1 **Investigating Influencing Factors of Gas-Particle Distribution of Oxygenated Organic**

2 **Molecules in Urban Atmosphere and Its Deviation from Equilibrium Partitioning Using**

3 **Random Forest Model**

4 Xinyu Wang[1], Nan Chen[2,3], Bo Zhu[2,3], Huan Yu[1,2] *

5 [1] School of Environmental Studies, China University of Geosciences, Wuhan, 430074, China

6 [2] Research Centre for Complex Air Pollution of Hubei Province, Wuhan 430074, China

7 [3] Hubei Ecological Environment Monitoring Center Station, Wuhan 430070, China

8 * Correspondence to: yuhuan@cug.edu.cn

9 **Abstract**

10   Gas-to-particle partitioning governs the fate of Oxygenated Organic Molecules (OOMs) and

11 the formation of organic aerosols. We employed a Chemical Ionization Mass Spectrometer equipped

12 with a Filter Inlet for Gases and AEROsol (FIGAERO-CIMS) to measure gas-particle distribution

13 of OOMs in a winter campaign in urban atmosphere. The observed gas to particle (G/P) ratios show

14 a narrower range than the equilibrium G/P ratios predicted from saturation mass concentration C^*

15 and organic aerosol content. The difference between observed and equilibrium G/P ratios could be

16 up to 10 orders of magnitude, depending on C^* parameterization selection. Our random forest

17 models identified relative humidity (RH), aerosol liquid water content (LWC), temperature and

18 ozone as four influential factors driving the deviations of partitioning from equilibrium state.

19 Random forest models with satisfactory performance were developed to predict the observed G/P

20 ratios. Intrinsic molecule features far outweigh meteorological and chemical composition features

21 in the model's predictions. For a given OOM species, particle chemical composition features

22 including pH, RH, LWC, organic carbon, potassium and sulfate dominate over meteorological and

23 gaseous chemical composition features in predicting the G/P ratios. We identified positive or

24 negative effects, as well as the sensitive ranges, of these influential features using SHapley Additive

25 exPlanations (SHAP) analysis and curve fitting with a generalized additive model (GAM). Our

26 models found that temperature does not emerge as a significant factor influencing the observed G/P

27 ratios, suggesting that other factors, most likely associated with particle composition, inhibit the

28 gas/particle partitioning of OOMs in response to temperature change.

1. Introduction

30 Oxygenated organic molecules (OOMs) are ubiquitous in the atmosphere. They are key

31 constituents of organic aerosols (OA) and play a critical role in particle formation and growth (Yuan

32 et al., 2024). The distribution of an OOM between gas and particle phases not only reflects its

33 volatility or water solubility, but also governs its formation pathways, atmospheric transport and

34 deposition. Therefore, understanding the phase distribution of OOMs is essential for gaining

35 insights into their volatility, transformation processes and environmental impacts in the atmosphere.

36 Gas-to-Particle (G/P) ratios of OOMs measured by laboratory (e.g., ozonolysis products from

37 Δ^3-Carene (Li et al., 2024a)) or field studies (e.g., in Hyytiälä forest, Finland (Lutz et al., 2019))

38 were sometimes used to derive saturation mass concentrations (C^*) or partitioning coefficients (K_i),

39 assuming that the observed G/P ratios represent an equilibrium partitioning state (Priestley et al.,

40 2024; Li et al., 2024a; Lutz et al., 2019; Stark et al., 2017). However, the G/P ratio of an OOM in

41 atmospheric conditions is influenced by not only intrinsic OOM physicochemical properties but

42 also external factors such as meteorological shifts (Hildebrandt et al., 2009), precursor oxidation

43 (Pankow, 1994; Seinfeld and Pankow, 2003), particle chemical composition, morphology and

44 particle-phase reactions (Jang et al., 2002; George et al., 2007). As a result, OOMs rarely achieve

45 equilibrium partitioning between the gas and particle phases (Roldin et al., 2014; Li et al., 2024b).

46 Gas/particle partitioning kinetics has been incorporated into many atmospheric aerosol models,

47 such as aerosol dynamics models (Liu et al., 2019; Zaveri et al., 2014) and kinetic multilayer models

48 (Fowler et al., 2018; Roldin et al., 2014), which accounted for molecular transfer rates, interphase

49 interactions, and environmental variability (Shiraiwa and Pöschl, 2021) in the gas-to-particle

50 transfer process. The development of these models has advanced our understanding of the

51 distribution and transport of organic compounds. However, existing theories and models often rely

52 on parameter simplifications or assumptions, and there is a lack of systematic studies examining the

53 factors influencing the phase distribution of OOMs under real atmospheric conditions. In recent

54 years, machine learning methods have been successfully applied for a variety of purposes including

compound identification (Franklin et al., 2022; Boiko et al., 2022), aerosol classification (Christopoulos et al., 2018; Bland et al., 2022), precursor apportionment (Pande et al., 2022; Wang et al., 2021) and property prediction (Gong et al., 2022; Ruiz-Jimenez et al., 2021). Machine learning has been proven to be a powerful, data-driven approach capable of uncovering complex and nonlinear relationships between variables. (Lin et al., 2022; Zhu et al., 2019). Unlike physical or chemical models, machine learning does not rely on predefined assumptions or simplifications, which enables it to unveil previously unrecognized interactions.

In this work, we employed a Chemical Ionization Mass Spectrometer equipped with a Chemical Ionization Mass Spectrometer equipped with a Filter Inlet for Gases and AEROsol (FIGAERO-CIMS) in an urban location to measure the concentrations of OOMs in both the gas and particle phases. By building data-driven machine learning models with the G/P ratio as the target variable, we explored the influencing factors of gas-particle distribution of OOMs and examined the factors that contribute to the deviations from equilibrium gas/particle partitioning. This study offered new insights and provided the foundation for future studies on the atmospheric behavior of OOMs.

2. Methodology

2.1 OOM measurement

Hourly measurements of OOMs in both gas and particle phases was conducted during a winter campaign from December 5th, 2022, to January 8th, 2023, using an iodide-based FIGAERO-CIMS (Aerodyne Research Inc., USA) at a suburban site in Wuhan, a megacity in central China (114.6157°E, 30.4577°N). The site is located in the campus of China University of Geosciences, which is surrounded by residential and agricultural mixed area. The nearest urban center and industrial area are about 25 km west to the measurement site. Nearest highways and major roads lie about 2 km north and south of the site. The site is the only provincial supersite operated by local environmental authority for monitoring air quality in Wuhan and can thus be regarded as a receptor site influenced by wide ranges of emission sources from neighboring regions. We obtained valid data of 594 hours, during which meteorological parameters (e.g., relative humidity (RH) and temperature (T)), particulate chemical components (e.g., organic carbon (OC) and sulfate ions

83 (SO_4^{2-})), and gaseous components (e.g., sulfur dioxide (SO_2) and ozone (O_3)) were routinely

84 monitored. Detailed information about those routine measurement is shown in the supplementary

85 materials (Text S1).

86   The design of FIGAERO-CIMS for hourly OOMs measurement has been described by

87 previous studies (Zhao et al., 2024; Lopez-Hilfiker et al., 2014; Lee et al., 2014). Briefly, the

88 FIGAERO operated in a measurement cycle of 1 hour alternating between gas-phase and particle-

89 phase modes. During the gas-phase mode, ambient air was drawn at a flow rate of 2 L/min directly

90 into the ion-molecule reactor (IMR), where gaseous molecules were ionized and subsequently

91 detected as adduct ions with the reagent ion I^-. Simultaneously, another flow of ambient air was

92 pulled through a $PM_{2.5}$ cyclone (URG-2000-30EN, URG Corp.) and then a PTFE filter (2 μm

93 Zefluor, 25 mm, Pall Corp.), where particles smaller than 2.5 μm were collected. During the

94 subsequent particle-phase mode, the molecules on the PTFE filter underwent thermal desorption in

95 a heated ultrahigh-purity (UHP) nitrogen flow, which kept at room temperature for 2 minutes,

96 increased to 200 °C over 15 minutes, held at 200 °C for an additional 15 minutes to ensure the

97 desorption of the majority of OOMs (Lopez-Hilfiker et al., 2014) and then cooled to room

98 temperature within 4 minutes. The desorbed molecules were directed into a turbulent flow IMR. A

99 field blank sample was collected every 24 hours. According to our earlier investigation (Wang et al.

100 2024), The OOM measured with the FIGAERO-CIMS stands for only those polar and moderate-

101 volatility organic species being desorbed below 200°C and accounted for only $26 \pm 8\%$ of the total

102 OA (OC×1.4) measured with the thermal-optical method using the IMPROVE protocol.

103 **2.2 OOMs Identification and Selection**

104   OOMs were identified using a non-target strategy. Mass calibration was performed using ions

105 such as NO_3^-, $C_2F_3O_2^-$, $IC_2H_2O_2^-$, $IC_2F_3HO_2^-$, $IC_3F_5HO_2^-$, and I_3^-, covering a mass range from 62

106 to 381 m/z. The spectra peaks were iteratively fitted with multiple peaks using a custom peak shape

107 until the residual was reduced to less than 5 % (Lee et al., 2014; Stark et al., 2015). Subsequently,

108 the exact masses of these multiple peaks were matched with the most probable elemental formulas

109 within the ranges of $C_{1-30}H_{1-60}O_{0-20}N_{0-2}S_{0-2}X_{0-2}I_{0-1}^-$, where X stands for halogen atoms, with mass

110 errors smaller than 10 ppm (mass resolution of ~6000). Isotope distribution was inspected to match

with theoretical isotope pattern. Elemental ratio and double bond equivalent (DBE) limits of the formulas were $0.3 \leq H/C \leq 3$, $N/C \leq 0.5$, $O/C \leq 3$, $S/C \leq 1$ and $0 \leq DBE \leq 20$ (Kind and Fiehn, 2007; Lee et al., 2018; Kind and Fiehn, 2006).

In order to obtain reliable concentrations and thus G/P ratios, only those OOMs with a unit mass peak area ratio of > 20 % and a sample-to-blank ratio of > 2 were included for further analysis. This filtered out the OOMs with small concentrations in the atmosphere, as well as those extremely high or low volatility OOMs that are predominantly in only one phase. Thermal desorption may cause OOM decomposition in the particle phase. According to our earlier study on the same dataset using a K-means clustering method (Wang et al., 2024), on average, 25.1% of particle-bound OOM species number and 26.8% of OOM mass detected by the FIGAERO-CIMS could be attributed to thermal decomposition fragments (Supplementary Materials Text S2). These fragments were excluded from the gas/particle partitioning analysis. The overlap of non-fragment particle-bound OOM species with those gas-phase OOM species resulted in 123 species, which were chosen as the target species for subsequent partitioning analysis. Based on our previous work (Figure S1) (Wang et al., 2024), these 123 OOM species were classified to 41 aromatic species (33.7%), 35 monoterpene-derived species (28.3%), 14 isoprene-derived species (11.4%), 11 aliphatic species (8.7%), 10 biomass burning tracers (8.1%), 3 sulfur-containing species (2.4%) and 9 other unknown species (7.3%).

2.3 Observed G/P ratios of OOMs

The concentrations of an OOM species in gas phase and particle phase are calculated as:

$$C_g = \frac{signal_g}{S \times t_g \times Q_g} \times 1000 \qquad (1)$$

$$C_p = \frac{signal_p}{S \times t_p \times Q_p} \times 1000 \qquad (2)$$

where C_g (ng m^{-3}) and C_p (ng m^{-3}) are average concentrations of a species in gas phase and particle phase, respectively, in a measurement interval (e.g., 1 hour in our campaign). $signal_g$ is the integrated signal (unit: counts) of this species during the 21-minute gas-phase measurement time (t_g) in a measurement interval. t_p is the particle sampling time (24 minutes) in a measurement

137 interval. $signal_p$ is the integrated signal of the particle-phase species during thermal desorption

138 (30 minutes) period. Q_g and Q_p are the sampling flow rates for the gas phase and particle phase,

139 respectively (Liter min^{-1}). S is the sensitivity of the species (counts per ng). The observed G/P ratio

140 $(\frac{G}{P})_{obs}$ can be calculated as:

141
$$\left(\frac{G}{P}\right)_{obs} = \frac{C_g}{C_p} = \frac{signal_g \times t_p \times Q_p}{signal_p \times t_g \times Q_g} \qquad (3)$$

142 **2.4 Comparison with equilibrium G/P ratios**

143 According to modified Raoult's Law, the saturation ratio of an organic species in gas phase (i.e.

144 $\frac{C_g}{C^*}$) equals the mass fraction of the species in organic aerosol with mass concentration C_{OA} ($i.e. \frac{C_p}{C_{OA}}$),

145 under the assumptions of equilibrium absorptive partitioning of the species over an ideal organic

146 solution and that the species has a molecular weight similar to that of the organic solution (Donahue

147 et al., 2009; Epstein et al., 2010). The equilibrium G/P ratio $(\frac{G}{P})_{eq}$ can thus be estimated from

148 saturated mass concentration C^* and mass concentration of organic aerosol C_{OA} ($C_{OA} = C_{OC} \times 1.4$)

149 using Eq. (4)

150
$$\left(\frac{G}{P}\right)_{eq} = \frac{C^*(T)}{C_{OA}} \qquad (4)$$

151 C^* at 300 K of OOMs was calculated using 4 different parameterizations reported by Mohr et

152 al. (2019) , Peräkylä et al. (2020), Ren et al. (2022) and Priestley et al. (2024). Based on the

153 saturation concentrations of HOM detected by Tröstl et al. (2016), Mohr et al. (2019) applied an

154 updated version of SIMPOL-type parameterization described by Donahue et al. (2011) to estimate

155 C^* from the numbers of carbon, oxygen, and nitrogen atoms of an organic species (n_C, n_O, and n_N),

156 but emphasizing the increased importance of OOH groups. This parameterization likely produces

157 C^* of pure compounds without considering the effect of particle matrix. Ren et al. (2022) obtained

158 C^* of OOMs via calibrated C^* $vs.$ T_{max} (thermal desorption temperature at which the maximum signal

159 intensity occurs) correlations in thermal desorption process. A similar parameterization formula

160 between C^* and n_C, n_O, and n_N was then derived using multivariate regression. Peräkylä et al.

161 (2020) parameterized the dependence of C^* on n_C, n_O, n_N and number of hydrogen atoms (n_H) by

162 comparing steady-state gas-phase concentrations of α-pinene ozonolysis products with and without

seed addition in a chamber. This parameterization predicts much smaller sensitivities of HOMs volatility to oxygen-containing functional groups than SIMPOL. The parameterization of Priestley et al. (2024) was based on measured gas and particle-phase concentrations, at an assumed equilibrium state, in residential wood-burning emissions. The C^* of the products were obtained via Eq. (4) and a parameterization was obtained between C^* and n_C, n_O, n_N and n_H. The parameterization of Priestley et al. (2024) was obtained similar to Peräkylä et al. (2020), but the gas and particle-phase concentrations of OOMs were measured in residential wood-burning emissions. The four C^* parameterizations are listed in Text S3. A temperature correction was made based on Eqs. (5) and (6) to convert $C^*(300K)$ to $C^*(T)$ at observed temperatures (Epstein et al., 2010; Li et al., 2024a):

$$C^*(T) = C^*(300K) \times \exp\left(\frac{\Delta H_{vap}}{R}\left(\frac{1}{300K} - \frac{1}{T}\right)\right) \qquad (5)$$

$$\Delta H_{vap} = -11 \times \log_{10} C^*(300K) + 129 \qquad (6)$$

where ΔH_{vap} is the enthalpy of vaporization. R is gas constant. T is the observed temperature in every hour. $C^*(T)$ was then used in Eq. (4) to estimate equilibrium G/P ratios.

2.5 Random forest model

2.5.1 Build random forest models

Complex interactions and potentially non-linear dependences exist among OOM gas-particle partitioning, atmospheric chemistry, and environmental variables. We employed random forest models to investigate the influencing factors of gas-particle partitioning.

Our selection of influencing factors (i.e. features) is based on a comprehensive literature review. We categorized 30 features into four groups: (1) 9 molecule features of the OOMs: n_C, n_O, n_N, n_H, molecular weight (Mw), double bond equivalent (DBE), hydrogen to carbon atom ratios (H/C), oxygen to carbon atom ratios (O/C) and oxidation state of carbon (OSc). (2) 7 meteorological features: RH, T, wind speed (WS), wind direction (represented by sine and cosine functions to preserve the periodicity, denoted as WD_sin and WD_cos), ultraviolet-A (UV-A), ultraviolet-B (UV-B), J_{HONO}. (3) 4 gaseous composition features: SO_2 concentration, O_3 concentration, nitrogen

189   dioxide (NO_2) concentration and ammonia (NH_3) concentration. (4) 10 particle composition features:

190   OC concentration, elemental carbon (EC) concentration, SO_4^{2-} concentration, nitrate ions (NO_3^-)

191   concentration, chloride ions (Cl^-) concentration, ammonium ions (NH_4^+) concentration, $PM_{2.5}$

192   concentration, potassium ions (K^+) concentration, as well as aerosol-phase pH and liquid water

193   content (LWC). Calculation details of pH and LWC using ISORROPIA-II model (Fountoukis and

194   Nenes, 2007) are provided in Text S4. This feature selection scheme guarantees a balanced

195   representation of pertinent factors, while preserving the simplicity and predictive efficacy of the

196   models. Unlike neural networks and other machine learning algorithms, the random forest model

197   used in this study is an ensemble model made up of multiple decision trees. During training, each

198   tree splits using a randomly chosen subset of features. Because each tree uses different feature

199   subsets, this randomness in feature selection reduces the model's reliance on any single feature,

200   making it less likely to be severely impacted by multicollinearity. To further ensure model stability,

201   we also conducted five-fold cross-validation to confirm the robustness of the model.

202   First, we developed a multi-species model involving 123 OOM species to predict the $\left(\frac{G}{P}\right)_{obs}$

203   of OOMs from molecule features, meteorological features, gas and particle composition features. A

204   total of 73062 $\left(\frac{G}{P}\right)_{obs}$ values for 123 species with hourly resolution were collected in the winter

205   campaign. Outliers can indeed exacerbate modeling errors and potentially affect the model's

206   outcomes. Therefore, they should be removed (Leong et al., 2020). Outlier removal is described in

207   Text S5. The data used for modeling were randomly divided into training data (85% of the total) for

208   model training and test data (15% of the total) for evaluating model generalization.

209   Second, we selected six typical OOMs, including more volatile ($C_5H_8O_4$, $C_6H_{10}O_4$, $C_6H_5NO_3$,

210    $C_7H_7NO_3$, C^* range: $10^{3.90} \sim 10^{6.53}$ µg m^{-3}) and less volatile species ($C_{10}H_{16}O_4$, $C_{12}H_{21}NO_9$, C^* range:

211    $10^{-4.73} \sim 10^{1.18}$ µg m^{-3}) according to the C^* parameterization of Mohr et al. (2019). $C_5H_8O_4$ (glutaric

212   acid (Lee et al., 2014; Reyes-Villegas et al., 2018)) and $C_6H_{10}O_4$ (adipic acid (Ye et al., 2021; Lee

213   et al., 2014)) are small dicarboxylic acids (C $\leqslant$ 6) typically formed through photochemical

214   degradation of reactions of alkenes, aldehydes, longer-chain acids (Kawamura and Sakaguchi, 1999)

215   or other low-oxygen organic compounds (Grosjean and Friedlander, 1980) in urban atmosphere

216   (Kawamura and Ikushima, 1993). $C_6H_5NO_3$ (Huang et al., 2019; Cai et al., 2022) and $C_7H_7NO_3$

217 (Huang et al., 2019; Cai et al., 2022) are nitrophenols either directly emitted from vehicle exhaust

218 (Tremp et al., 1993), coal and wood combustion (Huang et al., 2019), industrial processes (Harrison

219 et al., 2005) or being formed through the nitration of phenol in gas or liquid phase (Lüttke and

220 Levsen, 1997). $C_{10}H_{16}O_4$ is primarily derived the oxidation of monoterpenes (Ye et al., 2019;

221 Barreira et al., 2021). $C_{12}H_{21}NO_9$ is an organic nitrate from long-chain alkane oxidation under high-

222 NO_x conditions (Wang and Ruiz, 2018).

223   Third, single-species models were tailored to predict the gas/particle partitioning behaviors of

224 these six individual OOMs under varying meteorological and gas-particle composition conditions.

225 We also built random forest models to investigate how $\left(\frac{G}{P}\right)_{obs}$ of the six OOMs deviate from $\left(\frac{G}{P}\right)_{eq}$

226 under varying meteorological conditions and gas/particle compositions. In this study, we did not

227 build random forest model to predict absolute gas or particle phase concentrations of OOMs, due to

228 their strong dependences on diverse emission sources from neighboring regions. We lack reliable

229 features for quantifying the variable strengths of unknown sources and atmospheric aging processes

230 during transport, which are key factors influencing the OOM concentrations.

231 **2.5.2 Model optimization, evaluation and feature importance analysis**

232   To optimize and evaluate the model's performance, we applied a combination of Grid Search

233 and Cross-Validation methods. First, we employed Grid Search to tune the hyperparameters of the

234 Random Forest model. The search space included the following hyperparameters: n_estimators (the

235 number of decision trees), max_depth (the maximum depth of each tree), and max_features (the

236 number of features considered for splitting at each node) and min_samples_split (the minimum

237 number of samples required to split an internal node). For each combination of hyperparameters,

238 we used 5-fold Cross-Validation on the training set with coefficient of determination (R^2) as the

239 primary metric to assess model performance and identify the best configuration. The specific

240 hyperparameter settings used in the Grid Search are provided in the supplementary materials, in

241 Table S1.

242   After selecting the optimal hyperparameters, we further evaluated the final model using 5-fold

243 cross-validation to assess the model's generalization ability and ensure it was not overfitted. In this

evaluation, both R² and Root Mean Square Error (RMSE) were used as metrics: R^2 indicates the proportion of variance in the G/P ratio explained by the model. RMSE, on the other hand, quantifies the average prediction error and is calculated as the square root of the average squared differences between the predicted and actual values. The final model performance was determined by averaging the R^2 and RMSE values across the 5 validation sets. All model tuning and evaluation were conducted using Python (v.3.8).

To quantify the influence of each feature on the G/P ratio, we computed SHAP (SHapley Additive exPlanations) value of each feature for each sample (i.e., at each hour) using the SHAP package (v.0.40.0) in Python (v.3.8). A positive SHAP value indicates that the feature contributes positively to the G/P ratio, while a negative SHAP value means it has a negative contribution. The SHAP values versus feature values were then fitted with a generalized additive model (GAM) using the pygam package (v.0.8.0) to further identify the sensitive ranges where the changes of feature values significantly affect the SHAP values. For more details, please refer to Text S6. We utilized two-way Partial Dependence Plots (PDPs) (Chen et al., 2024a; Shi et al., 2023; Zhang et al., 2022) to analyze the joint effects of T and RH on the predicted G/P ratio. This analysis yielded a comprehensive understanding of how simultaneous changes of T and RH affect the observed G/P ratio, thereby unveiling the complex dynamics among these variables. For more details, please refer to the Text S7.

3. Results and Discussion

Despite the overall improvement in air quality in recent years, $PM_{2.5}$ episodes still occur frequently in December and January in most Chinese cities, contributing to the majority of $PM_{2.5}$ exceedance days of a year. During the winter observation period of this study, $PM_{2.5}$ concentrations ranged from 20 to 150 $\mu g\ m^{-3}$, spanning both clean and severe pollution conditions. Organic aerosol ($C_{OA} = C_{OC} \times 1.4$) comprised 10%–76% of $PM_{2.5}$, emerging as a critical bottleneck of eliminating $PM_{2.5}$ episodes. Time series of other criteria pollutants and key meteorological parameters are presented in Figure S2. The data collected during the observation period herein is considered representative of winter $PM_{2.5}$ pollution characteristics in Wuhan.

 3.1 Observed G/P ratios of OOMs and comparison with equilibrium partitioning

272

273

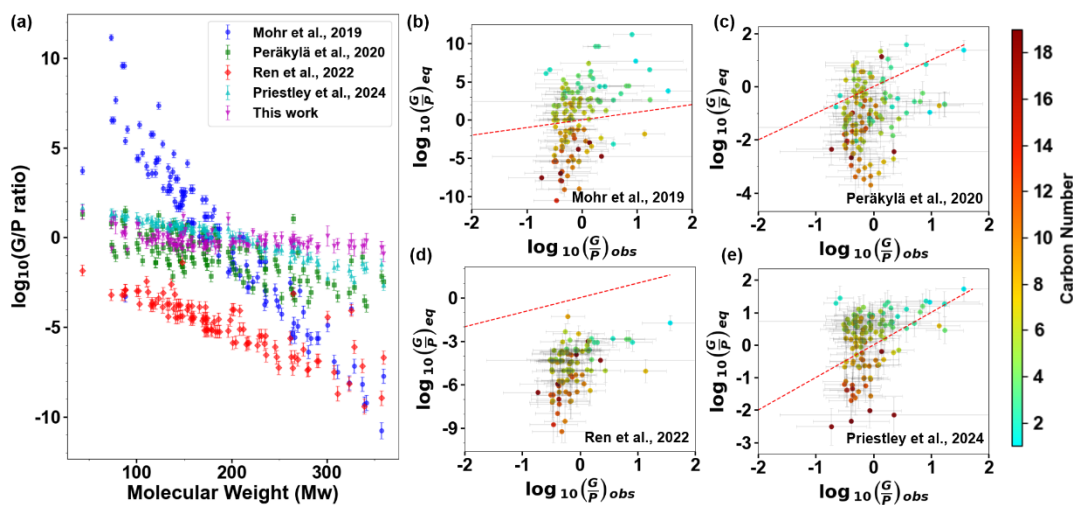



Figure 1. Comparison of $\left(\frac{G}{P}\right)_{obs}$ of 123 OOMs with corresponding $\left(\frac{G}{P}\right)_{eq}$ predicted by Eq. (4). C^* was estimated from the parameterizations of Mohr et al. (2019), Peräkylä et al. (2020), Ren et al. (2022) and Priestley et al. (2024), respectively. Error bars of $\left(\frac{G}{P}\right)_{obs}$ denote the range of G/P ratios observed under varying conditions for 594 samples (i.e. 594 hours). Error bars of $\left(\frac{G}{P}\right)_{eq}$ denote the variations with temperature and C_{OA}. Color scales in (b-e) denote carbon number of OOM species. Dashed red lines in (b-e) denote a 1:1 correspondence.

As shown in Figure 1a, although G/P ratios generally decrease with increasing molecular weight, the observed G/P ratios $\left(\frac{G}{P}\right)_{obs}$ show a narrower range ($10^{-1}\sim10^{1.5}$) than the equilibrium G/P ratios $\left(\frac{G}{P}\right)_{eq}$ predicted from Eq. (4). The differences could be up to 10 orders of magnitude, depending on C^* parameterization. Among all the methods, Mohr et al. (2019) predicts the steepest dependence of $\left(\frac{G}{P}\right)_{eq}$ on MW. Their $\left(\frac{G}{P}\right)_{eq}$ are higher than $\left(\frac{G}{P}\right)_{obs}$ for the OOMs with $n_C = 2\text{-}5$ and lower than $\left(\frac{G}{P}\right)_{obs}$ for the OOMs with $n_C > 8$ (Figure 1b). It has been recognized by Kurtén et al. (2016) and subsequent publications that SIMPOL-derived parameterizations predict a too steep dependence of C^* on Mw and oxygen content. Moreover, the parameterization of Mohr et al. (2019) likely produces C^* of pure compounds. Without considering the effect of particle matrix, it may be

289   unrealistic to predict G/P ratios using their C^* parameterization. On the basis of thermal desorption

290   temperature, Ren et al. (2022) predicts lower equilibrium G/P ratios than all other parameterizations

291   and our observation. The weakness of Ren et al. (2022) is thermal desorption may result in the

292   formation of decomposed fragments, which could be misidentified as OOM species. As a result, the

293   T_{max} of OOM formulas tends to be overestimated and the C^* tends to be underestimated in their

294   parameterization. Although Peräkylä et al. (2020) also predicted lower G/P ratios, their ratios are

295   much closer to our observation than Ren et al. (2022). Among all the predictions, the prediction

296   from Priestley et al. (2024) is most close to our observation. This is because their C^*

297   parameterization is based on the measured gas and particle-phase concentrations of OOMs in fresh

298   or aged residential wood-burning emissions. Their predicted G/P ratio is thus inherently consistent

299   with the observed G/P ratios in our study. This also highlights the risks of estimating volatility (C^*)

300   using the partitioning method, which is based on measuring equilibrium gas- and particle-phase

301   concentrations of OOMs. Two key issues arise: (1) OOMs may not achieve the assumed equilibrium

302   state in real atmospheric or chamber conditions, introducing substantial uncertainty into calculations

303   of $\left(\frac{G}{P}\right)_{eq}$; (2) The method fails for the compounds with extremely high or low volatility, as their gas-

304   or particle-phase concentrations often fall below the detection limit of mass spectrometers. These

305   limitations explain why the partitioning method typically reports a narrow volatility range (Voliotis

306   et al., 2021; Chen et al., 2024b).

307   In theory, no matter which C^* parameterization is used in Eq. (4), the temporal variation of

308    $\left(\frac{G}{P}\right)_{eq}$ for an OOM species depends solely on C_{OA} and temperature. Therefore, we are able to obtain

309   a normalized $\left(\frac{G}{P}\right)_{eq}$, which is independent of C^* parameterization, by dividing the $\left(\frac{G}{P}\right)_{eq}$ of an

310   OOM by its maximum value. Diurnal variations of normalized $\left(\frac{G}{P}\right)_{eq}$ of $C_5H_8O_4$ and $C_7H_7NO_3$ are

311   shown in Figure 2a-2b and those of other four selected OOMs are shown in Figure S3. We found

312   similar diurnal variations for all six OOMs: a peak G/P ratio appeared in the afternoon, which is

313   attributed to higher temperature. In contrast, we observed different patterns of $\left(\frac{G}{P}\right)_{obs}$ diurnal

314   variations for the six OOM species during the campaign, as shown in Figure 2c-2h. This indicates

315   that the extent of deviation of actual gas/particle partitioning from equilibrium state fluctuates

randomly over time, driven by other unknown factors. In this study, we will first examine the influencing factors of gas-particle distribution of OOMs in urban atmosphere during the winter campaign (Section 3.2), followed by an investigation into the factors contributing to the discrepancies between observed and equilibrium G/P ratios (Section 3.3).

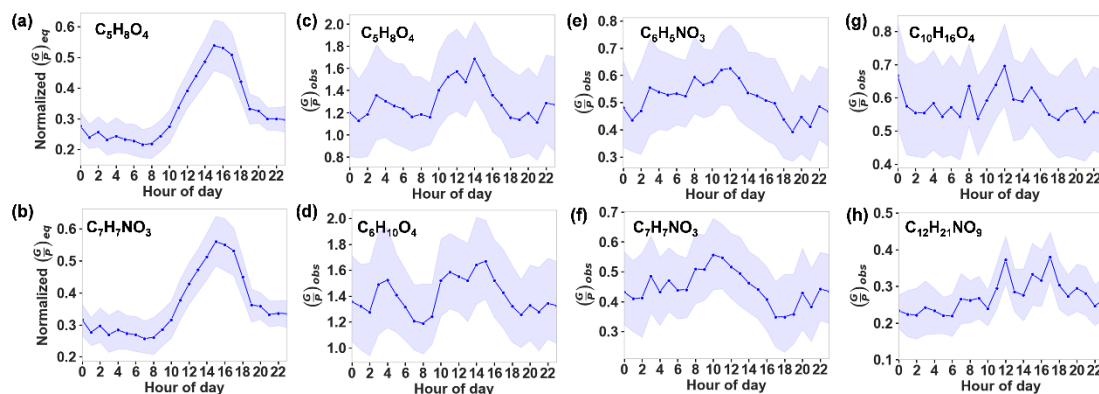



Figure 2. Diurnal variations of (a-b) Normalized equilibrium G/P ratios for the selected species ($C_5H_8O_4$ and $C_7H_7NO_3$) and (c-h) observed G/P ratios during the campaign. Solid line denotes the average value and filled area denotes the 95% confidence intervals of the mean.

3.2 Influencing Factors of the observed G/P ratios of OOMs

3.2.1 Multi-species model performance and key features

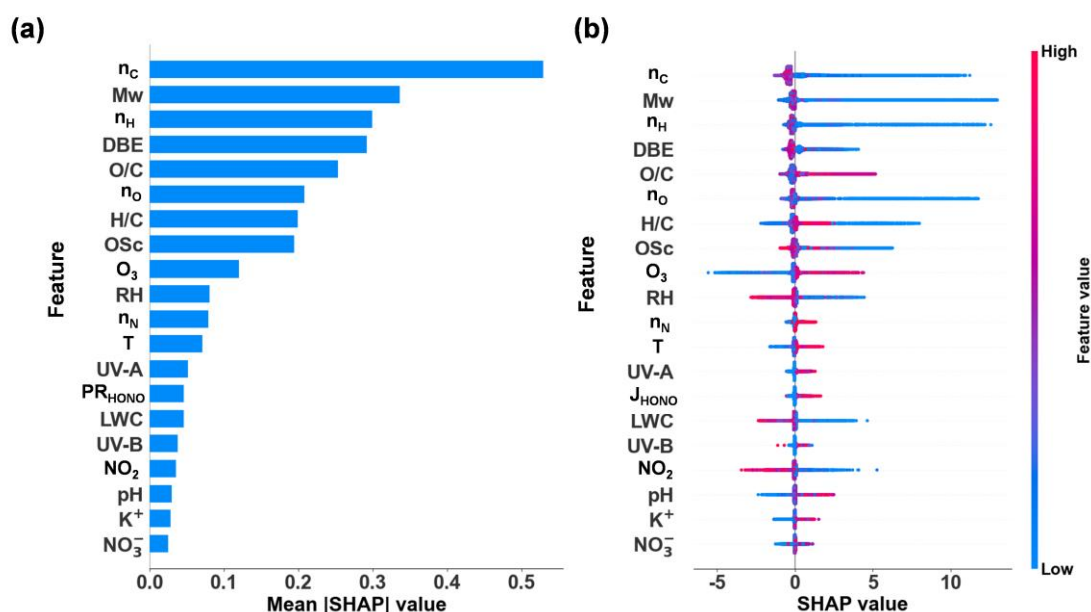



Figure 3. Multi-Species Model: (a) Feature importance based on the mean of absolute SHAP values

328 calculated for 594 samples (i.e. 594 hours) to predict the G/P ratio. (b) Distribution of SHAP values

329 in 594 samples for top 20 features.

330   The 5-fold cross-validation demonstrates that a predictive multi-species model with

331 satisfactory generalization performance was developed, achieving $R^2=0.88 \pm 0.02$ and RMSE = 1.76

332 ± 0.13 on the test set (Figure S4). Mean absolute SHAP values indicate the average importance of

333 each feature in predicting the observed G/P ratios (Figure 3a). The model highlights that intrinsic

334 molecule features, such as n_C, Mw, n_H, DBE, far outweigh meteorological and chemical composition

335 features in the model's predictions. Of the nine molecular features, eight are ranked as highly

336 important, with n_N being comparatively less influential.

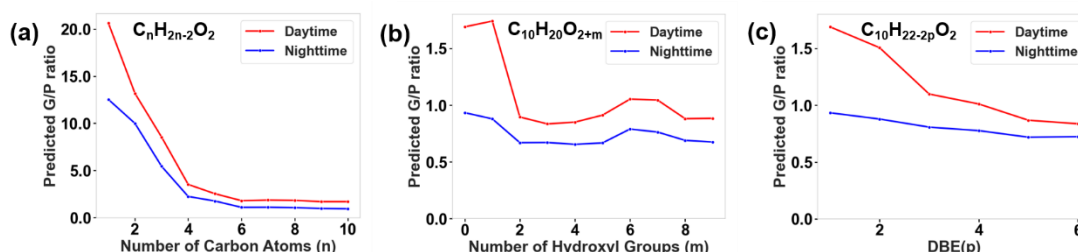
337   Figure 3b shows the SHAP value distribution for each feature. For molecule features, such as

338 n_C, Mw, n_H and n_O, high feature values are associated with negative SHAP, while low feature values

339 are associated with positive SHAP. This suggests that large molecules with high n_C, Mw, n_H and n_O,

340 and consequently lower volatility, are more likely to partition into the particle phase, thereby

341 reducing the G/P ratio.

342

343 **Figure 4.** Predicted G/P ratios using the developed multi-species model for (a) Monocarboxylic

344 acids as a function of the number of carbon atoms, (b) Modified 10-carbon monocarboxylic acids

345 as a function of the number of additional hydroxyl groups and (c) Modified 10-carbon

346 monocarboxylic acids as a function of DBE, under average daytime and nighttime environmental

347 and gas/particle composition conditions.

348   However, the molecule features related to oxidation state and unsaturation degree did not show

349 consistent effects on the observed G/P ratios. For example, OSc has a negative effect on the G/P

350 ratios, whereas O/C has a positive effect. DBE has a negative effect on the G/P ratios, whereas H/C

351 shows a mixed positive or negative effect. This is due to the fact that these features are dependent

variables as a function of n_C, n_H, n_N and n_O. To isolate the effect of oxidation and unsaturation-related features, we utilized the trained random forest model to predict G/P ratios of modified C_{10} monocarboxylic acid with varying number of hydroxyl group and DBE (Figure 4b and 4c). Other features in the model were fixed at average daytime or nighttime values observed during the campaign (see Table S2, S3). For comparison, the isolated effect of carbon atom number is also plotted (Figure 4a).

Figure 4 demonstrates that the number of carbon atoms exerts the most significant influence on the predicted G/P ratio, which decreases sharply as the carbon atom number increases from 1 to 4. Beyond this point, the ratio levels off. For modified 10-carbon monocarboxylic acids, G/P ratios are high when there is one or no hydroxyl group (Figure 4b). The G/P ratio levels off when the number of hydroxyl group exceeds 2. The G/P ratio decreases with increasing DBE value (Figure 4c). When DBE value exceeds 5, the G/P ratio change becomes minimal. In all the subplots, the G/P ratio during nighttime is consistently lower than that during daytime, which could be attributed to enhanced partitioning from gas to particles at lower nighttime temperature.

3.2.2 Identification of key features and sensitive analysis in single-species models

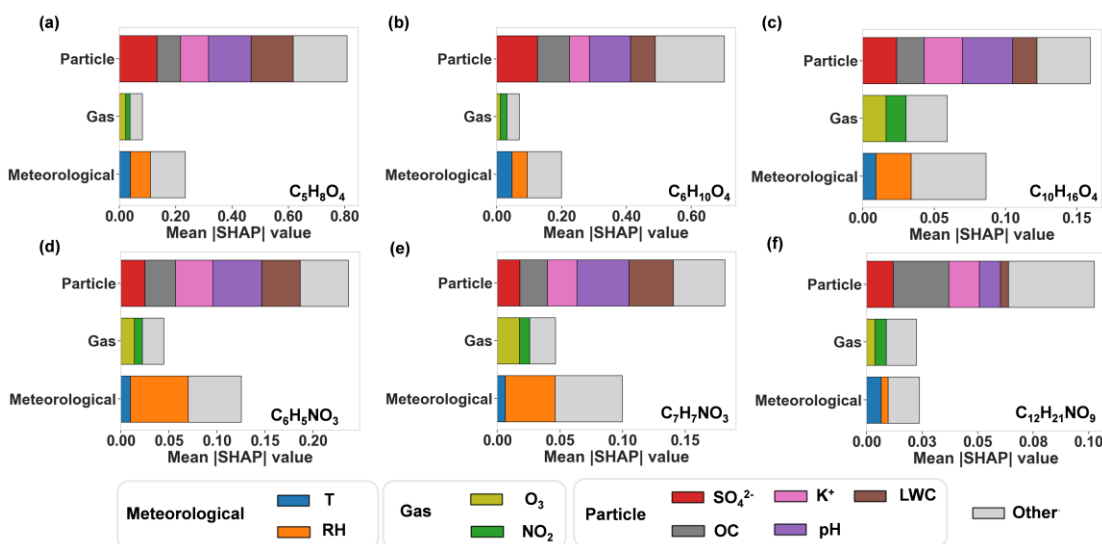



Figure 5. SHAP value analysis of three categories of features. Mean |SHAP| denotes the mean absolute SHAP values calculated for 594 samples (i.e. 594 hours): (a) glutaric acid ($C_5H_8O_4$), (b) adipic acid ($C_6H_{10}O_4$), (c) monoterpene oxidation products ($C_{10}H_{16}O_4$), (d, e) nitrophenol ($C_6H_5NO_3$

371   and $C_7H_7NO_3$), and (f) nitrated aliphatic acid ($C_{12}H_{21}NO_9$).

372   By excluding molecule features, single-species models focus on the prediction of observed

373   gas/particle partitioning behaviors of individual OOMs from meteorological and gas/particle

374   composition features. The evaluation results and optimal parameters of the six single-species models

375   on the test set are presented in Table S4. All models show acceptable generalization ability (R^2 =

376   0.51-0.88). For all six OOMs, particle composition features dominate over meteorological and

377   gaseous composition features in predicting the G/P ratios (Figure 5). Particle composition features

378   LWC, OC, K^+, SO_4^{2-} and pH, as well as RH, consistently play important roles in influencing the G/P

379   ratios of these species. This is roughly in line with the correlation analysis between the features and

380   the observed G/P ratios of the selected 6 OOMs (Figure S5), which show that pH, RH, LWC, and

381   SO_4^{2-} exhibited strong positive or negative correlations with the G/P ratios. Below, we (1) examined

382   the positive or negative effects of these features one by one (Figure 6a), and (2) identified the

383   sensitive ranges of these features by fitting SHAP values against feature values using a GAM (Figure

384   7).

385   pH is among the two most influential factors for the gas/particle partitioning of five species

386   ($C_5H_8O_4$, $C_6H_{10}O_4$, $C_6H_5NO_3$, $C_7H_7NO_3$ and $C_{10}H_{16}O_4$) with a sensitive range of 3.5–4.5 (as

387   illustrated for $C_6H_{10}O_4$ in Figure 7a). Within this range, the contribution to the G/P ratio decreases

388   by 0.5 from pH 3.5 to 4.5. Beyond pH 4.5, the G/P ratio stabilizes at -0.1. An increase in pH results

389   in a pronounced decrease of the G/P ratio. This phenomenon can be attributed to the enhanced

390   partitioning of OOMs with acidic functional groups from gas to particles with elevated pH (Su et

391   al., 2020).

392   RH has a positive effect, ranking among the top 5 significant features, on the G/P ratios of three

393   OOMs $C_6H_5NO_3$, $C_7H_7NO_3$, and $C_{10}H_{16}O_4$ (Figure 6a). SHAP value is sensitive to RH across the

394   full RH range (20%-80%, illustrated by an example $C_6H_5NO_3$ in Figure 7b). LWC also has a

395   significant positive effect for $C_5H_8O_4$, $C_6H_{10}O_4$, $C_6H_5NO_3$, and $C_7H_7NO_3$. For example, in the case

396   of $C_5H_8O_4$, a sharp increase of 0.35 in the G/P ratio is observed within the LWC range below 20 µg

397   m^{-3}. Above 20 µg m^{-3}, the contribution to the G/P ratio stabilizes at 0.15 (Figure 7c). The underlying

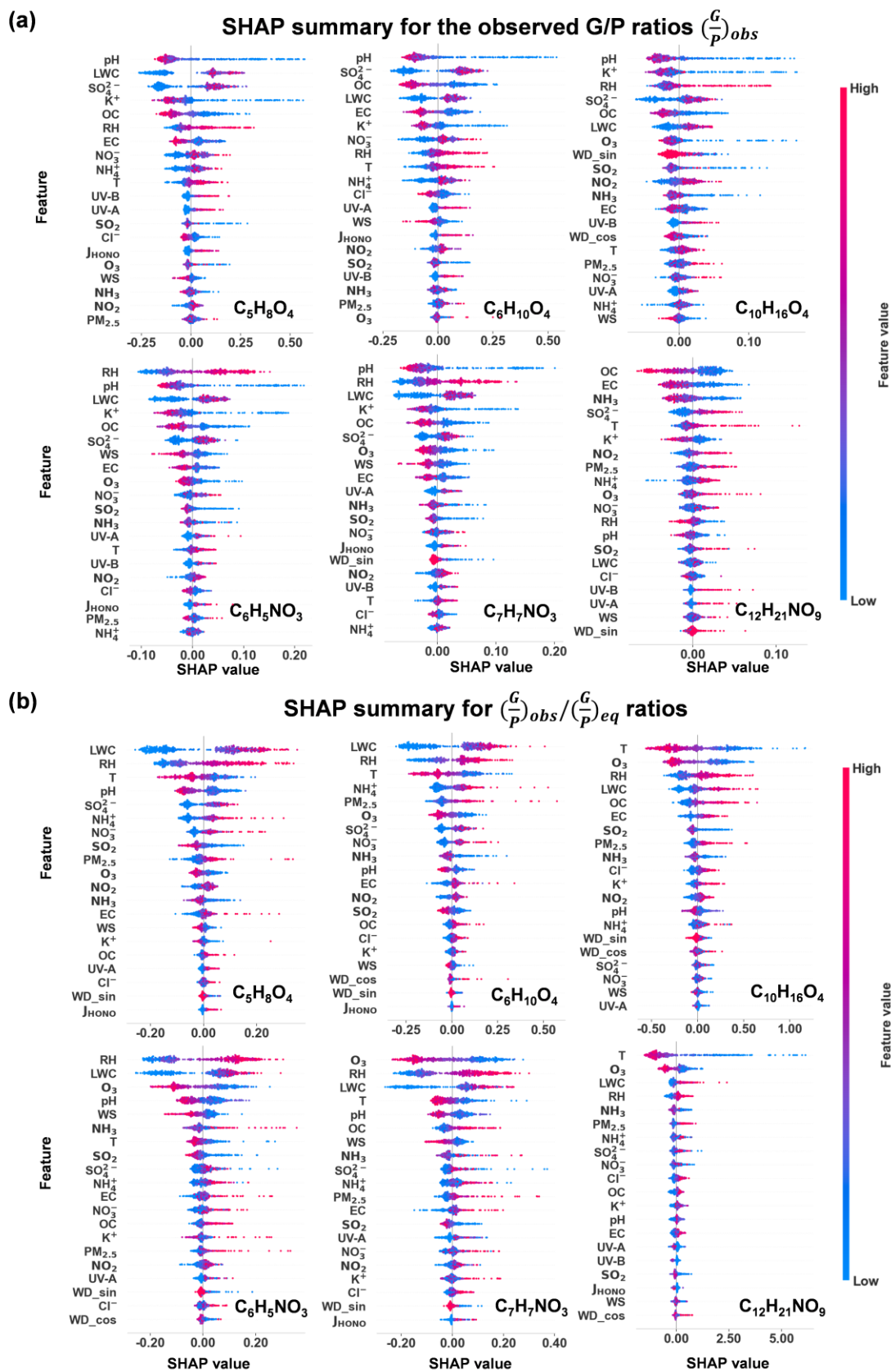
398   mechanism of this behavior is unclear. One explanation is that the low RH and LWC in particles

399   may facilitate the reversible formation of oligomers (Shen et al. (2018) and suppress their hydrolysis

400   (Liu et al., 2012), thereby increasing the concentration of these OOMs in particle phase. It is also

401   possible that the thermal desorption and subsequent detection of particle-bound OOMs were

402   inhibited in aerosol particles with more moisture.

403

404 **Figure 6**. SHAP summary plots for feature importance based on the random forest model for glutaric

405   acid ($C_5H_8O_4$), adipic acid ($C_6H_{10}O_4$), monoterpene oxidation product ($C_{10}H_{16}O_4$), nitrophenol

406   ($C_6H_5NO_3$ and $C_7H_7NO_3$), and nitrated aliphatic acid ($C_{12}H_{21}NO_9$). Features are prioritized in

407   descending order based on their importance. (a) SHAP summary for the observed G/P ratios $(\frac{G}{P})_{obs}$.

408   (b) SHAP summary for $(\frac{G}{P})_{obs}/(\frac{G}{P})_{eq}$ ratios.

409   OC has a significant negative impact (i.e., rank among the top 5) on the G/P ratios of all six

410   species, being consistent with Eq. (4), where the equilibrium G/P ratios are inversely proportional

411   to C_{OA}. Taking $C_{12}H_{21}NO_9$ as example (Figure 7d), the SHAP values decrease monotonically with

412   C_{OA} by 0.08 in the entire C_{OA} range (5-25 µg m^{-3}). For this compound, EC ranks as the second most

413   influential factor, exerting a notable negative impact below 4 µg m^{-3}. A significant G/P decrease of

414   0.05 was observed in this range (Figure 7e).

415   SO_4^{2-} has a positive effect (i.e., rank among the top 5) on the G/P ratios of $C_5H_8O_4$, $C_6H_{10}O_4$,

416      $C_{10}H_{16}O_4$ and $C_{12}H_{21}NO_9$. For example, in the case of $C_6H_{10}O_4$, the G/P ratio rises rapidly by 0.30

417   with increasing SO_4^{2-} concentrations below 6 µg m^{-3} (Figure 7f). Above 6 µg m^{-3}, the contribution

418   to the G/P ratio stabilizes at 0.1. This may be partly related to the fact that SO_4^{2-} is a highly

419   hydrophilic component (Thaunay et al., 2015), which makes its effect similar to that of LWC. In

420   addition, an increase of sulfate in aerosols is often associated with enhanced acidity and a decrease

421   in pH (Zhang et al., 2007), which drives OOM from particle to gas phase as we explained above.

422   K^+ has a negative effect on the G/P ratios of $C_5H_8O_4$, $C_{10}H_{16}O_4$, $C_6H_5NO_3$ and $C_7H_7NO_3$.

423   Taking $C_{10}H_{16}O_4$ as example, the G/P ratio decreases rapidly by 0.15 with K^+ in the concentration

424   range of below 1 µg m^{-3}. Above 1 µg m^{-3}, its contribution to the G/P ratio stabilizes at -0.03 (Figure

425   7g). K^+ is considered as a tracer of biomass burning. The increase of K^+ is generally associated with

426   higher pollution levels and higher OC concentrations in the study region (Zhao et al., 2024). The

427   effect of K^+ on the G/P ratio is thus similar to that of OC.

428   In general, temperature is supposed to be an important influential factor of G/P ratio, because

429   saturation vapor pressure of OOMs increases with temperature. Temperature ranged from -1.6 °C

430   to 14.9 °C during the winter campaign. Although temperature increase tends to elevate the G/P ratios

431   as expected (Figure 6a), the models show that temperature did not rank as important feature for 5

432 out of the 6 OOM species. We evaluated the effect of temperature on G/P ratios using two-way

433 partial dependence plots (Figure S6). G/P ratio is sensitive to temperature change only for two

434 dicarboxylic acids ($C_5H_8O_4$ and $C_6H_{10}O_4$, Figure S6a-S6b) and for $C_{12}H_{21}NO_9$ in a narrow

435 temperature range of 10-13 °C (Figure S6f and Figure 7h). The G/P ratios of $C_6H_5NO_3$, $C_7H_7NO_3$

436 and $C_{10}H_{16}O_4$ are not sensitive to temperature across most of the RH range. This behavior may be

437 attributed to the aerosol coating of inorganic salts and other aerosol components that hinder the rapid

438 equilibrium partitioning of OOMs when temperature changes. In addition, the influence of

439 temperature may be obscured due to the dominant effect of particle composition features (e.g., LWC,

440 pH, OC, SO_4^{2-}, and K^+) as discussed above.

441 As shown in Figure 6a, wind speed and direction rank relatively low in terms of feature

442 importance for the six OOMs. This suggests that while wind direction and speed might influence

443 the source areas of OOMs, they have a minimal impact on the G/P ratios of OOMs.

444

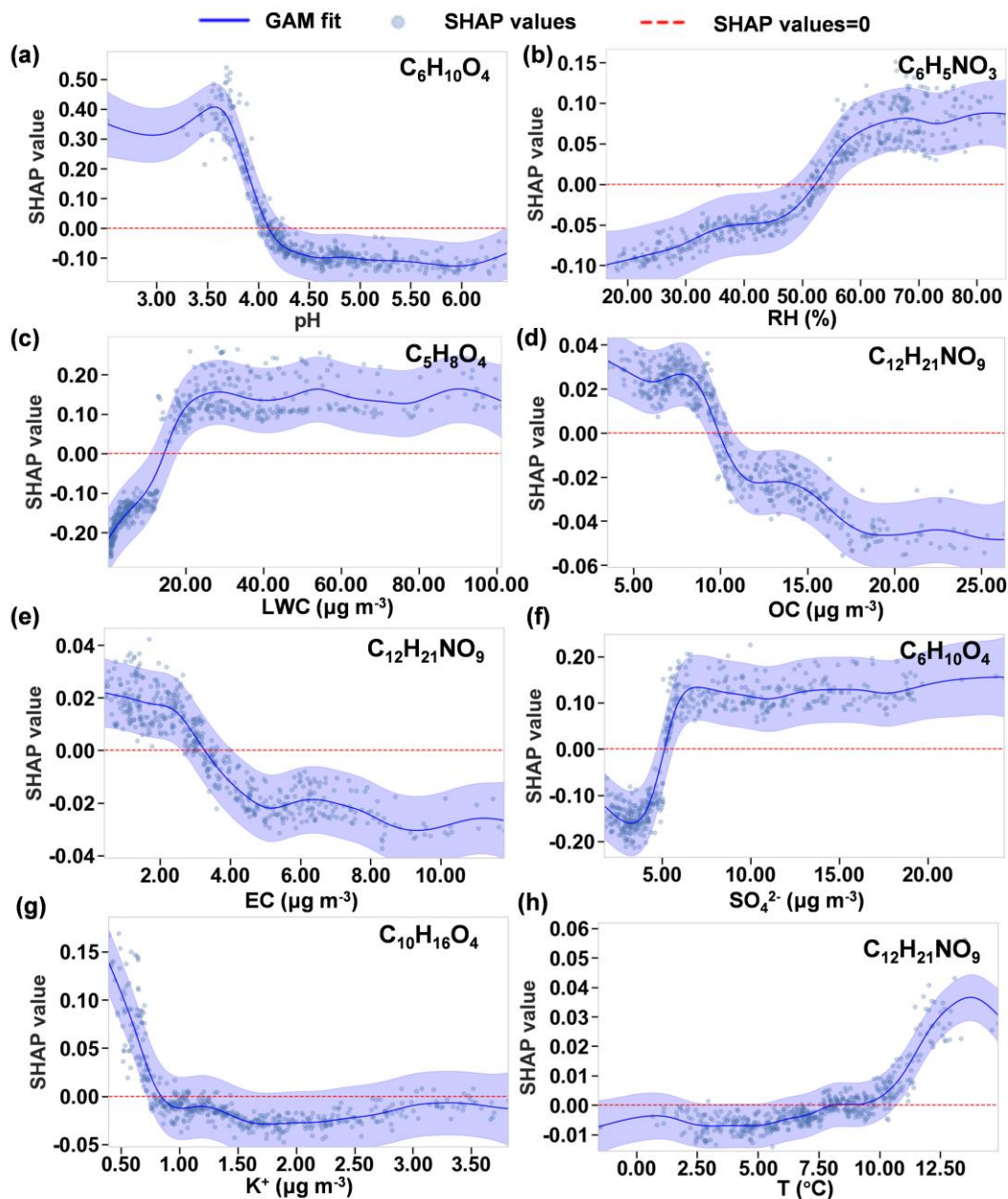



Figure 7. Curve fitting of SHAP values versus features using a GAM, illustrating the sensitive ranges where the changes of feature values significantly affect the SHAP values. Only the most affected OOM species by the eight features are shown. (a) RH for $C_6H_5NO_3$. (b) LWC for $C_5H_8O_4$. (c) pH for $C_6H_{10}O_4$. (d) Temperature for $C_{12}H_{21}NO_9$. (e) OC for $C_{12}H_{21}NO_9$. (f) EC for $C_{12}H_{21}NO_9$. (g) SO4^{2-} for $C_6H_{10}O_4$. (h) K$^+$ for $C_{10}H_{16}O_4$. Blue line denotes the GAM fit. Shaded area indicates 95% confidence interval. Dots are the SHAP values for 594 samples (i.e. 594 hours). Red dashed line denotes SHAP value of 0.

3.3 Identifying key factors driving the deviations of gas/particle partitioning from equilibrium state

To investigate the deviations of observed gas/particle partitioning from equilibrium state, we first calculate the ratios of $(\frac{G}{P})_{obs}$ over normalized $(\frac{G}{P})_{eq}$ in every hour for the selected six OOM species. Normalized $(\frac{G}{P})_{eq}$ was used here in order to offset the effect of the C^* parameterization selection. We then developed new random forest models to investigate the effects of meteorological and gas/particle composition features on the $(\frac{G}{P})_{obs}/(\frac{G}{P})_{eq}$ ratios. All the models show acceptable generalization performance ($R^2 = 0.52$-0.83) (Table S5) on the test set.

Figure 6b presents the SHAP analysis results for the $(\frac{G}{P})_{obs}/(\frac{G}{P})_{eq}$ ratios of the six OOMs. The models identify RH, LWC, O_3 and temperature as four influential factors driving the deviations from equilibrium partitioning. Positive correlations are observed between the SHAP values of $(\frac{G}{P})_{obs}/(\frac{G}{P})_{eq}$ and RH and LWC for all six compounds. This indicates that RH and LWC have stronger positive effect on $(\frac{G}{P})_{obs}$ than their effect on $(\frac{G}{P})_{eq}$, which should be negligible according to Eq. (4). Temperature is shown to be a negative factor driving the deviation from equilibrium partitioning, suggesting that temperature has a stronger influence on $(\frac{G}{P})_{eq}$ than $(\frac{G}{P})_{obs}$. This is consistent with our earlier result that $(\frac{G}{P})_{obs}$ is not sensitive to temperature. Surprisingly, O_3 is identified as an important influential factor with negative effect, particularly for the three nitrophenols and monoterpene oxidation product $C_{10}H_{16}O_4$. Since O_3 is not expected to change $(\frac{G}{P})_{eq}$, the negative impact of O_3 on $(\frac{G}{P})_{obs}/(\frac{G}{P})_{eq}$ ratio could be explained by the fact that high O_3 concentrations are likely to deplete gas-phase OOMs at a faster rate than particle-phase OOMs, thereby reducing $(\frac{G}{P})_{obs}$.

Conclusions

We measured the G/P ratios of OOM species using a FIGAERO-CIMS in urban atmosphere in a winter campaign. The observed G/P ratios show a narrower range than the equilibrium G/P ratios

476   predicted from C^* and C_{OA}. The difference between observed and equilibrium G/P ratios could be

477   up to 10 orders of magnitude, depending on C^* parameterization. Our observed G/P ratio is

478   inherently closer to the equilibrium G/P ratios predicted from the C^* parameterization by Priestley

479   et al., which was derived from measured G/P ratios in wood-burning emissions. Our random forest

480   models identified RH, LWC, O_3 and temperature as four influential factors driving the deviations of

481   gas/particle partitioning from equilibrium state.

482   Random forest models with satisfactory performance were developed to predict observed G/P

483   ratios. Intrinsic molecule features, such as n_C, Mw, n_H, DBE, far outweigh meteorological and

484   chemical composition features in the model's predictions. Large molecules with high n_C, Mw, n_H

485   and n_O, and consequently lower volatility, are more likely to partition into the particle phase, thereby

486   reducing the G/P ratio. As dependent variables, oxidation state and unsaturation do not show

487   consistently positive or negative effects on the observed G/P ratios. If other variables are fixed, the

488   model predicts that G/P ratios generally decrease with the addition of oxygen atom and DBE.

489   Particle composition features dominate over meteorological and gaseous composition features

490   in predicting the G/P ratio of a given OOM species. Among those particle features, pH, RH, LWC,

491   OC, K^+ and SO_4^{2-} consistently play important roles in influencing the G/P ratios of the six selected

492   OOM species, showing either positive or negative effect. We also identified the sensitive ranges

493   where the changes of these features significantly affect the SHAP values and provided valuable

494   insights for future research in atmospheric chemistry. It is surprising that temperature does not

495   emerge as an important factor influencing the G/P ratios for five out of the six selected OOM species.

496   Our model suggests that other factors, most likely associated with the particle composition, inhibit

497   the gas/particle partitioning of OOMs in response to temperature change.

498   At last, the random forest models developed in this study have certain limitations. (1) Aerosol

499   particle coating may serve as an inhibitory factor of gas/particle partitioning. However, the mixing

500   state and morphology of aerosol particles were not considered in the model due to the challenges in

501   quantifying these features with high resolution. (2) The OOMs with extremely high or low volatility

502   might be underrepresented in this study, because their gas- or particle-phase concentrations often

503   fall below the limit of quantification of FIGAERO-CIMS. (3) Isomers were not differentiated in the

measurement of FIGAERO-CIMS in this study. The observed G/P ratio was contributed by isomers sharing the same chemical formula. The machine learning model built in this study did not account for the effect of isomerization on gas-particle distribution of OOMs. (4) The model was based solely on the data collected during the winter season and for specific groups of OOM species present in urban atmosphere. To enhance the robustness of the gas-to-particle partitioning model, future data collection under a broader range of atmospheric conditions is recommended.

Data availability

The data used in this article are available in the public data repository Zenodo https://doi.org/10.5281/zenodo.15428774.

Author contributions

HY designed the experiment. XW, BZ and NC contributed to data collection. XW and HY analyzed the data and wrote the manuscript.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Financial support

This research was supported by the National Natural Science Foundation of China (42175131) and the Fundamental Research Funds (No. G1323523063) for the Central Universities, China University of Geosciences (Wuhan).

References

Barreira, L. M. F., Ylisirnio, A., Pullinen, I., Buchholz, A., Li, Z. J., Lipp, H., Junninen, H., Horrak, U., Noe, S. M., Krasnova, A., Krasnov, D., Kask, K., Talts, E., Niinemets, U., Ruiz-Jimenez, J., and Schobesberger, S.: The importance of sesquiterpene oxidation products for secondary organic aerosol formation in a springtime hemiboreal forest, Atmos. Chem. Phys., 21, 11781-11800, https://doi.org/10.5194/acp-21-11781-2021, 2021.

Bland, G. D., Battifarano, M., Pradas del Real, A. E., Sarret, G., and Lowry, G. V.: Distinguishing Engineered TiO2 Nanomaterials from Natural Ti Nanomaterials in Soil Using spICP-TOFMS and Machine Learning, Environ. Sci. Technol., 56, 2990-3001, https://doi.org/10.1021/acs.est.1c02950, 2022.

Boiko, D. A., Kozlov, K. S., Burykina, J. V., Ilyushenkova, V. V., and Ananikov, V. P.: Fully Automated Unconstrained Analysis of High-Resolution Mass Spectrometry Data with Machine Learning, J. Am. Chem. Soc., 144, 14590-14606, https://doi.org/10.1021/jacs.2c03631, 2022.

Cai, J., Wu, C., Wang, J., Du, W., Zheng, F., Hakala, S., Fan, X., Chu, B., Yao, L., Feng, Z., Liu, Y., Sun, Y., Zheng, J., Yan, C., Bianchi, F., Kulmala, M., Mohr, C., and Daellenbach, K. R.: Influence of organic aerosol molecular composition on particle absorptive properties in autumn Beijing, Atmos. Chem. Phys., 22, 1251-1269, https://doi.org/10.5194/acp-22-1251-2022, 2022.

Chen, Q., Shao, K., and Zhang, S.: Enhanced PM2.5 estimation across China: An AOD-independent two-stage approach incorporating improved spatiotemporal heterogeneity representations, J. Environ. Manage., 368, 122107, https://doi.org/10.1016/j.jenvman.2024.122107, 2024a.

Chen, W., Hu, W., Tao, Z., Cai, Y., Cai, M., Zhu, M., Ye, Y., Zhou, H., Jiang, H., Li, J., Song, W., Zhou, J., Huang, S., Yuan, B., Shao, M., Feng, Q., Li, Y., Isaacman-VanWertz, G., Stark, H., Day, D. A., Campuzano-Jost, P., Jimenez, J. L., and Wang, X.: Quantitative Characterization of the Volatility Distribution of Organic Aerosols in a Polluted Urban Area: Intercomparison Between Thermodenuder and Molecular Measurements, J. Geophys. Res. Atmos., 129, e2023JD040284, https://doi.org/10.1029/2023JD040284, 2024b.

Christopoulos, C. D., Garimella, S., Zawadowicz, M. A., Möhler, O., and Cziczo, D. J.: A machine learning approach to aerosol classification for single-particle mass spectrometry, Atmos. Meas. Tech., 11, 5687-5699, https://doi.org/10.5194/amt-11-5687-2018, 2018.

Donahue, N. M., Robinson, A. L., and Pandis, S. N.: Atmospheric organic particulate matter: From smoke to secondary organic aerosol, Atmospheric Environment, 43, 94-106, https://doi.org/10.1016/j.atmosenv.2008.09.055, 2009.

Donahue, N. M., Epstein, S. A., Pandis, S. N., and Robinson, A. L.: A two-dimensional volatility basis set: 1. organic-aerosol mixing thermodynamics, Atmos. Chem. Phys., 11, 3303-3318, https://doi.org/10.5194/acp-11-3303-2011, 2011.

Epstein, S. A., Riipinen, I., and Donahue, N. M.: A Semiempirical Correlation between Enthalpy of Vaporization and Saturation Concentration for Organic Aerosol, Environ. Sci. Technol., 44, 743-748, https://doi.org/10.1021/es902497z, 2010.

Fountoukis, C. and Nenes, A.: ISORROPIA II: a computationally efficient thermodynamic equilibrium model for K^+–Ca^{2+}–Mg^{2+}–NH_4^+–Na^+–SO_4^{2-}–NO_3^-–Cl^-–H_2O aerosols, Atmos. Chem. Phys., 7, 4639-4659, https://doi.org/10.5194/acp-7-4639-2007, 2007.

Fowler, K., Connolly, P. J., Topping, D. O., and O'Meara, S.: Maxwell–Stefan diffusion: a framework for predicting condensed phase diffusion and phase separation in atmospheric aerosol, Atmos. Chem. Phys., 18, 1629-1642, https://doi.org/10.5194/acp-18-1629-2018, 2018.

Franklin, E. B., Yee, L. D., Aumont, B., Weber, R. J., Grigas, P., and Goldstein, A. H.: Ch3MS-RF: a random forest model for chemical characterization and improved quantification of unidentified atmospheric organics detected by chromatography–mass spectrometry techniques, Atmos. Meas. Tech., 15, 3779-3803, https://doi.org/10.5194/amt-15-3779-2022, 2022.

George, I. J., Vlasenko, A., Slowik, J. G., Broekhuizen, K., and Abbatt, J. P. D.: Heterogeneous oxidation of saturated organic aerosols by hydroxyl radicals: uptake kinetics, condensed-phase products, and particle size change, Atmos. Chem. Phys., 7, 4187-4201, https://doi.org/10.5194/acp-7-4187-2007, 2007.

Gong, X., Wex, H., Müller, T., Henning, S., Voigtländer, J., Wiedensohler, A., and Stratmann, F.: Understanding aerosol microphysical properties from 10 years of data collected at Cabo Verde based on an unsupervised machine learning classification, Atmos. Chem. Phys., 22, 5175-5194, https://doi.org/10.5194/acp-22-5175-2022, 2022.

Grosjean, D. and Friedlander, S.: Formation of organic aerosols from cyclic olefins and diolefins, Adv. Environ. Sci. Technol.;(United States), 9, 1980.

Harrison, M. A. J., Barra, S., Borghesi, D., Vione, D., Arsene, C., and Iulian Olariu, R.: Nitrated phenols in the atmosphere: a review, Atmos. Environ., 39, 231-248, https://doi.org/10.1016/j.atmosenv.2004.09.044, 2005.

Hildebrandt, L., Donahue, N. M., and Pandis, S. N.: High formation of secondary organic aerosol from the photo-oxidation of toluene, Atmos. Chem. Phys., 9, 2973-2986, https://doi.org/10.5194/acp-9-2973-2009, 2009.

Huang, W., Saathoff, H., Shen, X., Ramisetty, R., Leisner, T., and Mohr, C.: Seasonal characteristics of organic aerosol chemical composition and volatility in Stuttgart, Germany, Atmos. Chem. Phys., 19, 11687-11700, https://doi.org/10.5194/acp-19-11687-2019, 2019.

Jang, M., Czoschke, N. M., Lee, S., and Kamens, R. M.: Heterogeneous Atmospheric Aerosol Production by Acid-Catalyzed Particle-Phase Reactions, Science, 298, 814-817, https://doi.org/10.1126/science.1075798, 2002.

Kawamura, K. and Ikushima, K.: Seasonal changes in the distribution of dicarboxylic acids in the urban atmosphere, Environ. Sci. Technol., 27, 2227-2235, 1993.

Kawamura, K. and Sakaguchi, F.: Molecular distributions of water soluble dicarboxylic acids in marine aerosols over the Pacific Ocean including tropics, J. Geophys. Res. Atmos., 104, 3501-3509, https://doi.org/10.1029/1998JD100041, 1999.

Kind, T. and Fiehn, O.: Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm, BMC Bioinformatics, 7, 234, https://doi.org/10.1186/1471-2105-7-234, 2006.

Kind, T. and Fiehn, O.: Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC Bioinformatics, 8, 105, https://doi.org/10.1186/1471-2105-8-105, 2007.

Kurtén, T., Tiusanen, K., Roldin, P., Rissanen, M., Luy, J.-N., Boy, M., Ehn, M., and Donahue, N.: α-Pinene Autoxidation Products May Not Have Extremely Low Saturation Vapor Pressures Despite High O:C Ratios, J. Phys. Chem. A, 120, 2569-2582, http://doi.org/10.1021/acs.jpca.6b02196, 2016.

Lee, B. H., Lopez-Hilfiker, F. D., Mohr, C., Kurten, T., Worsnop, D. R., and Thornton, J. A.: An Iodide-Adduct High-Resolution Time-of-Flight Chemical-Ionization Mass Spectrometer: Application to

Atmospheric Inorganic and Organic Compounds, Environ. Sci. Technol., 48, 6309-6317, https://doi.org/10.1021/es500362a, 2014.

Lee, B. H., Lopez-Hilfiker, F. D., D'Ambro, E. L., Zhou, P., Boy, M., Petäjä, T., Hao, L., Virtanen, A., and Thornton, J. A.: Semi-volatile and highly oxygenated gaseous and particulate organic compounds observed above a boreal forest canopy, Atmos. Chem. Phys., 18, 11547-11562, https://doi.org/10.5194/acp-18-11547-2018, 2018.

Leong, W. C., Kelani, R. O., and Ahmad, Z.: Prediction of air pollution index (API) using support vector machine (SVM), J. Environ. Chem. Eng., 8, 103208, https://doi.org/10.1016/j.jece.2019.103208, 2020.

Li, L., Thomsen, D., Wu, C., Priestley, M., Iversen, E. M., Tygesen Skønager, J., Luo, Y., Ehn, M., Roldin, P., Pedersen, H. B., Bilde, M., Glasius, M., and Hallquist, M.: Gas-to-Particle Partitioning of Products from Ozonolysis of Δ3-Carene and the Effect of Temperature and Relative Humidity, J. Phys. Chem. A, 128, 918-928, https://doi.org/10.1021/acs.jpca.3c07316, 2024a.

Li, Y., Cai, R., Yin, R., Li, X., Yuan, Y., An, Z., Guo, J., Stolzenburg, D., Kulmala, M., and Jiang, J.: A kinetic partitioning method for simulating the condensation mass flux of organic vapors in a wide volatility range, J. Aerosol Sci., 180, 106400, https://doi.org/10.1016/j.jaerosci.2024.106400, 2024b.

Lin, G.-Y., Chen, H.-W., Chen, B.-J., and Chen, S.-C.: A machine learning model for predicting PM2.5 and nitrate concentrations based on long-term water-soluble inorganic salts datasets at a road site station, Chemosphere, 289, 133123, https://doi.org/10.1016/j.chemosphere.2021.133123, 2022.

Liu, S., E., S. J., Chen, S., Naruki, H., A., Z. R., and and Russell, L. M.: Hydrolysis of Organonitrate Functional Groups in Aerosol Particles, Aerosol Sci. Technol., 46, 1359-1369, http://doi.org/10.1080/02786826.2012.716175, 2012.

Liu, X., Day, D. A., Krechmer, J. E., Brown, W., Peng, Z., Ziemann, P. J., and Jimenez, J. L.: Direct measurements of semi-volatile organic compound dynamics show near-unity mass accommodation coefficients for diverse aerosols, Commun. Chem., 2, 98, https://doi.org/10.1038/s42004-019-0200-x, 2019.

Lopez-Hilfiker, F. D., Mohr, C., Ehn, M., Rubach, F., Kleist, E., Wildt, J., Mentel, T. F., Lutz, A., Hallquist, M., Worsnop, D., and Thornton, J. A.: A novel method for online analysis of gas and particle composition: description and evaluation of a Filter Inlet for Gases and AEROsols (FIGAERO), Atmos. Meas. Tech., 7, 983-1001, https://doi.org/10.5194/amt-7-983-2014, 2014.

Lüttke, J. and Levsen, K.: Phase partitioning of phenol and nitrophenols in clouds, Atmos. Environ., 31, 2649-2655, https://doi.org/10.1016/S1352-2310(96)00228-2, 1997.

Lutz, A., Mohr, C., Le Breton, M., Lopez-Hilfiker, F. D., Priestley, M., Thornton, J. A., and Hallquist, M.: Gas to Particle Partitioning of Organic Acids in the Boreal Atmosphere, ACS Earth Space Chem., 3, 1279-1287, https://doi.org/10.1021/acsearthspacechem.9b00041, 2019.

Mohr, C., Thornton, J. A., Heitto, A., Lopez-Hilfiker, F. D., Lutz, A., Riipinen, I., Hong, J., Donahue, N. M., Hallquist, M., Petäjä, T., Kulmala, M., and Yli-Juuti, T.: Molecular identification of organic vapors driving atmospheric nanoparticle growth, Nat. Commun., 10, 4442, https://doi.org/10.1038/s41467-019-12473-2, 2019.

Pande, P., Shrivastava, M., Shilling, J. E., Zelenyuk, A., Zhang, Q., Chen, Q., Ng, N. L., Zhang, Y., Takeuchi, M., Nah, T., Rasool, Q. Z., Zhang, Y., Zhao, B., and Liu, Y.: Novel Application of Machine Learning Techniques for Rapid Source Apportionment of Aerosol Mass Spectrometer Datasets, ACS Earth Space Chem., 6, 932-942, https://doi.org/10.1021/acsearthspacechem.1c00344, 2022.

Pankow, J. F.: An absorption model of gas/particle partitioning of organic compounds in the atmosphere, Atmos. Environ., 28, 185-188, https://doi.org/10.1016/1352-2310(94)90093-0, 1994.

Peräkylä, O., Riva, M., Heikkinen, L., Quéléver, L., Roldin, P., and Ehn, M.: Experimental investigation into the volatilities of highly oxygenated organic molecules (HOMs), Atmos. Chem. Phys., 20, 649-669, https://doi.org/10.5194/acp-20-649-2020, 2020.

Priestley, M., Kong, X., Pei, X., Pathak, R. K., Davidsson, K., Pettersson, J. B. C., and Hallquist, M.: Volatility Measurements of Oxygenated Volatile Organics from Fresh and Aged Residential Wood Burning Emissions, ACS Earth Space Chem., 8, 159-173, https://doi.org/10.1021/acsearthspacechem.3c00066, 2024.

Ren, S., Yao, L., Wang, Y., Yang, G., Liu, Y., Li, Y., Lu, Y., Wang, L., and Wang, L.: Volatility parameterization of ambient organic aerosols at a rural site of the North China Plain, Atmos. Chem. Phys., 22, 9283-9297, https://doi.org/10.5194/acp-22-9283-2022, 2022.

Reyes-Villegas, E., Bannan, T., Le Breton, M., Mehra, A., Priestley, M., Percival, C., Coe, H., and Allan, J. D.: Online Chemical Characterization of Food-Cooking Organic Aerosols: Implications for Source Apportionment, Environ. Sci. Technol., 52, 5308-5318, https://doi.org/10.1021/acs.est.7b06278, 2018.

Roldin, P., Eriksson, A. C., Nordin, E. Z., Hermansson, E., Mogensen, D., Rusanen, A., Boy, M., Swietlicki, E., Svenningsson, B., Zelenyuk, A., and Pagels, J.: Modelling non-equilibrium secondary organic aerosol formation and evaporation with the aerosol dynamics, gas- and particle-phase chemistry kinetic multilayer model ADCHAM, Atmos. Chem. Phys., 14, 7953-7993, https://doi.org/10.5194/acp-14-7953-2014, 2014.

Ruiz-Jimenez, J., Okuljar, M., Sietiö, O. M., Demaria, G., Liangsupree, T., Zagatti, E., Aalto, J., Hartonen, K., Heinonsalo, J., Bäck, J., Petäjä, T., and Riekkola, M. L.: Determination of free amino acids, saccharides, and selected microbes in biogenic atmospheric aerosols – seasonal variations, particle size distribution, chemical and microbial relations, Atmos. Chem. Phys., 21, 8775-8790, https://doi.org/10.5194/acp-21-8775-2021, 2021.

Seinfeld, J. H. and Pankow, J. F.: Organic atmospheric particulate material, Annu. Rev. Phys. Chem., 54, 121-140, https://doi.org/10.1146/annurev.physchem.54.011002.103756, 2003.

Shen, H., Chen, Z., Li, H., Qian, X., Qin, X., and Shi, W.: Gas-Particle Partitioning of Carbonyl Compounds in the Ambient Atmosphere, Environ. Sci. Technol., 52, 10997-11006, http://doi.org/10.1021/acs.est.8b01882, 2018.

Shi, H., Yang, N., Yang, X., and Tang, H.: Clarifying Relationship between PM2.5 Concentrations and Spatiotemporal Predictors Using Multi-Way Partial Dependence Plots, Remote Sens., 15, 358, https://doi.org/10.3390/rs15020358, 2023.

Shiraiwa, M. and Pöschl, U.: Mass accommodation and gas–particle partitioning in secondary organic

aerosols: dependence on diffusivity, volatility, particle-phase reactions, and penetration depth, Atmos. Chem. Phys., 21, 1565-1580, https://doi.org/10.5194/acp-21-1565-2021, 2021.

Stark, H., Yatavelli, R. L. N., Thompson, S. L., Kimmel, J. R., Cubison, M. J., Chhabra, P. S., Canagaratna, M. R., Jayne, J. T., Worsnop, D. R., and Jimenez, J. L.: Methods to extract molecular and bulk chemical information from series of complex mass spectra with limited mass resolution, Int. J. Mass Spectrom., 389, 26-38, https://doi.org/10.1016/j.ijms.2015.08.011, 2015.

Stark, H., Yatavelli, R. L. N., Thompson, S. L., Kang, H., Krechmer, J. E., Kimmel, J. R., Palm, B. B., Hu, W. W., Hayes, P. L., Day, D. A., Campuzano-Jost, P., Canagaratna, M. R., Jayne, J. T., Worsnop, D. R., and Jimenez, J. L.: Impact of Thermal Decomposition on Thermal Desorption Instruments: Advantage of Thermogram Analysis for Quantifying Volatility Distributions of Organic Species, Environ. Sci. Technol., 51, 8491-8500, https://doi.org/10.1021/acs.est.7b00160, 2017.

Su, H., Cheng, Y., and Pöschl, U.: New Multiphase Chemical Processes Influencing Atmospheric Aerosols, Air Quality, and Climate in the Anthropocene, Acc. Chem. Res., 53, 2034-2043, https://doi.org/10.1021/acs.accounts.0c00246, 2020.

Thaunay, F., Clavaguéra, C., and Ohanessian, G.: Hydration of the sulfate dianion in cold nanodroplets: $SO_4^{2-}(H_2O)_{12}$ and $SO_4^{2-}(H_2O)_{13}$, Phys. Chem. Chem. Phys., 17, 25935-25945, https://doi.org/10.1039/C5CP02557E, 2015.

Tremp, J., Mattrel, P., Fingler, S., and Giger, W.: Phenols and nitrophenols as tropospheric pollutants: Emissions from automobile exhausts and phase transfer in the atmosphere, Water Air Soil Pollut., 68, 113-123, https://doi.org/10.1007/BF00479396, 1993.

Tröstl, J., Chuang, W. K., Gordon, H., Heinritzi, M., Yan, C., Molteni, U., Ahlm, L., Frege, C., Bianchi, F., Wagner, R., Simon, M., Lehtipalo, K., Williamson, C., Craven, J. S., Duplissy, J., Adamov, A., Almeida, J., Bernhammer, A.-K., Breitenlechner, M., Brilke, S., Dias, A., Ehrhart, S., Flagan, R. C., Franchin, A., Fuchs, C., Guida, R., Gysel, M., Hansel, A., Hoyle, C. R., Jokinen, T., Junninen, H., Kangasluoma, J., Keskinen, H., Kim, J., Krapf, M., Kürten, A., Laaksonen, A., Lawler, M., Leiminger, M., Mathot, S., Möhler, O., Nieminen, T., Onnela, A., Petäjä, T., Piel, F. M., Miettinen, P., Rissanen, M. P., Rondo, L., Sarnela, N., Schobesberger, S., Sengupta, K., Sipilä, M., Smith, J. N., Steiner, G., Tomè, A., Virtanen, A., Wagner, A. C., Weingartner, E., Wimmer, D., Winkler, P. M., Ye, P., Carslaw, K. S., Curtius, J., Dommen, J., Kirkby, J., Kulmala, M., Riipinen, I., Worsnop, D. R., Donahue, N. M., and Baltensperger, U.: The role of low-volatility organic compounds in initial particle growth in the atmosphere, Nature, 533, 527-531, http://doi.org/10.1038/nature18271, 2016.

Voliotis, A., Wang, Y., Shao, Y., Du, M., Bannan, T. J., Percival, C. J., Pandis, S. N., Alfarra, M. R., and McFiggans, G.: Exploring the composition and volatility of secondary organic aerosols in mixed anthropogenic and biogenic precursor systems, Atmos. Chem. Phys., 21, 14251-14273, https://doi.org/10.5194/acp-21-14251-2021, 2021.

Wang, D. S. and Ruiz, L. H.: Chlorine-initiated oxidation of n-alkanes under high-NOx conditions: insights into secondary organic aerosol composition and volatility using a FIGAERO-CIMS, Atmos. Chem. Phys., 18, 15535-15553, https://doi.org/10.5194/acp-18-15535-2018, 2018.

Wang, F., Yu, H., Wang, Z., Liang, W., Shi, G., Gao, J., Li, M., and Feng, Y.: Review of online source

apportionment research based on observation for ambient particulate matter, Sci. Total Environ., 762, 144095, https://doi.org/10.1016/j.scitotenv.2020.144095, 2021.

Wang, X., Zhao, Y., Hu, K., Wang, J., Wang, Q., Chen, N., Zhu, B., Zhang, H.-H., and Yu, H.: Linking Precursors and Volatility of Ambient Oxygenated Organic Aerosols Using Thermal Desorption Measurement and Machine Learning, ACS ES&T Air, https://doi.org/10.1021/acsestair.4c00076, 2024.

Ye, C. S., Yuan, B., Lin, Y., Wang, Z. L., Hu, W. W., Li, T. G., Chen, W., Wu, C. H., Wang, C. M., Huang, S., Qi, J. P., Wang, B. L., Wang, C., Song, W., Wang, X. M., Zheng, E., Krechmer, J. E., Ye, P. L., Zhang, Z. Y., Wang, X. M., Worsnop, D. R., and Shao, M.: Chemical characterization of oxygenated organic compounds in the gas phase and particle phase using iodide CIMS with FIGAERO in urban air, Atmos. Chem. Phys., 21, 8455-8478, https://doi.org/10.5194/acp-21-8455-2021, 2021.

Ye, Q., Wang, M. Y., Hofbauer, V., Stolzenburg, D., Chen, D. X., Schervish, M., Vogel, A., Mauldin, R. L., Baalbaki, R., Brilke, S., Dada, L., Dias, A., Duplissy, J., El Haddad, I., Finkenzeller, H., Fischer, L., He, X. C., Kim, C., Kurten, A., Lamkaddam, H., Lee, C. P., Lehtipalo, K., Leiminger, M., Manninen, H. E., Marten, R., Mentler, B., Partoll, E., Petaja, T., Rissanen, M., Schobesberger, S., Schuchmann, S., Simon, M., Tham, Y. J., Vazquez-Pufleau, M., Wagner, A. C., Wang, Y. H., Wu, Y. S., Xiao, M., Baltensperger, U., Curtius, J., Flagan, R., Kirkby, J., Kulmala, M., Volkamer, R., Winkler, P. M., Worsnop, D., and Donahue, N. M.: Molecular Composition and Volatility of Nucleated Particles from alpha-Pinene Oxidation between -50 °C and +25 °C, Environ. Sci. Technol., 53, 12357-12365, https://doi.org/10.1021/acs.est.9b03265, 2019.

Yuan, Y., Chen, X., Cai, R., Li, X., Li, Y., Yin, R., Li, D., Yan, C., Liu, Y., He, K., Kulmala, M., and Jiang, J.: Resolving Atmospheric Oxygenated Organic Molecules in Urban Beijing Using Online Ultrahigh-Resolution Chemical Ionization Mass Spectrometry, Environ. Sci. Technol., 58, 17777-17785, https://doi.org/10.1021/acs.est.4c04214, 2024.

Zaveri, R. A., Easter, R. C., Shilling, J. E., and Seinfeld, J. H.: Modeling kinetic partitioning of secondary organic aerosol and size distribution dynamics: representing effects of volatility, phase state, and particle-phase reaction, Atmos. Chem. Phys., 14, 5153-5181, https://doi.org/10.5194/acp-14-5153-2014, 2014.

Zhang, Q., Jimenez, J. L., Worsnop, D. R., and Canagaratna, M.: A Case Study of Urban Particle Acidity and Its Influence on Secondary Organic Aerosol, Environ. Sci. Technol., 41, 3213-3219, https://doi.org/10.1021/es061812j, 2007.

Zhang, Z., Xu, B., Xu, W., Wang, F., Gao, J., Li, Y., Li, M., Feng, Y., and Shi, G.: Machine learning combined with the PMF model reveal the synergistic effects of sources and meteorological factors on PM2.5 pollution, Environ. Res., 212, 113322, https://doi.org/10.1016/j.envres.2022.113322, 2022.

Zhao, Y., Hu, K., Wang, J., Wang, Q., Wang, P., Chen, N., Zhu, B., Zhang, H.-H., and Yu, H.: Quantifying New Sources of Ambient Organic Aerosol and Their Roles in Particle Growth Using Oxygenated Organic Molecule (OOM) Tracers, Environ. Sci. Technol. Lett., 11, 988-995, https://doi.org/10.1021/acs.estlett.4c00536, 2024.

Zhu, X., Wang, X., and Ok, Y. S.: The application of machine learning methods for prediction of metal

762   sorption onto biochars, J. Hazard. Mater., 378, 120727, https://doi.org/10.1016/j.jhazmat.2019.06.004,

763   2019.

764