Atmospheric Chemistry and Physics

Ref: egusphere-2025-4624

Driving Factors of Oxalic Acid and Enhanced Role of Gas-Phase Oxidation under Cleaner Conditions: Insights from 2007–2018 Field Observations in the Pearl River Delta

Response Letter to Reviewer's Comments

Dear reviewer:

We sincerely thank you for your time and valuable comments. We have added more information about machine learning to enhance the robustness of our results, and carefully revised the manuscript to improve its clarity and enhance the readers' understanding. Our point-by-point responses are marked in blue and the corresponding changes to the original text are shown below each response. We hope that these revisions adequately address the comments and concerns.

Comment 1: *Attribution to gas- vs aqueous-phase pathways is mechanistically simplified; aqueous production depends on pH, transition metals, oxidant availability, and organic composition, the author may consider adding more feature variables in the machine learning model.*

Response: Thanks for your valuable suggestions. We agree that there are other factors influencing the formation of aqueous-phase products other than pH and ALWC. This is same to gas-phase products. However, due to unavailability of related data in this study, such as transition metals and oxidant concentrations in aqueous phase, we can not quantify their contributions on variations in C_2.

Here, we add three feature variables in the machine learning model, including sulfate, photolysis frequencies of O_3 ($J(O^1D)$) and NO_2 ($J(NO_2)$), to make our results better reflect the impacts of gaseous and aqueous pathways. Sulfate is suggested as an important product from secondary aqueous-phase chemistry (Liu et al., 2021) and can be used as an indicator for aqueous reactions. $J(O^1D)$ represents the photolytic rate of ozone producing excited oxygen atoms $O(^1D)$, which subsequently react with water vapor to generate hydroxyl radicals (OH), the dominant oxidant driving daytime gas-phase oxidation processes. $J(NO_2)$ describes the photolysis rate of nitrogen dioxide, leading to the formation of NO and ground-state oxygen atoms O (^3P), which further participate in ozone formation. Therefore, they are key parameters characterizing the intensity of atmospheric photochemical activity (Ehhalt and Rohrer, 2000).

After the inclusion of these three variables, the results of the machine learning model remain highly consistent with those of the previous version, and the overall conclusions are unchanged. These results confirm the appropriateness of the selected variables and the robustness of the model outcomes. Specifically, from IT0 to IT4, the IF values associated with gas-phase oxidation processes increased from 37% to 55%, whereas those related to aqueous-phase oxidation processes decreased from 42% to 30%, indicating an increasing importance of gas-phase oxidation under cleaner

atmospheric conditions. In addition, the general impacts of changes in gas-phase oxidation (45%) and aqueous-phase oxidation (34%) are substantially higher than that of AVOC (14%) and BVOC emissions (7%). Although the ranking of feature importance changed, the indicators for gas-phase and aqueous-phase oxidation (such as O_x, $J(O^1D)$, sulfate, ALWC) still exhibit relatively high importance among all variables. As ALWC and pH were estimated by a thermodynamic equilibrium model, ISORROPIA II (Nenes et al., 1998), in which sulfate plays a crucial role and partly reflects variations in both pH and ALWC, sulfate ranks second in feature importance rather than ALWC. The comparison between new and old version can be seen below. To maintain consistency in the number of variables, we also added sulfate, $J(O^1D)$, and $J(NO_2)$ into correlation analysis. In addition, the appropriateness of the selected variables also needs to be clarified. We have updated the number and corresponding statement in manuscript, and added limitation of this study at the end of manuscript.

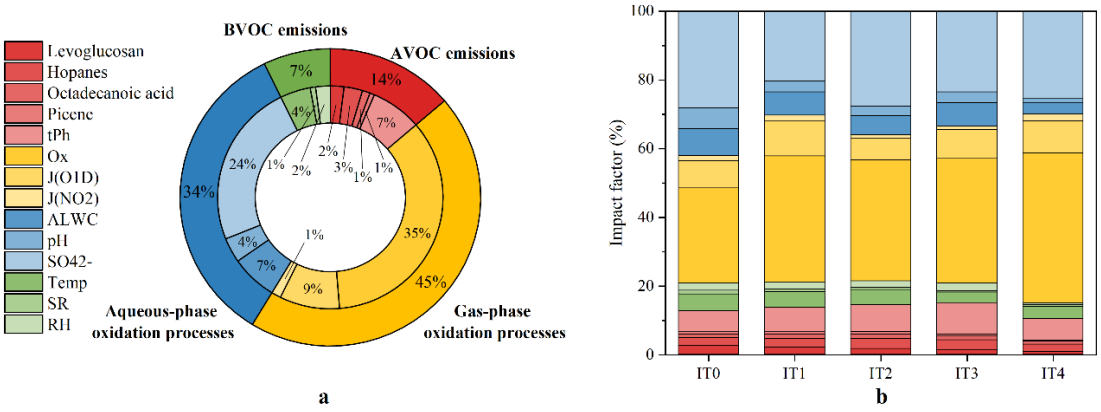

Figure 6 (new). (a) Impact of changes of each variable on C_2 variation during the whole study period. (b) Impact factor of individual variable under different pollution conditions.

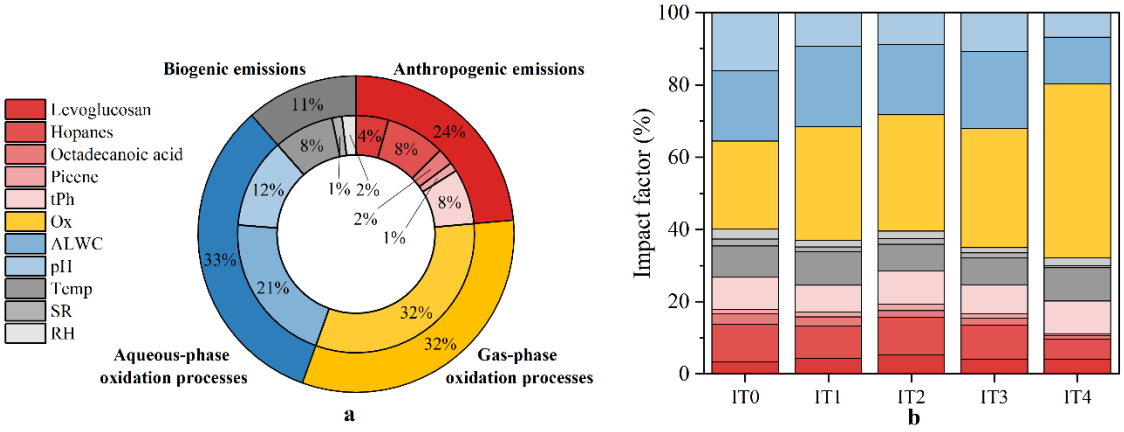

Figure 6 (old).

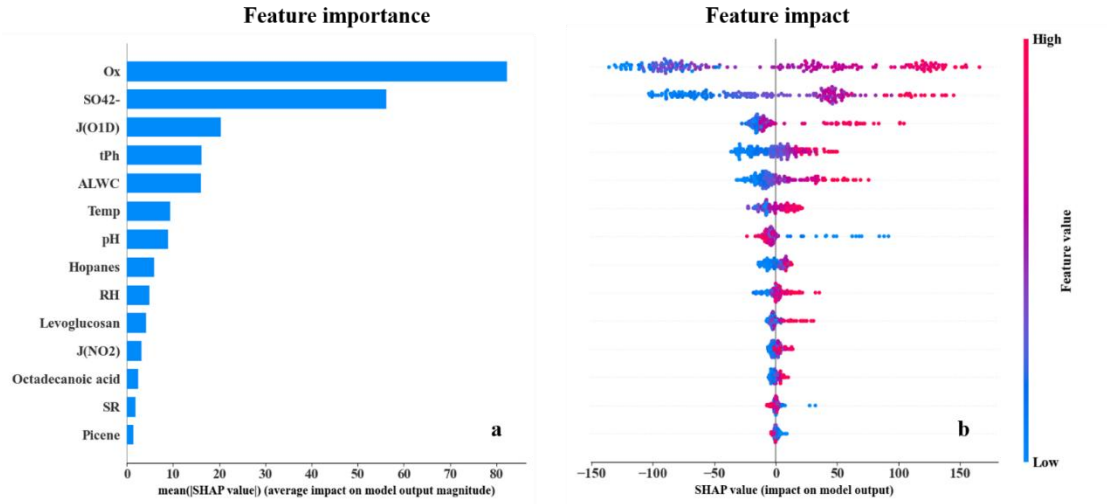

Figure 5 (new). (a) Bar plot of the mean |SHAP| values representing the overall importance of each feature in predicting C_2 concentrations. (b) Beeswarm plot of individual SHAP values for each feature across all samples. Red (blue) represents high (low) value in each feature. Positive (negative) SHAP values indicate that the feature contributes to an increase (decrease) in the C_2 prediction.

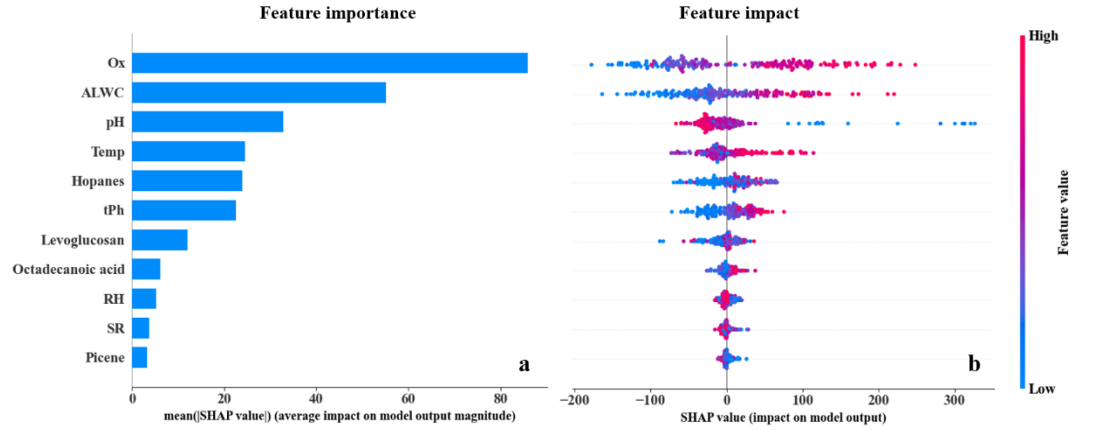

Figure 5 (old).

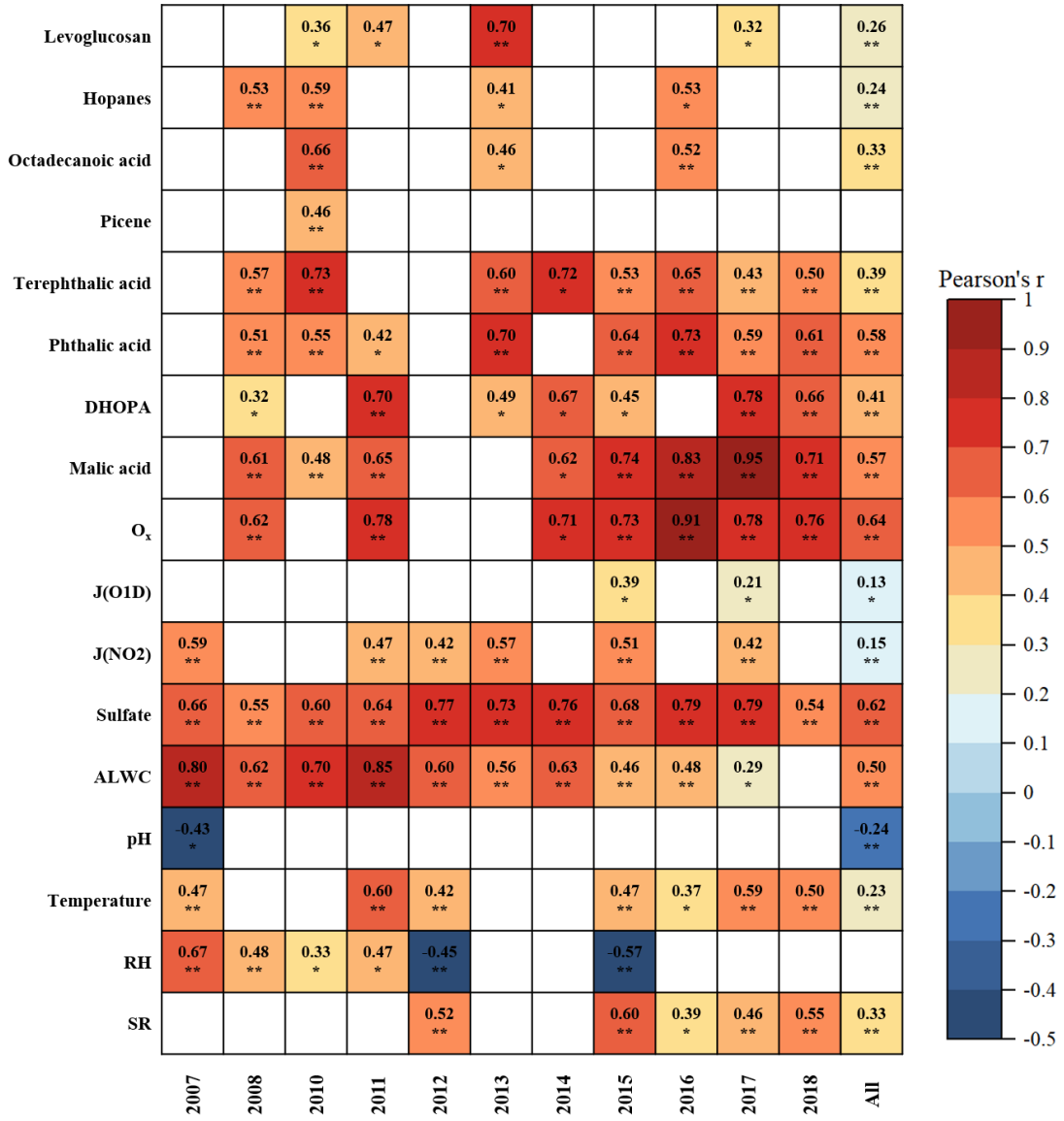

Figure 3. Correlations between C_2 and primary anthropogenic source markers, SOA markers, indicators for gas- and aqueous-phase oxidation, as well as meteorological parameters. Blank cells indicate no significant correlations. One asterisk, two asterisks denote p value < 0.05, 0.01, respectively. Due to the unavailability of O_x data in 2012 and 2013, correlation analysis was not conducted for these two years.

Extreme Gradient Boosting (XGBoost), an advanced ensemble machine learning method based on gradient boosting decision trees, is known for its high computational efficiency, robust predictive performance (Chen et al., 2016) and thus has been applied in air pollutant research recently (Hou et al., 2022; Peng et al., 2023; Liu et al., 2025). In this study, XGBoost was employed to assess the relative contributions of various factors to oxalic acid variation. The implementation and Python package of XGBoost algorithm are publicly available online (https://github.com/dmlc/xgboost). A total of 14 variables were used as input features to train the model, including levoglucosan, hopanes, octadecanoic acid, picene, terephthalic acid (tPh), O_x, photolysis frequencies of O_3 (JO^1D) and NO_2 (JNO_2), ALWC, pH, sulfate (SO_4^{2-}), temperature (Temp), solar radiation (SR), and relative humidity (RH). To avoid redundant and confounding explanations, the secondary organic molecular markers,

such as DHOPA, phthalic acid (Ph), and malic acid, were excluded in the model training. They are influenced by VOC emissions and secondary oxidation processes, which are already represented by the factors mentioned above. Our results showed that there were great agreements between the observations and simulations for C_2 and other DCA (Fig. S2), which indicated the model predictions were reliable.

ALWC not only regulates the gas-particle partitioning of semi-volatile VOCs and their reaction rates by acting as a medium (Nenes et al., 2021), but also serves as a nucleophile that participates in reactive uptake of SOA intermedium (Zhang et al., 2022b). Aerosol pH plays a crucial role in governing acid-catalyzed reactions during aqueous-phase processing (Cooke et al., 2024). In addition, sulfate is also an important secondary product formed through aqueous oxidation (Liu et al., 2021). O_x, a proxy of atmospheric oxidants, facilitates secondary photochemical oxidation of VOCs. $J(O^1D)$ and $J(NO_2)$ represent photolysis frequencies of O_3 and NO_2 (Ehhalt and Rohrer, 2000). Accordingly, ALWC, pH, and sulfate were employed as indicators for aqueous-phase oxidation , while O_x, $J(O^1D)$, and $J(NO_2)$ were used as indicators for gas-phase oxidation. In this study, C_2 showed strong correlations with ALWC ($r = 0.50$, $p < 0.01$), sulfate ($r = 0.62$, $p < 0.01$), and O_x ($r = 0.64$, $p < 0.01$) across the whole datasets, suggesting secondary oxidation processes were the dominant drivers of C_2 variability between 2007 and 2018.

The rationality for selecting the variables used to train the model need to be clarified to ensure the reliability of the results. Levoglucosan, hopanes, octadecanoic acid, picene, and tPh serve as source-specific molecular markers for biomass burning, vehicle emission, cooking, coal combustion, and waste incineration, respectively. These species are used to represent changes in AVOC emissions. As two of the most important BVOCs globally, isoprene emission is highly dependent on temperature and solar radiation, while monoterpenes emission is sensitive to temperature (Guenther et al., 1993). Their emissions rate can be estimated using equation 3-5 and equation 6, respectively:

$$E_i = I_S \cdot C_L \cdot C_T \quad (3)$$

$$C_L = \frac{\alpha c_{L1} L}{\sqrt{1+\alpha^2 L^2}} \quad (4)$$

$$C_T = \frac{\exp\frac{c_{T1}(T-T_S)}{RT_S T}}{1+\exp\frac{c_{T2}(T-T_M)}{RT_S T}} \quad (5)$$

where E_i is isoprene emission rate at a temperature T(K) and photosynthetically active radiation (PAR) flux L (μmol m^{-2} s^{-1}), I_s is isoprene emission rate at a standard temperature T_s and a standard PAR flux (1000 μmol m^{-2} s^{-1}). $\alpha = 0.0027$ and $c_{L1} = 1.066$ are empirical coefficients determined by measurements. L can be calculated as multiplying solar radiation (W m^{-2}) by photon flux efficacy (1.86 μmol J^{-1}). R is a constant -8.314 J K^{-1} mol^{-1}, and c_{T1} = 95000 J mol^{-1}, c_{T2} = 230000 J mol^{-1}, and T_M = 314 K are empirical coefficients estimated by measurements.

$$E_m = M_S \cdot \exp(\beta (T - T_S)) \quad (6)$$

where E_m is monoterpenes emission rate at temperature T (K), M_s is monoterpenes emission rate at a standard temperature T_s, β (K^{-1}) is an empirical coefficient ranging from 0.057 to 0.144 K^{-1}. In addition, inadequate moisture can have significantly decreased stomatal conductance and photosynthesis (Guenther et al., 2006). Therefore, RH is an important factor influencing BVOC emissions. $J(O^1D)$ and $J(NO_2)$ are photolysis frequencies of O_3 and NO_2, which are relevant to the generation of hydroxyl radical (an important oxidant in atmosphere) (Ehhalt and Rohrer, 2000). O_x is also commonly used as a proxy for ambient oxidizing capacity. ALWC and pH have important impacts on SOA formation in aqueous phase (Nguyen et al., 2015; Xu et al., 2016). Previous studies have shown that sulfate is a secondary species primarily produced through aqueous-phase oxidation (Yu et al., 2005; Liu et al., 2021). Thus, ALWC, pH, and sulfate are used as indicators of aqueous-phase processes. To avoid redundant and confounding explanations, the secondary organic molecular markers, such as DHOPA, Ph, and malic acid, were excluded from the model training. These species are influenced by both VOC emissions and secondary oxidation processes, which are already represented by the factors mentioned above.

The feature importance is presented in Fig. 5a. O_x, sulfate, and $J(O^1D)$, which represent secondary oxidation processes, exhibited the three highest |SHAP| values, indicating their dominant impacts on C_2 variation. Although pH and ALWC exhibited relatively high feature importance among all variables, their |SHAP| values were lower than sulfate. This is because pH and ALWC in this study was calculated by a thermodynamic equilibrium model, ISORROPIA II (Nenes et al., 1998), in which sulfate plays a crucial role and partly reflects variations in both pH and ALWC. In contrast, the feature importance of anthropogenic emission markers and meteorological parameters were relatively lower, suggesting that their influences were smaller compared to that of secondary processes. As shown in Fig. 5b, O_x, sulfate, and $J(O^1D)$ exhibited obviously positive correlations with their SHAP values, indicating that higher values of these variables contributed to increases in C_2 concentrations. However, pH showed a negative correlation with its SHAP values, suggesting that lower pH levels were associated with higher C_2 concentrations. Notably, the influence of extremely low pH on C_2 formation appeared to be more pronounced.

To further quantify the impacts of changes in all factors on C_2, IF (discussed in Section 2.4) was calculated and presented in Fig. 6. O_x accounted for the highest contribution (35%), followed by sulfate (24%) and $J(O^1D)$ (9%). All factors were classified into four groups according to their representativeness mentioned before: (1) AVOC emissions (levoglucosan, hopanes, octadecanoic acid, picene, and tPh); (2) BVOC emissions (Temp, SR, and RH); (3) gas-phase oxidation pathways (O_x, $J(O^1D)$, and $J(NO_2)$); (4) aqueous-phase oxidation pathways (ALWC, pH, and sulfate). Due to the minor fluctuations of meteorological conditions in each year, the impacts of changes in BVOC emissions on C_2 were small (7%). Although AVOC emissions showed an obvious decreasing trend over the study period, the impacts of these changes (14%) were significantly lower than that of gas-phase oxidation processes (45%) and aqueous-phase oxidation processes (34%). The results were consistent with correlation analysis, underscoring the dominant role of secondary oxidation

processes in C_2 formation.

The IF values for each variable are presented in Table S10. From IT0 to IT4, IF values for gas-phase oxidation processes increased from 37% to 55%, whereas those for aqueous-phase oxidation processes decreased from 42% to 30% (Fig. 6b). Meanwhile, IF values for AVOC (10%–15%) and BVOC emissions (5%–8%) remained at a low and stable level. These findings indicated that the gas-phase oxidation pathway became increasingly important as pollution levels decreased. A possible explanation is that under cleaner conditions, lower ALWC levels favored the partitioning of semi-volatile C_2 precursors (e.g., Gly and mGly) from the particle phase into the gas phase. In addition, less ALWC participates in heterogeneous reactions of SOA intermedium as a nucleophile. Thus, their aqueous-phase pathway was hindered, and more C_2 was formed via photochemical degradation of longer-chain DCA (Kawamura and Bikkina, 2016; Meng et al., 2023). This indicated the growing importance of gas-phase oxidation processes in the formation of C_2 and SOA under cleaner conditions. Although O_x did not exhibit a clear trend at our measurement station, Cao et al. (2024) reported a rapid increase in O_3 concentration across the PRD region over the past decade. This may promote SOA formation through enhanced gas-phase oxidation pathways. Therefore, coordinated control of VOCs and NO_x should be emphasized (Wang et al., 2021b) in the future to reduce ozone pollution and further mitigate SOA formation.

Meanwhile, we acknowledge there are several limitations in this study. First, our measurements were mainly conducted in wintertime, which may not represent summertime conditions when photochemical activity is higher. Second, there are other factors influencing the formation of aqueous-phase products other than pH and ALWC. This is same to gas-phase products. However, due to unavailability of related data in this study, such as transition metals and hydroxyl radical in aqueous phase, we were unable to quantify their contributions on variations in C_2, which may introduce uncertainties.

Comment 2: *line 84: you should spell out an abbreviation (ALWC) the first time it appears in the main text even if you already defined it in the abstract.*

Response: Thanks for reminding this. We have added statement of ALWC (aerosol liquid water content) in line 84.

During COVID-19, lower aerosol liquid water content (ALWC) and elevated O_3 shifted the dominant formation pathway of C_2 from aqueous-phase oxidation of ωC_2 and Pyr to gas-phase photochemical decomposition of longer-chain DCA (malonic (C_3) and succinic (C_4)).

Comment 3: *Figure 1: there's almost no exact content in the figure. The author may consider adding back-trajectories or removing this figure to the SI.*

Response: Thanks for your suggestion. We have moved Figure 1 to SI because we don't have

discussion about back-trajectories in this part.

Comment 4: *line 209: Malic acid is a plausible product of biogenic VOC photooxidation, but it is not a unique tracer. Given the winter, urban-influenced atmosphere, anthropogenic VOCs and combustion sources could contribute substantially.*

Response: Thank for this insightful comment. We agree that malic acid is a typical secondary product originating from the photooxidation of both biogenic and anthropogenic precursors, and thus should not be considered a unique tracer for BSOA. We should clarify that in our manuscript. The contributions from biogenic and anthropogenic VOCs on malic acid formation are different. Sato et al. (2021) conducted a chamber study to investigated mass fractions of malic acid in SOA produced from biogenic and anthropogenic sources. Based on chamber results, they estimated that malic acid produced through the oxidation of BVOCs (α-pinene and isoprene) accounted for 63%, which was higher than that formed by AVOCs (toluene and naphthalene). Given that α-pinene only accounts for 34% in monoterpenes (Sindelarova et al., 2014) and BVOC emissions are about eight times higher than that AVOC emissions globally (Glasius and Goldstein, 2016), malic acid produced from biogenic sources may dominate over that from anthropogenic sources. In addition, malic acid was found to be strongly correlated (N = 49, R^2 = 0.95) with monoterpene tracers (3-Hydroxyglutaric acid, 3-Hydroxy-4,4-dimethylglutaric acid, 3-Methyl-1,2,3-butanetricarboxylic acid, 3-Isopropylpentanedioic acid, 3-Acetyl pentanedioic acid) in one-year field measurements (Cheng et al., 2021). Another research also observed such strong correlation between malic acid and monoterpene tracers in both summer (R^2 = 0.92) and winter (R^2 = 0.87) (Hu and Yu, 2013).

Due to low level of human activities, traffic and industrial emissions in the surrounding area, this site experiences limited anthropogenic influence. Furthermore, there is no residential heating in the PRD region, which is a major source of AVOCs during the wintertime. Consequently, although anthropogenic emissions may increase in winter, the rise is less pronounced than in urban areas. The PRD region is situated in a subtropical zone, characterized by mild winter temperatures averaging around 20 °C (Table S5). This climatic condition sustains considerable biogenic emissions even in winter. Therefore, these evidences indicated that malic acid in our sampling site could be formed mainly by photodegradation of BVOCs, especially monoterpenes.

Furthermore, as shown in Table S7, the correlation between oxalic acid and malic acid strengthens with pollution levels decreasing, while the correlation between oxalic acid and ASOA tracers weakens. This divergent pattern indicates that anthropogenic precursors were not the dominant source of malic acid. In general, biogenic sources had more contribution to malic acid formation than anthropogenic in this study.

Because we don't have unique BSOA tracers in this study, we used malic acid concentrations to reflect BSOA variations. When we quantified impact of BVOCs on oxalic acid by machine learning, we used meteorological parameters (e.g., temperature, solar radiation, and relative humidity), which

can determine BVOC emissions, as proxies for BVOC emissions instead of malic acid. This will avoid potential confusion of AVOC and BVOC emissions. We acknowledge that the original phrasing in the manuscript was imprecise and have revised the relevant sentences accordingly to prevent any misunderstanding.

Phthalic acid has been identified as a SOA tracer derived from naphthalene (Kleindienst et al., 2012), while DHOPA is a tracer for SOA formed from aromatic hydrocarbons (Ding et al., 2017). Given the substantial anthropogenic sources of naphthalene and aromatic hydrocarbons, phthalic acid and DHOPA can be used as anthropogenic SOA (ASOA) markers. Malic acid is a typical secondary product formed through photooxidation of both anthropogenic and biogenic VOCs (AVOCs and BVOCs). However, a recent study estimated that malic acid produced through the oxidation of BVOCs (α-pinene and isoprene) was higher than that formed by AVOCs (toluene and naphthalene) (Sato et al., 2021). In addition, malic acid was also found to be strongly correlated with monoterpene tracers (R^2 = 0.87-0.95) in field measurements (Hu and Yu, 2013; Cheng et al., 2021). Given high BVOC emissions (Wang et al., 2021) and relatively high temperature (~20 ℃, Table S5) in the PRD region, malic acid was mainly produced from biogenic precursors in this study, especially monoterpenes. Thus, we used malic acid to reflect the variations of SOA (BSOA). Although phthalic acid, DHOPA, and malic acid decreased from 51.9 ± 14.9, 1.85 ± 1.35, and 24.2 ± 19.4 ng m^{-3} to 16.7 ± 5.7, 1.05 ± 0.88, 5.9 ± 4.9 ng m^{-3}, respectively, their declining trends were not statistically significant ($p > 0.05$). This indicated that the influence of reductions in emissions of anthropogenic organic precursors on SOA was limited.

Notably, the correlation between C_2 and malic acid strengthened progressively with the reductions in anthropogenic emissions. This trend became more apparent when the data were categorized by pollution levels, with the correlation coefficients increased from 0.33 (IT0) to 0.72 (IT4) (Table S7). Meanwhile, the correlations between C_2 and ASOA markers weakened. As discussed previously, malic acid can be produced by photooxidation of both anthropogenic and biogenic precursors. However, this divergent pattern of correlations supported that anthropogenic precursors were not the dominant source of malic acid in this study. Thus, these results suggested that the relative contributions of biogenic sources to SOA become more important under cleaner conditions.

Comment 5: *line 226: The authors normalize oxalic acid and related species by $PM_{2.5}$ to reduce dilution effects. I would rather recommend using primary and inertia tracers such as ΔCO as a more appropriate normalizer for removing dilution.*

Response: Thank you for the valuable suggestion. Indeed, using CO as a normalization tracer for dicarboxylic acids and oxalic acid is more reasonable for evaluating the influence of atmospheric dilution. Accordingly, we have added a figure in the Supplement showing that the temporal trends of dicarboxylic acids and oxalic acid normalized by CO are consistent with their original trends.

This result indicates that meteorology-driven atmospheric dilution had a limited influence on their observed variations. The related discussion has been incorporated into the same paragraph in the revised manuscript. Here, we showed an increase in the the ratio of $C_2/PM_{2.5}$ to reflect the relative importance of SOA is increasing as pollution levels decrease.

C_2 was the most abundant compound among aliphatic DCA, accounting for 80%–91%, followed by C_4 (4%–13%), and C_9 (1%–4%). Therefore, the overall trend of aliphatic DCA was primarily driven by C_2 (Fig. 2), and subsequent discussions will focus on C_2. Its concentration declined from 692 ± 243 (2007) to 274 ± 114 (2018), but did not exhibit a clear trend ($p > 0.05$). Carbon monoxide (CO) can be used as a normalization tracer to assess the influence of atmospheric dilution. As shown in Fig. S3, the temporal trends of DCA and C_2 normalized by CO are consistent with their original trends, indicating that atmospheric dilution had a limited influence on their observed variations. To further explore the changes of SOA formation under different pollution conditions, our samples were divided into five categories according to interim targets recommended by the Worle Health Organization (WHO) in 2021 (World Health Organization, 2021): IT0 ($PM_{2.5}$ > 75 μg m^{-3}), IT1 (75 μg m^{-3} > $PM_{2.5}$ > 50 μg m^{-3}), IT2 (50 μg m^{-3} > $PM_{2.5}$ > 37.5 μg m^{-3}), IT3 (37.5 μg m^{-3} > $PM_{2.5}$ > 25 μg m^{-3}), and IT4 ($PM_{2.5}$ < 25 μg m^{-3}). We found that the molecular markers and C_2 decreased significantly ($p < 0.01$) from IT0 to IT4 (Table S6). However, the ratio of C_2 to $PM_{2.5}$ ($C_2/PM_{2.5}$) increased from 6.8 × 10^{-3} to 10.3× 10^{-3} ($p < 0.01$, Fig. S4), suggesting that the relative importance of SOA increased as pollution levels decreased.

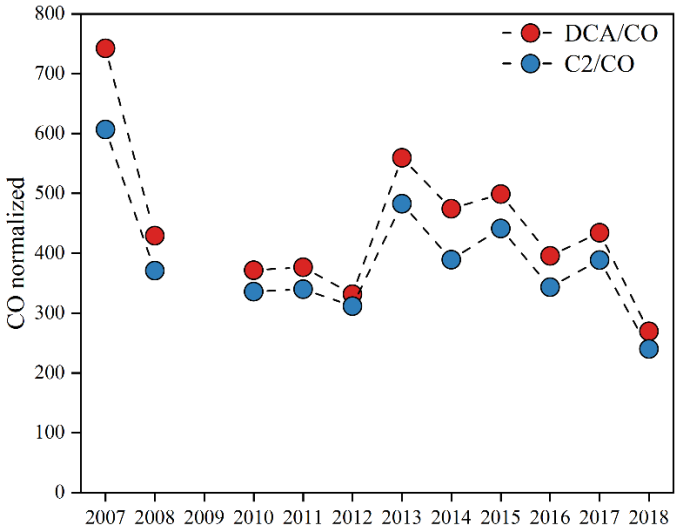

Figure S3. The concentrations of DCA and C2 normalized by carbon monoxide (CO, ppm). Due to the lack of in situ CO measurements at the sampling site, monthly CO data were obtained from the Copernicus Atmosphere Monitoring Service (CAMS) global reanalysis product (EAC4), provided by the European Centre for Medium-Range Weather Forecasts (ECMWF, https://cds.climate.copernicus.eu/datasets). The dataset has a horizontal resolution of approximately 0.75° × 0.75°.

Comment 6: *Figure 3: add oxalic acid data in this figure.*

Response: Thanks for suggestion. We have added oxalic acid data in this figure.

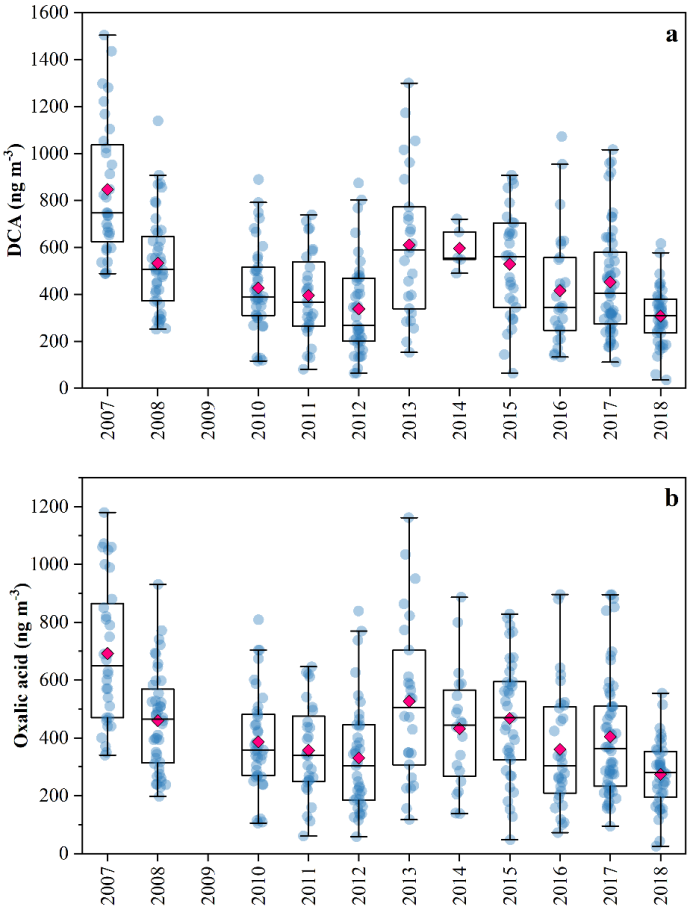

Figure 3. (a) Annual variations in aliphatic DCA. (b) Annual variations in oxalic acid. The concentrations of DCA decreased from 864 ± 283 ng m^{-3} (2007) to 307 ± 122 ng m^{-3} (2018), and the concentrations of oxalic acid decreased from 692 ± 243 ng m^{-3} (2007) to 274 ± 114 ng m^{-3} (2018), but the trends were not statistically significant ($p > 0.05$). Due to the absence of oxalic acid measurements in 2009, the concentrations of aliphatic DCA for that year are not presented.

Comment 7: *line 255-258: I do not find enough evidence supporting the two sentences claiming the limited contribution of anthropogenic VOCs and meteorology.*

Response: Thanks for this valuable comment. The AVOCs and meteorology play important roles in oxalic acid formation. However, what we want to discuss here is that the influences of **changes** in AVOCs and meteorology on C_2 variations. We apology for the imprecise statement, which confuses the concept of "absolute contributions of AVOCs and meteorology" with "impacts of **changes** in AVOCs and meteorology". For example, although contributions from AVOCs to C_2 is important, their impacts on C_2 variations could be limited when AVOCs remain at a stable level. We have revised the relevant sentences to prevent any misunderstanding.

As shown in Figure 3 (see above), oxalic acid exhibited weak correlations with primary anthropogenic source markers across the entire dataset. Although anthropogenic sources

experienced substantial reductions during the campaign period (discussed in Section 3.1), oxalic acid did not show a corresponding significant decreasing trend. In addition, a recent study observed an unexpected increase in oxalic acid when anthropogenic emissions were substantially reduced during the COVID-19 pandemic (Meng et al., 2023). These evidences implied that the reductions in anthropogenic emissions were not the driving factor for oxalic acid variations.

Because our field measurements were conducted in the same season each year (from October to December), the inter-annual differences in meteorological conditions were small. This resulted in the consistently weak correlations observed between oxalic acid and key meteorological parameters such as temperature, solar radiation, and relative humidity (Figure 3). Therefore, we conclude that the changes in meteorology were too small to be driving factor for oxalic acid formation.

The results of correlation analysis are presented in Fig. 3. The correlation coefficients between C_2 and SOA tracers (phthalic acid, DHOPA, and malic acid) were relatively higher ($r = 0.58$, 0.41, and 0.51, respectively; $p < 0.01$), further supporting that C_2 was primarily formed via secondary oxidation processes. Meng et al. (2023) reported an unexpected enhancement of C_2 during the COVID-19 pandemic, when anthropogenic emissions were substantially reduced. This reflected limited influence of reductions in anthropogenic organic precursors on formation of C_2. Similarly, we found that anthropogenic emissions experienced substantial reductions during our campaign period (discussed in Section 3.1), while C_2 did not show a corresponding significant decreasing trend. Although strong correlations between C_2 and primary anthropogenic source markers were observed in certain individual years, the correlations remained weak across the entire dataset. These findings implied that the changes in anthropogenic emissions were not the driving factor for oxalic acid formation in this study. Because our field measurements were conducted in the same season each year (from October to December), the inter-annual differences in meteorological conditions were negligible. This resulted in consistently weak correlations observed between C_2 and meteorological parameters such as temperature, SR, and RH. Therefore, we concluded that the changes in meteorology were too small to be the driving factor for C_2 formation.

Comment 8: *Table 1: how may data points are in each category?*

Response: Thanks for your suggestion. We have moved Table 1 to Table S5, which shows correlations between C_2 and various factors under different pollution levels. In addition, we have added number of samples in each category and each year.

Table S5. Correlations between C_2 and various factors under different pollution levels.

	IT0	IT1	IT2	IT3	IT4
Levoglucosan	0.17 (-0.05, 0.37)	-0.03 (-0.22, 0.16)	-0.10 (-0.36, 0.16)	0.01 (-0.23, 0.26)	-0.29 (-0.61, 0.11)
Hopanes	-0.04 (-0.25, 1.08)	-0.21 (-0.38, -0.01) *	-0.05 (-0.31, 0.22)	0.29 (0.05, 0.49) *	0.41 (-0.01, 0.70) *
Octadecanoic acid	0.54 (0.36, 0.68) **	-0.03 (-0.22, 0.16)	0.01 (-0.26, 0.27)	-0.03 (-0.26, 0.22)	0.17 (-0.23, 0.52)

	IT0	IT1	IT2	IT3	IT4
Picene	0.06 (-0.17, 0.28)	-0.28 (-0.46, -0.07) *	-0.18 (-0.45, 0.12)	0.08 (-0.25, 0.39)	0.02 (-0.62, 0.64)
Terephthalic acid	0.40 (0.20, 0.57) **	0.23 (0.04, 0.40) *	0.43 (0.19, 0.62) **	0.34 (0.11, 0.54) *	0.41 (0.04, 0.69) *
Phthalic acid	0.63 (0.47, 0.74) **	0.28 (0.01, 0.45) **	0.44 (0.20, 0.63) **	0.34 (0.11, 0.54) **	0.31 (0.01, 0.54) **
DHOPA	0.19 (-0.13, 0.30) *	0.49 (0.29, 0.60) **	0.45 (0.21, 0.64) **	0.42 (0.20, 0.61) **	0.32 (-0.01, 0.65) **
Malic acid	0.33 (0.13, 0.52) *	0.53 (0.38, 0.66) **	0.66 (0.48, 0.77) **	0.69 (0.44, 0.75) **	0.72 (0.45, 0.87) **
O_x	0.28 (0.05, 0.48) *	0.54 (0.37, 0.68) **	0.56 (0.25, 0.70) **	0.51 (0.42, 0.75) **	0.68 (0.39, 0.84) **
J(O1D)	0.366 (0.15, 0.53) **	0.17 (-0.03, 0.36)	0.33 (0.05, 0.56) *	0.13 (-0.12, 0.37)	-0.09 (-0.49, 0.34)
J(NO2)	0.29 (0.08, 0.48) **	0.14 (-0.07, 0.33)	0.49 (0.24, 0.68) **	0.22 (-0.03, 0.45)	0.02 (-0.40, 0.44)
Sulfate	0.49 (0.28, 0.62) **	0.29 (0.12, 0.46) **	0.60 (0.43, 0.74) **	0.42 (0.21, 0.59) **	0.55 (0.24, 0.76) **
ALWC	0.48 (0.31, 0.65) **	0.36 (0.19, 0.50) **	0.32 (0.09, 0.53) **	0.30 (0.08, 0.49) **	0.15 (-0.01, 0.31)
pH	-0.19 (-0.39, 0.03)	-0.15 (-0.32, 0.03)	-0.38 (-0.57, -0.16) **	-0.01 (-0.24, 0.22)	-0.19 (-0.54, 0.21)
Temperature	0.24 (0.02, 0.43) *	0.42 (0.27, 0.56) **	0.50 (0.30, 0.67) **	0.40 (0.19, 0.58) **	0.63 (0.35, 0.81) **
RH	0.15 (-0.06, 0.36)	0.28 (0.11, 0.44) **	-0.03 (-0.21, 0.26)	-0.03 (-0.19, 0.26)	-0.03 (-0.39, 0.33)
SR	-0.01 (-0.23, 0.21)	0.13 (-0.06, 0.30)	0.43 (0.21, 0.61) **	0.42 (0.21, 0.59) **	0.53 (0.22, 0.75) **

The values in brackets indicate the 95% confidence intervals (CIs) of the correlation coefficients. One, two asterisks denote p values less than 0.05, 0.01, respectively. No asterisk denotes the correlations are not statistically significant.

Table S6. Meteorological parameters, PM2.5 main components, organic molecular tracers, diacids, pH, and ALWC in the PRD (IT0-IT4).

	IT0 *N*=129	**IT1** *N*=144	**IT2** *N*=72	**IT3** *N*=84	**IT4** *N*=33
I. Meteorological parameters					
Temperature (℃)	20.2 ± 2.9	21.5 ± 3.6	21.6 ± 3.4	22.8 ± 3.1	20.8 ± 4.8
Relative humidity (%)	56 ± 12.4	56 ± 13	62 ± 10	67 ± 9	66 ± 7
Solar radiation (W m^{-2})	148.0 ± 43.9	145.6 ± 42.6	118.0 ± 46	115.5 ± 43.4	112.0 ± 50.5
Boundary layer height (m)	578 ± 159	578 ± 134	613 ± 167	583 ± 142	626 ± 154
II. Molecular tracers (ng m^{-3})					
Levoglucosan	333 ± 225	194 ± 131	114 ± 79	96 ± 74	63 ± 34
Hopanes	3.4 ± 2.6	2.0 ± 1.6	1.3 ± 1.9	0.88 ± 0.70	0.54 ± 0.30
Octadecanoic acid	37.5 ± 21.0	28.4 ± 17.2	22.3 ± 14.8	17.3 ± 8.7	11.3 ± 0.93
Picene	0.26 ± 0.20	0.22 ± 0.15	0.18 ± 0.11	0.17 ± 0.10	0.10 ± 0.04
Terephthalic acid	50.0 ± 46.8	48.9 ± 30.7	32.1 ± 31.3	27.9 ± 27.1	14.5 ± 12.4
Phthalic acid	40.3 ± 17.8	29.2 ± 16.0	22.7 ± 10.2	19.6 ± 10.1	14.1 ± 8.8
DHOPA	2.52 ± 2.28	2.27 ± 2.07	1.42 ±1.06	1.05 ± 1.01	0.78 ± 0.43
Malic acid	19.0 ± 19.0	16.6 ± 16.4	9.6 ± 8.3	7.4 ± 6.1	3.9 ± 2.3
III. Aliphatic Diacids (ng m^{-3})					
Oxalic acid (C_2)	619 ± 290	483 ± 200	329 ± 158	293 ± 125	189 ± 102
Succinic acid (C_4)	55.0 ± 49.5	29.3 ± 28.5	18.5 ± 14.2	16.7 ± 12.7	12.9 ± 12.1
Glutaric acid (C_5)	12.5 ± 10.5	6.4 ± 5.9	4.8 ± 2.7	4.2 ± 4.2	4.5 ± 5.6
Adipic acid (C_6)	7.1 ± 4.2	4.9 ± 3.4	4.0 ± 2.7	3.4 ± 2.5	2.9 ± 2.6
Pimelic acid (C_7)	1.9 ± 1.3	1.4 ± 0.8	1.1 ± 0.7	1.1 ± 0.9	0.7 ± 0.5
Suberic acid (C_8)	3.0 ± 2.2	2.5 ± 1.5	2.2 ± 1.3	2.0 ± 1.3	1.4 ± 1.0
Azelaic acid (C_9)	13.5 ± 12.3	11.9 ± 8.3	10.4 ± 7.0	9.6 ± 6.1	6.7 ± 3.8

	IT0	IT1	IT2	IT3	IT4
	N=129	*N*=144	*N*=72	*N*=84	*N*=33
Sebacic acid (C_{10})	2.0 ± 1.8	1.7 ± 1.2	1.6 ± 1.3	1.5 ± 1.1	1.0 ± 0.9
Subtotal	734 ± 337	540 ± 218	358 ± 163	325 ± 135	208 ± 67
IV. Other species					
pH	2.04 ± 0.96	2.40 ± 0.61	2.48 ± 0.43	2.36 ± 0.58	2.11 ± 0.71
ALWC (μg m^{-3})	20.9 ± 11.0	15.1 ± 9.9	13.1 ± 6.9	13.1 ± 8.0	7.2 ± 3.0
O_x (μg m^{-3})	136.7 ± 31.7	134.9 ± 34.4	111.9 ± 27.1	98.5 ± 25.0	72.7 ± 19.1

Table S2. Information of $PM_{2.5}$ samples.

Year	Duration	Number of samples
2007	October to November	32
2008	November to December	45
2009	November to December	25
2010	October to December	69
2011	November to December	28
2012	November to December	39
2013	November to December	29
2014	October to November	20
2015	October to November	37
2016	October to November	33
2017	October to December	55
2018	October to December	50

Comment 9: *Figure 6: The author should consider using the same features to predict other di-acids to see if these features can well capture the variation of other di-acids.*

Response: Thank you for the valuable suggestion. We have used the same features to predict other DCA. Our results show great agreements between measurement data and prediction (R^2=0.72–0.82), which further verify the reliability of our machine learning model.

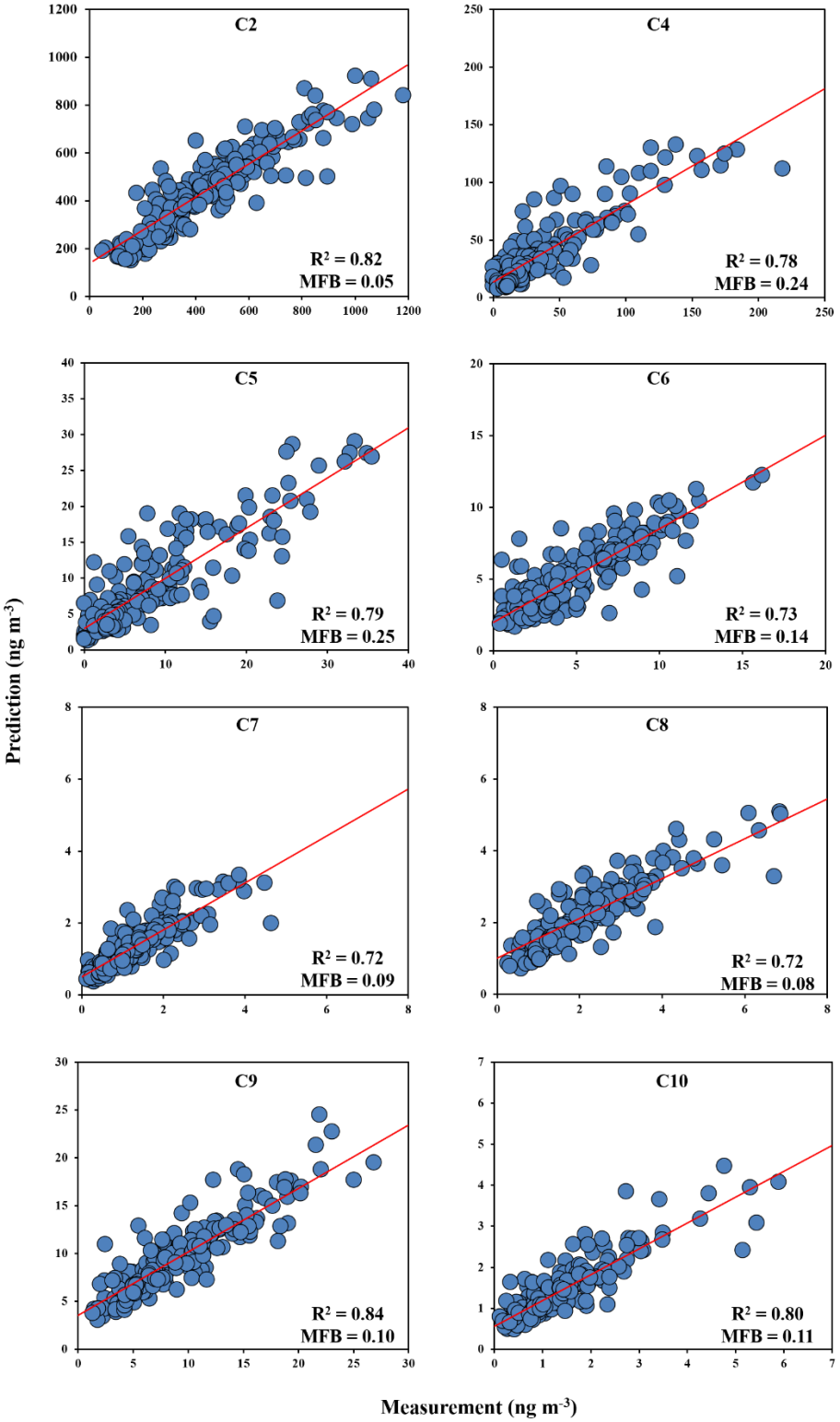

Figure S2. Observations and simulations of DCA

References:

Cheng, Y., Ma, Y., and Hu, D.: Tracer-based source apportioning of atmospheric organic carbon and the influence of anthropogenic emissions on secondary organic aerosol formation in Hong Kong, Atmos. Chem. Phys., 21, 10589-10608, https://doi.org/10.5194/acp-21-10589-2021, 2021.

Ding, X., Zhang, Y. Q., He, Q. F., Yu, Q. Q., Wang, J. Q., Shen, R. Q., Song, W., Wang, Y. S., and Wang, X. M.: Significant Increase of Aromatics-Derived Secondary Organic Aerosol during Fall to Winter in China, Environ. Sci. Technol., 51, 7432-7441, https://doi.org/10.1021/acs.est.6b06408, 2017.

Ehhalt, D. H. and Rohrer, F.: Dependence of the OH concentration on solar UV, Journal of Geophysical Research-Atmospheres, 105, 3565-3571, https://doi.org/10.1029/1999jd901070, 2000.

Glasius, M. and Goldstein, A. H.: Recent Discoveries and Future Challenges in Atmospheric Organic Chemistry, Environ. Sci. Technol., 50, 2754-2764, https://doi.org/10.1021/acs.est.5b05105, 2016.

Hu, D. and Yu, J. Z.: Secondary organic aerosol tracers and malic acid in Hong Kong: seasonal trends and origins, Environ. Chem., 10, 381-394, https://doi.org/10.1071/en13104, 2013.

Kleindienst, T. E., Jaoui, M., Lewandowski, M., Offenberg, J. H., and Docherty, K. S.: The formation of SOA and chemical tracer compounds from the photooxidation of naphthalene and its methyl analogs in the presence and absence of nitrogen oxides, Atmos. Chem. Phys., 12, 8711-8726, https://doi.org/10.5194/acp-12-8711-2012, 2012.

Liu, T. Y., Chan, A. W. H., and Abbatt, J. P. D.: Multiphase Oxidation of Sulfur Dioxide in Aerosol Particles: Implications for Sulfate Formation in Polluted Environments, Environ. Sci. Technol., 55, 4227-4242, https://doi.org/10.1021/acs.est.0c06496, 2021.

Meng, J., Wang, Y., Li, Y., Huang, T., Wang, Z., Wang, Y., Chen, M., Hou, Z., Zhou, H., Lu, K., Kawamura, K., and Fu, P.: Measurement Report: Investigation on the sources and formation processes of dicarboxylic acids and related species in urban aerosols before and during the COVID-19 lockdown in Jinan, East China, Atmos. Chem. Phys., 23, 14481-14503, https://doi.org/10.5194/acp-23-14481-2023, 2023.

Sato, K., Ikemori, F., Ramasamy, S., Fushimi, A., Kumagai, K., Iijima, A., and Morino, Y.: Four- and Five-Carbon Dicarboxylic Acids Present in Secondary Organic Aerosol Produced from Anthropogenic and Biogenic Volatile Organic Compounds, Atmosphere, 12, 1703, https://doi.org/10.3390/atmos12121703, 2021.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Müller, J. F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years, Atmos. Chem. Phys., 14, 9317-9341, https://doi.org/10.5194/acp-14-9317-2014, 2014.

Wang, H., Wu, Q., Guenther, A. B., Yang, X., Wang, L., Xiao, T., Li, J., Feng, J., Xu, Q., and Cheng, H.: A long-term estimation of biogenic volatile organic compound (BVOC) emission in China from 2001–2016: the roles of land cover change and climate variability, Atmos. Chem. Phys., 21, 4825-4848, https://doi.org/10.5194/acp-21-4825-2021, 2021.

World Health Organization: . WHO Global Air Quality Guidelines in 2021, https://www.who.int/news-room/questions-and-answers/item/who-global-air-quality-guidelines, (last access: 17 June 2024), 2021.