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

We sincerely thank reviewers for your time and constructive comments. We have 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.

Anonymous Referee #1

General comments

The manuscript reports long-term field observations of di-acids and its related primary and secondary markers from anthropogenic and biogenic sources at a site in the PRD region, China. It also combines these observations with machine-learning methods to investigate and quantify potential contributions of major drivers to the variation of oxalic acid. Their major findings highlight the increasing importance of gas-phase oxidation in forming SOA. Overall, the topic is valuable with good-quality datasets, but the manuscript needs clearer methodological descriptions, stronger validation of the machine-learning attribution, and more mechanistic and systematic support before publication. For the machine learning methodology part, the attribution is potentially interesting, but I am concerned about robustness given the relatively small dataset (~400 observations) and 11 features. With this sample size there is a substantial risk of overfitting and unstable feature attributions, especially if the data are temporally autocorrelated. I would suggest a major revision.

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.

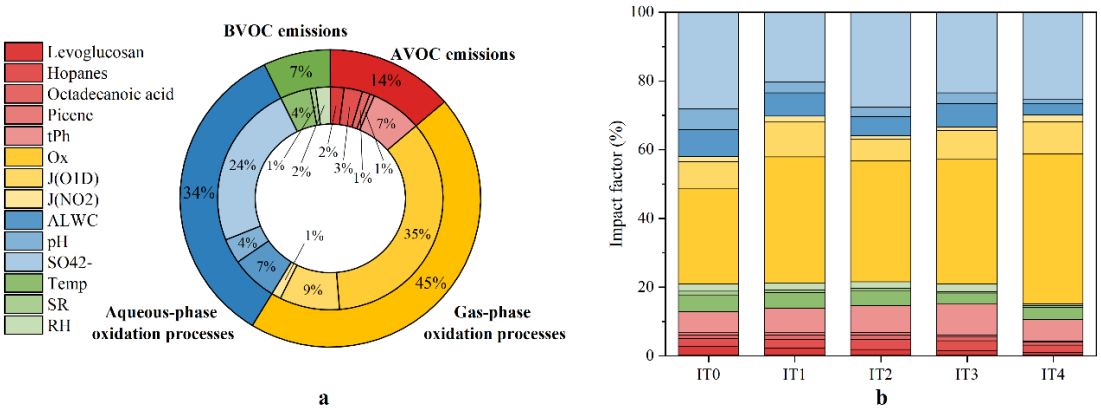

Figure 6 (new). (a) Impact of changes of each variable on C_2 variation during the whole study period. (b) Impact factor of individual variable under different pollution conditions.

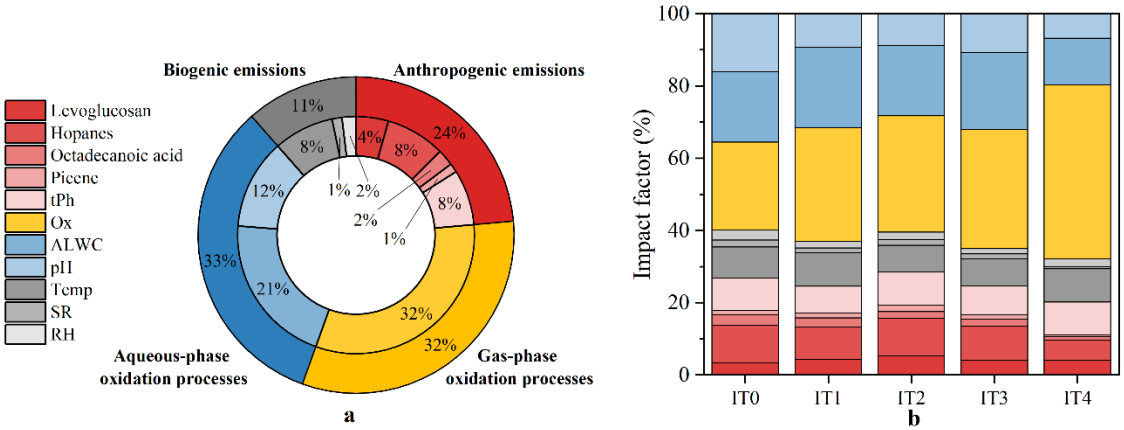

Figure 6 (old).

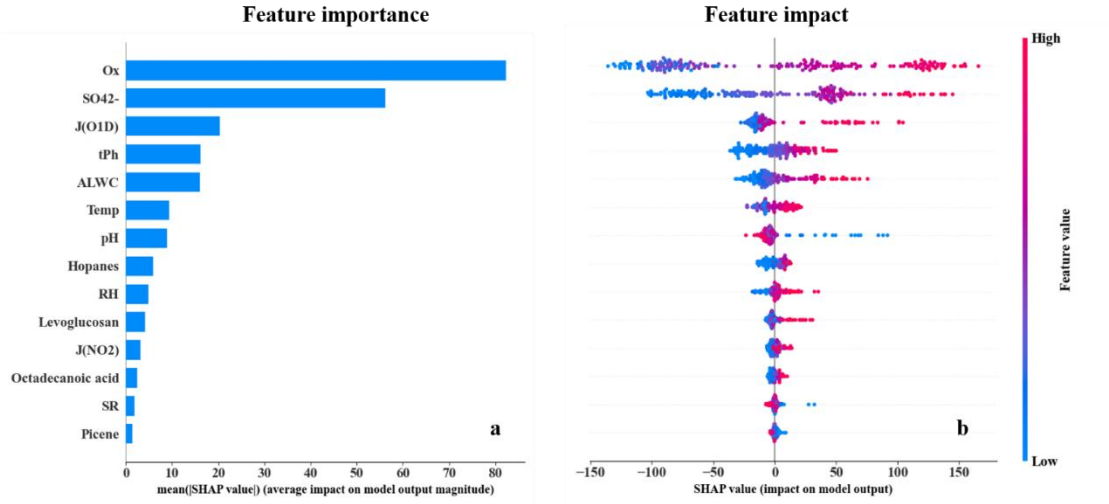

Figure 5 (new). (a) Bar plot of the mean |SHAP| values representing the overall importance of each feature in predicting C_2 concentrations. (b) Beeswarm plot of individual SHAP values for each feature across all samples. Red (blue) represents high (low) value in each feature. Positive (negative) SHAP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

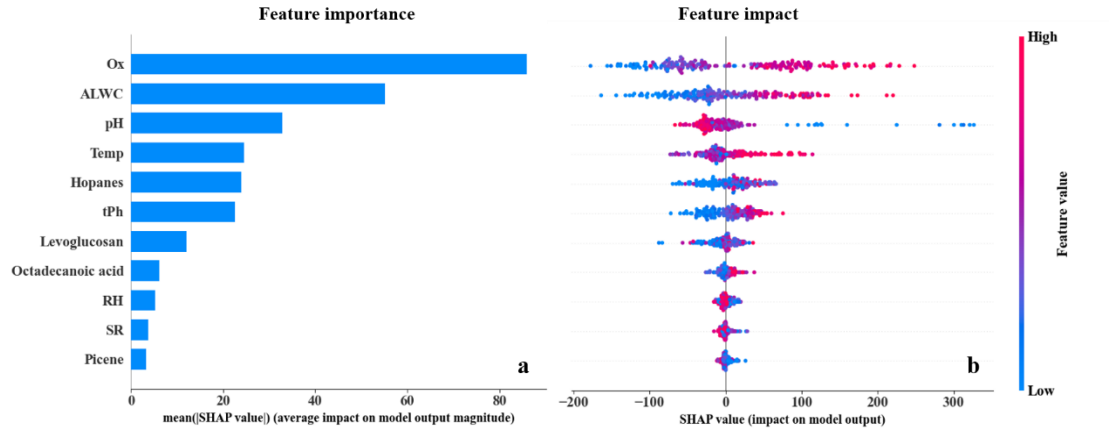

Figure 5 (old).

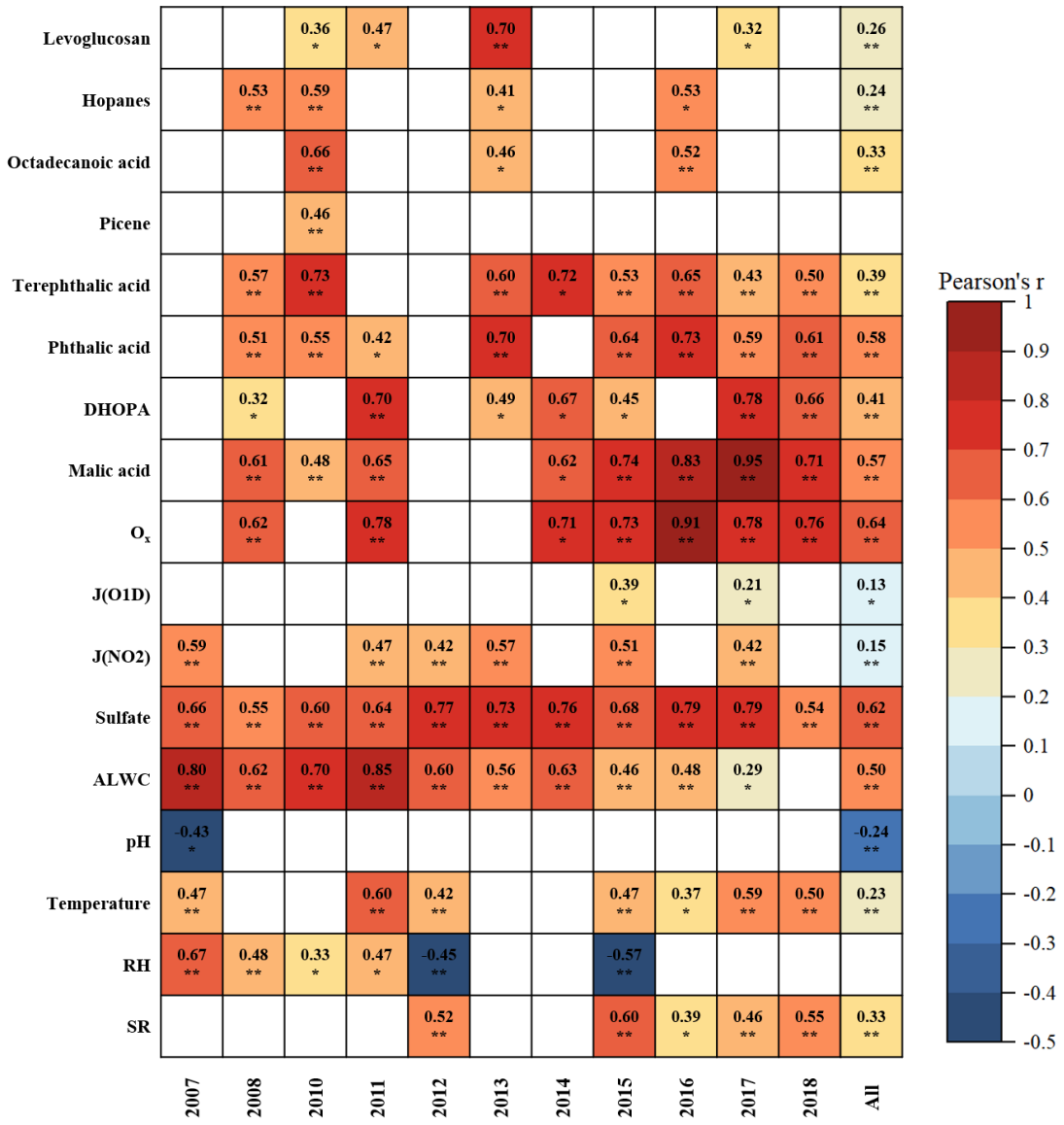

Figure 3. Correlations between C_2 and primary anthropogenic source markers, SOA markers, indicators for gas- and aqueous-phase oxidation, as well as meteorological parameters. Blank cells indicate no significant correlations. One asterisk, two asterisks denote p value < 0.05, 0.01, respectively. Due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

Line 168-179

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.

Line 283-292

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.

Line 347-369

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.

Line 370-375

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.

Line 385-393

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.

Line 394-399

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.

Line 437-439

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.

Line 84-86

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.

Line 214-220

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 °C, 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).

Line 278-282

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.

Line 236-244

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.

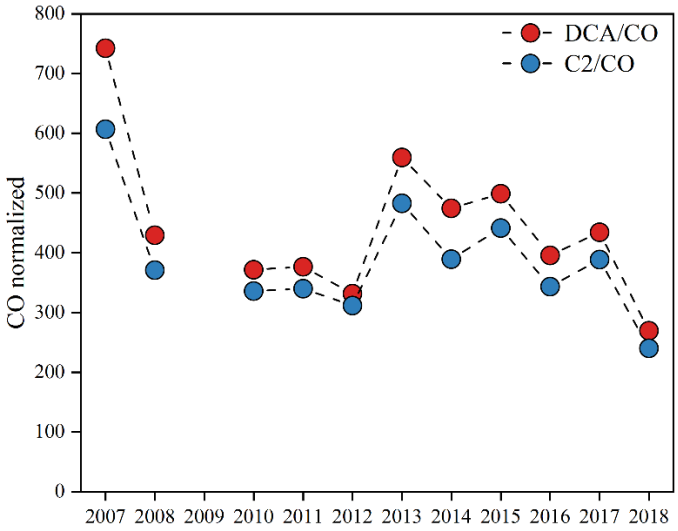

Figure S3. The concentrations of DCA and C2 normalized by carbon monoxide (CO, ppm). Due to the lack of in situ CO measurements at the sampling site, monthly CO data were obtained from the Copernicus Atmosphere Monitoring Service (CAMS) global reanalysis product (EAC4), provided by the European Centre 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.

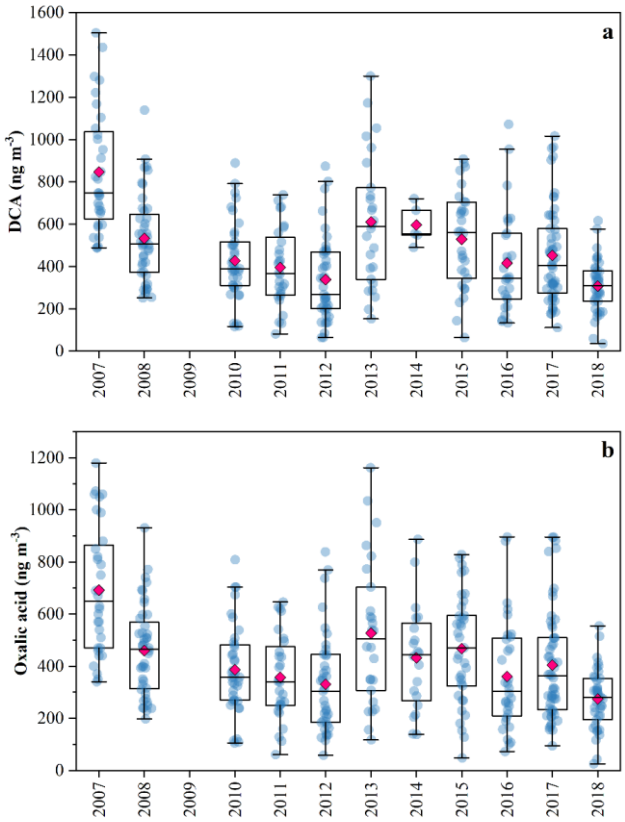

Figure 3. (a) Annual variations in aliphatic DCA. (b) Annual variations in oxalic acid. The concentrations of DCA decreased from 864 ± 283 ng m^{-3} (2007) to 307 ± 122 ng m^{-3} (2018), and the concentrations of oxalic acid decreased from 692 ± 243 ng m^{-3} (2007) to 274 ± 114 ng m^{-3} (2018), but 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.

Line 265-275

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)
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)

	IT0	IT1	IT2	IT3	IT4
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 (°C)	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
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.

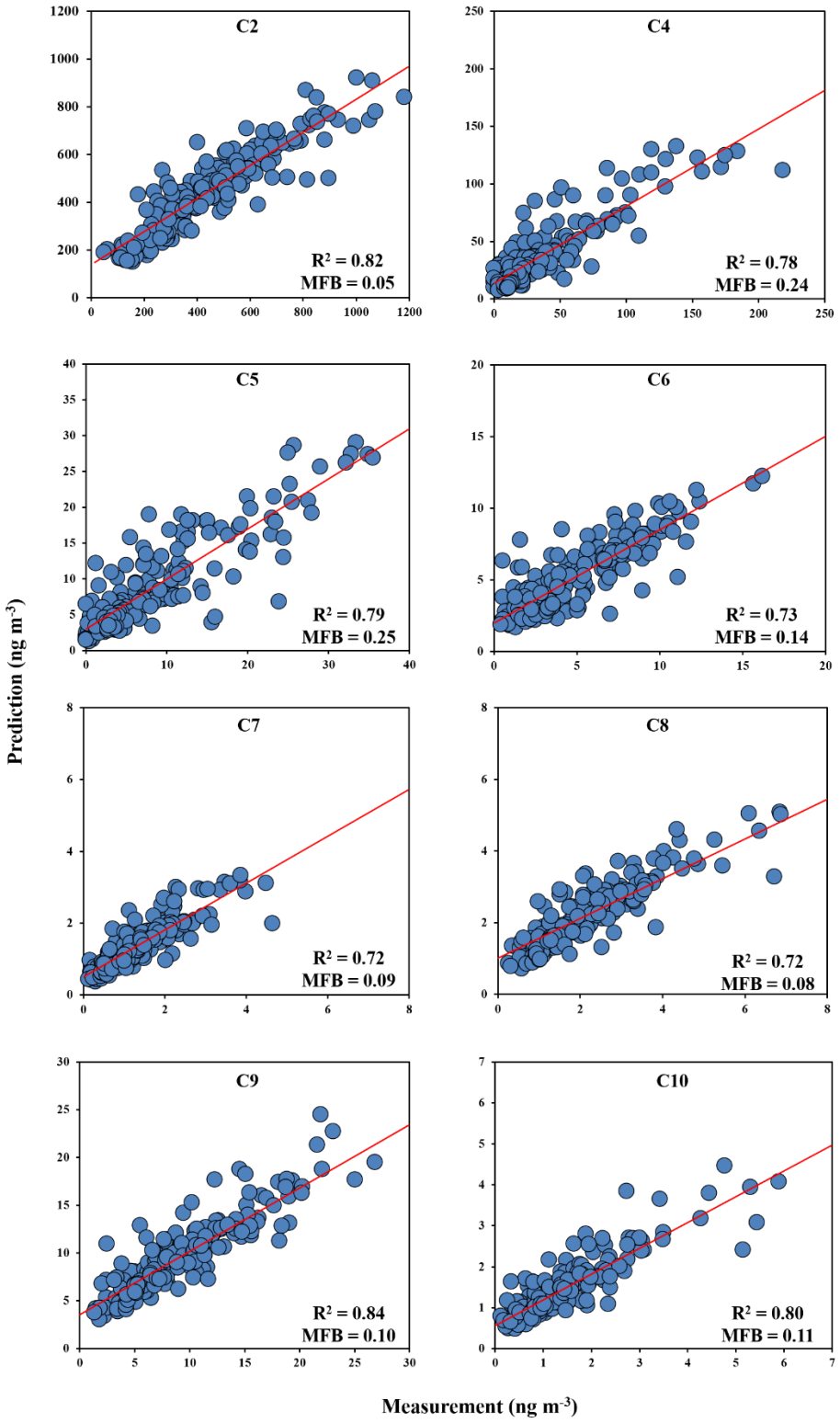

Figure S2. Observations and simulations of DCA

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.

Anonymous Referee #2

General comments

This work analyzed a long-term variation of oxalic acid in atmospheric aerosols in the Pearl River Delta from 2007 to 2018. Aerosol liquid water and Ox are the driving factors of C2 formation, and gas-phase oxidation would play a more important role than aqueous-phase oxidation as air pollution decreases. The long-term data is valuable and informative. However, the data interpretation needs to be more rigorous. I may suggest a major revision before publication.

Major comments:

Comment 1: *The ALWC and Ox are identified as the drivers of C2 variation. The authors also highlight the increased contribution of gas-phase oxidation and decreased contribution of aqueous-phase oxidation to C2 formation as pollution levels declined. However, high air pollution level is usually accompanied by high humidity and low Ox.*

Response: Thanks for comment. In Table S6, we provided information under different pollution levels during our study period. Because the sampling was conducted in the same season (mainly from October to December), the inter-annual differences in meteorological condition are small. Relative humidity fluctuated within a narrow range (56% ~ 67%) under different pollution levels. Due to strict emission control measures during the past decades in the Pearl River Delta (PRD) region (Bian et al., 2019), dramatic reductions were identified in sulfate and nitrate, which are the highly hygroscopic compounds in $PM_{2.5}$. As a result, aerosol liquid water content (ALWC) decreased significantly from 20.9 μg m^{-3} (high pollution, IT0) to 7.5 μg m^{-3} (low pollution, IT4).

In many urban environments, high pollution levels are often accompanied by suppressed O_x concentrations. This results from that heavy pollution events are typically characterized by stagnant meteorological conditions, high relative humidity, and elevated NO emissions. Under such conditions, strong NO titration efficiently removes O_3, leading to lower O_3 and thus lower O_x. However, previous WRF-Chem simulations have demonstrated that the response of O_3 to anthropogenic emission changes exhibits strong regional heterogeneity in China (Wang et al., 2021). For example, emission reduction due to the COVID-19 lockdown increased O_3 concentrations in the Yangtze River Delta (YRD) and central China, while it led to O_3 decrease in most parts of the PRD. This could be explained by the regional disparities of the O_3–NO_x–VOC regime. In PRD, most rural areas were NO_x-limited (our sampling site is located in a rural area), which means that the reductions in anthropogenic emissions would lead to lower O_3. This is consistent with our observations showing a decreasing trend in O_x concentrations from 163.7 μg m^{-3} (high pollution, IT0) to 72.7 μg m^{-3} (low pollution, IT4).

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

	IT0	IT1	IT2	IT3	IT4
I. Meteorological parameters					
Temperature (°C)	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
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

Comment 2: *How did the authors exclude the impacts of different emission levels when addressing the influence of RH and Ox on C2 variation?*

Response: Thanks for the comment. Anthropogenic volatile organic compounds (AVOCs) are indeed important precursors of C_2, and their influences should be carefully examined. In this study, instead of excluding impacts of changes in anthropogenic emission, we demonstrated the impacts were limited from the several complementary perspectives.

First, long-term trends showed that anthropogenic emissions (biomass burning, vehicle emission, and cooking) decreased significantly from 2007 to 2018, whereas C_2 did not experience corresponding decline. This initial finding was further supported by correlation analysis, which revealed the consistently weak correlations between C_2 and primary anthropogenic markers

throughout the entire dataset. These findings indicated that changes in anthropogenic emissions were not the dominant driving factors for C_2 variations. In contrast, O_x and ALWC displayed strong correlations with C_2, suggesting they are main drivers of C_2 variability.

However, we acknowledge that when investigating the influence of O_x and ALWC on C_2 variations under different pollution levels (IT0-IT4), the original manuscript did not include the corresponding correlations between C_2 and other factors within each category. The absence of this information indeed weakens the logical completeness of our argument. To address this, we have added the relevant results to the Supplement (Table S7). It shows the correlations between C_2 and primary anthropogenic markers (Levoglucosan, hopanes, octadecanoic acid, picene, and terephthalic acid) remain generally weak across all pollution categories. This indicates that changes in anthropogenic emissions exert only limited influence on C_2 variations. Moreover, fluctuations in anthropogenic markers within each category were not substantial. Thus, it is reasonable that we apply the correlation coefficients between C_2 and O_x/ALWC to reflect their influences on C_2 formation.

Additionally, instead of simply excluding anthropogenic emissions from consideration, we quantitatively assessed their influence using a machine learning model. The results show that changes in anthropogenic emissions contribute only a small portion (~14%) of the overall variability in C_2. This quantitative evidence strengthens our conclusion that variations in gas-phase and aqueous-phase oxidation are the dominant drivers of C_2 variations.

Table S7. 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)
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.

Line 265-275

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 C_2 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 small. 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.

Line 315-329

The opposite trends implied the roles of gas-phase and aqueous-phase oxidation in C_2 formation might change. However, the precision and stability of Pearson's r values are strongly influenced by sample size when the variations appear small or when sample sizes differ among groups. Therefore, the differences in Pearson's r values do not necessarily imply statistically significant changes, especially when they are very close (IT1–IT4). To assess the statistical significance of these differences, we compared correlation coefficients between groups using the method described in Text S1. As shown in Table S8–S9, significant differences in the C_2-O_x correlation were observed only between IT0 and the other pollution levels. For the C_2-ALWC correlation, a significant difference was found only between IT0 and IT4. Given that IT1–IT4 represents a continuous evolution of atmospheric conditions, rather than discrete and independent regimes, large differences in correlation coefficients among these categories are not expected. Although the correlation between C_2 and sulfate was strong, it did not show the similar trends as that between C_2 and ALWC. In contrast, the correlations between C_2 and primary anthropogenic markers remained generally weak across all pollution categories (Table S7), indicating that changes in anthropogenic emissions exert only limited influence on C_2 variations. Therefore, the significant and opposite changes in correlations of C_2 with O_x and ALWC between high pollution level (IT0) and low pollution level (IT4) suggested a shift in the dominant C_2 formation pathway from aqueous-phase oxidation to gas-phase photochemical oxidation under lower pollution conditions.

Comment 3: *In Fig. 5, the correlation between C2 and Ox is always higher than that between C2 and ALWC under IT1, 2, 3, and 4. The correlation between C2 and ALWC is weak under any IT*

condition.

Response: Thanks for the comment. In this study, both O_x and ALWC display significant positive correlations with C_2 under most pollution levels. Our results show that the correlation coefficients between C_2 and O_x are consistently higher than that between C_2 and ALWC under IT1–IT4, whereas the C_2-ALWC correlations is higher than that of C_2-O_X under IT0 (Figure 4, see below). This pattern likely results from the reduction in pollution levels (accompanied by decreases in ALWC), which weakens the role of aqueous-phase oxidation in C_2 formation, while the contribution from gas-phase oxidation becomes relatively more important.

We note that the Pearson correlation coefficients reflect the strength of linear associations but do not directly represent the quantitative contribution of each factor. Thus, the lower Pearson's r values between C_2 and ALWC than those between C_2 and Ox under IT1-IT4 do not necessarily imply a smaller contribution of aqueous-phase pathways compared with gas-phase pathways. Here, correlation analysis is used primarily to identify potential drivers and their changing patterns, while a machine learning model is further applied to quantify their contributions to C_2 variations.

Although the correlations between C_2 and ALWC is relatively weak under IT1-IT4, they are still stronger than those for other factors (Table S7, see above). This weak correlation may arise from several reasons: (1) reductions in pollution levels (and ALWC) weaken the influence of aqueous-phase oxidation; (2) the effect of ALWC on C_2 may be non-linear and not fully captured by linear correlation; (3) ALWC serves as an indicator of aqueous-phase processes in this study, but aqueous production also depends on other factors such as transition metals and oxidant availability.

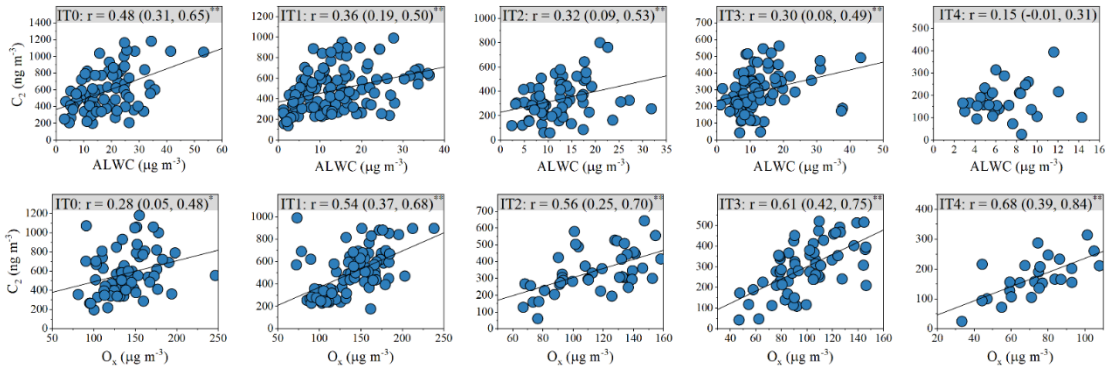

Figure 4. The correlation between C_2 and ALWC, as well as O_x. The values in brackets indicate the 95% confidence intervals (CIs) of the correlation coefficients. One, two asterisks denote p values less than 0.05, 0.01, respectively. With decreasing pollution levels, the correlation between C_2 and ALWC weakens, whereas that between C_2 and O_x strengthens.

Line 330-334

The Pearson correlation coefficients reflect the strength of linear associations but do not directly represent the quantitative contribution of each factor. Thus, the lower Pearson's r values

between C_2 and ALWC than those between C_2 and O_x under IT1-IT4 do not necessarily imply a smaller contribution of aqueous-phase pathways compared with gas-phase pathways. Here, correlation analysis is used primarily to identify potential drivers and their changing patterns, while a machine learning model is further applied to quantify their contributions to C_2 variations.

Comment 4: *Section 3.4: For the machine learning analysis, the authors quantify the contributions of different sources using some input parameters. The rationality of this approach needs to be elaborated. For example, air temperature, solar radiation, and relative humidity are used to represent the emission of biogenic precursors (lines 318-319). Do all these meteorological factors promote the emission of biogenic precursors? For biogenic emissions, is there a synergistic or antagonistic mechanism between these factors? Please explain in detail. The reasonability of using the input parameters to represent other sources also needs to be elaborated.*

Response: Thanks for this valuable comment. We apologize for the missing explanation regarding the selection of input parameters in the original manuscript. Each variable used in the machine learning model should be chosen rationally to ensure the reliability of the results. We have now added a detailed discussion on the selection of input parameters in Section 3.4.

Line 347-369

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. 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}). α = 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, temperature, solar radiation, and RH are important factors determining BVOC emissions. As discussed previously, O_x, $J(O^1D)$, and $J(NO_2)$ can be employed as indicators of gas-phase oxidation, while 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.

Comment 5: *For the contribution of gas-phase oxidation versus aqueous-phase oxidation, is the result here obtained based on machine learning comparable to those reported in published literature?*

Response: A previous stable carbon isotopic study conducted in North China (urban site) reported that the contributions of gas-phase and aqueous-phase pathways to C_2 formation accounted for 12.3% and 47.2% during average days, but shifted to 50.5% and 16.1% during the COVID-19 lockdown.

However, in this study, we investigate the "impacts of **changes** in gas-phase oxidation and aqueous-phase oxidation" by machine learning model, rather than the "**absolute** contributions of gas-phase oxidation and aqueous-phase oxidation". The former is independent of VOCs precursors because SHAP values reflect the marginal impact of a unit change in each variable on the predicted C_2 concentration while keeping other variables constant (discussed in Section 2.4). In contrast, the latter is dependent on VOCs precursors. Therefore, our machine learning approach provides a more appropriate and meaningful assessment of process-driven changes in C_2 formation. Our updated results show that gas-phase and aqueous-phase pathways account for 45% and 34% (after including additional factors following Reviewer #1's suggestion) for C_2 variations generally. In addition, after inclusion of new factors, the results and general conclusions are similar to our previous version. This further enhances the reliability of our method. The comparison between results of new and old versions is presented below. We apology for some unclear statements in original text, which lead to misunderstanding. We have revised them now.

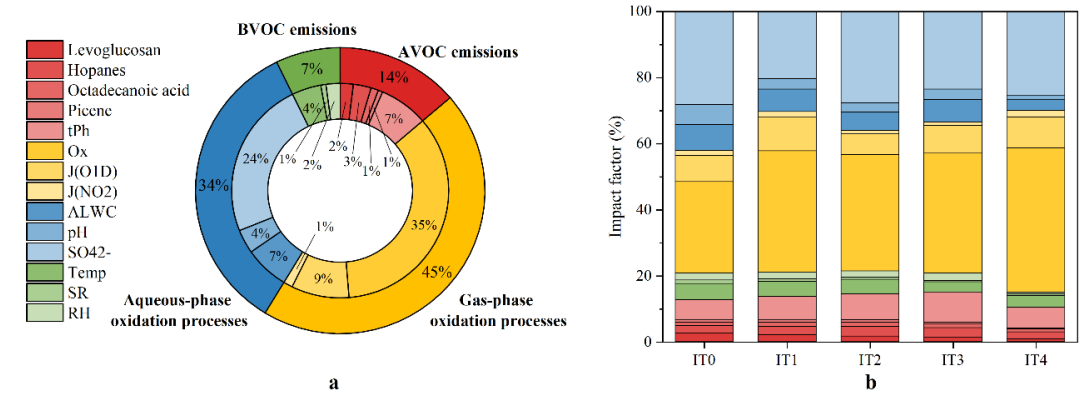

Figure 6 (new). (a) Impact of changes of each variable on C_2 variation during the whole study period. (b) Impact factor of individual variable under different pollution conditions.

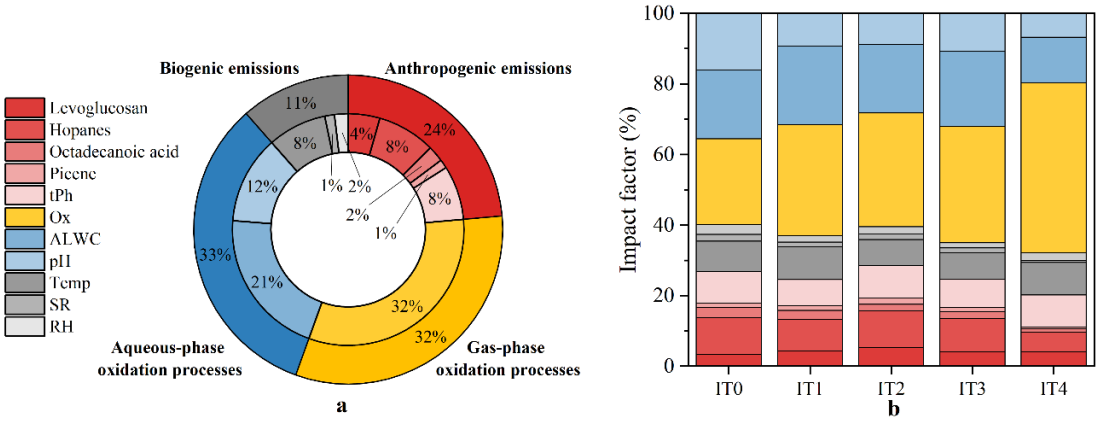

Figure 6 (old).

Line 28-42

Secondary organic aerosol (SOA) is a dominant constituent of fine particulate matter, exerting significant impacts on both climate and human health. Oxalic acid (C_2), a key end-product formed from the oxidation of volatile organic compounds, can provide insights into the formation mechanism of SOA. Thus, long-term measurements of C_2 and related compounds help understand the changes in SOA formation with decreasing pollutant levels. In this study, C_2 and its homologs, along with five primary anthropogenic source markers and three SOA markers, were measured in the Pearl River Delta (PRD) during 2007–2018. The concentrations of C_2 did not exhibit significant downward trends, despite substantial reductions in anthropogenic emissions, such as biomass burning (-11% yr^{-1}), vehicle emissions (-17% yr^{-1}), and cooking emissions (-7% yr^{-1}). Correlation analysis revealed that aerosol liquid water content (ALWC) and O_x ($O_3 + NO_2$) were the main drivers of C_2 variations. Moreover, the relative contribution of biogenic SOA increased under cleaner conditions. A machine learning model was applied to quantify the impacts of changes in anthropogenic precursor emissions, biogenic precursor emissions, aqueous-phase oxidation processes, and gas-phase oxidation processes on C_2 variability. As pollution levels declined, the

impacts of gas-phase oxidation increased from 37% to 55%, whereas that of aqueous-phase oxidation declined from 42% to 30%. This shift indicated a transition from aqueous-phase to gas-phase pathways in C_2 and SOA formation. Our findings highlight the increasing importance of gas-phase oxidation under low-pollution conditions and underscore the need for effective ozone control strategies to further reduce SOA in the future.

Line 385-393

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.

Comment 6: *Figures 2 and 3: I am curious about the high levels since 2013. Please explain the reasons.*

Response: Thanks for the comment. In our previous study, we observed that there was also a rebounce in $PM_{2.5}$ and its main components since 2013 (Figure 7, see below) (He et al., 2025). Importantly, that analysis employed a measurement technique distinct from the analytical procedures used for organic molecular markers in this study, suggesting that the observed rebound is unlikely to arise from methodological artifacts. We recognize that examining long-term trends of molecular markers is inherently challenging, as even minor changes in sampling or analytical protocols may introduce discontinuities. Consequently, it was necessary to assess the stability of both the sampling procedures and the analytical system over the entire study period.

Because background concentrations of many organic molecular markers were below detection limits, we examined the long-term patterns of $PM_{2.5}$ main components in blank filter samples to evaluate whether sampling or environment biases were related to this phenomenon. As shown in Figure 8 (see below), these species exhibited minimal variability in the blank filter samples and no obvious increase was observed since 2013, indicating that the potential sampling bias was negligible.

To further assess analytical stability, we examined the response factors (RFs) derived from annual

calibration curves for all quantified compounds (Table S4, see below). The RF values remained highly consistent across years, indicating that neither instrument sensitivity nor analytical performance experienced significant drift during the measurement period.

Meteorological variability was also considered as a potential driver. However, interannual differences in temperature, relative humidity, solar radiation, and boundary layer height were small (Table S5), suggesting that meteorology alone cannot explain the observed rebound.

In addition, another study covering multiple stations in the PRD region also reported a similar increase in $PM_{2.5}$ and its main components since 2013 (Figure 9) (Yan et al., 2020), supporting that the rebound was not due to newly emerging pollution sources near our sampling location. However, the underlying cause of this rebound was not elaborated in that study.

Unfortunately, until now, we can not fully explain the rebound since 2013. We have incorporated the relevant information and clarifications in Materials and Method section, which confirms the stability and reliability of our long-term measurements and demonstrate that the observed trends in molecular markers were not driven by methodology.

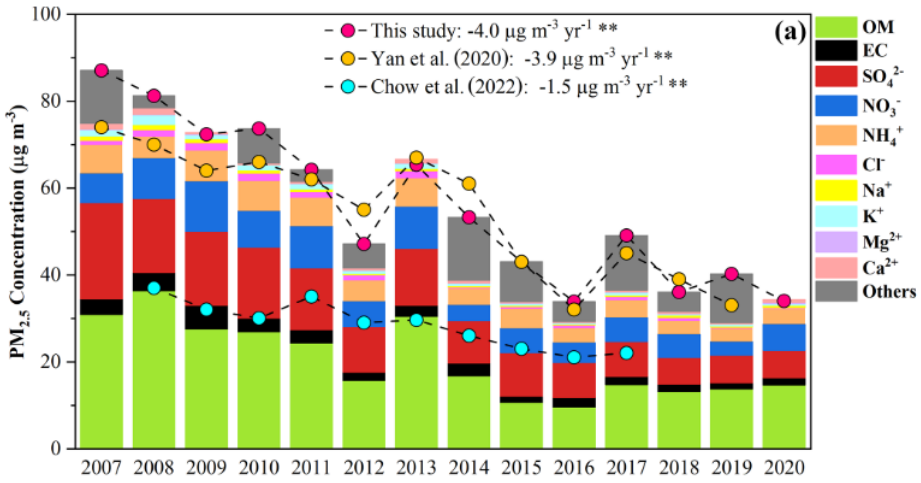

Figure 7. Trend of PM2.5 and its major components (He et al., 2025).

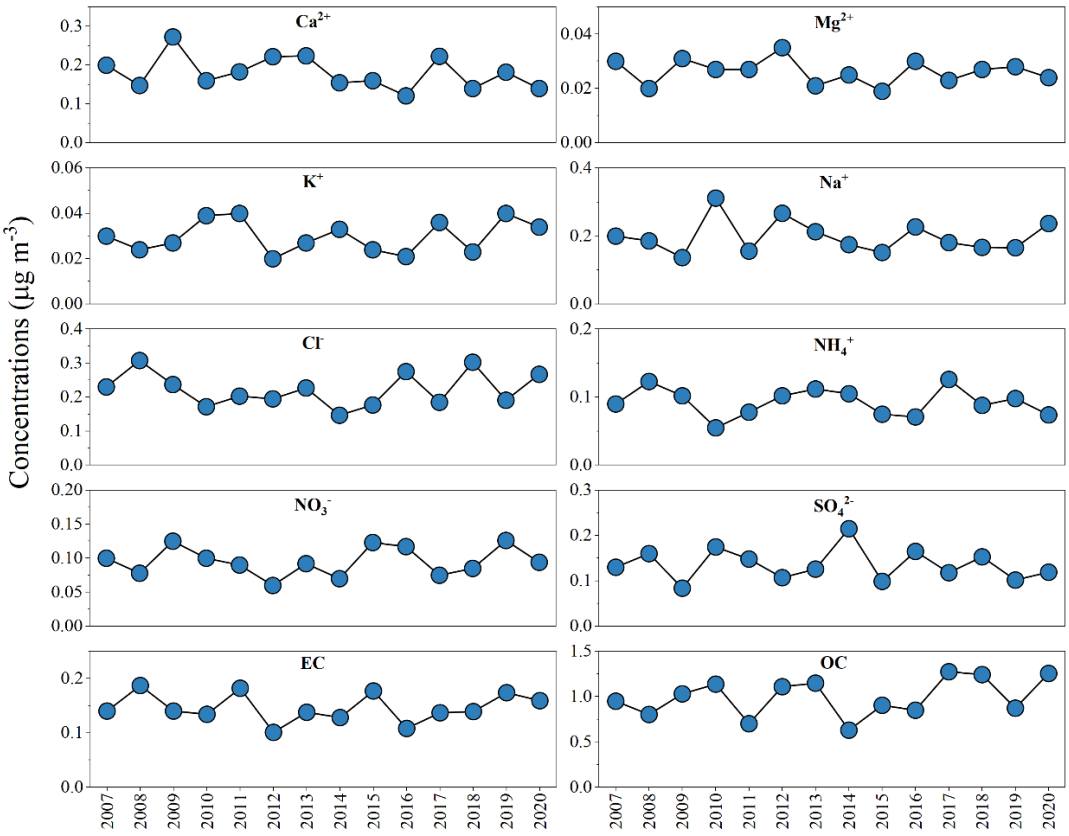

Figure 8. Annual variations in measured compounds of blank filter samples (He et al., 2025).

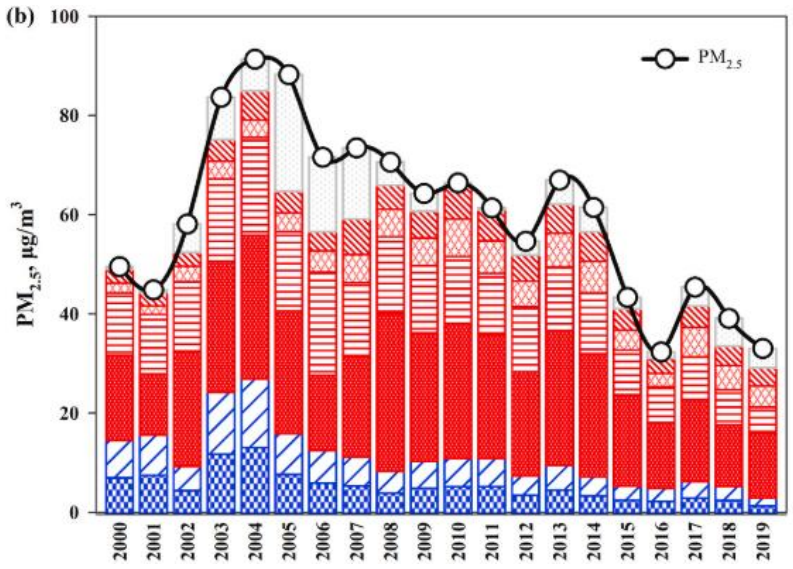

Figure 9. The concentration of PM2.5 and its chemical compositions from 2000 to 2019 in PRD (Yan et al., 2020).

Table S4. Response factors (RFs) derived from the annual calibration curves.

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Succinic acid	1.79	3.66	0.90	0.93	0.93	1.47	2.11	1.47	1.47	1.65	2.41	1.52
Glutaric acid	2.88	4.07	3.31	2.30	2.30	2.93	2.70	2.93	0.81	3.95	5.53	2.38
Adipic acid	2.64	3.87	2.82	2.81	2.81	3.70	3.66	3.76	1.54	5.86	3.76	3.24
Pimelic acid	2.16	2.53	2.30	2.19	2.19	3.97	3.90	3.97	3.96	5.78	3.76	4.54
Suberic acid	2.23	2.47	2.70	2.68	2.68	4.6	3.06	4.60	5.41	7.36	4.39	6.45
Azelaic acid	2.03	2.03	2.42	2.41	2.41	3.99	3.12	3.98	5.07	6.37	3.44	6.87
Sebacic acid	3.13	3.27	4.31	4.29	4.29	5.79	5.54	5.79	5.79	4.83	4.93	4.45
Phthalic acid	1.01	1.07	0.86	0.66	0.98	1.92	1.86	1.93	1.88	1.88	1.98	1.91
Terephthalic acid	1.01	1.07	0.86	0.66	0.98	1.92	1.86	1.93	1.88	1.88	1.98	1.91
17α(H)-22,29,30-trisnorhopane	0.79	0.79	0.70	0.79	0.55	0.83	0.99	0.82	0.91	0.53	0.67	0.76
17α(H),21β(H)-30-norhopane	0.73	0.73	0.63	0.73	0.53	0.74	0.87	0.75	0.78	0.46	0.62	0.78
17α(H),21β(H)-30-hopane	0.78	0.78	0.74	0.78	0.52	0.78	0.88	0.76	0.42	0.70	0.59	0.97
17α(H),21β(H)-22R-homohopane	1.26	1.26	0.72	1.26	0.45	1.11	1.50	1.12	0.42	0.73	0.91	1.75
17α(H),21β(H)-22S-homohopane	1.55	1.55	0.72	1.55	1.33	1.25	1.90	1.19	0.42	1.83	1.01	1.79
Levoglucosan	1.35	1.30	1.22	0.92	1.04	0.93	0.97	0.93	0.97	0.96	0.95	0.96
Octadecanoic acid	0.50	0.56	0.66	0.68	0.96	0.89	1.29	0.89	1.22	0.71	0.89	1.09
Picene	0.84	0.80	0.84	0.77	0.78	0.66	0.80	0.66	0.79	1.11	0.81	1.06
Citramalic acid	1.35	1.95	1.30	3.70	3.56	2.60	2.10	2.59	2.63	2.07	2.04	2.09
Malic acid	1.46	1.46	1.16	0.95	0.91	0.87	0.86	1.02	1.15	0.85	0.76	1.05

Line 158-162

Examining long-term trends of molecular markers is inherently challenging, as even minor changes in sampling or analytical protocols may introduce discontinuities. Consequently, it was necessary to assess the stability of both the sampling procedures and the analytical system over the entire study period. Table S4 presents the response factors (RFs) of all measured species derived from the annual calibration curves. The RF values remained highly consistent across years, indicating that neither instrument sensitivity nor analytical performance experienced significant drift during the measurement period.

Line227-232

Unexpectedly, a rebound in the concentrations of molecular markers and aliphatic DCA was observed in 2013, which was also observed in $PM_{2.5}$ and its major component (He et al., 2025). This rebound was unlikely driven by meteorological variability, as inter-annual differences in key meteorological parameters were relatively small (Table S5). Similarly, another study covering multiple stations across the PRD also report a post-2013 increase in $PM_{2.5}$, supporting that the rebound was not attributable to newly emerging pollution sources near our sampling site (Yan et al., 2020). Unfortunately, until now, the underlying cause of this rebound remained unclear.

Comment 7: *Lines 250-252: "Meanwhile, the correlations between C2 and ASOA markers became weaker. These results suggested ...." Please explain this statement. In Table 1, I did not find an obvious decreasing trend for the correlation coefficients between C2 and Phthalic acid (changing from 0.28 under IT1 to 0.31 under IT4) or DHOPA (changing from 0.49 under IT1 to 0.32 under IT4) from IT1 to IT4.*

Response: Thanks for the comment. We apologize for this imprecise statement. Indeed, the decreases in correlations between C_2 and Phthalic acid, as well as DHOPA relatively small. But the correlations between C_2 and malic acid displayed an obvious increase, suggesting that the relative contribution of BVOCs becomes more important under cleaner environment. We have moved Table 1 to Supplement (Table S7), and removed "Meanwhile, the correlations between C_2 and ASOA markers became weaker" in the manuscript.

Table S7. 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)
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) **

	IT0	IT1	IT2	IT3	IT4
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) **

Line 278-282

As discussed previously, malic acid can be produced by photooxidation of both anthropogenic and biogenic precursors. However, no corresponding increasing trends were observed in the correlations between C_2 and ASOA tracers (Ph and DHOPA), supporting that anthropogenic precursors were not the dominant source of malic acid in this study. Thus, these results indicated that the relative contributions of biogenic sources to SOA become more important under cleaner conditions.

Comment 8: *Lines 287-290: I did not see an obvious difference in the correlation efficiency between C2 and Ox or between C2 and ALWC from IT1 to IT4. The change of Pearson r values seems small.*

Response: Thanks for your comment. Although Pearson correlation coefficients provide a measure of association strength, the differences in Pearson r values do not necessarily imply statistically significant changes, especially when the variations appear small or when sample sizes differ among groups. We agree that the significance of the differences in correlation coefficients between different categories need to be further verified, because they are very close from IT1 to IT4.

To further evaluate whether the observed differences in correlation coefficients across the different groups are statistically meaningful, we applied the Fisher *r*-to-*z* transformation test. This method converts Pearson r values into approximately normally distributed z-scores, enabling a rigorous statistical comparison between two independent correlations. Thus, it allows us to determine whether the correlation strength between C_2 and ALWC or between C_2 and O_x differs significantly under different pollution levels, thereby providing a more robust basis for our interpretation. We have added the methodological description and the corresponding results in the Supplement (Text S1, Table S8-S9; see below). In addition, we calculated the 95% confidence intervals (95% CIs) of Pearson r values for each group and included them in Figure 4.

The results show that, for the correlation between C_2 and O_x, significant differences in Pearson r values are only observed between IT0 and other pollution levels. For the correlation between C_2 and

ALWC, significant difference in Pearson r values is only observed between IT0 and IT4. Although r value itself is not a function of sample size, its precision and stability are strongly influenced by sample size. Smaller sample size leads to wider confidence intervals and greater variability in Pearson r, which may mask the differences between groups. In addition, IT-IT4 represents a continuous evolution of atmospheric conditions, rather than discrete and independent regimes. Consequently, the differences in correlation coefficients among IT1-IT4 are not expected to be very large. We acknowledge that the differences in correlation coefficients from IT1 to IT4 are not statistically significant due to the limit of our relatively small sample size in each group, and it is imprecise to claim an increase/decrease trend in correlation between C_2 and O_x/ALWC with reductions in pollution levels. But the significant differences between IT0 and IT4 for O_x (increasing) and ALWC (decreasing) suggests that there could be a potential shift in the dominant formation pathways in C_2 formation from high to low pollution stage. To verify this hypothesis, we conducted machine learning in the next section, which generated more evidences to support our conclusions.

We have corrected our statements in manuscript to make them more rigorous.

Table S5. Significance (p values) of the difference between correlation coefficients in different categories (C_2-ALWC).

	IT0	IT1	IT2	IT3
IT1	0.25			
IT2	0.21	0.76		
IT3	0.15	0.64	0.89	
IT4	< 0.05	0.22	0.37	0.43

Table S6. Significance (p values) of the difference between correlation coefficients in different categories (C_2-O_x).

	IT0	IT1	IT2	IT3
IT1	< 0.01			
IT2	< 0.05	0.84		
IT3	< 0.01	0.47	0.65	
IT4	< 0.01	0.22	0.33	0.55

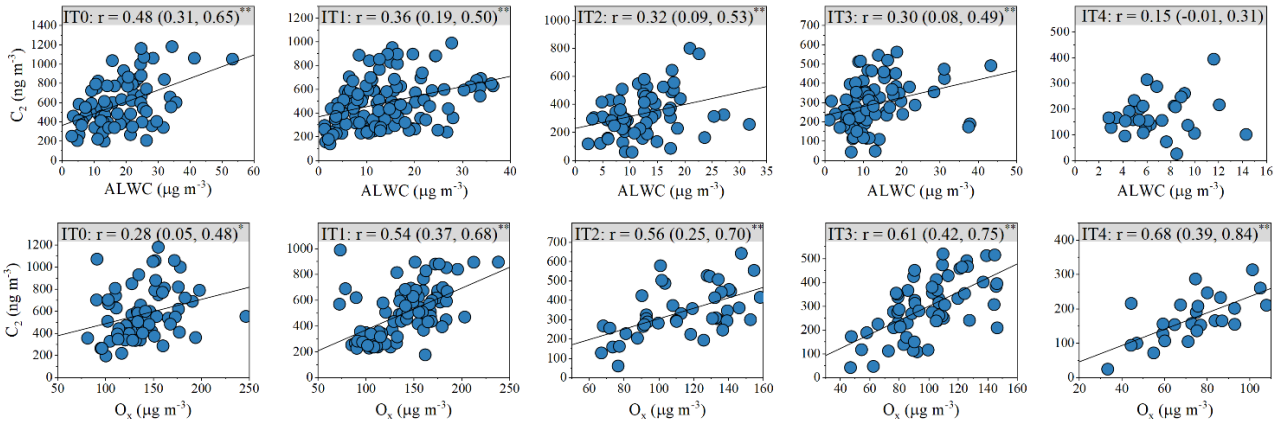

Figure 4. The correlation between C_2 and ALWC, as well as O_x. The values in brackets indicate the 95% confidence intervals (CIs) of the correlation coefficients. One, two asterisks denote p values less than 0.05, 0.01, respectively. With decreasing pollution levels, the correlation between C_2 and ALWC weakens, whereas that between C_2 and O_x strengthens.

Text S1. Descriptions of Fisher *r*-to-*z* transformation.

The Pearson correlation coefficient *r* is widely used to quantify the strength and direction of linear relationships between two variables. However, the sampling distribution of *r* is not normally distributed, especially when the true correlation is far from zero or the sample size is small. To address this issue, Fisher (1921) proposed a transformation of *r* to a variable *z*, known as the Fisher *r*-to-*z* transformation, defined as:

$$z = \frac{1}{2} ln \frac{1+r}{1-r} \quad (1)$$

This transformation converts *r* into an approximately normal distribution, allowing for more accurate estimation of confidence intervals and hypothesis testing. The standard error of *z* is given as:

$$SE = 1/\sqrt{n-3} \quad (2)$$

where *n* is the sample size. After calculating the confidence interval in the *z*-space, it can be back-transformed to the original *r* scale, providing a robust measure of uncertainty for correlation estimates.

Furthermore, the Fisher *r*-to-*z* transformation can also be used to test whether two correlation coefficients from independent samples differ significantly. For two correlations r_1 and r_2 with sample sizes n_1 and n_2, their corresponding *z* values are calculated as above, and the standard error of the difference is calculated as:

$$SE = 1/\sqrt{\frac{1}{n1-3} + \frac{1}{n2-3}} \quad (3)$$

The difference is then standardized as:

$$z = \frac{z1-z2}{SE} \quad (4)$$

A two-tailed p value can be derived from the standard normal distribution to determine whether the

difference between r_1 and r_2 is statistically significant. This approach provides a rigorous method for comparing correlation strengths across independent datasets.

Line 315-329

The opposite trends implied the roles of gas-phase and aqueous-phase oxidation in C_2 formation might change. However, the precision and stability of Pearson's r values are strongly influenced by sample size when the variations appear small or when sample sizes differ among groups. Therefore, the differences in Pearson's r values do not necessarily imply statistically significant changes, especially when they are very close (IT1–IT4). To assess the statistical significance of these differences, we compared correlation coefficients between groups using the method described in Text S1. As shown in Table S8–S9, significant differences in the C_2-O_x correlation were observed only between IT0 and the other pollution levels. For the C_2-ALWC correlation, a significant difference was found only between IT0 and IT4. Given that IT1–IT4 represents a continuous evolution of atmospheric conditions, rather than discrete and independent regimes, large differences in correlation coefficients among these categories are not expected. Although the correlation between C_2 and sulfate was strong, it did not show the similar trends as that between C_2 and ALWC. In contrast, the correlations between C_2 and primary anthropogenic markers remained generally weak across all pollution categories (Table S7), indicating that changes in anthropogenic emissions exert only limited influence on C_2 variations. Therefore, the significant and opposite changes in correlations of C_2 with O_x and ALWC between high pollution level (IT0) and low pollution level (IT4) suggested a shift in the dominant C_2 formation pathway from aqueous-phase oxidation to gas-phase photochemical oxidation under lower pollution conditions.

Line 330-334

The Pearson correlation coefficients reflect the strength of linear associations but do not directly represent the quantitative contribution of each factor. Thus, the lower Pearson's r values between C_2 and ALWC than those between C_2 and O_x under IT1-IT4 do not necessarily imply a smaller contribution of aqueous-phase pathways compared with gas-phase pathways. Here, correlation analysis is used primarily to identify potential drivers and their changing patterns, while a machine learning model is further applied to quantify their contributions to C_2 variations.

Comment 9: *The dataset collected during 2007-2018 is valuable and informative. My concern is the uncertainty caused by long-term storage. How long after sampling were these samples analyzed? How much of the C2 organic acid could change during storage?*

Response: Thank you for raising this important concern. We fully agree that long-term storage may introduce uncertainties in the quantification of molecular markers. In this study, all filter samples were immediately wrapped in aluminum foil and stored at -20 °C after collection. The samples were typically analyzed within several months after sampling.

Previous studies have shown that low temperature storage can largely preserve the chemical composition of ambient organic aerosols. For example, Resch et al. (2023) reported that ambient

aerosol samples stored at -20 °C for more than one month largely retained their molecular profiles. High-intensity peaks, such as carboxylic acids with molecular-weight (MW) = 172, 184, 186, and 200, showed changes in peak area within ±25%, indicating good stability under low temperature storage. In addition, this study demonstrates that low MW carboxylic acids are more stable than high MW carboxylic acids during storage. Thus, as a typical low MW carboxylic acid, C_2 is expected to remain largely stable during storage.

Although we cannot quantify decomposition of dicarboxylic acid during storage in this study, the storage duration was generally consistent across each year, minimizing potential inter-annual biases.

Line 113-121

$PM_{2.5}$ samples were collected using prebaked (450℃, 4h) quartz filters (8in. × 10in., QMA, Whatman, UK). Each sample lasted for 24h using a high-volume air sampler (HVPM2.5, Tisch Environmental Inc., USA) at an airflow rate of 1.1 m^3 min^{-1}. Field blank samples were also collected by mounting the blank filter onto the sampler for 10 min without turning on the sampler. In this study, a total of 462 $PM_{2.5}$ samples were collected mostly during the wintertime (October, November, and December) of each year from 2007 to 2018. The detailed information about sampling can be found in Table S2. After the collection, each filter was wrapped in an aluminum foil, zipped in Teflon bags, and stored in a freezer (-20°C) prior to analysis. Resch et al. (2023) reported that ambient aerosol samples stored at -20 °C for more than one month largely retained their molecular profiles. In addition, this study demonstrates that low MW carboxylic acids are more stable than high MW carboxylic acids during storage. Thus, the aliphatic DCA measured in this study are expected to remain largely stable during storage.

Comment 10: *In addition, in lines 115-116, I may suggest adding a table in the supplementary to detail the sample information.*

Response: Thanks for your suggestion. We have added $PM_{2.5}$ samples information in the Supplement.

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

Year	Duration	Number of samples
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

Specific comments:

Comment 1: *Please specify the data source of solar radiation in the method section.*

Response: We apologize for not including this information. We apologize for not specifying the data source in the original manuscript. The solar radiation data used in this study were obtained from the ERA5 reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) via the Copernicus Climate Data Store (CDS, https://cds.climate.copernicus.eu/datasets/).

Line 124-127

The surface net solar radiation (SR) and boundary layer height (BLH) data used in this study were obtained from the ERA5 reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) via the Copernicus Climate Data Store (CDS, https://cds.climate.copernicus.eu/datasets/). The concentrations of $PM_{2.5}$ and its main components, as well as ALWC and pH, can be found in our previous study (He et al., 2025).

Comment 2: *In Figure 2, 3, or other similar figures, modify the name of the y-axis to the corresponding species. It would be easier for readers.*

Response: Thanks for suggestion.

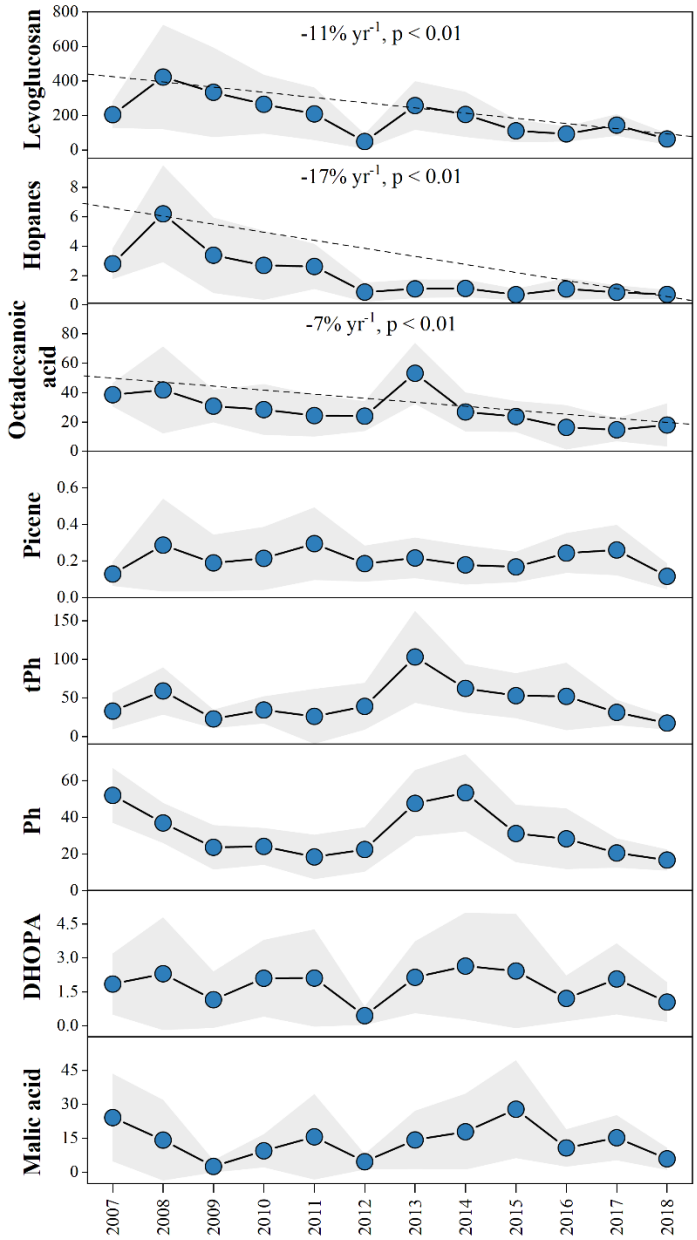

Figure 2. Annual variations in different molecular markers (ng m^{-3}) in the PRD during 2007 to 2018. The shaded area represents the 95% prediction band.

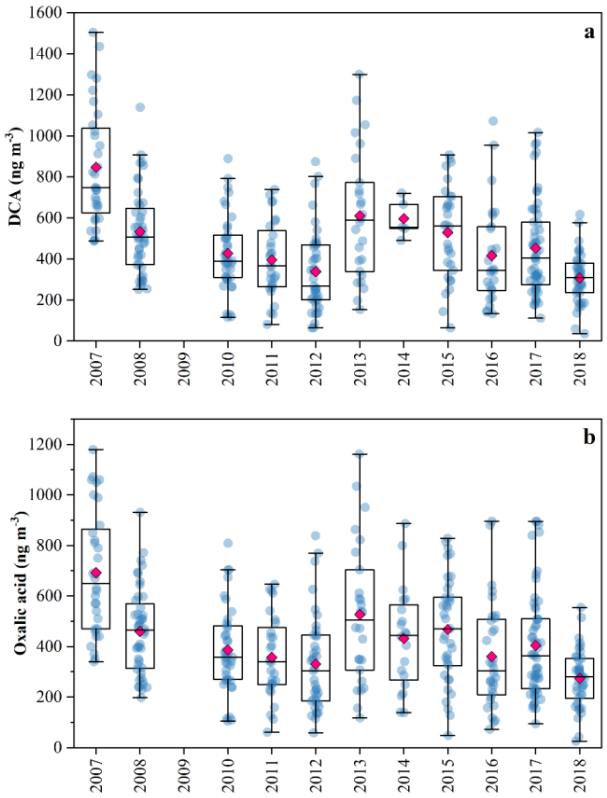

Figure 3. (a) Annual variations in aliphatic DCA. (b) Annual variations in oxalic acid. Due to the absence of aliphatic DCA and oxalic acid measurements in 2009, their concentrations for that year are not presented.

Comment 3: *Figure captions need to be revised. For example, "The concentrations decreased from 864 ± 283 ng m-3 (2007) to 307 ± 122 ng m-3 (2018), ….", "Pearson's r values between C2 and ALWC decreased from 0.43 to 0.15, while those between C2 and Ox increased from 0.28 to 0.68." or similar statements should not be described in the figure caption.*

Response: Thanks for suggestion. We have revised that.

Comment 4: *Line 225 and Table 1: Change "IT4 (25 ug/m^3 >PM2.5)" to "$PM_{2.5}$ < 25 ug/m^3".*

Response: Thanks for suggestion. We have moved Table 1 to Supplement (Table S7) and changing corresponding statements in manuscript.

Comment 5: *Lines 244-245: Please show the data or other evidence on the higher temperature, solar radiation, or humidity in PRD.*

Response: Thanks for your comments. The PRD region is located in southern China and is characterized by a subtropical monsoon climate, featuring warm temperatures, high humidity, and abundant solar radiation throughout the year. During wintertime, the temperatures were above 20 °C,

while the RH were close to 60%. Meanwhile, solar radiation ranged from 95.3 ± 49.1 to 161.3 ± 41.3 W m^{-2} (Table S5). It is necessary to include relevant evidences to support our statement. We have added them in revised manuscript.

Line 260-261

This was likely related to elevated temperature (above 20 °C), solar radiation (95.3–161.3 W m^{-2}), and relative humidity (~60%) in the PRD (Table S5), which led to a higher degree of aerosol aging.

Comment 6: *Lines 330-332: Please show evidence on the statement that lower ALWC favors the C2 compounds from the particle-phase to the gas-phase.*

Response: Thank you for this comment. We apologize for inappropriate wording in the original manuscript. According to Hu et al. (2022), more than 90% of the gas–particle partitioning of glyoxal (Gly) and methylglyoxal (mGly) proceeds through an irreversible pathway. Moreover, this irreversible uptake was found to be positively dependent on relative humidity (RH) and the abundance of secondary inorganic aerosols (SNA; sulfate, nitrate, and ammonium), which are key determinants of aerosol liquid water content (ALWC). Therefore, we should state it as "lower ALWC levels would suppress the partitioning of semi-volatile C_2 precursors (e.g., Gly and mGly) from the gas-phase into the particle-phase". The statement has been revised accordingly, and the relevant reference has been added to the manuscript.

Line 394-399

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 would suppress the partitioning of semi-volatile C_2 precursors (e.g., Gly and mGly) from the gas-phase into the particle-phase (Hu et al., 2022).

References:

Bian, Y. H., Huang, Z. J., Ou, J. M., Zhong, Z. M., Xu, Y. Q., Zhang, Z. W., Xiao, X., Ye, X., Wu, Y. Q., Yin, X. H., Li, C., Chen, L. F., Shao, M., and Zheng, J. Y.: Evolution of anthropogenic air pollutant emissions in Guangdong Province, China, from 2006 to 2015, Atmos. Chem. Phys., 19, 11701-11719, https://doi.org/10.5194/acp-19-11701-2019, 2019.

Fisher, R. A. S.: On the "Probable Erro" of a Coefficient of Correlation Deduced from a Small Sample, Metron, https://digital.library.adelaide.edu.au/server/api/core/bitstreams/d40cfd09-0ddc-490a-89c1-437bf6ae26ef/content, 1921.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181-3210, https://doi.org/10.5194/acp-6-3181-2006, 2006.

Guenther, A. B., Zimmerman, P. R., Harley, P. C., Monson, R. K., and Fall, R.: Isoprene and monoterpene emission rate variability: Model evaluations and sensitivity analyses, Journal of Geophysical Research-Atmospheres, 98, 12609-12617, https://doi.org/10.1029/93JD00527, 1993.

He, Y., Ding, X., He, Q., Zhang, Y., Chen, D., Zhang, T., Yang, K., Wang, J., Cheng, Q., Jiang, H., Wang, Z., Liu, P., Wang, X., and Boy, M.: Long-term Trends in PM2.5 Chemical Composition and Its Impact on Aerosol Properties: Field Observations from 2007 to 2020 in Pearl River Delta, South China, Atmos. Chem. Phys., 25, 13729-13745, https://doi.org/10.5194/acp-25-13729-2025, 2025.

Hu, J., Chen, Z., Qin, X., and Dong, P.: Reversible and irreversible gas–particle partitioning of dicarbonyl compounds observed in the real atmosphere, Atmos. Chem. Phys., 22, 6971-6987, https://doi.org/10.5194/acp-22-6971-2022, 2022.

Kawamura, K. and Bikkina, S.: A review of dicarboxylic acids and related compounds in atmospheric aerosols: Molecular distributions, sources and transformation, Atmos. Res., 170, 140-160, https://doi.org/10.1016/j.atmosres.2015.11.018, 2016.

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.

Resch, J., Wolfer, K., Barth, A., and Kalberer, M.: Effects of storage conditions on the molecular-level composition of organic aerosol particles, Atmos. Chem. Phys., 23, 9161-9171, https://doi.org/10.5194/acp-23-9161-2023, 2023.

Wang, N., Xu, J. W., Pei, C. L., Tang, R., Zhou, D. R., Chen, Y. N., Li, M., Deng, X. J., Deng, T., Huang, X., and Ding, A. J.: 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, Environ. Sci. Technol., 55, 5721-5730, https://doi.org/10.1021/acs.est.0c08383, 2021.

Yan, F. H., Chen, W. H., Jia, S. G., Zhong, B. Q., Yang, L. M., Mao, J. Y., Chang, M., Shao, M., Yuan, B., Situ, S., Wang, X. M., Chen, D. H., and Wang, X. M.: Stabilization for the secondary species contribution to PM2.5 in the Pearl River Delta (PRD) over the past decade, China: A meta-analysis, Atmos. Environ., 242, https://doi.org/10.1016/j.atmosenv.2020.117817, 2020.