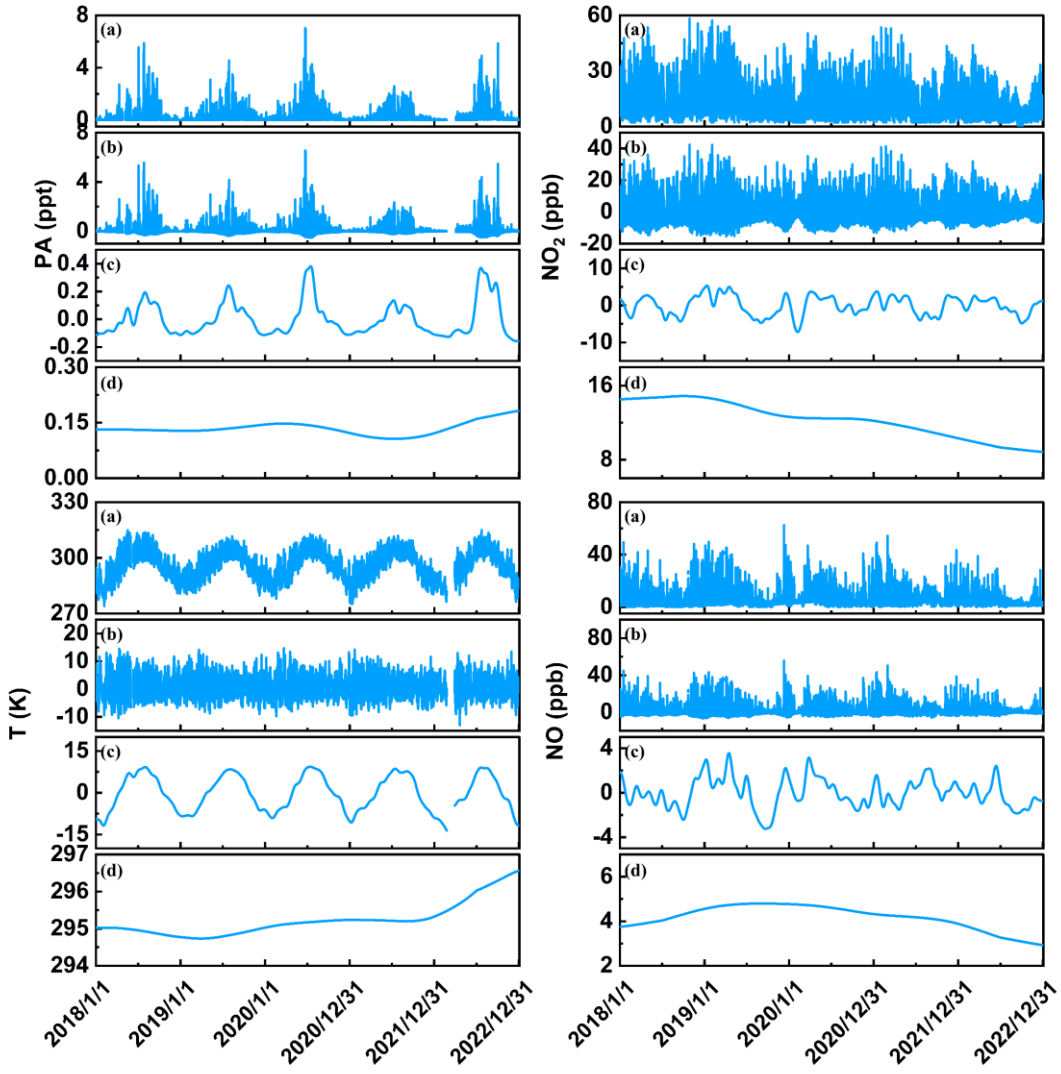
The Theil-Sen method obtain the variation rate by calculating all slopes between all pairs of points and selecting the median slope among all these slopes.[20, 21] Its advantages include yielding accurate confidence intervals even with non-normal data and heteroscedasticity, resisting the outliers, and getting robust result by bootstrap-resampling. Bootstrap resampling also provides the interval of slope and the significant test result of the slope. In this study, the Theil-Sen estimator was implemented by Theil-Sen function in the openair package in RStudio.



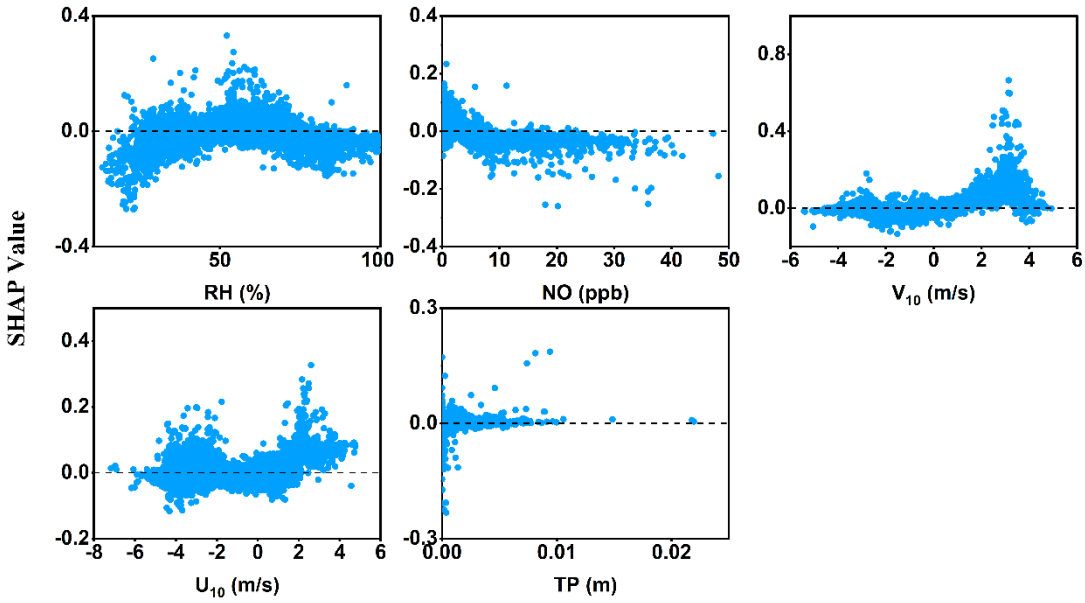
Supplementary Figure 1. Time series of monthly mean value of pollutants and meteorological parameters during the observation period 2018–2022. Shade represents a standard deviation.



Supplementary Figure 2. Diurnal patterns of PAN, PA, and O_3 with a standard error (shaded) **(a)** and seasonal patterns of PAN, O_3 and τ(PAN) with a standard deviation **(b)**. Results are all five-year averages.



Supplementary Figure 3. The summary of KZA filter results of PA, NO_2, T, and NO. (a) original time series; **(b)** short-term variations; **(c)** seasonal variations; **(d)** long-term trend.

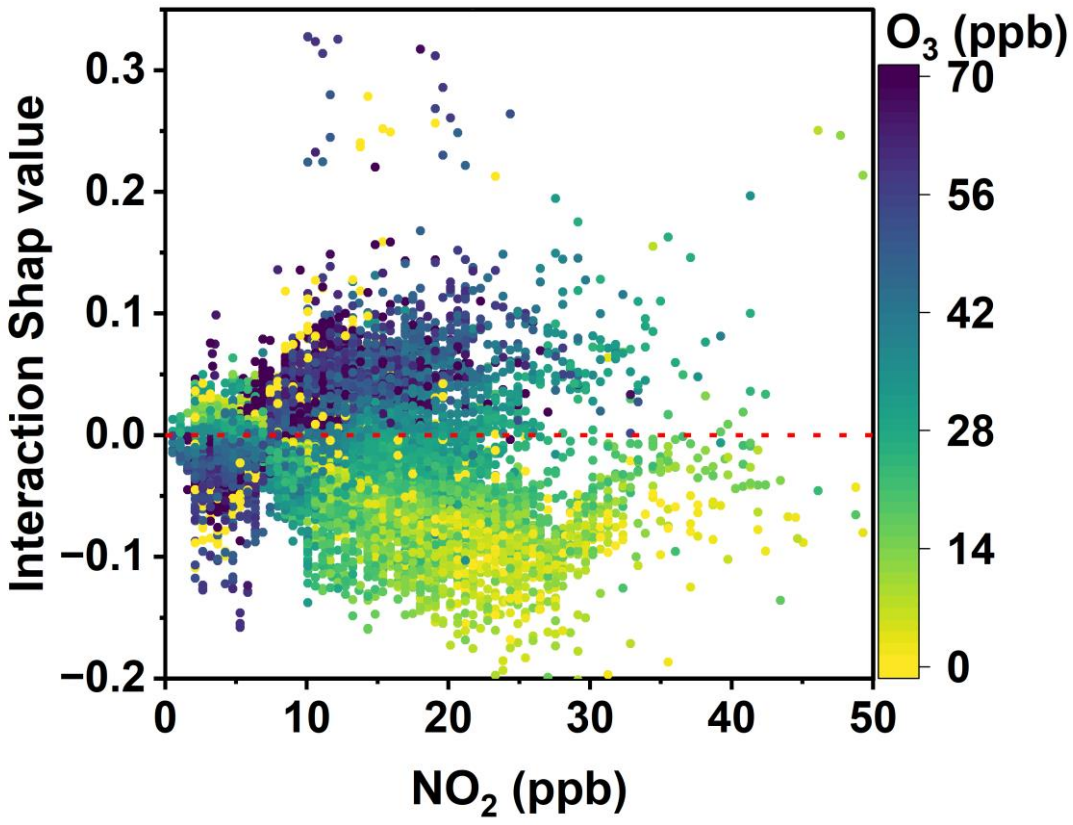


204

205 **Supplementary Figure 4. Main effect of other factors on PAN** obtained by XGBoost-

206 SHAP method.

207



208

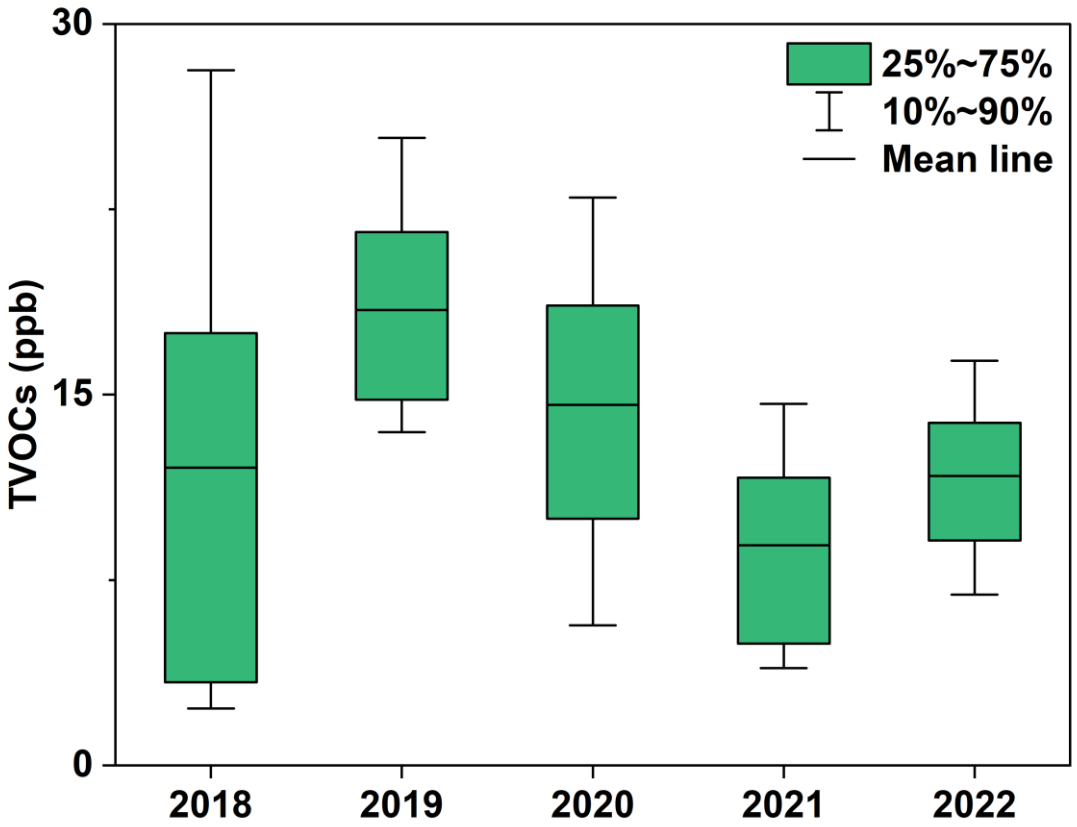
209 **Supplementary Figure 5. The bivariate interaction scatter plot.** The Y-axis of

210 interaction scatter plot is the SHAP values of PAN calculated by features of NO_2 and

211 O_3. Color depth represents the magnitude of feature value and red dash line indicates

212 the interaction of SHAP value is 0.

213

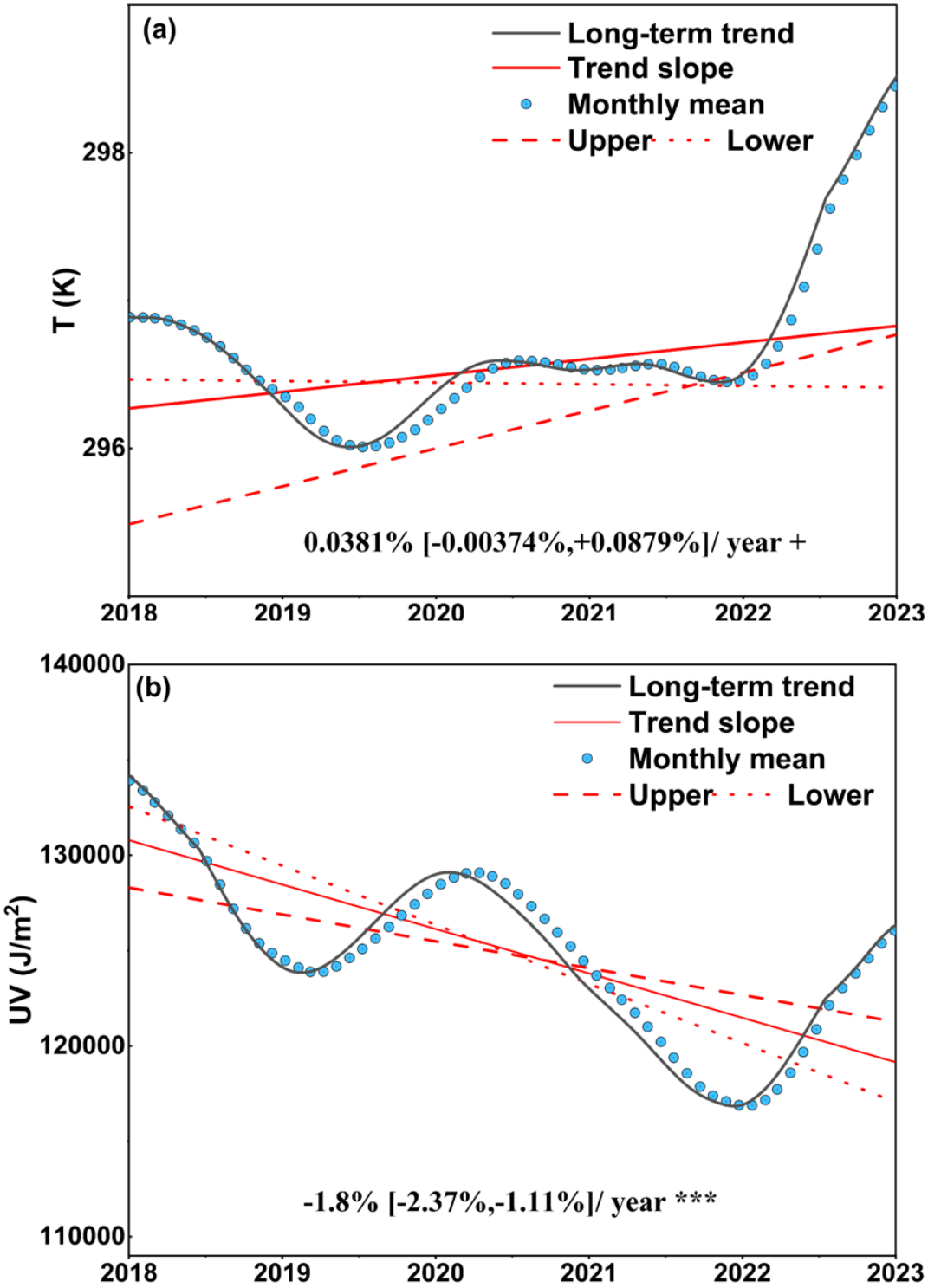


214

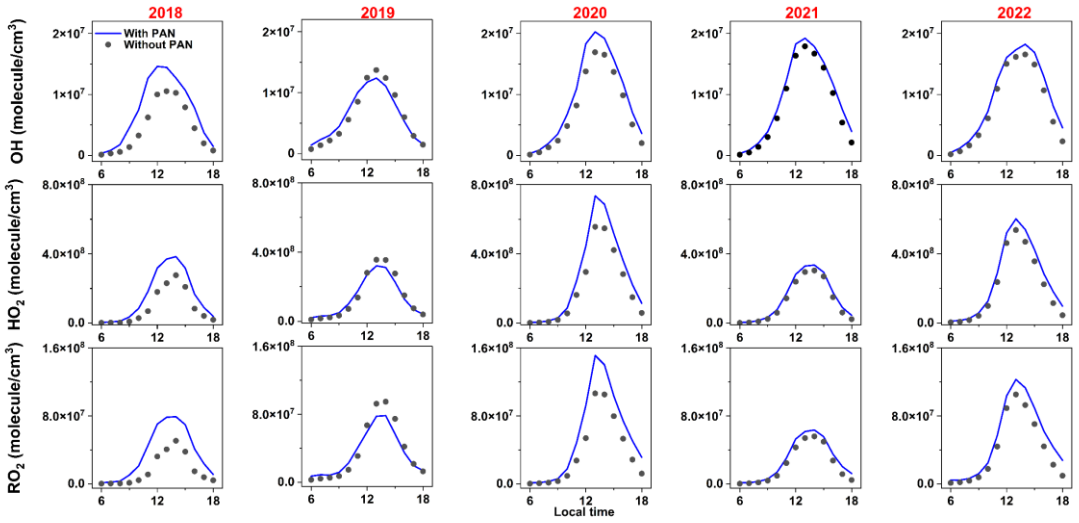
215 **Supplementary Figure 6. Box plots of total VOCs (TVOCs) concentrations** during

216 each summer of 2018–2022.

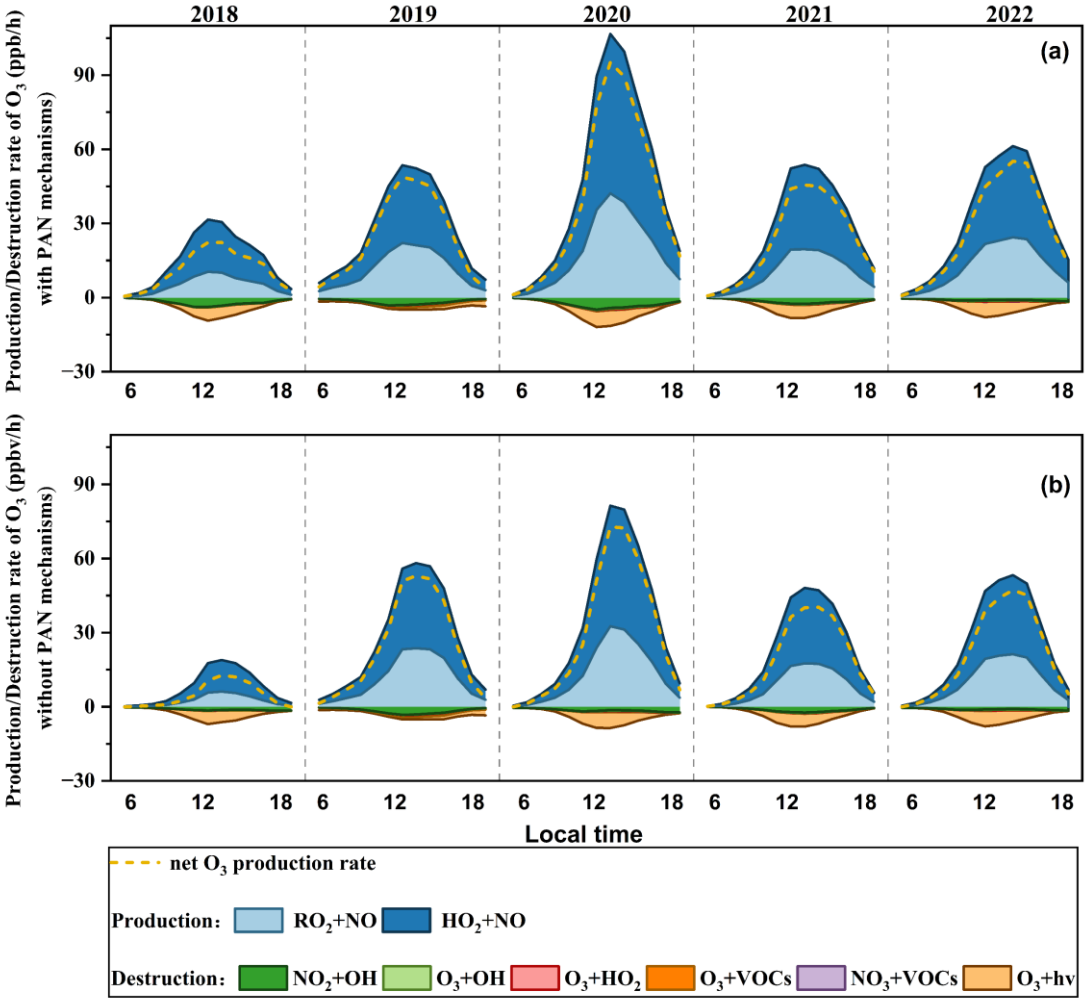
217



Supplementary Figure 7. Long-term trend of T (a) and UV (b) in UV > 0 condition. The long-term trend was decomposed from original time series by KZA filter. The + and *** show the trend significance level $p < 0.1$ and $p < 0.001$, respectively.

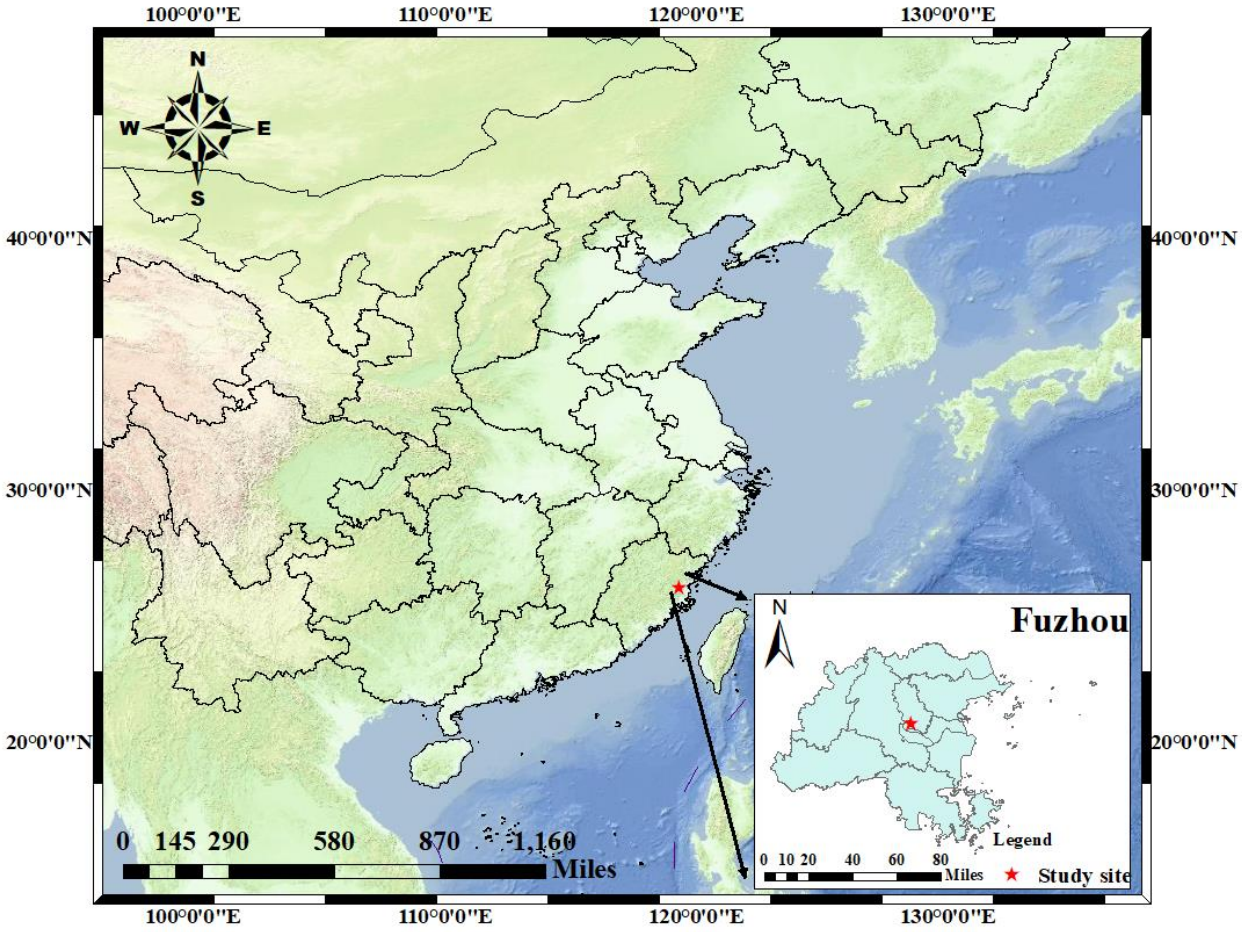


Supplementary Figure 8. Simulation results of RO*x* with and without PAN mechanisms. The simulated levels of OH, HO_2 and RO_2 from 2018 to 2022.

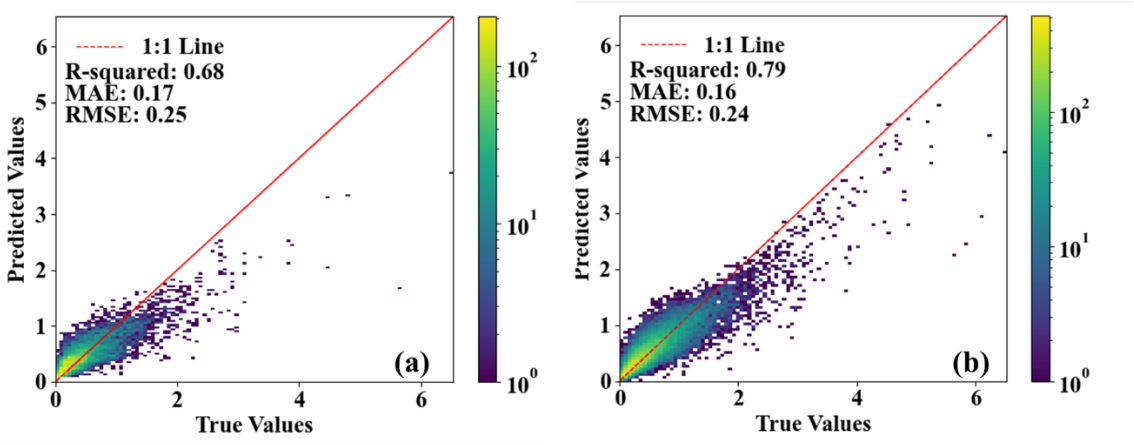


Supplementary Figure 9. Simulated formation and loss rates of O_3 with **(a)** and

without **(b)** PAN mechanisms from 2018 to 2022.



Supplementary Figure 10. Measurement site located at Environmental Monitoring Center Station of Fujian Province (26°6′29″N, 119°18′6″E) in Fuzhou city of Southeast China.

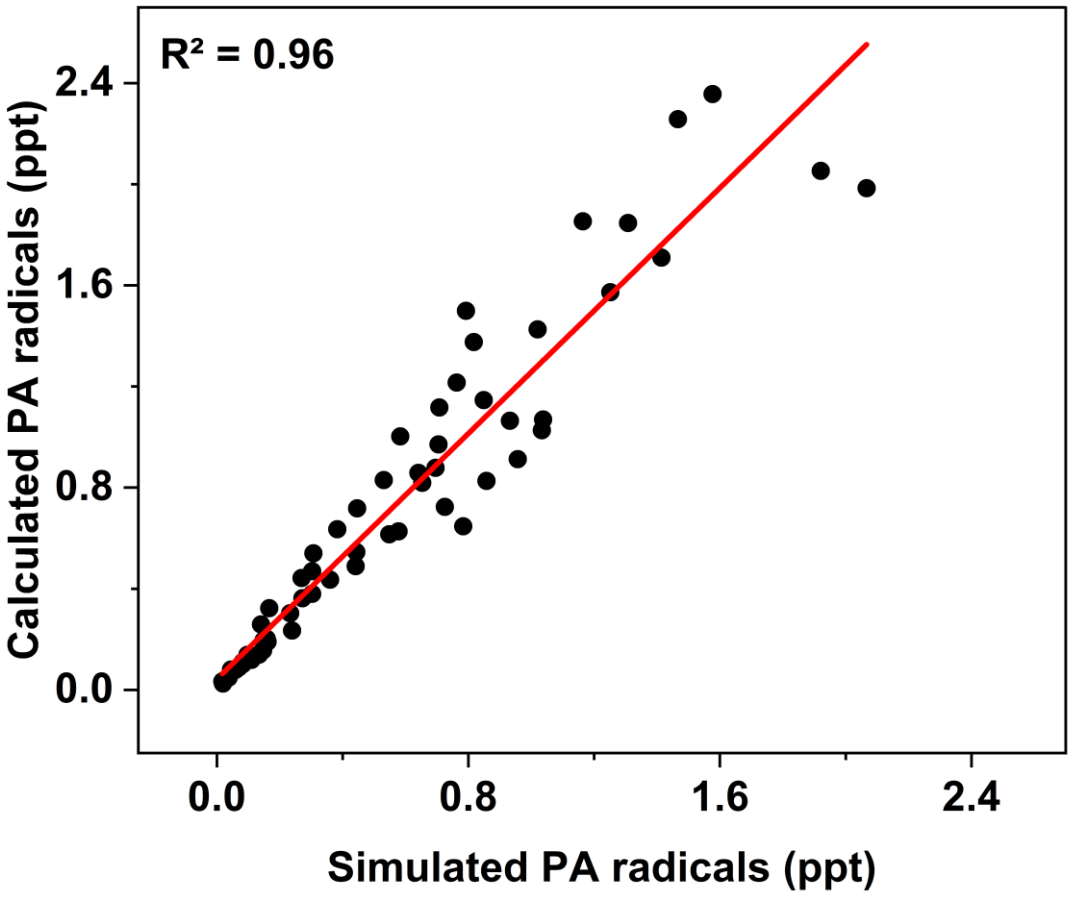


235

236 **Supplementary Figure 11. Model performance** of **(a)** RF model and **(b)** XGBoost

237 model. The color bar is colored with number density.

238



239

240 **Supplementary Figure 12. The performance of F0AM Model.** The relationship

241 between calculated PAN and simulated PA.

242

243 **Supplementary Table 1.** PAN concentrations at different types of sites in China and
244 other countries.

Region	Site type	Study period	Mean PAN (ppb)	Reference
Fuzhou, China	Urban	Jan, 2018 – Dec, 2022	0.50 (0.20 - 1.50)	This study
Beijing, China	Urban	Aug, 2022 – Sep, 2022	1.00 ± 0.97	[22]
Shenzhen, China	Urban	Oct, 2021 – Sep, 2022	0.54	[15]
Tianjin, China	Urban	Jul, 2017 – Aug, 2017	0.73	[23]
Jinan, China	Urban	Nov, 2015 – Jul, 2016	1.89 ± 1.42	[5]
Zhengzhou, China	Urban	Sep, 2018 – Aug, 2019	1.96 ± 1.55	[24]
Shanghai, China	Urban	Dec, 2016 – Nov, 2017	1.30	[25]
Zhengzhou, China	Suburban	Sep, 2018 – Aug, 2019	2.01 ± 1.54	[24]
Qingdao, China	Suburban	Oct, 2018 – Jul, 2018 and Jul, 2019 – Aug, 2019	0.75 ± 0.79 - 0.81 ± 0.87	[26]
Beijing, China	Suburban	Oct, 2018, Jan, 2019, Apr, 2019, and Jul, 2019	2.44 - 5.34	[27]
Xiamen, China	Suburban	Jan, 2018 – Dec, 2018	0.55	[28]
Beijing, China	Background	Aug, 2015 – Feb, 2019	0.94 (0.33 - 2.41)	[29]
South-Central, China	Background	Sep, 2018 – Nov, 2018	0.52	[4]
Tibetan Plateau, China	Remote	May, 2019 – Jul, 2019	0.27 ± 0.08	[30]
Seoul, South Korea	Urban	Jan, 2011 – Dec, 2011	0.61	[31]
Jungfraujoch and Zugspitze, Swiss	Background	Jan, 2008 – Dec, 2008	< 0.6	[32]
California, USA	Urban	1994 – 1995	0.47 ± 0.44	[33]

245

Supplementary Table 2. Descriptive statistics of the SHAP proportion for original time series and long-term trend of PAN. The long-term component was separated from original time series by KZA filter.

Factor	Mean SHAP proportion (original series)	Mean SHAP proportion (long-term)
NO_2	16.4% ± 10.9%	31.5% ± 16.1%
O_3	31.0% ± 16.5%	12.3% ± 7.5%
CO	5.9% ± 5.0%	11.2% ± 5.7%
T	10.1% ± 6.6%	10.9% ± 6.2%
UV	10.8% ± 6.7%	10.8% ± 6.6%
SP	7.4% ± 5.9%	6.8% ± 4.9%
RH	4.6% ± 4.8%	6.5% ± 3.6%
V10	5.2% ± 3.8%	2.9% ± 1.7%
TP	2.2% ± 1.9%	2.8% ± 1.9%
U10	3.1% ± 3.1%	2.4% ± 2.2%
NO	2.5% ± 2.3%	2.0% ± 1.2%

Supplementary Table 3. Descriptive statistics of the measured VOCs during each summer of 2018 – 2022. Concentrations are in ppbv unless stated.

Species	2018		2019		2020		2021		2022	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ethylene	1.13	1.17	1.65	0.56	0.57	0.43	0.66	0.46	0.61	0.29
Acetylene	1.01	1.00	0.99	0.46	0.50	0.53	0.49	0.37	0.47	0.23
Ethane	1.84	1.51	0.87	0.43	0.86	0.49	0.92	0.47	0.86	0.35
Propylene	0.22	0.28	2.17	1.15	0.12	0.09	0.15	0.08	0.15	0.09
Propane	2.07	2.60	0.18	0.14	0.94	0.85	0.98	0.81	1.39	1.04

Isobutane	0.95	1.16	0.02	0.01	0.42	0.37	0.50	0.37	0.62	0.56
1-Butene	0.04	0.04	0.00	0.00	0.04	0.04	0.05	0.02	0.04	0.01
Butane	1.08	1.10	0.91	0.50	0.61	0.53	0.64	0.50	0.81	0.59
Cis-2-Butene	0.05	0.06	0.08	0.04	0.07	0.07	0.03	0.02	0.02	0.02
Trans-2-Butene	0.03	0.06	0.68	0.25	0.03	0.09	0.03	0.03	0.02	0.01
Isopentane	0.45	0.79	0.01	0.00	0.49	0.41	0.74	0.67	0.63	0.33
1-Pentene	0.02	0.02	0.03	0.02	0.51	0.62	0.02	0.01	0.02	0.01
Pentane	0.12	0.26	0.08	0.04	0.23	0.25	0.23	0.27	0.22	0.16
2-Pentene	0.01	0.02	0.06	0.04	0.01	0.02	0.00	0.01	0.00	0.00
Isoprene	0.24	0.32	0.01	0.01	0.98	0.81	0.47	0.30	0.67	0.45
Cis-2-pentene	0.01	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.00	0.00
2,2-Dimethylbutane	0.10	0.35	0.69	0.41	0.01	0.01	0.04	0.02	0.02	0.01
2,3-Dimethylbutane	0.07	0.07	0.03	0.02	0.02	0.02	0.02	0.02	0.03	0.01
2-Methylpentane	0.08	0.08	0.15	0.16	0.03	0.03	0.13	0.09	0.13	0.06
3-Methylpentane	0.12	0.08	0.83	0.43	0.10	0.08	0.08	0.06	0.10	0.05
1-Hexene	0.01	0.04	0.52	0.39	0.11	0.11	0.02	0.01	0.02	0.01
N-Hexane	0.08	0.21	0.03	0.02	0.03	0.03	0.22	0.27	0.37	0.36
Benzene	0.14	0.07	0.06	0.01	0.36	0.56	0.15	0.08	0.18	0.09
Cyclohexane	0.10	0.04	0.00	0.00	0.09	0.19	0.05	0.03	0.06	0.03
2-Methylhexane	0.03	0.02	0.02	0.01	0.05	0.04	0.04	0.02	0.05	0.02
N-Heptane	0.03	0.04	1.08	0.46	0.06	0.05	0.07	0.03	0.08	0.03
Toluene	0.54	0.51	0.18	0.10	0.05	0.04	0.58	0.50	0.64	0.42
N-Octane	0.02	0.01	0.22	0.12	0.03	0.02	0.04	0.02	0.03	0.02
Ethylbenzene	0.15	0.17	0.04	0.02	0.03	0.02	0.19	0.13	0.36	0.19
Styrene	0.02	0.02	0.09	0.04	0.03	0.03	0.17	0.13	0.24	0.15
O-Xylene	0.14	0.19	0.01	0.00	0.02	0.02	0.18	0.13	0.23	0.12

Cumene	0.02	0.01	0.21	0.36	0.02	0.02	0.03	0.01	0.04	0.01
Propylbenzene	0.02	0.01	0.07	0.03	0.62	0.60	0.03	0.01	0.04	0.01
O-Ethyl toluene	0.02	0.01	0.07	0.03	0.02	0.01	0.03	0.01	0.03	0.01
M-Ethyl toluene	0.02	0.01	0.00	0.00	0.03	0.02	0.03	0.01	0.02	0.01
1,3,5-Trimethylbenzene	0.02	0.01	0.12	0.05	0.47	0.55	0.04	0.01	0.03	0.01
P-Ethyl toluene	0.01	0.01	0.03	0.01	0.19	0.19	0.05	0.01	0.04	0.01
N-Decane	0.03	0.01	0.10	0.05	0.00	0.00	0.06	0.02	0.05	0.02
1,2,4-Trimethylbenzene	0.03	0.02	0.14	0.05	0.04	0.02	0.04	0.01	0.02	0.01
1,2,3-Trimethylbenzene	0.02	0.01	0.65	0.55	0.22	0.27	0.02	0.01	0.01	0.00
N-Undecane	0.05	0.03	0.08	0.04	0.19	0.19	0.10	0.04	0.06	0.02
Acrolein	0.09	0.05	0.14	0.09	0.03	0.02	0.14	0.05	0.12	0.04
2-Butanone	1.18	2.10	0.00	0.00	0.05	0.02	0.13	0.12	0.32	0.15
1,3-Butadiene	0.02	0.05	0.17	0.09	0.03	0.02	0.02	0.01	0.03	0.01
1,1-dichloroethylene	0.02	0.01	0.04	0.04	0.05	0.03	0.01	0.00	0.01	0.01
Dichloromethane	0.82	0.43	0.04	0.12	0.05	0.03	0.75	0.68	1.37	1.27
Methyl tert-butyl ether	0.11	0.08	0.02	0.03	0.00	0.00	0.03	0.06	0.04	0.05
1,2-Dichloroethane	0.74	0.75	0.01	0.00	0.00	0.00	0.32	0.26	0.26	0.15
Trichloroethylene	0.04	0.04	0.02	0.02	0.00	0.00	0.03	0.04	0.02	0.01
CL12PROP	0.11	0.09	0.01	0.00	0.00	0.00	0.10	0.07	0.11	0.06

Supplementary Table 4. Incorporating reaction mechanism in the F0AM model.

Reaction	Rate constant and emission factor	Reference

$NO_x + emission \rightarrow HONO$	0.79%	34
$pNO_3^- + hv \rightarrow HONO$	$j_{pNO_3^-} = EF \times j_{HNO_3,g}$	35
$2NO_2 + ground + H_2O \rightarrow HONO + HNO_3$	$\gamma_{NO_2,ground} = 1 \times 10^{-6}$	36, 37
$2NO_2 + ground + H_2O \xrightarrow{hv} HONO + HNO_3$	$\gamma_{NO_2,ground+hv} = 2 \times 10^{-5}$	36, 37
$2NO_2 + aerosol + H_2O \rightarrow HONO + HNO_3$	$\gamma_{NO_2,aerosol} = 1 \times 10^{-6}$	36, 37
$2NO_2 + aerosol + H_2O \xrightarrow{hv} HONO + HNO_3$	$\gamma_{NO_2,aerosol+hv} = 2 \times 10^{-5}$	36, 37
$NO + OH \rightarrow HONO$	7.2×10^{-12}	38

$j_{HNO_3,g}$ is the photolysis frequency of HNO_3 in its gaseous form the model simulated and was adjusted by the measured by the observed and simulated photolysis rate of NO_2. The enhancement factor (EF) is set as 30 in this work.

Supplementary Table 5. Major PA production and loss reactions extracted from MCM v3.3.1 in F0AM model.

PA formation pathways	Reaction
PAN	$PAN = CH_3CO_3 + NO_2$
CH_3CHO	$NO_3 + CH_3CHO = HNO_3 + CH_3CO_3$
	$OH + CH_3CHO = CH_3CO_3$
MGLY	$MGLYOX = CH_3CO_3 + CO + HO_2$
	$NO_3 + MGLYOX = CH_3CO_3 + CO + HNO_3$
	$OH + MGLYOX = CH_3CO_3 + CO$
Radicals	$HMVKBO = CH_3CO_3 + HCHO$
	$CH_3COCH_2O = CH_3CO_3 + HCHO$
	$MEKBO = CH_3CHO + CH_3CO_3$
	$MCOCOMOXO = CH_3CO_3 + HCHO$
	$C726CO5O = C5DICARB + CH_3CO_3$
	$C23O3MO = CH_3CHO + CH_3CO_3$

	DMKOHO = $CO_2H_3CHO + CH_3CO_3$
	C4MCO2O = MGLYOX + CH_3CO_3
	MC6CO2O = C5DICARB + CH_3CO_3
	C7BDCO = C5CO243OH + CH_3CO_3
	C7CO2M5O = C4DBDIKET + CH_3CO_3
	$CH_3C_2H_2O_2 = CH_3CO_3$ +HCHO
Other OVOCs	$CH_3COCH_3 = CH_3CO_3 + CH_3O_2$
	MVK = CH_3CO_3 + HCHO + CO + HO_2
	MEK = CH_3CO_3 + HCHO + CO + HO_2
	BIACET = $CH_3CO_3 + CH_3CO_3$
	BIACETOH = $CH_3CO_3 + HOCH_2CO_3$
	ACETOL = CH_3CO_3 + HCHO + HO_2
	$CH_3CO_3H + OH = CH_3CO_3$
	HYPERACET = CH_3CO_3 + HCHO + OH
	CO23C5 = $CH_3CO_3 + C_2H_5CO_3$
	$CH_3COCO_3H = CH_3CO_3$ + OH
	MXYEPOXMUC = EPXMC4DIAL + CH_3CO_3 + HO_2 + CO
	MXYEPOXMUC = EPXMC4DIAL + CH_3CO_3 + CO
	MXYMUCCO = CH_3CO_3 + EPXMC4DIAL + CO + HO_2
	C4DBDIKET = MGLYOX + CH_3CO_3 + HO_2 + CO
	DMKCOOH = CO23C3CHO + CH_3CO_3 + HO_2
	DMKCOOH = CO2H3CO3 + CH_3CO_3
	TM124OXMUC = EPXM2C4DAL + CH_3CO_3 + HO_2 + CO
	O_3 + MPAN = HCHO + CH_3CO_3 +NO_3

	$MVKNO_3 = CH_3CO_3 + HOCH_3CHO + NO_2$
PA loss pathways	Reaction
$CH_3CO_3 + NO$	$CH_3CO_3 + NO = NO_2 + CH_3O_2$
$CH_3CO_3 + NO_2$	$CH_3CO_3 + NO_2 = PAN$
$CH_3CO_3 + NO_3$	$CH_3CO_3 + NO_3 = NO_2 + CH_3O_2$
$CH_3CO_3 + HO_2$	$CH_3CO_3 + HO_2 = CH_3CO_2H + O_3$
	$CH_3CO_3 + HO_2 = CH_3CO_3H$
	$CH_3CO_3 + HO_2 = CH_3O_2 + OH$

REFERENCES:

1. Mellouki A, Wallington TJ, Chen J. Atmospheric Chemistry of Oxygenated Volatile Organic Compounds: Impacts on Air Quality and Climate. *Chemical Reviews* **115**, 3984-4014 (2015).

2. von Ahsen S, Willner H, Francisco JS. Thermal decomposition of peroxy acetyl nitrate CH3C(O)OONO2. *Journal of Chemical Physics* **121**, 2048-2057 (2004).

3. Fischer EV, et al. Atmospheric peroxyacetyl nitrate (PAN): a global budget and source attribution. *Atmospheric Chemistry and Physics* **14**, 2679-2698 (2014).

4. Gong DC, et al. Characteristics of peroxyacetyl nitrate (PAN) in the high-elevation background atmosphere of South-Central China: Implications for regional photochemical pollution. *Atmospheric Environment* **254**, (2021).

5. Liu L, et al. Understanding unusually high levels of peroxyacetyl nitrate (PAN) in winter in Urban Jinan, China. *Journal of Environmental Sciences* **71**, 249-260 (2018).

6. Tuazon EC, Carter WPL, Atkinson R. Thermal-decomposition of peroxyacetyl nitrate and reactions of acetyl peroxy-radicals with NO and NO2 over the temperature-range 283-313-K. *Journal of Physical Chemistry* **95**, 2434-2437 (1991).

7. Atkinson R, et al. Evaluated kinetic, photochemical and heterogeneous data for atmospheric chemistry .5. IUPAC Subcommittee on Gas Kinetic Data Evaluation for Atmospheric Chemistry. J*ournal of Physical and Chemical Reference Data* **26**, 521-1011 (1997).

8. Rao ST, Zurbenko IG. Detecting and tracking changes in ozone air-quality. *Journal of the Air & Waste Management Association* **44**, 1089-1092 (1994).

9. Zurbenko I, Porter PS, Rao ST, Ku JY, Gui R, Eskridge RE. Detecting discontinuities in time series of upper-air data: Development and demonstration of an adaptive filter technique. *Journal of Climate* 9**,** 3548-3560 (1996).

10. Rao ST, Zurbenko IG, Neagu R, Porter PS, Ku JY, Henry RF. Space and time scales in ambient ozone data. *Bulletin of the American Meteorological Society* **78**, 2153-2166 (1997).

11. Breiman L. Random forests. *Machine Learning* **45**, 5-32 (2001).

12. Kotsiantis SB. Decision trees: a recent overview. *Artificial Intelligence Review* **39**, 261-283 (2013).
13. Pedregosa F, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830 (2011).
14. Chen TQ, Guestrin C, Assoc Comp M. XGBoost: A Scalable Tree Boosting System. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)) (2016).
15. Xia SY, et al. Seasonal variation characteristics of atmospheric peroxyacetyl nitrate (PAN) and its source apportionment in a megacity in southern China. *Science of the Total Environment* **892**, (2023).
16. Qian X, Shen HQ, Chen ZM. Characterizing summer and winter carbonyl compounds in Beijing atmosphere. Atmospheric Environment 214, (2019).
17. Shapley LS. A value for n-person games. Contributions to the Theory of Games, (1953).
18. Lundberg SM, et al. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* **2**, 56-67 (2020).
19. Liu TT, et al. Atmospheric oxidation capacity and ozone pollution mechanism in a coastal city of southeastern China: analysis of a typical photochemical episode by an observation-based model. *Atmospheric Chemistry and Physics* **22**, 2173-2190 (2022).
20. Theil H. A Rank-Invariant Method of Linear and Polynomial Regression Analysis. Springer Netherlands (1992).
21. Sen, Kumar P. Estimates of the Regression Coefficient Based on Kendall's Tau. *Publications of the American Statistical Association* **63**, 1379-1389 (1968).
22. Zhang H, et al. A comprehensive observation on the pollution characteristics of peroxyacetyl nitrate (PAN) in Beijing, China. *The Science of the total environment* **905**, 166852 (2023).
23. Yao Q, et al. Transport Characteristics of PAN and O3 in the Lower Atmosphere of the Boundary Layer in Tianjin in Summer. *Huanjing Kexue* **40**, 67-75 (2019).
24. Sun M, et al. Seasonal discrepancies in peroxyacetyl nitrate (PAN) and its correlation with ozone and PM2.5: Effects of regional transport from circumjacent industrial cities. *Science of the Total Environment* **785**, (2021).
25. Zhang G, et al. Simultaneous observation of atmospheric peroxyacetyl nitrate and ozone in the megacity of Shanghai, China: Regional transport and thermal decomposition. *Environmental Pollution* **274**, (2021).
26. Liu YH, et al. Formation of peroxyacetyl nitrate (PAN) and its impact on ozone production in the coastal atmosphere of Qingdao, North China. *Science of the Total Environment* **778**, (2021).
27. Wei W, Zang JX, Wang XQ, Cheng SY. Peroxyacetyl nitrate (PAN) in the border of Beijing, Tianjin and Hebei of China: Concentration, source apportionment and photochemical pollution assessment. *Atmospheric Research* **246**, (2020).
28. Hu BY, et al. Characteristics of peroxyacetyl nitrate (PAN) in a coastal city of southeastern China: Photochemical mechanism and pollution process. *Science of the*

Total Environment **719**, (2020).
29. Qiu YL, et al. A study of peroxyacetyl nitrate at a rural site in Beijing based on continuous observations from 2015 to 2019 and the WRF-Chem model. *Frontiers of Environmental Science & Engineering* **14**, (2020).
30. Xu W, et al. O3 and PAN in southern Tibetan Plateau determined by distinct physical and chemical processes. *Atmospheric Chemistry and Physics* **23**, 7635-7652 (2023).
31. Lee JB, et al. Peroxyacetyl nitrate (PAN) in the urban atmosphere. *Chemosphere* **93**, 1796-1803 (2013).
32. Deolal SP, et al. Analysis of elevated springtime levels of Peroxyacetyl nitrate (PAN) at the high Alpine research sites Jungfraujoch and Zugspitze. *Atmospheric Chemistry and Physics* **14**, 12553-12571 (2014).
33. Roberts JM, et al. Measurements of PAN, PPN, and MPAN made during the 1994 and 1995 Nashville Intensives of the Southern Oxidant Study: Implications for regional ozone production from biogenic hydrocarbons. *Journal of Geophysical Research-Atmospheres* **103**, 22473-22490 (1998).
34. Liu YL, et al. Semi-quantitative understanding of source contribution to nitrous acid (HONO) based on 1 year of continuous observation at the SORPES station in eastern China. *Atmospheric Chemistry and Physics* **19**, 13289-13308 (2019).
35. Zare A, Romer PS, Nguyen T, Keutsch FN, Skog K, Cohen RC. A comprehensive organic nitrate chemistry: insights into the lifetime of atmospheric organic nitrates. *Atmospheric Chemistry and Physics* **18**, 15419-15436 (2018).
36. Gonçalves M, Dabdub D, Chang WL, Jorba O, Baldasano JM. Impact of HONO sources on the performance of mesoscale air quality models. *Atmospheric Environment* **54**, 168-176 (2012).
37. Hu BY, et al. The effect of nitrous acid (HONO) on ozone formation during pollution episodes in southeastern China: Results from model improvement and mechanism insights. *Science of the Total Environment* **891**, (2023).
38. Pagsberg P, Bjergbakke E, Ratajczak E, Sillesen A. Kinetics of the gas phase reaction OH+NO(+M)->HONO(+M) and the determination of the UV absorption cross sections of HONO. *Chemical Physics Letters* **272**, 383-390 (1997).