Author Comments in Response to Referee Comments 1

RC1: 'Comment on egusphere-2026-116', Anna Feerick, 14 Mar 2026

Citation: https://doi.org/10.5194/egusphere-2026-116-RC1

General Comments

The preprint is of high quality. The paper addresses relevant scientific questions within the scope of AMT, and, importantly, provides a much-desired resource within the gas chromatography nontarget analysis and atmospheric communities. While the current data in the UCB-GLOBES library is most useful for derivatized compounds separated through GCxGC, the development of a platform that can be contributed to by the greater community and provides an open-source option for semi-volatile organic aerosols will greatly expand the depth to which researchers can explore their environmental sampling sets. The worth of the UCB-GLOBES libraries in retrospective analysis is demonstrated through the re-classification of sourcing for both the GoAmazon and Southeast US SOAS studies. Most importantly, for the study of organic aerosols, it provides a framework for comparing different SOA sources and informing future research directions for SOA characterization. Areas of improvement include adding additional ions to the MS spectra related to TMS derivatization to improve the likelihood of a correct library match and to provide a quantitative measure of similarity between source profiles.

We thank Anna Feerick, Referee 1, for their thorough and detailed review of this manuscript and helpful comments. We have addressed all comments in line following those of Referee 1 with author responses in italics and blue. All references to line numbers are associated with those in the originally submitted manuscript for discussion.

Specific Comments

All current MS spectra in the UCB-GLOBES library are of derivatized compounds. Adding this point to the abstract or the conclusion would improve clarity and understanding of the current use cases for other GC/EI-MS practitioners.

Thank you for this helpful suggestion. We agree that it would improve clarity and understanding. As such, Abstract line 40 has been modified to, "To advance non-targeted analyses of environmental samples, we have catalogued approximately 27,000 mass spectra (MS) of the trimethylsilyl derivatives of semi-volatile organic aerosol (OA) analytes

in the open-access University of California Berkeley Goldstein Library of Organic Biogenic Environmental Spectra (UCB-GLOBES). Analytes were observed in ambient samples from the U.S. and the Central Amazon and/or laboratory simulations of secondary OA (SOA) formation. These samples are representative…"

Line 254: Does the simple neighbor comparison find peaks that are considered local maxima in addition to those of true maxima?

Yes, the simple neighbor comparison finds several pairs of peaks that may include local maxima or true maxima, while looking for the amu difference between m/z's at which those peak maxima occur to look for the characteristic 15 amu difference indicating m/z = [M-$CH_3^{\cdot}$]$^+$ and potential molecular ion, [$M^{\cdot}$].$^+$

Section: Automated metadata entries from MS featurization: MW prediction, base peak, five highest intensity m/z ions

- Line 257: Does an exceedance of the 75th intensity percentile mean that comparisons for the candidate molecular ion had to be at least 75% of the highest intensity ion for the spectra at that point? If so, why pick 75%? Were these intensities considered after blank subtraction?

  The 75th intensity percentile was picked to eliminate the potential of erroneously assigning high m/z ions with small signals as the molecular ion peak. While there were blank subtractions performed in terms of eliminating artifactual chromatographic peaks and MS baseline subtraction was performed there can still exist some low signal chromatographic peaks with noisier mass spectra.

  I'd highly recommend including the M-15 peak in saved mass spectra, in addition to the five highest intensity ions. Since trimethylsilyl derivatives in EI-MS have a characteristic [M-15]+ ion, this ion is very likely to be associated with the mass spectra of the assumed molecular ion. Having more unique ions to the feature of interest would improve the cosine similarity scores and increase the likelihood of a correct library match or a correct rejection.

  We have not altered the mass spectra in any way even if there was a prediction of the molecular ion. All mass spectra still contain the M-15 peak when a molecular ion is predicted, and the accompanying metadata for each mass spectral entry still has a listing of the top five highest intensity ions and predicted MW (from predicted molecular ion).

Line 298: How were Retention index tolerances checked?

Manual review of mass spectral match hitlists for a few hundred query mass spectra was utilized for the basis of development of retention index tolerances. We found that for intra-UCB-GLOBES MS matches, a tight RI tolerance of 10 units or less should be maintained to prevent false positive MS matches, especially in the case of closely eluting isomers. For identified MS that had matches to the NIST MS Database, we learned that a wider RI tolerance of 30 provided accurate MS match attribution in addition to match factor scores of a certain tolerance (as described in Section 2.4 on MS Matching Criteria).

Line 455-456: The authors state that the copaiba oil spread shown in Figure 6 is similar to the sesquiterpene system in a) and the monoterpene system in c). I agree up until Nc > 20, where I believe the copaiba oil lacks the data points to confidently say it overlaps with monoterpenes. Since "similar spread" seems to be determined by a qualitative rather than a quantitative measure and can leave room for debate, I'd recommend including a quantitative metric for the degree of similarity for these different databases.

Thank you for raising this point of confusion. We have adjusted these lines to specify quantitatively the ranges in terms of Nc where there is qualitative overlap between the copaiba oil and monoterpene systems. This line now reads, "…b) copaiba oil (an essential oil comprised mainly of sesquiterpenes and some monoterpenes) has a similar spread in Avg. OS_c and N_c as the sesquiterpene systems in a) and as the monoterpene ($C_{10}H_{16}$) systems in c) for $N_c < 20$; c) α-pinene reaction with NO_3 radical…"

Technical corrections:

Table 1, Dataset Description:

- Acedox equation has a parenthesis that needs to be removed.

  Thank you, this has been corrected.

- Copox equation may need a comma after O3

  Thank you, this has been corrected.

- Apinox and Apinox2 equations need a space removed after alpha

  Thank you, this has been corrected.

- Limonox, Myrcox, and Isopox equations need compound names to be all lowercase

  Thank you, this has been corrected.

- Isopox equation may need “as OH source)” to not be a subscript

  Thank you, this has been corrected.

Table 2, Description column

- Remove periods from the end of the descriptions for Synonyms and Contributor_ID

  Thank you, this has been corrected.

- For Description of Column_Type, add a space before the -, or change to a colon

  Thank you, this has been corrected by using a colon instead of a dash.

Figure 1: A boarder should be present separating the bars of Fire Science Lab and Napa, CA Fires 2017

Thank you for catching this. We have added the border between the two bars.

Figure 3: ISOP and BBOA could use a color change to make them more distinct in black and white

Thank you for noting this. The ISOP category transparency has been altered to make it more distinct from the BBOA category.

Figure 4: The number overlap with the border on the right side is very difficult to see in black and white. I’d recommend removing the numbers on the right side of each graphic to reduce clutter.

Thank you for catching this. The number labels have been moved away from the right-side border for better visual clarity. We would like to keep them, however, to accompany the text descriptions in Section 3.2.2 that refers to these values.

Figure 5: Some of the borders between the matches and unmatched blocks look thicker than others. If possible, I’d recommend unifying the border size. Additionally, there is a light blue block at the top of the 2-methyfuran/OH bar. If this is part of the No Matches, Unknown category, I’d recommend combining the two.

Thank you for noting this. We have corrected the plotting to unify the border sizes and have addressed the 2-methylfuran/OH bar to unify the erroneous light blue block on top within the No Matches, Unknown category.

Figure 6: These figures are difficult to parse in black and white, and sub-figures a, c, and d would benefit from more color variations. To improve the clarity of sub-figure a, I'd recommend O3 and NOx be split into dark colors for one and light colors for the other. This would help visualize the similar spreads of OSc and Nc. Lightening the apin+NO3 in sub-figure c could bring a similar clarity. For sub-figure d, darkening 3-me-fur+OH would help it stand out against the FSL_FIREX data. Capitalize the last word in the title of sub-figure d "burn".

Thank you for bringing this issue to light. We have adopted the reviewer's suggested changes in colors, and in addition we have varied the makers by oxidation system within subpanels a, c, and d to further enhance readability of the figures when viewed in black and white. The last word in the title of sub-figure d, "burn" has also been revised to "Burn."

Line 91: "have been born using" feels awkward. One option for alteration is to add "from" between born and using or replace with "have been made using"

This line now reads, "…have been born from using…"

Line 106: re-evaluate the use of "/" in this sentence. Are you using "/" as "or", or is it implying that "known chemical identity" and "MS generated from authentic standards readily available" are the same thing in the previously mentioned resources?

Thank you. We have selected "or" instead of the use "/" and revised the line accordingly, "Still, most of these resources only document compounds of known chemical identity or MS generated from authentic standards readily available…"

Line 498-500: Clarify the sentence "In contrast…chemical categorization". I do not understand it. This makes it difficult to understand what the following example is trying to highlight.

We appreciate this comment and have revised the sentence and discussion to broaden its meaning for those less familiar with aerosol mass spectrometry and positive matrix factorization. We have altered this sentence to, "In contrast, statistical techniques such as positive matrix factorization (PMF) have been applied to aerosol mass spectrometry data, in which a timeline of composite mass spectra (also with 70 eV ionization) of all non-refractory flash-vaporized aerosol compounds is generated. In other words, "bulk" OA chemical composition is measured, rather than mass spectra of each individual OA analyte. These timelines of composite mass spectral data are separated into statistical factors, each with an average characteristic mass spectrum generated from the weighted signal contributions of the aerosol compounds' mass spectral fragments assigned to this factor. Thus, a PMF factor mass spectrum will never be a match to any one compound's mass spectrum. Determination of a "specific" source often relies on factor timeline

correlation with known chemical tracers measured by complementary contemporaneous techniques and/or evidence of one or a few enriched characteristic ions within the factor mass spectrum that are also enriched in the MS of an authentic standard or known chemical tracers."

Line 167: used a hyphen when writing high-resolution in the following paragraph. Either add a hyphen here or remove one from line 76 and 187

We have added a hyphen in Line 167, now reading "High-Resolution."

Line 178: remove space between "N-methyl-N-" and "trimethylsilyl"

This edit has been implemented.

Line 227: The current phrasing implies that low polarity peaks were the contaminants. Is this correct? If not, and PFMD was the "additional known contaminants" make "contaminants" in this sentence singular.

Thank you for noting this. The sentence has been revised to read, "Additional known contamination influencing MS quality included convolution of low polarity peaks coinciding with perfluoromethyldecalin (PFMD, an internal pulsed calibrant during chromatographic runs for MS calibration)."

Line 238: remove "be able to"

This edit has been implemented.

Line 329: May benefit from a paragraph break between "column." and "Using UCB-GLOBES..."

Thank you for this helpful suggestion. This has been implemented.

Line 456: Add c) after the semi-colon to connect the following statement with the correct sub-figure.

Thank you for noting this. This has been implemented as shown in response to RC1, Specific Comment Line 455-456.