Reply on anonymous Referee #2

We thank referee #2 for the constructive comments that allowed us to improve the manuscript. Original comments are written in black, our replies in blue as well as comprehensible excerpts from the text including the old and the new version.

General comments:

It would be useful to say a few words about the historical origin of this approach in the introduction, as the atmospheric community may not be aware about this development.

A brief introduction to the topic of metabolomics has been included with the introduction.

> ~~Therefore, we initialized a database for compound matching,~~ Inspired by metabolomics, a tool widely used in the life
> sciences to identify metabolites, metabolic pathways, and biomarkers (Fiehn, 2002; Johnson et al., 2016), we created an *aerosolomics*
> 50 database for database-assisted identification of marker compounds (without having the need for authentic standards), and hence
> enables the investigation of atmospheric transformation pathways of VOCs under different oxidation conditions. The database
> enables compound matching based on filters from potential aerosol mass (PAM) oxidation flow reactor (OFR) experiments of
> nine biogenic and anthropogenic VOCs. We applied the database to ambient air filter samples collected in summer 2018 near

Specific comments:

Line 162: Here, a reference to the paper in which MBTCA was first reported would be appropriate: Szmigielski R., et al., 2007.

The citation was added.

> 165 that this parameterization comprises a large molecular corridor and thus leads to a wide range of $\log_{10} C^*$. A bias has been
> reported for nitrogen containing compounds (Isaacman-VanWertz and Aumont, 2021), but also for CHO compounds it appears
> to be biased. For example, $\log_{10} C^*$ of the atmospheric tracer 3-methyl-1,2,3-butanetricarboxylic acid ($C_8H_{12}O_6$, MBTCA,
> Szmigielski et al. (2007)) results in $1.97\,\mu g\,m^{-3}$, while with SIMPOL.1 (Pankow and Asher, 2008) we find $\log_{10} C^*$ at $298\,K$

Line 179: The authors mention here that they were not able to determine the individual chemical structure of the different SOA compounds, but, in my opinion, it is feasible to assign most of them taking into account available knowledge, certainly for SOA related to alfa- and beta-pinene, on the basis of MS/MS and reversed-phase LC retention data. See below.

In general, we decided to use the identification scheme of Schymanski et al. (2014). Therefore, we speak only of unambiguously identified (level 1) compounds when we compared to authentic standards. In this case, this applies only for pinic acid, MBTCA, and phthalic acid. Certainly we are able to make level 2 identifications based on available knowledge from literature (MS/MS spectra for oxidation products of other terpenes and of accretion products). These level 1 and level 2 identifications help us to validate formation mechanisms of certain compound clusters of the hierarchical cluster analysis. For example MBTCA appears in our daytime OH-chemistry cluster. In our revised manuscript we include the majority of the suggested papers on single compound identification, although the aerosolomics approach does not rely on level 1 identifications of every single compound by authentic standards.

Lines 188 - 207: In this section, the oxidation products of alfa-pinene are discussed, which have been well documented in the literature. References to the early literature, in which these SOA products were first characterized on the molecular level, are appropriate. Many of the products mentioned have even been unambiguously assigned, not only tentatively. For example, the following molecular markers related to alfa- and beta-pinene:

Please consider the comment above.

$C_9H_{14}O_4$ (MW 186): characterized as pinic acid (e.g., Yu et al., 1999; Glasius et al., 2000)

The citation was added.

> in the mass range between 140 and 210 Da, dimers are in the range between 300 and 400 Da. The major products dur-
> 195 ing ozonolysis are (pinic acid ($C_9H_{14}O_4$ at 8.79 min), (, Yu et al. (1999), level 1), terpenylic acid ($C_8H_{12}O_4$ at 6.67 min),
> (, Claeys et al. (2009), level 2), pinyl-diaterpenyl ester ($C_{17}H_{26}O_8$ at 11.28 min, Kahnt et al. (2018), Yasmeen et al. (2010)
> , level 2), $C_8H_{14}O_5$ (at 5.84 min), and $C_8H_{14}O_6$ (at 6.56 min). Oxidation by OH reduces the absolute signal intensity of

$C_{10}H_{16}O_4$ (MW 184): characterized as hydroxypinonic acid (e.g., Glasius et al., 1999)

We mention $C_{10}H_{16}O_4$ (MW 200) in the 3-carene system which has a different retention time and MS/MS spectra than hydroxypinonic acid from alpha-pinene. According to Larsen et al. (2001) we added hydroxy-3-caronic acid and 3-caronic acid ($C_{10}H_{16}O_3$).

> Panel (d) shows the results of the 3-carene oxidation experiments. Three monomers are the most prominent products in both
> systems: (caric acid ($C_9H_{14}O_4$ at 9.62 min), (, Yasmeen et al. (2011), level 2), hydroxy-3-caric acid ($C_{10}H_{16}O_4$ at 8.72 min,
> 225 Larsen et al. (2001), level 3), and (3-caronic acid ($C_{10}H_{16}O_3$ at 10.27 min, Larsen et al. (2001), level 3). The four dimers

$C_8H_{12}O_6$ (MW 204): characterized as MBTCA (Szmigielski et al., 2007)

Please consider the comment "Line 162".

$C_8H_{12}O_4$ (MW 172): characterized as terpenylic acid (Claeys et al., 2009; Yasmeen et al., 2011)

The citation was added. Please consider the comment "$C_9H_{14}O_4$".

$C_8H_{12}O_5$ (MW 188): characterized as hydroxyterpenylic acid isomers (Kahnt et al., 2014)

The citation was added.

> formed, which is in general agreement with Hammes et al. (2019). The ozonolysis shows three major products, (hydroxyterpenylic acid ($C_8H_{12}O_5$ at 5.57 min), (, Kahnt et al. (2014), level 2), ketolimononic acid ($C_9H_{14}O_4$ at 6.44 min, Yasmeen et al. (2011), level 2), and $C_{10}H_{16}O_5$ (at 6.85 min). In the OH system (at 6.44 min) ketolimononic acid becomes the major compound
> 220 whereas the intensity of (at 5.57 min) hydroxyterpenylic acid increases clearly. Analogous to the β-pinene oxidation, the $C_9H_{14}O_4$ isomer at 6.44 minutes ketolimononic acid can be used as specific limonene tracer due to the missing appearance of

$C_{17}H_{26}O_8$ (MW 358): characterized as cis-pinyl-diaterpenyl ester (Yasmeen et al., 2010; Kahnt et al., 2018)

The citations were added. Please consider the comment "$C_9H_{14}O_4$".

The same comment applies to molecular markers related to delta-3-carene and d-limonene, and to other molecular markers discussed in the text.

Designations and references have been added in some places where identification was possible, e.g., hydroxyterpenylic acid ,caric acid or ketolimononic acid.

Legend Figure 4: Molecular weight has no dimensions; delete "Da" in the x-axis.

Please consider the comment "Line 360"

Lines 279-283: Here, mention is made of CHNO biomass burning markers, such as nitrosalicylic acid and methylnitrophenol. It would be appropriate to cite the early papers in which products like methylnitrophenol were first reported in ambient fine aerosol by the group of Grgic, e.g., Kitanovski, Z., et al., J. Chrom. A 2012.

In order to limit our number of citation we decided to cite the first paper, and a more recent one.

> 290 despite proximity to an airport and a large refinery. Even though, several CHNO tracers for biomass burning could be iden-
> tified, they only play a minor role with regard to noticeably lower signal intensities: $C_7H_5NO_5$ (nitrosalicylic acid, level 3),
> $C_7H_7NO_4$ (methylnitrocatechol, level 3), $C_7H_6N_2O_5$ (methyldinitrophenol/dinitrocresol, level 2), $C_7H_6N_2O_6$, $C_6H_5NO_4$
> (nitrocatechol, level 2), $C_6H_4N_2O_5$ (dinitrophenol, level 2), $C_8H_7NO_4$ (methylnitrobenzoic acid, level 3), and $C_8H_9NO_5$
> (Ikemori et al., 2019; Mohr et al., 2013; Salvador et al., 2021)(Kitanovski et al., 2012; Salvador et al., 2021).

Line 344: The abundant $C_8H_{12}O_5$ product can be assigned to the alfa-pinene-related SOA markers, characterized as hydroxyterpenylic acid isomers. See Kahnt et al., 2014.

The citation has already been added, please consider the comment "$C_8H_{12}O_5$". At this point in the text we use the suggested compound name.

Line 360 and many places elsewhere: Molecular weight has no dimensions but molecular mass has; thus: ''The molecular mass reaches up to 440 Da and ….'' or "The molecular weight reaches up to 400 and …..".

Throughout the text, molecular mass with the unit Da has now been used uniformly.

Lines 371-374: Here, alfa-pinene-related dimers $C_{17}H_{26}O_8$ and $C_{19}H_{28}O_7$ are discussed; see the detailed chemical characterization study by Kahnt et al., 2018. It is also relevant to mention that the $C_{17}H_{26}O_8$ dimer, characterized as cis-pinyl-diaterpenyl ester, was first reported in nighttime ambient fine aerosol (Yasmeen et al., 2010).

The citation to Yasmeen et al. (2010) and their findings were added which support the correctness of the statistical approach. The citation to Kahnt et al. (2018) was added previously.

> 385 experiments described in Sect. 3.1. These 14 dimers are responsible for 72 % of the mean dimer signal intensity. Among them
> are pinyl-diaterpenyl ester $C_{17}H_{26}O_8$ (*aerosolomics*-database library: α-pinene/β-pinene + O_3) and $C_{19}H_{28}O_7$ (*aerosolomics*-
> database library: α-pinene/3-carene + O_3), both of which are described as esters between pinic acid and terpenylic acid (Gao
> et al., 2010) or pinic acid and hydroxypinonic acid (Müller et al., 2008), respectively. Furthermore, Yasmeen et al. (2010)
> reported pinyl-diaterpenyl ester in nightime ambient aerosol.

Line 395: Here, mention is made of a CHNOS compound with a mass of 295 Da. New intriguing insights about abundant MW 295 nitrooxy organosulfates related to alfa-pinene have more recently been gained by the group of Yu; see Wang, Y.-C., et al., Environ. Sci. Technol., 2021.

The citation was added.

> tion times indicates a broad mixture of polarities. Among the most common signals are ~~nitrooxy-organosulfates~~ terpene-derived nitrooxy organosulfates like $C_{10}H_{17}NO_7S$ (Wang et al., 2021), isoprene-derived CHOS like $C_2H_4O_6S$ (Claeys and Maenhaut, 2021) as well as monoterpene oxidation products like ~~(~~MBTCA, terpenylic acid, ~~level 3), (MBTCA, level 1), and (pinic acid,~~
>
> 275 ~~level 1)~~ and pinic acid.

Technical comments:

Line 47: PM2.5 (2.5 in subscript)

Has been corrected.

> 45 isotopic pattern. Furthermore, MS^2-spectra can be compared to fragmentation libraries and enable database-assisted identification of compounds (Ditto et al., 2018; Ma et al., 2022; Pereira et al., 2021; Pleil et al., 2018). However, there are currently no established databases of atmospheric SOA tracers, which can be applied on measurements of ambient $PM_{~~2.5~~2.5}$ filter samples.

Line 59 and many places elsewhere in the text: .... (98%, Alfa ....; remove the space before "%"

You are right, in English, no space is used, but the authors adhere to the AMT submission information (section "Mathematical notation and terminology"), which is based on the 9th edition of "The international System of Units (SI)" (2019). For this reason, a non-breaking space has been used.

Line 156: .... of this mode .....??

In this context a "node" is defined as a knot of a non-target analysis workflow.

Line 221: .... ß-norcaryophyllonic acid ??

Has been corrected.

In addition to the four monoterpenes, we investigated the composition of sesquiterpene-SOA from *trans*-caryophyllene

230 $(C_{15}H_{24})$. During ozonolysis we find one major and four minor products ~~with molecular weights~~ in the mass range be-

tween 198 and 302 Da (Fig. 2e). The major compound is tentatively identified as β-~~nocaryophyllonic~~ norcaryophyllonic

acid $(C_{14}H_{22}O_4$ at 11.82 min, van Eijck et al. (2013), Jaoui et al. (2003), level 3)~~(van Eijck et al., 2013; Jaoui et al., 2003)~~. In

Legend Fig. 3: .... The most abundant compounds .....

Has been changed in the captions of Fig. 2 and Fig. 3 and in the text in line 191.

190 the ozonolysis products under dark conditions, while the lower spectra show the products from OH oxidation (254 nm UV).

~~Compounds with the five highest intensities~~ The most abundant compounds are labeled with the predicted formula and their

retention time, however, the database contains these entries of all compounds down to 1 % relative peak intensity.

Figure 2. Mass spectra of the detected products from the OFR experiments of five biogenic precursors (a) α-pinene, (b) β-pinene, (c) limonene, (d) 3-carene, (e) and *trans*-caryophyllene. The intensity is normalized to the highest signal of each chemical system. The ~~five~~ most ~~intensive~~ abundant compounds of each experiment are labeled with their predicted composition and the according retention time.

Figure 3. Mass spectra of detected products from OFR experiments of four anthropogenic precursors (a) 1,2,4-trimethylbenzene, (b) toluene, (c) *o*-xylene, (d) and naphthalene. The intensity is normalized to the highest signal of each chemical system. The ~~five~~ most ~~intensive~~ abundant compounds of each experiment are labeled with their predicted composition and the according retention time.

Line 350: However, it needs .....

Has been changed.

It is also remarkable that compounds appearing in this cluster show low standardized values in sample cluster II. Low stan-

dardized values indicate lower signal intensities of the MS measurements. However, it ~~need~~ needs to be considered that signal

365 intensities are not directly quantitative to its concentration. It is well known, that the ionization efficiency of HESI varies

References

Larsen, B.R., Di Bella, D., Glasius, M. et al. Gas-Phase OH Oxidation of Monoterpenes: Gaseous and Particulate Products. Journal of Atmospheric Chemistry 38, 231–276 (2001). https://doi.org/10.1023/A:1006487530903.