Guidelines for off-line and on-line source apportionment procedures using aerosol mass spectrometry including a protocol how to report such data in EBAS

J. Slowik, F. Canonaco, V. Moschos, Y. Tong, R. Casotto, A. Prévôt, I. El Haddad, U. Baltensperger
Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, 5232 Villigen PSI, Switzerland

Villigen, 10 May 2019

WP 1, Ground-based component for SLCFs

Task 1.4: Sources and sinks of atmospheric pollution in the polar areas
D1.4.1, Guidelines for off-line and on-line source apportionment procedures using aerosol mass spectrometry including a protocol how to report such data in EBAS

Document version 1

1. Overview

Source apportionment describes aerosol concentration and composition in terms of its contributing sources and processes, facilitating intelligent responses to aerosol climate and health effects. Aerosol mass spectrometry (AMS) using the Aerodyne time-of-flight aerosol mass spectrometer (ToF-AMS) and aerosol chemical speciation monitor (ACSM) has become one of the most useful and important data sources for source apportionment analyses, which are then commonly conducted using the positive matrix factorization (PMF) receptor model. Initial efforts focused on the use of AMS data from short-term (approximately month-long) on-line intensive campaigns, while in recent years measurement strategies have diversified to also include long-term or perpetual on-line measurements and off-line measurements of collected aerosol filter samples. These different measurement strategies give rise to different sampling considerations, as due to the differences in AMS data streams (where the ToF-AMS optimizes chemical specificity and ACSMs optimize measurement duration). Finally, advances in PMF analysis techniques open the possibility for diverse source apportionment strategies. The diversity of measurement and analysis options has greatly increased the power of AMS-PMF source apportionment, but the accompanying increase in complexity of an already challenging analysis causes significant variation in the quality of individual AMS-PMF-based studies.

To address this issue, this document details the major considerations in AMS-PMF analysis and provides implementation guidelines. We focus specifically on source apportionment of organic aerosol (OA), which has the largest uncertainties and has thus been the focus of most research[1]. For clarity, we distinguish between PMF analyses of on-line and off-line measurements (sections 2 and 3, respectively), which although similar in many respects have also some significant differences. On-line datasets that are long-term (year or more) or analyzed near-synchronously with data collection likewise require special approaches, as discussed in section 2.4. We do not present a detailed discussion of instrument operation and standard data analysis procedures, as these are covered extensively in the literature[2-7], although certain issues of particular relevance to source apportionment analyses are highlighted. Section 4 discusses the protocols for archiving AMS/ACSM data in the EBAS database.

2. On-line source apportionment

2.1. Measurement and data preparation

Source apportionment analyses do not fundamentally require any divergence from standard AMS/ACSM operating procedures. However, the source apportionment analysis will be greatly facilitated by the availability of supporting data such as NO_x, black carbon (preferably from a multi-wavelength optical measurement), and high-quality meteorological data. The instrument size cut (either PM_1 or $PM_{2.5}$) should be considered in terms of the local mass distribution and desired relationship to regulated quantities.

Two inputs are required for PMF-based source apportionment: a mass spectral time series, denoted as the data matrix **X** (having matrix elements $x_{i,j}$ where rows (i) denote the filter collection time stamp and columns (j) denote the mass to charge ratio m/z), and its corresponding uncertainty matrix, **S**, which corresponds point-by-point to **X**. The uncertainties in **S** correspond solely to measurement precision and depend on ion counting statistics, variability in detector response to a single ion, and electronic noise within the instrument[6]. For most AMS-type instruments, standard data analysis procedures calculate the organic contributions to CO^+, H_2O^+, OH^+, and O^+ as a constant fraction of CO_2^+ for high resolution instruments, or alternatively m/z 28, 18, 17, and 16 as a constant fraction of m/z 44 at low resolution. These dependent ions should be removed from **X** and **S** prior to PMF analysis to avoid overweighting of CO_2^+. To optimize factor deconvolution, it is recommended to downweight ions with signal-to-noise ratios below 2 and to remove ions with signal-to-noise ratios below 0.2[8].

2.2. Source apportionment model and solution selection

PMF describes the input data is described as a linear combination of static factor profiles (i.e. characteristic mass spectra) and their time-dependent intensities (i.e. mass concentration). These factors are typically related either to specific primary emissions sources or to secondary organic aerosol, which is classified according to its extent of oxygenation. PMF is implemented using the Multilinear Engine[9] (ME-2), with the Source Finder (SoFi) toolkit used for model configuration and post-analysis[10]. This allows for selected factor profiles to be constrained to chemical signatures characteristic of wellseparated sources, facilitating retrieval of environmentally optimized solutions in which mathematical mixing between factors is minimized. Constraining factors is particularly important for off-line AMS source apportionment because of the decreased time resolution relative to online measurements; for example, the time series of traffic and cooking emissions have diurnal patterns that are quite different (peak concentrations at rush hours and mealtimes, respectively), but are often correlated on a 24-hr basis. Because of this temporal correlation, constrained profiles can be necessary for clear factor resolution. Constraints are applied using the a-value approach, shown in Eq. 1:

$$(1 - a)f_{z,j} = f'_{z,j} = (1 + a)f_{z,j} \qquad (1)$$

Here $f'_{z,j}$ represents the modelled factor profile (z denotes a specific factor and j the individual ions comprising the factor mass spectrum), $f_{z,j}$ represents a reference "anchor" profile derived externally, and a determines the tightness of constraint ($0 \leq a \leq 1$). In practice, the values of $f'_{z,j}$ may slightly exceed the limits indicated by Eq. 1 due to post-solution renormalization. Anchor profiles may be obtained using a variety of methods, including previous source apportionment studies, source apportionment of a subset of the analyzed dataset (selected such that the compositional variability of OA is dominated by the source in question), or direct source/plume measurements.

Traditional PMF analyses focused on obtaining a single optimized solution, followed by the analysis of uncertainties relative to this solution. This is no longer recommended. Instead, a two-part analysis is suggested where traditional exploratory methods (including possible constraints via a-value) are used to generate an environmentally plausible solution. This plausible solution is then used as the base case for comprehensive uncertainty and sensitivity analyses, as discussed in section 2.3, which will generate a large set of potential solutions, which must then be individually accepted or rejected. The final source apportionment output then consists of the average of all accepted solutions, with their variation used as a measure of uncertainty.

Selection of the base case requires determination of the appropriate number of factors and their identity, together with a rough assessment of the extent to which profile constraints via the a-value technique are needed to ensure clean separation. These considerations are determined *in situ* by the scientist and depend strongly on site location and characteristics. Therefore it is not possible to provide a universal set of detailed guidelines describing this stage of the analysis. Methods for solution selection/factor identification of on-line AMS data are discussed in detail in the literature[1, 8, 10, 11]. A set of general principles for consideration is presented here.

1. The PMF model is solved by minimizing the objective function Q, defined as the sum of the squares of the uncertainty-weighted residuals. Increasing the number of factors will nearly always decrease Q due to the greater number of degrees of freedom provided. However, a large decrease in Q in response to increasing the number of factors suggests a large improvement in the mathematical description of the dataset, while a small change suggests that the additional factor yields only incremental improvements and/or primarily allows the model to better describe noise.

2. Similar to (1), investigation of the *m/z*- or time-dependent contributions to Q, and the change in these contributions with an increasing number of factors indicates more precisely the aspects of the dataset where the model description has improved. The user can inspect these changes to see whether there is a discrete change in the model's ability to describe certain features, or rather an incremental lessening of the same unexplained events (which is less likely to be physically interpretable). Taken together, (1) and (2) provide a good first estimate of the appropriate number of factors for the base case.
3. Factors that are highly correlated in time or profile should be treated with extra caution due to the possibility of mathematical splitting. In the case two chemically distinct but temporally similar factors are known to be present, use of the *a*-value method to constrain factor profiles is strongly recommended. Likewise, factors comprising a low (< 5%) fraction of the total signal should be treated cautiously, and *a*-value constraints may be needed to ensure separation.
4. Wherever possible, factor time series should be correlated with tracer species expected to result from the same source. For example, strong correlations are often observed between the AMS-derived hydrocarbon-like organic aerosol (HOA), which is closely related to traffic, and NO_x or traffic-derived black carbon estimated from aethalometer measurements[12], while AMS-derived biomass burning often correlates with aethalometer-estimated black carbon from wood burning. Less oxygenated SOA may correlate with nitrate or anticorrelate with temperature if its time series is largely partitioning-driven, while more oxygenated may correlate with long-lived secondary species such as sulfate (in the summer) or the sum of sulfate and nitrate (in the winter). When tracer/factor relationships are unsatisfactory, *a*-value constraints on factor profiles may be needed to ensure clean separation. Many other possible factor-tracer relationships exist; conversely, situations can be proposed where one or more of the relationships detailed above will fail (e.g. the presence of coal burning complicates a source-based interpretation of black carbon). Detailed qualitative knowledge of the sources and processes affecting a site and their effects on atmospheric composition are therefore integral to successful PMF analysis.
5. Many sources exhibit a characteristic repeating temporal pattern based on human activities and/or meteorological processes. For example, HOA often peaks during rush hours, cooking-related aerosol during mealtimes, and (in Europe) wood burning during overnight hours due to domestic heating. Secondary aerosol may respond to changes in temperature and/or photochemistry. Local sources may exhibit strong concentration changes due to boundary layer cycling, while regional or long-range transport is less affected. Vehicle operation patterns may lead to concentration or temporal differences between weekdays and weekends. Inconsistent diurnal patterns may result from incompletely resolved factors, which may be improved with *a*-value constraints.
6. It is generally recommended to apply *a*-value constraints only to primary factors, given that SOA undergoes continuous chemical evolution in the atmosphere. Ideally, the anchor profile chosen should have direct relevance to the measurement site (e.g., a wood burning profile from a city in central Europe is likely applicable to another central European site, but less likely to perform well in an Asian city).

7. When applying the a-value constraints, it is recommended to conduct a sensitivity analysis where all a-values are varied and the effects on the above considerations are assessed. The extent to which factor diurnal patterns, time series, correlations with external tracers, and characteristic profiles provide a consistent description of each factor is the key criterion in determining a factor's reliability, and thus the appropriateness of the base case. The precise form of this optimization analysis can vary based on the particular measurement site, but can include maximization of factor/tracer correlations, retrieval of a clean diurnal pattern, or exclusion of spectral features inconsistent with factor identity.

2.3. Uncertainty analysis

Having determined an environmentally plausible base case as described above, this base case is leveraged for a comprehensive exploration of the environmentally reasonable solution space. In practice, this combines a sensitivity test in which the a-values of all constrained profiles are systematically varied with a simultaneous bootstrap analysis in which the time axis is randomly resampled. This generates a large number (typically a minimum of several hundred) of potential solutions. These "test" solutions are then individually accepted or rejected according to the framework of Vlachou et al.[13], which is summarized below:

1. For each test solution, a correlation matrix was calculated between the base case factor time series represented in rows and the test factor time series in columns.
2. Factors are sorted according to the highest correlation of the test time series with the base case.
3. Solutions are discarded if any of the correlation coefficients in the matrix diagonal are not higher by a statistically significant margin than all other correlation coefficients in the respective row or column.
4. The criteria in step (3) must likewise be satisfied for correlations with the selected external tracers.

Under this method, acceptance of a solution implies that every factor in the test solution must uniquely correspond to a factor in the base case, and vice versa. Rejected solutions can optionally be further classified for inspection according to the reasons for failure (e.g. two specific factors are not significantly different) to facilitate identification of other local minima in the solution space and/or likely mixing characteristics[14]. Note that the algorithm above can in theory be applied to profiles, though in practice the performance for time series has been better.

The final reported solution is the average of all accepted test solutions. The full set of test solutions can be used for a statistical description of the overall uncertainties in the solution (e.g. standard deviation, percentiles).

2.4. Long-term on-line data

Special considerations exist for long-term and/or continuous source apportionment analyses. First, factor profiles can vary over time depending on sources. A clear example is shown by SOA in central Europe, where wood burning influences are elevated in winter and biogenic influences in summer[15]. However, this seasonally driven variability is convoluted with diurnal variability due to daily temperature and/or photochemical cycling. Simultaneous deconvolution of these two processes in a single PMF analysis has not previously been successful. Further, the size of long-term or continuously growing datasets can make the repetition-heavy analyses described above impractical. Both of these concerns are satisfied by an alternate approach, in which a small (approximately 2 week) window is translated across the dataset. Within each step of this window, PMF analyses are conducted and a set of acceptable solutions are automatically extracted. This allows for a dynamic and time-dependent approach to long-term analysis. This algorithm is implemented within SoFi Pro (Canonaco et al., in preparation) and has been tested successfully on datasets in Zurich, Magadino (Switzerland), and Paris.

The rolling window algorithm essentially consists of two parts. First, conventional PMF analyses are conducted on individual seasons. These analyses follow the methods described above for obtaining a base case. The primary goals of these pre-analyses are to (1) determine the number and identity of the key factors for each season; (2) identify simple criteria (e.g. factor/tracer correlations, signatures in the diurnal patterns) that can be used for rapid classification of solutions for acceptance/rejection; and (3) establish the range of *a*-values for constrained factors that are likely to yield acceptable solutions. Once these pre-analyses are completed, the rolling analysis can be initialized.

The rolling window is of a sufficiently large size to enable effective PMF analysis, but small enough to allow rapid computation. In tests to date, 2 weeks has performed well. The window moved by an increment much smaller than the total window size (1 day in tests). Within the window, a combined bootstrap/*a*-value sensitivity test is performed as described in the previous section, although acceptance/rejection criteria are determined based on the seasonal pre-analysis. Crucially, acceptance/rejection is here not implemented as an optimization problem, but rather as a series of binary evaluations. Failure of a single criterion causes the entire solution to be rejected. Thresholds for acceptance/rejection are set by the user and sensitivity tests can be performed during post-analysis. All accepted solutions are retained for further analysis. Note that these accepted solutions do not consist only of those obtained from a single window position, as the small window increment means that in the example above each day will be analyzed 14 times.

Similar to the conventional on-line analysis, the final output of this long-term rolling source apportionment is not a single solution but an average time series of each factor concentration, combined with the ability to calculate relative statistics to describe its uncertainty. However, a key difference is that in this analysis the factor profiles are not static, but rather adapt to long-term changes in source characteristics.

3. Offline-AMS source apportionment

3.1. Measurement and data preparation

The offline-AMS source apportionment technique was originally described by Daellenbach et al.[16] and has since been refined and implemented for filters collected from around the world. Samples are collected on quartz fiber filters using a high-volume sampler (~500 L min^{-1}), typically with either 12 or 24 h resolution. From each filter sample, 4 punches of 16 mm diameter are extracted in 15 mL of ultrapure water in an ultra-sonic bath for 20 min at 30° C. In case of low filter loadings (e.g. in pristine environments), additional punches and or filter samples from multiple days can be combined to provide sufficient OA mass. The extracts are aerosolized in Ar using an Apex Q nebulizer (Elemental Scientific Inc., Omaha, NE, USA) operating at 60 °C and passed through a Nafion dryer to remove excess water. In testing, this nebulizer was preferred to other models due to the optimal overlap of the generated particle size distribution with the AMS transmission window.

For efficient measurement of large numbers of filter samples, the nebulizer is connected to an auto-sampler, in which the extracts are mounted. The auto-sampler programmatically alternates measurement of a filter extract with ultra-pure water (denoted "blank"), providing a continuous measurement of any contamination/background in the nebulizer and/or particle transfer lines. It is recommended to reserve one slot in the auto-sampler for the extract from an un-loaded filter ("field blank") and one for a solution of NH_4NO_3, to allow characterization of the CO_2^+ signal generated by the interaction of NH_4NO_3 with carbon deposited on the surface of the AMS vaporizer[17], which can otherwise be spuriously considered as OA mass. AMS mass spectra are typically collected every 30 s, with 12 min spent on the extract and 12 min on the blank. After removing transitional data points, this yields approximately 20 mass spectra per sample; these replicates are an essential component of calculations required to generate the source apportionment input data and also useful in uncertainty analysis of the source apportionment outputs.

For quantitative results, the measured organic mass spectra are multiplied by the mass of water-soluble organic carbon (WSOC), measured externally, times the OM:OC ratio determined from high-resolution analysis of individual AMS mass spectra[18]. Otherwise, calculation of the **X** matrix for PMF input is identical to the on-line analysis. The calculation of the uncertainties in **S** differs slightly in that these uncertainties incorporate not only the AMS measurement uncertainties ($\delta_{i,j}$), but also the standard deviation of the blank obtained directly before and after a measured extract ($\sigma_{i,j}$). The elements of **S** ($s_{i,j}$) are then calculated according to Eq. 2:

$$s_{i,j} = \sqrt{\delta_{i,j}^2 + \sigma_{i,j}^2} \qquad (2)$$

3.2. Source apportionment model and solution selection

The conceptual framework described above for the on-line solution largely applies as well to off-line analysis, with a few exceptions which are discussed here. The main differences arise from the lower time resolution and the need for water extraction prior to measurements.

Low time resolution typically prevents the use of diurnal patterns as a criterion for solution evaluation. However, it also has the more fundamental effect that variability in SOA composition is driven by season rather than daily temperature or photochemical cycles. This can be an advantage, as it is typically more directly connected to emissions sources[15]. It also affects selection of useful factor/tracer correlations; in particular, an exponential relationship between summer SOA and temperature has been observed at many sites, consistent with SOA generation from biogenic emissions. However, the low time resolution also complicates separation of local sources such as traffic and cooking, which if integrated over 24 h are strongly correlated (though are emitted at different times of the day). Separation of these factors, or others with similar characteristics, nearly always requires profile constraints.

As noted above, the OA measured by the AMS in filter samples is scaled to ambient OC using the measured WSOC. However, in principle each factor may have a different extraction efficiency, ranging from approximately 1 for SOA factors to 0.1 for HOA[16]. Two options are then possible. Either the extraction efficiencies can be ignored, in which case the source apportionment applies to WSOA, or factor-specific extraction efficiencies can be applied. Probability distributions of efficiencies for commonly observed factors have been determined[16], while for unusual factors it may be possible to determine an extraction efficiency via mass balance if the total OC is measured:

$$OC = \sum_n \left(\frac{WSOC_n}{R_n}\right)_{known} + \sum_m \left(\frac{WSOC_m}{R_m}\right)_{unknown} \tag{3}$$

Here R_n and R_m denote the extraction efficiencies for factors for which these efficiencies are *a priori* known and unknown, respectively. This is done via a pseudo-Monte Carlo approach in which a large number of combinations (~10^6) of known R_n are selected randomly according their predetermined probability distributions and combined with randomly selected PMF solutions from the sensitivity/bootstrap “test” analysis described in section 2.3[19]. This analysis is most likely to succeed (i.e., generate a probability distribution for R_m with manageably low uncertainties) when the number of unknown R is low and their mass fraction is relatively high.

4. Reporting data to the EBAS database

It is often desirable for AMS-type data and source apportionment outputs to be permanently archived in a public, searchable database. The EBAS Export Tool for ACSM/AMS (EETA) is a software utility within IGOR Pro (Wavemetrics, Inc.) developed to facilitate the archiving of AMS-type data in the EBAS database. EETA operates on fully processed data analyzed according to the best practices of the respective instruments. It handles a variety of tasks including documentation of instrument properties, site characteristics, and analysis methods, data conversion, metadata generation, file formatting, and upload. At present, the EBAS database accepts only the standard data products from AMS-type instruments (nitrate, sulfate, organics, ammonium, and chloride). EETA supports this for all AMS-type instruments, with the quadrupole-ACSM (Q-ACSM) being somewhat more automated. EETA also supports the archiving of PMF inputs and can be readily extended to PMF outputs and standard operating procedures (SOPs) once these analyses become available (the above sections represent a step towards such SOPs).

Section 4 provides guidelines and basic instructions for use of the EETA tool. In doing so, it draws heavily on the more comprehensive and continuously updated documentation packaged with the tool, which should serve as the primary point of reference for proper use. In the discussion below, the syntax <txt> is used to refer to user selection of a specific field visible in EETA (e.g. <Button#5>).

4.1. Output file contents

Output data consists of two corresponding files at EBAS data levels 0 and 2. Level 0 contains data for species of interest in native time resolution together with important metadata and is intended for specialist users, while Level 2 contains only the species of interest in hourly time resolution and is intended for general users. These files contain:

1. An extensive header providing information on instrument operation and users, sampling site, and analysis methods (Level 0 and 2);
2. Data time series for organics, sulfate, nitrate, ammonium, chloride, and the organic contributions to *m/z* 43, 44, and 60 (Levels 0 and 2), and the nitrate contributions to *m/z* 30 and 46 (Level 0 only);
3. Time series of uncertainties for the above species (Level 0 and 2);
4. Time series of species-dependent flags denoting measurement quality and special conditions (e.g. normal, compromised, missing, or unusual data) according to EBAS specifications (Level 0 and 2);
5. Time series of important metadata (airbeam, collection efficiency, lens pressure, ambient, pressure, temperature and relative humidity, species-dependent relative ionization efficiencies) (Level 0 only);
6. Flags for metadata as in (4) (Level 0 only);

Species concentrations are reported in $\mu g/m^3$, $\mu g\ S/m^3$, or $\mu g\ N/m^3$ (depending on the species) and are corrected to STP.

4.2. Instructions for Use

After standard analysis of the collected data is complete, archiving within EBAS using EETA proceeds via 4 steps: (1) background information on the site, instrument, and personnel are assembled; (2) data range and export method are defined; (3) formatted data is previewed and flagged; (4) files are generated for final review and upload. A more detailed description of each step follows.

Step 0: Data analysis

Data must be fully processed in IGOR before using EETA. Depending on the instrument, this may include determination of collection efficiency, modifications to fragmentation patterns, airbeam and/or flowrate corrections, RIE values, ion transmission corrections, etc. The EETA procedure file is then added to the active IGOR experiment. The user then follows the following protocol for data archiving.

Step 1: Background information

This section allows the recording of various information regarding instrument identification, sampling site characteristics, and operation/analysis personnel. In EETA, this information is divided into Instrument, Site, and Personnel categories, each treated within a different panel. Once set, this information is unlikely to significantly change across sequential EBAS submissions from the same site and can therefore be written to an external file and subsequently re-loaded for new analyses. The entries within these panels are treated in detail in the EETA documentation. A few key points follow:

1. All sites must be registered with GAWSIS in order to receive their identifying set of station IDs. Contact instructions are provided. Once these IDs are received, the station status in the Site panel may be changed to "Registered" and the IDs entered. This is a required step prior to archiving with EBAS.
2. In the Personnel panel, an <EBAS lab code> denotes the primary hosting organization for the instrument and can be obtained from ebas@nilu.no.
3. In the Instrument panel, the <Instrument type> field indicates the specific instrument model in use, which in turn controls the operation of the EETA software with respect to the extent of automation.
4. The Instrument panel includes a field denoted <Additional ID>, which is a unique identifier representing the specific parameter(s) being measured at the station. This field should remain unchanged if another ACSM is switched in for identical measurements, but would change if the measurement concept changed (e.g. dry vs. wet aerosol). Closely related is the <Analysis method> field, which denotes the laboratory-specific standard operating procedure used to generate the final dataset.

Step 2: Define data for export

In this step, the user defines the type of data requested (i.e., standard AMS outputs or PMF inputs) and generates the resulting time series. This procedure differs slightly by instrument (as defined in the previous step), with the <Automatic> method applicable to Q-ACSM only, and <Manual> applicable to all other systems. Here <Automatic> means that EETA will attempt to generate the requested data products, while <Manual> indicates that the user will explicitly provide them by providing full paths to waves and/or numeric constant.

<Manual> operation: Selecting this option opens a sub-panel in which three categories of data must be declared: time axis, data/error waves, and metadata waves. The following types of entries are possible: (1) wave declaration including the full path to the desired data; (2) "Auto", which indicates that EETA will attempt an automatic calculation of this data product; or (3) a numerical constant (metadata only), e.g. 0.5, in which case this value will be used for all time points. In <Manual> operation, all data are assumed to be final, requiring only conversion to STP, formatting, and time stamp conversions. That is, no calculations are performed and changing for example the value of the collection efficiency will not affect the corresponding mass concentrations.

<Automatic> operation: This option utilizes current panel configuration of the Q-ACSM data analysis software to generate the required data and metadata products. The user should ensure that this panel is configured for proper analysis, and will be requested to the input gain, dwell time per *m/z*, and *m/z* to use for electronic noise determination, but no other setup is required.

Step 3: Preview data and set flags

At this point, a time period for archiving can be defined, and the data generated. The Preview panel then appears, showing time series of data at Level 0 and 2 (with or without uncertainties) and Level 0 metadata, as well as automatically generated flags. Automatic flags identify only present or missing data; others must be set by the user. Other flags can be selected in the Preview panel by a combination of the <Active flag> and <Apply to> popup menus together with the graph marquee.

Step 4: Finalize and upload

The user may optionally add acknowledgment instructions and/or additional comments relevant to measurement, analysis, or other issues. The <Write EBAS file> button is then used to generate the text files for upload. It is recommended to view these files in e.g. MS Excel (imported as a delimited text file with no selected delimitors).

The Level 2 version of the most recently created files is stored in the <Last file saved> field. The file pair can be automatically uploaded to EBAS/PSI using the <Upload to EBAS/PSI database> button. (A manual upload of both Level 0 and Level 2 files is also possible through the EBAS ftp site.)

5. References

1. Zhang, Q.; Jimenez, J. L.; Canagaratna, M. R.; Ulbrich, I. M.; Ng, N. L.; Worsnop, D. R.; Sun, Y., Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review. *Anal. Bioanal. Chem.* **2011,** *401*, (10), 3045-67.
2. Canagaratna, M. R.; Jayne, J. T.; Jimenez, J. L.; Allan, J. D.; Alfarra, M. R.; Zhang, Q.; Onasch, T. B.; Drewnick, F.; Coe, H.; Middlebrook, A.; Delia, A.; Williams, L. R.; Trimborn, A. M.; Northway, M. J.; DeCarlo, P. F.; Kolb, C. E.; Davidovits, P.; Worsnop, D. R., Chemical and microphysical characterization of ambient aerosols with the Aerodyne aerosol mass spectrometer. *Mass Spec. Rev.* **2007,** *26*, (2), 185-222.
3. Aiken, A. C.; DeCarlo, P. F.; Kroll, J. H.; Worsnop, D. R.; Huffmann, J. A.; Docherty, K.; Ulbrich, I. M.; Mohr, C.; Kimmel, J. R.; Sueper, D.; Zhang, Q.; Sun, Y.; Trimborn, A.; Northway, M.; Ziemann, P. J.; Canagaratna, M. R.; Onasch, T. B.; Alfarra, R.; Prévôt, A. S. H.; Dommen, J.; Duplissy, J.; Metzger, A.; Baltensperger, U.; Jimenez, J. L., O/C and OM/OC ratios of primary, secondary, and ambient organic aerosols with high resolution time-of-flight aerosol mass spectrometry. *Environ. Sci. Technol.* **2008,** *42*, 4478-4485.
4. Ng, N. L.; Herndon, S. C.; Trimborn, A.; Canagaratna, M. R.; Croteau, P. L.; Onasch, T. B.; Sueper, D.; Worsnop, D. R.; Zhang, Q.; Sun, Y. L.; Jayne, J. T., An aerosol chemical speciation monitor (ACSM) for routine monitoring of the composition and mass concentrations of ambient aerosol. *Aerosol Sci. Technol.* **2011,** *45*, (7), 780-794.
5. Fröhlich, R.; Cubison, M. J.; Slowik, J. G.; Bukowiecki, N.; Prévôt, A. S. H.; Baltensperger, U.; Schneider, J.; Kimmel, J. R.; Gonin, M.; Rohner, U.; Worsnop, D. R.; Jayne, J. T., The ToF-ACSM: a portable aerosol chemical speciation monitor with TOFMS detection. *Atmos. Meas. Tech.* **2013,** *6*, (11), 3225-3241.
6. Allan, J. D.; Jimenez, J. L.; Coe, H.; Bower, K. N.; Williams, P. I.; Worsnop, D. R., Quantitative sampling using an Aerodyne aerosol mass spectrometer. Part 1: Techniques of data interpretation and error analysis. *J. Geophys. Res. - Atmos.* **2003,** *108*, (D3), 4090.
7. Jimenez, J. L.; Jayne, J. T.; Shi, Q.; Kolb, C. E.; Worsnop, D. R.; Yourshaw, I.; Seinfeld, J. H.; Flagan, R. C.; Zhang, X.; Smith, K. A.; Morris, J.; Davidovits, P., Ambient aerosol sampling using the Aerodyne aerosol mass spectrometer. *J. Geophys. Res. - Atmos.* **2003,** *108*, (D7), 8425.
8. Ulbrich, I. M.; Canagaratna, M. R.; Zhang, Q.; Worsnop, D. R.; Jimenez, J. L., Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data. *Atmos. Chem. Phys.* **2009,** *9*, 2891-2918.

9. Paatero, P., The multilinear engine - a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J. Comput. Graph. Stat.* **1999,** *8*, 854-888.
10. Canonaco, F.; Crippa, M.; Slowik, J. G.; Baltensperger, U.; Prévôt, A. S. H., SoFi, an IGOR-based interface for the efficient use of the generalized multilinear engine (ME-2) for the source apportionment: ME-2 application to aerosol mass spectrometer data. *Atmos. Meas. Tech.* **2013,** *6*, (12), 3649-3661.
11. Crippa, M.; Canonaco, F.; Lanz, V. A.; Äijälä, M.; Allan, J. D.; Carbone, S.; Capes, G.; Ceburnis, D.; Dall'Osto, M.; Day, D. A.; DeCarlo, P. F.; Ehn, M.; Eriksson, A.; Freney, E.; Hildebrandt Ruiz, L.; Hillamo, R.; Jimenez, J. L.; Junninen, H.; Kiendler-Scharr, A.; Kortelainen, A.-M.; Kulmala, M.; Laaksonen, A.; Mensah, A. A.; Mohr, C.; Nemitz, E.; O'Dowd, C.; Ovadnevaite, J.; Pandis, S. N.; Petäjä, T.; Poulain, L.; Saarikoski, S.; Sellegri, K.; Swietlicki, E.; Tiitta, P.; Worsnop, D. R.; Baltensperger, U.; Prévôt, A. S. H., Organic aerosol components derived from 25 AMS data sets across Europe using a consistent ME-2 based source apportionment approach. *Atmos. Chem. Phys.* **2014,** *14*, 6159-6176.
12. Sandradewi, J.; Prevot, A. S. H.; Szidat, S.; Perron, N.; Alfarra, M. R.; Lanz, V. A.; Weingartner, E.; Baltensperger, U., Using aerosol light absorption measurements for the quantitative determination of wood burning and traffic emission contributions to particulate matter. *Environ. Sci. Technol.* **2008,** *42*, (9), 3316-3323.
13. Vlachou, A.; Tobler, A.; Lamkaddam, H.; Canonaco, F.; Daellenbach, K. R.; Jaffrezo, J.-L.; Minguillón, M. C.; Maasikmets, M.; Teinemaa, E.; Baltensperger, U.; El Haddad, I.; Prevot, A. S. H., Development of a versatile source apportionment analysis based on positive matrix factorization: a case study of the seasonal variation of organic aerosol sources in Estonia. *Atmos. Chem. Phys. Discuss.* **2018**.
14. Stefenelli, G.; Lopez-Hilfiker, F. D.; Pospisilova, V.; Vogel, A.; Hüglin, C.; Baltensperger, U.; Prévôt, A. S. H.; Slowik, J. G., Source apportionment of ambient organic aerosol by online extractive electrospray ionization time-of-flight mass spectrometry (EESI-TOF). *Atmos. Chem. Phys. Discuss.* **in press**.
15. Daellenbach, K. R.; Stefenelli, G.; Bozzetti, C.; Vlachou, A.; Fermo, P.; Gonzalez, R.; Piazzalunga, A.; Colombi, C.; Canonaco, F.; Hueglin, C.; Kaspar-Giebl, A.; Jaffrezo, J.-L.; Bianchi, F.; Slowik, J. G.; Baltensperger, U.; El Haddad, I.; Prévôt, A. S. H., Long-term chemical analysis and organic aerosol source apportionment at nine sites in Central Europe: source identification and uncertainty assessment. *Atmos. Chem. Phys.* **2017,** *17*, 13265-13282.
16. Daellenbach, K. R.; Bozzetti, C.; Krepelova, A.; Canonaco, F.; Wolf, R.; Zotter, P.; Fermo, P.; Crippa, M.; Slowik, J. G.; Sosedova, Y.; Zhang, Y.; Huang, R.-J.; Poulain, L.; Szidat, S.; Baltensperger, U.; El Haddad, I.; Prévôt, A. S. H., Characterization and source apportionment of organic aerosol using offline aerosol mass spectrometry. *Atmos. Meas. Tech.* **2016,** *9*, 23-39.
17. Pieber, S. M.; El Haddad, I.; Slowik, J. G.; Canagaratna, M. R.; Jayne, J. T.; Platt, S. M.; Bozzetti, C.; Daellenbach, K. R.; Frohlich, R.; Vlachou, A.; Klein, F.; Dommen, J.; Milijevic, B.; Jimenez, J. L.; Worsnop, D. R.; Baltensperger, U.; Prevot, A. S. H., Inorganic salt interfence on CO_2^+ in Aerodyne AMS and ACSM organic aerosol composition studies. *Environ. Sci. Technol.* **2016,** *50*, (19), 10494-10503.
18. Canagaratna, M. R.; Jimenez, J. L.; Kroll, J. H.; Chen, Q.; Kessler, S.; Massoli, P.; Hildebrandt Ruiz, L.; Fortner, E.; Williams, L.; Wilson, K.; Surratt, J.; Donahue, N.; Jayne, J. T.; Worsnop, D. R., Elemental ratio measurements of organic compounds using aerosol mass spectrometry: characterization, improved calibration, and implications. *Atmos. Chem. Phys.* **2015,** *15*, 253-272.
19. Bozzetti, C.; Daellenbach, K. R.; Hueglin, C.; Fermo, P.; Sciare, J.; Kaspar-Giebl, A.; Mazar, Y.; Abbaszade, G.; El Kazzi, M.; Gonzalez, R.; Shuster-Meiseles, T.; Flasch, M.; Wolf, R.; Krepelova, A.; Canonaco, F.; Schnelle-Kreis, J.; Slowik, J. G.; Zimmermann, R.; Rudich, Y.; Baltensperger, U.; El Haddad, I.; Prévôt, A. S. H., Size-resolved identification, characterization, and quantification of primary biological organic aerosol at a European rural site. *Environ. Sci. Technol.* **2016,** *50*, (7), 3425-3434.