Title: A searchable database and mass spectral comparison tool for aerosol mass spectrometry (AMS) and aerosol chemical speciation monitor (ACSM)

Response to Referee#1

The authors appreciate the overall comments of the Referee #1 and we would like to thank for his/her constructive comments. In the following, the Referee suggestions (in bold) and answers are addressed (plain text for response, quotes for new text added to manuscript).

Specific comments:

The impact of the following factors on cosine score needs to be discussed in detail, since the scores in tables seem to be pretty close.

In response to the reviewer's question about the impact of certain factors on the cosine score, it is important to clarify the role and purpose of the tool described in the paper. The tool presented in the paper is designed to assist AMS/ACSM users in efficiently comparing their mass spectra with reference mass spectra in the database. Since the tool treats only mass spectra among AMS data, in lines 405-407, we already stated that users should carefully review the published manuscript of the reference mass spectra listed in the results table to gain insights into the variations and possibilities of other AMS data such as time series, sample condition, etc. The detailed and accurate interpretation of the data remains the responsibility of the users. This tool intends to conveniently provide potential candidates with a high correlation to the target mass spectrum, not to identify a single, distinct mass spectrum with the highest correlation. However, we agree that the impact of the following factors on cosine score needs to be discussed in the manuscript. We have added the following sentence to start the discussion about this in the discussion section:

"However, we highlight considerations that users should keep in mind when using this tool for more accurate AMS and ACSM data interpretation."

(1) Maximum m/z inconsistency

The maximum m/z for HR AMS varied from 120 to 200, and some studies fitted the discrete PAH signals. Such differences will impact the correlation coefficients when comparing with the global database. How would the authors take such influence into consideration when applying mass spectra with different m/z ranges?

In the tool here, all calculations for comparison are performed based on the user's target mass spectrum. When a user adjusts the mass range, the tool applies the same mass range to the reference mass spectra in the database. If there is significant signal and distinctive peaks at the higher m/z range, including the higher range may help identify spectra in the database that may be similar if they also contain the higher m/z range. To clarify this, we added the following note in the manuscript:

"Initially, when users set a new mass range, users should take into account the maximum m/z value of their target mass spectrum for comparison. As the maximum m/z values (having none-zero value) of reference mass spectra in the database may vary, adjusting the mass range can impact the cosine score. In cases where there are significant signals and distinctive peaks at the higher m/z range for the target spectrum, including the higher range may help identify spectra in the database that may be similar if they also contain the higher m/z range."

(2) HR AMS mode inconsistency

The differences in MS between W mode and V mode for the same aerosol samples have been observed in previous field/laboratory campaigns, since more ions are identified in the W mode due to its high resolution. Hence, the correlation would change when the mass spectra of W mode or V mode are used as references.

We agree with the referee's comment. Since W mode analysis may result in being able to more confidently fit ions due to its higher spectral resolution than V mode, the correlation would change when the mass spectra of W mode or V mode are used as references. This is the main reason why we made a filtering option on the panel below. Therefore, users can select the specific instrument option to filter the results for more accurate data interpretation. However, in the paper, since we had a limited number of V-mode spectra (53 out of 442 mass spectra) in the current database, we carried out mass spectra comparisons with the entire database to demonstrate the tool's functionality. To address this, we have added text in the discussion section as follows:

"For especially AMS users, it's important to consider the instrumental conditions of reference mass spectra in the database, such as the instrument analysis mode (e.g. W or V mode) and particle size range, as these factors can impact the mass spectrum during AMS data processing. For instrument analysis mode, in this paper, we had a limited number of V-mode spectra which was the mode of our target spectra, so we conducted comparisons using the entire database to demonstrate the tool's functionality. However, since W-mode analysis may result in being able to more confidently fit ions due to its higher resolution than V-mode, the correlation may be different when comparing to W-mode vs V-mode. To address this, our tool offers an instrument filtering option, which we recommend to use for more precise AMS data interpretation."

(3) Particle size inconsistency

The MS of same OA factors are different for particles smaller than 2.5 μm (PM2.5) and 1 μm (PM1). However, these differences differ between different factors due to their varied size distributions. For example, the authors used the HOA MS of PM2.5 samples in Xi'an and Beijing, while the HOA MS of PM1 samples in other cities in Fig. 5. Such difference in aerosol size would cause uncertainties in comparison. How would the authors address this issue?

Since this database tool is built based on the existing webpage-based mass spectral database, particle size is beyond the scope of this paper. However, the tool provides metadata and easy access to the reference paper on the internet. Therefore, users can obtain particle size information by reviewing the reference paper or reaching out to authors. To clarify the impact of this parameter, we added the following note:

"Particle size range (due to different aerodynamic lens use), although not provided by this tool, can also influence the mass spectrum. To access this information, users can utilize the metadata and reference paper available online via the panel. We encourage users to carefully review the reference paper for particle size details and, if needed, additional information such as a fragmentation table for their data interpretation."

2. Line 388, What is the difference in the cosine score between the SOA spectra caused by different precursors under the same conditions (i.e., RH,T and seed)? Are there any further support other than the mass spectra?

Line 388 described the potential source of a new combination of OOA factors based on non-ambient comparison results. We compared the cosine score between our input OOA factor mass spectrum and reference mass spectrum in the database. Since this tool is designed to compare mass spectrum data, we concluded all the results in the manuscript based on the mass spectrum and its correlation results. The information about sample conditions (i.e., RH, T, and seed) is not provided through the tool. If users want to confirm the SOA sample conditions, they need to review the details in the reference paper. If some of the reference SOA mass spectra in the database have the same conditions, then users can select the 'Existing MS' option in section 2(a) to compare existing MS in the database with reference MS.

3. Consider using transparent marks in all diagrams.

We assume the reviewer is referring to the m/z value labels on the mass spectrum plot. Thank you for the suggestion. We modified the mass spectrum plot with transparent m/z value labels in the manuscript.

Title: A searchable database and mass spectral comparison tool for aerosol mass spectrometry (AMS) and aerosol chemical speciation monitor (ACSM)

Response to Referee #2

The authors appreciate the overall comments of Referee #2 and we would like to thank for his/her constructive comments. In the following, the Referee suggestions (in bold) and answers are addressed (plain text for response, quotes for new text added to manuscript).

General comments:

1. **The software tool also allows for a weighting of the mass spectra. This introduces additional uncertainty and I wonder if it is necessary. The user can vary the weighting until the result matches with the desired (or expected) results. I am not sure if this is a valid approach.**

As mentioned in lines 155-158 in the method section, reweighting the mass spectrum can improve correlations by varying their mass weighting and intensity scaling factors (Stein and Scott, 1994). It has been commonly used as a statistical method to analyze correlation between mass spectra. The weighting option is applied to all m/z, not individual m/z values, so if the mass spectrum of interest doesn't have any correlation with the reference mass spectrum, users cannot obtain desired (or expected) results.

2. **I wonder if it is possible to optimize this weighting by taking different source spectra from the data base (e.g. all BBOA spectra, all diesel exhaust spectra) and adjust the weighting factors until best scores between all MS from the same source are reached. Then, these weighting factors can be used as default or recommended settings. This should reduce the degree of free and subjective choice of the user.**

The database here has been built based on the existing web-based mass spectral database and its metadata does not include a classification category based on different sources. As a result, the tool currently does not have the capability to categorize each mass spectrum by its source. Furthermore, the primary purpose of the tool is to provide correlated reference mass spectra to users, rather than selecting specific reference mass spectra to compare. Given these limitations, adding the suggested function is challenging. This would likely be on the level of scope of a separate analysis and manuscript. Nevertheless, we are grateful for the valuable suggestions and will continue to try to enhance the tool's functionality.

Minor comments:

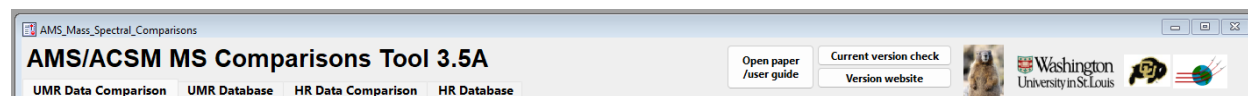
1. **This software tool should also be accessible through the ToFAMS Software Page (https://cires1.colorado.edu/jimenez-group/ToFAMSResources/ToFSoftware/index.html)**

Thank you for the suggestion. Since the ToFAMS Software page is currently in the process of being moved to Github, this software tool might not be accessible through the webpage. However, users can access the tool through the existing mass spectral database (https://cires1.colorado.edu/jimenez-group/AMSsd/).

2. Is there any interaction with Datalystica (the SoFi tool by Canonaco and coworkers)? I have the feeling this is more and more used by AMS/ACSM users for their PMF analysis.

We have not directly worked with Datalystica on the development of this tool or writing this manuscript. However, we expect that they and their users will learn of the tool (e.g. we presented this database at the Annual AMS Users Meeting in October 2023 which included several European attendees).

3. Does the AMS MS Comparison Tool regularly check for updates or changes in the database? Does "Download the newest version of DB" load new data from the database or does it load the updated Igor procedures only?

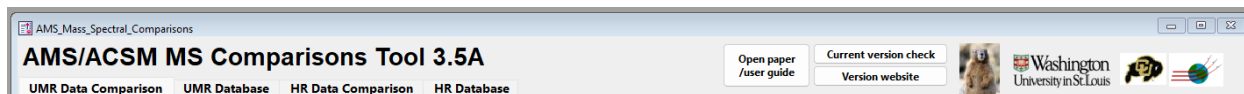
For this, we updated the tool and added the button ("Current version check") so that users can check the version of the database and tool they are using. The "Version website" button (previously "Download the newest version of DB") is linked to GitHub where users can download both the newest database and Igor procedures.

To clarify this, we have modified and added the sentences as follows:

"Users can download the latest version of the database as an .h5 file and the procedure file for this tool through the GitHub link on the existing AMS database webpage (https://cires1.colorado.edu/jimenez-group/AMSsd/). Users can confirm the version of the database and procedure file they are using on the panel and update them by downloading files at the link."

4. I noticed that some URL are not given or are not valid anymore. Also, group identification and persons doing the measurements need updates (at least in our case).

Since all the metadata including URLs in the database are derived from the existing webpage (https://cires1.colorado.edu/jimenez-group/AMSsd/), some URLs could not be valid or be not given if the original webpage has the same issue. We are in the process of checking if each URL is valid manually. Group identification and persons doing the measurements are newly added metadata for this tool. It would be good to update them for all the reference mass spectra, but as the webpage database has long since been developed, there are practical difficulties to do it. If user submissions to the database have wrong or incomplete information, we can attempt to rectify this, but again we encourage users to review the submitted details for a specific spectrum of interest. If users who previously submitted their data to the website need to correct or update their specific information for this tool, they can contact the database website manager for modifications.

5. Should/can users update their entries in the data base to keep it up to date? Or can this be done via the AMS MS Comparison Tool?

Users should update their entries in the database via the GitHub downloading the newest database. As mentioned above, we updated the tool and users can check the current version of the database they are using with the "Current version check" button below.

6. **Last but not least: You need a good acronym (such as SPARROW, SQURREL, PIKA etc).**

We appreciate this comment. The acronym of this tool is decided as MARMOT (AMS/ACSM mAss spectRal coMparisOn Tool).

Specific comments:

1. **lines 68-74: BBOA and COA are also primary OA. HOA as surrogate for POA is too simple.**

We agree with the referee's comment. As we are able to identify BBOA and COA with the improvement of AMS data analysis, HOA is currently identified as a surrogate for POA directly emitted from fossil fuel combustion not simply combustion sources. To clarify this, we have modified the manuscript like below:

"a surrogate for POA directly emitted from fossil fuel combustion"

2. **line 123: "Igor-Pro"**

Thank you for the correction. We have modified the word to "Igor-Pro".

3. **line 148: Is the uncentered correlation coefficient also the default setting in PMF analysis? (both PET and SoFi).**

In PET, both the uncentered correlation coefficient and the Pearson R are used in PMF analysis (Ulbrich et al., 2009). But normally the uncentered correlation coefficient is used for the mass spectra, and Pearson R for the times series. In SoFi, it provides various correlation coefficients such as Pearson R, Spearman R, and Kendall tau including also the uncentered correlation coefficient for PMF analysis.

4. **lines 155-165: Is there an option to reweight only selected m/z values? e.g., only downweight mz 44?**

The reweighting option applies to all m/z values, not selected m/z values.

Title: A searchable database and mass spectral comparison tool for aerosol mass spectrometry (AMS) and aerosol chemical speciation monitor (ACSM)

Response to Referee #3

The authors appreciate the overall comments of the Referee #3 and we would like to thank for his/her constructive comments. In the following, the Referee suggestions (in bold) and answers are addressed (plain text for response, quotes for new text added to manuscript).

Minor comments:

1. **My major confusion is that if you have HR MS, why is not directly compare with other HR MS other than comparison with family ions (line 290-298). The family ions is shown the exact composition of fragments. In addition, the case in the manuscript is also not presented the HR MS comparison which in my image, is more precise than UMR MS.**

Since this database tool has been designed to serve both AMS and ACSM users, we decided to focus on UMR comparison analysis even if we had HR MS. The primary objective of our paper is to introduce and demonstrate the functionality of this database tool, so we thought UMR MS comparison analysis would be more effective to show the function of the tool for both AMS and ACSM users.

2. **The high correlation of BBOA with OOA factors (line 306-308) is also not accept by me. Maybe this is not a good case which BBOA is not good separated which included many CHO ions.**

As mentioned in lines 308-311, users don't always obtain a 'perfectly separated' mass spectrum identifying the mass spectrum as HOA, BBOA, OOA, etc. In addition, the identification of each factor is executed by considering other AMS data such as time series, sample conditions, etc., as well as factor mass spectrum. As a consequence of this complexity, our BBOA factor may have high correlations with OOA factors. However, since our BBOA was not perfectly separated, we believe we could show the potential application of this database tool during AMS data processing.

Title: A searchable database and mass spectral comparison tool for aerosol mass spectrometry (AMS) and aerosol chemical speciation monitor (ACSM)

Response to Referee #4

The authors appreciate the overall comments of Referee #4 and we would like to thank for his/her constructive comments. In the following, the Referee suggestions (in bold) and answers are addressed (plain text for response, quotes for new text added to manuscript).

Comments:

1. **The reason why the Cosine score 'only', among other possible methods, has been chosen shall be clarified.**

Cosine score was used in this panel since it has been proven to show better performance in calculating correlations between MS compared to other methods and therefore is commonly used to compare the similarity between MS in analyses of AMS spectra. To clarify this in the manuscript, we have modified the manuscript as follows:

"In this comparison tool, we chose cosine similarity to estimate mass spectrum similarity. Cosine similarity has been proved to show better performance in calculating correlations between MS compared to other methods (Stein and Scott, 1994; Ulbrich et al., 2009). Therefore, it is commonly used to analyze the similarity between MS in analyses of AMS spectra, referring to it as the dot product with normalized spectra input or uncentered correlation coefficient (e.g. Marcolli et al., 2006; Lambe et al., 2015; Day et al., 2022)."

2. **Similar databases using such comparison tools already exist for other types of datasets. In particular, cf. SPECIEUROPE mainly for PM source profiles obtained from receptor models applied to chemical species (https://publications.jrc.ec.europa.eu/repository/handle/JRC96463), which could be mentioned here, and possibly seen as a source of inspiration for further development.**

Thank you for your suggestion. We have added the given reference in the discussion section as follows:

"As a variety of databases and tools (e.g., SoFi, SPECIEUROPE, and ICARUS) have been developed to enhance data analysis efficiency in the atmospheric field (Canonaco et al., 2013; Pernigotti et al., 2016; Nguyen et al., 2023), we anticipate providing a valuable database and tool for users as well."

3. **As different fragmentation tables may be used to retrieve OA mass spectra from AMS/ACSM measurements, information on the used fragmentation table should be stated as a metadata for each mass spectra archived in this new database.**

As mentioned in the manuscripts, all the metadata for each mass spectrum in this database is from the existing mass spectral database (https://cires1.colorado.edu/jimenez-group/AMSsd/)). As the web-based database didn't provide fragmentation tables, this database also doesn't offer them. Also since the

webpage database has long since been developed, there are practical difficulties to do so. In addition, it's not clear that differences in the fragmentation table would exert a major uncertainty or variability in the spectra (compared to other factors) since most analysis start with a default fragmentation table and then only make alterations to more accurately correct for other interferences specific to that instrument and sampling (i.e. mainly gas-phase interferences). Those adjustments are meant to make the spectra more accurately reflect the aerosol concentration and the specific values are not important. However, if users feel they need the fragmentation table information, as this database allows users to access the paper easily via the panel, we encourage users to review the corresponding paper or reach out to the authors for fragmentation table details (if they are not in the paper). To clarify this comment, we added the following note:

"*We encourage users to carefully review the reference paper for particle size details and, if needed, additional information such as a fragmentation table for their data interpretation.*"

4. The meaning of the y-label used in every figures (i.e., 'fraction of signal' vs. 'relative abundance') should be better explained (e.g., in the paragraph starting line 130 ?).

We appreciate your detailed observation. All the mass spectra in the database are stored and used on the panel after normalization. So both should indicate 'relative abundance'. We have changed the y-label of the figures and the panel to 'Relative abundance'.

5. The authors might also consider any interest in keeping the exact same ranges for each y-axis in Figure 7 (??).

Since the spectra are already normalized, we don't feel that the scaling is relevant for comparison and would rather scale them to utilize the full y-axis range for better comparison of signals across the range.

6. (How) could we imagine to store and compare mass spectra associated with some kind of uncertainties for each m/z? (e.g., for deconvoluted mass spectra obtained from bootstrap analyses).

We agree that this information could be useful for further statistical evaluation. However, it is beyond the scope of this work. For most spectra, we don't have uncertainty information. Also even if people provided them for some spectra, they may not always mean the same thing (e.g. bootstrap uncertainties for a PMF component, vs. variabilities for a standard or changes in time for SOA in a chamber etc.). Therefore, it would take some effort to think through how to use the uncertainty information in a way that improves the process.

7. What about mass spectra obtained from measurements achieved with the new ToF-ACSM:X instrument ?

If users have the required waves (spectra, mz value, and spectra name (for 2D)) for comparison on the panel in Igor-Pro, it's also possible to use this tool. If such spectra are posted on the MS database, this info

would be included in the metadata and a filter for those types of spectra can be added to the comparison tool. Generally, it seems that ToF-ACSM X data would be fairly comparable to HR-AMS V-mode data.