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

Sohyeon Jeon1, Michael J. Walker1, Donna T. Sueper2,3, Douglas A. Day2, Anne V. Handschy2, Jose L. Jimenez2, Brent J. Williams1

1Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, MO, USA
2Department of Chemistry, and Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA
3Aerodyne Research Inc, Billerica, MA

Correspondence to: Brent J. Williams (brentw@wustl.edu)

Abstract. The Aerodyne Aerosol Mass Spectrometer (AMS) and Aerosol Chemical Speciation Monitor (ACSM) are the most widely applied tools for in-situ chemical analysis of the non-refractory bulk composition of fine atmospheric particles. The mass spectra (MS) of many AMS and ACSM observations from field and laboratory studies have been reported in peer-reviewed literature and many of these MS have been submitted to an open-access website. With the increased reporting of such data sets, the database interface requires revisions to meet new demands and applications. One major limitation of the web-based database is the inability to automatically search the database and compare previous MS with the researcher’s own data. In this study, a searchable database tool for the AMS and ACSM mass spectral dataset was built to improve the efficiency of data analysis using Igor Pro, consistent with existing AMS and ACSM software. The database tool incorporates the published MS and sample information uploaded on the website. This tool allows the comparison of a target mass spectrum with the reference MS in the database, calculating cosine similarity, and provides a range of MS comparison plots, reweighting, and mass spectrum filtering options. The aim of this work is to help AMS users efficiently analyze their own data for possible source or atmospheric processing features by comparison to previous studies, enhancing information gained from past and current global research on atmospheric aerosol.

1. Introduction

Atmospheric aerosol particles have adverse effects on human health and impact visibility, the hydrological cycle, and climate changes through direct and indirect radiative properties (Ramanathan et al., 2001; Bäumer et al., 2008; Kampa and Castanas, 2008; IPCC, 2021). Globally, the dominant constituent of dry fine-mode respirable aerosol particles is organic matter (OM), commonly referred to as organic aerosol (OA). Understanding the characteristics, sources, and processes of OA is key in developing aerosol-related control policies and air quality and climate models. However, OA is chemically complex, with thousands of different compounds detected in single samples (Goldstein and Galbally, 2007). The compositional complexity of OA arises from diverse primary sources and reactions of organic species in the atmosphere that produce secondary OA (SOA) material from gas-to-particle conversion or chemically aged OA (De Gouw and Jimenez, 2009).

Aerosol Mass Spectrometry (AMS) has widely been applied to atmospheric science research for analysis of the bulk chemical composition of fine particles. It allows one to measure the non-refractory components of the particles with high time resolution, most typically reporting mass concentrations of total fine OA, sulfate, nitrate, ammonium, and chloride (e.g. Jayne et al., 2000; Jimenez et al., 2003; Allan et al., 2004; DeCarlo et al., 2006; Canagaratna et al., 2007; Baltensperger et al., 2010). The AMS has frequently been used in both field and laboratory studies. In field studies, the AMS has characterized atmospheric particles in environments such as urban, rural, remote, forested, ocean, and agricultural regions (e.g. Allan et al., 2004, 2006; Phinney et al., 2006; Aiken et al., 2009; He et al., 2011; Bates et al., 2012; Cleveland et al., 2012; Dall’Osto et al., 2013; Hao et al., 2014; Xu et al., 2014; Lee et al., 2015; Modini et al., 2015; Xu et al., 2015; Young et al., 2016; Kim et al., 2017). In addition, AMS has been used on many platforms such as mobile labs, aircraft, and ships as well as stationary sites (e.g. Bahreini et al., 2014; Xu et al., 2014; Lee et al., 2015; Modini et al., 2015; Xu et al., 2015; Young et al., 2016; Kim et al., 2017).

Correspondence to: Brent J. Williams (brentw@wustl.edu)
et al., 2003; Zorn et al., 2008; Mohr et al., 2011; Coggon et al., 2012; Drewnick et al., 2012; Claeys et al., 2017; Pirjola et al., 2017; Robinson et al., 2018; Shilling et al., 2018). For laboratory studies, particles generated from a variety of sources, including wood burning, cooking, trash burning, coal and fuel combustion, etc., have been characterized. DeCarlo et al., 2006; Canagaratna et al., 2004; Schneider et al., 2006; Aiken et al., 2008; Weimer et al., 2008; Mohr et al., 2009; Chirico et al., 2010; He et al., 2010; Adam et al., 2011; Heringa et al., 2011; Wang et al., 2013; Collier et al., 2015; Fortenberry et al., 2018). Furthermore, AMS has been employed to analyze SOA formed from different oxidation experiments using environmental chambers or flow tube reactors with varying oxidants, aging processes, concentrations, humidity levels, temperatures, precursors, etc. (e.g. Bahreini et al., 2005; Kroll et al., 2005, 2006; Murphy et al., 2007; Ng et al., 2007, 2008; Chhabra et al., 2010; Lambe et al., 2011, 2012; Loza et al., 2014; Riva et al., 2016; Boyd et al., 2017; Lim et al., 2019).

The Aerosol chemical speciation monitor (ACSM) also has been used to analyze the bulk chemical composition of non-refractory components with a very similar sampling and detection technology as the AMS. The key difference is that the ACSM was developed with reduced complexity (e.g. no particle size measurement) and performance (Ng et al., 2011). A major advantage of ACSM over AMS is that it provides a smaller size, lower cost, simpler operation, and less attention from site managers than AMS. In addition, ACSM data are analyzed with the same techniques that are used for the AMS. Therefore, the ACSM has been deployed for long-term monitoring at locations including urban (e.g. Ng et al., 2011; Sun et al., 2011; Budisulistiorini et al., 2013; Carbone et al., 2013; Budisulistiorini et al., 2014; Aurela et al., 2015; Petit et al., 2015; Budisulistiorini et al., 2016; Reyes-Villegas et al., 2016; Rattanavaraha et al., 2017; Sun et al., 2018; Kommula et al., 2021), suburban (Zhang et al., 2018), rural (e.g. Tiitta et al., 2014; Canonaco et al., 2015; Fröhlich et al., 2015a; Parworth et al., 2015; Bessi et al., 2016; Budisulistiorini et al., 2016; Schlag et al., 2016; Zhao et al., 2020), remote (e.g. Budisulistiorini et al., 2015; Fröhlich et al., 2015b; Ripoll et al., 2015; Michoud et al., 2017), and forested areas (e.g. Fröhlich et al., 2013; Minuillón et al., 2015; Heikkilä et al., 2020).

According to the broad and diverse application of the AMS and ACSM, data analysis procedures have also been advanced to improve efficiency and accuracy to interpret the data. These advances have made it possible to divide the quantified OA mass concentrations obtained from AMS and ACSM into hydrocarbon-like OA (HOA), a surrogate for primary OA (POA) directly emitted from combustion sources and oxygenated OA (OOA, a surrogate for SOA generated from a chemical reaction or phase partitioning). Furthermore, OOA can be subdivided into low-volatility OOA (LV-OOA or more-oxidized OOA, MO-OOA) and semi-volatile OOA (SV-OOA or less-oxidized OOA, LO-OOA) when combined with methods such as thermal-denuding inlets and positive matrix factorization (PMF) analysis (Paatero and Tapper, 1994; Zhang et al., 2005; Lanz et al., 2007; Ulbrich et al., 2009; Zhang et al., 2011). In addition, these deconvolved OA have been specifically characterized depending on source types such as biomass burning (BBOA) and cooking (COA), increasing the utility of source apportionment of atmospheric aerosols (Paatero and Tapper, 1994; Ulbrich et al., 2009; Mohr et al., 2009). This improvement in data analysis, in addition to the higher performance and mass spectral resolution of modern time-of-flight (ToF) MS detectors often deployed with the AMS and ACSM, has led to a growing number of reported measurements under various sample configurations and conditions.

Mass spectra (MS) from the AMS and ACSM provide powerful information to interpret the OA data. A mass spectrum with specific marker ions, or specific ratios of ions, can often be related to a source or chemical process. Here we broadly classify MS types by mass resolution (unit mass resolution or UMR vs. high mass resolution or HR) and extent of MS separation through data processing (non-deconvoluted/deconvoluted). UMR indicates that the signal at each mass is separated from the next 'integer' m/z. In contrast, HR is able to quantify multiple HR ion fragments at a nominal m/z, which enables more detailed characterization than UMR. DeCarlo et al., (2006) has demonstrated the ability to characterize HR ion fragments below a nominal mass, typically m/z 120. These HR ion fragment signals can be grouped into different ion classes such as Cn H2n, CnH2Ox, etc., which are called 'HR families'. A HR mass spectrum is often plotted in a color-coded stacked bar chart of these HR families. Deconvolution indicates whether or not the MS are generated from an additional analysis step like PMF. If the mass spectrum is obtained from PMF, it is called a 'Deconvoluted' mass spectrum and is identified as HOA, SV-OOA, LV-OOA, etc. If not separated, the mass spectrum is referred to as 'Non-deconvoluted'. A non-deconvoluted MS may be representative of multiple contributing sources or aging processes if an ambient sample or can be a single source or standard observed in a lab study. Figure 1 presents various examples of AMS OA MS classified by resolution and extent of MS processing.
Figure 1. Various examples of AMS OA mass spectral data divided by resolution and deconvolution. Each mass spectrum was plotted via the developed AMS MS comparison panel here and original data were from Alfarra et al., (2004); Canagaratna et al., (2004); Katrib et al., (2004); Bahreini et al., (2005); Li et al., (2012); Loza et al., (2012) for non-deconvoluted MS and Mohr et al., (2012) for deconvoluted MS.

With existing datasets growing in complexity and number according to the advancement and accessibility of technology in the aerosol science field, it is increasingly important to develop customizable databases that store previous observations and act as a reference for future studies. Previous AMS and ACSM data have often been provided in text and graphical form, such as peer-reviewed journal publications. However, as the amount of data grows, finding the appropriate data in a usable format is increasingly challenging. Previous efforts to build a database for AMS and ACSM have been made by Ulbrich et al., (2009). They collected published AMS and ACSM spectral data and metadata (e.g., instrument operating conditions, sample, and experiment details), and posted it on an open-access website (https://cires1.colorado.edu/jimenez-group/AMSsd/). The digital data was uploaded as an Igor Pro software (Wavemetrics, Portland, OR) text file format, with an .itx file name extension. This allows users to load or export the data in the Igor Pro software commonly used for AMS and ACSM data analysis. This is a useful repository to advertise and distribute AMS and ACSM data. However, this repository is not programmed or formatted to systematically search for appropriate results and compare them with the researcher’s own data. Beyond the need to program a search method, formatting needs to be standardized compared to the current method where each spectral data is individually uploaded on the webpage with some variability in format (e.g., variable mass-to-charge (m/z) ranges).

In this study, we introduce a searchable database tool for the AMS and ACSM mass spectral dataset. Our aim is to improve the efficiency and utility of the AMS and ACSM data analysis process, building on the existing database from Ulbrich et al., (2009). We converted the web-based database to a software-based database format using Igor-pro and developed the Igor Pro visualization interface. The interface is called ‘AMS MS comparison panel’ or simply ‘panel’ in the following sections. To demonstrate the practical application of this tool, we compared our AMS MS with reference MS in the database. This comparison demonstrates how our tool can be useful in practical applications. We believe that incorporating this comparison tool will enhance the ability of AMS and ACSM users to conveniently compare their data with previously reported studies.

2. Methods

The developed database is based on the existing open-access website by Ulbrich et al., (2009), providing published AMS and ACSM mass spectral datasets. The website mainly consists of 3 separate web pages such as unit mass resolution...
(UMR, standard vaporizer), high resolution (HR, standard vaporizer), and capture vaporizer (CV, both UMR, and HR). Capture vaporizer indicates that a more recently developed particle vaporizer was installed, which can lead to different spectra primarily from increased thermal decomposition (Hu et al., 2018a). For each sample, the website provides metadata in a table format including spectra identification, sources, research groups, AMS instruments used, electron ionization (EI) energy, and vaporizer temperature. It also provides citation information and the original figure number in the publication, sometimes with additional comments. The mass spectrum is uploaded in a digital form for analysis as an .itx file, a plain text file that can be directly loaded into the data analysis software Igor-pro (Wavemetrics, Portland, OR).

Here, to convert this web-based information and individual mass spectrum files to the software-based database in Igor-Pro, the given information was fetched from HTML using the ‘fetchURL’ function in Igor-Pro to convert the contents of the HTML to strings. Metadata was extracted from these HTML strings and saved in Igor Pro. Each mass spectrum was automatically downloaded to the computer first and then saved in the Igor-Pro database. When saving the mass spectrum, it was aligned and normalized in a uniform format with a consistent m/z range from 1 to 600 and a sum of the spectrum summing to a value of 1. In the case of high resolution MS, they were converted to a UMR mass spectrum to make the HR mass spectrum directly comparable to the UMR mass spectrum, and then also saved into the Igor-Pro database. HR ion family data was also converted to UMR form and saved using the same method. For example, if the HR mass spectrum has three ions as CO$_2^+$ (m/z 43.9898), C$_2$H$_4$O$^+$ (m/z 44.0262), and C$_3$H$_8$+ (m/z 44.0626), all these three ions are considered as m/z 44 on the database panel in both UMR and HR. However, for HR ion families, these ions are respectively saved to corresponding HR ion families such as CHOgt1, CHO1, and CH (a detailed description of HR families is presented in Table S1). After creating the software-based database, the visualization interface (i.e., AMS MS comparison panel) was constructed in Igor Pro to analyze the correlation between MS.

2.1 Mass spectrum correlation calculation

The main goal of this comparison panel is to provide a convenient function to analyze similarities between the mass spectrum of interest and the reference MS in the database. Cosine similarity is a useful method to estimate mass spectrum similarity (Stein and Scott, 1994). It has been widely used to analyze the similarity between AMS MS, 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). In this comparison tool, cosine similarity (referred to as ‘Cosine score’ on the panel) is used to compare MS. It measures the cosine of the angle between two vectors and is calculated by using the equation below.

\[
\text{Cosine score} = \cos(\theta) = \frac{A \cdot B}{||A|| ||B||} \quad (\text{Eqn. 1})
\]

where each vector A and B are corresponding to the mass spectrum of interest and each reference mass spectrum in the database, respectively. $A$ and $B$ denote the magnitudes of vectors A and B, and $A \cdot B$ indicates the dot product of A and B. The possible range of cosine similarity (score) is from 0 to 1, and the higher the score value, the higher the similarity between the MS.

In addition, the panel supplies the option to reweight the mass spectrum using Equation 2 below. Regarding the comparison of MS, instrument operation parameters can cause m/z-dependent differences in the amount or detection of ions, and preprocessing spectral intensities by reweighting can have a beneficial effect on improving correlations between acquisition methods (Stein and Scott, 1994). The score can be adjusted by varying their mass weighting and intensity scaling factors. Increasing the relative significance of the lower-abundance high m/z values can enhance the match-weighting of these more distinctive ions (i.e., molecular fragments). This is achieved by increasing the mass exponent (mass weighting) or decreasing the peak intensity exponent (intensity scaling factors) that corresponds to the $m$ and $n$ in Equation 2, respectively.

\[
\text{Weighted intensity} = [m/z]^m[\text{Peak intensity}]^n \quad (\text{Eqn. 2})
\]

The weighted exponents default to $m=0$ and $n=1$ on the panel. After reweighting, the mass spectrum is normalized by dividing each value by the sum of the reweighted mass spectrum (thus summing to unity again). Finally, the panel calculates the score using the scaled mass spectrum.

For HR mass spectral data, the score can be calculated with only the selected HR families on the panel. For example, if one selects three HR families out of 17, then it combines these three HR family MS of interest into one mass spectrum and sums to unity for normalization. It then calculates the score with a normalized mass spectrum utilizing the same method described above. Detailed HR family information will be described in the next section. This score is referred to as ‘Score with HR family’ on the HR data comparison tab of the panel.
2.2 Igor AMS mass spectral comparison panel

2.2.1 Data Comparison tab

Figure 2 shows a screenshot of the AMS MS comparison panel. The panel is divided into four tabs (dashed line in orange of Figure 2): i) UMR Data comparison, ii) UMR Database, iii) HR Data comparison, and iv) HR Database. Both data comparison tabs for UMR and HR data consist of three regions. Figure panel section 2a (highlighted in blue) is where a new or existing mass spectrum is selected for comparison against database MS. Section 2b (highlighted in red) is where the user will set search parameters and view results, and section 2c (highlighted in black) is where metadata of selected database components can be viewed. Screenshots of other tabs are shown in the supplementary material (Figs. S1-3).

In panel section 2a, users are able to select MS to be compared. There are two options for users to choose a mass spectrum of interest. One is to choose their own mass spectrum, and the other is to select an existing mass spectrum in the database. In the case of selecting their own mass spectrum, the mass spectrum has to be in UMR with a standard m/z range.

The current panel and database use a mass range of m/z 1-600. Detailed information on how to upload the user’s own data to the database panel is described in the user guide section (Chapter 2 for UMR and Chapter 3 for HR) of the supplementary information.

Panel section 2b contains the results table, mass spectrum plots window, scaling mass spectrum options, and comparison filter options. The results table shows the list of reference spectra sorted by the calculated score (cosine similarity) in descending order. Users are able to see the reference spectrum of interest on the mass spectrum plots window by clicking on the results table. For the UMR data comparison tab, the mass spectrum plot window provides five types of mass spectrum plots (selectable with the “Plot Type” drop-down above the plot shown in Fig. 2: sample mass spectrum of interest, reference mass spectrum selected, mirror-image sample and reference MS, subtraction of sample and reference MS, and the scatter plot between the sample and reference with the markers corresponding to the specific m/z value with regression information (Fig. 3). For the HR data comparison tab, the panel plots a stacked mass spectrum with HR ion families such as Cx, CH, CHO1, CHOgt1, etc (Fig. S2b). These HR ion families are the same names generated by PIKA v1.23B which is an AMS data analysis tool for HR (DeCarlo et al., 2006; Sueper, 2021, https://cires1.colorado.edu/jimenez-group/ToFAMSResources/ToFSoftware/).

The chemical formula for each HR ion family is described in Table S1. Users can select the HR ion family to be viewed on the window and can calculate a ‘score with HR family’ with these selected HR families to compare the mass spectrum specifically.
These various types of plots will help users visually observe the similarities and differences between the MS. The scaling mass spectrum option can reset the mass range and reweight the mass spectrum. Users are able to choose whether to apply the scaling options to all reference MS in the database (below the results table in panel section 2b) or to only one selected reference mass spectrum on the list (top right of the panel in panel section 2b) in UMR data comparison tab. For HR data, it only provides the scaling option for all reference MS. Furthermore, the panel supplies comparison constraints to filter the reference MS in the database depending on the metadata. Comparison constraint categories include sample type, perturbation, deconvolution, instrument, MS resolution, and CO and CO$_2$ contribution sections. Detailed descriptions of these constraints are provided in the supplementary material (Section S1).

![Figure 3. Five types of plots provided by the panel to inform mass spectral comparisons: (a) sample mass spectrum of interest, (b) reference mass spectrum selected, (c) mirrored-image sample and reference MS, (d) subtraction (difference) of sample and reference MS, and (e) scatter plot between the sample and reference mass spectrum with the markers corresponding to the specific m/z values with regression information. Original data were from Ulbrich et al., (2009) and Mohr et al., (2009). In this example (as observed in panel d) the sample MS (HOA) contains additional m/z 43 and 57 (both of which have primary or secondary contributions and would need to further explore the HR results to distinguish), and the reference MS (laboratory, chicken cooking without skin) contains additional m/z 60 and 73 (common biomass markers or carboxylic acids).](image)

Finally, in panel section 2c, users can find detailed information on the sample mass spectrum of interest and selected reference MS in the results table. The “Open metadata table” button in panel section 2c enables a new window to pop up for users to obtain information. The new window shows sample type, perturbed type, analysis, instrument type, resolution, vaporizer type, EI energy value, vaporizer temperature, experimenter’s name, group, citations, citation URLs, figure numbers.
in the citation, and related comments (Fig. S4). In addition, users can directly open the relevant published paper in a web browser via the panel (“Open article URL” button in panel section 2c) if the URL address of the paper is saved in the database.

2.2.2 Database tab

Both database tabs of UMR and HR (Figs. S1, S3) display reference MS in the database. Users can simply observe the reference mass spectrum in the database with metadata without correlation calculation. Users can scan the MS stored in the database through this tab and easily obtain the corresponding reference metadata.

3. Application of the AMS MS database and comparison panel

Here, we used PMF factor mass spectral data from the DAURE campaign (Determination of the sources of atmospheric Aerosols in Urban and Rural Environments) to display the utility of the developed database and comparison panel. Data for this study were acquired in high resolution during the intensive DAURE field campaign in Montseny, Spain, during February-March 2009. An 8-factor solution (FPEAK=0) was chosen since this was the lowest number of factors at which the HOA and BBOA factors showed a clear separation from each other and OOA. Six factors were recombined to make up the OOA factor. The BBOA and HOA factors used here are those from the 8-factor solution. More details about an overview of the DAURE campaign and a summary of the results can be found in Minguillón et al., (2011), Pandolfi et al., (2014), and Zhang et al., (2022). For comparison, the acquired PMF factor mass spectral data in HR was converted to a UMR mass spectrum. Figure 4 shows the converted PMF factor MS of HOA, BBOA, and combined OOA in UMR and stacked HR ion families, respectively. In this section, we will introduce examples of ways to utilize the new comparison panel by applying functions via the UMR data comparison tab and present a potential for the HR data comparison tab based on HR families and ‘score with HR family’ to obtain more information to interpret the data depending on our factor MS.

Figure 4. PMF factor MS from Montseny, Spain during the DAURE campaign (Minguillón et al., 2011) in UMR and stacked HR ion families. The acquired PMF factor MS in high resolution were converted to a UMR mass spectrum and plotted by the AMS MS comparison panel here.

As introduced above, the developed database panel provides different ways to compare the MS with filtering or scaling options. Among these options, the ‘sample type’ category in the filtering option (Fig. 2. section b) enables researchers to select ‘ambient’ or ‘non-ambient’ samples in the database. The ‘Ambient’ option indicates direct measurement of ambient air and the ‘non-ambient’ option includes analysis of chamber studies, source emission studies, standards, etc. The scaling option is used to scale the mass spectrum using the mass range, mass exponent, and intensity exponent based on Eqn. 2. The default setting of the scaling option on the panel is $m=0$ and $n=1$ on Eqn. 2. The comparison process and results for each method will be described in this section with the DAURE campaign PMF factor MS as an example.
Selecting the ‘ambient’ option is useful when the user wants to compare their mass spectrum only with previous ambient measurements. For example, after the final PMF factors are identified, we may want to confirm if the identified PMF factors have a similar MS compared with previous studies. In this case, users can apply the ‘ambient’ sample filter to compare.

Table 1 shows the comparison results with the cosine similarity when we executed a UMR comparison of the HOA factor identified from the DAURE campaign (Montseny, Spain) with the database as an example. In the results table, most of the samples in the database with a score greater than 0.9 were shown as ‘HOA’. An interesting observation that appears is that some cooking organic aerosol (COA) factors also showed a high correlation with our HOA factor. The COA factor is often identified by its characteristic mass fragment of \( m/z \) 55, and it also has the same hydrocarbon ion series (\( m/z \) 29, 43, 57, 71, for \( C_nH_{2n+1}^+ \) and \( m/z \) 41, 55, 69, for \( C_nH_{2n-1}^+ \)) as the HOA (Mohr et al., 2009). The HOA factor we compared has a large \( m/z \) 55 in the spectrum and that may be the main cause for some COA factors to score a high match. Therefore, we would conclude that our HOA factor might be mixed with cooking sources. However, in this case, if the user has HR family information of the mass spectrum, as we do here, one could have more information to interpret the result.

Table 1. Top matches from “Ambient” sample UMR comparison results with HOA factor from the DAURE campaign (\( m/z \) 1-200)

<table>
<thead>
<tr>
<th># in DB</th>
<th>Sample</th>
<th>Score</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>311</td>
<td>Beijing urban area_2010_COA</td>
<td>0.9640</td>
<td>(Hu et al., 2016)</td>
</tr>
<tr>
<td>279</td>
<td>Xian and Beijing_HOA</td>
<td>0.9497</td>
<td>(Elser et al., 2016)</td>
</tr>
<tr>
<td>279</td>
<td>Shanghai urban area_2011_HOA</td>
<td>0.9492</td>
<td>(Hu et al., 2016)</td>
</tr>
<tr>
<td>253</td>
<td>MILAGRO Campaign_2006_HOA</td>
<td>0.9483</td>
<td>(Alfarra et al., 2004)</td>
</tr>
<tr>
<td>277</td>
<td>Paris Summer_2009_HOA</td>
<td>0.9477</td>
<td>(Crippa et al., 2013)</td>
</tr>
<tr>
<td>312</td>
<td>Beijing urban area_2010_HOA</td>
<td>0.9403</td>
<td>(Hu et al., 2016)</td>
</tr>
<tr>
<td>293</td>
<td>CalNex campaign_2010_HOA</td>
<td>0.9361</td>
<td>(Hayes et al., 2013)</td>
</tr>
<tr>
<td>270</td>
<td>SPC Research Station Po Valley_2008_HOA</td>
<td>0.9327</td>
<td>(Saarikoski et al., 2012)</td>
</tr>
<tr>
<td>314</td>
<td>Beijing urban area_2010_CCOA</td>
<td>0.9305</td>
<td>(Hu et al., 2016)</td>
</tr>
<tr>
<td>101</td>
<td>Paris Winter_2010_HOA</td>
<td>0.9269</td>
<td>(Crippa et al., 2013)</td>
</tr>
</tbody>
</table>
Figure 5. Segments of the MS (m/z 50-60) for our HOA and top 3 reference samples that correlated highly with our HOA in the “Ambient” comparison for m/z 55 HR ion families

‘Non-ambient’ option is convenient to compare the mass spectrum of interest with laboratory samples in the database. For example, this option is helpful when comparing a mass spectrum derived from a specific source such as BBOA. When applying the non-ambient filtering option, several non-ambient reference MS were highly correlated with our BBOA MS (Table 2). The characteristic mass fragment of BBOA is m/z 60 attributed to C_2H_4O_2 (from anhydrous sugars such as levoglucosan) and this specific ion was observed in most of the MS in the list of high score matches. On the HR data comparison tab, we are able to confirm if the m/z 60 is from the oxygenated HR family (CHOgt1). In Table 2, only the mosquito coil sample included HR ion family data and we confirmed the m/z 60 was mostly from CHOgt1 ion family which is pink (Fig. 6). Since the main component of mosquito coil is biomass such as sawdust, coconut shell flour, pyrethrum, potato starch, etc, it seems to show a high correlation with our BBOA and similar HR ion family contribution in m/z 60.

Table 2. Top matches from “Non-ambient” sample UMR comparison results with the BBOA factor from the DAURE campaign (m/z 1-200)

<table>
<thead>
<tr>
<th># in DB</th>
<th>Sample</th>
<th>Score</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>Ponderosa Pine Duff</td>
<td>0.9404</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Puerto Rican Mixed Wood</td>
<td>0.9337</td>
<td>(FLAME)</td>
</tr>
<tr>
<td>40</td>
<td>Southern Pine Needles</td>
<td>0.9335</td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>Mosquito Coil</td>
<td>0.9303</td>
<td>(Li et al., 2012)</td>
</tr>
<tr>
<td>38</td>
<td>Puerto Rican Fern</td>
<td>0.9296</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Alaska Core Tundra Duff</td>
<td>0.9275</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Ceanothus Leaves and Berries</td>
<td>0.9193</td>
<td>(FLAME)</td>
</tr>
<tr>
<td>27</td>
<td>Ponderosa Pine Needles and Sticks</td>
<td>0.9184</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Wax Myrtle</td>
<td>0.9038</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Lignin Powder</td>
<td>0.8951</td>
<td></td>
</tr>
</tbody>
</table>
An interesting observation from BBOA example is that when carrying out the ambient sample comparison with the BBOA factor, other identifications such as HOA, SVOOA, LOOOA, COA, etc were observed as well as BBOA (Table 3) even though our BBOA mass spectrum was highly correlated with biomass burning laboratory references. Since ion separation of AMS mass spectrum via PMF is largely impacted by the mix of sources present at a given time, users don’t always obtain a ‘perfectly separated’ mass spectrum identifying the mass spectrum as HOA, BBOA, COA, etc, but a possibly mixed mass spectrum. It indicates that even if the UMR score on the panel is lower, it could be possible that there is a good match with a single HR family or some families, and vice versa. For this case, ‘Score with HR family’ can be a useful function to gain more information on similarities. When users select the HR families of interest next to the HR plot window and click the calculate button below HR family selection, it combines only the selected HR families into one MS, calculates a cosine similarity in the same way, and displays the results named ‘score with HR family’ on the results table. Table 3 shows top 10 matches from ambient sample comparison results sorted by the UMR score and top 10 matches from ‘score with HR family’. For BBOA, the separation of the oxygenated ions provides a lot of information and is a key signature of BBOA spectra. On Table 3, we were able to observe more BBOA factor mass spectrum on the list when sorting the results by the score with CHOgt1 family including m/z 60. Most of the new matches showed BBOA, CCOA (coal combustion OA), HOA, and COA, not OOA like when sorting the list by UMR score. UMR match scores for these reference MS are included alongside the HR Family match score for comparison. We also carried out a score comparison utilizing additional HR families and family combinations (CH, CHO1, and combined), but it did not provide additional insight beyond what was observed with CHOgt1 alone (Table S2).

Table 3. Top matches from HR “Ambient” sample comparison results with the BBOA factor from the DAURE campaign (m/z 1-200) depending on UMR score and score with HR family (CHOgt1)

<table>
<thead>
<tr>
<th># in DB</th>
<th>Sample</th>
<th>UMR score</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>293</td>
<td>Changdao island_2011_biomass burning</td>
<td>0.9632</td>
<td>(Hu et al., 2013)</td>
</tr>
<tr>
<td>310</td>
<td>SOAS campaign_2013_SOA</td>
<td>0.9445</td>
<td>(Hu et al., 2015)</td>
</tr>
<tr>
<td>300</td>
<td>Beijing urban area_2011_HOA</td>
<td>0.9367</td>
<td>(Hu et al., 2016)</td>
</tr>
<tr>
<td>245</td>
<td>SOAR-1_Campaign_2005_SVOOA</td>
<td>0.9346</td>
<td>(Docherty et al., 2011)</td>
</tr>
<tr>
<td>313</td>
<td>SOAS campaign_2013_LOOOA_II</td>
<td>0.9287</td>
<td>(Hu et al., 2015)</td>
</tr>
<tr>
<td>296</td>
<td>Changdao island_2011_CCOA</td>
<td>0.9239</td>
<td>(Hu et al., 2013)</td>
</tr>
<tr>
<td>315</td>
<td>KORUS-AQ study_2016_LOOOA</td>
<td>0.9164</td>
<td>(Hu et al., 2018b)</td>
</tr>
<tr>
<td>269</td>
<td>Paris Summer_2009_SVOOA</td>
<td>0.9149</td>
<td>(Crippa et al., 2013)</td>
</tr>
<tr>
<td>301</td>
<td>Beijing urban area_2011_COA</td>
<td>0.9097</td>
<td>(Hu et al., 2016)</td>
</tr>
<tr>
<td>265</td>
<td>SPC Research Station Po Valley_2008_HOA</td>
<td>0.9050</td>
<td>(Saarikoski et al., 2012)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># in DB</th>
<th>Sample</th>
<th>UMR score</th>
<th>Score with CHOgt1</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>254</td>
<td>DAURE campaign_2009_BBOA</td>
<td>0.8348</td>
<td>0.9527</td>
<td>(Mohr et al., 2012)</td>
</tr>
</tbody>
</table>
Lastly, users can use the “scaling mass spectrum” feature, which is especially helpful for up-weighting signal intensity of larger and more unique ions by increasing the mass exponent value or decreasing the intensity exponent value. When we compared the OOA factor MS with ambient samples, we observed that our OOA factor MS was highly correlated with LVOOA (or MOOOA) MS from previous studies (Table S3) showing a score of more than 0.97. We also explored cosine scores of the OOA factor with non-ambient spectra within the database. However, when we compared the OOA factor MS with non-ambient samples, top scores in the list were shown as less than 0.9 (Table 4). For OOA, the abundances of fragment ions m/z 28 (CO\(_2\)) and 44 (CO\(_4\)) are extremely dominant compared to other ions in the mass spectrum (typically, the m/z 28 signal is constrained to be equal to the m/z 44 signal). These prominent abundances of a few spectral peaks may affect score calculation. In the case of ‘ambient’ comparison with deconvoluted MS dominated by m/z 28 and 44, signals from other ions may not highly influence the results as shown in Table S3. On the other hand, in the case of ‘non-ambient’ comparison with non-deconvoluted MS dominated by m/z 28 and m/z 44 signals of laboratory samples, signals from other ions are more unique to the specific laboratory conditions and may reflect the lower scores as shown in Table 4. In this case, we have two options to reduce these impacts on score results: i) increase the m/z exponent value, or iii) both i and ii in Eqn 2.

<table>
<thead>
<tr>
<th># in DB</th>
<th>Sample</th>
<th>Score</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>141</td>
<td>Oxalic acid</td>
<td>0.8493</td>
<td>(Takegawa et al., 2007)</td>
</tr>
<tr>
<td>334</td>
<td>Chamber m-Xylene aged SOA</td>
<td>0.8458</td>
<td>(Loza et al., 2012)</td>
</tr>
<tr>
<td>321</td>
<td>Incense Coil</td>
<td>0.8385</td>
<td>(Li et al., 2012)</td>
</tr>
<tr>
<td>145</td>
<td>Adipic acid</td>
<td>0.8196</td>
<td>(Takegawa et al., 2007)</td>
</tr>
<tr>
<td>149</td>
<td>Glyoxylic acid</td>
<td>0.8132</td>
<td>(Takegawa et al., 2007)</td>
</tr>
<tr>
<td>331</td>
<td>Citric Acid (C6H8O7)</td>
<td>0.7661</td>
<td>(Hu et al., 2018b)</td>
</tr>
<tr>
<td>143</td>
<td>Succinic acid</td>
<td>0.7649</td>
<td>(Takegawa et al., 2007)</td>
</tr>
<tr>
<td>234</td>
<td>Diesel_Exhaust_2</td>
<td>0.7629</td>
<td>(Sage et al., 2008)</td>
</tr>
<tr>
<td>322</td>
<td>Mosquito Coil</td>
<td>0.7580</td>
<td>(Li et al., 2012)</td>
</tr>
<tr>
<td>333</td>
<td>Chamber m-Xylene peak growth SOA</td>
<td>0.7529</td>
<td>(Loza et al., 2012)</td>
</tr>
</tbody>
</table>

Table 4. Top matches from UMR “Non-ambient” sample comparison results with the OOA factor from the DAURE campaign (default setting, m/z 1-200)

Figure 7 shows the reweighted OOA mass spectrum generated by two exponent options. When we increased the m/z exponent from 0 to 1 (Figure 7b), we observed that the relative abundance of lower mass fragments decreased and that of larger mass fragments increased in a similar scale of relative abundance as the original MS (Figure 7a). When we decreased the peak intensity exponent from 1 to 0.5 (Figure 7c), the maximum value of relative abundance was decreased. Signal intensities other than the dominant fragments (m/z 28 and 44) were more balanced than in Figure 7a. Figure 7d is the case where we applied both mass and peak intensity exponent options. It shows the reweighted spectra with relatively smaller contributions from lower masses and higher contributions from larger masses. Table 5 shows the list of reference MS with a high correlation with our reweighted OOA factor MS with various pairs of m/z and peak intensity exponent. The scores less than 0.9 were not shown in the table and the reference samples were sorted by the maximum score among the values from each scenario. It shows that for our OOA factor MS, decreasing the intensity exponent option resulted in more scores greater than 0.9 compared to increasing the m/z exponent. Also, we observed that the reweighted OOA factor mass spectrum is mostly correlated with coils, terpene
ozonolysis, aged SOA, etc. Correlations to the coils may be because both incense coil and mosquito coil have fragrances that contain aromatic compounds that can produce oxygenated aromatics during combustion.

Figure 7. The reweighted mass spectrum of DAURE OOA factor ((a): No adjustment, (b): increase m/z exponent from 0 to 1, (c): decrease intensity exponent from 1 to 0.5, and (d): apply (b) and (c) simultaneously). Enlarged mass spectrum with a restricted mass range (m/z 45-200) shown in the dashed line box.
Table 5. Top matches from “Non-ambient” UMR sample comparison results with the OOA factor from the DAURE campaign depending on decreasing peak intensity exponent and increasing mass exponent. The maximum score is shown in bold.

<table>
<thead>
<tr>
<th>m/z exponent</th>
<th>Sample</th>
<th>Score</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># in DB</td>
<td>0</td>
<td>0.9113 (Li et al., 2012)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td>0.9316 (Loza et al., 2012)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>0.5</td>
<td>0.9252 (Bahreini et al., 2005)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td>0.9199 (Li et al., 2012)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>0.5</td>
<td>0.9176 (Bahreini et al., 2005)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td>0.9174 (FLAME, 2007)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>0.5</td>
<td>0.9157 (Bahreini et al., 2005)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td>0.9151 (FLAME, 2007)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>0.5</td>
<td>0.9142 (Alfarran, 2004)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Top matches from “Non-ambient” sample UMR comparison results with the reweighted OOA factor from the DAURE campaign depending on the pairs of mass and intensity exponent value with restricted mass range (m/z 45-200).

<table>
<thead>
<tr>
<th>m/z exponent</th>
<th>Sample</th>
<th>Score</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># in DB</td>
<td>0</td>
<td>0.9097 (Bahreini et al., 2005)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>0.5</td>
<td>0.9587 (FLAME, 2007)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td>0.9563 (Bahreini et al., 2005)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>0.5</td>
<td>0.9277 (Bahreini et al., 2005)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td>0.9185 (Bahreini et al., 2005)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>0.5</td>
<td>0.9121 (FLAME, 2007)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td>0.9199 (Bahreini et al., 2005)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>0.5</td>
<td>0.9174 (FLAME, 2007)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td>0.9157 (Bahreini et al., 2005)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>0.5</td>
<td>0.9142 (Alfarran, 2004)</td>
</tr>
<tr>
<td></td>
<td># in DB</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

In addition to the comparison of the combined OOA mass spectrum, we used these scaling options to check if each of the original 6 OOA-related factors had potential characteristics to be identified before they were combined. As a result of the application of a variety of paired scalings of mass range, mass exponent value, and intensity exponent value, we observed potential characteristics between factors when using the mass range is m/z 45-200 and default exponent value (m=0, n=1) (Table S4). In this case, factor 1 and factor 4 were highly correlated with the oxidation of m-xylene (an anthropogenic VOC common in fossil fuels) and other potential ozonolysis products, and factor 5 and factor 6 were mostly correlated with terpene ozonolysis. Factor 2 and factor 3 showed relatively lower scores but nitrogen-related MS was shown in the results table in both factors. These results suggest that we could combine factors 1 and 4, factors 2 and 3, and factors 5 and 6 as potential OOA subcategories. When we calculated cosine similarities between these pairs of factors, it showed a high correlation with the score of 0.9576, 0.9469, and 0.9838, respectively, which supports their similarities in composition. Factors 5 and 6, in particular, had characteristic fragment ion of m/z 81 which can be indicative of monoterpane oxidation matching the results we obtained. Therefore, based on a new combination of factors, we may conclude that our OOA may be mainly derived from m-xylene.
(anthropogenic VOC) oxidation, terpene ozonolysis, and nitrogen-related reactions. For HR data comparison, unfortunately, all non-ambient samples in this match list were originally submitted only in UMR, so HR family comparison was not available in this case. Importantly, future MS submissions from the user community should include HR MS when available.

4. Conclusions

The existing AMS and ACSM mass spectral database in web-based form has been converted into a searchable, filterable software library coupled with a comparison panel in Igor Pro, the main analysis program currently used for AMS and ACSM data. The comparison panel provides the functionality to compare the mass spectrum of interest with MS in the database statistically and visually by cosine similarity and several types of mass spectral plots as UMR or stacked HR ion families. Furthermore, the option to exponentially reweight the mass spectrum and filter the samples depending on their measurement information (e.g. lab vs field data) can help users understand and analyze their data relative to a growing list of past observations. We believe this new database can be used to improve the efficiency of data interpretation and provide new insights for AMS and ACSM studies. However, the AMS and ACSM users’ network should further populate this digital mass spectral database in order to fully realize the potential strength of such a tool. Users can download the procedure file for this database and search the tool through the GitHub link on the existing AMS database webpage (https://cires1.colorado.edu/jimenez-group/AMSsd/). Users can also follow the listed procedures for submitting new MS to the database and refer to the user manual (see user guide section in supplementary material) to use this tool. While MS are one of the AMS data to be used for data interpretation, users should continue to consider all supporting chemical and meteorological measurements together with the AMS MS in determining factor identities. Our intention is that this database continues to grow and develop with the various types of AMS systems in operation, and can serve as a valuable tool to help AMS users frame their observations against a wealth of previous observations obtained globally.

Code availability

The code of AMS mass spectral database in this study was developed in the Igor Pro environment. The codes are available at https://github.com/ActlabW/AMS_MS_DB.

Data availability

AMS mass spectra data used in this study is available at https://cires1.colorado.edu/jimenez-group/AMSsd/.

Author contribution

SJ developed the software code and prepared the manuscript with constructive feedback from DTS and DAD. MJW conceptualized the project and framed the initial code. DTS and DAD contributed to inspecting the tool and the database. DTS, DAD, AVH, and JLJ supplied AMS mass spectral data of DAURE campaign (Montseny). BJW supervised the project. All authors participated in reviewing and editing the manuscript.

Competing interests

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

Acknowledgments

BJW acknowledges NSF CBET (award number 1554061) for supporting this study. DTS, AVH, DAD, JLJ acknowledge support from NASA 80NSSC21K1451 and NSF AGS AGS-2131914 and 2206655.
References


FLAME: http://chem.atmos.colostate.edu/FLAME/.


Parmar, Y., Worsnop, D. R., Kolb, C. E., and Davidovits, P.: Effect of oxidant concentration, exposure time, and seed


Sueper: https://cires1.colorado.edu/jimenez-group/ToFAMSResources/ToFSoftware/, last access: 17 January 2021.


