Reply on anonymous Referee #1

We thank referee #1 for the constructive comments, which improved the manuscript especially with regard to the OH exposure estimates in the oxidation flow reactor. Original comments are written in black, our replies in blue as well as comprehensible excerpts from the text highlighting the tracked changes.

1. Can the authors explicitly comment on whether application of their aerosolomics data base is limited to PM$_{2.5}$ samples that are analyzed specifically with a Thermo Fisher Orbitrap mass spectrometer and the related Compound Discovery software? I think that this is the case, but if I am wrong, it would be useful to clarify how/where else it can be applied.

For individual substances, we see good comparability of fragmentation patterns with other instruments (e.g. QToF, Zhao et al. (2022) or linear ion trap LXQ, Yasmeen et al. 2010). Whether this comparability applies to the entire database and other mass spectrometers needs to be (and will be) investigated in future studies.

The application of the database, however, can be done with the related Compound Discoverer software, with the open source software MZmine 3 (https://mzmine.github.io/) as with every self-build program to match mass spectra. For this we provide the database as db-, msp-, and csv-files on our homepage. We added this information to the conclusion.

> 465   We would like to encourage the community to apply the database on their own samples. Therefore we provide the database as db-files, msp-files, and csv-files which allows the application of the database with Compound Discoverer, MZmine 3 or every self-build solution. As a community effort, further input to the database is desirable to improve our understanding of sources and formation of secondary organic aerosol.

2. L75. In addition to the irradiance, the external OH reactivity (OHR) also significantly influences the OH exposure (e.g. Li et al., 2015; Peng et al., 2015; Rowe et al., 2020). 50 ppb SO$_2$ was used in offline calibration experiments to constrain the OH exposure (Fig. S1). The corresponding OHR in the SO$_2$ calibration experiments is 50 ppb * 2.5e10 molec/cm3 * 9e-13 cm3/molec/s = 1 s^{-1}. However, the OHR in the SOA studies was most likely considerably larger than 1 s^{-1}. Consequently, the OH exposure was probably lower than what was suggested from the calibration data due to OH suppression. For example, in the α-pinene OFR experiment, ~83.9 μg m^{-3} SOA was generated; assuming an SOA yield of approximately 0.3 (e.g. Lambe et al., 2015) and complete consumption of the α-pinene, the initial α-pinene concentration was approximately 83.9/0.3 = 280 μg m^{-3} ≈ 51

ppb, with a corresponding OHR ~ 51*2.5e10*5.33e-11 = 68 s^{-1}. At these conditions, along with the experimental conditions provided in Section 2.1, I estimate that the corresponding OH exposure was approximately 1*10^{11} molec/cm^{-3}-sec using the OFR254 OH exposure estimation equation introduced by Peng et al. (2016) and published online here: https://sites.google.com/site/pamwiki/estimation-equations?authuser=0. This calculated OH exposure value is about 10 times lower than that obtained from their SO$_2$ calibration. This analysis should be applied to the other OFR experiments as well, and a column should be added to Table S1 with the corresponding calculated OHexp for comparison purposes. While the overall conclusions of the paper will remain unchanged, the (likely) lower photochemical age may provide higher confidence in applying the SOA tracers in the authors' aerosolomics database to ambient PM2.5 samples that may also be subject to lower aging timescales.

Thank you very much for this detailed comment. Based on your approach we calculated the external OH reactivity, the OH exposure as well as the equivalent atmospheric OH exposure. We changed the according section in the main part and in the SI.

> was ~1 ppm, decreasing to 0.8 ppm under OH conditions. ~~As an approximation of the mean~~ Based on these experimental conditions we calculated the external OH ~~exposure, we carried out a calibration experiment, measuring the decay of at different irradiances in the range between 0 and 200, adapted to Li et al. (2015). The calculated exposures and the fitted curve is shown in Fig.~~ reactivity (Eq. S1~~. The experimental settings resulted in a mean exposure of 1.06×10^{12} molecules cm^{-3} s^{-1},~~
>
> 80 ~~corresponding to approximately 11~~) and the OH exposure using the OFR exposure estimator (Peng et al., 2015, 2016). The resulting OH exposures (Table S1) correspond to approximately 0.1–6 days of equivalent atmospheric OH exposure, based on the assumption of an averaged tropospheric OH concentration of 1.09×10^6 molecules cm^{-3} (Li et al., 2018).

Table S1. Conditions during oxidation flow reactor experiments.

Precursor	254 nm lamp	Carrier gas	Source temperature	Mean mass concentration ± standard deviation		OH exposure
				Blank	Sample	
	V	ml min^{-1}	°C	μg m^{-3}		molec cm^{-3}s^{-1}
α-Pinene	2	37.5	26	0.34 ± 0.09	83.9 ± 3.8	1.1×10^{11}
α-Pinene	-	37.5	26	0.34 ± 0.09	42.5 ± 1.6	~
β-Pinene	2	16.6	35	0.27 ± 0.08	184.4 ± 10.4	1.5×10^{10}
β-Pinene	-	16.6	39	0.27 ± 0.08	61.5 ± 8.9	~
Limonene	2	93.6	27	0.06 ± 0.07	104.3 ± 12.6	9×10^{9}
Limonene	-	93.6	27	0.06 ± 0.07	55.1 ± 3.2	~
3-Carene	2	16.6	28	1.3 ± 0.4	62.5 ± 4.8	8.1×10^{10}
3-Carene	-	12.9	29	1.3 ± 0.4	90.1 ± 10.6	~
trans-Caryophyllene	2	37.5	32	0.09 ± 0.05	52.5 ± 6.7	1.5×10^{11}
trans-Caryophyllene	-	71.6	32	0.09 ± 0.05	47.3 ± 4.8	~
Toluene	2	16.6	23	0.08 ± 0.03	66.2 ± 1.8	5.7×10^{11}
o-Xylene	2	25.4	22	0.42 ± 0.14	66.0 ± 2.2	3.2×10^{11}
1,2,4-Trimethylbenzene	2	37.5	40	0.21 ± 0.08	24.2 ± 1.2	3.9×10^{11}
Naphthalene	2	93.6	25	2.9 ± 0.7	35.9 ± 5.7	4.1×10^{11}

We used the OFR exposure estimator (Peng et al., 2015, 2016) to approximate the OH exposure. Therefore, the OH reactivity (OHR_i) must be calculated (Eq. (S1)) using the mixing ratio r_i (in ppb) and the rate constant k_i from Atkinson and Arey (2003) of the respective precursor. The conversion factor 2.46×10^{10} results from the ideal gas law (1 atm and 298.15 K).

$$OHR_i = r_i \cdot 2.46 \times 10^{10} \cdot k_i \tag{S1}$$

5   The mixing ratio r_i (in ppb) is calculated with Eq. (S2) from the mean mass concentration c_i given in Table S1. The SOA yields (y_i) are estimated to be 0.3 for monoterpenes (Lambe et al., 2015), 0.6 for caryophyllene (Xavier et al., 2019), and 0.2 for aromatic precursors (Peng et al., 2022). In this approach, we assume complete consumption of the precursor. The molar volume is 24.47 L mol^{-1} (1 atm and 298.15 K).

$$r_i = \frac{c_i}{y_i} \cdot \frac{\text{molar volume}}{\text{molecular weight}_i} \tag{S2}$$

3. L228: The authors state: "Furthermore, ion source dimerization is a known phenomenon that hinders the unambiguous identification of atmospheric dimers, or leads to misinterpretation of results from direct-injection HESI." Can they provide references for this statement?

This phenomenon is actually visible in the raw data. As an example, Fig. S3 shows a high signal intensity of the dimer mass trace at the retention time of the monomer. This signal can be allocated to ion source dimerization.

4. L264: The formulas that are listed for terpenylic acid and MBTCA are incorrect: terpenylic acid should be $C_8H_{12}O_4$, MBTCA should be $C_8H_{12}O_6$.

Yes, the formulae were incorrect. However, due to changes based on the comments of the second referee no chemical formulas are used here anymore.

5. L307 – L316: I find the discussion of the "unassigned" signals vague and unsatisfactory. It seems clear from the text that 53.5% of the signal in the ambient samples does not correspond to the SOA tracers obtained from the laboratory OFR studies. However, all that is stated is that "isoprene as well as other monoterpenes to be promising candidates closing this gap". Further, Figure 7 conflates "unassigned" with "unidentified", because information about compound MW, carbon number, and O/C is shown here, though it is difficult to interpret from the figure. At the least, they should list the formulas of the "few compounds with high signal intensities [that] remain unassigned" in the text, and perhaps add a table to the supplement with the corresponding information that is shown in Figure 7, rather than interpreting the result without providing the data to support the hypothesis as was done here.

With "unidentified" we mean signals, for which we could not determine a molecular formula. "Unassigned" means that we determined a molecular formula, but the compound is not included in the database (we cannot assign a VOC precursor).

Figure 7 belongs to the results from the hierarchical cluster analysis and of these unidentified compounds only the molecular mass (not a molecular formula!) and retention time are available.

Figure S4 (= Fig. S5 in the old manuscript) shows all assigned and unassigned CHO compounds. Assigned means we have a match with the database. Unassigned means we can determine a CHO-formula, but we don't have a hit with the database. We have noticed that the identical coloring of unidentified (e.g., in Fig 7) and unassigned (in Fig. S5, old manuscript) can be confusing. For this reason, all unassigned CHO compounds in Fig. S4 (new manuscript) are white and hatched as in Fig. 5. Still we think that a discussion of the most abundant unassigned compounds is beneficial for interpretation and has been added to the text.
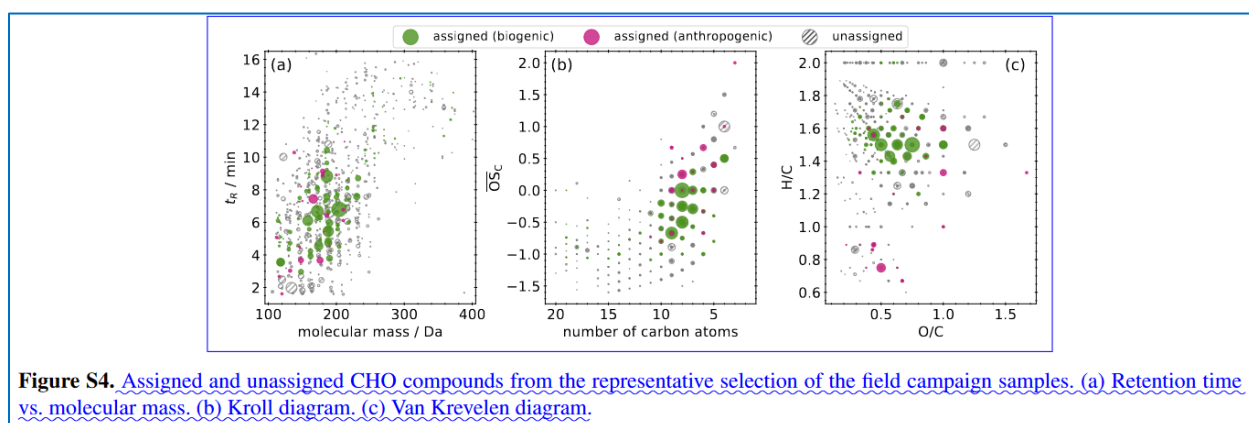

Figure S4. Assigned and unassigned CHO compounds from the representative selection of the field campaign samples. (a) Retention time vs. molecular mass. (b) Kroll diagram. (c) Van Krevelen diagram.

The majority of the assigned compounds from the representative selection of the Vienna field campaign have ~~a molecular weight~~ molecular masses smaller than 250 Da and a retention time lower than 10 minutes (Fig. ~~S5a~~S4a). The mean number

320 of carbon atoms is 9 and the mean $\overline{OS}_C$ is -0.4 (Fig. ~~S5b~~S4b). SOA originated from biogenic precursors is located in the H/C area between 1.2 and 1.8 (Fig. ~~S5c~~S4c), while compounds with H/C < 1.2 can be of aromatic character. The observation that 19 % of the CHO compounds (number-wise) in the database are responsible for nearly 50 % of the mean signal intensity demonstrates the high relevance of the investigated VOCs in SOA formation. Nevertheless, a few compounds with high signal intensities remain unassigned, such as $C_8H_{14}O_5$, $C_7H_6O_2$, $C_9H_{16}O_4$ or the tentatively isoprene derived $C_4H_6O_5$ and $C_4H_8O_4$

325 (Claeys and Maenhaut, 2021; Krechmer et al., 2015). Considering the ~~molecular weight~~ retention behaviour and the molecular mass of these unassigned compounds (Fig. S4a) we expect isoprene as well as other monoterpenes to be promising candidates closing this gap. In addition, the precursors already used should be investigated under varying chemical conditions, like further oxidants or more complex mixtures of VOCs. A comprehensive study of isoprene oxidation is planned and the outcome will be uploaded to the *aerosolomics* database in the near future.

References

Zhao, Y., Yao, M., Wang, Y., Li, Z., Wang, S., Li, C., and Xiao, H.: Acylperoxy Radicals as Key Intermediates in the Formation of Dimeric Compounds in α-Pinene Secondary Organic Aerosol, Environ. Sci. Technol., Article ASAP, https://doi.org/10.1021/acs.est.2c02090, 2022.

Yasmeen, F., Szmigielski, R., Vermeylen, R., Gómez-González, Y., Surratt, J. D., Chan, A. W. H., Seinfeld, J. H., Maenhaut, W., and Claeys, M.: Mass spectrometric characterization of isomeric terpenoic acids from the oxidation of α-pinene, β-pinene, d-limonene, and Δ3-carene in fine forest aerosol, J. Mass Spectrom., 46, 425–442, https://doi.org/https://doi.org/10.1002/jms.1911, 2011.